

# Abschließender Sachbericht

## **LZI+DBLP: Konsolidierung der bibliometrischen Datenbasis in der Informatik**

Leibniz-Einrichtung: Schloss Dagstuhl – Leibniz-Zentrum für Informatik  
Aktenzeichen: SAW-2011-LZI-3  
Projektlaufzeit: 01.06.2011-31.10.2013  
Ansprechpartner: Prof. Dr. Dr. h.c. Reinhard Wilhelm

# LZI+DBLP: Konsolidierung der bibliometrischen Datenbasis in der Informatik

---

## ABSCHLIEßENDER SACHBERICHT, APRIL 2014

Ausführende Stelle: Schloss Dagstuhl – Leibniz-Zentrum für Informatik  
Ansprechpartner: Prof. Dr. Dr. h.c. Reinhard Wilhelm, wissenschaftlicher Direktor  
Aktenzeichen: SAW-2011-LZI-3  
Förderzeitraum: 01. Juni 2011 – 31. Oktober 2013

## 1 Ausgangsfragen und Zielsetzung

Die Evaluierung wissenschaftlicher Literatur erfordert eine belastbare Datenbasis für wissenschaftliche Publikationen, der eine nachhaltige Infrastruktur zugrunde liegt. Die einzigartige Publikationskultur in der Informatik, die einen starken Schwerpunkt auf Konferenzpublikationen legt, stellt dabei besondere Anforderungen, denen allgemeine Literaturdatenbanken in der Regel nicht gerecht werden. Die Literaturdatenbank dblp leistet seit mehr als 20 Jahren einen wertvollen Dienst für die Informatik, indem personen-, zeitschriften- und konferenzbezogene Publikationsdaten von gesicherter Qualität zur Verfügung gestellt werden. Allerdings ist die thematische Breite der dblp-Datenbank in Teilen noch unzureichend.

Schloss Dagstuhl – Leibniz-Zentrum für Informatik (LZI) und die Universität Trier arbeiten seit November 2010 zusammen mit dem gemeinsamen Ziel, die Literaturdatenbank dblp gemäß ihrer nationalen und internationalen Bedeutung als zentralen Nachweis für wissenschaftliche Veröffentlichungen in der Informatik und der Informatik nahestehenden Gebieten inhaltlich, organisatorisch und technisch zu stärken und auszubauen. Zu diesem Zweck wurde das im Leibniz-Wettbewerb geförderte Projekt „LZI+DBLP: Konsolidierung der bibliometrischen Datenbasis in der Informatik“ im Juni 2011 offiziell gestartet. Ziel dieses Kooperationsprojektes ist die Bildung einer substantiellen, offenen und transparenten Infrastruktur, die langfristig die Bereitstellung einer umfassenden Datenbasis für aussagekräftige Analysen im Bereich der Informatik anvisiert.

## 2 Entwicklung der durchgeführten Arbeiten

Die Datenbank dblp besteht seit 1993 und wird seitdem von ihrem Gründer Dr. Michael Ley gepflegt und erweitert. Zunächst als Ein-Personen-Projekt gestartet und später durch gelegentliche Spendenmittel, wissenschaftliche Mitarbeiter und studentische Hilfskräfte unterstützt, entwickelte sich dblp über die Jahre zu der internationalen Referenzdatenbank bibliographischer Metadaten in der Informatik. Die Personalstärke von dblp blieb dabei jedoch stets deutlich unterhalb der gemäß Aufgabe und Bedeutung angemessenen Notwendigkeit. Um den Nutzen von dblp für die nationale und internationale Informatik-Forschung langfristig zu konsolidieren und zu stärken, wurde Anfang 2010 eine Zusammenarbeit

zwischen dblp und Schloss Dagstuhl -- Leibniz-Zentrum für Informatik initiiert. Der wissenschaftliche Beirat, das industrielle Kuratorium und das wissenschaftliche Direktorium von Schloss Dagstuhl sprachen sich dabei nachdrücklich für eine Kooperation zwischen Schloss Dagstuhl und dblp aus. Ebenso positiv war die Resonanz durch den Fakultätentag Informatik und den Beirat der Universitätsprofessoren der Gesellschaft für Informatik (GIBU), denen die Zusammenarbeit 2010 vorgestellt wurde.

Schloss Dagstuhl hat deshalb im März 2010 beim Senatsausschuss Wettbewerb der Leibniz-Gemeinschaft einen Antrag auf Förderung gestellt, der im Dezember 2010 bewilligt wurde und im Wesentlichen die Finanzierung von zwei wissenschaftlichen Stellen für die Dauer von zwei Jahren vorsah. Dank der Finanzierung war es möglich, ein Team unter der Koordination von Dr. Michael Ley an der Universität Trier zu formieren. Mit zunächst einer durch Dipl.-Inform. Oliver Hoffmann besetzten Stelle begann das Projekt offiziell im Juni 2011 und wurde ab August 2011 durch Dr. Marcel R. Ackermann verstärkt. Die Datenerfassung und Qualitätskontrolle wird zudem seit Juli 2011 halbtags von Frau Stefanie von Keutz unterstützt.

Bereits vorab wurde das Vorhaben durch eine Spende in Höhe von EUR 25.000 von der Klaus Tschira Stiftung gefördert. Die Spende ging im November 2010 ein und diente der Finanzierung der Stelle von Herrn Oliver Hoffmann während der Projektvorbereitungsphase von November 2010 bis Mai 2011. Dadurch konnten Vorarbeiten für das Projekt bereits vor dem offiziellen Projektbeginn initiiert werden.

Dank einer weiteren Spende der Klaus Tschira Stiftung in Höhe von insgesamt EUR 120.000 konnte im Januar 2012 mit Dr. Michael Wagner eine weitere Vollzeitkraft zunächst für zwei Jahre für das Projekt gewonnen werden.

Ab März 2012 wurde das dblp-Team bei der Administration der dblp-Infrastruktur von Herrn Christopher Perrin (studentische Hilfskraft) unterstützt.

## 3 Projektergebnisse

### 3.1 Wissenschaftliche Aufsicht

Ein wesentliches Ziel des Projektes war die Formierung eines wissenschaftlichen Beirates unter dem Dach von Schloss Dagstuhl, welcher dblp in Entscheidungen hinsichtlich der strategischen Entwicklung berät und Rahmenrichtlinien vorgibt. Das dblp Advisory Board hat sich im November 2011 in Saarbrücken konstituiert. Dabei ist es gelungen, für das Board namhafte Persönlichkeiten aus verschiedenen Disziplinen der Informatik zu gewinnen. Dem dblp Advisory Board gehören an:

- Prof. Dr. Andreas Butz (LMU München)
- Prof. Dr. Dietmar Saupe (Universität Konstanz)
- Prof. Dr. Hannah Bast (Universität Freiburg, Sprecherin)
- Prof. Dr. Hans-Peter Lenhof (Universität des Saarlandes)
- Prof. Dr.-Ing. Jürgen Teich (Universität Erlangen-Nürnberg)
- Prof. Dr. Mila Majster-Cederbaum (Universität Mannheim)
- Prof. Oliver Günther, Ph.D. (Universität Potsdam)
- Prof. Dr. Dr. h.c. Otto Spaniol (RWTH Aachen)

- Prof. Dr. Dr. h.c. Reinhard Wilhelm (Universität des Saarlandes)
- Prof. Dr.-Ing. Rüdiger Dillmann (Karlsruher Institut für Technologie)
- Prof. Dr. Rüdiger Reischuk (Universität zu Lübeck)

Zu den Aufgaben des dblp Advisory Boards gehören insbesondere die Festlegung von Richtlinien, Qualitäts- und Prioritätsvorgaben zur inhaltlichen Ausrichtung von dblp, die wissenschaftliche Aufsicht über die inhaltliche Qualität des Datenbestandes, die ideelle Förderung von dblp durch das Einbringen von Fachkompetenz und die Repräsentation von dblp gegenüber der wissenschaftlichen Gemeinschaft.

Nach seiner Konstituierung im November 2011 hat das dblp Advisory Board 2012 seine Arbeit voll aufgenommen. Dabei standen insbesondere zunächst zwei Themen im Mittelpunkt: Die Vorgabe von Rahmenrichtlinien für den Aufbau einer bibliometrischen Infrastruktur sowie die Erarbeitung eines Kataloges an Mindeststandards für die Indizierung von Publikationsreihen in dblp. Zudem hat das dblp Advisory Board das dblp-Team bei der Erweiterung von dblp beraten und durch das Einbringen von Fachwissen unterstützt. So konnte auf Anregung des dblp Advisory Boards die Abdeckung gerade in multidisziplinären Teilgebieten wie Wirtschaftsinformatik, Bioinformatik und weiteren Gebieten verbessert und ausgeweitet werden.

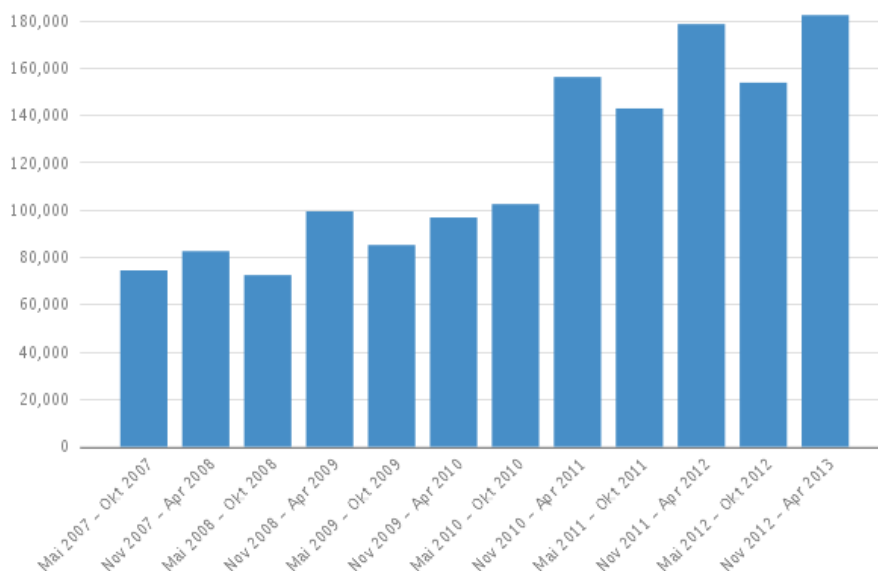
### 3.2 Weiterentwicklung der Daten-Wrapper

Die Literaturdatenbank dblp verzeichnet wissenschaftliche Literatur in der Informatik auf der Ebene einzelner Beiträge. Sie konzentriert sich bei der Wahl der Datenquellen auf Verlage und Bibliotheken und verfolgt das Ziel, ganze Reihen und Serien und damit Themengebiete wissenschaftlicher Literatur möglichst vollständig zu erfassen. Über Jahre hinweg wurden diese Daten stets mit hohem manuellen Anteil erfasst. Dieses Vorgehen führte zu einer anerkannt hohen Qualität des Datenbestandes, stellt aber einen hohen Arbeitsaufwand dar und kann der sich stetig ausweitenden Publikationslandschaft nicht adäquat Rechnung tragen.

Bereits in der Projektvorbereitungsphase wurde daher eine spezialisierte Software, so genannte Wrapper, entwickelt, die bibliographische Rohdaten von Web-Seiten wissenschaftlicher Verlage und Bibliotheken sammelt. Dies entspricht dem Vorgehen einer Suchmaschine wie Google, die durch Crawler Web-Seiten automatisch besucht und zur Indexierung vorbereitet. Diese Software wurde im Rahmen des Projektes stetig erweitert und ausgebaut. Gegen Projektende (Stand Ende Mai 2013) kommen Wrapper bei bereits 112 verschiedenen digitalen Bibliotheken und Verlagssystemen zum Einsatz und stellen heute einen integralen Teil des Arbeitsablaufes bei der Neuaufnahme von Daten dar. Die abschließende Qualitätskontrolle und Fehlerbeseitigung hingegen wird zwar von Hilfssoftware unterstützt; die Endabnahme erfolgt jedoch auch weiterhin bewusst von Hand.

### 3.3 Produktivität und Aktualität der Datenakquise

Bei Beginn der Zusammenarbeit von Schloss Dagstuhl und dblp, Anfang November 2010, indexierte dblp etwa 1,49 Millionen Publikationen. Das Aufnahmevolumen in den Jahren davor betrug dabei zwischen 100.000 und 150.000 Publikationen pro Jahr. Dem gegenüber ist dblp zum offiziellen Projektende im Mai 2013 um etwa 53% gewachsen und indexierte damit über 2.3 Millionen Publikationen. Dabei konnte bereits im ersten Projektjahr ein Aufnahmevolumen von nun über 300.000 neuen Publikationen pro Jahr etabliert und verstetigt werden. Das anvisierte Projektziel von etwa 200.000 Neuaufnahmen pro Jahr



**Abbildung 1: Produktivitätsentwicklung dblp.** Die Grafik gibt eine Übersicht über die Anzahl neu aufgenommener Publikationseinträge in den Sechsmonatsintervallen vor und seit Projektbeginn (November 2010). Der nur scheinbare Produktivitätsrückgang in den Sommermonaten ist durch saisonale Schwankungen im Publikationskalender und Urlaubszeiten begründet.

Anzahl neuer Publikationen in DBLP														
Nov 2009 bis Apr 2010		Mai 2010 bis Okt 2010		Nov 2010 bis Apr 2011		Mai 2011 bis Okt 2011		Nov 2011 bis Apr 2012		Mai 2012 bis Okt 2012		Nov 2012 bis Apr 2013		
Anzahl	Jahr	Anzahl	Jahr	Anzahl	Jahr	Anzahl	Jahr	Anzahl	Jahr	Anzahl	Jahr	Anzahl	Jahr	Alter
0	2011	544	2011	1	2012	704	2012	0	2013	1.097	2013	0	2014	zukünftig
16.753	2010	59.079	2010	26.154	2011	87.869	2011	38.078	2012	93.146	2012	37.014	2013	aktuell
55.830	2009	25.333	2009	68.647	2010	21.557	2010	68.107	2011	12.146	2011	66.766	2012	-1 Jahr
9.564	2008	6.211	2008	18.441	2009	8.733	2009	15.181	2010	6.363	2010	11.188	2011	-2 Jahre
3.269	2007	2.979	2007	6.911	2008	5.064	2008	9.512	2009	5.460	2009	5.872	2010	-3 Jahre
11.483	älter	8.422	älter	36.114	älter	19.064	älter	47.765	älter	35.690	älter	61.576	älter	älter
96.899	total	102.568	total	156.268	total	142.991	total	178.643	total	153.894	total	182.416	total	

**Tabelle 1: Aktualität von Neuaufnahmen dblp.** Die Tabelle gibt eine Übersicht über die neu aufgenommener Publikationseinträge, aufgeschlüsselt nach ihrem jeweiligen offiziellen Publikationsjahr. Es ist zu beachten, dass auf Grund der „online first“-Strategie einiger Verleger in dblp auch Publikationen gelistet sind, deren Publikationsdatum bereits in einem kommenden Jahr liegt.

wurde demnach deutlich übertroffen. Die Produktivitätsentwicklung von dblp ist in Abbildung 1 dargestellt.

Die Literaturdatenbank dblp soll neue Literatur möglichst schnell und vollständig erfassen, um die große Nachfrage insbesondere bei der Recherche von aktuellen Publikationen befriedigen zu können. Ein Maß für die Aktualität einer Veröffentlichung ist das Alter bei ihrer Erfassung, d.h., der Zeitraum zwischen dem Erscheinen der Veröffentlichung und ihrem Eintrag in dblp. Um die Entwicklung der Aktualität zu messen, haben wir das Alter von neu erfassten Veröffentlichungen verglichen. Eine Zusammenfassung der Ergebnisse kann Tabelle 1 entnommen werden. Dabei lassen sich im Wesentlichen zwei Beobachtungen machen. Zum einen ist die Anzahl an Publikationen, die binnen eines Jahres in dblp

Anzahl Publikationen/Reihen in DBLP						
	Journale			Konferenzen/Workshops		
	Reihen	Ausgaben	Artikel	Reihen	Bände	Papiere
Okt 2010	852	14.031	577.743	2.877	16.184	877.320
Apr 2011	1.044	15.874	664.129	2.950	16.939	937.194
Okt 2011	1.091	16.591	722.161	3.094	18.138	1.019.094
Apr 2012	1.179	18.139	806.739	3.204	19.311	1.100.778
Okt 2012	1.213	19.345	879.053	3.279	20.464	1.174.361
Apr 2013	1.275	21.671	989.474	3.363	21.675	1.284.735
Zuwachs	+50%	+54%	+71%	+16%	+34%	+46%

**Tabelle 2: Journale und Konferenz-/Workshopreihen in dblp. Die Tabelle beschreibt die Trendentwicklung bei der Anzahl von Publikationsreihen, Ausgaben und Artikeln in dblp. Es ist zu beachten, dass die Aufteilung in Reihen und Bände insbesondere bei Konferenzen und Workshops nicht immer eindeutig ist. Die vorliegende Zählung betrachtet daher die Aufteilung in Übersichtsseiten (Reihen), Inhaltsverzeichnisse (Ausgaben/Bände) und bibliographische Einträge (Artikel/Papiere) wie sie auf der dblp-Webseite abgebildet ist, und spiegelt damit die allgemeine Tendenz wieder.**

aufgenommen werden, signifikant gestiegen, während gleichzeitig die Anzahl der im zweiten oder dritten Jahr aufgenommenen Publikationen gesunken ist. So wurden zwischen Mai und Oktober 2012 etwa 55% mehr Publikationen binnen eines Jahres indiziert als noch zwei Jahre zuvor. Dies legt eine deutliche Steigerung der Aktualität der in dblp verfügbaren Daten nahe. Des Weiteren ist zu erkennen, dass die Aufnahme von Publikationen nach mehr als 3 Jahren ebenfalls sehr stark angestiegen ist. Dies spiegelt die erfolgreichen Bemühungen des dblp-Teams wieder, gezielt fehlende, ältere Ausgaben von wichtigen Publikationsreihen zu vervollständigen.

### 3.4 Neuaufnahme von Publikationsreihen

Trotz der hohen Reputation von dblp erwies sich in einer Studie von 2010 sowohl die thematische Breite in den Randgebieten der Informatik als auch die Abdeckung einzelner Themengebiete der Kern-Informatik noch als unzureichend. In der Vergangenheit führten vor allem fehlende Ressourcen dazu, dass viele Publikationsreihen nicht berücksichtigt werden konnten. Als Auswahlkriterien zur Aufnahme neuer Reihen wurden lange Zeit vor allem die eigene Erfahrung sowie persönliche Kontakte herangezogen. Diesen Prozess galt es, auf ein solideres Fundament zu stellen.

Da dblp bisher noch keine eigene Infrastruktur zur Bewertung von Publikationsreihen etabliert hat, wurde hierzu verstärkt externe Sachkompetenz herangezogen. Zum einen konnte mit dem neu konstituierten dblp Advisory Board erstmals eine Runde von Experten aus verschiedenen Disziplinen der Informatik direkt ihre Expertise in den Auswahlprozess einbringen. Mitglieder des Boards beraten das dblp-Team bei Entscheidungsfragen und weisen zudem aktiv auf Fehlstände im Datenbestand hin. Das dblp Advisory Board hat dabei einen Katalog an Mindeststandards definiert, die i.d.R. für die Aufnahme in dblp vorausgesetzt werden sollen:

- *Aspekte der Publikationsreihe:* Die Reihe soll vorwiegend Themen der Informatik behandeln; die Reihe soll eine klar definierte thematische Ausrichtung innerhalb der Informatik besitzen; die Reihe soll etabliert sein und regelmäßig erscheinen; die Reihe soll durch einen etablierten Herausgeber oder eine etablierte Fachgesellschaft unterstützt werden.

- *Aspekte der Editoren und Autoren:* Das Editorial Board bzw. die Steuerungs- und Programmkomitees sollen mit innerhalb des thematischen Schwerpunktes herausragenden Wissenschaftlern besetzt sein; unter den Autoren sollen sich herausragende Wissenschaftler des thematischen Schwerpunktgebietes befinden; der Autorenkreis der Reihe soll international sein.
- *Aspekte der Beiträge:* Beiträge der Reihe sollen wissenschaftliche Originalarbeiten sein, die nicht bereits zuvor in anderen Reihen publiziert wurden; Beiträge der Reihe müssen in einem seriösen Peer-Reviewing-Verfahren überprüft werden; Beiträge der Reihe sollen eine adäquate typographische Aufbereitung und Struktur besitzen; Beiträge der Reihe sollen den gängigen Konventionen im Bezug auf Quellenangaben, Literaturverweise und der Darstellung des Standes der Forschung folgen; Beiträge der Reihe sollen sich an eine internationale Leserschaft richten.
- *Aspekte der Daten-Zugänglichkeit:* Metadaten und Zusammenfassungen aller Beiträge sollen frei zugänglich sein; Volltexte aller Beiträge sollen frei oder als bezahlte Dienstleistung zugänglich sein; elektronische Versionen aller Beiträge der Reihe sollen in einer digitalen Bibliothek im Web derart aufbereitet sein, so dass diese auch noch in Jahrzehnten verfügbar sind; alle Beiträge sollen einen eindeutigen und dauerhaften elektronischen Identifikator (z.B. DOI) besitzen; digitale Bibliotheken sollten die Metadaten der Ausgaben und Beiträge vollständig und fehlerfrei anbieten und das elektronische Auslesen und Sammeln von Metadaten erlauben.

Darüber hinaus konnten in vielen Einzelgesprächen mit Teilnehmern von Dagstuhl-Seminaren Einschätzungen und Empfehlungen von anerkannten Fachexperten gewonnen werden. In diesem Austausch offenbarte sich ein weiterer, großer Vorteil der Kooperation von dblp und Schloss Dagstuhl. Zudem wurde als weiteres, regelmäßiges Werkzeug eine Befragung unter allen Seminarteilnehmern mittels eines Fragebogens etabliert. Die Auswertung der Fragebögen liefert wertvolle Hinweise die helfen, Fehlstände im dblp-Datenbestand auszumerken. Der Fragebogen hat zudem als Erprobung der groß angelegten Befragung unter Informatik-Experten im Juni 2013 fungiert.

Neben der Integration von Expertenmeinungen in den Auswahlprozess haben sich Entscheidungen zudem an anerkannten Informatik-Rankings orientiert. Hier sind vor allem das CORE-Ranking der *Computing Research and Education Association of Australasia* sowie das CAPES-Ranking des brasilianischen Forschungsministeriums hervorzuheben. Beide Rankings bieten — wenngleich auch regional geprägt — eine wertvolle Hilfestellung bei der Suche nach wichtigen Publikationsreihen. Um eine angemessene Datengrundlage aus Sicht der deutschen Informatik zu erhalten, wurde zudem in Zusammenarbeit mit der *Gesellschaft für Informatik e.V.* und dem *Fakultätentag Informatik* eine Studie zur Qualität und Relevanz wissenschaftlicher Publikationsreihen aus Sicht der deutschen Informatik-Experten durchgeführt.

Insgesamt konnten seit Projektbeginn über 900 neue Reihen identifiziert und in dblp aufgenommen werden.

### 3.5 Vollständigkeit bestehender Reihen

Neben der Aufnahme neuer Publikationsreihen liegt ein weiterer Schwerpunkt auf der Vervollständigung bereits indexierter Zeitschriften und Konferenzserien. Hier wurden im Verlaufe des Projekts verstärkt Anstrengungen unternommen, insbesondere die wissenschaftlich bedeutendsten Reihen lückenlos in dblp zu erfassen.

<b>Anzahl Monographien in DBLP</b>		
	Bücher und Sammelbänder	Dissertationen
Okt 2010	1675	91
Apr 2011	7.484	808
Okt 2011	8.851	829
Apr 2012	9.263	6.898
Okt 2012	9.502	6.909
Apr 2013	9.841	6.927
Zuwachs	+588%	+7.612%

**Tabelle 3: Monographien in dblp.** Die Tabelle fasst die Anzahl in dblp indexierter Monographien im Sechsmonatsabstand zusammen. Der Sprung von Oktober 2010 zu April 2011 geht auf die Integration des Ist-Bestandes der Informatik-Fachbibliothek von Schloss Dagstuhl zurück. Der Anstieg bei den Dissertationen zum April 2012 wurde durch das gezielte Sammeln von Metadaten über die Webseiten deutscher Hochschulen und der Deutschen Nationalbibliothek erreicht.

Als Auswahlkriterien wurden dabei ebenfalls die etablierten CORE- und CAPES-Rankings sowie die Eingaben des dblp-Beirats und der Besucher der Dagstuhl-Seminare herangezogen. Ziel der Bemühungen war dabei insbesondere, die Reihen der Spitzenkategorien vollständig, sowie hochklassige Reihen zumindest bis in das Jahr 2000 vollständig zu ergänzen. Hier konnten inzwischen große Lücken in der Abdeckung geschlossen werden. Die Vervollständigung aller prestigeträchtigen Reihen ist dabei ein Prozess, der noch immer andauert.

### 3.6 Neuaufnahme von Monographien

Ein weiteres Ziel des Projektes war die Verbreiterung der Datenbasis im Bereich von Monographien und Sammelbänden. Hierbei wurde bereits in der Projektvorbereitungsphase ab November 2010 eine Zusammenarbeit mit der Informatik-Fachbibliothek von Schloss Dagstuhl initiiert. Nachdem im Zeitraum vor offiziellem Projektbeginn zunächst die Integration des Ist-Bestandes der Bibliothek im Vordergrund stand, wurde seitdem ein regelmäßiger Datenaustausch zwischen Bibliothek und dblp installiert, bei dem die monatlichen Bibliothek-Neuzugänge weitgehend in dblp integriert werden. Diese Zusammenarbeit wurde im Laufe des Projektes gepflegt und intensiviert. Insgesamt wurden so bis zum Mai 2013 über 8.100 neue Bände in den Datenbestand übernommen.

Ähnlich zum Fall der Journale und Konferenzen stellt die Aufnahme von Fachbüchern und Sammelbänden die Frage nach den Kriterien für die Auswahl geeigneter Bände. Eine Aufnahme ganzer Bibliotheks- oder Verlagskataloge scheint nicht sinnvoll, da auch hier Bedeutung und wissenschaftliche Qualität sehr inhomogen sind. Zudem erfordert die Bearbeitung bei der Neuaufnahme im Vergleich zu Journalen und Konferenzen einen überproportional hohen Arbeitsaufwand des dblp-Teams bei relativ geringer Anzahl aufzunehmender Datensätze. Nichts desto trotz hat sich das dblp Advisory Board für eine Ausweitung der Aufnahmebemühungen ausgesprochen. Die genaue Ausgestaltung der Arbeitsprozesse und Qualitätskriterien sind jedoch zum formalen Ende des Projektes ein offenes Problem, dem sich dblp noch stellen muss.

Darüber hinaus bildet die Neuaufnahme von Dissertationen einen weiteren Schwerpunkt. Dank der Freigabe der Katalogdaten der *Deutschen Nationalbibliothek* als Open-Data unter der „Creative



Commons Zero 1.0“-Lizenz konnte dblp im Laufe des Projektes um über 6.800 Dissertationen von zumeist deutschen Hochschulen erweitert werden. Es bleibt ein Ziel von dblp, die Aufnahme von Dissertationen zu intensivieren und zudem auf den internationalen Bereich auszudehnen. Jedoch stellt zu Projektende insbesondere die Akquise von internationalen Dissertationen noch immer ein offenes Problem dar.

### 3.7 Kontakte mit Verlagen als Datenlieferanten

Der Kontakt zu Verlagen und Fachgesellschaften ist notwendig, um über Neuerscheinungen informiert zu bleiben, sowie um möglichst zeitnah an bibliographische Metadaten zu gelangen. Im Idealfall erweisen sich Verlage dabei als kontinuierliche Datenlieferanten.

Zu den neu gewonnenen Partnern von dblp gehören zum Zeitpunkt des formalen Projektendes folgende Partner: Bereits im Mai 2011 konnte mit *Springer Science+Business Media* einer der für die Informatik wichtigsten Verleger als Partner und Datenlieferant gewonnen werden. Seitdem beliefert Springer dblp regelmäßig mit den Metadaten zu allen Neuerscheinungen. Die Daten werden tagesaktuell in den dblp-Datenbestand eingepflegt. Seit Oktober 2011 ist es zudem gelungen, das *Institute of Electrical and Electronics Engineers (IEEE)* als regelmäßigen Datenlieferant für dblp zu gewinnen. dblp erhält wöchentlich alle Neuaufnahmen der digitalen Bibliothek *IEEE Xplore* und hat zudem auch vollen Zugriff auf deren Altbestand. Im Dezember 2012 konnte ferner mit *IOS Press* eine Vereinbarung über regelmäßige Datenlieferungen getroffen werden. Die Metadaten der Informatik-Sparte des Verlagsprogrammes werden seit dem über eine FTP-Schnittstelle bereitgestellt. Ferner werden auch die Open-Access-Publikationen von *Dagstuhl Publishing* von dblp indexiert. Über eine eigens bereitgestellte Schnittstelle werden bereits seit 2011 die Metadaten an dblp übermittelt. Darüber hinaus befinden sich zum Zeitpunkt des formalen Projektendes Kooperationen mit der *Association for Computing Machinery (ACM)* und der *USENIX Association* in Verhandlung. Beide Partner haben sich bereits positiv über die Einrichtung regelmäßiger Datenlieferungen geäußert.

Die etablierten Partnerschaften erwiesen sich für dblp als äußerst vorteilhaft und trugen erheblich zur Steigerung von Produktivität und Aktualität der Neuaufnahmen bei. Es ist auch in Zukunft beabsichtigt, Kontakte mit weiteren Großverlagen zu intensivieren. Aber auch für die Autoren der Verlage dürfte sich die verbesserte Indexierung als Vorteil erweisen: Nicht selten werden inzwischen Publikation in dblp indexiert, noch bevor sie in den digitalen Bibliotheken der Verlage gelistet werden.

### 3.8 Aufbau einer bibliometrischen Infrastruktur

Die hohe Datenqualität der Literaturdatenbank dblp ist das Ergebnis eines arbeitsintensiven Datenpflegeprozesses. Für ein ressourcenbeschränktes Projekt wie dblp ist es daher unvermeidlich, sich bei der Auswahl zu indexierender Reihen auf eine repräsentative Teilmenge der Zehntausende an existierenden Publikationsreihen zu beschränken. dblp verfolgt dabei den Ansatz, zentrale und wissenschaftlich hochwertige Reihen bei der Aufnahme zu bevorzugen. Ein wesentliches Ziel des Projektes war daher auch die Bildung einer bibliometrischen Infrastruktur, die mittelfristig eine möglichst objektive Auswahl zu indexierender Publikationsreihen ermöglicht.

Im ersten Projektjahr wurden in enger Absprache mit dem dblp Advisory Board sowie dem Fakultätentag Informatik die Anforderungen an eine solche Infrastruktur spezifiziert:

- Einheit der Evaluation sind Publikationsreihen, nicht Autoren oder Institutionen.
- Grundlage der Evaluation soll die Beurteilung durch die Forschenden in der Informatik sein, und kein starres, auf statistischen Daten basierendes Formelgerüst.
- Die Beurteilung soll auf der Basis von qualifizierten Mindeststandards erfolgen.
- Das Resultat der Evaluation soll kein Ranking oder numerisches Rating sein, sondern eine Klassifikation in eine geringe Anzahl grober Qualitätskategorien (z.B.: Klasse A bis Klasse D).
- Die Beurteilung der Reihen soll relativ zum jeweilig einschlägigen Teilfeld der Informatik erfolgen.
- Die Ergebnisse und Grundlagen der Evaluation sollen frei zugänglich und für jedermann nachvollziehbar sein.
- Die Evaluation soll in periodischen Abständen wiederholt werden.

Das dblp Advisory Board sprach sich zu diesem Zweck für die Durchführung einer großflächigen Umfrage unter Informatik-Forschenden aus. Die erhobenen Daten sollen die Grundlagen für eine Klassifikation nach wissenschaftlichem Stellenwert und thematischer Ausrichtung bilden.

### 3.9 Erhebung zu wissenschaftlichen Publikationsorganen

Ab Mitte 2012 wurde in Zusammenarbeit mit der *Gesellschaft für Informatik e.V. (GI)* und dem *Fakultätentag Informatik* mit der Planung der ersten deutschlandweiten Erhebung zum Stellenwert von nationalen und internationalen wissenschaftlichen Publikationsreihen in der Informatik begonnen. Ziel der geplanten Erhebung war das Erlangen einer qualifizierten Übersicht über die in der deutschen Informatik genutzten Publikationskanäle in Form von internationalen und nationalen wissenschaftlichen Konferenzen und Zeitschriften.

Im Juni 2013 wurde schließlich die Online-Befragung aller Mitglieder der Gesellschaft für Informatik unter der Adresse unter <http://survey.dagstuhl.de> gestartet. GI-Mitglieder wurden dazu mit persönlichen, anonymisierten Zugangsdaten auf dem Postweg eingeladen. Inhaltlich ergründet die Befragung das persönliche Publikationsverhalten der Befragten (Welche Publikationsreihen werden verwendet?) sowie deren Einschätzung der bekannten Reihen bezüglich verschiedener Qualitätsdimensionen (Einzigartigkeit, Originalität der Ergebnisse, wissenschaftliche Methodik, Qualität der Präsentation, Einfluss auf die eigene Forschungsarbeit, Bedeutung als Kommunikationsorgan) und ihrer thematischen Kategorisierung. Die Teilnahme an der Befragung war freiwillig; der Online-Fragebogen war bis Ende September 2013 erreichbar.

Zur Qualitätssicherung der Umfrage und zur Auswertung der Ergebnisse wurde ein Editorial Board, bestehend aus Vertretern der Kooperationspartner, etabliert:

- Prof. Dr. Hannah Bast (Sprecherin dblp Advisory Board)
- Prof. Dr.-Ing. Peter Liggesmeyer (Vizepräsident Gesellschaft für Informatik)
- Prof. Dr. Rüdiger Reischuk (stellvertretender Vorsitzender Fakultätentag Informatik)
- Prof. Dr. Dr. h.c. Reinhard Wilhelm (Wissenschaftlicher Direktor Schloss Dagstuhl)
- Dr. Michael Wagner (dblp-Team, Koordination der Umfrage)

Das Editorial Board wird zudem von Prof. Dr. Uwe Brinkschulte (Gesellschaft für Informatik, FB Technische Informatik), Prof. Dr. Oliver Deussen (Gesellschaft für Informatik, FB Graphische Datenverarbeitung), Prof. Dr. Gregor Engels (Gesellschaft für Informatik, FB Softwaretechnik), Prof. Dr. Erhard Rahm (Gesellschaft für Informatik, FB Datenbanken und Informationssysteme) und Prof. Dr.-Ing. Dr. h.c. Manfred Nagl (Fakultätentag Informatik) und Alexander Rabe (Leiter Hauptstadtbüro GI) beraten und unterstützt.

Von den 18.034 eingeladenen GI-Mitgliedern haben insgesamt 2.356 an der Umfrage teilgenommen. Von diesen Personen haben 2.009 Befragte die Umfrage in vollem Umfang bearbeitet; 347 Fragesätze wurden nur unvollständig beantwortet und mussten daher bei der Bewertung aussortiert werden. Die Rücklaufquote an vollständig erfassten Befragungen liegt damit bei etwa 11%.

Im Rahmen der Befragung wurden insgesamt 1.421 verschiedene Konferenzserien und 844 verschiedene Journale angegeben. Dies entspricht in etwa dem 60%-zu-40%-Verhältnis zwischen Konferenzen und Journalen, welches auch im dblp-Datensatz zu beobachten ist. Eine detaillierte Aufschlüsselung der gegebenen Reihen zeigt, dass 71% der genannten Konferenzen und 63% der genannten Journale bereits in dblp gelistet waren. Eingeschränkt auf alle Reihen, die wenigstens drei Mal von verschiedenen Befragten genannt wurden, zeigt sich sogar eine Abdeckung von 94% bzw. 97%. Dies weist bereits auf eine offenbar hervorragende Abdeckung von dblp innerhalb der Kerninformatik hin. Eine genauere Betrachtung der Reihen, die nur von einzelnen Befragten genannt wurden, zeigt zudem, dass deren überwältigende Mehrheit nicht zum Zuständigkeitsbereich von dblp gehört.

Die angegebenen Reihen wurden zudem von den Befragten an Hand einer Reihe von wissenschaftlichen Qualitätskriterien wie Originalität und Methodik der Arbeiten, Einzigartigkeit der Reihe innerhalb der Disziplin und anderen Aspekten beurteilt. Unglücklicherweise wurden einzelne Reihen im Durchschnitt von nur 2,8 Befragten bewertet, weswegen diese Daten in der Breite nur zu statistisch zweifelhaften Ergebnissen führen dürften. Das Editorial Board ist derzeit noch mit der Diskussion um die Nachverwendung der Ergebnisse sowie der Veröffentlichung der anonymisierten Rohdaten befasst.

### **3.10 Anbindung an das semantische Web als Open-Data**

Open-Source, Open-Access und Open-Data haben sich in den vergangenen Jahren als wegweisende Konzepte der Wissensvermittlung etabliert. Das semantische Web mit seinem Bedarf an Linked Open Data (LOD) stellt dabei eine der großen Zukunftschancen für eine freie Wissensgesellschaft dar. Obwohl dblp bereits seiner Gründung 1993 seine Daten frei im Web zur Verfügung stellt, gab es bisher keine technische Anbindung der dblp Stammdaten an die LOD-Cloud. Die existierenden dblp-Komponenten des semantischen Webs basieren viel mehr auf veralteten, statischen Exporten des dblp-Datensatzes, was zwangsläufig zu Inkonsistenzen gegenüber dem aktuellen, „lebendigen“ dblp-Datensatz zur Folge hat.

Entsprechend der Empfehlungen der *Open Knowledge Foundation* und der *Deutschen Initiative für Netzwerkinformationen* zu offenen bibliographischen Daten wurde der komplette dblp-Datensatz daher seit November 2011 unter der „Open Data Commons Attribution License v1.0“ nun auch formal als Open-Data veröffentlicht. Die ohnehin seit Jahren etablierte Nutzung des dblp-Datensatz steht somit der Allgemeinheit in Zukunft auch rechtssicher zur Verfügung.

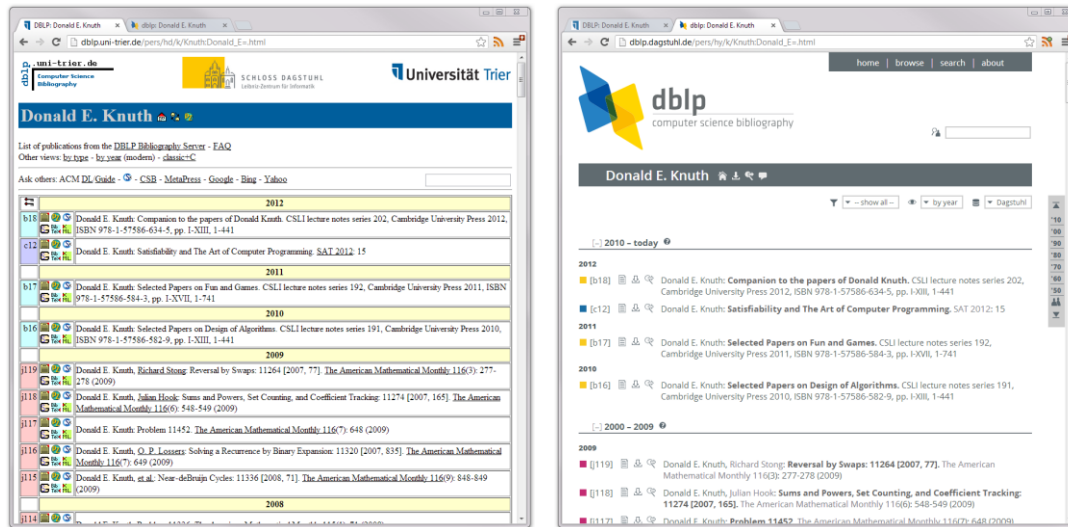


Abbildung 2: Das alte und das neue Layout des dblp-Webservice.

Des Weiteren wurde mit der Etablierung einer RDF/XML-basierten Datensicht die Anbindung an das semantische Web mit der Bereitstellung von Linked Open Data in die Wege geleitet. Mit Hilfe dieser neuen Infrastruktur können die dblp-Daten nun mit den Werkzeugen des semantischen Webs (z.B. SPARQL-Browsern) erschlossen werden. Die RDF-Schnittstelle von dblp befindet sich zu Projektende zwar noch in der internen Erprobungsphase, wird aber im ersten Quartal 2014 in den öffentlichen Betrieb übergehen.

### 3.11 Nutzerorientierte Neugestaltung der Webseiten

Die dblp-Webseiten wurden 1993 als einfache Sammlung statischer HTML-Seiten geschaffen. Seitdem hat sich die Erscheinung des Web-Frontends von dblp in Prinzip nur wenig gewandelt. Das rustikale Tabellenlayout wirkt heute veraltet und fällt im optischen Vergleich und in Sachen Benutzerführung gegenüber modernen Websystemen wie Google Scholar oder Microsoft Academic Search zurück.

Aus diesem Grund wurde das Web-Frontend überarbeitet und im Sinne eines leichtgewichtigen, modernen Designs unter Beibehaltung der inhaltlichen Aussagekraft erneuert. Neben einer optischen Generalüberholung stand dabei insbesondere auch die nutzungsgerechte Aufbereitung der bibliographischen Daten im Vordergrund. So wird einer häufig geäußerten Kritik Rechnung getragen und neben der bisherigen chronologischen Sortierung der Autorensiten auch eine kategorisierte Sortierung nach Publikationstyp angeboten. Die neuen Leistungsmerkmale umfassen die folgenden Punkte:

- Volle HTML5- und CSS2-Konformität.
- Neues dblp-Logo und ein klares und aufgeräumtes Layout. Das klassische Layout ist als Variante weiter verfügbar.
- Integration der bisher extern betriebenen CompleteSearch-Suchmaschine, inklusive erweiterter Such- und Filteroptionen.
- Neue Kategorien-Sicht auf Autorensiten als Standard; die herkömmliche chronologische Sicht ist weiterhin verfügbar.

- Einfache Seitennavigation durch Sektionsmenüs, sowie ein- und aufklappbare Abschnitte.
- Neue Datenexport-Optionen, wie etwa einen RSS-Feed und eine neue XML-Sicht für Autorensseiten.
- Anbindung an das semantische Web als Linked Open Data mittels RDF/XML-API für Autoren- und Publikations-Entitäten.
- Neue Statistik-Seiten geben tagesaktuell Auskunft über die Entwicklung von dblp.
- Die F.A.Q.-Seiten wurden optisch und inhaltlich überarbeitet. Kontextuell relevante Fragen sind nun auch direkt aus jeder Webseitensicht heraus erreichbar.

Des Weiteren ist bei der Restrukturierung auf einen modularen Aufbau geachtet worden, der in Zukunft die Erweiterung um zusätzliche Komponenten und Verlinkungen, sowie die Anpassung an unterschiedlichste Ausgabegeräte (z.B. Drucker oder Mobilgeräte) erleichtern wird. Seit Anfang 2013 hat das neue Layout die alten Seiten abgelöst und befindet sich nun im öffentlichen Testbetrieb

### 3.12 Verstetigung des Engagements von Schloss Dagstuhl für dblp

Die Zusammenarbeit von Schloss Dagstuhl LZI und dblp fand von Anfang an in der internationalen Wissenschaftsgemeinschaft großen Anklang. In Gesprächen mit Seminarteilnehmern auf Schloss Dagstuhl wurde uns immer wieder positiver Zuspruch zugetragen. Sowohl Schloss Dagstuhl als auch dblp gelten in der internationalen Informatikforschung als zwei wichtige Institutionen mit tadellosem Ruf, die sich zudem hervorragend ergänzen.

Wie die Projektergebnisse berichten, konnte dank der personellen Verstärkung des dblp-Teams die Literaturdatenbank sowohl inhaltlich weiterentwickelt als auch organisatorisch gestärkt werden. Dabei konnten vor allem bei der Weiterentwicklung der organisatorischen Struktur vielerlei Synergien genutzt werden. Zum Beispiel konnte beim Aufbau des dblp Advisory Boards von der hervorragenden Einbindung Schloss Dagstuhls in die deutsche Informatikgemeinschaft profitiert werden. Die Gespräche und Befragung der Seminarteilnehmer von Schloss Dagstuhl haben zudem wertvolle Hinweise für die inhaltliche Ausgestaltung von dblp geben können. Umgekehrt ist dblp ein Werkzeug, das Organisatoren, Teilnehmer und den wissenschaftlichen Stab von Schloss Dagstuhl bei der Vorbereitung und Durchführung der Dagstuhl-Seminare unterstützt.

Zur Verstetigung der Zusammenarbeit über das formale Projektende hinaus wurden deshalb unter dem Dach von Schloss Dagstuhl drei feste Mitarbeiterstellen für die Arbeit an dblp geschaffen. Aufgabe der Mitarbeiter ist es, die dblp-Infrastruktur von Schloss Dagstuhl zu betreiben und somit den dblp-Dienst auch auf lange Sicht für die internationale Informatikforschung zu garantieren. Diese Entwicklung wurde ebenfalls allenthalben positiv aufgenommen. Die für alle Seiten vorteilhafte und fruchtbare Zusammenarbeit mit Michael Ley von der Universität Trier soll dabei fortgesetzt werden.

## 4 Ausblick

Die Zusammenarbeit von Schloss Dagstuhl mit der Literaturdatenbank dblp hat bereits nach nur zweieinhalb Jahren substantielle Erfolge vorzuweisen. Die durch Personaverstärkung und technische Weiterentwicklung ermöglichte Produktivitätssteigerung hat schon jetzt die ursprünglichen Erwartungen

übertrafen. Dank der Verstärkung des Engagements von Schloss Dagstuhl für dblp besteht auch die Absicht, dieses Niveau in Zukunft zu halten und darauf aufzubauen.

Gleichzeitig wurde aber nicht nur die Quantität sondern auch die Vollständigkeit des Datenbestandes verbessert. Die Etablierung des dblp Advisory Board bedeutet einen wesentlichen Schritt vorwärts, um die wissenschaftliche Aufsicht über den dblp-Datenbestand auf solide Beine zu stellen. Die für die Datenaufnahme nötige Priorisierung von Publikationsreihen hat durch die Einbeziehung von Experten und externer Rankings eine qualitative Verbesserung erfahren. Die Hinzunahme von externem Sachverstand hat zudem zu einer verbesserten Abdeckung gerade in den multidisziplinären Randgebieten der Informatik geführt. Außerdem wurden bestehende Reihen der Kerninformatik bezüglich ihrer Abdeckung in dblp geprüft und Fehlstände konsequent behoben.

Die Entwicklung von dblp ist mit dem Ende des Projektes „LZI+DBLP“ aber selbstverständlich noch nicht abgeschlossen. Neben der tagtäglichen Neuaufnahme von Publikationen und dem steten Streben nach einer möglichst vollständigen Abdeckung bleibt noch eine Vielzahl von konzeptionellen Herausforderungen bestehen.

So gilt die Gewinnung von weiteren Verlagen als Partner bei der Datenakquise unverändert als ein Ziel von dblp. Die etablierten Kooperationen haben sich als sehr förderlich für die Effizienz und Aktualität der Datenakquise erwiesen. Auf den im Rahmen des Projektes gemachten ersten Erfahrungen soll weiter aufgebaut werden.

Ein offenes Problem bleibt zudem die Behandlung von Monographien in dblp. Das Selektions- und Priorisierungsproblem von Monographien steht dem von Journalen und Konferenzreihen in nichts nach, jedoch sind die jeweiligen Arbeitsabläufe bei Monographien notwendigerweise um einiges individueller und deren Ertrag um Größenordnungen geringer als bei den periodisch erscheinenden Reihen. Die Bedeutung der Indexierung von Monographien ist jedoch unbestritten. Hier gilt es in Zukunft eine Infrastruktur zu finden, die eine effiziente Selektion und Akquise analog zu den für Journale und Konferenzen etablierten Arbeitsabläufen ermöglicht. Zudem sollen (sowohl nationale als auch internationale) Dissertationen in Zukunft eine wichtigere Rolle bei dblp spielen. Erste Erfahrungen wurden mit Hilfe von Daten der Deutschen Nationalbibliothek gemacht; auf diesen Erfahrungen soll aufgebaut werden.

Die Befragung der Seminarteilnehmer von Schloss Dagstuhl liefert bereits regelmäßig wertvolle Hinweise auf übersehene oder unterschätzte Reihen. Mit der großflächigen Befragung unter den Mitgliedern der Gesellschaft für Informatik (GI) e.V. wurde zudem eine wertvolle Datenbasis für die zukünftige Neuaufnahme und Priorisierung von Publikationsreihen erhoben. Mit der deutschlandweiten Befragung der GI-Mitglieder wurde ein großer, erster Schritt zur Bildung einer eigenen bibliometrischen Infrastruktur geleistet. Diese Entwicklung gilt es zu verstärken und auszuweiten. Es ist bereits vorgesehen, die Erhebung im selben oder ähnlichen Expertenkreis mit etwas zeitlichem Abstand periodisch zu wiederholen. Die Ergebnisse der Evaluation sollen so kontinuierlich aktualisiert und weiter entwickelt werden. Aufbauend auf den gemachten Erfahrungen soll aber auch insbesondere eine Einbeziehung von zunächst europäischen (und später auch internationalen) Experten das Ziel sein.

Eine große Herausforderung bibliographischer Datenbank ist und bleibt die Sicherung der Datenqualität. Ein besonderes Alleinstellungsmerkmal von dblp liegt in der hochqualitativen Identifikation von Autoren-Entitäten an Hand von in der Regel mehrdeutigen Metadaten. Die Routinen und Arbeitsabläufe, die diesem Merkmal zu Grunde liegen, haben jedoch ihre Grenzen und Eigenarten. So stellt insbesondere der asiatische Namensraum dblp noch vor große Herausforderungen. Eine Disambiguierung asiatischer Namen ist derzeit meist nur durch sehr arbeitsintensive manuelle Recherche möglich.

Es wird ein zentrales Ziel von dblp sein, mittelfristig leistungsfähigere und dem Stand der Forschung entsprechende Methoden zu installieren, die die Qualität der Autorenidentifikation noch weiter steigern und eine effiziente Disambiguierung unterstützen. Methoden des maschinellen Lernens und der Computerlinguistik haben in den vergangenen Jahren auf diesem Gebiet bereits einige Fortschritte verzeichnen können. Diese theoretischen Verfahren sind aber in der Regel nicht auf die Bedürfnisse und Besonderheiten eines Einsatzes in einer sich kontinuierlich verändernden und entwickelnden Datenbank ausgerichtet. Um diese Lücke zwischen Theorie und Praxis zu schließen, soll der Kontakt mit einschlägigen Experten verstärkt und das Disambiguierungsproblem mit forschungsnahen Methoden praktisch angegangen werden. Ferner gilt es auch, die Zusammenarbeit mit anderen bibliographischen Informationsdiensten und digitalen Bibliotheken zu intensivieren, um diese große, gemeinsame Herausforderung auch gemeinsam zu meistern.

Zudem gibt es neue Trends in der wissenschaftlichen Kommunikation, für welche die eher traditionelle Struktur von dblp derzeit noch nicht aufgestellt ist. So haben verschiedene Reihen in der Informatik damit begonnen, hybride Strukturen zwischen Journalen und Konferenzen zu bilden, die von der dblp-eigenen Typisierung nicht eindeutig erfasst werden. Mega-Journale nach dem Modell von *PLOS One* publizieren statt Ausgaben einen kontinuierlichen Strom von Einzelartikeln und entziehen sich zudem gängigen Fächerkategorien. Online-Repositories wie arXiv.org verstehen Publikationen als lebendige Dokumente, die sich über die Zeit von Version zu Version wandeln können. Forschungsdaten werden künftig mit Metadaten versehen und als zitierfähige Publikationseinheiten verfügbar sein. Eine ähnliche Situation ist auch bei dem bibliographischen Nachweis von Programm-Quellcode gegeben. In all diesen Bereichen gilt es, zunächst beobachtend die aktuellen Entwicklungen zu begleiten und sich für die Zukunft richtig zu positionieren.

## **5 Wirtschaftliche Verwertbarkeit der Ergebnisse**

Alle Ergebnisse der Zusammenarbeit wurden im Sinne des Open-Data-Gedankens unter der „Open Data Commons Attribution License v1.0“ veröffentlicht und stehen damit der Allgemeinheit zur freien (auch wirtschaftlichen) Nachnutzung zur Verfügung.

## **6 Beiträge der Kooperationspartner**

Das Projekt LZI+DBLP wurde in Zusammenarbeit mit der Universität Trier durchgeführt. Neben der Expertise und Arbeitszeit des dblp-Gründers und Projektpartners Michael Ley sind zudem auch in bedeutendem Maße Beiträge der wissenschaftlichen und studentischen Mitarbeiter des Lehrstuhls „Datenbanken und Informationssysteme“ (Prof. Dr. Bernd Walter) in das Projekt eingeflossen. Ferner

profitierten die Projektmitarbeiter vor Ort in Trier allgemein von der Infrastruktur des Fachbereiches Informatik der Universität.

Es ist hervorzuheben, dass die Förderung durch die Klaus Tschira Stiftung ihren signifikanten Anteil an dem Erfolg des Projekts hat. Durch die Förderung der Projektvorbereitungsphase von November 2010 bis Mai 2011 konnten bereits erhebliche Teilziele im Bereich der technischen Infrastruktur vor dem offiziellen Projektbeginn realisiert werden. Zudem hat erst die Förderung der dritten Vollzeitstelle die Fortschritte insbesondere beim Aufbau der eigenen bibliometrischen Infrastruktur in dem erreichten Maße möglich gemacht.

## 7 Qualifikationsarbeiten

- keine

## 8 Publikationen

- Florian Reitz, Oliver Hoffmann: *Learning from the Past: An Analysis of Person Name Corrections in the DBLP Collection and Social Network Properties of Affected Entities*. The Influence of Technology on Social Network Analysis and Mining 2013: 427-453
- Florian Reitz, Oliver Hoffmann: *Did They Notice? - A Case-Study on the Community Contribution to Data Quality in DBLP*. TPDL 2011: 204-215

## 9 Pressemitteilungen und Medienberichte

- 10.12.2010: Open-Source Literaturnachweis in der Informatik: DBLP-Datenbank erhält Unterstützung von Schloss Dagstuhl  
<http://www.dagstuhl.de/de/ueber-dagstuhl/presse/pressemitteilungen/detail/meldung/345/>
- 14.12.2010: DBLP-Datenbank der Uni Trier wird von Schloss Dagstuhl unterstützt  
[http://www.uni-trier.de/index.php?id=14187& tx\\_ttnews\[tt\\_news\]=10821](http://www.uni-trier.de/index.php?id=14187& tx_ttnews[tt_news]=10821)
- 12.11.2012: Mehr als zwei Millionen Einträge —Datenbank DBLP wächst als Kooperationsprojekt weiter  
<http://www.uni-trier.de/fileadmin/organisation/Presse/Unijournal/Umbruch-UJ-3-12-WEB.pdf#page=34>
- 05.06.2013: Erhebung zu den Publikationsorganen der deutschen Informatik geplant  
<http://www.dagstuhl.de/de/ueber-dagstuhl/presse/pressemitteilungen/detail/meldung/489/>
- 26.06.2013: Datenbank dblp ist in 20 Jahren rasant gewachsen  
[http://www.uni-trier.de/index.php?id=14187& tx\\_ttnews\[tt\\_news\]=15028](http://www.uni-trier.de/index.php?id=14187& tx_ttnews[tt_news]=15028)
- 04.07.2013: Datenbank dblp ist in 20 Jahren rasant gewachsen und für Informatiker unverzichtbar  
<http://www.dagstuhl.de/de/ueber-dagstuhl/presse/pressemitteilungen/detail/meldung/496/>
- 28.07.2013: Der virtuelle Zettelkasten  
<http://www.volksfreund.de/nachrichten/region/trier/Heute-in-der-Trierer-Zeitung-Der-virtuelle-Zettelkasten;art754,3597087>



- *14.11.2013*: Eine Herkulesaufgabe im Dienst der Informatik — Vor 20 Jahren gründete Dr. Michael Ley die Datenbank dblp  
[http://www.uni-trier.de/fileadmin/organisation/Presse/Unijournal/Ab\\_2010/Unijournal\\_3\\_2013.pdf#page=14](http://www.uni-trier.de/fileadmin/organisation/Presse/Unijournal/Ab_2010/Unijournal_3_2013.pdf#page=14)
- *22.01.2014*: Datenbank dblp knackt 2,5-Millionen-Marke  
[http://www.uni-trier.de/index.php?id=14187&tx\\_ttnews\[tt\\_news\]=15914](http://www.uni-trier.de/index.php?id=14187&tx_ttnews[tt_news]=15914)
- *22.01.2014*: Datenbank dblp knackt 2,5-Millionen-Marke  
<http://www.dagstuhl.de/de/ueber-dagstuhl/presse/pressemitteilungen/detail/meldung/517/>