

Sven Bittner, Stefan Hornbostel, Frank Scholze (Hg.)

# **FORSCHUNGSIONFORMATION IN DEUTSCHLAND: ANFORDERUNGEN, STAND UND NUTZEN EXISTIERENDER FORSCHUNGSIONFORMATIONSSYSTEME**

## **WORKSHOP FORSCHUNGSIONFORMATIONSSYSTEME 2011**

iFQ-Working Paper No.10 | Mai 2012



**iFQ – Institut für Forschungsinformation und Qualitätssicherung**

Schützenstraße 6a

10117 Berlin

Telefon 030-206 41 77-0

Fax 030-206 41 77-99

E-Mail [info@forschungsinfo.de](mailto:info@forschungsinfo.de)

Internet [www.forschungsinfo.de](http://www.forschungsinfo.de)  
[www.research-information.de](http://www.research-information.de)

ISSN 1864-2799

Mai 2012

Sven Bittner, Stefan Hornbostel, Frank Scholze (Hg.)

**Forschungsinformation in Deutschland:  
Anforderungen, Stand und Nutzen  
existierender Forschungsinformationssysteme**

Workshop Forschungsinformationssysteme 2011

iFQ-Working Paper No. 10 | Mai 2012



## Inhalt

<b>Einführung</b>	7
-------------------	---

### **Sicht ausgewählter Akteure des Wissenschaftsbetriebs**

*Dietrich Nelle*

Wissenschaftsinformationssysteme aus der Sicht des BMBF	11
---	----

*Jürgen Güdler*

Forschungsinformationssysteme – Services und Anforderungen der DFG	19
--	----

*Frank Schüller*

Die amtliche Forschungsstatistik	29
----------------------------------	----

### **Sicht der Universitäten & Einrichtungen**

*Ulrike Küsters, Michael Erben-Russ*

Forschungsinformationssysteme bei Fraunhofer	37
--	----

*Sebastian Hernig, Jörg Becker*

Einführung eines Forschungsinformationssystems an der Westfälischen Wilhelms-Universität Münster – Von der Konzeption bis zur Implementierung	41
--	----

*Regine Tobias, Volker Karl*

Einführung eines integrierten Forschungsinformationssystems am Karlsruher Institut für Technologie	55
---	----

*Barbara Ebert, Alexander Kujath, Joachim Holtorf, Karsten Holmberg, Thomas Rupp*

Erfahrungen aus der Einführung des Forschungsinformationssystems Pure an der Leuphana Universität Lüneburg	65
---	----

*Annette Lewerentz*

Forschungsdatenbank der Freien Universität Berlin	79
---	----

*Astrid Lubinski*

Forschung präsentieren, fördern, verwalten – die Forschungsdatenbank der Universität Rostock	91
---	----

## **Ein Blick in den internationalen Raum**

*Brigitte Jörg*

Übersicht Systeme in Europa .....	103
-----------------------------------	-----

*Josh Brown*

Current Research Information Systems in the UK: an evolving network .....	115
---	-----

*David Baker*

CASRAI and Research Impacts .....	127
-----------------------------------	-----

*Jonathan Corson-Rikert, Dean B. Krafft, Brian J. Lowe*

VIVO: A Semantic Network of Researchers and Research Information as Linked Open Data .....	139
---	-----

<b>Verzeichnis der Autorinnen und Autoren .....</b>	<b>155</b>
---	------------

## Einführung

Das Thema „Forschungsinformationen und Forschungsinformationssysteme“ ist keineswegs neu. Es begleitet die Wissenschaft mit konjunkturellen Schwankungen seit mehr als dreißig Jahren, ohne dass sich Standards, Interoperabilität und Mehrfachnutzbarkeit der Daten in der Praxis wirklich etabliert hätten. Es scheint, als sei die derzeitige Entwicklung durch zwei gegenläufige Trends gekennzeichnet: Einerseits nimmt der Bedarf an validen, vergleichbaren und mit begrenztem Aufwand erzeugbaren Informationen zum In- und Output von Lehr- und Forschungseinrichtungen angesichts wachsender Wettbewerbsintensität, performanzabhängigen Steuerungssystemen, zunehmend indiktorisierter Berichterstattung und erhöhten Anforderungen an „accountability“ deutlich zu. Andererseits scheinen die Schwierigkeiten, einen geeigneten technischen, organisatorischen und rechtlichen Rahmen zu finden, eher gestiegen zu sein.

Gleichwohl bestimmen forschungsbezogene Informationen unterschiedlichster Provenienz und Güte den Alltag der verschiedenen Akteure des Wissenschaftsbetriebs in immer stärkerem Ausmaß. Unter anderem Benchmarkings, Rankings, Ratings, formelbasierte Mittelverteilungen, Berichtssysteme, Patent-, Publikations- und Drittmittelindikatoren spielen auf unterschiedlichen Aggregations-ebenen eine immer stärkere Rolle in der strategischen Planung und der eigenen Positionsbestimmung. Die verschiedenen Akteure benötigen und nutzen Forschungsinformationen aus sehr unterschiedlichen Gründen, stellen daher auch verschiedenste Anforderungen an diese Informationen und formulieren dementsprechend teilweise orthogonale Erwartungen. Insofern ist das alte Thema inzwischen in einem stark veränderten Kontext angekommen, der nicht nur von positiven Erwartungen, sondern auch von Ängsten und Sorgen hinsichtlich der Datennutzung geprägt ist, und hat dabei deutlich an Relevanz gewonnen. Vor diesem Hintergrund arbeiten derzeit viele Einrichtungen an einer Überarbeitung bestehender Systeme oder an den Planungen für den Erst- oder Neuaufbau von Informationssystemen.

Um Erfahrungen mit derartigen Planungs- und Aufbauprozessen nutzbar zu machen, um aus Fehlern lernen zu können und um zukünftige Anforderungen zu antizipieren, hat das iFQ (Institut für Forschungsinformation und Qualitätssicherung) zusammen mit der DINI (Deutsche Initiative für Netzwerkinformation) im November 2011 einen Workshop „Forschungsinformationssysteme“ durchgeführt und dazu verschiedene Akteure des Wissenschaftsbetriebs im deutschsprachigen Raum eingeladen. Gastgeber der Veranstaltung war das Karlsruher Institut für Technologie (KIT). Ziele des Workshops waren die Bestandsaufnahme der Situation in Deutschland in Bezug auf Forschungsinformationen, die Darstellung der diesbezüglichen Interessen und Anforderungen der verschiedenen Akteure im Wissenschaftsbetrieb (Forschungsförderer, Universitätsleitungen, Bund und Länder etc.) sowohl die Entwicklung von Anregungen, wie das Thema gemeinsam weiterentwickelt werden kann. Weiterhin wurden die Strategien und Prozesse verschiedener Einrichtungen und einige der derzeit genutzten Forschungsinformationssysteme mit ihren Vor- und Nachteilen vorgestellt.

Der Workshop mit seinen 14 inhaltlichen Vorträgen und einer Podiumsdiskussion stieß mit über 160 Besuchern auf eine große Resonanz.

Für diesen Beitragsband haben sich einige der Referenten des Workshops bereit erklärt, ihre Erfahrungen und Ansichten in Form eines erweiterten Textbeitrages zur Verfügung zu stellen. Darüber hinaus konnten drei Autoren gewonnen werden, die für Großbritannien, die USA und Kanada internationale Perspektiven der Forschungsinformation beleuchten. Das vorliegende Working Paper ist daher nicht nur eine Tagungsdokumentation im engeren Sinne.





## **Sicht ausgewählter Akteure des Wissenschaftsbetriebs**



## **Wissenschaftsinformationssysteme aus der Sicht des BMBF**

### **1 Einleitung**

Systematische Informationen über Forschungseinrichtungen gewinnen zunehmend an Bedeutung. Das BMBF betrachtet dabei solche Forschungsinformationssysteme aus einer doppelten Perspektive, nämlich:

- zum einen als Forschungsförderer, der Forschungsinformationssysteme unterstützen will und
- zum anderen aber auch als Nutzer von Forschungsinformationen für die Praxis des Ministeriums, z.B. als Material für strategische Planungen oder Berichterstattung gegenüber dem Parlament.

Die Autonomie der Wissenschafts- und Forschungseinrichtungen in Deutschland ist insbesondere in Folge von Reformen in Länderhochschul- und Hochschulrahmengesetzgebung sowie der Wissenschaftsfreiheitsinitiative (seit 2009) deutlich gewachsen. Die Zunahme der Autonomie führt einerseits zu einer Zunahme der Eigenverantwortung der Wissenschafts- und Forschungseinrichtungen, welche von diesen effektiv und überzeugungskräftig wahrzunehmen ist. Andererseits steigt damit zugleich der Bedarf an Transparenz, Qualitätssicherung und Rechenschaftslegung gegenüber wissenschaftlichen Peers ebenso wie gegenüber der Öffentlichkeit als Mittelgeber. Dies erfordert zeitgemäße Ansätze, welche sowohl die Binnensteuerung der Einrichtungen als auch die externe Legitimation unterstützen. Während traditionelle Governance der Wissenschaft überwiegend Input-orientiert war, wandelt sich dies im Zuge dieses Prozesses zunehmend zu einer Output-Orientierung mit dem Fokus auf Qualität und Forschungsleistungen. Damit steigen die Informationsbedarfe und die Ansprüche an Evaluationen. Auch braucht die qualitative Bewertung von Output-Eigenschaften eine verbesserte Unterstützung durch quantitativ erfassbare Parameter: Quantitative Parameter bieten gute Ansatzpunkte, um Veränderungen sichtbar hervortreten zu lassen und nach den zugrundeliegenden Ursachen zu fragen. Quantitative Parameter sind zudem sehr geeignet zur Veranschaulichung qualitativer Aussagen. Die Bedeutung solcher Angaben für die politische Kommunikation zeigt anschaulich das Beispiel PISA. Während viele hervorragende Studien im Bildungsbereich früher eher selten von einer breiteren Öffentlichkeit wahrgenommen wurden, hat sich die Diskussionslage mit der Verfügbarkeit quantitativer Daten fundamental geändert. Ein bedauerlicher Nachteil für die Überzeugungskraft der Wissenschaftspolitik ist es deshalb, dass zu wichtigen Aspekten des Wissenschaftssystems, beispielsweise zu Studienverläufen und Promotionen, noch viel zu wenig quantitative, überregional kompatible Daten vorliegen.

Zu berücksichtigen ist aber stets auch, dass die Aussagekraft von Forschungsinformationen davon abhängt, dass sie eine angemessene Berücksichtigung der unterschiedlichen Aufgaben und Missionen der jeweiligen Institution, unterschiedliche Gegebenheiten in den verschiedenen Disziplinen usw. ermöglichen. Ein Beispiel für besondere Herausforderungen ist die Erzeugung sinnvoller bibliometrischer Aussagen im Bereich der Geistes- und Sozialwissenschaften. Es besteht also noch ein umfangreicher Bedarf an Forschung und Weiterentwicklung. Deshalb ist es erfreulich, dass auf vielen Ebenen an neuen Ansätzen der Berichterstattung und der Leistungsmessung gearbeitet wird und neue Instrumente erprobt werden und dass der Workshop Forschungsinformationssysteme hierzu beigetragen hat.

## 2 Förderung von Forschungsinformation

Aus Platzgründen will ich auf die Rolle des BMBF als Förderer von Forschungsinformation hier nur sehr knapp eingehen.

1. Mit der Förderung der Pilotstudien zum Rating des Wissenschaftsrates wird die Erprobung von Verfahren unterstützt, wie man eine methodisch und konzeptionell gut durchdachte Bewertung der Forschungsleistung von einzelnen Disziplinen durchführen kann.
2. In der Förderinitiative „Neue Governance der Wissenschaft“ werden Projekte gefördert, die anhand empirisch angelegter Untersuchungen das Verständnis der Transformationsprozesse im Wissenschaftssystem erweitern und Konsequenzen der Globalisierung, der zunehmenden Europäisierung von Innovationspolitiken und der Reorganisation von Hochschulen und Forschungseinrichtungen aufzeigen.
3. Das BMBF fördert am Forschungsverbund „Kompetenzzentrum Bibliometrie“ den Aufbau einer Datenbank für bibliometrische Analysen.
4. Darüber hinaus unterstützt das BMBF den „Rat für Sozial- und Wirtschaftsdaten“ RatSWD. Auf diesen werde ich in Verlauf dieses Beitrags noch einmal zurückkommen, weil er auch ein gutes Beispiel dafür ist, wie die Perspektiven von Datenerzeugern und Datennutzern überzeugend zusammen geführt werden können. Außerdem verbessern **Forschungsdatenzentren** die Zugänglichkeit (anonymisierter) Datensätze für die Forschung; **Datenservicezentren** leisten Unterstützung für empirisches Arbeiten.

Ausführlicher eingehen möchte ich dagegen auf drei unterschiedlich gelagerte Aspekte der Nutzung von Forschungsinformationen im Rahmen meines eigenen Aufgabenkreises, nämlich erstens auf die konkrete Umsetzung von Forschungsinformationen im Monitoring des Paktes für Forschung und Innovation, zweitens – wie bereits erwähnt - auf den Rat für Sozial- und Wirtschaftsdaten und drittens auf die aktuellen Empfehlungen des Wissenschaftsrates zur Steuerung und Bewertung von Forschungsleistungen.

## 3 Nutzung von Forschungsinformationen im Rahmen des Paktes für Forschung und Innovation

Bund und Länder sowie die Wissenschafts- und Forschungsorganisationen Deutsche Forschungsgemeinschaft (DFG), Fraunhofer-Gesellschaft (FhG), Helmholtz-Gemeinschaft (HGF), Max-Planck-Gesellschaft (MPG) und Leibniz-Gemeinschaft (WGL) haben 2005 einen „Pakt für Forschung und Innovation“ geschlossen und 2009 durch einen Pakt II verlängert. Der Pakt II sieht vor, dass Bund und Länder ihre Förderung um 5% jährlich steigern. Dies bedeutet im Zeitraum 2011 bis 2015 von Bund und Ländern 4,9 Milliarden Euro zusätzliche Mittel. Damit erhalten die beteiligten Wissenschaftsorganisationen finanzielle Planungssicherheit, im Gegenzug verpflichten sie sich auf folgende forschungspolitischen Ziele:

- das Wissenschaftssystem dynamisch entfalten und zukunftsweisende Forschungsthemen frühzeitig aufgreifen
- die Vernetzung im Wissenschaftssystem leistungssteigernd und dynamisch gestalten
- neue Strategien der europäischen und internationalen Zusammenarbeit entwickeln und umsetzen
- nachhaltige Partnerschaften zwischen Wirtschaft und Wissenschaft stärken
- die Besten dauerhaft für die deutsche Forschung gewinnen, Förderung von Nachwuchs und insbesondere der wissenschaftlichen Karriere von Frauen.

Auf der Grundlage dieser forschungspolitischen Zielsetzung haben die Wissenschafts- und Forschungsorganisationen in konkreten Selbstverpflichtungen dargestellt, welche Maßnahmen sie er-

greifen wollen, um die genannten forschungspolitischen Ziele zu erreichen und dazu beizutragen, die internationale Wettbewerbsfähigkeit des deutschen Wissenschaftssystems weiter auszubauen.

Im Pakt für Forschung und Innovation ist hierzu ein jährliches Monitoring vereinbart. Ziele der Berichterstattung sind es, den Blick auf wichtige Entwicklungen der letzten Periode zu lenken, qualitative Aussagen quantitativ zu untermauern und Erfolge sowie fortbestehende Handlungsnotwendigkeiten besser zu kommunizieren. Die Wissenschaftsorganisationen legen dazu der gemeinsamen Wissenschaftskonferenz von Bund und Ländern jährlich Rechenschaftsberichte zu den forschungspolitischen Zielen vor, die qualitative und quantitative Indikatoren enthalten. Auf dieser Grundlage erfolgt eine Gesamtbewertung durch eine Bund-Länder-Arbeitsgruppe. Dieser „Monitoring –Bericht“ wird von den Ministerinnen und Ministern von Bund und Ländern in der Gemeinsamen Wissenschaftskonferenz (GWK) verabschiedet, dem Haushaltsausschuss des Parlaments vorgelegt und veröffentlicht.

Neben den Textdarstellungen der Entwicklung der Forschungsorganisationen werden insgesamt 86 Indikatoren erhoben, einige davon für alle beteiligten Organisationen, andere sind spezifischen Aufgaben einzelner Organisationen angepasst. Die Daten stammen aus verschiedenen Quellen: Manche Daten werden gemeinsam für alle Organisationen im Rahmen von Studien erhoben (Bibliometrie); auch werden Daten berücksichtigt, die bereits in anderem Kontext erhoben wurden, z.B. vom Statistischen Bundesamt oder von der GWK, wohingegen andere Daten direkt bei den Forschungsorganisationen ermittelt werden müssen.

Zur Nutzung dieser Daten hier einige – stark verkürzte – Beispiele und Ihre Interpretation:

1. Deutschland konnte seine **Stellung in der Spitzengruppe** der weltweit leistungsfähigsten Wissenschaftssysteme behaupten. Die Anzahl der Publikationen aus Deutschland wird nur von den USA, China und Großbritannien übertroffen. Die Publikationen der Forschungsorganisationen liegen hinsichtlich ihres Impacts, berechnet anhand der Zitierhäufigkeit der Publikationen, je nach Forschungsfeld 20 – 60 % über dem weltweiten Durchschnitt. Der Erfolg bei der Einwerbung von ERC Grants stieg zwischen 2007/8 bis 2010 kontinuierlich an (bei den Starting Grants von 10,4% auf 15,7%, bei den „Advanced Grants“ 10,0% auf 17,0%).
2. Die **Vernetzung innerhalb des deutschen Wissenschaftssystems** schreitet voran und hat eine neue Qualität erreicht. So gibt es vielfältige Forschungsk Kooperationen von außeruniversitären Forschungseinrichtungen mit Hochschulen, die regional, institutionell oder inhaltlich fokussiert sind. Beispiele herausragender regionaler Vernetzungen sind neben dem Karlsruher Institut für Technologie (KIT) und der Jülich-Aachen Research Alliance (JARA) auch das Göttinger Research Council (GRC), das Dresden concept oder Pearls Potsdam. Diese zunehmende Vernetzung zeigt sich auch deutlich am Anstieg gemeinsamer Berufungen auf W2- und W3-Stellen um 28 % von 2005 bis 2010.
3. Das **europäische Engagement wurde weiter ausgebaut**. Deutsche Forschungseinrichtungen, Hochschulen und Unternehmen konnten bis November 2010 18,6 % der Mittel einwerben (= über 2,6 Mrd. Euro), die im Europäischen Forschungsrahmenprogramm für die Forschungsförderung zur Verfügung stehen, also mehr als Deutschland rechnerisch in dieses Programm einzahlt. Die außeruniversitären Forschungseinrichtungen tragen maßgeblich hierzu bei. Zudem obliegt ihnen in einem Drittel als Projekte an denen sie beteiligt sind die Koordination.
4. Mit ihrem **internationalen Engagement** stärken die Wissenschaftsorganisationen die heimischen Standorte und das deutsche Wissenschaftssystem insgesamt. Den außeruniversitären Wissenschaftseinrichtungen gelingt es zunehmend besser, herausragende Wissenschaftlerinnen und Wissenschaftler auf verschiedenen Karrierestufen für eine

Tätigkeit in Deutschland zu rekrutieren. So sind 30 % aller Wissenschaftlerinnen und Wissenschaftler der MPG ausländische Staatsbürger, die DFG fördert 60 Internationale Graduiertenkollegs, die WGL hat seit 2006 gemeinsam mit deutschen Hochschulen 18 Leibniz International Graduate Schools etabliert.

5. Die **Kooperation mit der Wirtschaft** ergibt ein inhomogenes Bild bei Drittmittel-einwerbungen aus der Wirtschaft, Schutzrechtsvereinbarungen und Ausgründungen, wobei allerdings die Kooperation der FhG mit der Wirtschaft als exzellent eingestuft wird. Insgesamt haben die außeruniversitären Einrichtungen die Zahl ihrer bestehenden Schutzrechtsvereinbarungen bzw. Patente in den letzten Jahren erheblich gesteigert. Allein im Jahr 2010 haben die außeruniversitären Forschungseinrichtungen 874 neue Schutzrechtsvereinbarungen abgeschlossen, eine Steigerung um knapp 29 % im Vergleich zum Vorjahr. Gleichwohl wird noch erhebliches weiteres Potenzial gesehen.
6. „**Die besten Köpfe**“: Die Attraktivität der Wissenschaft als Beruf ist hoch. Der Anstieg der Zahl der Beschäftigten von 2005 bis 2010 um mehr als 23 % ist ein Indiz hierfür. Angesichts des demographischen Wandels gilt es diese Attraktivität weiter auszubauen. Mit einer Steigerung der in den Einrichtungen der außeruniversitären Forschungsorganisationen betreuten Promotionen um 50 % seit 2006 leisten die außeruniversitären Forschungseinrichtungen hierzu einen erheblichen Beitrag. So wurden die Möglichkeiten des selbständigen wissenschaftlichen Arbeitens von Postdoktorandinnen und -doktoranden erfolgreich erweitert: die Zahl der selbständigen Nachwuchsgruppen in den außeruniversitären Forschungseinrichtungen wurde seit 2005 verdoppelt und die Anzahl der von der Deutschen Forschungsgemeinschaft bewilligten Einzelmaßnahmen in der direkten Nachwuchsförderung seit 2005 um 45 % gesteigert. Die Gewinnung von Frauen für die Wissenschaft bleibt weiterhin eine vordringliche Aufgabe. Zwar haben die außeruniversitären Forschungseinrichtungen in den Jahren 2005 bis 2010 den Frauenanteil am wissenschaftlichen Personal um knapp 5 % gesteigert, doch liegt der Anteil deutlich niedriger als an Hochschulen. Hier gelingt es noch nicht in ausreichendem Maße das große Potential zu nutzen.

Weitere quantitative Indikatoren, die im Bericht verwendet werden, sind z.B.

- Mittel, die für den organisationsinternen Wettbewerb zur Verfügung stehen
- Anzahl an Leibniz-Preis-Trägern, die an den Forschungsorganisationen angesiedelt sind, darunter die Anzahl Frauen
- Anzahl Innovationscluster (FhG)
- Anzahl Max-Planck Fellowships (MPG)

Ein wichtiges Merkmal der Monitoring-Berichterstattung ist, dass es sich um ein lernendes Instrument handelt. Auch wenn sich schon aus Gründen der Konstanz und Vergleichbarkeit möglichst wenig am Kanon der herangezogenen Indikatoren ändern soll, so gibt es doch hinreichend Spielraum für Justierungen und methodische Änderungen, z.B. bei der Bibliometrie.

Lässt man die jährlichen Berichte noch einmal Revue passieren, wird deutlich, dass die indirekte Steuerung über qualitative, auf Leistungssteigerung ausgerichtete Ziele verbunden mit Transparenz über die Fortschritte bei der Zielerreichung ein ausgesprochen wirksames Instrument ist. Die Organisationen verfolgen die vereinbarten Ziele konsequent und haben vielfach bemerkenswerte Erfolge vorzuweisen.

Der Bericht etabliert sich als übersichtliche Informationsquelle und die Ergebnisse des Monitoring-Berichts werden verstärkt aus dem Bundestag und der Landespolitik nachgefragt. In diesem Sinn ist der Pakt-Monitoring Bericht auch ein wichtiges Instrument zur Legitimation unserer Forschungsausgaben.

## **4 Der Rat für Sozial und Wirtschaftsdaten**

Auch in der Analyse der Wissenschaft müssen die Perspektiven von Erzeugern und Nutzern zusammengebracht werden, um eine effektive Datenversorgung und deren Nutzung zu erreichen. Ein instruktives Modell hierfür ist der – wie bereits erwähnt - vom BMBF geförderte Rat für Sozial und Wirtschaftsdaten (RatSWD). Der Rat wurde 2004 eingerichtet mit der Zielsetzung die Forschungsdateninfrastruktur für die empirische Forschung nachhaltig zu verbessern und somit zu ihrer internationalen Wettbewerbsfähigkeit beizutragen. Er ist ein unabhängiges Gremium und besteht aus sechzehn Mitgliedern, je acht besetzt durch Datenproduzenten und Servicezentren sowie je acht besetzt durch empirisch arbeitende Wissenschaftlerinnen und Wissenschaftlern aus Hochschulen und anderen Forschungseinrichtungen. Er hat sich, wie ihm der Wissenschaftsrat vor nicht allzu langer Zeit bescheinigt hat, als institutionalisierter Ort des Austauschs und der Vermittlung zwischen den Interessen von Wissenschaft und Datenproduzenten bewährt und übernimmt in den Sozial-, Verhaltens- und Wirtschaftswissenschaften in Bezug auf die Standardsetzung und Qualitätssicherung sowie die weitere Entwicklung bei den Forschungsdatenzentren und Datenservicezentren eine beratende, initiiende und qualitätssichernde Rolle.

Für die Datenerzeuger hat dies den Vorteil, dass sie den Datenbedarf der Wissenschaft in einer strukturierten, effizienten Weise erfüllen können und mittelbar von den Ergebnissen einer guten Forschung profitieren können. Für die Wissenschaft besteht der Nutzen in einem geringeren Erhebungsaufwand, einer besseren Datenqualität und einer gesicherten Verfügbarkeit. Dies halte ich für ein exzellentes Modell einer Kooperation von Datenerzeugern und -nutzern.

## **5 Die Empfehlungen des Wissenschaftsrates zur Bewertung und Steuerung von Forschungsleistung**

Auf die zum Tage des Workshops erst wenige Tage alten Empfehlungen des Wissenschaftsrates zur Bewertung und Steuerung von Forschungsleistung möchte ich etwas ausführlicher eingehen, da diese sicher viel Stoff für weitere Diskussionen bieten. Dort finden sich wertvolle Hinweise zu einer umsichtigen Nutzung von Indikatoren und damit auch von Forschungsinformationssystemen. Im Kontext der Leistungsdifferenzierung und des Wettbewerbs greift das Papier die Debatte darüber auf, wie eine adäquate Leistungsmessung und –bewertung aussehen soll, damit sie sinnvoll und nutzbringend in der Steuerung von Forschung umgesetzt werden kann. Zu einem angemessenen Verhältnis von Aufwand und Nutzen der zunehmenden Evaluationen gibt es naturgemäß unterschiedliche Sichtweisen. Zielspannungen werden z.B. gesehen zwischen innerwissenschaftlicher Prioritätensetzung und der Berücksichtigung außerwissenschaftlicher Relevanz, zwischen Breiten- und Spitzenförderung, aber auch zwischen der Aufwandsreduktion der Bewertungsverfahren und der Komplexität des Bewertungsgegenstandes, sowie zwischen dem wissenschaftlichen Anspruch auf Freiheit der Forschung und dem öffentlichen Anspruch auf Rechenschaftslegung und Effizienz.

Der Wissenschaftsrat empfiehlt den Scientific Communities, sich auf fachinterne Qualitätsstandards zu verständigen, damit diese in die Interpretation und Weiterentwicklung von Indikatoren eingehen können. Indikatoren sollten vor diesem Hintergrund mit Fachkenntnis, Methodenkenntnis und kritischer Reflexion interpretiert werden, auch und besonders in Peer Review Verfahren. Zudem empfiehlt der Wissenschaftsrat, dass der Begriff der „Leistung“ möglichst breit verstanden werden sollte und neben den Wissens- und Transferleistungen auch Forschungsdienstleistungen und Gremienarbeit einschließen sollte.

Auch sollten Steuerungszugänge vielfältig ausgestaltet sein und nicht eindimensional auf monetäre Anreize setzen, sondern auch die erhebliche Anreizwirkung z.B. der Gewährung individueller Freiheiten bei der Schwerpunktsetzung in der Karriereplanung oder einer indirekten Steuerung über einen Reputationswettbewerb in Betracht ziehen.



Der Wissenschaftsrat hat Leitlinien zu den folgenden 11 Punkten herausgearbeitet:

1. Selbstreflexivität
2. Verfahrensüberprüfung
3. Aufwandsbegrenzung
4. Angemessene Taktung
5. Vermeidung einer „Tonnenideologie“
6. Adäquanz
7. Einbeziehung aller Akteurs-Ebenen
8. Finanzielle Voraussetzungen
9. Attraktivität sichern (Gewinnung wissenschaftlichen Nachwuchses)
10. Berücksichtigung der Lehre
11. Vielfältigere Steuerungszugänge

Im Kontext der Tagung möchte ich die folgenden *Feststellungen und Empfehlungen* des Wissenschaftsrates besonders hervorheben:

- *Selbstreflexivität: es müssen institutionelle Mechanismen etabliert sein, damit die Einrichtungen über sich selbst informiert sind und Rechenschaft ablegen können.*  
Das heißt der Wissenschaftsrat geht davon aus, dass in weiten Teilen die Einrichtungen selbst die Datenlieferanten sind.
- *Verfahrensüberprüfung: Bewertungs- und Steuerungsverfahren müssen regelmäßig überprüft werden, damit sie sich nicht verselbständigen oder „unerwünschte Nebenwirkungen“ hervorrufen.*  
Dieser empfohlene Überprüfungsprozess würde durch eine wissenschaftliche Begleitung und eine systematische Forschung über das Wissenschaftssystem erleichtert und ggf. verbessert.
- *Vermeidung einer „Tonnenideologie“: ein Zuwachs an Quantität bedeutet nicht zwangsläufig einen Zuwachs an Qualität*  
Diese Feststellung des Wissenschaftsrates mahnt zu einem differenzierten Umgang mit quantitativen Indikatoren und zu ihrer Einbettung in einen Funktionskontext. Z.B. werden Promotions- und Habilitationszahlen als Messgröße für Forschungsleistung kritisch diskutiert, da sie nichts darüber aussagen, ob mit der Steigerung der Quantität auch ein gleichbleibend hohes Qualitätsniveau einhergeht. Vielmehr bestätigt sich auch in wissenschaftlichen Kontexten, dass beim Überschreiten eines gewissen Optimums der Grenznutzen sinken kann.
- *Zur Aufwandsbegrenzung sollte eine bessere Koordinierung von Evaluationen angestrebt werden.*  
Hier sind zwei Aspekte von Bedeutung: eine Koordinierung verschiedener Verfahren (z.B. interne und externe auf verschiedenen Ebenen) und eine sinnvolle Frequenz bzw. angemessene Rhythmisierung der Zeittakte.
- *Zur Aufwandsbegrenzung sollte eine Standardisierung der Datenerhebung angestrebt werden.*  
Dazu ist es dringend erforderlich, sich über Mindeststandards und Harmonisierung der Berichtssysteme der Hochschulen und außeruniversitären Einrichtungen zu verständigen. Die Verbesserung der Berichtssysteme soll aber nicht zu Lasten der Forschenden gehen. Zur Entlastung bei Datenvorhaltung, -bereitstellung und -pflege wird ein gutes Dokumentationswesen angemahnt, das von jeder Einrichtung vorgehalten werden soll.



Darüber hinaus formuliert der Wissenschaftsrat Empfehlungen zu spezifischen **Indikatoren**. Auch hierzu einige Beispiele:

Der Nutzen von **bibliometrischen Indikatoren** wie Publikationslisten, Publikationszahlen und Zitationszahlen wird in der Unterstützung von Peer-Review-Verfahren gesehen. Besonders die Berücksichtigung fachspezifischer Publikationskulturen ist nach Ansicht des Wissenschaftsrates nur durch Peers möglich. Ihre Nutzung erfordert eine Einordnung in den Fachgebetsdurchschnitt sowie auf Seiten der Peers Methodenkenntnis und kritische Reflexion von Chancen und Risiken der Bibliometrie.

Ein weiterer gängiger Indikator für Forschungsqualität (der bisherigen Forschungsleistungen) sind die **eingeworbenen Drittmittel**. Drittmittel sollten anhand der Vergabep Praxis gewichtet werden: diejenigen aus kompetitiven Begutachtungsverfahren sollten höher gewichtet werden als diejenigen aus nicht kompetitiven Verfahren. Die Projekte können eine Drittmittelstrategie illustrieren und damit die quantitativen Indikatoren ergänzen.

**Patentzahlen und die Erträge aus Patenten** sind aus Sicht des Wissenschaftsrates für sich genommen keine hinreichenden Indikatoren für den Anwendungsbezug von Forschung, sie sollten in Verbindung mit weiteren aussagekräftigen Indikatoren genutzt werden.

## 6 Ausblick

Damit hat der Wissenschaftsrat eine sehr differenzierte Einschätzung zum derzeitigen Stand der Bewertung- und Steuerung von Forschung gegeben, deren Ziele gewürdigt, aber in weiten Teilen auch Entwicklungspotentiale aufgezeigt. Die weitere Umsetzung der Empfehlungen erfordert eine Diskussion, welcher **konkrete** Handlungsbedarf zur Erfüllung der „11 Gebote“ besteht, und wo hierbei Prioritäten gesetzt werden sollten. Die Umsetzung erfordert aber auch ein kooperatives Vorgehen unterschiedlicher Akteure, von den individuellen Forschenden über die Scientific Communities bis zu Einrichtungen und Organisationen bis zur Systemebene.

Evaluationen und Monitoring von Forschungs-Input und -Output werden uns in Zukunft begleiten.

Welche Veränderungen sind absehbar, welche müssen auch tatsächlich berücksichtigt werden? Welche Stärken haben die Forschungsinformationssysteme und welche Schwächen behindern eine effiziente Nutzung? Ich bin sehr gespannt, welche Erkenntnisse die weiteren Diskussionen während und nach dieser Tagung hierzu liefern werden.

## Referenzen

*Gemeinsame Wissenschaftskonferenz (Hg)*, 2011: Pakt für Forschung und Innovation, Monitoring-Bericht 2011. Materialien der GWK, Heft 23. Bonn: GWK.

Online: <http://www.gwk-bonn.de/fileadmin/Papers/GWK-Heft-23-PFI-Monitoring-Bericht-2011.pdf>

*Wissenschaftsrat (Hg.)*, 2011: Empfehlungen zur Bewertung und Steuerung von Forschungsleistung. Drs. 1656-11. Köln: WR. Online: <http://www.wissenschaftsrat.de/download/archiv/1656-11.pdf>



## Forschungsinformationssysteme – Services und Anforderungen der DFG

Die Deutsche Forschungsgemeinschaft ist die Selbstverwaltungsorganisation der deutschen Wissenschaft. Getragen von Bund und Ländern ist ihr Auftrag die Förderung exzellenter Wissenschaft in allen ihren Zweigen. Im Jahr 2011 erfolgte dies unter Zugriff auf ein Budget von 2,3 Milliarden Euro. Pro Jahr werden mit diesen Mitteln mehr als 30.000 laufende Projekte gefördert, an denen schätzungsweise 20.000 Promovierende beteiligt sind. Die Deutsche Forschungsgemeinschaft ist damit auch der größte Förderer des wissenschaftlichen Nachwuchses in Deutschland.

Für das Informationshandeln der DFG ergeben sich daraus drei Erfordernisse: Zum einen ist sie als zentrale Akteurin in der Forschungslandschaft selbst Informationsanbieterin, die mit statistischen und weiteren Informationsangeboten über ihr Förderhandeln Auskunft gibt; zum zweiten ist sie selbst auf Informationsservices Dritter angewiesen, um ihr eigenes Förderhandeln bestmöglich an den sich laufend verändernden Bedarf anzupassen. Schließlich tritt sie zum dritten gegenüber ihren Mitgliedshochschulen mit spezifischen Services auf und ist beratend tätig.

Dieser Beitrag setzt zwei Akzente: Er beschreibt zunächst die DFG-eigenen Informationsdienstleistungen; anschließend skizziert er bestehende Herausforderungen und Lösungsansätze, welche die DFG bezüglich der Informationsbereitstellung auf Seiten der Hochschulen sieht.

### 1 Allgemeine Einführung

Die Deutsche Forschungsgemeinschaft spielt als Förderin für grundlagenorientierte Forschung insbesondere an Hochschulen seit jeher eine wichtige Rolle im Wissenschaftssystem. Mit dem steigenden Stellenwert von Drittmitteln für die Forschung an Hochschulen ergeben sich für die DFG auch eine Reihe informatorischer Verpflichtungen.

Dass Drittmittel für die Forschung an Hochschulen heute nahezu unverzichtbar sind, belegt eine im Auftrag der DFG durchgeführte Studie des Instituts für Forschungsinformation und Qualitätssicherung (iFQ). Die Befragung von 3.100 Professorinnen und Professoren an deutschen Universitäten, die insbesondere die Rahmenbedingungen von Forschung thematisierte, behandelte auch das Thema Drittmittel. 89 Prozent der Befragten gaben an, in den letzten fünf Jahren Drittmittelanträge mit einem Gesamtvolumen von über 25.000 Euro gestellt zu haben. Als „*primären Drittmittelgeber*“ nannten die Interviewten die DFG (vgl. Böhmer et al, 2011: 37ff). DFG-interne Berechnungen, die Personalzahlen des Statistischen Bundesamtes und die in den internen Datenbanken dokumentierten Zahlen zu Antragstellenden gegenüberstellen, bestätigen das Ergebnis in der Größenordnung: Etwa zwei Drittel aller Professorinnen und Professoren an Universitäten haben in einem Fünfjahreszeitraum (2006 bis 2010) Anträge bei der DFG eingereicht. Ein weiterer Befund dieser internen Analyse bestätigt die besondere Stellung der DFG aus einer anderen Perspektive: Von 100 Universitätsprofessorinnen und -professoren haben im selben Fünfjahreszeitraum 48 mindestens ein Gutachten für die DFG erstellt. Während die erste Zahl den hohen (und nach wie vor steigenden) Stellenwert von DFG-Drittmitteln illustriert, belegt die zweite Zahl, dass deren Vergabe durch breite Partizipation der scientific community unterstützt wird: Der Status der DFG als zentrale Selbstverwaltung der deutschen Wissenschaft spiegelt sich in der beträchtlichen ehrenamtlichen Beteiligung an der Beurteilung und Bewertung von Förderanträgen wider.

Mit der zentralen Positionierung der DFG im deutschen Wissenschaftssystem gehen eine Reihe informatorischer Verpflichtungen einher:

- **Öffentliches Interesse:** Forschung und deren wettbewerblich ausgestaltete Förderung

stößt auf zunehmendes öffentliches Interesse. Informationen hierzu müssen in einer Form bereitgestellt werden, die dem Anspruch an Allgemeinverständlichkeit genügt und transparent macht, welche Forschungsziele die von der DFG mit öffentlichen Mitteln geförderten Projekte verfolgen.

- **Rechenschaft:** Geldgeber und Steuerzahler fordern ein hohes Maß an Transparenz über die Verwendung von Mitteln ein.
- **Planungsdaten:** Die DFG ist zunehmend gefordert, Aktivitäten strategisch auszurichten und faktenbasiert zu begründen.
- **Erfolgsbewertung:** Die Frage nach Fakteninformationen über den Erfolg von Programmen und Projekten gewinnt an Bedeutung.

Als Einrichtung, die von ihren Mitgliedern, insbesondere von Hochschulen, getragen wird, sieht sich die Deutsche Forschungsgemeinschaft dabei nicht nur gegenüber ihren Entscheidungsgremien und Mittelgebern in der Verantwortung, sondern auch und insbesondere gegenüber diesen Mitgliedseinrichtungen, die ebenfalls Adressat und Abnehmer der von der DFG entwickelten Informationsservices sind.

## 2 Datengrundlage der DFG-Informationsservices

Das Spektrum datengestützter Informationsdienstleistungen ist dabei weit gefasst. Es reicht von der Bereitstellung von Statistiken, die vornehmlich das Förderhandeln der DFG abbilden, über die Planung und Durchführung programmevaluativer Studien bis hin zum Aufbau und laufenden Weiterentwicklung kennzahlenbasierter Systeme (z.B. Förder-Ranking, Gleichstellungs-Monitoring). Ebenso umfasst es das in großen Teilen datenbankgestützte Informationsportfolio des DFG-Internets sowie die auf den DFG-Webseiten angebotenen Informationssysteme GEPRIS und Research Explorer.

Eine Hauptquelle für diese Services bildet die DFG-interne Datenbank zur Antragsbearbeitung. Bereits vor nunmehr 30 Jahren hat die DFG begonnen, ihr Förderhandeln elektronisch zu erfassen – zunächst in erster Linie für statistische Zwecke, zum zweiten aber beispielsweise auch, um auf Basis der Daten als Anlage zum Jahresbericht ein Verzeichnis der im Berichtszeitraum geförderten Projekte zu erstellen. Erfolgte bis 2004 die Datenerfassung durch eine zentrale Arbeitseinheit, wurde diese mit der Einführung des Systems ElektrA (elektronische Antragsbearbeitung) dezentralisiert. Zu dem Zweck der Erfassung statistischer Daten kam mit dieser Umstellung die Verwendung der Daten für maschinell erstellte Formschriften hinzu, also z.B. Briefe an Antragstellende, die über die Entscheidung zu einem Antrag informieren, Briefe an Wissenschaftlerinnen und Wissenschaftler, in denen diese um die Begutachtung eines Antrags gebeten werden, Sitzungsunterlagen für Beratungs- und Entscheidungsgremien und vieles andere mehr. Für die Formschriften besteht ein eigener Workflow, mit dem sehr genau geprüft wird, dass dort erfolgte Angaben, etwa zur bewilligten Summe, tatsächlich korrekt sind. Seit der Einführung von ElektrA sind circa 400 Mitarbeitende der Geschäftsstelle praktisch täglich mit der Erfassung von Daten befasst.

Die Qualität der Daten wird aber nicht nur über die im Prozess der Antragsbearbeitung produzierten Formschriften gewährleistet. Zusätzlich verantwortet ein eigenes Team die Qualitätssicherung der ElektrA-Daten, etwa indem regelmäßig geprüft wird, ob Personen als Dubletten angelegt wurden oder ob Termine (z.B. von Gutachtersitzungen) vollständig und korrekt eingetragen sind. Das zentrale Qualitätssicherungsteam hat bei seiner Arbeit vor allem die aus den Daten generierten Informationsservices der DFG im Blick.

### **3 Ausgewählte Informationsprodukte**

#### **3.1 DFG bearbeitet jährlich mehr als 400 statistische Anfragen**

Die DFG bearbeitet auf Basis der beschriebenen Daten jährlich circa 400 statistische Anfragen – Tendenz steigend. Etwa ein Drittel dieser Anfragen stammen von Externen, also zum Beispiel Ministerien, Hochschulverwaltungen, Wissenschaftlerinnen und Wissenschaftlern oder Journalistinnen und Journalisten. Das Spektrum der Fragen reicht von der Zahl der Heisenberg-Professuren in einem einzelnen Bundesland über die Förderung bestimmter Themen- oder Fachgebiete bis hin zur Nachwuchsförderung oder der Beteiligung von Frauen und Männern an bestimmten Programmen. Interne Anfragen dienen der Vorbereitung von Sitzungen (etwa der Fachkollegien), der Berichtslegung gegenüber den Gremien der DFG (z.B. zur Entwicklung eines Programms sowie der demographischen Zusammensetzung seiner Klientel) oder der Öffentlichkeitsarbeit.

#### **3.2 Das DFG-interne Management hat Zugriff auf über hundert datenbankgestützte Reports**

Neben diesen sogenannten *ad-hoc-Statistiken* hat die DFG in den letzten Jahren ein umfassendes internes Berichtswesen etabliert, das zwei Linien unterscheidet: Zum einen das datenbankgestützte Reporting und zum zweiten analyseförmige Berichte zu ausgewählten Fragestellungen.

Beim datenbankgestützten Reporting handelt es sich um listenförmige Übersichten sowie um Statistiken, die von Mitarbeiterinnen und Mitarbeitern der Geschäftsstelle quasi „auf Knopfdruck“ im Intranet der DFG abgerufen werden können. Etwa hundert dieser Reports unterstützen die Fachabteilungen bei der Budgetüberwachung, informieren über statistische Entwicklungen in allen von der DFG geförderten Fachgebieten (in der Differenzierung nach über 200 Fächern) sowie über die Nachfrageentwicklung in den verschiedenen Programmlinien der DFG. Personenbezogene Reports werden etwa herangezogen, um vor Auswahl eines Gutachters oder einer Gutachterin deren bisherige Begutachtungsbelastung zu prüfen – um bei zu hoher Beanspruchung nach Möglichkeit auf alternative Kandidaten oder Kandidatinnen auszuweichen. Mit über 32.000 Aufrufen jährlich werden diese Reports in der Geschäftsstelle sehr gut angenommen.

#### **3.3 Die Entscheidungsgremien nutzen analyseförmige Berichte zur Entwicklung von DFG-Programmen und deren Nutzerkreisen**

Analyseförmige Berichte werden vor allem von den Entscheidungsgremien der DFG in Auftrag gegeben. In ihnen wird etwa die Entwicklung eines Programms (z.B. Heisenberg-Stipendien) abgebildet, also zum Beispiel die Nachfrage nach diesem Programm, die Erfolgsquote bei der Antragstellung, die Beteiligung von Männern und Frauen, die regionale und fachliche Verteilung sowie die Altersstruktur der Antragstellenden. Die Analysen dienen der Programmplanung, finden aber auch in der Öffentlichkeitsarbeit Verwendung – etwa indem auf sogenannten Vernetzungstreffen von Geförderten im Heisenberg-Programm den Veranstaltungsteilnehmerinnen und -teilnehmern entsprechende Analysen zu „ihrem“ Programm zur Verfügung gestellt werden. Im Format *DFG Infobrief* werden entsprechende Analysen auch auf den Internetseiten der DFG publiziert, zuletzt etwa zur Frage der demographischen Zusammensetzung der DFG-Gutachterschaft (Reinhardt 2009), zur Heisenberg-Professur (Imöhl 2010) oder zum Förderinstrument *Eigene Stelle* (Imöhl 2011).

### 3.4 Der DFG-Förderatlas adressiert vor allem die Mitgliedshochschulen der DFG

Das wohl bekannteste Statistikprodukt der DFG ist das sogenannte Förder-Ranking. Die erste Ausgabe des Rankings erschien 1997, damals noch unter dem etwas sperrigen Titel *Bewilligungen nach Hochschulen – Bewilligungsvolumen 1991 bis 1995, Anzahl kooperativer Projekte 1996* (DFG 1997). Thema war also der institutionelle Erfolg im Einwerben von DFG-Drittmitteln; zurück ging der Bericht auf eine Initiative von zehn großen Hochschulen, welche die DFG explizit um die Erstellung einer entsprechenden Rangreihe gebeten hatten. Auch heute noch sind die Mitgliedshochschulen der DFG der primäre Adressat dieses Monitoring-Services, auf große Nachfrage stößt er aber auch bei den Wissenschaftsministerien des Bundes und der Länder, bei Fachgesellschaften sowie, in seiner englischen Ausgabe, bei ausländischen Forschungs- und Förderorganisationen.

Im Mai 2012 ist die sechste Ausgabe der Reihe erschienen – erstmals unter dem Titel *Förderatlas* (vgl. DFG 2012). Die Umbenennung des seit der Ausgabe 2003 im 3-Jahres-Turnus als *Förder-Ranking* erscheinenden Berichts war vor allem damit begründet, dass sich die Ausrichtung des Berichts von Ausgabe zu Ausgabe deutlich vom reinen Bewilligungsvolumen weg zur Frage der sich anhand dieser Bewilligungen ablesbaren Profilentwicklungen von Hochschulen und regionalen Forschungsstandorten verschoben hat. Die Umbenennung folgt dabei auch einem politischen Impetus: Nicht das „immer mehr und immer höher“ ist das, was bei dieser statistischen Betrachtung von Drittmiteleinwerbungen im Vordergrund steht, sondern das „wo und wofür“.

Standen dabei in den ersten beiden Ausgaben des Berichts tatsächlich DFG-Bewilligungen im alleinigen Fokus, sind seither eine Reihe weiterer Quellen und damit auch Kennzahlen hinzugekommen, die es erlauben, die Frage nach den Forschungsprofilen auch nach Drittmittelgebern zu differenzieren. Dem Förderatlas 2012 liegen Daten zur Förderung im 7. Forschungsrahmenprogramm der EU zugrunde, ein eigenes Kapitel beschäftigt sich im europäischen Vergleich mit der Beteiligung deutscher Wissenschaftlerinnen und Wissenschaftler an den Förderungen durch das 2007 neu gegründete European Research Council (ERC). Die direkte Projektförderung durch den Bund geht ebenso in den Bericht ein, wie Zahlen zur Förderung von Gastwissenschaftleraufenthalten durch die Alexander von Humboldt-Stiftung (AvH) oder den Deutschen Akademischen Austauschdienst (DAAD). In allen Fällen werden die herangezogenen Daten in fachlich differenzierter Form ausgewertet, für die DFG etwa nach 48 Forschungsfeldern (in Anlehnung an die 48 DFG-Fachkollegien), für AvH und DAAD erfolgt eine Aggregation der dort eingesetzten Fachklassifikationen nach den 12 Fachgebieten der DFG-Fachsystematik. Bund und EU gehen mit ihren thematischen Fördergebieten in die Statistik ein. Auf diese Weise entstehen fachlich äußerst differenzierte Möglichkeiten der Profildarstellung, die neben der Fokussierung auf Einrichtungen (sowohl Hochschulen wie auch außeruniversitäre Einrichtungen insbesondere der großen Forschungsorganisationen) auch Regionen und sich dort manifestierende Vernetzungen und Clusterbildungen in den Blick nehmen.

Gegenüber den vorherigen Ausgaben wurde das Themenspektrum um eine Betrachtung des Stellenwerts internationaler Zusammenarbeit am Beispiel einer einzelnen Disziplin (der Chemie) auf Basis bibliometrischer Daten erweitert. Ein weiteres Sonderkapitel beschäftigt sich mit der Beteiligung universitätsmedizinischer Einrichtungen an der DFG-Förderung. Den besonderen Stellenwert, den die DFG Fragen der Gleichstellung in der Wissenschaft beimisst, bringt schließlich die umfassende Differenzierung der berichteten Kennzahlen nach Geschlecht zum Ausdruck.

Alleinstellungsmerkmal und weiteres Kennzeichen für den besonderen Servicecharakter des Förderatlas ist, dass alle dort berichteten Kennzahlen aus bereits bestehenden Datensammlungen generiert werden. Wie die DFG verfügen auch die anderen im Bericht erscheinenden Förderinstitutionen über Datenbanken, die deren Förderhandeln abbilden und unterstützen. Daten des Statistischen Bundesamtes (Personal- und Drittmittelstatistik) werden in einem seit Jahren eta-



blierten und gesetzlich legitimierten Prozess bei den Hochschulen erhoben, bibliometrische Daten stammen von kommerziellen Anbietern bzw. vom Kompetenzzentrum Bibliometrie (vgl. [www.bibliometrie.info](http://www.bibliometrie.info)), das diese in aufwändigen Prozeduren für Analysezwecke aufbereitet und qualitätssichert. Kein Mitarbeiter und keine Mitarbeiterin einer Hochschulverwaltung muss für diesen Bericht einen Fragebogen ausfüllen, keine Wissenschaftlerin und kein Wissenschaftler wird mit Datenerhebungen behelligt. Indem für den Förderatlas nahezu ausschließlich prozess-produzierte Daten herangezogen werden, ist gleichzeitig eine hohe Qualität der berichteten Kennzahlen gewährleistet: Sie bilden faktisches Förderhandeln ab und sind so frei von wissenschaftlichen oder unwissenschaftlichen Manipulationen, wie sie für surveygestützte Verfahren typisch sind.

### **3.5 Online zugängliche Informationssysteme der DFG**

Eine die statistischen Services ergänzende Informationslinie bilden die internetgestützten Informationssysteme der DFG, namentlich die Projektdatenbank GEPRIS (German Project Information System, vgl. [www.dfg.de/gepris](http://www.dfg.de/gepris)) und das Institutionenverzeichnis Research Explorer ([www.research-explorer.de](http://www.research-explorer.de)). Wie den zuvor beschriebenen Dienstleistungen liegen auch diesen Online-Datenbanken nahezu ausschließlich Daten zugrunde, die im Prozess der Antragsbearbeitung sowie durch die zentrale Datenpflegeeinheit der DFG generiert bzw. qualitätsgesichert wurden. Für die Projektdatenbank GEPRIS finden dabei neben strukturierten Datenbankinformationen (Namen der Projektverantwortlichen, Fach, Ort, Einrichtung, Institut etc.) auch Texte Verwendung, die im zentralen Dokumentenmanagementsystem (DMS) der DFG elektronisch vorgehalten werden. Es handelt sich hierbei zum einen um das Abstract, mit dem Antragstellende die Kernziele ihres Forschungsprojekts in möglichst allgemein verständlicher Form beschreiben. Ursprüngliche Zielgruppe dieser Abstracts bilden die Entscheidungsgremien der DFG (und etwa nicht: die Begutachtenden). Seit nunmehr 12 Jahren dienen diese Abstracts via GEPRIS auch der Öffentlichkeitsarbeit. Zum anderen werden seit 2010 in GEPRIS auch die Abstracts aus den Abschlussberichten von DFG-Projekten sowie (teilweise in Auszügen) die dort enthaltenen Listen der aus einem Projekt hervorgegangenen Publikationen der Öffentlichkeit zugänglich gemacht. GEPRIS informiert so nicht nur über die Zielsetzungen DFG-geförderter Projekte, sondern gibt auch über deren Ergebnisse Auskunft.

Mit mehr als 80.000 Projekten unter Beteiligung von über 42.000 Antragstellenden aus knapp 20.000 Instituten an Hochschulen und außeruniversitären Forschungseinrichtungen ist GEPRIS heute eines der größten disziplinenübergreifenden Forschungsinformationssysteme weltweit – seit 2011 auch mit einer englischsprachigen Nutzerführung, die ausländischen Interessierten den strukturierten Zugang zu dieser Informationsquelle erleichtert.

Der Research Explorer ist seit 2008 online. Zielsetzung des gemeinsam mit dem DAAD entwickelten Informationssystems ist der einfache Zugang zu Informationen zu den „Stätten der Forschung“ in Deutschland (vgl. Abbildung 1). Das System ist durchgängig zweisprachig (deutsch/englisch) konzipiert und bietet im Sinne eines Portals Zugang zu den Homepages von knapp 20.000 Hochschulinstituten und außeruniversitären Forschungseinrichtungen. Die Namen dieser Institute sind per Stichwortsuche recherchierbar, die sehr fein gegliederte Fachklassifikation des Statistischen Bundesamtes erlaubt die gezielte Suche nach Instituten einer bestimmten fachlichen Ausrichtung. Die Nutzung regionaler Filtermöglichkeiten grenzt das Suchergebnis auf ausgewählte Bundesländer oder Orte ein. Seit Mitte 2011 ist es dank einer Zusammenarbeit mit der Hochschulrektorenkonferenz und der Verknüpfung mit dem dort betreuten Hochschulkompass auch möglich, über den Research Explorer Informationen zu den Promotionsmöglichkeiten an den einzelnen Standorten sowie in den dort angebotenen Fächern abzurufen. Durch Verlinkung mit dem Projektinformationssystem GEPRIS bietet der Research Explorer darüber hinaus ein nach Einrichtungen und Instituten gegliedertes Verzeichnis DFG-geförderter Projekte.

Abbildung 1: Das Informationssystem Research Explorer



Eine Kooperation zwischen:



Deutscher Akademischer Austausch Dienst  
German Academic Exchange Service



Deutsche Forschungsgemeinschaft

[Startseite](#)
[Geographische Suche](#)
[Fächerbezogene Suche](#)
[Erweiterte Suche](#)
[Tipps / Hinweise](#)
[Kontakt / Impressum](#)

## Research Explorer

Das Forschungsverzeichnis der Deutschen Forschungsgemeinschaft (DFG) und des Deutschen Akademischen Austauschdienstes (DAAD) in Zusammenarbeit mit der Hochschulrektorenkonferenz (HRK).

**Der Research Explorer erschließt über 19.000 Institute an deutschen Hochschulen und außeruniversitären Forschungseinrichtungen nach geografischen, fachlichen und strukturellen Kriterien.**

Sie haben auf den folgenden Seiten unterschiedliche Möglichkeiten, den Research Explorer für Ihre Recherchen zu nutzen. Für einen ersten Einstieg können Sie auch direkt unten auf der Karte einen Forschungsstandort anwählen oder ihn über Anklicken in der Liste aufrufen.



Stadt	Institutionen
Aachen	20
Aalen	2
Alfter	1
Alt Ruppin	1
Amberg	1
Ansbach	1
Aschaffenburg	1
Augsburg	4
Bad Harzburg	1
Bad Honnef	1
Bad Münstereifel	1
Bad Nauheim	1
Bad Sooden-Allendorf	1
Bad Wurzach	1
Bakum	1
Bamberg	4
Bayreuth	6
Benediktbeuern	1
Bennewitz	1



Auch beim Research Explorer handelt es sich genau genommen um eine Zweitverwertung von Daten, deren ursprünglicher Zweck die Unterstützung der Antragsbearbeitung in den Fachabteilungen der DFG ist. Das dem System zugrunde liegende Adressverzeichnis wird von den Sachbearbeiterinnen und Sachbearbeitern der DFG für die datenbankgestützte Generierung von Adressfeldern in Briefen oder zum Beispiel auch von Teilnehmerlisten bei Sitzungen genutzt. Die Daten sind so strukturiert, dass sie in der Bewilligungsstatistik der DFG für beliebig aggregierte Analysen (nach Instituten, Fakultäten, Einrichtungen, Orten, aber etwa auch nach Instituten bestimmter Fachrichtungen) heranzuziehen sind. Die tägliche Verwendung der Daten für Zwecke der Antragsbearbeitung gewährleistet eine hohe Datenaktualität und -qualität vor allem bei jenen Instituten, an denen Wissenschaftlerinnen und Wissenschaftler regelmäßig als Antragstellende oder Begutachtende mit der DFG in Kontakt treten. Die Aktualisierung oder Korrektur von Daten ist darüber hinaus über ein einfach zu bedienendes Formularfeld innerhalb des Research Explorers möglich.

Bei der Entwicklung des Systems wurde darauf geachtet, dass es einen Einstieg nicht nur über die in Abbildung 1 veranschaulichte bundesweite geographische Suchmaske erlaubt, sondern auch



regionen- oder ortsspezifische Zugänge<sup>1</sup>. Dies nutzen etwa die Wissenschaftsministerien von Baden-Württemberg und Nordrhein-Westfalen, um auf ihren Webseiten via Research Explorer die Forschungseinrichtungen ihrer Länder recherchierbar zu machen. Hochschulen wiederum haben die Möglichkeit, über ihre Webseiten direkt auf die Einrichtungen am jeweiligen Ort oder auf die Institute der jeweiligen Hochschule zu verlinken<sup>2</sup>. Eine solche Verlinkung bietet sich etwa dort an, wo Hochschulen nach einer einfachen Möglichkeit suchen, ein umfassendes Bild der dort durch die DFG geförderten Forschungsvorhaben zu präsentieren: Viele Hochschulen weisen auf ihren Webseiten bisher nur ihre Beteiligungen an den „großen“ DFG-Programmen aus, also etwa erfolgreich eingerichtete Sonderforschungsbereiche, Forschergruppen, Graduiertenkollegs oder Beteiligungen an Fördermaßnahmen im Rahmen der Exzellenzinitiative. Research Explorer und GEPRIS machen darüber hinaus auch das eigentliche Rückgrat der DFG-Förderung, die Einzelförderung, sichtbar.

#### **4 Herausforderungen und Lösungsansätze auf Seiten der Hochschulen**

Die bisherigen Ausführungen haben verdeutlicht, dass die zentrale und strukturierte Erschließung organisationsinterner Informationsbestände – in diesem Falle der DFG – eine Vielzahl an Nutzungsmöglichkeiten entsprechender Daten erlaubt. Die Palette reicht von der Bedienung statistischer ad-hoc-Anfragen über den Aufbau monitoringförmiger Kennzahlensysteme bis hin zur datenbankgestützten Generierung von Informationssystemen. Die DFG hat in den Aufbau ihrer Datenhaltung viel investiert. Diese Investitionen amortisieren sich, weil die Daten sowohl für die maschinengestützte Erstellung antragsbearbeitungsrelevanter Dokumente als auch für die Generierung von Daten für interne und externe Forschungsmanagement unterstützenden Statistik-Services sowie für die Öffentlichkeitsarbeit genutzt werden.

Hochschulen stehen vor ähnlichen Herausforderungen wie die Deutsche Forschungsgemeinschaft. Auch hier werden unter dem Zeichen zunehmender Hochschulautonomie und eines wachsenden auch institutionellen Wettbewerbs Daten benötigt, welche die eigene Performanz widerspiegeln und Auskunft zur Profilentwicklung geben. Dies können Daten zu den Projekten sein, die über erfolgreich eingeworbene Drittmittel ermöglicht wurden, aber ebenso auch Daten zur grundfinanzierten Forschung. Auf der Outputseite sind es vor allem Publikationsverzeichnisse, mit denen Hochschulen auf die Ergebnisse ihrer Forschung aufmerksam machen. Auch der Nachweis von an einem Standort erfolgreich abgeschlossenen Qualifizierungsarbeiten (etwa in Form von Dissertationen) ist eine denkbare Option.

Entsprechende Daten werden für hochschulinterne Zwecke benötigt, werden häufig aber auch von Externen und hierbei vor allem für evaluative Zwecke nachgefragt. Sei es das Rating des Wissenschaftsrats (vgl. Wissenschaftsrat 2007) oder das Ranking des Centrums für Hochschulentwicklung (CHE), sei es für Zwecke der leistungsorientierten Vergabe von Landesmitteln (vgl. KMK 2011) oder für ein internationales Benchmarking europäischer Hochschulverbünde (etwa der EUA – European University Association). Oft behandeln diese Erhebungen ähnliche Themen mit allerdings (oft nur graduell) divergierenden Berichtsgegenständen und -zeiträumen. Antworten „aus einem Guss“ sind so häufig schwierig, mit der Konsequenz, dass jede Erhebung zu einer solitären Kraftanstrengung der Hochschulverwaltung führt.

---

1 Daneben bietet der Research Explorer eine fachspezifische Suche sowie eine erweiterte Suche an, die etwa auch die Eingrenzung auf Institute bestimmter Einrichtungsarten (z.B. Mitgliedseinrichtungen der Max-Planck-Gesellschaft) erlaubt.

2 Der Link

[http://research-explorer.dfg.de/research\\_explorer.de.html?ort=1&institution\\_id=10284&schritt=detail](http://research-explorer.dfg.de/research_explorer.de.html?ort=1&institution_id=10284&schritt=detail) macht in der Spezifikation „ort=1“ beispielsweise Einrichtungen am Standort Bonn zugänglich, mit „institution\_id=10284“ wird die Suche weiter auf Institute der Rheinischen Friedrich-Wilhelms-Universität Bonn eingegrenzt.

Eine Alternative zu einer rein nachfrageorientierten Informationserhebung und -aufbereitung bietet der Aufbau hochschulweiter Informationssysteme. Das Themenspektrum, das solche Systeme abzudecken haben, ist bekannt – und skalierbar: Ein System, das etwa damit beginnt, den Publikations-Output der an einer Hochschule beschäftigten Wissenschaftlerinnen und Wissenschaftler zu erfassen, stellt bereits eine Herausforderung dar. Ist diese gemeistert, kann ein Ausbau des Systems in Angriff genommen werden.

Oft verfügen Hochschulen bereits über Daten, die für ein hochschulweit zugreifbares Informationssystem nutzbar sind: Durch Verbindung mit einem elektronischen Personalverzeichnis ist etwa die Grundlage geschaffen für eine Datenerhebung personenbezogener Informationen. Per Zugriff auf die Haushaltsdatenbank werden Informationen zu eingenommenen Drittmitteln (und die damit geförderten Projekte) nutzbar. Mancherorts existieren bereits seit Jahrzehnten etablierte Workflows der Publikationsdatenerhebung, etwa weil diese für die Erstellung einer jährlich aktualisierten Hochschul-Bibliographie genutzt werden. An anderen Orten hat sich wiederum eine strukturierte Forschungsberichterstattung etabliert, die aber nur den Bericht, nicht aber die dahinter stehenden Daten und deren Nützlichkeit für aussagekräftige Analysen zum Forschungshandeln in den Blick nimmt.

In vielen Ländern aber auch bundesweit gab es immer wieder Überlegungen, eine Datenbank aufzubauen, die an zentraler Stelle Daten zur Forschung an Hochschulen und außeruniversitären Forschungseinrichtungen sammelt und aufbereitet. Bisher sind alle Versuche in diese Richtung gescheitert bzw. nur in Teilen gelungen – etwa in Form von Informationssystemen, die zwar dem Nachweis der in einem Land durchgeführten Forschungsprojekte dienen, aber keine Daten bereit halten, die auch für statistische Zwecke nutzbar sind. Auch im Kontext der Überlegungen zur weiteren Entwicklung des Ratings des Wissenschaftsrats beschäftigt sich eine Unterarbeitsgruppe *Standardisierung der Datenerhebung* mit der Frage, ob und wie zukünftig eine Datenerhebung zu organisieren wäre, die mit einem angemessenen Kosten-Nutzen-Verhältnis weitere Ratings ermöglicht und auch weitere Informationsbedürfnisse nachhaltig unterstützt (vgl. <http://www.wissenschaftsrat.de/arbeitsbereiche-arbeitsprogramm/forschungsrating/standardisierung/>). Der Aufbau einer zentralen „Super-Datenbank“ wird auch hier kritisch gesehen, die Diskussion geht eher in die Richtung einer Einigung auf einrichtungsübergreifende Standards. Diese reichen von der Verwendung einheitlicher Identifikationsnummern (IDs) für Personen<sup>3</sup> oder Institute (ID der DFG-Institutsdatenbank) bis hin zu einheitlichen Klassifikationssystemen für Zeitschriften und andere Verlagspublikationen oder Publikationsnummern für Online- und Verlagsveröffentlichungen (DOI, ISBN)<sup>4</sup>.

Weit fortgeschritten und zum Teil schon an Ergebnissen ablesbar sind entsprechende Entwicklungen in den USA. Dort wurde in jüngster Vergangenheit das Projekt *Star Metrics* aufgesetzt, das auf Basis bundesweit einheitlicher Standards den Aufbau eines anspruchsvollen Kennzahlensystems anstrebt. Übergeordnetes Ziel von Star Metrics ist es, die Wirksamkeit öffentlicher Forschungsinvestitionen zu messen. Kernelement des Projekts bildet ein einheitliches Schlüssel-System, das den Austausch von Daten zwischen Forschungsförderern (z.B. NSF und NIH) und geförderten Einrichtungen erlaubt (vgl. <https://www.starmetrics.nih.gov/>). Die Nutzung der Daten dient dem Ausbau bereits prototypisch zugänglicher Informationssysteme wie dem R+D Dashboard, das derzeit in einer Beta-Version zugänglich ist (<http://rd-dashboard.nitrd.gov/>) und den Zugriff auf einrichtungs-, orts- und regionenspezifische statistische Informationen ebenso ermöglicht wie die Darstellung konkreter Forschungsprojekte.

---

3 Vgl. hierzu Projekte kommerzieller Anbieter wie ResearcherID.com von Thomson Reuters, oder das internationale Netzwerk ORCID.org (Open Research Contributor ID).

4 Ein wichtiger Ansprechpartner für darüber hinaus gehende Datenaustauschformate ist das europäische Netzwerk EuroCris (European Current Research Information Systems), das für diesen Zweck das CERIF-Format entwickelt hat (vgl. [www.eurocris.org](http://www.eurocris.org)), siehe dazu auch den Beitrag von Brigitte Jörg in diesem Band.

## 5 Fazit

Am Beispiel der Informationsservices der DFG wurde in diesem Beitrag veranschaulicht, welche Vorteile aus einer zentralen Datenhaltung und -verwaltung für die Erbringung vielfältigster Informationsdienstleistungen resultieren. Der Aufwand, der in die Entwicklung eines solchen Systems zu investieren ist, ist beachtlich. Auch die laufenden Betriebskosten im Anschluss an die Entwicklung sind nicht unerheblich. Dabei darf nicht übersehen werden, dass neben den Kosten für technischen Betrieb und Weiterentwicklung vor allem (Personal-) Kosten für die Pflege und Qualitätssicherung entstehen: Noch das technisch ambitionierteste System bleibt wertlos, wenn keine Sorge dafür getragen wird, dass es sowohl von den adressierten Personenkreisen genutzt wird als auch die dort abgelegten Daten von einer hierauf spezialisierten zentralen Arbeitseinheit laufend qualitätsgesichert werden. Dies gilt umso mehr im Falle dezentraler Pflege der Daten. Der Aufwand amortisiert sich nur dann, wenn entsprechende Systeme keinem single-item-Auftrag folgen (also etwa „nur“ Informationen über laufende Projekte oder nur über erfolgte Publikationen bereitstellen), sondern universell einsetzbar sind und genutzt werden.

Die DFG unterstützt ihre Mitgliedshochschulen beim Aufbau entsprechender Systeme durch die Bereitstellung von Standard-Schlüsseln (z.B. Instituts-ID), die etwa verwendet werden können, um hochschuleigene Systeme mit der DFG-Projektdatenbank GEPRIS oder mit dem Research Explorer zu verknüpfen. Auch die Bereitstellung hochschulspezifischer Bewilligungsübersichten, wie sie seit 2008 den DFG-Vertrauensdozenten und Leitungen von Mitgliedshochschulen zur Verfügung gestellt werden, soll weiter ausgebaut werden.

## Literatur

- Böhmer, Susan / Neufeld, Jörg / Hinze, Sybille / Klode, Christian / Hornbostel, Stefan*, 2011: Wissenschaftler-Befragung 2010: Forschungsbedingungen von Professorinnen und Professoren an deutschen Universitäten. iFQ-Working Paper No.8. Bonn.  
Online: [www.forschungsinform.de/Publikationen/Download/working\\_paper\\_8\\_2010.pdf](http://www.forschungsinform.de/Publikationen/Download/working_paper_8_2010.pdf)
- Borgwardt, Angela*, 2011: Rankings im Wissenschaftssystem – zwischen Wunsch und Wirklichkeit. Publikation zur Konferenz der Friedrich-Ebert-Stiftung am 17. März 2011. Schriftenreihe des Netzwerk Exzellenz an Deutschen Hochschulen. Berlin: Friedrich-Ebert Stiftung.  
Online: [library.fes.de/pdf-files/studienfoerderung/08407.pdf](http://library.fes.de/pdf-files/studienfoerderung/08407.pdf)
- Deutsche Forschungsgemeinschaft*, 1997: DFG-Bewilligungen nach Hochschulen – Bewilligungsvolumen 1991 bis 1995, Anzahl kooperativer Projekte im Jahr 1996. Bonn. Online: [www.dfg.de/foerderatlas/archiv](http://www.dfg.de/foerderatlas/archiv)
- Deutsche Forschungsgemeinschaft*, 2009: Förder-Ranking 2009. Institutionen – Regionen – Netzwerke. Fachliche Profile von Hochschulen und außeruniversitären Forschungseinrichtungen im Licht öffentlich geförderter Forschung. Bonn. Online: [www.dfg.de/foerderatlas/archiv](http://www.dfg.de/foerderatlas/archiv)
- Deutsche Forschungsgemeinschaft*, 2012: Förderatlas 2012. Kennzahlen zur öffentlich finanzierten Forschung in Deutschland. Bonn. Online: [www.dfg.de/foerderatlas/](http://www.dfg.de/foerderatlas/)
- Güdler, Jürgen*, 2011: German Funding Ranking as a Tool for Self-Management. International Higher Education, 63, 27-29. Online: [www.bc.edu/research/cihe/ihe.html](http://www.bc.edu/research/cihe/ihe.html)
- Imöhl, Isabell*, 2010: Die Heisenberg-Professur – ein statistisches Profil. DFG Infobrief 2/2010. Bonn. Online: [www.dfg.de/dfg\\_profil/evaluation\\_statistik/programm\\_evaluation/dfg\\_infobrief/](http://www.dfg.de/dfg_profil/evaluation_statistik/programm_evaluation/dfg_infobrief/)
- Imöhl, Isabell*, 2011: Die „Eigene Stelle“ – ein DFG-Förderinstrument im Spiegel der Statistik. DFG Infobrief 1/2011, Bonn.  
Online: [www.dfg.de/dfg\\_profil/evaluation\\_statistik/programm\\_evaluation/dfg\\_infobrief/](http://www.dfg.de/dfg_profil/evaluation_statistik/programm_evaluation/dfg_infobrief/)
- Kultusministerkonferenz (KMK)*, 2011: Instrumente der Qualitätsfeststellung in der Hochschulforschung – Erfahrungen der Länder. Beschluss der Kultusministerkonferenz vom 12.5.2011. Online: [www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2011/2011\\_05\\_12-Instru-](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2011/2011_05_12-Instru-)

mente-Qualitätsfeststellung.pdf

*Reinhardt, Anke*, 2009: Gutachterinnen und Gutachter in der DFG – Begutachtungen und Demografie. DFG Infobrief 2/2009. Bonn.

Online: [www.dfg.de/dfg\\_profil/evaluation\\_statistik/programm\\_evaluation/dfg\\_infobrief/](http://www.dfg.de/dfg_profil/evaluation_statistik/programm_evaluation/dfg_infobrief/)

*Wissenschaftsrat*, 2007: Forschungsleistungen deutscher Universitäten und außeruniversitärer Einrichtungen in der Chemie. Köln, 18.12.2007.

Online: [www.wissenschaftsrat.de/download/Forschungsrating/Dokumente/Pilotstudie\\_Forschungs-rating\\_Chemie/pilot\\_ergeb\\_chemie.pdf](http://www.wissenschaftsrat.de/download/Forschungsrating/Dokumente/Pilotstudie_Forschungs-rating_Chemie/pilot_ergeb_chemie.pdf)

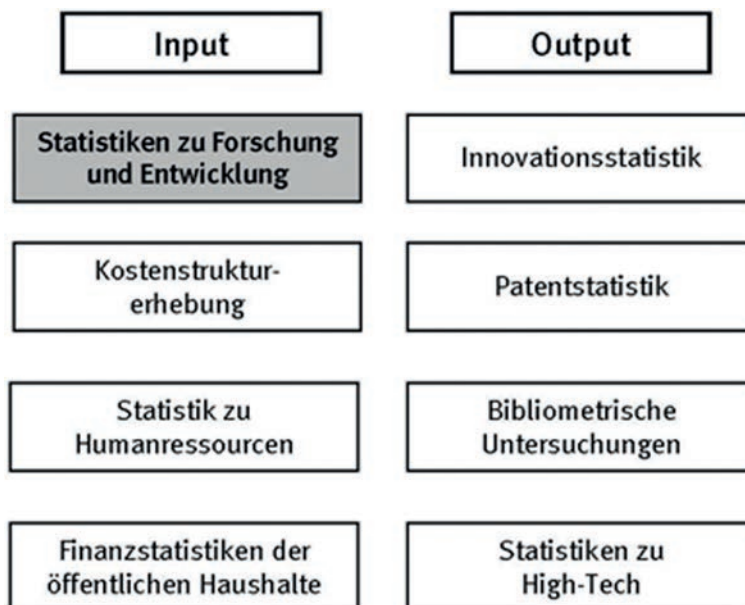
*Wissenschaftsrat*, 2011: Empfehlungen zur Bewertung und Steuerung von Forschungsleistung, Drs. 1656–11, Halle, November 2011. Online: [www.wissenschaftsrat.de/download/archiv/1656-11.pdf](http://www.wissenschaftsrat.de/download/archiv/1656-11.pdf)

## Die amtliche Forschungsstatistik

### 1 Die Forschungsstatistik im Kontext der Wissenschafts- und Technologiestatistiken

Wissenschaft und Technologie sind Schlüsselthemen moderner Gesellschaften und ihrer zukünftigen Entwicklung. Zur Berichterstattung über diese Bereiche existiert eine Reihe von Statistiken, darunter sowohl amtliche als auch nicht-amtliche Statistiken. Es werden auch Statistiken verwendet, die primär anderen Zwecken dienen, die aber forschungsrelevante Informationen enthalten. Gemeinsam geben die Statistiken einen umfassenden Überblick über den Bereich Wissenschaft und Technologie und dienen als Grundlage für wissenschaftliche und forschungspolitische Fragestellungen. Die Statistiken lassen sich zwei verschiedenen Gruppen zuordnen. Während die Statistiken zu Forschung und Entwicklung, die Kostenstrukturerhebung, die Statistik zu den Humanressourcen und die Statistik zu den Ausgaben der öffentlichen Haushalte den Input in den Forschungsprozess messen, haben die Innovationsstatistik, die Patentstatistik, die bibliometrischen Untersuchungen und die Statistik zu High-Tech die Messung des Outputs der Forschungsaktivitäten zum Gegenstand (*Abbildung 1*).

*Abbildung 1: Wissenschafts- und Technologiestatistiken*



Die Statistiken über Forschung und Entwicklung (FuE) sind zentraler Bestandteil der Berichterstattung zu Wissenschaft und Technologie. Sie werden auf der Basis einer international harmonisierten Methodik erhoben. Grundlage zur Sicherung international vergleichbarer Daten zu Ausgaben und Personal für Forschung und Entwicklung ist das sogenannte Frascati Handbuch der OECD zur Erhebung von Daten über Forschung und experimentelle Entwicklung. Forschung und Entwicklung wird darin definiert als systematische, schöpferische Arbeit zur Erweiterung des Kenntnisstands, einschließlich der Erkenntnisse über den Menschen, die Kultur und die Gesellschaft, sowie dessen Verwendung mit dem Ziel, neue Anwendungsmöglichkeiten zu finden.<sup>1</sup>

<sup>1</sup> Siehe OECD: Proposed Standard Practice for Surveys of Research and Experimental Development - Frascati Manual. Paris 2002, S. 30.

Nicht zu Forschung und Entwicklung zählen Gutachten, Beratungsleistungen, Versuchsarbeiten, die der Lehre dienen, wissenschaftliche Lehre, chemische Untersuchungen etc., wenn kein nennenswertes Element der Weiterentwicklung vorhanden ist. Die internationale Ausrichtung der Forschungsstatistik zeigt sich unter anderem auch daran, dass die Europäischen Staaten seit dem Jahr 2004 durch eine EU-Verordnung zur Bereitstellung von Daten zu Forschung und Entwicklung verpflichtet sind.

## **2     Sektoren in der Forschungsstatistik**

In der Forschungsstatistik wird differenziert zwischen der Durchführungs- und der Finanzierungsbetrachtung. Die Durchführungsbetrachtung beschreibt, in welchem Bereich Forschungsleistungen erbracht werden. Unterschieden werden dabei die Sektoren Unternehmen, Hochschulen und Staat und private Einrichtungen ohne Erwerbszweck. In der Forschungsstatistik in Deutschland werden die Informationen für alle Sektoren mit Hilfe von Erhebungen bei den forschenden Stellen gewonnen. In Anlehnung an das Konzept der Volkswirtschaftlichen Gesamtrechnungen werden in der FuE-Statistik die Sektoren „Wirtschaft“ und „Staat, private Organisationen ohne Erwerbszweck“ dargestellt. Die Hochschulen werden aufgrund ihrer großen Bedeutung im Bereich Forschung und Entwicklung als zusätzlicher dritter Sektor ausgewiesen.

Neben der Durchführungsbetrachtung gibt es auch eine Finanzierungsbetrachtung, in der dargestellt wird, welcher Bereich die Forschungsleistungen finanziert bzw. woher die benötigten Mittel stammen. Damit lassen sich Aussagen treffen, wie die finanziellen Lasten verteilt sind, beispielsweise zwischen öffentlichem und privatem Bereich. Innerhalb des öffentlichen Bereichs ist die Trennung zwischen Bund, Ländern und Gemeinden von Bedeutung. Die FuE-Finanzmittel können bei der durchführenden oder bei der finanzierenden Institution (z.B. in der Statistik der öffentlichen Haushalte) erfasst werden.

Im Folgenden werden die Erhebungen in den genannten Sektoren beschrieben, in denen Forschung und Entwicklung durchgeführt wird.

## **3     Erhebungen**

### **3.1   Unternehmen**

FuE-Daten im Wirtschaftssektor werden von der Stifterverband Wissenschaftsstatistik GmbH, einer Einrichtung des Stifterverbandes für die Deutsche Wissenschaft e. V., im Auftrag des Bundesministeriums für Bildung und Forschung im Rahmen einer freiwilligen Befragung erhoben. Erfasst werden Unternehmen und Einrichtungen ohne Erwerbszweck, die in erster Linie für Unternehmen arbeiten. Entsprechend dem Erhebungskonzept der Stifterverband Wissenschaftsstatistik GmbH wird alle zwei Jahre eine Totalerhebung bei den Unternehmen durchgeführt, die FuE betreiben. Die Daten für die Zwischenjahre werden anhand von Plandaten und einer Erhebung bei ausgewählten Unternehmen geschätzt. Die Statistik der Stifterverband Wissenschaftsstatistik GmbH ist die Datenbasis für die nationale und internationale Berichterstattung zu Forschung und Entwicklung in Unternehmen. Die Ergebnisse liegen in einer Gliederung nach Bundesländern und Regierungsbezirken vor.

### **3.2   Staat, private Einrichtungen ohne Erwerbszweck (außeruniversitäre Einrichtungen)**

Die Rechtsgrundlage für diese Erhebung mit Auskunftspflicht bildet das Finanz- und Personalstatistikgesetz. Im Rahmen dieser zentralen Erhebung des Statistischen Bundesamtes werden Bundes-, Landes- und andere öffentliche Forschungseinrichtungen, rechtlich selbständige Organi-



sationen ohne Erwerbszweck für Wissenschaft, Forschung und Entwicklung, sofern sie öffentliche Zuwendungen von mehr als 160 000 € erhalten, und rechtlich selbständige Institute an Hochschulen befragt. Jährlich werden das Personal sowie die Einnahmen und Ausgaben nach Arten und Wissenschaftszweigen, die Art der Einrichtung, ihre Aufgaben und der Anteil der FuE an der Gesamttätigkeit des wissenschaftlichen Personals erfasst. Alle vier Jahre werden darüber hinaus die Ausgaben nach sozioökonomischen Forschungszielen und Technologiebereichen, die Einnahmen nach Mittelgebern und die Art der Forschungstätigkeit erhoben.<sup>2</sup> Ergebnisse liegen in einer Gliederung nach Bundesländern und Regierungsbezirken vor.

### 3.3 Hochschulen

Die Basisdaten über die FuE-Aktivitäten der Hochschulen stammen aus der Hochschulfinanz- und der Hochschulpersonalstatistik, die auf der Grundlage des Hochschulstatistikgesetzes durchgeführt werden. Unabhängig von der Trägerschaft zählen zu den Hochschulen Universitäten, Gesamthochschulen, Pädagogische Hochschulen, Theologische Hochschulen, Kunsthochschulen sowie Verwaltungsfachhochschulen. Ebenfalls den Hochschulen zugeordnet werden die Hochschulkliniken, die Sonderforschungsbereiche sowie die Einrichtungen der Hochschulen, deren Ausgaben und Einnahmen in den Hochschulhaushalt einbezogen werden. In der Hochschulfinanzstatistik werden Einnahmen und Ausgaben der Hochschulen in fachlicher und organisatorischer Zuordnung und in haushaltsmäßiger Gliederung erhoben. Die Hochschulpersonalstatistik stellt das gesamte an Hochschulen haupt- und nebenberuflich tätige Personal nach Personalgruppen, nach Fächergruppen, nach Art der Finanzierung, nach Alter und nach Geschlecht dar.

Das Verfahren zur Berechnung von FuE-Ausgaben und FuE-Personal im Hochschulsektor basiert auf den internationalen Konventionen des Frascati Handbuchs. Es geht grundsätzlich von der Trennung von "Grundmittelforschung" und "Drittmittelforschung" aus. Während aus Drittmitteln finanzierte Ausgaben und Personal insgesamt der FuE zugeordnet werden, werden die FuE-Ausgaben und das FuE-Personal aus den Grundmitteln der Hochschulen mit Hilfe von FuE-Koeffizienten ermittelt. Dabei wird zunächst der Zeitaufwand für nicht FuE-Tätigkeiten bestimmt (Lehre, Overheadzeit, Korrekturzeit, Prüfungszeit, Erzielung von Verwaltungseinnahmen). Die FuE-Koeffizienten werden dann durch ein detailliertes Verfahren ermittelt, indem indirekt der Zeitaufwand des wissenschaftlichen und künstlerischen Personals für Forschung und Entwicklung berechnet wird.<sup>3</sup> Die Koeffizienten werden nach Hochschularten und Fächergruppen differenziert berechnet. Das Statistische Bundesamt veröffentlicht die Ergebnisse auf Bundes-, Landes- und Regierungsbezirksebene.

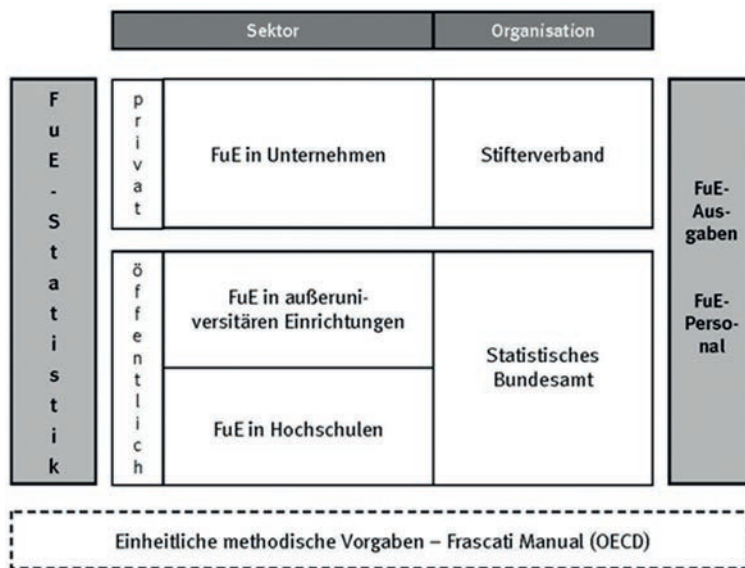
Alle drei Erhebungen entsprechen den methodischen Empfehlungen des Frascati Handbuchs der OECD. Die Sektoren und damit auch die Berichtskreise der einzelnen Erhebungen sind methodisch aufeinander abgestimmt und definitorisch so abgegrenzt, dass die Daten zu einem Gesamtergebnis für die FuE-Aktivitäten der Bundesrepublik Deutschland zusammengefasst werden können. Sie liefern damit die Datengrundlage für die nationale und internationale Berichterstattung zu Forschung und Entwicklung (*Abbildung 2*).

---

2 Siehe Brugger, P.: Einnahmen und Ausgaben der öffentlichen und öffentlich geförderten Einrichtungen für Wissenschaft, Forschung und Entwicklung. In *Wirtschaft und Statistik* 10/1996.

3 Siehe Hetmeier, H.-W.: Methodik der Berechnung der Ausgaben und des Personals der Hochschulen für Forschung und experimentelle Entwicklung ab dem Berichtsjahr 1995. In *Wirtschaft und Statistik* 2/1998, aktualisiert in Fachserie 11, Reihe 4.3.2: Monetäre hochschulstatistische Kennzahlen. Statistisches Bundesamt, Wiesbaden, verschiedene Jahrgänge.

Abbildung 2: Organisation der Forschungsstatistik in Deutschland



### 3.4 Nationale und Europäische Zielsetzungen

Der Bereich Wissenschaft und Technologie ist wegen seiner Bedeutung für die Zukunftsfähigkeit einer rohstoffarmen Volkswirtschaft Gegenstand verschiedener politischer Programme und Zielsetzungen. Im Bereich Forschung und Entwicklung ist hier in erster Linie das sogenannte 3 Prozent-Ziel zu nennen.

Der Europäische Rat der Europäischen Union hatte das strategische Ziel gesetzt, zum wettbewerbsfähigsten und dynamischsten wissensbasierten Wirtschaftsraum der Welt zu werden. Der Übergang zu einer wissensbasierten Wirtschaft und Gesellschaft sollte unter anderem durch Intensivierung von Forschung und Entwicklung gefördert werden. Im Rahmen der sogenannten Lissabon-Strategie wurde als europaweites Ziel für das Jahr 2010 ein Anteil der FuE-Ausgaben von 3 Prozent am Bruttoinlandsprodukt formuliert. Dieses Ziel wurde auch in die Wachstumsstrategie für die Europäische Union „Europa 2020“ für das Jahr 2020 übernommen. Die Regierungschefs von Bund und Ländern haben die Zielmarke auch auf nationaler Ebene in das sogenannte 10 Prozent-Ziel (10 Prozent des Bruttoinlandsproduktes für Bildung und Forschung) für das Jahr 2015 aufgenommen.

## 4 Ergebnisse

Die amtliche Statistik ist primär darauf ausgerichtet, Daten zu Aggregaten zusammenzufassen und verdichtete Informationen bereitzustellen. Dazu werden häufig aus den erhobenen Basisdaten Indikatoren zum Vergleich von Staaten, Bundesländern, Regionen oder Sektoren berechnet. Die Bereitstellung von Ergebnissen für einzelne Einrichtungen, Hochschulen oder Unternehmen, etwa zum Zwecke interinstitutioneller Vergleiche, gehört in der Regel nicht zum Aufgabenprogramm der amtlichen Statistik.

Mit den Daten der Forschungsstatistik werden verschiedene Indikatoren berechnet. Der prominenteste Indikator ist die FuE-Intensität. Der Indikator beschreibt den Anteil der FuE-Ausgaben am Bruttoinlandsprodukt. Da die Teilgrößen zur Berechnung des Indikators nach einer einheitlichen Methodik berechnet werden, ermöglicht der Indikator sowohl nationale als auch internationale Vergleiche bei der Messung der Fortschritte auf dem Weg zum 3 Prozent-Ziel.



Abbildung 3: Anteil der FuE-Ausgaben am Bruttoinlandsprodukt 2005 – 2010

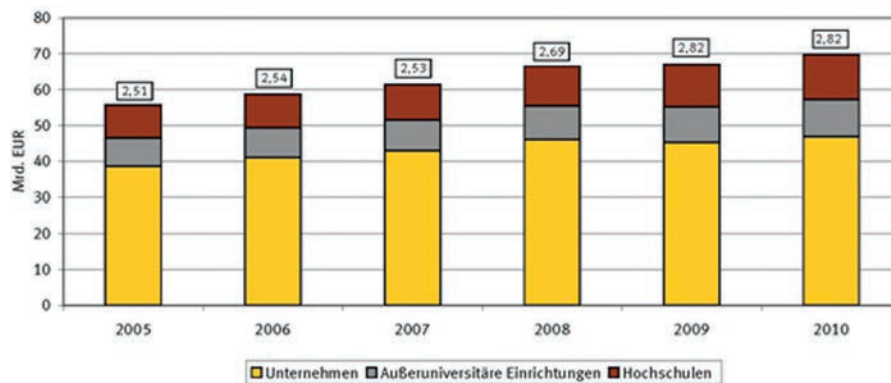
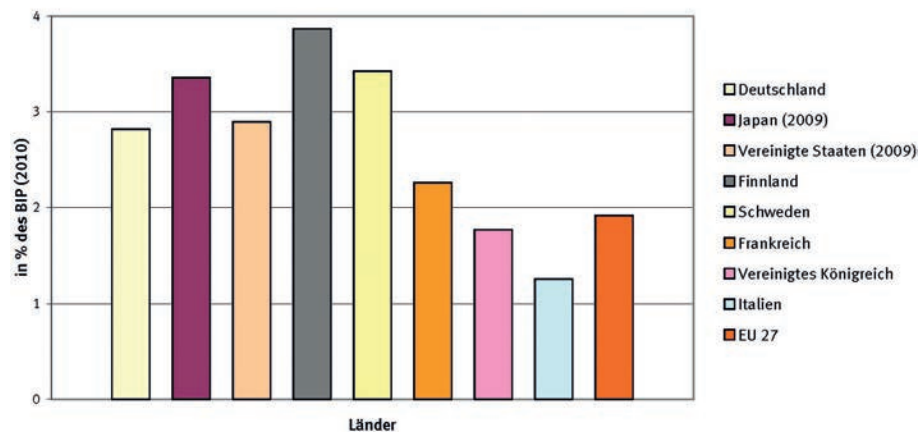


Abbildung 3 zeigt, dass der Anteil der FuE-Ausgaben am Bruttoinlandsprodukt in Deutschland seit 2005 kontinuierlich gestiegen ist und nach vorläufigen Berechnungen im Jahr 2010 den Wert von 2,82 Prozent erreicht. Dabei ist auffällig, dass im Jahr 2009, in dem die Wirtschafts- und Finanzkrise die stärkste negative Auswirkung auf das Bruttoinlandsprodukt hatte, die FuE-Intensität deutlich gestiegen ist. Bei getrennter Betrachtung beider Teilgrößen wird deutlich, wie dieser Effekt zustande kommt. Zwar hat die Wirtschafts- und Finanzkrise auch die Entwicklung der FuE-Ausgaben gebremst. Die negative Auswirkung auf die Wirtschaftsleistung insgesamt, ausgedrückt im Bruttoinlandsprodukt, war aber deutlich stärker. Im Ergebnis ist das Verhältnis von FuE-Ausgaben zur Wirtschaftsleistung, die FuE-Intensität, gestiegen. Im Jahr 2010 hat sich die Wirtschaftsleistung wieder positiv entwickelt und auch die FuE-Ausgaben sind im gleichen Maße gestiegen. Der Anteil der FuE-Ausgaben am Bruttoinlandsprodukt bleibt folglich konstant bei 2,82 Prozent.

Abbildung 4: Anteil der FuE-Ausgaben am Bruttoinlandsprodukt 2010



Im internationalen Vergleich (Abbildung 4) wird deutlich, dass Deutschland im Jahr 2010 die größte FuE-Intensität unter den großen Industriestaaten in Europa aufweist, deutlich höher als Frankreich (2,26 Prozent), dem Vereinigten Königreich (1,77 Prozent) und Italien (1,26 Prozent). Deutlich wird auch, dass die skandinavischen Länder Finnland (3,87 Prozent) und Schweden (3,43 Prozent), mit großem Abstand die Spitzenposition einnehmen. In der Europäischen Union insgesamt (EU 27) liegt die FuE-Intensität bei 1,91 Prozent. Die Vereinigten Staaten (2009: 2,90 Prozent) und Japan (2009: 3,36 Prozent) weisen ebenfalls hohe Forschungsintensitäten auf.

## **5 Ausblick**

Nicht zuletzt durch die Formulierung von Zielwerten, wie dem 3 Prozent Ziel, ist die Forschungsstatistik zentrale Statistik im Rahmen der Berichterstattung zu Wissenschaft und Technologie. Durch die Übernahme des Zielwertes in die Strategie „Europa 2020“ der Europäischen Union wird auch auf internationaler Ebene die künftige Bedeutung der Ergebnisse sehr hoch sein. Grundlage für die Qualität und Vergleichbarkeit der Ergebnisse ist die seit vielen Jahren etablierte harmonisierte Methodik, die im Frascati Handbuch der OECD dokumentiert ist. Hier wird es entscheidend sein, bei der Weiterentwicklung der Methodik die richtige Abstimmung zu finden zwischen der einerseits notwendigen Anpassung an eine sich wandelnde Realität und der andererseits erforderlichen Wahrung der Kontinuität und Vergleichbarkeit der Ergebnisse im Zeitverlauf.

## **Sicht der Universitäten & Einrichtungen**



## **Forschungsinformationssysteme bei Fraunhofer**

Die Fraunhofer-Gesellschaft ist die größte Organisation für angewandte Forschung in Europa. In mehr als 80 Forschungseinrichtungen in Deutschland, davon 60 Fraunhofer-Institute, arbeiten über 18 000 Mitarbeiterinnen und Mitarbeiter, überwiegend mit natur- oder ingenieurwissenschaftlicher Ausbildung. Das Forschungsvolumen beträgt 1,66 Milliarden Euro jährlich, davon 1,40 Milliarden Euro im Leistungsbereich Vertragsforschung. Über 70 Prozent dieses Bereichs werden mit Aufträgen aus der Industrie und mit öffentlich finanzierten Forschungsprojekten erwirtschaftet. Knapp 30 Prozent werden von Bund und Ländern als Grundfinanzierung beigesteuert. Dabei kommt das so genannte Fraunhofer-Modell zum Tragen: Seit 1973 bemisst sich die Grundfinanzierung weitgehend erfolgsabhängig nach der Höhe der Wirtschaftserträge.

Die Forschungsfelder der Fraunhofer-Institute richten sich nach den Bedürfnissen der Menschen: Gesundheit, Sicherheit, Kommunikation, Mobilität, Energie und Umwelt. Die Entwicklung von anwendungsreifen Lösungen steht dabei stets im Vordergrund. Vertragspartner und Auftraggeber sind Industrie- und Dienstleistungsunternehmen sowie die öffentliche Hand.

### **1 Bedarf an zentralen und dezentralen Informationsmanagementsystemen**

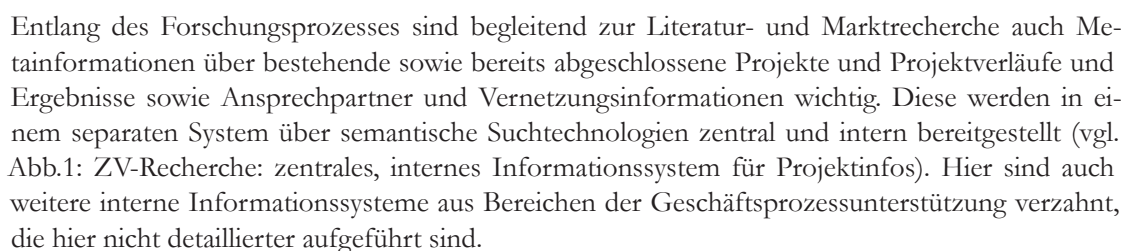
Die Fraunhofer-Institute agieren in ihrer Forschungsprojektplanung und -durchführung weitgehend selbständig. Zwar gibt es vielseitige Vernetzungen sowie Allianz- und Verbundbildungen auf inhaltlicher Ebene, doch keinen übergeordneten Bedarf nach einer zentralen Bereitstellung der Forschungsinformationen aller Institute. Da in der Zusammenarbeit mit Industriepartnern häufig Vertraulichkeitsaspekte zu berücksichtigen sind, bleibt die Speicherung von Projekt- oder Forschungsdaten oder das Publikationsmanagement mit internen Freigabeverfahren Institutschoheit. So sind die Forschungsinformationen bei Fraunhofer heterogen erfasst und es existieren verteilte Informationskreisläufe. Im Folgenden werden daher nur die institutsübergreifenden Aspekte eines zentralen Informationsmanagements dargestellt.

Als beste Näherung erscheint die Darstellung der vorhandenen bzw. geplanten Systeme, die einen idealisierten Forschungsprozess flankieren:

Den Input für ein neues Forschungsprojekt liefern entweder Aufträge oder eigene Ideen und Strategien. Jedes Forschungsprojekt beginnt aber zunächst mit einer eingehenden Recherche: Der schnelle und vollständige Zugang zu allen Quellen der Fachinformation ist dabei für Fraunhofer-Wissenschaftler ein wichtiger Faktor für die Priorität bei Forschungsprojekten. Unter Fachinformationsmanagement versteht man bei Fraunhofer die umfassende Versorgung mit allen relevanten Informationen, so findet neben der klassischen Literaturrecherche immer auch eine umfangreiche Markt- und Patentrecherche statt. Hierzu wird derzeit mit dem Projekt eLib („enhanced Library Services“) ein neues Informationssystem und eine zukunftsfähige Infrastruktur implementiert (vgl. Abb.1: eLib: zentrales, internes Informationssystem für Fachinformationsdiscovery).

Mit eLib soll die Suche nach lokalen Fachinformationsquellen und zentral lizenzierten Zeitschriften und Datenbankzugängen künftig einfacher und zeitgemäßer werden. Für die Wissenschaftler als „Endkunden“ wird ein zentraler Suchschlitz zum einfachen und übersichtlichen Einstieg in alle Suchen nach Literatur und Fachinformation. Erreicht wird dieses durch eine verbesserte „Discovery Technologie“ – einem zentralen Index mit direkter Verlinkung zu den vorhandenen Volltexten. Unterstützt wird dies durch den Einsatz einer modernen Bibliothekssoftware, die auch die lokalen Literaturbestände und Bibliotheksdienste direkt in das Suchportal einbindet. Das eLib-Portal soll ab Mitte 2012 bereitstehen und als Suchschnittstelle flexibel in die unterschiedlichen Instituts-

Abbildung 1: Idealisierter Forschungsprozess und die ihn flankierenden Informationssysteme bei Fraunhofer



Für die Außenwelt sichtbar ist die Publikationsdatenbank Fraunhofer-Publica [<http://publica.fraunhofer.de>], eine seit 1980 betriebene zentrale bibliographische Nachweisdatenbank aller Publikationen und Patente, die aus der Forschungstätigkeit der Fraunhofer-Institute resultieren (vgl. Abb.1: Publica, ePrints: zentrale, externe Nachweisdatenbank aller Fraunhofer Veröffentlichungen und Patente). Die Fraunhofer-Publica ist nicht nur Informationssystem für die Außenwelt. Als zentrale Sammelstelle bietet die Publica den Instituten bequeme Aufbereitungsformen für ihre Publikations- oder Autorenlisten. Außerdem werden durch die einheitliche Erfassung und den Abgleich mit den Zitationsdatenbanken Web of Science und Scopus bibliometrische Publikationsanalysen möglich.

38

Als letzter Teil des Forschungsprozesses besteht als bisher noch nicht realisiertes Teil-Informationssystem die Beschreibung und Veröffentlichung von Forschungsdaten (vgl. Abb.1: Forschungsdaten: zentrale, extern verfügbare Metadaten über Forschungsdaten aus Fraunhofer-Projekten; noch nicht realisiert). Die Anforderung, neben der textuellen Ergebnisdokumentation in Form von Publikationen auch die strukturierte Veröffentlichung von Metadaten zu den verwendeten Forschungsdaten aus öffentlich geförderten Projekten (Daten über Daten) zu dokumentieren, ergibt sich insbesondere aus den von den Forschungsförderern EU und DFG formulierten Strategien für die kommenden Forschungsprogramme. So sieht u. a. das 8. Europäische Forschungsrahmenprogramm (FP8) bereits dezidierte Datenmanagementpläne und Konzepte zur Langzeitsicherung von Forschungsdaten vor.

## **2 Besteht Bedarf für ein „Forschungsinformations-Supersystem?“**

Die im Rahmen des Symposiums Forschungsinformationssysteme vorgestellten Beispiele für Anforderungskataloge an Forschungsinformationssysteme zeigten deutlich, dass es zurzeit noch sehr schwierig ist, die unterschiedlichen Informationsdimensionen unter einen Hut zu bringen. Klar zu unterscheiden waren zwei unterschiedliche Blickwinkel: Die Sicht der Organisation, die der Informationsflut Herr werden und Informationen bündeln, zentralisieren und aufbereiten muss und zu diesem Zweck gebündelte Informationssysteme entwirft. Hier sind Steuerungsbedarf und Berichtspflichten das strukturgebende Element bei der Ausgestaltung. Die andere Seite vertreten die Wissenschaftler, die als „Enddatenproduzenten- und -nutzer“ die eigentliche Forschung betreiben und verschiedenste Formen und Informationssysteme als Input für ihre Forschungstätigkeit brauchen.

Die hier beschriebene Informationslandschaft von Fraunhofer nimmt eher die Sicht der Wissenschaftler an, als die der Organisation mit ihrem Bedarf nach Informationssystemen für interne Steuerung. Die Fraunhofer-Wissenschaftler benötigen einen schnellen, intuitiven Zugang zu Informationen (Literatur, Daten, Marktrecherche); am besten eine Gesamtsicht über unterschiedliche Systeme hinweg (zentraler Sucheinstieg). Sie fordern eine Entlastung vom Such- und Verwaltungsaufwand und auch einen „direkten Draht“ zu den Infrastrukturbetreibern und Hilfe vor Ort. Wünschenswert wäre, wenn sich der Daten- und Informationsfluss jederzeit an den Stand eines Projekts anpasste. Zugleich werden durch Publikationsunterstützung und Präsentation der Forschungs- und Publikationsaktivitäten Anreize für die persönliche Karriere geschaffen.

Aus Fraunhofer-Organisationssicht sind anzustrebende Eckpfeiler in der Ausgestaltung der Informationslandschaft sicherlich immer in erster Linie die Datensicherheit und eine Architektur, die Schutz vor unerlaubtem Zugriff gewährleistet. Trotzdem muss die Bereitstellung der Informationen möglichst einfach gestaltet und durch Attribute wie Einmalserfassung, Aggregationsfähigkeit und Nachnutzung charakterisiert sein. Die Ausgestaltung solcher Informationsinput- und -Outputsysteme zu reifen Instrumenten der Wissenschaftskommunikation erfordert ein vielschichtiges Ineinandergreifen nicht nur der technischen Schnittstellen.

Das Fraunhofer Organisationsmodell und die Heterogenität der Informationsstrukturen bei Fraunhofer lassen keinen Bedarf für ein „All-in-One“-Forschungsinformationssystem erkennen oder auch nur im Rahmen des Möglichen erscheinen. Konkreten Entwicklungsbedarf gibt es bei der Herausforderung, Forschungsdaten qualifiziert zu beschreiben und nachnutzbar zu machen. Hierfür ist zunächst eine geeignete, bestenfalls bereits standardisierte Infrastruktur für Datenpublikationen und Erfassungsmethodik erforderlich. Dies könnte praktisch durch eine entsprechende Erweiterung der Fraunhofer-Publica erreicht werden, doch sind entsprechende Standards noch nicht umfassend realisiert. 2012 wird der Ausbau der Fraunhofer-Publica zu einem Datenveröffentlichungsrepository in einem internen Evaluierungsprojekt untersucht.

Fraunhofer betreibt gemeinsam mit der Alexander von Humboldt-Stiftung, der Deutschen Forschungsgemeinschaft (DFG), dem Deutschen Akademischen Austauschdienst (DAAD), der Helmholtz-Gemeinschaft, der Hochschulrektorenkonferenz (HRK), der Leibniz-Gemeinschaft,

der Max-Planck-Gesellschaft und dem Wissenschaftsrat die Schwerpunktinitiative „Digitale Information“ zur Verbesserung der Informationsversorgung in Forschung und Lehre. Im Fokus der Aktivitäten der sechs Arbeitsgruppen steht das Ziel, Wissenschaftler mit der bestmöglichen Informationsinfrastruktur auszustatten, die sie für ihre Forschung brauchen.

Die Arbeitsgruppe hat eine gemeinsame Forschungsdaten-Policy entwickelt und veröffentlicht (<http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsätze>), um bei Wissenschaftlern das Bewusstsein für den Handlungsbedarf und für den Nutzen von Forschungsdaten-Infrastrukturen zu schärfen. Außerdem zielen die Aktivitäten darauf ab, die Entstehung geeigneter Strukturen zu stimulieren, das heißt, die Zusammenarbeit zwischen Fachwissenschaftlern und Informationswissenschaftlern anzustoßen und zu begleiten sowie Förderangebote für Modellprojekte anzubieten.

Deshalb muss flankierend eine intensive Beratung stattfinden, die der erforderlichen Kultur mit einer positiven Einstellung zu „Data Sharing“ den Weg bereitet. Im Hinblick auf die oft einzuhaltende Vertraulichkeit und Sicherheit von Daten in der Vertragsforschung bedeutet dies bei Fraunhofer sicherlich eine anspruchsvolle Aufgabe.

Hier kann man aus den positiven Erfahrungen mit der Open Access Förderung und dem entsprechenden Beratungsaktivitäten begleitend zum Open Access Infrastrukturausbau schöpfen. Dies bedeutet nicht zuletzt, die Vernetzung der zentralen und dezentralen Dienstleister zu verbessern und Anreize für eine Vereinheitlichung von Prozessen zu bieten. Bei Fraunhofer findet dies stets unter dem Motto „Best Practice statt zentraler Vorgabe“ statt.



## **Einführung eines Forschungsinformationssystems an der Westfälischen Wilhelms-Universität Münster – Von der Konzeption bis zur Implementierung**

### **1    Forschungsinformationssysteme als Instrument für ein integriertes Management von Forschungsinformationen**

„Science is a global enterprise“. Die weltweite Wissenschaftslandschaft wird zunehmend durch Kollaborationen geprägt (Royal Society 2011). Um der Komplexität und Vielschichtigkeit der im Rahmen von Forschungsvorhaben betrachteten Fragestellungen begegnen zu können, ist eine kollaborative Herangehensweise oftmals unabdingbar. Wissenschaftler agieren daher mehr und mehr in selbstorganisierten oder institutionalisierten Forschungsverbünden und -netzwerken, welche nicht selten international und vornehmlich interdisziplinär aufgestellt sind (Carayol, Matt 2004).

In Folge dieser Trendentwicklung wächst auch die Menge und Vielfalt der Informationen zu Forschungsaktivitäten und -ergebnissen, die im Rahmen von (kollaborativen) Forschungsverbünden generiert, ausgetauscht, dokumentiert und präsentiert werden, fortwährend. Innerhalb von Forschungskonsortien entstehen daher vielfältige Sammlungen an Forschungsinformationen, die jedoch nicht selten nur auf den Bereich eines einzelnen Vorhabens oder Verbunds beschränkt bleiben und dort isoliert voneinander in projekt- oder institutionsspezifischen Dokumenten- und Datenbeständen oder Webportalen gespeichert und verwaltet werden.

Universitäten und Forschungseinrichtungen sehen sich daher nicht zuletzt im Hinblick auf hochschulweite Aufgaben wie die strategische Planung und Steuerung sowie eine kontinuierliche Forschungsberichtserstattung und Wissenschaftskommunikation zunehmend vor der Herausforderung einer erschwerten Gewinnung, Konsolidierung und Kommunikation der hierfür benötigten Informationen über die in den verschiedenen Disziplinen und Einrichtungen unternommenen Forschungsaktivitäten und erlangten Ergebnisse (Krücken, Meier 2009; Bovelet 2006; Schimank 2005; Fox 1992). Um diese Aufgaben informationstechnisch effizient und effektiv zu unterstützen, werden an deutschen Universitäten und Forschungseinrichtungen vermehrt Forschungsinformationssysteme eingeführt, welche gezielt eine integrierte Verwaltung, Nutzung und Kommunikation von Forschungsinformationen über die Grenzen einzelner Disziplinen und Forschungsverbünde hinweg unterstützen (Jörg et al. 2012; euroCRIS 2012).

Nicht zuletzt vor diesem Hintergrund hat die Leitung der Westfälischen Wilhelms-Universität (WWU) Münster im Jahr 2008 den Beschluss zur Einführung eines universitätsweiten Forschungsinformationssystems gefasst, um die vielfältigen dezentralen Datenbestände im Sinne eines integrierten Forschungsinformationsmanagements zusammenzuführen und auf diese Weise neue Nutzenpotenziale zu erschließen. Rückblickend auf das 2011 abgeschlossene Einführungsprojekt diskutiert dieser Beitrag den Projektverlauf von der Konzeption bis hin zur Implementierung und Einführung eines Forschungsinformationssystems am Beispiel der WWU Münster. Hierzu wird zunächst die Ausgangssituation und Zielsetzung eines Forschungsinformationssystems an der WWU Münster charakterisiert, um hierdurch entsprechende Anforderungen und Rahmenbedingungen für das Einführungsprojekt herauszustellen (Abschnitt 2). Auf Basis der identifizierten Anforderungen und Anwendungsfälle wird in einem nächsten Schritt eine fachliche Spezifikation eines Forschungsinformationssystems vorgeschlagen (Abschnitt 3) sowie die Einbindung des Forschungsinformationssystems in die bestehende Anwendungslandschaft und Organisationsstruktur der WWU Münster thematisiert (Abschnitt 4). Im Sinne des „lessons learned“ schließt der Beitrag mit einer Zusammenfassung der grundlegenden Erfahrungen, die über die verschiedenen Phasen des Einführungsprojektes hinweg gesammelt wurden (Abschnitt 5).

## **2 Ausgangssituation und Zielsetzung**

### **2.1 Charakteristika der WWU Münster**

Mit aktuell ca. 37.000 Studierenden in mehr als 110 Studienfächern und 250 Studiengängen zählt die WWU Münster zu den fünf größten Universitäten Deutschlands. Mit 15 Fachbereichen (inklusive Medizin), über 500 Professoren, ca. 3800 wissenschaftlichen Beschäftigten und einem breiten Spektrum an Wissenschaftsdisziplinen (außer Ingenieurwissenschaften) weist die WWU Münster die klassischen Merkmale einer Volluniversität auf (Universität Münster 2010; Statistisches Bundesamt 2010). Das Forschungsprofil der WWU Münster wird sowohl durch disziplinenbezogene als auch interdisziplinär ausgerichtete Forschungsaktivitäten geprägt, die sich über die Bereiche der Geistes- und Sozialwissenschaften, Wirtschafts- und Rechtswissenschaften, Lebenswissenschaften, Naturwissenschaften sowie der Mathematik erstrecken und diese miteinander verknüpfen. Ein wesentliches Charakteristikum der Forschungslandschaft der WWU ist eine intensiv ausgeprägte „Zentren-Kultur“, die durch entsprechende Rahmenbedingungen wie eine flexible Bildung bewusst gefördert wird und durch eine Vielzahl interdisziplinärer Forschungsverbünde innerhalb der WWU Münster zum Ausdruck kommt. Die Größe und vor allem die Disziplinenvielfalt der WWU werden hier als Chance genutzt, um auf diesem Weg auch ungewöhnliche Kooperationen zwischen verschiedenen Disziplinen etablieren zu können. Die WWU Münster wird somit maßgeblich durch die verfolgte Breitenforschung und deren interdisziplinäre Ausrichtung charakterisiert.

Wie vielen Universitäten und Forschungseinrichtungen ist es auch der WWU Münster mehr und mehr ein Anliegen, die unternommenen Forschungsaktivitäten und daraus hervorgegangenen Ergebnisse den verschiedenen Interessengruppen zu kommunizieren und hierdurch Verbreitung, Akzeptanz und Anwendung zu erzielen (Cheng et al. 2008). In Sinne der Wissenschaftskommunikation und Forschungsberichterstattung werden an der WWU Münster kontinuierlich Informationen für verschiedene Gruppen von Rezipienten über vielfältige Kanäle und in unterschiedlichen Formen zur Verfügung gestellt. Von zentraler Ebene der Universitätsleitung und -verwaltung wird bspw. ein Statistisches Jahrbuch veröffentlicht, welches die Forschungsaktivitäten der WWU Münster vorrangig aus einer quantitativen Perspektive beleuchtet. Eine vielmehr inhaltliche und qualitative Kommunikation von Forschungsaktivitäten und -ergebnissen erfolgt zumeist auf dezentraler Ebene durch die einzelnen Forschungseinrichtungen der WWU Münster in Form von einrichtungs- und projektbezogenen Webauftritten, Jahresberichten und anderen Print- bzw. Online-Medien.

Durch einen derartig dezentral ausgerichteten Ansatz der Forschungsberichterstattung gefördert, liegen Informationen zu den Forschungsaktivitäten und -ergebnissen der WWU mehrheitlich in unterschiedlichen dezentralen und zumeist isoliert voneinander existenten Datenbeständen vor. Verbunden mit dem breiten Spektrum an Forschungsdisziplinen ergibt sich für eine universitätsweit einheitliche und Disziplinen überspannende Wissenschaftskommunikation und Forschungsberichterstattung hieraus die Herausforderung, dass Forschungsinformationen in unterschiedlichen Darstellungsformen und Umfängen bereitgestellt sowie über verschiedene Medien zugänglich gemacht werden. Infolge einer derartigen Heterogenität wird ein Disziplinen übergreifender Informationszugang zu den einzelnen Forschungsaktivitäten und -ergebnissen entscheidend erschwert und nicht zuletzt hierdurch die Wahrnehmung des Forschungsprofils einer solch großen Universität wie der WWU Münster als Gesamteinstitution nachhaltig gehemmt.

### **2.2 Zielsetzung und Anforderungen an ein Forschungsinformationssystem**

Um sich im nationalen wie internationalen Wettbewerb um wissenschaftliche Reputation, Talente und Ressourcen entsprechend positionieren zu können, sehen sich Universitäten und Forschungseinrichtungen zunehmen vor der Herausforderung, die Forschungsstärken bzw. Profil bildenden Schwerpunktbereiche mit den darin unternommenen Forschungsaktivitäten möglichst weitrei-

chend und zielgerichtet zu kommunizieren und hierdurch nachhaltig sichtbar zu machen (Cheng et al. 2008). Vor diesem Hintergrund verfolgt die WWU Münster mit der Einführung eines universitätsweiten Forschungsinformationssystems ein integratives Konzept zur Verwaltung und Bereitstellung der vielfältigen Forschungsinformationen, um auf diese Weise der mit der dezentralen Forschungsberichterstattung einhergehenden Heterogenität zu begegnen. Infolgedessen sind mit der Einführung eines zentralen Forschungsinformationssystems an der WWU Münster die folgenden Zielsetzungen verbunden:

- Universitätsweite Harmonisierung und Zusammenführung der unterschiedlichen dezentralen Bestände an Forschungsinformationen in einem zentralen Forschungsinformationssystem
- Etablierung des Forschungsinformationssystems als zentrales Instrument zur einheitlichen und kontinuierlichen Kommunikation und Dokumentation der vielfältigen Forschungsaktivitäten und -ergebnisse
- Stärkung der Wahrnehmung und Sichtbarkeit des gesamtuniversitären Forschungsprofils und der interdisziplinären Forschungsaktivitäten

Ausgehend von diesen Zielstellungen ergeben sich vielfältige Anforderungen, die sich insbesondere auf die inhaltliche Ausgestaltung eines entsprechenden Systems, die daraus resultierenden Anwendungsmöglichkeiten sowie den Prozess der Einführung beziehen.

Um einer breiten und interdisziplinär ausgerichteten Forschungslandschaft wie die der WWU Münster gerecht zu werden, gilt es bei der Gestaltung eines Forschungsinformationssystems die vielfältigen Forschungskulturen mit den disziplinspezifischen Charakteristika und Dimensionen von Forschungsinformationen zu berücksichtigen und abzubilden. Ein Aspekt, in dem die Vielfältigkeit an disziplinenbezogenen Charakteristika deutlich zum Ausdruck kommt, ist die Publikationskultur (Alexander von Humboldt-Stiftung 2008). Publikationen werden gemeinhin als ein bedeutendes Medium genutzt, um erlangte Erkenntnisse zu verbreiten und – damit verbunden – Reputation und Status innerhalb der jeweiligen wissenschaftlichen Community zu erlangen (Albert 2004; Leonard 2002; Walstad 2002). Darüber hinaus werden Publikationen auch als ein bedeutender Indikator zur Beurteilung der Qualität der Forschung herangezogen (Münch 2006). Werden in einigen Fachdisziplinen Monographien und Beiträge in Herausgeber- oder Sammelwerken als anerkannte Publikationsformate angesehen, gibt es dem entgegen Disziplinen, wo diese Formate weniger üblich sind. Stattdessen wird in diesen Bereichen Publikationen, die auf wissenschaftlichen Konferenzen oder in wissenschaftlichen Zeitschriften (Journals) veröffentlicht werden, ein höheres Gewicht beigemessen.

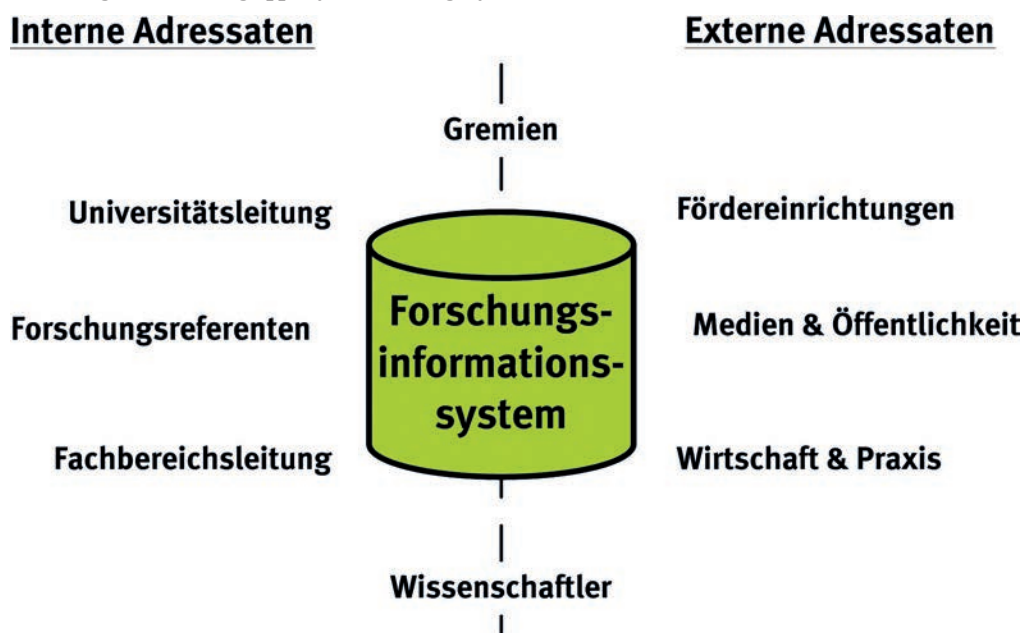
Neben Publikationen zeichnet sich – nicht zuletzt im Rahmen der Entscheidung über die Bemessung und Verteilung von finanziellen Ressourcen – auch in Bezug auf Forschungsprojekte eine disziplinspezifische Betrachtung ab (Berghoff et al. 2009; Hornbostel 2001). Während in einigen Disziplinen ausschließlich die Fördereinrichtung wie bspw. die Deutsche Forschungsgemeinschaft (DFG), das Bundesministerium für Bildung und Forschung (BMBF) oder die Europäischen Kommission (EU) sowie die Höhe der finanziellen Unterstützung als Kriterien zur Beurteilung von Forschungsprojekten herangezogen werden, erfordern andere Disziplinen auch die Berücksichtigung von solchen Projekten, die allein aus dem Grundetat einer Einrichtung finanziert werden und ebenfalls von hoher wissenschaftlicher Qualität sein können.

Im Hinblick auf die Vielfalt der Forschung, die an einer Volluniversität wie der WWU Münster betrieben wird, erscheint eine alleinige Berücksichtigung von Publikationen und/oder Forschungsprojekten in einem Forschungsinformationssystem als nicht ausreichend. Um dem Anspruch der verschiedenen Disziplinen gerecht zu werden, ist es vielmehr notwendig, einen *multidimensionalen Ansatz* bei der Betrachtung von Forschungsinformationen zu verfolgen, in den neben Publika-

onen und Forschungsprojekten unter anderem auch erhaltene Auszeichnungen und Ehrungen, die Förderung des wissenschaftlichen Nachwuchses im Sinne von Promotionen und Habilitationen, erteilte Patente, sowie die Einbindung von Wissenschaftlern in wissenschaftliche Netzwerke und Gremien Berücksichtigung finden.

Im Sinne einer *multiperspektivischen Betrachtung* von Forschungsinformationen ist weiterhin zu beachten, dass im Umfeld der universitären Forschung verschiedene Interessensgruppen existieren, die als Rezipienten in Betracht kommen und die es infolgedessen mit den für sie relevanten Informationen zu bedienen gilt (vgl. Abbildung 1). Hierbei handelt es sich aus Sicht einer Universität sowohl um interne als auch um externe Adressatenkreise (vgl. auch im Folgenden Jörg et al. 2012; euro-CRIS 2012; Landry et al. 2001; Closs, Cheater 1994; Beyer, Trice 1982). Innerhalb einer Universität greifen zu Evaluations- und Steuerungszwecken beispielsweise die Universitäts- und Fachbereichsleitung auf Forschungsinformationen zurück, wohingegen Forschungsreferate und einzelne Wissenschaftler diese Informationen zum Wissenstransfer, zur Kommunikation sowie zur Initiierung und Förderung zukünftiger (interdisziplinärer) Forschungsvorhaben nutzen. Aus der Perspektive externer Adressaten sind Informationen zu Forschungsaktivitäten und daraus hervorgegangenen Forschungsergebnissen einerseits im Kontext des Managements von Forschungsprojekten sowie zur Unterstützung förderpolitischer Entscheidungen für Fördereinrichtungen und -gremien wie bspw. die DFG, das BMBF oder die EU von Interesse. Weiterhin bieten Forschungsinformationen im Kontext des Forschungs- und Technologietransfers sowohl der unternehmerischen Praxis als auch anderen Wissenschaftlern eine geeignete informatorische Basis. Darüber hinaus leistet die mediale Kommunikation von Forschungsinformationen einen nicht unbedeutenden Beitrag zur Verbreitung, Akzeptanz und Anwendung von Forschungsergebnissen in der breiten Öffentlichkeit. (Cheng et al. 2008)

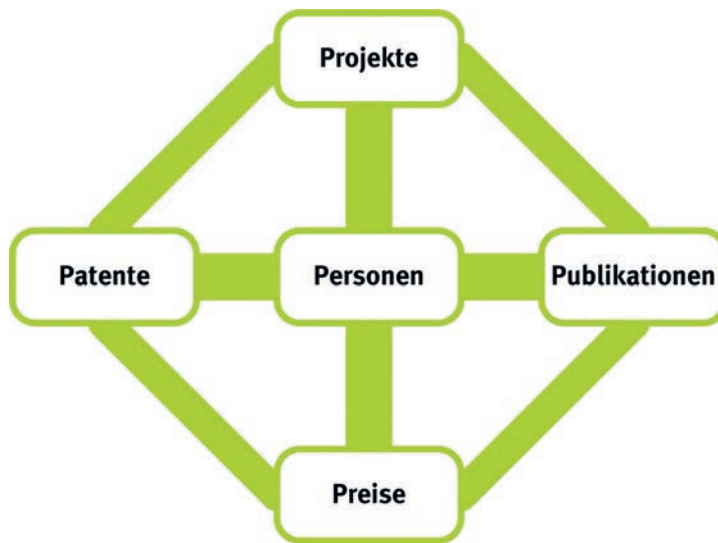
Abbildung 1: Interessengruppen für Forschungsinformationen



Übertragen auf die Charakteristika der WWU Münster lassen sich die grundlegenden inhaltlichen Kategorien von Forschungsinformationen und deren Beziehungen, die es in einem Forschungsinformationssystem abzubilden gilt, anhand von fünf Dimensionen strukturieren und in einem Ordnungsrahmen zusammenfassen (vgl. Abbildung 2). Im Mittelpunkt stehen hierbei *Personen*, die als Wissenschaftler an den verschiedenen Einrichtungen der WWU Münster forschen bzw. promovieren oder habilitieren. Neben *Projekten* können Personen über die von ihnen verfassten *Publikationen*

die wissenschaftliche Reputation ebenso erhöhen wie über erhaltene *Preise* oder *Patente*. Während in einigen Fachrichtungen Patente einen hohen Stellenwert genießen, sind in anderen Bereichen Forschungsprojekte – sowohl solche, die durch externe Mittelgeber gefördert werden, als auch solche, die aus dem Grundetat finanziert werden – relevant. Um den Kontext herauszustellen, in dem Forschungsaktivitäten unternommen und Forschungsergebnisse erzielt werden, ist es entgegen einer isolierten Betrachtung der einzelnen Dimensionen von Forschungsinformationen von besonderer Bedeutung, die genannten Inhaltsdimensionen über die Grenzen einzelner Disziplinen hinweg miteinander in Beziehung zu setzen und zu verknüpfen. Eine solch integrative Darstellung ermöglicht es nicht zuletzt, dass Zusammenhänge zwischen interdisziplinär angelegten Forschungsaktivitäten herausgestellt werden und dass das daraus resultierende Informationsnetz von den unterschiedlichen Adressaten zielgruppenspezifisch über verschiedene Navigationspfade durchlaufen werden kann.

Abbildung 2: Dimensionen von Forschungsinformationen



### 3 Konzeption eines Forschungsinformationssystems

Zur Entwicklung einer fachlichen Konzeption eines Forschungsinformationssystems für die WWU Münster wurde der grundlegende Ordnungsrahmen (vgl. nochmals Abbildung 2) als elementares Anforderungsgerüst aufgegriffen und hinsichtlich der einzelnen Dimensionen auf fachlicher Ebene weiter konkretisiert. Als Ausgangspunkt für die Entwicklung der detaillierten fachkonzeptionellen Spezifikation dienten neben dem europäischen Datenaustauschstandard für Forschungsinformationen CERIF (Jörg et al. 2012) vor allem die in Gesprächen mit unterschiedlichen Nutzergruppen erhobenen Anforderungen und Anwendungsszenarien, die es mit einem entsprechenden Forschungsinformationssystem zu adressieren gilt. Die hierbei einbezogenen Nutzergruppen erstrecken sich sowohl über die verschiedenen Fachdisziplinen als auch über die jeweiligen Ebenen der Fachbereiche (von der Fachbereichsleitung bis auf die Ebene des einzelnen Wissenschaftlers) sowie der Universitätsverwaltung hinweg. Ein auf die grundlegenden Elemente reduzierter Ausschnitt der entwickelten fachlichen Konzeption des Forschungsinformationssystems für die WWU Münster ist in Form des Entity-Relationship Models (Chen 1976) in (min, max)-Notation (ISO 1982) in Abbildung 3 gegeben.

Dem zugrundeliegenden Ordnungsrahmen folgend ist das zentrale Element des fachkonzeptionellen Datenmodells die *Person*, die als Angehörige(r) der WWU Münster wissenschaftliche Aktivitäten betreibt sowie Forschungsergebnisse erzielt und damit unterschiedliche Beziehungen zu



Abbildung 3: Konzeptionelle Spezifikation eines Forschungsinformationssystems (Ausschnitt)



Informationen zu Forschungsaktivitäten werden nicht allein auf der Ebene einzelner Personen betrachtet, sondern ebenfalls auch für die unterschiedlichen Einrichtungen einer Universität wie bspw. Fachbereichen, Instituten, Lehrstühlen bzw. Arbeitsgruppen sowie Forschungszentren. Um dies zu ermöglichen, werden Personen *Organisationseinheiten* zugeordnet. Die sich hieraus ergebende *Zugehörigkeit* gibt an, in welcher *Funktion* (bspw. Leiter(in), wissenschaftliche(r) Mitarbeiter(in) oder studentische Hilfskraft) und für welche *Zeitspanne* eine Person einer Organisationseinheit angehört. Im Sinne der Aufbauorganisation ist eine Organisationseinheit wiederum in eine *Organisationsstruktur* eingebettet. Da sich die Struktur der Organisationseinheiten im Zeitverlauf verändern kann, wird hierfür weiterhin die *Zeitspanne* angegeben, in der die jeweilige Organisationsstruktur gültig ist. Während der Zugehörigkeit zu einer Organisationseinheit kann eine Person für eine bestimmte Zeit an einem *Projekt* mitwirken. Die Art der Mitwirkung kann unterschiedlicher Natur sein. Exemplarisch kann es sich hierbei um die *Mitwirkungsart* in Form einer Sprecher-, Projektleiter-, Projektmitarbeiter- oder einer Hauptantragstellerrolle handeln. Anhand von Teilvorhaben oder Arbeitspaketen kann ein Projekt in weitere Teilprojekte untergliedert werden, wodurch sich aus Sicht des Gesamtprojektes eine *Teilprojekthierarchie* ergibt. Somit ist die Möglichkeit gegeben, Verbund- und Kooperationsprojekte wie bspw. DFG geförderte Sonderforschungsbereiche mit ihrer jeweiligen internen Struktur darzustellen.

Sowohl für Zwecke der Kommunikation von Forschungsprojekten als auch für Evaluations- und Steuerungszwecke ist es von gesonderter Bedeutung herauszustellen, ob ein Projekt in Form eines Drittmittelprojektes durch externe Fördereinrichtungen oder durch Landesmittel finanziert wird bzw. ob das Projekt als sog. Eigenmittelprojekt aus dem internen Haushalt der Universität getragen wird. Um zu markieren, dass es sich bei einem Projekt um ein Drittmittelprojekt handelt, können diesem die entsprechenden *Fördereinrichtungen mit den jeweiligen Förderinstrumenten* zugeordnet werden. Ein Förderer bzw. ein Instrument wird hierzu aus einer vorgegebenen *Förderersystematik* ausgewählt, die neben öffentlichen Förderinstitutionen (z. B. die DFG oder die EU) auch die jeweiligen Förderinstrumente und -programme (bspw. Sonderforschungsbereiche und Forschergruppen als Instrumente der DFG oder die einzelnen Forschungsrahmenprogramme der EU) beinhaltet. Ferner kann eine derartige Systematik ebenfalls privatwirtschaftliche Förderinstitutionen (z. B. Unternehmen) sowie Stiftungen und weitere Förderer beinhalten. Dies ermöglicht nicht nur eine detailliertere Darstellung von Projekten sondern auch eine differenzierte Betrachtung dieser für Evaluations- und Steuerungszwecke.

Personen können neben der Mitwirkung bei Forschungsprojekten und deren Beteiligung als *Erfinder* bei *Patenten* auch als Autoren von *Publikationen* in Erscheinung treten. Vergleichbar zu Projekten wird die *Autorenschaft* einer Person bei einer Publikation nicht direkt über die Person, sondern über deren organisatorische Zugehörigkeit hergestellt. Dies ist insbesondere bei Personen von Belang, die mehreren Organisationen angehören, um angeben zu können, welcher der Organisationen diese Publikation zuzuordnen ist. Eine Publikation kann anhand verschiedener bibliografischer Eigenschaften wie dem Publikationstyp (bspw. Monografie, Sammelband, Konferenzbeitrag oder Zeitschriftenartikel) und in Abhängigkeit davon mithilfe unterschiedlicher Merkmale wie bspw. Titel, Veröffentlichungsjahr, Zeitschrift, Verlag, Seitenzahlen etc. charakterisiert werden. Inwieweit es diese Publikationsmerkmale separiert oder als ein zusammenhängender Zitationsstring in einem Forschungsinformationssystem zu hinterlegen und zu verwalten gilt, ist in Abhängigkeit der Anwendungsszenarien festzulegen, die es mit dem Forschungsinformationssystem zu adressieren gilt. Auf eine detaillierte bibliografische Betrachtung von Publikationen kann aufgrund des Umfangs der Thematik an dieser Stelle jedoch nicht eingegangen werden.

Wenngleich eine Dissertations- oder Habilitationsschrift eine Publikation darstellt, wird das damit verbundene Verfahren einer *Promotion* und *Habilitation* als separates Informationsobjekt aufgefasst. Eine Person kann hierbei einerseits als *Kandidat* oder andererseits als *Betreuer* mit einem Promotions- oder Habilitationsverfahren in Verbindung stehen. Aus Gründen der Übersichtlichkeit wurden in



der Spezifikation (vgl. Abbildung 3) die Promotion und Habilitation zu einem Informationsobjekt zusammengefasst. Ferner wurde auf eine explizite Darstellung von Verknüpfungsrelationen zu weiteren Informationsobjekten wie Projekten, Publikationen oder Preisverleihungen verzichtet.

Nicht zuletzt als Würdigung besonderer Forschungsaktivitäten kann einer Person ein *Preis* verliehen werden. Unter einem Preis wird hierbei die generelle Ehrung wie bspw. der Gottfried Wilhelm Leibniz-Preis oder der ERC Advanced Investigators Grant verstanden, die zu einem *Zeitpunkt* (bspw. im Jahr 2010) an eine oder mehrere Personen in Form einer *Preisverleihung* vergeben wird. Erhält eine Person einen Preis in Verbindung mit einem Forschungsprojekt (bspw. ein Transferpreis), einer Publikation (bspw. ein Best Paper Award) oder einer Promotion (bspw. ein Dissertationspreis), kann dies über eine Verknüpfung der entsprechenden Inhalte mit der jeweiligen Preisverleihung hinterlegt werden.

Um die im Zusammenhang mit der *multidimensionalen und multiperspektivischen Betrachtung von Forschungsinformationen* (vgl. Abschnitt 2.2) angeführte Navigation zwischen den einzelnen Inhalten eines Forschungsinformationssystems zu adressieren, bestehen über die bereits explizit angesprochenen Verknüpfungen hinaus ebenfalls wechselseitige Relationen zwischen Projekten, Publikationen, Preisverleihungen, Promotionen und Habilitationen sowie Patenten. Hierüber kann letztlich dargestellt werden, dass aus einem Projekt eine Publikation hervorgegangen ist, wofür eine Person, die ebenfalls in dem Projekt mitarbeitet, einen Preis erhalten hat.

#### **4 Umsetzung und Einbettung in die Anwendungslandschaft der WWU**

Eine effektive und effiziente Verwaltung und Bereitstellung von Informationen über die an einer Universität unternommenen Forschungsaktivitäten und erlangten Ergebnisse entwickelt sich zunehmend zu einer wesentlichen Grundbedingung für eine zukunftsorientierte Forschung. Im Sinne eines integrierten Informationsmanagements rückt daher die verzahnte Bereitstellung von wissenschaftlichen und organisatorischen Informationen in Form einer einheitlichen Informationsbasis in den Mittelpunkt der Betrachtung. In Bezug auf die fachkonzeptionelle Spezifikation eines Forschungsinformationssystems wird die Notwendigkeit einer derartig integrierten Verwaltung und Bereitstellung von Forschungsinformationen insbesondere durch den Umstand deutlich, dass wissenschaftliche Inhalte wie Publikationen, Forschungsprojekte und Patente sowohl untereinander als mit organisatorischen Inhalten wie die Zugehörigkeit von Personen zu universitären Einrichtungen und deren Einbettung in die Aufbauorganisation der Universität eng verwoben sind.

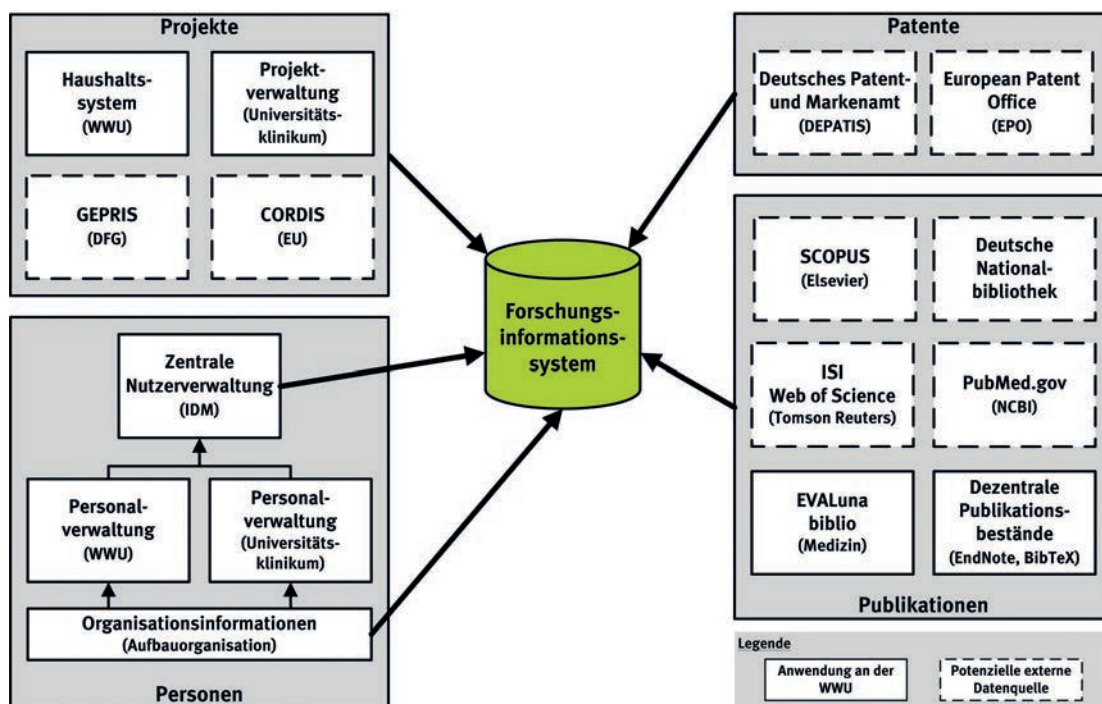
Aus Sicht der Anwendungslandschaft einer Universität liegen diese Informationen jedoch zu meist in unterschiedlichen anwendungsbezogenen Systemen vor, die nicht notwendigerweise auf Funktions- oder Datenebene miteinander verknüpft sein müssen. Vor dem Hintergrund eines integrierten Managements von Forschungsinformationen kann ein entsprechendes Forschungsinformationssystem nicht allein als eigenständige Anwendung betrachtet werden, sondern nimmt vielmehr eine integrierende Rolle zwischen verschiedenen Einzelanwendungen ein. Im Rahmen der Einführung eines Forschungsinformationssystems gilt es daher auf bestehenden Verknüpfungen zwischen Anwendungssystemen aufzusetzen sowie zusätzliche Integrationspunkte zwischen einzelnen Anwendungen zu etablieren.

An der WWU Münster werden Stammdaten einer Person wie Name, Vorname, akademischer Grad bzw. Titel sowie die Organisationszugehörigkeit in Systemen zur Personalverwaltung gepflegt. Hierbei ist anzumerken, dass die Personalstammdaten der Angehörigen des Universitätsklinikums und die Stammdaten des nicht-medizinischen Bereichs der WWU getrennt voneinander in zwei autonomen Personalsystemen verwaltet werden. Um eine konsistente Angabe von Organisationszugehörigkeiten zu gewährleisten, greifen diese beiden Systeme auf ein gemeinsames Verzeichnis mit Organisationsinformationen zurück, welches die gesamtuniversitäre Aufbauorganisation der WWU Münster – sowohl den nicht-medizinischen Bereich als auch das Universitätsklinikum – um-

fasst (vgl. auch im Folgenden Abbildung 4). Die Integration von medizinischen und nicht-medizinischen Disziplinen in einem Forschungsinformationssystem ist bislang in Deutschland einzigartig.

Um einen einheitlichen Zugang zu den relevanten Informationen und den bereitstellenden Anwendungen zu gewährleisten, werden die Stammdaten und Organisationszugehörigkeiten aus den beiden Personalverwaltungssystemen in einem zentralen Identitätsmanagementsystem zusammengeführt und auf dieser Basis universitätsweit eindeutige Nutzeridentitäten herausgebildet. Über eine Anbindung zum Identitätsmanagementsystem werden diese Nutzerangaben ebenfalls im Forschungsinformationssystem genutzt, um auf Basis der Personenstammdaten ein grundlegendes Personenprofil zu erstellen sowie anhand der für eine Person mitgelieferten Organisationen die entsprechenden Organisationszugehörigkeiten zu erzeugen. Die hierfür benötigten Organisationsinformationen werden aus dem zentralen Einrichtungsverzeichnis in das Forschungsinformationssystem übernommen. Über die Anbindung des Identitätsmanagementsystems wird ferner der Zugriff auf das Forschungsinformationssystem ermöglicht, sodass die einzelnen Personen die zu ihnen hinterlegten Inhalte eigenständig mit Hilfe ihrer universitätsweiten Nutzererkennung bearbeiten und pflegen können. Auf diesem Weg kann sichergestellt werden, dass jeder einzelne Nutzer die Hoheit über die auf die eigene Person bezogenen Inhalte behält und über deren Sichtbarkeit nach außen selbst entscheiden kann.

Abbildung 4: Integration des Forschungsinformationssystems in die Anwendungssystemlandschaft der WWU Münster



Informationen zu drittmittelgeförderten Forschungsprojekten sind in der Regel über ein zentrales Finanz- bzw. Haushaltssystem verfügbar. Vergleichbar zu den Personeninformationen werden an der WWU Münster die Forschungsprojekte des Universitätsklinikums ebenfalls getrennt von den Projekten des nicht-medizinischen Bereichs in separaten Haushaltssystemen verwaltet. Um bei der Verwaltung von Projekten im Forschungsinformationssystem eine konsistente Datenbasis gegenüber den Angaben in den Haushaltssystemen zu schaffen, werden die beiden Haushaltssysteme ebenfalls an das Forschungsinformationssystem angebunden und die hierüber verfügbaren Informationen zu Projekten laufend übernommen. Hierbei ist anzumerken, dass in einem Haus-

haltssystem Forschungsprojekte insbesondere aus Sicht der finanziellen Abrechnung und weniger aus einer inhaltlich beschreibenden Sicht betrachtet werden. Daher ist zumeist nur eine begrenzte Auswahl der hierüber verfügbaren Informationen wie bspw. Langtitel, Förderkennzeichen, Laufzeit und Mittelgeber unmittelbar für die Belange eines Forschungsinformationssystems relevant. Weiterführende Angaben wie bspw. eine Kurzbeschreibung oder die am Projekt beteiligten Personen und daraus hervorgegangenen Publikationen sind in der Regel nicht Gegenstand eines Haushaltssystems und müssen daher direkt im Forschungsinformationssystem gepflegt und den entsprechenden Projekten zugeordnet werden.

Über die innerhalb einer Universität verfügbaren Datenbestände hinaus betreiben auch Förderinstitutionen wie die DFG mit GEPRIS (<http://gepris.dfg.de/>) oder die EU mit CORDIS (<http://cordis.europa.eu/>) umfangreiche Datenbanken, worüber Informationen zu den jeweils geförderten Projekten verfügbar sind. Derartige Projektdatenbanken bieten ebenfalls aussichtsreiche Datenquellen für ein Forschungsinformationssystem. Einerseits können Universitäten und Forschungseinrichtungen auf einen qualitätsgesicherten Stand an Projektinformationen zurückgreifen. Andererseits bietet eine Verknüpfung mit derartigen Projektdatenbanken auch Förderinstitutionen die Möglichkeit auf weiterführende Projektinformationen wie zugehörige Publikationen, Preise oder Promotionen zu verweisen bzw. zuzugreifen. Ein derartiges Zusammenspiel von internen und externen Datenquellen leistet nicht zuletzt einen entscheidenden Beitrag zur Verbesserung der Qualität von und zur Vereinfachung des Zugangs zu Forschungsinformationen. Technische oder organisatorische Möglichkeiten zum Austausch von Projektdaten zwischen universitären Forschungsinformationssystemen und den Datenbeständen der Förderinstitutionen sind derzeit jedoch noch nicht bekannt.

## 5 Zusammenfassung und Fazit

Heutige Forschungsaktivitäten, insbesondere wie sie an einer Volluniversität wie die WWU Münster unternommen werden, sind in zunehmendem Maße von einer inter- und transdisziplinären Herangehensweise geprägt. In Folge dieser Entwicklung wird insbesondere die Verwaltung und Bereitstellung von Informationen zu Forschungsaktivitäten und -ergebnissen, die im Rahmen eines Forschungsprozesses unternommen bzw. generiert, ausgetauscht, dokumentiert und präsentiert werden, vor neue Herausforderungen gestellt. In Verbindung mit einer multidimensional und multiperspektivisch ausgerichteten Betrachtung von Forschungsinformationen stellen Forschungsinformationssysteme ein geeignetes Instrumentarium dar, um die Kommunikation und Steuerung von Forschungsaktivitäten gezielt zu unterstützen und somit diesen Herausforderungen zu begegnen.

Ausgehend von unterschiedlichen Anwendungsszenarien von Forschungsinformationssystemen ergeben sich vielfältige Fragestellungen und Hindernisse, die es sowohl bei der inhaltlichen Ausgestaltung und technischen Umsetzung als auch im Rahmen der Projektgestaltung und -organisation zu berücksichtigen gilt. Aus den im Verlauf der Einführung eines Forschungsinformationssystems an der WWU Münster gesammelten Erfahrungen lassen sich insbesondere generelle Erkenntnisse zur Gestaltung und Organisation eines Einführungsprojektes zusammenfassen.

- *Ausrichtung der inhaltlichen Ausgestaltung an der jeweiligen Wissenschaftslandschaft:* Als Volluniversität weist die WWU Münster ein umfassendes Spektrum an Forschungsdisziplinen (außer Ingenieurwissenschaften) auf, die ihrerseits wiederum unterschiedliche Forschungskulturen pflegen. Verbunden mit dem Ziel, ein universitätsweites Forschungsinformationssystem zu etablieren, gilt es, die jeweiligen Fachkulturen und die jeweils daraus erwachsenden Anforderungen an die Verwaltung von Forschungsinformationen bei der Entwicklung einer Konzeption und der späteren Umsetzung zu vereinbaren und zu adressieren. Infolgedessen nimmt die jeweilige Forschungslandschaft einer Universität einen entscheidenden Einfluss

auf die Ausgestaltung eines Forschungsinformationssystems.

- *Enge und frühzeitige Einbindung der verschiedenen Nutzerkreise und Betroffenen in den Projektverlauf:* Um eine adäquate Verwaltung und Bereitstellung von Forschungsinformationen zu gewährleisten, gilt es, die jeweiligen Nutzerkreise und Betroffenen von Forschungsinformationssystemen frühzeitig bereits im Rahmen der fachlichen Spezifikation des Forschungsinformationssystems einzubeziehen. Einerseits ermöglicht die frühzeitige Beteiligung eine aktive Mitwirkung an der Ausgestaltung des Systems, um somit ein frühes Bewusstsein für das System und dessen Funktionsweise zu schaffen. Andererseits kann durch den frühen Einbezug möglichst aller Nutzerkreise eine breite Abdeckung hinsichtlich der unterschiedlichen Forschungskulturen sowie von künftigen Anwendungsfällen eines Forschungsinformationssystems sichergestellt werden. Werden Nutzergruppen erst zu einem späteren Zeitpunkt (bspw. im Nachgang zur Umsetzung eines Forschungsinformationssystems) einbezogen, können Änderungen sowie zusätzliche Anforderungen nur in Verbindung mit hohen Aufwendungen und Kosten berücksichtigt werden.
- *Kommunikation der Anwendungsbereiche und Vorteile eines Forschungsinformationssystems:* Die enge Einbindung der Anwender in die frühe Phase der Konzeption stellt eine Maßnahme dar, um den Nutzerkreisen die Anwendungsbereiche des Forschungsinformationssystems offen zu legen und darüber hinaus frühzeitig Bewusstsein für das Forschungsinformationssystem zu schaffen. Weiterhin trägt die Kommunikation von Anwendungsszenarien insbesondere dazu bei, dass eventuelle Vorbehalte gegenüber der Nutzung eines Forschungsinformationssystems bspw. als Instrument zur Evaluation abgemildert werden. Für die Förderung der breiten Akzeptanz ist eine offensive Kommunikation von Vorteilen eines Forschungsinformationssystems von nicht zu unterschätzender Bedeutung, da diese insbesondere dem einzelnen Nutzer nicht per se bewusst werden und mehrheitlich durch Befürchtungen über einen möglichen Mehraufwand überlagert werden. Die Darstellung von Vorteilen sollte jedoch an konkreten Anwendungsfällen ausgerichtet sein und sich weniger auf einer generellen Ebene möglicher Nutzenpotenziale bewegen.
- *Rechtliche Absicherung des Betriebs eines Forschungsinformationssystems:* In engem Zusammenhang mit Vorbehalten über die Verwendung der in einem Forschungsinformationssystem strukturiert hinterlegten Forschungsinformationen ist die rechtliche Legitimation des Betriebs eines Forschungsinformationssystems zu sehen. Eine Veranlassung hierzu ergibt sich zum einen aus dem Umstand, dass mit einem Personenprofil und dem damit verbundenen Lebenslauf explizit personenbezogene Daten erhoben werden. Darüber hinaus lässt sich weiter argumentieren, dass Informationen wie Publikationen oder Forschungsprojekte, die grundlegend keinen direkten Bezug zu Personen aufweisen, durch die direkte Verknüpfung mit einem Personenprofil in einem Forschungsinformationssystem jedoch einen direkten Personenbezug erhalten. Neben der rechtlichen Absicherung des Betriebs eines IT-gestützten Systems ist es daher auch von besonderer Bedeutung, dass die Hoheit über die eigenen Daten beim Nutzer verbleibt. Letztlich muss es insbesondere bei schutzwürdigen Forschungsinformationen dem einzelnen Wissenschaftler überlassen werden, ob diese der Geheimhaltung unterliegen oder der Öffentlichkeit zugänglich gemacht werden können.
- *Datenintegrations- und Organisationsprojekt:* Im Hinblick auf die Umsetzung und Einführung eines Forschungsinformationssystems wird deutlich, dass ein solches System nicht als eigenständige Anwendung betrachtet werden kann, sondern aufgrund der angestrebten Verknüpfung von wissenschaftlichen und organisationsbezogenen Informationen vielmehr eine integrative Stellung innerhalb der Anwendungslandschaft einer Universität einnimmt. Infolgedessen liegt eine wesentliche Aufgabe der Umsetzung darin, die unterschiedlichen Datenbestände zu harmonisieren und innerhalb des Forschungsinformationssystems zusammenzuführen. Die damit verbundenen Problemstellungen beschränken sich jedoch nicht allein auf die Etablierung technischer Schnittstellen, sondern adressieren vor allem Fragestellungen der Datenstruktur und -qualität. Wenngleich über Personal- oder Haus-



haltssysteme grundlegende Informationen zu Personen oder Projekten verfügbar sind, werden diese durch die Übernahme in ein Forschungsinformationssystem in einen neuen Verwendungskontext mit anderen Anforderungen gebracht, denen die ursprünglichen Daten nicht notwendigerweise gerecht werden. Unabhängig von der technischen Realisierbarkeit muss bei der Einbindung des Forschungsinformationssystems in die jeweilige Anwendungslandschaft für jedes System separat geprüft werden, ob die Datenstruktur und -qualität sowie die jeweiligen Pflegeprozesse den Anforderungen des Forschungsinformationssystems gerecht werden.

## Literatur

- Albert, Max*, 2004: Methodologie und die Verfassung der Wissenschaft – Eine institutionalistische Perspektive, in: *Held, Martin / Kubon-Gilke, Gisela / Sturn, Richard (Hrsg.): Ökonomik des Wissens. Jahrbuch Normative und institutionelle Grundfragen der Ökonomik*, Bd. 3. Marburg: Metropolis, 127-150.
- Alexander von Humboldt-Stiftung (Hrsg.)*, 2008: Publikationsverhalten in unterschiedlichen Disziplinen – Beiträge zur Beurteilung von Forschungsleistung, Diskussionspapier Nr. 12. Bonn: Alexander von Humboldt-Stiftung.
- Berghoff, Sonja / Federkeil, Gero / Giebisch, Petra / Hachmeister, Cort-Denis / Hennings, Mareike / Roessler, Isabel / Ziegele, Frank*, 2009: CHE Hochschulranking. Vorgehensweise und Indikatoren 2009. Arbeitspapier Nr. 119. Gütersloh: CHE Centrum für Hochschulentwicklung
- Beyer, Janice M. / Trice, Harrison M.*, 1982: The Utilization Process: A Conceptual Framework and Synthesis of Empirical Findings. *Administrative Science Quarterly* 27(4), 591-622.
- Bovelet, Daniel*, 2006: The Use of Research Funding Databases for Research Assessment Information Systems, in: *Asserson, Anne Gams Steine / Simons, Eduard J. (eds.): Enabling Interaction and Quality: Beyond the Hanseatic League. 8th International Conference on Current Research Information Systems*. Belgium: Leuven University Press, 41-54.
- Carayol, Nicolas / Matt, Mireille*, 2004: Does research organization influence academic production? Laboratory level evidence from a large European university. *Research Policy* 33(8), 1081-1102.
- Chen, Peter Pin-Shan*, 1976: The Entity-Relationship Model – Toward a Unified View of Data. *ACM Transactions on Database Systems* 1(1), 9-36.
- Cheng, Donghong / Claessens, Michel / Gascoigne, Toss / Metcalfe, Jenni / Schiele, Bernard / Shi, Shunke (eds.)*, 2008: *Communicating Science in Social Context – New models, new practices*. Dordrecht: Springer Netherlands.
- Closs, José S. / Cheater, F.M.*, 1994: Utilization of nursing research: culture, interest and support. *Journal of Advanced Nursing* 19(4), 762-773.
- euroCRIS*, 2012: CRIS concept and CRIS benefits. euroCRIS website. Online: [http://www.eurocris.org/Index.php?page=concepts\\_benefits&t=1](http://www.eurocris.org/Index.php?page=concepts_benefits&t=1). Abrufdatum: 2012-03-19.
- Fox, Mary Frank*, 1992: Research, Teaching, and Publication Productivity: Mutuality Versus Competition in Academia. *Sociology of Education* 65(4), 293-305.
- Hornbostel, Stefan*, 2001: Third Party Funding of German Universities. An Indicator of Research Activity? *Scientometrics* 50, 523-537.
- ISO/TC97/SC5/WG3*, 1982: *Concepts and Terminology for the Conceptual Schema and the Information Base*. New York: American Standards Institute.
- Jörg, Brigitte et al.*, 2012: CERIF 1.3 Full Data Model (FDM): Introduction and Specification. Technical Report. Online: [http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3\\_FDM.pdf](http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_FDM.pdf). Abrufdatum: 2012-03-19.
- Krücken, Georg / Meier, Frank*, 2006: Turning the University into an Organizational Actor, in: *Drori, Gili S. / Meyer, John / Hwang, Hokeyu (eds.): Globalization and Organization*. Oxford: Oxford University Press, 241-257.

- Landry, Rejean / Amara, Nabil / Lamari, Moktar*, 2001: Utilization of social science research knowledge in Canada. *Research Policy* 30(2), 333-349.
- Leonard, Thomas C.*, 2002: Reflection on Rules in Science: An Invisible-Hand Perspective. *Journal of Economic Methodology* 9(2), 141-168.
- Münch, Richard*, 2006: Drittmittel und Publikationen. *Forschung zwischen Normalwissenschaft und Innovation. Soziologie* 35(4), 440-461.
- Royal Society*, 2011: Knowledge, Networks and Nations: Global scientific collaboration in the 21st century. RS Policy document 03/11.
- Schimank Uwe*, 2005: New public management and the academic profession: Reflections on the German situation. *Minerva* 43(4), 361-376.
- Statistisches Bundesamt*, 2010: Personal an Hochschulen - Fachserie 11 Reihe 4.4 – Daten zum Hochschulpersonal, Personalstellen und Stellenäquivalente sowie Habilitationen an deutschen Hochschulen im Berichtsjahr 2009. Wiesbaden: Statistisches Bundesamt.
- Universität Münster*, 2010: Statistisches Jahrbuch der Westfälischen Wilhelms-Universität Münster 2010. Online: <http://www.uni-muenster.de/wwu/statistik/> Abruf am 2012-03-10.
- Walstad, Allan*, 2002: Science as a Market Process. *Independent Review* 8(1), 5-45.





## Einführung eines integrierten Forschungsinformationssystems am Karlsruher Institut für Technologie

### 1 Ausgangslage am KIT

Das Karlsruher Institut für Technologie (KIT)<sup>1</sup> verfügt über eine lange Tradition der dezentralen Erfassung und des Nachweises der Veröffentlichungen und Forschungsaktivitäten der Wissenschaftlerinnen und Wissenschaftler. Das KIT ging 2009 aus den beiden Vorgängereinrichtungen Universität Karlsruhe (TH) und Forschungszentrum Karlsruhe hervor. An beiden Standorten des KIT haben sich aufgrund der langjährigen Erfahrung der zentralen Bibliotheken und Forschungsfördereinrichtungen im Umgang mit zentralen Nachweisinstrumenten zwei unterschiedliche Forschungskulturen entwickelt, die den Umgang mit Forschungsmetadaten im Grundsatz ähnlich, aber in den Details unterschiedlich gestalten. Daher gibt es am KIT standortspezifische Unterschiede bei der Dokumentation von Forschungsmetadaten hinsichtlich der Datenmodelle und der Erfassungsmethodik. Aber der Unterschied der beiden Forschungskulturen betrifft nicht nur den Einsatz verschiedener Softwaresysteme und Datenmodelle. Er wird besonders deutlich im Umgang mit der bisher praktizierten Nachweispflicht von Veröffentlichungen der Wissenschaftlerinnen und Wissenschaftler beider Forschungseinrichtungen. Der heutige „Campus Nord“ des KIT ist Teil des Großforschungsbereichs der Helmholtz-Gemeinschaft und unterliegt daher der sogenannten *Programmorientierten Förderung*. Die Mittelzuweisung an die Institute basiert auf einem Leistungsnachweis, für den die zentrale Auflistung der Publikationen und Vorträge in einschlägigen Fachzeitschriften und Kongressen ein wichtiger Baustein ist. Daher ist die Abdeckungsrate der Veröffentlichungen in der Publikationsdatenbank des Forschungszentrums nahezu vollständig. Eine zentrale Publikationsordnung regelt im Detail den Genehmigungsworkflow, den jedes Publikationsvorhaben und jeder eingereichte Vortrag durchlaufen muss. Im Gegensatz dazu folgte die Universität Karlsruhe (TH) dem Leitbild der Freiheit von Forschung und Lehre und es gab weder eine zentrale Nachweisverpflichtungen noch ausreichende Anreize für Forschende, ihre Publikationen in das Veröffentlichungsverzeichnis zu melden. Die Abdeckungsrate dieser Universitätsbibliographie ist entsprechend geringer als die der Publikationsdatenbank des Forschungszentrums und beruht zum großen Teil auf den Rechercheanstrengungen der Bibliothek. Nach der Gründung des KIT wurden die beiden Veröffentlichungsdatenbanken durch die KIT-Bibliothek virtuell vereinigt und im Katalog *KIT-Publikationen*<sup>2</sup> zu einer gemeinsamen Abfrage im Internet zusammengeführt.

Die Einführung eines neuen, zentralen Forschungsinformationssystems (FIS) für das KIT erfolgt also vor dem Hintergrund dieser unterschiedlichen Ausgangssituationen der beiden Vorgängereinrichtungen und mit der Anforderung, die existierenden großen Datenbestände in das neue System zu integrieren.<sup>3</sup>

### 2 Ziele des Projekts Forschungsinformationssystem

Vor dem Hintergrund der langjährigen Heterogenitäten und Prozessbrüche der KIT-Publikations- und Forschungsdokumentationen soll am KIT eine konsistente Übersicht über die Forschungsleistungen erreicht werden. Im März 2010 beschloss das Präsidium, im Rahmen des Projekts *KIM-*

---

1 <http://www.kit.edu/index.php>

2 <http://www.bibliothek.kit.edu/cms/kit-publikationen.php>

3 Stand März 2012: 160.500 Publikationen des KITs.

FIS ein integriertes Forschungsinformationssystem am KIT schrittweise aufzubauen. Es soll die Brüche der Vergangenheit überwinden und einheitliche Nachweisprozesse für alle Forschenden des KIT implementieren. Der Begriff *KIM* steht dabei für das *Karlsruher Integrierte Informationsmanagement*<sup>4</sup>. Dahinter steckt ein Konzept, das seit einigen Jahren am KIT versucht, die in der Vergangenheit häufig isoliert entstandenen IT-Strukturen und -Dienste zu bündeln. Im Rahmen eines einheitlichen Informationsmanagements wird eine ganzheitliche Betrachtungsweise verfolgt. Durch den Einbezug verschiedener Akteure und die über KIM institutionalisierte Zusammenarbeit zwischen den verschiedenen Einrichtungen des KIT sollen effizientere Servicestrukturen entstehen als in der Vergangenheit.

Das neue Forschungsinformationssystem wird für jeden Wissenschaftler des KIT seine Publikationen, Projekte, Patente und in weiteren Ausbaustufen betreute Promotionen, Preise und sonstige veröffentlichungswerte Forschungsinformationen wiedergeben. Hinter der Einführung eines Forschungsinformationssystems am KIT steht die Vision, dass aus einem föderativ organisierten Verbund von Systemen in der Phase der Datengewinnung durch die Abfrage externer und interner Datenquellen sowie in Kooperation mit den Wissenschaftlerinnen und Wissenschaftlern selbst Informationen gewonnen werden.

Verstanden wird unter einem Forschungsinformationssystem die Gesamtheit der Prozesse und Instrumente zur Gewinnung, Verknüpfung, Darstellung und Nutzung von Forschungs-Metadaten der am KIT tätigen Forschenden. KIM-FIS soll einerseits die Forschenden, andererseits die Organe der Leitung und zentralen Dienste von Dokumentationstätigkeiten entlasten sowie die Verknüpfung, Aggregation und Darstellung von Forschungsmetadaten erleichtern beziehungsweise in einigen Fällen überhaupt erst ermöglichen. Im eigentlichen System des Forschungsinformationssystems werden diese Informationen zu den Einzelpersonen des KIT und auch zu den Organisationseinheiten zugeordnet. Diese Daten können vielfältig und strukturiert genutzt werden, indem sie den Forschenden, der Leitung und auch den zentralen Diensten wie beispielsweise das Public Webangebot des KIT für verschiedenste Zwecke und in unterschiedlichster Form zur Verfügung stehen und nach innen und außen dargestellt und visualisiert werden können.

Erklärtes Ziel des Projekts ist es, dass alle Wissenschaftlerinnen und Wissenschaftlern des KIT in ausgewählten Phasen des Scientific Lifecycle beim Aufwand der Dokumentation von Projektkoordinaten und Publikationen unterstützt werden. Langfristig hat das Forschungsinformationssystem das Ziel, als verlässliche Grundlage für die Beurteilung der wissenschaftlichen Leistungsfähigkeit des KIT zu dienen und kann so auch in der Zukunft in die Erarbeitung von strategischen Entscheidungen für Forschung und Innovation, Controlling, Berichtswesen und Qualitätsmanagement einfließen.

### 3 Teilprojekt Bibliothek

Ein wichtiger Teilbereich des umfassenden Forschungsinformationssystems am KIT ist die Weiterentwicklung des Publikationsmanagements, den die KIT-Bibliothek<sup>5</sup> verantwortet. Wie bereits eingangs erwähnt hat die KIT-Bibliothek auf diesem Gebiet große Vorarbeiten geleistet und bringt ihre Expertise in das Projekt mit ein.<sup>6</sup> Das Hauptziel der Weiterentwicklungen des Publikationsmanagements am KIT im Rahmen des Projekts KIM-FIS<sup>7</sup> ist es, direkt am Forschungszyklus des Wissenschaftlers anzusetzen, die Datensammlung aus externen Fachdatenbanken und Repositorien zu integrieren und somit die Arbeitsprozesse weitgehend zu automatisieren. Dabei werden die

---

4 <http://kim.cio.kit.edu/>

5 <http://www.bibliothek.kit.edu/cms/index.php>

6 <http://www.bibliothek.kit.edu/cms/publizieren.php>

7 <http://www.ubka.uni-karlsruhe.de/eva/index.html>; [http://opac.fzk.de:81/de/fzk\\_oai\\_qsim\\_frm.html](http://opac.fzk.de:81/de/fzk_oai_qsim_frm.html)

bestehenden Publikationsdatenbanken und Volltextrepositorien EVA STAR und KAROLA an Campus Nord und Süd des KIT abgelöst und integriert. Daten im Zusammenhang mit Publikationen sollen nur noch einmal erfasst und bei der Zuordnung aus den externen Quellen nur einmal validiert werden. Ein großer Vorteil des künftigen Forschungsinformationssystems liegt darin, dass die Daten nach der Eingabe und Aggregation in den unterschiedlichsten Zusammenhängen – Webseiten, Curriculum Vitae, Publikationslisten, Literaturverwaltungsprogramme etc. – wieder verwendet werden können. Durch diesen neuen Nachnutzungsdienst entsteht ein entscheidender Mehrwert, der Autorinnen und Autoren von wissenschaftlichen Veröffentlichungen motivieren soll, ihre Daten in das neue zentrale System einzugeben. Ein weiterer wichtiger Baustein dieses Teilprojekts ist die elektronische Abbildung des bislang im Großforschungsbereich des KITs existierenden formularbasierten Genehmigungsworkflows für Publikationen. Parallel zum Abbildungsprozess im System wird in den entsprechenden Gremien des KIT eine einheitliche Publikationsordnung diskutiert. Das Projekt KIM-FIS ist hier ein entscheidender Impulsgeber für die Diskussion innerhalb der Forschungseinrichtung.

## 4 Teilprojekt Forschungsförderung

Die Forschungsförderung (FOR)<sup>8</sup> versteht sich als der zentrale Dienstleister des KIT bei der Einwerbung von Drittmitteln. Das von FOR betriebene derzeit noch aktive Datenbanksystem zur Erfassung von Forschungsmetadaten wird im Zuge der Neukonfigurierung von KIM-FIS abgelöst. Das Teilprojekt hat zum Ziel, die genannten Anforderungen im System zu implementieren und damit eine durchgängige Datenerfassung, Weiterverwertung der Informationen und ein leistungsfähiges Reporting in einer Multiuserumgebung zu ermöglichen. Die Metadaten zu Forschungsprojekten bilden die Grundlage vieler Recherchen und Berichte und sind letztlich unumgänglich, um dem Präsidium einen Überblick über die Diversität und thematische Relevanz der Forschung am KIT zu verschaffen. Bei der Einwerbung von Drittmitteln entstehen umfassende Sammlungen von Forschungsanträgen, die als Grundlage für die Erfassung von Metainformationen dienen können. Letztlich hilft die sogenannte Drittmittelanzeige in Baden-Württemberg unter anderem dabei, die erfassten Daten durch die Wissenschaftlerinnen und Wissenschaftler vervollständigen und validieren zu lassen. Forschende möchten im Rahmen ihrer Portfolio- und Forschungsschwerpunktdarstellung natürlich auch ihre eigenen Forschungsprojekte im Public Web präsentieren. Durch die Entwicklung des neuen Systems KIM-FIS werden viele Prozesse elektronisch abgebildet. Die Daten werden bereits bei der Antragstellung erfasst und stehen sowohl zentralen Dienstleistungseinheiten als auch den Wissenschaftlerinnen und Wissenschaftlern jederzeit zur weiteren Verwendung (z.B. nach einer Bewilligung) zur Verfügung. Mithilfe von Webservices und auf die jeweiligen Bedürfnisse angepassten Templates können die Informationen auf den Institutswebseiten oder auf persönlichen Seiten der Forschenden individuell aufbereitet dargestellt werden.

## 5 Projektplanung

Das Gesamtprojekt KIM-FIS wird modular und sukzessive eingeführt. In den ersten Phasen stehen der Aufbau eines effizienten und komfortablen Publikationsmanagements im Fokus sowie die Etablierung eines Systems zur Dokumentation von Forschungskompetenzen, Forschungsprojekten und Forschungsk Kooperationen. Diese Themenbereiche stellen die beiden Teilprojekte der ersten Projektphase, die im Juni 2010 begann. Der endgültige Zuschlag für die Einführung des FIS am KIT erfolgte jedoch erst im Dezember nach der Entwicklung eines Demonstrators. 2011 wurde im Zuge eines agilen Entwicklungskonzepts die Systemkonfiguration und Systemanpassung und -weiterentwicklung für das KIT vorangetrieben. Dieser agile Ansatz wurde bewusst gewählt, um statt einer umfangreichen und detaillierten Leistungsbeschreibung im Rahmen eines kleinschrittigeren Ansatzes

---

<sup>8</sup> <http://www.for.kit.edu/>

die Bedarfsermittlung und Prozessoptimierung in die Phase der Implementierung zu vergeben.

KIM-FIS war von Beginn an kein reines Softwareprojekt sondern muss vor dem Hintergrund der jungen Forschungsorganisation des KIT in erster Linie als ein Organisationsentwicklungsprojekt gesehen werden. Neben der Konfiguration der Software und Abbildung auf etablierte Prozesse müssen im KIT projektbegleitend die Prozesse und Instrumente zur Gewinnung, Verknüpfung, Darstellung und Nutzung von Forschungsmetadaten analysiert und optimiert wurden. In einigen Fällen stößt das KIM-FIS-Projekt auf ungeklärte Zuständigkeiten, in vielen Fällen wirkt es als ein Katalysator für Entscheidungen, die im KIT anstehen. KIM-FIS ist folglich in mehrfacher Hinsicht ein wichtiges Integrationsprojekt für das KIT.

## 6 Projektumsetzung und Organisation

Nach Ansicht des Präsidiums sollte ein umfassendes FIS nicht top down eingeführt werden, ohne die Anwender in Form der Institute und Wissenschaftlerinnen und Wissenschaftler zu integrieren. Der Grundsatz des Projekts lautet daher „Think big in small steps“ und das Forschungsinformationssystem wird langsam und schrittweise eingeführt. Um die spätere Akzeptanz im KIT zu erhöhen, basiert das Projekt auf dem Grundsatz der Partizipation. So wurden bereits bei der Erstellung des Leistungskatalogs für die Projektausschreibung sowie bei der Zusammenstellung der Auswahlkommission Wissenschaftlerinnen und Wissenschaftler der KITs einbezogen. Im Verlauf des Projekts stehen sie der Projektleitung und den Projektpartnereinrichtungen mit ihrem Rat und den Anwendungsfällen aus der Praxis zur Seite.

Schon bei der Organisation des Projekts sieht man die starke Vernetzung mit den Anwendern nach dem Grundsatz der Partizipation.

Abbildung 1: Projektorganisation



Auftraggeber der Projektleitung ist der Chief Information Officer des KIT. Neben den beiden Hauptprojektpartnern KIT-Bibliothek und Forschungsförderung gibt es verschiedene assoziierte Partner im KIT, die von Beginn an beratend das Projekt begleiten. Das sind zentrale Stellen wie Finanzmanagement, Innovationsmanagement und Pressestelle.

Das Projekt ist in der ersten Phase in zwei Teilprojekte untergliedert – Publikationen und Projekte. Die Teilprojekte werden von den jeweils zuständigen Fachabteilungen geleitet. Weitere Teilprojekte werden sich anschließen. In regelmäßigen Abstimmungsterminen innerhalb der Projektgruppe sowie mit dem externen Anbieter werden die Produktentwicklung und Anpassungen an die Bedürfnisse des KIT abgestimmt.

## 7 Datenmodell

Eine der Besonderheiten der Software Converis der AVEDAS AG<sup>9</sup> ist der modulare Aufbau: Die Datenhaltung der Teilmengen des Forschungsinformationssystems Publikationen, Projektinformationen sowie Patente werden getrennt gehalten und sind über die Personen aus dem zentralen Identitätsmanagement des KIT verknüpft. Die Personen werden zentral am KIT geführt und sind in Converis integriert. An den Personenstammdaten hängt die organisationale Zugehörigkeit des betreffenden Wissenschaftlers. Jede Person erhält für den Zugang zum System eine sogenannte „Visitenkarte“, die die Rechte und Rollen des zugrundeliegenden Datenmodells abbilden. Da einzelne Personen des KIT verschiedene Funktionen innehaben, können im System auch mehrere Visitenkarten vergeben werden. So hat beispielsweise ein Institutsleiter des KIT zwei Visitenkarten, an denen verschiedene Berechtigungen hängen. Wählt er sich im System mit seiner ersten Visitenkarte „Wissenschaftler“ ein, kann er seine gesamten Publikationen, Projekte und sonstigen persönlichen Daten im Forschungsinformationssystem einsehen und bearbeiten. Mit dem Zugang über die zweite Visitenkarte, in diesem Beispiel als Leiter einer Organisationseinheit am KIT, hat er die Berechtigung nicht nur seine, sondern alle Publikationen und Projektinformationen seines Instituts einzusehen.

Von Beginn an waren der Datenschutzbeauftragte des KIT und der Personalrat am Projekt KIM-FIS beteiligt. Die Berechtigungen aller Rollen im System wurden detailliert erläutert und sind mit diesen Einrichtungen des KIT abgesprochen. Ein hoher Stellenwert hat der Schutz personenbezogener Daten der einzelnen Wissenschaftlerinnen und Wissenschaftler: Es herrscht Konsens, dass auch ein Institutsleiter mit seiner erweiterten Sicht auf die Daten seiner Organisationseinheit im neuen Forschungsinformationssystem keine persönlichen Daten der Mitarbeiter einsehen kann, die im System hinterlegt sind.

Das KIT beabsichtigt, das neue Forschungsinformationssystem auch auf europäischer Ebene kompatibel zu gestalten und ist ein aktives Mitglied der europäischen Organisation EuroCRIS<sup>10</sup>, die das Ziel hat, einen europäischen Standard zur Beschreibung von Forschungsinformationen zu pflegen und weiter zu entwickeln. Ein wichtiger Grundbaustein dafür ist das Datenmodell des Common European Research Information Format (CERIF), das die Software Converis sowohl nativ als auch über einen standardkonformen CERIF- Import und -Export unterstützt.

## 8 Umsetzungen im Teilprojekt Publikationen

Die KIT-Bibliothek war bereits im Vorfeld der Entscheidung für KIM-FIS ein wichtiger Treiber für die Einführung eines Forschungsinformationssystems. Die Bibliothek plant mit Converis nicht nur die heterogenen Altsysteme der aus den beiden Zentralbibliotheken an Campus Nord und Süd vereinigten KIT-Bibliothek abzulösen. Mit der Einführung eines neuen, einheitlichen Systems eröffnet sich auch für die bibliotheksinternen Prozesse die Chance, die einzelnen Workflows zu vereinheitlichen und die neu vereinigten Abteilungen weiter zu integrieren.

Das Forschungsinformationssystem soll eine hohe Datenqualität, -quantität und -konsistenz beinhalten, um als eine verlässliche Grundlage für die wissenschaftliche Leistungsfähigkeit des KIT fungieren zu können. Eine zentrale Rolle kommt dabei der Bibliothek zu, die - wie in der Vergangenheit - die Rolle einer Qualitätssicherungsinstanz innehat. Gerade vor dem Hintergrund der langjährigen Erfahrung der Bibliothek auf diesem Gebiet waren die Anforderungen an den Systemanbieter bei der Umsetzung des Workflowmanagements von Seiten der Bibliothek sehr hoch.

Converis arbeitet nach dem Prinzip der verteilten Dateneingabe: sowohl die Mitarbeiter und Mitarbeiterinnen der Bibliothek als auch die Forschenden des KIT können Publikationsdaten eingeben.

---

<sup>9</sup> <http://www.avedas.com/de/converis.html>

<sup>10</sup> <http://www.eurocris.org>



Bei der Konfigurierung des Systems kamen auch die langjährigen Erfahrungen aus der Praxis des KIT zu tragen, und es wurde eine neue Rolle „Publikationsbeauftragter“ im Forschungsinformationssystem eingerichtet. Dieser Person kommt eine Vermittlerrolle zu, indem sie für Kolleginnen und Kollegen eines Instituts Daten erfasst. Der Grundgedanke für diese neue Rolle im System ist, dass auch in der Vergangenheit Institute des KIT in der Regel Personen benannt haben, die eine Übersicht über die Vollständigkeit der Daten verantworten - in manchen Fällen sind das Wissenschaftler, in einigen Fällen aber auch Sekretariate. Diese Personen erhalten im künftigen System eine gesonderte Visitenkarte und verfügen über ausgedehntere Rechte in Bezug auf Dateneingabe und -kontrolle. Sie werden gezielt von den Mitarbeiterinnen und Mitarbeitern der KIT-Bibliothek geschult und stehen in engem Kontakt zur dort zuständigen Fachabteilung Publikations- und Mediendienste.

Einer der großen Vorteile von Converis ist die Möglichkeit, Fremddaten aus externen Datenbanken (Scopus, Web of Science, PubMed etc.) nachnutzen zu können. Allerdings ist der Datenabzug stark vom bereit gestellten Datenmodell der entsprechenden Anbieter abhängig und Converis fungiert hier in erster Linie als eine Art Vermittler. Im Laufe des Projekts zeigte sich, dass nicht alle für das KIT relevanten Datenbanken angegliedert werden können, weil der automatisierte Datenabzug der bibliographischen Daten einen zu großen Nachbearbeitungsaufwand und in einigen Fällen auch ungewollte falsche Einträge verursacht. Die Datenbank Web of Science ermöglicht für Converis nur eine „Light“-Version des eigentlichen Gesamtverzeichnisses und kann keine Publikationen in Abhängigkeit des Publikationstyps importieren. Die genaue Zuordnung und Validierung ist jedoch für ein konsistentes und qualitativ hochwertiges Forschungsinformationssystem am KIT unerlässlich, denn im Bereich der Publikationen werden im Rahmen der Programmorientierten Förderungen die einzelnen Publikationstypen unterschiedlich gewichtet.

In das Softwaresystem der Firma AVEDAS wurden weitere Prozessschritte integriert, die in erster Linie vor dem Hintergrund der langjährigen Bibliotheksexpertise zu sehen sind: Der neu eingeführte Prozessschritt „Validierung durch die Bibliothek“ sichert vor der Datenfreigabe ins Public Web die Qualität der Angaben. Im Idealfall wird der Wissenschaftler komplett entlastet, wenn automatisiert ins System integrierte bibliographische Daten bereits von der Bibliothek abgearbeitet wurden, und er sie nicht erst selbst sichten muss.

Die Weiterentwicklungen des Publikationsmanagements von KIM-FIS sind vollständig eingebettet in die Open Access- Strategie des KIT<sup>11</sup>. Hier baut das Projekt auf fundierten organisatorischen und strategischen Vorarbeiten auf: Das KIT ist die erste Hochschule in Baden-Württemberg und auch bundesweit eine der wenigen, die sich in der illustren Reihe der Wissenschaftseinrichtungen platziert, welche die international vielbeachtete Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen unterzeichnet hat. Seit 2010 hat der KIT-interne Diskussionsprozess in den Gremien außerdem eine eigene Open Access Policy verabschiedet. Das KIT folgt damit konsequent dem Trend der Beförderung von neuen Formen der Wissenschaftskommunikation. Auch bei der Infrastruktur werden in Karlsruhe seit Jahren hohe Maßstäbe gesetzt: Neben langjährig etablierten Servicestrukturen für Open Access wie zentrale Publikationsserver und ein eigener Open Access Verlag nimmt das KIT auch an bundesweiten Modellprojekten teil: als eine der ersten Forschungseinrichtungen hat das KIT den Zuschlag zum neuen Förderprogramm Open Access Publizieren der Deutschen Forschungsgemeinschaft erhalten, das nachhaltige Strukturen der modernen Wissenschaftskommunikation befördert. Am KIT besteht daher seit 2010 ein eigener KIT-Publikationsfonds für die Finanzierung von Open Access-Publikationen. Um diese Entwicklungen zu verstetigen, hat das Präsidium des KIT im Oktober 2011 als erste deutsche Forschungseinrichtung den internationalen Compact for Open Access Publishing Equity (COPE) unterzeichnet.

---

11 <http://www.bibliothek.kit.edu/cms/open-access.php>

Die Open Access-Dokumente sowohl des KIT-Verlags KIT Scientific Publishing<sup>12</sup> sowie alle Dokumente der Repositorien sind organisatorisch mit den bibliographischen Daten der Publikationsdatenbanken verknüpft. Die KIT-Bibliothek stand folglich mit der Einführung des Forschungsinformationssystems vor der Frage, wie künftig Dokumente und bibliographische Nachweisinformation über die Systeme verknüpft werden sollen. Die KIT-Bibliothek ist daher mit der Firma AVEDAS AG parallel zum KIM-FIS-Projekt eine strategische Partnerschaft eingegangen mit dem Ziel, im Rahmen eines weiteren Projekts das Forschungsinformationssystem um Repository-Funktionalitäten zu erweitern. Für diesen integrierten Ansatz wird das Datenmodell neben den bibliographischen Daten um die Angaben zur Verwaltung von Volltexten erweitert. Das neue Projekt *Forschungsinformationssystem - Open Access Repository (FIS-OAR)*, das vom Ministerium für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg gefördert wird, startete Mitte 2011. Wichtige Ziele sind die Weiterentwicklung des Produkts Converis sowie die Standardisierung und DINI-Zertifizierung<sup>13</sup>. Nach diesem international anerkannten Verfahren werden Standards für die Vernetzung von digitalen Archiven definiert, die sich von technischen bis zu organisatorischen Anforderungen erstrecken. Diese Entwicklungen gehen über die klassischen Anwendungen eines Forschungsinformationssystems hinaus und bringen den Forschenden des KIT in vielfacher Hinsicht große Mehrwerte. Denn durch diesen integrierten Ansatz kann die an vielen Orten historisch gewachsene Trennung der Forschungsdaten von den Repositorien überwunden werden, indem der gesamte Forschungszyklus des Wissenschaftlers unterstützt wird. Die entstehende eResearch-Infrastruktur fungiert als Backend, in dem ein Forscher sein gesamtes Forschungsportfolio verwaltet - dazu gehören neben den Publikationen auch Informationen, die für die externe Kommunikation bestimmt sind. In Abhängigkeit der Open Access-Politik von Verlagen und von sonstigen Benutzerrechten können am KIT Volltexte künftig direkt aus dem Forschungsinformationssystem weltweit veröffentlicht werden.

## 9 Reporting und Außendarstellung im Forschungsinformationssystem des KIT

Wie eingangs erwähnt, ist ein funktionierendes und einheitliches Reporting, wie es bereits im Großforschungsbereich im Rahmen der Programmorientierten Förderung praktiziert wird, eine der wichtigsten Anforderungen an das neue System und steht bereits in der Pilotphase für die ersten Testinstitute in der vollen Funktionalität zur Verfügung. Darüber hinaus soll KIM-FIS aus einer gesamtstrategischen Sicht langfristig zur zentralen Anlaufstelle für die Datenauswertung rund um Publikationen und Projekte für das gesamte KIT werden. Dazu sind weitere Ausbaustufen des Projekts in den Folgejahren geplant und bereits beantragt. Parallel zum laufenden Projekt ist im Rahmen des weiteren Ausbaus von KIM-FIS geplant, weitere Module wie Innovation, Studium & Lehre und Berichtswesen zu integrieren. Zukünftig wird man ein Modul nach dem anderen sinnvoll integrieren. Nach angemessenen, iterierenden Testszenarien, Feedbackrunden und Migration bereits vorhandener Datenbestände in KIM-FIS werden die Systeme am KIT zunehmend ihren Platz finden und den Charakter eines allumfassenden Instrumentariums zur Außen- und Berichts-darstellung des KIT annehmen.

Über spezielle Reporting-Features hinaus gibt es im neuen System spezielle Mehrwerte in Form von Webdiensten, die für Institute und auch für Wissenschaftlerinnen und Wissenschaftler die Nachnutzung der Daten in einer moderneren und komfortableren Form ermöglichen als in den Vorgängerdatenbanken. Die an vielen Stellen selbst gepflegten Darstellungen und Internetseiten der Forschenden und Institute in Form von Forschungsprofilen, Publikationslisten und Lebensläufen werden durch sogenannte Webservices, die die Daten online aus KIM-FIS extrahieren und nutzerspezifisch angepasst aufbereiten, abgelöst. Künftig können Publikations- und Projektdaten

12 <http://www.ksp.kit.edu/shop/>

13 <http://www.dini.de/dini-zertifikat/>



direkt in persönliche und Institutswebseiten integriert und Lebensläufe direkt *on the fly* aus der Datenbank generiert werden. KIM-FIS wird außerdem einen zentralen Webauftritt auf den Einstiegsseiten des KIT anbieten, der einen Forschungsüberblick für die interessierte Öffentlichkeit sowie einfachere Suchanfragen zur Verfügung steht. Die Informationen werden dreidimensional verlinkt und nach Personen, Einrichtungen und den Modulen Publikationen und Projekte bereitgestellt.

Darüber hinaus ist die Möglichkeit eines Zugriffs mittels SQL-Abfragen auf das vollständige Datenangebot für Administratoren geplant. KIM-FIS wird darüber hinaus mit der bestehenden Förderangebotsdatenbank verknüpft, um dem System seinerseits Metainformationen über die zu Grunde liegenden Förderungen zur Verfügung zu stellen und entsprechende Auswertungen zu ermöglichen.

## 10 Betriebskonzept

Im Verlauf des ersten Projektjahrs wurden vielfache Gespräche mit zentralen Einrichtungen des KIT geführt, um den langfristigen Betrieb des Forschungsinformationssystems über ein nachhaltiges Betriebskonzept zu etablieren. Der Betrieb von KIM-FIS wird in drei Schichten gewährleistet. Die unterste technische Betriebsebene (Anwendungsbetrieb) wird die Hardwarebetreuung, das Monitoring, Wartungsarbeiten und die Datensicherheit umfassen. Der Anwendungsbetrieb wird vom KIT eigenen Steinbuch Centre for Computing (SCC)<sup>14</sup> unterstützt. Im Rahmen der Anwendungsbetreuung werden Softwarepflege und Change Management des Produktes mit dem Anbieter abgestimmt. Letztlich findet die Anwenderbetreuung im Rahmen eines Service Help Desks statt, das in der KIT-Bibliothek angesiedelt wird. Mit Unterstützung der jeweiligen Fachabteilungen werden hier der KIT-weite Service Support sowie Schulungen und Marketing angesiedelt.

## 11 Erfahrungen aus der Einführung des Forschungsinformationssystems am KIT - *think big in small steps*

Die hier geschilderte Ausstattung mit technisch basierten Informationssystemen entspricht dem Zeitgeist der Hochschul- und Wissenschaftslandschaft. Im nationalen und internationalen Wettbewerb, der geprägt ist von Außendarstellungen, Vergleichen, Rankings und konkurrierenden Qualitäten von Forschung, Lehre und Innovation bilden sie die Grundlage zur Erfassung der Qualität und Quantität von Leistungskenngrößen und merkmalspezifischen, diversifizierten Inhalten. Ein funktionsfähiges und reaktionsschnelles Berichtswesen stärkt die Hochschulleitungen und sollte zuverlässige und zeitnahe Auswertungen bieten. Informationstechnische Systeme können dafür die notwendigen Hilfsmittel zur Verfügung stellen. Auf der anderen Seite der Anwender dieser Systeme stehen die Wissenschaftlerinnen und Wissenschaftler, die zunehmend die Rolle von modernen Forschungsmanagern einnehmen und sich dazu mit möglichst minimalem Aufwand in der Außen- und Innendarstellung gut positionieren möchten.

Das KIT hat sich für eines dieser Systeme entschieden und durchläuft derzeit die verschiedenen Phasen der notwendigen Anpassung in den Modulen Publikationen und Projekte. Von Anfang an stehen die Interessen der Forschenden am KIT im Vordergrund, welche letztlich eine besondere Berücksichtigung bei den Anforderungen und der Umsetzung von Mehrwerten und Synergien des Produktes erfahren.

Eine der wesentlichen Erkenntnisse der Einführung des Forschungsinformationssystems am KIT ist die Bestätigung der Bedeutung von Prozessen und Strukturen sowie Zugang und Pflege von Stammdaten im Sinne des *business developments* und die Berücksichtigung der weitreichenden Auswirkungen auf andere datenführende Systeme, Schnittstellen und Dienstleistungsanbieter im KIT. Letztlich zeigen die Erfahrungen des Projekts KIM-FIS, dass mit den Grundsätzen des Karlsruher

---

<sup>14</sup> <http://www.scc.kit.edu/>

Integrierten Informationsmanagements KIM eine valide Basis für dieses Organisationsentwicklungsprojekt vorhanden ist. Neue Kooperationsmodelle unter den Dienstleistern des KIT ermöglichen, dass unter Berücksichtigung des Datenschutzes und der persönlichen Belange der Wissenschaftscommunity ein Rechte- und Rollenkonzept entwickelt werden konnte, das den tagesaktuellen, individualisierten und funktionsabhängigen Zugriff auf den diversen Organisationsebenen erlaubt. Durch die Schaffung von Schnittstellen zu den Personal verwaltenden Systemen kann so die Authentifizierung der Nutzer in KIM-FIS über zentrale Zugangsberechtigungen abgewickelt werden.

In den kommenden Monaten wird das System zunächst an einzelnen Testinstituten und später am gesamten KIT eingeführt, um sich nach einer entsprechenden Bewährungsphase zunehmend etablieren zu können. In den weiteren Projektphasen der nächsten Jahre werden die Module Innovation, Studium & Lehre und Berichtswesen in KIM-FIS integriert und in den Produktivbetrieb überführt. Das KIT wird so die Chancen der informationstechnischen Unterstützung seiner Forschenden auf bestmöglichem Weg nutzen.



**Barbara Ebert, Alexander Kujath, Joachim Holtorf, Karsten Holmberg,  
Thomas Rupp**

## **Erfahrungen aus der Einführung des Forschungsinformationssystems Pure an der Leuphana Universität Lüneburg**

### **1 Einführung**

Die Leuphana Universität Lüneburg betreibt seit 2011 ein integriertes Forschungsinformationssystem. Pure steht für „Publications and Research“ und führt die bisher fragmentierten Informationen über Projekte, Publikationen und Expertise der Universität zusammen. So soll eine einheitliche, zwischen Wissenschaft und Verwaltung abgestimmte Datenbasis entstehen, die die bisherige Praxis der dezentralen, teils redundanten Datensammlungen ersetzt und Administration wie Wissenschaft in der Forschungsberichterstattung entlastet. Die Einrichtung einer Forschungsdatenbank ist Teil der 2007 vom Senat verabschiedeten Forschungspolitik.

Die Leuphana Universität ging 2005 aus der Fusion der Fachhochschule Nordostniedersachsen und der Universität Lüneburg hervor. Mit 7.500 Studierenden, 155 Professorinnen und Professoren sowie 300 wissenschaftlichen Mitarbeiterinnen und Mitarbeitern gehört sie zur Gruppe der mittelgroßen Universitäten in Deutschland.

In den vergangenen sieben Jahren hat sich am Standort eine Modelluniversität entwickelt, die besonders in der Lehre Neuland beschreitet und mehrfach für ihre visionären Konzepte ausgezeichnet wurde. Das wissenschaftliche Profil der Leuphana bilden vier fachübergreifende, transdisziplinäre Wissenschaftsinitiativen, die auch den thematischen Rahmen für die vier im Oktober 2010 neu gegründeten Fakultäten abstecken: Bildung, Kulturwissenschaften, Wirtschaftswissenschaften und Nachhaltigkeit.

Ein wichtiger Meilenstein in der Neuausrichtung war 2009 die Genehmigung des EU-Großprojekts „Innovations-Inkubator Lüneburg“. Für sechs Jahre werden aus dem Europäischen Fonds für regionale Entwicklung sowie vom Land Niedersachsen Gesamtinvestitionen von 100 Mio. Euro zur Verfügung gestellt, um aus der Universität heraus in einem europaweit einmaligen Vorhaben die regionale Wirtschaftsentwicklung substanziell zu stärken.

#### **1.1 Besondere Anforderungen an die Forschungsberichterstattung**

Die fusionierte Universität hatte in den vergangenen Jahren eine Reihe von Herausforderungen zu bewältigen und stand bzw. steht im besonderen Fokus des Landes. Eines der Ziele, an denen der Erfolg der Neuausrichtung gemessen wird, ist die Entwicklung des akademischen Renommées und eine spürbare Steigerung der zentralen Kennzahlen für die Forschung. Der Forschungsberichterstattung kommt an der Leuphana daher eine besondere Bedeutung zu.

Forschung gehört neben der Lehre zu den wichtigsten Aufgaben der Universität. Informationen über Aufgabenerfüllung und Leistungen in diesem Bereich sollten unaufwändig und verlässlich zur Verfügung stehen. In der Realität sind diese Informationen jedoch über viele Stellen und Systeme verteilt – teils inneruniversitär, teils in externen Publikations- und Projektdatenbanken. Intern wurden vorab rund zwei Dutzend Prozesse identifiziert, die regelmäßig Angaben zu Publikationen, Projekten, wissenschaftlichen Aktivitäten, Kooperationen usw. erheben – sei es auf Organisationsebene (Internetseiten, Mittelvergabe), sei es thematisch (Fächerevaluationen des Landes Niedersachsen, Teilnahme an Rankings) oder individuell, wenn Wissenschaftlerinnen und Wissenschaftler Anträge einreichen oder persönliche Webseiten pflegen.

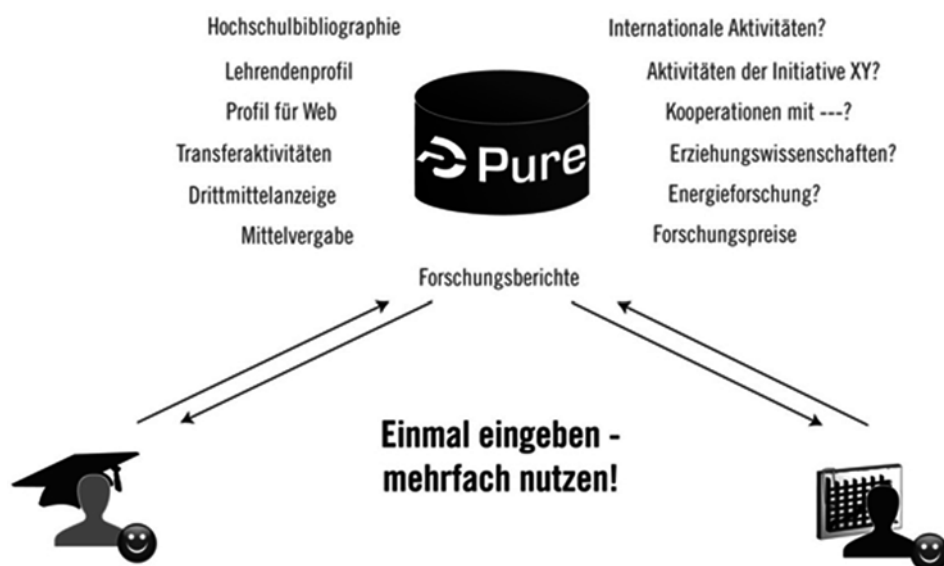
Die benötigten Informationen sind in der Regel ähnlich, jedoch sind sie aufgrund der individuellen

Formate oft nur mit großem Aufwand wieder verwendbar und durch die dezentrale Organisation schlecht verfügbar. Daraus resultieren unweigerlich Mehrfachabfragen und erhöhter administrativer Aufwand für die einzelnen Berichtsanslässe. Daher werden im Projekt Forschungsdatenbank vorrangig zwei Einzelziele verfolgt:

1. Schaffung einer einheitlichen, zwischen Wissenschaft und Verwaltung abgestimmten Datenbasis für die Forschungsberichterstattung
2. Schaffung von Mehrwerten für die Wissenschaftlerinnen und Wissenschaftler durch verbesserte Verfügbarkeit der Informationen für eigene Zwecke

In der Implementierungsphase wurden u.a. mehrere Tausend Publikationen und Hunderte von Autorinnen und Autoren der seit 2007 geführten Hochschulbibliographie in das neue System übernommen. Informationen über Organisationen und Personal sowie Drittmittel werden täglich aus dem SAP-System zur Verfügung gestellt. Wissenschaftlerinnen und Wissenschaftler steuern über ihr individuelles Nutzerkonto die Informationen über ihre Projekte, Publikationen und Aktivitäten. Zur Unterstützung können sie auf ein Netzwerk geschulter Redakteurinnen und Redakteure in den Instituten und Zentren zurückgreifen. Einmal registriert, stehen die Informationen regelhaft für interne Zwecke zur Verfügung, so dass die bislang praktizierten Mehrfachabfragen reduziert werden können. Nutzerinnen und Nutzer können für jeden Datenbankeintrag gesondert festlegen, ob er auch für den öffentlichen Forschungskatalog verfügbar sein soll.

*Abbildung 1: Ziel der Datenbankeinführung – eine autorisierte Sammlung von Forschungsinformationen schaffen, Mehrfachmeldungen und redundante Datenhaltung verringern*

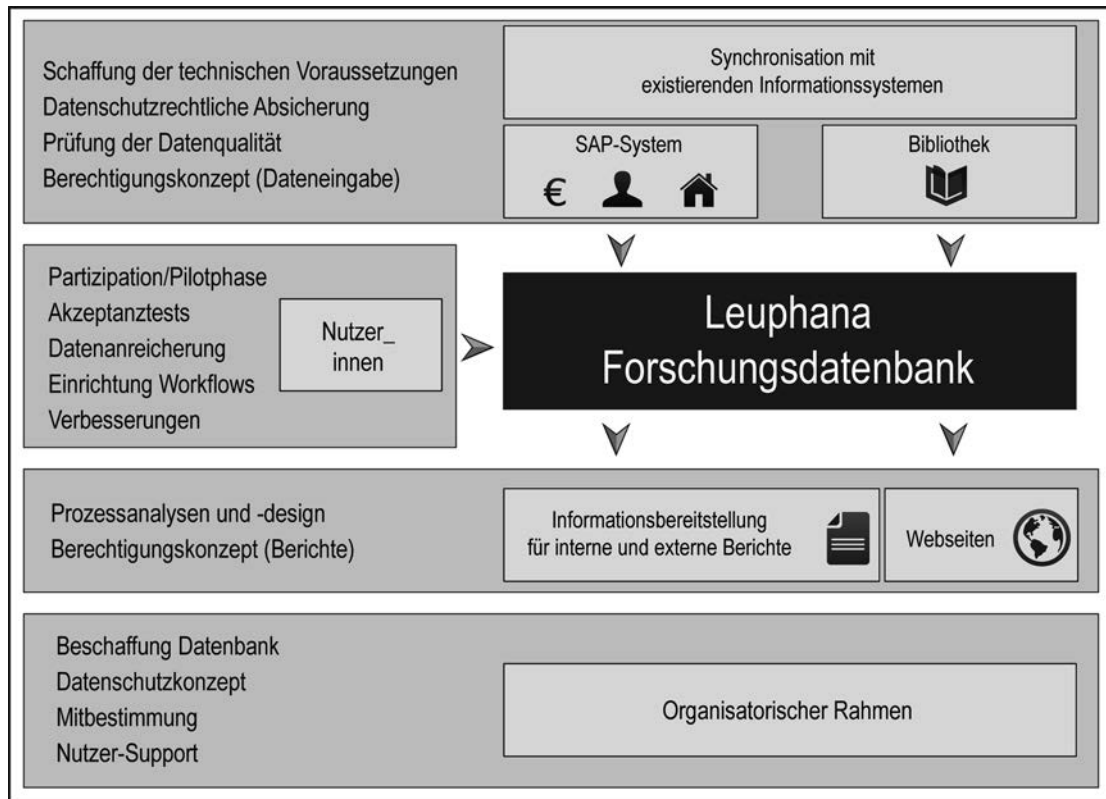


Für unseren Beitrag in diesem Tagungsband berichten wir auszugsweise über den bisherigen Verlauf des Projektes Forschungsdatenbank und ziehen eine vorläufige Bilanz auf Basis der gewonnenen Erfahrungen.

## 2 Das Datenbankprojekt

Eine Übersicht der Aufgabenpakete für die Einführung und Betrieb gibt Abbildung 2.

Abbildung 2: Projekt Forschungsdatenbank



### 2.1 Verwendetes System

Pure ist ein integriertes Forschungsinformationssystem (*Current Research Information System* – CRIS), in dem Expertenprofile, wissenschaftliche Aktivitäten, Projekte und Publikationen gepflegt werden. Diese werden verknüpft mit Datensätzen zu Personen (Autorinnen/Autoren, Projektleitungen etc.) und Organisationseinheiten der Leuphana. Das System übernimmt ausgewählte Informationen aus anderen internen Informationssystemen und wird dezentral durch berechtigte Nutzerinnen und Nutzer gepflegt. Das Datenmodell orientiert sich an dem europäischen Datenformat CERIF (Common European Research Information Format).

Pure wurde vor neun Jahren von dem dänischen Start-up Unternehmen Atira entwickelt. In Dänemark wird Pure auf Beschluss der Hochschulrektoren an allen staatlichen Universitäten eingesetzt und speist die nationale Forscherdatenbank. Neben der Leuphana gehören in Europa außerdem die Universität Helsinki, 17 Universitäten in Großbritannien sowie die flämische Regionalregierung zu den Kunden. Pure wird von Atira gemeinsam mit den beteiligten Forschungseinrichtungen laufend weiterentwickelt und ist dadurch konsequent an den Erfordernissen der Wissenschaft orientiert.

### 2.2 Das Ziel: Eine autorisierte Sammlung von Informationen schaffen

In Pure werden Informationen zu Forschungsleistungen aus mehreren Quellen zusammengeführt. Um Mehrfacheingaben und Unstimmigkeiten zu vermeiden, wurde Pure von Anfang an mit vor-

handenen Informationssystemen der Verwaltung synchronisiert: Ausgewählte Aufträge der Drittmittelverwaltung - SAP FI/CO

- Informationen über personelle Wechsel und organisatorische Veränderungen – SAP HR
- Hochschulbibliographie - PICA

Hochschulmitglieder melden sich aktiv zur Nutzung der Forschungsdatenbank an und geben damit ihre Einwilligung zur Datenverarbeitung. In Pure registrierte Hochschulmitglieder erhalten ein Nutzerkonto, so dass sie die mit ihrer Person verknüpften Daten einsehen und bearbeiten können. Zur Unterstützung können sie auf geschulte Redakteurinnen und Redakteure in den Instituten und Zentren zurückgreifen, die lokal mit der Pflege der Daten betraut werden.

## **2.3 Umgang mit personenbezogenen Daten**

Die Natur eines Forschungsinformationssystems und der Anschaffungsgrund für Pure an der Leuphana Universität legen es nahe, dass nicht nur aggregierte oder öffentliche Informationen in der Forschungsdatenbank gespeichert und verarbeitet werden. Grundsätzlich werden alle in Pure eingetragenen Daten mit Personen und Organisationseinheiten verknüpft. Dies ermöglicht es beispielsweise dem wissenschaftlichen Personal in Pure ein persönliches Archiv der wissenschaftlichen Aktivitäten wie Vorträge, Posten und Ämter, Preise und dergleichen zu pflegen. Pure bietet eine Reihe technischer Funktionalitäten zur Kontrolle der eigenen Daten. Nutzerinnen und Nutzer können die Verwendung der personenbezogenen Einträge in Pure auf mehreren Ebenen steuern:

1. Umfang der Dateneingabe – registriert wird nur, was zur Berichterstattung vorgesehen ist.
2. Sichtbarkeitseinstellung für Einzeleinträge (entscheidet über die Verwendbarkeit des Eintrags):
  - Datenbank – nur für autorisierte Nutzerinnen und Nutzer sichtbar
  - Campus – Darstellung auf Intranet und Campus-Systeme beschränkt
  - Öffentlich – Nutzung für die Öffentlichkeitsarbeit, z.B. Leuphana Webseite

Über neue Einträge und Änderungen der mit ihrer Person verknüpften Daten werden die jeweiligen Nutzerinnen und Nutzer informiert. Dies geschieht auch, wenn die Nutzerinnen und Nutzer das Eintragen von Informationen an andere Personen delegieren.

## **2.4 Regeln für die Auswertungen von Daten**

Um Mehrfachabfragen bei Forschenden und redundante Datenhaltung zu reduzieren, werden die in Pure registrierten Datensätze autorisierten Nutzerinnen und Nutzern für anerkannte und legitime Berichtsverfahren zur Verfügung gestellt. Art und Umfang der Berichte werden bestimmt durch die Beschlüsse, Richtlinien, Ausführungspapiere, Rechtsgrundlagen oder sonstige legitime Anforderungen der jeweiligen Verfahren. Ein Berichtsleitfaden legt fest, dass Nutzerinnen und Nutzer vorab über die Auswertungen zu informieren sind und ausreichend Gelegenheit zur Aktualisierung erhalten müssen. Personen mit Reporting-Rechten und ihre Vorgesetzten werden schriftlich auf die Einhaltung der Richtlinie verpflichtet.

Der Begriff „Bericht“ umfasst alle Darstellungen von Forschungsleistungen auf den verschiedenen Organisationsebenen der Leuphana Universität:

- Öffentlichkeitsarbeit, insbesondere
  - die Darstellung von Forschungsleistungen auf Webseiten
  - Tätigkeitsberichte
- Erstellung aggregierter Statistiken und Auswertungen für die Leitungsebenen der Universität
- Evaluation interner Fördermaßnahmen und –programme



- Leuphana Forschungspreise Erstellung von Forschungs- und Leistungsberichten für organisations- und fachbezogene Evaluationsverfahren

## 2.5 Arbeitsorganisatorische Auswirkungen

Während der Datenbankeinführung entsteht projektbezogene Mehrarbeit a) für die Organisation der Datenbereitstellung, b) für die Entwicklung der Dokumentation und c) für die Organisation der Prozesse zum Betrieb des Systems. Im Forschungsservice der Leuphana, der die Systemeinführung koordiniert, wurde zur Unterstützung der Projektleitung eine temporäre Stelle geschaffen und mit einem Informatiker besetzt. Für den Regelbetrieb entstehen durch die neu eingeführte Qualitätssicherung (Vier-Augen-Prinzip zur Validierung von Projekten und Publikationen) in Bibliothek und Drittmittelbetreuung neue Aufgaben. In den Instituten, Arbeitsgruppen und Zentren übernehmen v.a. die Sekretariate neue Aufgaben in der Datenpflege (Redakteursnetz, s. Abschnitt 4). In der Forschungsberichterstattung wirkt sich der Minderaufwand für das Erstellen von Statistiken und Berichten, bedingt durch bessere Datenverfügbarkeit, entlastend aus.

Tabelle 1 gibt eine Übersicht der beteiligten Organisationseinheiten und ihrer Aufgaben.

*Tabelle 1: Übersicht der internen Schnittstellen und ihrer Aufgaben*

Organisation	Pilotphase	Regelbetrieb (Neue Aufgaben)
Forschungsservice	Projektkoordination, Schulung, Dokumentation, Anschaffung	Datenbankadministration, Nutzersupport, Qualitätssicherung
Universitätsbibliothek	Übertragung der PICA-Hochschulbibliographie, bibliographische Fachkompetenz, (manuelle) Nachpflege der Publikationsdaten	Qualitätssicherung durch Validierung der Publikationseinträge (Vier-Augen-Prinzip)
Personal/Finanzen	Begleitung der Schnittstellenimplementierung, Beratung bei Fragen, Kontrolle der importierten Daten	Qualitätssicherung der laufenden Synchronisation
Universitätsmarketing	Entwicklung der Webintegration von Forschungsdaten aus Pure (u.a. Typo3-Plugin)	Technische Betreuung der Webintegration, inkl. der persönlichen Webseiten und des Online-Forschungskatalogs, Weiterentwicklung von Webanwendungen
Medien- und Informationszentrum	Bereitstellung einer Systemumgebung	Technische Betreuung und Administration, Backups
Wissenschaftliche Bereiche	Mitwirkung einer Pilotnutzergruppe an der Erprobung des Systems	Datenpflege durch Sekretariate bzw. andere Mitarbeiterinnen und Mitarbeiter (Redakteursrolle), Datenerhebung und Reporting mit Pure (Dekanate)
Personal- und Organisationsentwicklung	Projektsupervision und Koordination mit Einführung SAP-HR Abschluss Dienstvereinbarung mit Personalrat	
Datenschutzbeauftragter	Beratung und Begleitung	
Personalrat	Beratung und Begleitung	

### **3 Implementierung und Anbindung in der IT-Landschaft der Universität**

Für den Einsatz an der Leuphana Universität Lüneburg wurde das Pure Datenmodell durch Atria weiterentwickelt und an die lokalen Anforderungen angepasst. Nach der formalen Abnahme erfolgte unter Mitwirken des Medien- und Informationszentrums (MIZ) die grundständige Inbetriebnahme des Forschungsinformationssystems. Dem gesteckten Ziel „vorhandene Daten nutzen“ entsprechend wurden in der Implementierungsphase die vorab identifizierten Datenquellen für die Integration in die Pure Forschungsdatenbank vorbereitet und an die Datenbank angeschlossen (vgl. Abschnitt 2.2).

Aus dem SAP-System wurden die Organisationsstruktur der Leuphana mit 170 Organisationen sowie ein ausgewählter Bestand von rund 650 Drittmittelaufträgen übernommen. Für die Datenübernahme aus dem Bibliothekssystem PICA wurden rund 400 Personendatensätze von Autorinnen und Autoren mit ihren PICA-Normnummern in Pure erfasst. Damit gelang die Übernahme und Zuordnung von ca. 6.000 Publikationseinträgen der Hochschulbibliographie zu 95%. Die aus den verbundenen Systemen übernommenen Daten wurden durch vorab definierte Qualitätsstandards bewertet und der jeweilige Nachbearbeitungsbedarf formuliert.

Nachbearbeitung war insbesondere für Projekteinträge nötig, um die Datensätze für die Forschungsdatenbank zu erschließen, da die importierten SAP-Datensätze nur Finanzinformationen, eine Organisationszuordnung sowie eine für finanztechnische Zwecke abgekürzte Vorhabensbezeichnung enthielten.

Auch der Bestand an Publikationseinträgen in der Hochschulbibliographie musste nach der Übertragung in die Forschungsdatenbank nochmals nachbearbeitet werden, um Übertragungsfehler zu korrigieren und die Typisierung der Einträge als Monographie, Sammelwerke, Arbeits- oder Projektberichte anzupassen. Zuvor hatten bereits umfassende Arbeiten am Datenbestand in PICA stattgefunden, um die Datenübertragung vorzubereiten (vgl. Abschnitt 3.1).

#### **3.1 Hochschulbibliographie: Von PICA zu Pure**

Vor der Einführung von Pure als zentrale Publikationsdatenbank an der Leuphana Universität Lüneburg wurden an ihren Vorläufereinrichtungen bereits verschiedene Systeme für Hochschulbibliographien verwendet. Dabei handelte es sich um Eigenentwicklungen und OpenSource-Systeme, die nicht mit bibliographischen Standardformaten arbeiteten.

Bei Aufkommen des Wunsches nach einem einheitlichen System zur Speicherung von Hochschulschriften im Jahr 2007 wurde die Schaffung eines einheitlichen Datenmodells als wesentliche Voraussetzung für die Konsolidierung der unterschiedlichen Datenbestände erkannt. Dieses Ziel sollte über die Nutzung des lokalen Bibliothekssystems PICA erreicht werden. In PICA werden die Datensätze mithilfe von Normnummern und sogenannten Kategorien anstelle von Eingabefeldern organisiert. Damit wird das System flexibel gehalten und ist für die Erfassung eines breiten Spektrums an Medien geeignet. Die Standardisierung erfolgt durch ein komplexes Regelwerk für die Katalogisierung, das auch Spielraum für Auslegungen und lokale Nutzung einzelner Kategorien zulässt.

In 2007 wurden die Mitglieder der Hochschule aufgerufen, alle Publikationen ab dem Jahr 2000 an die Bibliothek zur Eingabe zu geben. Nach der Eingabe wurde ab August 2008 die Hochschulbibliographie der Universität in PICA veröffentlicht. Hauptanwendungszweck war die Präsentation der Publikationen für lokale Studierende und der Nachweis im Gesamtverbundkatalog. Darüber hinaus wurden die Daten auch zur Erstellung fakultätsbezogener Statistiken, u.a. für die leistungsbezogene Mittelvergabe und die Verleihung der Forschungspreise der Universität genutzt. Für diesen Zweck wurden zusätzlich benötigte Informationen wie Fakultätszugehörigkeit, Fächer, Dokumententypen, Verlagsinformationen, Rolle der Autoren und Medienformen in lokale Schlag-

wortkategorien des Bibliothekskatalogs aufgenommen. Der Export erfolgte über einen Webservice, der die Daten in einem standardisierten Format ausgab. Dieses konnte mit gängigen Tabellenkalkulationsprogrammen weiterverarbeitet werden.

Nach der Entscheidung der Universitätsleitung für die Einführung eines integrierten Forschungsinformationssystems wurde überlegt, die PICA-Hochschulbibliographie als Datenquelle für Pure zu verwenden, um den etablierten Prozess der Meldung und Erfassung von Forschungsschriften zu erhalten und von Anreicherungen der Datenbestände im Gesamtverbundkatalog profitieren zu können. Dafür wurde von Atira in Zusammenarbeit mit der Bibliothek eine Schnittstelle programmiert, um das PICA+ Format in das von Pure verwendete CERIF-basierte Datenmodell zu übersetzen. Über diese Schnittstelle sollte durch einen kontinuierlichen Synchronisationsprozess ein Abgleich von Neuerfassungen und Änderungen in der PICA-Datenbank mit den Daten in Pure durchgeführt werden.

Zudem wurde der auf 7.000 Literaturnachweise angewachsene Bestand der Hochschulbibliographie in PICA grundlegend überprüft und überarbeitet. Unter anderem wurden doppelt aufgenommene elektronische Ausgaben von Printwerken entfernt und mit den Schwestereinträgen zusammengeführt. Neben den Korrekturen wurden zahlreiche Attribute nachgepflegt, um die reibungslose Übertragung in die Forschungsdatenbank abzusichern.

Zur Übernahme der Publikationsdaten haben der Hersteller Atira und das Schnittstellenteam der Bibliothek erheblichen Aufwand investiert, um die Belegung der PICA-Kategorien zu dokumentieren und Regeln für die Datenübernahme zu entwickeln. Für viele Sonderfälle konnte nur unter zu Hilfenahme bibliothekarischer Fachkompetenz eine Lösung entwickelt werden.

Nach der initialen Übertragung der rund 6.000 verbliebenen Publikationseinträge von PICA nach Pure machten sich grundlegende Inkompatibilitäten zwischen dem bibliographischen Datenformat PICA+ und dem relationalen, an CERIF orientierten Datenmodell von Pure bemerkbar. Sie führten zu manuellen Nacharbeiten in erheblichem Umfang. Schlussendlich wurde deutlich, dass der Synchronisationsprozess den Anforderungen der Anwender nicht gerecht werden konnte.

Insbesondere zwei Attribute konnten über die Integration nicht aufgelöst werden. Eine ist die PICA-Kategorie "Umfang" einer Publikation, in der sowohl Seitenzahl, Seitenumfang als auch Dateigröße oder Gewicht eines Druckwerks erfasst werden. Für 18 Standardbelegungen dieser Kategorie wurden Integrationsregeln entwickelt. Trotzdem konnte eine korrekte Identifikation der Seitenangabe durch die Pure Integration nicht sicher gewährleistet werden. Die zweite Besonderheit betrifft Sammelwerke, bei denen in PICA nicht nur die Herausgeber, sondern auch der erste beitragende Autor erfasst sein können. Dies wird von den Verbundbibliotheken nicht einheitlich gehandhabt, und leider ist der Status der Personen anhand der Kategorienbelegung nicht eindeutig entscheidbar. Eine Berichtigung erfolgte manuell in Pure, war aber im Hinblick auf eine laufende Synchronisation problematisch.

Da für eine integrationsfähige Erfassung der Literaturnachweise in PICA letztendlich in beiden Systemen erheblicher Aufwand betrieben werden muss, entschied die Steuerungsgruppe sich für eine Ersterfassung mit allen für Statistiken notwendigen Attributen in Pure. Auf den Nachweis der Leuphana Hochschulbibliographie im Gesamtverbundkatalog (GVK) wird aktuell verzichtet, da die direkte Einbindung lokaler Systeme wie Pure in die GVK-Suche technisch nicht möglich ist.

### **3.2 Projekte, Organisationen und Personen – Anbindung an SAP**

Für die Anbindung und Belieferung des Systems Pure mit personalwirtschaftlichen und projektspezifischen Daten aus dem SAP-System wurde das Niedersächsische Hochschulkompetenzzentrum für SAP (CCC) in das Projekt mit eingebunden. Das CCC unterstützt und betreut 18 niedersächsische Hochschulen beim Einsatz und der Weiterentwicklung des SAP-Systems in den

Bereichen Rechnungswesen, Personalwesen, Logistik, Business Warehouse und Folders Management (elektronische Drittmittelakte).

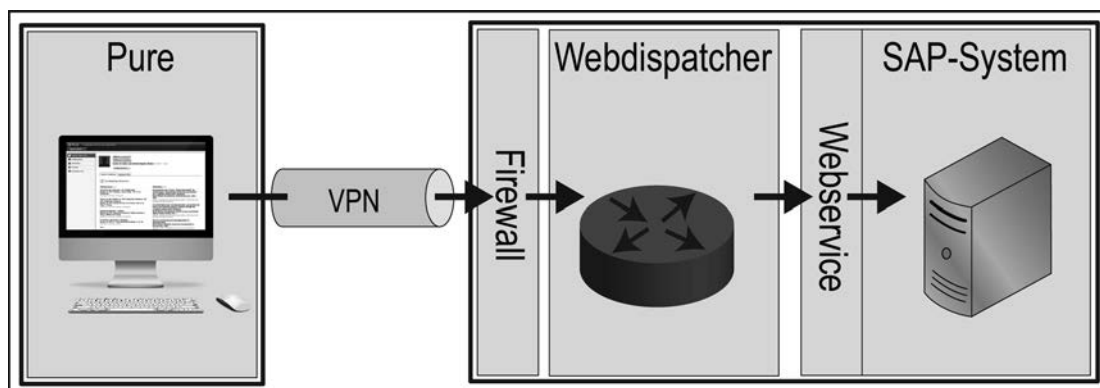
Im gemeinsam mit Universität und Datenbankanbieter entwickelten Fachkonzept wurden die zu übergebenen Datenstrukturen abgestimmt und die technische Infrastruktur festgelegt. Der Betrieb von SAP ist aufseiten der Leuphana durch eine Dienstvereinbarung geregelt. Zur Nutzung ausgewählter SAP-Daten für den Anwendungsfall Pure wurde an der Leuphana eine Ergänzung zur Dienstvereinbarung abgeschlossen.

Die Anforderungen aus dem Projekt Pure wurden auf SAP-Seite im CCC über die Bereitstellung zweier Webservices realisiert, über welche ausschließlich die von Pure benötigten Attribute abgerufen werden. Die Datenübermittlung erfolgt über eine gesicherte VPN-Verbindung. Eine zusätzliche Absicherung des SAP-Systems wurde durch den Einsatz des SAP Webdispatchers mit URL-Filterung und einer restriktiven Einschränkung der Berechtigungen des SAP-Nutzers, über den der Zugriff vorgenommen wird, erreicht. Eine schematische Darstellung des Zugriffs ist in Abbildung 3 dargestellt.

Der Webservice wurde so gestaltet, dass die spezifizierten Informationen entweder für eine einzelne Person oder aber für alle Personen übertragen werden können. Durch die Aufteilung der Datenlieferungen in diese drei Tabellen konnte die häufig vorkommende Zuordnung einer Person zu verschiedenen Organisationseinheiten transparent bereitgestellt werden.

Für die Bereitstellung von Projektinformationen wurde ein weiterer Webservice entwickelt, der zum Stichtag ausgewählte Stammdaten zu den Drittmittelprojekten aus dem Modul SAP-PSM selektiert und diese ebenfalls als Tabelle übergibt.

*Abbildung 3: Infrastruktur der Pure-SAP-Schnittstelle*



Der Webservice, der die Daten zum Stichtag aus dem Bereich Personalwirtschaft bereitstellt (SAP HR), liefert als Ergebnis drei Tabellen mit folgenden Inhalten an Pure zurück:

- eine Tabelle mit der im SAP-HCM-System hinterlegten hierarchischen Organisationsstruktur
- eine Tabelle mit den Zuordnungen von Personen zu Organisationseinheiten und
- eine Tabelle mit ausgewählten Personalinformationen.

Um es den SAP-Key Usern zu ermöglichen, die über die Webservices bereitgestellten Daten zu plausibilisieren und zu kontrollieren, wurde eine Transaktion für den SAP GUI entwickelt, mit der auf die selben Daten zugegriffen werden kann, wie sie auch über die Webservices bereitgestellt werden.

Zur Implementierung des Webserviceaufrufs in Pure wurden den Entwicklern von Atira entspre-

chende WSDL-Dateien bereitgestellt. Hier zeigte sich auch der Vorteil eines webserviceorientierten Schnittstellenansatzes, da bereits durch die Architektur eindeutige Verantwortungsbereiche abgegrenzt werden: Zum einen die Bereitstellung der Webservices durch das CCC und zum anderen die Implementierung des Webserviceaufrufs und der weiteren Verarbeitung in Pure durch den Datenbankanbieter Atira. In der Testphase wurden noch kleinere Anpassungen und Weiterentwicklungen an der Schnittstelle vorgenommen, nach positivem Testabschluss schloss sich die Produktivsetzung der Schnittstelle an. Die erfolgreiche Implementierung der Schnittstelle zeigt sich vor allem auch daran, dass es im produktiven Betrieb bisher keine Probleme oder Störungen gegeben hat.

### **3.3 Rechner und Routen: Einbettung im Medien- und Informationszentrum**

Das Hosting für das Forschungsinformationssystem der Hochschule erfolgt im Leuphana Medien- und Informationszentrum (MIZ). Server-Maintenance und Updates der Anwendung erfolgen durch den Datenbankanbieter Atira, Hardware, Betriebssystem und Backups durch das MIZ. Teile der Daten werden aus dem Niedersächsischen Hochschulkompetenzzentrum für SAP (CCC) bezogen. Mitarbeiterinnen und Mitarbeiter der Leuphana können über eigene Zugänge mit der Datenbank arbeiten. Der Pure Server ist unter anderem durch eine Firewall umfassend geschützt.

Das Miteinander aller am Betrieb Beteiligten koordiniert ein Regelwerk, die “Pure Policy”, in der die Zuständigkeiten, die Zuverlässigkeit, die Sicherheit, die Performanz und die Wiederherstellbarkeit geregelt werden. Enthalten sind die Punkte:

- Redundanzsystem für Updates und Bugfixes sowie als Fallback für Notfälle
- nächtliches Backup mit ADSM Tivoli Storage Manager
- Monitoring mit Nagios (Server, Dienst und Datenbank)
- Datenschutzbestimmungen
- Wartungsfenster und Informationspolitik bei Störungen
- Physikalischer Schutz
- Firewall-Schutz und Administrationszugänge

#### **3.3.1 Überführung in eine virtuelle Infrastruktur**

Als Basissystem wurde ein Debian System mit Festplattenspiegelung installiert, der Datennetzanschluss erfolgt über die Cisco Catalyst- und ASA-Infrastruktur. Zur IT-Infrastruktur Strategie des MIZ gehört es, dass Pure in 2012 in eine vmWare-Virtualisierungsumgebung überführt wird. Bei Ausfall eines vmWare-Gastgebers werden im Hochverfügbarkeitsmodus virtualisierte Server automatisch migriert, Pure wird also ohne spürbare Unterbrechung für die Nutzer auf anderer Hardware weiter ausgeführt. Die Datenspeicherung erfolgt in einem redundanten SAN-Speichersystem, an das alle Gastgeber angeschlossen sind. Das MIZ fördert mit dieser IT-Infrastruktur-Strategie die nachhaltige Entwicklung der Universität, indem sie die optimale Nutzung der Ressourcen ermöglicht und gleichzeitig die Ausfallsicherheit des ganzen Systems signifikant erhöht.

## **4 Going Live: Einführung in der Universität**

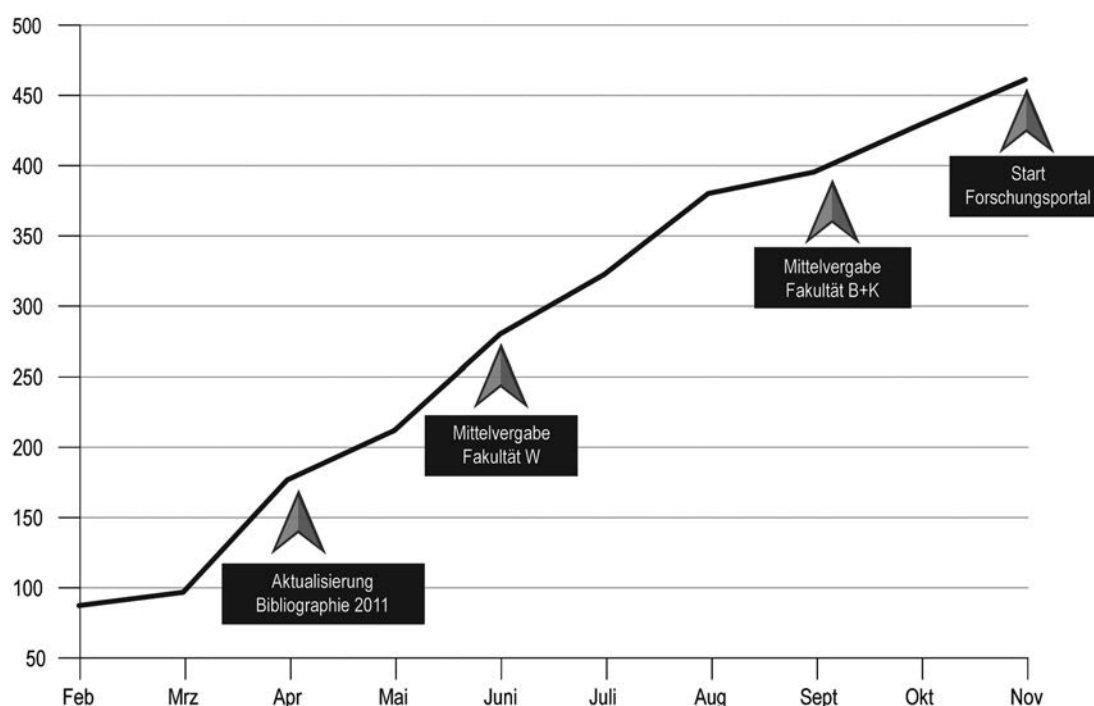
Als Auftakt für die Einführung der neuen Forschungsdatenbank wurde Anfang 2011 die jährliche Aktualisierung der Hochschulbibliographie gewählt. Dieser etablierte und weithin akzeptierte Prozess sieht vor, dass die Mitglieder der Universität ihre Publikationen zur Erfassung in der Universitätsbibliothek einreichen. Dieses Verfahren wurde in Grundzügen beibehalten, so dass sich die Neuerung zunächst auf den Systemwechsel von PICA nach Pure beschränkte. Die Autorinnen und Autoren wurden informiert, dass sie bereits in PICA erfassten Publikationen in das neue System Pure überführt wurden und erhielten die Gelegenheit, sich zur Datenbank anzumelden, um die Einträge prüfen und falls nötig korrigieren zu können.



Im zweiten Quartal 2011 begann die Fakultät Wirtschaft mit der Vorbereitung der internen Mittelvergabe. Für den Indikator Publikationen sollten Einträge der Hochschulbibliographie zentral ausgewertet werden. Forschungsservice und Dekanat organisierten gemeinsam eine Informationskampagne zur Durchsicht der Datenbestände und die Meldung noch fehlender Publikationen. Diese Kampagne führte noch einmal zum Anstieg der Pure Anmeldungen, zur Vervollständigung des Redakteursnetzes und zu einer Verbesserung des Datenbestands in Pure aufgrund zahlreicher Korrekturen, die die Nutzerinnen und Nutzer vornahmen.

Die beiden Auftaktprojekte führten innerhalb von fünf Monaten zu einer Steigerung der Nutzerzahlen von 93 (Pilotphase) auf 370. Im Lauf des Jahres sorgten weitere Anwendungsprojekte für einen kontinuierlichen Anstieg der Anmeldungen (s. Abbildung 4).

*Abbildung 4: Regelmäßige Kampagnen zur Nutzung von Pure lassen die Anmeldezahlen stetig steigen. Dargestellt ist die Entwicklung der Nutzerzahlen im Jahr 2011 mit den Meilensteinen Hochschulbibliographie, Mittelvergaben und Start Forschungsportal.*



Mit der Einführung von Pure begann der systematische Aufbau eines Redakteursnetzes in der Universität. Die Redakteursrolle in Pure ermöglicht die Vergabe von Berechtigungen zur Datenpflege in Instituten, Arbeitsgruppen und Zentren. Diese Rolle ist in der Regel bei den Sekretariaten angesiedelt, wobei die Unterstützung durch einen Wissenschaftler oder eine Wissenschaftlerin als gute Praxis empfohlen wird. Das Redakteursnetz umfasst ein Jahr nach Einführung knapp 80 Personen und deckt 74 % der Professuren ab. Zu den Unterstützungsmaßnahmen für Redakteure zählen eine individuelle Einführung am Arbeitsplatz, wöchentliche Redakteurstreffen mit kurzen Tutorials (30 min) und anschließenden betreuten Arbeiten in der Gruppe.

#### 4.1 Anwendungsprojekte – Nutzung des Datenbestands

In Vorbereitung der Datenbankeinführung war bereits eine Liste von knapp zwei Dutzend internen Prozessen erstellt worden, die regelmäßig Forschungsinformationen benötigen – von der Mittelvergabe über das Forschungsmarketing bis hin zu Tätigkeitsberichten und externen Evaluationen. Mithilfe der Forschungsdatenbank Pure soll eine autorisierte Sammlung entstehen, aus der

die benötigten Informationen entweder von den beteiligten Individuen selbst oder bei entsprechender Legitimation von den zuständigen Organisatoren abgerufen werden können.

Anwendungsprojekte im ersten Jahren waren neben den Mittelvergabeverfahren der Fakultäten z.B. die Bewerbungen für die Leuphana Forschungspreise. Für den vom Universitätsmarketing betreuten Internetauftritt wurde ein Online-Forschungskatalog entwickelt sowie ein Typo3-Plugin zur Integration von Publikations- und Projektlisten auf die persönlichen Webseiten der Wissenschaftlerinnen und Wissenschaftler.

Die Erstellung eines Querschnittsberichts für die Evaluation der Genderforschung in Niedersachsen sei hier exemplarisch vorgestellt, um den Mehrwert des Instruments Forschungsdatenbank zu illustrieren:

Die Wissenschaftliche Kommission Niedersachsen (WKN) begann in 2011 mit einer Evaluation der Genderforschung in Niedersachsen. In diesem Zuge waren die zur Genderforschung beitragenden Forschungseinheiten zu benennen und Fachgebiete jenseits der Geistes- und Gesellschaftswissenschaften einzubeziehen. Dies bedeutete faktisch eine Datenerhebung in der Universität über alle Organisationseinheiten hinweg.

Zunächst wurden Forschungseinheiten und Personen identifiziert, die in den letzten fünf Jahren Forschungsarbeiten zu Geschlechterfragen durchgeführt haben. Dieser Personenkreis wurde gebeten, relevante Publikationen, Projekte, Beratungs- und Beiratstätigkeiten sowie ihre Expertenprofile in der Forschungsdatenbank zu vervollständigen und sie dem Fachgebiet „Gender-Diversity“ zuzuordnen. Auf diese Weise wurden in einem sechswöchigen Zeitraum über 400 Publikationen sowie mehr als 60 Projekte und Expertenprofile erschlossen.

Für die Ausgabe der Informationen wurde mithilfe des Pure Berichtsmoduls eine spezielle Berichtsvorlage konfiguriert. Filterkriterium war wiederum das Fachgebiet „Gender-Diversity“. So konnten bereits während der Dateneingabe wöchentliche Zwischenbilanzen erstellt und Korrekturfassungen an die teilnehmenden Wissenschaftlerinnen und Wissenschaftler versandt werden. Durch die standardisierten Berichtsvorlagen sind diese Berichtsversionen gut kontrollierbar (Datum und Erzeuger werden jeweils vermerkt).

Vorteile des Instruments Forschungsdatenbank gegenüber der herkömmlichen Datenerhebung waren zum einen die strukturierte Erfassung über die Eingabeformulare der Datenbank, zum anderen entfiel der sonst übliche Aufwand zur Vereinheitlichung von Publikationslisten usw. Die Erzeugung einer einheitlich formatierten „Gender-Diversity-Bibliographie“ aus der zugehörigen Berichtsvorlage im Word-Format ist beispielsweise innerhalb einer Minute möglich. Darüber hinaus stehen die so erschlossenen Einträge unmittelbar für alternative Anwendungen zur Verfügung, z.B. für die Fachgebietssuche im Online-Forschungskatalog<sup>1</sup>. Die Redaktion des „Gender-Diversity Portals“ plant derzeit, auf dieser Datenbasis die Leistungen seines Forschungsnetzwerks besser zu präsentieren.

## **5 Lessons learned und Ausblick**

### **5.1 Fazit nach dem ersten Jahr der Einführung**

In der Vorbereitungsphase war insbesondere die Altdatenübernahme zeitintensiv. Beispielsweise war der Aufwand für die Übernahme der PICA-Daten trotz sorgfältiger Vorbereitung kaum planbar, und die Korrekturen am Datenbestand führten zu erheblichen Zeitverzögerungen. Im Nachgang bildet die komplette Durchsicht und Korrektur der Hochschulbibliographie eigentlich ein eigenes Teilprojekt im Gesamtkonzept. Weitere Nebenprodukte der Datenbankeinführung

---

<sup>1</sup> <http://www.leuphana.de/forschungsdatenbank> (nur Einträge, die in Pure als „öffentlich“ gekennzeichnet sind)

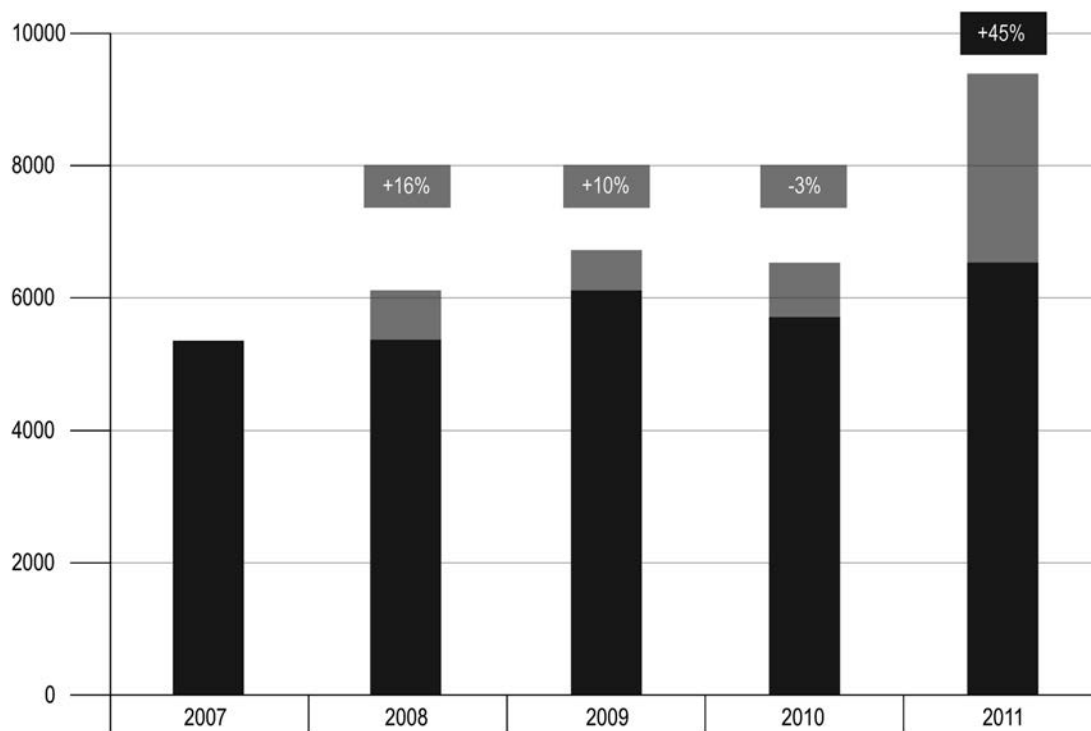


waren die Einrichtung der ersten Webservice-Schnittstelle im niedersächsischen SAP-System des CCC, der Aufbau eines internen Verzeichnisses der Mittelgeber und Kooperationspartner, die Weiterentwicklung des Fachgebietskatalogs aus der Hochschulbibliographie und die Verschlagwortung des Projektbestands.

Wie steht es um die Zielerreichung? Mit der Einführung eines integrierten Forschungsinformationssystems wurden vorrangig zwei Ziele verfolgt: Die Schaffung einer einheitlichen, zwischen Wissenschaft und Verwaltung abgestimmten Datenbasis für die Forschungsberichterstattung und die Erzeugung von Mehrwerten für die Wissenschaftlerinnen und Wissenschaftler durch verbesserte Verfügbarkeit der Informationen für eigene Prozesse.

Nach dem ersten Jahr des Produktivbetriebs kann konstatiert werden, dass die Schaffung einer soliden Datenbasis und die Motivation der Datenbanknutzung sich gegenseitig bedingen: Durch die systematische Korrektur und Anreicherung der übernommenen Publikationen im Pure Redaktionsteam konnte die Datenbank relativ schnell ersten Nutzen für Mittelvergabeverfahren und andere Prozesse generieren. Den Wissenschaftlerinnen und Wissenschaftlern wurden die bereits vorhandenen Daten zur Verfügung gestellt, für umfangreiche Nacherfassungen steht ein Eingabeservice zur Verfügung.

*Abbildung 5: Meldung von Publikationen für die Leuphana Hochschulbibliographie. Der konsequente Einsatz von Pure für die Forschungsberichterstattung hat die Meldebereitschaft 2011 stark erhöht.*



Wie Abbildung 5 zeigt, ist im Jahr der Pure Einführung beispielsweise die Meldebereitschaft für die Hochschulbibliographie stark gestiegen. Wurden in den Vorjahren 800-1.000 Publikationen an die Bibliothek gemeldet, so belief sich der Zuwachs 2011 auf 3.000 neue Einträge. Die Forschungsdatenbank hat sich hier schnell als zentrales Instrument etabliert, nicht zuletzt wegen der besseren Zugänglichkeit für die Beteiligten und einer Selbststeuerung der Einträge.

Die gute Abdeckung im Bereich Publikationen und die deutlich einfacheren Auswertungsmöglichkeiten überzeugten Dekane, ihre zuvor praktizierten Datenerhebungsverfahren für die Mittelvergabe zugunsten einer Nutzung der Forschungsdatenbank aufzugeben. Dies ist nicht für jeden

Prozess selbstverständlich. Für einzelne Kolleginnen und Kollegen erscheint es zunächst einfacher, bei den bewährten Word- und Excel-Listen zu bleiben und diese Jahr für Jahr in den gewohnten Umlaufverfahren zu aktualisieren. Demgegenüber erfordert die Nutzung des integrierten Forschungsinformationssystems für komplexere Berichte eine gewisse Systemkompetenz und eine gute Kenntnis der Daten. Insbesondere im ersten Jahr entsteht teils erheblicher zusätzlicher Aufwand durch Prüfungen und Korrekturen der Einträge sowie durch die Schulungen bzw. Beratungen der Nutzerinnen und Nutzer.

In der Erstellung von Berichten sind jedoch bereits deutliche Effizienzgewinne realisierbar. Da jede Korrektur die Datensätze dauerhaft verbessert und die Nutzerkompetenz laufend steigt, werden die Vereinfachungen beim zweiten Mal noch deutlicher ausfallen.

Das Pure Redaktionsteam arbeitet daran, für viele der vorhandenen Berichtsanlässe Effizienzgewinne zu realisieren. Dies erfordert eine intensive inhaltliche und prozessuale Begleitung. Daher investiert die Leuphana auch im zweiten Jahr nach der Einführung in zusätzliche Kapazitäten für die Forschungsberichterstattung.

Insgesamt beobachten wir, dass die Idee des integrierten Forschungsinformationssystems bereitwillig aufgegriffen wird. Allerdings begegnen wir auch überhöhten Erwartungen an die à la carte Verfügbarkeit der Einträge aus dem Forschungsinformationssystem. Datenqualität wie Nutzerkompetenz werden ebenfalls gern überschätzt. Die Botschaft, dass die an Forschungsberichten interessierten Einrichtungen sich aktiv an der Motivation der Wissenschaftlerinnen und Wissenschaftler und der Meldung von Korrekturbedarfen beteiligten müssen, ist ein wichtiger Baustein in der Kommunikation.

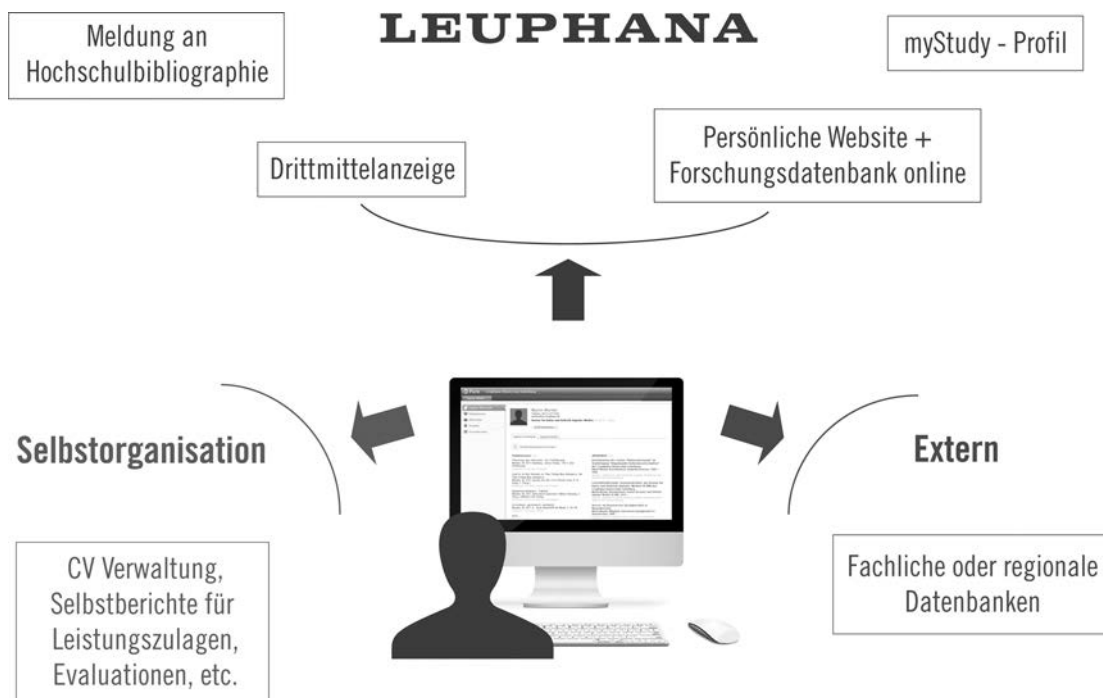
## **5.2 Aus unseren Erfahrungen: Anregungen für die Entwicklung von universitären Forschungsinformationssystemen in Deutschland**

Eines der wichtigen Ziele für die Leuphana war es, die über viele Stellen und Systeme verteilten Informationen zu sichten und aus ihnen eine Datenbasis für die Forschungsberichterstattung zu schaffen. Dazu gehören nicht zuletzt Informationen, die exklusiv oder mit höherer Qualität in externen Datenbanken gepflegt werden. Wichtig war hier vor allem der Abgleich mit den Daten der Förderer – z.B. mit den Datenbanken GEPRIS (DFG) und CORDIS (europäische Forschungsrahmenprogramme). Die Übernahme und der Abgleich erfolgten manuell – hier wäre es insbesondere für größere Universitäten von Vorteil, wenn die Daten in elektronischer Form ausgetauscht werden könnten.

Inspiration für Services, die beim Aufbau lokaler Systeme hilfreich wären, bietet ein Blick in andere Länder. Beispielsweise nutzen die lokalen Systeme in Norwegen ein nationales Verzeichnis von Universitäten und Forschungseinrichtungen als Datenquelle. So können die institutionellen Kooperationsbeziehungen auf einer einheitlichen Datenbasis analysiert werden. In Dänemark können lokale Forschungsinformationssysteme ein Verzeichnis mit mehreren Tausend Zeitschriftentiteln nutzen, und Vergleichsdaten aus der Forschungsevaluation stehen für Benchmarking zur Verfügung.

Als Universität verfolgen wir das Ziel, Informationen über die Organisation zunächst vor Ort zu erfassen und die Forschungsdatenbank als lokalen „One stop shop“ für die Wissenschaftlerinnen und Wissenschaftler zu etablieren (Abbildung 6). Erstes Projekt für einen Export von Forschungsdaten ist die Repräsentation von Profilen der Forschungseinheiten im Portal „Forschung in Niedersachsen“.

Abbildung 6: Vision - Forschungsdatenbank als lokaler „One stop shop“ für die Wissenschaftlerinnen und Wissenschaftler



Für die zukünftige Entwicklung wünschen wir uns vor allem eine bessere Wiederverwendbarkeit vorhandener Daten und eine bessere Interoperabilität für institutionelle Systeme und Fachdatenbanken – vor allem um Barrieren für Wissenschaftlerinnen und Wissenschaftler zu senken. Sie kritisieren zu Recht die Vielfalt der Abfragen, die seitens der Arbeitgeber, der Fachgemeinschaften und der Förderer an sie gestellt werden.

Der europäische Datenstandard CERIF bietet für die (Weiter-)entwicklung von Informationssystemen wertvolle Anregungen. CERIF wird von engagierten Wissenschaftlerinnen und Wissenschaftlern sowie kommerziellen Datenbankanbietern im Verband euroCRIS entwickelt.<sup>2</sup> Wertvoll ist auch der Austausch mit den vielfältigen Mitgliedseinrichtungen, die von renommierten Forschungseinrichtungen über Bibliotheken bis hin zu Forschungsförderern reichen.

<sup>2</sup> [www.eurocris.org](http://www.eurocris.org)

## Forschungsdatenbank der Freien Universität Berlin

### 1 Die Freie Universität Berlin in Zahlen

An der 1948 gegründeten Freien Universität Berlin<sup>1</sup> sind derzeit 32.373 Studierende immatrikuliert. 2010 belief sich die Anzahl der Absolventinnen und Absolventen auf 4.212, davon fanden 616 Promotionen statt. Inklusive Juniorprofessuren waren 2010 392 Professorinnen und Professoren, 733 wissenschaftliche Mitarbeiterinnen und Mitarbeiter in 14 Fachbereichen und Zentralinstituten sowie 1.884 sonstige Mitarbeiterinnen und Mitarbeiter tätig. Zusätzlich wurden 846 Wissenschaftlerinnen und Wissenschaftler aus Drittmitteln finanziert.

Seit 2007 wird die Freie Universität Berlin mit ihrem Zukunftskonzept einer „Internationalen Netzwerkuniversität“ durch die Exzellenzinitiative des Bundes und der Länder gefördert – unter anderem mit dem Ziel, neue, zukunftsorientierte und fachgebietsübergreifende Forschungsfelder zu identifizieren und die Forschung auf internationaler Ebene wettbewerbsfähig zu etablieren.

In den vergangenen Jahren ist die Freie Universität Berlin von einer Reduzierung der Finanzmittel durch Haushaltskürzungen und Sparvorgaben des Landes Berlin und damit einhergehend von einer Verringerung der Studienplätze und des Personals betroffen.<sup>2</sup> So ging zum Beispiel die Zahl der Professuren in den letzten 15 Jahren um nahezu die Hälfte zurück.<sup>3</sup> Dennoch lässt sich bei den Drittmiteleinwerbungen eine kontinuierliche Steigerung – unabhängig von der Förderung durch Mittel der Exzellenzinitiative des Bundes und der Länder – verzeichnen. Die Höhe der Drittmiteleinnahmen und -ausgaben hat sich von 2002 bis 2010 mehr als verdoppelt.<sup>4</sup>

### 2 Forschungsinformationen an der Freien Universität Berlin: Ziele und Anforderungen

Die ansteigenden Drittmiteleinwerbungen sind folglich von einer Zunahme an Drittmittelprojektdaten begleitet. Zudem bestehen durch das Berichtswesen für unterschiedliche Nutzerkreise,<sup>5</sup> durch Informationsbedarf an Forschungsprojektdaten für das universitäre Management wie auch durch die Umstrukturierung der institutionellen Forschungslandschaft hinsichtlich einer fächerunabhängigen Zusammenarbeit von Wissenschaftlerinnen und Wissenschaftlern spezifische Anforderungen an ein IT-gestütztes Forschungsinformationssystem. Die Veränderung der universitären Forschungsstruktur geschieht zum Beispiel im Rahmen des Zukunftskonzepts der Exzellenzinitiative des Bundes und der Länder, zu der unter anderem die Bildung von interdisziplinären, nationalen und internationalen Forschungsverbünden und die Generierung neuer Forschungsfelder von gesellschaftlich, politisch und wirtschaftlicher Relevanz gehören, in denen fach- und institutionenübergreifend geforscht wird. Zur Identifizierung zukünftiger Forschungspotenziale wiederum ist die Beschaffung qualitativer Forschungsprojektdaten notwendig. Informationen zum universitären Forschungsprofil sind dabei zielgruppen- und themenorientiert, schnell und nutzerfreundlich aufzubereiten. Auch die Veränderung der letzten Jahre hinsichtlich wissenschaftli-

---

1 Siehe auch <http://www.fu-berlin.de> und <http://www.fu-berlin.de/forschung/>.

2 Zum Beispiel 2003: 38.868 Studierende, 2005: 36.428, 2007: 33.879, 2009: 32.539.

3 Zum Beispiel 1996: 792 Professuren, 2010: 392.

4 Drittmiteleinnahmen 2002: 51.262 Mio. €, 2010: 111.459 Mio. € (Steigerung von 117,43 %); Drittmittelausgaben 2002: 42.508 Mio. €, 2010: 106.290 Mio. € (Steigerung von 150,04 %).

5 Berichtswesen für den internen und externen Nutzerkreis, wie zum Beispiel Forschungs-, EU-, Leistungs-, Rechenschafts- oder Kuratoriumsbericht, Forschungsprojektdaten für den Wissenschaftsrat etc.

cher Leistungsbemessung erfordert bestimmte Auswertungsmechanismen innerhalb eines Forschungsinformationssystems, zu dem nicht nur Drittmittelprojekte gehören. So müssen Parameter für interne Mittelzuweisungen und Zielvereinbarungen mit den Fachbereichen recherchierbar sein, zu denen zum Beispiel – bezogen auf die Forschungsdatenbank – Forschungsprojekt-, Publikationsdaten und statistische vergleichende Zeitreihenanalysen zählen. Um hierfür eine gute Vergleichbarkeit zu garantieren, müssen die Daten möglichst standardisiert erfasst sein.

Ziel der Forschungsdatenbank<sup>6</sup> der Freien Universität Berlin ist es, auf administrativer Ebene kostensparende Optimierungseffekte im Arbeitsablauf durch die integrative und effiziente Verwaltung der Drittmittel- wie auch der Forschungsvorhaben außerhalb von Drittmittelförderungen zu erzielen, auf inhaltlicher Ebene formale und qualitative Forschungsinformationen als Basis für das Berichtswesen zu präsentieren und unterschiedlichste Rechercheanforderungen, für universitäre leitungsbezogene Entscheidungen, für die öffentliche Dokumentation der universitären Forschungslandschaft und -leistung sowie für statistische Auswertungen anzubieten.

### **3 Integration der Forschungsdatenbank in die Software SAP**

Bis zur Implementierung der Forschungsdatenbank in SAP im Jahr 2005 existierten für das administrative und inhaltliche Management drittmittelgeförderter Forschungsprojekte heterogene und zum Teil unkompatible Datenbanksysteme. Die Folgen waren eine Redundanz und Inkonsistenz der Daten, deren fehlende Aktualität, Sicherheit und Qualitätskontrolle, mehrfache Arbeitsabläufe und nicht genau kontrollierbare Zugriffe auf die Daten, da es kein spezifisches Berechtigungskonzept gab. Die SAP-Forschungsdatenbank löste somit die technisch veralteten IT-Systeme ab und optimierte zugleich die internen Geschäftsabläufe. Die Entscheidung, für die Forschungsdatenbank SAP zu nutzen, hing damit zusammen, dass diese Software in den auch für die Forschungsprojektdaten relevanten Bereichen – Personal- und Finanzwesen – an der Freien Universität Berlin bereits verwendet wurde; seit 2004 im Personal- und Haushaltswesen und seit 2005 in der Studierendenverwaltung. Im Rahmen einer Homogenisierung der IT-Struktur und Verbesserung interner Arbeitsprozesse sollte ebenso für das Management von Forschungsprojektdaten ein möglichst kompatibles und einheitliches Datenbanksystem eingesetzt werden. Zudem bietet SAP durch Module und Standardobjekte die Möglichkeit, bestimmte Daten wie Personaldaten in einem Modul zu pflegen, auf das andere Arbeitsbereiche referentiell zugreifen können, wodurch Datenhaltung und -pflege nur einmal erfolgen und mehrfache Arbeitsgänge und Dateninkonsistenzen vermieden werden können.

#### **3.1 Struktur und Nutzung der Forschungsdatenbank**

Die Forschungsdatenbank<sup>7</sup> besteht aus folgenden, SAP-integrierten Teilapplikationen:

1. Datenbank mit Drittmittelprojekten (sog. Drittmittelprojektdatenbank).
2. Datenbank zur Erfassung von Forschungsaktivitäten außerhalb von Drittmittelförderungen (sog. Profildatenbank), in der zudem Mitgliedschaften und wissenschaftliche Ämter, Gastaufenthalte, Auszeichnungen, Preise, Promotionen sowie Publikationen aller Wissenschaftlerinnen und Wissenschaftler der Freien Universität Berlin zur Abbildung ihres wissenschaftlichen Profils verzeichnet werden können.

Systemvoraussetzung für die Forschungsdatenbank ist mindestens Windows XP. Die ABAP-

---

6 Präsentationsfolien zum Vortrag, der diesem Beitrag zugrunde liegt, unter: [http://www.dini.de/fileadmin/workshops/forschungsinformationssysteme/Lewerentz\\_Annette\\_Folien-Forschungsdatenbank-FU-Berlin.ppt](http://www.dini.de/fileadmin/workshops/forschungsinformationssysteme/Lewerentz_Annette_Folien-Forschungsdatenbank-FU-Berlin.ppt).

7 Leitfaden zur Forschungsdatenbank als PDF-Download unter: <http://www.fu-berlin.de/forschung/profil/datenbanken/Leitfaden-SAP-Forschungsdatenbank-FU-Berlin-pdf.pdf>.

Weitere Informationen unter: <http://www.fu-berlin.de/forschung/service/forschungsdatenbank.html>.

programmierte Datenbank mit MS-Office-Schnittstellen (mySQL) in der derzeit verwendeten Komponentenversion SAP ECC 6.0 ist so angelegt, dass eine zukünftige Erweiterung des Bedienkonzepts keine Veränderungen in der Applikation erfordert.

Um die Anforderungen von Datenschutz und -sicherheit zu gewährleisten, ist der dezentrale Mehrfachzugriff auf die Forschungsdatenbank durch spezifische Berechtigungskonzepte bzw. -objekte in derzeit 11 Sammelrollen zusammengefasst, in denen der nutzerbezogene Zugriff auf die kaufmännischen und inhaltlichen Daten durch differenzierte Lese- und Schreibrechte sowie die Auswahl und Anzeige bestimmter Dateninhalte festgelegt sind. Des Weiteren werden Kriterien des Datenschutzes, der Risikobewertung, der Art, der Übermittlung, der Verfügbarkeit, der Verarbeitung und der Sicherheit von Daten, der Arbeitsabläufe und der personellen Ressourcen, der eingesetzten Soft- und Hardware sowie deren Betreuung durch standardisierte IT-Verfahrensbeschreibungen, Dienst- und Datenschutzvereinbarungen definiert und geregelt.

### 3.2 Referenzen der Forschungsdatenbank auf SAP-Module

Zur Zeit der Implementierung der Forschungsdatenbank 2005 bot SAP keine Möglichkeit, qualitative Forschungsinformationen abzubilden. Daher wurde die Forschungsdatenbank als separate SAP-Applikation mit referentiellm Zugriff auf Stammdaten einzelner SAP-Module und Standardobjekte<sup>8</sup> generiert und die Bedienung der herkömmlichen Benutzeroberfläche des SAP-Systems angepasst.

Zu den integrierten SAP-Modulen gehören folgende (Abb. 1):

- SAP-FI/CO/PSM zur Finanzverwaltung der Drittmittelprojekte (Referenz auf Haushaltsdaten [PSM], Finanzierung und Controlling, Fonds und Innenaufträge [FI/CO]),<sup>9</sup>
- SAP-HCM zur Personalverwaltung,<sup>10</sup>
- SAP-Organisationsmanagement zur Einbindung der Einrichtungen in die Organigraphie der Freien Universität Berlin und Zuordnung der Projektleitung und -mitarbeiter/-innen zur jeweiligen Einrichtung,<sup>11</sup>
- SAP-Geschäftspartnerverwaltung zur Erfassung externer Kooperationspartner,
- SAP-Textverwaltung für die Verzeichnung von projektbezogenen Abstracts und Schlagworten (beide mehrsprachig möglich),
- SAP-Statusverwaltung für den aktiven oder beendeten Drittmittelberechnungsstatus, das heißt welche betriebswirtschaftlichen Vorgänge erlaubt oder verboten sind. Über die Statusverwaltung werden Fonds aktiviert, beendet oder gelöscht und Innenaufträge angelegt.

Hinsichtlich der Datenqualität und -aktualität bedeutet die Referenz zu diesen verschiedenen Modulen zugleich eine Abhängigkeit zu anderen Arbeitsbereichen, da die Daten der Module in unterschiedlichen Abteilungen der Freien Universität Berlin gepflegt werden und die Datenpflege in einen festen Workflow eingebunden ist.

---

8 Dies sind die Geschäftspartner-, Langtext- und Statusverwaltung. Hierbei werden Objekte und Zugriffstechniken aus dem Business Object Repository genutzt, vor allem SAP-eigene Funktionsbausteine.

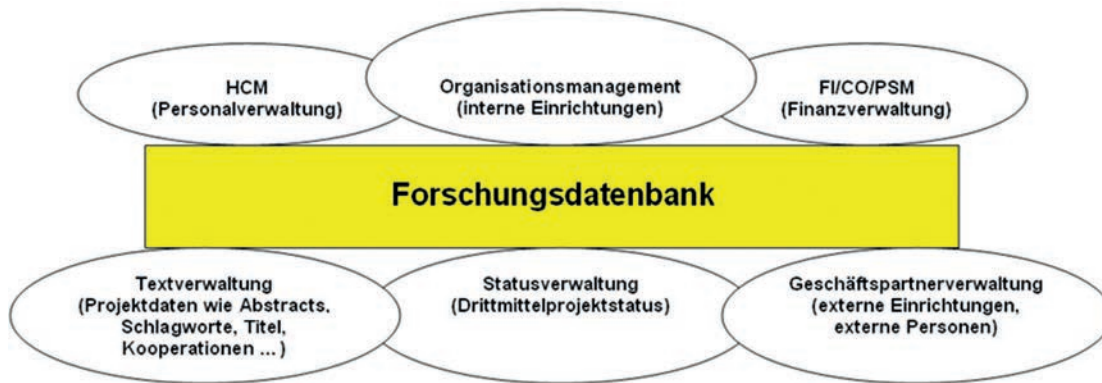
9 FI: Finanz-/Rechnungswesen, CO: Controlling (Kostenrechnung), PSM: Public Sector Management (Haushaltsmanagement).

10 HCM (Human Capital Management): Personalmanagement und -wirtschaft.

11 Untermodul von HCM.



Abbildung 1: Forschungsdatenbank: Referenzen auf SAP-Module und Standardobjekte

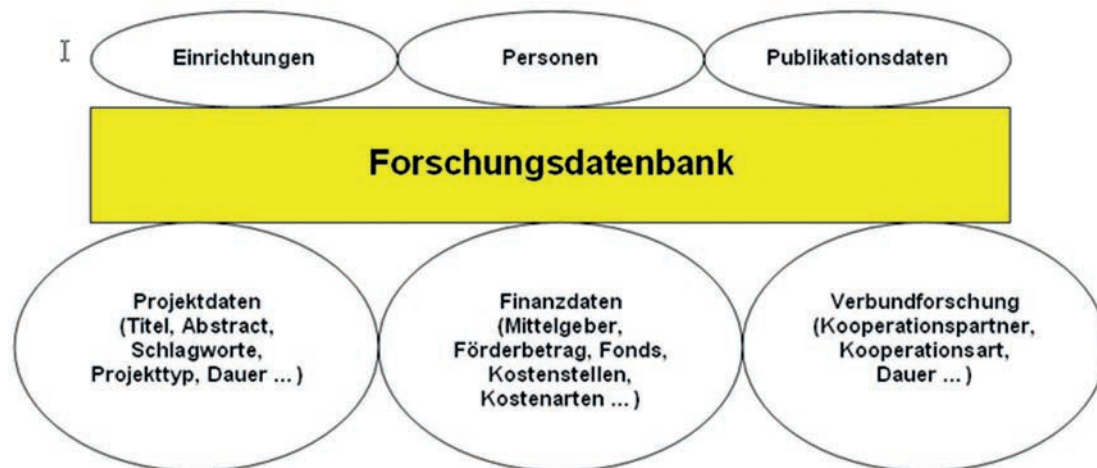


## 4 Erste Teilapplikation der Forschungsdatenbank: Drittmittelprojektdatenbank

### 4.1 Datenkategorien der Drittmittelprojektdatenbank

Drittmittelprojekte werden mit Formal- und Inhaltsdaten in der Datenbank verzeichnet (Abb. 2).

Abbildung 2: Datenkategorien der Forschungsdatenbank für Drittmittelprojekte (Drittmittelprojektdatenbank) und Forschungsaktivitäten außerhalb von Drittmittelförderungen (Profildatenbank)



Zu den Formaldaten gehören Angaben zur Projektleitung, zur Einrichtung, an der das Projekt angesiedelt ist, und zur Projektgesamtdauer. Hinzu kommen die Finanzdaten, die im Fonds und Innenauftrag hinterlegt sind. Zu diesen zählen Geldgeber, Finanzierungszweck und Finanzposition, Bewilligungssumme für die jeweilige Projektlaufzeit sowie die Dauer der einzelnen Projektbewilligungen, Projekt- und Mehrwertsteuerstatus, Fondsnummern und -gruppierung, Kostenstellen, Overheadverrechnung und das Geschäftszeichen sowie weitere buchungs- und abrechnungsrelevanten Kriterien. Die Inhaltsdaten umfassen den Projekttitle, sofern vorhanden Webseiten und Forschungsverbund inklusive Nennung der Art und des Titels der Kooperation,<sup>12</sup> der Kooperationspartner, differenziert nach Art, geographischer Herkunft der kooperieren-

<sup>12</sup> Kooperationsarten sind zum Beispiel DFG-Programmtypen wie Sonderforschungsbereiche oder Forschergruppen, ferner Verbundprojekte des Bundesministeriums für Bildung und Forschung, EU-Projekte u.s.w.



den Einrichtung<sup>13</sup> und Funktion wie in Form von Beteiligung oder Sprecherfunktion in einem Forschungsverbund. Ferner gibt es ein thesaurusähnliches Schlagwortverzeichnis, das inhaltlich und hierarchisch nach Fachgebieten und -themen mit Ober- und Unterbegriffen aufgebaut ist und kontinuierlich durch neue Begriffe erweitert wird. Die Projekte wiederum sind typologisch nach Forschungsarten differenziert.<sup>14</sup> Projekthinhalte stehen in der Projektbeschreibung. Diese Erfassung von inhaltlichen Projektdaten soll Recherchen ermöglichen, durch die beispielsweise bestimmte Forschungsthemen und -schwerpunkte identifiziert werden, nach denen zudem übergeordnete Forschungsleitlinien erkennbar und das universitäre Forschungsprofil präsentiert werden.

#### **4.2 Bereitstellung der Informationen der Drittmittelprojektdatenbank: SAP-Report**

All diese Daten sind in einem SAP-Bericht mit kombinierbaren Selektionskriterien und Mehrfachabfragen recherchierbar.<sup>15</sup> Die Ergebnisliste lässt sich in die MS-Officeprogramme Excel und Word zur weiteren Bearbeitung und Speicherung exportieren. Der Export der Daten des Rechercheergebnisses in MS-Word erfolgt über das Einfügen von Seriendruckfeldern, die den Datenfeldern entsprechen und mit den Feldinhalten gefüllt werden. Die ausgewählten Seriendruckfelder lassen sich standardisiert und mit individuellem Layout als Mustervorlage in MS-Word abspeichern, damit sie nicht bei jedem Datenexport von Neuem eingefügt werden müssen und somit als Vorlage für ähnliche Abfrageinhalte wieder verwendbar sind.

#### **4.3 Arbeitsprozessoptimierung in der Administration von Drittmittelprojekten**

Bewilligte Drittmittelvorhaben werden über ein Formular<sup>16</sup> der Drittmittelverwaltung der Freien Universität Berlin mitgeteilt. Dort werden in der Drittmittelprojektdatenbank die Grund- und Finanzdaten des Projekts erfasst und aus der Datenbank unter anderem Formulare für Wirtschaftserbefugnisse, Spendenbescheinigungen und Projektgenehmigung zur Generierung des Drittmittelfonds erstellt. In der Abteilung Personal- und Finanzwesen (Bereich Haushalt) der Freien Universität Berlin werden die Finanzdaten qualitätssichernd auf Richtig- und Vollständigkeit überprüft und das Projekt genehmigt,<sup>17</sup> um dieses in das SAP-Modul FI/CO zu integrieren und weitere FI/CO-Objekte anzulegen; das heißt, es werden hierdurch automatisch Fonds, Innenaufträge und Budgetstrukturpläne in SAP generiert. Im Fonds und Innenauftrag sind Informationen zum Budget, zur Mittelverfügbarkeit, zu Obligos und Kosten hinterlegt. Die Bewirtschaftung von Personal- und Sachmitteln, Mittelanforderung, Budgetierung und Budgetkontrolle, Zwischen- und Schlussverwendungsnachweiserstellung und Vertragsprüfung geschehen – auch mithilfe dieser datenbankgestützten SAP-Funktionalitäten – in der Drittmittelverwaltung.

Die Pflege der Inhaltsdaten obliegt der Abteilung Forschung. Hierzu zählt die Verschlagwortung der Projekte mit Begriffen der thesaurusähnlichen Systematik. Des Weiteren werden die Kooperationen und Kooperationspartner angelegt, die Projekte nach ihrer Art klassifiziert und die Projektbeschreibungen eingefügt. Eine redaktionelle Durchsicht aller verzeichneten Projektdaten

---

13 Die Kooperationspartnerarten sind zum Beispiel klassifiziert nach Universitäten, Max-Planck-Einrichtungen, Einrichtungen der Helmholtz-Gemeinschaft, Fraunhofer-Instituten u.s.w. Zudem werden immer die Orte und das Land erfasst, um nach regionalen, nationalen und internationalen Partnern suchen zu können.

14 Zum Beispiel Forschungsprojekte, Tagungen, Reisen, EU-Forschung, Workshops, Publikationsbeihilfen, Spenden, Preisauszeichnungen, Auftragsforschung, Nachwuchsförderung / Graduiertenkollegs, Projekte der Exzellenzinitiative, Overheads.

15 Verknüpfte Abfragen (Boole'scher Operator „und“).

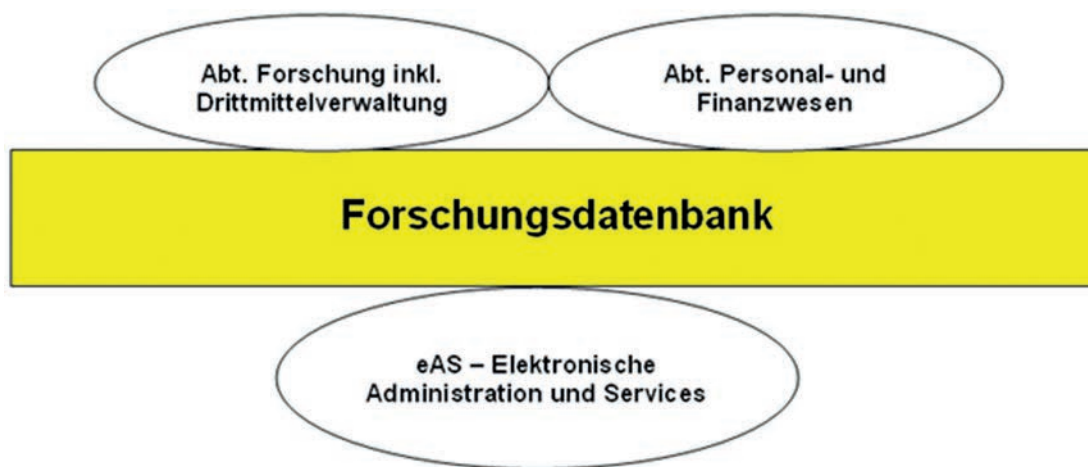
16 Sog. Anzegebogen für Drittmittelprojekte. Als Download unter:  
[http://www.fu-berlin.de/forschung/service/dmv/formular/formular\\_2/anzegebogen\\_doc.doc](http://www.fu-berlin.de/forschung/service/dmv/formular/formular_2/anzegebogen_doc.doc).

17 Das heißt, der Projektstatus wird auf „aktiv“ gesetzt.

wird zur Qualitätskontrolle durchgeführt. Zudem werden vor allem in der Abteilung Forschung Recherchen vorgenommen. Aber auch andere Arbeitsbereiche der Freien Universität Berlin haben dezentral einen benutzerspezifischen Zugriff auf Projektdaten und Recherchebericht.

Positive Folgen dieser integrierten Geschäftsprozessoptimierung (Abb. 3) und des mit den einzelnen SAP-Modulen verzahnten Forschungsinformationssystems (Abb. 1) sind die Verbesserung von Informationsqualität, Datentransparenz und Konsistenz, die schnellere Verfügbarkeit und Standardisierung der Informationen, die Vereinfachung von Arbeitsabläufen sowie die Vermeidung von Arbeits- und Systemredundanzen. Nachteil kann aber auch die Abhängigkeit zwischen den Abteilungen sein, in denen die Daten gepflegt werden; denn die Qualität der Daten hängt von der Qualität ihrer Pflege ab. Daten, die aus anderen Bereichen referentiell übernommen werden, können ggf. nicht direkt korrigiert werden. Kommunikation und Bereitschaft zur einheitlichen qualitativen Datenpflege sind Voraussetzung für ein qualitätsvolles und konsistentes System.

*Abbildung 3: Dateneingabe und -pflege der Forschungsdatenbank: beteiligte Einrichtungen der Freien Universität Berlin*



#### 4.4 Internetdarstellung von Drittmittelprojekten

Die drittmittelgeförderten Forschungsaktivitäten der Freien Universität Berlin sind seit 2007 im Internet zugänglich.<sup>18</sup> Es handelt sich um eine Online-Datenbank mit Web-Dynpro-Programmierung, die direkt auf die Daten der SAP-Forschungsdatenbank zugreift. Dies hat den Vorteil, dass Drittmittelvorhaben sofort im Internet verfügbar sind und den Daten der universitätsinternen SAP-Forschungsdatenbank entsprechen. Zugleich bedeutet diese zeitgleiche Datenübertragung in beide Systeme, dass die Bearbeitung der Projektdaten in den einzelnen Arbeitsbereichen zügig erfolgen sollte, damit die Online-Informationen schnell vollständig abgebildet werden und keine Eingabelücken aufweisen.

Die Online-Version der Datenbank umfasst die Projektgrunddaten, wie Angaben zur Projektleitung, Einrichtung, Projektdauer, ferner Inhaltsdaten, zu denen Projekttitle, Art des Projektes, Schlagworte, Projektbeschreibungen und Kooperationsangaben gehören. Finanzdaten sind aus Datenschutzgründen auf Angaben der Drittmittelgeber beschränkt. Insgesamt werden Drittmittelvorhaben, Forschungsbeziehungen in der nationalen und internationalen Verbundforschung sowie Forschungsschwerpunkte und -themen fächerübergreifend, fachspezifisch und transparent abgebildet. Wissenschaftlerinnen und Wissenschaftler sowie die interessierte Öffentlichkeit können sich dadurch über die universitäre Forschungslandschaft informieren und auch zur interdisziplinären Forschung und Forschungskooperation angeregt werden.

<sup>18</sup> <https://eforschung.ecampus.fu-berlin.de/dmv>.

## **5 Zweite Teilapplikation der Forschungsdatenbank: Profildatenbank**

### **5.1 Datenkategorien, -erfassung und -aufbereitung in der Profildatenbank (inkl. Workflow)**

Die zweite Teilapplikation der Forschungsdatenbank, die sog. Profildatenbank, soll Forschungsaktivitäten außerhalb von Drittmittelförderungen sowie Publikationen aller an der Freien Universität Berlin beschäftigten Wissenschaftlerinnen und Wissenschaftler abbilden (Abb. 2). Ziel ist es, ein umfassendes Informationssystem anzubieten, in dem über Drittmittelförderungen hinausgehende Forschungsleistungen und wissenschaftliche Profile der an der Freien Universität Berlin angestellten Forschenden dargelegt und im Internet<sup>19</sup> präsentiert werden. Diese Internetpräsentation ermöglicht eine zentrale Positionierung der Forschungsprofile, die zuvor nur auf den einzelnen Instituts- oder persönlichen Webseiten der Wissenschaftlerinnen und Wissenschaftler verteilt waren, und stellt Forschungsaktivitäten transparent dar; von der Profildatenbank wiederum wird auf die dezentralen Instituts- oder persönlichen Webseiten verlinkt. Forschungsaktivitäten lassen sich fach- und themenspezifisch, fächerübergreifend, nach Personen und Einrichtungen recherchieren. Zudem werden Auszeichnungen und Preise, Ämter und Mitgliedschaften, Publikationen sowie Gastaufenthalte dargestellt, die neben der Forschung Indikatoren wissenschaftlicher Leistungsbemessung bilden können.

Im Gegensatz zur Drittmittelprojektdatenbank beruht die Datenerfassung auf Selbsteintrag durch die Wissenschaftlerinnen und Wissenschaftler. Hierfür ist ein kennwortgeschütztes Formular im Internet vorhanden,<sup>20</sup> in das Daten zu wissenschaftlichen Aktivitäten eingetragen werden können. Zu den erhobenen Daten gehören Angaben zur Projektleitung, Einrichtung, eine Betitelung und Zeitphase der Forschungsaktivitäten und deren Klassifizierung, wie zum Beispiel Forschungsprojekt, Preis, Gastdozentur oder Mitgliedschaft, eine Verschlagwortung übergeordnet nach Forschungsschwerpunkten und spezifisch nach Forschungsthemen, die Projektbeschreibung, Kooperationspartner und eine Einbindung von Publikationen. Zudem kann entschieden werden, ob die Daten online veröffentlicht werden oder nur innerhalb der SAP-Profildatenbank; denn die Daten des Webformulars werden in dieser intern gespeichert. Der weitere Workflow besteht darin, dass die Daten innerhalb der SAP-Profildatenbank auf Vollständigkeit redaktionell nochmals überprüft, gegebenenfalls korrigiert und von dort für die öffentlich zugängliche Online-Datenbank freigegeben werden.

Um Redundanzen der Daten und des IT-Systems sowie mehrfache Arbeitsprozesse bei der Datenpflege zu vermeiden, verwendet die Profildatenbank dieselben SAP-Module und -Funktionalitäten wie die Drittmittelprojektdatenbank. Dies betrifft die Klassifizierung der Vorhaben in Projekttypen, die Eingabe von Kooperationen, die systematische Verschlagwortung und der Zugriff auf SAP-Module für die Erfassung der Wissenschaftlernamen und Einrichtungen. Das rollenbasierte Berechtigungskonzept regelt die spezifischen Lese- und Schreibrechte.

### **5.2 Bereitstellung der Informationen der Profildatenbank: SAP-Report**

Informationen der Profildatenbank sind neben der Internetdarstellung ebenso im SAP-Report recherchierbar. Die kombinierbare Datenselektion und Recherchemöglichkeiten entsprechen jenen, die bereits bei der Drittmittelprojektdatenbank beschrieben wurden. Die Ergebnisliste lässt sich ebenso in die MS-Office-Programme Word und Excel exportieren und weiterverarbeiten.

---

<sup>19</sup> <https://eforschung.ecampus.fu-berlin.de/ndm>.

<sup>20</sup> <https://eforschung.ecampus.fu-berlin.de/pdb>. Weitere Informationen zum Webformular siehe unter: <http://www.fu-berlin.de/forschung/service/projekterfassung.html>.

## 6 Übergeordnetes Reporting zur Forschung der Freien Universität Berlin

Um Recherchen über die universitäre Forschung unabhängig von der Finanzierungsart zu ermöglichen, ist ferner ein gemeinsamer SAP-Bericht für Drittmittel- und Nicht-Drittmittelprojekte implementiert. Hierfür werden strukturell beide Teilapplikationen der Forschungsdatenbank ausgewertet. Dementsprechend sind die Möglichkeiten bei der Selektion und bei der Darstellung des Suchergebnisses in der Ergebnisliste auf diejenigen Merkmale beschränkt, die sowohl in der Drittmittelprojekt- als auch in der Nicht-Drittmittelprojektdatenbank (sog. Profildatenbank) enthalten sind. Darüber hinausgehende Merkmale können jedoch in der jeweiligen Detailsicht zu einzelnen Projekten eingesehen werden.

## 7 Verknüpfung zwischen der Forschungsdatenbank und weiteren Informationssystemen der Freien Universität Berlin

Die Forschungsdatenbank ist mit beiden Teilapplikationen, Drittmittelprojekt- und Profildatenbank, in weitere SAP-Applikationen eingebunden (Abb. 4).

Abbildung 4 : Forschungsdatenbank: Schnittstellen zu weiteren Datenbanken der Freien Universität Berlin



### 7.1 Drittmittelprojektdatenbank und Statistikdatenbank

Eine weitere Anforderung an die Forschungsdatenbank besteht in der Gewinnung ausschließlich quantitativer Daten wie zum Beispiel Drittmittelstatistiken für verschiedene Berichte und für universitätsinterne Steuerungsprozesse.<sup>21</sup> Für diese Anforderung ist die Teilapplikation der Forschungsdatenbank, die Drittmittelprojektdatenbank, mit einem weiteren Informationssystem der Freien Universität Berlin verknüpft: der sog. Statistikdatenbank.

Die Statistikdatenbank ist ebenfalls eine SAP-Applikation, die Daten von dezentralen universitären Verwaltungsbereichen für Auswertungen aufbereitet, zum Beispiel in der Personal-, Haushalts-, Forschungs- und Studierendenabteilung,<sup>22</sup> und verschiedenen Nutzerinnen und Nutzern zentral zugänglich macht. Diese Datentransparenz schafft Wissen über universitäre Vorgänge<sup>23</sup> und entlastet Arbeitsbereiche, die ihre Daten nicht mehr einzeln zuliefern müssen. Die in der Statistikdatenbank aggregierten Zahlen sind keine tagesaktuellen, sondern Zahlen auf Semester- oder Jahresbasis. Ziel ist es, die Mitarbeiterinnen und Mitarbeiter der Freien Universität Berlin, die

<sup>21</sup> Zum Beispiel statistische Daten als Parameter für die interne Mittelverteilung an Fachbereiche und Institute.

<sup>22</sup> Zum Beispiel Daten zur Buchhaltung und zu Immatrikulationen.

<sup>23</sup> Zum Beispiel im Personal- und Haushaltswesen, in Studium, Lehre, Forschung, Internationalität und Gleichstellung.

zentral oder dezentral mit Entscheidungs- oder Controllingaufgaben oder mit Angelegenheiten für die Hochschulsteuerung befasst sind, durch die Statistikdaten zu unterstützen sowie Anfragen zur Hochschulstatistik, wie zum Beispiel durch das Statistische Bundesamt,<sup>24</sup> bedienen zu können. Der Datenzugriff basiert auf einem nutzer- und aufgabenbezogenen Berechtigungskonzept ohne öffentlichen Zugang zur Datenbank.

Statistisch ausgewertet werden aus der Forschungsdatenbank jahresweise Kooperationsdaten, zum Beispiel die zahlenmäßige Beteiligung oder Sprecherfunktion der Freien Universität Berlin in Forschungsverbünden, die Art der Forschungsprojekte<sup>25</sup> sowie Finanzdaten nach Geldgebern, Fonds und Kostenstellen. Die Daten lassen sich zum Beispiel jahresweise als Zeitreihe, fachbereichsspezifisch, nach Fächergruppen und nach Projektleitungen differenziert erheben. Über eine weitere Schnittstelle zwischen Forschungsdatenbank und Statistikdatenbank werden die jährlichen Drittmittelstatistiken erstellt. Hierfür greift die Statistikdatenbank innerhalb von SAP auf Daten der Forschungsdatenbank und auf Haushaltsdaten zurück, in denen Drittmittel nach Einnahmen und Ausgaben aus Zuwendungen der Freien Universität Berlin oder aus persönlichen Zuwendungen, sofern sie über die Universitätskonten abgerechnet werden, enthalten sind.

## **7.2 Profildatenbank und Publikationsdatenbank**

In der Profildatenbank können die Veröffentlichungen eingebunden werden, die jährlich in der Bibliographie der Universitätsbibliothek der Freien Universität Berlin erfasst sind. Die Wissenschaftlerinnen und Wissenschaftler tragen ihre Publikationen über ein Webinterface ein. Von hier werden die bibliographischen Daten an das bibliographische Informationssystem APELH der Universitätsbibliothek weitergegeben und dort zur Qualitätssicherung geprüft. Des Weiteren erfolgt ein Import der Publikationsdaten in die sog. Publikationsdatenbank, die einen Bestandteil der SAP-Statistikdatenbank bildet.<sup>26</sup> In der Statistikdatenbank können die Veröffentlichungen der Wissenschaftlerinnen und Wissenschaftler der Freien Universität Berlin statistisch ausgewertet und als verdichtete Publikationsdaten auf Ebene der wissenschaftlichen Einrichtungen, der Autorinnen und Autoren und der Forschungsprojekte für Potenzialanalysen verwendet werden. Über eine Schnittstelle zur Statistikdatenbank werden zudem die bibliographischen Publikationsdaten in die Profildatenbank referentiell eingebunden. Wenn Publikationen im Rahmen von Drittmittelprojekten entstanden sind, ist diese Zuordnung über den SAP-Innenauftrag möglich.

## **8 Ausblick: Weiterentwicklungen der Forschungsdatenbank**

Um einer effizienten Datenaufbereitung für zukünftig sich verändernde Anforderungen des Berichtswesens oder universitärer Steuerungsmodelle gerecht werden zu können, sind weitere Entwicklungen der Forschungsdatenbank vorgesehen.

### **8.1 Drittmittelprognose**

Die Drittmittelprognose soll als eines der Managementinstrumente der Universitätsleitung für die leistungsbasierte Hochschulfinanzierung verwendet werden. Die Prognose prospektiver Drittmittelausgaben bildet daher einen Indikator zum Nachweis tatsächlich erbrachter und zukünftig kalkulierbarer Drittmittel der Universität – getrennt nach den Kostenarten Personal- und Sachmittel.

In der Drittmittelprojektdatenbank sind im Innenauftrag nur die Bewilligungssummen für die

---

24 Siehe auch <http://www.statistik-berlin-brandenburg.de/>.

25 Das heißt, untergliedert nach Forschung allgemein und Forschung des wissenschaftlichen Nachwuchses.

26 Für den Import wird eine LIS-Eingangsschnittstelle verwendet, der eigentliche Import der Daten in das SAP-System erfolgt über Intermediate Documents (IDOC's).



gesamte Laufzeit eines Projektes, derzeit ohne Trennung der Kostenarten, hinterlegt. Regelmäßige Buchungen von Personal-<sup>27</sup> und Sachkosten sind bei den Haushaltsdaten – im SAP-FI/CO-Modul – recherchierbar. Für die Drittmittelprognose ist es notwendig, die verausgabten Mittel aus der noch zur Verfügung stehenden Bewilligungssumme eines Projektes herauszurechnen, das heißt noch nicht verausgabte Mittel von bereits gebuchten Mitteln rechnerisch zu trennen. Dies kann fondsgenau erfolgen, aus dessen Struktur sich Mittelherkunft und -empfänger feststellen lassen.<sup>28</sup> Die verbleibende Summe ist für die prospektive Drittmittelschätzung linear auf den noch vorhandenen Bewilligungszeitraum eines Projektes zu verteilen.

Das technische Konzept zur Erstellung der Drittmittelprognose sieht vor, dass bereits verausgabte Mittel pro Drittmittelprojekt in einem eigenen SAP-Report gespeichert werden, der monatlich aktualisiert wird. Über diesen ersten Report können die gebuchten Mittel herausgerechnet werden, um die noch zur Verfügung stehenden Mittel projektweise zu erfassen, die ihrerseits für die Erstellung der Drittmittelprognose über einen zweiten Report recherchiert werden. Eine dynamische Abfrage prognostizierbarer Daten ist zu jedem und für jeden Zeitpunkt möglich.

## 8.2 Vertragscontrolling

Eine weitere, technisch implementierte, aber noch nicht regelmäßig gepflegte Anwendung innerhalb der Forschungsdatenbank ist das sog. Vertragcontrolling. Hierüber werden Daten zu Verträgen in Drittprojekten verzeichnet, die wiederum über einen SAP-Bericht recherchierbar sind. Zu den verzeichneten Daten zählen Vertragsart, auch Musterverträge, Vertragsdauer und -fristen, Angaben zur Einrichtung und zur Projektleitung, sowie Vertragsbestandteile. Des Weiteren werden die im Vertrag benannten Kooperationspartner erfasst. Diese Daten werden referentiell mit den Drittmittelprojekten in der Drittmittelprojektdatenbank verknüpft, um Redundanzen und doppelte Eingabearbeiten zu vermeiden.<sup>29</sup>

Eines der Ziele dieser Anwendung ist es, Informationen über alle vorhandenen Drittmittelverträge der Freien Universität Berlin zu erhalten.

## 9 Zusammenfassung

Die Forschungsdatenbank der Freien Universität Berlin umfasst Drittmittelprojekte und Forschungsaktivitäten außerhalb von Drittmittelförderungen mit dem Ziel, verschiedene interne und externe Berichtspflichten abzudecken, Daten für universitäre Entscheidungsträger bereitzustellen, Forschungsleistung und -schwerpunkte transparent der Öffentlichkeit zu dokumentieren sowie die Arbeitsprozesse innerhalb der Verwaltung von Drittmittelprojekten effizient abzuwickeln. Hierfür wird ein sowohl inhaltlich bzw. qualitativ orientiertes als auch ein finanz- und controllingbezogenes bzw. quantitatives Berichts- und Erfassungssystem bereitgestellt, in dem die Daten vielfältig kategorisiert, strukturiert und möglichst standardisiert gepflegt und recherchiert werden.

Die Integration der Forschungsdatenbank in das SAP-System mit referentiellen Zugriffen auf die einzelnen SAP-Module beugt einer Redundanz der Daten vor, unterstützt die Homogenisierung der universitären IT-Landschaft, verschlankt die administrativen Vorgänge durch einen zwischen den Zuständigkeitsbereichen aufeinander abgestimmten Arbeitsprozess in der Verwaltung von Forschungsvorhaben. Zudem werden in einem Informationssystem durch Zugriffe auf gleiche SAP-Module verschiedene Facetten inhaltlicher, quantitativer, statistischer Forschungs- und

---

<sup>27</sup> Für die Ermittlung zukünftiger Personalkosten ist es notwendig, auf die Personalkostenhochrechnung (Personalplanung) in den SAP-Funktionalitäten CO und Obligo zuzugreifen.

<sup>28</sup> Einrichtung, Projektleitung und Drittmittelgeber.

<sup>29</sup> Die Anwendung „Vertragscontrolling“ greift ebenfalls auf dieselben SAP-Module wie die Drittmittelprojekt- und die Profildatenbank zurück.

Publikationsinformationen verzahnt und zugleich ein hoher Grad an Datenaktualität und -validität gesichert. Differenzierte Berechtigungskonzepte lassen eine transparente Sicht auf die Daten zu und schützen zugleich vor Missbrauch.

Um Anforderungen verschiedener Adressaten zeitnah gerecht zu werden, werden Forschungsprojektdaten inhaltlich möglichst vielfältig verzeichnet und häufige Abfragemuster sind voreingestellt abspeicherbar.

Universitätsintern kann die Forschungsdatenbank als Informationsinstrument der Forschungs- und Nachwuchsförderung zur Analyse gegenwärtiger Forschungsschwerpunkte oder -kooperationen<sup>30</sup> für die Projektplanung, die für eine Antragstellung zur Drittmittelförderung relevant ist, sowie für den weiteren Ausbau von Forschungsfeldern und deren Konsolidierung verwendet werden. Sie wirkt so als Informationsgrundlage bei der Entwicklung von Forschungsstrategien, bei Stärken-Schwächen-Analysen bezüglich Drittmittelinwerbungen, als Informationsquelle über Forschung im wissenschaftlichen Nachwuchsbereich, über nationale und internationale Verbundforschung sowie über Interdisziplinarität zur Bündelung von Forschungskompetenzen mit.

Ein weiterer Mehrwert der Forschungsdatenbank besteht darin, dass aus dieser inhaltliche Daten für die Präsentation der Forschung auf den Internetseiten der Freien Universität Berlin und für den Expertendienst ihrer Presse- und Kommunikationsstelle als Informationsquelle für Medien und Öffentlichkeit<sup>31</sup> generiert werden.

---

<sup>30</sup> Zum Beispiel anhand des systematischen Schlagwortverzeichnisses und der Suche mit verschiedenen inhaltlichen Abfragekriterien nach Kooperationspartnern und Verbundforschung.

<sup>31</sup> Mehr zum Expertendienst unter <http://www.fu-berlin.de/presse/expertendienst/index.html>.





## **Forschung präsentieren, fördern, verwalten – die Forschungsdatenbank der Universität Rostock**

Die Universität Rostock ist mit aktuell etwa fünfzehntausend Studierenden und etwa 290 Professuren eine mittelgroße Universität. Steigende Studierendenzahlen verweisen auf eine zunehmende Attraktivität der Universität an der Ostseeküste, die in Rankings mit einem breiten Fächerspektrum, guten Ausstattungswerten und Betreuungsrelationen auf sich aufmerksam macht. Ebenso gibt es eine schwunghafte forschungsseitige Entwicklung der Universität mit einem bemerkenswerten Anstieg der Drittmittelausgaben - seit 2005 um 83 Prozent.

Für diese Entwicklung und zur strategischen Ausrichtung der Universität ist eine stärkere Präsenz der Forschungsgebiete und -ergebnisse der Universität und ihrer Professuren ein maßgeblicher Erfolgsfaktor. Entwicklungspotential von Forschung liegt insbesondere in inner- und interuniversitärer interdisziplinärer Forschung. Hier kann ein Informationssystem, das die vielfältigen Aspekte von Forschung beleuchtet, Informationen bieten und Kommunikation sowie Kooperation anregen.

An der Universität Rostock wurde zu diesem Zweck schon vor einigen Jahren damit begonnen, eine Forschungsdatenbank zu konzipieren und zu installieren. Die wenigen zum damaligen Zeitpunkt auf dem Markt verfügbaren Forschungsinformationssysteme erfüllten Grundanforderungen an ein solches System nicht, so dass nur eine neue Lösung infrage kam. Auf der Suche nach einer langfristig nutzbaren, aber extern gepflegten Systemlösung lag für die Universität Rostock, in deren Verwaltungsbereich überwiegend HIS-Systeme verwendet werden, die Idee nahe, mit der HIS GmbH eine Lösung zu implementieren. Diese bestand in der Weiterentwicklung des zwar in HIS-LSF (Lehre-Studium-Forschung) angelegten, aber für die umfassende Darstellung von Forschungsaspekten damals unbrauchbaren Forschungsteils. Was mit einer Rostocker Konzeption einer Forschungsdatenbank startete, ist nun für alle HIS-LSF-nutzenden Hochschulen zugänglich und weiter entwickelbar.

Eine Arbeitsgruppe Forschungsdatenbank, die sich aus an der Forschungsdatenbank interessierten Verwaltungs-Mitarbeitenden der Universität Rostock aus dem Referat Forschung und dem Controlling gründete, entwarf mit Unterstützung eines Mitarbeiters zur IT-seitigen Umsetzung initial ein Konzept. Diese Kernarbeitsgruppe aus Controlling, Referat Forschung und IT- und Medienzentrum besteht bis dato. Das Konzept resultierte aus der Sicht der Forschenden, der Sicht der Forschungssteuerung und aus den Erfahrungen derjenigen Bereiche einschließlich der Universitätsbibliothek, die Forschungsaspekte managen und verwalten. Es wurden viele Gespräche mit dem umfangreichen Kreis der Stakeholder geführt: Darunter waren diejenigen Fakultäten, die sich bereits eigene kleine Lösungen auch für Teilaspekte von Forschung geschaffen hatten, und die Medizinische Fakultät, die zusammen mit dem Klinikum separat verwaltet wird und seit langem ein ausgefeiltes Mittelverteilungssystem u.a. auf Basis von Forschungsleistungen praktiziert. Anhand dessen, was als dringend gebraucht kommuniziert wurde, und auf Basis der unterschiedlichen disziplinären Sichten wurde die Konzeption immer wieder überprüft und verfeinert. Ergebnis ist eine Konzeption a) basierend auf einem breiten Verständnis von Forschungsaspekten, b) deren Abhängigkeiten untereinander, c) dem Weg der Informationen in die zu schaffende Forschungsdatenbank und d) den unterschiedlichen Sichtweisen bei der Nutzung der Forschungsdatenbank.

Diese Konzeption und deren Umsetzung werden im Weiteren als Erfahrungsbericht beschrieben.

## 1 Anforderungen an die Forschungsdatenbank

Selbstverständlich ist die Anforderung, dass die Forschungsdatenbank Forschung in einem breiten Spektrum ihrer Aspekte darstellen soll, von Publikationen bis zu Gutachtertätigkeit, von Promotionsbetreuungen bis zu Kooperationsangeboten. Die Forschungsdatenbank ist kein Verwaltungsinstrument zur Bürokratisierung sondern ein die Forschung unterstützendes Informationsangebot, das im günstigsten Fall auch zur Verwaltung eingesetzt werden kann.

Die Inhalte müssen dabei mit einem **vertretbaren Eingabeaufwand** in das Informationssystem gelangen, was bedeutet, dass Daten, die bereits anderswo erfasst sind, nicht noch einmal erfasst werden. Auch wenn die Anwendungszwecke voneinander abweichen, sollen die für die Forschungsdatenbank notwendigen und sinnvollen Daten so weit wie möglich aus anderen Systemen importiert werden bzw. über automatisierte Schnittstellen angebunden werden. Eine **Einbettung der Forschungsdatenbank in das weitere Systemumfeld** ist also mehr als nur „nice to have“. Verursacht wird diese Forderung nicht zuletzt dadurch, dass die Universität Rostock seitens des Landes Mecklenburg-Vorpommern mit einem bis 2017 umzusetzenden Stellenabbau konfrontiert ist. Zusätzliche dauerhafte Aufwände und Personal allein für die Forschungsdatenbank sind vor diesem Hintergrund nicht realisierbar. Doppelter Pflegeaufwand, z.B. für Projekte, die im Finanzsystem ohnehin erfasst werden, muss und kann - allein schon aus Gründen der Praktikabilität - vermieden werden. Außerdem hieße doppelte Eingabe von Daten unnötig Eingabefehler und Inkonsistenzen zu produzieren. Die Überwachung, dass Datenänderungen in den Originalsystemen ebenfalls in den abhängigen Systemen erfolgen, ist ein vermeidbarer Aufwand. Einmalige Dateneingabe und kontrollierte Replikate in HIS-LSF mittels Schnittstellen garantieren zusätzlich die **Konsistenz** gleicher Daten über mehrere Systeme hinweg. Welche Daten dies betrifft, wird weiter unten erläutert. Als positiver Nebeneffekt der Mehrfachnutzung von Daten stellte sich sofort ein, dass die Forschungsdatenbank und generell HIS-LSF als weiteres Anwendungssystem von in anderen Systemen verwalteten Daten diese zusätzlich validiert bzw. für deren Vervollständigung sorgt.

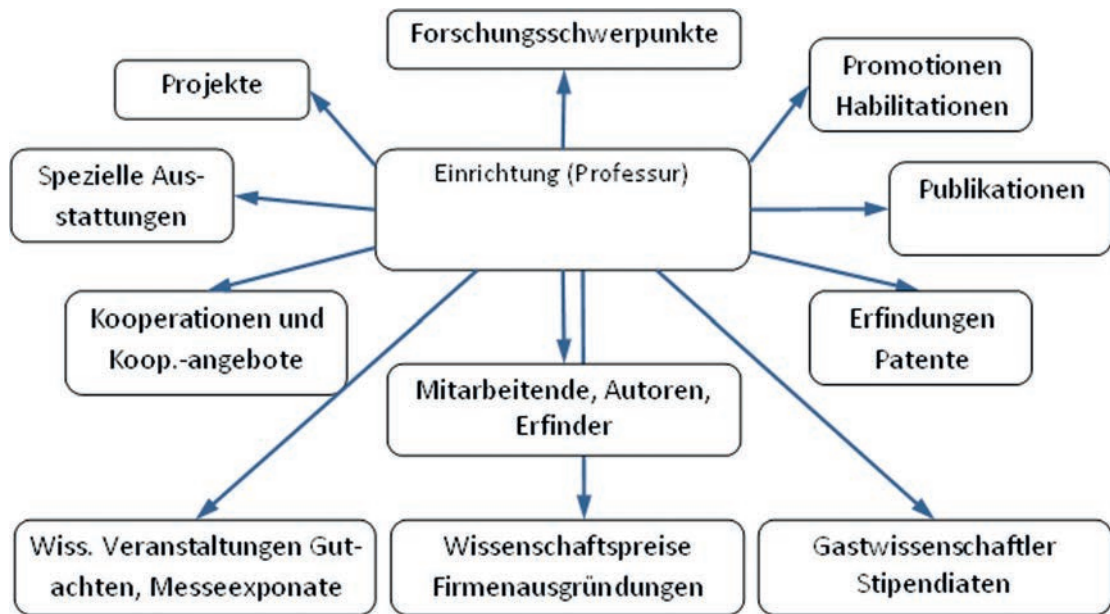
Die **Validität der Daten** wird in der Forschungsdatenbank der Universität Rostock neben der Nutzung von Schnittstellen zudem dadurch erreicht, dass die jeweils Daten verwaltenden Stellen „ihre“ Daten in die Forschungsdatenbank eingeben. Zum Beispiel gibt nicht ein Promovend oder die jeweilige Fakultät sondern die Promotionsstelle der Universität Promotionsdaten ein. Damit sind sie über die Fakultäten hinweg einheitlich und vergleichbar enthalten, die Daten werden regelmäßig eingegeben und aktuell gehalten und die Fakultäten sind von dieser Tätigkeit entlastet. Dies leistet neben einer hohen **Datenqualität** auch einen Beitrag zur **Datenaktualität**. Ein Informationssystem mit veralteten Daten schadet eher als dass es nützt. Eine Befüllung über Schnittstellen und die Aufnahme der Datenpflege in die existierenden Verwaltungs-Geschäftsprozesse setzen diese Forderung bestmöglich um. „Alte“ forschungsrelevante Daten beispielsweise ehemaliger Mitarbeitender werden jedoch nicht aus der Forschungsdatenbank entfernt.

Die Forschungsdatenbank wird eine Nutzerzahl im vierstelligen Bereich haben. Eine grundsätzliche Anforderung besteht darin, dass dieses System einfach im Umgang (**anwenderfreundlich**), weitestgehend selbsterklärend und **mehrsprachig** ist und durch Zugriffsbeschränkungen **Sicherheit** und Datenschutz realisiert.

## 2 Einheiten der Forschungsdatenbank

Als zentrale Einheit der Forschungsdatenbank wurden die **Einrichtungen** bestimmt. Sie bilden eine Hierarchie; von der Universität über ihre Fakultäten bis hin zur Professur (vgl. Abbildung 3 im Anhang). Der Einrichtungsbaum wird aus den Kostenstellen erstellt. Sobald eine der nachfolgend erläuterten Forschungsaspekte einen Eintrag erhält, wird dieser für die Forscher und Forscherinnen bzw. die Einrichtung ausgewiesen.

Abbildung 1: Forschungseinheiten im Rostocker Konzept der Forschungsdatenbank



Das heißt, es werden im Überblick die **Projekte** der Professur / des Forschenden, die **Publikationen**, die von ihm bzw. ihr betreuten **Promotionen**, **Erfindungen** und **Patente**, etc. unter der jeweiligen Überschrift angezeigt (vgl. Anhang, Abbildung 4). Mit **speziellen Ausstattungen** können **Kooperationsangebote** veröffentlicht werden. Hieraus oder aus Hochschulzusammenarbeiten ergeben sich dann **Kooperationen**. Die zusätzlichen, eher „weichen“ Faktoren der Forschungsdatenbank: die Organisation bzw. Mitarbeit in Programmkomitees **wissenschaftlicher Veranstaltungen**, ausgestellte **Messeexponate**, angefertigte Gutachten, erhaltene **Wissenschaftspreise**, **Firmenausgründungen**, **Gastwissenschaftler** und **Stipendiaten**, die ansonsten nicht erfasst werden, komplettieren die Aspekte von Forschung. Ein Teil der letztgenannten, ebenfalls attributgenau konzipierten Einheiten sind in HIS-LSF zur Zeit leider nur unter dem Punkt Forschungsparameter subsummiert umgesetzt, was nicht befriedigend ist. Großer Vorteil dieser Überblicksdarstellung für einzelne Einrichtungen und Personen ist, dass sowohl die Informationen der Lehre – aus dem Vorlesungsverzeichnisanteil in HIS-LSF – als auch die der Forschung gemeinsam dargestellt werden.

Ungeachtet dieser Einrichtungszentrierung gibt es zwischen den Forschungsaspekten weitere Verknüpfungen. So können alle Aspekte mit Projekten in Verbindung gebracht oder z.B. Promotionen und Patente einer dazu gehörigen Publikation zugeordnet werden. Es lassen sich auf diese Weise ebenso übergreifende Darstellungen auf Basis von Projekten oder Publikationen und anderen Einheiten der Forschungsdatenbank erstellen.

### 3 Datenquellen der Forschungsdatenbank

Die Kerninformationen zu Einrichtungen und deren Personen als Projektmitarbeiter, Autoren, Kooperationspartner etc. und die Projektgrunddaten werden an der Universität Rostock aus anderen Systemen übernommen, und zwar aus den Finanz- bzw. KLR- und dem Personalsystem (mit Umweg über das MetaDirectory). Die Grunddaten beinhalten Schlüsselnummern und Bezeichnungen, hierarchische Beziehungen (auch: Projekt - Teilprojekt-Beziehungen), Projekt-Förderdaten etc. Das heißt, an der Universität Rostock werden die Kostenstellen und Kostenträger universitätsweit und systemübergreifend identisch verwendet. In den Systemen, die für HIS-LSF ein Vor-System darstellen, wird für jeden Datensatz bei dessen Eingabe ausgewählt, in welchem System dieser genutzt werden kann. Hierzu dienen die sogenannten „Relevanzhäkchen“, die somit

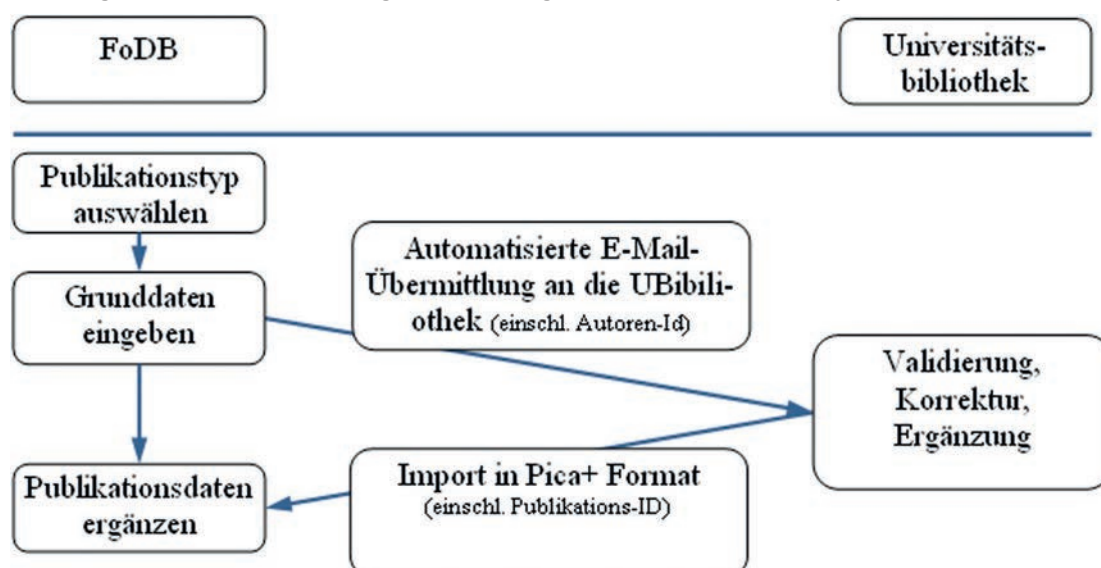
entscheiden, welche Datensätze nach HIS-LSF zu übertragen sind. Dadurch nicht übertragene Datensätze fehlen ggf. in der Hierarchie der Einrichtungen. Die entstandenen Lücken werden beim Import automatisch überbrückt.

Für die Forschenden sind über diese übertragenen Stammdaten hinaus weitere Informationen wie Projektbeschreibungen, Links zu Projektseiten, Sachgebietszuordnungen, Schlüsselwörter etc. ergänzbar. Die zu ergänzenden Informationen sind in HIS-LSF in Karteireitern organisiert. Projekten lassen sich die in diesen Projekten erstellten Publikationen, Promotionen, Erfindungen und Ausprägungen der weiteren Forschungsaspekte zuordnen. Im Überblick zu einem Projekt ergibt sich eine Übersicht der mit diesem Projekt zusammenhängenden Forschungsergebnisse. Projekte wiederum können Forschungsschwerpunkten zugeordnet werden, so dass sich die innerhalb von Forschungsverbünden erarbeiteten Forschungsleistungen gebündelt darstellen lassen.

Von Verwaltungseinrichtungen direkt in die Forschungsdatenbank eingegeben und teils auch dort verwaltet werden die Forschungsschwerpunkte, die Promotionen und die Habilitationen, ebenfalls ergänzbar um Kurzbeschreibungen, die dazu gehörigen Publikationen und anderes mehr.

Publikationen haben betreffs der Datenquelle eine Sonderstellung inne. Sie werden seitens der Fakultäten durch die Forschenden eingegeben. Publikationen mit in der Universitätsbibliographie erfassten Publikationstypen werden danach automatisch an die Universitätsbibliothek übermittelt, von den Mitarbeitenden dort (in ihre Systeme eingepflegt,) geprüft, ggf. ergänzt und korrigiert und die korrigierten bibliometrischen Daten an die Forschungsdatenbank zurück übermittelt (vgl. Abb. 2). Bei mehreren Autoren und daraufhin mehreren Eingaben ein- und derselben Publikation wird auf diese Weise eine Mehrfacherfassung vermieden. Grundlage für eine Eindeutigkeit der Autoren z.B. bei identischen Vor- und Zunamen ist allerdings, dass jede in der Forschungsdatenbank enthaltene Person die in der Universitätsbibliothek verwendete eindeutige Autoren-ID erhält. Auch bei Publikationen ist eine Vielzahl von ergänzenden Informationen wie Abstracts, Schlagworte und Links zum Volltext eintragbar. Außerdem lassen die mitgelieferten ID's eine Verbindung und weitere Recherche in den Systemen der Bibliothek und der dort verwalteten elektronischen Veröffentlichung zur Publikation (z.B. in einer Online-Zeitschrift) zu.

Abbildung 2: Schematische Darstellung der Validierung der bibliometrischen Daten zu den Publikationen



Zu den Kooperationen enthält das Konzept der Forschungsdatenbank eine Vielzahl von Beschreibungs-Attributen, die sie auf unterschiedlichste Weise charakterisieren können. Denkbar sind z.B.:

- a) national/international/Hochschul-intern,
- b) mit regionalen Bezügen (wie z.B. innerhalb der EU, im Ostseeraum),
- c) die Kooperationsbasis (per Kooperationsvertrag/innerhalb eines Projekts ),
- d) aufgrund der Kooperationspartner (mit weiteren Hochschulen/mit der Industrie),
- e) der Inhalt der Kooperation (Beratung/gemeinsame Forschung/Gerätenutzung),
- f) die Branche, mit der zusammengearbeitet wird.

Für das Instrument Forschungsdatenbank ist es also unerheblich, wie eine Kooperation definiert wird bzw. wann es sich um eine Kooperation handelt. Ebenso sind (auch im Zeitverlauf) unterschiedliche Vorgehen bei der Eingabe möglich. Ob alle Kooperationen von den Forschenden eingetragen werden oder ob es – ähnlich wie bei Publikationen – bestimmte Kooperationstypen gibt, die von einer verwaltenden Stelle, die die Kooperationsverträge prüft und verwaltet, eingetragen werden, muss jede Hochschule für sich klären und hängt von den jeweiligen Korrektheits- und Aktualitätsanforderungen ab. Auch die Kooperationsangebote und damit verbunden die Informationen über spezielle Forschungs-Ausstattungen werden zumindest an der Universität Rostock eher seitens der Forschenden selbst eingegeben, was Instrumente zur Aktualitätsprüfung (z.B. ein Ablaufdatum für ein Kooperationsangebot) bzw. Anreize für die Aktualität (wie regelmäßige automatische Erinnerungs-E-mails zur Aktualisierung, Anreiz durch Leistungsbewertungen, Aufnahme in die leistungsorientierte Mittelvergabe) erfordert.

Erfindungen und ggf. daraus beantragte und eingeschriebene Patente sind in der Konzeption so angelegt, dass sie mit HIS-LSF verwaltet werden könnten. Die Verbindung von einem Patent zur dazu gehörigen Erfindung lässt sich herstellen. Da es in Mecklenburg-Vorpommern einen Verwertungsverbund M-V gibt, der eine eigene Datenbank von der Erfindung bis zum Patent und dessen kommerzieller Verwertung für die Hochschulen des Landes führt, wird derzeit über eine Schnittstellen-Anbindung mit nachfolgender Datenpflege durch die verantwortliche Abteilung nachgedacht.

Alle weiteren „weichen“ Forschungsaspekte, gegenwärtig überwiegend unter „Forschungsparameter“ subsummiert, sind an der Universität Rostock mit der Eingabe durch die Forschenden mit den damit in Kauf zu nehmenden oben erwähnten Problemen der Validität und Aktualität geplant.

## 4 Perspektiven auf die Forschungsdatenbank

Die beschriebenen Forschungseinheiten genügen den Perspektiven der Stakeholder der Forschungsdatenbank.

1. Die erste und wichtigste Aufgabe ist die **Darstellung** von Forschungsausrichtungen, -aktivitäten und -ergebnissen, sowohl für universitätsexterne als auch für universitätsinterne Interessenten. Recherchen werden möglich, Forschung initiiert und Kooperationen angebahnt. Dabei gelten Einschränkungen, denn nicht jede Information ist geeignet, veröffentlicht zu werden, insbesondere wenn Interessen von Dritten dem entgegenstehen. Mittels eines Rollen- und Rechtemodell von HIS-LSF auf Basis universitätsweit gültiger Nutzeraccounts aus dem MetaDirectory der Universität Rostock wird sowohl die Ansicht auf Forschungsaspekte und deren Einzelinformationen (Attribute) eingeschränkt bzw. Informationen zu den Forschungsaspekten wie z.B. die vereinnahmten und verausgabten Projektmittel aus der Ansicht exkludiert. Zudem gibt es für einige der Forschungseinheiten Freigabekennzeichen, so dass jedes Projekt beispielsweise zur Ansicht erst frei geschaltet werden muss. Berichte unterstützen die unterschiedlichen Sichten und die Rechtedurchsetzung. Die Universität Rostock arbeitet aktuell an der Schnittstelle, um die Informationen aus HIS-LSF auf die mit TYPO3 erstellten Webseiten der Fakultäten, Professuren und



Wissenschaftler zu transferieren. Eine Ausgabeschnittstelle der Publikationen in z.B. BibTeX ist angedacht. Dies ist naheliegend, da sich die Publikationserfassung in HIS-LSF stark an BibTeX orientiert.

2. Eine weitere Sicht ist die der **Steuerung**, um Forschung gezielt zu fördern. In der Forschungsdatenbank sind spezielle Anforderungen, die aus Leistungsbewertungs- und Steuerungstools erwachsen, umgesetzt: Als Beispiel dafür sei hier die Berücksichtigung der Autorenreihenfolge in den Publikationen genannt, unabhängig davon, ob die Autoren Universitätsmitglieder sind oder nicht. Diese wird zur Bewertung einer Publikation als Leistung gebraucht. Es sind mehrere Wichtungsfaktoren für Publikationen vorgesehen und implementiert, die neben der Autorenreihenfolge z.B. den Publikationstyp, Seitenanzahlen, den jeweiligen Impact-Faktor oder beliebige andere Charakteristika bewertbar und vergleichbar machen. Mit Standardberichten können viele der steuerungsrelevanten Sichten abgedeckt werden.
3. Die **forschungsverwaltenden** Stellen nehmen in dem Konzept die wichtige Rolle ein, valide Daten in der Forschungsdatenbank zu pflegen. Mittels Rollen- und Rechtemodell werden auch diese Zugriffe genau angepasst. Ziel der Forschungsdatenbank ist zudem eine Unterstützung der Verwaltung bei der Erstellung von Forschungsberichten und der Ermittlung von Informationen für Rankings, Ratings, Presseinformationen, Anfragen der Landesregierung etc. Auch wenn die Forschungsdatenbank eine umfangreiche Sicht auf Forschung realisiert, kann und soll sie nicht die vollständige Verwaltung sämtlicher Forschungsaspekte unter deren weiteren Zielstellungen realisieren. Die Forschungsdatenbank eignet sich also nur bedingt zur alleinigen Verwaltung der Forschungsaspekte und es sind nur wenige Aspekte wie die Erfindungen, Patente oder die „weichen“ Parameter so umfangreich konzipiert, dass sie dies leisten könnten. Daneben werden immer Spezialsysteme z.B. in den Bibliotheken bestehen.

## 5 Kritische Betrachtung

Die Forschungsdatenbank der Universität Rostock befindet sich derzeit in der Pilotierungsphase. Dabei werden die Forschungsaspekte einzeln mit unterschiedlichen Fakultäten, Instituten oder Professuren eingeführt. Die Fakultäten sollen mit studentischen Hilfskräften dabei unterstützt werden, die ggf. bereits gesammelten Daten der Fakultät in die Datenbank einzupflegen oder zu übertragen. Damit werden ggf. kleinere technische Änderungen oder Korrekturen an der Forschungsdatenbank wie Begriffsänderungen einhergehen. Konzeptionell und in der Implementierung ausstehend sind Historisierungs- und Archivierungskonzepte, denn neben aller wünschenswerten Aktualität sollen einmal eingegebene Forschungsinformationen nicht verloren gehen, falls sich Strukturen ändern oder Personen aus der Universität Rostock ausscheiden. Die HIS versichert, diese in der Nachfolgesoftwaregeneration HISinOne realisiert zu haben.

Die Beispiele der Charakterisierung von Kooperationen oder auch der Publikationen weiter oben zeigen, dass die Forschungsdatenbank ein Instrumentarium ist, das erst mit Leben gefüllt werden muss und dies auf viele disziplinäre Arten. Aber nicht für alles ist in der Forschungsdatenbank auch eine Struktur vorgesehen, denn auch eine Forschungsdatenbank kann keine Semantikprobleme lösen.

Informationen, die nicht in die Forschungsdatenbank eingegeben wurden, lassen sich auch nicht extrahieren. Was sich simpel anhört, erweist sich in der Praxis als Balance zwischen dem Eingabe- bzw. Pflegeaufwand einerseits und dem Umfang an Abfragemöglichkeiten andererseits.

Als Beispiel für Semantikprobleme sei auch die Zahl der Promovierenden genannt. Während die Promotionen in der Forschungsdatenbank konzipiert und umgesetzt sind, wurden die Promo-



vierenden, deren Anzahl ggf. nach Geschlecht und Internationalität unter unterschiedlichen Gesichtspunkten von externen Organisationen erfragt werden, nicht modelliert. Ihre Zählung ist ein semantisches Problem und zöge ggf. einen erheblichen Pflege-Aufwand nach sich. Hintergrund sind die sehr unterschiedlichen Arten des Zustandekommens von Promotionen. Definitiv zählbar sind die Promovierenden erst bei Eröffnung des Promotionsverfahrens mit der Einreichung einer Promotionsschrift. Das ist für Aussagen über die Nachwuchsförderung der Universität, Promotionsdauern oder „Abbruchquoten“ von Promotionsvorhaben nicht hinreichend, da zeitlich zu spät. Promovierende sollen aber müssen sich laut LHG MV nicht als Studierende einschreiben. Die Universität Rostock greift bei solchen Anfragen dennoch eher auf die schwer interpretierbare, möglicherweise unvollständige oder möglicherweise auch zu große Zahl aus der Studierenden-Datenbank zurück als eine Verwaltungsmöglichkeit in der Forschungsdatenbank dafür vorzusehen.

Dem Nachteil, dass der Universität Rostock ein hoher (Zeit-)Aufwand für die Entwicklung, Tests, Erläuterungen zum Verständnis von Forschung gegenüber der HIS und ein finanzieller Aufwand zur Implementierung von HIS-LSF bei der HIS entstanden ist, steht der Vorteil gegenüber, dass das System nach den Wünschen der Universität gestaltet wurde, die sich als keineswegs Rostock-spezifisch herausstellten. Auch wenn es ein langer Weg bis hierhin war, was ohne Frage auch an der Stellung des Forschungsteils in der Prioritätenliste der HIS lag, ist das Ergebnis (von ausstehenden Umsetzungen der oben beschriebenen „weichen“ Forschungsaspekte abgesehen) insbesondere in Verknüpfung mit dem Vorlesungsverzeichnis ein Gewinn, da die Aspekte von Forschung und Lehre gemeinsam dargestellt werden.

Jedes System hat Grenzen. Eine Grenze von HIS-LSF, die nach Aussage der HIS mit dem Konzept zur nächsten Softwaregeneration der HIS grundsätzlich behoben sein soll, ist, dass Studierende nicht gleichzeitig mit den Mitarbeitenden in der Personentabelle geführt werden können. Somit ist z.B. das Forschungspersonal der Graduiertenkollegs – die Stipendiaten – oder auch die Autoren einer Promotion nur in einem Textfeld abgelegt. Das behindert die Suche und Auswertung sowie Berichtserstellungen, da diese Personen nicht auswählbar sind.

Die Struktur von HIS-LSF ist ebenso wie dessen generische Schnittstellen offen gelegt. Jede Bezeichnung in HIS-LSF kann an Hochschulspezifika auch mehrsprachig angepasst und Attribute und Karteireiter in der Such- und Eingabemaske rollenbasiert ausgewählt werden. Die Offenheit ist Vor- und Nachteil zugleich. Es lässt sich vieles selbst gestalten aber deren Anpassung und Pflege ist den Hochschulen überlassen und erfordert intime Kenntnisse des Systems und ggf. der angebundenen weiteren HIS-Systeme, da sich deren Philosophien nicht immer treffen. Die schlechte Dokumentation erleichtert die Arbeit leider nicht. Einen IT-Administrator für HIS-LSF insgesamt einzusetzen hat sich zumindest für die Einführungsphase als unerlässlich erwiesen. Von Vorteil für die Universität Rostock war, dass grundlegende (Schlüssel-) Daten bereits in allen Verwaltungseinrichtungen einheitlich verwendet werden, so dass die Datenübertragung ohne zusätzliche semantische Probleme umgesetzt werden konnte.

Zu den „lessons learned“ des Projektes gehört unbedingt, dass ein engagiertes Projektteam als Projekttreiber wichtig ist, aber zu einem so großen, komplexen Projekt unbedingt eine auskömmliche Zahl Projektpersonal gehört. Die große Zahl der „Stakeholder“ und aller in irgendeiner Weise Beteiligten über einen längeren Entwicklungs- und Einführungszeitraum im Boot zu behalten, bedarf großer zeitlich-kommunikativer Anstrengungen. Zu bedenken ist bei der Einführung auch, dass die Forschungsdatenbank an dessen Vorsysteme Anforderungen stellt, die dort teils als neue, zusätzliche Aufgaben nur für HIS-LSF erfüllt werden. Die Forschungsdatenbank ist also generell auf ein Mitdenken und Mitarbeiten vieler angewiesen.

## 6 Anhang

Abbildung 3: Forschungsdatenbank mit HIS-LSF (Administratorsicht) mit einem Ausschnitt aus der Einrichtungshierarchie der Universität Rostock (Bildschirmabzug)

The screenshot displays the administrative interface of the HIS-LSF research database. At the top, a navigation bar includes links for 'Meine Funktionen', 'Veranstaltungen', 'Einrichtungen', 'Räume und Gebäude', 'Personen', and 'Forschung'. Below this, a breadcrumb trail indicates the current location: 'Sie sind hier: Startseite > Forschung > Hochschulstruktur'.

On the left side, a vertical menu lists various functions such as 'Hochschulstruktur', 'Projekte', 'Neues Forschungsprojekt', 'Publikationen', 'Neue Publikation', 'Forschungsschwerpunkte', 'Kooperationsangebote', 'Kooperationen', 'Mittelgeber', 'Mittelgeberstruktur', 'Mittelgeberstruktur bearbeiten', 'Förderprogramme', 'Erfindungen', 'Patente', 'Tagungen/Workshops', 'Forschungspreise', 'Forschungsparameter bearbeiten', 'Promotionen/Habilitationen', 'externes Personal bearbeiten', 'Auswertungen und Berichte', 'Schlüsselstabellen Forschung', 'Berichtsübersicht mit Suchmaske', and 'Navigation ausblenden'.

The main content area features a warning banner: 'Hinweis: Sie befinden sich im Sommer 2012 und nicht im neu zu planenden Semester!'. Below the banner, it shows the current path: 'Sie sind hier: Einrichtungen, Forschungsprojekte und Publikationen'. It also provides options for 'Seitenansicht wählen: > kurz > mittel > lang' and 'Navigation: --> ganzen Baum anzeigen' and 'Link --> nächste Ebene anzeigen'.

The central part of the screen displays a hierarchical tree structure of the University of Rostock. The root node is 'Universität Rostock', which branches into 'Fakultät für Informatik und Elektrotechnik (IEF)'. This faculty further branches into 'IEF/LFE Elektrotechnik', which then branches into 'IEF/Institut für Automatisierungstechnik'. This institute branches into 'IEF/IAT/Automatisierungstechnik / Life Science Automation', which then branches into 'IEF/IAT/Prozeßmeßtechnik', which finally branches into 'IEF/IAT/Regelungstechnik'. Each node in the hierarchy is accompanied by a tree icon and buttons for 'PDF' and 'PUB'.

Abbildung 4: Beispielhafte Überblicksdarstellung aus HIS-LSF (Bildschirmabzug)

The screenshot shows the HIS-LSF web interface in a Mozilla Firefox browser. The address bar displays the URL: <https://lsf.uni-rostock.de/qisserver/rds?state=verpublish&status=init&vmfile=no&>. The page title is "Universität Rostock - PHF/IG".

The top navigation bar includes links: Startseite | Abmelden | Frau Astrid Lubinski | Sie sind angemeldet als: | in der Rolle: Administrator | Debug |. Below this, there are tabs for: Meine Funktionen, Veranstaltungen, Einrichtungen, Räume und Gebäude, Personen, and **Forschung**.

The left sidebar contains a tree view of functions, including: Hochschulstruktur, Projekte, Neues Forschungsprojekt, Publikationen, Neue Publikation, Forschungsschwerpunkte, Kooperationsangebote, Kooperationen, Mittelgeber, Mittelgeberstruktur, Mittelgeberstruktur bearbeiten, Förderprogramme, Erfindungen, Patente, Tagungen/Workshops, Forschungspreise, Forschungsparameter bearbeiten, Promotionen/Habilitationen, externes Personal bearbeiten, Auswertungen und Berichte, Schlüsseltabellen, Forschungsübersicht mit Suchmaske, and Navigation ausblenden.

The main content area displays the following information:

- Hinweis:** Sie befinden sich im Sommer 2012 und nicht im neu zu planenden Semester!
- Einrichtung:** PHF/Institut für Germanistik (IG)
- Funktionen:** EDIT
- Seiteninhalt:** Grunddaten | Administrator | Weitere Angaben | Strukturbaum
- Grunddaten:**

Bezeichnung	PHF/Institut für Germanistik (IG)	Einrichtungsart	Institut
Institutsnummer	273100	Kostenstelle	273100
- Weitere Angaben:**
  - Veranstaltungen
  - Räume
  - Mitarbeiter
  - Projekte
  - Publikationen
  - Promotionen
  - Kooperationen
  - Patente
- Strukturbaum:**
  - Universität Rostock
    - Philosophische Fakultät (PHF)
      - PHF/Institut für Germanistik (IG) ...
- Administrator:**



## **Ein Blick in den internationalen Raum**





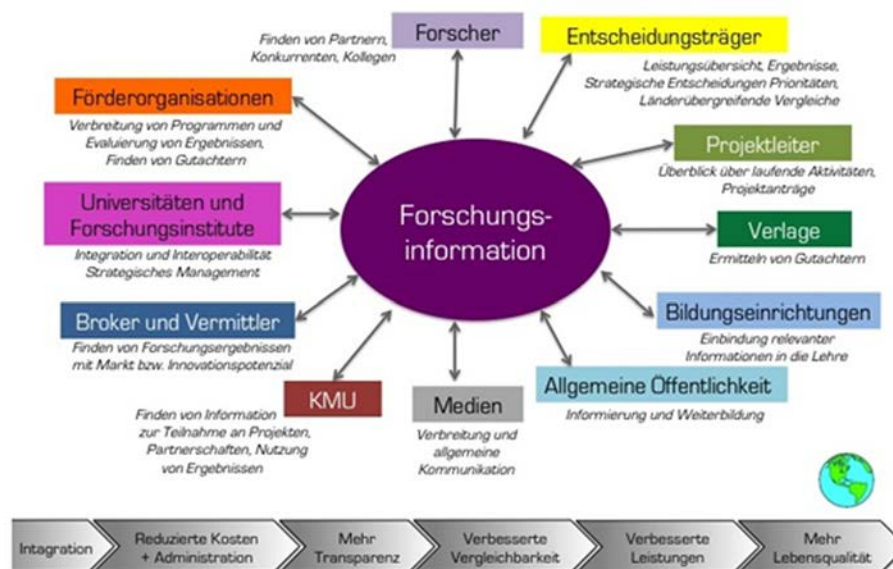
## Übersicht Systeme in Europa

Forschungsinformationssysteme sind von besonderer Bedeutung, wenn es um die Wiederverwendung wissenschaftlicher Information geht. Sie unterstützen sowohl deren Management, als auch Pflege und Bereitstellung, und ermöglichen damit einen vielfachen Zugang zu Daten und Informationen, welche im weiteren Forschungsumfeld von den Akteuren laufend benötigt werden. Akteure operieren dabei nicht nur in einem nationalen, sondern ebenso in einem europäischen sowie internationalen Kontext, und befinden sich zunehmend im Wettbewerb. Der vorliegende Beitrag leistet eine Übersicht zu Systemen in Europa, betrachtet im Umfeld von euroCRIS – einer gemeinnützigen Organisation, die sich der Entwicklung und Interoperabilität von Forschungsinformationssystemen widmet. Zunächst wird daher die Organisation euroCRIS vorgestellt und der von ihr entwickelte Standard CERIF kurz erläutert, bevor eine Übersicht europäischer Systeme aus verschiedenen Blickwinkeln erfolgt. Darüber hinaus wird auf internationale und verwandte Aktivitäten und ergänzende Initiativen hingewiesen.

### 1 Einleitung

Forschungsinformationssysteme sind von besonderer Bedeutung, wenn es um die Wiederverwendung wissenschaftlicher Information geht [Hornbostel 2006, Ivanovic 2011, Zimmermann 2002, Jeffery and Asserson 2010]. Sie unterstützen sowohl deren Management, als auch Pflege und Bereitstellung, und ermöglichen damit einen vielfachen Zugang zu Daten und Informationen, welche im weiteren Forschungsumfeld von den Akteuren – Forscher, Entscheidungsträger, Förderorganisationen, Universitäten, Forschungsinstitute, Kleinere und mittlere Unternehmen, Manager, Verlage, Medien, Vermittler, Bildungseinrichtungen, sowie der allgemeinen Öffentlichkeit – laufend benötigt werden (siehe Abbildung 1).

Abbildung 1: Nutzungsszenarien von Forschungsinformation



Die Akteure operieren dabei nicht nur in einem nationalen, sondern ebenso in einem europäischen sowie internationalen Kontext, und befinden sich zunehmend im Wettbewerb [EC 2010, Jacobs 2010]. Das Interesse und die Forderung nach Forschungsinformationssystemen reichen zurück

bis in die früheren 80er Jahre. Eine gemeinsame Arbeitsgruppe von UNESCO<sup>1</sup> und ICSU<sup>2</sup> untersuchte die Umsetzbarkeit eines globalen Forschungsinformationssystems und kam damals zu dem Ergebnis – *Study report on the feasibility of a world science information system* [UNISIST 1971] – dass ein solches System *sowohl notwendig als auch umsetzbar sei*<sup>3</sup>.

In Europa wurde wenig später CERIF<sup>4</sup> als Format für Forschungsinformationssysteme entwickelt. Sein Ursprung geht auf die Konferenz europäischer Universitätsrektoren (Conference of European University Rectors) zurück sowie zeitgleich auf die Empfehlung von Forschungsförderorganisationen der G7 Länder [Jeffery 2010]. Das erste CERIF Model wurde 1991 als eine EU-Empfehlung für Mitgliedsstaaten von der Europäischen Kommission (EC) veröffentlicht<sup>5</sup>, dabei stand die Entität *Projekt* als strukturiertes Objekt im Vordergrund. Dazugehörige Personen-, Organisationen- und weitere Information war unstrukturiert dem Projekt untergeordnet; *CERIF91* wurde damals als nicht genug formal und daher unattraktiv empfunden [Jeffery 2010]. Im Jahr 1997 beauftragte die Europäische Kommission eine neue Arbeitsgruppe, welche dann *CERIF2000* entwickelte. *CERIF2000* war bereits für Multilingualität und Einbindung eines Klassifikationssystems angelegt. Es erlaubte zudem Rollen und Typen sowie kontextuelle Beschreibungen durch zusätzliche Entitäten wie Person, Organisation oder Publikation und war als Anleitung für alle gedacht, die mit Forschungsinformationen und Forschungsinformationssystemen befasst sind<sup>6</sup>. 2002 übertrug die Europäische Kommission ihre Verantwortung für CERIF an euroCRIS, zur weiteren Pflege und Entwicklung, und zur Verbreitung seiner Nutzung. Seitdem wurde CERIF erweitert und verbessert, und mit *CERIF2004*, *CERIF2006*, *CERIF2008*, und jüngst *CERIF1.3* (2012) neuere Versionen veröffentlicht. CERIF wurde und ist in Systemen mehrerer europäischer Länder implementiert und genutzt und wird zunehmend auch jenseits von Europa beachtet. Im internationalen Diskurs spricht man auch von Current Research Information System(s) oder CRIS(s) für Forschungsinformationssysteme. Folgend wird nun zuerst euroCRIS vorgestellt und CERIF kurz erläutert, bevor eine Übersicht von Systemen in Europa erfolgt – im euroCRIS Umfeld – und auf aktuelle, weitergehende diesbezügliche Aktivitäten auch über Europa hinaus hingewiesen wird.

## 2 euroCRIS

euroCRIS ([www.eurocris.org](http://www.eurocris.org)) ist eine gemeinnützige Organisation – registriert in den Niederlanden – die sich der Entwicklung von Forschungsinformationssystemen und deren Interoperabilität widmet. Mit dem Präsidenten wird euroCRIS aktuell durch zwölf Vorstandsmitglieder repräsentiert, das Sekretariat finanziert sich über die Mitgliedsbeiträge. Neben sog. *Exekutiven* wie Strategie und Außenbeziehungen, Schatzmeister, Konferenzen und Website, sind *Arbeitsgruppenleiter*

---

1 The United Nations Educational Scientific and Cultural Organization (UNESCO).

2 The International Council of Scientific Unions (ICSU).

3 “[T]he Unesco-ICSU Central Committee has reached the unanimous conclusion that a world science information system, considered as a flexible network based on the voluntary co-operation of existing information services, is both necessary and feasible” [Wysocki and Tocatlán 1971]. Das aus der damaligen Studie resultierende, allgemeine Modell für wissenschaftliche und technische Kommunikation wurde später von [Søndergaard et. al. 2003] mit Blick auf neuere Informationstechnologien aktualisiert und als Rahmenwerk zur weiteren Analyse und zum Vergleich mit den Human- und Sozialwissenschaften, aber auch für weitergehende, die Disziplinen übergreifende Studien vorgeschlagen.

4 CERIF – The Common European Research Information Format – eine EU Empfehlung an Mitgliedstaaten: <http://cordis.europa.eu/cerif/>, <http://www.euroCRIS.org>

5 “Commission Recommendation concerning the harmonisation within the Community of research and technological development databases (6 may 1991). Official Journal L189, 13/07/1991 p. 0001-0034”: <ftp://ftp.cordis.europa.eu/pub/cerif/docs/cerif1991.htm>

6 CERIF 2000 – Common European Research Project Information Format (2000) European Commission: <http://cordis.europa.eu/cerif/> (last visited: March 2012)

für Projekte, CERIF, Architektur, Best Practice, Integration mit Repositorien sowie Linked Open Data aktiv. Der Vorstand wird von den euroCRIS Mitgliedern für jeweils zwei Jahre gewählt.

Einmal im Jahr organisiert euroCRIS ein *Strategisches Seminar* in Brüssel zur Kommunikation und zum Austausch mit strategischen Partnern. Im zwei-jährigen Turnus finden zudem die CRIS Konferenzen statt und zweimal jährlich werden Mitgliedertreffen von einzelnen Mitgliedern organisiert<sup>7</sup>. Im Rahmen der jeweiligen Arbeitsgruppen finden außerdem zusätzliche Treffen statt – finanziert und vermehrt initiiert durch eine euroCRIS- und euroCRIS-Mitglieder-Beteiligung in Drittmittel-geförderten europäischen sowie auch nationalen Projekten. Neue Mitglieder profitieren insbesondere vom fachlichen Austausch bei den Mitgliedertreffen. Aktive Mitglieder diskutieren über Mailinglisten und Foren und engagieren sich in Arbeitsgruppen. Per Newsletter werden alle Mitglieder regelmäßig über neueste Entwicklungen und Ereignisse informiert. Aktuell hat euroCRIS mehr als 100 institutionelle Mitglieder und unterhält Mitgliedschaften<sup>8</sup> in mehr als 40 Ländern. Unter den Mitgliedern befinden sich neben Universitäten, Forschungsinstituten und Akademien, auch Regierungs- und Förderorganisationen, sowie Firmen und Verlage. Über Europa hinaus sind auch in Australien, Kanada, China, Iran, Israel, Malaysia, Mexiko, Nigeria, Südkorea, und USA Mitglieder affiliert.

*Tabelle 1: euroCRIS – Strategische Partnerschaften mit Organisationen.*

Akronym	Name der Organisation
ESF	European Science Foundation
ERCIM	European Research Consortium for Informatics and Mathematics
CODATA/ICSU	International Council for Science / Commission on Data Access
EARMA	European Association of Research Managers and Administrators
JISC	The Joint Information Systems Committee
APA	Alliance for Permanent Access
ALLEA	All European Academies
CASRAI	Consortia Advancing Standards in Research Administration Information
VIVO	Enabling National Networking of Scientists

Strategisch strebt euroCRIS eine Verbreitung der CERIF-basierten Systeme auf der europäischen sowie auf internationaler Ebene an. Eine wichtige Rolle spielen dabei Mehrwerte für Mitglieder und die Pflege von Partnerschaften. Darunter fällt auch insbesondere die Zusammenarbeit mit der Open Access Repositorien (OAR) Gemeinschaft, weil diese Forschungsinformation in Publikationen bereits in großem Umfang zugänglich macht. Zur Demonstration des CRIS-Potentials nimmt die Teilnahme in Drittmittel-geförderten Projekten für euroCRIS eine immer wichtigere Rolle ein. Aktuell ist euroCRIS als Partner in folgenden EC-geförderten Projekten aktiv:

- Open Access Infrastructure for Research in Europe (OpenAIREPlus) ([www.openaire.eu](http://www.openaire.eu))
- An Infrastructure for Open, Linked Governmental Data Provision towards Research Communities and Citizens (ENGAGE) ([www.engage-project.eu](http://www.engage-project.eu))

Im Vereinigten Königreich wurden und werden unter anderem auch im Rahmen des sog. Forschungsinformationsmanagement (*Research Information Management* (RIM)) Programme Projekte gefördert, um CERIF in Bildungs- und Forschungseinrichtungen weiter zu verbreiten: “Projects to expand the community of the HE institutions and organisations which are using the CERIF

<sup>7</sup> euroCRIS Mitgliedertreffen (Übersicht): [http://www.eurocris.org/Index.php?page=members\\_meetings&t=1](http://www.eurocris.org/Index.php?page=members_meetings&t=1)

<sup>8</sup> euroCRIS Mitglieder: <http://www.eurocris.org/MembersListAll.php?order=cfName>

standard.”<sup>9</sup> Jüngst wurde ein sogenannter *Landschaftsbericht* über die Annahme von CERIF veröffentlicht<sup>10</sup>.

Desweiteren spielt CERIF in folgenden laufenden EU-Projekten eine Rolle:

- Virtual Open Access Agriculture & Aquaculture Repository (VOA3R) ([voa3r.eu](http://voa3r.eu))
- Mapping of the European Research Infrastructure Landscape (MERIL) (<http://www.esf.org/activities/science-policy/research-infrastructures/meril-mapping-of-the-european-research-infrastructure-landscape.html>)
- European Research Infrastructures Network of National Contact Points (EuroRIS-Net) ([www.euroris-net.eu](http://www.euroris-net.eu))

CERIF wird in kommerziellen Systemen und von CRIS Herstellern unterstützt:

- Aтира A/S vertreibt PURE ([www.atira.dk](http://www.atira.dk)) in Dänemark, Nordirland, England, Schottland, Finnland, Deutschland, Schweden, Belgien.
- Avedas AG vertreibt Converis ([www.avedas.com](http://www.avedas.com)) in Schweden, Deutschland, Niederlande, Belgien, Frankreich, Luxembourg, England, Schottland, Portugal, Schweiz, Botswana.
- Symplectic Ltd. vertreibt Seamless Research Management ([www.symplectic.co.uk](http://www.symplectic.co.uk)) in England.
- InterWay s.r.o. implementiert aktuell SK CRIS ([www.interway.sk](http://www.interway.sk)) in der Slowakei.
- uniCRIS operiert in den Vereinigten Staaten.

Thomson Reuters mit Research In View ([researchanalytics.thomsonreuters.com/researchinview/](http://researchanalytics.thomsonreuters.com/researchinview/)) war Partner im JISC-geförderten CERIFy<sup>11</sup> Projekt. Elsevier mit SciVal ([www.info.scival.com](http://www.info.scival.com)) war Partner im Projekt Snowball<sup>12</sup>.

### 3 CERIF – the Common European Research Information Format

CERIF ist eine EU-Empfehlung an Mitgliedstaaten. 2002 hat die Europäische Kommission euroCRIS autorisiert, CERIF zu pflegen und weiter zu entwickeln, sowie dessen Nutzung weiter zu verbreiten. Im vorliegenden Beitrag soll daher ein ganz kurzer Einblick in das Format gewährt werden.

CERIF kann auf verschiedenen Ebenen betrachtet werden. Wichtig ist es, das CERIF Konzept zu verstehen. Diesem folgen dann die Ebenen formaler (computer-lesbarer) Beschreibungssprachen. Zeichnet man die ARIS-(Architektur integrierter Informationssysteme)-Ebenen [Scheer 1992] nach, so spricht man von *Fachkonzept – Datenverarbeitungskonzept – Implementierung*. Diese Ebenen spiegeln sich auch in den CERIF Spezifikationen und öffentlichen Lehrmaterialien, als *Conceptual Level – Logical Level – Physical Level*. Der vorliegende Beitrag soll einen Anreiz schaffen sich weitergehend mit CERIF zu befassen, er reduziert sich hiermit auf Abbildung 2. Für weitergehende Information sei auf Spezifikationen [Jörg et. al. 2012 a, b] und Publikationen sowie auf die euro-

---

9 JISC infrastructure for education and research programme:  
[http://www.jisc.ac.uk/fundingopportunities/funding\\_calls/2010/10/grant1510.aspx](http://www.jisc.ac.uk/fundingopportunities/funding_calls/2010/10/grant1510.aspx)

10 Adoption of CERIF in Higher Education Institutions in the UK: A Landscape Study by Rosemary Russel: <http://www.ukoln.ac.uk/isc/reports/cerif-landscape-study-2012/CERIF-UK-landscape-report-v1.1.pdf>

11 CERIFy: <http://cerify.ukoln.ac.uk/>

12 Snowball: <http://www.projectsnowball.info/>

CRIS Website verwiesen<sup>13</sup>.

Abbildung 2: CERIF Entitäten und deren Verknüpfungen.

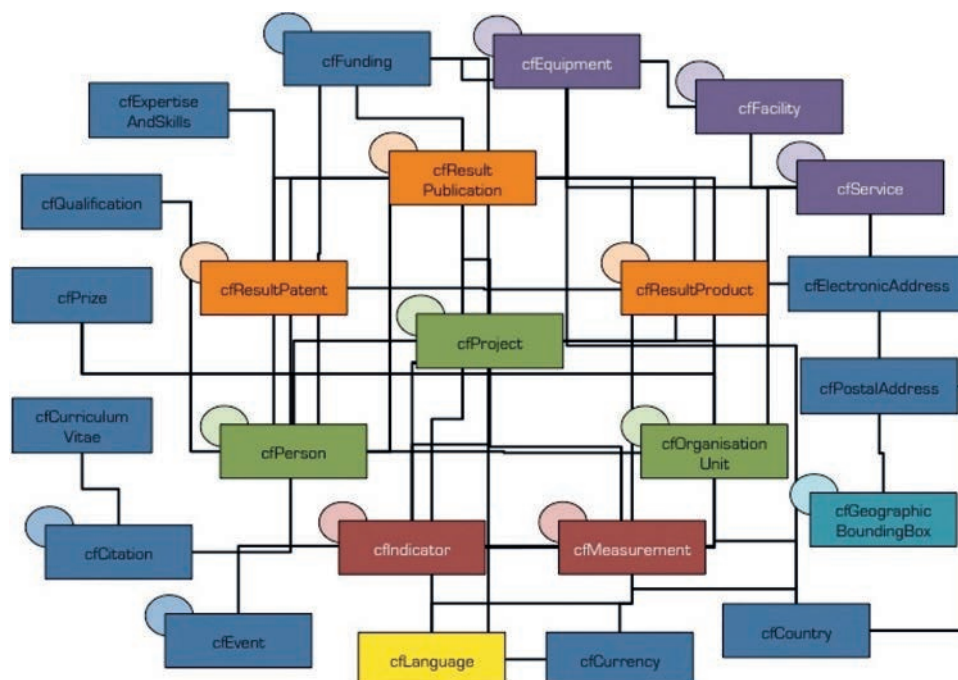


Abbildung 2 soll die CERIF-CRIS Idee vermitteln und zeigt im konzeptionellen Sinne diejenigen Entitäten, welche im Forschungsumfeld relevant sind. Die Abbildung gibt desweiteren einen Eindruck von den zahlreichen Verknüpfungen, die diese Entitäten miteinander unterhalten. CERIF ermöglicht es, diese Komplexität konsistent zu beschreiben und systematisch zu formalisieren, und trägt damit substantiell nicht nur zur Interoperabilität, sondern auch zur Nachhaltigkeit der Information und von Systemen bei. Vermehrt werden – nicht nur, aber auch – im Forschungsumfeld (formale) semantische Beschreibungen nachgefragt, deren Erörterung begrenzt sich jedoch im vorliegenden Beitrag auf die hinlänglich bekannten Stichworte: Linked Open Data (LOD), formale Vokabularen oder Wissensorganisationssysteme (KOS).

Nach der Vorstellung von euroCRIS und dem CERIF Format, werden im nächsten Abschnitt Systeme und Aktivitäten im europäischen CERIF-CRIS Umfeld präsentiert. Dabei werden sowohl nationale Systeme als auch Initiativen von Universitäten und Förderorganisationen vorgestellt, so wie diesen verwandte internationale Aktivitäten.

## 4 Nationale CERIF-basierte Systeme

Die Tabelle 2 gibt einen Überblick über CERIF-basierte Systeme auf nationaler Ebene. Es soll hierbei ein erster Eindruck vermittelt werden, über die Initiatoren, die inhaltliche systemische Abdeckung und deren Ausrichtung. Die Beschreibungen in der Tabelle sind in englischer Sprache gehalten und wurden weitestgehend von den Webseiten extrahiert. Für weiterführende Information sei auf diese verwiesen – eine entsprechende Linkliste befindet sich daher im Anhang.

<sup>13</sup> CERIF Releases: <http://www.eurocris.org/Index.php?page=CERIFreleases&t=1>



Tabelle 2: Nationale CERIF-basierte Systeme

Land	Akronym	Initiator	Inhaltliche Abdeckung	CERIF	Ausrichtung
NO	CRISin	Ministry of Education & Research; Ministry of Health & Care	Licensing/Consortium Agreements; Institutional Repositories and OA; Research Documentation: Results; Researchers; Projects; Units; open to all.	2011  FRIDA 1991	Value for Society by facilitating cooperation across areas; Compilation of information; High quality data as a base for Funding.
NL	NARCIS	SURF (Tender); Formerly University Initiatives led by Radboud University, Nijmegen	Researchers; Organizations; Publications; Thesis; Data-sets; NARCIS (Repository); Link to DAI (Digital Author Identifier) and DANS (Data Archiving and Networked Services); others.	Former (1991)	Formerly Inter-University Cooperation; Support Researchers and Research Managers; Decision for a single CRIS (2009); SURF went for EU Tender (2010).
DK	For- sknings databasen	Is part of Denmark's Electronic Research Library (DEFF) in cooperation with the Ministry of Science, Technology and Innovation, Ministry of Education, Ministry of Culture	People, Organisations, Publications to identify published research: journal articles, conference contributions, phd theses, conference presentations, lecture notes, published information, data sets, plans, projects.	(2008) (2006) with all Universities	Harvests data from universities and research institutions to document research output. Provide central collected access to all local CRISs (PURE-(CERIF)). The initiative behind also seeks to expose Danish research to a wider and international audience.
BE	FRIS	Flemish Ministry; Economy, Science and Innovation	Publications; Projects; Organizations; Researchers.	2008 2006	Accelerating the innovation value chain; Improved customer services (E-Government); efficiency and effectiveness of R&D policy; contribution to Flemish RI Space.
SE	ScienceNet	Joint leading Universities' Initiative	Research Activities funded by Vetenskapsrådet, FAS, Formas, VINNOVA, the Swedish Foundation for Strategic Research and Riksbankens Jubileumsfond, soon to expand to publications and patents.	2011	Further Integration with Data from internal University systems (contract or project DBs) and other relevant aspects of research activities.
SK	SK CRIS	Ministry for Education and Science, Research and Sports	R&D Subjects; University Institutes; Private or non-profit R&D Orgs; Projects; Registry of Researchers; R&D Results; EU Structural Funds	2011	Reporting; Decision taken in Autumn 2011; Planned going live in Autumn 2012.



HU	HUN-CRIS	Government Decision	R&D Projects; Researchers; Institutes.	1991	Information Supply by every R&D organisation receiving budgetary funding; CRIS a part of the National Innovation System.
SL	SICRIS	Slovene Research Agency	Research Organizations; Groups; Researchers; Projects; Programs;	2000	Data Collection by electronic forms; Future Plans: User Accounts; Mandatory Data for Research Organizations.
IS	ICERIS	Icelandic Research Council; University of Iceland	Currently offline; first ever CERIF-CRIS.	1991	Access to Information on Icelandic Research Projects and R&D Results.

## 5 CERIF in Universitäten und Förderorganisationen

Tabelle 3 gibt Einblick in nationale Initiativen und ausgewählte CERIF-verwandte Aktivitäten. Die Beschreibungen in der Tabelle sind teilweise in englischer Sprache und wurden weitestgehend von den Webseiten extrahiert, für weiterführende und ggf. aktuellere Informationen sei daher auch immer auf diese verwiesen. Beeindruckend sind aktuell sicherlich die Initiativen und Programme im Vereinten Königreich.

*Tabelle 3: CERIF in Universitäten und Förderorganisationen*

Land	Initiativen	Ausgewählte, CERIF-verwandte Aktivitäten	Weiterführende Links
DE	Exzellenz-initiative (2005)	<ul style="list-style-type: none"> <li>- Einführung integrierter Forschungsinformationssysteme (CRISs) an Universitäten (z.B. Kassel, Lüneburg, Münster, Karlsruhe).</li> <li>- Strategische Zusammenarbeit von Avedas und KIT - Workshop Forschungsinformationssysteme Institut für Forschungsinformation und Qualitätssicherung (iFQ) und Deutsche Initiative für Netzwerkinformation (DINI), organisieren FIS Workshop am KIT, Karlsruhe, in 11/2011)</li> <li>- Converis Anwendertreffen Deutschland an der Uni Münster</li> <li>- Kommerzielle Vertreiber: Avedas (Converis), Atira (PURE), Qleo (FactScience)</li> </ul>	<a href="http://www.exzellenz-initiative.de">http://www.exzellenz-initiative.de</a> ; <a href="http://gepris.dfg.de">http://gepris.dfg.de</a> ; <a href="http://www.his.de">http://www.his.de</a> ; <a href="http://www.dini.de/veranstaltungen/workshops/forschungsinformationssysteme/">http://www.dini.de/veranstaltungen/workshops/forschungsinformationssysteme/</a>  <a href="http://www.avedas.de">www.avedas.de</a> ; <a href="http://www.atira.dk">www.atira.dk</a> ; <a href="http://www.qleo.de">www.qleo.de</a>
AT	Wissensbilanz (2002)	Universität Wien: <ul style="list-style-type: none"> <li>- Research Activities Documentation (RAD)</li> <li>- bereitet EU Tender für CERIF-CRIS an der Uni Wien vor</li> <li>- Vormals AURIS-MM (Austrian Research Information System) basierend auf CERIF 2000</li> <li>- Aktuell CampusOnline (seit 1998 an der TU Graz entwickelt, bei 35 Unis und Hochschulen in Österreich und Deutschland im Einsatz (Campusmanagement))</li> </ul>	<a href="http://radinfo.univie.ac.at">http://radinfo.univie.ac.at</a>  <a href="http://www.campusonline.tugraz.at">http://www.campusonline.tugraz.at</a>

ES	National Research Evaluation	<p>SIGMA</p> <ul style="list-style-type: none"> <li>- gemeinnütziges Universitätskonsortium</li> <li>- entwickelt ARGOS (Research Information System)</li> <li>- Unterstützung: Spanish Government's Science and Innovation Ministry (MICINN)</li> <li>- GrandIR organisierte Workshop Serie im November 2011: Technical Session on CRIS Systems and Repositories at Open University Catalonia mit mehr als 60 Teilnehmern</li> <li>- CVN (Normalized CV Project)</li> </ul> <p>Kommerzielle Vertreter:</p> <ul style="list-style-type: none"> <li>- GrandIR</li> <li>- OCU</li> </ul> <p>Organisiert das nächste euroCRIS Mitgliedertreffen in Madrid, November 5th-6th, 2012</p>	<p><a href="http://www.gestionuniversitariaisigma.com">http://www.gestionuniversitariaisigma.com</a></p> <p>Workshop Serie:  <a href="http://www.gestionuniversitariaisigma.com/en/ultimas-noticias/236-jornada-tecnica-sobre-cris-y-repositorios.html">http://www.gestionuniversitariaisigma.com/en/ultimas-noticias/236-jornada-tecnica-sobre-cris-y-repositorios.html</a></p> <p><a href="http://www.grandir.com">http://www.grandir.com</a>  (Technical Support);  <a href="http://www.ocu.es">http://www.ocu.es</a>  (Universitas XXI CRIS);</p> <p>euroCRIS Members-Meeting:  <a href="http://eurocris.grandir.com">http://eurocris.grandir.com</a></p>
IT	Research Evaluation – Combine Research with Education	<p>CINECA</p> <ul style="list-style-type: none"> <li>- ein gemeinnütziges Konsortium (50 italienische Universitäten, plus National Institute for Oceanography and Experimental Geophysics (OGS); the CNR (National Research Council); the Ministry of Education, University and Research (MIUR))</li> <li>- entwickelt Universitätssystem U-GOV (Studentenmanagement Repository; HR, ERP);</li> <li>- Konvertierung nach CERIF für bestimmte Systemmodule</li> <li>- Organisierte das euroCRIS Mitgliedertreffen in Bologna im Mai 2011</li> </ul> <p>CILEA</p> <ul style="list-style-type: none"> <li>- ein gemeinnütziges Konsortium (11 italienische Universitäten plus Ministry of Universities and Research)</li> <li>- entwickelt Universitäts- und Forschungsinformationssystem</li> </ul>	<p><a href="http://www.cineca.it">http://www.cineca.it</a>;</p> <p><a href="http://eurocris.cineca.it">http://eurocris.cineca.it</a>;</p> <p><a href="http://www.cilea.it">http://www.cilea.it</a></p>
FR	Research Documentation – Teaching	<p>National Center for Scientific Research (CNRS)</p> <ul style="list-style-type: none"> <li>- INRA (French National Institute for Agricultural Research) veröffentlichte Information über CERIF and euroCRIS auf deren Blog im Februar 2012</li> <li>- Vortrag über CERIF und euroCRIS in einem französischen Seminar über Open Access und Interdisziplinarität am “Conservatoire National des Art et Métiers” in Paris</li> <li>- die Universität Lille 3 organisierte das euroCRIS Mitgliedertreffen in Lille im November 2011</li> </ul> <p>Kommerzielle Vertreter</p> <ul style="list-style-type: none"> <li>- Expernova</li> </ul>	<p><a href="http://www.cnrs.fr/comiternational/english/cn/index.htm">http://www.cnrs.fr/comiternational/english/cn/index.htm</a></p> <p><a href="http://ist.blogs.inra.fr/technologies/2012/02/27/cerif-unstandard-europeen-pour-decrire-les-informations-de-recherches-organisations-laboratoires-chercheurs-production-scientifique/">http://ist.blogs.inra.fr/technologies/2012/02/27/cerif-unstandard-europeen-pour-decrire-les-informations-de-recherches-organisations-laboratoires-chercheurs-production-scientifique/</a></p> <p><a href="http://dicen.cnam.fr/open-access-services-interdisciplinaire-et-expertise-oasie--479308.kjsp?RH=dicen">http://dicen.cnam.fr/open-access-services-interdisciplinaire-et-expertise-oasie--479308.kjsp?RH=dicen</a>  <a href="http://univ-lille3.fr/fr/">http://univ-lille3.fr/fr/</a></p> <p><a href="http://expernova.com">http://expernova.com</a></p>

CZ	Research Evaluation	<p>Zentrales CRIS koordiniert durch den R&amp;D&amp;I Council</p> <ul style="list-style-type: none"> <li>- entwickelt durch InfoScience Praha</li> <li>- Fokus auf Budgetierung, R&amp;D Programme, Projekte, Ergebnisse (Evaluation)</li> <li>- Leistungsbasierte Förderung</li> </ul> <p>Organisierte das euroCRIS Mitgliedstreffen in Prag, im November 2010, über InfoScience Praha.</p> <p>InfoScience Praha organisiert die CRIS 2012 Konferenz in Prag, 6.-9. Juni, 2012</p>	<p><a href="http://www.vyzkum.cz/FrontC-lanek.aspx?idsekce=633">http://www.vyzkum.cz/FrontC-lanek.aspx?idsekce=633</a></p> <p><a href="http://infoscience.cz">http://infoscience.cz</a></p> <p><a href="http://eurocris.infoscience.cz">http://eurocris.infoscience.cz</a></p> <p><a href="http://cris2012.org">http://cris2012.org</a></p>
FI	Centers of Excellence Programs (CoE)	<p>Academy of Finland</p> <p>University of Helsinki CRIS project</p> <ul style="list-style-type: none"> <li>- Initiative after 2005 Research Evaluation</li> <li>- University CRISs über Atira.</li> </ul>	<p><a href="http://www.aka.fi/eng">www.aka.fi/eng</a></p> <p><a href="http://www.aka.fi/en-GB/A/Centres-of-Excellence-/tuhat.halvi.helsinki.fi/portal/">www.aka.fi/en-GB/A/Centres-of-Excellence-/tuhat.halvi.helsinki.fi/portal/</a></p>
IE		<p>Trinity College Dublin CRIS</p> <ul style="list-style-type: none"> <li>- CERIF-based CRIS with DSpace Repository Integration</li> <li>- Partner in JISC-funded CERIFy project</li> </ul>	<p><a href="http://sonexworkgroup.blogspot.com/2010/04/crisir-integration-at-trinity-college.html">sonexworkgroup.blogspot.com/2010/04/crisir-integration-at-trinity-college.html</a></p>
RU		<p>CRIS of the Russian Academy of Sciences: RAS-CRIS</p>	<p><a href="http://www.asurid.ru">www.asurid.ru</a></p>
UK	Research Excellence Framework (REF)	<p>Ziel für REF und die Research Councils (eingeführt durch HEFCE) ist die Evaluation von Forschungsergebnissen</p> <ul style="list-style-type: none"> <li>- JISC: Commissioned technical study (EXRI): recommended CERIF</li> <li>- CERIF ein UK Standard (2010): HEFCE, UUK, RCUK, ARMA</li> </ul> <p>JISC fördert Serie von Projekten durch sog. RIM Programme:</p> <ul style="list-style-type: none"> <li>- RIM 3: CiA; IRIOS-2</li> <li>- RIM 2: CERIFy; BRUCE (was BRIM); MICE; IRIOS</li> <li>- RIM 1: CRISPool; BRIM; R4R</li> </ul> <p>Bericht veröffentlicht im März 2012: "Adoption of CERIF in UK HEIs: A Landscape Study"</p> <p>Kommerzielle Vertreiber:</p> <ul style="list-style-type: none"> <li>- Atira, Avedas, Symplectic</li> </ul>	<p><a href="http://www.hefce.ac.uk/research/ref/">http://www.hefce.ac.uk/research/ref/</a></p> <p><a href="http://www.jisc.ac.uk/media/documents/publications/reports/2010/Businesscasefinal-report.pdf">http://www.jisc.ac.uk/media/documents/publications/reports/2010/Businesscasefinal-report.pdf</a></p> <p><a href="http://www.jisc.ac.uk/whatwedo/themes/informationenvironment/researchinfomgt.aspx">http://www.jisc.ac.uk/whatwedo/themes/informationenvironment/researchinfomgt.aspx</a></p> <p><a href="http://isc.ukoln.ac.uk/2012/03/15/adoption-of-cerif-in-uk-heis-report-just-published/">http://isc.ukoln.ac.uk/2012/03/15/adoption-of-cerif-in-uk-heis-report-just-published/</a></p> <p><a href="http://www.symplectic.co.uk">http://www.symplectic.co.uk</a></p>
Europa und darüber hinaus		<ul style="list-style-type: none"> <li>- Weitere Aktivitäten und Informationen sind über die euroCRIS Website verfügbar. <a href="http://www.eurocris.org">http://www.eurocris.org</a></li> <li>- Consortia Advancing Standards in Research Administration Information (CASRAI) eine gemeinnützige Organisation in Ottawa, Kanada – initiiert durch Förderorganisationen mit dem Ziel die Wiederverwendung von Information zu verbessern: <a href="http://www.casrai.org">http://www.casrai.org</a></li> <li>- Enabling National Networking of Scientists (VIVO): <a href="http://www.vivoweb.org">http://www.vivoweb.org</a></li> <li>- Plataforma Nacional de Ciencia e Tecnologia: <a href="http://www.degois.pt/index.jsp?lang=en">http://www.degois.pt/index.jsp?lang=en</a></li> <li>- Plataforma Lattes: <a href="http://www.lattes.cnpq.br/english/conteudo/aplataforma.htm">http://www.lattes.cnpq.br/english/conteudo/aplataforma.htm</a></li> </ul>	

Verwandte Aktivitäten und Partnerschaften	Im November haben die CRIS Gemeinschaft und die Open Access Gemeinschaft eine gemeinsame Deklaration <i>Rome Declaration on CRIS and OAR</i> veröffentlicht: <a href="http://www.openaire.eu/en/about-openaire/publications-presentations/public-project-documents/doc_details/308-rome-declaration-on-cris-and-oar">http://www.openaire.eu/en/about-openaire/publications-presentations/public-project-documents/doc_details/308-rome-declaration-on-cris-and-oar</a>
---	--

## 6 Zusammenfassung

In der Summe lässt sich zusammenfassen, dass a) in Europa zahlreiche Systeme mit heterogenen Funktionen und Formaten als auch unterschiedlicher Ausrichtung existieren b) eine Wiederverwendung von Information und Systemen übergreifend gewünscht wird c) forschungsinformationsrelevante Prozesse vielfach repliziert werden. Eine nachhaltige Organisation von Forschungsaktivitäten und –ergebnissen auf institutioneller, nationaler, europäischer oder internationaler Ebene ist daher dringend notwendig. Um die gewünschte Interoperabilität von Systemen zu erreichen und damit eine Wiederverwendung von Forschungsinformationen zu ermöglichen, sollten Standardisierungsaktivitäten noch mehr Aufmerksamkeit zukommen.

### Anerkennung

Nicht zuletzt möchte ich mich vor allem beim DFKI (Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI GmbH)) für die Unterstützung meiner langjährigen Tätigkeiten im Rahmen von euroCRIS und CERIF bedanken. Darüber hinaus möchte ich dem euroCRIS Vorstand meinen Dank aussprechen für den immer umfänglichen Austausch im Rahmen der diesbezüglichen Tätigkeiten. Außerdem möchte ich dem Institut für Forschungsinformation und Qualitätssicherung (iFQ) und der Deutschen Initiative für Netzwerkinformation (DINI) meinen Dank zukommen lassen, für die Einladung zum Workshop Forschungsinformationssysteme, und damit überhaupt, die Initiierung des vorliegenden Beitrages.

## Referenzen

- Hornbostel, Stefan, 2006: From CRIS to CRIS. Integration and Interoperability, in: *Asserson, Anne Gams Steine / Simons, Eduard J. (Hg.): Enabling Interaction and Quality. Beyond the Hanseatic League. 8th International Conference on Current Research Information Systems. Leuven: University Press 2006, 29-38.*
- Ivanović, Dragan, 2011: Data Exchange between CRIS UNS, Institutional Repositories and Library Information Systems, in: *Arsovski, Slavko / Lazji, Miodrag / Stefanovi, Miladin (eds.): International Quality Conference, (5; 2011; Kragujevac) Conference Manual, Kragujevac: Faculty of Mechanical Engineering, Center for Quality, 371-378.*
- Zimmermann, Eric H., 2002: CRIS-Cross: Current Research Information Systems at a Crossroads, in: *Simons Eduard / Asserson, Anne Gams Steine (eds.): CRIS 2002 - Gaining Insight from Research Information. 6th International Conference on Current Research Information Systems. Leuven: University Press, 11-20.*
- Jeffery, Keith G. / Asserson, Anne Gams Steine, 2010: Special Issue: CRIS for European E-Infrastructure. *Data Science Journal*, Vol. 9, 2010.
- European Commission (ed), 2010: EUR24187 - Assessing Europe's University-based Research (EC 2010) - Expert Group on Assessment of University-based Research. Luxembourg: Publications Office of the European Union, 2010. DOI: 10.2777/80193.
- Jacobs, Neil, 2010: New platform is positive sign for research (Last visited: Jan 2012): Online: [http://www.researchinformation.info/news/news\\_story.php?news\\_id=711](http://www.researchinformation.info/news/news_story.php?news_id=711)
- UNESCO (ed), 1971: UNISIST; Synopsis of the Feasibility Study on a World Science Information System. Paris: Unesco, 1971.
- Wysocki, Adam / Tocatlán, Jacques, 1971: A World Science Information System: Necessary and Feasible. *Taxon*, Vol. 20, No. 4, August 1971, 603 -608.
- Søndergaard, Trine Fjordback / Andersen, Jack / Hjørland, Birger, 2003: Documents and the communication of scientific and scholarly information: Revising and updating the UNISIST model. *Journal of Documentation*, Vol. 59, No. 3, 278 - 320.
- Jeffery, Keith G., 2010: The CERIF Model as the Core of a Research Organisation. *Data Science Journal*, Vol. 9, 2010.
- Scheer, August-Wilhelm, 1992: Architektur integrierter Informationssysteme: Grundlagen der Unternehmensmodellierung. Berlin: Springer.
- Jörg, Brigitte / Jeffery, Keith / Dvorák, Jan / Houssos, Nikos / Asserson, Anne / van Grootel, Geert / Gartner, Richard / Cox, Mark / Rasmussen, Henrik Steen / Vestdam, Thomas / Strijbosch, Luc / Clements, Anna / Brasse, Valerie / Zendulkova, Danica / Höllrigl, Thorsten / Valkovic, L. / Engfer, Andreas / Jägerhorn, Martin / Mabey, Mahendra / Brennan, Niamb / Sicilia, Miguel-Angel / Ruiz-Rube, Ivan / Baker, David / Evans, Ken / Price Adrian / Zielinski Manfred, (2012a). CERIF 1.3 Full Data Model (FDM): Introduction and Specification (euroCRIS 2012).
- Jörg, Brigitte / Dvorák, Jan / Vestdam, Thomas / van Grootel, Geert / Jeffery, Keith / Clements, Anna, 2012b: CERIF 1.3 XML Data Exchange Format Specification. Towards the next Release (euroCRIS 2012).
- Jörg, Brigitte / Jeffery, Keith / van Grootel, Geert, 2011: Towards a Sharable Research Vocabulary (SRV) – A Model-driven Approach, in: *García-Barriocanal, Elena / Cebeci, Zeynel / Okur, Mehmet C. / Öztürk, Aydın (eds.): Metadata and Semantic Research. 5th International Conference, MTSR 2011, Izmir, Turkey, October 12-14, 2011. Proceedings. Berlin, Heidelberg: Springer, 256-268.*

## Anhang / Link-Listen

- *CRISin*, [www.cristin.no](http://www.cristin.no)
- *Pure Danish Research Database*, [www.forskningsdatabasen.dk](http://www.forskningsdatabasen.dk)
- *Converis*, [www.avedas.com](http://www.avedas.com)
- *PURE*, [www.atira.dk](http://www.atira.dk)
- *NARCIS*, [www.narcis.nl](http://www.narcis.nl)
- *FRIS*, [www.researchportal.be](http://www.researchportal.be)
- *ScienceNet*, [www.sciencenet.se](http://www.sciencenet.se)
- *SICRIS*, [sicris.izum.si](http://sicris.izum.si)
- *HUNCRIIS*, [nkr.info.omikk.bme.hu/ff004/show.do](http://nkr.info.omikk.bme.hu/ff004/show.do)
- *UK RIM Activities*, [www.jisc.ac.uk/whatwedo/themes/informationenvironment/researchinfomgt.aspx](http://www.jisc.ac.uk/whatwedo/themes/informationenvironment/researchinfomgt.aspx)
- *Scholarly Output Notification and Exchange (SONEX)*, [sonexworkgroup.blogspot.de/2010/05/summary-of-ongoing-deposit-related\\_9901.html](http://sonexworkgroup.blogspot.de/2010/05/summary-of-ongoing-deposit-related_9901.html)
- *RIS in European Countries*, [nispez.cvtisr.sk/userfiles/file/Aktivita%204.1/Annex%20-Information\\_system\\_as\\_an\\_implement\\_of\\_the\\_state\\_science\\_policy\(2\).pdf](http://nispez.cvtisr.sk/userfiles/file/Aktivita%204.1/Annex%20-Information_system_as_an_implement_of_the_state_science_policy(2).pdf)
- *CASRAI in 2 minutes*, [www.youtube.com/watch?v=Gmx7U9-i3Gg](http://www.youtube.com/watch?v=Gmx7U9-i3Gg)
- *European Science Foundation (ESF)*, [www.esf.org](http://www.esf.org)
- *European Research Consortium for Informatics and Mathematics (ERCIM)*, [www.ercim.eu](http://www.ercim.eu)
- *International Council for Science/ Commission on Data Access (CODATA/ICSU)*  
[www.codata.org](http://www.codata.org), [www.icsu.org](http://www.icsu.org)
- *European Association of Research Managers and Administrators (EARMA)*, [www.earma.org/](http://www.earma.org/)
- *The Joint Information Systems Committee (JISC)*, [www.jisc.ac.uk](http://www.jisc.ac.uk)
- *Alliance for Permanent Access (APA)*, [www.alliancepermanentaccess.org](http://www.alliancepermanentaccess.org)
- *All European Academies (ALLEA)*, [www.allea.org](http://www.allea.org)
- *Consortia Advancing Standards in Research Administration Information (CASRAI)*, [casrai.org](http://casrai.org)
- *Enabling National Networking of Scientists (VIVO)*, [vivoweb.org](http://vivoweb.org)



## Current Research Information Systems in the UK: an evolving network

### 1 Introduction

The Joint Information Systems Committee (JISC) has made a commitment to “help institutions to improve the quality, impact and productivity of academic research”<sup>1</sup> and this commitment has never meant more. Higher Education research in the UK has a major role that reaches a long way beyond our universities. Government research spending (£4.6 billion<sup>2</sup>, or approximately €5.5 billion, of government funding in 2012 for instance) is complemented by funding from charities (such as the Wellcome Trust) and private companies to produce a large and vibrant research community. The UK’s share of research papers published and citations (6.4% and 10.9% respectively<sup>3</sup>) demonstrate the intellectual impact of this community.

However, the benefits of this activity go beyond the purely intellectual. Research can improve human lives in many ways, from better educational practice to better quality of life for the ill, and is a significant source of innovation and economic growth<sup>4</sup>. This recognition has fuelled an increasing focus on research impact amongst policy makers and research funders which has shaped the research reporting landscape and become a growing concern as the economic downturn has continued.

Given the scale of investment in the UK research base, evidence of that impact, and evidence of efficiency savings and improved practices are crucial to protect that investment in years to come. This concern has sharpened in some quarters as the government’s higher education reforms begin to be implemented. From the academic year 2012-13, the funding awarded for teaching will be dramatically reduced, to be replaced by fees charged to students which will add a great deal of uncertainty and unpredictability to university finances while the new system is established. To some in the higher education community then, research has become their most stable source of income which has moved it even further up the agenda.

In order for universities and other research organisations to make the very best of their research, and to capture and demonstrate its intellectual, economic and social impact they have to interoperate with a large number of national systems. The Higher Education Funding Councils will be gathering information for their periodic review of research, the Research Excellence Framework<sup>5</sup> from January 2013. The seven Research Councils, each with a distinct disciplinary focus, have their own reporting systems to gather information about the outcomes from the projects and activities they fund, such as the Research Councils UK (RCUK) Research Outcomes System<sup>6</sup> or the Medi-

---

1 <http://www.jisc.ac.uk/aboutus/strategy.aspx>

2 <http://www.bis.gov.uk/assets/biscore/science/docs/a/10-1356-allocation-of-science-and-research-funding-2011-2015.pdf>

3 <http://www.bis.gov.uk/assets/biscore/science/docs/i/11-p123es-international-comparative-performance-uk-research-base-2011-summary.pdf>

4 <http://www.bis.gov.uk/assets/biscore/innovation/docs/i/11-1387-innovation-and-research-strategy-for-growth.pdf>

5 <http://www.hefce.ac.uk/research/ref/>

6 <http://www.rcuk.ac.uk/research/Pages/ResearchOutcomesProject.aspx>

cal Research Council<sup>7</sup> and Science and Technology Facilities Council's e-Val<sup>8</sup> systems, which collect information from researchers and their institutions. This means that it is quite plausible that for a single large piece of research, an individual researcher may have to interact with three distinct systems in order to meet reporting requirements.

However, the challenge does not end there. Just to give a sense of the complexity of the research information landscape in the UK, researchers can also rely on the RCUK Joint e-Submission (Je-S) system<sup>9</sup>, the Higher Education Statistics Agency (HESA)<sup>10</sup>, UK PubMed Central<sup>11</sup> and the Royal Society e-GAP<sup>12</sup> as systems that should be 'taken for granted' in the medium term, and there are others in use and being built.

In this context, JISC's role has been to seek to increase the flow of validated research information across institutions, to reduce the burden on researchers and administrators and to create efficiency savings wherever possible. Given the sheer scale of the research sector in the UK, this is quite a challenge.

## **2 Research Information Management: a programme of change**

The Research Information Management (RIM) programme at JISC has been undertaken as part of the Digital Infrastructure team, which in turn forms a part of JISC's innovation (as opposed to services, for instance) activity. The Digital Infrastructure team focuses on funding projects to develop new or emerging technologies, to embed existing technologies in institutions and to explore the ways in which new technologies can form the basis for new systems and services for the UK higher education sector.

As a part of this group of activities, RIM has benefitted from the sizable human network surrounding research activity in the UK. In fact, it is clear to see that one of the defining features of the strand is partnership, as exemplified by the membership of the UK Research Information Management Group (RIM Group)<sup>13</sup>. This group represents many of the organisations most active in research in the UK, and has played an exceptionally productive role in drawing together the ideas and support that have enabled the RIM programme to progress.

The other defining feature of the programme is one often associated with large partnerships and that is complexity. As noted in the introduction above, researchers have to interact with many different organisations and systems in the course of their careers, and the systems have not historically interoperated on any large scale. Also, of course, research is an international activity and there are international partnerships to factor in to the equation. Some of the key organisations for the RIM programme have been international standards bodies working in the area of research information, such as euroCRIS<sup>14</sup> or the Canadian Consortia Advancing Standards in Research Administration Information (CASRAI)<sup>15</sup> organisation.

There are also different priorities and needs across the group of stakeholders on the RIM Group

---

7 <http://www.mrc.ac.uk/Achievementsimpact/Outputsoutcomes/e-Val/index.htm>

8 <http://www.stfc.ac.uk/About+STFC/18664.aspx>

9 <https://je-s.rcuk.ac.uk/>

10 <http://www.hesa.ac.uk/>

11 <http://www.ukpmc.ac.uk/>

12 <https://e-gap.royalsociety.org/>

13 <http://rim.jisc.ac.uk/>

14 <http://www.eurocris.org>

15 <http://casrai.org/>

and balancing these and building consensus on each step the programme has taken has been essential. Managing this complexity and these partnerships is both time-consuming and rewarding, and has led to some remarkable projects and progress in a relatively short period of time.

Many of the projects have been centred on the use, implementation and evolution of the Common European Research Information Format (CERIF), which has added another layer of complexity to the programme. CERIF is an extremely complex and flexible standard, which is necessary as it has to capture the reality of research activity, which is itself, of course, extremely complex.

This focus on CERIF dates back to the earliest discussions of the RIM Group. One of the first priorities of the group was to identify an efficient way of standardising the exchange of research information. It needed to function at every level of the research process, from researchers recording and managing information about their work within universities, to institutions sharing information with one another during a research collaboration, for example, or as researchers move around during their careers, to institutions reporting information in bulk to funding bodies.

JISC commissioned an investigation on the options available on behalf of the group, and the Exchanging Research Information in the UK report<sup>16</sup> was published in 2010 and found that CERIF was the best available option for all these use cases. The group then asked JISC to conduct a further exploration of the potential cost savings to be gained by using CERIF, as a solid business case would need to be made to justify what could be a substantial investment across the sector in adapting old systems as well as procuring or developing new ones.

The resulting report, the *Business case for the adoption of a UK standard for research information interchange*<sup>17</sup>, examined the costs of research reporting to the UK higher education sector, and used examples of CERIF CRIS (Current Research Information System) to CERIF CRIS exchange from other European countries to estimate the scale of the potential savings. The predecessor to the REF, the Research Assessment Exercise 2008, cost the UK more than £47 million (£56 million) or the equivalent of £1,127 (£1,350) per researcher submitted during the exercise. The analysis presented in the report proposed that, by using CERIF to CERIF exchange, savings could be made in the region of 25 to 30% for most institutions. This presented a powerful case for investment in CERIF.

Prior to this, the most significant implementations of CERIF in the UK were the University of St Andrews' instance of the Pure CRIS system, and the CERIF4REF plug-ins for popular repository software platforms developed by the Readiness for REF project<sup>18</sup>, led by King's College London, which had developed a CERIF schema for the reporting of publications data to meet REF requirements. It was clear that much more work needed to be done to make the leap from a few isolated projects and systems to a nationwide engagement with the standard.

The first formal group of RIM projects funded by JISC sought to explore both the UK research information landscape as it was, and to experiment with and assess the use of CERIF. As an example, the CRISPool project<sup>19</sup>, led by the University of St Andrews, experimented with the use of CERIF-XML to integrate heterogeneous research information from several Scottish Universities into a single portal. One of the key findings of the project was that CERIF-XML as an exchange format carried significant overheads which needed to be understood and reduced.

Other projects examined other aspects of research information management. The "Developing

---

16 Rogers, N. and Huxley, L. and Ferguson, N. (2010) Exchanging Research Information in the UK: final report. Project Report. Available at: [http://ie-repository.jisc.ac.uk/448/1/exri\\_final\\_v2.pdf](http://ie-repository.jisc.ac.uk/448/1/exri_final_v2.pdf) [Accessed 13/2/12]

17 Bolton, S. (2010) Business case for the adoption of a UK standard for research information interchange. Report to JISC. Available at: <http://www.jisc.ac.uk/media/documents/publications/reports/2010/Businesscasefinalreport.pdf> [Accessed 13/2/12]

18 <http://r4r.cerch.kcl.ac.uk/>

19 <http://www.crispool.org/>

tools to inform the management of research and translating existing good practice” project<sup>20</sup>, led by Imperial College London in partnership with Elsevier, undertook research which found that the market in software and tools to support research information management activities was fragmented, and that few institutions in the UK had mature and comprehensive systems to manage their research administration. This clearly represented a major barrier to the realisation of efficiency savings, and made the case for improved tools and systems to be developed in a more comprehensive way.

This project went on to spark an ongoing collaboration between the team and Elsevier, which has since resulted in the launch of Project Snowball<sup>21</sup> which aims to help universities benchmark their performance across a range of research activities and is developing standardised metrics to enable this to happen.

The findings of the first round of RIM projects were vital in assessing the current condition and extent of UK research information management practices, and in prioritising areas of development for CERIF. As a result, the second group of projects, which were conducted in 2011, explored the use of CERIF for particular business purposes.

These included a range of reporting requirements, such as recording research impact indicators (the Measuring Impact under CERIF project<sup>22</sup> (the outputs of which extended CERIF and were adopted by euroCRIS)), and the CERIFy project<sup>23</sup>, led by UKOLN at the University of Bath with a partnership that included Thomson Reuters who ‘CERIFied’ their InCites<sup>24</sup> tool. These projects also exemplify a common theme across the strand which is partnerships with commercial organisations, either vendors of informational products or research management systems. This has been a crucial component of JISC’s work with the higher education sector in this area, as working with both the supply and demand side of research information has provided participants in the programme with a detailed understanding of the needs, concerns and aims of all their partners.

In order to share this information and understanding with the wider UK higher education community, most of whom have not taken part in JISC funded projects in this area, work was undertaken to improve the dissemination of project findings and lessons. These ‘synthesis’ projects draw together key lessons from RIM projects and activities and package them for easier understanding for a variety of audiences.

These projects are based at JISC infoNet, who have created a research information management resource<sup>25</sup> aimed at both senior managers in universities and those with responsibility for managing or administering research information systems and activities, and at UKOLN, which has taken on the task of providing more technically focussed information for implementers and developers of CERIF systems<sup>26</sup>. The work of these projects has been essential in getting the messages and lessons from the projects out to the UK research community, and has also helped to shape the ongoing work of the programme by providing JISC with additional analyses of the programme’s progress and impact. Without this focus on communicating the experiences of the projects to a wider audience, the programme would be much less effective in driving change forward.

---

20 Green, J. (2010) Research information management: Developing tools to inform the management of research and translating existing good practice. Project report. Available at: <http://www.researchdatatools.com/downloads/2010-research-information-management-2.pdf> [Accessed 13/2/12]

21 <http://www.projectsnowball.info/>

22 <http://mice.cerch.kcl.ac.uk/>

23 <http://cerify.ukoln.ac.uk/>

24 <http://researchanalytics.thomsonreuters.com/incites/>

25 <http://www.jiscinfonet.ac.uk/infokits/research>

26 <http://www.ukoln.ac.uk/rim/>

This emphasis on driving forward change is exemplified by current projects, such as the Research Management and Administration System (RMAS)<sup>27</sup> project, which is a collaborative venture between the Universities of Exeter, Kent, and Sunderland to procure, develop and implement a cloud-based research management and administration system. Once again, partnerships with commercial system suppliers is a central part of the project's approach, as vendors of particular products have joined a procurement framework, agreeing to meet certain conditions, such as the use of CERIF for improved interoperability, in order to join. Universities seeking to replace all or part of their research information management systems can use the framework to identify approved suppliers and can purchase systems from those suppliers much more quickly than when using ordinary procurement rules. This means that they can deploy systems more efficiently, and has increased the awareness of CERIF amongst commercial software developers.

The ongoing emphasis on CERIF extended to the third round of RIM funding, which was designed to implement CERIF to CERIF exchange at a group of institutions. These projects are designed to generate real evidence of the benefits of CERIF in support of the original business case report and once more emphasise broad partnerships with research funders, commercial companies and institutions.

The Integrated Research Input and Output System 2 (IRIOS 2) project<sup>28</sup> builds on the previous IRIOS<sup>29</sup> project, led by the University of Sunderland during the second phase of RIM activity, which utilised an existing platform for collaborative engagement with the local business community, the Universities for the North East Information System (UNIS). IRIOS produced a proof-of-concept demonstrator, which supported the import and export of Research Council data in CERIF, and the linking of research grant information and publications together. In IRIOS the UNIS platform was successfully adapted and extended to include basic research grant information taken directly from the Research Councils.

The aims of the follow-up IRIOS-2 project are to go beyond the work done in the first project by extending the IRIOS proof of concept demonstrator to better handle non-standard research outputs, such as new forms of publication and non-textual materials, as well as multiple authors and roles. In order to do this, the project is developing a CERIF based business to business, two-way communication platform using enterprise serviced bus technology (which is also being used by the RMAS project), designed to enable universities to better collaborate by improving the communications between their systems. The project will also enable Research Councils and universities to reconcile financial information about projects held in their various systems, reducing the time and effort taken to ensure that information across research partners is accurate and consistent.

Another current project which seeks to build directly on the previous work of the RIM programme is the CERIF in ACTION (CiA) project<sup>30</sup>. This project draws together the partners from the CRISPool project with commercial companies (Atira, EPrints services and Symplectic) to develop a standard CERIF-XML schema and agreed vocabularies building on work done in earlier JISC projects such as Readiness4Ref, CRISPool, CERIFy, and IRIOS to ensure that wherever in the UK CERIF is implemented it can be done in a consistent, valid way. This will help the UK to make the best use of the power of CERIF by ensuring that its inherent flexibility does not result in a different version of CERIF in every institution.

The project will also build standard, production-ready plug-ins to import and export data in this agreed format for CRIS, Repository and Research Council software and demonstrate the use of

---

27 <http://www.exeter.ac.uk/research/rmas/>

28 <http://irios2.wordpress.com/>

29 <http://www.irios.sunderland.ac.uk/>

30 <http://cerifinaction.wordpress.com/>



these plug-ins in live systems by exchanging data between partner institutions, for example when an researcher moves to a new Institution, and by uploading grant-level information to the RCUK Research Outcomes System.

CiA aims to create a roadmap setting out the costs and benefits for adoption of this standard CERIF-XML approach for other Higher Education Institutions and other use cases, in particular the development of an open source aggregator and portal for use by research pools and other consortia. This should not only support more efficient collaborative working, but should help other universities and funders to implement and exploit information and software systems using CERIF.

These current projects demonstrate the pace at which the programme has evolved. In just two years, projects have gone from exploring the research information landscape and testing limited exchanges using CERIF to building live systems and services, engaging with the majority of the software vendors actively promoting products and systems for research information management and building production ready tools to address some of the most common research information exchange formats. At the same time, the number of institutions in the UK using CERIF has gone up from a handful in 2010 to more than 40 in 2012, and many more will begin to use CERIF once the current projects have completed their development.

It is worth noting, however, that while there has been a strong focus on CERIF in the evolution of the RIM programme, this has not been the only area of activity. The intelligence gathered during the first phases of the programme revealed pressing issues that could not be addressed simply by improving interoperability, and this is why the programme has also taken on projects designed to address particular issues and challenges.

### **3 Research Information challenges: addressing the needs of UK higher education**

Whilst driving forward development and adoption of CERIF, the projects and the support teams at JISC, infoNet and UKOLN have also been exploring key related issues that have emerged as pressing concerns for the UK higher education sector. For example, the projects in the second RIM phase demonstrated that research reporting, from institutions to funders, was one of the key use cases for research management systems. Research reporting certainly represents one of the major causes of the bulk exchange of structured research information, and the exchange of information was the primary issue explored at the beginning of the programme.

In light of this, JISC has funded a feasibility study to investigate the possibility of a national shared service to facilitate reporting of research information. While it is too early to anticipate the results of this study (it is not due to report until November 2012), it is clear that any reduction in the burden on institutions, either directly financial or in terms of staff time, could offer substantial savings.

The study is being undertaken by a partnership including King's College London, Brunel university, Exeter University, The British Library, Cottagelabs Ltd. and euroCRIS, and is called the UK Research Information Shared Service (UKRISS)<sup>31</sup>. The project will perform a detailed feasibility study of the requirements for a national research information management service, which will provide an integrated approach to the gathering, analysis and dissemination of this information as required by funders. This will begin with a requirements gathering study on a representative sample group of research organisations, and funding bodies including UK funding councils and charities. The project team will then scope the metadata requirements for such a service, basing this on CERIF standard to ensure interoperability with existing and emerging systems. This scoping will be used to generate a detailed specification of the data mappings to the CERIF standard

---

<sup>31</sup> <http://ukriss.cerch.kcl.ac.uk>



necessary to enable reporting of research information at a national level and detailed specifications for the technological requirements for a shared service, and the resource implications that such a service would imply.

The feasibility study will recommend the most appropriate, sustainable and cost-effective solution for research reporting across UK Research Organisations. If the study findings suggest that it would be appropriate, then the team will go on to build and evaluate a pilot, proof-of-concept system for national research information reporting. In order to ensure that it as easy as possible for institutions to engage with any resulting service or system, JISC has asked the team to base their work on open standards, and where practicable open source, and to provide for a wide range of institutions to make sure that every participant in UK research can benefit from the project.

The system will build directly on the technologies used or developed in previous projects when this is the most cost effective solution, particularly RMAS, which has already developed a framework for exporting data in CERIF format for processing by applications running in the cloud, although the exchange tools and CERIF mappings and implementations from IRIOS 2 and CiA will also prove useful.

The aim of this project is ultimately to create a sustainable, scalable service that will facilitate research reporting and support both large research institutions with well developed CRIS systems and smaller institutions with fewer active researchers and no special software in place to manage their research information. The cost of research reporting is higher in the large institutions, but the burden is proportionally much greater in smaller institutions where the ratio of staff time to research income is less favourable. If the final product of UKRISS can scale up to a national system, and can lower the cost and burden of reporting, it will release a great deal of resource for research itself.

A related concern amongst the UK research community is that of research impact. As noted above, all the major research funders in the UK have explicitly defined impact as a component of their reporting requirements, and the potential for academic research to galvanise economic activity has become a source of some interest amongst policy makers, as demonstrated by the UK Government's 'Strategy for Innovation and Growth'<sup>32</sup> and such international undertakings as the European Union and United States roadmap to measuring the results of investments in science ('Bellagio Statement')<sup>33</sup>.

While the Bellagio statement focuses on measurable interactions between the research and commercial worlds, using intellectual property such as patents, the Higher Education Funding Council for England has defined impact as "any social, economic or cultural impact or benefit beyond academia that has taken place during the assessment period, and was underpinned by excellent research produced by the submitting institution within a given timeframe. Submissions will also include information about how the unit has supported and enabled impact during the assessment period."<sup>34</sup> This impact will be evidenced via case studies, and researchers are allowed to cite evidence going as far back in time as 1993 to support their claims.

This has led to a scramble across the UK to gather and assess a range of evidence for demonstrating impact, and has fuelled an ongoing debate as to whether qualitative evidence, rich in context, is the best evidence or whether qualitative measures or metrics which can be said to offer some objectivity are better. Views vary across disciplines, as do the kinds of evidence available. This is why JISC sought to fund a research project to examine the available evidence for research impact.

Definitions, Evidence and Structures to Capture Research Impact and Benefits (DESCRIBE) is a

---

32 See footnote number 4 above.

33 <http://www.iscintelligence.com/event.php?id=48>

34 [http://www.hefce.ac.uk/research/ref/pubs/2011/01\\_11/](http://www.hefce.ac.uk/research/ref/pubs/2011/01_11/)

12 month project based at the University of Exeter. Given that this is an emerging area of practice DESCRIBE will undertake a rigorous assessment of current standards relating to the evidence of impacts arising from research. This assessment should lay the foundations for an understanding of those areas where research impact is less well understood or demonstrated and make specific recommendations for their future development. The aim of the project is to embed a comprehensive and more ‘nuanced’ understanding of impact and its evidence.

The project will seek to draw together leading expertise by working with specialists in the field from both the commercial and academic sectors. Working with the major UK research funders, JISC intends this project to both support universities in meeting current reporting requirements and to help research funders demonstrate the value of their enormous investment in research with increasingly strong evidence of its value to society, culture and the economy.

Another substantial issue that touches upon many of the activities under the research information banner is that of identifying individual researchers consistently, across different organisational systems and in different contexts. The RIMG named unique identifiers as a pressing concern in 2010, a concern which has been echoing around the world for some time.

Unique identifiers for individuals involved in research are a central component of many other processes. An identifier solution would enable the comprehensive disambiguation of authors. If this were to be achieved then assigning publications to a specific individual would be possible automatically on both a national and global scale. Identifiers would also support the improved career tracking of researchers, from post-graduate research to emeritus professor, across institutions. This would serve funders as well as individuals, as it would enable funders to identify disciplines where the active researcher population is aging, for instance, meaning that a concerted effort to support younger researchers and encourage students to enter the research profession could be necessary. Researchers also stand to gain substantially from a reduced administrative burden, as identifiers could be used to enable easier entry on forms for funding and appraisal and to support the consistent clarification of staff institutional affiliations. This could, for example, be used to confirm eligibility for access to international facilities.

In light of this, JISC convened a ‘task and finish’ group, consisting of representatives from key organisations and bodies from the UK higher education sector, alongside experts and researchers, to draft a set of recommendations for the UK in how to implement an identifier solution that will meet the needs of the sector now and in the future. Organisations represented on the group included the Association of Research Managers and Administrators (ARMA), the Higher Education Funding Council for England, the Higher Education Statistics Agency (HESA), the Medical Research Council (MRC), the Natural Environment Research Council (NERC), the Universities and Colleges Information Systems Association (UCISA) (an organisation representing senior IT professionals in higher education) and the Wellcome Trust.

The group also benefitted from the input of several researcher viewpoints which were offered by Dr Simon Coles (Head of the National Crystallography Service, University of Southampton) and Dr Cameron Neylon (Senior Scientist, Science and Technology Facilities Council). The group was supported by staff from the Innovation Support Centre at UKOLN, University of Bath and JISC.

The group met in May 2011, and discussed the information they would need to arrive at a set of recommendations. This led to the creation of a series of eight reports which were used to inform and guide discussion of the challenges, issues and opportunities facing the sector in developing a unique identifier solution for research activity. One of the primary challenges identified in the course of developing this process was that of achieving sufficient political support for any solution to gather momentum and gain a critical mass of users. In order to facilitate this support, the group will conduct a sector wide consultation on its recommendations. The feedback from the research community will be used to ensure that the recommendations really do reflect the needs

and concerns of the research community in the UK, and will be used to help to guide and prioritise the resulting work to integrate existing systems, such as those mentioned above, with any new infrastructure that may be needed to make the chosen solution work.

What these issues, and the proposed solutions to them, have in common is both ubiquity (every institution is facing these challenges in one way or another) and scale (they are all too big for one institution to solve on its own). This is where JISC is ideally placed to contribute to solving these problems, both in terms of building on its existing investment in the area of research information management and in terms of its position in the community, having spent years developing connections with the key stakeholders in UK research and facilitating collaborations between them.

The task now is to turn to the future, and to assess the likely shape of the network of current research information systems and related services and tools that is beginning to emerge in the UK.

## **4 Looking to the future**

The network is growing rapidly. Many institutions have, or have plans to implement a CRIS. Services that will interact with these are being developed, and key infrastructure components, such as an identifier solution are being built. The current position in the UK has been described as being at a tipping point, and the outputs of the current RIM projects, from RMAS, CiA and IRIOS 2 are specifically designed to make it easier for universities to adopt the CERIF standard and to make their systems more interoperable and more efficient.

In order to sustain the momentum required to get past this notional ‘tipping point’, JISC will fund a group of small, rapid innovation projects (i.e. short projects aimed at implementing a particular technology using a relatively lightweight, agile project management approach) to spread the use of the tools and technologies developed in the RIM programme.

JISC has also created a CERIF Support Project, based at UKOLN, which will offer hands-on advice and assistance for universities implementing a CRIS via site visits, online guidance and advocacy materials. This project will increase the level of CERIF CRIS expertise in the UK higher education sector by offering training and support, but will also raise awareness of work that has been done in this area amongst senior managers in universities with targeted messages about the benefits to individual institutions of improved research information management processes and systems.

Meanwhile, within JISC, the RIM programme has come to be aligned with the ongoing, long-term investment in institutional repositories and newer investments in managing research data (a group of projects which focuses on research datasets as a product of research activity across the disciplines, rather than the administrative ‘information about research’ approach which makes up the RIM programme). These programmes and projects are converging under the general heading of ‘Research Management’<sup>35</sup>, and this represents a more defined focus for JISC’s investment and change programmes in this area: supporting institutions in improving the quality, impact and productivity of their academic research by addressing the research lifecycle as a whole and building on established and evolving information networks.

---

<sup>35</sup> [http://www.jisc.ac.uk/whatwedo/programmes/di\\_researchmanagement.aspx](http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement.aspx)

## References

- Adams, Sam / Jones, Richard / MacGillivray, Mark / O'Steen, Ben / Ramsey, Malcolm*, 2011a: Researcher Identifiers: Technical interoperability report. Technical Report. Available at: <http://ie-repository.jisc.ac.uk/571/> [Accessed 28/01/12].
- Adams, Sam / Jones, Richard / MacGillivray, Mark / O'Steen, Ben / Ramsey, Malcolm*, 2011b: Researcher Identifiers: Data sources report. Technical Report. Available at: <http://ie-repository.jisc.ac.uk/572/> [Accessed 28/01/12].
- Bolton, Stuart*, 2010: Business case for the adoption of a UK standard for research information interchange. Report to JISC. Available at: <http://www.jisc.ac.uk/media/documents/publications/reports/2010/Businesscasefinalreport.pdf> [Accessed 13/2/12]
- Brown, Josh*: 2011: Identity Management and Identifiers. Workshop Report. Available at: <http://ie-repository.jisc.ac.uk/566/> [Accessed 28/01/12].
- Clements, Anna / Lockhart, Niall*, 2010: CRISPool Project: using CERIF-XML to integrate heterogeneous research information from several institutions into a single portal. Project Report. Available at <http://www.crispool.org/> [Accessed 12/02/12]
- Department for Business, Innovation and Skills*, 2010: The allocation of science and research funding 2011/12 to 2014/15. Available at: <http://www.bis.gov.uk/assets/biscore/science/docs/a/10-1356-allocation-of-science-and-research-funding-2011-2015.pdf> [Accessed 10/02/12]
- Department for Business, Innovation and Skills*, 2011a: International comparative performance of the UK research base – 2011. Available at: <http://www.bis.gov.uk/assets/biscore/science/docs/i/11-p123es-international-comparative-performance-uk-research-base-2011-summary.pdf> [Accessed 10/02/12]
- Department for Business, Innovation and Skills*, 2011b: Innovation and research strategy for growth. Available at: <http://www.bis.gov.uk/assets/biscore/innovation/docs/i/11-1387-innovation-and-research-strategy-for-growth.pdf> [Accessed 10/02/12]
- Ferguson, Nicky*, 2011a: Stakeholder use cases and identifier needs: Report One – use cases, needs, requirements, roles. Technical Report. Available at: <http://ie-repository.jisc.ac.uk/568/> [Accessed 28/01/12].
- Ferguson, Nicky*, 2011b: Stakeholder use cases and identifier needs Report Two – What technical systems would need to interoperate with any identifier infrastructure? When does an individual become a ‘researcher’? Technical Report. Available at: <http://ie-repository.jisc.ac.uk/570/> [Accessed 28/01/12].
- Ferguson, Nicky*, 2011c: Stakeholder use cases and identifier needs: Executive summary – reports 1 and 2. Project Report. Available at: <http://ie-repository.jisc.ac.uk/569/> [Accessed 28/01/12].
- Green, John*, 2010: Research information management: Developing tools to inform the management of research and translating existing good practice. Project report. Available at: <http://www.research-datatools.com/downloads/2010-research-information-management-2.pdf> [Accessed 13/2/12]
- Hill, Amanda*, 2011: Names Project Phase Two Final Report. Project Report. Available at: <http://ie-repository.jisc.ac.uk/573/> [Accessed 28/01/12].
- Hill, Amanda*, 2011: Report on National Approaches to Researcher Identification Systems. Technical Report. Available at: <http://ie-repository.jisc.ac.uk/567/> [Accessed 28/01/12].
- O'Steen, Ben*, 2011: Observations on progress and issues with ORCID. Blog posts. Available at: <http://benosteen.wordpress.com/category/orcid/> [Accessed 28/01/12]
- Rockefeller Foundation*, 2011: EU/US. Roadmap to Measuring the Results of Investments in Science: The Bellagio Statement. A Report following the “EU/US Science of Science Policy” Rockefeller Foundation, Bellagio Centre Workshop, 27 – 30 June 2011.
- Rogers, Nicky / Huxley, Lesley / Ferguson, Nicky*, 2010: Exchanging Research Information in the UK: final report. Project Report. Available at: [http://ie-repository.jisc.ac.uk/448/1/exri\\_final\\_v2.pdf](http://ie-repository.jisc.ac.uk/448/1/exri_final_v2.pdf) [Accessed 13/2/12]
- Russell, Rosemary*, 2011a: JISC Research Information Management Programme: phase 1 technical

- synthesis. Report to JISC. Available at: <http://www.ukoln.ac.uk/rim/dissemination/2011/rim1-technical-synthesis.pdf> [Accessed 12/02/12]
- Russell, Rosemary*, 2011b: An introduction to CERIF. Briefing paper, UKOLN, Bath, UK. Available at: [http://www.ukoln.ac.uk/rim/documents/Introduction\\_to\\_CERIF\\_1.0.pdf](http://www.ukoln.ac.uk/rim/documents/Introduction_to_CERIF_1.0.pdf) [Accessed 12/02/12]
- Russell, Rosemary*, 2011c: Research Information Management in the UK: a supporting document for the JISC RIM3 call, July 2011. Briefing Paper, UKOLN, Bath, UK. Available at: <http://www.ukoln.ac.uk/rim/dissemination/2011/rim-cerif-uk.pdf> [Accessed 12/2/12]





## CASRAI and Research Impacts

### 1 Introduction to CASRAI

Conducting and administering research today is an increasingly multi-stakeholder and multi-disciplinary endeavor. The stakeholders include the highly qualified personnel doing the actual research (researchers and students) and the organizations that facilitate and support the research (universities, colleges and funding agencies). This ‘ecosystem’ of independent but collaborating stakeholders depends on an increasingly fragmented, duplicative and complex set of information about research inputs, activities, outputs, outcomes and impacts. Recording, maintaining, analyzing and sharing this information is difficult and places a tremendous administrative burden on researchers. This results in valuable research time being spent on repetitive administrative tasks and valuable data being left untapped and under-utilized.

The Consortia Advancing Standards in Research Administration Information (CASRAI) is an international not-for-profit organization with a mandate to solve this problem by providing the forum and the mechanisms for the research community to standardize on a common platform of semantics and formats for the sharing and aggregation of research administration information.

We envision researchers maintaining a single, authoritative source of information on themselves and their activities and being able to easily reuse this single source, as needed, with all stakeholders using any software hosted at any location. Research institutions and funding agencies can then have confidence in the quality and comparability of data that is generated, maintained, and exchanged in research administration processes. Any stakeholder should have a frictionless capacity to share information with any other stakeholder within our research community.

Organizations participate in CASRAI based on the following value proposition:

- **Stewardship.** Research information and data are assets. Standardization improves the integrity and accessibility of those assets.
- **Innovation.** CASRAI activities in developing and promoting standards for research administration information position Canada as a first mover and innovator on a global scale.
- **Communication.** CASRAI plays an important role as interlocutor among funders, institutions and researchers on research administration information and management.
- **Efficiency and effectiveness.** Standardization reduces duplication and streamlines the process of applying for funds through (a) elimination of the need to re-enter data and (b) inter-operability of information systems without the need to standardize software. Researchers will spend less time on administration and focus more on their research. Institutions and funders will gain better access to better quality information.
- **Returns on investment.** Saving just one day per year for each researcher in Canada (a conservative estimate of the benefit of standards) translates into between 8 and 10 million dollars channelled to research rather than administration. A recent survey found that 42% of researcher time is spent on admin rather than the research.
- **Accountability.** Standardization greatly facilitates the transparent collection and comparison of data for accountability, including the reporting of key performance indicators (KPI).
- **Incentive.** Standardization of data serve as an incentive to industry (software vendors, other industry players) through the development and implementation of a common information platform with a predictable life-cycle. Greater incentives to industry translate into greater diversity of suppliers and this into lower costs to the research institutions and funders that depend on such software.

## 2 Applying CASRAI to Research Impacts

### 2.1 The Impact Challenge

Organizations that support important research activity (both funders and hosts) must report on the impact of that activity if they are to maintain support momentum in an environment that is competitive for available resources. Two factors make impact reporting increasingly difficult. First, excellence in research activity means a multiplicity of disciplines, institutions and funders in collaboration. This means research teams have multiple organizations seeking impact reports on the same research activity but each organization asks for different information. Second, no single harmonized impact report can serve the diverse needs of the entire research community.

The goal of the CASRAI Research Impact Standards Committee is to reach agreement on the indicators that measure the impact of research activity and to encode that agreement in an open data specification for software providers to implement in their solutions. As a result of achieving this goal researchers will be able to maintain a single source of data from which they can automatically generate impact reports for all of their stakeholders.

### 2.2 Inputs to the Standard Framework

There is a great deal of commonality of structure and content in the various frameworks that are employed for measuring impact, even while there are superficial differences. The table below compares the five frameworks used as inputs to this combined framework.

To approach the impact standards in an organized and comprehensive manner, the CASRAI framework should be inclusive of the subcomponents outlined below, and structured for clarity so that diverse users can appreciate easily how their research impacts categories map onto the standard.

Payback CIHR	Payback ESRC Future of Work	CFI OMS (re-ordered)	UBC (slightly reframed)	SSHRC Workshop Output
<i>Knowledge Production</i>	<i>Knowledge Production</i>	<i>Research Productivity</i>	<i>Knowledge Production</i>	<i>Knowledge Production</i>
- the contributions to knowledge from a research project or a body of research involving multiple projects. Knowledge production is usually measured through contributions to scientific publications and patents or invited presentations (e.g. conferences) but includes knowledge fed more directly to users through commissioned reports etc.	- research quality and quantity  - journal articles; conference presentations; books; book chapters; research reports	- research quality  - research intensity  - external research linkages and collaboration  - sharing of infrastructure  (Note – does not include patents and other forms of KT in this category, but places them under innovation)	- research quality  - research activity and intensity  - collaboration and outreach  - inter- and multi- disciplinary  - new methods and databases	- quality and intensity in various venues:  - academic metrics (Journals, books, conferences, awards) - Performance - Exhibits - Catalogues - DVD's - Websites - Community hand-books - Popular press - Databases (publicly accessible) - Policy documents

<i>Research Targeting and Capacity Building</i>	<i>Research Targeting and Capacity Building</i>	<i>Research Capacity</i>	<i>Research Capacity</i>	<i>Research Training and Capacity Building</i>
<ul style="list-style-type: none"> <li>- benefits to future research activity. This includes the use of research information to improve targeting of future research; individual and group development of research skills and research capacity;</li> <li>- development of the capability to use existing national or international research.</li> </ul>	<ul style="list-style-type: none"> <li>- better targeting of future research;</li> <li>- development of research skills, personnel and overall research capacity;</li> <li>- staff development and educational benefits</li> </ul>	<ul style="list-style-type: none"> <li>- physical capacity: technical and operational</li> <li>- human capacity: numbers and critical mass of researchers</li> <li>- multi-disciplinarity</li> <li>- funding</li> </ul>	<ul style="list-style-type: none"> <li>- quality of research personnel: professional and technical support</li> <li>- physical facilities: operational and technical capacity and currency</li> <li>- funding adequacy and diversity</li> <li>- efficiency of knowledge exchange: within institution, national, international, inter-sectoral</li> <li>- anticipatory capacity in research cadre</li> </ul>	<ul style="list-style-type: none"> <li>- post-graduate students and professionals</li> <li>- Postdoctoral fellows</li> <li>- preparedness</li> <li>- networks</li> <li>- mentoring</li> <li>- civilizing (take a role in society)</li> <li>- domestic capacity</li> <li>- social transformation</li> </ul>
<i>Informing Policy and Product Development</i>	<i>Informing Policy and Product Development</i>	<i>Highly Qualified Personnel</i>	<i>Education and Training</i>	<i>Quality of Life</i>
<ul style="list-style-type: none"> <li>- clinical and administrative benefits, including the development of informed information bases upon which to make decisions, and the application of research findings in policy development (at all levels of policy). Includes development of clinical practice guidelines. Benefits for product or process development where research findings feed into commercial decisions and developments.</li> </ul>	<ul style="list-style-type: none"> <li>- improved information bases for political and executive decisions</li> <li>- development of pharmaceutical products and therapeutic techniques</li> </ul>	<ul style="list-style-type: none"> <li>- number</li> <li>- quality</li> <li>- knowledge transfer through highly qualified personnel (HQP)</li> </ul>	<ul style="list-style-type: none"> <li>- quality of graduate students</li> <li>- production of graduates and skilled post-doctoral fellows</li> <li>- quality of training environment</li> <li>- relevance of training to employment outcomes</li> <li>- extent of transfer of tacit knowledge through highly qualified personnel</li> </ul>	<p>Social:</p> <ul style="list-style-type: none"> <li>- improved equity</li> <li>- ethical debates</li> <li>- measures of democracy</li> <li>- Improved social cohesion (tolerance)</li> <li>- health</li> <li>- social mobility</li> </ul> <p>Environmental:</p> <ul style="list-style-type: none"> <li>- reduce carbon footprint</li> <li>- green economy</li> <li>- sustainability (species)</li> </ul> <p>Cultural:</p> <ul style="list-style-type: none"> <li>- preservation - heritage</li> <li>- creation of new culture</li> <li>- state of cultural institutions</li> <li>- language preservation</li> <li>- sustainability of existing cultures</li> </ul> <p>All three:</p> <ul style="list-style-type: none"> <li>- improved policy (could be under engagement)</li> </ul>

<i>Health and Health Sector Benefits</i>	<i>Employment Sector Benefits</i>	<i>Strategic Plan/ Action</i>	<i>Strategic and Structural Change</i>	<i>Engagement</i>
<ul style="list-style-type: none"> <li>- improvements in life expectancy and quality of life through advances in prevention, diagnosis or treatment made possible by research.</li> <li>- increased efficiency of service organization.</li> <li>- improved equity in the health sector.</li> </ul>	<ul style="list-style-type: none"> <li>- improved working conditions;</li> <li>- higher participation in workforce;</li> <li>- more effective regulation</li> </ul>	<ul style="list-style-type: none"> <li>- process</li> <li>- impact on others</li> <li>- others impact on strategic directions</li> <li>- alignment of investments (e.g. human resources, research, infrastructure)</li> </ul>	<ul style="list-style-type: none"> <li>- effective strategic research planning process and plan that engages appropriate players and informs resource allocation</li> <li>- influence on external stakeholders of research activities</li> <li>- facility effect*</li> <li>- organizational effect**</li> </ul>	<ul style="list-style-type: none"> <li>- policy, practice, public discourse, public engagement:</li> <li>- knowledge mobilization /exchange/ transfer (2 way process)</li> <li>- Modes of collaboration/interface</li> </ul>
<i>Broader economic benefit</i>	<i>Societal and broader economic benefits</i>	<i>Innovation</i>	<i>Innovation outcomes and impacts</i>	<i>Economic</i>
<ul style="list-style-type: none"> <li>- benefits to the economy that result from health research. These benefits can include economic returns from commercialization and contributions to the economy from improvements in workforce health.</li> </ul>	<ul style="list-style-type: none"> <li>- lower stress among workers;</li> <li>- improved public health; improved mental health through decreased unemployment;</li> <li>- greater productivity;</li> <li>- improved equity</li> </ul>	<ul style="list-style-type: none"> <li>- partnerships &amp; collaborations with end users</li> <li>- technology transfer (patents, spin off companies etc)</li> <li>- knowledge transfer of other types (wide range)</li> <li>- socio-economic benefits</li> </ul>	<ul style="list-style-type: none"> <li>- economic &amp; industrial innovation</li> <li>- social, cultural &amp; environmental innovation</li> <li>- health &amp; health services innovation</li> <li>- policy, regulation and public discourse informed by research</li> <li>- economic growth &amp; competitiveness</li> <li>- health &amp; well-being</li> <li>- sustainability</li> <li>- world class research environment</li> </ul>	<p>Direct:</p> <ul style="list-style-type: none"> <li>- foreign direct investment</li> <li>- increased gross domestic product (GDP)</li> <li>- sales/profits</li> <li>- exports of services and products</li> <li>- static effects</li> <li>- wages, disposable income</li> </ul> <p>Indirect:</p> <ul style="list-style-type: none"> <li>- improved standard of living</li> <li>- skilled workforce</li> <li>- regional regeneration</li> <li>- benefits to businesses and non-governmental organisations (NGO's)</li> <li>- dynamic effects</li> </ul>

\* *Facility effect* – the collective power of integrated suites of state-of-the-art equipment, usually housed in purpose built facilities, and deliberately sited to maximize their accessibility and multi-disciplinary, multi-sectoral effects.

\*\* *Organization effect* – the impact of deliberate and strategic integration of infrastructure acquisitions with the organizational policies and structures for research (e.g. institutes, networks and other means of creating synergies), educational curricula and research training (e.g. undergraduate and graduate programs that reflect research strengths and synergies), and innovation and technology transfer activities.

## 2.3 Standard Catalogue of Indicators

Impact indicators are the fundamental input to all impact measurement. Some indicators are common to most impact reporting (i.e. change in number of publications in refereed journals as a result of the research activity) while some indicators are used only in very subject-specific reporting (i.e. change in the adequacy of operational support of research infrastructure as a result of the research activity).

A review of the five impact measurement methodologies outlined above produced what we could term a “standard catalogue” of impact indicators. Each framework uses slightly different labels and groupings and not all frameworks include all indicators.

A standard catalogue of impact indicators can be seen as falling into three main categories and thirteen sub-categories:

- **Capacity Impacts**  
Human, Leadership, Physical, Integration.
- **Productivity Impacts**  
Direct/quantitative, Direct/qualitative, Derived/quantitative.
- **Societal Impacts**  
Direct economic, Indirect economic, Derived social, Derived health, Derived environmental, Derived cultural.

The tables below list the specific indicators included in each sub-category. Changes to any of the following indicators that can be attributed to a research activity represent a measure of the overall impact of that research activity.

## 2.4 Research Impact on Capacity

### 2.4.1 Human Capacity

number of faculty in a theme	number of faculty hired	number of faculty per grad student
percentage of students holding competitive scholarships	Grade Point Average of entering students	number of masters students enrolled
number of PhD students enrolled	number of post-doc fellows enrolled	number of resident associates enrolled
number of technician students enrolled	number of enrolment applicants for advanced training	number of masters students produced
number of PhD students produced	number of post-doc fellows produced	number of resident associates produced
number of technician students produced	number of new programs created	number of new courses created
number of new seminar series created	number of research support staff	number of awards for best presentations
number of awards for best papers	number of students securing jobs in their study sector	number of research chairs

### 2.4.2 Leadership Capacity

level of research planning preparedness	quality of research management	quality of research leadership
---	--------------------------------	--------------------------------

### 2.4.3 Physical Capacity

percentage of infrastructure that is nationally or internationally competitive	level of operational funding for deployed infrastructure	stage in life-cycle of partially or fully deployed infrastructure
--	--	---

number of technical support personnel for infrastructure	level of infrastructure utilization	operational support of infrastructure
number of library holdings	number of large-scale, specialized facilities	quality of research space (facility condition index)

#### 2.4.4 Integration Capacity

specialization index	level of multi-disciplinarity in research activities	number of shared facilities
efficiency of knowledge exchange (internal, national, international, sectoral)	number of larger research network clusters	number of smaller network linkages (performers, users, suppliers)
number of international collaborations	new research activity triggered	amount of lead time for private sector access to research groups
level of access for private sector to research groups	level of open access to findings	

## 2.5 Research Impact on Productivity

### 2.5.1 Quantity of direct knowledge and innovation production

number of publications in refereed journals	number of books published	number of other scholarly outputs
number of other contributed outputs	number of conference outputs	number of artistic outputs
number of artistic performances	number of patents	number of licenses
number of disclosures	number of registered copyrights	number of trademarks
number of contracts with private sector	number of contracts with public sector	number of cases of direct participation in major external research & development activities

### 2.5.2 Quality of direct or indirect knowledge and innovation production

impact factor of publications in refereed journals	number of citations per publications in refereed journals	number of citations per publication geared for broader audiences
--	---	--

### 2.5.3 Quantity of derived knowledge and innovation production

number of new standards	number of new policies	number of new regulations
number of new codes of practice	number of ethical debates	number of publications geared for broader public audiences
number of publications geared for broader professional audiences	number of scholarly services contributed to broader audiences	level of dissemination of best practices
number of improvements in best practices	level of tacit development of knowledge and expertise in collaborating non-research organizations	level of research & development focus of collaborating non-research organizations
number of improvements in professional practice (e.g. better teaching methods)	number of spin-off companies	amount of public information and interaction
number of formal collaborations community and professional organizations	number of presentations to community and professional organizations	

## 2.6 Research Impact on Society

### 2.6.1 Direct economic benefits to society

level of direct cost savings	level of cost savings via outcome implementation	number of new products
number of new processes	number of new services	level of sales
level of profits	number of exported products	number of exported services



number of jobs created	Annual value of royalties/licenses from university spawned technology	Cumulative equity (held) in first generation university spin-offs
Cumulative value of shares sold / derived from university spin-offs		

## 2.6.2 Indirect economic benefits to society

level of direct foreign investment	level of standard of living	level of GDP
number of skilled workers	level of wages	level of disposable income
level of regional regeneration	number of direct benefits to businesses and NGOs	level of dynamic effects
number of pieces of specialized equipment provided to industry	number of testing services contributed to industry	number of engineering design services contributed to industry
number of data collection services contributed to industry	number of analysis services contributed to industry	number of interpretation services contributed to industry
level of indirect passive marketing support contributed to industry (references, connections, seals-of-approval etc)	number of cases where research outputs triggered unintended value in an un-related sector	

## 2.6.3 Broader social benefits

level of equity in society	quality of working conditions	number of public safety improvements
level of participation by under-represented groups in the workforce	level of effectiveness of regulations	number of informed ethical debates
number of measures for democracy levels	level of social cohesion (tolerance)	level of social mobility

## 2.6.4 Broader health benefits

life expectancy	quality of life	number of treatment improvements using existing methods applied in new ways
level of health-related productivity in workforce	level of equity in health sector	

## 2.6.5 Broader environmental benefits

level of carbon footprint	number of green economy initiatives	level of species sustainability
---------------------------	-------------------------------------	---------------------------------

## 2.6.6 Broader cultural benefits

degree of preservation of a heritage artifact	number of new cultural trends	quality of state of cultural institutions
level of preservation of a language	level of sustainability of existing cultures	

### 3 Considerations

In developing an effective standard framework the committee must consider and discuss the following:

1. Capturing impacts
2. Linking impact to people and activity
3. Attributing impact to investment

#### 3.1 Capturing Impacts

Impact indicators can be simple counts of one type of data or compound comparisons of multiple types of data. But all impact indicators have one thing in common: they are calculations performed upon source data. The CASRAI Research Impact Data Standards will govern the structure, semantics and format of that source data.

Measuring the impact of research is measuring the change caused by that research. To measure this each piece of source data **MUST** include two elements: the change and its cause. To fulfill this requirement in a standardized manner we propose a new concept: **Indicator Events**.

A CASRAI Indicator Event is a single recordable occurrence that changes the count of one or multiple impact indicators and that can be related to one or more research personnel or activity.

Most indicator events have a one-to-one relationship with a corresponding single indicator. For example, publishing a new article in a journal is an indicator event. Recording it once in a CASRAI CV would result in the “number of publications in refereed journals” indicator to increase by one if the CV data of that researcher were included in an aggregate impact report.

Some individual indicator events trigger an incremental change to multiple indicators. For example, winning a research contract on a highly collaborative and multi-disciplinary activity is an indicator event. Recording it once in a CASRAI Research Activity Profile could result in an incremental “up-tick” to the counts of: funding received, Highly-qualified-personnel-in-training, number of international collaborators, level of multi-disciplinarity, smaller network linkages, and level of access of private sector to research groups.

The potential is getting the most indicator data from the fewest administrative steps. The challenge is to standardize the minimum number of indicator events that, if recorded once, will produce all the required impact indicators.

Fortunately, many of the indicator events are already captured in some form as they occur. Researchers already capture some (in their various CVs) and institutions already capture others (in their various information management systems). All that is required for the systems now used is to capture this data to have key data elements added that make the data-capture “CASRAI-conformant”. This will enable easier “roll-up” reporting on data collected by multiple researchers and teams and departments and aggregated for periodic impact reports.

What follows is a summary breakdown of:

- a. those indicator events currently captured by researchers and students,
- b. those indicator events currently captured by institutions, and
- c. those indicator events not currently captured (but that need to be).

### **3.1.1 Indicator events that researchers and students already record**

These events are already captured when updating a CV or a reference manager tool. The current CASRAI Standard CV and the upcoming Canadian Common CV support the capture of these events.

Examples of such indicator events include:

- outputs published or presented
- citations (academic)
- mentoring services provided
- degrees
- distinctions
- funding
- new intellectual property
- employment
- program/course creation
- participation in research activity

### **3.1.2 Indicator events that institutions already record**

These events are already captured in some form at institutions. With the adoption of open data standards for capturing these events (the work of this committee) institutions could capture this data using the systems tool of their choice. This source data could be periodically aggregated and become an automated input to a standard Institutional Research Impact Report.

Examples of such indicator events include:

- admissions applications
- employment applications
- new hires
- infrastructure budget approval
- new library holdings
- infrastructure becoming operational

### **3.1.3 Indicator events not currently recorded**

These are events that may not be currently captured systematically by either the Researchers or the Institutions. But this data is required by impact measurement and so it needs to be captured in a CASRAI-conformant manner. The community should decide who is responsible for capturing it.

Examples of such indicator events include:

- citations (non-academic)
- recognition of a derived output
- post-research identification of a derived societal impact

### 3.2 Linking Impacts

As stated above, source data for impact indicators must include two components: the change to an indicator and, the cause of that change.

There are two ways to link an indicator event to a Research Activity:

1. **Direct link:** In recording the indicator event (i.e. a new publication) the event was a clear outcome of an identifiable research activity.
2. **Indirect link:** In recording the indicator event (i.e. getting a new job in your field) it is not feasible to make a direct link between the event and a single research activity. But a link can be inferred through the connection of this highly qualified person and the activities they participated in.

It is therefore crucial that the standard data structures for capturing indicator events not only include fields that describe the event (Journal Title, Article Title, Job Title, Date, etc.) but also a single field that sets a link to a specific research activity (Activity ID).

With this single field (Activity ID) an indicator event is linked to a research activity. Any reports about the overall impact of that research activity will be able to include all events linked to that activity.

In the cases where a direct link cannot be made from an indicator event to a specific research activity there CAN be a link to the unique ID of the researcher (Researcher ID).

### 3.3 Attributing Impacts

A major problem in measuring the impact of research activity is that of attribution of the impacts to the investments made. This is not a simple problem to solve but open data standards that enable recorded impacts to be linked to people and activities will allow every funder of each activity to draw a link between their investments and the resulting impacts.

## 4 Looking Forward

CASRAI is making strides with its approach in a number of key jurisdictions around the world. It is an endorsed approach within the euroCRIS community and is being adopted as a business semantics layer in CERIF. It is also being applied in various JISC-funded projects in the UK; a recent job posting by JISC included CASRAI as a key partner external to the UK for consultation. CASRAI is also advancing the standard dictionary of terminology within the US.

Next steps for CASRAI include recruiting more countries to become participants in this technology-agnostic business dictionary of research terms. In addition to Canada, the UK and US, it is hoped that other research-intensive countries get involved so that we are advancing this harmonization effort worldwide. Over the next few years we hope to see even less friction between any two systems holding research information, regardless of vendor or storage format choices.

## References

*CASRAI website:* [casrai.org](http://casrai.org)

*CASRAI online dictionary:* [dictionary.casrai.org](http://dictionary.casrai.org)

*CASRAI video:* <http://www.youtube.com/watch?v=Gmx7U9-i3Gg>

*euroCRIS website:* [eurocris.org](http://eurocris.org)

*euroCRIS/CASRAI announcement:*

<http://cordis.europa.eu/wire/index.cfm?fuseaction=article.Detail&rcn=28533>





**Jonathan Corson-Rikert, Dean B. Krafft, Brian J. Lowe**

## **VIVO: A Semantic Network of Researchers and Research Information as Linked Open Data**

### **VIVO Collaboration\***

Cornell University

Indiana University

Ponce School of Medicine

The Scripps Research Institute

University of Florida

Washington University School of Medicine in St. Louis

Weill Cornell Medical College

**\*VIVO Collaboration:** Cornell University: Dean Krafft (Cornell PI), Manolo Bevia, Jim Blake, Nick Cappadona, Brian Caruso, Jon Corson-Rikert, Elly Cramer, Medha Devare, Elizabeth Hines, Huda Khan, Deepak Konidena, Brian Lowe, Joseph McEnerney, Holly Mistlebauer, Stella Mitchell, Anup Sawant, Christopher Westling, Tim Worrall, Rebecca Younes. University of Florida: Mike Conlon (VIVO and UF PI), Beth Auten, Michael Barbieri, Chris Barnes, Kaitlin Blackburn, Cecilia Botero, Kerry Britt, Erin Brooks, Amy Buhler, Ellie Bushhousen, Linda Butson, Chris Case, Christine Cogar, Valrie Davis, Mary Edwards, Nita Ferree, Rolando Garcia-Milian, George Hack, Chris Haines, Sara Henning, Rae Jesano, Margeaux Johnson, Meghan Latorre, Yang Li, Jennifer Lyon, James Pence, Hannah Norton, Narayan Raum, Nicholas Rejack, Alexander Rockwell, Sara Russell Gonzalez, Nancy Schaefer, Dale Scheppler, Nicholas Skaggs, Matthew Tedder, Michele R. Tennant, Alicia Turner, Stephen Williams. Indiana University: Katy Borner (IU PI), Kavitha Chandrasekar, Bin Chen, Shanshan Chen, Ryan Cobine, Jeni Coffey, Suresh Deivasigamani, Ying Ding, Russell Duhon, Jon Dunn, Poornima Gopinath, Julie Hardesty, Brian Keese, Namrata Lele, Micah Linnemeier, Nianli Ma, Robert H. McDonald, Asik Pradhan Gongaju, Mark Price, Michael Stamper, Yuyin Sun, Chintan Tank, Alan Walsh, Brian Wheeler, Feng Wu, Angela Zoss. Ponce School of Medicine: Richard J. Noel, Jr. (Ponce PI), Ricardo Espada Colon, Damaris Torres Cruz, Michael Vega Negrón. The Scripps Research Institute: Gerald Joyce (Scripps PI), Catherine Dunn, Sam Katkov, Brant Kelley, Paula King, Angela Murrell, Barbara Noble, Cary Thomas, Michaelleen Trimarchi. Washington University School of Medicine in St. Louis: Rakesh Nagarajan (WUSTL PI), Kristi L. Holmes, Caerie Houchins, George Joseph, Sunita B. Koul, Leslie D. McIntosh. Weill Cornell Medical College: Curtis Cole (Weill PI), Paul Albert, Victor Brodsky, Mark Bronnimann, Adam Cheriff, Oscar Cruz, Dan Dickinson, Richard Hu, Chris Huang, Itay Klaz, Kenneth Lee, Peter Michelini, Grace Migliorisi, John Ruffing, Jason Specland, Tru Tran, Vinay Varughese, Virgil Wong.

This project is funded by the National Institutes of Health, U24 RR029822, “VIVO: Enabling National Networking of Scientists”.

## **1 Introduction**

The Cornell University Library embarked on the VIVO project in 2003 in response to an ambitious multi-disciplinary, faculty-driven initiative in genomics and life sciences. In advocating for increased investment in new faculty hiring, a number of faculty members expressed concern that Cornell was not being successful in communicating the full breadth and depth of research at the institution, especially given the increase in interdisciplinary activities not confined to single departments or colleges. By integrating information from multiple source systems across the University,

VIVO<sup>1</sup> provided a unique focus on research and academics that soon expanded beyond the life sciences to reflect all disciplines at Cornell. In 2009, the VIVO project expanded dramatically through a National Institutes of Health (NIH) grant award to transform VIVO into a platform capable of networking researcher profile information across multiple institutions as a resource at the national level. VIVO is now evolving into an open source community<sup>2</sup> offering multiple tools with broad international participation by institutions including governmental units and private enterprises as well as universities.

VIVO has been designed to communicate information about public aspects of research, as a means of increasing awareness, promoting collaboration, and contributing to the more effective conduct of research within one university or among many. The dissemination of public information about research is much more appropriate for a library, and relates naturally to the library's more general mandate for information discovery, access, and preservation. Advances in large-scale, collaborative management of library information resources mirror today's Web environment, where users routinely search and navigate information on a global scale.

VIVO fulfills only a subset of the functions often included in the scope of a Current Research Information System, or CRIS, in that it does not address the administrative, compliance, or financial aspects of research. There are a number of research administrative platforms represented in the U.S. as well as many home-grown systems; Cornell University has begun a process of conversion from a local-developed "legacy" system to the Kuali Coeus<sup>3</sup> platform supported through the Kuali Foundation<sup>4</sup>, a consortium of universities pooling resources to address common enterprise-level information technology systems needs. Research information systems with financial, operational, regulatory compliance, and administrative reporting functions must of necessity deal with private and sometimes even quite sensitive information.

We would also note that the United States government does not yet conduct anything approaching the United Kingdom's Research Assessment Exercise<sup>5</sup>, or have any history of national-level reporting and review of research outputs across multiple federal funding sources. In the absence of a coordinated national assessment strategy, bibliometric services offered by publishers as an overlay to bibliographic databases, including citation metrics, have assumed a prominent role in university assessments of research. One long-term goal of the VIVO community and related projects, including the NIH-funded eagle-i project<sup>6</sup> led by Harvard University, is to expand the scope of information about research to encompass additional activities, track additional forms of research output including research datasets, and recognize new forms of individual and group contributions to knowledge.

Rather than recount the history of VIVO, this article will focus on the transition of VIVO from a single institution to a multi-institutional community of adopters and users; while far from mature, this community and the challenges and opportunities it faces offer insights into how CRIS systems may change in the coming years.

We begin with a discussion of semantic technologies, since VIVO is a leading practical implementation of the much larger transformation in information infrastructure known as the Semantic Web. A second section focuses on data and branding of VIVO at Cornell and opportunities that have motivated VIVO to look beyond a single institution. In the third and final section, we discuss

---

1 <http://vivo.cornell.edu>

2 <http://vivo.sourceforge.net>

3 <http://kuali.org/kc>

4 <http://kuali.org/>

5 <http://www.rae.ac.uk/>

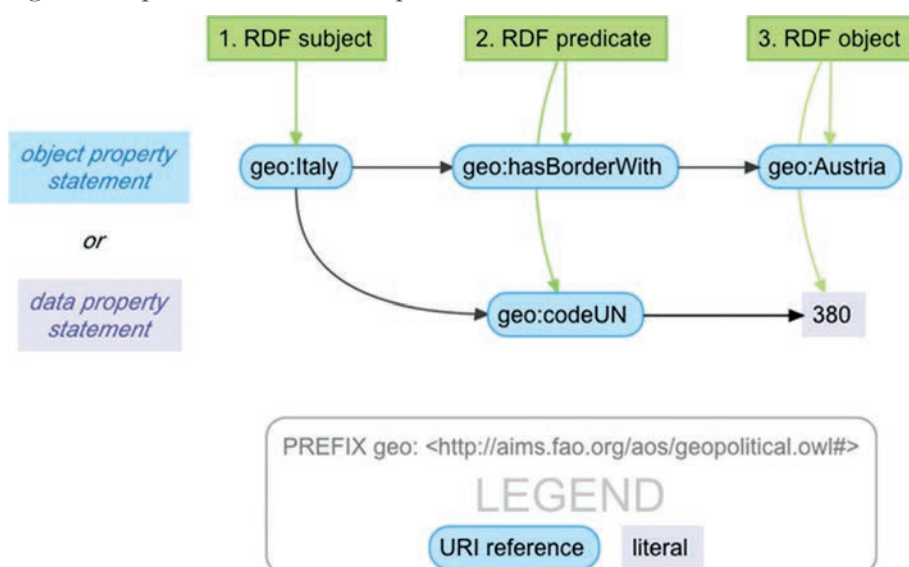
6 <https://www.eagle-i.net/>

how we currently envision the VIVO community evolving over the next few years.

## 2 VIVO as a Semantic Application

In calling VIVO a natively semantic application, we refer to VIVO's structure and function rather than its ability to reference one or more semantic vocabularies or classification schemes. VIVO stores all data using the Resource Description Framework (RDF), an international standard for interchange of data on the Web<sup>7</sup>. The primary unit of information in RDF is a statement, or triple, containing a subject, predicate (verb), and object, loosely corresponding to the structure of a simple sentence in human language (Figure 1). The subjects of semantic triples reference entities of defined types, known as classes; classes in turn are described and inter-related through defined relationships known as properties. The classes and properties referenced in semantic data are defined in an ontology, or specification for a data model in a domain<sup>8</sup>, written in the Web Ontology Language, OWL<sup>9</sup>, or in the RDF vocabulary description language, RDF Schema or RDFS<sup>10</sup>. The VIVO application draws on many existing ontologies and organizes them with new classes and properties of its own to create the VIVO Ontology<sup>11</sup>.

Figure 1: Simple RDF statements or triples



VIVO has elected to use semantic technologies to maintain flexibility in the information about research that it represents and stores; to leverage a growing set of standards and tools; to be able to adopt best practices from other projects; and to make VIVO data fully interoperable on the Web. By storing all data in semantic format, VIVO not only provides human-readable HTML page views but also gives direct access to its underlying data to any application issuing requests for the same information in RDF format. A growing body of public data is being made available as RDF through standard Web request protocols at distributed locations around the Internet, referred to collectively as Linked Open Data<sup>12</sup>.

<sup>7</sup> <http://www.w3.org/RDF/>

<sup>8</sup> see <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>

<sup>9</sup> <http://www.w3.org/TR/owl-features/>

<sup>10</sup> <http://www.w3.org/TR/rdf-schema/>

<sup>11</sup> <http://sourceforge.net/apps/mediawiki/vivo/index.php?title=Ontology>

<sup>12</sup> <http://linkeddata.org/>

Linked Open Data integrates two distinct concepts that together provide a new and direct mode of publishing data on the Web. All the data and ontologies that collectively form the Semantic Web application have URIs, or Universal Resource Identifiers, which can be directly accessed on the Web via standard Hypertext Transfer Protocol<sup>13</sup> requests using the namespace of the home institution of the data or the publishers of the ontology. Semantic data are self-documenting, at least to the extent that definitions are embedded in ontologies; the reuse of common ontologies helps to make data directly interoperable as well, although the Semantic Web also provides a means for mapping data from one ontology to another by asserting that one URI refers to the same data element, class, or property as another.

The second essential attribute of Linked Open Data is openness – by making data freely accessible without restriction, governments and cultural organizations and corporations take a calculated risk that if they publish interesting data, entrepreneurial developers will build the requisite applications to entice users, thereby increasing the return on the initial data investment and helping to justify maintaining that data over time. The principles are familiar from linking HTML documents on today's Web, but the richer structure of Semantic Web data enables more sophisticated applications for both automated and human use.

The VIVO application includes an ontology editor, a content editor, and a simple content management system for displaying data in HTML on the Web. This tri-level functionality makes VIVO very adaptable when local requirements require extensions to the standard VIVO data model, or ontology. The application in most cases reflects data model changes in the editing interface without additional programming, allowing for iterative cycles of testing and populating new data in the full context of the running application. The engine for VIVO without the full VIVO ontology, called Vitro<sup>14</sup>, is maintained and distributed<sup>15</sup> explicitly for rapid prototyping of semantic applications.

VIVO is built with and runs on open source tools, including Java<sup>16</sup>, the Jena Semantic Web framework<sup>17</sup>, the Tomcat servlet engine<sup>18</sup>, the FreeMarker java templating engine<sup>19</sup>, the MySQL database<sup>20</sup>, the Pellet reasoner<sup>21</sup>, and the Apache web server<sup>22</sup>.

## 2.1 VIVO as a Semantic CRIS

The first version of VIVO used a conventional relational database but emulated the flexibility of a fully semantic, ontology-driven model. While we knew we needed to model people, organizations, events, publications, grants, places, and many other types of entities, we were also certain that additional requirements would arise, and furthermore that we could not anticipate in advance the full scope of relationships among these various types of entities. Experience over time has confirmed that flexibility and extensibility are key elements for a successful CRIS, since academic and research domains do not stay static and new circumstances may motivate or even require supporting very different kinds of data. A new project called CTSaconnect directly addresses the challenges of

---

13 <http://www.w3.org/Protocols/rfc2616/rfc2616.html>

14 <http://vitro.mannlib.cornell.edu>

15 <http://sourceforge.net/p/vivo/vitro/>

16 <http://www.oracle.com/technetwork/java/index.html>

17 <http://incubator.apache.org/jena/>

18 <http://tomcat.apache.org/>

19 <http://freemarker.sourceforge.net/>

20 <http://www.mysql.com/>

21 <http://clarkparsia.com/pellet>

22 <http://httpd.apache.org/>

new types of data by seeking to align the VIVO ontology even more closely with the eagle-i ontology, and representing as many components of technical, scholarly, and medical expertise and research resources as possible through a set of carefully aligned ontologies<sup>23</sup>.

From the beginning we have recognized the importance of temporal and other contextual information about relationships between people and organizations (positions), people and publications (authorships), and the roles that people and organizations play in a variety of activities. We need to be able to record more than the simple fact that a specified relationship exists; we may need to know when it started and/or ended, the relative strength or ordering of the relationship, or additional descriptive information such as the title associated with a person's position in an organization. Following terminology from the Mesur project<sup>24</sup>, we call these relationships *context nodes* and model them as any other explicit type in our ontology, with either shared or unique properties. Modeling relationships explicitly (a form of reification<sup>25</sup>) not only makes it possible to classify relationships but also encourages consistent data modeling across institutions.

The VIVO project ontology team holds biweekly open conference calls to discuss proposed additions to the VIVO ontology<sup>26</sup>; the agenda frequently involves new requirements at one university that could well arise at another, and the team works together to compare experiences across institutions and make a recommendation either for an extension to the core ontology shared by every VIVO instance or a local extension reflecting either special terminology or a more specific modeling need. Whenever possible, local extensions are made as subclasses or sub-properties of core VIVO ontology classes and properties, to allow data from any institution to “roll up” into a common, more general representation at the national or international level.

The VIVO ontology team is careful to avoid assumptions that may prove too restricted to a single worldview, whether that of a U.S. research university or of academia in general. The team implementing VIVO for five agencies within the U.S. Department of Agriculture has faced a number of challenges in representing organization types and hierarchies that do not match patterns from academia. We expect to need to modularize parts of the VIVO ontology as needs diversify, and would look in every case to existing ontologies as good templates or direct imports. VIVO has recently submitted its ontology to the National Center for Biomedical Ontology's BioPortal<sup>27</sup> in an effort to make our own ontology more visible and to solicit feedback from other semantic web projects.

It becomes increasingly important to maintain and improve the VIVO ontology as that ontology takes on a life of its own as a framework for common data representation and exchange independent of the VIVO software. The Harvard Profiles software has adopted the VIVO ontology as its native data model<sup>28</sup>, and the University of Iowa's Loki system has been extended to map its internal data model to VIVO RDF.<sup>29</sup> The growth of VIVO adoption at U.S. universities has been accelerated by a vote by the principal investigators of the sixty NIH Clinical and Translational Science Awards in October, 2011 to “encourage their institution(s) to implement research networking tool(s) institution-wide that utilize RDF triples and an ontology compatible with the VIVO ontology.”<sup>30</sup>

---

23 <http://ctsconnect.org>

24 [http://mesur.lanl.gov/Documentation\\_files/JCDL07\\_bollen.pdf](http://mesur.lanl.gov/Documentation_files/JCDL07_bollen.pdf)

25 [http://en.wikipedia.org/wiki/Reification\\_\(computer\\_science\)](http://en.wikipedia.org/wiki/Reification_(computer_science))

26 [http://sourceforge.net/apps/mediawiki/vivo/index.php?title=Ontology\\_Calls](http://sourceforge.net/apps/mediawiki/vivo/index.php?title=Ontology_Calls)

27 <http://biportal.bioontology.org/>

28 <http://profiles.catalyst.harvard.edu/>

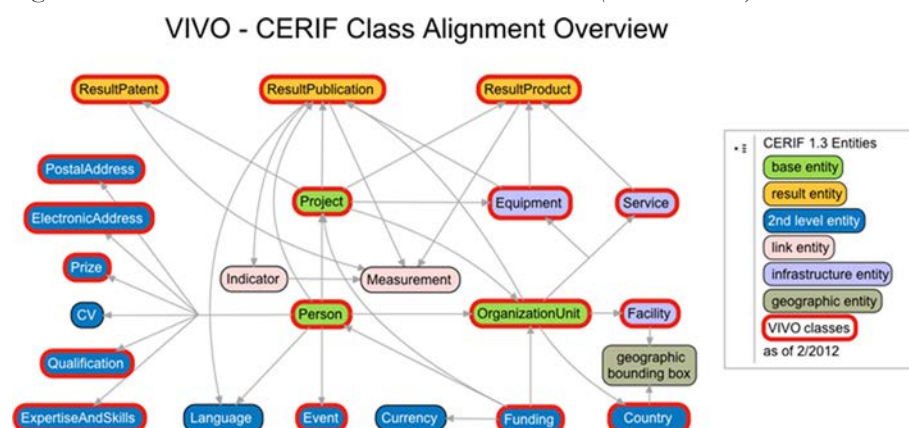
29 <https://www.icts.uiowa.edu/confluence/display/ICTSit/Loki-Vivo+Alignment>

30 <http://www.ctsacentral.org/recommendations-and-best-practices-research-networking>



The VIVO project has recently joined euroCRIS as a strategic partner “with the aim of convergence to allow VIVO semantic web representations and CERIF information system representations to interoperate.”<sup>31</sup> Figure 2 illustrates the initial high-level alignment of VIVO ontology classes with CERIF 1.3<sup>32</sup>. VIVO team members will participate in the euroCRIS Linked Open Data Task Group<sup>33</sup> and coordinate during 2012 with the VOA3R project<sup>34</sup> to create a VIVO-CERIF mapping. Discussions are also underway with CASRAI<sup>35</sup> to establish a partnership focusing on alignment of the VIVO ontology with profiles in the CASRAI data dictionary<sup>36</sup>.

Figure 2: Classes in common between VIVO and CERIF (outlined in red)



### 3 Data and Branding in a University Web Ecosystem

While the scale and administrative structure of American universities varies widely, the challenges and opportunities VIVO has faced at Cornell University have been reflected in the experiences of other institutions adopting VIVO. Cornell’s installation of VIVO draws its data from multiple sources, and each source presents unique challenges – complicated by the fact that all of Cornell’s major systems of record are migrating to new platforms over a four-year period. In some cases VIVO has read-only access to data warehouses, usually when views have been created that scrub private or sensitive data. In other cases we process periodic data exports in CSV format, access RSS feeds, or request data via XML web services. Data transformation is done at the VIVO end through techniques appropriate to each source, including transformations using XSLT<sup>37</sup> and the use of intermediate databases and tools such as D2RQ.<sup>38</sup>

Core data elements from most systems of record carry robust unique identifiers, but whenever alignment depends solely on string comparisons the challenges increase exponentially. Several, but not all, colleges within Cornell license an externally-hosted faculty reporting tool<sup>39</sup> that allows free-text input for many fields, including author and journal names on publications, conference

31 <http://www.eurocris.org/Uploads/Web%20pages/newsflash/Newsflash%2048.pdf>

32 [http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3\\_FDM.pdf](http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_FDM.pdf)

33 <http://www.eurocris.org/Index.php?page=taskgroups&t=1>

34 <http://voa3r.eu/>

35 <http://casrai.org/>

36 <http://dictionary.casrai.org/>

37 <http://www.w3.org/TR/xslt20/>

38 <http://www4.wiwiw.fu-berlin.de/bizer/d2rq/>

39 <http://www.digitalmeasures.com/faculty-activity-reporting-software/>



names, keywords and research areas, and the names of collaborating or sponsoring organizations. Since data are most commonly entered and reported by one person at a time, resulting discrepancies in the free text entries are often not visible until they appear in VIVO. VIVO's interlinked data structure pulls together common references, but we have had to invest considerable effort in data disambiguation to date, and we have much further to go.

In looking ahead to estimate future operational costs for VIVO, it is clear that we need to improve the automatic processing of data feeds, at least to a level requiring only minimal human involvement to curate edge cases. Initial steps include feeding errors back to the source system to reduce the amount of alignment required in VIVO. We are also improving VIVO's use of named RDF graphs, a feature of many triple stores to tag RDF triple statements using a fourth URI identifier. This will allow us to record provenance and to segregate data from different systems of record, and even by date of ingest, so that when we perform retraction and replacement of data in bulk, we can reduce the risk that changes to data from one source will affect data from another. We ultimately hope to implement dynamic linkages to retrieve data on demand for search indexing or display, most likely with a short-term caching option to maintain predictable performance.

VIVO first established an informal presence in Cornell's Web space at its initial soft launch in 2004. Successive improvements to the VIVO website and expansion beyond the life sciences have increased visibility and impact, but VIVO remains something of an anomaly as a single service integrating data from multiple systems of record with annual faculty reporting information and direct end-user editing.

We have sought to convey our role as an information substrate by prominently featuring links to and even images of the primary web pages for units and people, but control over Web branding is pushing us to consider repackaging VIVO pages to look more like search results than professional or organizational profiles.

Our longer-term goal is to increase the reuse of VIVO data on personal, organizational, and university-wide websites through support for filtered feeds, and to phase out VIVO landing pages except when requested, or when no other appropriate page exists elsewhere at Cornell. Data reuse improves the return on the original investment and provides a tangible benefit from efforts to clean data. Purpose-built, "walled garden" websites are still more common than shared databases, in part for administrative and technical expediency, but also due to a lack of awareness of alternatives.

We have developed a number of independent tools<sup>40</sup> including semantic web services<sup>41</sup> to deliver VIVO content as XML and JSON feeds for inclusion in common off-the-shelf content management systems such as CommonSpot<sup>42</sup> or Drupal<sup>43</sup>, and new tools based on the popularity of Linked Open Data promise to make sharing data from VIVO even easier.<sup>44</sup> Figure 3 illustrates a map interface developed to highlight international projects by faculty members in the College of Agriculture and Life Sciences at Cornell.

---

40 <http://sourceforge.net/p/vivo/tools/home/Home/>

41 <http://semanticsservice.svn.sourceforge.net/>

42 <http://www.paperthin.com/>

43 <http://drupal.org/sandbox/miles/1085078>

44 [http://code.google.com/p/linked-data-api/wiki/API\\_Rationale](http://code.google.com/p/linked-data-api/wiki/API_Rationale)

# International

671 projects focusing on 158 countries

Afghanistan ▾ See projects

The world map displays project locations across various countries, each labeled with its three-letter ISO code. The map uses a color-coded system where orange indicates a higher density of projects and yellow indicates a lower density. A legend in the top right corner shows a black box with white text indicating 'Germany' has '25 projects'. On the left side of the map, there are two small black boxes containing a white plus sign (+) and a white minus sign (-), likely for zooming in and out.

Legend: Germany 25 projects

### 3.1 VIVO, Scholarly Communication, and Repositories

Cornell has not yet directly coupled VIVO with its eCommons Digital Repository<sup>47</sup> in a comprehensive way, in large part because the current eCommons platform supports searching by author but does not maintain a stable author URLs. Connecting VIVO's people pages to eCommons and to arXiv.org e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quanti-

47 <http://ecommons.cornell.edu>

tative Finance and Statistics<sup>48</sup>, also hosted by the Cornell University Library, would provide a common semantic underlay connecting library repositories with information on the researchers at Cornell responsible for creating repository content. The University of Rochester Libraries pioneered customizable Web pages for individual faculty members as a feature of their institutional repository<sup>49</sup>, and the VIVO platform offers a similar functionality with the additional benefit of internal connections with research grants, courses, seminars, and other scholarly activities across the university.

VIVO is also being extended through a grant from the Institute for Museum and Library Services<sup>50</sup> to support metadata about research datasets and offer an optional add-on module to provide research data registry functions. This effort continues work at the Cornell University Library on an innovative research support effort for documenting, discovering, and sharing research datasets, called DataStaR<sup>51</sup>. The advent of the National Science Foundation's Data Management Plan Requirement<sup>52</sup> has stimulated awareness of the need for tracking and preserving access to outputs of research beyond text-based publications, and many libraries across the U.S. are developing services to support researchers' needs in documenting and publishing datasets<sup>53</sup>.

The Australian National Data Service (ANDS) has progressed further toward a coherent national data policy in creating a unified discovery service for Australian research data as a partnership with research institutions throughout that country<sup>54</sup>. ANDS promotes improved local research data management and collects metadata on research datasets through a standardized protocol,<sup>55</sup> and a community of institutions implementing VIVO has emerged<sup>56</sup> in concert with addressing ANDS program goals.

Working with research data and the metadata required to support discovery and reuse of that data has provided a number of ontology challenges, most notably the need to represent the often more complex derivation, attribution, documentation, and usage patterns of research datasets and non-traditional forms of publication including blogs and online databases. Brian Lowe from the VIVO team participated in the 2011 *Identity in the Research Infrastructure and Scientific Communication* workshop<sup>57</sup> in Finland as an opportunity to learn more about researchers' roles in diverse forms of research output, especially as new requirements may inform future VIVO ontology and application development.

### 3.2 Looking Beyond One Institution

Weill Cornell Medical College is located in New York City, some 280 kilometers southeast of the main campus in Ithaca. Cornell's VIVO included information about people and departments from Weill Cornell through 2009, when a second VIVO instance was created<sup>58</sup> to allow Weill's IT and

---

48 <http://arxiv.org>

49 <https://urresearch.rochester.edu/viewResearcherBrowse.action>

50 [http://www.imls.gov/news/national\\_leadership\\_grant\\_announcement.aspx#NY](http://www.imls.gov/news/national_leadership_grant_announcement.aspx#NY)

51 <https://sites.google.com/site/datastarsite/>

52 <http://www.nsf.gov/eng/general/dmp.jsp>

53 e.g., <https://confluence.cornell.edu/display/rdmsgweb/Home>,  
<http://www.cdlib.org/services/uc3/datamanagement/>, <http://www.library.illinois.edu/lldata/dmp/dmp.html>

54 <http://services.andis.org.au/home/orca/rda/>

55 <http://ands.org.au/about/approach.html>

56 <http://blogs.unimelb.edu.au/vivoands/>

57 <http://irisc-workshop.org/>

58 <http://vivo.med.cornell.edu>

Library staff to maintain and update content using Weill's internal systems of record. A search in Cornell Ithaca VIVO will return results from either VIVO, and pages are rendered from Ithaca or Weill depending on the source as indicated by the namespace of the URI. Interoperability across distributed repositories also allows Weill Cornell to use their VIVO data in connection with other partnerships that may have almost no overlap with the rest of Cornell, such as the Clinical and Translational Science Center<sup>59</sup>, a NIH-funded partnership with 6 institutions in the New York City area, several of which are also adopting VIVO. As described by Curtis Cole, Chief Information Officer of Weill Cornell Medical College, "VIVO's flexible structure allows us to provide filtered views of the appropriate people and programs in each partnership while providing seamless linking to the full VIVO context at the participating institutions."<sup>60</sup>

Weill Cornell was also a partner in the NIH-funded "VIVO: Enabling National Networking of Scientists" project<sup>61</sup>, a grant awarded in September, 2009 to the University of Florida as prime contractor for implementing VIVO at a total of seven institutions in the United States<sup>62</sup>, with funds from the American Recovery and Reinvestment Act of 2009<sup>63</sup>. This award immediately brought the multi-institutional aspects of VIVO into sharper focus, and it has also transformed the way the project operates and engages a wider community.

Despite VIVO's roots in a single institution, many of the partner institutions had a VIVO installation operating by the end of 2009, and the project made its first open release of the VIVO software in January of 2010, under the BSD software license<sup>64</sup>. Our proposal specified a single software upgrade during the project, but the project team has delivered four major releases and migrated the ontology and data forward at each adopting institution with each release.

In addition to funding implementation, the VIVO project includes major components supporting ontology design and development at the University of Florida and Indiana University as well as at Cornell. These universities have contributed many important new features to the VIVO ontology and code base, including a visualization framework developed in Professor Katy Börner's<sup>65</sup> Information Visualization Lab<sup>66</sup> that supports sparklines, temporal graphs, co-author and co-investigator networks, geospatial maps, and maps of science.

Under the overall guidance of Dr. Michael Conlon<sup>67</sup> and the technical supervision of Christopher Barnes<sup>68</sup>, the University of Florida's Clinical and Translational Research Informatics Program<sup>69</sup> has developed the VIVO Harvester<sup>70</sup>, a general semantic Extract, Transform, and Load (ETL)<sup>71</sup> tool supporting the conversion of multiple forms of input data to RDF, including CSV files, relational databases, and XML. The University of Florida team has also been instrumental in establish-

---

59 [http://en.wikipedia.org/wiki/Morrill\\_Land-Grant\\_Colleges\\_Act](http://en.wikipedia.org/wiki/Morrill_Land-Grant_Colleges_Act)

60 Private correspondence, January 22, 2012

61 <http://vivoweb.org>

62 [http://www.ncrr.nih.gov/publications/ncrr\\_reporter/winter2010/funding\\_matters.asp](http://www.ncrr.nih.gov/publications/ncrr_reporter/winter2010/funding_matters.asp)

63 <http://www.recovery.gov/>

64 <http://www.opensource.org/licenses/BSD-3-Clause>

65 <http://vivo.iu.edu/individual/person25557>, <http://ella.slis.indiana.edu/~katy/>

66 <http://ivl.cns.iu.edu/>

67 <http://vivo.ufl.edu/individual/n25562>

68 <http://vivo.ufl.edu/individual/n64866>

69 <http://www.ctrip.ufl.edu/>

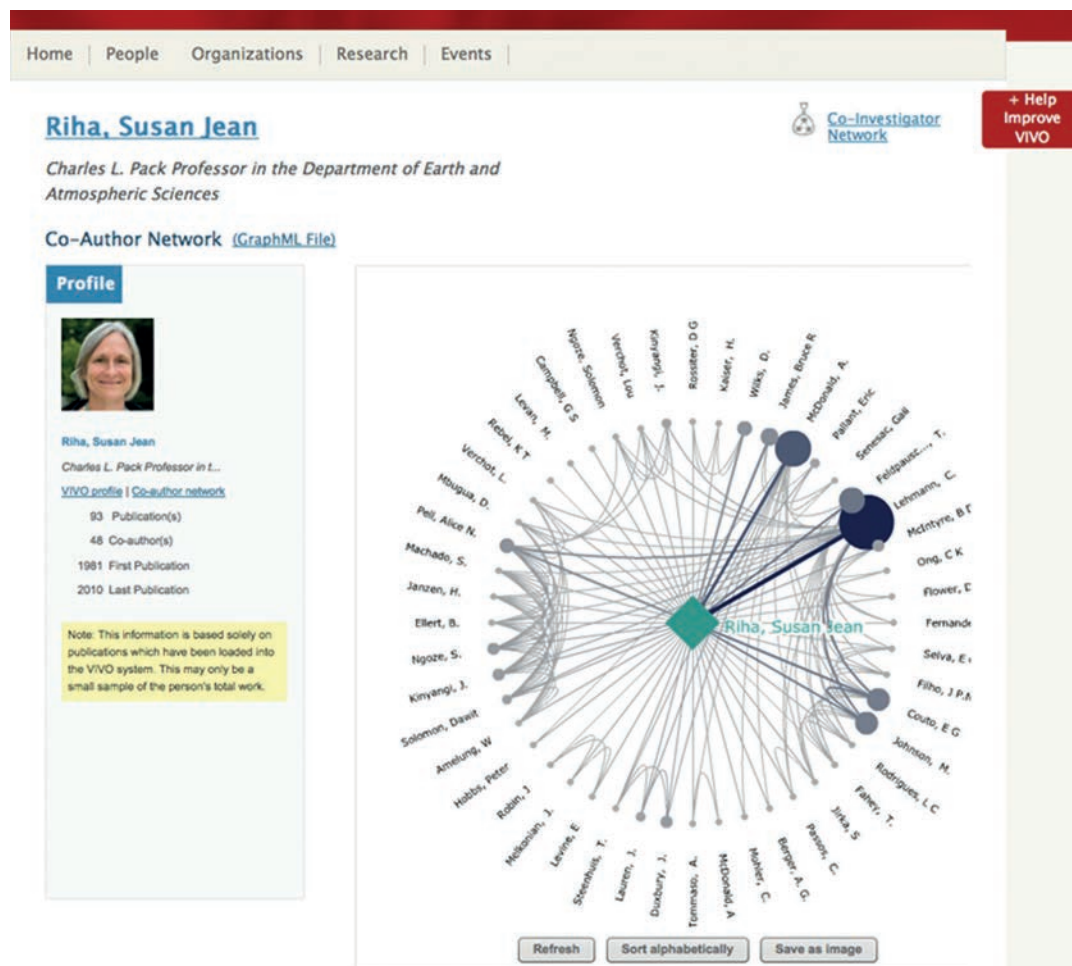
70 <http://sourceforge.net/apps/mediawiki/vivo/index.php?title=Harvester>

71 [http://en.wikipedia.org/wiki/Extract,\\_transform,\\_load](http://en.wikipedia.org/wiki/Extract,_transform,_load)

ing and building out the VIVO source code repository, download center, and documentation wiki on Sourceforge.<sup>72</sup>

VIVO's outreach team, led by Dr. Kristi Holmes<sup>73</sup> of Washington University in St. Louis, has been instrumental in raising the profile of VIVO nationally and internationally through participation in eScience and research networking committees, workshops, and conferences throughout the project.

Figure 4: Visualization of co-author relationships in VIVO



## 4 The Emerging VIVO Community

The first VIVO conference was held in New York City in August, 2010 and served as the effective launch for a larger VIVO community extending well beyond the funded VIVO partners and the boundaries of the United States. The 2011 VIVO Conference featured a number of international sponsors and speakers, and we expect VIVO 2012<sup>74</sup> to continue the pattern.

VIVO's impact in terms of evaluation and adoption is most visible through daily traffic on the

<sup>72</sup> <http://vivo.sourceforge.net/>

<sup>73</sup> <http://vivo.wustl.edu/individual/n4754>

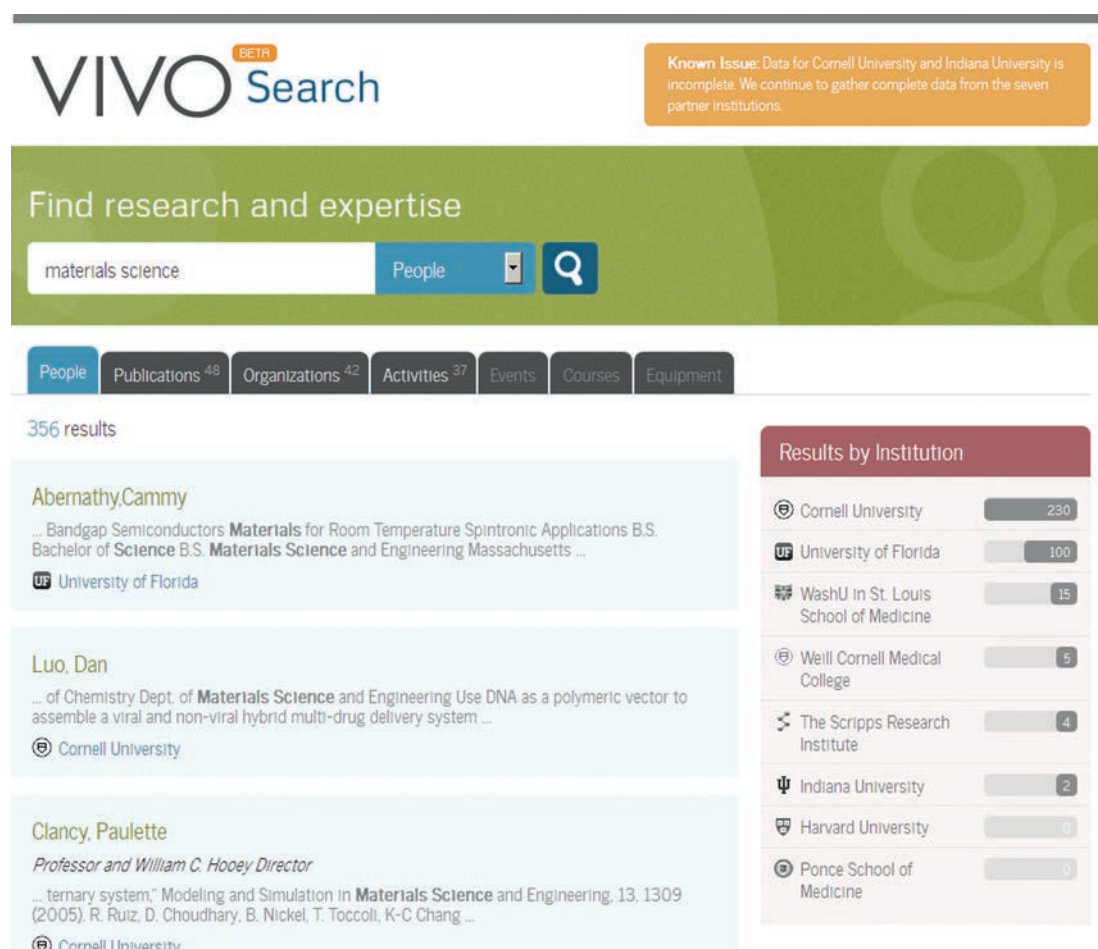
<sup>74</sup> <http://vivoweb.org/conference2012>



VIVO development, ontology, and implementation mailing lists<sup>75</sup>, augmented by weekly open conference calls and biweekly ontology and implementation calls. Implementation calls are organized and led by Alex Viggio from the University of Colorado, an institution outside the VIVO grant that will also be hosting a VIVO Implementation Fest in May, 2012.

VIVO's potential as a common open-source software platform is significant, and the VIVO ontology may gain additional traction as a common data model for research networking. The potential for services leveraging both the platform and ontology across multiple institutions is best illustrated by the VIVO search prototype developed for the 2011 VIVO Conference, demonstrating a unified search interface across the seven VIVO partner sites and Harvard Profiles.<sup>76</sup> Data were harvested from each VIVO instance via Linked Open Data requests and indexed into a single Apache Solr<sup>77</sup> search index exposed through a Web interface in Drupal.<sup>78</sup> Figure 5 shows the first page of results from a search across all eight sites for the term, "materials science."

Figure 5: Search results across institutions in the *vivosearch.org* prototype



<sup>75</sup> <http://vivo.sourceforge.net/communication/>

<sup>76</sup> <http://vivosearch.org>

<sup>77</sup> <http://lucene.apache.org/solr/>

<sup>78</sup> <http://drupal.org>, <https://github.com/milesworthington/vivosolr>



#### 4.1 Scalability, sameAs, and Sustainability

A typical VIVO database at a research university will not stretch the capacity of today's RDF databases, some of which can handle billions of RDF statements, or triples.<sup>79</sup> Apache Solr is also used in production for projects exceeding the aggregate amount of data harvested from any foreseeable collection of VIVO instances. This does not mean that delivering fully responsive interactive editing in a very large VIVO or efficiently updating a centralized index across hundreds of VIVO instances would not present technical challenges, but technical issues pale in comparison to the social, organizational, and funding requirements of achieving any large scale CRIS integration at a national or international scale.

We mentioned earlier the difficulty of disambiguating names within one institution; the problem is far greater at a national or international level. Until recently, each VIVO instance has worked in isolation. With version 1.4, released in December, 2011, VIVO offers a search interface for concept terms that connects to web services for selection of concepts from the National Library of Medicine's Unified Medical Language System<sup>80</sup> or the GEMET Thesaurus.<sup>81</sup> Both vocabularies are available in RDF via web services, the former through Stony Brook University<sup>82</sup> and the latter through Eionet<sup>83</sup>. We plan to extend this to Agrovoc web services<sup>84</sup>, and have begun investigating authority services for people and organizations such as the experimental VIAF International Authority File RDF service.<sup>85</sup>

Matching incoming data from interactive or batch processes to authoritative sources would somewhat reduce the name disambiguation problem, but only published monograph authors could be expected to have name authority records in VIAF or to have been assigned an International Standard Name Identifier, a system designed for "parties involved throughout the media content industries in the creation, production, management, and content distribution chains."<sup>86</sup> In recognition of the global nature of the problem and the need to establish unique identifiers for any person engaged in research, VIVO has collaborated with the Open Researcher and Contributor ID (ORCID)<sup>87</sup> initiative and funded work by ORCID<sup>88</sup> to explore using VIVO as a vehicle for bulk institutional submission of researcher name and employment information to ORCID. ORCID will provide a stable URI for registered researchers that incorporates a unique identifier; VIVO can assert an owl:sameAs relationship<sup>89</sup> between the VIVO URI for the person and the ORCID URI, and potentially provide crosswalking services to a person's institutional or other identifiers.

As the number of databases serving Linked Open Data increases, the number of URIs referencing the same concepts, organizations, people, events, and individuals of any other defined type will clearly also increase; services such as sameAs.org<sup>90</sup> are beginning to address the scope of the problem, but many policy as well as practical questions remain. We sometimes have enough lo-

---

79 <http://challenge.semanticweb.org/>

80 <http://www.nlm.nih.gov/research/umls/>

81 <http://www.eionet.europa.eu/gemet>

82 <http://link.informatics.stonybrook.edu/umls/>

83 <http://www.eionet.europa.eu/gemet/webservices?langcode=en>

84 <http://aims.fao.org/standards/agrovoc/linked-open-data>

85 <http://www.oclc.org/developer/services/viaf>

86 <http://www.isni.org/> (text from Introduction)

87 <http://orcid.org>

88 <http://about.orcid.org/content/orcid-awarded-collaborative-research-grant-vivo>

89 <http://www.w3.org/TR/owl-ref/#sameAs-def>

90 <http://sameAs.org/about.php>

cal information to definitively prove or disprove the assertion that two URIs in a VIVO instance refer to the same person, and it may be advantageous to merge all RDF statements to refer to a single URI and drop the other. If a common identity cannot be asserted with certainty, asserting an owl:sameAs relationship between the two URIs may be preferable since statements pertaining to each original URI can still be distinguished; if the asserted sameAs relationship is later proven false, removing the incorrect assertion is a clean and simple way to undo the error. Geoffrey Bilder of CrossRef<sup>91</sup> has published a detailed white paper on this general topic, entitled “Disambiguation without de-duplication: Modeling authority and trust in the ORCID system.”<sup>92</sup>

## 4.2 Organizational Models for Sustaining Multi-Institutional CRIS Systems

The VIVO project itself has enjoyed strong National Institutes of Health sponsorship as well as significant institutional commitments of time and resources at the local level. Much of the effort in the larger VIVO community depends on volunteer participation by both technical and policy advocates at early adopting institutions. These advocates have been willing to articulate the vision for VIVO and CRIS in general at their own institutions and at national or international events, and to live with the gaps remaining in the relatively new VIVO system.

The VIVO project leadership has embarked on discussions about incubator status as an open source project under the legal and organizational umbrella of the DuraSpace organization<sup>93</sup>. DuraSpace would provide ongoing hosting for VIVO code repositories, documentation wikis, and blogs, while also providing legal status and branding for VIVO independent of any single university – an important factor in encouraging commercial investment in software or services related to VIVO. DuraSpace is committed to open source models and would require transparency in VIVO’s governance, strategic planning, and procedures to set priorities for new development. No financial agreements have been reached, but DuraSpace sustains itself through a combination of sponsorships, grants, and fees for certain services including cloud storage.

The future of CRIS in the United States has been the subject of a series of workshops at the federal level under the general auspices of STAR METRICS, “a multi-agency venture led by the National Institutes of Health, the National Science Foundation (NSF) and the White House Office of Science and Technology Policy (OSTP).”<sup>94</sup> Projects underway include a Science Experts Curriculum Vitae (SciENCv) pilot data platform that “will be evaluated to determine if the shared researcher profile system creates value for the science research community.”<sup>95</sup>

The Lattes Platform<sup>96</sup>, developed in Brazil as a national database of Curriculum Vitae information, is a required system for posting CV information for any researcher seeking federal research funding in Brazil. The Lattes System is predicated on full researcher control over the information in his or her profile; however, information that has been confirmed by an authoritative source such as a publisher or employer is graphically distinguished on the Lattes website.<sup>97</sup>

---

91 <http://crossref.org>

92 <http://about.orcid.org/content/disambiguation-without-de-duplication-modeling-authority-and-trust-orcid-system>, posted March 16, 2011

93 <http://duraspace.org>

94 <https://www.starmetrics.nih.gov/>

95 [http://rbm.nih.gov/profile\\_project.htm](http://rbm.nih.gov/profile_project.htm)

96 <http://lattes.cnpq.br/english/index.htm>

97 Personal communication from José Salm, August, 2011

## 5 Summary

Predicting the future of Current Research Information System adoption and integration into mainstream enterprise information technology practice at universities in the United States or by U.S. federal government funding agencies is well beyond the scope of this article and the knowledge of its authors. This paper has touched on experience at Cornell as one case study, while describing factors we have encountered in growing VIVO from a single-university focus to an active community of users representing small and large institutions in academia, government, and the private sector.

VIVO is well positioned to play a significant role in the future of CRIS adoption in the US based both on its use of Semantic Web technologies and its openness and transparency. By providing the software and ontology under a standard open-source license and pursuing development, implementation and outreach in transparent ways, we hope to gain a critical mass of adopting institutions and committed contributors working to move CRIS adoption in the U.S. toward open platforms, and most importantly, open data.

We do believe we can make progress toward these goals through collaboration with other significant international organizations such as euroCRIS, CASRAI, and Lattes. While technologies differ, we have found receptive partners in early discussions and believe that even aligning data models at the conceptual level will lead to improved data interoperability. No single approach needs to be agreed upon before valuable collaboration can begin; CASRAI prefers to work at a level independent of specific implementations, while euroCRIS has defined and evolved a shared relational data model (CERIF) and has recently established a Linked Open Data Task Group that will actively explore mappings of CERIF to RDF and a bidirectional mapping to VIVO.

## References

- Bilder, Geoffrey*, 2011: Disambiguation without de-duplication: Modeling authority and trust in the ORCID system. White paper. Online: <http://about.orcid.org/documents>.
- Devare Medha / Corson-Rikert, Jon / Caruso, Brian / Lowe, Brian / Chiang, Kathy / McCue, Janet*, 2007: VIVO: Connecting People, Creating a Virtual Life Sciences Community. D-Lib Magazine, Volume 13, No. 7/8.
- Foster, Nancy Fried / Gibbons, Susan*, 2005: Understanding Faculty to Improve Content Recruitment for Institutional Repositories. D-Lib Magazine, Volume 11, No. 1, 2005. Online: <http://hdl.handle.net/1802/1292>.
- Khan, Huda / Caruso, Brian / Corson-Rikert, Jon / Dietrich, Dianne / Lowe, Brian / Steinbart, Gail*, 2010: Using the Semantic Web Approach for Data Curation. Proceedings of the 6th International Conference on Digital Curation. December 6-8, 2010. Chicago, IL. Online: <http://hdl.handle.net/1813/22945>.
- Krafft, Dean B. / Cappadona, Nicholas A. / Caruso, Brian / Corson-Rikert, Jon / Devare, Medha / Lowe, Brian J.*, 2010: VIVO: Enabling National Networking of Scientists, in: Proceedings of the WebSci10: Extending the Frontiers of Society On-Line, April 26-27, 2010, Raleigh, NC, US. Online: <http://journal.webscience.org/316/>
- Lowe, Brian*, 2009: DataStaR: Bridging XML and OWL in Science Metadata Management. Metadata and Semantics Research 46, 141-150. Online: <http://www.springerlink.com/content/q0825vj78ul38712/>
- Lowe, Brian / Caruso, Brian / Corson-Rikert, Jon*, 2007: VIVO Development Roadmap: Enhancing an Ontology-Based Research portal with OWL and Rules, in: *Golbreich, Christine / Kalyanpur, Aditya / Parsia, Bijan (eds)*: Proceedings of the OWLED 2007 Workshop on OWL: Experiences and Directions, Innsbruck, Austria, June 6-7, 2007. Online: <http://ceur-ws.org/Vol-258/paper05.pdf>.
- Mitchell Stella / Tornia, Carlo / Lowe, Brian / Corson-Rikert, Jon / Wilson, Melanie / Ahmed, Mansoor /*

- Chen, Shanshan / Ding, Ying / Rejack, Nicholas / Haendel, Melissa*, 2011: Aligning research resource and researcher representation: the eagle-i and VIVO use case. Poster presentation at: International Conference on Biomedical Ontology 2011; July 26–30, 2011; Buffalo, NY.
- Shotton, David*, 2012: The Five Stars of Online Journal Articles – a Framework for Article Evaluation. D-Lib Magazine, Volume 18, Number 1/2. DOI:10.1045/january2012-shotton.
- Steinhart, Gail*, 2010: Using the Semantic Web Approach for Data Curation. Proceedings of the 6th International Conference on Digital Curation. December 6-8, 2010. Chicago, IL.  
Online: <http://hdl.handle.net/1813/22945>
- Torniai, Carlo / Brush, Matt / Vasilevsky, Nicole / Segerdell, Erik / Haendel, Melissa / Shaffer, Chris / Johnson, Tenille / Corday, Karen*, 2011: Developing an application ontology for biomedical resource annotation and retrieval: Challenges and lessons learned, in: International Conference of Biomedical Ontologies (ICBO) 2011, Buffalo, New York, USA.  
Online: <http://ceur-ws.org/Vol-833/paper14.pdf>.
- Weber, Griffin M. / Barnett, William / Conlon, Mike / Eichmann, David / Kibbe, Warren / Falk-Krzesinski, Holly / Halaas, Michael / Johnson, Layne / Meeks, Eric / Mitchell, Donald / Schleyer, Titus / Stallings, Sarah / Warden, Michael / Kahlon, Maninder / Direct2Experts Collaboration*, 2011: Direct2Experts: a pilot national network to demonstrate interoperability among research-networking platforms. Journal of the American Medical Informatics Association, December 2011, Suppl 1, i157-60.

## **Verzeichnis der Autorinnen und Autoren**

### **David Baker**

Consortia Advancing Standards in Research Administration Information (CASRAI)  
440 Laurier Avenue, West; Ottawa ON K1R 7X6, Kanada  
<http://www.casrai.org/>

### **Prof. Dr. Jörg Becker**

Westfälische Wilhelms-Universität Münster  
Prorektor für strategische Planung und Qualitätssicherung  
Schlossplatz 2; 48149 Münster  
<http://www.uni-muenster.de/>

### **Josh Brown**

Joint Information Systems Committee  
Digital Infrastructure  
Brettenham House, 5 Lancaster Place; London, Großbritannien  
<http://www.jisc.ac.uk/>

### **Jonathan Corson-Rikert**

Cornell University  
Albert R. Mann Library  
Ithaca, NY 14853; USA  
<http://vivo.cornell.edu/individual/individual22972>

### **Dr. Barbara Ebert**

Leuphana Universität Lüneburg  
Forschungsservice  
Scharnhorststraße 1; 21335 Lüneburg  
<http://www.leuphana.de/forschung/services.html>

### **Dr. Michael Erben-Russ**

Fraunhofer-Gesellschaft  
Informationsmanagement  
Hansastraße 27c; 80686 München  
<http://www.fraunhofer.de/>

### **Dr. Jürgen Güdler**

Deutsche Forschungsgemeinschaft  
Gruppe Informationsmanagement  
Kennedyallee 40; 53175 Bonn  
<http://www.dfg.de/>

**Sebastian Herwig**

Westfälische Wilhelms-Universität Münster  
Dezernat für Forschungsangelegenheiten  
Röntgenstraße 19; 48149 Münster  
<http://www.uni-muenster.de/>

**Karsten Holmberg**

Leuphana Universität Lüneburg  
Medien- und Informationszentrum  
Scharnhorststraße 1; 21335 Lüneburg  
<http://www.leuphana.de/services/rmz.html>

**Joachim Holtorf**

Leuphana Universität Lüneburg  
Universitätsbibliothek  
Scharnhorststraße 1; 21335 Lüneburg  
<http://www.leuphana.de/bibliothek.html>

**Brigitte Jörg**

euroCRIS, Board Member, CERIF Task Group Leader  
Cor van Osnabruggeaan 61; 2251 RE Voorschoten, Niederlande  
<http://www.eurocris.org/>

DFKI GmbH  
Language Technology Lab  
Alt Moabit 91c, 10559 Berlin

**Volker Karl**

Karlsruher Institut für Technologie  
Forschungsförderung  
Kaiserstraße 12; 76131 Karlsruhe  
<http://www.for.kit.edu/>

**Dean B. Krafft**

Cornell University  
201 Olin Library  
Ithaca, NY 14853; USA  
<http://vivo.cornell.edu/individual/individual8772>

**Alexander Kujath**

Leuphana Universität Lüneburg  
Forschungsservice  
Scharnhorststraße 1; 21335 Lüneburg  
<http://www.leuphana.de/forschung/services.html>



**Ulrike Küsters**

Fraunhofer-Informationszentrum Raum und Bau  
Competence-Center Fraunhofer-Online: Publikationssupport & Bibliothekssupport  
Nobelstraße 12; 70569 Stuttgart  
<http://www.irb.fraunhofer.de/>

**Dr. Annette Lewerentz**

Freie Universität Berlin  
Abteilung Forschung  
Kaiserswerther Straße 16–18; 14195 Berlin  
<http://www.fu-berlin.de/forschung/>

**Brian J. Lowe**

Cornell University  
Albert R. Mann Library  
Ithaca, NY 14853; USA  
<http://vivo.cornell.edu/individual/individual24052>

**Astrid Lubinski**

Universität Rostock  
Stabstelle Controlling  
18051 Rostock  
<http://www.uni-rostock.de/struktur/verwaltung/stabstellecontrolling/>

**Dr. Dietrich Nelle**

Bundesministerium für Bildung und Forschung  
Unterabteilung Forschungsorganisationen  
Heinemannstraße 2; 53175 Bonn  
<http://www.bmbf.de/>

**Dr. Thomas Rupp**

Niedersächsisches Hochschulkompetenzzentrum für SAP (CCC)  
Welfengarten 1; 30167 Hannover  
<http://www.ccc.uni-hannover.de/>

**Frank Schüller**

Statistisches Bundesamt  
Referat Forschung und Entwicklung, Kultur, Berufsbildung  
Gustav-Stresemann-Ring 11; 65189 Wiesbaden  
<https://www.destatis.de/>

**Regine Tobias**

Karlsruher Institut für Technologie  
KIT-Bibliothek, Publikations- und Mediendienste  
Kaiserstraße 12; 76131 Karlsruhe  
<http://www.bibliothek.kit.edu/>





