

Der Schritt zurück als Schritt nach vorn – Macht der Siegeszug des Open Access Bibliotheken arbeitslos?

7. Konferenz der Zentralbibliothek, Forschungszentrum Jülich, 14. – 16. Juni 2016

Bernhard Mittermaier (Hrsg.)

Proceedingsband

WissKom
2016



Forschungszentrum Jülich GmbH
Zentralbibliothek, Verlag

Der Schritt zurück als Schritt nach vorn – Macht der Siegeszug des Open Access Bibliotheken arbeitslos?

7. Konferenz der Zentralbibliothek
Forschungszentrum Jülich

Bernhard Mittermaier (Hrsg.)

14. – 16. Juni 2016
Proceedingsband

Schriften des Forschungszentrums Jülich
Reihe Bibliothek / Library

Band / Volume 22

ISSN 1433-5557

ISBN 978-3-95806-146-0

Bibliografische Information der Deutschen Nationalbibliothek.
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte Bibliografische Daten
sind im Internet über <http://dnb.d-nb.de> abrufbar.



Herausgeber
und Vertrieb: Forschungszentrum Jülich GmbH
Zentralbibliothek, Verlag
52425 Jülich
Tel.: +49 2461 61-5368
Fax: +49 2461 61-6103
E-Mail: zb-publikation@fz-juelich.de
www.fz-juelich.de/zb

Umschlaggestaltung: Grafische Medien, Forschungszentrum Jülich GmbH

Titelbild: © Coloures-pic/Fotolia.com

Druck: Grafische Medien, Forschungszentrum Jülich GmbH

Copyright: Forschungszentrum Jülich 2016

Schriften des Forschungszentrums Jülich
Reihe Bibliothek / Library, Band / Volume 22

ISSN 1433-5557
ISBN 978-3-95806-146-0

Vollständig frei verfügbar über das Publikationsportal des Forschungszentrums Jülich (JuSER)
unter www.fz-juelich.de/zb/openaccess.



This is an Open Access publication distributed under the terms of the [Creative Commons Attribution License 4.0](https://creativecommons.org/licenses/by/4.0/),
which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Konferenzkomitees	iii
Vorwort	1
Festvortrag	
Negotiating your requirements in the transition to gold open access – best practices	5
Estelle Lorraine	
Wandel im Publikationsprozess	
Gold Open Access verändert Bibliotheken – ein Call-To-Action	17
Bernhard Mittermaier	
Publizieren ist nicht genug	23
Alexander Wagner	
Wandel aktiv gestalten: Das Projekt INTACT- Transparente Infrastruktur für Open-Access-Publikationsgebühren	47
Kai Geschuhn, Dirk Pieper	
Nachweissysteme	
Gemeinsam sind wir stärker – die neue Digitale Bibliothek Thüringen (DBT) als kooperative Publikationsplattform Thüringer Hochschulen	73
Thomas Mutschler	
Autorenidentifikation als (neue) Dienstleistung für wissenschaftliche Spezialbibliotheken	87
Annette Polly, Ariane Streicher	
Recommender-System für Projektkollaborationen basierend auf wissenschaftlichen Publikationen und Patenten	103
Christoph Quix, Sandra Geisler, Rihan Hai	
Repositorien	
Open Access in der Leibniz-Gemeinschaft – Die Rolle der ZBW als Infrastrukturdienstleister	119
Ralf Flohr, Jan B. Weiland	
PUBLISSO – Das Open-Access-Publikationsportal der Lebenswissenschaften: Wissenschaftsunterstützung durch Open-Access-Publikationsangebote	135
Ursula Arning	
Fachrepositorium Lebenswissenschaften: Ausbau des Repositoriums zur Publikation von Forschungsdaten	147
Birte Lindstädt	
Green Open Access	
E-Periodica: die Plattform für digitalisierte Schweizer Zeitschriften	161
Regina Wanger, Michael Ehrismann	
Green Open Access im Bibliothekskatalog - Chancen & Risiken	177
Martin Blenke, Rachel Ellis, Elmar Haake, Manfred Nölte	
DeepGreen - Entwicklung eines rechtssicheren Workflows zur effizienten Umsetzung der Open-Access-Komponente in den Allianz-Lizenzen für die Wissenschaft	189
Markus Putnings, Thomas Dierkes	

Forschungsdaten

Serviceleistungen der Bibliothek des Deutschen Krebsforschungszentrums im Bereich Forschungsdatenmanagement	203
Dagmar Sitek	

Ein kooperativ nutzbares Werkzeug zur Erstellung von Datenmanagementplänen	213
Franziska Ackermann, Dennis Wehrle	

Wissenschaftsevaluierung

Wie messen wir unseren Open-Access-Anteil?	231
Marco Tullney	

Vorstellung des JOIN2 Statistikmoduls mit seinen Differenzen und Problemen zum kommenden Kerndatensatz Forschung	245
Robert Thiele, Katrin Große	

Autorenliste.....	257
Sponsoren und Aussteller	265

Konferenzkomitees

Programmkomitee

Dr. Rafael Ball	ETH Zürich, ETH-Bibliothek
Dr. Ulrike Eich	RWTH Aachen, Universitätsbibliothek
Dr. Antje Kellersohn	Albert-Ludwigs-Universität Freiburg, Universitätsbibliothek
Dr. Bernhard Mittermaier	Forschungszentrum Jülich, Zentralbibliothek
Heinz Pampel	Helmholtz Open Science Koordinationsbüro
Dr. Ralf Schimmer	Max-Planck-Gesellschaft, Max Planck Digital Library (MPDL)
Frank Scholze	Karlsruher Institut für Technologie, KIT-Bibliothek

Organisationskomitee

Thomas Arndt	Forschungszentrum Jülich, Zentralbibliothek
Hilde Dobbelsstein	Forschungszentrum Jülich, Zentralbibliothek
Charlotte Esser	Forschungszentrum Jülich, Zentralbibliothek
Yasmin Fattah	Forschungszentrum Jülich, Zentralbibliothek
Dr. Christoph Holzke	Forschungszentrum Jülich, Zentralbibliothek
Dr. Bernhard Mittermaier	Forschungszentrum Jülich, Zentralbibliothek

**Der Schritt zurück als Schritt nach vorn –
Macht der Siegeszug des Open Access Bibliotheken arbeitslos?**

Die klassische Aufgabe der Bereitstellung lizenzierter Literatur verliert zumindest im Bereich wissenschaftlicher Zeitschriften umso mehr an Bedeutung, je größer der Anteil der Open Access-Zeitschriften wird. Wenn der von vielen angestrebte Transformationsprozess im wissenschaftlichen Publikationswesen gelingt, dann droht in Bibliotheken ein großes Aufgabengebiet wegzubrechen. Statt dieser am Ende des wissenschaftlichen Publikationsprozesses angesiedelten Aufgaben ergeben sich jedoch möglicherweise neue Betätigungsfelder, die weiter vorn im Publikationsprozess angesiedelt sind: Ist die Begleitung wissenschaftlicher Autoren im Publikationsprozess von A bis Z eine zeitgerechte Aufgabe für Bibliotheken? Welche technischen und administrativ-strukturellen Voraussetzungen müssen geschaffen werden, um einen Publikationsprozess begleiten zu können und Bibliotheken in die Lage zu versetzen als Dienstleister gegenüber Lehre, Wissenschaft und Forschung auftreten zu können? Und was gibt es in puncto Forschungsdatenmanagement und Evaluierung zu tun?

Mit der WissKom2016 greift die Zentralbibliothek des Forschungszentrums Jülich erneut Themen im Spannungsfeld von „Bibliothek - Information – Wissenschaft“ interdisziplinär auf und versucht, neue Handlungsfelder für Bibliotheken aufzuzeigen. Diese siebte Konferenz der Zentralbibliothek thematisiert den sich abzeichnenden Wandel im Publikationswesen primär aus der Sicht der Bibliotheken, die vor ähnlich großen Umbrüchen stehen wie sie seit den 1990er Jahren durch die Digitalisierung der Literatur hervorgerufen wurden.

Ich danke allen herzlich, die zum Gelingen der Konferenz beigetragen haben: Den Vortragenden und Moderatoren, den Ausstellern und Sponsoren, den Mitgliedern des Programmkomitees, den Organisatoren und last but not least allen Teilnehmerinnen und Teilnehmern.

Dr. Bernhard Mittermaier
Forschungszentrum Jülich, Zentralbibliothek

Festvortrag

Negotiating your requirements in the transition to gold open access – best practices

Estelle Lorraine

Director COUNTER
COUNTER, Winchester, Hampshire, United Kingdom

In July 2012, the UK Government accepted the e recommendations of the Working Group on Expanding Access to Published Research Findings, chaired by Dame Janet Finch. This report became known as the “Finch Report”. The working group, first set up in October 2011, had membership from the UK higher education sector and publishing industry. The report recommended a clear policy direction in the UK towards support for ‘Gold’ open access publishing, the model where publishers receive their revenues from authors rather than readers, and so research articles become freely accessible to everyone immediately upon publication.

In response the Research Councils UK, which invests around £3billion in research across the full spectrum of academic disciplines announced a new open access policy to come from 1 April 2013. The policy meant a requirement for all research articles submitted for funding to be published in open access, with a preference for gold open access.

The Research Councils UK provided some block funding to some universities to meet the extra cost of Gold open access, but it became clear that the even with the block grants the total cost of publication for UK universities during a period of transition would mean a significant increase in cost. Universities would be required to pay article processing charges APCs to the same publishers to whom they already paid significant subscription fees for journals. These publishers had transitioned their journals to the hybrid open access model, in which predominately subscription journals have some of the articles are open access.

Early modelling by Jisc Collections, the library consortium for UK universities, indicated that if academics were fully compliant with the Research Councils UK by policy 2018, and subscriptions to journals maintained, expenditure for some publishers would more than double. Table 1 below illustrates the anticipated extra cost for one UK university with one of the publishers to which it subscribes:

Institution A: Annual subscription expenditure with publisher 1 in 2014	£28,895
Publisher 1's average Article Processing Charge (APC)	£1,800
Anticipated number of open access article to be published by authors from Institution A in 2014	12
Total expenditure for APCs with Publisher 1 in 2014	£21,600
Total Cost of Publication (subscription + APCs)	£50,459

Table 1

As Chris Banks, the library director at Imperial College London, has so succinctly put it, “From the librarian’s perspective, OA may have started as a means of using the disruptive innovation that came about with the development of the internet in order to seek to drive down journal subscription costs, but it is increasingly seen as a mechanism by which publicly funded research can be communicated to the public at the earliest possible opportunity and without having to pay at the point of use. OA means moving the costs of publishing upstream, from the point of consumption to the point of production. At the moment, though, we are in a transition period where costs of publishing as well as the costs of subscribing are incurred by institutions. This is widely referred to as ‘double dipping’ and it is not sustainable.”¹

It was clear, that without publisher agreement to introduce offset systems, a transition to Gold open access would not be sustainable for UK universities. In this context, offset systems would be the method of compensating the expenditure on APCs with a reduction in the cost of subscriptions or vis-a-versa. Jisc Collections, set about negotiating with scholarly publishers for the introduction of offset systems.

¹ Banks, C., (2016). Focusing upstream: supporting scholarly communication by academics. Insights. 29(1), pp.37–44. DOI: <http://doi.org/10.1629/uksg.292>

The issue of double-dipping

There were some challenges to be overcome in these negotiations. A number of publishers responded to Jisc Collections by stating that they did not “double dip”. Indeed, this is true to the extent that publishers considered “double dipping” to mean being paid twice by not considering the number of open access articles when setting subscription prices. Many already had “no double dipping” policies in place. These operate on principle that in a hybrid journal the publisher will reduce the global subscription price for all subscribers as the proportion of Gold open access articles in the journal increases. Such policies, designed to operate on a global scale, would probably be satisfactory if all countries moved to Gold open access at the same speed. However, the Research Councils UK policy, put the UK at the forefront of a transition to gold open access and it was clear that the publishers’ “no double dipping” policies, while welcome, would have little impact in reducing the local total cost of publication.

The UK represents just 9.5% of global research article downloads. Assuming Gold open access articles were predominately funded by UK universities, it might be expected that on average subscription fees for journals would reduce by around the same amount. Table 2 models the negligible benefit of a 9.5% global price reduction in subscriptions for the same university and publisher modelled in Table 1.

Institution A: Annual subscription expenditure with publisher 1 in 2014 £28,895 less a 9.5% reduction for open access articles.	£26,149
Publisher 1's average Article Processing Charge (APC)	£1,800
Anticipated number of open access article to be published by authors from Institution A in 2014	12
Total expenditure for APCs with Publisher 1 in 2014	£21,600
Total Cost of Publication (subscription + APCs)	£47,749

Table 2

The issue was not about “double-dipping” on a global scale, but the huge increase in costs at a local level. Therefore, Jisc Collections focused the discussion on local offsetting to reduce the total cost of publication and easing the transition to open access, rather than double-dipping.

At the early stages of these discussions some publishers responded by telling Jisc Collections that as a library consortium it had no mandate to negotiate, because APCs costs are not library expenditure. Jisc Collections gathered support from UK higher universities to negotiate on their behalf, and indeed were also supported by Rt. Hon. David Willetts MP, who published an open letter to Dame Janet Finch. In it he said: “Universities, Jisc, funders and publishers should work together, within the constraints of completion law, to consider whether and how, expenditures and revenues for APCs might be offset against each other.”

Data Collection

The other rebuttal that Jisc Collections received from publishers was that its modelling of the increase in the total cost of publication was supposition and that it had no data to support its argument. That was true, this was a new area and no data about the spend by UK universities on APCs was available. Jisc Collections set about collecting data to support the argument for offset systems.

The first set of data was collected from 22 universities. The data collected showed that in 2014, expenditure on Gold open access accounted for over 17% of total expenditure with the largest publishers. Stephen Pinfield and his colleagues at the University of Sheffield analysed the data. They found that APC payments in the UK had increased since 2012 (the year the Research Councils UK policy was implemented), with payments by the 23 HEIs rising from less than £1 million in 2012 to more than £4 million in 2013 (see Table 3)².

² Pinfield, Stephen, Salter, Jennifer, Bath, Peter A., The “total cost of publication” in a hybrid open-access environment: Institutional approaches to funding journal article-processing charges in combination with subscriptions, Journal of the Association for Information Science and Technology, J Assn Inf Sci, 2330-1643, <http://dx.doi.org/10.1002/asi.23446>, 10.1002/asi.23446, 2015

Year	APC numbers (number of institutions making payments)	Total cost	Mean	Median	Minimum	Maximum
2007	31 (1)	£40,595	£1,310	£1,250	£235	£2,827
2008	67 (1)	£108,442	£1,619	£1,432	£456	£4,022
2009	99 (3)	£177,200	£1,790	£1,725	£246	£4,023
2010	380 (8)	£641,798	£1,689	£1,761	£115	£4,800
2011	469 (9)	£818,150	£1,744	£1,800	£175	£5,280
2012	570 (14)	£977,848	£1,716	£1,738	£183	£4,800
2013	2445 (23)	£4,097,981	£1,676	£1,680	£82	£4,955
2014	1081 (23)	£1,784,879	£1,651	£1,554	£107	£4,660
Total	5142	£8,646,892	£1,682	£1,674	£82	£5,280

Table 3 **Total number and cost of APCs per year for the 23 institutions, 2007-2014**

Source: [Pinfield, Salter & Bath \(2015\)](#).

The universities provided their subscription expenditure and so Pinfield et al were able to model this against and APC expenditure to calculate their total cost of publication. They found that overall the total cost of publication for those universities (excluding admin costs) was £32.7 million, of which £29.4 million was for subscriptions (90%) and £3.3 million for APCs (10%) (see Table 4). APCs were then already 10% of the total cost of publication. However, it should be noted that some institutions were able to comply with the new policy more quickly and in some cases the APCs were then already a much higher percentage of the total cost of publication.

Institution	Total Subscription Costs (%)	Total APC Costs (%)	Total Cost
1	£578,708 (94%)	£34,186 (6%)	£612,894
2	£1,053,260 (93%)	£73,777 (7%)	£1,127,037
3	£2,274,060 (90%)	£242,601 (10%)	£2,516,661
4	£1,756,783 (89%)	£206,404 (11%)	£1,963,187
5	£2,816,456 (91%)	£275,148 (9%)	£3,091,604
7	£2,025,761 (86%)	£332,363 (14%)	£2,358,124
8	£2,781,917 (85%)	£473,557 (15%)	£3,255,474
9	£1,815,342 (91%)	£189,200 (9%)	£2,004,542
10	£934,655 (95%)	£54,165 (5%)	£988,820
11	£1,403,884 (99%)	£10,209 (1%)	£1,414,093
12	£1,821,589 (96%)	£68,078 (4%)	£1,889,667
13	£264,492 (61%)	£170,246 (39%)	£434,738
14	£2,194,903 (90%)	£239,940 (10%)	£2,434,843
15	£865,998 (93%)	£63,678 (7%)	£929,676
16	£139,168 (95%)	£6,691 (5%)	£145,859
17	£44,875 (72%)	£17,603 (28%)	£62,478
19	£887,186 (97%)	£23,421 (3%)	£910,607
21	£829,924 (98%)	£18,444 (2%)	£848,368
22	£3,271,535 (81%)	£763,602 (19%)	£4,035,137
23	£1,631,646 (97%)	£49,366 (3%)	£1,681,012
Totals	£29,392,142 (90%)	£3,312,679 (10%)	£32,704,821

Table 4 **Total Cost of Publication, 2013 (excluding administrative costs) for 20 institutions**
Source: [Pinfield, Salter & Bath \(2015\)](#).

Implementation of offset systems

By the autumn of 2014, some considerable success had been achieved with a number of publishers including, SAGE, Institute of Physics, Royal Society of Chemistry, Wiley, and Taylor & Francis implementing offsetting systems. In October 2015, Jisc and Springer were able to announce the Springer Compact agreement, a pilot that combines open access publishing and subscription access in one annual fee which will run from October 2015 until December 2018.

Initially, Jisc was not prescriptive about the type of offset system publishers introduced, providing the proposed system had potential to reduce the total cost of publication to UK institutions. The systems implemented in the early days differ in their design:

- Offsetting APC spend against subscription costs
- Credits to be spent on APCs based on a university's total expenditure
- APC vouchers based on a university's amount of subscription spend
- "Flipped models" – subscription fees pay for APCs, with a modest additional fee to cover access to content

However, in April 2015, having gained some experience of the different systems in practice, Jisc Collections published its *Principles for Offset Agreements*³. The five principles reflect UK universities' expectations about what will drive the design of effective offset systems:

"1. Systems should operate in the context of a transition to fully gold open access and support that transition. To meet this principle, a system should be inclusive, remove barriers (both to authors and their institutions) to open access and ensure that all the outputs of a subscribing institution are immediately open on publication under licences and other conditions which meet funders' mandates and other requirements. These systems will be designed to enable the publisher's journals to become fully open access as soon a tipping point is reached.

2. Systems should ensure that publishers do not charge the same institutions twice, through the payment of subscriptions and the payment of APCs.

3. Offset systems should not be restricted to institutions that subscribe to large collections of journals (the big deal) but should also apply to all institutions that subscribe to individual journals with a hybrid OA offering from a publisher

4. Offset systems should also apply, at the level of each subscribing institution, to ensure that the cost incurred by each institution for scholarly publishing is contained. Publishers should also apply global reductions to the cost of subscription in respect of increasing volumes of open access articles in hybrid journals in line with their stated "No Double Dipping Policies".

5. Systems should operate on a "cash basis" and avoid the additional administration and work involved in handling vouchers, particularly if those vouchers have an expiry date. Where an offset system does operate on the basis of vouchers, they must be available to the institution (which processes the transactions) rather than to individual authors.

Jisc pointed out that these principles are not set in stone, and that they will be refined and updated as data is gathered on the effectiveness of each offsetting system and to enable the systems to be re-shaped and improved.

³ <https://www.jisc-collections.ac.uk/Global/News%20files%20and%20docs/Principles-for-offset-agreements.pdf> [accessed 27 May, 2016]

APCs: The cost of administration versus an effective market

It is perhaps worth exploring one aspect of offset systems in terms of their impact on a long held assumption that Gold open access would encourage researchers to develop price sensitivity in the selection of their publication venue. Although the cost of administration has not been a major feature in the argument for offset systems, the experience of the UK showed that the cost to a university of administering individual APC payments is not insignificant, at an average of £88 per article⁴. Agreements such as the Springer Compact, which combines open access publishing and subscription access in one annual fee, has the advantage of removing the extra cost of administration for the university and the publisher. The cost of open access, in effect, becomes invisible to the researcher in the same way that the cost of journal subscriptions has always been invisible. “Flipped models” such as Springer Compact, mean that researcher submits her or his paper in the usual way with no financial transaction required to ensure open access. There those who might see this as regrettable side effect of such an offset system, although others, such as Martin Eve argue, “To believe that price for the actual service rendered would trump this perceived brand-value will remain a naïve position while scholarly communication retains its dual roles in dissemination *and* assessment.”⁵

Development and implementation of standards, metadata and workflows

Organisations such as Jisc are now not only concerned with the cost of APC and offsetting systems, but also that publishers adopt the standards, metadata and workflows that will satisfy the requirements of research funders and enable a smooth transition to open access. For example, Jisc has been talking with Crossref about its proposed metadata schemas and application programming interfaces (APIs) that will enable funders, universities and publishers to research more effectively. Clearly, such dialogue between all the stakeholders in scholarly communications is required to address the issues. In conclusion will again quote from Chris Banks:

⁴ Pinfield, Stephen., Current Research into the Total Cost of Publication – UKSG <http://www.uksg.org/sites/uksg.org/files/PresentationPinfield.pdf> [accessed 27 May, 2016]

⁵ Eve, M.P., (2015). Co-operating for gold open access without APCs. Insights. 28(1), pp.73–77. DOI: <http://doi.org/10.1629/uksg.166>

“HEIs now think about total cost of ownership and those costs no longer simply comprise the subscription costs to journal content. They include the support services that the publisher gives to the academic; the infrastructure that the publisher puts in place in order to minimize – or, ideally, all but eliminate – the transactional costs, particularly at the article level; the extent to which the publisher, alongside funders and institutions, adopts and implements the many standards which now exist to support easy and timely exchange of data; the extent to which they work with academics, funders and institutions.”⁶

Lorraine Estelle, May 2016

⁶ Banks, C., (2016). Focusing upstream: supporting scholarly communication by academics. *Insights*. 29(1), pp.37–44. DOI: <http://doi.org/10.1629/uksg.292>

Wandel im Publikationsprozess

Gold Open Access verändert Bibliotheken – ein Call-To-Action

Bernhard Mittermaier

Forschungszentrum Jülich, Zentralbibliothek

Seit zwei Dekaden folgt der Anteil der Gold Open Access-Publikationen am Gesamt-aufkommen der Zeitschriftenartikel einem exponentiellen Wachstum (Abbildung 1). Extrapoliert man die Entwicklung aus der Vergangenheit in die Zukunft, so kann man für das Jahr 2019 einen Gold OA-Anteil von 25% annehmen und im Jahr 2023 von 50%. Dem Kurvenverlauf folgend, wäre im Jahr 2028 jede Publikation Gold OA. Unabhängig davon, ob der Kurvenverlauf in Zukunft genau so sein wird (insbesondere für Gold OA-Anteile nahe 100% ist dies sicher diskutabel), sollten sich alle Beteiligten auf ein solches Szenario einstellen. Zwei Drittel des Weges (in Jahren gerechnet) bis zu einem OA-Anteil von über 50% sind jedenfalls bereits zurückgelegt.

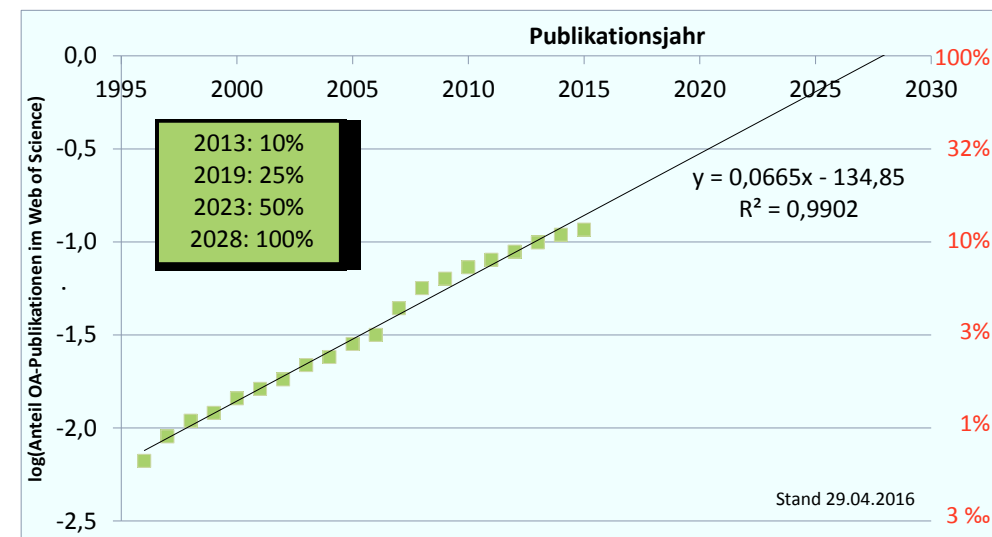


Abbildung 1: Entwicklung des Anteils der Open Access-Artikel an der Gesamtheit der Artikel im Web of Science mit Extrapolation in die Zukunft (logarithmische Darstellung). Datengrundlage: Alle Dokumente im Science Citation Index (SCIE), Social Science Citation Index (SSCI) und Arts&Humanities Index (A&HI)

Treiber der Entwicklung waren bislang vor allem das Eigeninteresse der Autoren und Anforderungen der Forschungsförderung. Letztere werden allem Anschein nach in Zukunft weiter steigen. So hat die Europäische Kommission im April 2016 einen Europäischen Aktionsplan für eine offene Wissenschaft¹ veröffentlicht, in dem u.a. gefordert wird: „Vollständiger offener Zugang (Open Access) zu allen öffentlich finanzierten wissenschaftlichen Publikationen im Jahr 2020.“ Die Bundesregierung plant für den Sommer 2016 ebenfalls die Veröffentlichung einer Open Access-Strategie; konkrete Zielvorgaben für die öffentlich finanzierte Forschung sind dabei nicht auszuschließen. Beschleunigend könnte in Zukunft auch die zunehmende Zahl von Offsetting-Verträgen² wirken. Dabei vereinbaren Forschungseinrichtungen (z.B. Max-Planck-Gesellschaft, Forschungszentrum Jülich) oder nationale Konsortien (z.B. in Großbritannien, den Niederlanden und in Österreich) mit Verlagen den Zugriff auf das gesamte Verlagsportfolio und gleichzeitig die Möglichkeit für alle dem Vertragspartner zuzurechnenden Autoren die Gold Open Access-Publikation ohne zusätzliche Kosten in den Zeitschriften des betreffenden Verlags. Diese sind dann genau genommen Hybrid Open Access-Publikationen, die beispielsweise im Web of Science nicht als Gold OA gezählt werden³. Nichtsdestotrotz sind sie ohne Embargofrist unmittelbar frei verfügbar. In Deutschland gibt es ab 2017 erste Allianz-Lizenzen im Offsetting-Modell, und auch das DEAL-Projekt der bundesweiten Lizenzierung von Zeitschriften großer Verlage soll eine Open Access-Komponente enthalten.

Relevant für die weitere Entwicklung ist noch zusätzlich die von der Max-Planck-Gesellschaft angestoßene Initiative OA2020⁴. Verschiedene internationale Forschungsorganisationen und Forschungsförderer haben sich der Initiative angeschlossen und fordern, den Weg für eine Umstellung des Publikationswesens auf Open Access bis zum Jahr 2020 zu bereiten.

¹ <http://deutsch.eu2016.nl/aktuelles/nachrichten/2016/04/05/europaischer-aktionsplan-fur-eine-offene-wissenschaft>, aufgerufen am 30.04.2016.

² <https://www.jisc-collections.ac.uk/Global/News%20files%20and%20docs/Principles-for-offset-agreements.pdf>, aufgerufen am 30.04.2016.

³ Nur Publikationen in Gold OA-Zeitschriften werden im Web of Science als „Gold Open Access“ verzeichnet. Die Zuordnung erfolgt also auf Ebene der Zeitschrift und nicht auf Ebene des individuellen Artikels.

⁴ <http://www.oa2020.org>, aufgerufen am 30.04.2016.

Bei vielen Bibliotheken und ihren Trägern ist das Bewusstsein über die Dynamik der Entwicklung und über die Konsequenzen der Transformation von der Subskription zu OA für das eigene Handeln sehr gering ausgeprägt.⁵ Nachstehend sind einige Handlungsfelder benannt; die Aufzählung erhebt keinen Anspruch auf Vollständigkeit.

1) Finanzströme:

In einer Gold OA-Welt zahlen wissenschaftliche Einrichtungen für das Publizieren an die Verlage, während sie in der Subskriptionswelt für das Lesen bezahlt haben. Zahlungsempfänger sind weiterhin in jedem Fall die Verlage, aber die Rolle der Zahlungspflichtigen könnte sich eventuell ändern. In der Subskriptionswelt waren es die Bibliotheken (z.T. auch Fakultäts- und Institutsbibliotheken), in der Gold OA-Welt könnten es, neben den Bibliotheken, auch die Autoren selbst resp. deren Institute sein. So postuliert der Open Access-Aktivist Stevan Harnad: *„It's authors who publish, and authors who pay to publish (if they wish, or must). Author payment is not a subscription matter, not a library matter, and not a library aggregator matter.“*⁶ Abgesehen vom Umstand, dass Publizieren ein Teil der Forschungsarbeit ist (Lesen allerdings auch), spricht für diesen Ansatz vor allem, dass er eine (gewisse) Preissensitivität bei den Autoren erzeugen kann: Wenn sie die Publikationsgebühren selbst entrichten müssen (wobei „selbst“ in den seltensten Fällen das persönliche Vermögen meint, sondern meist den Institutsetat), dann sind Autoren u.U. weniger bereit, sehr hohe Publikationsgebühren zu bezahlen, sondern sehen sich ggf. nach einer fachlich ebenso geeigneten und ähnlich renommierten Zeitschrift mit niedrigeren Gebühren um. Andererseits hat de facto nur die Bibliothek die Möglichkeit, über die Höhe der Publikationsgebühren zu verhandeln und beispielsweise durch Depositzahlungen Rabatte zu erzielen. Auch ist die betriebs- und bibliothekswissenschaftliche Erkenntnis nicht neu, dass die Bündelung derartiger Aufgaben in einer Hand Synergieeffekte gegenüber einer Dezentralisierung bringt. Die Bibliothek kann die Abwicklung der Publikationsgebühren rationeller vornehmen, als dies bei einer verteilten Einzelabwicklung der Fall wäre. Letztlich wird es eine Existenzfrage für Bibliotheken sein, ob Ihnen die Re-Kontextualisierung des Bibliothekserwerbungssetats⁷ gelingt. Andernfalls werden mit den Erwerbungsmitteln auch die Personalmittel schwinden.

2) Abwicklung von Publikationsgebühren

Schritte der Rationalisierung und Vereinheitlichung bei der Abwicklung von Publikationsgebühren sind in jedem Fall nötig. Man kann abschätzen, dass die Zahl der zu bearbeitenden Rechnungspositionen in einer Gold OA-Welt etwa fünfmal größer ist, als in der Subskriptionswelt. Dies ist nur zu bewältigen, wenn die Zahlungen aus Deposits abgebucht werden oder statt Einzelrechnungen Sammelabrechnungen erstellt werden. Es ist weiter zu fordern, dass mittelfristig nicht mehr jeder Verlag seinen eigenen Standard entwickelt, sondern dass ein einheitlicher Standard für alle Verlage gilt. Standardisierte, maschinenlesbare elektronische Rechnungen mit den kaufmännischen Daten sind ebenso nötig, wie standardisierte Lieferungen

⁵ Strohschneider, Peter: Kein »entweder – oder«. Positionen der Deutschen Forschungsgemeinschaft zu Open Access. <http://www.forschung-und-lehre.de/wordpress/?p=17955>, aufgerufen am 30.04.2016.

⁶ <http://listserv.crl.edu/wa.exe?A2=LIBLICENSE-L;f530d4fc.1307>, aufgerufen am 30.04.2016.

⁷ Schimmer, R (2012): Open Access und die Re-Kontextualisierung des Bibliothekserwerbungssetats. Bibliothek, Forschung und Praxis 36, 293-299.

der bibliographischen Daten einschließlich des Volltexts. Die ESAC-Initiative⁸ der Max Planck Digital Library befasst sich mit diesen Fragestellungen.

3) Höhe der Publikationsgebühren

Vielfach wird befürchtet, dass in einer Gold Open Access-Welt das gleiche eintritt wie in der Subskriptionswelt: Immer stärker steigende Preise weit jenseits der Inflationsrate, die in den Bibliotheksbudgets nicht mehr abgebildet werden können. Was in der Zeitschriftenkrise zu massiven Abbestellungen von Abonnements geführt hat, könnte in einer Gold Open Access-Welt dazu führen, dass nach Ausschöpfen der verfügbaren Finanzmittel nicht mehr publiziert werden kann. Diese Gefahr muss sicher entgegen getreten werden. Es gibt jedoch einige Aspekte, die auf Unterschiede in beiden Szenarien hinweisen und die Situation als lösbar erscheinen lassen: Preissteigerungen bei Zeitschriften führen dazu, dass vor Beginn eines Lizenzierungsjahres abbestellt wird und die Zeitschrift somit ganzjährig nicht im Zugriff ist – letztlich eine im Voraus zu treffende ja/nein-Entscheidung. In einer Gold Open Access-Welt würden die für Publikationen zur Verfügung stehenden Mittel im Laufe des Jahres zur Neige gehen. Dieses vor Augen würden sich Autoren bemühen, im Rahmen der Forschungsförderung zusätzliche Mittel für Publikationsausgaben zu akquirieren (für den lesenden Zugriff nicht möglich) und sie würden sich überlegen, ob es nicht auch kostengünstigere oder sogar kostenfreie, alternative Publikationsmöglichkeiten gibt. Angesichts dessen, dass der Publikationsoutput eine maßgebliche Kennzahl bei Evaluationen darstellt, wird es die Leitung einer Einrichtung wohl nicht zulassen, dass die Veröffentlichungen von Forschungsergebnissen an mangelnden Finanzmitteln scheitern (während Zusatzmittel für Erwerbungsausgaben eher selten sind). Dies kann jedoch keine Dauerlösung darstellen, sondern müsste der Einstieg sein in die Entwicklung eines Kostenbewusstseins bei den Akteuren. Es wäre hierfür in jedem Fall wichtig, die Kostenentwicklung im Publikationsbereich im Blick zu behalten, und zwar sowohl lokal für die eigene Einrichtung als auch global⁹. Schließlich ist zu überlegen, ob es sinnvoll ist als Einrichtung Obergrenzen festzulegen, bis zu welchem Höchstbetrag Publikationskosten gezahlt werden.¹⁰ Dies würde in der Summe sicher preisdämpfend wirken. Und schließlich ist es auch eine zukünftige Aufgabe von Bibliotheken, mit Verlagen in Verhandlungen über die Höhe der Publikationsgebühren einzutreten.

4) Zeitschriftenmanagement

Derzeit weisen Bibliotheken in ihren Katalogen diejenigen Zeitschriften nach, die sie lizenziert haben; manche zusätzlich frei verfügbare Zeitschriften, die in der EZB „grün“ geschaltet sind. Die Mühe einer Reklamation beim Verlag in Fällen von Zugangsproblemen wird man sich aber wohl nur bei den selbst lizenzierten Zeitschriften machen. Wie wird man hier in Zukunft vorgehen? Keinerlei Interaktion mit dem Verlag ist wohl ebenso wenig eine Lösung wie das „Kümmern“ um alle Zeitschriften.¹¹

⁸ <http://esac-initiative.org/>, aufgerufen am 30.04.2016

⁹ Morrison H, Mondésir G, Salhab J, Villamizar C, Calvé-Genest A, Desautels L. (2016): Open Access Article Processing Charges (OA APC) Longitudinal Study 2015 Preliminary Dataset. Data 1(1), 4.

¹⁰ So können aus Publikationsfonds, die mit Unterstützung der DFG eingerichtet wurden, nur Publikationskosten bis maximal 2000 € übernommen werden http://www.dfg.de/formulare/12_20/12_20_de.pdf, aufgerufen am 30.04.2016.

¹¹ Die EZB verzeichnet derzeit 85.553 Titel, davon 52.983 „grüne“ Zeitschriften <http://rzblx1.uni-regensburg.de/ezeit/about.phtml>, aufgerufen am 30.04.2016

Man wird hierfür eine Auswahl treffen müssen, deren Methodik erst noch zu entwickeln ist. Was für die Sicherstellung der Verfügbarkeit gesagt wurde, gilt analog für die Verzeichnung in Linkresolver-Systemen und Discovery-Services und für die Auswertung von Nutzungsstatistiken: Bibliotheken müssen neu entscheiden, was sie als „ihre“ Zeitschriften betrachten. Neben diesen individuellen Fragestellungen muss auch generell die Frage der Langzeitverfügbarkeit der Zeitschriften geregelt werden. Derzeit erwerben Bibliotheken mit ihren Lizenzverträgen in der Regel dauerhafte Zugriffsrechte; für die faktische Umsetzung dieses Rechts – auch im Fall des Verkaufs der Zeitschrift sowie des Untergangs des Verlags – gibt es verschiedene technische und organisatorische Lösungen. Bei Open Access-Zeitschriften gibt es keinen Vertragspartner, der entsprechende Rechte erwerben könnte. Die Autoren werden zwar in vielen Fällen ihre eigenen Artikel in Repositorien zweitveröffentlichen, dies ersetzt aber nicht den Zugang zur ganzen Zeitschrift. Man kann weiter annehmen, dass Suchmaschinen / Webcrawler die Inhalte indexieren, aber dies wird nicht zu befriedigenden Ergebnissen führen, mindestens unter Layout-Gesichtspunkten. Letztlich wird keine Alternative zu organisierten Hosting-Lösungen auf nationaler und/oder internationaler Ebene bestehen. Dies kann die Partizipation des Verlags an Sicherungssystemen wie Portico und LOCKSS sein. Da hiermit eine 100%ige Abdeckung aber wohl nicht zu erreichen ist, werden noch zusätzliche Anstrengungen nötig sein.

5) Personal

Mit der Verlagerung vom Subskriptionsgeschäft hin zur Abwicklung von Publikationsgebühren entstehen neue Aufgaben, für die möglicherweise Qualifizierungsmaßnahmen notwendig sind. Es steigt der Bedarf nach finanzbuchhalterischen Kenntnissen ebenso wie die Notwendigkeit, intensiv mit Autoren zu kommunizieren (was traditionell in einer Zeitschriftenstelle nicht unbedingt der Fall ist). Aber auch andere Bereiche der Bibliothek sind von der Transformation zu Open Access mitbetroffen: So ist zu erwarten, dass während des Übergangsprozesses, in dem massive Abbestellungen durchgeführt werden müssen, um Publikationskosten finanzieren zu können, der Bedarf nach Dokumentlieferungen zunächst ansteigt. Dieser erhöhte Arbeitsanfall wird später wieder abnehmen, wenn immer mehr Zeitschriften Gold Open Access verfügbar sind. Da dies in vielen Fällen jedoch nur die neuen Jahrgänge betreffen wird, wird der Rückgang des Arbeitsaufkommens nur langsam erfolgen und sich über einen langen Zeitraum erstrecken. Der wichtigste Aspekt ist allerdings die erwähnte indirekte Sicherung der Personalmittel dadurch, dass die Erwerbsmittel – mit anderer Zweckbindung – erhalten bleiben.

Publizieren ist nicht genug

Alexander Wagner

Deutsches Elektronen-Synchrotron DESY
Library and Documentation

Zusammenfassung

Im *Einreichungsprozess* einer Veröffentlichung erfassen Wissenschaftler traditionell die primären Metadaten ihrer Artikel, welche ggf. von Verlagen durch einige weitere Daten (Band, Seiten, DOI etc.) ergänzt werden, die eine genaue Identifikation und Zitation der jeweiligen Publikation erlauben.

Zahlreiche neue Anforderungen v. a. aus den Bereichen Reporting und Evaluation führen dazu, dass diese Verlagsmetadaten nicht mehr ausreichen um die die daraus resultierenden komplexen Anfragen zu beantworten. Das führt i. d. R. zu einer erneuten Erfassung der Publikationen in speziellen Nachweissystemen (Publikationsdatenbanken) und zur Anreicherung der Metadaten. Da die Publikation auf diese Weise von deren Nachweis entkoppelt wird stellt sich bereits hier das Problem des (hinreichend) vollständigen Nachweises, resultierend in einem z. T. immensen manuellen Rechercheaufwand an den Einrichtungen, an denen diese Publikationen entstanden sind. Häufig erfolgt zusätzlich eine weitere Erfassung zur Darstellung der Inhalte in den Repositorien.

Der Nachweis in Fachdatenbanken, kommerzieller wie nichtkommerzieller Natur, erfolgt i. d. R. wiederum aus Verlagsdaten. Hier fehlen daher alle in den Publikationsdatenbanken der Einrichtungen aufwändig nachgepflegten Werte wie z. B. Projekt-, Instituts- oder Personenzuordnungen. Somit ist die Darstellung der Ergebnisse sowohl für Wissenschaftler, aber v. a. Einrichtungen und Förderer häufig nicht optimal.

Auch auf Verlagsseite führen die derzeitigen Prozesse, vor allem im Licht APC-basierter Modelle, häufig zu Problemen. So liegen Daten zur Fakturierung falsch oder unvollständig vor und eine einzeltransaktionsbasierte Abrechnung ist aufwändig und teuer.

Ausgehend von den in den unterschiedlichen Kontexten notwendigen Daten und einem optimierten Prozess wird der Frage nachgegangen, wie man durch

möglichst einfache Änderungen an den derzeitigen Prozessen einen Mehrwert für alle Beteiligten schaffen kann.

Abstract

Scientists usually generate meta data for their publications upon submission. Publishers enrich those by additional data (volume, pages, doi etc.) allowing for a precise identification and citation.

Today, new requirements especially for reporting and evaluation, require additional data associated with each publication. As the publisher generated meta data is not sufficient, this usually results in cataloguing each publication again in specialised systems (publication databases) and enrichment of the data provided by those required for the specialized needs. However, this decouples publications from reporting and easily results in a lack of completeness. Tackling this requires huge amounts of manual work at the institutions to identify own publications and add them to their respective bibliographies. Sometimes, yet another step is required to feed the publications to the institutional Open Access repositories.

Furthermore, bibliographic databases generate their content from publishers meta data. Thus, their records usually lack all the additional values from the publication databases like project assignments, affiliations or author disambiguation. From an institutional as well as a funders perspective this is far from optimal.

As soon as APCs enter the game, the current processes are a source of trouble even on the publishers end. APCs in principle require individual billing on an article basis and thus additional meta data containing funding and affiliation information. Nowadays, incomplete or even plainly wrong data result in expensive and time consuming work flows.

Starting out from a short description how articles are born today and keeping in mind the data necessary in different contexts this article investigates possible improvements. The overall goal is to suggest a properly modularized work flow that can be adopted to the current system and realize gains for all stakeholders while minimizing the necessary changes.

Einleitung: Publikationsprozess und Außendarstellung

"A mathematician is a machine for turning coffee into theorems" (Alfréd Rényi, ungarischer Mathematiker).

Hierfür benötigen, nicht nur Mathematiker, zunächst Zugriff auf die bereits publizierte Literatur ihrer Kollegen. Diese Literatur besteht heute, je nach Arbeitsfeld, nicht nur aus Artikeln und Büchern (egal welcher Medienform), sondern ebenso aus publizierten Daten oder auch der für deren Bearbeitung nötigen Software. So stellt das Sammeln relevanter Quellen einen nicht unwesentlichen Teil der wissenschaftlichen Arbeit dar, und eine gut gepflegte und strukturierte Quellensammlung ist entsprechend wertvoll. Da ferner heute Aufsätze i. d. R. mit mehreren Koautoren erstellt werden¹, muss diese Sammlung, oder zumindest die für das konkrete Projekt bzw. die geplante Publikation relevanten Teile, eben diese Kollegen zugänglich sein². Dies gilt ebenso für die gemeinsam erzielten Ergebnisse wie eigene Publikationen der Gruppe (auch Qualifikationsschriften), Daten und/oder Software, die diese Sammlung ergänzen. Da zahlreiche Arbeitsgruppen für die Verwaltung dieser Quellensammlungen lokale Nachweissysteme in Eigenregie aufbauen, mithin Ressourcen in Entwicklung und Betrieb investieren, ist die Relevanz dieser Sammlungen offensichtlich.

Zum eigentlichen Erstellen neuer Publikationen sind weitere Werkzeuge notwendig³, die ferner den Koautoren Zugriff auf die Publikation und alle ihre relevanten Teile erlauben müssen. Idealerweise steht diese mit der Quellensammlung in engem Austausch so dass ein einfaches, korrektes Zitieren möglich ist⁴.

Spätestens im Falle (oft international) verteilt arbeitender Autorentteams drängt sich eine Implementation im Web auf, ist aber derzeit noch wenig

¹ Der Sonderfall des Einzelautors welcher v. a. in den Geisteswissenschaften häufig und bei Abschlussarbeiten die Regel ist, ergibt sich hier zwanglos als Grenzfall der "Ein-Personen-Kollaboration", und muss im Folgenden nicht explizit betrachtet werden.

² Es kann heute i. A. nicht davon ausgegangen werden, dass dies bei Verwendung kostenpflichtiger Angebote automatisch der Fall ist. Die Publikation im *Open Access* senkt hier offensichtliche Hürden.

³ z.B. Satzsysteme wie LaTeX oder Textverarbeitungen, Hilfsmittel zur Formatierung von Referenzen, Erstellung graphischer Darstellungen, Versionskontrollsysteme, etc.

⁴ Ein Institut am Forschungszentrum Jülich verwendet hier eine eigens entwickelte Datenbank die auf Basis von BibTeX und subversion (svn) als Versionskontrollsystem allen Koautoren die Pflege und Nutzung der Quellensammlung erlaubt.

gebräuchlich⁵. Häufig praktiziert ist der Dokumentaustausch per Email oder, institutsintern über gemeinsame Netzwerklaufwerke, und die Bearbeitung mit lokal vorhandener Software. Bei institutionsübergreifenden Kollaborationen finden (z. T. kommerzielle) cloudbasierte Speicher⁶ zunehmend Nutzer. Hier werden mittlerweile auch Web-basierte (z. T. ebenfalls kommerzielle) Lösungen zur Bearbeitung der Publikation eingesetzt⁷. All dies soll hier unter den Begriff der kollaborativen *Schreibumgebung* zusammengefasst werden. Gemein ist allen diesen Lösungen, dass sie eine (i. d. R. nicht nachnutzbare, da auf lokalen IDs basierende) *Affiliations-* und *Autorenidentifikation* implizieren, um Zugriffsrechte zu regeln⁸.

Ist nach mehreren internen Iterationen innerhalb der kollaborativen *Schreibumgebung* die von allen Beteiligten genehmigte finale Fassung (*Preprint*) entstanden wird diese heute i. d. R. über ein web-basiertes *Submission Interface* (SI) beim gewählten Publikationsorgan eingereicht. Hier werden zahlreiche Metadaten erfasst, z. T. solche, die die *Schreibumgebung* beim Erstellen des Dokuments in einer nicht-nachnutzbaren Form bereits vorliegen hatten (z. B. Autoren, Titel, Kurzfassung, Affiliationen). Ist die Publikation des Preprints, z. B. auf [arXiv](#) (Cornell University Library 2016), vorgesehen um Ergebnisse *schnell* publizieren zu können und sich so ggf. die Priorität zu sichern ist bereits hier eine Doppelerfassung von Metadaten meist unvermeidlich, da bereits hier zweimal eingereicht wird: einmal auf dem Preprint-Server und einmal beim Verlag. Um dies zumindest teilweise zu vereinfachen gehen einige Zeitschriften (z. B. JHEP) dazu über die eigentliche Submission über [arXiv](#) (Cornell University Library 2016) abzuwickeln, so dass ein Großteil der Metadaten von dort übernommen werden kann.

⁵ Angewandt wird ein solches Verfahren z. B. bei den großen LHC-Kollaborationen (ATLAS, CMS, Alice, LHC-B), die am CERN koordiniert werden. Es sei ferner erwähnt, dass derartige Prozesse bei der Erstellung von Software seit langem etabliert sind. Durch die immer stärkere Einbindung von Software in den Forschungsalltag ist eine Nutzung der dort verwendeten Werkzeuge auch zum Erstellen von Publikationen häufig naheliegend, da sich viele Wissenschaftler mit diesen für ihre tägliche Arbeit heute auseinandersetzen müssen.

⁶ z.B. [BSCW](#) (OrbiTeam Software GmbH & Co. KG 2016), [OwnCloud](#) (ownCloud 2016), [Dropbox](#) (Dropbox 2016), [Google Drive](#) (Google, Inc. 2016a)

⁷ z.B. Wikis, [Google Docs](#) (Google, Inc. 2016b), [Etherpad](#) (The Etherpad Foundation 2016) [Authorea](#) (Authorea 2016), s. a. (Perkel 2014).

⁸ Email-Adressen, Logins etc. Die immer stärkere Nutzung von [OAuth](#) (OAuth Community 2016) z. B. über [ORCID](#) (ORCID 2016) könnte perspektivisch die Wiederverwendbarkeit deutlich verbessern.

Je nach Publikationsorgan/-modell fallen, z. T. erhebliche, Kosten an⁹, deren Deckung zu klären und einzuleiten ist. Daneben entsteht auf Grund der Abrechnung pro Artikel, ein nicht zu verachtender Verwaltungsaufwand¹⁰. Erschwerend kommt die Unkenntnis vieler Wissenschaftler zu Möglichkeiten der Finanzierung der Publikationskosten hinzu. So sind die im Bereich Gold Open Access oft vorhandenen Publikationsfonds und ihre Modalitäten häufig unbekannt. Bei institutionsübergreifenden Publikationen wird ferner, insbesondere bei hohen Kosten, häufig eine Teilung angestrebt. Beides verursacht derzeit nicht unerheblichen Aufwand sowohl bei den Autoren als auch bei den beteiligten Verlagen, und wird im Kontext Open Access v. A. von letzteren gerne als Argument für die Beibehaltung des Subskriptionsmodells verwendet.

Nach Einreichung und je nach Fach und Organ verschieden ausgestalteten Zwischenschritten wie z. B. Review Prozessen, ggf. Ablehnung und Einreichen bei einer weiteren Zeitschrift, wird der Beitrag schließlich publiziert. Neben der hierdurch eintretenden Verlangsamung der Kommunikation ist der damit einhergehende Zeitverzug zwischen Ergebnis und Publikation oft gerade für junge Wissenschaftler ein Problem. Nicht selten dauert der Prozess länger als die Projektlaufzeit bzw. die für die Einreichung des Abschlussberichts vorgegebenen Fristen. So manches Projekt wird also "ohne Publikation", oder mit "in Print" bzw. "submitted" beendet. Vor allem letzterer Status, der ja dazu führen kann, dass finale Ergebnisse schlussendlich doch noch anderweitig, und mithin nicht nachvollziehbar, erscheinen ist nicht nur für Förderorganisationen unbefriedigend.

Bis vor einigen Jahren war der Publikationsprozess nach Übersendung einiger *Sonderdrucke* an die Autoren abgeschlossen.

Im Rahmen der Berichtspflicht kommen für Autoren heute jedoch zunehmend wichtiger werdende Aufgaben hinzu. Die Meldung an eine einrichtungsspezifische Publikationsdatenbank, Hochschulbibliographie oder

⁹ Bei Publikation in Closed Access Journals können dies Kosten für Überlänge (*page charges*) oder Farbabbildungen (*colour charges*) sein. Im Bereich Open Access fallen ggf. *Article Processing Charges* (APC) an. Zur Diskussion diverser Kosten s. a. (Bruch u. a. 2015, Abschn. 2.2) und die dortigen Referenzen und (Björk und Solomon 2014).

¹⁰ Zur komplexen Gemengelage alleine bei Open Access-Modellen s. a. (Bruch u. a. 2015), insbesondere die Abschnitte 2.2 und 2.3 sowie die dort angegebenen Referenzen.

ein Forschungs-informationssystem (FIS) wird immer häufiger verpflichtend und dient zunehmend für *leistungsorientierte Mittelvergaben* bzw. zur Evaluation von Bereichen oder Projekten, mithin also für eine ggf. beantragte Weiterfinanzierung. Daher ist ein *vollständiger* Nachweis der eigenen Leistungen im eigenen Interesse der Wissenschaftler¹¹. Es ist unmittelbar einsichtig, dass die Berichtspflicht bereits bei nur zwei beteiligten Institutionen zur Doppelerfassung führt, und dass ferner die bereits bei der Publikation erzeugten Metadaten erneut erfasst werden müssen. Neben den reinen bibliographischen Daten (die bei Zeitschriftenaufsätzen oder Büchern häufig noch über gezielte Importe übernommen werden können) sind für die Berichte *zusätzliche* Daten erforderlich wie z. B. Projekt-, Instituts-zuordnungen, Autorenidentifikation, Rollen und Fachzuordnungen¹². Beispiele für solche Anforderungen liefert der *Kerndatensatz Forschung* (Institut für Forschungsinformation und Qualitätssicherung, Fraunhofer-Institut für Angewandte Informationstechnik und Wissenschaftsrat, Geschäftsstelle 2015) der auf Empfehlungen des Wissenschaftsrats basiert (Wissenschaftsrat 2016). In bestimmten Förderschemata sind zusätzlich die Publikationen einschließlich Volltexte dem Förderer z. T. auf vorgeschriebenem Weg zu übermitteln bzw. entsprechende Open Access-Richtlinien einzuhalten (z. B. Horizon 2020 (European Commission 2016a) und OpenAIRE (European Commission 2016b).)

Um dies leisten zu können müssen für die Meldungen zu den Nachweissystemen bibliographische Verlagsdaten z. T. erheblich mit zusätzlichen Informationen, die meist nur den Autoren bekannt sind, angereichert und derzeit häufig vielfach erfasst werden.

Gefühlt wesentlich niederschwelliger als das Reporting scheinen die Anforderungen an die eigene Außendarstellung im Internet. Im allgemeinen

¹¹ Das Eintragen von Publikationen durch zentrale Stellen, z. B. die Bibliotheken, ist zwar oft gefordert aber nicht zielführend: hier kann systembedingt nur ein Bruchteil der erbrachten Leistung erfasst werden, da nur Publikationen gefunden werden können, die bereits anderweitig nachgewiesen sind. Publikationsformen jenseits des klassischen Zeitschriftenaufsatzes und der Monographie werden in diesem Verfahren i. d. R. nur schlecht oder gar nicht berücksichtigt. Je nach Fachgebiet führt dies zu einer z. T. signifikanten systematischen Verzerrung nicht nur im Nachweis sondern ebenso in den ggf. abgeleiteten Indikatoren.

¹² In der Helmholtz-Gemeinschaft reicht z. B. auch die "einfache" Affiliationszuordnung häufig nicht aus. Publikationen müssen hier auf Teilinstitusebene abgebildet werden können. Eine Projektzuordnung ist ferner neben Drittmitteln wie z. B. EU, DFG, BMBF außerdem noch auf Basis der Helmholtz-internen sog. programmorientierten Förderung (POF) notwendig, die darüber hinaus weitere Indikatoren pro Publikation erfordert (Große 2014).

Fall kann eine Einrichtung die gewünschten Darstellungen hier nur schwer vorhersehen. Ein häufig nachgefragter Dienst sind jedoch *Publikationslisten* auf den jeweiligen Webseiten in verschiedensten Granularitäten. Einmal ist hier der persönliche Lebenslauf zu ergänzen (was automatisch nur über eine eindeutige Autorenidentifikation möglich ist), daneben die Gesamtliste des Instituts und/oder der Gruppe. Häufig sind projektbasierte Webseiten zu pflegen und bei größeren Projekten (z. B. SFB, EU-Förderung) ist die jeweilige Projektleitung über die Publikationen der Unterarbeitsgruppen zu informieren. All dies geschieht auch heute häufig manuell durch erneute Erfassung in einem bestimmten Layout, also einem unstrukturierten Anzeigeformat. Die Mehrarbeit kann hier immens sein¹³.

Anders als die periodischen Berichte wird für die Außendarstellung zusätzlich eine hohe Aktualität gefordert, die praktisch nur zu leisten ist, wenn Publikationen unmittelbar nach Einreichen auch den entsprechenden Systemen gemeldet werden. Fast alle Publikationslisten stellen daher praktisch ein *Reporting in Echtzeit* dar, d. h. beruhen auf erfassten Metadaten die nach diversen z. T. komplexen Kriterien selektiert und gruppiert werden.

Vision

Aus den vorstehenden Ausführungen ergeben sich einige Folgerungen für Infrastrukturen. Im nachfolgenden soll der Versuch unternommen werden, ausgehend von derzeit bestehenden Systemen zu signifikanten Verbesserungen zu kommen.

Sammeln, Archivieren, Strukturieren - Literaturverwaltung global

"Die Bibliothek ist eine Einrichtung, die unter archivarischen, ökonomischen und synoptischen Gesichtspunkten publizierte Information für die Benutzer sammelt, ordnet und verfügbar macht."(Ewert und Umstätter 1997)

Ausgehend von dieser Definition drängt sich auf, dass die beschriebene *Quellensammlung*, die die Grundlage wissenschaftlichen Arbeitens ist, zumindest eine große Schnittmenge mit den klassischen Aufgaben von

¹³ Dem Autor wurde hier ein "Institutsworkflow" geschildert der nicht weniger als vier unabhängige Erfassungen der gleichen Publikation und das Speichern von drei unabhängigen, verschieden benannten PDF-Kopien des gleichen Beitrags im Institut umfasst. Nicht eingerechnet die erneuten Erfassungen für notwendige Projektberichte, da die unstrukturierten Daten hier nicht nachnutzbar sind.

Bibliotheken hat. Zieht man ferner in Betracht, dass die überwiegende Mehrzahl von relevanten Publikationen in diesem Kontext textueller Natur sind und heute elektronisch zur Verfügung stehen (i. d. R. sogar *born digital*) liegt es nahe, Systeme genauer in den Blick zu nehmen, die originär geschaffen wurden um Literatur elektronisch zu archivieren und verfügbar zu machen¹⁴. Fasst man den Aufgabenbereich des, immer noch häufig als *Dissertationsserver* missbrauchten, institutionellen Repositoriums ein klein wenig weiter und ergänzt dieses System um private Bereiche, die das Sammeln von Literatur in geschlossenen Gruppen ermöglichen, kann man, leistungsfähige Schnittstellen vorausgesetzt, bereits zahlreiche Anforderungen erfüllen, die sich aus der geschilderten Sammlungsproblematik ergeben. Schließlich wurden diese Systeme geschaffen um genau dieses zu leisten: Literatur zu sammeln, zu erschließen, und dauerhaft zugänglich zu machen. Der Aspekt von privaten, von außen nicht zugänglichen Bereichen ist hier deshalb relevant, da zwar die Wissenschaft von Austausch lebt, zugleich aber kompetitiv aufgestellt ist.

In der Struktur der JOIN²-Repositorien (JOIN² 2016) ist daher der private Bereich auf Instituts- bzw. Gruppenebene eines der grundlegenden Designprinzipien (Wagner 2013, Wagner (2016)). So wird jedes eingetragene Dokument zunächst nur in diesem Bereich übernommen und erst danach ggf. weiter freigegeben. Dies gilt sowohl für die bibliographische Beschreibung als auch für ggf. vorhandene Volltexte. Grundsätzlich sollte sich jede beliebige Repository-Software so konfigurieren lassen, dass entsprechende Bereiche vorgesehen sind.

Einzig limitierender Faktor mag hier häufig die interne Verwendung des Dublin Core (Dublin Core Metadata Initiative 2016) als Datenmodell darstellen, da sich dieser nur sehr bedingt eignet um z. B. Zeitschriftenreferenzen korrekt strukturiert abzubilden und so bibliographischen Exportformaten (BibTeX, RIS etc.) sinnvoll verfügbar zu machen.

Kollaborative Texterstellung

Hat das lokale Repository die relevanten Funktionen wie private Bereiche und Exportschnittstellen für bibliographische Daten lassen sich dieses bereits

¹⁴ Für den Bereich der wichtiger werdenden Datensammlungen greifen analoge Überlegungen für die jeweiligen Datenrepositorien und Datenzentren.

mit der in der Einleitung angesprochenen *Schreibumgebung* verknüpfen. Die einfachste Art sind strukturierte Exporte in die jeweils verwendeten Formatierungswerkzeuge (z. B. BibTeX oder EndNote/RIS). Das ist sicherlich nicht die komfortabelste, aber die heute sofort verfügbare Möglichkeit der Einbindung des Repositoriums in eine *Schreibumgebung 1.0*. Normiert man die Schnittstelle inkl. Login-Prozedur wäre über entsprechende Web-Schnittstellen eine komplette Integration leicht möglich.¹⁵ An den am Erstellen der Publikation beteiligten Prozessen ändert sich hier praktisch nichts, lediglich der Ort der Sammlung würde verlagert. In Hinblick auf die von Arbeitsgruppen gepflegten eigenen Lösungen ließe sich hier zusammen mit den Bibliotheken eine nachhaltige Lösung realisieren.

Derzeit existieren einige Ansätze um die heute verwendeten lokalen Bearbeitungswerkzeuge ebenfalls web-basiert zu ersetzen (Perkel 2014). Dies hat v. a. bei verteilten Kollaborationen einige Vorteile. Sollten solche Lösungen globale Autoren-IDs (z. B. [ORCID](#) (ORCID 2016)) bereits in den Login-Routinen ermöglichen, was z. B. via [OAuth](#) (OAuth Community 2016) mit [ORCID](#) (ORCID 2016) als Identity Provider technisch leicht möglich ist, ließen sich aus den Loginkennungen der beteiligten Personen bereits hinreichend gute Metadaten für die Autoren beim späteren Einreichungsprozess gewinnen. Da ferner für diese Systeme ebenfalls Metadaten erforderlich sind (z. B. Titel oder Kurzfassung) können auch diese für den folgenden Einreichungsprozess nachgenutzt werden und eine erneute Erfassung entfielen. Dies sichert neben der Arbeitersparnis eine automatische Konsistenz der Daten.

Da einige dieser Schreibumgebungen bereits heute Zitationen über DOI (via CrossRef) einbetten (z. B. [Authorea](#) (Authorea 2016)) ist ein analoger Service der statt CrossRef auf den lokalen privaten Sammlungen eines Repositoriums basiert, und so weitere Dokumentarten (Bücher, Qualifikationsschriften, Vorträge, Daten etc.) erschließt, kein technisches Problem. Repositorien müssen hierzu lediglich den strukturierten Export ihrer Metadaten über eine Web-Schnittstelle erlauben, die sich extern authentifiziert ansprechen lässt. Verstärkt zu adressieren sind in bei Web-basierten Lösungen, vor allem bei

¹⁵ So erlauben [JOIN²](#)-Repositorien (JOIN² 2016) bereits den direkten, automatischen Download strukturierter Formate aus manuell erstellten Teilsammlungen, sog. *Körben*. Technisch ließe sich dies über einen einfachen http-get auch nahtlos einbinden.

Cloud-Services, datenschutzrechtliche Fragestellungen. Auch kann *Vertrauen in den Dienst* problematisch sein. Es sollen schließlich keine unfertige Arbeitsversionen der eigenen Publikationen kursieren, bevor die Autoren diese freigeben haben, und die Autoren sollen alle Rechte an ihrer Publikation behalten. Dies kann für eine On-Site Implementation sprechen, und sollte sich in entsprechenden Policies widerspiegeln.

Generell scheint es an dieser Stelle erstrebenswert offene Schnittstellen und Formate zu spezifizieren, um eine breite Akzeptanz zu ermöglichen und ein Vendor Lock-In zu vermeiden. Um eine einfache Adaption zu ermöglichen scheint es geboten, sich stark an derzeit bereits verfügbaren Standards zu orientieren. Kann z. B. eine Schreibumgebung aus einer DOI automatisch eine entsprechende Referenz erstellen, so scheint es sinnvoll einen Export anzubieten, der sich weitgehend an den CrossRef Metadaten schemata (Publishers International Linking Association, Inc. 2016a) orientiert. In der Schreibumgebung wäre dann lediglich eine URL zu ersetzen.

Publizieren: mehr als "Einreichen"

Ist das Dokument geschrieben und liegt die von den Autoren zur Einreichung vorgesehene Fassung vor, steht die Publikation an. Betrachtet man die in der Einleitung beschriebenen *weiteren Verwendungen* genauer, wird schnell ersichtlich, dass es sich hier um den zentralen Punkt des Prozesses handelt, der außerdem das höchste Optimierungspotential bietet:

- Es muss mehr als ein Zielsystem angesprochen werden, und dies ggf. zu verschiedenen Zeitpunkten (z. B. Preprint Server, Repository, Bibliographie/FIS, Verlag)
- Der Eingabeaufwand soll minimiert werden
- Zahlreiche Details sind von den Autoren abhängig und oft nur diesen bekannt (häufig auch nur zum Zeitpunkt des Entstehens der Publikation)
- Die bibliographischen Daten werden vielfach und normiert benötigt, häufig aber erst von Verlagen vervollständigt
- Z. T. werden einrichtungsspezifische Daten benötigt
- Es gelten Randbedingungen aus den Förderrichtlinien
- Datenschutz

Es scheint daher sinnvoll das *Submission Interface* (SI) genauer zu untersuchen. Hierbei soll vom gewünschten Ergebnis ausgegangen werden, und nicht von derzeit realisierten Verfahren, da sich diese für die beschriebenen Anforderungen nur sehr bedingt eignen und ein Festhalten an diesen viele wünschenswerte Verbesserungen verhindert.

Das Submission Interface (SI)

Anforderungen und Potentiale

Diskussionen mit Wissenschaftlern zeigen den Wunsch nach einer Standardisierung des *SI*. Selbst wenn eine reine Verlagspublikation angestrebt ist, taucht der Wunsch nach einer verlagsübergreifend einheitlichen Maske auf, da niemand z. B. nur in einer einzigen Zeitschrift publiziert und der Einreichungsprozess u. A. durch die verschiedenen Formulare als sehr zeitaufwendig wahrgenommen wird.

Eine Harmonisierung scheint also selbst dann erstrebenswert, wenn keine weiteren Verwendungen angedacht wären. Da eine neue Struktur für alle Publikationen und alle Zielsysteme nutzbar sein soll, muss sie mehr Daten aufnehmen können, als eine heutiges verlagsseitiges Submission Form.

Aus Sicht der Wissenschaftler ist es wünschenswert, dass die notwendigen Eingaben auf ein Minimum reduziert und ggf. notwendige Normierungen transparent erfolgen. Es sollen also nur die für das jeweilige Medium relevanten Daten abgefragt werden. Die Feldauswahl und ggf. die möglichen Inhalte lassen sich über entsprechende Definitionen steuern. Darüber hinaus ist eine maximale Nachnutzung von bereits im Schreibprozess anfallenden Metadaten sinnvoll, was eine Anbindung an die Schreibumgebung wünschenswert macht. So können z. B. Felder wie Titel, Kurzfassung, Volltext automatisch befüllt werden. Erlaubt die *Schreibumgebung* bereits eine Autorenidentifikation so können auch diese Daten übergeben werden. Ein wesentlicher Aspekt an dieser Stelle ist, dass Daten, die in der *Schreibumgebung* nicht strukturiert vorliegen, weiterhin manuell eingegeben werden können und müssen. Dies ist dann nicht besser, *aber auch nicht schlechter* als im derzeitigen Verfahren.

Werden verlagsseitig weitere Daten erforderlich (z. B. Publikationsorgan, Lizenz, Einräumung von Nutzungsrechten) können die dafür möglichen Werte über eine Web-Schnittstelle der Maske verfügbar gemacht und dem Wissenschaftler zur Auswahl angeboten werden.

Gleiches gilt für einrichtungslokale Daten wie z. B. institutionelle Zugehörigkeiten (z. B. Teilinstitute oder Gruppen) oder Projektzuordnungen (lokale Förderschemata, Zuordnung zu Großgeräten etc.). Auch hier können normierte Webservices gültige Werte der Maske zur Verfügung stellen. Je nach Ausgestaltung der *Schreibumgebung* sind diese Daten aber bereits zur nötigen Rechteverwaltung dort vorhanden. Es scheint auch die Anbindung eines *lokalen* Helpdesks sinnvoll, um z. B. auftauchende Fragen zu Finanzierungsmöglichkeiten (die heute häufig erst *nach* der Publikation auftauchen) bereits im Vorfeld klären zu können.

Auch Fördervorgaben könnten über ein solches Interface transparenter transportiert werden. Z. B. hätte auf diese Weise das *SI* Autoren auf eine vorhandene *Special Clause 39*¹⁶ der FP7-Grants hinweisen können, eine Frage

¹⁶ "[...]In addition to Article II.30.4, beneficiaries shall deposit an electronic copy of the published version or the final manuscript accepted for publication of a scientific publication relating to foreground published before or after the final report in an institutional or subject-based repository at the moment of publication.[...]" (European Commission 2012)

die aktuell häufig erst im Kontext der Abschlussberichte auftaucht¹⁷. Die Maske könnte ebenfalls automatisch auf entsprechende Publikationsfonds verweisen (im Beispiel ein Refunding durch die EU, oder auch ggf. verfügbare Fonds innerhalb eines DFG-Projekts oder der Einrichtung), wenn der Verlag APCs oder andere Gebühren geltend macht¹⁸. Ein entsprechendes Budget könnte bereits an dieser Stelle blockiert werden¹⁹. Ist das zu belastende Konto bereits im Einreichungsprozess definiert, sind die Kosten für die Einrichtung und die Förderer kalkulierbar und es eröffnet sich die Möglichkeit für entsprechende Etatverteilungsmodelle (analog zu z. B. (Trott 2004)), so dass auch *Orchideenfächer* keine Sorge haben müssen ihre Publikationen zum Jahresende nicht mehr finanzieren zu können. Gleichzeitig ermöglicht dies den Wissenschaftlern den einfachen Umgang mit diesen neuen Instrumenten.

Auch Verlage würden von einer solchen Harmonisierung des *SI* stark profitieren. Systemimmanent wäre die Kostendeckung bereits *vor* Einreichung der Publikation final geklärt. Da die nötigen Fakturierungsdaten aus den jeweiligen relevanten Verwaltungssystemen zur Verfügung gestellt werden und die Deckung garantiert ist, kann von einer stets korrekten Rechnungslegung und einer mithin zügigen Bearbeitung ausgegangen werden. Die Teilung von Publikationsgebühren zwischen beteiligten Einrichtungen würde ebenfalls bereits vor Einreichung festgelegt und deren Kostendeckung geprüft.

Da ggf. nötige Kommunikation hier mit zentralen Einrichtungen stattfindet verbliebe man bei dem bisher im Bereich *Subskription* vorhandenen Modell weniger Experten als Partner bei der Abwicklung. Eine Umstellung von Einzelrechnungen auf Sammelabrechnungen wie auch in (Bruch u. a. 2015) angeregt wäre damit deutlich einfacher realisierbar. Der Verwaltungsaufwand sowohl auf Verlags- als auch auf Einrichtungsseite würde sich signifikant reduzieren.

¹⁷ Erfreulicherweise führt dies zumindest am DESY derzeit einer retrospektiv ansteigenden Open Access-Quote aus diesen Programmen, da über diese Vorgabe zumindest vom Zweitveröffentlichungsrecht bzw. dem Grünen Weg gebraucht gemacht wird. Allerdings hatte sich der Förderer sicherlich eine zeitnähere Open Access-Publikation erhofft.

¹⁸ Die in (Bruch u. a. 2015) geforderte Einführung einer zentralen Stelle zur Verwaltung von Publikationsgebühren setzt ein derartiges Verfahren praktisch voraus, da sich diese zentralen Fonds nach derzeitigem Stand aus verschiedenen, z. T. zweckgebundenen, Quellen speisen werden.

¹⁹ Es *kann* und *soll* nicht von Wissenschaftlern erwartet werden, dass sie sich gleichzeitig zu Verwaltungsfachleuten entwickeln, die diese gesamte komplexe Gemengelage im Detail überblicken, insbesondere da die hierfür nötigen Verwaltungsstrukturen bereits alle existieren.

Anders als in bisherigen Prozedere, bei denen *nach* der Publikation aus i. d. R. unnormierten Freitexten mit Affiliationsangaben oder gar der Danksagung die Einrichtungen und Förderer mit viel Aufwand *ihre* Publikationen ermitteln (wobei systemimmanent nur diejenigen ermittelt werden können, die in anderen Datenbanken nachgewiesen sind) stünde im vorgeschlagenen Verfahren die Affiliationszuordnung bereits bei Einreichung fest. Die Forderung an Verlage diese Daten strukturiert zu liefern, ist damit obsolet. Das *SI* würde nicht nur das Verlagssystem mit Daten versorgen, sondern die Daten um die vom Verlag vergebenen Werte (z. B. DOI, Zeitschrift, Volume, Jahr, Seiten etc.) ergänzt direkt der Bibliographie und einem ggf. angeschlossenen FIS verfügbar machen. Gleiches gilt für das Einpflegen des Volltextes in die institutionellen Repositorien²⁰. Es ist hervorzuheben, dass in diesem Verfahren verlagsseitig *keine* zusätzlichen Arbeiten anfallen. Reicht momentan der eigene Webserver die Formulardaten an die Verlagsdatenbank weiter, würde dies künftig über ein normiertes Format ggf. von einem anderen Host aus erfolgen.

Vom Monolith zum Web-Service

Da es beim Einsatz von Web-Technologien nicht relevant ist, *wo* das *SI* realisiert wird, ist hier die Definition von offenen, normierten Datenformaten entscheidend. Die Übertragung kann dann über heutige Web-Technologien erfolgen. Da zahlreiche Werte einrichtungsinterne vorliegen, scheint eine Ansiedlung des *SI* an den Forschungseinrichtungen selbst sinnvoll. Dies kommt u. U. auch den Autoren entgegen, die der eigenen Einrichtung häufig mehr vertrauen als einem fremden System.

Auch ein entsprechendes Branding bzw. die Anpassung der Feldbezeichnungen an lokale Gegebenheiten (z. B. gliedern sich manche Institutionen in Institute, andere in Gruppen oder Departments) scheint auf diese Weise einfacher und käme den Benutzern entgegen.

Da aber für eine Publikation im Allgemeinen einrichtungsinterne Daten *verschiedener* Institutionen relevant sind²¹ und daher die Befüllung der

²⁰ Da die nötigen Schnittstellen hierbei aus Sicht des *SI* immer 1-n ausgelegt sind ist auch der Fall abgedeckt, dass mehrere Institutionen gemeinsam publizieren oder das FIS-Modul, das sich um Publikationen kümmert, noch nicht das Repository darstellt.

²¹ Man denke an die Kollaborationen von Autoren verschiedener Einrichtungen.

jeweiligen Vorschlagslisten über Webschnittstellen erfolgen muss, ist eine Ansiedlung an der eigenen Einrichtung nicht strikt notwendig. Bei genauer Betrachtung kapselt das *SI* aus systemischen Gründen den Prozess an sich in einem strikt privaten Bereich. So kann man sich durchaus vorstellen, dass zur Einreichung das *SI* irgendeiner Einrichtung verwendet wird, selbst einer, die an der konkreten Publikation nicht beteiligt ist²². Selbst eine Auslegung als verteiltes System, welches den zu verwendenden Host zufällig auswählt ist denkbar. Neben entsprechenden Richtlinien ist dies ein wirksames Mittel, datenschutzrechtlichen Vorbehalten entgegenzutreten. Per se haben zwar Einrichtungen oder Förderer am Procedere des Einreichungsvorgangs kein Interesse, es könnte aber der Eindruck entstehen, dass sich z. B. Ablehnungen von Publikationen negativ auf Bewertungen auswirken, wenn diese der Einrichtung bekannt sind²³.

In allen diesen Fällen sammelt das *SI* zunächst an zentraler Stelle Metadaten, sowie die für die Submission vorgesehenen Volltexte und legt diese in einem temporären Speicher ab²⁴.

An dieser Stelle sollten bereits die für spätere Verwendungen nötigen zusätzlichen Metadaten wie Projektzuordnungen²⁵ oder Affiliationen²⁶ erfasst werden, da diese i. d. R. nur den Autoren bekannt sind.

Es ist ferner zu berücksichtigen, dass Publikationen im *SI* optional noch Genehmigungsprozessen innerhalb einer Institution unterliegen können, bevor diese zur Publikation freigegeben werden. Z. B. müssen in den meisten

²² Möchte man trotzdem an einem einrichtungsspezifischen Branding oder Wording festhalten müssen diese Parameter über entsprechende Webschnittstellen transportiert werden.

²³ Allerdings wäre das natürlich nicht sachgerecht, da ein üblicher Ablehnungsgrund ist, dass der Herausgeber einen Artikel als zu weit vom Kernbereich der Zeitschrift entfernt ansieht (*out of scope*). Hier wird also aus eher formalen denn aus wissenschaftlichen Gründen abgelehnt. Überhaupt ist die Bewertung der Qualität von Aufsätzen auf Basis der Zeitschriften in denen diese erscheinen fragwürdig.

²⁴ Einige Repository-Lösungen bieten hierfür bereits fertige Lösungen z. B. der HoldingPen in invenio (Invenio Development Team 2016).

²⁵ Eine mögliche Quelle wäre z. B. FundRef (Publishers International Linking Association, Inc. 2016b). Allerdings müsste hier die Granularität noch deutlich ausgebaut und die Lieferung der relevanten Grants durch die Förderer sichergestellt werden. Da hier allerdings immer Web-Schnittstellen relevant sind kann hier auch eine Anbindung an interne Projektdatenbanken der an einer Publikation beteiligten Institutionen erfolgen. (Große 2014)

²⁶ Derzeit fehlen noch hinreichend granulare und frei zugängliche Datenbanken ausreichender Qualität, um diesen Teil abzubilden. Da sich Affiliationen, insbesondere auf Teilinstitutsebene, häufig ändern scheint es, bis automatische normierte Datenlieferungen an einen zentralen Hub erfolgen können sinnvoller, diese Werte aus den internen Daten der beteiligten Einrichtungen zu beziehen. Da auch hier immer Web-Schnittstellen relevant sind ist ein zentraler Host aber wie bei Projekten *keine* Voraussetzung. (Große 2014)

Einrichtungen der Helmholtz-Gemeinschaft die jeweiligen Instituts-/Gruppenleiter als Vorgesetzte die Publikation an sich genehmigen²⁷.

Um ein solches System zu realisieren sollte ein Baukastenprinzip angestrebt werden. Liegen Web-Technologien zu Grunde ist dies vergleichsweise einfach realisierbar. So ist es möglich nach und nach neue Funktionen zu integrieren und sich flexibel an neue Aufgaben anzupassen.

So wäre aus heutiger Sicht zunächst eine Abstraktion der vorhandenen verlagsspezifischen Schnittstellen ein Fortschritt. Dies würde zunächst lediglich die Definition des Datenaustauschformats sowie die Auslegung von Login-Prozeduren und der verwendeten Protokolle erfordern. In einem ersten Schritt würde sich also, bis auf die technische Basis nicht viel ändern.

In weiteren Ausbaustufen können zusätzliche Module eingebaut werden, um z. B. eine Autorenidentifikation via ORCID (ORCID 2016), die Anbindung von FundRef (Publishers International Linking Association, Inc. 2016b) und weiterer Normdatensysteme zu realisieren.

Da aus jetziger Sicht weder ORCID (ORCID 2016) noch FundRef (Publishers International Linking Association, Inc. 2016b) geeignet sind die Anforderungen zu erfüllen, schafft dies auch die Möglichkeit nach und nach Verbesserungen an den angeschlossenen Systemen vorzunehmen. Die Herausforderungen liegen hier eher in der Datenqualität als im technischen Aufsetzen entsprechender webprotokoll-basierter Schnittstellen. Es sei hervorgehoben, dass in den derzeit diskutierten Modellen solche Anbindungen an die Verlage als Forderungen herangetragen werden, so dass jeder Verlag solche Schnittstellen in Eigenregie erstellen muss. Es scheint daher auch im verlagsinteresse, wenn hier eine abstrakte, allgemeine Lösung geschaffen wird, die als OpenSource nachgenutzt werden kann.

Sollten sich, was zu erwarten ist, mittelfristig web-basierte Systeme zum Erstellen von Publikationen etablieren, kann als nächster Schritt eine Schnittstelle zu diesen Systemen geschaffen werden, die in zunächst nur die Übernahme der Volltexte erlaubt und sukzessive um die Übergabe weiterer Metadaten je nach

²⁷ Gründe hierfür sind z. B. Wahrung von Patentmöglichkeiten, Einhaltung guter wissenschaftlicher Praxis, Haftung etc. Man kann diese Prozesse natürlich auch außerhalb des *SI* abbilden. Sie weisen aber strukturell hohe Ähnlichkeiten mit Reviews und der Lieferung von Referee Reports auf.

Fähigkeit des angeschlossenen Systems erweitert wird (Titel, Kurzfassung, Autoren, Affiliationen etc.). Auf Seiten des *SI* sollte vorgesehen sein, dass fehlende Daten manuell ergänzt werden bis sie dem zu definierenden Austauschformat entsprechen. Die Anzahl der Pflichtfelder ist hier, um eine einfache Anbindung zu ermöglichen, weitestgehend zu reduzieren.

Bereits heute steht mit SWORD (swordapp.org 2016) ein Protokoll zur Verfügung das nach dem *Push*-Prinzip wesentliche Teile für die Übertragung eines Volltexts aus einer Schreibumgebung an ein *SI* zur Verfügung stellt. Auch zahlreiche Versionskontrollsysteme wie *git* stellen Möglichkeiten zur Übertragung zur Verfügung. Definiert man die Metadaten in einer entsprechenden Datei (Filialdateien oder sidecar files (Wikipedia-Autoren 2015)) ist auch eine solche Anbindung mit heutigen Mitteln einfach möglich, sofern das Datenschema und -format geklärt sind.

Eine alternative Implementation würde statt des *Push* aus der Schreibumgebung ein *Pull* von diesem System realisieren, also einen *Datenimport* in das *SI*. Dies ist analog dem *harvesten* eines Einzeldokuments, so dass OAI-PMH hier bereits heute die nötigen Funktionen realisiert. Ein solches Verfahren setzt allerdings web-basierte Schreibumgebungen voraus, während man bei Push-Schnittstellen auch eigenständige Clients realisieren kann, die auf lokalen Dateien arbeiten. Solche eigenständigen Clients könnten ggf. auch die Umsetzung lokaler Autorenidentifikationssysteme (Logins, Email-Adressen) in globale Systeme wie ORCID (ORCID 2016)) übernehmen. Die Ausgestaltung und der Umfang der Daten läge hier auf Seiten des Clients.

Soll die Publikation schlussendlich eingereicht werden leitet das *SI* die dort gesammelten Metadaten weiter. Im Falle der Publikation auf einem Repositorium entspricht dies einem einfachen Deposit. Zusätzlich soll der Autor die Möglichkeit haben weitere Repositorien (z. B. arXiv) zur Einreichung auszuwählen. Diese bekämen ebenfalls den normierten Metadatensatz und könnten ihn übernehmen. Ist eine Publikation in einem Verlag angestrebt, soll das *SI* hier das jeweilige Organ zur Auswahl anbieten, und über eine Web-Schnittstelle die jeweils geltenden Bedingungen des Verlags für den Autor darstellen. Dies versetzt den Wissenschaftler in die Lage an dieser Stelle eine informierte Entscheidung zu treffen und im Sinne des Marktgedankens Angebote abzuwägen und auszuwählen.

Ist die Publikation kostenpflichtig würde dies in einer ersten Stufe lediglich mit allen anfallenden Kosten aufgelistet und angezeigt. Mit weitergehender Automatisierung könnte hier die Deckung über lokale Fonds geprüft werden. Im einfachsten Falle würde lediglich auf den entsprechenden Helpdesk verwiesen, und die Kostenübernahme manuell durch den Autor bestätigt. Sodann wird das Dokument dem Verlag übermittelt, der dieses ggf. in den jeweils üblichen Prozess übernimmt. Entsprechende Antworten können den Autoren ebenfalls über das *SI* verfügbar gemacht werden. Hierzu müssen diese nicht in das *SI* übertragen werden, sondern können online vom jeweiligen Verlag abgerufen werden. Dies stellt sicher, dass das *SI* keine Bewertungen des jeweiligen Aufsatzes kennt, was aus Datenschutzgründen sinnvoll sein kann. Soll nach einer endgültigen Ablehnung die Publikation ggf. nach einer Überarbeitung bei einem anderen Organ eingereicht werden, kann dies ohne Neueingaben lediglich durch eine neue Auswahl geschehen.

Damit Überarbeitungen einfach möglich sind ist eine möglichst gute Anbindung an die Schreibumgebung hilfreich. Im einfachsten Fall erlaubt diese nur Austausch des Volltextes durch eine neue Version, aber auch eine komplette Versionierung inkl. der Aktualisierung der Metadaten ist denkbar. Realisiert mit eindeutigen Identifiern für die Publikation und über offene Webprotokolle separiert dies auch die Schreibumgebung vom *SI*.

Bei Annahme der Publikation übernimmt das *SI* die automatische Speicherung der endgültigen Fassung die ggf. notwendige Aktualisierung der bereits im initialen Prozess angestoßenen Deposits auf Fachrepositorien durch Übermittlung des finalen, normierten Metadatensatzes sowie auf dem jeweiligen Repositorium. Fallen für die Publikation Kosten an können die jeweiligen Modalitäten direkt mit korrekten Daten an den jeweiligen Verlag weitergereicht werden, so dass je nach Verlagsprozess eine entsprechende Rechnungslegung mit korrekten Daten erfolgt.

Durch Übernahme in das eigene Repositorium steht die eigene Publikation ferner für neue Arbeiten zur Zitation zur Verfügung. Dass das *SI* auch die Einreichungsformen des Dokuments bis hin zum *post print* kennt, lässt sich auch der *Grüne Weg* für den Autor einfach realisieren. Sind mit den Verlagen z. B. im Kontext der *Allianz-Lizenzen* zusätzliche Open Access-

Vereinbarungen getroffen können diese an dieser Stelle, verbunden mit der damit einher gehenden Steigerung der Visibilität, bereits realisiert werden.

Da das *SI* bereits bei Einreichung, also *vor* dem Entstehen der finalen Publikation, die entsprechenden Daten erfasst ist eine Vollständigkeit prinzipiell möglich. Ferner ist an dieser Stelle bereits die Affiliationszuordnung per Prozess geklärt, eine Herausforderung die in fast allen anderen Kontexten derzeit als unmöglich gilt²⁸.

Ist das Repositorium der Einrichtung nicht gleichzeitig die jeweilige Publikationsdatenbank bzw. der für Publikationsdaten zuständige Teil eines FIS(Große 2014) können die normierten Daten auch hier direkt deponiert werden. Dies stellt größt mögliche Konsistenz und Vollständigkeit der Meldungen im Sinne der Wissenschaft sicher²⁹. Über die Normierung der Depositschnittstellen und Metadatenformate können alle beteiligten Institutionen von einer Stelle aus versorgt werden. Als gewünschter Nebeneffekt sorgt die automatische Anbindung an die jeweiligen institutionellen Repositorien für eine entsprechende Indexierung durch Suchdienste wie z. B. BASE (Universität Bielefeld 2016) und erhöht so die Sichtbarkeit.

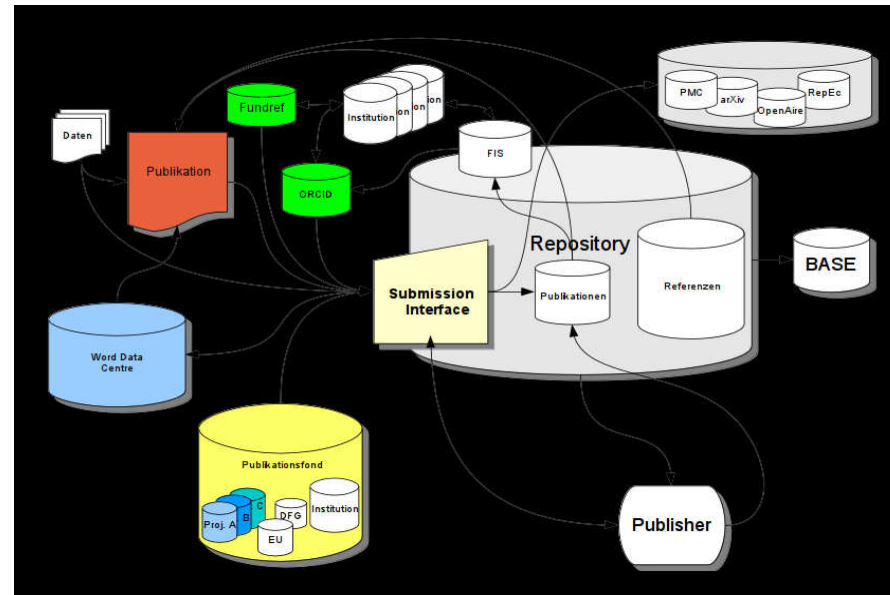
Heute ist nach der Publikation vor dem Bericht, d. h. es müssen, je nach Bericht z. T. extrem aufwendig, alle Publikationen einer Einrichtung, eines Projekts, einer Person zusammengestellt, und dafür heute i. d. R. zunächst erfasst werden (Große 2014). Diese aufwendige Sammlung entfällt im vorgeschlagenen Modell, da das *SI* automatisch die entsprechenden Systeme beliefern kann. Projektzuordnungen müssen nicht aus textuellen Danksagungen intellektuell ermittelt werden, sondern liegen bereits ab Einreichung strukturiert vor. Dies ermöglicht den beteiligten Wissenschaftlern ein korrektes Reporting auch wenn ein Projekt mehrere Jahre läuft. Ebenso ermöglicht sie den Förderern mit Hilfe der normierten Projektzuordnungen Publikationen zuzuordnen selbst wenn diese erst nach dem Projektende final vorliegen. Verknüpft man daher Meldung mit der Einreichung spart dies

²⁸ Vgl. den wesentlich aufwendigeren Prozess(Koch u. a. 2016), der ohne eine solche Schnittstelle notwendig ist insbesondere die diskutierte Problematik der dort relevanten *retrospektiven* Affiliationszuordnung.

²⁹ Ist die Schreibumgebung das *übliche* Werkzeug zum Erstellen von Dokumenten können hier auch Lehrmaterialien wie z. B. Vorlesungsskripte direkt verfügbar gemacht werden. Es findet lediglich keine Übergabe an einen Verlag statt, sondern die Publikation über das lokale Repository.

einerseits die erneute Erfassung und sichert andererseits eine zeitnahe Verfügbarkeit konsistenter Daten an den gewünschten Stellen. Dies gilt auch für die eingangs erwähnte Außendarstellung.

Eine schematische Darstellung der jeweiligen Datenflüsse könnte wie folgt aussehen:



Möglicher Datenfluss und Verbindungen von Systemen, die im Kontext einer wissenschaftlichen Publikation relevant sind.

Schlussfolgerungen

In der Vergangenheit dienten bibliographische Daten im Wesentlichen zum Austausch von Informationen in einer kleinen Fachöffentlichkeit. Sie dienten dazu, dass Wissenschaftler relevante Publikationen in ihrem eigenen Forschungsfeld leicht finden konnten und die dortigen Ergebnisse in ihren eigenen Arbeiten verwenden zitieren konnten. Hierfür sind vergleichsweise wenige Daten erforderlich, zumal je nach Feld die Anzahl relevanter

Publikationsorgane übersichtlich war und die Zuordnung immer intellektuell erfolgte.

Neben diese klassische Verwendung bibliographischer Daten treten zunehmend weitere Anwendungen. So hat heute auch die allgemeine Öffentlichkeit Interesse an den Ergebnissen der Forschung. Durch den gestiegenen Wettbewerb in der Wissenschaft wird ferner eine Darstellung der eigenen Leistung immer wichtiger.

Dies gilt sowohl für den einzelnen Wissenschaftler als auch für entsprechende wissenschaftliche Einrichtungen, die sich alle im Wettbewerb um endliche Ressourcen befinden. Da hier außerdem überwiegend automatische Verfahren zum Einsatz kommen sollen, steigen die Anforderungen an die vorhandenen Metadaten und deren Qualität deutlich.

Vor diesem Hintergrund scheint es geboten den Prozess der Publikation zu überdenken. Zusammen mit den heute im Web verfügbaren Technologien können hier neue Überlegungen bisherige Verfahren ersetzen und einen entsprechenden Mehrwert für alle beteiligten Schaffen. Insbesondere die heute verwendeten *SI*, die lediglich die Systeme der Verlagshäuser beliefern, erscheinen nicht mehr zeitgemäß.

Es wird daher vorgeschlagen, das eigentliche *Submission Interface* abstrakter zu gestalten und auf offenen Protokollen und Datenformaten basieren zu lassen. Dies würde erlauben, Publikationen bereits beim Einreichen mit den notwendigen Metadaten zu versehen, die es erlauben die verschiedenen Anforderungen zu erfüllen, die heute an bibliographische Daten gestellt werden (Institut für Forschungsinformation und Qualitätssicherung, Fraunhofer-Institut für Angewandte Informationstechnik und Wissenschaftsrat, Geschäftsstelle 2015). Damit können die aktuell noch notwendigen Mehrfacherfassungen entfallen. Die gemachten Vorschläge hierzu basieren ausschließlich auf heute bereits vorhandenen Technologien. Der vorgeschlagene Prozess erlaubt eine sukzessive Überführung des bestehenden Systems, wobei mit möglichst geringen Mitteln Vorteile für alle Beteiligten (Wissenschaftler, Verlage, Einrichtungen, Förderorganisationen) realisierbar werden.

Bibliographie

Authorea. 2016. Authorea. <https://www.authorea.com>.

Björk, Bo-Christer und David Solomon. 2014. Developing an Effective Market for Open Access Article Processing Charges. <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Open-access/Guides/WTP054773.htm>.

Bruch, Christoph, Gernot Deinzer, Kai Geschuhn, Petra Hätscher, Kristine Hillenkötter, Ulrike Kreß, Heinz Pampel, u. a. 2015. Positionen zur Schaffung eines wissenschaftsadäquaten Open-Access Publikationsmarktes : Positionspapier der ad-hoc-AG Open-Access-Gold im Rahmen der Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen. Potsdam: Ad-hoc-Arbeitsgruppe Open-Access-Gold der Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen. doi:10.2312/allianzoa.008,.

Cornell University Library. 2016. ArXiv.org. <http://arxiv.org>.

Dropbox. 2016. Dropbox. <https://dropbox.com>.

Dublin Core Metadata Initiative. 2016. Dublin core metadata element set, version 1.1. <http://dublincore.org/documents/dces/>.

European Commission. 2012. *Special clause 39*. http://ec.europa.eu/research/press/2008/pdf/annex_1_new_clau.

---. 2016a. Horizon 2020 - The EU Framework Programme for Research and Innovation. <https://ec.europa.eu/programmes/horizon2020>.

---. 2016b. OpenAIRE. <http://openaire.eu>.

Ewert, Gisela und Walther Umstätter. 1997. *Lehrbuch der Bibliotheksverwaltung*. Hg. von Wilhelm Krabbe und Wilhelm Martin Luther. [4. Aufl.]. Stuttgart: Hiersemann.

Google, Inc. 2016a. Google Drive. <https://www.google.com/drive/about>.

---. 2016b. Google Docs. <https://www.google.com/docs/about>.

Große, Katrin. 2014. Forschungsevaluation aus dem Repositorium. In: OpenAccess Tage 2014, Köln (Germany), 8 Sep 2014 - 9 Sep 2014, 8. September. <http://repository.gsi.de/record/67551>.

Institut für Forschungsinformation und Qualitätssicherung, Fraunhofer-Institut für Angewandte Informationstechnik und Wissenschaftsrat, Geschäftsstelle. 2015. Spezifikation des Kerndatensatz Forschung - Version 1.0. <https://bib-pubdb1.desy.de/record/296180>.

Invenio Development Team. 2016. Invenio. <http://inveniosoftware.org>.

JOIN². 2016. JOIN² - just anOther iNvenio iNstance. <http://join2.de>.

Koch, Thorsten, Rolf Griebel, Konstanze Söllner, Jürgen Christof und Roland Bertelmann. 2016. DeepGreen - Entwicklung eines rechtssicheren Workflows zur effizienten Umsetzung der Open-Access-Komponente in den Allianz-Lizenzen für die Wissenschaft. ZIB. <http://nbn-resolving.de/urn:nbn:de:0297-zib-56799>.

OAuth Community. 2016. OAuth. <http://oauth.net>.

OrbiTeam Software GmbH & Co. KG. 2016. BSCW. <https://public.bscw.de/pub/>.

ORCID. 2016. ORCID. <http://orcid.org>.

ownCloud. 2016. OwnCloud. <https://owncloud.org/>.

Perkel, Jeffrey M. 2014. Scientific writing: The online cooperative. *Nature* 514, Nr. 7520 (Oct): 127–128. doi:10.1038/514127a,.

Publishers International Linking Association, Inc. 2016a. Crossref deposit schema. http://help.crossref.org/deposit_schema.

---. 2016b. FundRef. <http://www.crossref.org/fundingdata>.

swordapp.org. 2016. Sword. <http://swordapp.org/about>.

The Etherpad Foundation. 2016. Etherpad. <http://etherpad.org/>.

Trott, Sabine. 2004. *Entwicklung von Erwerbsprofilen für die Fächer Mathematik und Physik an der UB Ilmenau*. Bd. 133. Berliner handreichungen zur bibliothekswissenschaft. Institut für Bibliothekswissenschaft der Humboldt-Universität zu Berlin. <http://www.ib.hu-berlin.de/~kumlau/handreichungen/h133/>.

Universität Bielefeld. 2016. BASE - bielefeld academic search engine. <https://www.base-search.net>.

Wagner, Alexander. 2013. JuSER – Bibliographie, OpenAccess-Server, Literaturverwaltung... In: *Zentrale hochschulbibliothek luzern*. Luzern. <http://hdl.handle.net/2128/5470>.

---. 2016. JOIN² - What it is and what it does. In:.. Privatuniversität Witten/Herdecke. <https://bib-pubdb1.desy.de/record/296178>.

Wikipedia-Autoren. 2015. Filialdateien. *Wikipedia*. <https://de.wikipedia.org/w/index.php?title=Filialdatei&oldid=144778980>.

Wissenschaftsrat. 2016. Empfehlungen zur Spezifikation des Kerndatensatz Forschung. Wissenschaftsrat. http://www.wissenschaftsrat.de/arbeitsbereiche-arbeitsprogramm/kerndatensatz_forschung.html.

Wandel aktiv gestalten: Das Projekt INTACT- Transparente Infrastruktur für Open-Access-Publikationsgebühren

Kai Geschuhn¹, Dirk Pieper ²

¹Max Planck Digital Library¹

²Universitätsbibliothek Bielefeld

Zusammenfassung

Während sich die Open-Access-Debatte jüngst auf die Forderung nach einer großflächigen Umstellung der traditionellen Zeitschriften zuspitzt (Open Access Transformation), zeigt sich auch anhand der internationalen Aufmerksamkeit, die die illegale Schattenbibliothek „SciHub“ derzeit erlangt, wie sehr das Subskriptionsmodell für wissenschaftliche Zeitschriften unter Druck steht. Eine Umstellung ihres Geschäftsmodells von Subskription auf Open Access hätte zur Folge, dass Verlage künftig ihre Umsätze zu einem großen Teil aus Open-Access-Artikelgebühren generieren (Article Processing Charges, APC). In diesem Szenario werden Bibliotheken wesentliche Anteile ihrer Budgets nicht mehr für den Erwerb von Subskriptionen und Lizenzen verwenden, sondern für die Übernahme der APC für die Autoren ihrer Einrichtung, wodurch sie nicht nur eine völlig neue Erwerbslogik verinnerlichen müssen- vielmehr verändert sich ihre Rolle für die wissenschaftliche Informationsversorgung grundlegend.

Das seit Oktober 2015 operierende DFG-Projekt INTACT entwickelt auf drei Arbeitsbereichen eine Infrastruktur, die über internationale Initiativen und den Betrieb eines Datenservices zur Offenlegung von Publikationsgebühren die Transparenz des neuen Geschäftsmodells, die Vernetzung der Bibliotheken und der beteiligten Akteure unterstützt und somit aktiv die nachhaltige Ausgestaltung des Wandels hin zu Open Access befördert.

¹ Die vorliegende Arbeit entstand im Rahmen des Projekts INTACT unter Mitwirkung des Projektteams an der Universitätsbibliothek Bielefeld (Christoph Broschinski, Najko Jahn und Vitali Peil), der Universität Bielefeld (Christine Rimmert, Matthias Winterhager und Michael Wohlgemuth) und der Max Planck Digital Library (Ralf Schimmer, Michael Schlachter und Adriana Sikora).

Abstract

Recent open access initiatives focus on a large-scale transition of the existing corpus of scholarly journals to an open access business model. Moreover, the intense international debate on SciHub, a pirate repository of over 51,000,000 scientific academic papers, shows the growing pressure on the subscription business. Once open access becomes the default model for scientific publishing, libraries will have to adapt a new logic of acquisition when providing financial resources and services instead of licenses and subscriptions in order to support the authors of their institutions.

The project INTACT, which is funded by the German Research Society (DFG), aims at establishing transparent and efficient procedures to manage article processing charges (APC) for open access publications. The project is based on the understanding that the transition from a subscription based market towards open access can only be achieved if higher education and research institutions and their libraries collaborate and work with publishers on both parameters and lean workflows.

1. Der disruptive Wandel in der wissenschaftlichen Informationsversorgung erfordert eine neue Infrastruktur

Der digitale Wandel hat dazu geführt, dass Bibliotheken heute wesentliche Anteile ihrer Erwerbungsbudgets für elektronische Verlagsprodukte ausgeben. Mit der Entstehung der elektronischen Zeitschriften ab circa Mitte der neunziger Jahre und nochmals verstärkt durch das Aufkommen elektronischer Monographien etwa zehn Jahre später haben Bibliotheken ihre Workflows und Services zunehmend an eine „Online-Welt“ angepasst, während ihr Ressourcen-Einsatz für die „Print-Welt“ stetig zurückgeht. Ihr Kerngeschäft der Literaturversorgung hat sich dabei zwar strukturell gewandelt, blieb aber in seinem Grundsatz unberührt.

Dieses Kerngeschäft wird derzeit jedoch zunehmend durch die ubiquitäre Verfügbarkeit von wissenschaftlichen Publikationen infrage gestellt.

Das Angebot an frei verfügbaren Dokumenten auf Repositorien von wissenschaftlichen Einrichtungen oder auf Plattformen von Open-Access-Verlagen wird dabei durch eine Grauzone aus sozialen Netzwerken und illegalen, digitalen Schattenbibliotheken erweitert. Die jüngste Diskussion um SciHub (Bohannon 2016) ist dabei das sichtbarste Zeichen dafür, dass der traditionelle Markt der wissenschaftlichen Literatur- bzw. Informationsversorgung im Zeitschriftenbereich nicht mehr funktioniert.

SciHub hat - ähnlich wie Napster Ende der neunziger Jahre für die Musikindustrie - das Potenzial einer disruptiven Innovation (Christensen/Raynor/McDonald 2015), welche die Spielregeln und Geschäftsprozesse für alle Akteure, und damit auch für wissenschaftliche Bibliotheken, grundlegend verändern kann. Die universelle, bequeme und kostenfreie Verfügbarkeit des bisher größten - wenngleich illegalen - Gesamtkorpus wissenschaftlicher Artikel lässt die aufwändige und teure Lizenzierung elektronischer Zeitschriften sowie den Betrieb komplexer bibliothekarischer Zugangssysteme sinnfrei erscheinen. Das auf Verknappung ausgelegte Subskriptionsmodell wird daher - zumindest für den STM-Bereich - zunehmend unter Druck geraten und schließlich gänzlich obsolet werden.

Somit sind neue Wege zur Finanzierung des wissenschaftlichen Publikationswesens erforderlich, sollen die traditionellen, reputationsbildenden Verlags- und Zeitschriftenmarken weiterbestehen. Das Open-Access-Konzept gewinnt vor diesem Hintergrund nun auch jenseits der bisher vorwiegend politisch-moralisch geführten Diskussion als Geschäftsmodell weiter an Bedeutung.

Die Notwendigkeit einer Mittelumschichtung von Subskription zu Open Access ist zwar erkannt und als möglich beschrieben worden (Schimmer/Geschuhn/Vogler 2015) und wird durch Transformationsinitiativen wie OA2020 sowie politische Forderungen nach Open Science verstärkt. Trotzdem wird der Einsatz von Bibliothekserwerbungsmitteln für die Finanzierung des Open-Access-Publizierens noch kritisch diskutiert. Entwicklungen wie SciHub verdeutlichen jedoch, dass das Ende des Subskriptionswesens auch ohne das Zutun der Bibliotheken unmittelbar

bevorsteht. Bibliotheken, und ebenso Dienstleister wie etwa der Buchhandel, sind daher gefordert, ihre Rolle neu zu finden und auszugestalten.

Auch wenn traditionelle Dienstleistungen wie der Erwerb oder die Ausleihe von Büchern fortbestehen, wird sich der Ressourceneinsatz deutlich in die Richtung der Finanzierung und des Managements des institutionellen Publikations-Outputs verschieben müssen, wenn wissenschaftliche Bibliotheken weiterhin eine relevante Rolle für den Wissenschaftsbetrieb wahrnehmen wollen.

Der begonnene Pfad dieser disruptiven Innovation im gesamten wissenschaftlichen Publikationswesen erfordert den Aufbau einer neuen Infrastruktur, die den Akteuren Verlage, Handel und Bibliotheken die Anpassung an und die Entwicklung von neuen Geschäftsmodellen und Services ermöglicht: Damit Bibliotheken die Open-Access-Transformation aktiv und erfolgreich gestalten und ihr Service-Portfolio neu ausrichten können, benötigen sie valide Daten über das Publikationsaufkommen ihrer Einrichtungen und über die damit verbundenen Kosten. Die Finanzierung von Open-Access-Publikationsgebühren muss verankert werden, Geschäftsgänge und Prozesse zur Einreichung und Abrechnung von Artikeln in kostenpflichtigen Open-Access-Zeitschriften sowie ein transparentes Berichtswesen sind möglichst effizient zu organisieren. Das seit Oktober 2015 durch die Deutsche Forschungsgemeinschaft (DFG) geförderte Projekt „Transparent Infrastructure for Article Charges“ (INTACT) leistet einen substantiellen Beitrag zum Aufbau dieser neuen Infrastruktur.

INTACT² ist ein gemeinschaftliches Projekt der Universitätsbibliothek Bielefeld, des „Institute for Interdisciplinary Studies of Science“ (I SoS) an der Universität Bielefeld und der Max Planck Digital Library (MPDL). Es arbeitet eng mit deutschen Hochschulen, außeruniversitären Forschungseinrichtungen und Verlagen zusammen und wird von der DINI AG Elektronisches Publizieren unterstützt. Das Projekt vereint drei Handlungsstränge: Der Projektbereich „OA Analytics“ forscht über das (Open-Access)-Publikationsaufkommen auf institutioneller sowie auf nationaler Ebene. Mit der Initiative „Open APC“³ wird

² <http://www.intact-project.org>

³ <https://github.com/OpenAPC/openapc-de>

im INTACT-Projekt eine Plattform etabliert, über die die Zahlungen für Open-Access-Publikationsgebühren der Wissenschaftseinrichtungen offengelegt werden. Auf einer dritten Ebene adressiert die Initiative „ESAC- Efficiency and Standards for Article Charges“⁴ die notwendige Optimierung der Geschäftsprozesse für APC sowohl an Bibliotheken, als auch zwischen Bibliotheken und Verlagen.

Die ersten Projektphase umfasste schwerpunktmäßig eine Erfassung des Status Quo der Praxis des APC-Managements an wissenschaftlichen Einrichtungen sowie die Weiterentwicklung der Plattform „Open APC“. Die folgende Darstellung zeigt somit die derzeitige Praxis und die Herausforderungen des APC-Geschäftsmodells im nationalen und internationalen Kontext auf und präsentiert entsprechend erste Lösungsansätze aus dem Projekt INTACT.

2. APC-Geschäftsmodell: Praxis und Herausforderungen

2.1 Deutschland

Im Vergleich zu vielen europäischen Ländern gibt es in Deutschland derzeit keine nationale Open-Access-Strategie. Mit dem schon 2009 begonnenen DFG-Förderprogramm „Open Access Publizieren“⁵ sind jedoch an vielen Universitäten bereits Open-Access-Publikationsfonds entstanden. Im außeruniversitären Bereich ist die MPDL der mit Abstand größte Akteur in Deutschland, wenn es um die Finanzierung von kostenpflichtigen Open-Access-Publikationen geht. Seit Januar 2016 bietet zudem die Leibniz-Gemeinschaft einen zentralen Open-Access-Publikationsfonds an.⁶ Mit der „Handreichung Open-Access-Publikationsfonds“ der Arbeitsgruppe Open Access der Schwerpunktinitiative Digitale Information der Allianz der deutschen Wissenschaftsorganisationen liegen Empfehlungen zur Etablierung solcher Fonds vor.⁷

⁴ <http://esac-initiative.org/>

⁵ http://www.dfg.de/foerderung/programme/infrastruktur/lis/lis_foerderangebote/open_access_publizieren

⁶ <http://www.leibniz-gemeinschaft.de/infrastrukturen/open-access/open-access-publikationsfonds/>

⁷ <http://doi.org/10.2312/allianzoa.006>

Von den derzeit 108 Universitäten in Deutschland konnten etwas mehr als ein Drittel bislang von der DFG-Förderung profitieren, an einigen großen Universitäten mit hohem Publikationsaufkommen konnten allerdings bislang noch keine Open-Access-Publikationsfonds etabliert werden. Mit dem voraussichtlichen Ende des DFG-Förderprogramms „Open Access Publizieren“ 2020 und der stärker werdenden politischen Forderung nach Open Science stehen alle Universitäten, ebenso wie die außeruniversitären Forschungseinrichtungen, vor der Herausforderung einer strukturellen Verankerung der Finanzierung von kostenpflichtigen Open-Access-Publikationen. Im Folgenden werden die Ansätze in den Universitäten sowie an der MPDL kurz umrissen und die Herausforderungen dieser neuen Arbeitsfelder aufgezeigt.

2.1.1 DFG-Publikationsfonds

Das 2009 eingeführte und 2014 bis 2020 verlängerte Förderprogramm „Open Access Publizieren“ unterstützt Hochschulen bei der Einführung von Open-Access-Publikationsfonds mit dem Ziel, nachhaltige Strukturen zur Finanzierung von Open-Access-Publikationen zu schaffen. Neben SCOAP3 stellen die DFG-geförderten Open-Access-Publikationsfonds derzeit das zentrale Instrument zur Beförderung der Open-Access-Transformation dar. Das DFG-Programm sieht folgende Förderkriterien verbindlich vor:

- Die zu veröffentlichenden Artikel erscheinen in Zeitschriften, deren Beiträge sämtlich unmittelbar mit Erscheinen im Sinne der Open-Access-Idee frei im Internet zugänglich sind: entgeltfrei und ohne technische oder rechtliche Hürden.
- Die Zeitschriften wenden im jeweiligen Fach anerkannte, strenge Qualitätssicherungsverfahren an.
- Die Artikelbearbeitungsgebühr darf eine Höhe von max. 2.000 Euro pro Aufsatz nicht übersteigen.
- Ein Angehöriger der jeweiligen Universität ist als „submitting author“ oder „corresponding author“ für die Publikation verantwortlich.
- Die Autorinnen sind gehalten, die geförderten Publikationen mit dem Hinweis auf Förderung durch den von der DFG und der jeweiligen Universität geförderten Open-Access-Publikationsfonds zu versehen (Acknowledgement).

Der Ausschluss sogenannter „hybrider“ Publikationskosten von der Förderung (Zahlung von APC für die Open-Access-Freischaltung einzelner Artikel in Subskriptionszeitschriften) sowie die Preisgrenze haben dazu geführt, dass im universitären Bereich Publikationen in echten Open-Access-Zeitschriften gefördert werden und die durchschnittlichen APC-Kosten deutlich unter der Preisgrenze liegen. Allerdings werden in manchen Fällen Ko-Finanzierungen zwischen Publikationsfonds und Autoren, die die Differenzkosten zu hochpreisigen APCs dann aus eigenen Fakultäts- oder Institutsmitteln aufbringen, realisiert.

In fast allen Fällen verwaltet die jeweilige Universitätsbibliothek den Publikationsfonds. Die Universitätsbibliotheken schließen Rahmenverträge mit Verlagen ab, um bei regelmäßig hohem Publikationsaufkommen Rabatte zu realisieren und die Effizienz des APC-Managements zu steigern. Darüber hinaus nehmen die Universitätsbibliotheken entsprechende Berichtspflichten wahr. Über den bis 2013 erreichten Stand des DFG-Förderprogramms haben Fournier/Weihberg (2013) berichtet.

Im Rahmen des INTACT-Projekts können Open-APC-Daten nun über einen OLAP-Server zur Verfügung gestellt. Mit einer Abfrage über die von 24 Universitäten gemeldeten Daten kann das Volumen der APC-Ausgaben beispielsweise für die Jahre 2014 und 2015 einfach bestimmt werden: Mit Stand Mitte Mai 2016 haben die 24 Universitäten im genannten Zeitraum insgesamt 2.640 Open-Access-Zeitschriftenartikel im Umfang von 3.582.872 EURO finanziert. Der Durchschnittspreis pro Artikel betrug in diesem Ausschnitt 1.266 EURO. Im Vergleich dazu haben die datengebenden außeruniversitären Forschungseinrichtungen im selben Zeitraum 1.207 Artikel im Umfang von 1.641.196 EURO finanziert, allerdings lag der Durchschnittspreis mit 1.553 EURO deutlich über dem der Universitäten.

2.1.2 Max Planck Digital Library

Während an den Universitätsbibliotheken die Open-Access-Publikationsfonds separate Finanztöpfe darstellen, die überwiegend losgelöst von den sonstigen Erwerbungsausgaben bestehen und verwaltet werden, hat die MPDL die Mittel zur Finanzierung von Open-Access-Publikationskosten seit vielen Jahren in ihren regulären Erwerbungssetat integriert.⁸ Dabei erfolgt die Kostenübernahme der Open-Access-Publikationsgebühren für Max-Planck-Autorinnen durch die MPDL ausschließlich auf Grundlage von Rahmenverträgen mit Open-Access-Anbietern. Konkrete Bedarfsäußerungen seitens der wissenschaftlichen Autoren und regelmäßige Analysen der Verteilung der Max-Planck-Publikationen über einzelne Zeitschriften und Verlage hinweg gehen dem Abschluss eines Rahmenvertrages voraus. Der Vertrag regelt die zentrale Rechnungslegung an die MPDL sowie weitere Bedingungen, wie etwa den Prozess zur Identifikation und Bestätigung der berechtigten Autoren und die unter den Vertrag fallenden Publikationstypen sowie Preismodelle (Geschuhn/Sikora 2015). Mittlerweile übernimmt die MPDL jährlich zentral die Kosten für rund 500 Open-Access-Artikel aus der Max-Planck-Gesellschaft. Sie ist damit derzeit in Deutschland im Hinblick auf das APC-basierte Geschäftsmodell die Einrichtung mit dem größten Fallaufkommen und Finanzeinsatz.

2.1.3 Praktische Herausforderungen

INTACT hat im Rahmen von Workshops und Fachveranstaltungen zahlreiche Gespräche und Diskussionen mit Betreibern von Publikationsfonds und Open-Access-Managerinnen geführt. Auch an der MPDL wird das Artikelgebührenmanagement seit Projektbeginn systematisch begleitet und in der Projektgruppe regelmäßig diskutiert. Die folgende Darstellung gibt einen systematischen Überblick über die derzeitigen strukturellen, finanziellen und operativen Herausforderungen an deutschen Bibliotheken in diesem Kontext.

⁸ Die MPDL ist eine zentrale Serviceeinrichtung der Max-Planck-Gesellschaft. Ihre Aufgabe ist u.a. die zentrale Versorgung der über 80 Max-Planck-Institute mit digitalen Informationsressourcen.

Finanzierung

Die nachhaltige finanzielle Absicherung von APC-Budgets stellt derzeit in den meisten Bibliotheken die größte Herausforderung dar. Da die Förderkriterien der DFG für Publikationsfonds an Universitäten eine stetige Erhöhung des Eigenanteils vorsehen, erhöht sich zunehmend der Druck auf die Einrichtungen, stabile Finanzierungskonzepte zu entwickeln, will man diesen eingeführten Service nicht wieder einstellen. Dabei sind insbesondere die zweischichtigen Hochschulbibliothekssysteme mit strukturellen Hindernissen konfrontiert. In zweischichtigen Systemen verfügen die Institutsbibliotheken häufig über eigene Literaturerwerbungssetats und finanzieren die zentral über die Universitätsbibliothek bereitgestellten Medien und Dienste in der Regel nur anteilig. Zahlreiche Universitätsbibliotheken haben somit keine uneingeschränkte Verfügungsgewalt über den gesamten Medienerwerbungssetat. Die DFG- geförderten Publikationsfonds werden hingegen allen wissenschaftlich Publizierenden einer Hochschule zur Verfügung gestellt, unabhängig von ihrer Zugehörigkeit zu einem Fachbereich oder Institut. Die Bereitschaft von Institutsbibliotheken, einen Publikationsfonds zu bezuschussen und eventuell dafür sogar noch eigene Erwerbungsmittel umzuschichten bzw. eigene Erwerbungsausgaben zugunsten von Open Access einzuschränken ist oft nicht gegeben. Darüber hinaus ist die Ko-Finanzierung von Zentral- und Institutsbibliotheken in der praktischen Umsetzung mit administrativen Aufwänden verbunden. Eine weitgehende Zentralisierung von Erwerbungsmitteln zumindest für den Bereich der Zeitschriften ist für die Open-Access-Transformation jedoch von essentieller Bedeutung, da nur so eine Verbindung zwischen Subskriptionsausgaben und APC hergestellt werden kann. Sie kann insbesondere auch in zweischichtigen Bibliothekssystemen eine Möglichkeit zur Stärkung von ohnehin notwendigen Zentralisierungsbemühungen sein. Im Moment fördern die Publikationsfonds ausschließlich Artikel in genuinen Open-Access-Journalen. Für die finanzielle Nachhaltigkeit und im Sinne der Open-Access-Transformation ist es jedoch auch hier erforderlich, Mittel von der Subskriptionsseite in Open-Access-Modelle umzuschichten, um einen Anstieg der Gesamtkosten für das wissenschaftliche Publikationswesen zu vermeiden. Stellen schließlich auch

Subskriptionsverlage auf ein Open-Access-Modell um (siehe auch Abschnitt „Offset Deals“), ist ein umfassendes aggregiertes Bild über die Publikationen einer Einrichtung und Subskriptionszahlungen an einzelne Verlage unabdingbar.

An der MPDL ist diese Aggregation durch die zentrale Budgetierung und Erwerbung der elektronischen Ressourcen bereits gelungen. Dennoch ist auch hier das Kostenbild noch nicht ganz vollständig. Im Gegensatz zu den zentralen Lizenzkosten sind jene Zahlungen, die lokal von den einzelnen Einrichtungen der Max-Planck-Gesellschaft an die wissenschaftlichen Verlage gehen, noch eine weitgehend unbekannte Größe. Hierzu gehören insbesondere die Zahlungen für hybrides Open Access sowie weitere Gebühren im Publikationsprozess wie zum Beispiel die sogenannten „Color“- oder „Pages-Charges“, Zahlungen für Sonderdrucke oder auch der Rechteerwerb für die Weiternutzung von Bild- und Datenmaterialien.

Datenevidenz

Um die benötigten Finanzmittel eines Fonds abschätzen zu können, werden verlässliche Kenngrößen über den Publikations-Output einer Einrichtung benötigt. Das heißt also, in welchen Zeitschriften und bei welchen Verlagen publizieren die Wissenschaftler einer Universität oder einer Forschungseinrichtung, und welche Trends und Entwicklungen lassen sich hieraus ablesen? Die Erhebung und Analyse von Publikationsdaten stellt derzeit allerdings noch für viele Einrichtungen eine Hürde dar.

Selbst wenn der Zugang zu den einschlägigen Datenbanken besteht, reicht die Datenqualität darin vielfach nicht aus, um beispielsweise alle Institutionenbezeichnungen von einer Einrichtung als zugehörig zu identifizieren und zusammenzuführen. Des Weiteren stehen viele Einrichtungen vor der Herausforderung, dass sie mit den ihnen zur Verfügung stehenden Mitteln nicht alle Open-Access-Publikationen erfassen können. Eine genaue Publikations-datenanalyse zusammen mit festen Kriterien für die Kostenübernahme (zum Beispiel Festlegung der berechtigten Autorenschaft, Preisobergrenzen, Open-Access-Lizenz und Publikationstyp) sind für

Budgetplanungen jedoch ebenso relevant, wie die Modellierung von verlagsbezogenen Transformationsszenarien.

Generelle Verfahren der Kostenübernahme

Die Kostenübernahme für Open-Access-Publikationsgebühren handhaben wissenschaftliche Bibliotheken derzeit ganz unterschiedlich. Folgende Verfahren sind dabei im Einsatz:

- Die Erstattung von Einzelrechnungen an die Autoren.
- Vorauszahlungsmodelle bei einzelnen Verlagen. Hier wird ein bestimmter Betrag an einen Verlag vorausgezahlt, von dem die Gebühren für einzelne Publikationen aus einer Einrichtung nach und nach abgezogen werden. Häufig gewährt der Verlag im Gegenzug einen Rabatt auf die APC. Dieses Modell wird auch häufig als „Membership-Modell“ bezeichnet.
- Rahmenverträge mit Open-Access-Anbietern, die die zentrale Rechnungslegung an die Bibliothek beziehungsweise an die Publikationsfondsbetreiber regeln. Die Publizierenden erhalten somit keine Rechnungen.

Welche Verfahren gewählt werden, hängt zum einen von dem Publikationsaufkommen einer Einrichtung ab, aber auch davon, welche Modelle verlagsseitig angeboten werden. Handelt es sich nur um wenige Open-Access-Publikationen pro Jahr, ist eine Erstattung von Einzelrechnungen sicherlich noch handhabbar. Einige Bibliotheken setzen begleitend Online-Formulare ein, über die die Autoren ihre Publikation zunächst anmelden müssen und die Möglichkeit der Kostenübernahme zunächst anfragen. Steigen die Fallzahlen jedoch stark an, lassen sich solche manuelle Verfahren nur mit hohem Personaleinsatz bewältigen. Vorauszahlungsmodelle sind immer dann sinnvoll, wenn mit einer bestimmten Mindestanzahl an Publikationen bei einem Verlag gerechnet werden kann. Hier berichten Einrichtungen allerdings, dass die von den Verlagen geforderten Vorauszahlungssummen oft nicht dem tatsächlichen Publikationsaufkommen einer Einrichtung in einer Periode entsprechen, oder dass diese vielmehr unabhängig von entsprechenden Abschätzungen verlangt werden. Eine Vorauszahlungsoption kann zwar zur Budgetflexibilität beitragen, die Finanzierung von noch nicht erbrachten

Leistungen wird aber auch als haushaltsrechtliches Problem diskutiert und erschwert darüber hinaus die Möglichkeit der Ko-Finanzierung zwischen zentraler und dezentraler Einrichtung, sofern diese praktiziert werden soll. Ebenso sind Mitgliedschaften nicht für alle Einrichtungen unproblematisch. Diskutiert wird auch, ob die Verfügbarkeit von Vorauszahlungs- und Mitgliedschaftsmodellen die Publizierenden in der Auswahl eines Publikationsorgans beeinflusst. Vor diesem Hintergrund ist es insgesamt als problematisch einzustufen, wenn Verlage ihre Geschäftsbeziehungen ausschließlich über Mitgliedschaften und Vorauszahlungen organisieren und den Institutionen keine Wahlmöglichkeiten anbieten.

Kritisch hinterfragt werden müssen auch verlagsseitig angebotene Rabatte. Selbstverständlich fließen diese in die APC-Preisgestaltung eines Verlages mit ein und bieten insbesondere dort keinen Anreiz, wo keine Alternativen zur Mitgliedschaft oder Vorauszahlung bestehen. Da die eigentliche „Kaufentscheidung“ auf dem APC-Markt zudem den Publizierenden zufällt, sind die den Institutionen angebotenen Rabattmodelle letztlich weitgehend artifiziell und können für die angestrebte Preis- und Markttransparenz kontraproduktiv sein.

Reine Rahmenverträge mit Open-Access-Anbietern erweisen sich hingegen als ein vergleichsweise unproblematisches Instrument zur Organisation der Publikationskostenübernahme, da sie in der Regel keinerlei finanzielle Verpflichtungen jenseits der tatsächlich publizierten Artikel beinhalten. Neben den Vereinbarungen zur zentralen Rechnungsstellung definiert der Rahmenvertrag die berechtigten Autoren, die unter den Vertrag fallenden Zeitschriften, Artikeltypen und -preise sowie das Verfahren der Autorenidentifikation und -bestätigung. Jedoch lohnt sich gerade für kleinere Einrichtungen mit wenigen Publikationen das Aufsetzen eines solchen Vertrages kaum. Oft ist es schon schwierig, mit den großen Anbietern diesbezüglich ins Gespräch zu kommen. Aus diesem Grund besteht bei kleineren Einrichtungen der Wunsch nach einer Partizipation an zentral oder konsortial organisierten Vertragsmodellen.

Autorenidentifikation

Eine zentrale Herausforderung im Abrechnungsprozess zwischen den Verlagen und wissenschaftlichen Einrichtungen stellt die Identifikation der berechtigten Autoren dar. Wird die Kostenübernahme über Rahmenverträge oder Mitgliedschaftsmodelle organisiert, muss der Verlag in der Lage sein, Autorinnen über das Einreichungs- und Publikationssystem der zahlenden Einrichtung zuzuordnen. Diese Verfahren sind aufgrund der Verschiedenheit der eingesetzten Publikationssysteme bislang wenig standardisiert. In der Regel erfolgt die Zuordnung der Autoren über die Erkennung von institutionellen E-Mail-Adressen, über IP-Authentifizierung oder über die entsprechenden Angaben der Autorinnen bei Einreichung. Als problematisch erweist sich in diesem Zusammenhang auch, wenn ein Verlag nicht sein gesamtes Zeitschriftenportfolio über ein zentrales Publikationssystem administriert, sondern einzelne Titel beispielsweise von Fachgesellschaften über andere Publikationsworkflows herausgegeben werden.

Wichtig ist hierbei, dass Verlage die einreichenden Autorinnen in ihren Systemen auf entsprechende Vereinbarungen aufmerksam machen und über die jeweiligen Konditionen, Ansprechpartner und Verfahren informieren. Beispielsweise setzt sich zunehmend durch, dass Open-Access-Publikationsfonds nur die Kosten für Publikationen der korrespondierenden Autoren (corresponding authors) aus ihren Einrichtungen übernehmen. Sind Autorinnen mit mehreren Institutionen affiliert, etwa mit einer Universität und einer außeruniversitären Forschungseinrichtung gleichzeitig, muss die zahlende Einrichtung korrekt gewählt und auch in der Publikation an sich ausgewiesen sein. Hierfür müssen Verlage entsprechende Prüfmechanismen etablieren.

Möglichst schon bei Einreichung eines Artikels sollte die zahlende Einrichtung informiert werden und die Möglichkeit haben, die Kostenübernahme zu bestätigen bzw. gegebenenfalls abzulehnen. Verlage versenden hierfür in der Regel E-Mail-Alerts an zuvor vereinbarte Account-Manager-Adressen. Zunehmend werden diese Prozesse aber auch in sogenannte Dashboards

verlagert. Dies sind Account-Administratoren-Portale, die den Institutionen eine Übersicht über die Einreichungen bieten. Bislang bietet jedoch keines der von den Verlagen angebotenen Dashboards die notwendigen Funktionalitäten, um die APC-Administration an den zahlenden Einrichtungen tatsächlich zu erleichtern. Eine zentrale Mindestanforderung für solche Systeme wäre zum Beispiel die Möglichkeit, den aktuellen Kontostand automatisch abrufen und exportieren zu können, was insbesondere bei geleisteten Vorauszahlungen von Bedeutung ist. Darüber hinaus stellen die Administratoren-Plattformen oftmals nicht alle Informationen zur Verfügung, die für die Kostenübernahme benötigt werden. Dazu gehören unter anderem

- vollständige Angaben zu den beteiligten Autoren, zum Autorenstatus (corresponding author) und genaue Affiliationsangaben (zum Beispiel genaue Institutsbezeichnung),
- der Artikel als solcher zur Überprüfung des Autorenstatus in der Publikation selbst,
- DOI und gegebenenfalls Link zu dem publizierten Artikel,
- Preisangaben,
- exportfähige Titellisten der zu einem Vertrag gehörenden Zeitschriften inklusive APC.

Rechnungslegung

Auch in Bezug auf die Rechnungsstellung berichten die deutschen Publikationsfondsbetreiber von zahlreichen Hindernissen, die eine reibungslose Abwicklung der Zahlungsvorgänge erschweren. Nicht alle Open-Access-Anbieter haben die zentrale Rechnungsstellung an wissenschaftliche Einrichtungen in ihre Abläufe integriert und fordern dementsprechend weiterhin die Autorinnen zur Zahlung auf. Weitere Hürden sind unter anderem

- mangelnde Konformität der Rechnungen mit den haushaltsrechtlichen Rahmenbedingungen deutscher Einrichtungen, wie zum Beispiel zu kurze Zahlungsziele und fehlende Steueridentifikationsnummern (VAT)
- fehlende Angaben auf den Rechnungen, wie zum Beispiel Verlagsnamen, Artikel-Metadaten, detaillierte Preis- und Rabattinformationen, Mehrwertsteuer
- Kreditkartenzahlungen

Die Rechnungsstellung erfolgt je nach Publikationsaufkommen über Einzel- oder Sammelrechnungen. Bei manche Verfahren fällt die Überprüfung der Autorenberechtigung und damit die Bestätigung der Kostenübernahme mit der Rechnungslegung zusammen. Die Reklamation von eventuell fehlerhaft berechneten Publikationen ist zu diesem Zeitpunkt im weit fortgeschrittenen Publikationsprozess einerseits sehr aufwändig. Da jedoch an dieser Stelle ohnehin eine Rechnungsprüfung stattfindet, vereinfacht diese Praxis im Normalfall den Prozess. Die nachfolgende Grafik der MPDL veranschaulicht die unterschiedliche Handhabung von Bestätigung und Rechnungsstellung für die APC-Kostenübernahme innerhalb der jeweiligen Publikationsprozesse der Open-Access-Verlage.

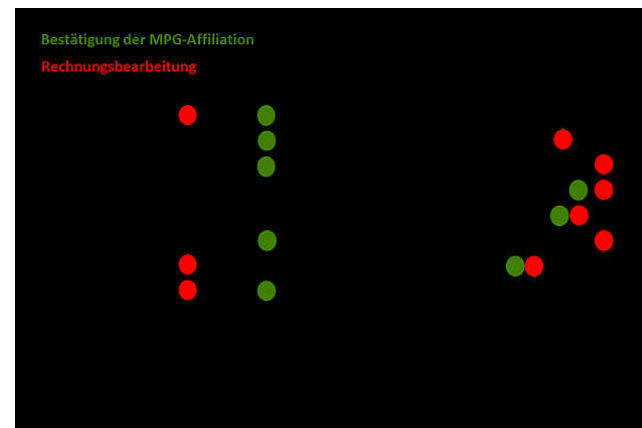


Abbildung 1: Zeitpunkte der Affiliationsbestätigung und Rechnungsbearbeitung nach Verlag

Reporting und Transparenz

Die Berichtsanforderungen der DFG an die Betreiber von Open-Access-Publikationsfonds sind in einem Datenblatt des Förderprogramms zusammengefasst.⁹ Für die Diskussion mit Hochschulleitungen über die nachhaltige Finanzierung von kostenpflichtigen Open-Access-Publikationen oder zur Entwicklung von Verteilungsmodellen werden außerdem zusätzlich Daten- bis hin auf Artikelebene- benötigt. Neben den Forschungsförderern und den Leitungen wissenschaftlicher Einrichtungen fordert die Öffentlichkeit

⁹ http://www.dfg.de/formulare/12_201/12_201_de.pdf

zunehmend Transparenz über die von Bibliotheken verausgabten Mittel, und letztlich sind Daten über das Open-Access-Publizieren selbst inzwischen Gegenstand wissenschaftlicher Forschung. Neben der Verfügbarkeit von Kosteninformationen ist die Preis- und Kostentransparenz ein zentrales Postulat der Open-Access-Transformation. In Unterscheidung zum Subskriptionsmarkt, auf dem über Vertraulichkeitsklauseln und komplexe Lizenzmodelle ein Vergleich der Ausgaben und Kostenstrukturen zwischen den Einrichtungen verhindert werden soll, liegt die Herausforderung in Bezug auf neue Open-Access-Geschäftsmodellen darin, funktionierende Markt- und Wettbewerbsmechanismen wiederherzustellen. Insbesondere muss der Gefahr eines erneuten unverhältnismäßigen Kostenanstiegs begegnet werden.

2.2 International: Offsetting-Verträge

Im internationalen Kontext ist bislang Großbritannien durch eine starke Förderung des goldenen Open-Access-Weges in den Vordergrund getreten. Den Empfehlungen des sogenannten Finch Report (Finch 2012) folgend stellten die britischen Research Councils 2013 umfangreiche Mittel zur Finanzierung des Open-Access-Publizierens auch über den hybriden Weg zur Verfügung, wodurch das APC-Geschäftsmodell flächendeckend Eingang in die britischen Wissenschaftseinrichtungen und Bibliotheken fand.

Die unter dem Schlagwort „double dipping“ viel kritisierte Praxis einer zweifachen Finanzierung wissenschaftlicher Publikationen im hybriden Open Access führte jedoch zu einem erheblichen Anstieg der Gesamtkosten des wissenschaftlichen Publikationssystems, sodass in Großbritannien empfohlen wurde, die Kosten für das hybride Open Access mit den Subskriptionskosten in Verbindung zu setzen. Österreich war dabei das erste Land, das sich mit dem Anbieter IOP auf ein neues Vertragsmodell einigen konnte, welches eine Verrechnung der Hybrid-APC mit den geleisteten Subskriptionszahlungen vorsieht (sogenanntes „Offsetting“).

Nach und nach haben mehrere Subskriptionsverlage diesen Ansatz adaptiert. Die Verabschiedung zahlreicher nationaler Open-Access-Strategien im

europäischen Raum (Österreich¹⁰, Niederlande¹¹, Schweden¹², Dänemark¹³, Norwegen¹⁴ und Finnland¹⁵) trug dabei zur Beförderung von Offsetting bei. Der Vorteil im Hinblick auf die Realisierung von Open Access liegt auf der Hand: Die bestehenden Subskriptionszeitschriften stellen ihre Geschäftsmodelle um, gleichzeitig bleiben ihre für die Wissenschaft zentralen Kommunikations- und Reputationsfunktionen erhalten. Im Vergleich zu dem mit rechtlichen Unsicherheiten behafteten grünen Weg oder einem wie auch immer gearteten parallelen Aufbau neuer Publikationswege birgt dieser Ansatz die Möglichkeit eines schnellen und großflächigen Umstiegs auf Open Access.

Offsetting kann- je nach Verlag und nach Publikationsverhalten einer wissenschaftlichen Einrichtung zu einem sprunghaften Anstieg der APC-Fallzahlen führen. Das neue Vertragsmodell vermittelt von daher einen ersten Eindruck des hier thematisierten grundlegenden Wandels der Geschäftsbereiche in Bibliotheken. Neben den oben geschilderten Anforderungen an eine praktische Implementierung stehen beim Offsetting-Modell jedoch zunächst grundsätzliche, vertragsstrategische Herausforderungen im Vordergrund. In der öffentlichen Diskussion werden die bisherigen Vertragsabschlüsse häufig kritisiert, ein dysfunktionales, monopolistisch organisiertes Verlagssystem bedenkenlos weiter zu stützen.¹⁶ Bibliotheken und Wissenschaftseinrichtungen sind hier zu einer nachhaltigen Ausgestaltung der Rahmenbedingungen herausgefordert. Aus diesen Überlegungen heraus initiierte INTACT im Rahmen der Initiative ESAC im März 2016 einen zweitägigen internationalen Workshop zum Thema Offsetting, zu dem Vertreter von Konsortien sowie Wissenschafts-, Bibliotheks- und Informationseinrichtungen geladen wurden, die sowohl mit der Verhandlung als auch mit der praktischen Implementierung der derzeit laufenden oder zur Verhandlung anstehenden Verträge betraut sind.

¹⁰ https://zenodo.org/record/34079/files/OANA_OA-Recommendations_30-11-2015.pdf

¹¹ <http://www.openaccess.nl/en/in-the-netherlands/what-does-the-government-want>

¹² <https://publikationer.vr.se/en/product/proposal-for-national-guidelines-for-open-access-to-scientific-information/>

¹³ <http://ufm.dk/en/research-and-innovation/cooperation-between-research-and-innovation/open-science/open-access-to-research-publications>

¹⁴ http://www.forskningradet.no/en/Newsarticle/A_boost_for_open_access_to_research/1253997204282

¹⁵ <http://openscience.fi/>

¹⁶ Siehe hierzu auch <http://esac-initiative.org/offsetting-under-construction/>

Über den Vergleich der individuellen verhandlungsstrategischen Herangehensweisen an das Offsetting-Modell in den jeweiligen Ländern, der Vertragsstrukturen der Anbieter sowie durch die Diskussion grundsätzlicher Ziele und Anforderungen ließen sich folgende Kernproblematiken herauskristallisieren:

- Die Verhandlung von Offsetting-Verträgen ist in den meisten Ländern zunächst von dem Wunsch motiviert, die Anteile der im Open Access verfügbaren Publikationen möglichst rasch zu erhöhen, um den politischen Zielsetzungen Rechnung zu tragen. Richtungsweisende Überlegungen im Hinblick auf eine globale Open-Access-Transformation, die eine grundsätzliche Überwindung des Subskriptionswesens zur Folge hat, treten hierbei eher in den Hintergrund. Dementsprechend fehlt es an konkreten vertraglichen Anreiz- oder Sanktions-Mechanismen, die diesen Weg befördern.
- Die von den Verlagen angebotenen Modelle unterscheiden sich stark zum einen im Hinblick auf den Umfang der Verrechnung von Subskriptions- und Publikationskosten (volle Anrechnung versus prozentuale, jährlich steigende Anteile) und zum anderen in Bezug auf die Mechanismen der Verrechnung (zum Beispiel Gutscheinsysteme oder nachträgliche Verrechnung).
- Alternative Szenarien im Falle eines Scheiterns der Verträge sowie Konzepte für eine nachhaltige Fortführung über den Vertragszeitraum hinaus fehlen in der Regel.
- In vielen Fällen steigen die Vertragskosten trotz der Zusammenlegung von APC und Lizenzkosten insgesamt an, da vielfach zunächst eine einheitliche Lizenz-, also Zugangslage, geschaffen werden musste, wodurch sich die in den Verträgen enthaltenen Subskriptionsanteile erhöhen.
- In den Verhandlungen stellt die mangelnde Verfügbarkeit von verlässlichen, verlagsunabhängigen Publikationsdaten ein Problem dar.

Die derzeitigen Offsetting-Verträge wurden überwiegend auf nationaler Ebene oder über große Konsortien abgeschlossen. Zum einen bietet dieses Setting grundsätzliche strategische Vorteile, insbesondere dann, wenn hochrangige Repräsentantinnen aus der Wissenschaft oder der politischen Ebene die Verhandlungen führen. Andererseits ergeben sich hieraus oft komplexe

Anforderungen an eine praktische Implementierung, die auf der Ebene der lokalen Wissenschaftseinrichtungen erfolgen muss, jedoch zentral zu koordinieren ist.

3. INTACT als Transparente Infrastruktur für das APC-Geschäftsmodell

Die Implementierung des APC-Geschäftsmodells ist mit zahlreichen Herausforderungen verbunden. Der Bedarf nach gemeinsamen Prinzipien, nach Standards und Kennzahlen, nach Kostentransparenz und effizienten, skalierbaren Prozessen wird klar erkennbar.

Für eine Verbesserung der praktischen Prozesse zwischen den Verlagen und wissenschaftlichen Einrichtungen ist es zunächst wichtig, dass Bibliotheken ihre Anforderungen gemeinsam formulieren. Insbesondere kleineren Einrichtungen gelingt es nicht immer, ihre Anliegen durchzusetzen. Die in INTACT integrierte ESAC-Initiative trägt, zum Beispiel über Workshops, entsprechende Fragen und kritischen Punkte zusammen, um sie gebündelt zu kommunizieren. Die ESAC-Website wird damit auch Anlaufpunkt für Verlage, die APC-basierte Prozesse aufsetzen. Für den weiteren Projektverlauf sind deshalb auch Kooperationen mit Verlagen geplant. Damit wird ESAC perspektivisch zur Kommunikationsschnittstelle zwischen Verlagen und Bibliotheken im Kontext der Open-Access-Transformation.

Im Fokus des Projektes stehen jedoch nicht nur die rein praktischen Prozesse des APC-Handlings. Die se werden sich, analog zum Subskriptionsmodell, zukünftig auch mithilfe von Zwischendienstleistern effizienter gestalten lassen. Wo jedoch neue Vertragsmodelle entstehen, wie zum Beispiel Offsetting oder APC-Rahmenverträge, ist vor dem Hintergrund der angestrebten Open-Access-Transformation ein Austausch über Zielsetzungen und Mechanismen unerlässlich. Die ESAC-Initiative treibt die Vernetzung der Akteure voran und bietet eine Plattform zur Kommunikation von Anforderungen, Richtlinien und Best Practices. In Bezug auf das Offsetting-Modell wurden dementsprechend gemeinsame Zielsetzungen formuliert und veröffentlicht, welche eine klare Ausrichtung der Verträge auf die Open-Access-Transformation fordern. Dazu

gehören der schrittweise und letztlich vollständige Abbau von Zugangskosten in den Verträgen, ausdifferenzierte APC Preisstrukturen und eine vollständige Ablösung der jetzigen Garantiezahlungen durch ein publikationsbasiertes Zahlungsmodell.

In diesem Zusammenhang und auch im Kontext der Publikationsfonds sind verlässliche und verlagsunabhängige Zahlen über den Publikations-Output wissenschaftlicher Einrichtungen und deren Open-Access-Anteile die entscheidende Voraussetzung für Kostenprojektionen und für die Schaffung von Verhandlungsgrundlagen. Im Hinblick auf die angestrebte möglichst vollständige Open-Access-Transformation bilden sie zudem die Basis für die Modellierung möglicher Ausgleichsmechanismen, etwa zwischen publikationsstarken und weniger publikationsstarken wissenschaftlichen Einrichtungen. Das Teilprojekt „Open Analytics“ im Rahmen von INTACT begegnet diesen Fragestellungen über bibliometrische Analysen.

Mit der integrierten Open-APC-Initiative können Kennzahlen über das Publikationsaufkommen einzelner Einrichtungen mit den entsprechenden Kosteninformationen verbunden werden. Damit entsteht eine verlagsunabhängige, transparente empirische Basis für Vertragsverhandlungen sowohl für Offsetting-Modelle, als auch für nach erfolgter Open-Access-Transformation rein publikationsbezogene Modelle. Insbesondere in Kombination mit bibliometrischen Analysen ergibt sich hier die Diskussionsgrundlage über differenzierte Preisstrukturen, die zurzeit in Offsetting-Verträgen noch nicht realisiert sind.

Schon in der Frühphase des INTACT-Projekts kommt „Open APC“ eine wichtige Rolle in der Darstellung und Offenlegung von APC-Zahlungsströmen zu, die inzwischen auch international wahrgenommen wird. So ist in Schweden mit Unterstützung durch INTACT geplant, eine analoge Initiative aufzubauen, ebenso beabsichtigen JISC und der „FP7 post-grant Open Access publishing funds pilot“ im Rahmen von OpenAIRE2020 eine Zusammenarbeit mit INTACT. Darüber hinaus referenziert Science Europe ebenfalls die Open-APC-Initiative.

Neben der Entwicklung einer grundlegenden Qualitätssicherung der APC-Daten sowie des Aufbaus eines OLAP-Servers, ist die Darstellung aller

gemeldeten APC-Daten über eine interaktive Treemap eine erste Umsetzung der ohnehin öffentlich über GitHub zugänglichen APC-Daten. Die Treemap ermöglicht ausgehend von zu wählenden verschiedenen Aggregations-Ebenen einfache Auswertungen, bis hin zu den einzelnen Journals oder den über APCs finanzierten Artikeln.

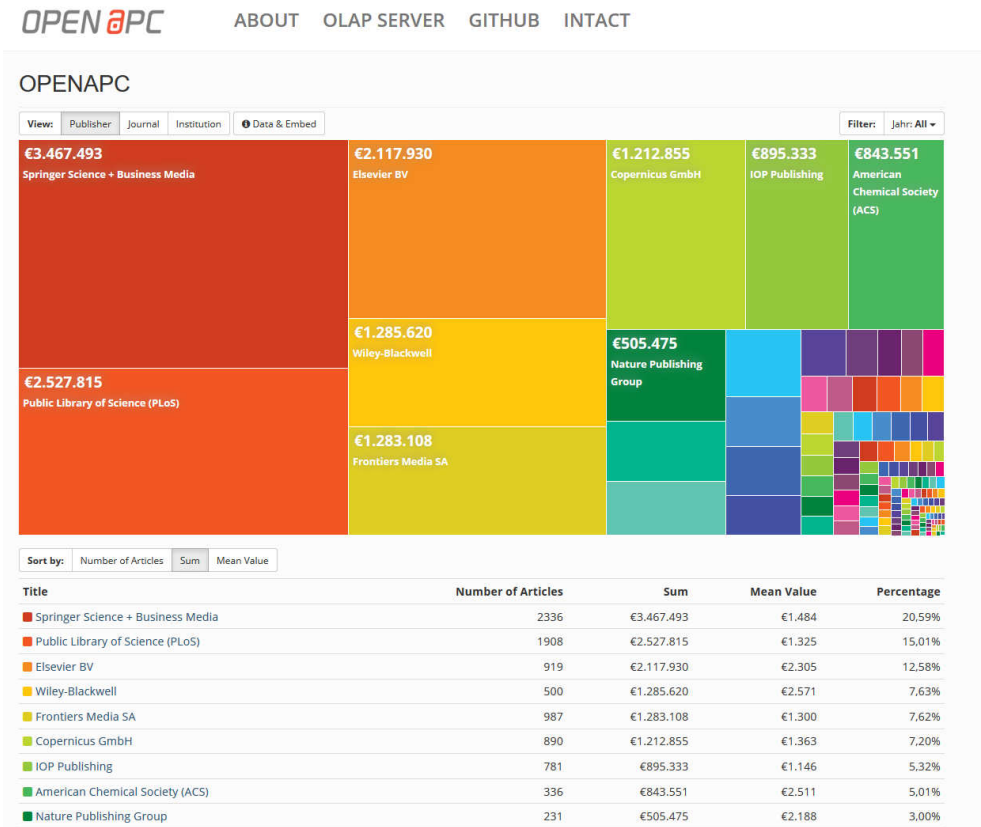


Abbildung 2: Treemap-Darstellung der APC-Ausgaben aller datengegebenden Institutionen nach Verlagen

Über eine solche Darstellung ist die rasche Information über Zahlungsströme von Institutionen an Verlage oder der rasche Vergleich von Preisen und deren Entwicklung möglich. Nicht zuletzt können auf einem Markt gebildete APC-Durchschnittspreise Richtpreise bei Verhandlungen von Offsetting-Verträgen darstellen.

INTACT kann über das Zusammenspiel der dargestellten Teilbereiche die Open-Access-Transformation wirkungsvoll unterstützen. Wichtig sind dabei die

internationale Vernetzung sowie die Kooperation mit Verlagen und ähnlichen Initiativen.

4. Ausblick

Die Bewährungsprobe für Bibliotheken, die mit dem neuen Geschäftsmodell einhergeht, liegt in der Überwindung der Unzulänglichkeiten des Subskriptionsmodells durch eine aktive, gemeinschaftlich organisierte Ausgestaltung von nachhaltigen Rahmenbedingungen für Open Access. Dies bedeutet insbesondere, die aus dem Subskriptionsmarkt bekannten Preissteigerungsraten auf dem APC-Markt nicht zu akzeptieren und von Beginn an transparente Verfahren für den Umgang mit Preisen und Kosten zu etablieren. Offsetting-Verträge können dabei die Open-Access-Transformation als Zwischenschritt weiter befördern. Verlage müssen allerdings ihre Kosten senken, da die Verknappung des Zugangs zu wissenschaftlichen Publikationen nicht mehr aufrechterhalten werden kann.

Auch in Bezug auf Vertragsstrukturen und Geschäftsprozesse haben Bibliotheken nun die Möglichkeit, die Schieflagen des wissenschaftlichen Publikationsmarktes, die sich mit dem digitalen Wandel eingestellt haben, offensiv anzugehen. Die zentrale Rolle, die Bibliotheken hier zufällt, ist offensichtlich: Bibliotheken verfügen über umfangreiche Erfahrungen im Umgang mit wissenschaftlichen Verlagen. Sie stehen in engem Kontakt mit den Publizierenden ihrer Einrichtungen. Über ihre Erwerbungsbudgets haben sie darüber hinaus ein strategisch wichtiges Instrument in der Hand, um dieser Aufgabe zu begegnen.

Für Bibliotheken steht aber noch mehr auf dem Spiel. Der Wandel des wissenschaftlichen Publikationswesens zu immer mehr Open Access, die weiter zunehmende Verfügbarkeit von Publikationen sowie die Konkurrenz durch einfache Zugangssysteme zwingt Bibliotheken, ihr Serviceportfolio neu auszurichten und die Unterstützung des kostenpflichtigen Open-Access-Publizierens an möglichst allen Universitäten und außeruniversitären Forschungseinrichtungen auszubauen.

Literatur:

(Bohannon 2016) John Bohannon: Who's downloading pirated papers? Everyone, in: Science Mag, Apr. 28, 2016. <http://www.sciencemag.org/news/2016/04/whos-downloading-pirated-papers-everyone>

(Christensen/Raynor/McDonald 2015) Clayton M. Christensen, Michael E. Raynor, Rory McDonald: What Is Disruptive Innovation?, in: Harvard Business Review, December 2015 Issue. <https://hbr.org/2015/12/what-is-disruptive-innovation>

(Finch 2012) Finch, Dame Janet (Hrsg.): Accessibility, sustainability, excellence: how to expand access to research publications. Report of the Working Group on Expanding Access to Published Research Findings. <http://www.researchinfonet.org/wp-content/uploads/2012/06/Finch-Group-report-FINAL-VERSION.pdf>

(Fournier/Weihberg 2013) Fournier, Johannes/ Weihberg, Roland: Das Förderprogramm „Open Access Publizieren“ der Deutschen Forschungsgemeinschaft. Zum Aufbau von Publikationsfonds an wissenschaftlichen Hochschulen in Deutschland, in: ZfBB 60 (2013) 5, S. 236-243, http://zs.thulb.uni-jena.de/servlets/MCRFileNodeServlet/jportal_derivate_00234000/j13-h5-auf-1.pdf

(Geschuhn/Sikora 2015) Geschuhn, Kai/ Sikora, Adriana: Management von Article Processing Charges - Herausforderungen für Bibliotheken, in: o-bib Das offene Bibliotheksjournal 1 (2015) 2, S. 27-34, <http://dx.doi.org/10.5282/o-bib/2015H1S27-34>

(Schimmer/Geschuhn/Vogler 2015) Schimmer, R., Geschuhn, K. K., Vogler, A.: Disrupting the subscription journals' business model for the necessary large-scale transformation to open access. <http://doi:10.17617/1.3>

Nachweissysteme

Gemeinsam sind wir stärker – die neue Digitale Bibliothek Thüringen (DBT) als kooperative Publikationsplattform Thüringer Hochschulen

Thomas Mutschler

Thüringer Universitäts- und Landesbibliothek Jena

Abstract

Ausgehend von der Betrachtung des Informations- und Medienmarkts zeigt der Beitrag neue Wege des elektronischen Publizierens im Rahmen der Digitalen Bibliothek Thüringen (DBT) auf. Die DBT wird von der Thüringer Universitäts- und Landesbibliothek Jena (ThULB) im Verbund mit weiteren Thüringer Hochschulbibliotheken als kooperative Publikationsplattform betrieben. Die DBT fungiert als Repositorium für den Grünen Weg des Open Access und lädt Wissenschaftlerinnen und Wissenschaftler dazu ein, ihre Lehr- und Forschungsmaterialien online zu veröffentlichen. Im Frühjahr 2016 fand der Relaunch der DBT auf das Open-Source-Framework MIR (MyCoRe Institutional Repository) statt. Damit reagierte die DBT zum einen auf die gestiegenen Ansprüche der Wissenschaftscommunity und zum anderen auf die komplexen Anforderungen hinsichtlich der Aufbereitung wissenschaftlicher Informationen und deren Erschließung im digitalen Zeitalter. Im Rahmen des Beitrags werden Funktionalitäten, Inhalte, Geschäftsprozesse, Partnerschaften sowie Entwicklungsperspektiven vorgestellt.

Starting from reflections on current developments on the information- and mediamarket, the article illustrates new ways of electronic publishing in the context of the “Digital Library Thuringia” (Digitale Bibliothek Thüringen, DBT). As a platform for electronic publishing the DBT is provided by the Thuringian University and State Library of Jena in cooperation with other university libraries. The DBT serves as an institutional repository for the green way of Open Access. Members of the participating universities and researchers are invited to use the DBT for publishing their articles and papers. In spring 2016 the DBT was relaunched to the open-source-framework MIR (MyCoRe Institutional Repository). In this context the DBT meets the demands of the scientific community and the requirements on processing and indexing scientific information in the digital age. Focusing on the DBT, the article introduces functions, content, workflows, cooperation models and future developments of electronic publishing.

Angesichts der hochdynamischen Entwicklung auf dem Medien- und Informationsmarkt können sich Bibliotheken im digitalen Zeitalter längst nicht

mehr auf ihre Bewahrfunktion zurückziehen. Mehr denn je sind sie heutzutage als eigenständige Akteure im Prozess der Wissensproduktion und -vermittlung gefragt. Allerdings bringt das vielzitierte Ende der „Gutenberg-Galaxis“ (Marshall McLuhan) für Bibliotheken nicht nur Herausforderungen mit sich, sondern bedeutet für sie auch die Chance, sich neue Geschäftsfelder und Services zu erschließen. Das elektronische Publizieren und Open Access führen dies eindrucksvoll vor Augen. Gleichzeitig ist erkennbar, dass die digitale Herausforderung des 21. Jahrhunderts nicht allein zu bewältigen ist, sondern nur im Verbund mehrerer Bibliotheken.

Einem solchen kooperativen Ansatz folgt auch die Digitale Bibliothek Thüringen (DBT)¹, welche als kooperative Publikationsplattform von der Thüringer Universitäts- und Landesbibliothek Jena (ThULB) im Verbund mit weiteren Thüringer Hochschulbibliotheken seit inzwischen mehr als zehn Jahren betrieben wird. Dabei erfüllt die DBT zweierlei Funktion: Hochschulangehörige nutzen die DBT als Plattform für die Online-Veröffentlichung ihrer Lehr- und Forschungsmaterialien, die beteiligten Bibliotheken als Repositorium für den digitalen Bestandsaufbau.

Im Frühjahr 2016 fand der Relaunch der DBT auf das Open-Source-Framework MIR (MyCoRe² Institutional Repository) statt. Dieser Umzug auf eine neue Systemumgebung lädt dazu ein, Resümee zu ziehen, und Inhalte der DBT sowie Funktionalitäten, Partnerschaften und Entwicklungsperspektiven vorzustellen. Dabei können die im Folgenden aufzugreifenden Befunde aus Thüringen als durchaus repräsentativ gelten für die Entwicklung institutioneller Repositorien an deutschen Hochschulen. Zur weiteren Einordnung des Themas lohnt deshalb zunächst der übergreifende Blick auf die Herausforderung des elektronischen Publizierens.

Elektronisches Publizieren als Herausforderung

Das elektronische Publizieren hat das akademische Kommunikations- und Publikationswesen in den vergangenen zwanzig Jahren grundlegend verändert: Das reicht von der Art und Weise, wie neue Erkenntnisse aufbereitet

¹ <http://www.db-thueringen.de/>

² <http://www.mycore.de>

werden, bis zu deren Veröffentlichung, Vermittlung, Rezeption, Bewertung und Bewahrung. Damit ist das Publizieren wissenschaftlicher Information im digitalen Zeitalter zu einem hochkomplexen Vorgang geworden. Dies umso mehr, als sich sämtliche an der traditionellen Publikationskette beteiligte Akteure vor neue Anforderungen gestellt sehen. Vor allem das Phänomen des Open Access, also die Idee vom „freien“, entgeltlosen Zugang zu wissenschaftlichen Veröffentlichungen (mit dem elektronischen Publizieren häufig in einem Atemzug genannt), justiert das akademische Publikationswesen völlig neu, so der in Zürich lehrende Wissenschaftshistoriker und Mediziner Michael Hagner.³

Vor mehr als zehn Jahren ins Leben gerufen, bestand das ursprüngliche Ziel der Open Access-Bewegung darin, den Wissenschaftsverlagen alternative Publikationsmöglichkeiten gegenüberzustellen. Zum einen sollten auf diese Art die kurzen Wege des Internets für schnelleres Publizieren genutzt werden, zum anderen sollten Kostensteigerungen bei Periodika im STM-Bereich abgefangen werden („Zeitschriftenkrise“). Bereits in den neunziger Jahren entstanden die ersten disziplinspezifischen Repositorien, wie beispielsweise ArXiv⁴ für Online-Publikationen in der Physik oder PubMedCentral (PMC)⁵ in der Medizin. Als bald kamen die sogenannten institutionellen Repositorien hinzu, welche sich im Umfeld von Universitäten oder Forschungsgesellschaften konstituierten, sowie die (reinen) Open Access-Journals. Publizieren ließ sich hier zwar schnell, vielfach jedoch ohne Qualitätsprüfung (Peer Review), und zumeist außerhalb etablierter Verlagsstrukturen.

Der zweite Entwicklungsstrang des elektronischen Publizierens verlief innerhalb etablierter Bahnen: Mit dem Siegeszug des Internets überführten die Wissenschaftsverlage einen stetig wachsenden Anteil ihrer Inhalte in das digitale Medium und boten sie seitdem als subscriptionspflichtige E-Medien an. Motiviert durch neue förderpolitische Initiativen, entdeckte sodann auch die Verlagswelt einschließlich der akademischen Fachgesellschaften das Thema „Open Access“ für sich: Sie ermöglichten den Autorinnen und Autoren, ihre

³ Vgl. Michael Hagner: Zur Sache des Buches, Göttingen 2015, S. 63ff.

⁴ <http://arxiv.org/>

⁵ <http://www.ncbi.nlm.nih.gov/pmc/>

Artikel gegen Zahlung einer Article Processing Charge (APC) über den „Verlagsserver“ frei zugänglich zu machen. Kosten entstehen bei diesem neuen Geschäftsmodell nicht mehr bei der Subskription der Medien, also am Ende der Publikationskette, sondern am Anfang, bei der Veröffentlichung des einzelnen Artikels – in diesem Sinne ist auch die Titelüberschrift des vorliegenden Sammelbands zu verstehen: „Der Schritt zurück als Schritt nach vorn“.

Die neue bibliothekarische Farbenlehre sprach fortan vom Goldenen und vom Grünen Weg des Open Access-Publizierens: also der Erstveröffentlichung wissenschaftlicher Texte in Open Access-Medien (Gold) oder der parallelen bzw. nachträglichen Archivierung digitaler Inhalte auf einem institutionellen oder disziplinspezifischen Dokumentenserver (Grün). Der Platz reicht hier nicht aus, um auf die zahlreichen Farbvarianten auf der Palette des Open Access einzugehen. Es sei nur soviel hinzugefügt: Auch hier existieren diverse Geschäftsmodelle, sowohl was die Finanzierung von Open Access-Publikationen bei Verlagen betrifft (einzeln, konsortial, per Mitgliedschaft oder Vouchers), als auch bezüglich des Grünen Wegs, also der Variante, Publikationen nach Ablauf entsprechender Embargofristen über Repositorien frei zugänglich zu machen.

Längst ist Open Access ein zentraler Bestandteil der wissenschaftlichen Publikationspraxis geworden und dank der Unterstützung seitens der jeweiligen nationalen Forschungsfördergesellschaften weltweit stark auf dem Vormarsch. So fördert und fordert die Europäische Union Open Access in ihrem Forschungsrahmenprogramm Horizon 2020.⁶

Die Bundesregierung erklärte Open Science vor zwei Jahren zum Leitgedanken der digitalen Agenda 2014-2017.⁷ Und auch die Regierungschefs der G8-Staaten haben bei ihrem Gipfeltreffen 2013 ein Open Access-Statement abgegeben.⁸ Es existieren zahlreiche Initiativen auf nationalstaatlicher Ebene,

⁶ Vgl. z. B. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

⁷ Vgl. https://www.digitale-agenda.de/Webs/DA/DE/Home/home_node.html

⁸ Vgl. <https://www.gov.uk/government/news/g8-science-ministers-statement>

aber auch in einzelnen Bundesländern, zum Beispiel in Baden-Württemberg.⁹ Die Argumentation der Wissenschaftspolitik lautet: Öffentlich finanzierter Forschungsoutput ist per se als Allgemeingut zu betrachten und muss daher im Rahmen des Open Access für die Öffentlichkeit frei zugänglich sein, was besonders unter deutschen Professorinnen und Professoren höchst umstritten ist. In Thüringen ist Open Access Gegenstand des Koalitionsvertrags der 2014 zusammengetretenen rot-rot-grünen Landesregierung: „Gemäß dem Open Access-Ansatz sollen zukünftig insbesondere wissenschaftliche Informationen und wissenschaftlich erhobene Daten, die mit öffentlichen Geldern durch staatliche Stellen, Forschungseinrichtungen oder private Unternehmen gewonnen werden, der Allgemeinheit frei zur Verfügung gestellt werden, wenn nicht rechtliche Gründe dagegen sprechen“.¹⁰

Was tun wir in Thüringen?

Wie reagieren wir in Thüringen auf diese Herausforderung? Um diese Frage zu beantworten, lohnt zunächst ein Rückblick auf die Historie: So beauftragte der Thüringer Landeshochschulplan bereits im Jahr 2001 die wissenschaftlichen Bibliotheken des Freistaats, „auch multimediale Informationsquellen in ihr Sammelspektrum einzubeziehen“¹¹. Gleichzeitig stellte er fest: „Als Einzelinstitutionen (...) werden die wissenschaftlichen Bibliotheken Thüringens ihre Leistungen in diesem Bereich nicht aufrechterhalten oder gar ausüben können“¹².

Um diese Anforderung umzusetzen, konstituierte sich alsbald eine Projektgruppe, die sich vor allem aus Mitarbeiterinnen und Mitarbeitern der ThULB Jena und des Rechen- und Multimediazentrums der Friedrich-Schiller-Universität Jena (FSU) zusammensetzte: Auf der Basis von Miless und in

⁹ Vgl. <https://mwk.baden-wuerttemberg.de/de/forschung/forschungslandschaft/e-science/>

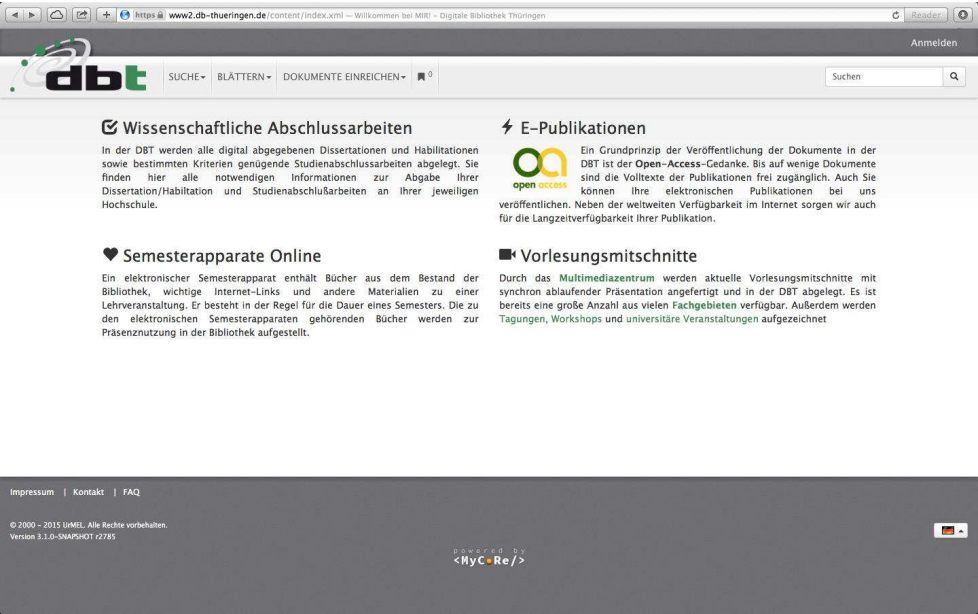
¹⁰ Der Koalitionsvertrag ist online beispielsweise über die Webpräsenz der Thüringer Grünen verfügbar: Thüringen gemeinsam voranbringen – demokratisch, sozial, ökologisch. Koalitionsvertrag zwischen den Parteien Die Linke, SPD, Bündnis 90/Die Grünen für die 6. Wahlperiode des Thüringer Landtags, 2014 (<http://gruene-thueringen.de/sites/gruene-thueringen.de/files/r2g-koalitionsvertrag-final.pdf>).

¹¹ Zitat gefunden bei Sabine Wefers: Thüringer Modell(e) für die Hochschulbibliotheken, in Zeitschrift für Bibliothekswesen und Bibliographie 63 (2016) H 1, S. 31.

¹² Ebenda.

Partnerschaft mit der Universität Duisburg-Essen entstand so die erste Version der DBT, zunächst als klassischer Hochschulschriftenserver.

Seither haben sich weitere Einrichtungen gefunden, welche die DBT als Repositorium nutzen: die Universitätsbibliothek der Technischen Universität Ilmenau, die Universitäts- und Forschungsbibliothek Erfurt/Gotha, die Ernst-Abbe-Hochschule Jena sowie weitere Hochschul- und Wissenschaftseinrichtungen in Thüringen. Die UB Ilmenau ist von Anfang an nicht nur Anwenderin, sondern bringt sich auch in die Pflege und Weiterentwicklung der DBT ein. Insofern war es nur konsequent, wenn auch der Gesetzgeber bei der Novellierung des Thüringer Hochschulgesetzes 2006 mit Blick auf die Hochschulbibliotheken des Freistaats feststellte: „Sie fördern durch die Bereitstellung einer geeigneten Infrastruktur das elektronische Publizieren und den Aufbau digitaler Bibliotheken.“¹³



Webseite der DBT (www.db-thüringen.de)

¹³ §38 Abs. 1 ThürHG.

Bald nach der Implementierung der DBT wurden schnell jedoch auch deren Grenzen erkennbar. Das lag einerseits an der rasant fortschreitenden technischen Entwicklung und andererseits am Reichtum der in Thüringer Bibliotheken, Archiven und Museen verwahrten kulturellen Überlieferung. Vor diesem Hintergrund sah sich die ThULB Jena herausgefordert, nach neuen Lösungen zu suchen, und startete den Aufbau weiterer Repositorien in Ergänzung zur DBT: ab 2005 zunächst für digitale und digitalisierte Zeitschrifteninhalte¹⁴, ein Jahr später auch für historische Bestände und Sammlungen¹⁵. Diese ebenfalls MyCoRe-basierten Plattformen firmierten fortan unter dem Label Universal Multimedia Electronic Library (UrMEL)¹⁶ und werden ähnlich wie die DBT von zahlreichen Partnereinrichtungen nachgenutzt. Besonders der „Zeitschriftenserver“ der ThULB Jena, Journals@UrMEL, verdient hier Beachtung, da er neben der DBT einen zentralen Baustein innerhalb der Open Access-Aktivitäten der ThULB Jena bildet: Inzwischen nutzen etliche universitäre wie nichtuniversitäre Einrichtungen diese Plattform für die Veröffentlichung eigener Open Access-basierter Zeitschriften.

Doch auch die DBT blieb in ihrer Entwicklung keineswegs stehen, sondern erweiterte ihr Dienstleistungsspektrum in den Folgejahren deutlich. Hinzu kamen neue Dokumenttypen: so beispielsweise nichtsubskriptionspflichtige, elektronische Pflichtexemplare, zu deren Sammlung die ThULB Jena seit 2008 per Gesetz verpflichtet ist. Hinzu kamen ferner elektronische Semesterapparate, welche sich insbesondere bei den geistes- und sozialwissenschaftlichen Fachbereichen der FSU Jena großer Beliebtheit erfreuen und das klassische Dienstleistungsspektrum der DBT im Bereich des elektronischen Publizierens ergänzen. Heutzutage profitiert die DBT von den Vorzügen einer deutschlandweiten Entwicklercommunity im Kontext von MyCoRe, besonders auch von der Zusammenarbeit der ThULB Jena mit dem Repository Service (Reposi) der Verbundzentrale des Gemeinsamen Bibliotheksverbunds (VZG).¹⁷

¹⁴ <http://zs.thulb.uni-jena.de/content/below/index.xml>

¹⁵ http://archive.thulb.uni-jena.de/hisbest/templates/master/template_collections/index.xml

¹⁶ <http://www.urmel-dl.de/>

¹⁷ <https://www.gbv.de/Verbundzentrale/serviceangebote/reposis-repository-service>

Welche Inhalte bietet die DBT?

Vom Umfang ihrer digitalen Bestände her betrachtet, muss die DBT den Vergleich mit kommerziellen Produkten nicht scheuen, sondern kann sich im Gegenteil mit diesen durchaus messen: Aktuell stehen rund 20.000 Dokumente/Objekte zur Verfügung, welche in ihrer Gesamtheit den digitalen Sammlungsbestand der DBT umfassen (Stand: April 2016). Darin bildet sich die gesamte Bandbreite wissenschaftlicher Publikationstypen ab: angefangen von klassischen Hochschulschriften, insbesondere Dissertationen, über Zeitschriftenartikel, Forschungs- und Konferenzberichte, Arbeitspapiere und Präsentationen, bis hin zu multimedialen Vorlesungs- und Veranstaltungsmaterialien einschließlich Videomitschnitten.

Das umfangreichste Bestandssegment bilden Vorlesungs- und Veranstaltungsmaterialien (7.465). Hierunter befinden sich nicht nur freizugängliche Publikationen, sondern auch zugangsbeschränkte Materialien als Teil der Elektronischen Semesterapparate. Rund ein Fünftel des Gesamtbestands der DBT nehmen sodann klassische Hochschulschriften ein, also Dissertationen, einige Habilitationen, jedoch auch Master-, Bachelor- und sonstige Studienarbeiten (4.282). Hinzu kommen typische Open Access-Materialien aus dem Forschungsalltag, also Artikel, sowie sonstiges akademisches Kleinschrifttum, abgespeichert in der Regel als Textdokumente im PDF-Format (4.088). Abgerundet wird das Medienangebot um weitere Materialien akademischer Provenienz (2.873), beispielsweise amtliches Schrifttum aus den Hochschulen, sowie audiovisuelle Medienangebote (1.683).

Am häufigsten werden in der DBT Interessierte der technisch-naturwissenschaftlichen sowie medizinischen Fachbereiche fündig (6.092). Dieser fachliche Schwerpunkt überrascht nicht weiter, da er nur den allgemeinen Trend im Bereich des elektronischen Publizierens widerspiegelt. Auf dem zweiten Platz folgen die Sozialwissenschaften, deren Anteil im E-Medien-Bereich allgemein stark anzieht (2.310), während das geisteswissenschaftlich interessierte Publikum hier ganz klar das Nachsehen hat (653). Institutionell lässt sich der Großteil des Sammlungsbestands der ThULB bzw. der FSU Jena zuordnen (12.785), gefolgt von der TU Ilmenau (4.316), der

Universität Erfurt (1.363) sowie weiteren Thüringer Hochschulen (192) und diversen Kultur-, Wissenschafts- und behördlichen Einrichtungen (1.385).



DBT mit responsive Design auf dem Smartphone

Erfreulich ist auch der hohe Grad an Akzeptanz der DBT seitens der Nutzerinnen und Nutzer weltweit. Dies spiegelt sich in einer entsprechend starken Nutzungsfrequenz wider – mit jährlichen Zugriffsraten im mittleren sechsstelligen Bereich. Ein überaus positiver Befund, wenn dieser auch differenziert zu bewerten ist, da die starke Nutzung im Wesentlichen auf die Lehrmaterialien zurückzuführen ist, leider weniger auf die eigentlichen E-Publikationen. Auch im Vergleich mit der Nutzung digitalisierter Sammlungsbestände aus dem Bereich der kulturellen Überlieferung (historische Bestände) kann sich die DBT durchaus sehen lassen und schneidet hier durchaus günstiger ab.

Wie gelangen die Inhalte in die DBT?

Die Feststellung, dass elektronisches Publizieren ein hochkomplexes Verfahren ist, gilt umso mehr, wenn man auch die technisch-administrativen Prozesse in Betracht zieht. Schon für den Ingest der Dokumente/Objekte werden je nach Anforderung unterschiedliche Workflows bedient. Zum einen können Wissenschaftlerinnen und Wissenschaftler ihre Materialien im Selfpublishing-Verfahren bequem von ihrem Arbeitsplatz in die DBT hochladen – solche Dokumente/Objekte werden zunächst von der jeweils zuständigen Bibliothek geprüft, gegebenenfalls um Metadaten ergänzt und anschließend zur Veröffentlichung freigeschaltet. Zum anderen existieren gesonderte Geschäftsgänge für Materialien, welche von der Bibliothek gleichsam eigenständig erworben werden und einer tiefergehenden fachqualifizierten Medienbearbeitung und Erschließung bedürfen. Das gilt beispielsweise für elektronische Dissertationen oder Pflichtexemplare, welche zudem höhere Anforderungen an die Langzeitarchivierung stellen. Dabei kann zukünftig auch auf SWORD v2¹⁸ als Schnittstelle für den standardisierten technischen Ablauf bei der Aufnahme der Sammlungsobjekte zurückgegriffen werden.

Der Nachweis der Materialien findet zunächst in der Verbunddatenbank des GBV statt. Damit ist gewährleistet, dass Titelinformationen sowohl im lokalen Nachweissystem der jeweiligen Bibliothek als auch im Verbundrahmen und darüber hinaus in übergreifenden Informations- und Recherchesystemen erfasst sind. Weiterhin erhalten die digitalen Objekte im Rahmen der Medienbearbeitung eine Notation der Dewey Decimal Classification (DDC) als Grundlage für fachbasierte Navigationsmöglichkeiten.

Wie werden die Inhalte auffindbar gemacht?

Sind Metadaten in der Verbunddatenbank erst einmal erstellt, erfolgt in einem nächsten Schritt deren Import in die DBT, wo sie mit den zugehörigen digitalen Dokumente/Objekte verknüpft werden. Hier erhält jedes Objekt auch seinen Persistent Identifier, einen Handle, welcher in die Katalogisate (sowohl in der DBT als auch im GVK) eingebettet wird: Sammlungsobjekte sind nun dauerhaft

¹⁸ <http://swordapp.org/>

zitier- und referenzierbar und erfüllen damit ein wesentliches Anforderungskriterium für das elektronische Publizieren.¹⁹ Um die Inhalte aus der DBT optimal auffindbar zu machen, sollen Metadaten sowie Volltexte, soweit technisch und rechtlich umsetzbar, perspektivisch in den Suchmaschinenindex GBV Zentral (Findex) sowie in Indexe von Anbietern kommerzieller Discoverysysteme integriert werden. Zum Auffinden in Google (Scholar) sind Dokumente/Objekte der DBT zusätzlich per Highwire Press Tag suchmaschinenoptimiert aufbereitet. Darüber hinaus unterstützt die DBT das Open Access Initiative Protocol for Metadata Harvesting (OAI-PMH)²⁰, und erreicht damit, dass Metadaten aus der DBT von übergreifenden Portalen „geharvested“ und in diesen aufgefunden werden können. Auf jeden Fall ist gewährleistet, dass die DBT den Vorgaben der europäischen Initiative OpenAIRE²¹ entspricht und Daten aus Thüringen per OAI-PMH auch in BASE²² einfließen.

Daten und Prozesse der DBT müssen also auf die Anforderungen einer zusehends interoperablen Umwelt reagieren. Voraussetzung für den möglichst barrierefreien Datenflow ist die Anwendung international gängiger Normen und Standards. Was die Wahl des technischen Formats für die Kodierung der Metadaten betrifft, so setzt die neue DBT auf das Metadata Object Description Schema (MODS)²³, welches von der Library of Congress entwickelt wurde. Auch Normdaten aus der Gemeinsamen Normdatei (GND)²⁴ der Deutschen Nationalbibliothek (DNB) sind in der DBT verfügbar und stellen im Rahmen der Workflows genormtes Vokabular zur Verfügung. Dank MODS nähert sich die DBT den übrigen UrMEL/MyCoRe-Anwendungen der ThULB Jena an, welche ebenfalls auf dem MODS-Standard basieren. Durch das einheitliche Metadatenmanagement lassen sich fortan Synergien sowie die Vorteile der betriebsorganisatorischen Matrix-Struktur der ThULB Jena noch wirkungsvoller ausschöpfen.

¹⁹ Die DBT unterstützt URNs, neuerdings auch DOIs.

²⁰ <https://www.openarchives.org/pmh/>

²¹ <https://www.openaire.eu>

²² <https://www.base-search.net/about/en/>

²³ <http://www.loc.gov/standards/mods/>

²⁴ http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html

Entwicklungsperspektiven

Perspektiven für die Weiterentwicklung der DBT bestehen für die folgenden Bereiche bzw. Arbeitsfelder:

- **Repositoriumsfunktion für E-Ressourcen:** In ihrer Rolle als klassischer Hochschulschriftenserver für die beteiligten Thüringer Hochschulen wird die DBT auf dem jeweils aktuellen Stand der Technik fortgeführt. Weiteres Entwicklungspotential besteht vor allem im Bereich des Open Access (green way). Hier gilt es, den digitalen Bestandsaufbau im Kontext des Zweitveröffentlichungsrechts voranzutreiben, um die Attraktivität der DBT im Sinne eines Schaufensters für den forschungsmäßigen Output Thüringer Hochschulen noch weiter zu stärken. Open Access-Mehrwertkomponenten im Rahmen bestehender Allianz-, Konsortial- und sonstiger Lizenzen sollen (soweit vorhanden) für den digitalen Bestandsaufbau stärker in Anspruch genommen werden. Für die Profilierung der DBT im Zusammenhang mit dem Open Access werden derzeit noch stärkere Akzente in der Vermittlung gesetzt einschließlich eines Akquise-, Rechte- und Embargomanagements. Dies setzt die Untersetzung mit konkreten Geschäftsgängen voraus, wie sie momentan an der ThULB Jena entwickelt und erprobt werden.
- **Forschungsdatenrepositorium:** Darüber hinaus ist die DBT als Repositorium für die Speicherung, Erschließung und Präsentation primärer Forschungsdaten gefragt. Aktuell wird an der ThULB Jena evaluiert, inwiefern etablierte Standards und Normen auch für den Bereich der Forschungsdaten genutzt werden können, beispielsweise MODS für das Metadatenmanagement oder Persistent Identifier für die dauerhafte Referenzierung von Forschungsdaten. Weiterhin wird nach Lösungen gesucht, in welcher Form primäre Datenbestände in publizierten Materialien, also in Artikeln oder sonstige Veröffentlichungen, einfließen können. Hier sieht sich die Bibliothek einmal mehr als Kompetenzpartner der Forschung gefordert, indem sie ihr fachliches, technisches, juristisches und bibliothekarisches Know-how als Serviceleistung zur Verfügung stellt.
- **Digital Humanities:** Da zusehends auch die geistes- und kulturwissenschaftlichen Fachbereiche von den neuen digitalen Angeboten profitieren und diese in stetig wachsendem Maß in Forschung und Lehre einbinden, sind diese ebenfalls zu berücksichtigen. Das bezieht sich zum einen auf die spartenübergreifende Bereitstellung digitalisierter Sammlungen außerhalb der DBT, zum anderen auf die Entwicklung entsprechender Tools, um digitalisierte historische Sammlungen stärker in den Forschungsbetrieb zu integrieren. Hier ist noch viel Platz für künftige Entwicklungen. Ein erster Schritt ist damit getan, dass mit der neuen

Version der DBT innovative Viewertechnologie zur Verfügung steht, welche sich auch für Editionsprojekte nutzbar machen lässt.

- **Altmetrics:** Sodann gilt es, die Bemühungen im Hinblick auf das Thema Altmetrics voranzutreiben. Es sollen die technischen Voraussetzungen geschaffen werden, um alternative Metriken für die Ermittlung bibliometrischer Kennzahlen in die DBT zu implementieren und damit Wissenschaftlerinnen und Wissenschaftlern Messinstrumente hinsichtlich der Bewertung ihrer Forschungsergebnisse an die Hand zu geben. Altmetrics stellen eine alternative Methode zur Ermittlung traditioneller bibliometrischer Kennzahlen wie dem Impact Factor und dem h-Index dar. Während letztere sich in erster Linie auf Zitierungen beziehen, fließen unter dem Stichwort „Altmetrics“ Informationen zur Nutzungshäufigkeit, der Verlinkung auf Websites von Autorinnen und Autoren oder Institutionen, der Rezeption in sozialen Netzwerken oder der Referenzierung in webbasierten Literaturmanagementsystemen (z. B. Mendeley) zusammen.

Abschließend muss der Korrektheit halber darauf hingewiesen werden, dass die Zukunft der DBT derzeit Verhandlungsgegenstand im Rahmen der Reorganisation der Thüringer Hochschulbibliotheken ist.

Autorenidentifikation als (neue) Dienstleistung für wissenschaftliche Spezialbibliotheken

Annette Polly¹, Ariane Streicher²

¹ Max Rubner-Institut Karlsruhe, Bibliothek

² HeBIS-Verbundzentrale Frankfurt am Main

Zusammenfassung

Wissenschaftler publizieren während ihrer wissenschaftlichen Laufbahn in unterschiedlichen beruflichen Kontexten und für unterschiedliche Einrichtungen. Im Zuge der Internationalisierung von Wissenschaft und Forschung gewinnt die eindeutige Identifikation ihrer Person und die eindeutige Zuordnung ihrer Werke eine immer größere Bedeutung.

Derzeit pflegen Wissenschaftler mehrere ID's, weil sie proprietär und nicht in jedem Fall international verbreitet sind. Dies steht dem Konzept einer eindeutigen, weltweit anerkannten ID derzeit noch entgegen. Seit einigen Jahren etabliert sich ORCID (Open Researcher and Contributor ID) als neue Möglichkeit, Personen eindeutig zu identifizieren und mit ihren Werken zu verknüpfen.

Die vorliegende Arbeit leistet einen Beitrag zur aktuellen Diskussion um Autorenidentifikation als Aufgabe einer Spezialbibliothek, indem sie einen Bogen zwischen der Perspektive der Wissenschaftler und der Perspektive der Bibliotheken spannt. Sie liefert ein Beispiel dafür, dass sich für Bibliotheken neue und interessante Themengebiete eröffnen, um ihren Wissenschaftlern nützliche Dienstleistungen anzubieten.

Abstract

Author ID as a new area of responsibility for research libraries

During their academic careers scientists publish in different professional contexts and for different institutions. As part of the internationalization of science and research, a unique ID (identification of the individual) and a reliable attribution of the research work is becoming increasingly important.

Currently scientists maintain multiple ID's as the latter are proprietary and not spread internationally in each case. This is currently opposed to the concept of a unique, internationally recognized ID. For several years, the ORCID (Open Researcher and Contributors ID) has been established as a new way to clearly identify individuals and link them with their research work.

This paper contributes to the current debate on author ID regarding the task of a research library, namely by covering the spectrum between the perspective of scientists and the perspective of libraries. It provides an example that will open up for libraries new and interesting topics to offer useful services to their scientists.

Keywords: Autorenidentifikation, wissenschaftliche Spezialbibliothek, author identification, research library

Einleitung

Wissenschaftler¹ veröffentlichen während ihrer wissenschaftlichen Laufbahn in unterschiedlichen Kontexten und für unterschiedliche Einrichtungen. Sie publizieren unterschiedliche Werke, neben Zeitschriftenartikeln und Monographien auch Forschungsdaten, Konferenzbeiträge, Blogbeiträge und dergleichen mehr. „Die Wissenschaftsnetzwerk, die Wissenschaftler dringend für ihre gute Arbeit brauchen, findet, abgesehen von den Räumlichkeiten der eigenen Einrichtung, im Internet statt. Wissenschaftler, die in diesem Wissenschaftssystem arbeiten, profitieren von der normierten Ansetzung ihrer Personennamen“ (Streicher 2015, S. 6). Für Wissenschaftler besteht der Nutzen der eindeutigen Identifikation in der Unterscheidung gleichlautender Namensformen, der verbesserten Auffindbarkeit und Recherchierbarkeit im World Wide Web.

Für Bibliotheken stellt die Normierung unterschiedlicher Entitäten eine etablierte Aufgabe dar. Insbesondere bei Personennamen werden große Anstrengungen geleistet, um die Normdatensätze in den Bibliothekskatalogen zu individualisieren und korrekte Verknüpfungen zu den Werken der Autoren herzustellen.

Seit einigen Jahren etabliert sich ORCID (Open Researcher and Contributor ID) als neue Möglichkeit, Personen eindeutig zu identifizieren und mit ihren Werken zu verknüpfen.² Aktuell startet ein DFG-Projekt, das die Einführung von ORCID in Deutschland vorantreiben soll.³

Die vorliegende Arbeit leistet einen Beitrag zur aktuellen Diskussion um Autoridentifikation als Aufgabe einer Spezialbibliothek, indem sie einen Bogen zwischen der Perspektive der Wissenschaftler und der Perspektive der Bibliotheken spannt. Es wird der Frage nachgegangen, ob und unter welchen Bedingungen die Implementierung von ORCID gegenüber dem traditionellen Handling von Normdaten durch Bibliotheken einen Mehrwert für Wissenschaftler und Unterhaltsträger generiert. Gleichzeitig stellt die Arbeit eine Momentaufnahme aus Sicht einer wissenschaftlichen Spezialbibliothek dar.

¹ Im vorliegenden Aufsatz wird aus Gründen der Vereinfachung auf geschlechtsspezifische Formulierungen verzichtet und stets die männliche Formulierung verwendet. Bei allen Nennungen (Wissenschaftler, Bibliothekare etc.) sind stets Männer und Frauen gleichermaßen gemeint.

² <http://www.orcid.org>

³ <http://www.dini.de/projekte/orcid-de/>

Sie spiegelt die aktuelle Situation wider und weist auf die Chancen und Herausforderungen hin, die mit der Implementierung einer neuen Dienstleistung wie der Autorenidentifikation verbunden sein können.

Persistente Identifier für die eindeutige Identifizierung von Autoren

Das Deutsche Institut für Normung (DIN) definiert den persistenten Identifier (PI) als einen weltweit eindeutigen Namen, "der zur persistenten Identifizierung adressierbarer Objekte, wie zum Beispiel Dokumente, Bilder, Tonaufnahmen, Animationen oder Metabeschreibungen eingesetzt wird.[...] Im Grunde kann aber ein PI jegliche Art von Entitäten identifizieren, z.B. Personen [...]“ (DIN 31464 2013, S. 4)

Im Zuge der Internationalisierung von Wissenschaft und Forschung gewinnen die eineindeutige Identifikation der Person eines publizierenden Wissenschaftlers sowie die eindeutige Zuordnung zu seinen Werken eine immer größere Bedeutung. Der digitale Wandel, die verstärkte Evaluierung der Wissenschaftler und die damit verbundene Vergabe von Projektmitteln tragen ebenfalls dazu bei. Gleichzeitig wird die eindeutige Identifikation von einzelnen Personen durch dublette Namenseinträge immer komplizierter.

Wissenschaftler ergreifen unterschiedliche Maßnahmen, um sich eindeutig zu identifizieren und dublette Namenseinträge auszuschließen. Viele Forscher verwenden in ihren Publikationen mehrere Initialen, um sich von anderen Forschern gleichen Namens zu unterscheiden. Bei Namensänderungen können frühere Publikationen oftmals nicht mehr dem aktuellen Namen zugeordnet werden. Gleiches gilt bei Verwendung verschiedener Namensformen. Außerdem pflegen Autoren häufig mehrere Profile in interdisziplinären Zitationsdatenbanken, z.B. ResearcherID⁴ oder Scopus AuthorID⁵, aber auch in akademischen sozialen Netzwerken wie Research Gate⁶ oder Mendeley.⁷ Diese Profile werden Teil von proprietären Angeboten und sind auf die Anwendung in den Produkten der Anbieter beschränkt. Der Aufwand für die Pflege der Profile erfordert Zeit und Aufmerksamkeit der Wissenschaftler.

⁴ www.researcherid.com

⁵ www.elsevier.com/solutions/scopus/support/authorprofile

⁶ www.researchgate.net/

⁷ www.mendeley.com/

Doch nicht nur die Autoren selbst haben ein großes Interesse an einer eindeutigen Identifizierung ihrer Person und der Verknüpfung mit ihren Werken. Auch die Träger-einrichtungen sowie die Forschungsförderer verlangen regelmäßige Berichte über Forschungsaktivitäten und die Publikation der Ergebnisse. Diese Evaluationen be-ziehen sich z.B. auf eine ganze Einrichtung, auf deren Teilbereiche oder auf einrich-tungsübergreifende Projekte. Um dem steigenden Informationsbedarf und den hete-rogenen Anfragen zu begegnen, wurde vom Wissenschaftsrat im Januar 2016 der Kerndatensatz Forschung verabschiedet. Unter anderem wird die Verwendung "ein-deutiger Identifikatoren für alle Objekte des Kerndatensatz Forschung“ empfohlen. Neben anderen Identifikatoren wie der DOI, ISBN oder ISSN wird die ORCID als „einrichtungsübergreifender Identifier und offener Standard zur Zuordnung von Per-sonen“ ausdrücklich genannt. (Wissenschaftsrat 2016 S. 22)

Personennormdaten in Bibliotheken

Normierung unterschiedlicher Entitäten, z.B. von Personen und Körperschaften, stellt seit mehreren Jahrzehnten eine etablierte Aufgabe von Bibliotheken dar. Normdaten-sätze werden mit individualisierten Merkmalen in der Gemeinsamen Normdatei (GND) durch geschulte Bibliothekare erstellt und bei Bedarf angereichert. An der Normdatenerfassung in der GND sind Bibliothekare der sechs deutschen Bibliotheks-verbünde und die Deutsche Nationalbibliothek (DNB) kooperativ beteiligt. Für Per-sonen wird ein GND-Satz erstellt, sobald in einer deutschsprachigen Bibliothek eine selbstständig erschienene Publikation eines Autors vorliegt. Die eindeutige Identi-fizierung eines Autors, die die Erfassung eines personalisierten Normdatensatzes ermöglicht, ist mit identifizierenden Merkmalen wie z.B. dem Geburtsjahr gegeben. Dann werden Publikation und Normdatensatz verknüpft.

Auf internationaler Ebene kombiniert „VIAF™ (Virtual International Authority File) mehrere Normdateien in einem vom Online Computer Library Center (OCLC) ge-hosteten Normdatendienst. Ziel dieses Dienstes ist die Steigerung des Nutzens der bibliothekarischen Normdateien bei gleichzeitiger Kostenreduzierung, indem welt-weitweit genutzte Normdateien zusammengeführt, verlinkt und so die Daten im Web zugänglich gemacht werden.“ (VIAF 2016) Von allen teilnehmenden Bibliotheken ist die DNB größter Datenlieferant nach VIAF und lieferte bisher 7,7 Millionen Perso-nendatensätze. (OCLC 2014, S. 6-7)

Für viele wissenschaftliche Autoren erfolgt die Erfassung eines GND-Satzes im Rahmen der Pflichtablieferung ihrer Dissertation an die Bibliothek. Bei unselbständig erscheinenden Publikationen ist die Erfassung eines GND-Satzes nicht in allen Bibliotheken obligatorisch, was insbesondere für Nachwuchswissenschaftler von Nachteil sein kann. Sowohl die Auffindbarkeit als auch die Recherchierbarkeit der Autoren samt ihren unselbständigen Werken in deutschen Bibliothekskatalogen sind begrenzt. Weder die DNB noch VIAF erfüllen den Wunsch der Wissenschaftler, ihre Werke vollständig, sofort bei Erscheinen und ohne ihre manuelle Beteiligung eindeutig nachzuweisen.

Die Open Researcher and Contributor ID

Die Open Researcher and Contributor ID (ORCID) ist seit 2009 als gemeinnütziges Unternehmen am Markt mit dem Ziel, Autoren jeglicher Art das Anlegen eines individuellen Profils zu ermöglichen. Angesprochen werden damit Wissenschaftler, für die sich ORCID das Ziel gesetzt hat, das Problem der mehrdeutigen Personennamensansetzung in der wissenschaftlichen Kommunikation zu lösen. Darüber hinaus ermöglicht ORCID die Abbildung des gesamten wissenschaftlichen Lebenslaufs (Ausbildung, Arbeitsstellen, Projekte, Werke aller Art) eines Forschers und die eindeutige Verknüpfung mit seinen Werken. Unterschiedliche Namensformen einer Person können erfasst werden und sind indexiert.

Jeder Autor bzw. jeder Wissenschaftler kann für sich selbst einen eigenen Datensatz anlegen und sich damit eine persönliche, 16-stellige ORCID erstellen. Er ist jedoch nicht befugt, dies für andere Personen zu tun.⁸ Der Wissenschaftler selbst entscheidet, welche Angaben er in das Profil einpflegt und welche er für die Öffentlichkeit freigibt. „Jede Person, die im wissenschaftlichen Arbeitsprozess einen Beitrag leistet, kann sich über ihre ORCID mit ihren Publikationen, Forschungsdaten und anderen Produkten des Forschungsprozesses (zum Beispiel Software) eindeutig vernetzen. Damit werden diese Objekte sichtbar und technisch verlässlich mit ihren Erschafferrinnen und Erschaffern verbunden.“ (DINI 2016) ORCID ist als nicht-proprietäres System international anerkannt und wird u.a. durch Mitgliedschaften von Universitäten,

⁸ <http://www.support.orcid.org/knowledgebase/articles/188278-link-works-website-user>

Verlagen oder Forschungsförderern finanziert. Für Wissenschaftler ist die Pflege ihrer Profile kostenfrei. Für die weitergehende Nutzung von ORCID, z.B. die Pflege der Profile durch bibliothekarische Mitarbeiter oder den Import von Metadaten aus anderen Informationssystemen, fallen jährliche Mitgliedsgebühren an.⁹

Die Verbreitung der ORCID wird durch die Einbindung in die unterschiedlichen Systeme verschiedenster Informationsanbieter stark gefördert: So kann sich ein Autor beim Einreichen eines Manuskripts gegenüber dem Verlag mit seiner ORCID authentifizieren. Diese Angabe erscheint auch auf der Publikation. Im Nachgang kann die ORCID als Selektionskriterium in bibliographischen Datenbanken verwendet werden. Bislang waren für diese Vorgänge unterschiedliche Authentifizierungsmerkmale erforderlich.

Es stehen verschiedene Schnittstellen zu Informationsanbietern wie Web of Science, Scopus, DataCite oder CrossRef zur Verfügung, um Metadaten zu importieren.¹⁰ Die Kosten einer einzelnen Mitgliedschaft bei ORCID sind relativ hoch, so dass bislang das Forschungszentrum Jülich als einzige außeruniversitäre Forschungseinrichtung Mitglied bei ORCID war. (Bertelmann 2015, S. 5) Als erste Universitäten in Deutschland haben die TU Dortmund und die RUB Bochum ihre Mitgliedschaft in ORCID bekanntgegeben.¹¹

Im Frühjahr dieses Jahres wurde das DFG-Projekt „ORCID DE“ bewilligt. Ziel ist es, „die vielerorts erwogene Implementierung der ORCID an Hochschulen und außeruniversitären Forschungseinrichtungen durch einen übergreifenden Ansatz nachhaltig zu unterstützen. Dabei stehen organisatorische, technische und rechtliche Fragen gleichermaßen im Fokus. Neben der Schaffung einer zentralen Anlaufstelle für Hochschulen und außeruniversitäre Forschungseinrichtungen sind die Vernetzung und Verbreitung der ORCID im Bereich von Open-Access-Repositories und Publikationsdiensten im Kontext der Bielefeld Academic Search Engine (BASE) sowie die Verzahnung mit der Gemeinsame Normdatei (GND) wesentliche Aspekte des Projekts“. (DINI 2016)

⁹ <http://www.orcid.org/about/membership>

¹⁰ <https://orcid.org/organizations/integrators/current>

¹¹ http://www.tu-dortmund.de/uni/Uni/aktuelles/meldungen/2016-04/16-04-19_orcid/index.html

Die Prüfung eines möglichen nationalen ORCID-Konsortiums, wie es in Großbritannien¹² oder Australien¹³ bereits realisiert wird, steht erst am Ende der dreijährigen Laufzeit, die am 1. Mai 2016 beginnt. (Bertelmann 2015)

Wissenschaftliche Spezialbibliotheken außeruniversitärer Forschungseinrichtungen

Im deutschen Wissenschaftssystem arbeiten Wissenschaftler an den Einrichtungen des Wissenschaftssystems, an den Hochschulen und außeruniversitären Forschungseinrichtungen. Sie kooperieren und vernetzen sich mit ihrer Arbeit und stellen so Wissen sicher, das zur Entwicklung und weitergehend zum Beispiel bei der Produktion dieser Entwicklungen genutzt werden kann. Für eine erfolgreiche Arbeit nutzen Wissenschaftler die Forschungs- und Informationsinfrastruktur ihrer Einrichtungen, was ihnen ein effizientes Arbeiten garantiert und die Leistungsfähigkeit sichert. Sie müssen für ihre jeweiligen Bedarfe die geeigneten Rahmenbedingungen vorfinden, um effizient und engagiert arbeiten zu können. (Wissenschaftsrat 2013)

Im Vergleich zu Hochschulbibliotheken stellt in wissenschaftlichen Spezialbibliotheken die Versorgung der dort wissenschaftlich Tätigen einen zentralen Aufgabenschwerpunkt dar. Ihr Vorteil liegt in den kurzen Wegen zu ihren Kollegen sowie in der Nähe zur Leitungsebene der Einrichtung. Je nach Größe der Trägereinrichtung verfügen die Bibliotheken häufig über geringe personelle Ressourcen. Die Herausforderungen, die sie zu bewältigen haben, sind dennoch genauso komplex wie diejenigen von Universitätsbibliotheken, denn die Erwartungen ihrer Nutzer sind die gleichen.

¹² <http://www.orcid.org/blog/2015/06/23/national-consortium-orcid-set-improve-uk-research-visibility-and-collaboration>

¹³ <http://www.orcid.org/blog/2016/02/19/australian-orcid-consortium-officially-launched>

Autorenidentifikation als Dienstleistung wissenschaftlicher Spezialbibliotheken

Für wissenschaftliche Spezialbibliotheken außeruniversitärer Forschungseinrichtungen stellt sich die Frage, ob überhaupt und unter welchen Bedingungen die Autorenidentifikation über die rein bibliothekarische Anwendung in der GND hinaus als Dienstleistung für die Wissenschaftler der Einrichtungen angeboten werden kann.

Grundlegende Voraussetzung ist die Zustimmung der Wissenschaftler zur Einrichtung ihres Profils, falls dieses noch nicht vorhanden ist. Als mögliche Hindernisse gelten einerseits Bedenken des Datenschutzes und andererseits die Befürchtung der Wissenschaftler, noch mehr Verwaltungsarbeit bewältigen zu müssen. Diesen Befürchtungen kann jedoch begegnet werden, indem einerseits auf die o.g. Privacy Policy von ORCID hingewiesen wird, nach der jeder Wissenschaftler selbst entscheidet, welche Daten er freigibt. Zur Erstellung eines Profils genügen die Angabe des Namens und einer E-Mail-Adresse. Darüber hinaus soll der Verwaltungsaufwand insgesamt minimiert werden, indem in der Folge die ORCID als Identifikationsmerkmal z.B. bei der Einreichung einer Publikation bei einem Verlag oder bei der Beantragung von Projekten bei Forschungsförderern verwendet wird. Manche Verlage gehen derzeit dazu über, die Angabe einer ORCID obligatorisch zu erwarten, wodurch sowohl das Bewusstsein als auch die Zustimmung der Autoren zu ORCID deutlich ansteigt.¹⁴

Wegen der begrenzten personellen Ressourcen besteht auch für die Bibliotheken ein wesentliches Ziel darin, von Verwaltungsarbeit soweit wie möglich entlastet zu werden. Deshalb ist es wichtig, die Daten nur ein einziges Mal und konsistent zu erfassen. Die Pflege der Profile sollte soweit wie möglich automatisiert erfolgen. Die manuelle Aktualisierung von Profilen durch Mitarbeiter der Bibliothek ist ORCID-Mitgliedern vorbehalten. Die Umsetzung kann in der Regel nicht von den Bibliotheken alleine geleistet werden, sondern stellt eine Querschnittsaufgabe dar, die nur in Zusammenarbeit mit der Verwaltung sowie den IT- und Forschungsreferaten der jeweiligen Forschungseinrichtungen gemeinsam gelingen kann.

¹⁴ <http://www.orcid.org/blog/2016/01/07/publishers-start-requiring-orcid-ids>

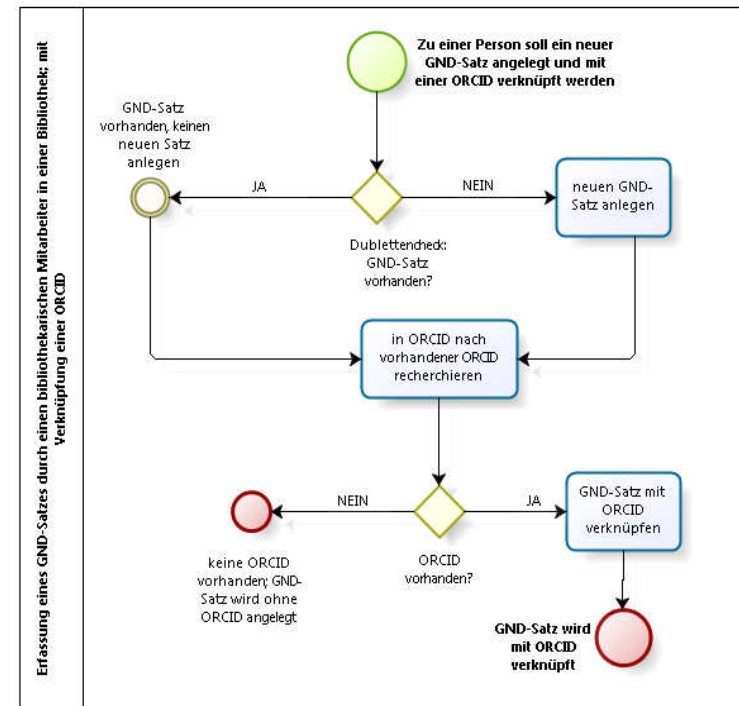
Szenario 1: Anlegen und Verknüpfen ORCID GND durch die Bibliothek

Die Beschreibung der Funktionalitäten von GND und ORCID legt die Schlussfolgerung nahe, dass eine gemeinsame Verwendung von GND und ORCID sowohl im Sinne der Bibliotheken als auch der Wissenschaftler ist. Die normierte Ansetzung von Personen in der GND durch Bibliothekare in Verbindung mit einem stets aktuell gehaltenen ORCID-Profil durch die Autoren würde zwei Identifier bündeln. Automatisierte, technische Abgleichmechanismen würden die Forderung berücksichtigen, die Verwaltungsarbeit soweit wie möglich zu reduzieren. Davon ausgehend werden drei modellhafte Entwürfe beschrieben, die aufzeigen, wie Bibliotheken, Wissenschaftler und auch Verlage bei der Erfassung und weiteren Verwendung persistenter Identifier für Personennormdaten gut miteinander kooperieren können. Wegen der Kosten der Mitgliedschaft werden die Beispiele für Aktivitäten differenziert erläutert, je nachdem, ob sie auch ohne eine Mitgliedschaft realisierbar sind.

Im ersten Prozess (Abb. 1) wird die Aktivität durch einen bibliothekarischen Mitarbeiter (im weiteren Verlauf als Bibliothekar bezeichnet) ausgelöst. Ein Bibliothekar möchte für einen Autor einen neuen GND-Satz erstellen.

Zu Beginn der Aktivität führt er einen Dublettencheck durch, um sicherzugehen, dass für den Autor noch kein GND-Satz existiert. Nach negativem Dublettencheck wird ein neuer GND-Satz für den Autor erstellt. Anschließend erfolgt die Recherche nach einer vorhandenen ORCID. Sofern eine passende ORCID vorhanden ist, wird diese in einem weiteren Schritt mit dem GND-Satz verknüpft. Existiert keine ORCID, erfolgt keine Verknüpfung.

Äquivalent verläuft die Erfassung einer ORCID, wenn der Dublettencheck positiv ist, wenn also bereits ein GND-Satz existiert. Es erfolgt ebenfalls die Recherche nach einer vorhandenen ORCID und eine Verknüpfung der ID mit dem bereits vorhandenen GND-Satz.



Powered by
bizagi
Modeler

Abbildung 1: Erfassung eines GND-Satzes durch einen bibliothekarischen Mitarbeiter in einer Bibliothek; mit ORCID-Verknüpfung

Die Mitgliedschaft bei ORCID erlaubt der Bibliothek die Pflege von ORCID-Profilen ihrer Wissenschaftler. Die Verantwortung für Inhalt und Aktualität des Profils liegt beim Wissenschaftler. Verlässt oder wechselt er die Einrichtung, nimmt er das Profil mit. Je mehr Einrichtungen dieses Vorgehen unterstützen, desto mehr wird die einrichtungsübergreifende Zusammenarbeit gefördert. Der Aufwand für die einzelne Einrichtung sinkt. Der Prozess gesamte Prozess der Verknüpfung von GND-Nummer und ORCID wird in jedem Fall durch den Bibliothekar abgeschlossen.

Szenario 2: Anlegen und Verknüpfen ORCID GND durch Wissenschaftler

Abb. 2 zeigt den zweiten Prozess, der deutlich umfangreicher ist. Ein Autor legt zunächst eine neue ORCID für sich an und erfasst seine Publikationen. Zur Verknüpfung der ORCID mit einem GND-Satz ruft der Autor seine GND-Nummer auf, sofern er eine besitzt und diese ihm bekannt ist. Der GND-Satz kann dann mit der ORCID verknüpft werden. Existiert für den Autor noch kein GND-Satz, hat er die Möglichkeit, einen GND-Satz für sich selbst bei seiner Bibliothek zu beantragen. Anhand seiner dafür erforderlichen Angaben wie dem Namen erfolgt ein Dublettencheck, vergleichbar mit dem Dublettencheck im ersten Prozess. Ergibt dieser, dass bereits ein GND-Satz vorhanden ist, wird dem Autor der Satz angeboten. Dieser bestätigt daraufhin, dass die Angaben im GND-Satz seinen persönlichen Angaben entsprechen. Daraufhin kann der Autor den GND-Satz mit der ORCID verknüpfen. Sollte der Autor den dubletten GND-Satz nicht als seinen akzeptieren, wird für ihn ein neuer GND-Satz angelegt. Der korrekte bzw. der korrigierte GND-Satz kann anschließend vom Autor mit der ORCID verknüpft werden.

Die automatisierte Anreicherung des GND-Satzes mit ORCID-Daten ist technisch grundsätzlich möglich, sofern die ORCID-Daten für die Öffentlichkeit freigegeben werden. Im Rahmen des DFG-Projekts ORCID DE ist die Realisierung einer solchen Schnittstelle vorgesehen.(Bertelmann 2015, S. 15) Für die GND bedeutet dieses Vorgehen einen signifikanten Beitrag zur Verbesserung der Datenqualität, da auch ohne Vorliegen selbständig erschienener Publikationen Personennormsätze individualisiert und aktualisiert werden.

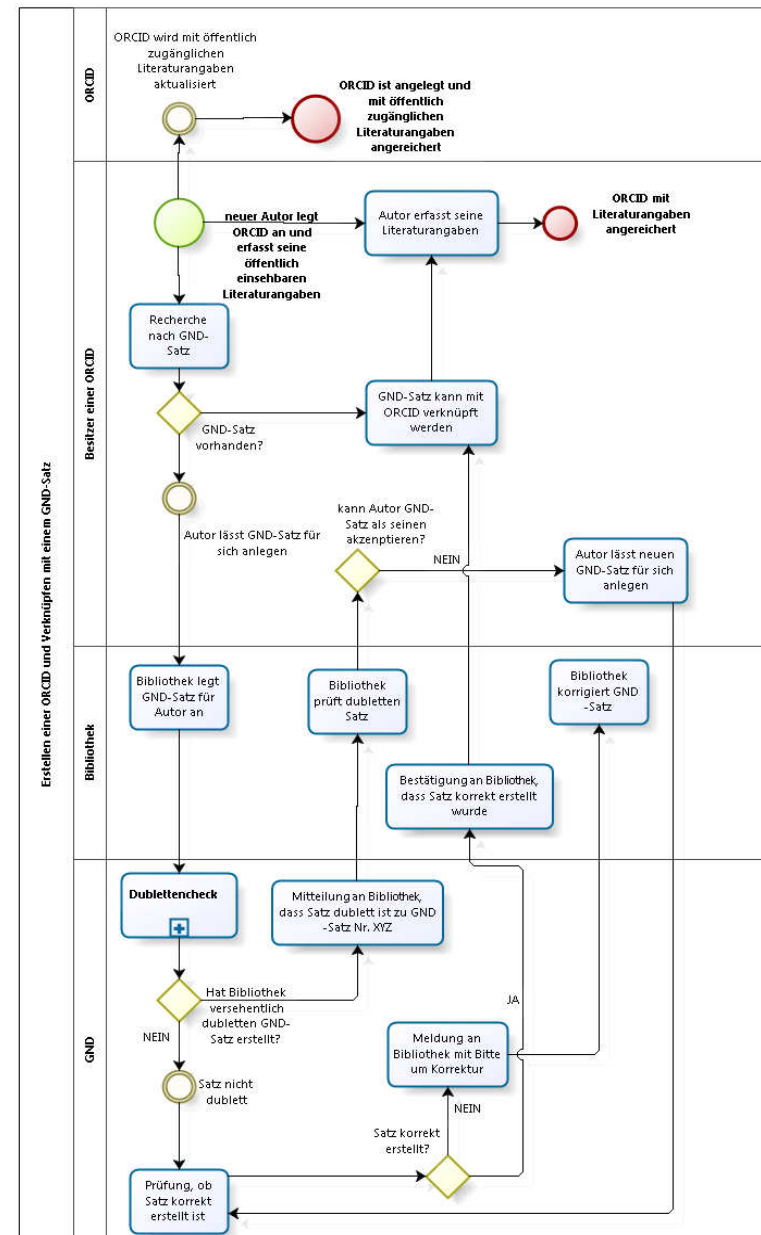


Abbildung 2: Erstellen einer ORCID und Verknüpfen mit GND-Satz

Szenario 3: Automatische Aktualisierung eines ORCID-Eintrags durch einen Verlag

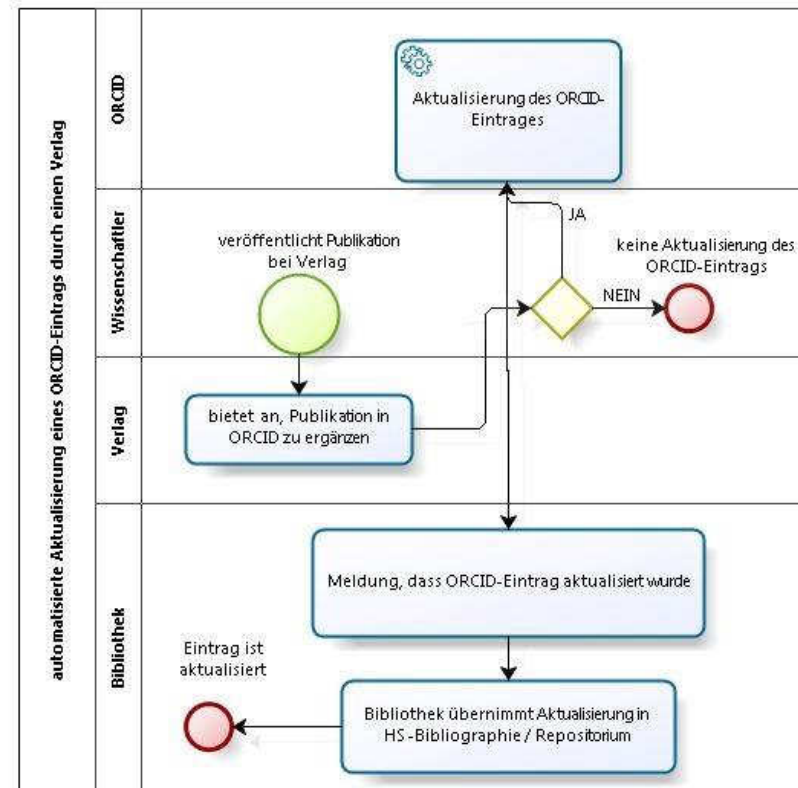


Abbildung 3: Automatische Aktualisierung eines ORCID-Eintrags durch einen Verlag

Der dritte Prozess (Abb. 3) zeigt eine mögliche Kooperation zwischen Wissenschaftlern, ORCID, Verlagen und der Bibliothek mit einem durch sie bereitgestellten Repositoryum. Immer mehr Verlage bieten Wissenschaftlern an, sich auf der verlagseigenen Homepage mit der ORCID anstelle einer proprietären ID zu registrieren. Dann ist auch ein automatischer Abgleich zwischen den vom Autor auf der Verlagshomepage hochgeladenen Publikationen und ORCID möglich. Damit kann der Autor einer automatisierten Aktualisierung des ORCID-Accounts zustimmen und muss Publikationen nicht mehr manuell erfassen. Ebenso wäre es möglich, dass aktualisierte ORCID-Einträge an ein Repositoryum gemeldet werden. Die Meldung neuer Publikationen erfolgt demnach zweimal: vom Verlag zu ORCID und von ORCID zum Repositoryum.

Dies stellt sowohl für den Wissenschaftler als auch für die Bibliothek als Betreiberin des Repositoriums einen eklatanten Mehrwert dar, der eine Mitgliedschaft der Bibliothek in ORCID rechtfertigen würde. Dies garantiert die eingangs beschriebene Forderung nach automatisierten Arbeitsabläufen.

Für Publikationen, die in keinem der Informationssysteme erfasst werden, ist eine automatisierte Aktualisierung des ORCID-Profiles nur bei Mitgliedschaft möglich¹⁵.

Ohne die Mitgliedschaft liegt der Verwaltungsaufwand beim Wissenschaftler selbst.

Fazit

Die eindeutige Zuordnung und der Nachweis von Veröffentlichungen zu ihren Autoren gelten als eine der klassischen Aufgaben von Bibliotheken. Doch darüber hinaus besteht die Möglichkeit, die Pflege von Profilen als eine Dienstleistung auf hohem Niveau für Bibliotheken zu etablieren. Für die Bibliothek besteht der Mehrwert im engen Kontakt mit den Wissenschaftlern, die von kleinteiliger Verwaltungsarbeit entlastet werden. Der Nutzen liegt sowohl bei den Wissenschaftlern als auch bei den Einrichtungen, womit wiederum ein Beitrag zur Legitimation der Bibliothek geleistet wird.

Darüber hinaus ist eine ORCID für Institutionen im Sinne eines nicht-proprietären Identifiers sinnvoll. Nach wie vor werden Evaluationen ganzer Institutionen auf Ebene aller Publikationen der betreffenden Institutionen durchgeführt. Beispiele hierfür sind das Leiden Ranking¹⁶ oder Journals & Highly Cited Data¹⁷, die beide auf Auswertungen des Web of Science basieren. Dafür sind korrekte und stets aktuelle Ansetzungen nötig. Bibliotheken erfassen zwar GND-Sätze für Körperschaften – diese sind auch eindeutig und dank bibliothekarischer Gründlichkeit werden die GND-Sätze regelmäßig bei Bedarf aktualisiert (es erfolgen beispielsweise vorher-nachher-Verknüpfungen oder Namensänderungen von Körperschaften). Aber erstens erfolgt die Verlinkung von Publikationen mit dem GND-Satz nur bei selbständigen Publikationen. Zweitens garantiert nur eine normierte Ansetzung der Körperschaft im Web of Science eine korrekte Auswertung. GND-Körperschaftsansetzungen gibt es aber nicht im Web of Science.

¹⁵ <https://members.orcid.org/repositories>

¹⁶ <http://www.leidenranking.com/>

¹⁷ <http://ip-science.thomsonreuters.com/jhcd-fb/>

Fazit der Autorenidentifikation für wissenschaftliche Spezialbibliotheken ist deshalb: Dieser Service lohnt sich für die Bibliothek der Forschungseinrichtung am ehesten, wenn die Bibliothek nicht nur Autoren normiert erfasst, sondern auch die damit verbundene Affiliation, auf der das Institutionenranking basiert. Der Mehrwert durch die automatisierten Verknüpfungen kommt letztlich allen Beteiligten zugute. Es bleibt zu wünschen, dass ein nationales ORCID-Konsortium möglichst bald zustande kommt, so dass für die teilnehmenden Einrichtungen die Kosten der Mitgliedschaft deutlich sinken.

Literatur

Alle Links wurden zuletzt am 29.4.2016 geprüft.

[Bertelmann 2015] Bertelmann, R., Niggemann, E., Pieper, D., Elger, K., Fenner, M., Hartmann, S., Höhnnow, T., Jahn, N., Müller, U., Pampel, H., Schirrwagen, J., Summann, F. (2015): ORCID DE – Förderung der Open Researcher and Contributor ID in Deutschland, 24 p. DOI: <http://doi.org/10.2312/lis.16.01>

[DIN 31464] Normenausschuss Bibliotheks- und Dokumentationswesen (NABD) im DIN: Information und Dokumentation – Anforderungen an die langfristige Handhabung persistenter Identifikatoren (Persistent Identifier). DIN 31646:2013-01. Berlin : DIN, 2013.

[DINI 2016] Deutsche Initiative für Netzwerkinformation: ORCID DE – Förderung der Open Researcher and Contributor ID in Deutschland. Elektronisch verfügbar unter: <http://www.dini.de/projekte/orcid-de>

[ISO 2013] Information technology -- Object Management Group Business Process Model and Notation. ISO 19510, S. 1. Elektronisch verfügbar unter: https://webstore.iec.ch/preview/info_isoiec19510%7Bed1.0%7Den.pdf

[OCLC 2014] 2014 Annual Report to VIAF Council. Elektronisch verfügbar unter <http://www.oclc.org/content/dam/oclc/viaf/OCLC-2014-VIAF-Annual-Report-to-VIAF-Council.pdf>

[Streicher 2015] Streicher, A.: Weltweite Personen-Identifizierung in deutschen Bibliotheken: mögliche Szenarien ihrer Integration und Nutzung. – Stuttgart: Hochschule der Medien, 2015. – unveröff. Masterarbeit.

[VIAF 2016] Virtual International Authority File. Elektronisch verfügbar unter: <http://www.viaf.org/>

[Wissenschaftsrat 2013] Perspektiven des deutschen Wissenschaftssystems. Drs. 3228-13 vom 12.07.2013. Elektronisch verfügbar unter: <http://www.wissenschaftsrat.de/download/archiv/3228-13.pdf>

[Wissenschaftsrat 2016] Empfehlungen zur Spezifikation des Kerndatensatz Forschung. Drs. 5066-16 vom 22.01.2016. Elektronisch verfügbar unter: <http://www.wissenschaftsrat.de/download/archiv/5066-16.pdf>

Recommender-System für Projektkollaborationen basierend auf wissenschaftlichen Publikationen und Patenten

Christoph Quix ^{1,2}, Sandra Geisler ¹, Rihan Hai ¹

¹ Informatik 5, RWTH Aachen University

² Fraunhofer Institut für Angewandte Informationstechnik FIT

Kurzfassung

Die erfolgreiche Durchführung von Entwicklungs- und Forschungsprojekten hängt von vielen Faktoren ab. Innovationspotential und Zukunftsorientierung helfen bei der Antragsbewilligung. Doch genauso wichtig ist die Zusammensetzung des Projektteams. Insbesondere bei interdisziplinären Projekten ist man auf ein Team angewiesen, das aus hervorragenden Experten der jeweiligen Teilgebiete besteht. Die Medizintechnik ist ein gutes Beispiel für ein sehr innovatives und gleichzeitig hoch interdisziplinäres Feld. Aber gerade die Interdisziplinarität macht die Suche nach Experten schwierig und langwierig, da man sich erst in fremden Domänen zurechtfinden muss und eventuell nicht zum gewünschten Ergebnis kommt. In diesem Beitrag stellen wir unsere Arbeiten an einem Recommender-System für Projektpartner im Rahmen des mi-Mappa- Projekts vor, das basierend auf Informationen aus Patenten, wissenschaftlichen Publikationen und Produktinformationen Experten für ein Projekt innerhalb eines Innovationsfeldes der Medizintechnik empfehlen kann.

Abstract

Successful research and development projects start with finding the right partners for the venture. Especially for interdisciplinary projects, this is a difficult and tedious task as experts from foreign domains are not known. Furthermore, the transfer of knowledge from research into practice is becoming more important in research projects to enable the quick application of research results. This is in particular relevant for projects in medical engineering. Patents and publications contain technical knowledge which can be exploited to find suitable experts. Patents are usually more product-oriented as the inventors have to describe an application area and products might be protected by patents. On the other hand, scientific publications represent the state-of-the-art in research. The challenge is finding the right mixture of research- or application-oriented experts from different domains. Hence, we propose a recommender system for finding experts for a certain topic based on patent topic clustering, ontologies, and ontology matching, which maps patents to corresponding innovation fields. The medical engineering domain serves as a first test bed, since projects in this area are highly interdisciplinary.

1 Einleitung

Bei der Zusammenstellung eines Konsortiums für ein Forschungs- und/oder Entwicklungsprojekt wird oft auf die existierenden Netzwerke und Kontakte und weniger auf objektive Kriterien bzw. Hilfsmittel für die Auswahl von potenziellen Projektpartnern zurückgegriffen. Bei interdisziplinären Projekten, wie sie in der Medizintechnik üblich sind, müssen aber Experten, die außerhalb bestehender Netzwerke zu suchen sind, aus unterschiedlichen Bereichen zusammengeführt werden. Darüber hinaus spielen die innovativen Fähigkeiten der Projektpartner eine zunehmend größere Rolle, da die Produktentwicklungszyklen immer kürzer werden. Daher ist die Bereitschaft und Fähigkeit der potenziellen Partner zur schnellen Umsetzung der Projektergebnisse in ein Produkt ein wichtiges Kriterium für die Auswahl von Projektpartnern in der Medizintechnik.

Studien haben gezeigt, dass die Kooperation zwischen Forschungs-einrichtungen und Unternehmen sowohl Produkt- als auch Prozessinnovation begünstigen können (Robin und Schubert, 2013). Gerade in interdisziplinären Projekten kann die Suche nach geeigneten Partnern sehr zeitaufwändig, komplex und am Ende vielleicht nicht so erfolgreich wie erwartet sein. Daher wäre eine Unterstützung der Suche in Form eines Recommender-Systems für Projektpartner wünschenswert. Ein solches System könnte den Suchprozess erheblich beschleunigen und neue Kooperationsmöglichkeiten eröffnen.

Für ein solches Recommender-System können Patente als sehr wichtige Informationsquelle dienen, da sie sehr viele technische Informationen zur Entwicklung eines Produkts beinhalten. Allerdings ist ihre Analyse schwierig, da die Sprache (Begriffe und Formulierungen) von anderen wissenschaftlichen Texten stark abweichen (Aras et al., 2014; Zhang et al., 2015). Da Autoren von Patenten (Erfinder) nicht nur Experten in ihrem Gebiet sind, sondern auch eine produktorientierte Sicht auf ihr Forschungsfeld haben, stellen Erfinder gerade für innovative Forschungsbereiche wie die Medizintechnik interessante potentielle Projektpartner dar. In diesem Beitrag stellen wir einen Ansatz vor, der Techniken aus den Bereichen Patentanalyse und -Mining, Ontologie-Mappings und -Matching (Shvaiko und Euzenat, 2013) kombiniert und in einem System zur Empfehlung von Kooperationspartnern implementiert wird.

Im Projekt mi-Mappa¹ wird derzeit an geeigneten Methoden zur Suche von Projektpartnern geforscht. Aufgrund des Projektes konzentrieren wir uns auf den Anwendungsbereich Medizintechnik, jedoch sollten die erforschten Methoden letztlich auch auf andere Bereiche übertragbar sein. Eine Voraussetzung für die Anwendung unseres Ansatzes in einem bestimmten Forschungsbereich ist allerdings die Identifikation von Innovationsfeldern, welche bereits in der Medizintechnik stattgefunden hat (Schlötelburg et al., 2008). Nach Schlötelburg et al. (2008) ist ein Innovationsfeld in der Medizintechnik ein Bereich, in dem es signifikante Innovationsaktivitäten, Zukunftspotenzial und eine möglichst vollständige Wertschöpfungskette gibt. Die Hauptinnovationsfelder der Medizintechnik umfassen derzeit (Schlötelburg et al., 2008; Deutsche Gesellschaft für Biomed. Technik im VDE, 2012)

- Bildgebende Verfahren,
- Prothesen und Implantate,
- Telemedizin und modellbasierte Therapien,
- Operative und interventionelle Geräte und Systeme,
- In-Vitro-Diagnostik,
- Regenerative Medizin

und auch Querschnittsthemen wie z.B. Patientensicherheit oder Gebrauchstauglichkeit.

Unser Ansatz kombiniert zwei komplementäre Wege. Einerseits erstellen wir für die Erfinder ein Profil auf Basis ihrer Patente, wissenschaftlichen Publikationen, Webseiten und anderen forschungsrelevanten Informationen. Dies beinhaltet die Identifikation eines Erfinders mit einem Autor einer Publikation. Wenn ein entsprechender Autor gefunden werden kann, können die Publikationen den Innovationsfeldern zugeordnet und somit das Spezialgebiet des Erfinders bzw. Autors bestimmt werden. Zur Abbildung von Publikationen auf Innovationsfelder nutzen wir Klassifikationsschemata von Publikationsdatenbanken und semiautomatische Ontology-Matching-Verfahren. Andererseits, sollte der Erfinder nicht als Autor wissenschaftlicher Publikationen identifiziert werden können, gruppieren wir die Patente eines Erfinders nach Themen und bilden diese Themen auf die

¹ <http://www.dbis.rwth-aachen.de/mi-Mappa>

Innovationsfelder ab. Auch für diese Verknüpfung zwischen Themen und Innovationsfeldern nutzen wir Ontology- Matching-Verfahren.

Im folgenden Abschnitt diskutieren wir zunächst verwandte Arbeiten im Bereich Patentanalyse und Recommender-Systeme für Forschungspartner. Unser Ansatz wird in Abschnitt 3 beschrieben. In Abschnitt 4 geben wir eine Zusammenfassung und einen Ausblick auf zukünftige Arbeiten.

2 Verwandte Arbeiten

Recommender-Systeme zur Suche von Experten bzw. Kooperationspartnern

Die Suche von Kooperationspartnern beinhaltet üblicherweise manuelle Schritte. Zum Beispiel müssen festgelegte Kriterien analysiert, bewertet und gewichtet werden (Geum et al., 2013; Awasthi et al., 2015). Systeme zur Suche von Experten für ein bestimmtes Thema basieren häufig auf Informationen, die von den Experten selbst veröffentlicht und verwaltet werden, auf vom Experten veröffentlichten Dokumenten oder auf Informationen aus sozialen Netzwerken (Wang et al., 2013). Diese Systeme erstellen entweder Expertenprofile (Expert Profiling) oder unterstützen den Suchprozess (Expert Finding) (Balog und De Rijke, 2007). Beispiele aus dem wissenschaftlichen Bereich sind ResearchGate², Google Scholar³ oder AMiner⁴. Aktuelle Arbeiten nutzen Algorithmen für die Analyse von sozialen Netzwerken (z.B. PageRank oder HITS (Rafiei und Kardan, 2015; Wang et al., 2013)) und graphbasierte Algorithmen (Rani et al., 2015). Wir konzentrieren uns in dieser Arbeit auf Systeme zur automatischen Expertensuche und -profiling auf Basis von veröffentlichten Dokumenten (z.B. Patenten und Publikationen). Viele der bestehenden dokumentbasierten Experten- Recommender-Systeme (ERS) beschränken ihren Suchbereich auf Dokumente und Personen innerhalb eines Unternehmens. Im Gegensatz dazu schlagen wir ein System vor, das Informationen aus beliebigen Datenquellen integrieren kann (Hai et al., 2016). Collaboration Spotting⁵ ist ein Projekt des CERN-Instituts, das basierend auf Patenten und Publikationen die Analyse eines wissenschaftlichen Feldes zu ermöglichen.

² <https://www.researchgate.net>

³ <https://scholar.google.com>

⁴ <https://aminer.org>

⁵ <http://collspotting.web.cern.ch>

Das Projekt verbindet sowohl die Analyse von Dokumenten als auch die soziale Netzwerkanalyse Skogstad et al. (2013). Es kann anhand von Begriffen gesucht werden und man kann sich verschiedene Netzwerke, z.B. von Unternehmen in einem Feld, visualisieren lassen. Ein Profiling der Experten wird jedoch nicht durchgeführt. Der DEMOIR-Ansatz (Yimam-Seid und Kobsa, 2003) nutzt ebenso wie unser Ansatz Ontologien, jedoch werden damit lediglich die Expertise der Experten und die Anwendungsdomäne modelliert. Ein 'Matching' der Ontologien findet nicht statt.

Patentanalyse mit Ontologien Die Nützlichkeit von Ontologien als formalisierte Domänenmodelle wurde auch für die Patentanalyse erkannt, insbesondere für die Patentsuche (Bonino et al., 2010). Beispielsweise nutzt das System PatExpert ein Netzwerk aus Ontologien und Wissensdatenbanken um Patentsuche, -klassifikation und -Clustering zu unterstützen. Trappey et al. (2009) stellen ein System vor, das Textfragmente mit einem Konzept in einer Ontologie verknüpft. Die Ähnlichkeit von zwei Patenten hängt dann von der Anzahl der Konzepte ab, die beide Patenten gemeinsam haben. Dieser Ansatz ist aber nur auf die Konzepte beschränkt, die innerhalb der Ontologie vorkommen.

Patent-Clustering Eine (graphische) Übersicht einer Menge von Patenten kann mithilfe von Clustering-Techniken realisiert werden (Tseng et al., 2007). Den berechneten Clustern kann ein Titel zugewiesen werden, der sich aus den häufigsten Begriffen der Patente innerhalb des Clusters zusammensetzt. Mogee und Kolar (1999) präsentieren einen bibliometrischen Ansatz, in dem Ko-Zitationsanalyse verwendet wird, um ko-zitierte Dokumente zu verknüpfen. Hierbei wird angenommen, dass diese Dokumente das gleiche Thema behandeln. Allerdings kann dieser Ansatz zu oberflächlichen Ergebnissen führen, da Details der Patente nicht berücksichtigt werden. Trappey et al. (2010) beschreiben eine Methodik für das Patent-Clustering, die zunächst die wichtigsten Phrasen aus einem Patent extrahiert (mithilfe von ontologiebasierten statistischen Methoden). Anschließend werden diese Phrasen zu Technologie-Clustern ('Technology Clusters') zusammengefasst und schließlich werden mit der gleichen Methode aus den Patentdokumenten Cluster gebildet.

Zusammenfassend kann man sagen, dass zwar einzelne Aspekte unseres Ansatzes bereits betrachtet wurden, die Kombination von verschiedenen Techniken zur Patent-Analyse, Clustering, Ontologie-Entwicklung und -Matching aber einen innovativen wissenschaftlichen Beitrag leistet.

3 Lösungsansatz

In der Patentanalyse werden quantitative und qualitative Ansätze unterschieden (Hong, 2004). Qualitative Ansätze extrahieren Metadaten aus den Patenten (z.B. Erfinder, Lite- raturreferenzen, Anmelder), während quantitative Ansätze die Häufigkeit von Begriffen im Patenttext zählen. Das Ziel unseres Ansatzes, dargestellt in Abb. 1, ist die Verknüpfung von Erfindern und ihren Patenten zu Innovationsfeldern. Wie bereits in der Einleitung dargestellt wurde, verfolgen wir dafür zwei parallele Ansätze:

- Verknüpfung der Erfinder mit Autoren von wissenschaftlichen Publikationen oder
- Clustering der Patente und Zuordnung eines oder mehrere Themen zu einem Cluster (Topic Clustering).

In beiden Ansätzen werden nur die Patente betrachtet, die das Ergebnis einer initialen Stichwortsuche sind. Die Stichworte sollten aus der Beschreibung des geplanten Projekts extrahiert werden. Die Ergebnisse beider Ansätze (Publikationen bzw. Themen) werden dann Konzepten aus existierenden medizinischen Ontologien zugeordnet. Die medizinischen Ontologien sind umfangreiche Beschreibungen der Domäne und können daher auch für die Charakterisierung der Innovationsfelder der Medizintechnik genutzt werden, d.h. den Innovationsfeldern werden also Konzepte der Ontologien zugeordnet. Schließlich können durch die Verkettung der Ergebnisse Erfinder Innovationsfeldern zugeordnet werden.

In den folgenden beiden Abschnitten werden wir die beiden Ansätze detailliert beschreiben. Im Anschluss daran werden wir auf den Entwurf der Ontologien und die Matching- Techniken eingehen.

3.1 Verknüpfung von Erfindern mit Autoren

Grundlage für die Verknüpfung ist die Annahme, dass Erfinder auch als Forscher tätig sind und wissenschaftliche Publikationen zum gleichen Thema publizieren. Zu den Publikationen existieren häufig Klassifikationsterme oder andere manuell festgelegte Schlagworte, die sich einfacher mit Ontologien verknüpfen lassen als die eher abstrakten Klassifikations terme der 'International Patent Classification' (IPC).

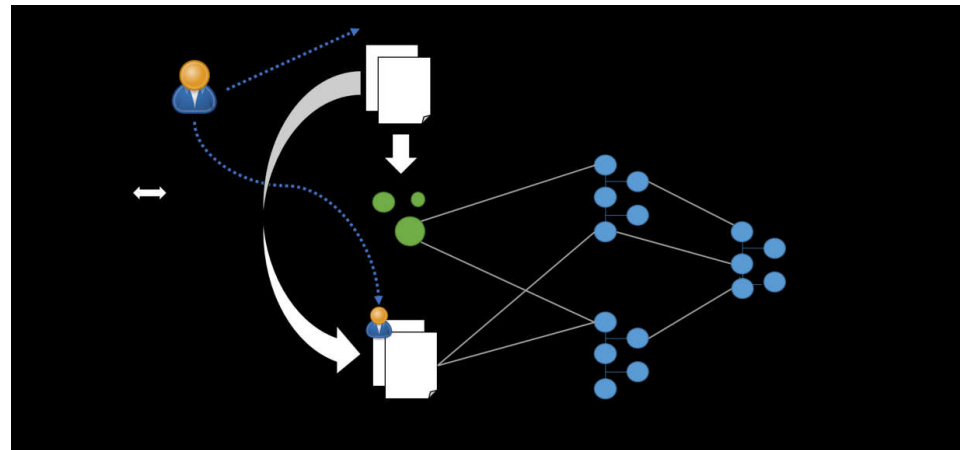


Abbildung 1: Architektur zur Abbildung von Patenten auf Innovationsfelder

In unserem ersten Prototyp haben wir Publikationsdatenbanken wie 'Web of Science' oder PubMed eingebunden (Hai et al., 2016), die Artikel anhand weitverbreiteter medizinischer Taxonomien klassifizieren, z.B. MeSH (Medical Subject Headings). Aus den Datenbanken werden alle Publikationen eines Autors abgefragt (ein Erfinder der im initialen Schritt gefundenen Patente). Da keine eindeutigen Identifikatoren für Erfinder bzw. Autoren zur Verfügung stehen, sondern nur Namen als Zeichenketten, ist ein mehrstufiger Prozess zur Identifizierung von Erfindern und Autoren notwendig. Hierfür extrahieren wir diverse Metadaten der Erfinder eines Patents (neben Name z.B. Firmenname, Adresse) und nutzen diese Informationen zusammen mit den inhaltlichen Informationen eines Patents zum Clustering der Erfinder. Im Idealfall sollten alle Patente eines Erfinders zu einem Cluster zugeordnet werden können, der dann eindeutig auf einen Autor von wissenschaftlichen Publikationen abgebildet werden kann.

3.2 Topic Clustering

In verschiedenen Studien wurde belegt, dass 70-90% des technischen Wissens nur in Patenten veröffentlicht werden (The Thompson Corporation, 2007; Stiftung Universität Hildesheim, 2014). Daher ist es wahrscheinlich, dass Erfinder nur Patente publizieren und bisher nicht als Autoren wissenschaftlicher Publikationen in Erscheinung getreten sind. Aus diesem Grund müssen wir einen alternativen Weg für die Zuordnung von Erfindern zu Innovationsfeldern zur Verfügung stellen. Für diesen zweiten Weg unseres Ansatzes nutzen wir Topic Clustering, das gleichzeitig auch für die Verifikation der ersten Lösungsmöglichkeit genutzt werden kann.

Das IPC-Klassifikationsschema für Patente ist zwar umfassend, aber für eine gute Patentanalyse nicht detailliert genug (Tseng et al., 2007). Genauere Kategorisierungen der Patente, wie sie mit Topic Clustering erstellt werden können, sind wünschenswert. Clustering gruppiert ähnliche Objekte miteinander, wobei die Ähnlichkeit durch die Distanz zwischen den Objekten in einem n-dimensionalen Vektorraum bestimmt wird.

Für die Merkmalsextraktion aus Textdokumenten nutzen wir die üblichen Vorverarbeitungsmethoden, wie Stoppwortentfernung, Wortstammbildung und Tokenisierung (Gonçalves et al., 2010). Dabei werden nur die wichtigsten Begriffe betrachtet, die durch die Berechnung der 'Inverse Document Frequency' (IDF) und 'Term Frequency' (TF) ermittelt werden. Den erkannten Clustern wird ein Titel bestehend aus den 3-5 wichtigsten Begriffen zugeordnet. Da die Merkmalsextraktion aus Textdokumenten aufwändig ist, betrachten wir nur den ersten Teil der Textabschnitte. Dies verbessert nicht nur die Verarbeitungsgeschwindigkeit sondern auch die Ergebnisqualität (Fall et al., 2003).

3.3 Auswahl, Entwurf und Matching der Ontologien

Auswahl der Ontologien: Unser Ansatz basiert auf der Verfügbarkeit von medizinischen Ontologien und deren Verknüpfung mit der Ontologie für Innovationsfelder. Da eine Vielzahl von Ontologien im biomedizinischen Bereich existieren, müssen zunächst eine oder mehrere Ontologien ausgewählt werden, die unsere Domäne am besten abdecken.

Dazu haben wir zunächst über Bioportal⁶ (eine Suchmaschine für biomedizinische Ontologien) nach passenden Ontologien gesucht. Des Weiteren nutzten wir den 'Ontology Lookup Service'⁷ und Ontobee⁸, um einen umfassenden Überblick über relevante Ontologien zu bekommen. Für die Suche extrahierten wir 174 Begriffe aus den Beschreibungen der sechs Innovationsfelder (Schlötzelburg et al., 2008; Deutsche Gesellschaft für Biomed. Technik im VDE, 2012). Die relevantesten Ontologien waren.

- National Cancer Institute Thesaurus (NCIT),
- Systematized Nomenclature of Medicine - Clinical Terms (SNOMEDCT),
- Medical Subject Headings (MeSH) und
- die 'Robert Hoehndorf Version of MeSH' (RHMeSH).

Für diese Ontologien analysierten wir die Abdeckung der Suchbegriffe durch einzelne Ontologien (vgl. Abb. 2). Die Y-Achse stellt die Prozentzahl der in der Ontologie vorkommenden Suchbegriffe dar.

Offensichtlich ist keine Ontologie in allen Bereichen besser als die anderen Ontologien. Generell ist die Abdeckung auch sehr niedrig. Daher versuchten wir die Abdeckungsrate zu verbessern, indem mehrere Ontologien betrachtet werden (vgl. Abb. 3). Durch die Verwendung von allen vier Ontologien kann die Abdeckungsrate im Schnitt um etwa 10 Prozentpunkte verbessert werden. Auch das 'Ontology Recommender Tool' des Bioportal kommt zu einem ähnlichen Ergebnis. Zusätzlich werden wir die gefundenen Ontologien noch mit Hilfe von Qualitätskriterien für Ontologien bewerten (Vrandečić, 2009; Gomez-Perez, 2004).

⁶ <http://bioportal.bioontology.org>

⁷ <http://www.ebi.ac.uk/ontology-lookup>

⁸ <http://www.ontobee.org>

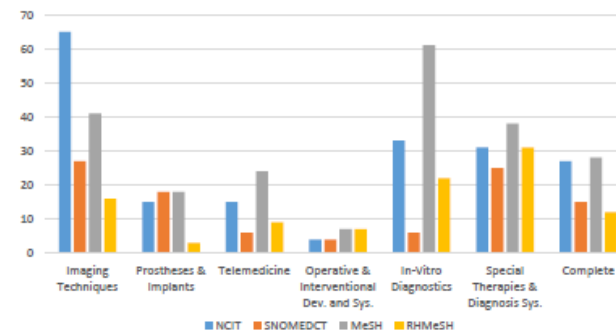


Abbildung 2: Abdeckung der Suchbegriffe durch einzelne Ontologien

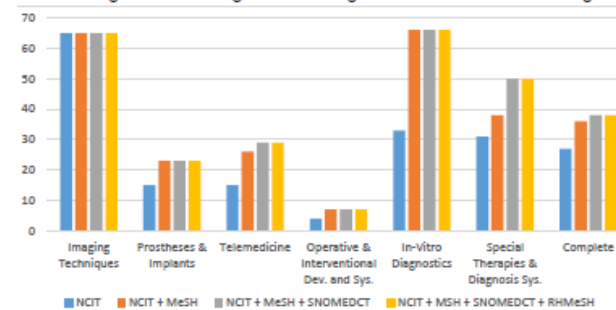


Abbildung 3: Abdeckung der Suchbegriffe durch Kombinationen von Ontologien

Entwurf der Ontologie für Innovationsfelder Um Publikationen oder Patent-Cluster mit Innovationsfeldern zu verknüpfen, ist eine Modellierung der Innovationsfelder durch Ontologien erforderlich. In dieser Ontologie sollten nicht nur die Innovationsfelder beschrieben sein, sondern auch wichtige Konzepte, die die wesentlichen Elemente eines Innovationsfelds darstellen. Dazu wurden Interviews mit Domänenexperten durchgeführt, existierende Ontologien analysiert und Fachliteratur zu den Innovationsfeldern untersucht. Der Entwurfsprozess orientiert sich an der NeOn-Methodik (Suárez-Figueroa, 2010). Die in der Ontologiesuche verwendeten 174 Begriffe stellen auch den Startpunkt für die Ontologie-modellierung dar. Die erste Version der Innovationsfeldontologie wurde durch Interviews mit Domänenexperten evaluiert und wird nun kontinuierlich auf Basis der Evaluierungsergebnisse weiterentwickelt.

Matching der Ontologien Um Erfinder oder Autoren Innovationsfeldern zuzuordnen, müssen die bisher gewonnenen Informationen miteinander verknüpft werden. Die Titel der Patent-Cluster und die Themengebiete der Publikationen müssen auf die Innovationsfeldontologie abgebildet werden. Die Verknüpfung der existierenden medizinischen Ontologien (z.B. NCIT) mit unserer Ontologie für Innovationsfelder wird manuell durchgeführt, da schon bei der Modellierung auf eine Verknüpfung mit existierenden Ontologien geachtet wurde. Da die Themenbereiche der gefundenen Publikationen bzw. Patente von der anfangs eingegebenen Beschreibung abhängen, können für diesen Fall die Verknüpfungen nicht vorab definiert werden. Hierzu wollen wir Methoden aus dem Ontology-Matching (Shvaiko und Euzenat, 2012) einsetzen, die z.B. in unserem GeRoMeSuite-System (Kensche et al., 2007) implementiert wurden. BioPortal bietet darüber hinaus auch bereits Verknüpfungen zwischen verschiedenen Ontologien an, die für diesen Fall genutzt werden können. Die gefundenen Verknüpfungen könnten von einem Experten verifiziert werden. Da aber in einem Recommender-System Fehler toleriert werden können, ist hier kein perfektes Ergebnis erforderlich. Daher könnte der Ontology-Matching-Schritt automatisch durchgeführt werden.

4 Fazit

In diesem Beitrag haben wir einen innovativen ontologiebasierten Ansatz vorgestellt, der die Expertensuche für Forschungsprojekte in der Medizintechnik unterstützt. In unserem Ansatz verwenden wir sehr viele Methoden aus dem Ontology-Engineering, z.B. Analyse, Erstellung und Matching von Ontologien, Definition von Anforderungen für Ontologien und schließlich Evaluierung der Ontologien. Wir verwenden zusätzlich aber auch Technologien aus anderen Bereichen, wie Text Mining und Patentanalyse. Derzeit konzentriert sich unsere Arbeit auf die Datenintegration, Modellierung und Auswahl der Ontologien und der Evaluierung der Clustering-Methoden. Für die bisher verwendete Clustering-Methode arbeiten wir derzeit an Verbesserungen der Performanz, da diese bei komplexen Text-Clustering-Aufgaben eine Herausforderung darstellt und wir ein interaktives System mit geringen Antwortzeiten benötigen.

Die Arbeiten sind derzeit noch nicht abgeschlossen, aber wir denken, dass die Integration von Text Clustering, Topic Modeling, Patentanalyse und Ontologie-Matching gut funktionieren wird. Die verschiedenen Methoden stellen an sich gesehen schon große Herausforderungen dar, aber deren Kombination könnte eine innovative Möglichkeit sein, sich unbekannte Forschungsfelder zu erschließen. Unser Ansatz ist nicht auf die Medizintechnik beschränkt - er kann auch auf andere Forschungsgebiete angewandt werden. Da für die Medizin schon viele Ontologien und Taxonomien verfügbar sind, ist dieses Gebiet für unseren ontologiebasierten Ansatz natürlich ideal.

Literatur

Aras, H., Hackl-Sommer, R., Schwantner, M., und Sofean, M. (2014). Applications and challenges of text mining with patents. In Proc. of 1st Intl. Workshop on Patent Mining and its Applications. Stiftung Universität Hildesheim.

Awasthi, A., Adetiloye, T., und Crainic, T. G. (2015). Collaboration partner selection for city logistics planning under municipal freight regulations. Applied Mathematical Modelling.

Balog, K. und De Rijke, M. (2007). Determining expert profiles (with an application to expert finding). In IJCAI, volume 7, pages 2657–2662.

Bonino, D., Ciaramella, A., und Corno, F. (2010). Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. World Patent Information, 32(1):30–38.

Deutsche Gesellschaft für Biomed. Technik im VDE (2012). Empfehlungen zur Verbesserung der Innovationsrahmenbedingungen für Hochtechnologie-Medizin. Technical report, VDE.

Fall, C. J., Töröcsvári, A., Benzineb, K., und Karetka, G. (2003). Automated categorization in the international patent classification. In ACM SIGIR Forum, volume 37, pages 10–25. ACM.

Geum, Y., Lee, S., Yoon, B., und Park, Y. (2013). Identifying and evaluating strategic partners for collaborative r&d: Index-based approach using patents and publications. Tech- novation, 33(6):211–224.

Gomez-Perez, A. (2004). Ontology evaluation. In Staab, S. und Studer, R., editors, Hand- book on Ontologies, Intl. Handbooks on Information Systems, pages 250–273. Springer.

Gonçalves, C. A., Gonçalves, C. T., Camacho, R., und Oliveira, E. C. (2010). The impact of pre-processing on the classification of medline documents. In PRIS, pages 53–61.

Hai, R., Geisler, S., und Quix, C. (2016). Constance: An intelligent data lake system. In Proc. SIGMOD. ACM. to appear.

Hong, S. (2004). The magic of patent information. World Intellectual Property Organization. http://www.wipo.int/sme/en/documents/patent_information_fulltext.html.

Kensche, D., Quix, C., Li, X., und Li, Y. (2007). GeRoMeSuite: A system for holistic generic model management. In Proc. VLDB, pages 1322–1325.

- Mogee, M. E. und Kolar, R. G. (1999). Patent co-citation analysis of eli lilly & co. patents. *Expert Opinion on Therapeutic Patents*, 9(3):291–305.
- Rafiei, M. und Kardan, A. A. (2015). A novel method for expert finding in online communities based on concept map and pagerank. *Human-centric Computing and Information Sciences*, 5(1):1–18.
- Rani, S. K., Raju, K., und Kumari, V. V. (2015). Expert finding system using latent effort ranking in academic social networks. *Intl. J. of Information Technology and Computer Science*, 2:21–27.
- Robin, S. und Schubert, T. (2013). Cooperation with public research institutions and success in innovation: Evidence from france and germany. *Research Policy*, 42(1):149–166.
- Schlötelburg, C., Weiß, C., Hahn, P., Becks, T., und Mühlbacher, A. C. (2008). Identifizierung von Innovationshürden in der Medizintechnik. Technical report, Bundesministeriums für Bildung und Forschung.
- Shvaiko, P. und Euzenat, J. (2013). Ontology matching: State of the art and future challenges. *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 1, pp. 158–176.
- Skogstad, E., Le Goff, J.-M., Fragkiskos, S., und Agocs, A. (2013). Visualizing collaborations and technology landscapes with interactive sociograms. In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 2013 IEEE, pages 1–6. IEEE.
- Stiftung Universität Hildesheim (2014). Womit beschäftigen sich Erfinder? <http://www.uni-hildesheim.de/en/fb3/institute/iwist>.
- Suárez-Figueroa, M. C. (2010). NeOn Methodology for building ontology networks: specification, scheduling and reuse. PhD thesis, Universidad Politecnica de Madrid.
- The Thompson Corporation (2007). Global patent sources - an overview of international patents.
- Trappey, A. J., Trappey, C. V., Hsu, F.-C., und Hsiao, D. W. (2009). A fuzzy ontological knowledge document clustering methodology. *IEEE Trans. on Systems, Man, and Cybernetics, Part B*, 39(3):806–814.
- Trappey, C. V., Trappey, A. J., und Wu, C.-Y. (2010). Clustering patents using non-exhaustive overlaps. *System Science and System Engineering*, 19(2):162–181.
- Tseng, Y.-H., Lin, C.-J., und Lin, Y.-I. (2007). Text mining techniques for patent analysis. *Information Processing & Management*, 43(5):1216–1247.
- Vrandečić, D. (2009). Ontology evaluation. In Staab, S. und Studer, R., editors, *Handbook on Ontologies*, chapter 13, pages 293–313. Springer.
- Wang, G. A., Jiao, J., Abrahams, A. S., Fan, W., und Zhang, Z. (2013). Expertrank: A topic-aware expert finding algorithm for online knowledge communities. *Decision Support Systems*, 54(3):1442–1451.
- Yimam-Seid, D. und Kobsa, A. (2003). Expert-finding systems for organizations: Problem and domain analysis and the demoir approach. *J. of Organizational Computing and Electronic Commerce*, 13(1):1–24.
- Zhang, L., Li, L., und Li, T. (2015). Patent mining: A survey. *ACM SIGKDD Explorations Newsletter*, 16(2):1–19.

Repositorien

Open Access in der Leibniz-Gemeinschaft – Die Rolle der ZBW als Infrastrukturdienstleister

Ralf Flohr, Jan B. Weiland

ZBW - Leibniz Informationszentrum Wirtschaft (Kiel und Hamburg)

Zusammenfassung

Open Access wird in der Leibniz-Gemeinschaft zunehmend als strategisches Ziel verankert. Den Infrastruktureinrichtungen und Institutsbibliotheken in der Leibniz-Gemeinschaft erwachsen daraus zahlreiche neue Aufgaben. Als Infrastruktureinrichtung in der Leibniz-Gemeinschaft sieht die ZBW – Leibniz-Informationszentrum Wirtschaft im Bereich Open Access seit längerem einen Arbeitsschwerpunkt und bietet inzwischen eine Reihe von Dienstleistungen, die es den Wissenschaftlerinnen und Wissenschaftlern der Leibniz-Gemeinschaft ermöglichen, mit ihren Forschungsergebnissen größere Sichtbarkeit zu erlangen. Der Beitrag stellt zunächst die einzelnen Ebenen der Umsetzung von Open Access in der Leibniz-Gemeinschaft dar, um dann im Detail auf den Grünen Weg und hier insbesondere auf die publikationsbegleitenden Dienste der ZBW für die Leibniz-Institute einzugehen. Abschließend wird erörtert, wie sich Open Access auf die Arbeitsteilung zwischen der ZBW als Infrastruktureinrichtung und den Forschungsinstituten in der Leibniz-Gemeinschaft ausgewirkt hat.

Abstract

Open Access is more and more conceived as a strategic target in the Leibniz Association. This has resulted in a number of new tasks for the research infrastructure institutions and libraries within the Leibniz institutes. As a research infrastructure the ZBW – Leibniz Information Centre for Economics places a special focus on Open Access and offers a range of services which help researchers to increase the visibility of their publications. This article describes the different levels of Open Access activities in the Leibniz Association and in particular discusses the Green Road of Open Access and ZBW's publication services for Leibniz institutes. Finally, we will show the impact Open Access has had on the division of labour between research infrastructures and research institutes within the Leibniz Association.

1. Einleitung

Die Leibniz-Gemeinschaft ist einer der großen Verbünde außeruniversitärer Forschungs- und Infrastruktureinrichtungen in Deutschland. Neben der gemeinsamen Finanzierung durch Bund und Länder ist für die Leibniz-Gemeinschaft eine „starke inhaltliche Zusammenarbeit - beispielsweise in Form der Leibniz-Forschungsverbünde und Leibniz-Netzwerke -, regelmäßiger Informations- und Erfahrungsaustausch, die Zusammenarbeit hinsichtlich gemeinsamer Interessen sowie die Wahrnehmung dieser Interessen nach außen, d.h. gegenüber der Politik, aber auch allgemein gegenüber der Öffentlichkeit kennzeichnend“ (Leibniz-Gemeinschaft, 2016).

Die Leibniz-Gemeinschaft gehört zu den Erstunterzeichnern der „Berliner Erklärung für den offenen Zugang zu Wissen“.¹ Seit der Unterzeichnung im Jahr 2003 hat sich Open Access zu einem zentralen Element der Zusammenarbeit zwischen den Leibniz-Einrichtungen entwickelt. Bereits 2005 wurde der Arbeitskreis „Open Access“ der Leibniz-Gemeinschaft gegründet, um die Institute bei der Umsetzung von Open-Access-Strategien zu unterstützen (Gebert et al., 2013). 2007 wurde von der Mitgliederversammlung der Leibniz-Gemeinschaft eine Leitlinie zu Open Access beschlossen, die neben einer grundsätzlichen Empfehlung der Veröffentlichung von Forschung auf dem „Goldenen Weg“ auch auf die Möglichkeiten des „Grünen Wegs“ eingeht. Entsprechend der Leitlinie soll der Zugang zu den Forschungsergebnissen aus Leibniz-Instituten über die frei zugänglichen elektronischen Archive innerhalb der Leibniz-Gemeinschaft erfolgen. Umsetzung und Weiterentwicklung der Open-Access-Leitlinie sollen sich dabei an nationalen und internationalen Aktivitäten der Wissenschaftsorganisationen und der Fach-Communities orientieren (Leibniz-Gemeinschaft, 2007).

2. Ebenen der Umsetzung von Open Access in der Leibniz-Gemeinschaft

Spätestens mit der Einsetzung des Arbeitskreises „Open Access“ arbeiten die Mitgliedseinrichtungen der Leibniz-Gemeinschaft intensiv an konkreten Maßnahmen zur Umsetzung von Open Access. Zum einen wurde auf dem Grünen Weg mit dem Portal „LeibnizOpen“² eine umfassende Repository-Infrastruktur aufgebaut. LeibnizOpen basiert auf einer Reihe von Fachrepositorien, die von den Informationsinfrastruktureinrichtungen in der Leibniz-Gemeinschaft aufgebaut wurden

¹ Vgl. <http://openaccess.mpg.de/Berliner-Erklärung>.

² Vgl. <http://www.leibnizopen.de/>.

und prinzipiell allen Leibniz-Einrichtungen für die Archivierung und Bereitstellung ihrer Publikationen im Open Access zur Verfügung stehen. Die Titeldaten aus den Fachrepositorien werden über standardisierte Schnittstellen im zentralen Portal von LeibnizOpen zusammengeführt und aufbereitet. Entsprechend ihrer fachlichen Ausrichtung sind die Leibniz-Einrichtungen einem der bestehenden Fachrepositorien zugeordnet. Die Repository-Betreiber leisten Unterstützung bei der Aufnahme der Publikationen und rechtlichen Fragen. Da die fachliche Zuordnung nicht für alle Einrichtungen zufriedenstellend erfolgen konnte, besteht zusätzlich die Möglichkeit, dass Institute ein eigenes Repositorium betreiben und an LeibnizOpen anbinden oder ihre Publikationen auf externen (also nicht-Leibniz-) Repositorien einspielen und über eine Schnittstelle mit LeibnizOpen verknüpfen (Gebert, 2013, S. 3). Der Online-Gang von LeibnizOpen erfolgte 2011.

Zum anderen engagierten sich die Mitgliedseinrichtungen auch auf dem Goldenen Weg, indem beispielsweise eigene Open-Access-Zeitschriften gegründet wurden. Viele der Institute haben Open-Access-Strategien entwickelt, Leitlinien verabschiedet oder Open-Access-Publikationsfonds eingerichtet. Die Entwicklungen von Open Access auf europäischer oder nationaler Ebene wurden aufgenommen und in die eigenen Strategien mit einbezogen, wie etwa die Open-Access-Klausel im 7. Forschungsrahmenprogramm der EU bzw. das Open-Access-Mandat im aktuellen Rahmenprogramm „Horizon 2020“. Für die von einer eigenen Förderlinie der Leibniz-Gemeinschaft unterstützten Projekte im Rahmen des sogenannten „Leibniz-Wettbewerbs“³ gilt seit 2015 ebenfalls ein Open-Access-Mandat. Wichtige Impulse wurden durch die Beteiligung an Gremien der „Allianz-Initiative der Wissenschaftsorganisationen“ aufgenommen. Hier sind unter anderem die Verankerung von Open-Access-Rechten in den Allianz- und Nationallizenzen und die Empfehlungen zu Open-Access-Strategien für wissenschaftliche Einrichtungen und zu Open-Access-Publikationsfonds zu nennen. Erstmals 2016 wurde auch von der Leibniz-Gemeinschaft im Rahmen eines Pilotprojekts ein zentraler Open-Access-Publikationsfonds⁴ eingerichtet, der allen Leibniz-Instituten zur Finanzierung der Publikationsgebühren von Open-Access-Zeitschriften offen steht.

Auch neuere Bestrebungen auf nationaler und internationaler Ebene konzentrieren sich stark auf den Bereich von „Open Access Gold“.

³ Vgl. <http://www.leibniz-gemeinschaft.de/ueber-uns/leibniz-wettbewerb/>.

⁴ Vgl. <http://www.leibniz-gemeinschaft.de/infrastrukturen/open-access/open-access-publikationsfonds>.

Die Leibniz-Gemeinschaft hat sich der „OA2020 Initiative for the large scale transtition to Open Access“⁵ angeschlossen, die auf einem Positionspapier der Max-Planck-Gesellschaft zur Transformation von Zeitschriften basiert. Demnach soll bei künftigen Lizenzverträgen von Subskriptionszeitschriften zunehmend ein „Offsetting-Modell“ verankert werden, bei dem zusätzlich zum Zugriff auf die lizenzierten Inhalte eine Open-Access-Publikationsmöglichkeit für die Forscherinnen und Forscher der beteiligten Einrichtungen angeboten wird. Sobald eine kritische Masse an Open-Access-Content erreicht ist, sollten diese Zeitschriften in reguläre Open-Access-Zeitschriften umgewandelt werden. In eine ähnliche Richtung gehen Initiativen der „Allianz der Wissenschaftsorganisationen“ (Allianz der deutschen Wissenschaftsorganisationen, 2016) und „Science Europe“, dem Europäischen Dachverband von Fördereinrichtungen und Forschungsorganisationen (Science Europe, 2015).

3. Publikationsbegleitende Dienste der ZBW für die Leibniz-Institute

Aufgrund der hier skizzierten Entwicklungen ist Open Access für die ZBW und die anderen Infrastruktureinrichtungen in der Leibniz-Gemeinschaft zu einem strategischen Arbeitsfeld erwachsen. Es gehört neben den Themen „Forschungsdatenmanagement“ und „Science 2.0“ zu den inhaltlichen Schwerpunkten, über die eine intensive Vernetzung der ZBW mit der Fach-Community in den Wirtschaftswissenschaften hergestellt wird. Dieser Beitrag konzentriert sich insbesondere auf den Grünen Weg des Open Access und beschreibt die publikationsbegleitenden Dienste, die die ZBW für die Wirtschaftsforschung in der Leibniz-Gemeinschaft anbietet.

Die ZBW betreibt mit EconStor einen der fachlichen Publikationsserver, die zusammen mit den anderen Repositorien in der Leibniz-Gemeinschaft die Grundlage für das zentrale Portal LeibnizOpen bilden. Im Rahmen der fachlichen Zuordnung kooperiert die ZBW mit den Wirtschaftsforschungsinstituten und einigen weiteren Leibniz-Einrichtungen mit wirtschaftswissenschaftlichen Forschungsbereichen. Die publikationsbegleitenden Dienste beziehen sich vor allem auf die Zweitveröffentlichung von Forschungsarbeiten in EconStor. Dazu gehören auf der Input-Seite die Aufbereitung der Metadaten, das Einspielen der Volltexte in das Repository und die Unterstützung bei rechtlichen Fragen, sowie auf der Output-Seite die Weiterverbreitung der Titeldaten über eine Reihe von fachlichen Datenbanken und

⁵ Vgl. <http://oa2020.org/mission/>.

Aggregatoren mit dem Ziel, den Publikationen die größtmögliche Sichtbarkeit in der fachlichen Community zu verschaffen.

In den Wirtschaftsforschungsinstituten der Leibniz-Gemeinschaft findet sich ein breites Spektrum an Publikationsarten. Möglichst viele dieser Publikationen sollen in Zusammenarbeit mit den Instituten auf dem Weg der Zweitveröffentlichung im Open Access verfügbar gemacht werden. Dabei lässt sich das Publikationsaufkommen der Institute grob in hauseigene und externe Forschungsveröffentlichungen unterteilen. Zu den Hauspublikationen zählen eigene Zeitschriften, Schriftenreihen und Monografien. In den Wirtschaftswissenschaften ist es seit langem üblich, Forschungsergebnisse auf dem Wege der Vor-Veröffentlichung frei verfügbar zu machen. Auch die Wirtschaftsforschungsinstitute der Leibniz-Gemeinschaft geben Reihen von Arbeits- und Diskussionspapieren heraus und machen diese auf ihren Webseiten zugänglich. Schwieriger gestaltet sich die Open-Access-Situation bei den externen Forschungsveröffentlichungen. Es handelt sich hier überwiegend um Zeitschriftenartikel, die zum großen Teil in referierten Subskriptionszeitschriften oder in Sammelwerken bei nationalen oder internationalen Verlagen erschienen sind.

3.1 EconStor-Input: Die Upload-Services der ZBW

3.1.1 Self-Archiving und “Mediated Deposit Service”

Jedem wissenschaftlichen Angehörigen der Leibniz-Gemeinschaft steht grundsätzlich der Self-Archiving-Bereich von EconStor zur Verfügung, um die eigenen, wirtschaftswissenschaftlich-relevanten Veröffentlichungen in EconStor einzubringen. Dieser Zugang ist vorzugsweise für Einzelveröffentlichungen vorgesehen, insbesondere für die Veröffentlichung sogenannter „Postprints“. Die Autorin bzw. der Autor überträgt dabei im Verlauf des Upload-Workflows durch die Zustimmung zur „EconStor Deposit License“ ein einfaches Nutzungsrecht für die jeweilige Publikation auf die ZBW. EconStor dient damit quasi als „Institutionelles Repositorium“ aller wirtschaftswissenschaftlichen Leibniz-Institute.

Die Erfahrung hat gezeigt, dass die einzelnen Forscherinnen und Forscher ihre Veröffentlichungen – wenn überhaupt – nur unregelmäßig bzw. unvollständig in institutionelle Repositorien einpflegen. Aus diesem Grund hat sich die ZBW gemeinsam mit den beteiligten Institutsbibliotheken für eine Art „Mediated Deposit Service“ entschieden, der die folgenden Elemente umfasst:

- (1) Die Institutsbibliotheken übernehmen das systematische Monitoring sämtlicher Neuerscheinungen der institutsangehörigen Forscherinnen und Forscher, und prüfen die jeweiligen Optionen einer Open-Access-Bereitstellung. In der Praxis werden dabei u.a. folgende Fragen geprüft:
- Greifen die Voraussetzungen des deutschen Zweitveröffentlichungsrechts? Wenn ja, ab wann darf das Dokument frei zugänglich gemacht werden?
 - Enthalten die Autorenverträge konkrete Aussagen zu einer möglichen Zweitverwertung? Wurde ggf. ein Vertragszusatz eingefügt, mit dem sich die Autorin bzw. der Autor ein einfaches Nutzungsrecht für eine spätere Online-Zweitveröffentlichung vorbehält („Author Addendum“)?⁶
 - Falls kein Autorenvertrag (mehr) vorliegt, wie ist die Standard-Policy der Zeitschrift bzw. des Verlages in Bezug auf die Zweitverwertung laut SHERPA/RoMEO-Datenbank?⁷
 - Liegt ggf. Content aus Allianz-Lizenzen vor, die eine Open-Access-Verwertung erlauben?⁸
- (2) Der Upload der Dokumente wird vollständig durch Bibliotheksmitarbeiterinnen und -mitarbeiter durchgeführt. Institutsintern werden dazu von der Institutsbibliothek schriftliche Vollmachten mit den jeweiligen Autorinnen und Autoren abgeschlossen, die es den zuständigen Bibliotheksangehörigen erlauben, die Publikationen im Auftrag der eigentlichen Rechteinhaber auf EconStor zu übertragen.
- (3) Die für den EconStor-Upload registrierten Bibliotheksangehörigen erhalten über das interne Rechtemanagement von EconStor erweiterte Rechte, die es ihnen ermöglichen, weitere Titel „ihrer“ Autorinnen und Autoren, die eventuell bereits über andere Quellen ihren Eingang in EconStor gefunden haben, mit ihrer Institution zu verknüpfen. So wird ein mehrfacher Upload der Dokumente vermieden, und die Titel können dennoch sowohl in EconStor also auch in LeibnizOpen der jeweiligen Leibniz-Einrichtung zugeordnet werden.

3.1.2 „Full Upload Service“ für größere Publikationsaufkommen

Wie oben bereits erwähnt, ist im Grunde jedes wirtschafts- und sozialwissenschaftliche Forschungsinstitut der Leibniz-Gemeinschaft zugleich Herausgeber eigener Schriftenreihen oder Zeitschriften, die entweder im Eigenverlag oder über private Verlage

⁶ Vgl. die Formulierungshilfen des Arbeitskreises „Open Access“ der Leibniz-Gemeinschaft unter <http://bibliothek.inm-gmbh.de/event/formulierungshilfen-publikationsvertrage/>.

⁷ Vgl. <http://www.sherpa.ac.uk/romeo/index.php>.

⁸ Vgl. Handreichungen der Allianz-Initiative unter <http://dx.doi.org/10.2312/allianzoa.004>.

vertrieben werden. Insbesondere bei den im Eigenverlag veröffentlichten Titeln liegen die Nutzungsrechte üblicherweise bei den Instituten.

Diese Reihen oder Zeitschriften sind zumeist mit einer höheren Anzahl an Publikationen verbunden, so dass es nicht praktikabel wäre, diese per Einzel-Upload auf EconStor übertragen zu lassen. Die EconStor zugrunde liegende Repository-Software „DSpace“ verfügt über eine Schnittstelle für den Massenimport, über die diese Titel importiert werden können. Im Gegensatz zum Self-Archiving werden vor dem Import mit den herausgebenden Instituten schriftliche Nutzungsvereinbarungen abgeschlossen, mit denen der ZBW einfache, nicht-ausschließliche Nutzungsrechte an allen Titeln einer Reihe oder Zeitschrift eingeräumt werden.

Ein wesentlicher Teil des kostenlosen „Full Services“ ist die ZBW-interne Übernahme sämtlicher Arbeitsschritte zur Aufbereitung der Titeldaten, inklusive einer Erfassung der Abstracts, Schlagwörter oder der JEL-Fachklassifikation bis hin zum Hochladen der PDF-Dokumente auf EconStor. Das Full-Service-Angebot der ZBW gilt sowohl rückwirkend für ältere Ausgaben, als auch für den fortlaufenden Upload aller Neuerscheinungen.

Das beschriebene Full-Service-Angebot kann für alle größeren Publikationsaufkommen angewendet werden, beispielsweise auch für Publikationen aus Digitalisierungsprojekten, auf die im Folgenden näher eingegangen wird.

3.1.3 Digitalisierung

Eine Vielzahl älterer Publikationen von Leibniz-Instituten wird durch retrospektive Digitalisierung verfügbar gemacht. Es handelt sich dabei um Veröffentlichungen, bei denen die Nutzungsrechte entweder in den Instituten verblieben sind oder im Zuge einer 2008 erfolgten Gesetzesänderung des § 137 I UrhG zurückerlangt werden konnten. Teilweise wurde darüber hinaus mit den Verlagen die Zustimmung zur Bereitstellung älterer Jahrgänge von Zeitschriften und Monografien („Out-of-Print“-Werke) ausgehandelt. Das Wissenschaftszentrum Berlin für Sozialforschung (WZB) hat in einem umfangreichen Projekt auf diesem Weg rund 1.500 Werke seiner Forscherinnen und Forscher digitalisiert und auf EconStor bereitgestellt. Auch ältere Schriftenreihen und Zeitschriftenbände aus dem DIW Berlin und dem ZEW Mannheim konnten auf diese Weise erstmalig in elektronischer Form zugänglich gemacht werden.

Vor allem die im Erscheinungszeitraum zum Teil weit zurückreichenden Schriftenreihen können nun vollständig ab der ersten Nummer online angeboten werden. Das Einscannen von Zeitschriftenbänden und Schriftenreihen wird teilweise im Digitalisierungszentrum der ZBW durchgeführt. Für das Katalogisieren und Einspielen der Digitalisate in EconStor wurden im Rahmen des Full-Service-Angebots eigene Geschäftsabläufe entwickelt.

3.2 „EconStor-Output“: Die Verbreitungsservices in weitere Datenbanken und Fachportale

Die konsistente und homogene Metadaten-Aufbereitung und die zuverlässige, stabile Verlinkung sämtlicher EconStor-Volltexte sind die Voraussetzung für die zusätzliche Verbreitung dieser Titeldaten über weitere fachliche und allgemeine bibliografische Datenbanken, Portale und Suchmaschinen. Diese weiteren Vertriebskanäle sorgen somit insgesamt für eine deutlich erhöhte Sichtbarkeit der im Open Access verfügbaren Leibniz-Publikationen. Neben den im Folgenden beschriebenen Datenbanken wird EconStor darüber hinaus von weiteren, sogenannten „OAI-Service-Providern“ regelmäßig über die offene OAI-PMH-Schnittstelle geharvestet.

3.2.1 LeibnizOpen

Wie schon erwähnt ist LeibnizOpen seit 2011 das zentrale Nachweisportal für die Open-Access-Publikationen der Leibniz-Gemeinschaft. Aktuell sind dort rund 30.000 Publikationen von 71 zuliefernden Leibniz-Instituten verzeichnet (Stand April 2016). Die derzeit zehn von der ZBW betreuten Institute haben demzufolge ein besonderes Interesse an einem möglichst schnellen und vollständigen Nachweis der institutseigenen Veröffentlichungen in dem Portal. Das Gros der Titel wird dabei im Rahmen des Full-Services in EconStor erfasst, wobei sämtliche Leibniz-Veröffentlichungen bevorzugt bearbeitet werden.

Die Weiterleitung der Titeldaten an LeibnizOpen erfolgt automatisch über die OAI-PMH-Schnittstelle, für die ein eigenes Metadaten-Ausgabeformat entwickelt wurde, um den spezielleren Metadaten-Anforderungen gerecht zu werden. Letzteres ergibt sich u.a. aus der Darstellung der Titel in LeibnizOpen, v.a. die eindeutige institutionelle Zuordnung der Titel ist in diesem Kontext zu benennen (vgl. Abb. 1).

Sport as a Common Property Resource: A Solution to the Dilemmas of Doping

Autor/-in: Bird, Edward J.; Wagner, Gert G.

Link zum Volltext: http://www.econstor.eu/bitstream/10419/66050/1/Bird_1997_Sport_Common_Property.pdf

Einrichtung DIW

Verlag Thousand Oaks: Sage Publications

Quelle Journal of Conflict Resolution; Vol. 41; Iss. 6; p. 749-766; <http://dx.doi.org/doi:10.1177/0022002797041006002>

Publikationsjahr 1997

Dokumentart Zeitschriftenartikel

Zusammenfassung / Abstract The use of drugs in high-performance sports (doping) is a common pool resource (CPR) dilemma: regardless of the number of other athletes who dope, the athlete with strong tastes for victory will find doping optimal; yet if all athletes dope, they all bear negative health

Abb. 1: Titeldarstellung in LeibnizOpen (Ausschnitt)

Technisch betrachtet agiert LeibnizOpen folglich als OAI-Service-Provider, der die angebundenen Open-Access-Repositoryn regelmäßig harvestet, und so den eigenen Datenbestand fortlaufend aktualisiert.

3.2.2 Google Scholar

Seit ihrem Start vor über zehn Jahren konnte sich die wissenschaftliche Suchmaschine Google Scholar in zunehmenden Maße gegenüber den etablierten Fachdatenbanken als alternativer Einstieg für die Suche nach wissenschaftlicher Literatur etablieren (Handreck und Mönnich, 2008, S. 406). Eine aktuelle Studie kann für akademische Suchmaschinen wie Google Scholar bei der Literatursuche sogar eine höhere Popularität gegenüber den allgemeinen Suchmaschinen nachweisen (Gardner und Inger, 2016, S. 20). Inzwischen ist aber nicht mehr nur die Recherche nach wissenschaftlicher Literatur der zentrale Service von Google Scholar. Die hohe Anzahl von Literatur-Nachweisen im Google-Scholar-Index, sowie die in vielen Fällen parallel ebenfalls verfügbaren Volltexte, ermöglichen zusammengekommen eine umfassende Zitationsanalyse zum Aufbau eines eigenen Zitationsindexes, der wiederum die Grundlage für weitere Metriken und Rankings innerhalb von Google Scholar bildet.⁹

⁹ Vgl. „Google Scholar Metrics“: <https://scholar.google.com/intl/en/scholar/metrics.html>.

Seit 2012 bietet Google Scholar zudem jedem wissenschaftlich Publizierenden die Möglichkeit, eine persönliche Profilseite einzurichten, die neben den üblichen persönlichen Angaben eine Auflistung sämtlicher in Google Scholar enthaltenen Veröffentlichungen, inklusive der aktuell ermittelten Zitationen ermöglicht.¹⁰ Der besondere Mehrwert aus der Sicht der Leibniz-Autorinnen und -Autoren, der im EconStor-Tagesgeschäft entsprechend oft direkt nachgefragt wird, ist daher v.a. ein schneller und vollständiger Nachweis ihrer Publikationen in Google Scholar, da nur die im Index verfügbaren Titel auch innerhalb der jeweiligen Autoren-Profile verlinkt werden können (vgl. Abb. 2).

Um eine weitestgehende Integration der EconStor-Titel in Google Scholar zu gewährleisten, werden sämtliche Metadaten gemäß der Google-Scholar-Vorgaben aus den „Inclusion Guidelines for Webmasters“ aufbereitet, die im Wesentlichen eine normierte Ausgabe der Metadaten mit speziellen „HTML Meta Tags“ der jeweiligen beschreibenden Abstract-Seite eines Dokuments verlangen.¹¹ In EconStor werden dazu die „Highwire Press Tags“ verwendet.

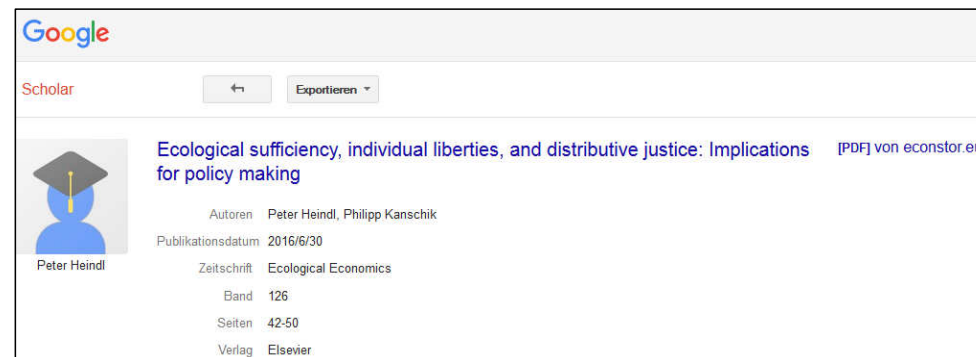


Abb. 2: Titelnachweis in einem Autoren-Profil von Google Scholar

3.2.3 RePEc – Research Papers in Economics

Mit aktuell mehr als zwei Millionen bibliographischen Nachweisen¹² ist RePEc weltweit eine der größten, frei zugänglichen Fachdatenbanken für die Wirtschaftswissenschaften, mit dem Schwerpunkt „Volkswirtschaftslehre“. Über 90% der Titel sind online verfügbar, wenn auch zum Teil zugangsbeschränkt, was v.a. für die Aufsätze aus den Zeitschriften der großen Verlage gilt.

¹⁰ Vgl. Google-Scholar-Blog-Beitrag vom 21. August 2014: <http://googlescholar.blogspot.de/2014/08/fresh-look-of-scholar-profiles.html>.

¹¹ Vgl. <https://scholar.google.com/intl/en/scholar/inclusion.html>.

¹² Stand April 2016.

Mit den verschiedenen, zum Teil bereits ab Anfang der 1990er Jahre entwickelten Diensten, und einem offenen und gut dokumentierten Datenmodell, das den herausgebenden Institutionen eine einfache Beteiligung ermöglicht (vgl. Krichel und Zimmermann, 2009), hat sich RePEc international zu einer der zentralen Datenbanken für die Suche nach und die Verbreitung von volkswirtschaftlichen Arbeitspapieren und Zeitschriftenaufsätzen entwickelt. Neben den Services IDEAS¹³ und EconPapers¹⁴, die in unterschiedlicher Form jeweils die vollständige RePEc-Datenbank abbilden, sorgt vor allem der Alerting-Dienst „New Economic Papers (NEP)“¹⁵ für eine unmittelbare Verbreitung neuer Arbeitspapiere in der wirtschaftswissenschaftlichen Community. Die Dienste LogEc¹⁶ und CitEc¹⁷ ergänzen die bibliografischen Daten um detaillierte Downloadstatistiken bzw. Zitationsanalysen. Ähnlich wie in Google Scholar können Wirtschaftswissenschaftlerinnen und -wissenschaftler über den „RePEc Author Service“¹⁸ persönliche Autoren-Profile erstellen, die sämtliche der aus Autorensicht relevanten Informationen auf einer Seite bündeln, neben vollständigen Titellisten v.a. Zitations- und Downloadangaben. Die Verknüpfung dieser Dienste bzw. der enthaltenen Informationen eröffnet wiederum zahlreiche Optionen für bibliometrische Analysen, die sich in unterschiedlichen Rankings niederschlagen (vgl. Zimmermann, 2013).¹⁹ Insbesondere diese Rankings werden in der Fach-Community aufmerksam verfolgt und regelmäßig diskutiert (vgl. Butz und Wohlrabe, 2016). Da über die Autoren-Affiliation auch Institutionen-Rankings erstellt werden können²⁰, hat neben den einzelnen Autorinnen und Autoren auch jede Forschungseinrichtung ein großes Interesse an einem möglichst umfassenden Nachweis ihrer Veröffentlichungen in der RePEc-Datenbank. EconStor verfügt dazu über eine flexibel konfigurierbare RePEc-Schnittstelle, die es erlaubt, die Metadaten ganzer Schriftenreihen, Zeitschriften oder Konferenz-Beiträge, separiert nach herausgebender Institution automatisch in RePEc einzuspielen.

¹³ Vgl. <https://ideas.repec.org/>.

¹⁴ Vgl. <http://econpapers.repec.org/>.

¹⁵ Vgl. <http://nep.repec.org/>.

¹⁶ Vgl. <http://logec.repec.org/>.

¹⁷ Vgl. <http://citec.repec.org/>.

¹⁸ Vgl. <http://nep.repec.org/>.

¹⁹ Siehe Übersicht der verfügbaren RePEc-Rankings unter <https://ideas.repec.org/top/>.

²⁰ Siehe z.B. das Institutionen-Ranking für Deutschland unter <https://ideas.repec.org/top/top.germany.html>.

3.2.4 EconBiz

EconBiz²¹ ist ein von der ZBW betriebenes Fachportal für die Wirtschaftswissenschaften. Das EconBiz-Portal bietet einen zentralen Einstieg für die Suche nach allen Arten wirtschaftswissenschaftlicher Fachinformation. Der EconBiz-Suchraum umfasst hierfür die wichtigsten deutschen und internationalen wirtschaftswissenschaftlichen Datenbanken, darunter die gesamte RePEc-Datenbank sowie ECONIS, den ZBW-eigenen Katalog mit über 4 Millionen Titelnachweisen von gedruckter und elektronischer wirtschaftswissenschaftlicher Literatur. Insgesamt summiert sich die Anzahl bibliografischer Nachweise auf mehr als 10 Millionen. Der gesamte EconStor-Bestand ist in EconBiz enthalten und wird regelmäßig per OAI-Harvesting aktualisiert.

3.2.5 OpenAIRE

Mit dem im Jahr 2014 gestarteten Forschungsrahmenprogramm „Horizon 2020“ verpflichtet die Europäische Union grundsätzlich alle Beteiligten, unter bestimmten Voraussetzungen sämtliche im Rahmen eines EU-Projektes entstandenen Publikationen auch im Open Access bereitzustellen („Horizon 2020 Mandate“).²² Diese Verpflichtung kann entweder direkt auf dem sogenannten „Goldenen Weg“ erfüllt werden, oder auf dem „Grünen Weg“ über die Bereitstellung einer akzeptierten Manuskript-Version in einem institutionellen oder fachlichen Open-Access-Repository. OpenAIRE ist das Open-Access-Portal, das diese frei verfügbaren Publikationen unter einer Oberfläche bündelt, mit detaillierten Projektinformationen verknüpft und zu jedem Projekt weitere Services anbietet (z.B. Literaturlisten oder die Verlinkung zu Forschungsdaten). EconStor ist als fachliches Repository direkt an OpenAIRE angebunden („OpenAIRE compliant“)²³, die EconStor-Titeldaten werden regelmäßig über die OAI-Schnittstelle aktualisiert. Damit bietet EconStor also auch jedem Leibniz-Angehörigen direkt die Möglichkeit, Publikationen aus EU-geförderten Projekten gemäß dem EU-Mandat im Open Access bereitzustellen.

3.3 Pflichtablieferung von Netzpublikationen an die Deutsche Nationalbibliothek

Im Oktober 2008 trat die „Verordnung über die Pflichtablieferung von Medienwerken an die Deutsche Nationalbibliothek“ in Kraft.²⁴

²¹ <http://www.econbiz.de/>.

²² „Open Access in Horizon 2020“: <https://www.openaire.eu/open-access-in-horizon-2020>.

²³ <https://www.openaire.eu/search/dataprovider?datasourceid=opendoar::7fea637fd6d02b8f0adf6f7dc36aed93>.

²⁴ <http://www.gesetze-im-internet.de/pflav/index.html>.

Demnach müssen der Deutschen Nationalbibliothek (DNB) neben Printmedien zusätzlich auch alle in Deutschland veröffentlichten Netzpublikationen im Wege der Pflichtablieferung zur Verfügung gestellt werden.²⁵

Insbesondere für außeruniversitäre Forschungseinrichtungen ohne eigene Repository-Infrastruktur ist diese Pflichtablieferung mit einem hohen Aufwand verbunden, da die intern verwendeten Publikationsmanagementsysteme oftmals nicht über die Standardschnittstellen verfügen, die eine automatische Ablieferung erlauben würden. Eine solche Schnittstelle müsste mühsam nachträglich programmiert werden. Vielfach wird dann aber zur Notlösung einer manuellen Einzelablieferung über das DNB-Webformular gegriffen.

Als DSpace-Repository verfügt EconStor über eine OAI-PMH-Schnittstelle, die grundsätzlich für die DNB-Ablieferung verwendet werden kann. Die Ausgabe der OAI-Schnittstelle wurde dazu um das von der DNB verlangte „xMetaDissPlus“-Metadatenformat erweitert²⁶. In EconStor lassen sich nun sämtliche Publikationen jeder einzelnen Einrichtung aggregiert unter einem bestimmten „OAI-Set“ direkt aufrufen. Das Harvesting-Tool der DNB kann somit gezielt bestimmte Einrichtungen in das automatische Pflichtablieferungsverfahren übernehmen.

Dieser DNB-Ablieferungsservice kann grundsätzlich von allen Forschungseinrichtungen genutzt werden, die über EconStor der gesetzlichen DNB-Pflichtablieferung für ihre Online-Veröffentlichungen nachkommen möchten. Faktisch wird sich die Nutzung allerdings auf die außeruniversitären Einrichtungen, allen voran die der Leibniz-Gemeinschaft beschränken.

3.4 Nutzungsanalyse nach COUNTER-Regeln

Die Bereitstellung von Zugriffsstatistiken durch EconStor ist ein wichtiger Mehrwert für die beteiligten Institute, weil diese Zahlen oft für die interne Leistungsmessung benötigt werden. Dazu ist es wichtig, eine konsistente, verlässliche und möglichst normierte Nutzungsanalyse durchzuführen.

Seit Oktober 2012 werden die EconStor-Nutzungszahlen nach dem „COUNTER Code of Practice for e-Resources“²⁷ ermittelt – einem international anerkannten Standard, der auch von den großen Datenbank Anbietern und zahlreichen Fachverlagen

²⁵ Das Ablieferungsverfahren der DNB wird hier beschrieben: <http://www.dnb.de/ablieferungsverfahren>.

²⁶ Siehe Übersicht der OAI-Metadatenformate von EconStor: <http://www.econstor.eu/dspace-oai/request?verb=ListMetadataFormats>.

²⁷ Vgl. http://www.projectcounter.org/code_practice.html

verwendet wird. Die COUNTER-Regeln gewährleisten nicht nur eine objektive Erhebung der Nutzungsdaten, sondern auch eine gute Vergleichbarkeit der Zahlen. Um sicherzustellen, dass die EconStor-Nutzungszahlen die tatsächlichen Zugriffe realistisch abbilden, werden die entsprechenden Logfiles regelmäßig auf weitere, auffällige Massen-Zugriffe geprüft und bei Unregelmäßigkeiten um „Ausreißer“ bereinigt.

Auf Wunsch wird jeder Herausgeber, dessen Schriften sich auf EconStor befinden, regelmäßig einmal im Monat per E-Mail über die jeweils angefallen Nutzungszahlen informiert. Alle Nutzungsdaten sind zudem offen auf der EconStor-Webseite²⁸ einsehbar. Basierend auf den Ergebnissen werden dort außerdem verschiedene Ranglisten angeboten.

4. Fazit

Wie eingangs beschrieben, unterstützt die Leibniz-Gemeinschaft mit Nachdruck den Wandel im Publikationssystem hin zu Open Access. Diese Ausrichtung auf Open Access hat in den letzten Jahren zu einer deutlich intensiveren Kooperation zwischen den Wirtschaftsforschungsinstituten und der ZBW als Informationsinfrastruktur-einrichtung geführt. Die Bibliotheken der Wirtschaftsforschungsinstitute spielen hier eine besondere Rolle, weil sie mit ihren Wissenschaftlerinnen und Wissenschaftlern in engem Kontakt stehen und am besten mit den institutionellen Gegebenheiten vertraut sind. Sie haben sich mit der Open-Access-Beratung von Autorinnen und Autoren und im Bereich der Klärung von Nutzungsrechten neue Arbeitsfelder erschlossen. Auf der anderen Seite verfügen die Institutsbibliotheken häufig nicht über die Ressourcen, dauerhaft eigenständige Open-Access-Angebote wie Repositorien aufzubauen und zu betreiben. Die ZBW hat frühzeitig in den Aufbau eines fachlichen Repositoriums und in die damit verbundenen technischen Kompetenzen investiert und bietet den beteiligten Leibniz-Instituten mit EconStor eine etablierte Plattform mit einer Reihe von Services und Funktionalitäten. Damit entlastet die ZBW die Institute von diesen ressourcenintensiven Tätigkeiten.

²⁸ Vgl. <https://www.econstor.eu/dspace/dsinfos/usagestatistics>

Literaturverzeichnis:

Allianz der deutschen Wissenschaftsorganisationen (2016): „Empfehlungen zur Open-Access-Transformation: Strategische und praktische Verankerung von Open Access in der Informationsversorgung wissenschaftlicher Einrichtungen. März 2016“, <http://dx.doi.org/10.3249/allianzoa.011>.

Butz, Alexander und Klaus Wohlrabe (2016): „Die Ökonomen-Rankings 2015 von Handelsblatt, FAZ und RePEc: Methodik, Ergebnisse, Kritik und Vergleich“, Ifo Working Paper No. 212, München: ifo Institut – Leibniz-Institut für Wirtschaftsforschung an der Universität München e.V., http://econpapers.repec.org/RePEc:ces:ifowps:_212.

Gardner, Tracy und Simon Inger (2016): „How Readers Discover Content in Scholarly Publications: Trends in reader behavior from 2005 to 2015“, Abingdon, UK: Renew Training, <http://www.simoningerconsulting.com/papers/How%20Readers%20Discover%20Content%20in%20Scholarly%20Publications.pdf>.

Gebert, Agathe, Anita Eppelin und Benjamin Zapilko (2013): „LeibnizOpen – das Portal für Open-Access-Publikationen der Leibniz-Gemeinschaft“, GMS Medizin – Bibliothek – Information, Vol. 13, Iss. (1-2), Doc. 17, doi:10.3205/mbi000281, <http://www.egms.de/static/pdf/journals/mbi/2013-13/mbi000281.pdf>.

Handreck, Franka und Michael W. Mönnich (2008): "Google Scholar als Alternative zu wissenschaftlichen Fachdatenbanken", B.I.T. online, Bd. 11, H. 4, S. 401-406.
Krichel, Thomas und Christian Zimmermann (2009): "The Economics of Open Bibliographic Data Provision", Economic Analysis and Policy (EAP), Vol. 39, Iss. 1, S. 143-52, [http://dx.doi.org/10.1016/S0313-5926\(09\)50049-5](http://dx.doi.org/10.1016/S0313-5926(09)50049-5).

Leibniz-Gemeinschaft (2007): "Leitlinie zu Open Access in der Leibniz-Gemeinschaft", http://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Infrastruktur/OpenAccess-Leitlinie.pdf.

Leibniz-Gemeinschaft (2016): „Geschichte der Leibniz-Gemeinschaft“, <http://www.leibniz-gemeinschaft.de/ueber-uns/geschichte/>, (abgerufen am 15.4.2016).

Science Europe (2015): "Science Europe Principles on Open Access to Research Publications. Updated May 2015", http://www.scienceeurope.org/uploads/PublicDocumentsAndSpeeches/WGs_docs/SE_POA_Pos_Statement_WEB_FINAL_20150617.pdf.

Zimmermann, Christian (2013): "Academic Rankings with RePEc", Econometrics, Vol. 1, Iss. 3, S. 249-280, <http://dx.doi.org/10.3390/econometrics1030249>.

**PUBLISSO – Das Open-Access-Publikationsportal der
Lebenswissenschaften: Wissenschaftsunterstützung
durch Open-Access-Publikationsangebote**

Ursula Arning

ZB MED – Leibniz-Informationszentrum Lebenswissenschaften

Abstract:

ZB MED ist mit PUBLISSO einen Schritt nach vorn gegangen. Es bietet seinen Zielgruppen, den Wissenschaftlerinnen und Wissenschaftlern wie auch den Bibliothekarinnen und Bibliothekaren, mit diesem Angebot einen wichtigen zusätzlichen Service zum traditionellen Bibliotheksgeschäft: Die Forschungserkenntnisse der einen können schnell und ohne kommerzielles Interesse veröffentlicht werden, die öffentlichen Gelder werden perspektivisch nicht durch teure Subskriptionszeitschriften doppelt beansprucht und die Publikationsplattform-PUBLISSO steht mit Fertigstellung interessierten Bibliotheken zur Erweiterung des eigenen Angebots zur Verfügung. ZB MED stellt zudem Beratungs- und Publikationsangebote zur Verfügung, die die Kolleginnen und Kollegen für ihre eigene Arbeit (nach)nutzen können. Durch das Open-Access-Angebot entspricht es seinem öffentlichen Auftrag, wissenschaftliche Ergebnisse allen zugänglich zu machen. Damit nimmt ZB MED seine Aufgabe als Informationszentrum Lebenswissenschaften und zentrale Fachbibliothek mit zukunftsfähigen Angeboten wahr.

With PUBLISSO, ZB MED succeeded in making a step forward and could provide for its target groups, the scientists and the librarians, an important additional service to the traditional librarian services: The results of the scientific researches can be published fast and without any commercial interests, the public resources don't need to be used for expensive abonnements for journals and the platform for publication, PUBLISSO, can be used after completion by interested libraries to improve their own offer. ZB MED offers also advisory and publishing services the colleagues can reuse for their own work. This Open Access offers correspond to the public task of ZB MED to make the scientific results public and accessible for all the people around the world. In this way, ZB MED performs its task as a centre for information for the life sciences and central library.

Ein Schritt zurück: Blick auf die jüngsten Entwicklungen im Bibliothekswesen

Um die aktuelle Situation einordnen zu können, ist ein Schritt zurück in die neunziger Jahre des 20. Jahrhunderts sinnvoll: Die Wissenschaftlerinnen und Wissenschaftler wurden vermehrt aufgefordert, ihre Forschungsergebnisse zu publizieren, um auf ihrem Karriereweg erfolgreich zu sein. Das schlägt sich in dem Schlagwort „Publish or Perish“ nieder. So werden tagtäglich neue Publikationen herausgebracht sowie auch neue Zeitschriften gegründet. Doch welche Zeitschrift ist relevant, welche muss unbedingt gelesen werden? Eine Auswahl zu treffen wird immer schwieriger. Trotz zunehmender Publikationsrate haben sich die Anforderungen nicht verändert, auf alle Publikationen Zugriff und Einsicht haben zu können. Die Wissenschaftlerinnen und Wissenschaftler benötigen alle Publikationen bzw. wollen alle, wenn nicht gelesen, so doch zumindest verfügbar haben, um den aktuellen Forschungsstand im Blick zu haben. Doch der Etat, den Bibliotheken zur Verfügung haben, steigt nicht in gleichen Maße. Einige Zeitschriften werden sogar immer teurer, weil in ihnen nur hochkarätige Forschende publizieren, so die gängige Meinung. Es bleibt den Bibliothekarinnen und Bibliothekaren also nur, einen Teil der von ihnen bezogenen Zeitschriften abzubestellen, in der Hoffnung, dass eine andere, größere Bibliothek diese Zeitschrift weiter vorhalten kann. Dies zieht den Protest der Forschenden nach sich, weil sie auf die gewohnten Zeitschriften nicht mehr auf dem eigenen Campus zugreifen können. Was also tun?

Durch die Entwicklung des Internets gibt es die Möglichkeit, zeitgleich und schnell auf der ganzen Welt zu kommunizieren. Mit der Zeit hat sich das Web zu einem partizipatorischen Web gewandelt, bei dem Nutzende nicht nur Konsumenten von Webinhalten sind, sondern zu Prosumern werden und Inhalte ohne viel Aufwand selbst erstellen können. Zunächst gestalten sich die Anfänge sehr einfach, doch schon bald wird das digitale Publizieren perfektioniert. PDFs bekommen ein Layout in Buchqualität, sie können unterstrichen und kommentiert werden, ganz wie ein „richtiges“ Buch. Zusätzlich bieten sie die Möglichkeit Texte zu verlinken und damit direkt aufeinander zu verweisen, Videos und andere audio-visuelle Dateien können direkt in die Texte eingebunden und abgespielt werden. Der Aufbau entsprechender Plattformen kostet jedoch auch Geld und sie müssen sich amortisieren. So haben die Verlage Geschäftsmodelle entwickelt, gemäß denen sie Gebühren für die Publikation von Artikeln (meistens) von den Autorinnen und Autoren verlangen, den sogenannten

Article Processing Charges (APCs). Das heißt, in diesem Geschäftsmodell fallen auf der anderen Seite des Publikationsprozesses Gebühren an.

Was bedeuten diese Szenarien für Bibliotheken und ihre tägliche Arbeit? Längst werden nicht alle wissenschaftlichen Publikationen im Open Access publiziert, weiterhin gibt es Printzeitschriften mit hohen Subskriptionsgebühren für Bibliotheken und sinkende Etats. Wer zahlt in dieser Situation die APCs für die Open-Access-Publikationen? Wer trägt die Doppelbelastungen in dieser Umbruchphase? Hat sich die Preisspirale der Subskriptionen nicht auf die Open-Access-Publikationen verlagert? Reicht es aus, in dieser Situation der Preisspirale durch Lizenzverhandlungen entgegen zu wirken?

Der Schritt nach vorn – der Schritt von ZB MED

ZB MED, die deutsche Zentralbibliothek für Medizin, Gesundheit, Ernährung, Umwelt und Agrarwissenschaften hat schon 2003 den Schritt nach vorn gewagt: Zu der Zeit, als die ersten Institutionen die Berliner Erklärung¹ unterschrieben, hat ZB MED in Kooperation mit der Arbeitsgemeinschaft Wissenschaftlich Medizinischer Fachgesellschaften (AWMF) und dem Deutschen Institut für Medizinische Dokumentation und Information (DIMDI) das Open-Access-Portal German Medical Science (GMS)² aufgebaut. Inzwischen ist GMS mit 16 Zeitschriften und über 60.000 Publikationen ein Flaggschiff im Bereich Open Access.

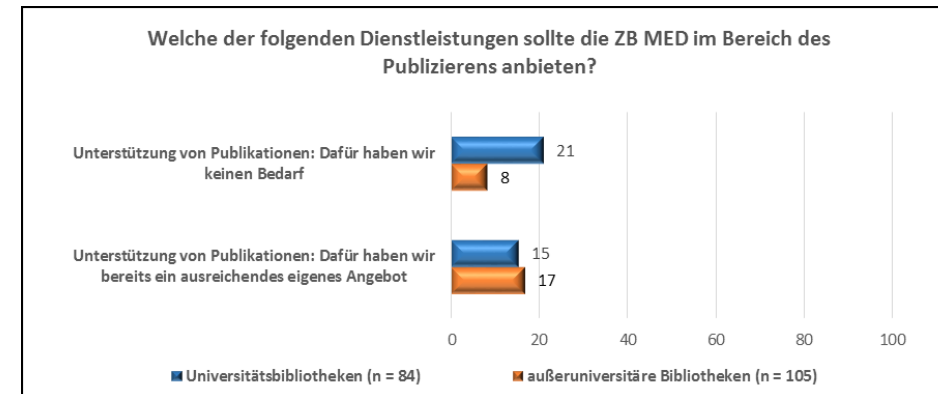
Im Rahmen der strategischen Neuausrichtung der letzten zwei Jahre, hat sich ZB MED entschlossen, den Bereich Open Access weiter zu stärken und auszubauen. Dieser Entschluss beruht einerseits auf den erwähnten positiven Erfahrungen mit GMS, andererseits auf der Marktstudie, die 2013 mit Hilfe einer externen Consulting-Firma durchgeführt wurde.³

¹ <http://openaccess.mpg.de/Berlin-Declaration>

² www.egms.de

³ Ehrhardt F. Heinold; Heinold/ Spiller & Partner Unternehmensberatung GmbH: Markt- und Zielgruppenstudie Gap-Analyse und Empfehlungen für die ZB MED – Endbericht. Köln; 2014. [dx.doi.org/10.4126/zbmed2014001](https://doi.org/10.4126/zbmed2014001).

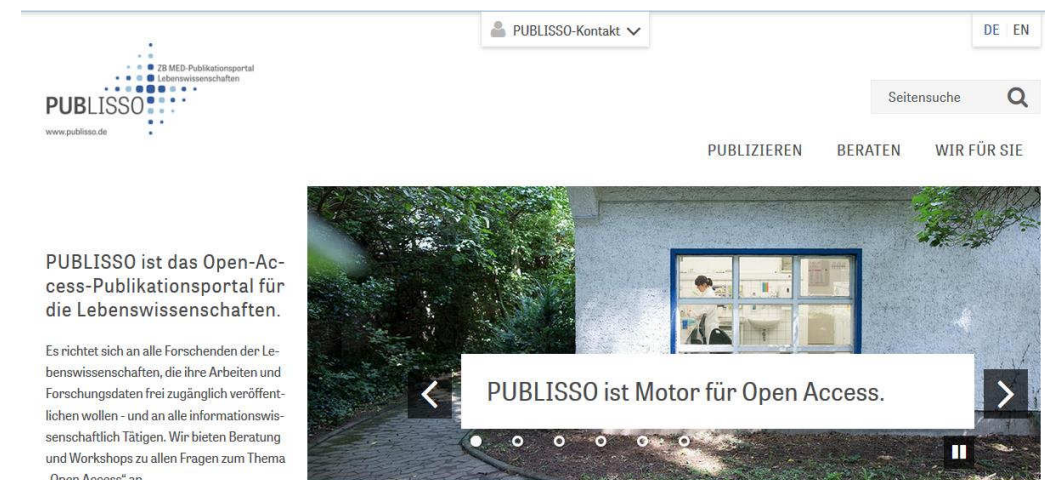
Aus der Studie hat sich ergeben, dass die Bedürfnisse der Zielgruppen, also der Forschenden und der Multiplikatorinnen und Multiplikatoren in Bibliotheken, in der Unterstützung bei der Publikation von Forschungsdaten und Forschungsergebnissen oder im Bereich von E-Learning-Medien liegen. Diese Unterstützung kann durch die Bereitstellung eines Publikationssystems wie auch durch Beratungstätigkeiten erfolgen.



(Abbildung 1: Für welche Dienstleistungen es keinen Bedarf gibt bzw. für welche es Bedarf gibt)

Abbildung 1 zeigt, dass nur 21% der universitären Bibliotheken angeben, **keinen** Bedarf an Unterstützung im Bereich Publizieren zu haben bzw. nur 17% der außeruniversitären Bibliotheken angeben, bereits über eigene Angebote zu verfügen. Daraus lässt sich schlussfolgern, dass ca. 80% der befragten Bibliotheksvertreterinnen und -vertreter ein Bedürfnis nach Unterstützung in diesem Bereich haben. Zumindest haben 80% noch kein eigenes Angebot aufgebaut, so dass sie in den kommenden Jahren, in denen mit wachsenden Open-Access-Publikationen zu rechnen ist, einen Bedarf entwickeln könnten, für den das ZB MED-Angebot zur Verfügung steht.

PUBLISSO – der nächste Schritt in die Zukunft



(Abb. 2: PUBLISSO, das Open-Access-Publikationsportal von ZB MED)

Seit 2015 ist der Name “Open-Access-Publizieren und -Beraten” Programm eines der drei Handlungsfelder von ZB MED. Es bildet eine Einheit mit den beiden benachbarten Handlungsfeldern „Suchen und Finden“ und „Forschen und Aufbereiten“. Allen dreien ist es gemein, den Forschungsprozess in allen Stadien zu begleiten und ihn darin zielgerichtet zu unterstützen. So hat ZB MED im Handlungsfeld „Suchen und Finden“ das Suchportal LIVIVO aufgebaut, das den Zugang zu der relevanten wissenschaftlichen Literatur und zu Forschungsdaten aus den Lebenswissenschaften ermöglicht. Gegebenenfalls nach Forschungsunterstützung durch das Handlungsfeld „Forschen und Aufbereiten“, können die Forschenden die eigenen Forschungsergebnisse über die PUBLISSO-Publikationsplattform veröffentlichen. Dort erhalten sie auch Unterstützung beim Publizieren und Speichern von Forschungsdaten. Diese Publikationen werden wiederum in LIVIVO eingespielt und sind so für andere Wissenschaftlerinnen und Wissenschaftler auffindbar.

ZB MED setzt sich für Open Access ein, da es sein Ziel ist, als Informationszentrum, den Wissenschaftlerinnen und Wissenschaftlern in aller Welt den bestmöglichen Zugang zu Forschungsergebnissen und Forschungsdaten zu ermöglichen.

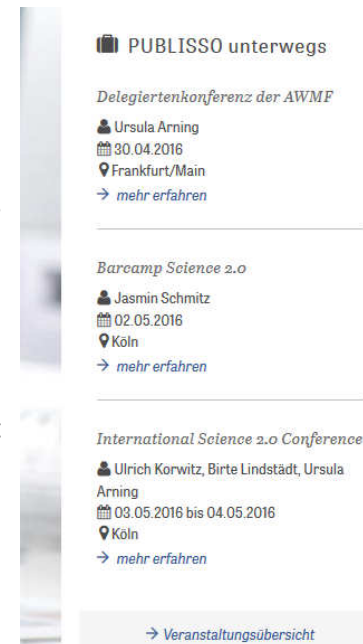
Damit ist eine schnellere Publikation von Forschungsergebnissen verbunden, die Möglichkeit, diese zeitnah zu aktualisieren, ihre weitreichende Sichtbarkeit und vor allem der kostenlose Zugang zu den neuesten wissenschaftlichen Erkenntnissen.

Um dieses zu ermöglichen, hat ZB MED das Open-Access-Publikationsportal PUBLISSO aufgebaut, das auf den drei Säulen Vernetzen-Beraten-Publizieren basiert, die eng miteinander verbunden sind und sich gegenseitig ergänzen. Mit diesem Service-Angebot versteht sich ZB MED als Motor für Open Access in den Lebenswissenschaften.

Vernetzen

Eine der drei Säulen von PUBLISSO ist die Vernetzungstätigkeit. PUBLISSO wird am Bedarf und mit der Wissenschaft entwickelt. Daher bringt sich das PUBLISSO-Team in Fachcommunities, Arbeitsgruppen und andere Gremien ein, die sich im Bereich Open Access engagieren und trägt neue Ideen in die konkrete Arbeit bzw. in das Publikationsportal zurück. Zu dieser Arbeit gehören auch Gespräche mit Vertretern aus Politik und Wissenschaft, die in ihren Gremien für die Verbreitung der Open-Access-Idee

(Abb. 3: *PUBLISSO unterwegs*:
Veranstaltungsankündigung und auswärtige Termine
des PUBLISSO-Teams auf www.publisso.de)



tätig werden können. ZB MED versteht sich dabei als Sprachrohr für den Bedarf der Wissenschaftlerinnen und Wissenschaftler aus den Lebenswissenschaften, das die Bedürfnisse aus der Praxis in die politische Arbeit einbringt. Für das PUBLISSO-Team bedeutet das aber auch, neue politische Entwicklungen zeitnah aufgreifen und im Publikationsportal umsetzen zu können.

Ein anderer Aspekt dieser Tätigkeit ist die Teilnahme an wissenschaftlichen Fachtagungen. Dort macht das PUBLISSO-Team auf die Vorteile des Open-Access-Publizierens aufmerksam, stellt seine eigenen Service-Leistungen vor und kommt mit den Forschenden ins Gespräch und erfährt so von ihren Bedarfen in der aktuellen wissenschaftlichen Arbeit bis hin zum Publizieren.

Gleichzeitig ist PUBLISSO auch in den Open-Access-Gruppen aktiv und vernetzt sich dort mit den Akteuren, schließt Kooperationen, um gemeinsam Überzeugungsarbeit für Open Access, Open Data und Open Science zu betreiben. Somit wird durch PUBLISSO die Publikation von Volltexten und Forschungsdaten im Open Access in den Lebenswissenschaften weiter verbreitet, dadurch bekannter gemacht und seine Akzeptanz gefördert.

Beraten

Ob letztlich verstärkt Open Access veröffentlicht werden wird, hängt auch vom Umgang mit dem Journal Impact Factor (JIF) ab, einer Metrik, die die Zitierhäufigkeit von Zeitschriften misst. Es kann über den JIF also keine Aussage über die Relevanz eines einzelnen Artikels in der jeweiligen Zeitschrift getroffen werden. Trotzdem hat der JIF bei der Forschungsevaluierung eine bedeutende Rolle, was die Forschenden quasi dazu zwingt, sich dieser Metrik zu beugen. Dass auch zahlreiche Open-Access-Zeitschriften einen JIF haben und über Renommee verfügen, kann durch Beratungsangebote aufgegriffen werden.

Daher gehört zur Überzeugungsarbeit für Open-Access sowohl eine intensive als auch breite Beratung. Das Beratungsangebot von PUBLISSO beruht zurzeit auf vier Formaten:

- 1) Informationstexte in Form von FAQs auf dem PUBLISSO-Open-Access-Portal (www.publisso.de/beraten),
- 2) Workshops für unterschiedliche Zielgruppen
- 3) Video-Tutorials zu ZB MED-Services,
- 4) Individuelle Beratung über das Kontaktformular, per E-Mail oder telefonisch.

In den PUBLISSO-FAQs werden Basisinformationen zum Thema „wissenschaftliches Publizieren in den Lebenswissenschaften“ mit Schwerpunkt Open Access aufbereitet. Dabei wird auf die Problematik und Möglichkeiten der Auswahl der Publikationswege eingegangen, unterschiedliche Open-Access-Wege (gold/grün) aufgezeigt, auf Finanzierungsmöglichkeiten hingewiesen oder Kriterien zur Einschätzung der Qualität von Publikationsorganen benannt. Bei aller Informationsdichte wird dabei auf die Lesbarkeit und Übersichtlichkeit der Inhalte geachtet, um einen schnellen Einblick in

die Thematik zu gewährleisten. Kürzere Auszüge der FAQs werden auch über Social-Media-Kanäle verbreitet. Zur Vertiefung stehen weitere Quellenangaben bzw. die persönliche Beratung durch das PUBLISSO-Team zur Verfügung. Eine objektive und neutrale, nicht projektspezifische Beratung ist dabei oberstes Gebot. Die FAQs werden regelmäßig aktualisiert und erweitert. Dies erfolgt einerseits durch Aufgreifen und Vertiefung der Fragen innerhalb des Beratungsangebots, durch Tagungsbesuche und Workshops, andererseits durch aufmerksames Verfolgen der neuen Literatur und Diskussionen in der Open-Access-Community.

Erweitert wird dieses Angebot wie oben erwähnt auch durch Workshops, die sich zum einen an Nachwuchswissenschaftlerinnen und Nachwuchswissenschaftler aus den Lebenswissenschaften richten, zum anderen an Multiplikatorinnen und Multiplikatoren. Dies entspricht den Ergebnissen der Marktstudie, die ergeben hat, dass es im Hinblick auf das Thema Open-Access-Publizieren einen großen Informationsbedarf gibt (s. Abbildung 1). Für die erste Zielgruppe werden Workshops in Graduiertenschulen und lebenswissenschaftlichen Fachbereichen der Hochschulen angeboten, um den jungen Forschenden bereits am Anfang ihrer Karriere die Vorteile von Open Access aufzuzeigen.

Im Rahmen der durchgeführten Fokusgruppeninterviews⁴ haben sich folgende Themen als relevant herauskristallisiert: Die Vermittlung umfassender Grundkenntnisse zu Open Access, Informationen zu Publikationskosten, Definitionen und Umgang von und mit Forschungsdaten sowie eine Einführung in die Rechte und Pflichten von Autoren. Auch die kritische Reflexion der Bewertung wissenschaftlicher Arbeiten über den Journal Impact Factor (JIF) und das Aufzeigen alternativer Metriken gehört dazu. Des Weiteren werden die Vorgaben von Forschungsförderungseinrichtungen und damit einhergehend finanzielle Fördermöglichkeiten (von Autorengebühren bzw. Transformationen von Zeitschriften) vermittelt. Auch die DOI-Vergabe für Forschungsdaten oder der Aufbau eines Forschungsdatenmanagements werden in Workshops oder in Vorträgen beleuchtet.

⁴ Die Fokusgruppeninterviews wurden am 25.6.2014 mit wissenschaftlichen Mitarbeitenden sowie am 3. und 4.12.2014 mit Professorinnen und Professoren geführt. Themenschwerpunkte waren LIVIVO, PUBLISSO und die Projekte aus dem Bereich Anwendungsorientierte Forschung und Innovation.

Neben der genannten Zielgruppe richtet sich das Beratungsangebot auch an die Kolleginnen und Kollegen (außer-) universitärer Bibliotheken, die sich in diesem Bereich fortbilden oder ähnliche Seminare für ihre Zielgruppe anbieten möchten. Für diese Zielgruppe werden insbesondere Train-the-Trainer-Workshops angeboten, die Multiplikatorinnen und Multiplikatoren in wissenschaftlichen Bibliotheken sowie an Graduiertenschulen und Fachbereichen darauf vorbereiten, inhaltlich wie methodisch Kenntnisse rund um Open Access an ihre Zielgruppe zu vermitteln. Für diese Arbeit hat das PUBLISSO-Team individuell gestaltete Workshops im Portfolio, die sowohl thematisch als auch zielgruppenspezifisch angepasst werden.

Tutorials zu Open Access runden das Angebot ab. Offen gebliebene Fragen werden in diesem Format aufgegriffen oder sperrige Inhalte visualisierend erklärt und den Zielgruppen nahe gebracht. Als Tutorial stehen bereits kurze Videos zur Verfügung, die erklären, wie man Forschungsdaten gemeinsam mit Publikationstexten bei GMS-Zeitschriften einreichen kann oder wie die DOI-Vergabe durch ZB MED erfolgt.⁵ Derzeit wird daran gearbeitet, auch Webinare als weiteres Format aufzunehmen.

Ergänzt wird dieses Angebot für Multiplikatorinnen und Multiplikatoren durch Veranstaltungen, die das Ziel haben, sich untereinander zu vernetzen, Probleme in der Praxis miteinander auf kurzem Wege zu besprechen und gemeinsam Lösungen zu finden.

Publizieren

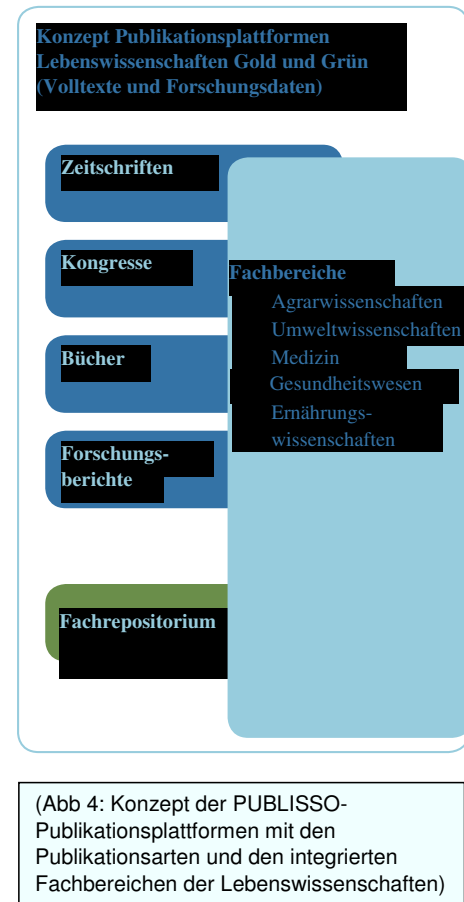
Um wirksam Alternativen aufzeigen zu können, muss es ein Angebot geben, das Open-Access-Publikationen ermöglicht, ohne von kommerziellen Anbietern abhängig zu sein und damit in eine neue Preisspirale zu geraten.

Deshalb ist 2003 die Open-Access-Plattform GMS online gegangen. Durch die Zusammenarbeit von GMS mit den medizinischen Fachgesellschaften, wird es Wissenschaftlerinnen und Wissenschaftlern dieser Fachbereiche möglich, unabhängig von kommerziellen Interessen, kostengünstig und zeitnah ihre Forschungserkenntnisse zu publizieren.

⁵ <https://repository.publisso.de/resource/frl%3A6399385>; <https://repository.publisso.de/resource/frl%3A6399500>; <https://repository.publisso.de/resource/frl%3A6399502>. Die Tutorials sind auch in englischer Sprache produziert: <https://repository.publisso.de/resource/frl%3A6399746>; <https://repository.publisso.de/resource/frl%3A6399748>; <https://repository.publisso.de/resource/frl%3A6399750>

Das Geschäftsmodell von GMS ist artikelbasiert, dabei tragen die Fachgesellschaften zumeist die Publikationsgebühren und sind für die wissenschaftliche Qualität der Beiträge verantwortlich. Seit 2014 ist das GMS-Angebot um die Veröffentlichung von Handbüchern (Living Handbooks)⁶ erweitert worden. Zudem hat ZB MED parallel eine neue Publikationsplattform aufgebaut, auf der Bücher, Artikel und Kongressbeiträge sowie auch Forschungsdaten aus den Bereichen der Lebenswissenschaften aus einer Hand publiziert werden können. Das Content-Management-System Drupal wurde dafür von ZB MED angepasst, so dass das System den Standards wissenschaftlicher Publikationen entspricht: Durch die Zuordnung von Rechten, kann der Peer-Review-Prozess im System gesteuert und dokumentiert werden. Nach erfolgreichem Abschluss können die Artikel, Kongressabstracts bzw. Kapitel mit einem Klick freigeschaltet werden und sind sofort weltweit sichtbar. So stehen die Forschungsergebnisse der Wissenschaftscommunity sehr schnell zur Verfügung. Die Sichtbarkeit der Publikationen wird zusätzlich gestärkt durch Querverweise zwischen den Veröffentlichungen und durch den interdisziplinären Charakter der Plattform. „Verwandte Publikationen“ verweisen auf Veröffentlichungen der PUBLISSO-Plattform mit ähnlichem Inhalt.

Auch im Bereich Forschungsdaten wird durch die Kooperation mit Dryad (einem Forschungsdaten-Repository mit Schwerpunkt auf den Lebenswissenschaften) eine „Schnittstelle“ von der Textpublikation zu Forschungsdaten angeboten, so dass die Forschungsergebnisse nachvollziehbar und transparent sind.



⁶ www.books.publisso.de bzw. www.gms-books.de

Nach dem erfolgreichen Aufbau der Plattform, wird das Publikationssystem auf einem Git als Open Source zur Verfügung gestellt. Andere Institutionen, zum Beispiel

Bibliotheken, die ihren Nutzerinnen und Nutzern auch diese Services zur Verfügung stellen und entsprechende Publikationsangebote aufbauen möchten, können das System nachnutzen.

Seit 2010 ist ZB MED Mitglied des internationalen Konsortiums DataCite, das sich politisch für die Verbreitung des Data Sharings einsetzt und Digital Object Identifier (DOI), einen persistenten Identifikator, vergibt. ZB MED ist auch Vergabestelle von DOIs, insbesondere für akademische Institutionen aus den Lebenswissenschaften. DOIs stellen die Auffindbarkeit und die Zitierfähigkeit von Texten und Forschungsdaten sicher.

ZB MED vergibt DOIs zudem auch für seine PUBLISSO-Publikationsplattform sowie für die Publikationen, die als Erst- oder Zweitveröffentlichung in das PUBLISSO-Fachrepository abgelegt werden.⁷ In Zusammenarbeit mit dem Hochschulbibliothekszenrum NRW (hbz) wird zurzeit ein neues Repositorysystem, basierend auf den Open-Source-Systemen Fedora und Drupal aufgebaut. Das Fachrepository steht sowohl für die Aufnahme der elektronischen medizinischen Dissertationen und Habilitationen (ZB MED veröffentlicht die elektronischen Dissertationen der Medizinischen Fakultät der Universität zu Köln) bereit, als auch zur Einbindung digitaler grauer Literatur und unselbständiger Werke wie Zeitschriftenartikel, Buchkapitel, Kongressabstracts usw. Neu ist, dass auch audiovisuelle Daten in hoher Qualität aus dem Repository durch vorinstallierte Viewer abgespielt werden können. Zudem ist das System geeignet, auch Forschungsdaten (auch singulär) in den gängigen Formaten aufzunehmen, für die spezielle Erfassungsmasken entwickelt wurden. Eine Besonderheit ist zudem, dass Institutionen eine eigene Institutsseite bekommen können, auf der die Publikationen der gesamten Institution abgebildet sind. Damit haben auch kleinere Organisationen, die über kein eigenes Repository verfügen, die Möglichkeit ohne größeren Aufwand und Kosten ihre Publikationen Open Access als Zweitveröffentlichung abzulegen und ihr Engagement für Open Access auf einen Blick sichtbar zu machen.

⁷ www.repository.publisso.de

Ausblick

Wissenschaftliche Bibliotheken haben die Möglichkeit und die Chance, sich durch Angebote im Bereich Open Access neu zu positionieren und dadurch Angebote zu schaffen, die über das traditionelle Bibliotheksgeschäft hinausgehen, es gleichzeitig aber auch unterfüttern. Durch Open-Access-Publikationsplattformen werden Publikationen ermöglicht, die eventuell durch profitorientiertes Verlagsgeschäft nicht realisiert werden könnten, aber dennoch für weitere Forschungen wichtig sind. Durch das Publikationsangebot wird indirekt auch der Content der Bibliotheken angereichert. Die eigenen Forschungsergebnisse durch das Publikationsportal verbreiten zu können, trägt zudem zur Bindung der Kooperationspartner bei oder führt zu neuen Kooperationen, die auch andere Bibliotheksbereiche betreffen können.

PUBLISSO ist ein Beispiel dafür. Nach nicht mal einem Jahr trägt das Engagement für Open Access Früchte. Die Beratungsanfragen sind seit dem Online-Gang von PUBLISSO im Oktober 2015 konstant gestiegen. Auch zu den Vernetzungs- und Beratungsworkshops gibt es sehr positive Rückmeldungen, so dass Fortsetzungen dazu bereits geplant sind. Zudem gibt es konkrete Anfragen von unterschiedlichen Fachgesellschaften aus den Lebenswissenschaften, ein Living Handbook zu veröffentlichen. Auch die Möglichkeit der Nachnutzung der DRUPAL-Plattform stößt auf Interesse der bibliothekarischen Community, wie zahlreiche Rückmeldungen auf Kongressen gezeigt haben. PUBLISSO setzt damit neue Standards für die Zukunft im Bereich des Bibliothekswesens.

Literaturhinweise:

Ehrhardt F. Heinold; Heinold/ Spiller & Partner Unternehmensberatung GmbH: Markt- und Zielgruppenstudie Gap-Analyse und Empfehlungen für die ZB MED – Endbericht. Köln; 2014. dx.doi.org/10.4126/zbmed2014001

Fachrepositorium Lebenswissenschaften: Ausbau des Repositoriums zur Publikation von Forschungsdaten

Birte Lindstädt¹

ZB MED Leibniz-Informationszentrum Lebenswissenschaften

Das Fachrepositorium Lebenswissenschaften von ZB MED wird von einem „klassischen“ Repositorium für Graue Literatur und elektronische Zweitveröffentlichungen zu einem Aufbewahrungsort für eine Vielzahl unterschiedlicher Publikationsarten, u.a. Forschungsdaten, ausgebaut. Die Konzeption und die erforderlichen Metadaten zur Erfassung und Publikation von Forschungsdaten müssen die Spezifika dieser Publikationsart berücksichtigen. Der Beitrag stellt den Prozess der Weiterentwicklung des Fachrepositoriums zu einem Forschungsdatenrepositorium dar und zeigt die im Zuge dessen erfolgten Entscheidungen, Arbeitsschritte und Diskussionen auf.

The ZB MED Repository Life Science will be developed from a repository for grey literature and for articles that have already been published to a deposit for a variety of possibilities for publishing research results, e.g. research data. The concept and the necessary metadata for registration and publication of research data have to refer to the specification of this kind of publishing. The article shows the enhancements from a repository mainly for text publications to a research data repository and points out the decisions, steps and discussions connected with this process.

Einführung

Angebote zum Open-Access-Publizieren sind aus Sicht des ZB MED Leibniz-Informationszentrum Lebenswissenschaften ein wichtiger Baustein im Leistungsspektrum einer Zentralen Fachbibliothek, um den Gedanken des offenen Zugangs zu Wissen zu verbreiten und entsprechende Publikationen zu ermöglichen. ZB MED bietet daher für alle Publikationsarten vom Zeitschriftenartikel über Kongressbeiträge und Handbücher bis zu Forschungsdaten die Möglichkeit der Open-Access-Publikation auf „goldenen“ oder „grünen“ Weg. Diese Angebote stellen eine

¹ Robin Rothe als studentischer Mitarbeiter bei ZB MED hat im Wesentlichen die grundlegenden Recherche- und Analysetätigkeiten für den Ausbau des Fachrepositoriums Lebenswissenschaften erledigt und somit auch die Grundlagen für diesen Beitrag geliefert.

Alternative zu den kommerziellen Verlagsangeboten dar, um allen qualitativ geeigneten Forschungsergebnissen eine Publikationsmöglichkeit zu eröffnen. Alle Services sind unter dem Publikationsportal PUBLISSO zusammengefasst.

Das Fachrepositorium Lebenswissenschaften von ZB MED bietet Autorinnen und Autoren derzeit die Möglichkeit der kostenfreien, elektronischen Zweitveröffentlichung („grüner“ Weg) ihrer wissenschaftlichen Texte aus den Fachgebieten Medizin, Gesundheitswesen, Ernährungs-, Umwelt- und Agrarwissenschaften nach den Grundsätzen des Open Access. Der Schwerpunkt lag bisher auf Monographien und sog. Grauer Literatur.

Mit der Umstellung und Weiterentwicklung der Software durch den technischen Partner hbz (Hochschulbibliothekszentrum Nordrhein-Westfalen) ging 2015 auch eine konzeptionelle Neuorientierung einher, die die Erweiterung des Fachrepositoriums zur Aufnahme von unselbstständigen Veröffentlichungen in Form von Zeitschriftenartikeln, Kongressbeiträgen, Buchartikeln und Forschungsdaten beinhaltet. Darüber hinaus ist es künftig möglich, eine Filterung nach Institutionen vorzunehmen, so dass es auch als institutionelles Repositorium genutzt werden kann.

Für die neu hinzukommenden Publikationsarten wurde jeweils ein eigenes Metadatenschema erarbeitet. Eine besondere Aufgabe stellte dabei die Erarbeitung eines Schemas für Forschungsdaten dar, da hierbei andere Kriterien als bei reinen Texten eine Rolle spielen.

So muss ggf. eine Verknüpfung zu einer Textpublikation möglich sein, falls es sich nicht um singuläre Forschungsdaten handelt. Die Erklärung und damit Nachnutzbarkeit der Daten muss durch Erfassung beispielsweise von Abstract, Erhebungszeitraum oder -punkt, den Erfassungsort, ggf. Koordinaten etc. gewährleistet sein. Bei der Beschreibung der Daten selbst spielt das Format, die Größe und der Ressourcentyp eine Rolle.

Die Softwareumstellung auf das Fedora-System kombiniert mit dem Frontendsystem Drupal bietet die Möglichkeit verschiedenste digitale Materialien in unterschiedlichen Formaten zu speichern und wiederzugeben. Auch Datenmengen bis zu einem Gigabyte stellen für das System im Hinblick auf Forschungsdaten kein Problem dar. Wichtig für die Funktion als Forschungsdatenrepositorium ist darüber hinaus, dass das System Relationen digitaler Objekte untereinander abbilden kann.

So können auch Objekte aufgenommen werden, die sich aus verschiedenen Komponenten zusammensetzen und verschachtelte hierarchische Strukturen aufweisen. Die Objekte können mit erweiterten Metadaten, z.B. einem DOI oder Lizenzbestimmungen, versehen werden und Zugangsbeschränkungen erhalten. Diese Funktionen wie auch das integrierte Viewer-Modul sind elementare Grundlagen für die Integration von Forschungsdaten, um beispielsweise eine Beziehung zwischen einem Artikel und den zugehörigen Daten herzustellen oder Audio- und Videodateien wiederzugeben (Quast, A., 2016).

Fachrepositorien und Forschungsdaten

Das Fachrepositorium Lebenswissenschaften war und ist für ZB MED ein wichtiger Baustein im Rahmen des Bestandsaufbaus im Bereich digitaler Fachinformation. Das Fachrepositorium für Open-Access-Publikationen hat dabei zwei wesentliche Ausgangspunkte:

Zum einen dient es als Veröffentlichungsserver der Medizinischen Fakultät der Universität zu Köln, der es den Promovierenden und Habilitanden der Medizin ermöglicht, ihrer Veröffentlichungspflicht nachzukommen. Zum anderen bietet es bisher überwiegend Grauer (digitalisierter) Literatur aus den Forschungs- und Infrastruktureinrichtungen des lebenswissenschaftlichen ZB MED-Fächerspektrums einen zentralen Ort zur Einsicht und dauerhaften Auffindbarkeit.

Hinzu kommt die Funktion als Publikationsort für elektronische Zweitveröffentlichungen überwiegend zunächst als Print erschienener Literatur.

Während die Integration weiterer unselbstständiger Publikationsformen wie Zeitschriftenartikel oder Kongressbeiträge von anderen Fachrepositorien² oder auch institutionellen Repositorien angeboten wird, ist die Publikation von Forschungsdaten im Rahmen eines Open-Access Repositoriums noch nicht weit verbreitet. Beispiele finden sich hier bei institutionellen Repositorien wie bei PUB Publikationen an der Universität Bielefeld³ oder Deposit Once der Technischen Universität Berlin⁴.

² Beispielsweise von EconStor, dem Open-Access-Server der Deutschen Zentralbibliothek für Wirtschaftswissenschaften / Leibniz-Informationszentrum Wirtschaft (ZBW). EconStor dient der freien Veröffentlichung wissenschaftlicher Literatur aus den Wirtschaftswissenschaften. ZBW: EconStor <https://www.econstor.eu/dspace/> (Zugriff: 21.04.16)

³ Universität Bielefeld: Publikationen an der Universität Bielefeld <https://pub.uni-bielefeld.de/> (Zugriff: 21.04.16)

⁴ TU Berlin: Repository for Research Data and Publications <https://depositonce.tu-berlin.de/> (Zugriff: 21.04.16)

Ziele und Strategien für die Publikation von Forschungsdaten im Fachrepositorium Lebenswissenschaften

Die Publikationsmöglichkeiten für Forschungsdaten im Rahmen von PUBLISSO bauen auf dem strategischen Ziel auf, bereits vorhandene Infrastrukturen zur Datenpublikation in den Lebenswissenschaften aufzuzeigen und an den Stellen eigene Angebote aufzubauen, wo Lücken identifiziert werden. Dies bezieht sich beispielsweise auf den sog. long tail der Forschungsdaten, also Daten, die ein geringes Datenvolumen aufweisen, in verschiedenen Datenformaten vorliegen und somit nur schwer standardisierbar sind, aber auch auf lebenswissenschaftliche Teildisziplinen, in denen Möglichkeiten zur Datenarchivierung und -publikation weitgehend fehlen.

Als einen ersten Schritt zur Erfüllung dieses Ziels wird das Fachrepositorium Lebenswissenschaften zur Aufnahme von Forschungsdaten ausgebaut. Dabei werden wiederum folgende Ziele verfolgt:

- dauerhafte Archivierung von Forschungsdaten aus den Lebenswissenschaften,
- Sicherstellung der Nachnutzbarkeit der Forschungsdaten,
- Publikation singulärer Forschungsdaten sowie von Forschungsdaten, die mit einem Volltext verknüpft sind,
- Lizenzierung der Daten im Sinne von Open Data, also der Möglichkeit zur Nachnutzung.

Obwohl die Abdeckung des lebenswissenschaftlichen Fächerspektrums Medizin, Gesundheit, Ernährung, Umwelt und Agrar einen eher generischen Ansatz bedeutet, sollen auch fachspezifische Dateneigenschaften berücksichtigt werden.

Wesentliche Schritte bei dem Ausbau des Fachrepositoriums sind:

1. Entwicklung eines Metadatenschemas (Erfassungsmaske),
2. technische Implementierung der Erfassungsmaske,
3. Erprobung durch Publikation realer Forschungsdaten.

Der Beitrag beschäftigt sich mit dem ersten dieser Schritte.

Bei der Erarbeitung der relevanten Metadaten sollten darüber hinaus in Pflicht- und optionale Kriterien unterschieden werden.

Prüfung vorhandener Metadatenschemata

Als Orientierungshilfe für die Entwicklung eines Metadatenschema zur Erfassung von Forschungsdaten wurden zunächst Metadatenschemata existierender (Daten-) Repositorien analysiert und auf eine Übertragbarkeit für die vorliegende Aufgabenstellung geprüft. Folgende Quellen und Beispiele sind dabei einbezogen worden (Tab. 1).

Institution	relevante Inhalte	Fachgebiet
Digital Curation Center (DCC) ⁵	Auflistung von bereits erstellten Metadatenstandards diverser Disziplinen (<i>Grundauswahl</i>)	generisch
Humboldt Universität Berlin ⁶	Informationen zur Dokumentation von Forschungsdaten sowie zu geeigneten Metadatenschemata	generisch / Beispiele für Disziplinen
DataCite ⁷	Auswahl zwischen festgelegten und optionalen Inhalten (<i>Möglichkeit auf Diversität der Disziplinen einzugehen</i>)	generisch
GESIS Datorium ⁸	Beispiel für allgemeine Aufnahme von Forschungsdaten	Sozialwissenschaften
Pangaea ⁹	Datenrepositorium für Geowissenschaften	Geo-/ Agrarwissenschaften
Dublin Core Metadata Initiative (DCMI) ¹⁰	weltweiter Metadaten-Standard	
Darwin Core ¹¹	Weiterentwicklung von DCMI zur Darstellung von biologischen Spezifikationen	Biologie / Geologie
PUB Uni Bielefeld ¹²	Beispiel zur Darstellung von Forschungsdaten im Repository	generisch

Tabelle 1: Quellen und Beispiele für Metadatenschemata

⁵ Digital Curation Center: <http://www.dcc.ac.uk/resources/metadata-standards> (Zugriff 28.04.2016)

⁶ Humboldt Universität Berlin: <https://www.cms.hu-berlin.de/de/ueberblick/projekte/dataman/teilen/dokumentation> (Zugriff 28.04.2016)

⁷ Data Cite: http://schema.datacite.org/meta/kernel-3/doc/DataCite-MetadataKernel_v3.1.pdf (Zugriff 28.04.2016)

⁸ gesis: <https://datorium.gesis.org/xmlui/> (Zugriff 28.04.2016)

⁹ PANGAEA Datapublisher for Earth and Environmental Science: <https://www.pangaea.de/> (Zugriff 28.04.2016)

¹⁰ Dublin Core Metadata Initiative: <http://dublincore.org/documents/dcmi-terms/> (Zugriff 28.04.2016)

¹¹ Biodiversity Information Standards: <http://rs.tdwg.org/dwc/terms/index.htm#livingspecimenindex> (Zugriff 28.04.2016)

¹² PUB Universität Bielefeld: <http://pub.uni-bielefeld.de/data/2777409> (Zugriff 28.04.2016)

Eine der wichtigsten Quellen stellte das DataCite-MetadatenSchema dar, das aktuell in der Version 3.1 vom Juni 2015 vorliegt¹³, im Laufe des Jahres 2016 jedoch aktualisiert wird. DataCite ist ein internationales Konsortium, das die Auffindbarkeit und Zitierbarkeit von frei zugänglichen Informationsressourcen, insbesondere Forschungsdaten, durch die Vergabe eines persistenten Identifiers fördert. ZB MED ist DataCite-Mitglied und fungiert als Vergabestelle für Digital Object Identifier (DOI) für lebenswissenschaftliche, akademische bzw. gemeinnützige Einrichtungen.

Auch die bisher bereits im Fachrepositorium Lebenswissenschaften veröffentlichten Informationen erhalten automatisch einen DOI.

Das in DataCite verwendete MetadatenSchema hat keinen fachlichen Fokus, da die Referenzierung verschiedenster Objekte aus allen Disziplinen angestrebt wird. Es versucht jedoch auch fachspezifische Aspekte einzubinden, z.B. durch das Feld GeoLocation. Außerdem setzt es die Hürde zur Registrierung und damit Publikation von Forschungsdaten durch lediglich sechs verpflichtende Metadaten recht niedrig an und unterscheidet darüber hinaus in „empfohlene“ und „optionale“ Felder.

Pflichtfelder	Empfohlene Felder	Optionale Felder
Identifier (with type Attribut)	Subject (with scheme sub-property)	Language
Creator (with type and name identifier attributes)	Contributor (with type, name identifier, and affiliation sub-properties)	Size
Title (with optional type attribut)	Date (with type sub-property)	Format
Publisher	RelatedIdentifier (with type and relation type sub-properties)	Version
Publication Year	Description (with type sub-property)	Rights
ResourceType (with general type description sub-property)	GeoLocation (with point and box sub-properties)	

Tabelle 2: aktuelles MetadatenSchema DataCite

¹³ Data Cite: http://schema.datacite.org/meta/kernel-3/doc/DataCite-MetadataKernel_v3.1.pdf (Zugriff 28.04.2016)

**Metadatenchema für Forschungsdaten im Fachrepositorium
Lebenswissenschaften**

Insbesondere auf dem DataCite-Schema, aber auch auf den anderen Beispielen sowie den Anforderungen der Forschungsdaten in den lebenswissenschaftlichen Disziplinen fußt das für das Fachrepositorium Lebenswissenschaften entwickelte Metadaten-schema:

Metadatum (übergeordneter Begriff)	Feldname	Feldname (untergeordnet)
Titel	Titel	
Urheberschaft	Autor*in	Nachname
		Vorname
		OrcID (optional)
		Affiliation (optional)
	Körperschaft (wenn kein Autor vorhanden)	
Dateiupload	Hochzuladende Datei	
	Format (xls, jpeg, etc.)	
	Medientyp (Bild, Video, Software, etc.)	
	Größe	
	Zugriffsrechte (open access, Embargo)	Embargofristende
	Copyrightjahr	
	Lizenz	
	DOI	Neu
		Vorhanden
	Zuletzt hochgeladen	
Erschließung	Abstract	Sprache
	Fachgruppenzuordnung	
	DDC-Klassifikation	
	Sprache	

Tabelle 3: Metadatenchema für Forschungsdaten im Fachrepositorium
Lebenswissenschaften (Pflichtfelder)

Metadatum (übergeordneter Begriff)	Feldname	Feldname (untergeordnet)
Beteiligte	Beteiligte Personen	OrcID
		Affiliation
	Förderer	Förder-ID
Erfassung	Schlagworte	Sprache
	Datenerhebungsform	
	Erhebungszeit	Zeitpunkt
		Zeitraum
	Erfassungsort	Koordinaten (Point)
		Kordinaten (Box)
Externe Referenzen	Verwendete Publikationen	
	Zugehörige Publikationen	
	Versionen	Vorgänger
		Nachfolger

Tabelle 4: Metadatenschema für Forschungsdaten im Fachrepositorium
Lebenswissenschaften (optionale Felder)

Dieses Schema war das Ergebnis eines längeren Entwicklungs- und Diskussionsprozess bei ZB MED. Die wichtigsten Punkte dieses Prozesses werden im Folgenden dargestellt.

Format und Medientyp

Auch wenn nicht alle Formate entsprechend im Fachrepositorium wiedergegeben werden können, ist es für die Nutzerinnen und Nutzer wichtig zu wissen welches Dateiformat vorliegt, um beurteilen zu können, ob eine entsprechende Wiedergabemöglichkeit vorliegt. Beispiele für Dateiformate von Forschungsdaten sind PDF, JPEG, XLS oder MPEG. Wichtig wird diese Angabe vor allem auch dann, wenn ZIP-Dateien, also Datencontainer, abgelegt werden.

Der Medientyp geht in der Regel mit einem bestimmten Dateiformat einher, ist jedoch als weiteres Feld aufgenommen worden, da Forschungsdaten sehr unterschiedlich definiert sein können und die Nutzerinnen und Nutzer dann sofort unterschiedliche Typen erkennen können. Hinzu kommt, dass beispielsweise für Audio- und Videodateien HTML-basierte Player direkt im Fachrepositorium zur Verfügung stehen. Als Medientypen können Audio, Video, Software, Text, Bild oder weitere angegeben werden.

Zugriffsrechte / Möglichkeit eines Embargos für die Publikation

Die Definition von Zugriffsrechten bietet die Auswahl zwischen „Open Access“ und „Embargo“. Eine Publikation der eigentlichen Daten ist demzufolge auf jeden Fall vorgesehen. Der Autor kann jedoch eine Embargofrist setzen und somit für eine gewisse Zeit lediglich die Metadaten für seine Publikation sichtbar machen.

Die Möglichkeit für das Setzen einer Embargofrist gründet auf einem wichtigen Vorbehalt der Forschenden in Bezug auf eine Datenpublikation: die Befürchtung, dass andere Wissenschaftler die Daten auswerten und Ergebnisse veröffentlichen bevor der Autor seine eigenen Forschungen abgeschlossen hat und ihm ggf. Möglichkeiten der Erstpublikation vorweggenommen werden (Fecher, B. Friesike, S., Hebing, M., Linek, S., Sauermann, A., 2015).

Vergabe von Lizenzen

Da die Datenpublikation im Fachrepositorium Lebenswissenschaften im Sinne von Open Data die Möglichkeit einer Nachnutzung impliziert, muss eine entsprechende Lizenz dafür vom Autor vergeben werden.

Obwohl Forschungsdaten vom Grundsatz her nicht urheberrechtlich geschützt sind, stellen sie im juristischen Sinne und im wissenschaftlichen Kontext vielfach ein sog. Datenbankwerk dar, da vom Forschenden eine „wesentliche Bearbeitung“ im Sinne einer systematischen Aufbereitung erfolgt.

Für Forschungsdaten kommt in der Regel die Open Data Commons Open Database License (ODbL)¹⁴, für Software die General Public Licence GNU¹⁵ in Frage.

Weit verbreitet für Textpublikationen, Präsentationen, Poster, etc. sind die Creative-Commons-Lizenzen¹⁶.

Da insbesondere hinsichtlich der Lizenzvergabe bei Forschungsdaten bei Autorinnen und Autoren eine große Unsicherheit besteht, stellt sich für die Publikation im Fachrepositorium Lebenswissenschaften die Frage, ob eine Empfehlung ausgesprochen werden soll oder die Forschenden selbst eine Auswahl treffen. Neben einer Auswahlmöglichkeit in Frage kommender Lizenzen für den publizierten Medientyp, erleichtert eine Empfehlung sicherlich die Publikation für den Autor.

¹⁴ Open Data Commons: <http://opendatacommons.org/licenses/odbl/> (Zugriff 26.04.2016)

¹⁵ Free Software Foundation: <http://www.gnu.org/licenses/gpl.html> (Zugriff 26.04.2016)

¹⁶ Creative Commons Deutschland: <http://de.creativecommons.org/was-ist-cc/> (Zugriff 26.04.2016)

Bevorzugung des DOI als persistenten Identifikator

Da ZB MED als DataCite-Mitglied DOIs als persistente Identifikatoren vergibt, verwendet es ihn auch in den eigenen Services. Ausgehend davon, dass es sich bei den künftigen Veröffentlichungen von Forschungsdaten um Erstpublikationen handelt, wird daher automatisch ein DOI vergeben. Sollte für Forschungsdaten bereits ein persistenter Identifikator existieren, wird dieser übernommen, gleichgültig ob es sich um einen DOI, eine URN oder ein Handle handelt. Dies muss allerdings manuell erfolgen.

Einbindung der DDC für die Sacherschließung

Neben der Fachgruppenzuordnung, die die lebenswissenschaftlichen Fächer Medizin, Gesundheit, Ernährung, Agrar, Umwelt und Biologie zur Auswahl stellt, war das Ziel einer Einbindung von Thesauri oder Klassifikationen zur Ableitung von Schlagworten für das Retrieval. Naheliegend ist die Einbindung der Dewey Decimal Classification (DDC) als weit verbreiteten Standard für eine detailliertere fachliche Zuordnung. Insbesondere bei digitalen Publikationen greift die Möglichkeit einer Mehrfachangabe von fachlichen Zuordnungen, um den Sucheinstieg auszuweiten. In der Praxis stellt sich dies für alle lebenswissenschaftlichen Fächer jedoch als schwer durchführbar heraus, da hier bis zu 35 Klassifikationen relevant sind. Eine richtige Zuordnung fällt daher schwer und aus Gründen der Qualitätssicherung wurde die Auswahl auf die Hauptsachgruppen 610 (Medizin, Gesundheitswesen), 610 & 640 (Ernährungswissenschaft), 630 (Agrarwissenschaft) und 333.7 (Umwelt) beschränkt.

Zulassung verschiedener Sprachen

Zu der Frage in welchen Sprachen die Publikation von Forschungsdaten erlaubt wird, wird eine recht weite Lösung angestrebt und im europäischen Kontext die fünf Sprachen Englisch, Deutsch, Spanisch, Französisch und Italienisch zugelassen. Das Argument dafür ist in erster Linie, dass die Forderung einer Übersetzung ins Englische oder Deutsche eine hohe Barriere für die Forschenden aufbaut. Das Datenrepositorium Dryad¹⁷ beispielsweise erlaubt lediglich die Ablage englischsprachiger Forschungsdaten.

¹⁷ Dryad Digital Repository: <http://datadryad.org/pages/policies> (Zugriff 28.04.2016)

Die Erfahrung in der Kooperation mit Dryad im Rahmen des Publikationsportals German Medical Science (GMS) von ZB MED, das die Möglichkeit einer Datenpublikation ergänzend zu einem Zeitschriftenartikel kostenfrei eröffnet, bestätigt dies. Die Zulassung mehrerer Sprachen bedeutet aber auch, dass Abstracts in allen fünf Sprachen akzeptiert werden. Ein englischer Abstract soll jedoch auf jeden Fall vorliegen. Dies ist bei Textpublikationen durchaus eine übliche Vorgehensweise. Inwieweit eine Einbindung von Übersetzungs-Tools möglich wäre, ist eine Frage, die künftig geklärt werden sollte, zum Start des Repositoriums jedoch keine Priorität hat.

Datenerhebungsform

Die Datenerhebungsformen, also die Art und Weise wie Daten gewonnen wurden, sollen anhand einer Dropdown-Liste wählbar sein. Das Problem hierbei ist, dass es kaum Vergleichsmöglichkeiten aus anderen Repositorien gibt und der Umfang aufgrund der Vielfalt in den Lebenswissenschaften schwer zu ermitteln ist. Eine Auswahl möglicher Erhebungsformen stellt die aktuelle Liste dar: Interview, Umfrage, Anamnese, Exploration, Probe, Gewebeprobe, Flächenmischprobe, Bodenbohrung, Experiment, apparative Untersuchung, körperliche Untersuchung, Feldbeobachtung, Laborbeobachtung, Analyse, Genomsequenzierung, Messung, Berechnung, Evaluation, Querschnittstudie, Langzeitstudie, Interventionsstudie, Kohortenstudie, Simulation, andere.

Verwendete Publikationen

Bei einer Erstpublikation von Forschungsdaten ist es derzeit eher unwahrscheinlich, dass sie auf anderen publizierten Daten aufbauen bzw. diese benutzen. Das Feld kann aber in Zukunft relevant werden, wenn das Aufkommen von publizierten Forschungsdaten steigt.

Ausblick

Das dargestellte Metadatenschema muss in einem nächsten Schritt technisch in das System integriert werden. Hierbei und bei der praktischen Erprobung im Rahmen der Publikation von Forschungsdaten wird sich zeigen, inwieweit die als „ideal“ oder „wünschenswert“ eingestuften Kriterien beibehalten werden können, welche weggelassen oder welche ergänzt werden müssen.

Neben dem Metadatenschema werden bei der Umsetzung noch weitere Faktoren eine Rolle spielen. Beispielsweise ist zu berücksichtigen, dass bei der Einreichung der Daten nicht notwendigerweise alle erforderlichen Informationen vorliegen, so dass es einer Rücksprache mit den Autorinnen oder Autoren bedarf. Hierfür sind entsprechende Prozesse zu entwickeln und Personalkapazität vorzuhalten. Dies und andere Punkte werden sich erst in den weiteren Phasen des Prozess ergeben.

Literaturhinweise

Quast, Andres: ZB MED Fachrepositorium Lebenswissenschaften - Open Access im Spiegel veränderter digitaler Publikationsformen. - In: Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen: Fakten und Perspektiven 2015. Jahresbericht. - 2016, in Press.

Fecher, B. Friesike, S., Hebing, M., Linek, S., Sauermann, A.: A Reputation Economy: Results from an Empirical Survey on Academic Data Sharing. Deutsches Institut für Wirtschaftsforschung, Discussion Paper 1454, 2015, S. 4-5

Green Open Access

E-Periodica: die Plattform für digitalisierte Schweizer Zeitschriften

Regina Wanger, Michael Ehrismann

ETH Zürich, ETH-Bibliothek

Abstract (D) :

E-Periodica ist eine Plattform der ETH-Bibliothek für digitalisierte Schweizer Zeitschriften aus den Bereichen Wissenschaft, Geschichte und Kultur. Seit der Aufschaltung 2007 wird das Angebot kontinuierlich mit neuen Publikationen ausgebaut. Aufgrund des enorm gewachsenen und immer noch steigenden Datenvolumens und um den langfristigen Betrieb gewährleisten zu können, erfolgte eine umfassende Überarbeitung der IT-Infrastruktur der Plattform. Im Rahmen dieser Weiterentwicklung wurde zudem ein Redesign des Webauftritts mit Fokus auf responsivem Design für mobile Geräte durchgeführt. Die Plattform E-Periodica wird im vorliegenden Text als bewährtes OpenAccess-Angebot einer Bibliothek vorgestellt. Dabei werden organisatorische und finanzielle Aspekte erläutert und – basierend auf der erfolgten Weiterentwicklung – Arbeitsprozesse sowie zentrale Funktionen des neuen Webauftritts beschrieben.

Abstract (E):

E-Periodica is an ETH-Bibliothek platform for digitized Swiss journals from the fields of science, history and culture. Since its launch in 2007, the service has been expanded continuously to include new publications. Due to the on-going growth in data and in order to keep the platform running in the long-term, E-Periodica's IT infrastructure underwent a complete overhaul. This makeover also included a redesign of its online presence with a focus on responsive design for mobile devices. This paper presents E-Periodica as a library's reliable open access service. It explains organizational and financial aspects and describes the work processes and central functions of the new-look website, based on the new design.

1 Einleitung

E-Periodica¹ ist eine Open Access-Plattform der ETH-Bibliothek² für digitalisierte Schweizer Zeitschriften aus den Bereichen Wissenschaft, Geschichte und Kultur. Der Zeitrahmen der aufgeschalteten Publikation erstreckt sich vom 18. Jahrhundert bis hin zu aktuellen Ausgaben. Zentrale Dienstleistung von E-Periodica ist es, einen einfachen und kostenlosen Zugriff auf die Zeitschriften zu gewährleisten. Oftmals können auch vergriffene Hefte und Jahrgänge auf diese Weise wieder zur Verfügung gestellt werden. Im Fokus stehen dabei Publikationen von nicht-kommerziellen Verlagen oder Herausgebern. Im April 2016 standen über 380 Zeitschriftentitel mit insgesamt rund 4.6 Millionen Seiten online zur Verfügung. Grundsätzlich ist bei sämtlichen Zeitschriften der Volltext frei zugänglich und kann als PDF heruntergeladen werden. Es werden keine Nutzungsgebühren erhoben. Je nach Vereinbarung mit dem jeweiligen Rechteinhaber können jedoch neuere Ausgaben bis zu maximal fünf Jahren gesperrt werden.

Die Zielgruppen von E-Periodica umfassen Personen aus unterschiedlichen Interessensbereichen und Altersklassen:

- Studierende, Wissenschaftler aus der ganzen Schweiz und aus Europa, die für wissenschaftliche Arbeiten und Studien recherchieren
- Journalisten und Journalistinnen auf der Suche nach Hintergrundinformationen
- Abonnenten und Abonnentinnen einer Zeitschrift, die die Publikation online konsultieren
- Allgemein Interessierte, die für Beruf persönliche Interessen und Hobbys auf Informationssuche sind

Im Folgenden soll E-Periodica als konkretes Praxisbeispiel im Bereich der Bereitstellung digitalisierter Zeitschriften vorgestellt werden. Besonderes Augenmerk liegt dabei auf zwei Neuerungen, die im Rahmen einer technischen Weiterentwicklung implementiert wurden:

- Vereinfachung der Arbeitsprozesse durch eine workflowgesteuerte Zeitschriftenaufbereitung
- ein umfassendes Redesign für die Gewährleistung eines zeitgemässen Auftritts

Zudem wird auf Arbeitsprozesse sowie auf strategische und organisatorische Belange eingegangen, um einen möglichst umfassenden Überblick über das Angebot E-Periodica zu gewährleisten.

¹ www.e-periodica.ch (bis 2015: <http://retro.seals.ch>)

² ETH-Bibliothek: <http://www.library.ethz.ch/>

2 Geschichte

Konzipiert wurde das Angebot mit dem ursprünglichen Namen retro.seals.ch im Rahmen des mittlerweile abgeschlossenen Innovations- und Kooperationsprojektes e-lib.ch³, Elektronische Bibliothek Schweiz. Die Durchführung des Projekts retro.seals.ch erfolgte zuerst durch das *Konsortium der Schweizer Hochschulbibliotheken*⁴ in Zusammenarbeit mit der ETH-Bibliothek. Seit 2013 hat die ETH-Bibliothek die Gesamtverantwortung für die Plattform vollständig übernommen und ist somit allein für deren Organisation und Betrieb zuständig.

Beim Start der Plattform im Jahr 2007 konzentrierte sich das Angebot auf Architekturzeitschriften der Schweiz – das so genannte „Baugedächtnis“ – sowie mathematische und naturwissenschaftliche Publikationen. Nach und nach erfolgten eine thematische Erweiterung und damit auch ein entsprechender Ausbau des Angebotes, so dass mittlerweile ein breites Spektrum an Themen und Inhalten zur Verfügung steht. [Retro.seals.ch](http://retro.seals.ch) war ursprünglich nicht für ein derart grosses und stetig wachsendes Zeitschriftenangebot konzipiert. Dieses enorme Datenvolumen stellte denn auch immer höhere Anforderungen an die IT-Infrastruktur. Um den Betrieb der Plattform langfristig gewährleisten zu können, hat die ETH-Bibliothek in Zusammenarbeit mit der Software-Betreiberfirma SRZ Berlin⁵ die Applikation umfassend weiterentwickelt. Ebenfalls im Rahmen dieser Weiterentwicklung erfolgten ein Redesign sowie die Umbenennung der Plattform von retro.seals.ch zu E-Periodica.

3 Organisation

3.1 Gesamtverantwortung ETH-Bibliothek

Sämtliche Arbeiten im Bereich Digitalisierung und Datenstrukturierung, Betrieb der IT-Infrastruktur der Plattform sowie die Speicherung der Daten erfolgen an der ETH-Bibliothek respektive an der ETH Zürich. Die Organisation und Koordination von E-Periodica ist im DigiCenter der ETH-Bibliothek angesiedelt. Die Betreuung der Infrastruktur erfolgt durch IT-Mitarbeitende der ETH-Bibliothek. Für die operativen Arbeiten im Bereichen Scannen, Qualitätskontrolle und Strukturierung sind hauptsächlich studentische Mitarbeitende im Einsatz.

³ Elektronische Bibliothek Schweiz: <http://www.e-lib.ch/>

⁴ Konsortium der Schweizer Hochschulbibliotheken: <http://lib.consortium.ch/>

⁵ SRZ Berlin: <http://www.srz.de/>

3.2 Kooperation mit der Schweizerischen Nationalbibliothek

Die Plattform wird in Zusammenarbeit mit der Schweizerischen Nationalbibliothek betrieben. Diese wählt selbst Zeitschriften aus ihrem Bestand aus, klärt die Rechte, lässt die Zeitschriften bei externen Dienstleistern digitalisieren und erschliesst die Daten wiederum intern mit dem bei E-Periodica eingesetzten XML-Editor. Schliesslich werden die komplett vorbereiteten Datensätze zur Aufschaltung an die ETH-Bibliothek geliefert.

3.3 Vertragspartner

Die Aufnahme einer neuen Zeitschrift in E-Periodica wird mit den involvierten Parteien – das sind beispielsweise Rechteinhaber, Herausgeber, Redaktionen – anhand eines schriftlichen Vertrages geregelt (s. Kap. 4.1). Ausnahmen werden diesbezüglich nur bei verwaisten älteren Publikationen gemacht, bei denen keine Rechteinhaber mehr ermittelt werden können.

3.4 Aufnahmekriterien

Während die Plattform zu Beginn allein auf wissenschaftliche Zeitschriften fokussierte, wurde diese Beschränkung schon bald zugunsten eines breiteren Angebotes gelockert. Entscheidungsgrundlage für die Aufnahme einer neuen Zeitschrift bilden grundsätzlich folgende Kriterien:

- zeitschriftenartiger Charakter der Publikation (regelmässige Erscheinungsweise)
- Einverständnis der Rechteinhaber (sofern es sich nicht um eine verwaiste Publikation handelt)
- Bezug zur Schweiz (z.B. thematisch, Sitz des Herausgebers)
- Grundlagenmaterial für die Forschung
- von allgemeinem Interesse
- nicht-kommerzieller Charakter
- kein gleichwertiges Online-Angebot an anderer Stelle vorhanden

3.5 Rechtliche Aspekte

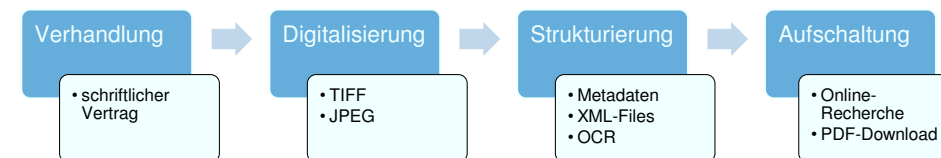
Die ETH-Bibliothek besitzt ausschliesslich die Rechte zur Digitalisierung und Aufschaltung der Zeitschriften im Rahmen von E-Periodica. Andere Rechte wie etwa Urheber- oder Nutzungsrechte verbleiben vollumfänglich bei den Rechteinhabern der Publikation – meist handelt es sich dabei um die Herausgeberschaft. Sollte ein Autor oder eine Autorin nach Aufschaltung der Publikation die Veröffentlichung seines oder ihres Artikels auf E-Periodica ablehnen, würde dieser einzelne Beitrag entfernt und durch einen Platzhalter ersetzt.

Sämtliche Autorenrechte im Vorfeld abzuklären, wäre jedoch unverhältnismässig und würde ein Zeitschriftenprojekt von vornherein zum Scheitern verurteilen.

Die angebotenen Inhalte stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Dateien oder Ausdrücke aus dem Angebot können zusammen mit den Nutzungshinweisen und den korrekten Herkunftsbezeichnungen weitergegeben werden. Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber. Das Setzen von Links auf Seiten aus dem Angebot von E-Periodica ist jedoch jederzeit möglich.

4 Prozesse E-Periodica

In den folgenden Abschnitten geht es um Prozesse, Standards und technische Lösungen von E-Periodica. Anhand eines Prozessmodells zur Aufnahme einer neuen Zeitschrift werden die einzelnen Schritte im Detail beleuchtet:



4.1 Verhandlung

Entspricht eine Zeitschrift den Kriterien und wird in E-Periodica aufgenommen, werden zusammen mit den Verantwortlichen der Zeitschrift die wichtigsten rechtlichen, organisatorischen und finanziellen Belange geklärt und in einem Vertrag festgehalten. Nach Unterzeichnung des Vertrages durch alle beteiligten Parteien ist eigentlicher Projektstart. Über das Digitalisierungsvorhaben wird auf *digicoord*⁶, einer Informationsplattform für schweizerische Digitalisierungsprojekte, informiert.

4.1.1 Kostenmodell

Die Kostenberechnung für ein Zeitschriftenprojekt erfolgt basierend auf der Gesamtseitenzahl aller erschienenen Jahrgänge.

⁶ Digicoord: Informationsplattform zu Schweizerischen Digitalisierungsprojekten: <https://www.digicoord.ch/>

Die Kosten werden unterteilt in:

- Einmalige Projektkosten: diese beinhalten die Aufschaltung des Retrobestandes einer Zeitschriften mit den entsprechenden Arbeitsschritten.
- Jährlich wiederkehrende Kosten: diese setzen sich zusammen aus Hosting der Daten und Aufschaltung der neuesten Ausgaben (sofern eine entsprechende Aktualisierung vereinbart wurde).

Die einmaligen Projektkosten werden zu einem grossen Teil von der ETH-Bibliothek mitfinanziert und die jährlich wiederkehrenden Kosten – die erheblich geringer ausfallen – vollständig den Vertragspartnern weiterverrechnet. Es gibt zudem Dachorganisationen⁷ oder andere Bibliotheken⁸, die die Kosten ganz oder teilweise übernehmen. Dieses Kostenmodell soll eine übermässige finanzielle Belastung der Vertragspartner verhindern. Dadurch können auch Interessenten mit geringerem Budget an E-Periodica teilnehmen, was wiederum zu einer grösseren Zeitschriftenvielfalt auf der Plattform führt.

4.2 Digitalisierung

Die Zeitschriften werden im DigiCenter der ETH-Bibliothek digitalisiert. Nach der Digitalisierung erfolgt eine standardisierte Qualitätskontrolle der Scans. Diese beinhaltet die Prüfung auf Vollständigkeit, Farbqualität und Schärfe. Nach Bedarf wird (pro Jahrgang) eine automatisierte Bildbearbeitung durchgeführt. Dazu werden zwei verschiedene Softwares – PageImprover⁹ und ScanTailor¹⁰ – eingesetzt. In Ausnahmefällen werden auch bereits von Dritten erstellte Digitalisate übernommen, sofern sie den grundlegenden Standards von E-Periodica entsprechen.

4.3 Rollenbasierter Workflow: Strukturierung

Im Anschluss an Digitalisierung und Qualitätskontrolle erfolgt die manuelle Strukturierung der Zeitschrift. Im Rahmen der Weiterentwicklung der IT-Infrastruktur von E-Periodica wurde in Zusammenarbeit mit dem SRZ Berlin ein maßgeschneiderter Workflow-Client entworfen, welcher direkt mit dem Workflow Server kommuniziert und eine zentrale Steuerung und Verwaltung der Zeitschriften- resp. Datenbearbeitung ermöglicht (s. Darstellung Server-Architektur im Anhang, Abb. 1).

⁷ Schweizerische Akademie der Geisteswissenschaften: <http://www.sagw.ch/sagw.html>

⁸ Universitätsbibliothek Bern: <http://www.unibe.ch/universitaet/dienstleistungen/universitaetsbibliothek/>,
Bibliothek am Guisanplatz, Bern: www.guisanplatz.ch/, Kantonsbibliothek Graubünden, Chur:
www.kantonsbibliothek.gr.ch

⁹ PageImprover, 4digitalbooks: http://www.4digitalbooks.com/_soft_imaget.html

¹⁰ ScanTailor: <http://scantailor.org/>

In den folgenden Kapiteln werden diese Arbeitsprozesse erläutert. Ein Diagramm des kompletten Strukturierungs-Workflows befindet sich ebenfalls im Anhang (Abb. 2).

4.3.1. Erfassen der Zeitschrift

Die Zeitschrift wird vom Administrator in der Applikation „Workflow-Client“ mit den wichtigsten Metadaten, wie zum Beispiel Titel der Zeitschrift, ISSN-Nummer, Herausgeber, Erscheinungsjahre, erfasst.

4.3.2 Laden der TIFF-Dateien

Ist die Zeitschrift im System erfasst und liegen die TIFF-Dateien für den Import bereit, kann der Administrator diese im Workflow-Client laden. Dazu öffnet er den Import-Dialog, welcher ihm sämtliche Jahrgangsorder der entsprechenden Zeitschrift auf dem *Network-Attached-Storage* (NAS)¹¹ zu Auswahl anzeigt. Die Auswahl eines oder mehrerer Ordner für den Import löst nun im Hintergrund verschiedene Prozesse aus:

- Für jeden Jahrgang wird zuerst ein Struktur-XML-File erstellt, welches die im vorherigen Schritt erfassten Metadaten der Zeitschrift enthält.
- Ein weiterer Prozess betrifft die Generierung von diversen JPEG-Derivaten aus den TIFF-Dateien, welche sowohl für die Strukturierung im XML-Editor, als auch für die Online-Präsentation, für Zoom-Funktion und PDF-Download benötigt werden.
- Parallel wird durch das Laden der TIFF auch der OCR-Server angestoßen. Dieser erstellt anhand der TIFF-Dateien die Volltexte. Je nach Konfiguration werden Antiqua-, Fraktur- oder gemischte OCR generiert.

Nachdem sämtliche Derivate eines Zeitschriften-Jahrgangs erstellt sind, befindet sich dieser Jahrgang im Status „Bereit zur Erschliessung“ und kann strukturiert werden.

4.3.3 Struktur erfassen

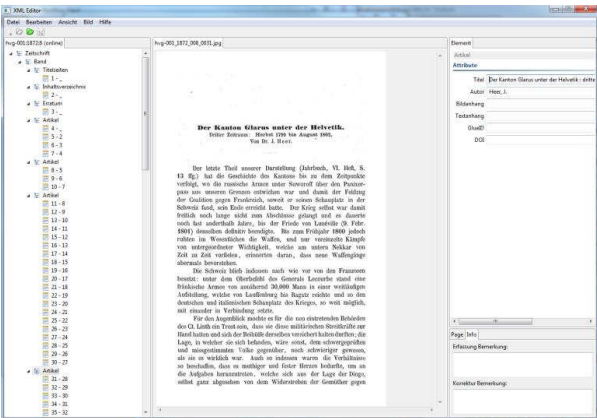
Zur Erfassung der Struktur wird der XML-Editor, ebenfalls eine Java-Anwendung, verwendet. Genau wie der Workflow-Client ist auch diese Applikation zentral auf einem NAS abgelegt und wird über ein EXE-File gestartet. Hat sich die Mitarbeiterin oder der Mitarbeiter mit den Benutzerdaten im XML-Editor angemeldet, baut dieser eine Verbindung zum Workflow-Server auf und kommuniziert – ebenfalls wie der Workflow-Client – mit diesem über eine REST-Schnittstelle.¹²

¹¹ Netzgebundener Speicher: In einem Rechnernetz zur Verfügung gestellter Speicherplatz

¹² REST – Representational State Transfer: Programmierparadigma für verteilte Systeme.

Über ein Popup-Menü kann nun die zu bearbeitende Zeitschrift ausgewählt werden. Dies kann eine neu zu strukturierende Zeitschrift, eine sich in Arbeit befindende oder eine zu korrigierende sein. Je nachdem, zu welcher Arbeitsgruppe die Mitarbeiterin oder der Mitarbeiter in der Strukturierung gehört, werden Zeitschriften in den verschiedenen Status zur Auswahl angeboten.

Nachdem eine Zeitschrift zur Bearbeitung ausgewählt wurde, präsentiert sich der XML-Editor wie folgt:



XML-Editor: Erfassen von Strukturmetadaten

Im linken Bereich sieht man die Baumstruktur des Inhalts. Im mittleren Fenster wird jeweils das JPEG der links selektierten Seite angezeigt. Und im rechten Bereich werden die Struktur-Metadaten erfasst. Die manuelle Strukturierung beinhaltet unter anderem:

- Eingabe von Autoren und Artikelstiteln
- Paginierung
- Kennzeichnung der Art des Inhaltes mit so genannten *Elementen* wie zum Beispiel Artikel, Vorwort, Buchbesprechung, Nachruf, Werbung, etc.
- Verknüpfung von Anhängen (beispielsweise Tabellen oder Tafeln) mit dem zugehörigen Artikel

Ist die Strukturierung beendet, befindet sich der Zeitschriften-Jahrgang im Status „Erschlossen“ und kann im nächsten Workflow-Schritt überprüft werden. Hier wird das Vier-Augen-Prinzip angewendet, das heisst, eine andere Mitarbeiterin oder ein Mitarbeiter überprüft die strukturierte Zeitschrift und gibt diese in der Folge entweder für die interne Aufschaltung frei (Status „Erschliessung geprüft“), oder beantragt eine Korrektur (Status „Korrektur beauftragt“).

4.4 Aufschaltung auf E-Periodica

Vor der definitiven Aufschaltung wird die Zeitschrift zuerst in einer Testumgebung anhand von Stichproben und mittels standardisierten Fragen zu verschiedenen Ansichten und Funktionen geprüft. Werden bei dieser Qualitätskontrolle noch Fehler entdeckt, werden diese im XML-Editor korrigiert. Nachdem die Fehler behoben sind, wird die Zeitschrift aufgeschaltet und steht ab sofort auf www.e-periodica.ch zur Verfügung.

4.4.1 Kommunikation, Nachweise, Recherche

Über Neuaufschaltungen wird jeweils auf ETH-News-Kanälen und über Mailinglisten informiert. Meist erfolgt auch eine Ankündigung der Herausgebenden selbst, die auf ihren Webseiten und/oder in einer aktuellen Zeitschriftenausgabe über die Aufschaltung ihrer Zeitschrift auf E-Periodica berichten.

Die Zeitschriftenbestände von E-Periodica sind über verschiedenen Kataloge und Datenbanken recherchierbar:

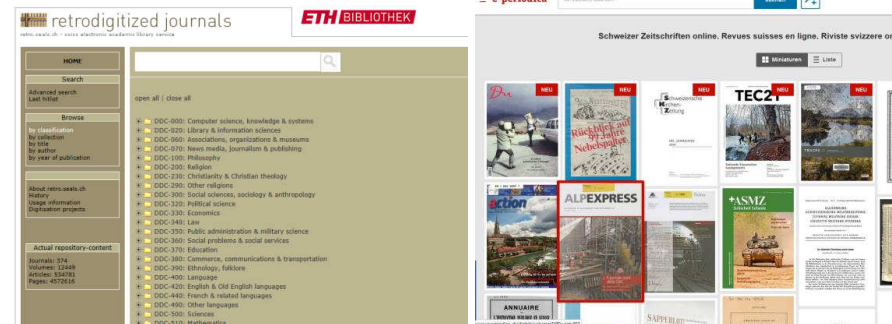
- Wissensportal der ETH-Bibliothek: <http://www.library.ethz.ch/>
- NEBIS-Katalog - Netzwerk von Bibliotheken und Informationsstellen in der Schweiz: <http://www.nebis.ch/>
- swissbib.ch - Katalog aller Schweizer Hochschulbibliotheken, der Schweizerischen Nationalbibliothek, zahlreicher Kantonsbibliotheken und weiterer Institutionen: <https://www.swissbib.ch/>
- Elektronische Zeitschriftenbibliothek: <http://rzblx1.uni-regensburg.de/ezeit/>

4.5 Erstellung der Archiv-Kapseln

Der letzte Schritt im ganzen Workflow-Prozess betrifft die Erstellung der Archivkapseln. Für jeden abgeschlossenen Zeitschriften-Jahrgang wird eine unkomprimierte ZIP-Datei, bestehend aus den originalen TIFF-Dateien, den OCR-Dateien und dem Struktur-XML, erstellt. Werden nach der Archivierung Korrekturen an einem Zeitschriften-Jahrgang vorgenommen – seien dies Modifikationen in den Struktur- oder Bibliographischen Metadaten, oder in den Bildern und OCR-Dateien – wird eine Delta-Archivkapsel generiert, die ausschliesslich die geänderten Daten enthält.

5 Webauftritt

Der ursprüngliche Webauftritt von retro.seals.ch war im Verlauf des Bestehens der Plattform nur marginal verändert worden und entsprach allmählich nicht mehr den zeitgemässen Standards.



Homepage retro.seals.ch (bis 2015)

Homepage E-Periodica (ab 2016)

Im Rahmen der technischen Weiterentwicklung der Plattform erfolgten deshalb eine Neugestaltung der Webseite sowie die Umbenennung der Plattform von retro.seals.ch zu E-Periodica. Zentrale Neuerungen des umfassenden Redesigns – entworfen und umgesetzt durch die Webdesign-Firma Cando¹³ – sollen an dieser Stelle etwas ausführlicher erläutert und die zentralen Features vorgestellt werden.

Kernkonzept ist die Aufbereitung der Inhalte in responsivem Design. Dieses Konzept berücksichtigt die geringere Grösse der Bildschirme mobiler Endgeräte – Tablets und Smartphones – im Gegensatz zu Desktop-Bildschirmen und enthält ein flexibles Gestaltungsraster für unterschiedliche Bildschirme. Inhaltlich wird dabei auf Kernfunktionen und auf die zentralen Aussagen der Webseite fokussiert. Weiterführende Informationen zur Webseite und deren Inhalte werden in so genannten „Off-Canvas-Elementen“ untergebracht, die bei Bedarf eingeblendet werden können. Dadurch können alle benötigten Informationen zur Verfügung gestellt werden, ohne die Gestaltung und damit die Lesbarkeit zu beeinträchtigen.

Wichtiges Kriterium war ausserdem die Barrierefreiheit der Webseite, deren Umsetzung erfolgt beispielsweise in Form von Alternativtiteln für Bilder, die sich auch als Listenelementen darstellen lassen. Benutzende mit Beeinträchtigungen können sich zudem die Texte vorlesen lassen.

¹³ Cando Image GmbH: <https://www.cando-image.com/>

5.1 Startseite

Die Startseite ermöglicht einen sehr visuellen Einstieg: Jede Zeitschrift wird mit dem jeweils neuesten Cover präsentiert. Um in diesem Zusammenhang Performanceprobleme beim Laden der Startseite zu vermeiden – es handelt sich mittlerweile doch um eine beträchtliche Menge an Bildern – wird ein so genanntes „lazy loading“ angewendet. Dabei handelt es sich um ein optimiertes Ladeverfahren:

„Eine spezielle Möglichkeit, den Seitenaufbau zu beschleunigen, bietet das sogenannte Lazy Loading. Dabei verzichtet man bei bestimmten Komponenten bewusst darauf, sie mit dem ersten Hauptladevorgang zu laden. Stattdessen lädt man sie lazy, also träge nach, sobald der Hauptladevorgang vorüber ist und der dadurch entstandene Aktivitätensturm sich gelegt hat.“ (Zillgens, S. 345).

Das heisst, die Bilder resp. Zeitschriftencovers werden auf E-Periodica erst geladen, wenn sie im Verlauf des Scrollens angezeigt werden. Es besteht aber auch die Möglichkeit, auf eine Listenansicht mit alphabetischer Titelliste ohne Abbildungen umzuschalten.

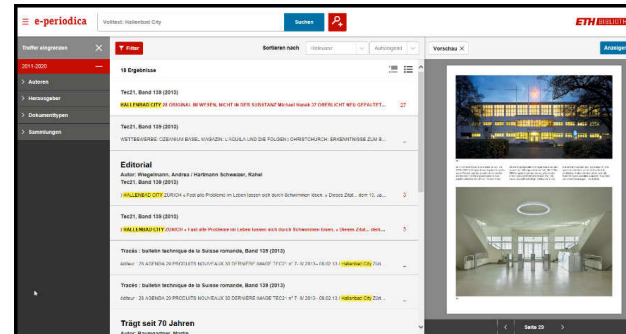
5.2 Recherche und Ansichtsoptionen

Das Recherchetool bietet eine kombinierte Such- und Filteroption. Das heisst, bei der einfachen Suche reagiert der Filter dynamisch, indem während der Eingabe der Suchbegriffe im Suchfeld die Auswahl der gefundenen Zeitschriften fortlaufend gefiltert und angeordnet wird („Masonry-Style“¹⁴). Bei Auswahl einer Zeitschrift gelangt man auf deren Bandübersicht (ein Band entspricht in der Regel einem Jahrgang) und von dort weiter auf das manuell erfasste Inhaltsverzeichnis des betreffenden Jahrganges mit der zugehörigen Seitansicht.

5.2.1 Split-View: Zerteilte Anzeige

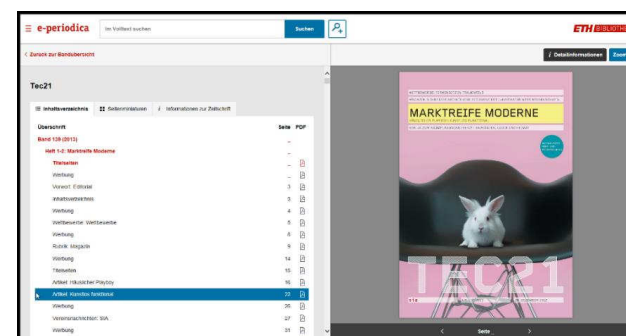
Über die Suche mittels Recherchetool gelangt man auf eine so genannte ‚Split-View‘. Hier wird die Trefferliste zugleich mit einer Bildvorschau angezeigt. Dabei sind die Suchbegriffe sowohl in der Trefferliste als auch in der Vorschau markiert. Für die Einschränkung der Anzahl Treffer stehen Facetten (Autoren, Zeiträume, Herausgeber, Dokumenttypen und Sammlungen) zur Verfügung.

¹⁴ Masonry-Style: <http://masonry.desandro.com/>



Split-View mit Trefferliste und Seitenansicht

Wählt man den Weg zum Volltext über die Browsing-Option auf der Homepage, gelangt man ebenfalls auf einen Split-View. In diesem Fall wird neben der Seitenansicht das im XML-Editor manuell erfasste Inhaltsverzeichnis des jeweiligen Zeitschriftenjahrgangs angezeigt. Hier kann in der Zeitschrift geblättert werden, während die Inhaltsübersicht gewährleistet bleibt und dadurch unmittelbar auch auf andere Artikel zugegriffen werden kann. Alle Seiten eines Bandes können zudem als Miniaturen dargestellt werden und es stehen in einem weiteren Register allgemeine Informationen zur jeweiligen Zeitschrift, wie zum Beispiel Herausgeber oder Kontaktmöglichkeiten, zur Verfügung.



Split-View mit Inhaltsverzeichnis und Seitenansicht

5.2.2 Vollbild- und Zoomansicht

Von der einfachen Seitansicht gelangt man auf die Vollbildansicht. Dazu wird der erweiterbare OpenSource-Viewer *OpenSeaDragon*¹⁵ eingesetzt. Der Zoom erfolgt stufenlos und auch in der gezoomten Detailansicht ist blättern möglich. Bei mobilen Endgeräten kann sowohl bei der einfachen Seitenansicht als auch in der Zoomansicht mittels „Swipen“ (Wischen) geblättert werden.

¹⁵ OpenSeaDragon: <https://openseadragon.github.io/>

5.3 Verwenden und Teilen der Inhalte

PDF-Download: Die Volltexte können als PDF heruntergeladen werden. Das PDF wird dabei jeweils direkt beim Download aus den vorhandenen JPEG-Dateien erstellt. (Aus Speicherplatzgründen sind auf der Plattform keine PDF-Dateien hinterlegt). Zusätzlich wird beim Download jeweils ein Deckblatt mit den wichtigsten Metadaten sowie den Nutzungsbestimmungen generiert und als erste Seite direkt in das PDF eingehängt.

DOI: Die bei der Strukturierung als *Artikel* gekennzeichneten Inhalte werden zur dauerhaften Auffindbarkeit mit einem Digital Object Identifier versehen. Registriert werden diese über den DOI-Desk der ETH Zürich.¹⁶

Social Media: Die Artikel können über Twitter, Facebook, Google+ und Pinterest geteilt werden.

5.4 Nachweis gesperrter Ausgaben

Wie weiter oben erwähnt, können neue Ausgaben einer Zeitschrift für maximal fünf Jahre gesperrt werden. Diese gesperrten Ausgaben werden in E-Periodica zwar aufgeschaltet, aber der Zugriff auf den Volltext wird nicht freigeschaltet. Gesperrte Ausgaben sind jeweils mit einem Schloss-Icon und einem entsprechenden Hinweis gekennzeichnet. Nach Ablauf der Sperrfrist werden die Volltexte automatisch freigeschaltet.

6 Fazit

E-Periodica hat sich als erfolgreiche Plattform im Bereich Zeitschriftendigitalisierung bewährt. Die stark angestiegenen Zugriffszahlen sowie zahlreiche Rückmeldungen von Nutzenden zeigen, dass das Angebot rege genutzt und geschätzt wird. Ursprünglich nicht für einen so grossen Zeitschriftenumfang entworfen, mussten jedoch Massnahmen getroffen werden, um dem stetig wachsenden Datenvolumen gerecht zu werden. Nach der erfolgten technischen Weiterentwicklung kann nun ein langfristiger Betrieb der Plattform gewährleistet werden. Der neue, rollenbasierte Workflow ermöglicht eine effiziente Durchführung sämtlicher Arbeitsschritte – von der Übernahme der Scans, der Generierung der Derivate und Volltexte, über die Strukturierung bis zur Aufschaltung und Archivierung der Daten. Das responsive Design schliesslich erlaubt einen modernen und zeitgemässen Zugriff auf die Inhalte.

7 Quellen

Zillgens, Christoph. Responsive Webdesign. München: Hanser, 2013.

¹⁶ DOI-Desk, ETH Zürich: <https://www.library.ethz.ch/de/Dienstleistungen/Publizieren-registrieren-verwalten/DOI-Desk-der-ETH-Zuerich>

8 Anhang

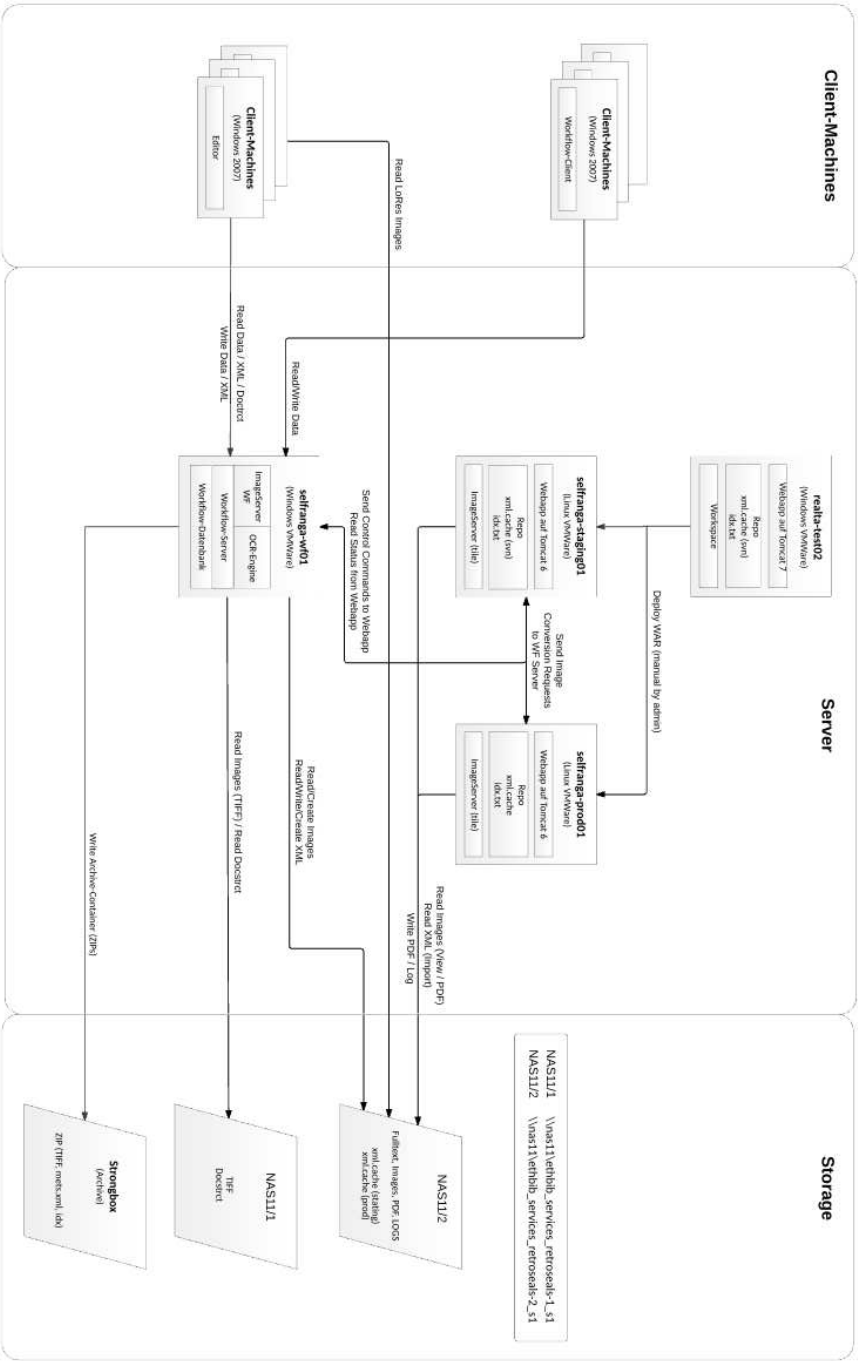


Abb. 1: Server-Architektur E-Periodica (ETH-Bibliothek, 2015)

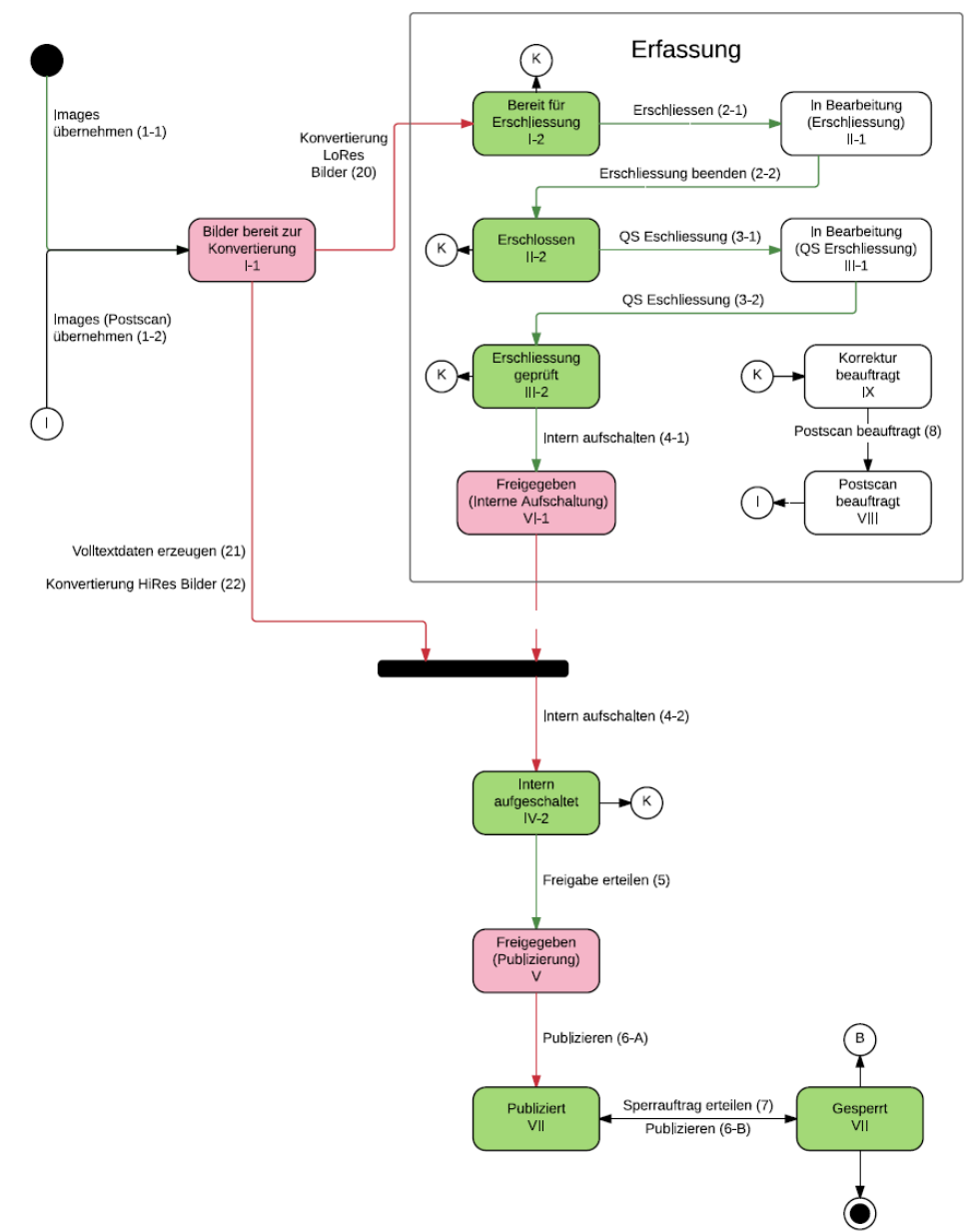


Abb. 2: Workflow-Client, Strukturierungsworkflow E-Periodica (ETH-Bibliothek 2015)

Green Open Access im Bibliothekskatalog - Chancen & Risiken

Martin Blenkle, Rachel Ellis, Elmar Haake, Manfred Nölte

Staats- und Universitätsbibliothek Bremen

Bibliotheken bewerben seit etwa 15 Jahren intensiv Open Access Modelle als neue offene Publikationsformen im Wissenschaftsbereich. Dennoch wird das Thema eines nutzerfreundlichen Nachweises solcher Open Access Dokumente in Bibliothekskatalogen immer noch kaum beachtet. Um die Akzeptanz grüner Publikationsmodelle nachhaltig zu fördern, ist es jedoch sinnvoll, Nutzer*innen die wertvollen Nachweise der frei verfügbaren wissenschaftlichen Dokumente auch im Gesamtkontext bibliothekarischer Kataloge bei der Suche anzubieten. Sollen die Förderung von Open Access Texten und die Gleichstellung mit lizenziertem Material aus klassischen Publikationsformen wirklich ernsthaft betrieben werden, ist die Integration in die Standardnachweissysteme der Bibliotheken unausweichlich, um die Sichtbarkeit dieser Medien für den Wissenschaftsbetrieb zu erhöhen. Dagegen besteht bisher immer noch vorherrschend das Konzept des universitären Dokumentenservers als Inselangebot mit eigener Suchoberfläche, das erfahrungsgemäß von Nutzer*innen aber nur selten direkt angesprochen wird.

Die Recherche über ein breites Gesamtangebot an grünen Open Access Medien ist bisher nur über gängige Internetsuchmaschinen sowie über die Spezialsuchmaschine BASE der UB Bielefeld möglich. Eine integrierte bibliothekarische Gesamtsicht auf lokal vorhandene Printmedien, lokal lizenzierte E-Medien und freies Open Access Material ist bundesweit in den Bibliotheken bisher kaum verfügbar. Zwar stehen die Metadaten von weltweit derzeit etwa 4300 Dokumentenservern¹ seit etwa 15 Jahren via OAI-PMH Schnittstellen prinzipiell frei zur Nutzung bereit, doch scheint es viele Vorbehalte gegen die Integration solcher Nachweise in ein gemeinsames Angebot mit bibliothekarisch kontrollierten Metadaten zu geben.

So ergibt sich die Situation, dass viele Bibliotheken erhebliche Aufwendungen in die Förderung von Open Access Publikationsformen investieren und damit wertvolle Überzeugungsarbeit an den Universitäten leisten, gleichzeitig aber die Auswirkungen dieser Maßnahmen in den bibliothekarischen Nachweisangeboten kaum sichtbar sind.

An der Staats- und Universitätsbibliothek (SuUB) Bremen ist der Nachweis frei verfügbarer wissenschaftlicher Texte bereits seit 2004 ein wichtiger Teil der Marketingstrategie von Open Access Publikationsmodellen. Das Angebot an Open Access Dokumenten im Rahmen der zentralen Bibliotheksuchmaschine erfolgt über das bereits seit 2004 an der SuUB Bremen verfügbare eigene Discoverysystem E-LIB.

¹ 4.285 Dokumentenserver – ca. 90 Mio. Nachweise Quelle BASE (Stand:21.04.2016)
https://www.base-search.net/about/de/about_sources_date_dn.php?menu=2 (Aufruf 21.04.16)

Die Elektronische Bibliothek Bremen (E-LIB) – ein Entwicklungssystem für neue Katalogdienste für wissenschaftliche Bibliotheken

Die E-LIB Bremen weist im Sinne moderner Discovery-Systeme E-Medien und Printmedien in einem lokalen Index nach und präsentiert die Titel gemeinsam in einer integrierten Trefferliste. Schon seit 2004 stehen auch Nachweise von Volltexten grüner Open Access Archive zur Verfügung. Die Open Access Daten werden dabei gemeinsam mit den bekannten Artikelangeboten klassischer Verlage völlig gleichberechtigt im Katalog angezeigt. Seit 2010 ersetzt das System den bisherigen klassischen Bibliotheks-OPAC als Zugangssystem komplett und wickelt derzeit etwa 13.000 bis 18.000 Suchanfragen täglich ab.^{2,3} Die E-LIB Bremen versteht sich dabei gleichzeitig als Produktivsystem und Entwicklungsumgebung für neue Services zur Unterstützung der Literatursuche von Nutzer*innen.

Im Gesamtindex der Bremer E-LIB finden sich derzeit ca. 60 Mio. Nachweise – davon mehr als 95% elektronische Volltexte. Die eingesetzte Katalogtechnik basiert auf einem lokal gepflegten Gesamtindex von heterogenen, aggregierten Metadaten. Der Index der Suchmaschine wird durch die Integration weiterer Metadatensammlungen im Sinne eines Data-Warehouse kontinuierlich erweitert. Die Auswahl der integrierten Metadatenbestände erfolgt nach inhaltlichen Kriterien und dient vordringlich der gezielten Informationsversorgung für den Standort der Universität und Hochschulen im Lande Bremen. Im Gegensatz zu der in Discoverysystemen meist eingesetzten Indexierungs- und Retrievalsoftware Apache Lucene nutzt die E-LIB Bremen das System CiXbase² zur internen Organisation der Metadaten, ein Indexierungs- und Recherchesystem unter Nutzung computerlinguistischer Verfahren.

² Blenkle, Martin; Ellis, Rachel; Haake, Elmar: Next-generation library catalogues: review of E-LIB Bremen. In: *Serials: The Journal for the Serials Community* 22 (2009), H. 2, S. 178-181. <http://uksg.metapress.com/content/040q4504p2565519/?id=040q4504p2565519> (Aufruf 21.04.16). Die Bibliothekssuchmaschine E-LIB ist in die Webpräsenz der SuUB Bremen vollständig integriert und verfügbar via <http://www.suub.uni-bremen.de/> (Aufruf 21.04.16).

³ Blenkle, Martin; Ellis, Rachel; Haake, Elmar: Only the first results count: user-feedback-modified relevance ranking in E-LIB Bremen. In: *Insights* 28 (2015), H. 2, S. 75–80. <http://doi.org/10.1629/22178> (Aufruf 21.04.16)

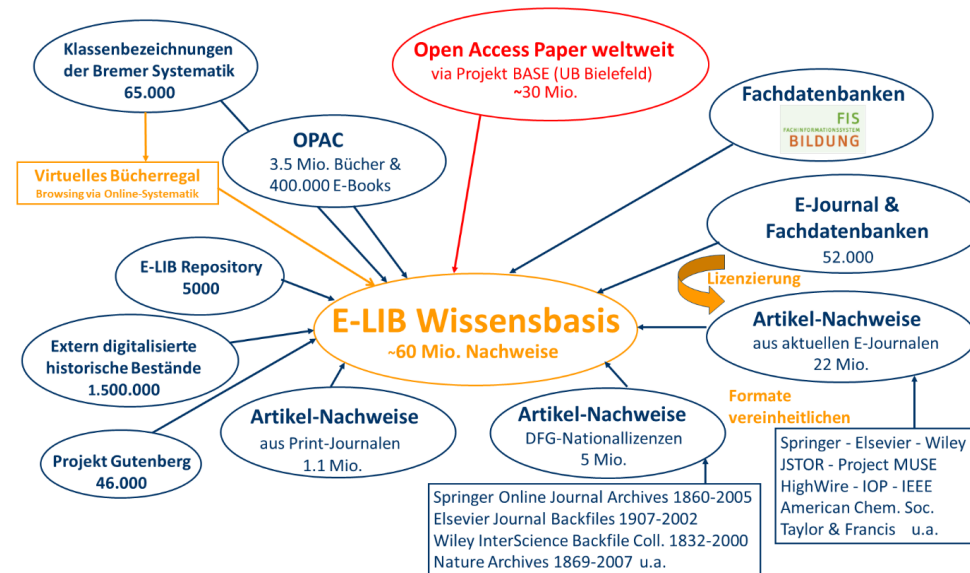


Abbildung 1: Wissensbasis der E-LIB Suchmaschine als Sammlung heterogener Metadatenpakete u.a. Open Access Nachweise

Integration von Metadaten von weltweit verfügbaren Open Access Medien

Seit der Bildung der Open Archives Initiative (OAI) im Jahr 2000 und dem zur Weitergabe von Archivmetadaten festgelegten, standardisierten Verfahren (OAI Protocol for Metadata Harvesting -OAI-PMH) ist es für Serviceanbieter möglich, Metadaten weltweiter Archive zu sammeln (harvesten) und Nutzer*innen zur Recherche anzubieten.

An der Staats- und Universitätsbibliothek wurde bereits im Jahr 2004 damit begonnen, den Suchindex der E-LIB um die Nachweise einzelner, für den lokalen Standort gezielt ausgewählter Archive selbst testweise zu harvesten. Dazu wurden die von etwa 80 Repositorien⁴ gesammelten Metadaten in den Suchindex der E-LIB integriert (ca. 3.5 Mio.) und neben dem lizenzierten Material größerer Verlagsanbieter bei der Recherche mitangeboten. Im Rahmen der experimentellen Nutzung ausgewählter Archive konnten viele nützliche Erfahrungen im Metadatenmanagement – sowohl beim Sammeln und aggregieren als auch beim Anbieten von Metadaten freier Dokumente innerhalb der Katalogsicht – gesammelt werden.

⁴ Für das Harvesten ausgewählt wurden damals regionale, nationale und internationale Repositorien wie z.B. Alfred-Wegener-Institut Bremerhaven, größere deutsche und internationale Universitätsarchive, CERN, ETH Zürich, PubMed- und BioMedCentral, ArXiv.org, CiteSeer, Cogprints, PhysNet u.a.

Obwohl es bereits früh erkennbar war, dass das Angebot an zusätzlichen Online Medien von den Nutzer*innen gut angenommen wurde, zeigte die Pilotphase, dass der investierte Aufwand an Personalressourcen nicht unerheblich war. Eine Ausweitung des Angebotes durch Vergrößerung der Anzahl an geharvesteten Archiven kam daher aufgrund der verfügbaren Personalkapazität zunächst nicht in Frage.

Eine deutliche Vergrößerung der Nachweisbasis an OAI Metadaten im Angebot der E-LIB gelang erst durch Nachnutzung der Metadatensammlung der wissenschaftlichen Suchmaschine BASE⁵, einem Projekt der UB Bielefeld. BASE verfolgt das Ziel, weltweit Metadaten aller wissenschaftlichen Repositorien, die Daten im Rahmen des Standards der Open Archives Initiative anbieten, zu sammeln, zu aggregieren und in einer Suchmaschine zur Recherche anzubieten.

Im Rahmen einer Kooperation steht seit 2012 der gesamte Metadatenbestand des Projekts BASE für die E-LIB zur Verfügung – insgesamt ca. 90 Mio. Metadaten. In der Bremer Suchmaschine wird derzeit eine Teilmenge von etwa 25 Mio. Nachweisen⁶ aus dem Gesamtbestand von BASE im integrierten Gesamtangebot von lokalen Printtiteln und lizenziertem Online-Material größerer Verlage angeboten (siehe beispielhaft Abbildung 2).

⁵ "BASE (Bielefeld Academic Search Engine) ist eine der weltweit größten Suchmaschinen speziell für frei im Sinne des Open Access zugängliche wissenschaftliche Dokumente im Internet." Derzeit umfasst die BASE Sammlung ca. 90 Mio. Nachweise von 4300 Repositorien – davon sind geschätzt 60% frei verfügbar (Open Access) (Stand April 2016). <https://www.base-search.net/about/de/index.php> (Aufruf 21.04.16)

⁶ Die Gesamtzahl an importierten BASE-Daten in der E-LIB Bremen schwankt kontinuierlich mit jeder Aktualisierung aufgrund der Filterung des Datenmaterials (siehe unten).

Open Access Medien im Bibliothekskatalog / Nutzungsaspekte

181

Nutzungshäufigkeit von Open Access Material in der Suchmaschine E-LIB

Eine Analyse der Nutzungshäufigkeit der importierten Metadaten im Vergleich zu anderen Angeboten des gemeinsamen Suchraumes der E-LIB Bremen ist nicht ganz einfach, da alle vergleichbaren Datentypen in unterschiedlicher Anzahl vorgehalten werden. Erschwerend kommt hinzu, dass die Metadaten eines jeweiligen Typs je nach Importkonditionen monatlich in schwankender Anzahl indexiert werden⁷.

Unter Berücksichtigung dieser Rahmenbedingungen erhält man einen ersten Hinweis auf die Nutzung der freien Inhalte durch den Vergleich der Gesamtnutzung aller angebotenen, insbesondere der lizenzierten E-Medien mit der Nachfrage nach den freien BASE/OAI-Angeboten (siehe Abbildung 3).

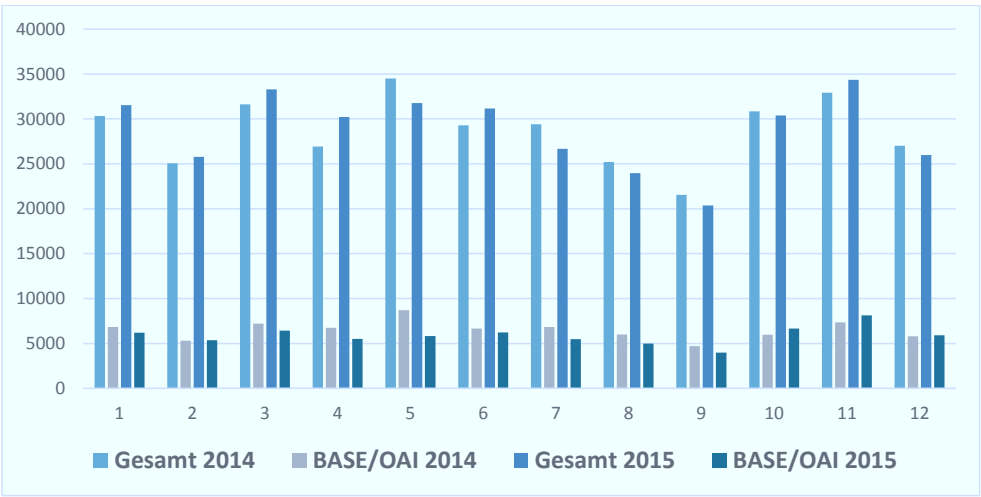


Abbildung 3: E-LIB Bremen: Informationen zur Nutzung freier Open Access Medien (Trend): Anteil an monatlichen Zugriffen auf freie Texte aus BASE/OAI-Metadaten im Vergleich zum Gesamtzugriff auf alle E-Ressourcen 2014/2015

Etwa 20% der E-Mediennutzung entfiel im Betrachtungszeitraum der letzten zwei Kalenderjahre auf die neuen kostenfreien Angebote – bei einer durchschnittlichen Nutzung von etwa 1000 täglichen Zugriffen auf alle E-Medien, lag der Anteil der Zugriffe auf BASE/OAI Daten damit bei etwa 200 am Tag.

⁷ Weiterhin sind nicht alle Indextypen der E-LIB Bremen - historisch bedingt - sauber differenzierbar: Die Metadaten des lokalen Bremer Bibliothekssystems (PICA) enthalten zwar überwiegend Printmaterial, aber auch zahlreiche E-Medien wie etwa E-Books und Metadaten von unselbständiger Literatur. Die Nutzung dieser E-Ressourcen bleibt bei dieser Betrachtung unberücksichtigt.

Probleme durch Metadaten mit falscher Open Access Zuordnung

Die Qualität der von vielen Anbietern bereitgestellten Metadaten bereitet den Anbietern von Suchdiensten in bestimmten Bereichen Probleme. Dies betrifft die Nachweise sowohl deutscher als auch weltweiter Open Access Archive - bezieht sich aber weniger auf klassische Qualitätsdefizite der formalen Beschreibung der Medien, sondern vielmehr auf den korrekten Status ihrer Verfügbarkeit.

Ein signifikanter Anteil der im Kontext der Open Archive Initiative von Repositorien angebotenen Metadaten beschreiben Medien, die nicht allgemein frei verfügbar sind und zudem oft auch keine Informationen über diesen Status ausweisen. Viele Repositorien dienen offensichtlich zugleich auch als lokale bibliografische Sammlungen, etwa zum Aufbau einer Universitätsbibliografie und weisen somit durchaus auch lizenzpflichtiges Material nach. Aus diesem Grunde finden sich in diesen Archiven Metadaten lizenzpflichtiger Verlagsangebote, die nicht frei zugänglich und mit der Grundphilosophie eines Repositoriums im Sinne der Open Access Initiative kaum in Einklang zu bringen sind. In vielen Fällen werden diese Daten dann über die OAI-PMH Schnittstellen irrtümlich als "frei verfügbar" definiert an die Datenharvester ausgegeben. Suchdiensteanbieter wie BASE oder E-LIB Bremen können diese Nachweise daher nicht sicher erkennen und gesondert anzeigen. Für Nutzer*innen von Suchmaschinen- und Discoverykatalogen sind diese *Pseudo-Open Access* Nachweise in der Regel nahezu wertlos.

Es gibt derzeit Ansätze auf Seiten der Dienstanbieter, solche fälschlich als frei verfügbar erhaltenen Metadaten durch intellektuelle Prüfung einzelner Repositorien in verschiedene Open Access Kategorien einzustufen und dann entsprechend zu behandeln. BASE kategorisiert Metadatenlieferanten nach ihrem prozentualen Anteil an frei verfügbarem Material und kennzeichnet dies in den Daten mit einem Attribut. An der Staats- und Universitätsbibliothek wird beim Import der BASE Daten dieses OA-Attribut ausgewertet. Archive mit einem geringen Anteil an frei verfügbaren Angeboten werden nicht in den E-LIB-Gesamtindex importiert und auf diese Weise ausgefiltert. Die Filterung ist nur auf Ebene der Archive möglich. Viele Archive mit einem hohen Anteil von lizenzpflichtigen Medien weisen jedoch auch viele freie Dokumente auf. Solange jedoch eine valide Kennzeichnung der einzelnen Medien nicht verfügbar ist, gehen viele interessante freie Nachweise durch ein zu restriktives Ausfiltern des gesamten Archivs verloren.

Viel sinnvoller wäre eine eindeutige Beschreibung der Medienverfügbarkeit in den von den Datenanbietern bzw. den Repositorien über die OAI-Schnittstellen ausgelieferten Metadaten.

Die Systeme sollten entweder nur Metadaten freier Dokumente über ihre Schnittstellen ausliefern, wie dies ursprünglich im Rahmen der Open Archive Initiative für offene Archive gedacht war. Falls dies nicht umsetzbar ist, sollten die Datenanbieter wenigstens ein entsprechendes Subset von Nachweisen nur freier Medien anbieten, wie dies im DINI Zertifikat 2013 als Mindestanforderung formuliert wird⁸. Zwar ist es sehr sinnvoll, die deutliche Kennzeichnung der Medienverfügbarkeit in den deutschen Standards zu verankern, eine wirkliche Verbesserung der Situation ist aber vorerst nicht in Sicht, da die weltweiten Archive sich kaum an derartigen nationalen Standards orientieren.

Um den Nutzer*innen möglichst wenig Nachweise mit unklaren Zugriffskonditionen anzubieten, werden die Metadaten vor der Integration in den Suchraum der E-LIB umfangreich analysiert und gefiltert (siehe unten).

Usability Probleme im Bereich der Medienbeschreibung / Repositorien Frontdoor Seite

Im Verlauf der Bereitstellung der Daten im Bibliothekskatalog machen sich weitere Nutzungshemmnisse beim Support bemerkbar, die durch ungünstige Gestaltung und fehlende Nutzerfreundlichkeit vieler Repositorien verursacht werden. Nach den in Bremen gemachten Erfahrungen im Zusammenhang mit den Dokumentenservern unterschiedlichster Anbieter werden Aspekte der Usability und der User Experience beim Design bibliothekarischer Repositorien noch zu wenig beachtet. Die oft ungünstig gestalteten Dokumenten-Frontdoor-Seiten⁹ zahlreicher Repositorien führen immer wieder zu Nachfragen.

⁸ Die AG Elektronisches Publizieren der deutschen Initiative für Netzwerkinformation erarbeitet mit dem DINI-Zertifikat, das inzwischen in der Version 2013 vorliegt, nationale Standards für den Betrieb wissenschaftlicher Dokumentenserver. <http://nbn-resolving.de/urn:nbn:de:kobv:11-100217162> (Aufruf 21.04.16). Aus dem **DINI Zertifikat 2013** S. 37: - A.2 OAI-PMH: Erweiterte Anforderungen - A.2.1 Set für Open-Access-Dokumente: Dienste veröffentlichen neben Open-Access-Dokumenten oftmals auch Dokumente, die nicht frei verfügbar, sondern beispielsweise nur für einen institutionell eingeschränkten Nutzungskreis zugänglich sind. Zur Erbringung von Mehrwertdiensten durch Service Provider ist es wichtig, Datensätze, die sich auf Open-Access-Dokumente beziehen, von solchen zu unterscheiden, für die das nicht gilt, und darüber hinaus eine vorab durchgeführte Selektion zu erlauben. **Mindestanforderung:** M.A.2-1 Es existiert ein Set mit der Bezeichnung (setSpec) „open_access“. Zu diesem Set gehören alle Metadatensätze, die sich auf Open-Access-Dokumente beziehen, d. h. bei diesen steht ein zugehöriger und verlinkter Volltext frei zur Verfügung.

⁹ Mit dem Begriff *Frontdoor* wird eine, in der Regel dem Volltext vorgeschaltete Eingangsseite eines Dokuments im Angebot eines Repositoriums bezeichnet, die alle wichtigen Metadaten und den Volltextlink bereitstellt.

Die aus Sicht der Nutzer*innen wichtigste Angabe, der Link zum Volltextdownload wird zuweilen durch umfangreiche Metadatenangaben verschleiert oder zu unauffällig in der Peripherie der Seite angeordnet.

Probleme durch schwankende Verfügbarkeit einzelner Archive

Durch die extreme Vielfalt der Datenlieferanten folgt auch eine große Bandbreite in der Zuverlässigkeit der Archive. Schwankende Verfügbarkeitsquoten einzelner stark nachgefragter Repositorien erhöhen den Supportaufwand in der Beratung signifikant.

Neue Dienste: Direktsupportoption am Dokument

Durch die besonderen Rahmenbedingungen der für Bibliotheken noch neuen Medienart der Open Access Texte ergeben sich auch Chancen für neue Formen des Bibliothekssupports bei der Katalognutzung. Die bereits geschilderten Nutzungshemmnisse einzelner Angebote führten an der Staats- und Universitätsbibliothek Bremen zur Entwicklung einer Direktsupportoption am jeweiligen Katalognachweis, um Fragen der Verfügbarkeit der Texte schnell und unkompliziert per Emailsupport von Bibliotheksmitarbeiter*innen klären zu können. Dazu wird in jeder Dokumentvollanzeige eines BASE/OAI-Nachweises ein Emailformular direkt beim Volltextlink angeboten, über den eine Anfrage generiert werden kann, die automatisch die angefragten Metadaten des Titels mitliefert.



Abbildung 4: Direktsupport-Link in der Medienvollanzeige von Open Access Medien in der E-LIB Bremen: "Probleme mit dem Link?"

Durch den kurzfristigen Support können oft Wege eröffnet werden, das vollständige Dokument aus anderer Quelle zu beschaffen.

Open Access Medien im Bibliothekskatalog / Technische Integration der Metadaten

Filterung der BASE Metadaten und Suchmaschinenindexierung

Die BASE/OAI-Daten liegen mit knapp 85.000 Dateien in einer flachen Verzeichnisstruktur in einem erweiterten OAI-DC-Format¹⁰ vor (180 GB unkomprimiert). Zur Identifikation der für den Import vorgesehenen Datensätze werden die Daten in Bezug auf die BASE-Verfügbarkeitsangabe im Tag "base_dc:oa" analysiert (1: open access; 0: unknown; 2: no open access) und Archive des Types 2 ausgefiltert. Darüber hinaus werden jeweils eine an der SuUB Bremen gepflegte Negativ- und Positiv-Liste mit weiteren Einträgen von Repositorien berücksichtigt, um auch die speziellen Anforderungen des eigenen Standortes berücksichtigen zu können. Die Blacklist korrigiert nachträglich den Wert von "base_dc:oa", so dass das Repository als "no open access" gewertet wird. Dagegen werden die Daten aller in der Whitelist enthaltenen Repositorien unabhängig vom "base_dc:oa" als open access Metadaten gewertet und importiert.

Im April 2016 wurden von ca. 85 Millionen Datensätzen etwa 25 Millionen für den Import in die E-LIB identifiziert. Ca. 6 Millionen wurden durch die Blacklist ausgefiltert, alle übrigen nicht berücksichtigten Archive gehören zum Typ BASE "no open access". Die Whitelist betrifft aktuell nur einen sehr geringen Anteil aller Datensätze (160.000).

Dieser Filterprozess wurde mit der Skriptsprache Perl und – nach ersten Laufzeittests mit den Bibliotheken XML::Twig¹¹ und XML::XPath¹² – endgültig mit XML::LibXML¹³ realisiert. Die Laufzeiten der Filterung unterscheiden sich bei dieser Datenmenge erheblich: So benötigte die erste Implementation auf Basis von XML::XPath ca. 80 Stunden, bei XML::Twig ca. 50 Stunden und mit XML::LibXML konnte die Laufzeit auf ca. 5 Stunden reduziert werden.

Im nächsten Schritt werden die Daten in den Index importiert. Alle verbliebenen relevanten 25 Mio. Metadatenätze aus den BASE/OAI-Daten werden mit Hilfe eines in ANSI-C geschriebenen, modular anpassbaren Importprogramms in 3 separate Index-Dateien¹⁴ in einer Laufzeit von 9 Stunden importiert.

¹⁰ OAI-DublinCore - <http://dublincore.org/documents/dces/> (Aufruf 21.04.16)

¹¹ XML::Twig <http://search.cpan.org/~mirod/XML-Twig-3.49/Twig.pm> (Aufruf 21.04.16)

¹² XML::XPath <http://search.cpan.org/~manwar/XML-XPath-1.36/lib/XML/XPath.pm> (Aufruf 21.04.16)

¹³ XML::LibXML <http://search.cpan.org/dist/XML-LibXML/LibXML.pod> (Aufruf 21.04.16)

¹⁴ Die CiXbase Indizes liegen in einer flat-file-Struktur mit sog. B*-Bäumen vor.

Die aufgebauten Indizes liegen – wie die jeweiligen Indizes anderer Datentypen – auf einem verteilten Cluster

aus 8 Servern – insgesamt etwa 60 Mio. Metadaten. Bei Suchanfragen werden alle verschiedenen Indizes parallel einzeln ausgewertet, anschließend gemischt, nach Relevanz sortiert und als ein aggregiertes Resultat über einen Frontend-Server an die Nutzer*innen ausgeliefert.

Fazit

Die Integration online frei verfügbarer wissenschaftlicher Texte besitzt großes Potential, das eigene Angebot um hochaktuelle, fremdsprachige aber auch historische Medien vielfältig zu erweitern und bisher wenig genutzte Quellen internationaler Dokumentenarchive in gleichzeitiger Sicht auf lokale Medien vor Ort verfügbar zu machen. Open Access Publikationen anderer Standorte werden damit an den Universitäten sichtbar und freie wissenschaftliche Publikationen genauso behandelt wie Zeitschriftenartikel aus klassischen Verlagsproduktionen.

Von besonderer strategischer Bedeutung ist zukünftig die zu verbessernde Nachweisqualität der Repositorienanbieter, um die Nachweise wirklich freier Dokumente von klassischen Verlagsmetadaten zu differenzieren.

Weitere Zukunftsprobleme stellen die meist noch fehlenden Verknüpfungen von Verlagsveröffentlichungen mit den potentiell vorhandenen parallelen Zweitveröffentlichungen auf freien Dokumentenservern und die fehlenden Verfügbarkeitsinformationen einzelner frei verfügbarer Open Accesspublikationen¹⁵ in ansonsten lizenzpflichtigen (Hybrid-)Zeitschriften dar.

¹⁵ Z.B. *OpenChoice* Publikationen beim Springer Verlag

DeepGreen - Entwicklung eines rechtssicheren Workflows zur effizienten Umsetzung der Open-Access-Komponente in den Allianz-Lizenzen für die Wissenschaft

Markus Putnings ^{1,3}, Thomas Dierkes ^{2,3}

¹Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Universitätsbibliothek

²Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV), Zuse Institute Berlin (ZIB)

³im Auftrag des DeepGreen-Projektkonsortiums:

Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV) – Projektleitung • Beate Rusch • Julia Alexandra Goltz • Thomas Dierkes • Jens Schwidder Technische Universität Berlin (TUB), Universitätsbibliothek • Jürgen Christof • Monika Kuberek • Dagmar Schobert • Pascal Becker	Bibliotheksverbund Bayern (BVB) • Klaus Ceynowa • Matthias Gross Bayerische Staatsbibliothek München (BSB) • Hildegard Schöffler • Lilian Landes • Michael Kassube Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Universitätsbibliothek • Konstanze Söllner • Markus Putnings • Oliver Schwab • Regina Heidrich	Helmholtz Open Science Koordinationsbüro am Deutschen GeoForschungs-Zentrum (GFZ) • Roland Bertelmann • Heinz Pampel • Paul Vierkant • Tobias Höhnnow
---	--	---

Zusammenfassung

Die Grundsätze für den Erwerb DFG-geförderter überregionaler Lizenzen (sog. Allianz-Lizenzen) beinhalten spezifische Regelungen zum Open Access. Die Erfahrung der seit 2011 getätigten Allianz-Abschlüsse zeigt allerdings, dass der Kreis berechtigter Autorinnen und Autoren eigenständig kaum Gebrauch ihrer hierdurch erhaltenen Open-Access-Rechte macht. Entsprechend liegt ein großer Schatz wissenschaftlicher Literatur bei den Verlagen, der noch zu heben ist. Das bewilligte DFG-Projekt DeepGreen (Ausschreibung „Open-Access-Transformation“ von 2014) zielt darauf ab, die vereinbarten Open-Access-Konditionen der Allianzlizenzen auf technischer Ebene komfortabel auszugestalten und wenn möglich zu automatisieren, so dass nicht mehr Autorinnen und Autoren oder die hierzu berechtigten Bibliotheken die Publikationen manuell in Open-Access-Repositorien einpflegen müssen, sondern

die Verlage selbst zyklisch über definierte Schnittstellen abliefern. Dazu bauen die Projektpartner (Universitätsbibliothek Erlangen-Nürnberg, Universitätsbibliothek TU Berlin, Helmholtz Open Science Koordinationsbüro am Deutschen GeoForschungsZentrum, Bayerische Staatsbibliothek München sowie die Verbünde Bibliotheksverbund Bayern und Kooperativer Bibliotheksverbund Berlin-Brandenburg) ein Dark Archive namens DeepGreen auf, in das teilnehmende Allianzlizenz-Verlage ihre Publikationen und Metadaten einspeisen. DeepGreen soll im Anschluss wiederum als Datendrehscheibe für berechnigte Open-Access-Repositorien dienen. Als Pilotpartner konnten die Verlage Karger und SAGE gewonnen werden. Die Universitätsbibliothek Erlangen-Nürnberg als Initiatorin des Projekts und der Kooperative Bibliotheksverbund Berlin-Brandenburg (KOBV) als Hauptentwickler von DeepGreen stellten das Projekt und den Stand der Umsetzung am 6. Bibliothekskongress 2016 in Leipzig vor.

Abstract

The acquisition standards of national licences supported by Deutsche Forschungsgemeinschaft (DFG) (referred to as alliance licences) include specific open access policies. Experience, however, has shown that entitled authors make almost no use of their right to publish openly by these alliance licences contracted since 2011. Thus, a tremendous amount of scientific literature is yet to uncover from publisher to be completely transformed into open access. The project DeepGreen, approved by DFG (based on the initiative „Open Access Transformation“, 2014), seeks to establish on a technical level that the existing open access agreements are put into practice, both conveniently and, most important, automatically as far as possible. In particular, publishers are required to deliver periodically all publications eligible for open access through defined interfaces, instead of authors (or their respective libraries) having to upload these items manually into corresponding open access repositories. To this end, the members of the project consortium (University Libraries Erlangen-Nuremberg and TU Berlin, Helmholtz Open Science Office at the German GeoResearch Centre, Bavarian State Library, and two Librarian Network Organizations, BVB and KOBV) build the platform DeepGreen, essentially a dark archive, into which publications and meta data are fed constantly by all publishers contracted alliance licences. In turn, the platform DeepGreen delivers then automatically these publications to the legitimate repositories. Two publishers, Karger

Publishers and SAGE Publications, have agreed to initially support the project as associated partners. The University Library Erlangen-Nuremberg (as initiator of the project) and the Cooperative Librarian Network Berlin-Brandenburg (KOBV), as responsible body of the project development) have presented the current state-of-the-art at the 6th Library Congress in Leipzig, 2016.

1. Ausgangslage

Über Allianz-Lizenzen werden im Rahmen von nationalen Opt-in-Konsortien dynamische Produkte, wie zum Beispiel Zeitschriftenpakete und Fachdatenbanken für den wissenschaftlichen Grundbedarf, häufig mit interdisziplinärem Charakter lizenziert und zum überwiegenden Teil von der Deutschen Forschungsgemeinschaft (DFG) finanziell bezuschusst.¹ Durch diese DFG-Förderung gelten für den Erwerb von Allianz-Lizenzen obligatorische Standards und Klauseln, die „Grundsätze für den Erwerb DFG-geförderter überregionaler Lizenzen (sog. Allianzlizenzen)“. Diese beinhalten auch eine Klausel zu Open Access:

„Regelungen zum Open Access

16. Autoren aus autorisierten Einrichtungen sind ohne Mehrkosten berechtigt, ihre in den lizenzierten Zeitschriften erschienenen Artikel in der Regel in der durch den Verlag publizierten Form (z.B. PDF) zeitnah in institutionelle oder disziplin-spezifische Repositorien ihrer Wahl einzupflegen und im Open Access zugänglich zu machen. Das gleiche Recht besitzen die autorisierten Einrichtungen, denen die jeweiligen Autoren angehören.

Der Anbieter erklärt sich bereit, autorisierte Einrichtungen bei der Identifizierung und Lieferung relevanter Artikeldaten und Volltexte, einschließlich der digitalen Objekte, ohne Mehrkosten beratend und technisch zu unterstützen. Das kann z.B. darin bestehen, die Volltexte inkl. der Metadaten in einem gängigen Format bzw. nach solchen Standards bereit zu stellen, die ein Einspielen in Repositorien erleichtern. Begrüßenswert wäre zudem, dass der Anbieter es selbst übernimmt, Artikel von Autoren aus autorisierten Einrichtungen in ein vereinbartes Repository einzupflegen, z.B. über eine SWORD-Schnittstelle.“ (DFG, o.J.)²

Diese Klausel erweitert also die Open Access-Rechte im sogenannten Grünen Weg von Open Access und erleichtert die Open Access-Veröffentlichung und Langzeitarchivierung der entsprechenden Artikel auf institutionellen oder disziplinären

¹ <https://www.nationallizenzen.de/ueber-nationallizenzen/allianz-lizenzen-2011-ff>.

² http://www.dfg.de/formulare/12_181/12_181_de.pdf

Open-Access-Repositorien: Üblicherweise ist im Grünen Weg bei vielen Verlagen die nachträgliche Open Access-Veröffentlichung auf Repositorien erst nach 12 bis 24 Monaten erlaubt, und auch nur in der Postprint-Version, das heißt der akzeptierten Autorenfassung, die vom finalen Layout der Verlagsversion abweichen kann. Durch die oben genannte Klausel darf dagegen in der Regel die Verlagsversion genutzt werden und in der überwiegenden Mehrzahl der Abschlüsse ganz ohne oder mit einer verkürzten Embargofrist von sechs Monaten. Eine Übersicht über die verhandelten Open Access-Rechte liefert die Seite <https://www.nationallizenzen.de/open-access/open-access-rechte.xls/view>.

2. Das Problem der Nutzung

Eine Umfrage der Bayerischen Staatsbibliothek hat 2013 jedoch ergeben, dass lediglich 16 Einrichtungen die oben beschriebene Open Access Komponente aus den Allianz-Lizenzen aktiv nutzten. Die Initiative zur Publikation ging dabei fast ausschließlich von den beteiligten Bibliotheken aus, die wissenschaftlichen Autorinnen und Autoren demonstrierten mangelndes Interesse. Das gemeldete Zahlenmaterial hat ergeben, dass von 2011 bis 2013 damit nur rund 600 Beiträge archiviert wurden.

Diese Zahl ist angesichts mehrerer Tatsachen enttäuschend: Aktuell sind in der Zeitschriftendatenbank 294 Produkte mit dem Sigel für National- oder Allianz-Lizenzen versehen. Hinter jedem Produkt finden sich viele verschiedene Zeitschriften, E-Books, et cetera. Multipliziert mit den jeweiligen Artikeln pro Zeitschrift und lizenzierten Jahren, die unter die benannten Open Access-Regelungen fallen, verbirgt sich hier ein Potential von tausenden Open Access-Artikeln, die aktuell noch hinter einer Paywall versteckt sind.

Außerdem resultieren viele der bei der Umfrage gemeldeten Beiträge allein von einer Einrichtung, der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). Hier werden seit 2011 die lizenzierten Inhalte von sieben Allianzlizenz-Verlagen auf Autoren der FAU hin ausgewertet und die gefundenen Artikel in einer speziellen Collection von OPUS FAU, dem institutionellen Repositorium eingestellt.³

Die Schattenseite davon ist, dass der entsprechende Aufwand vor Ort nicht unerheblich ist. Die FAU besitzt keine Hochschulbibliografie. Es müssen also alle

³ <https://opus4.kobv.de/opus4-fau/solrsearch/index/search/searchtype/collection/id/16213>

fraglichen Veröffentlichungen der Allianzlizenz-Verlage explizit auf Autorinnen und Autoren der FAU hin durchsucht werden. Dies geschieht über eine Abfrage der Universitätszugehörigkeit, sofern dieses Datenfeld "Affiliation" auf Verlagsseiten oder in Fachdatenbanken vorhanden und praktikabel nutzbar ist. Ein Problem hierbei ist, dass die entsprechenden Angaben nicht normalisiert sind, sondern schlicht den Freitexteintragungen der Autorinnen und Autoren beim Submission-Prozess entsprechen. Damit sind diverse Institute, Adressangaben und Eingaben in verschiedensten Sprachen und Abkürzungen enthalten; allein der offizielle Universitätsname "Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)" bietet an die 100 verschiedene Schreibweisen. Mit Trunkierungen wie Univ* Erlangen* lassen sich nicht alle potentiell zugehörigen Artikel ermitteln, da die Autorinnen und Autoren teilweise nur das Institut oder Department nennen, das Kürzel FAU oder aber auch gar keine Affiliation angeben. So kommt es jeweils zu einem gewissen „Schwund“ bei der Einbringung. Auch sind die Abfragen, die Extraktion der Volltexte und die Einbringung in OPUS FAU mit allen zugehörigen Metadaten so aufwändig, dass diese Aktionen nicht laufend durchgeführt werden können, sondern nur in konzertierten zusätzlichen kleinen Projekten neben dem Tagesgeschäft. Entsprechend gibt es lange Zeitverzögerungen bei der Einbringung, selbst wenn die Open-Access-Komponente der entsprechenden Allianz-Lizenz eine sofortige Einbringung ohne Embargofrist erlaubt.

3. Die Entstehung der Projektidee

An der FAU bestand also schon lange der Wunsch nach einem Workflow, um Artikel verlagsseitig automatisch in OPUS FAU einzuspielen, sofern z.B. durch Allianz-Lizenzen eine Berechtigung vorliegt. Triebfeder für einen konkreten Aktionsplan war dann die DFG-Ausschreibung "Open Access Transformation" im LIS-Programm vom Juni 2014. Erbeten wurden hier "innovative Lösungen [...], die sich auf unterschiedliche Bereiche des Publikationsprozesses beziehen, zum Beispiel

- auf technische und organisatorische Aspekte des Publikationsprozesses;
- auf die dem Open Access zugrunde liegenden Finanzierungs- und Geschäftsmodelle;
- auf die Überführung traditionell subskriptionsbasierter Zeitschriften in den Open Access;

- auf das Absichern der möglichst umfassenden
Nachnutzbarkeit der Publikationen [...]" (DFG, 2014)



Diese Potenziale waren schnell gefunden. Zum einen soll es kein hausspezifischer technischer Workflow werden, sondern nutzbar sein für alle berechtigten Allianzlizenzteilnehmer und deren institutionelle Repositorien.

Damit könnte der "Schatz" an tausenden Artikeln deutscher Autorinnen und Autoren, die sich hinter den 294 Produkten der National- oder Allianz-Lizenzen verbergen, gehoben und weltweit frei nachnutzbar gemacht werden. Gemäß den anfangs zitierten Regelungen zum Open Access sind hierzu die autorisierten Einrichtungen ohne Mehrkosten berechtigt. Diese Regelung kann zukünftig auch noch forciert werden, sofern die DFG das Projekt auch politisch unterstützt, und auch als Grundlage für ähnliche Lizenzverhandlungen im Bibliothekswesen dienen (zum Beispiel bei Konsortial- und Paketverhandlungen). Unter der Bedingung der verkürzten oder

idealerweise ausgesetzten Embargofrist und der erlaubten Nutzung der Verlagsversion könnte so zumindest national eine Überführung subskriptionsbasierter Zeitschriften in den Open Access durch die entsprechende "Vermischung" des Grünen und Goldenen Wegs von Open Access erzielt werden. Sofern die Verlage kooperieren und der technische Workflow die Kompatibilität der Schnittstelle(n) von Verlagen und Repositorien sicherstellt, ließe sich die Idee der Direktablieferung in Repositorien auch generell auf den Grünen Weg übertragen, indem der Autor bspw. gleich beim Autorenvertrag mit angibt, die Publikation nach Embargofrist abzuliefern oder indem die "Self-archiving"-Policy des Verlags dies standardmäßig vorsieht.

Die finalen Ziele der Projektidee waren also:

- Die Förderung von grünem Open Access.
- Eine automatisierte Ablieferung von Publikationen durch Verlage, unter Nutzung der lizenzrechtlich verhandelten Vorteile (zum Beispiel die zeitnahe oder sofortige Lieferung der Verlagsversion).
- Planung und Umsetzung einer technischen Datendrehscheibe, das heißt eines Intermediär, der idealerweise die Datenqualität sicherstellt, beispielsweise mit Blick auf die Zuordenbarkeit der Affiliation und die Kompatibilität der genutzten Vokabulare und Schnittstellen; denn Verlage nutzen im Gegensatz zu Bibliotheken in der Regel das Metadatenformat ONIX, NISO JATS oder CrossRef-XML, wohingegen auf Repositorien zum Beispiel Dublin Core oder Dublin Core-kompatible Importformate wie das DSpace Simple Archive Format verwendet werden.
- Die Schaffung eines möglichst offenen und anpassbaren Systems als Datendrehscheibe, um die verschiedenen technischen „Endsysteme“, also zum Beispiel EPrints, DSpace, OPUS et cetera bestücken zu können oder um passende Importer Scripts für die institutionellen Repositorien gestalten zu können.
- Die Ausgestaltung von Automatismen, die weitgehend unabhängig von individuellen Entscheidungsschritten und hierfür nötigen Beratungsbedürfnissen auf Autorenmseite ablaufen. In einer ersten Projektphase basieren diese vollständig auf den bereits verhandelten Open Access-Regelungen der Allianz-Lizenzen, welche auch die autorisierten Einrichtungen in Anspruch nehmen dürfen.
- Eine perspektivische, möglichst natürliche Erweiterbarkeit des implementierten Systems, etwa um weitere Lizenzverhandlungen oder Policy-Formulierungen oder ganz

generell den Grünen Weg von Open Access leicht umsetzen zu können.

4. Das DFG-Projekt DeepGreen

Die ursprüngliche kleine Idee des technischen Workflows, wissenschaftliche Artikel verlagsseitig automatisch in OPUS FAU einzuspielen, nahm konzeptionell wie oben beschrieben, eine größere Dimension ein, als dass sie alleine umsetzbar gewesen wäre. Zunächst wurde deshalb der Kooperative Bibliotheksverbund Berlin-Brandenburg (KOBV) ins Boot geholt, der für das Hosting von OPUS FAU verantwortlich ist. Nach und nach wurden weitere Partner hinzugezogen, um alle nötigen Kompetenzfelder abzudecken.

Das DeepGreen-Projektkonsortium besteht aus:

- Erfahrenen Vertragsverhandlern wie der Bayerischen Staatsbibliothek, der hier auch noch die Rolle eines Betreibers von Fachrepositorien wie z.B. OstDok - Osteuropa-Dokumente Online zukommt.
- Open-Access-Experten, hier insbesondere das GeoForschungsZentrum Potsdam, an dem das Open Science Büro der Helmholtz Gemeinschaft angesiedelt ist.
- Zwei Universitäten mit Anspruch und unterschiedlicher Repositorieninfrastruktur (TU-Berlin mit dSpace) und dem Ideengeber, der Friedrich-Alexander-Universität Erlangen-Nürnberg (OPUS)
- Zwei Verbünden, dem Bibliotheksverbund Bayern und dem Kooperativen Bibliotheksverbund Berlin-Brandenburg, die im Aufbau und Betrieb von bibliothekarischen Dienstleistungen geübt sind und in Fragen von Datenformaten und Schnittstellen gewandt sind.
- Der KOBV übernimmt in diesem Projekt die Koordination.

Zwei von 12 Verlagshäuser mit Allianzlizenzen konnten für die Mitarbeit gewonnen werden, das international aufgestellte Verlagshaus SAGE Publications, das nach eigenen Angaben über 950 Zeitschriften vor allem aus den Sozial-, Geistes- und Lebenswissenschaften verlegt und das Schweizer Verlagshaus Karger, welches sich mit seinen 105 Zeitschriften vor allem in der Disziplin Biomedizin bewegt.

Nachdem das Projektkonsortium Mitte 2015 die Förderzusage erreichte, wurde der offizielle Projektstart auf den 01.01.2016 festgelegt. Bei der DFG-Förderung handelt es sich um eine flexible Förderung, was bedeutet, dass Bedingungen gestellt wurden:

Ein Teil der bewilligten Mittel ist solange gesperrt, bis ein erster Prototyp der technischen Datendrehscheibe präsentiert werden kann. Dieser ist für den Herbst 2016 geplant.

Die Datendrehscheibe ist eine vermittelnde Instanz zwischen den Verlagen und den Repositorien, die einen Vertragsinhalt in die Realität umsetzt und die Geschäftsgänge in den Bibliotheken erleichtert. Die technische DeepGreen-Drehscheibe lässt sich als Schaltstelle, Poststelle, als ein Zwischenarchiv, ein öffentlich nicht zugängliches Repositorium und damit als ein unsichtbares, Dark Archive vorstellen.

Für Deutschland neu und damit innovativ ist der Ansatz, dass Verlagsdaten automatisiert an Repositorien geschickt werden. Auf der internationalen Ebene gibt es jedoch renommierte Fachrepositorien, z.B. Europe PubMed Central, die Lieferverfahren mit Verlagen seit Jahren erfolgreich in der Praxis einsetzen. Die Erfahrung dieser internationalen Vorbilder lehrt, dass eine erfolgreiche Zusammenarbeit mit Verlagen nur funktionieren kann, wenn keine neuen technischen Hürden aufgebaut werden und es verschiedene auch niedrigschwellige Möglichkeiten zur Ablieferung angeboten werden.

5. Die Ansätze zur technischen Umsetzung

Die Erwartungen an die technische DeepGreen werden formal in sogenannten User Stories aus Sicht der unterschiedlichen Beteiligten zusammengefasst: Aus der Perspektive von Verlagen, Bibliotheken in ihrer Rolle als Repositorienbetreiber, aber auch aus Sicht des Plattformbetreibers DeepGreen selber und nicht zuletzt mit Blick des Förderers, die z.B. Kennzahlen messen möchten, werden verschiedene Szenarien beschrieben, die ein Prototyp nach Möglichkeit bedienen können soll.

Die Szenarien werden nach dem agilen Prinzip MoSCoW (englisch: *Must have*, *Should have*, *Could have*, und *Would like but won't get*) von den Projektbeteiligten priorisiert. Eine Veröffentlichung auf der DeepGreen-Webseite ist geplant.

Parallel zu den Use-Cases wurde ein DeepGreen-Metadatenschema erarbeitet, das zur fachöffentlichen Kommentierung publiziert und im Rahmen eines Expertenworkshops im April vorgestellt wurde.^{4 5} Das Metadatenschema ist auf der Grundlage der Best-Practice-Empfehlungen von OCLC zur verteilten Nutzbarkeit

⁴ <https://deepgreen.kobv.de/veroeffentlichung-metadatenschema/>

⁵ <https://deepgreen.kobv.de/wp-content/uploads/2016/03/Metadata-Schema-DeepGreen.pdf>

zwischen verschiedenen Repositorien entwickelt worden.⁶ Einerseits sieht das Schema vielfältige Möglichkeiten vor, etwaige Verlags-Metadaten abzubilden. Gleichzeitig bietet es genügend Flexibilität, um den Bedürfnissen unterschiedlichster Repositorien entgegen zu kommen. Darüber hinaus werden sogar Metadaten-Felder vorgesehen, die eine weitere Verarbeitung, z.B. bei der Ablieferung an die Deutsche Nationalbibliothek erleichtern würden. Schließlich wird es in der Praxis darauf ankommen, was die Verlage tatsächlich liefern können. Eben dazu läuft im Moment eine Umfrage des Projektes bei Allianz-Lizenzverlagen.

Auf der Suche nach technischen Vorbildern zur Nachnutzung, beschäftigt sich das Projekt in der Zwischenzeit intensiv mit dem in Großbritannien schon vor Jahren aufgelegten „DeepGreen“-Projekt *Jisc Publications Router*.⁷ In diesem britischen Langzeitprojekt wurde ein digitales Verteilzentrum aufgebaut, das ebenso wie DeepGreen zwischen Verlagen und berechtigten Repositorien vermittelt. Es wäre geradezu töricht, wollte man die dabei geschaffene Infrastruktur und Frameworks ignorieren, um das Rad ein x-tes Mal neu zu erfinden.

Sicherlich sind die deutschen Rahmenbedingungen, die durch die existierenden Allianz-Lizenzen geschaffen wurden, sehr speziell. Gerade darin besteht jedoch eine hervorragende Möglichkeit, gemeinsam ein international gültiges Verteilsystem aufzubauen; zumal die Tätigkeit der Verlage sicherlich nicht an nationalen Grenzen Halt macht, sondern ebenso international ausgerichtet ist.

Der *Jisc Publications Router*, wie im Blog-Beitrag von Steve Byford erläutert, ist eine grundsätzliche Neuentwicklung basierend auf den Erfahrungen und Lehren der vorherigen Versionen dieses Projekts.⁸ Unter der Adresse <https://github.com/JiscPER> sind die vorläufigen Ergebnisse dieses erneuten, in der Skriptsprache Python implementierten Ansatzes zu finden. In insgesamt 10 Untermodulen werden sowohl die Kernfunktionalitäten des Verteilzentrums implementiert, als auch die gängigsten Schnittstellen wie SWORD zur Ein- und Ausgabe, OAI-PMH als Harvesting-Mechanismus, oder ein FTP-Client zur Entgegennahme von Verlagsablieferungen. Die interne Speicherung zur Abwicklung der Verteilfunktion geschieht dabei mittels einer eigenen, JSON-gestützten Datenbank, die durch die bekannte und etablierte Volltext-Suchmaschine „Elasticsearch“ indiziert wird. Zudem wird eine FLASK-basierte

⁶ <http://www.oclc.org/content/dam/support/wcdigitalcollectiongateway/MetadataBestPractices.pdf>

⁷ <https://pubrouter.jisc.ac.uk/>

⁸ <https://scholarlycommunications.jiscinvolve.org/wp/2015/07/01/jisc-publications-router-enters-a-new-phase/>

Software-Bibliothek zur Verfügung gestellt, um Webapplikationen mit der Suchmaschine „Elasticsearch“ als Endpunkt (engl. *Backend*) im Kontext des Publication Routers einfach aufzusetzen.

Das Projekt DeepGreen kann als ein Experiment verstanden werden mit dem Potential der Verteilung und Verbreitung nicht nur von Allianzlizenzdaten. Den Allianz-Lizenzen kommt an dieser Stelle vor allem Mustercharakter zu.

Literaturverzeichnis

Alle Verweise wurden am 26. April 2016 geprüft.

DeepGreen. "Veröffentlichung MetadatenSchema."
<https://deepgreen.kobv.de/veroeffentlichung-metadaten-schema/>

DFG. "Allianz-Lizenzen --- Nationallizenzen." <https://www.nationallizenzen.de/ueber-nationallizenzen/allianz-lizenzen-2011-ff.>

DFG. "Grundsätze für den Erwerb DFG-geförderter überregionaler Lizenzen (Allianz-Lizenzen)." *DFG-Vordruck 12.181 – 03/15*.
http://www.dfg.de/formulare/12_181/12_181_de.pdf

DFG. "Open-Access-Transformation." *Information für die Wissenschaft Nr. 29*, 2. Juni 2014.
http://www.dfg.de/foerderung/info_wissenschaft/2014/info_wissenschaft_14_29/index.html

Jisc. "Jisc publications router." <https://pubrouter.jisc.ac.uk/>

Jisc scholarly communications (Blog). <https://scholarlycommunications.jiscinvolve.org/>

OCLC. "Best Practices for CONTENTdm and other OAI-PMH compliant repositories: creating sharable metadata." *Version 3.1*, 20. Juni 2013.
<http://www.oclc.org/content/dam/support/wcdigitalcollectiongateway/MetadataBestPractices.pdf>

OPUS4-FAU. "OPUS 4 | Allianzlizenzen." <https://opus4.kobv.de/opus4-fau/solrsearch/index/search/searchtype/collection/id/16213>

Forschungsdaten

Serviceleistungen der Bibliothek des Deutschen Krebsforschungszentrums im Bereich Forschungsdatenmanagement

Dagmar Sitek

Deutsches Krebsforschungszentrum (DKFZ), Heidelberg

Abstract

Das Thema Forschungsdatenmanagement gewinnt zunehmend an Bedeutung. Die Entwicklung und Implementierung von disziplinspezifischen Lösungen für die Wissenschaft und Forschung sind essentiell, um die großen Herausforderungen, die sich in diesem Bereich ergeben, zu bewältigen. Innerhalb dieses Entwicklungsprozesses können auch Bibliotheken neue Aufgaben übernehmen, um sich als leistungsstarker Partner für die Wissenschaft zu positionieren. Die Bibliothek des Deutschen Krebsforschungszentrums hat in Kooperation mit den Wissenschaftlern „maßgeschneiderte“ Lösungen in diesem Kontext entwickelt, um adäquate Unterstützung anbieten zu können.

Forschungsdatenmanagement – ein Thema von zunehmender Relevanz

Der Umgang mit Forschungsdaten hat in den letzten Jahren zunehmend an Bedeutung gewonnen. Die Anforderungen, die von den verschiedenen Stakeholdern an die Forschenden in diesem Kontext gestellt werden, steigen zunehmend und werden immer komplexer. So fordern beispielsweise wissenschaftliche Zeitschriften, dass die Daten, die den eingereichten Publikationen zugrunde liegen, für den Review-Prozess zur Verfügung gestellt und mit dem Artikel veröffentlicht werden müssen. Dies erfolgt in der Regel in Form von „Supplementary Material“, welches in die Artikel integriert wird und damit die publizierten Ergebnisse überprüfbar und nachvollziehbar macht. Ebenso steigt bei den Zuwendungsgebern zunehmend das Bewusstsein, dass Daten, ebenso wie Publikationen, die aus öffentlichen Mitteln finanziert wurden, auch der Öffentlichkeit zur Verfügung stehen sollten, weil die dadurch ermöglichte Nachnutzung der Daten positive Auswirkungen auf die Forschungsprozesse hat. So hat unter anderem die Europäische Union in ihrem Rahmenprogramm für Forschung und

Innovation, Horizon 2020¹, einen „Open Research Data Pilot“ integriert, welcher beinhaltet, dass für bestimmte Projekte, die aus ausgewählten Programmen finanziert werden, Forschungsdatenmanagementpläne vorgelegt und die Daten Open Access zur Verfügung gestellt werden müssen.²

Auf der einen Seite gibt es also steigende Anforderungen an die Forschenden, ihre Daten effektiv zu managen und zur Verfügung zu stellen, auf der anderen Seite mangelt es aber zur Zeit noch häufig an geeigneten Systemen, Workflows und Mechanismen, um diese Aufgaben ohne allzu großen zusätzlichen Aufwand realisieren zu können.

In diesem Spannungsfeld bietet sich für Bibliotheken die Möglichkeit, neue Aufgaben zu erschließen und sich so als leistungsstarker Partner für die Wissenschaft zu positionieren.

Forschungsdaten im Deutschen Krebsforschungszentrum

Das Deutsche Krebsforschungszentrum (DKFZ)³ ist die größte biomedizinische Forschungseinrichtung in Deutschland und ist Mitglied in der Helmholtz-Gemeinschaft Deutscher Forschungszentren⁴. Seine Aufgabe besteht in der Erforschung von Krebsentstehung, Krebsrisikofaktoren und Krebsprävention sowie in der Entwicklung von neuen Behandlungsansätzen und Therapien.

Wie in den meisten Forschungseinrichtungen fallen auch im DKFZ sehr große Mengen von unterschiedlich strukturierten Forschungsdaten an. Die großen, sehr datenintensiven Projekte, wie z.B. bestimmte klinische Studien oder Genomsequenzierungsprojekte, verfügen in der Regel über Forschungsdatenmanagementpläne sowie geeignete Infrastrukturen und sind damit gut erschlossen. Ganz anders stellt sich die Situation bei den unzähligen einzelnen Datenelementen, wie z.B. den Excel-Tabellen, Abbildungen, Elektronenmikroskopieaufnahmen usw., die während eines Forschungsprozesses anfallen, dar. Diese vielen kleinen Datenmengen, die zusammen genommen durchaus den Umfang von Big-Data-Projekten erreichen, liegen häufig verteilt auf den Rechnern und Laptops der einzelnen

¹ <http://www.horizont2020.de/>

² <http://www.horizont2020.de/einstieg-open-access.htm>

³ <http://www.dkfz.de/de/index.html>

⁴ <http://www.helmholtz.de/>

Forschenden. Diese Situation in Verbindung mit der hohen Fluktuation, die in Forschungsinstitutionen herrscht, führt zu zahlreichen Problemen. Im günstigsten Fall erfolgt eine Übergabe dieser Forschungsdaten an den verantwortlichen Abteilungsleiter⁵, der sie speichern und entsprechend erschließen muss, um das Wiederauffinden zu gewährleisten. Es besteht aber die Gefahr, dass dies nicht passiert und auf diese Weise Daten verloren gehen, wenn zum Beispiel Forschende das DKFZ verlassen oder Abteilungen aufgelöst werden. Dies ist nicht nur problematisch, weil die Daten auf diese Weise für eine Nachnutzung verloren gehen, sondern weil so auch die Regeln der „Guten wissenschaftlichen Praxis“ nicht eingehalten werden. In ihrer Denkschrift „Sicherung guter wissenschaftlicher Praxis“, fordert die Deutsche Forschungsgemeinschaft aus Gründen der Nachprüfbarkeit: „Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, zehn Jahre lang aufbewahrt werden“.⁶ Somit ist ein effektives Forschungsdatenmanagement der vielen kleinen Datenelementen nicht nur sinnvoll sondern auch zwingend erforderlich.

Anders, aber ebenfalls problematisch, stellt sich die Situation bei den Forschungsdaten dar, die in speziell entwickelten Datenbanken gespeichert sind. Es gibt im DKFZ zahlreiche Inhouse-Datenbanken, die von einzelnen Abteilungen im Haus für ganz spezifische Forschungsdaten entwickelt wurden. Als Beispiel ist hier eine Datenbank mit Ultraschallbildern oder eine weitere Datenbank mit Röntgenbildern zu nennen. Dort liegen die Daten zwar in der Regel gut archiviert und erschlossen, jedoch unverbunden vor. Im Bedarfsfall müssen sie gegebenenfalls mühsam recherchiert und zusammengestellt werden.

Ein sinnvoller Ansatz in diesem Kontext wäre es, alle Forschungsdaten, die in einem Projekt anfallen, an einer Stelle zusammenzuführen. Da das Ergebnis eines Forschungsprozesses, zumindest momentan noch, als Zeitschriftenartikel veröffentlicht wird, liegt es nahe, diese Verknüpfung in der Publikationsdatenbank zu realisieren.

⁵ Aufgrund der einfacheren Lesbarkeit wird im Folgenden nur die männliche Form verwendet, die weibliche Form ist dabei jeweils mit eingeschlossen.

⁶ http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf S 21

Die DKFZ-Publikationsdatenbank enthält alle Veröffentlichungen der Mitarbeiter des DKFZ. Da die Bibliothek für die Entwicklung, Organisation und Pflege der Datenbank verantwortlich ist, konnte in diesem Bereich ein Serviceangebot implementiert werden, welches den Forschenden Unterstützung im Forschungsdatenmanagement anbietet. Ziel ist es, alle Forschungsdaten, die zu einer Publikation geführt haben, in der Publikationsdatenbank bei der entsprechenden Publikation zusammenzuführen und nachzuweisen. Dies wird, je nach Art der Forschungsdaten, unterschiedlich realisiert und soll zukünftig noch stark ausgebaut werden.

Verbindung Mausdatenbank - Publikationsdatenbank

Die Abteilung Kryokonservierung des DKFZ hat eine Datenbank für transgene Mauslinien entwickelt. In ihr werden Mausmodelle erfasst und detailliert beschrieben. Unter anderem werden dort auch Artikel, in denen das jeweilige Mausmodell publiziert wurde, eingetragen.⁷ Viele dieser Artikel wurden von Forschenden des DKFZ publiziert, aus diesem Grund wurde bei dem Aufbau der Mausdatenbank unter anderem in der Publikationsdatenbank recherchiert. Da beide Datenbanken Inhouse-Lösungen waren und sich beide noch in der Entwicklungsphase befanden, bot sich eine sehr gute Gelegenheit, beide miteinander zu verbinden. Es wurde eine Schnittstelle programmiert, die dazu führt, dass der Eintrag einer Publikation in der Mausdatenbank automatisch eine Meldung an die Bibliothek generiert. Handelt es sich um einen Artikel, der in der Publikationsdatenbank enthalten ist, so wird von der Bibliothek eine Verknüpfung durch einen Link zwischen den beiden Einträgen in beiden Datenbanken geschaffen. In der Mausdatenbank wird dadurch der verlinkte Hinweis zu der Publikation in der Publikationsdatenbank angezeigt. (Abb. 1)

⁷ Die Datenbank steht nur DKFZ intern zur Verfügung. Nähere Informationen dazu sind unter anderem in folgendem Artikel zu finden: Staudt M, Trauth J, El Hindi I, Galuschka C, Sitek D, Schenkel J. Managing major data of genetically modified mice – from scientific demands to legal obligations. Transgenic Res 2012; 21 (5): 959-966.

Intranet > Mausmutanten > Übersicht > Detailansicht

ein Leben ohne Krebs

EinleitungÜbersichtKontakt

Beschreibung der Mauslinie

Mauslinie#

ID: 483

Eingabedatum:

24.03.2009

Änderungsdatum:

26.11.2013

Short name: *

HPV11 E2 x URR11

Tierbase Nr.:

2390

Stock-Nr.:

Beschreibung der Mutante:

Mutationsart	mutiertes Gen / Transgen	Tg Spender	reg. Element	reg. Elem. Spender	verwendeter Vektor
Transgener Überexprimierer	HPV11 E2	HPV 11	Ubiquitin C promoter	Homo sapiens	pUC 18
Transgener Überexprimierer	LacZ	E. coli	HPV11 URR	HPV 11	pBluescript KS

Risikogruppe:
(GenTG) *

1

Gene ID# (MGI, OMIM):

Vater (Name der Mutante), Genotyp:

HPV11 E2 x URR11

Mutter (Name der Mutante), Genotyp:

HPV11 E2 x URR11 or WT

Genetischer Hintergrund:

B6D2Fn

Rückkreuzungsgeneration:

unknown/mixed

Full name:

B6;D2-Tg(HPV11-lacZ)1704Aal Tg(UBC-HPV11E2)613Josc

nützliche Links:

MGI Jax

ILAR Labcode

OMIM

Synonyme:

Fragestellung hinter der Entwicklung/
Kurzbeschreibung der Mutante: *

Mouse model for HPV-11 regulation

Belastungseinstufung:
(TierSchG)

keine

Beschreibung der Belastung inkl.
der belasteten Genotypen: *

Reportermaus, keine Belastung beobachtet

Welche Besonderheiten weisen
die Mäuse auf:

Homozygot Letal:

Nein

Besondere Körpermerkmale:

Besondere Verhaltensmerkmale:

Besondere Reproduktionsmerkmale:

Keywords:

Promoter, skin, transcription, hair follicle, animal model, E2 protein, Papillomavirus, transgenic mice Ubiquitin C promoter

Publikation:

Schenkel J, Gaissert H, Protopapa EE, Weiher H, Gissmann L, Alonso A (1999): The human Papillomavirus type 11 upstream regulatory region triggers hair-follicle-specific gene expression in transgenic mice. J Invest Dermatol 112, 893-898.

Leykauf,K., Kabsch,K., Gassler,N., Gissmann,L., Alonso,A., Schenkel,J.: Expression of the HPV11 E2 gene in transgenic mice does not result in alterations of the phenotypic pattern. Transgenic Research 17 (1), 1-8, 2008.

Abbildung 1

In der Publikationsdatenbank wiederum sind bei dem Artikel die Links zu den Mäusen in der Mausdatenbank aufgeführt, die darin publiziert sind. (Abb.2)

Journal-Artikel

Kostenstelle/n	W430, F020, F050
Allianzen / Kooperationen	
Programmatische Zuordnung	Krebsforschung
Artikel	Originalartikel
Titel	Expression of the HPV11 E2 gene in transgenic mice does not result in alterations of the phenotypic pattern
Autoren	Leykauf,K., Kabsch,K., Gassler,N., Gissmann,L., Alonso,A., Schenkel,J.
Journalname	Transgenic Research
Volume	17
Heftnummer	1
Seitenzahl von	1
Seitenzahl bis	8
Erscheinungsjahr	2008
Online First / epub ahead of print	Nein
Erscheinungsjahr	
in press	Nein
supplementary material	
Web of Science	000251869000001
PubMed	17701441
Laborbuchnummer	
Mausdatenbank	HPV11 E2 613 HPV11 E2 623 HPV11 E2 871 HPV11 E2 871 HPV11 E2 x URR11

Abbildung 2

Durch diese Verknüpfung bekommen die Nutzer die Information zusammenhängend präsentiert und können komfortabel zwischen beiden Datenbanken hin und her wechseln, ohne eine neue Recherche starten zu müssen. Dieses Konzept ist auf großes Interesse gestoßen und es sollen zukünftig weitere Datenbanken integriert werden.

Verbindung Laborbücher – Publikationsdatenbank

Neben der Anbindung von Datenbanken ist auch die Anbindung von Laborbüchern in der Planung. Im DKFZ sind zurzeit sowohl gedruckte Laborbücher als auch ein elektronisches Laborbuchsystem im Einsatz. Die elektronischen Laborbücher sollen, analog zu den Datenbanken, per Link mit den Veröffentlichungen in der Publikationsdatenbank verbunden werden. Für die gedruckten Laborbücher ist bereits ein Feld für die Laborbuchnummer angelegt (Abb. 2), so dass diese dort eingetragen werden kann. Da die Ausgabe und die Verwaltung der gedruckten Laborbücher durch die Bibliothek erfolgt, ist nicht nur nachzuvollziehen, in welchem Laborbuch Forschungsdaten zu der Publikation enthalten sind, sondern auch wo es sich befindet.

Verbindung Supplementary Plattform - Publikationsdatenbank

Für die bereits erwähnten vielen einzelnen Datenelemente, die nicht strukturiert in einer Datenbank abgelegt sind (Abbildungen, Tabellen usw.) wurde von der Bibliothek zusammen mit der Abteilung für Informationstechnologie (ITCF) eine Plattform für „Supplementary Material“ implementiert, auf der alle Arten von Daten gespeichert werden können. Der Wunsch nach einer solchen Möglichkeit, seine Daten schnell und komfortabel, unabhängig von den Servern der Zeitschriftenverlage, ablegen zu können, kam von Wissenschaftlern aus dem Haus. Die Verbindung der dort gespeicherten Materialien mit der entsprechenden Publikation in der Publikationsdatenbank erfolgt ebenfalls mittels eines Links.

Die Plattform wurde mittlerweile um einige Features erweitert und kann nun auch von den Forschenden dazu genutzt werden, Reviewern für einen befristeten Zeitraum Supplementary Material zu eingereichten Publikationen zugänglich zu machen.

Wird der Artikel angenommen und die Veröffentlichung in die Publikationsdatenbank eingetragen, dann können die Materialien schnell und komfortabel damit verbunden werden, ohne sie erneut hochladen zu müssen

Vorteile dieses Lösungsansatzes

Die Vorteile eines solchen pragmatischen Ansatzes sind sehr vielfältig. So lässt er sich schnell und unproblematisch mit allen Arten von Forschungsdaten, egal wie sie strukturiert sind und wo sie gespeichert sind, realisieren und damit erweitern und ausbauen. Dies beinhaltet zum Beispiel nicht nur Inhouse-Datenbanken sondern es kann auch auf externe Systeme verlinkt werden, wenn die Inhalte frei zugänglich sind. Aufgrund des Links in der Publikationsdatenbank auf die verschiedenen Systeme, die Forschungsdaten enthalten, bleiben deren „Besonderheiten“, wie zum Beispiel Zugangsbeschränkungen für bestimmte Nutzer, ohne zusätzlichen Aufwand erhalten. So kann sich ein externer Nutzer gegebenenfalls das Supplementary Material zu einer Publikation ansehen, weil es freigegeben ist, die damit ebenfalls verlinkten Daten aus einer Datenbank aber unter Umständen nicht, wenn diese nur DKFZ intern zur Verfügung steht. Für die Wissenschaftler bietet dieser Ansatz zwei große Vorteile. Erstens werden die Informationen zu allen Forschungsdaten, die einer Publikation zugrunde liegen, an einer Stelle, der Publikationsdatenbank, nachgewiesen. Zweitens sind die Forschungsdaten aufgrund der Verbindung zur Publikation durch diese gewissermaßen erschlossen. Die Aufbereitung von solchen Daten, die gewährleisten, dass sie schnell wiedergefunden werden können, ist sehr zeitaufwendig und wird von den Forschenden daher oft nicht praktiziert. Verlässt dann zum Beispiel ein Doktorand das DKFZ kommt es häufig vor, dass er eine große Menge von Dateien hinterlässt bei denen niemand auf den ersten Blick erkennen kann was sich genau dahinter verbirgt. Sind diese aber mit der Publikation verknüpft, so können sie den Projekten, in denen sie entstanden sind, zugeordnet werden.

Ergänzende und geplante Serviceangebote

Neben den bereits erläuterten Angeboten im Bereich Forschungsdatenmanagement hat die Bibliothek noch ergänzende Angebote etabliert. So ist zum Beispiel in den letzten Jahren ein zunehmender Beratungsbedarf in diesem Bereich festzustellen. Die Anzahl der teilweise sehr komplexen Fragestellungen an die Bibliothek zu diesem Themengebiet steigt stetig. Hier wird daher entsprechende Hilfestellung und Unterstützung angeboten. Weiterhin hat sich die Bibliothek für das DKFZ als Datenzentrum für Digital Object Identifiers (DOI) registrieren lassen. Falls gewünscht, können die Forschungsdaten mit DOIs versehen werden.

Ein Nachteil des bisherigen Systems besteht darin, dass die Forschungsdaten, erst nachdem es eine Publikation gibt, entsprechend eingetragen werden können. Das führt dazu, dass sie nach der Veröffentlichung des Artikels unter Umständen erst mühsam zusammengesucht und in der Publikationsdatenbank eingetragen werden müssen, was einen zusätzlichen Aufwand für die Nutzer bedeutet. Viel sinnvoller wäre es daher, ein System anzubieten, in dem die Daten schon im Vorfeld verzeichnet werden können, so dass sie nach der Veröffentlichung der Publikation schon vorhanden sind und nur noch mit ihr verbunden werden müssen. Die bisherige Publikationsdatenbank bietet eine solche Möglichkeit nicht. Die Bibliothek arbeitet zurzeit jedoch an der Implementierung einer neuen Publikationsdatenbank-Software. Sie ist Projekt-Partner im JOIN²-Projekt⁸ (Just an Other Invenio Instance) geworden. Das System basiert auf der INVENIO-Software, die vom CERN entwickelt wurde. Sie ist flexibel einsetzbar und kann an die unterschiedlichsten Anforderungen angepasst werden. Geplant ist, das System so auszubauen, dass das Management von Forschungsdaten auch unabhängig von der Veröffentlichung einer Publikation möglich sein soll. Eine solche Funktion wäre auch für zwei weitere Projekte, die bei der Bibliothek angefragt wurden, relevant. Zum einen wird im DKFZ momentan über die Gründung eines wissenschaftlichen Archivs nachgedacht. In ihm sollen relevante Forschungsdaten, die nicht oder noch nicht publiziert wurden, archiviert und erschlossen werden. Weiterhin soll eine Sammlung von Informationen zu Wissenschaftlern des DKFZ aufgebaut werden. Sie soll unter anderem deren wissenschaftlichen Lebenslauf abbilden und Publikationslisten, Informationen zu Preisen, Fotos usw. enthalten. Beide Projekte könnten sinnvoll in die neue Publikationsdatenbanksoftware integriert werden und sollen von der Bibliothek realisiert werden.

Dies macht deutlich, dass Bibliotheken schon jetzt und auch zukünftig wichtige Aufgaben zur Unterstützung der Wissenschaft im Kontext Forschungsdatenmanagement bieten können.

⁸ <https://join2-wiki.gsi.de/cgi-bin/view>

Referenz

Sicherung guter wissenschaftlicher Praxis: Denkschrift. – Ergänzte Auflage. – Bonn: Deutsche Forschungsgemeinschaft, 2013, S. 21
http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf

Ein kooperativ nutzbares Werkzeug zur Erstellung von Datenmanagementplänen

Franziska Ackermann¹, Dennis Wehrle²

¹Kommunikations- und Informationszentrum (kiz), Universität Ulm

²Professur für Kommunikationssysteme, Albert-Ludwigs-Universität Freiburg

Zusammenfassung

Wissenschaftler müssen sich zunehmend mit Richtlinien und Empfehlungen zum Datenmanagement auseinandersetzen, die von Drittmittelgebern, ihren Forschungseinrichtungen, von Fachdisziplinen, Verlagen und wichtigen Vertretern aus Wissenschaft und Politik ausgehen. Ein prominentes Beispiel ist das neue Element des Datenmanagementplans (DMP) in „Horizon 2020“, der auf verschiedene Aspekte wie z.B. Beschreibung, Backup und Nachnutzung der Daten eingeht. Projekte, die am „Open Data Pilot“ der Europäischen Kommission teilnehmen, müssen ein solches Dokument über das geplante Datenmanagement im Projekt einreichen und regelmäßig aktualisieren. Datenmanagementpläne haben sich im Vereinigten Königreich und den USA bereits zu einem Standard entwickelt. Dort können Wissenschaftler auf Software-Werkzeuge zurückgreifen (sog. DMP-Werkzeuge), die Vorlagen und Hilfestellung beim Erstellen eines passenden Datenmanagementplans beinhalten. Das vom Land Baden-Württemberg geförderte Projekt „Landesweit koordinierte Strukturen für Nachweis und effiziente Nachnutzung von Forschungsdaten“ (01/2014 – 02/2016) zwischen der Universität Ulm und der Universität Freiburg beinhaltete neben andere Zielen auch die Suche nach einer Lösung für ein DMP-Werkzeug, das von mehreren Einrichtungen in Baden-Württemberg kooperativ genutzt werden kann. Nach einem Vergleich bereits vorhandener Entwicklungen entschied man sich im Projekt dazu, den Quellcode des „DMPTool“ der California Digital Library nachzunutzen, das in den USA von einer Vielzahl von Einrichtungen gemeinsam genutzt wird. Die Software wurde von den Projektpartnern Ulm und Freiburg technisch und inhaltlich angepasst und beinhaltet z.B. eine Vorlage für „Horizon 2020“, die Wissenschaftler bei der Erstellung eines DMPs für ihr Projekt unterstützen kann. Über die Authentifizierung mittels Shibboleth/bwIDM soll das DMP-Werkzeug von Wissenschaftlern der teilnehmenden Einrichtungen unkompliziert für die Erstellung von Datenmanagementplänen genutzt werden können.

Abstract

Researchers increasingly face data management policies from funders, their research institutions, research communities, publishers and players from science and politics. A prominent example is the new element of a data management plan (DMP) in “Horizon 2020”, which includes aspects such as description, backup and re-use of data. Projects participating in the “Open Data Pilot” of the European Commission are asked to deliver such a document outlining their plan for data management in the project, and to keep it updated. Data management plans have become a standard in the UK and the USA, where researchers are offered tools that provide templates and guidance for writing a suitable data management plan. Among other goals, Ulm University and the University of Freiburg aimed to find a solution for a DMP tool that can be used cooperatively by several institutions in Baden-Württemberg in the project “Landesweit koordinierte Strukturen für Nachweis und effiziente Nachnutzung von Forschungsdaten”¹ (01/2014 – 02/2016), which was funded by the state Baden-Württemberg. After a comparison of existing DMP tools, the project partners decided to use the source code from “DMPTool” from the California Digital Library, which is used cooperatively by a large number of institutions in the USA. The project partners Ulm and Freiburg adapted “DMPTool” technically and content-wise. The adapted tool includes a template for “Horizon 2020”, for instance, which can support researchers when creating a data management plan for their project. It is intended to provide researchers from participating institutions with a comfortable way to use the tool for creating data management plans via Shibboleth/bwIDM authentication.

1. Motivation

Immer häufiger werden Wissenschaftler aufgefordert darzulegen, wie sie mit den erhobenen Daten in ihrem Forschungsvorhaben umgehen werden und ob und unter welchen Bedingungen die Daten nach Projektende für die Öffentlichkeit nachnutzbar sind. Nach der Praxis beim Datenmanagement fragen beispielsweise Drittmittelgeber wie die Deutsche Forschungsgemeinschaft (DFG) und die Europäische Union (EU).² Desweiteren verabschieden Forschungseinrichtungen in den letzten Jahren zunehmend Richtlinien und Empfehlungen zum Datenmanagement (Data Policies), in denen häufig die Erstellung eines Datenmanagementplans (DMP) gefordert wird.

¹ Statewide development of coordinated structures for indexing and re-use of research data

² DFG: http://www.dfg.de/formulare/54_01/54_01_de.pdf, Zugriff 25.04.2016

EU: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf, Zugriff 25.04.2016

Dabei handelt es sich um ein Dokument, das beschreibt, welche Art von Daten im Forschungsvorhaben genutzt und erhoben werden, wie sie gesichert werden, welche ethischen oder rechtlichen Aspekte eine Rolle spielen, wie die Nachnutzung der Daten geplant ist, etc. Neben Drittmittelgebern und Forschungseinrichtungen gibt es zudem einzelne Zeitschriften mit einer Data Policy, die Richtlinien und Empfehlungen für die Autoren zum Umgang mit den Daten enthält, die zu einer eingereichten Publikation gehören.³ Daneben gibt es weitere Empfehlungen zum Umgang mit Forschungsdaten, beispielsweise einzelner Fachdisziplinen oder von wichtigen Vertretern aus Wissenschaft und Forschung, wie der DFG⁴ und der Hochschulrektorenkonferenz (HRK)⁵.

Hinter der Forderung von Drittmittelgebern wie der EU und der DFG nach einem professionellen Datenmanagement und „Open Data“ steckt die Idee, dass öffentlich geförderte Forschung nach Möglichkeit wiederum öffentlich verfügbar und nachnutzbar sein soll. Wenn Daten stärker als bisher für eine Nachnutzung verfügbar wären, würde dies beispielsweise dazu beitragen, dass publizierte Ergebnisse besser nachvollzogen und verifiziert, dass kostspielige Neuerhebungen von Daten vermieden und einmalige Daten (z.B. Wetterdaten) mehrfach ausgewertet werden könnten.

1.1 Data Policies von Drittmittelgebern

Im aktuellen Rahmenprogramm für Forschung und Innovation „Horizon 2020“ der Europäischen Kommission gibt es als neues Element den Datenmanagementplan. Von Projekten, die dort am „Open Data Pilot“ teilnehmen, wird erwartet, dass sie spätestens ein halbes Jahr nach Projektbeginn einen Datenmanagementplan vorlegen und diesen aktuell halten. Die Europäische Kommission bietet einen Leitfaden zum Datenmanagement an, der unter anderem konkrete Fragen zum Umgang mit Forschungsdaten enthält, die von den Wissenschaftlern beantwortet werden sollen.⁶ Dieser Pilot der Europäischen Kommission zu offenen Daten ist insofern interessant, als dass der vorausgegangene Pilot zu Open Access für Publikationen inzwischen in einen Standard für sämtliche bewilligte Projekte überführt wurde.

³ Public Library of Science (PLOS), American Geophysical Unit (AGU) u.a.

⁴ http://www.dfg.de/foerderung/antragstellung_begutachtung_entscheidung/antragstellende/antragstellung/nachnutzung_forschungsdaten/, Zugriff 25.04.2016

⁵ http://www.hrk.de/uploads/tx_szconvention/Empfehlung_Forschungsdatenmanagement_final_Stand_11.11.2015.pdf, Zugriff 25.04.2016

⁶ Leitfaden: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf, Version 2.1 vom 15. Februar 2016, Zugriff 25.04.2016

Eine ähnliche Formulierung wie die der Europäischen Kommission findet sich im Leitfaden für die Antragsstellung der Deutschen Forschungsgemeinschaft (DFG):

„Wenn aus Projektmitteln systematisch Forschungsdaten oder Informationen gewonnen werden, die für die Nachnutzung durch andere Wissenschaftlerinnen und Wissenschaftler geeignet sind, legen Sie bitte dar, ob und auf welche Weise diese für andere zur Verfügung gestellt werden. Bitte berücksichtigen Sie dabei auch - sofern vorhanden - die in Ihrer Fachdisziplin existierenden Standards und die Angebote existierender Datenrepositorien oder Archive“ (DFG 2014, S.5)

Die Erwartungen der DFG sind detaillierter in den „Leitlinien zum Umgang mit Forschungsdaten“ formuliert, die im September 2015 vom Senat der DFG verabschiedet wurden.⁷ Diese Leitlinien zeigen zudem auf, wie die DFG die Wissenschaftler beim Datenmanagement unterstützt und beinhalten einen Appell an die wissenschaftlichen Fachgemeinschaften, disziplinspezifische Regularien zu entwickeln. Einzelne Fachdisziplinen sind dieser Aufforderung bereits nachgekommen, z.B. die Biodiversität, die Erziehungswissenschaften und die Digital Humanities (Digitale Editionen, Sprachkorpora).⁸

1.2 Data Policies an Forschungseinrichtungen in Deutschland

Neben Drittmittelgebern erwarten auch Forschungseinrichtungen verstärkt einen Datenmanagementplan oder ein vergleichbares Dokument. In Deutschland haben verschiedene Universitäten Richtlinien zum Datenmanagement verabschiedet (s. Tabelle). Data Policies gibt es zudem an weiteren, nicht-universitären Forschungseinrichtungen, wie z.B. dem Deutschen GeoForschungsZentrum in Potsdam und dem Robert Koch-Institut in Berlin.

⁷ http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf, Zugriff 25.04.2016

⁸ Der Appell an die wissenschaftlichen Fachgemeinschaften findet sich in den Richtlinien zum Forschungsdatenmanagement und zusätzlich auf der Webseite: http://www.dfg.de/foerderung/antragstellung_begutachtung_entscheidung/antragstellende/antragstellung/nachnutzung_forschungsdaten/, Zugriff 10.02.2016. Die bereits erarbeiteten disziplinspezifischen Regularien sind dort ebenfalls verlinkt.

Der Großteil der Universitäten mit einer Data Policy fordert, dass ein Datenmanagementplan erstellt wird:

Universität Bielefeld	„Ein „Data Management Plan“ und fachspezifische Richtlinien** für Einrichtungen und Projekte soll durch die jeweils verantwortliche Leitung bereitgestellt werden – vor allem in jedem neuen, Daten-intensiven Drittmittelantrag.“ (Universität Bielefeld 2011)
Universität Göttingen	„Forschungsprojekte mit Forschungsdaten erfordern einen Datenmanagementplan, der u.a. die Zugangsrechte und -vorbehalte der Forschungsdaten darlegt.“ (Universität Göttingen 2014)
Universität Heidelberg	„Teil jedes Forschungsprojektes ist ein Plan für das Datenmanagement, der explizit adressiert, wie die Akkuratheit, Vollständigkeit, Authentizität, Integrität, Vertraulichkeit, Veröffentlichung und der offene Zugang von Daten gehandhabt werden. Dabei werden fachspezifische Besonderheiten berücksichtigt.“ (Universität Heidelberg 2014)
Universität Kiel	„Das Forschungsdatenmanagement wird nach den aktuellen fachlichen Standards unter Einhaltung der Empfehlungen der DFG zur Sicherung der guten wissenschaftlichen Praxis durchgeführt und in einem Datenmanagement-Plan dokumentiert. Die Verantwortung für das Forschungsdatenmanagement liegt bei der Leitung der jeweiligen Forschungsaktivität.“ (Universität Kiel 2015)
TU Darmstadt	„Die Universität empfiehlt die Aufstellung eines Plans für den Umgang mit Forschungsdaten bereits vor Beginn eines Forschungsvorhabens, um einen systematischen und nachhaltigen Umgang mit den Daten zu gewährleisten. Ein Forschungsdatenmanagementplan enthält eine Beschreibung aller relevanten im Laufe des Vorhabens entstehenden Daten sowie ein Konzept für den Umgang mit ihnen im Hinblick auf Genauigkeit, Vollständigkeit, Authentizität, Integrität, Vertraulichkeit und Verbleib, sowie Urheber- und Nutzungsrechte. Dabei sind fachspezifische Besonderheiten und Standards zu berücksichtigen und der Plan dem jeweils aktuellen Verlauf des Vorhabens anzupassen.“ (TU Darmstadt 2015)

Es bleibt festzuhalten, dass Wissenschaftler zunehmend mit Richtlinien und Empfehlungen zum Datenmanagement konfrontiert werden, die für sie beispielsweise als Angehörige einer Forschungseinrichtung, als Antragsteller eines Drittmittelprojekts, oder als Angehörige einer Fachdisziplin gelten. Für die Erstellung von Datenmanagementplänen gibt es Werkzeuge, welche diese Bedingungen abbilden können und Wissenschaftler durch Beispieltexte, institutionsspezifische Hinweise, weiterführenden Links und zusätzliche Hilfestellung bei der Erstellung eines passenden DMPs unterstützen können.

2. Data Policies im anglo-amerikanischen Raum

Im anglo-amerikanischen Raum ist es bereits stark verankert, dass im Rahmen von Drittmittelprojekten Datenmanagementpläne erstellt werden. Die National Science Foundation ist hierfür ein prominentes Beispiel.⁹ Weitere Drittmittelgeber mit Vorgaben zu einem Datenmanagementplan sind die Alfred P. Sloan Foundation, das Department of Education, Department of Energy, und andere.¹⁰ In den USA und dem Vereinigten Königreich gibt es bereits etablierte Lösungen, die den Wissenschaftlern die Erstellung eines Datenmanagementplans erleichtern, sogenannte *DMP-Werkzeuge*. Sie nehmen den Projektteilnehmern nicht ab, sich mit den Themen des Datenmanagements auseinanderzusetzen, bieten jedoch verschiedene Vorlagen und zusätzliche Hilfestellungen in Form von Erläuterungen, Links, Beispiel- und Musterantworten an. Derzeit gibt es zwei etablierte DMP-Werkzeuge, die in den USA bzw. im Vereinigten Königreich kooperativ von jeweils mehr als 180 Einrichtungen genutzt werden. Diese sind das „DMPTool“¹¹ der California Digital Library (USA) und „DMPonline“¹² des Digital Curation Centers (UK). Die dort angebotenen Vorlagen und Hilfestellungen können von den teilnehmenden Einrichtungen jeweils an die eigene Situation angepasst werden, sodass die Wissenschaftler vor Ort die benötigte Unterstützung erhalten können.

⁹ <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>, Zugriff 18.02.2016

¹⁰ Alfred P. Sloan Foundation: <http://www.sloan.org/apply-for-grants/grant-proposals/>
Department of Education: http://ies.ed.gov/funding/datasharing_implementation.asp
Department of Energy: <http://www.energy.gov/datamanagement/doe-policy-digital-research-data-management>, Zugriff 26.04.2016

¹¹ <https://dmptool.org/>

¹² <https://dmponline.dcc.ac.uk/>

Das vom Land Baden-Württemberg geförderte Projekt „Landesweit koordinierte Strukturen für Nachweis und effiziente Nachnutzung von Forschungsdaten“ (01.01.2014 – 29.02.2016) beinhaltet als Teilaspekt die Suche nach einer Lösung für ein DMP-Werkzeug zur Erstellung von Datenmanagementplänen, das von mehreren Einrichtungen im Land Baden-Württemberg kooperativ genutzt werden kann.

3. Anforderungen an ein kooperatives DMP-Werkzeug

Ein wesentlicher Aspekt bei der Suche nach einem DMP-Werkzeug war der kooperative Charakter. Im Projekt wurde explizit nach einer existierenden Softwarelösung gesucht, die nur an einem Ort installiert und gepflegt werden muss und dennoch von einer Vielzahl von Einrichtungen genutzt werden kann. Dadurch muss nicht jede Einrichtung eine lokale Installation pflegen. Vielmehr können Ressourcen für die Bereitstellung und Weiterentwicklung des DMP-Werkzeugs gebündelt werden. Gleichzeitig sollte die Softwarelösung so flexibel sein, dass inhaltliche Anpassungen durch die teilnehmenden Einrichtungen möglich sind. So soll eine Einrichtung beispielsweise selbst Vorlagen für Datenmanagementpläne erstellen, anpassen und Hilfestellungen (Erläuterungen, Links, Beispiel- und Musterantworten) einbinden können, die auf die lokalen Gegebenheiten ausgerichtet sind.

Aus Sicht der Nachhaltigkeit ist eine möglichst große Community, welche das DMP-Werkzeug nutzt und unterstützt, von Vorteil. Eine Open Source Lösung ist ebenfalls von Vorteil, da hier zum einen durch die transparente Funktionsweise der Software bei Bedarf Anpassungen vorgenommen werden können und zum anderen einer Obsoleszenz entgegengewirkt werden kann.

Ein weiterer Aspekt, den es zu berücksichtigen gilt, betrifft die Datensicherheit und den Datenschutz. Die Universität Bielefeld berichtete, dass Wissenschaftler bei der Nutzung ihres DMP-Werkzeugs besonderen Wert auf diese beiden Punkte legten.¹³ Dies ist nachvollziehbar, da die Datenmanagementpläne Informationen zu den Daten im Projekt und den Beteiligten enthalten, die wiederum Rückschlüsse auf das Projektvorhaben ermöglichen.

¹³ Dieser Punkt findet sich im Fazit des Vortrags, der im Rahmen des Treffens der DINI/nestor AG Forschungsdaten im März 2015 von Johanna Vompras gehalten wurde. Folien: <http://www.forschungsdaten.org/images/3/31/04--vompras--lessons-learned-bielefeld.pdf>, Zugriff 10.02.2016

Damit das deutsche Datenschutzrecht Anwendung findet, wurde im Projekt die Prämisse gesetzt, dass die beiden Projektteilnehmer Ulm und Freiburg für das Testsystem eine eigene Installation an einem der beiden Standorte vornehmen, welche kooperativ genutzt werden kann. In Zukunft kann eine Migration auf einen anderen Server erfolgen, sofern dies keine Probleme bezüglich des Datenschutzes mit sich bringen würde.

Um Wissenschaftlern eine einfache Nutzung zu ermöglichen, wurde eine komfortable Authentifizierungsmethode angestrebt. Über bwIDM bzw. Shibboleth können sich Wissenschaftler der Universitäten in Baden-Württemberg beispielsweise mit ihren gewohnten institutionellen Login-Daten anmelden und müssen keinen separaten Account mit Passwort anlegen, was einen unkomplizierten Einstieg zur Erstellung eines Datenmanagementplans ermöglichen würde. Die Menüführung und Funktionen sollten ebenfalls weitestgehend selbsterklärend und schnell erlernbar sein.

4. Vergleich möglicher Kandidaten für ein kooperatives DMP-Werkzeug

Es gibt etablierte DMP-Werkzeuge, die bereits kooperativ genutzt werden und weitere Lösungen, die lokal an einer bestimmten Einrichtung zum Einsatz kommen. Bei den etablierten, kooperativ genutzten Angeboten handelt es sich wie eingangs erwähnt zum einen um „DMPonline“, dem DMP-Werkzeug des Digital Curation Center (UK) für Einrichtungen aus dem Vereinten Königreich, und zum anderen um das „DMPTool“, das von der California Digital Library (USA) für eine Vielzahl amerikanischer Universitäten und Einrichtungen bereitgestellt wird. Bei beiden DMP-Werkzeugen ist der Quellcode nachnutzbar, wovon teilweise bereits gebraucht gemacht wurde.¹⁴ In Deutschland gibt es an lokal eingesetzten DMP-Werkzeugen zum Beispiel das selbstentwickelte, in die Infrastruktur vor Ort eingebundene DMP-Werkzeug „TUB-DMP“ an der Technischen Universität Berlin, dessen Quellcode frei verfügbar ist.¹⁵ An der Universität Bielefeld kommt ebenfalls ein lokales DMP-Werkzeug zum Einsatz.¹⁶ Neben entsprechenden Softwareangeboten einzelner Einrichtungen gibt es auch den Ansatz, ein auf die Fachdisziplin zugeschnittenes DMP-Werkzeug anzubieten, wie es beispielsweise bei CLARIN-D der Fall ist.¹⁷

¹⁴ So gibt es z.B. eine spanische Version von DMPonline unter <http://dmp.consorcioadrono.es/>, Zugriff 21.04.2016

¹⁵ <https://github.com/tuub/TUB-DMP>, Zugriff 11.02.2016

¹⁶ <https://data.uni-bielefeld.de/de/data-management-plan>, Zugriff 11.02.2016

¹⁷ <http://www.clarin-d.de/de/aufbereiten/datenmanagementplan-entwickeln>, Zugriff 21.04.2016

Aufgrund der im vorausgegangenen Abschnitt genannten Kriterien an ein DMP-Werkzeug, insbesondere des kooperativen Charakters und des großen Nutzerkreises, rückten die beiden etablierten Angebote „DMPonline“ und „DMPTool“ in den Fokus und wurden genauer untersucht. Sie wurden zum Zeitpunkt des Vergleichs von jeweils weit über hundert Einrichtungen kooperativ genutzt – inzwischen sind es jeweils über 180 Einrichtungen – und ähneln sich in ihren Funktionen stark. Beide Softwarelösungen ermöglichen die Erstellung von weiteren Vorlagen durch die teilnehmenden Einrichtungen, sodass hier gezielt auf bestimmte Forschungsförderer und Fachdisziplinen eingegangen werden kann. „DMPTool“ bot jedoch einige interessante Features, die bei „DMPonline“ nicht zu finden waren, z.B. die Möglichkeit, einen Plan zu duplizieren, was für die Überarbeitung von Plänen hilfreich ist, oder die Option, einen Plan durch institutionelle Ansprechpartner begutachten zu lassen (Review-Prozess). Aufgrund der Möglichkeit die Software auf Basis der vorhandenen MIT-Lizenz den Bedürfnissen entsprechend anpassen zu können sowie des Vergleichs anhand der aufgestellten Kriterien wurde im Projekt entschieden, den Quellcode des „DMPTool“ der California Digital Library nachzunutzen und das DMP-Werkzeug technisch und inhaltlich anzupassen, damit es der Situation im deutschsprachigen Raum gerecht wird und von den Projektpartnern kooperativ genutzt werden kann. Für das im Projekt angepasste DMP-Werkzeug verwenden wir an dieser Stelle die Bezeichnung „DMP-Tool“, um es von den verschiedenen DMP-Werkzeugen und dem originalen „DMPTool“ begrifflich abzugrenzen.

Inzwischen haben sich die Datenmodelle der beiden DMP-Werkzeuge „DMPonline“ und „DMPTool“ stark aneinander angenähert, sodass das Digital Curation Center und die California Digital Library die Absicht erklärt haben, künftig gemeinsam an einer Lösung zu arbeiten. Es soll ein Prototyp auf Basis des Quellcodes von „DMPonline“ entstehen, der um Funktionalitäten von „DMPTool“ ergänzt wird, wie beispielsweise die Review-Funktion.¹⁸

¹⁸ Simms et al. *The Future of Data Management Planning: Tools, Policies, and Players*. International Digital Curation Conference (IDCC) 2016, Practice Paper, <https://dmptool.files.wordpress.com/2016/02/idcc16-simms-et-al-29-jan1.pdf>, S.7.

5. Was kann das DMP-Tool leisten?

Die folgende Beschreibung bezieht sich sowohl auf das „DMPTool“ der California Digital Library als auch auf die im Projekt vorgenommene Anpassung. Das DMP-Tool ist webbasiert und kann über eine einzige zentrale Installation von einer Vielzahl von Einrichtungen gemeinsam genutzt und den Wissenschaftlern zur Erstellung von Datenmanagementplänen angeboten werden. Der kooperative Charakter des DMP-Tools zeigt sich nicht nur in der technischen Komponente, sondern auch in der inhaltlichen. Teilnehmende Einrichtungen können bestehende DMP-Vorlagen nachnutzen und anpassen. Sie können Musterpläne erstellen und diese im DMP-Tool als Orientierung bereitstellen. Wissenschaftler wiederum können ihre Pläne als Beispiele auf der Webseite anbieten. Die Sichtbarkeit kann dabei auf die Wissenschaftler der eigenen Einrichtung beschränkt werden, komplett öffentlich oder komplett privat (Mitbesitzer mit eingeschlossen) sein.

Wissenschaftler können sich am DMP-Tool bequem mit ihren institutionellen Login-Daten anmelden. In der Weboberfläche können sie aus verschiedenen DMP-Vorlagen auswählen, z.B. zugeschnitten auf einen Forschungsförderer oder eine Fachdisziplin. Durch zusätzliche Hilfestellungen in Form von Links, Erläuterungen, Muster- und Beispielanworten werden die Nutzer bei der Erstellung des Plans unterstützt. Es können mehrere Personen an einem Plan schreiben, wobei eine Person der Hauptbesitzer ist. Diese kann weitere Personen, die bereits einen DMP-Tool Account¹⁹ besitzen müssen, als Mitbesitzer des Plans hinzufügen. Der Plan wird in der Applikation gespeichert und ist dadurch zentral abgelegt und für alle berechtigten Personen im Web einsehbar und veränderbar.

Es ist möglich, über das DMP-Tool bei Ansprechpartnern der Einrichtung Feedback zu den erstellten Plänen einzuholen. Diese Feedback-Funktion wird auf Ebene der Vorlagen festgelegt und bietet drei Möglichkeiten, die für die Pläne gelten, die mit der Vorlage erstellt wurden: kein Feedback, optionales Feedback oder obligatorisches Feedback. Hierdurch können die Chancen auf einen erfolgreichen Projektantrag und –verlauf erhöht werden. Datenmanagementpläne können als PDF oder DOCX exportiert werden.

¹⁹ Der Account wird beim erstmaligen Login angelegt.

Das DMP-Tool arbeitet mit einem differenzierten Rollenmodell. Neben einem oder mehreren *Administrator(en)* für die gesamte Webapplikation gibt es auch auf der Ebene der einzelnen Einrichtungen *institutionelle Administratoren*, die im DMP-Tool für die eigene Einrichtung die umfassendsten Berechtigungen haben. Wissenschaftler nehmen wie oben erwähnt die Rolle des Hauptbesitzers oder Mitbesitzers eines Plans ein. Feedback geben können sogenannte *Reviewer*. Reviewer erhalten die Pläne zur Begutachtung, welche von Wissenschaftlern der eigenen Einrichtung mit Vorlagen erstellt wurden, die die Review-Option anbieten. Die Erstellung der Vorlagen wird vom *Template Editor* vorgenommen, während die zusätzlichen Hilfestellungen (Links, Beispiel- und Musterantworten, Erläuterungen) vom *Resource Editor* erstellt werden. Eine Person kann selbstverständlich mehrere Rollen einnehmen.

6. Praktische Umsetzung und prototypischer Betrieb

Das von der California Digital Library angebotene „DMPTool“ wurde technisch und inhaltlich angepasst. Für die Anpassungen wurde ein *fork*²⁰ auf GitHub²¹ erstellt, in dem die Änderungen transparent nachvollziehbar sind. Von Anfang an fand eine Trennung bei der Entwicklung zwischen den optischen und den funktionalen Änderungen innerhalb von sogenannten *Branches* statt. Ziel dieser Trennung war es, dass die Erweiterung des Codes um die Mehrsprachigkeit unabhängig von etwaigen optischen Änderungen gepflegt wird, damit die funktionalen Änderungen so generisch wie möglich gehalten werden und somit für die Community von Interesse ist. Am Ende können die im Projekt vorgenommenen Anpassungen durch die Entwickler des ursprünglichen DMP-Werkzeugs in den Branch übernommen werden, der dem aktuellen Stand des von der California Digital Library verwendeten „DMPTool“ entspricht. Dies würde in Zukunft parallele Arbeiten am Code verhindern und bewirken, dass beide Installationen – die amerikanische und die im Projekt angepasste Softwarelösung – von Updates profitieren können. Für das angepasste DMP-Tool würde dies bedeuten, dass Neuerungen aus dem „DMPTool“ leicht übernommen werden können, da lediglich eine Integration der optischen Änderungen in die neue Version erfolgen muss.

²⁰ Ein fork ist eine Abspaltung und somit eine Kopie einer Software, die ab diesem Zeitpunkt unabhängig von der originalen Software entwickelt wird.

²¹ DMP-Tool, <https://github.com/derDaywalker/dmptool/>, Zugriff 22.04.2016

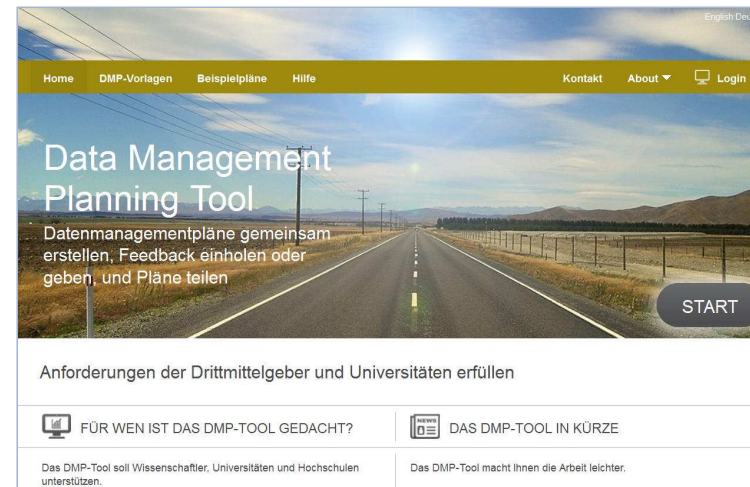


Abb.1: Startseite des DMP-Tools

Das DMP-Tool wurde im Rechenzentrum der Universität Freiburg installiert und wird dort technisch betreut. Im prototypischen Testsystem wird das DMP-Tool aktuell von den beiden Projektpartnern erprobt, der Universität Ulm und der Universität Freiburg. Statt dem Shibboleth-Login für amerikanische Einrichtungen wurde bwIDM angebunden, die Shibboleth-Förderung des Landes Baden-Württemberg. Aus Datenschutzgründen wurde eine zusätzliche Seite erstellt und so eingebunden, dass beim erstmaligen Login um Zustimmung zu den Hinweisen zum Datenschutz gebeten wird. Auch die Startseite wurde leicht modifiziert und das Menü verschlankt.

Ein wesentlicher Punkt war die Realisierung einer zweisprachigen Oberfläche (Englisch und Deutsch). Hierzu musste das DMP-Tool zunächst internationalisiert werden, d.h. die Textfragmente im Quellcode wurden herausgesucht und durch sogenannte Internationalisierungs-Variablen ersetzt. Die Texte wurde ins Deutsche übersetzt und – wo sinnvoll - auch inhaltlich angepasst. Insgesamt wurden mehr als 1050 Textstellen mit mehr als 5100 Wörtern sowie zehn komplett statische Seiten (Hilfeseiten, Datenschutzerklärung, ...) übersetzt. Durch die im Projekt vorgenommene Internationalisierung können nun auf leichte Weise weitere Sprachen hinzugefügt werden.

7. DMP-Vorlagen („Horizon 2020“ u.a.)

Für die Zielgruppe des DMP-Tools sind wie eingangs beschrieben verschiedene Richtlinien und Empfehlungen zum Datenmanagement relevant. Im Projekt wurden Vorlagen für „Horizon 2020“ erstellt, welche die Fragen und Hinweise aus dem Leitfaden der Europäischen Kommission zum Datenmanagement enthalten. Zusätzlich wurde eine generische Vorlage in das DMP-Tool eingepflegt, die für die Erstellung von Datenmanagementplänen für Projekte genutzt werden kann, bei denen keine konkreten Vorgaben durch Drittmittelgeber oder von anderer Seite existieren. Die Erstellung weiterer Vorlagen ist ein Teilziel von Projekten aus der E-Science-Ausschreibung des Landes Baden-Württemberg. Dort sollen beispielsweise in Zusammenarbeit mit Wissenschaftlern aus der Computerlinguistik (Replay-DH) und den Ingenieurwissenschaften/ der Informatik (bwFDM-Soft) jeweils geeignete Vorlagen erstellt werden, welche sich an den in den Projekten zu entwickelten Archivierungsabläufen orientieren. Ideal ist es, wenn die jeweiligen Fachdisziplinen und die Drittmittelgeber bzw. die Autoren der Data Policies in die Erstellung der Vorlagen einbezogen werden. Wenn es bereits eine detaillierte Anleitung für einen Datenmanagementplan gibt, wie es z.B. bei „Horizon 2020“ der Fall ist, können die entsprechenden Fragen und Hinweise des Forschungsförderers für die Erstellung einer Vorlage verwendet werden, die dann speziell von Projekten dieses Forschungsförderers genutzt werden kann. Die auf diese Weise erstellte Vorlage kann wiederum von jeder teilnehmenden Einrichtung angepasst werden, indem z.B. institutionsspezifische Hinweisen ergänzt werden.

Die im DMP-Tool angebotenen Vorlagen zu „Horizon 2020“ können durch jede teilnehmende Einrichtung um einrichtungsspezifische Hinweise wie Erläuterungen, Links, Muster- und Beispielantworten ergänzt werden. Gibt es ein Datenrepositorium an der Universität, kann hierfür beispielsweise ein Mustertext verfasst werden, den die Wissenschaftler der Einrichtung für ihren Plan nutzen können. Neben einrichtungsspezifischen Hinweisen bietet das kooperative DMP-Tool auch die Möglichkeit, Hilfestellungen zu erstellen, die global für alle Einrichtungen sichtbar sind. Hiervon wurde bei der Vorlage für „Horizon 2020“-Projekte Gebrauch gemacht, damit allgemeingültige Hilfestellungen nicht von jeder Einrichtung separat eingepflegt werden müssen.

Die Europäische Kommission sieht vor, dass Projekte im „Open Data Pilot“ den Datenmanagementplan aktuell halten und zu bestimmten Zeitpunkten erneut abgeben.

Das DMP-Tool ermöglicht ein leichtes Duplizieren von Plänen, sodass die ursprüngliche Version beibehalten werden kann. Die duplizierte Version beinhaltet bereits die bisherigen Antworten und kann so auf unkomplizierte Weise aktualisiert werden.

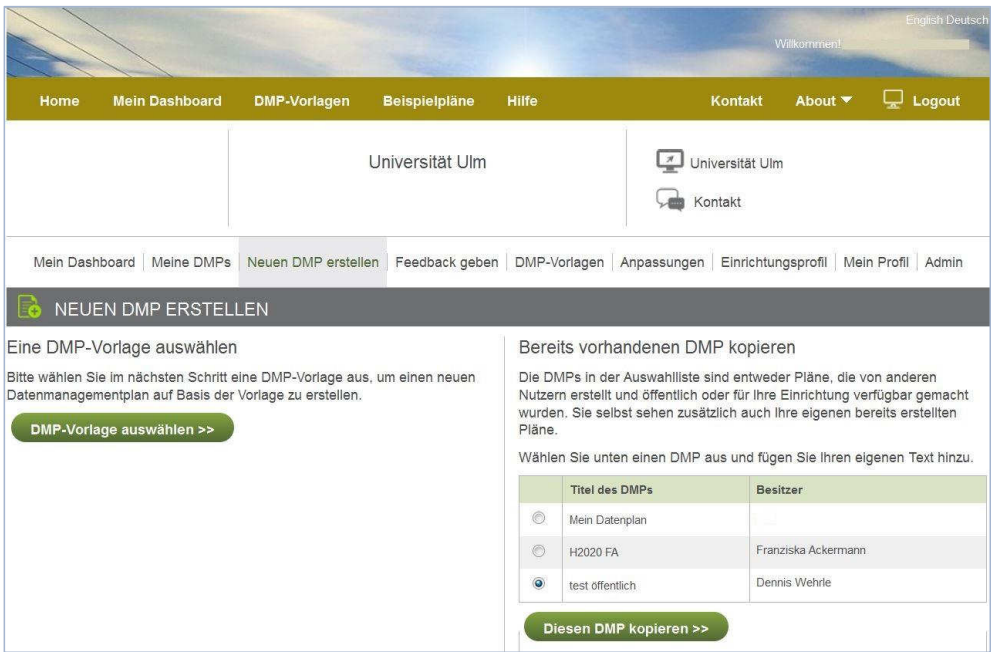


Abb.2: Einen neuen DMP erstellen

8. Fazit

Im Projekt wurde durch die Anpassung des in den USA weit verbreiteten „DMPTool“ eine Lösung für ein kooperatives DMP-Werkzeug geschaffen, das landesweit zum Einsatz kommen und Wissenschaftler bei der Erstellung von Datenmanagementplänen, z.B. für „Horizon 2020“, unterstützen kann. Es soll zunächst an den Standorten Ulm und Freiburg angeboten werden und kann über die Anbindung an Shibboleth / bwIDM darüber hinaus an weiteren Einrichtungen zum Einsatz kommen.

Der Vorteil des webbasierten DMP-Tools besteht unter anderem darin, dass es nur an einer Stelle installiert werden muss und dennoch von einer Vielzahl von Einrichtungen genutzt werden kann. Die Erstellung weiterer Vorlagen und die Anreicherung der Vorlagen durch geeignete Hilfestellungen, Musterantworten sowie einrichtungs- oder fachspezifische Gegebenheiten werden sicherlich eine bedeutende Rolle für die Akzeptanz des DMP-Tools bei den Wissenschaftler darstellen.

Durch den Aufruf der DFG zur Erarbeitung von disziplinspezifischen Data Policies und die zunehmende Wichtigkeit des Themas Forschungsdatenmanagement in Wissenschaft und Politik wird deutlich, dass ein kooperativ nutzbares DMP-Werkzeug ein hilfreicher Baustein in der Forschungsdaten-Infrastruktur des Landes darstellen kann.

Literaturverzeichnis

DFG 2014. Leitfaden für die Antragstellung: Projektanträge.
http://www.dfg.de/formulare/54_01/54_01_de.pdf, Zugriff 25.04.2016

TU Darmstadt, 2015. Leitlinien zum Umgang mit digitalen Forschungsdaten an der TU Darmstadt. http://www.tu-darmstadt.de/media/dezernatvi/relaunch_2015/gute_wiss_praxis/Leitlinien_Forschungsdaten_2015.pdf, Zugriff 10.02.2016

Universität Bielfeld, 2011. Resolution zu Forschungsdatenmanagement. <https://data.uni-bielefeld.de/policy>, Zugriff 10.02.2016

Universität Göttingen, 2014. Forschungsdaten-Leitlinie der Universität Göttingen (einschl. UMG). <http://www.uni-goettingen.de/de/01-juli-2014-forschungsdaten-leitlinie-der-universitaet-goettingen-einschl-umg/488918.html>, Zugriff 10.02.2016

Universität Kiel, 2015. Leitlinie zum Umgang mit Forschungsdaten: Empfehlungen zum Umgang mit Forschungsdaten der Christian-Albrechts-Universität zu Kiel. <http://www.uni-kiel.de/download/pm/2015/2015-408-leitlinie-forschungsdaten.pdf>, Zugriff 10.02.2016

Universität Heidelberg 2014. Research Data Policy: Richtlinien für das Management von Forschungsdaten. <https://www.uni-heidelberg.de/universitaet/profil/researchdata/>, Zugriff 10.02.2016

Wissenschafts- evaluierung

Wie messen wir unseren Open-Access-Anteil?

Marco Tullney

Technische Informationsbibliothek (TIB), Hannover

Einleitung

Open Access ist auf dem Siegeszug. Die „Open-Access-Transformation“ hat sich als neues Schlagwort etabliert, und Verlage konzentrieren sich mehrheitlich darauf, weitere Einnahmen durch Open Access zu generieren, statt die Entwicklung zu bekämpfen. Die von wissenschaftlichen Einrichtungen getragenen Kosten für Open-Access-Artikel steigen (ganz zu schweigen vom enormen Arbeitsaufwand, den manche der zugrundeliegenden Modelle mit sich bringen). Open Access ist ein Thema, das prominent in Auflagen von öffentlichen und privaten Förderern platziert wird. Wissenschaftliche Einrichtungen geben sich Richtlinien und verabschiedeten Strategien, die Open Access je nach Mut und Schwerpunktsetzung erzwingen oder herbeiwünschen.

Wenn von einer Erhöhung des Open-Access-Anteils die Rede ist oder spätestens dann, wenn Prozentzahlen vorgegeben werden, wird evident, dass die Zählgrundlage unklar ist. So haben wissenschaftliche Einrichtungen im Land Berlin folgende strategische Vorgabe erhalten:

„Bis 2020 sollen mindestens 60% der von wissenschaftlichen Einrichtungen des Landes Berlin veröffentlichten Zeitschriftenartikel über Green- oder Gold-Open-Access frei zugänglich sein.“ (Senat von Berlin, 2015).

Und bei der Fraunhofer-Gesellschaft ganz ähnlich:

„Konkret wird angestrebt, bis 2020 mindestens jede zweite wissenschaftliche Publikation eines Jahrgangs per Open Access frei zugänglich zu machen, davon mindestens ein Drittel über den goldenen Weg der Erstveröffentlichung in einer Open Access-Zeitschrift.“ (Fraunhofer-Gesellschaft, 2015).

Was aber wird dort gezählt? Offensichtlich geht es um Zeitschriftenartikel, die direkt (gold) oder parallel (grün) frei zugänglich gemacht werden sollen, aber wer kann wie entscheiden, ob dieses Ziel erreicht wird? Und wie sinnvoll ist ein Ziel, dessen Einhaltung man nicht überprüfen kann?

Was ist Open Access?

Open Access nach meinem Verständnis sind Publikationen, die

- am Ersterscheinungsort unmittelbar (ohne Zeitverzug) und ohne Barrieren kostenlos lesbar sind und die
- gemäß der Berliner Erklärung (Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003) frei von Beschränkungen der Nachnutzbarkeit sind.

De facto reden wir in heutigen Diskussionen, wenn es um Open-Access-Anteile geht, allenfalls über das erste Kriterium. Nachnutzbarkeitsanforderungen spielen häufig keine Rolle, obwohl Open-Access-Zeitschriften sich inzwischen weitgehend auf CC BY als Standardlizenz geeinigt haben, obwohl es Förderer gibt, die eine Veröffentlichung unter CC BY verlangen (vgl. (Bill & Melinda Gates-Stiftung, 2016)), und obwohl es Empfehlungen von Wissenschaftsorganisationen gibt zur Nutzung von CC BY (vgl. (Allianz der deutschen Wissenschaftsorganisationen, 2014)). Doch wenn es um die Verpflichtungen vieler öffentlicher Förderer geht, um die Strategiepapiere von Wissenschaftsorganisationen oder um Open-Access-Policies von Hochschulen, dann dominiert ein vager Open-Access-Begriff von „kostenlos lesbar“. Schon die Anforderung „unmittelbar“ gerät bisweilen unter Rechtfertigungsdruck.

Nun kann akzeptiert werden, dass für die meisten individuellen Leser/innen die meisten Wünsche erfüllt sind, wenn es einen nicht weiter behinderten direkten und legalen Zugriff auf eine Veröffentlichung gibt, und hier ist nicht der Ort, über die verpassten Chancen zu sprechen, die sich für Gesellschaft, Wissenschaft und Autor/innen aus der fehlenden freien Lizenzierung ergeben. Operieren wir also im folgenden unter der Minimalannahme, dass Open Access einen unmittelbaren kostenlosen Zugriff auf wissenschaftliche Publikationen meint.

Thesen

Ich gehe davon aus, dass die Frage nach dem Open-Access-Anteil von großer Bedeutung ist. Ich möchte dazu zehn Thesen diskutieren. Alle diese Punkte spielen eine Rolle bei der Messung des Open-Access-Anteils einer Institution.

- Open-Access-Anteil wird ein zentraler Indikator.
- Open Access ist ein eigenständiges Qualitätsmerkmal.
- Die weiche Open-Access-Formulierung ist ein großer Nachteil.
- Politischer Druck führt teilweise zu Fehlentwicklungen.

- Es braucht ein standardisiertes Vokabular.
- Green Open Access zu zählen, ist kompliziert.
- Gold Open Access zu zählen, ist manchmal schwierig.
- Es gibt verschiedene Wege zu Open Access, und es gibt mehr als Zeitschriftenartikel.
- Daten über Open Access müssen frei sein.
- Komplette Bibliografien sind der erste notwendige Schritt.

Open-Access-Anteil wird ein zentraler Indikator

Die zunehmenden Verweisen auf einen steigenden Open-Access-Anteil als wissenschaftspolitisches Ziel werden dazu führen, dass der Open-Access-Anteil einer Einrichtung berichtsrelevant wird. Das Monitoring des Open-Access-Anteils unterstützt auch die eigenen Aktivitäten und Planungen, aber es wird eben auch relevant sein in der Wahrnehmung der eigenen Einrichtung.

Analoge Entwicklungen gibt es auch bei Förderern – allerdings sind hier das Monitoring und auch das Verpflichten auf Open Access deutlich einfacher umzusetzen.

„Funding agencies will work together with publishers and the repository community to explore ways to not only count the actual number of open access research articles (be it on publishers' web sites, be it on open access repositories), but to also give an accurate estimate on the proportion of an agencies' portfolio that is available in open access“ (Global Research Council, 2013)

Spätestens dann, wenn Open-Access-Anteile auf Landesebene betrachtet werden, wird selbstverständlich geschaut werden, wie hoch der Open-Access-Anteil in den einzelnen Einrichtungen ist und wer wie stark zum aggregierten Anteil beiträgt. Dies mag umstritten sein, und es werden Besonderheiten von Institutionen und Disziplinen betrachtet werden, aber grundsätzlich ist die Entwicklung in dieser Form zu erwarten und auch inhaltlich zu begrüßen – wer mehr Open Access fordert, muss dies auch beobachten und messen können.

Und selbstverständlich ist es bereits jetzt möglich, Open-Access-Anteile auch für einzelne Wissenschaftler/innen anzuschauen. Dies kann durch automatisches Parsen von Publikationslisten (mit allen bekannten Mängeln solcher Verfahren) vorbereitet und von Hand angereichert werden.

Dabei wird es auch möglich sein, den Abstand zwischen potentielltem Open Access (z.B. Publikationen, für die Verlage akzeptable Zweitveröffentlichungserlaubnisse erteilen) und tatsächlichem Open Access (tatsächlich frei zugänglich bereitgestellte Publikationen) zu berechnen und als Kennzahl zu einer Person zu benennen. Solche Abstandsmessungen können Repository-Aktivitäten unterstützen und Anreize für mehr Open Access schaffen.

Der dänische *Open Access Indicator* ist ein gutes Beispiel dafür, wie technische Unterstützung und eine starke Open-Access-Mandatierung ineinandergreifen können:

The Open Access indicator monitors how the Danish universities fulfil the targets of the National Strategy for Open Access. Each year, the Indicator collects data about the Danish production of scientific publications and divides them into three categories: Realised Open Access (...) Unused Open Access potential (...) Unclear Open Access potential. (Danish Ministry of Higher Education and Science, 2016)

Das Resultat ist unter anderem ein Ranking der dänischen Forschungseinrichtungen nach Open-Access-Anteil (Danish National Research Database, 2016).

Open Access ist ein eigenständiges Qualitätsmerkmal

Dass Open-Access-Publikationen schlechter sind als nicht freie Publikationen, ist mittlerweile vermutlich ausgeräumt. Aber sie sind auch nicht automatisch besser – inhaltliche Fragen sind unabhängig von Open Access.

Jedoch ist Open Access eine weiteres Merkmal wissenschaftlicher Publikationen, das erfüllt sein kann – oder eben nicht. Ähnlich wie ein verständlicher Stil oder die Zugänglichkeit zugrundeliegender Daten definiert die Zugänglichkeit ein Qualitätsmerkmal, das in die Gesamtbeurteilung einfließt. Der gleiche Text, die gleiche Forschung im Open Access hat Vorteile gegenüber einer identischen Publikation, die nicht frei zugänglich ist. Sie ist einfacher überprüfbar, hat eine potentiell höhere Reichweite und steht günstigstenfalls unter freien Nachnutzungsbedingungen, die darauf aufbauende Forschung erleichtert und ermöglicht.

Die Messung von Open-Access-Anteilen und deren Abgleich mit wissenschaftspolitischen Zielsetzungen und Vorgaben unterstützt die Sicht auf Open Access als eine wichtige Eigenschaft wissenschaftlicher Publikationen. Wissenschaftler/innen werden sich häufiger mit Fragen nach dem Open-Access-Anteil an ihrer Publikationsliste konfrontiert sehen.

Wissenschaftliche Einrichtungen können ihren Teil dazu beitragen, dass Open Access nicht mehr als Qualitätsnachteil oder als „nice to have“ eingestuft wird, sondern als ein wichtiger Baustein der Gesamtsicht auf wissenschaftlichen Output.

Die weiche Open-Access-Formulierung ist ein zentraler Nachteil

Was nicht definiert ist, kann nicht gezählt werden. Die aktuellen Diskussionen und Vorgaben zu Open Access versäumen häufig eine sinnvolle Fundierung des vagen politischen Ziels „mehr Open Access“. Gründe hierfür liegen u.a. in der Rücksichtnahme auf unterschiedliche disziplinäre Publikationswege und in der Formulierung (schnell) erreichbarer Ziele. Damit wird mittel- und langfristig jedoch die Vergleichbarkeit untergraben, was die Versuche eines Open-Access-Monitorings auf nationaler, europäischer oder globaler Ebene gefährdet.

In den ersten Jahren der Open-Access-Bewegung war Open Access vor allem ein symbolischer Begriff, ein allgemeines Ziel. Den von Beginn an vorgenommenen Definitionsversuchen (und gerade die Berliner Erklärung, die auch von deutschen Wissenschaftseinrichtungen in großer Zahl unterzeichnet worden ist, ist relativ klar in ihren Anforderungen an Open Access) konnte man sich ohne große Konsequenzen entziehen. Doch bereits damals kam es zu Begriffsauslegungen, die vornehmlich im Sinn hatten, der neuen Forderung den Wind aus den Segeln zu nehmen, etwa durch den Verweis darauf, dass durch Bibliotheken doch sämtliche wissenschaftliche Literatur einfach zugänglich sei und damit quasi „Open Access“. Und auch heute kommt es zu munteren Neu- und Alternativinterpretationen. Beim Aufweichen des Open-Access-Begriffs geraten als erstes die Nachnutzungsrechte unter die Räder – gratis muss reichen. Als nächstes wird über den Zeitpunkt gesprochen – aus „unmittelbar“ wird dann „nach X Jahren“.

Die Vorgaben auf politischer Ebene oder in den Strategien und Open-Access-Richtlinien wissenschaftlicher Einrichtungen beziehen sich in aller Regel auf unmittelbar am Erstveröffentlichungsort frei zugänglich oder auf parallele gleichzeitige oder zeitverzögerte Bereitstellung, insbesondere via Repositorien. Dabei wird nur grob zwischen diesen zwei Wegen unterschieden, aber nicht immer genau ausgeführt, wann dies geschehen soll oder wann gemessen werden soll.¹

¹ Siehe Suber (2012), S. 49ff., für Definitionsversuche zu „gold“ und „grün“ sowie ähnlich Informationsplattform open-access.net (2016).

Wer den Open-Access-Anteil zählen (lassen) möchte, muss angeben, was hierbei gezählt werden soll. Ansonsten sind nicht vergleichbare Zahlen und eventuell vergeblicher Zählaufwand die Folge.

Politischer Druck führt teilweise zu Fehlentwicklungen

Rankings wie der genannte *Open Access Indicator* oder andere Versuche, durch die Politik den Open-Access-Anteil als Zielvorgabe zu setzen und zu messen, produzieren Anreize zur Erhöhung dieses Anteils. Diese Anreize können dazu führen, dass mehr Dokumente in den freien Zugriff gebracht werden. Sie können (und werden) aber auch dazu führen, dass Definitionen so gewählt werden, dass ohne zusätzliche Anstrengung möglichst viele Publikationen als „Open Access“ gezählt werden.

Die bereits angesprochene Begriffsunklarheit unterstützt diese negative Entwicklung. Im Kontext einer Entwicklung zu mehr Open-Access-Berichten und mehr Fragen nach dem Open-Access-Anteil erhält die Antwort auf „Was zählen wir für die Berechnung des Anteils?“ eine neue Qualität, sie entscheidet jetzt nämlich über die Erfüllung strategischer Ziele und Vorgaben. Deshalb ist die politische Fokussierung auf einen höheren Open-Access-Anteil nur dann sinnvoll, wenn möglichst genau geklärt ist, was hiermit gemeint ist. Andernfalls werden negative Anreize gesetzt, den Open-Access-Begriff pragmatisch anzupassen. Hierbei spielen zwei Faktoren eine Rolle:

- Um überhaupt mit einer Prozentzahl antworten zu können, muss der Open-Access-Anteil einfach berechenbar sein. Verstreut veröffentlichte Publikationen sind nur mit großem Aufwand feststellbar, und es könnte zu pragmatischen Anpassungen kommen, nach der nur an zentralen Orten (Datenbanken) Informationen erhoben werden.²
- Damit der zu berichtende Anteil möglichst hoch ausfällt, werden Publikationen als frei zugänglich angegeben, die dies nicht sind. Gründe hierfür können tatsächliche Fehler oder bewusste Falschangaben sein. Dies ist v.a. dann eine Gefahr, wenn die Autor/innen selbst angeben, ob ein Artikel Open Access ist.³

² Dies könnte kleine, institutionell veröffentlichte Zeitschriften treffen, aber auch z.B. verstreut auf Webseiten archivierte Publikationen (siehe z.B. Gutknecht et al. (2016)).

³ Beispiel hierfür könnten die Open-Access-Angaben in Forschungsinformationssystemen sein, wenn sie durch die Autor/innen gemacht werden. Dies geht z.B. in das Open-Access-Monitoring in den Niederlanden ein: „Whereas ‘Gold, DOAJ classified OA’ articles are determined by the nature of the journal in this case the author indicates that the article is OA within a restricted or not DOAJ journal. The university has to set up a control mechanism to ensure that these kinds of articles are really OA.“ (VSNU, 2016.) – Dies kann eine wichtige Aufgabe für Bibliotheken im Kontext von Forschungsinformationssystemen sein.

Wenn es möglich ist, zu manipulieren oder auch nur die Definition entgegen den Interessen der Fragenden zu verändern, wird dies auch geschehen. Wenn also nach konkreten Open-Access-Anteilen gefragt wird oder diese als Ziele vorgegeben werden, muss klar definiert werden, was gezählt werden soll. Generell gilt auch für diesen Bereich, dass Open-Access-Richtlinien so klar wie möglich formuliert sein sollten (Hunt and Picarra, 2016), um wirksam und erfolgreich zu sein.

Es braucht ein standardisiertes Vokabular

Um über verschiedene Aggregationen, national und international, verschiedene Disziplinen etc. hinweg diskussionsfähig zu sein und vergleichbare Daten zu produzieren, braucht es nicht nur ein im jeweiligen Fall klar definiertes Vokabular, sondern auch eine Form von Standardisierung. Sind 80% Open Access in den Niederlanden das gleiche wie 80% Open Access in Dänemark? Wie verhalten sich 60% Open Access im Bundesland Berlin zu 50% Open Access bei Fraunhofer?

Die Probleme internationaler Standardisierung sind nicht sofort auszuräumen, dies gilt erst recht in einem derart umkämpften Themenfeld. Insofern sind die genaue Darlegung der eigenen Messungen und die Offenlegung aller Daten (s.u.) wichtige Annäherungen an dieses Ziel. Doch können bereits jetzt auch sinnvolle Anleihen z.B. bei den Open-Access-Forderungen von Förderern gemacht werden. So sollte es ein leichtes sein, für die Zählung relevante Zeiträume zu benennen – wie schnell muss ein Artikel zugänglich gemacht werden, damit er (noch) als Open Access zählt? Wo werden solche Informationen hinterlegt, wo müssen die Artikel auffindbar sein?

Eine besondere Bedeutung könnte anderen normierten Informationen zukommen – so könnte etwa eine stärkere Verwendung von ORCID-Identifiern die Auffindbarkeit und insbesondere die Aggregation und Auswertbarkeit von Publikationslisten stark unterstützen. Die Verwendung etablierter und erfolgreicher Infrastruktur, insbesondere von SHERPA/RoMEO und DOAJ, spielt ebenfalls eine Rolle. Hier sehe ich auch stärkere Möglichkeiten für Institutionen und Bibliotheken, solche Initiativen zu unterstützen – nicht nur finanziell, sondern auch durch das Teilen von Informationen und das Mitwirken an qualitätssichernden Verfahren.

Nur ein klares und vergleichbares Vokabular liefert klare und vergleichbare Statistiken, und nur solche klaren und vergleichbaren Daten können sinnvoll Aufschluss geben über den erreichten Fortschritt und über den Erfolg auch unterschiedlicher Strategien und Wege zu Open Access.

Green Open Access zu zählen, ist kompliziert

Zweitveröffentlichungen über Repositorien sind in den Open-Access-Anforderungen von Förderern ähnlich prominent vertreten wie in den Open-Access-Strategien aus Wissenschaftspolitik und Wissenschaftseinrichtungen. Doch selten ist definiert, was hierbei eigentlich gezählt werden soll, explizit hier fehlt das standardisierte Vokabular. Wenn Zweitveröffentlichungen von Publikationen als frei zugänglich gezählt werden sollen, welche Anforderungen gelten? Aus welchem Publikationsjahr? Sind Publikationen mitgezählt, die flüchtig über institutionelle oder individuelle Homepages zugänglich gemacht werden oder gar über ResearchGate und ähnliche Dienste? Und insbesondere: Wann müssen sie frei zugänglich gemacht werden? Hierzu gehören zwei Aspekte:

- Gibt es eine maximale Verzögerung (Embargo), nach dessen Überschreiten eine Publikation nicht mehr mitgezählt wird? Ist eine Veröffentlichung *Open Access*, wenn sie nach fünf oder zehn Jahren in einem Repository auftaucht?
- Wann wird der Open-Access-Anteil für ein Publikationsjahr gemessen?

Speziell die letzte Frage wirft große praktische Probleme auf. Wenn im laufenden Jahr der Publikationsoutput des Vorjahres berichtet wird, was durchaus eine realistische Annahme ist, dann werden Publikationen, die erst nach zwei Jahren frei zugänglich gestellt werden, nicht mitgezählt. (Was aus Open-Access-Sicht auch nicht so falsch ist...) Hier wären Anstrengungen wissenschaftlicher Einrichtungen nötig, Publikationen schneller zu erfassen und auf kürzere Dauer zwischen Erstveröffentlichung und Zweitveröffentlichung zu drängen. Hierbei geht es nicht nur um das, was mit Verlagen vereinbart werden kann, sondern auch um das faktische Einstellen in Repositorien, mithin um originäre Leistungen von Wissenschaftler/innen und Einrichtungen wie Bibliotheken.

Ein wesentlicher Bezugspunkt für die Berechnung von Open-Access-Anteilen sollte das Publikationsjahr der Erstveröffentlichung sein. So wertvoll nachträgliche Digitalisierungen oder nachträgliche freie Veröffentlichungen sein mögen, sie stehen in einem unmittelbaren Zusammenhang mit der Erstveröffentlichung – nicht zuletzt deshalb, weil dieses Jahr der zeitliche Kontext ist, in dem die Forschungsarbeit zu verorten ist. Eine nach 10 oder 20 Jahren digitalisierte Arbeit kann, je nach Disziplin, durch ihre Zugänglichmachung immer noch einen gehörigen Mehrwert bringen, aber sie ist eben auch sehr viele Jahre nicht optimal zugänglich gewesen.

Wenn Zeitveröffentlichungen zum Open-Access-Anteil gerechnet werden, dann muss definiert sein, unter welchen Bedingungen sie dazugehören und wann ihre Verfügbarkeit gemessen wird.⁴

Gold Open Access zu zählen, ist manchmal schwierig

Gegenüber dem Messen von Zweitveröffentlichungen sollte das Messen von originären Open-Access-Veröffentlichungen deutlich einfacher sein. Doch auch hier lauern Probleme. Sie beziehen sich einerseits auf den Zeitpunkt der Messung, andererseits wiederum auf die Frage, wo gemessen wird.

Probleme mit einer zeitverzögerten Messung ergeben sich insbesondere beim nachträglichen Parsen von Literaturlisten. Die Tatsache, dass eine Publikation in einer Zeitschrift erschienen ist, die sich jetzt im DOAJ⁵ befindet, bedeutet nicht, dass die Zeitschrift schon zum Publikationszeitpunkt in diesem Verzeichnis gelistet oder überhaupt frei zugänglich gewesen ist. Auch andersherum: Dass eine Zeitschrift nun nicht mehr im DOAJ gelistet ist, schließt nicht aus, dass die Publikation zum Erscheinungszeitpunkt Open Access war. Noch komplizierter wird es bei Zeitschriften, die eine Moving Wall praktizieren, nach der alle Publikationen für eine bestimmte Dauer nicht frei zugänglich sind und nach Ablauf dieser Zeit die Zugriffssperre aufgehoben wird. Zählt die Publikation nun als Open Access, wenn sie zum Messzeitpunkt frei zugänglich ist, aber zum Veröffentlichungszeitpunkt nicht frei zugänglich war? Hier braucht es klare Definitionen (s.o.), aber auch ein sinnvolles Vorgehen, wann gemessen wird. Je näher am Veröffentlichungszeitpunkt gemessen wird, umso genauer die Ergebnisse.

Nicht alle Open-Access-Artikel erscheinen in Open-Access-Zeitschriften. Viele Artikel werden frei zugänglich gemacht, obwohl sie in nicht-freien Subskriptionszeitschriften erscheinen – gegen eine weitere Zahlung oder auf Basis anderer Vereinbarungen. Es ist einer von mehreren Nachteilen dieses „hybriden“ Modells, dass entsprechende Artikel nicht so einfach als Open Access auffindbar und sichtbar sind. Hier müssen geeignete Kombinationen gefunden werden aus besserer Markierung solcher Artikel auf Verlagsseite, genauem Reporting durch die Autor/innen und bibliothekarischer

⁴ Darüber hinaus ist es sicherlich auch interessant, Komplettberechnungen anzustellen: Wie viele der Publikationen einer Person, aus einer Institution etc. sind heute frei zugänglich – unabhängig vom Erscheinungsjahr? Dies ist jedoch immer eine Momentaufnahme, die schwer vergleichbar ist.

⁵ <http://doaj.org>

Expertise, entsprechende Zeitschriften und Artikel genau zu erkennen und zu beschreiben.

Es gibt verschiedene Wege zu Open Access, und es gibt mehr als Zeitschriftenartikel

Es gibt verschiedene Publikationskulturen in den Disziplinen. Dennoch gehe ich davon aus, dass Zeitschriftenpublikationen in allen Disziplinen zukünftig eine Rolle spielen werden. Dies gilt auch für Open Access: Genausowenig, wie Forschungsförderer Open-Access-Ausnahmen für Disziplinen vorsehen, werden Open-Access-Kennzahlen nur für einzelne Open-Access-affine Disziplinen berechnet werden. Die Forderung, wissenschaftliche Ergebnisse frei zugänglich zu machen, richtet sich an die ganze Wissenschaft.

Publikation in Open-Access-Zeitschriften und die Zweitveröffentlichung von Zeitschriftenartikeln auf Repositorien sind bisher die zahlenmäßig relevanten Open-Access-Wege. Doch auch Monographien, Artikel- und Berichtsveröffentlichungen jenseits von Zeitschriften, Forschungsdaten, multimediale und interaktive Publikationen fallen unter die Open-Access-Forderung, und es sollte beizeiten mitgedacht werden, wie dies gezählt werden kann.

Dies gilt auch für die retrospektive Digitalisierung von Werken, sei es nach zehn oder nach hundert Jahren: Wenn diese frei zugänglich gemacht werden, wird ein weiterer Nutzen möglich, und es sollte Einrichtungen möglich gemacht werden, diesen Beitrag zur Zugänglichkeit zu benennen. Dies sollte jedoch nicht verwechselt werden mit für den Open-Access-Anteil mitgezählten kurzfristigen Zweitveröffentlichungen – einerseits, weil der Zeitverlust bei sehr später Bereitstellung einen Nachteil darstellt und es eben nicht egal ist, ob ein Text nach einem oder nach fünf Jahren genutzt werden kann, andererseits, weil diese Werke nicht mehr ohne weiteres eingehen können in den Open-Access-Anteil ihres Publikationsjahres bei jahresweiser Berechnung (s.o.).

Den Open-Access-Begriff und die Berechnung von Open-Access-Anteilen auf alle Disziplinen und alle Publikationsarten anzuwenden, sollte auch als eine Chance begriffen werden, um verschiedene Open-Access-Modelle zu vergleichen und die Besonderheiten einzelner Felder weiterzuentwickeln, nicht im Sinne einer Vereinheitlichung, aber im Sinne einer Orientierung auf das allgemeine Ziel, einen möglichst großen Anteil an Publikationen so schnell wie möglich und so frei wie möglich zu veröffentlichen.

Daten über Open Access müssen frei sein

Nur dann, wenn die Daten zu Open Access frei sind, können Vergleichbarkeit und Korrektheit gewährleistet und Aggregationen über Institutionen hinweg gefördert werden. Abgesehen von der faktischen Anforderung, durch freie Veröffentlichung die Daten sinnvoll nutzbar zu machen, sollte auch der offenkundige Aberwitz vermieden werden, Open-Access-Informationen nicht im freien Zugriff bereitzustellen.

Die zu Zwecken des Monitorings und des Nachweises von Forschungsleistungen verstärkt implementierten Forschungsinformationssysteme („CRIS“, „FIS“) könnten eine gute Grundlage liefern, schließlich sind sie explizit auf die Listung von Publikationsdaten ausgerichtet. Doch leider sind die entsprechenden Systeme und ihre Implementierungen häufig auf interne Zwecke ausgerichtet und auf das Generieren selbst definierte Berichte. Ein taugliches Messen des Open-Access-Anteils erfordert aber, dass alle zur Berechnung herangezogenen Informationen frei zugänglich und frei nachnutzbar sind. Das bedeutet ausdrücklich auch, dass nachvollziehbar ist, auf welchem Wege eine Publikation frei zugänglich gemacht worden ist bzw. aus welchem Grund sie als *Open Access* gezählt worden ist.

Die Informationen zu Open-Access-Publikationen sollen also dezentral erhoben und frei zur Verfügung gestellt werden, damit sie dann übergreifend (aber nicht exklusiv an einem Ort) ausgewertet und verglichen werden können. Dazu sollten Bibliografien und ergänzende Daten auch mit einem entsprechenden Rechtshinweis versehen werden – in diesem Fall eine Veröffentlichung unter CC0⁶ (siehe Allianz der deutschen Wissenschaftsorganisationen, 2014).

Komplette Bibliografien sind ein notwendiger Schritt

Wissenschaftliche Einrichtungen brauchen Bibliografien, Verzeichnisse ihres Publikationsoutputs. Diese Bibliografien stellen die nötige Datengrundlage für das Messen des Open-Access-Anteils dar. Sie müssen zusammengestellt und frei zugänglich gemacht werden (s.o.). Dies ist insbesondere aus zwei Gründen elementar:

- Auf Basis der Rohdaten (vollständigen Publikationslisten) können berichtete Kennzahlen und Auswertungen überprüft werden.
- Auf dieser Basis können auch neuartige Auswertungen mit angepassten Definitionen, Fragestellungen etc. vorgenommen werden.

⁶ <https://creativecommons.org/publicdomain/zero/1.0/deed.de>

Nicht alle Einrichtungen werden einfach in der Lage sein, komplexe Datenanalysen vorzunehmen. Dies sollte aber nicht davon abhalten, dass andere diese Auswertungen vornehmen können. Doch was die Einrichtungen selbstständig tun können und sollen, ist die Zusammenstellung ihrer Publikationslisten. Publikationsinformationen sind ohnehin verstreut verfügbar. Hierbei treten häufig Probleme auf, insbesondere bei der Zuordnung zu Institutionen, aber auch hinsichtlich der Abdeckung in ansonsten herangezogenen Datenbanken. Durch die Einrichtung gelieferte Publikationsverzeichnisse bieten eine gute Gewähr dafür, die Daten so vollständig wie möglich zu erhalten und andererseits auch nur diejenigen Daten, die eine Einrichtung und ihre Wissenschaftler/innen sich für diesen Zweck zurechnen lassen möchten.

Vereinzelt gibt es Vorbehalte in wissenschaftlichen Einrichtungen, die sich auf den Mehraufwand zur Erstellung von Publikationslisten beziehen. Diese Vorbehalten werden m.E. keinen Bestand haben. Zum einen nehmen Berichtsanforderungen auch von anderen Seiten weiter zu, und die angesprochenen Forschungsinformationssysteme sind eine Reaktion darauf. Wissenschaftliche Einrichtungen müssen sich also ohnehin darauf einstellen, belastbar Auskunft über ihre Forschung zu geben.⁷ Zum anderen ist nicht sehr überzeugend, warum ausgerechnet das Zusammenstellen von Publikationslisten eine besondere Herausforderung darstellen sollte. Die notwendigen Publikationsmetadaten sind ohnehin frei zugänglich, urheberrechtliche Probleme sind ebenso wenig zu befürchten wie Datenschutzprobleme.⁸

Bibliografien sind notwendig, weil sie die Rohdaten zu jeder bibliometrischen Auswertung darstellen, ohne die keine dieser Auswertungen nachvollziehbar ist. Sie sind nur ein erster Schritt, weil Auswertungen, Kategorisierungen etc. noch darauf anzuwenden sind. Aber diese Aufgabe muss – im Hinblick auf den Mehraufwand – und sollte – im Hinblick auf die Vergleichbarkeit – auch nicht in jeder Einrichtung separat vorgenommen werden.

⁷ Hierzu gehören natürlich auch Angaben zum Publikationsaufkommen, also etwa „Wie viele Publikationen entstanden unter Beteiligung der Institutionsangehörigen im Jahr X?“ oder „Wie viele Prozent der Publikationen wurden im Jahr X in Open-Access-Zeitschriften veröffentlicht?“, aber auch detailliertere Aufschlüsselungen nach Verlagen, nach Kosten oder ähnlichem (vgl. Bruch et al., 2016).

⁸ Dies gilt insbesondere dann, wenn es den Autor/innen selbst überlassen ist, ihre Publikationen zu melden.

Bibliotheken im Prozess

Die Open-Access-Transformation stellt Bibliotheken vor große Herausforderungen – mehr vielleicht noch als Verlage oder Wissenschaftler/innen. Doch das hier diskutierte Problem gehört nicht zu den großen Schwierigkeiten. Es geht um den Umgang mit Publikationsmetadaten, ein Feld, das Bibliotheken vertraut ist. Insbesondere dort, wo aus Gründen der Akzeptanzsteigerung Alternativen zur individuellen, händischen Erfassung aller Publikationsdaten gesucht werden, können sich Bibliotheken mit all dem dort versammelten Wissen zu Publikationsdaten und -datenbanken sowie passenden Schnittstellen einbringen. Stärker als die Verwaltungen wissenschaftlicher Einrichtungen können Bibliotheken auf Vergleichbarkeit, Korrektheit, Offenheit bibliografischer Daten hinarbeiten, und sie sollten ihre Rolle im Feld von Forschungsinformationen einfordern.

Bibliotheken werden auch ein eigenes Interesse daran haben, dass Publikationsdaten komplett und offen zur Verfügung stehen. Schließlich erleichtern solche Listen nicht nur die Berechnung des Open-Access-Anteils, sondern sind elementar für viele weitere Schritte, mit denen Bibliotheken im Kontext der Transformation zu Open Access konfrontiert sind: Bedarfsplanung für Open-Access-Finanzierungsangebote und die sukzessive Umschichtung von Erwerbungsmitteln, Identifikation von zweitveröffentlichungsfähigen Publikationen für Repositorien, Langzeitsicherung von Publikationen, Unterstützung neuer bibliometrischer Analysen, Verknüpfungen von Publikationsdaten mit weiteren Informationen.

Insbesondere die finanzielle Seite der Open-Access-Transformation wird von offenen, nachvollziehbaren und nachnutzbaren Informationen zum Open-Access-Anteil profitieren. Wie schnell müssen Erwerbungsmittel auf die Finanzierung von Open Access ausgerichtet werden? Eine wesentliche Unsicherheit bei der Budgetierung für Open Access betrifft die Unklarheit über künftige Publikationen

Fazit

Die Umstellung auf Open Access sollte steuerbar sein durch wissenschaftspolitische Akteure: Hochschulen, Wissenschaftseinrichtungen, Förderer, einzelne Wissenschaftler/innen. Es existieren verschiedene Open-Access-Strategien, und für eine Erfolgsmessung und für Überlegungen zu Kosten, Aufwand, und Effizienz ist eine Messung des Open-Access-Anteils und eine Beobachtung seiner Entwicklung nötig. Dabei geht es nicht um den Ersatz der inhaltlichen Beurteilung von Wissenschaft durch eine (weitere) statistische Kennzahl, sondern um die Beobachtung einer (gleichwohl zentralen) Eigenschaft von Publikationen.

Der Open-Access-Anteil sollte nachvollziehbar angegeben werden können. Dazu braucht es offene Daten und vergleichbare Messungen. Auf dieser Grundlage berechnen wir vielleicht nicht die höchsten Open-Access-Anteile, aber dafür solche, die eine zügige Open-Access-Transformation unterstützen. Denn das Messen des Open-Access-Anteils ist nicht nur

Dokumentation einer Entwicklung, sondern unterfüttert Strategien und Maßnahmen. Und somit führt eine Vereinheitlichung der Datenerhebung und Messung an den zentralen Stellen nicht zu einer Verengung des Blickwinkels, sondern unterstützt im Gegenteil den Vergleich verschiedener Strategien und Ansätze zu Open Access.

Literatur

Allianz der deutschen Wissenschaftsorganisationen (2014). Appell zur Nutzung offener Lizenzen in der Wissenschaft. Available at: http://dfg.de/foerderung/info_wissenschaft/2014/info_wissenschaft_14_68/index.html.

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003). Available at: <https://openaccess.mpg.de/Berliner-Erklaerung>.

Bill & Melinda Gates-Stiftung (2016). Open-Access-Richtlinie der Bill & Melinda Gates-Stiftung. Available at: <http://www.gatesfoundation.org/de/How-We-Work/General-Information/Open-Access-Policy>.

Bruch, C., Geschuhn, K., Hanig, K., Hillenkötter, C., Pampel, H., Schäffler, H., et al. (2016). Empfehlungen zur Open-Access-Transformation: strategische und praktische Verankerung von Open Access in der Informationsversorgung wissenschaftlicher Einrichtungen. <http://doi.org/10.3249/ALLIANZOA.011>.

Danish Ministry of Higher Education and Science (2016). The Danish Open Access Indicator. Available at: <http://ufm.dk/en/research-and-innovation/cooperation-between-research-and-innovation/open-access/Publications/open-access-barometer>.

Danish National Research Database (2016). Available at: <http://www.Forskningsdatabasen.dk/en>.

Fraunhofer-Gesellschaft (2015). Fraunhofer Open Access-Strategie 2020. Available at: <https://www.fraunhofer.de/content/dam/zv/de/publikationen/fraunhofer-open-access-strategie-2020.pdf>.

Global Research Council (2013). Action Plan towards Open Access to Publications. Available at: http://www.dfg.de/download/pdf/dfg_magazin/internationales/130528_grc_annual_meeting/grc_action_plan_open_access.pdf.

Gutknecht, C., Graf, R., Kissling, I., Krämer, D., Milzow, K., Perini, L., et al. (2016). Monitoringbericht SNF. Open Access to Publications 2013-2015. Available at: http://www.snf.ch/SiteCollectionDocuments/Monitoringbericht_Open_Access_2015_d.pdf.

Hunt, M., and Picarra, M. (2016). Open Access Policy Alignment. PASTEUR4OA. Available at: http://pasteur4oa.eu/sites/pasteur4oa/files/resource/Briefing%20paper%20-%20policy%20alignment%20final_0.pdf.

Informationsplattform open-access.net (2016). Was bedeutet Open Access? Available at: <https://open-access.net/informationen-zu-open-access/was-bedeutet-open-access/>.

Senat von Berlin (2015). Open-Access-Strategie für Berlin: wissenschaftliche Publikationen für jedermann zugänglich und nutzbar machen (Drucksache 17/2512). Available at: <http://www.parlament-berlin.de/ados/17/IIIPlen/vorgang/d17-2512.pdf>.

Suber, P. (2012). *Open Access*. MIT Press. Available at: <https://mitpress.mit.edu/books/open-access>.

VSNU (2016). Definition framework monitoring Open Access. Available at: http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/Open%20access/Definitief%20Definition%20framework%20OA_VSNU-20160217.pdf.

Vorstellung des JOIN2 Statistikmoduls mit seinen Differenzen und Problemen zum kommenden Kerndatensatz Forschung

Robert Thiele^{1,3}, Katrin Große^{2,3}

¹Deutsches Elektronen-Synchrotron DESY

²GSI Helmholtzzentrum für Schwerionenforschung Darmstadt

³für JOIN²

JOIN² – Just anOther INvenio INstance

JOIN² ist eine kollaborativ entwickelte Repositorien-Infrastruktur basierend auf der INVENIO-Software vom CERN. Projekt-Partner sind die Bibliotheken des Deutschen Elektronensynchrotrons DESY (Hamburg/Zeuthen), des Deutschen Krebsforschungszentrums DKFZ (Heidelberg), des Forschungszentrums Jülich (Jülich), des GSI Helmholtzzentrums für Schwerionenforschung (Darmstadt), des Heinz Maier-Leibnitz-Zentrum MLZ (Garching) und der Rheinisch Westfälischen Technischen Hochschule RWTH (Aachen).

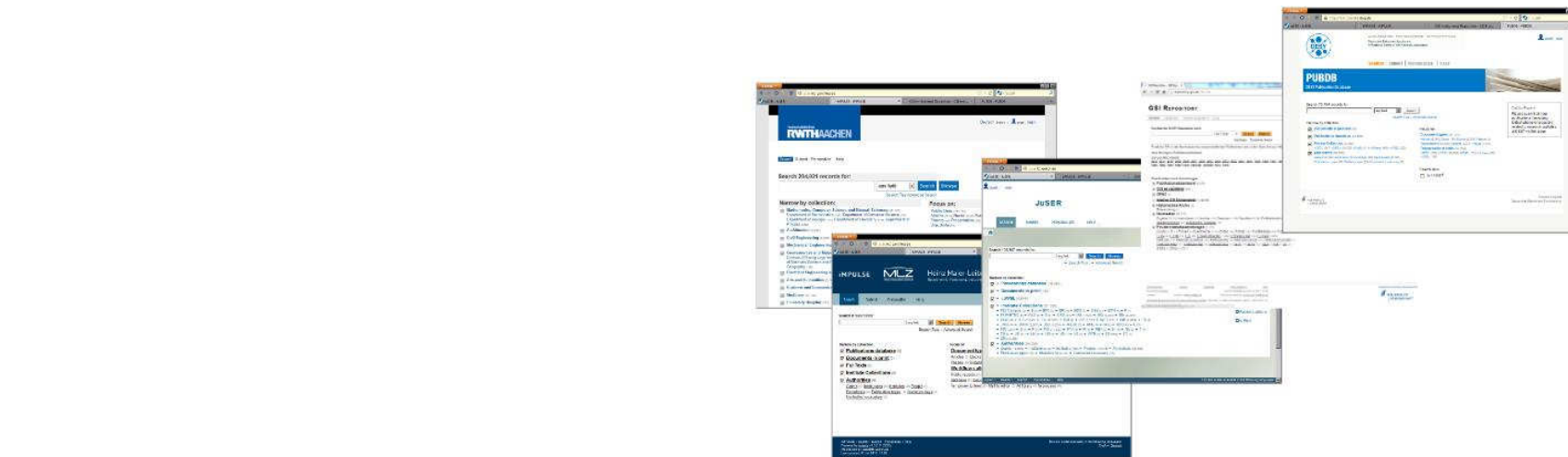


Abbildung 1: Startseite der verschiedenen JOIN² Instanzen

Derzeit laufen beim DESY, bei der GSI, im Forschungszentrum Jülich, dem MLZ, sowie an der RWTH die Systeme live mit derzeit knapp 170.000 gemeinsamen Normdatensätzen. Das DKFZ migriert gerade die Publikationsdaten vom alten System.

Das Deutsche Zentrum für Neurodegenerative Erkrankungen DZNE in Bonn evaluiert derzeit den Eintritt in unser Projekt. Abbildung 1 zeigt typische Startseiten von unseren Repositorien in den verschiedenen JOIN²-Instanzen. Durch die Normdatensätze können zum Beispiel EU-Grants von OpenAIRE <www.openaire.eu> eingesammelt werden. Im Rahmen von HORIZON2020, werden Publikationen, welche durch EU-Gelder gefördert werden, mit dem entsprechenden EU-Grant angereichert und dann automatisch zu OpenAIRE gepusht. Damit wird unseren WissenschaftlerInnen bei der Meldepflicht der Publikationen die Arbeit signifikant erleichtert.

Weitere gemeinsame Normdatensätze sind u.a. Experimente, über 65.000 Zeitschriften-Normdaten, sowie die komplette Programmorientierte Förderung (POF) der Helmholtz-Gemeinschaft (HGF) in allen drei Perioden.

HGFStatistic – Das Statistiktool für die HGF

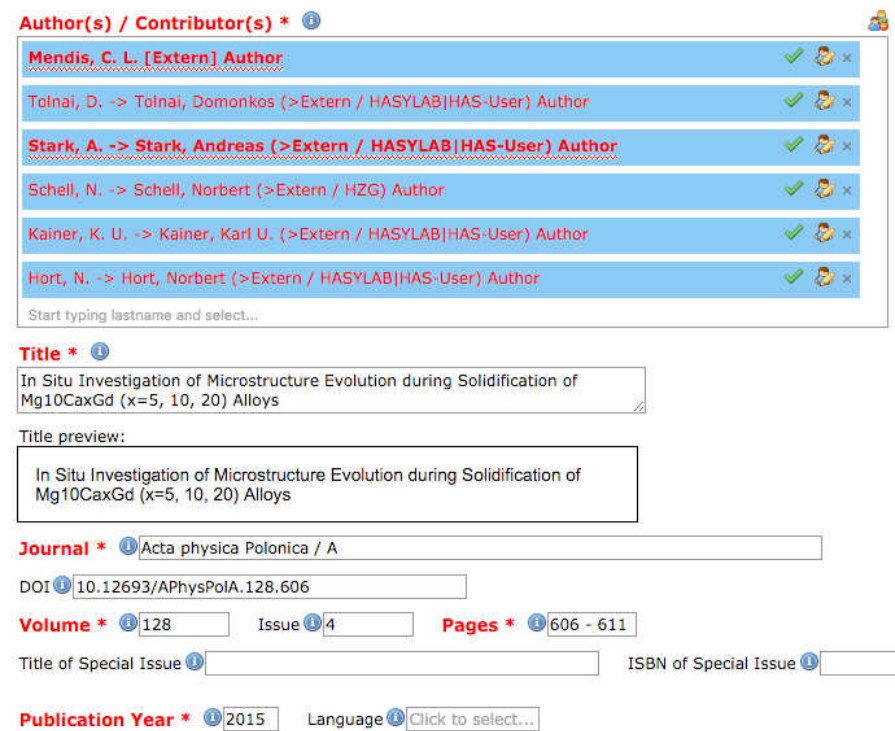
Die Grundlage belastbarer Statistiken sind gute Datensätze. In JOIN² werden Normdatensätze für Zeitschriften, wie auch für Grants zentral verwaltet. Das macht es den Helmholtz-Zentren möglich, ein gemeinsames Tool zur Erfassung der in der Helmholtz-Gemeinschaft benötigten Kennzahlen, wie für den Zentrenfortschrittsbericht zu benutzen. Hierbei werden nach den Richtlinien der POF die erforderlichen Publikationszahlen nach ISI und Scopus entnommen. Im Rahmen der POFIII wurde die Frage, welche Publikationen (Journals) als referiert gelten dürfen, mit der Auflistung in der Thomson Reuter Master Journal List (TRMJL) sowie bei SCOPUS beantwortet. Um diese Information in die Datensätze zu bekommen, werden einmal im Jahr die Zeitschriften-Normdaten aktualisiert und mit den zugehörigen Statistikschlüsseln (TRMJL, SCOPUS, etc.) versehen, sodass beim Eintragen einer neuen Publikation und der Auswahl des Journals die entsprechenden Schlüssel abgespeichert werden können.

```
915_ $$0StatID:(DE-HGF)0100$$2StatID$$aJCR$$bNUCL INSTRUM METH A : 2014
915_ $$0StatID:(DE-HGF)0200$$2StatID$$aDBCcoverage$$bSCOPUS
915_ $$0StatID:(DE-HGF)0300$$2StatID$$aDBCcoverage$$bMedline
915_ $$0StatID:(DE-HGF)0310$$2StatID$$aDBCcoverage$$bNCBI Molecular Biology Database
915_ $$0StatID:(DE-HGF)0199$$2StatID$$aDBCcoverage$$bThomson Reuters Master Journal List
915_ $$0StatID:(DE-HGF)0110$$2StatID$$aWoS$$bScience Citation Index
915_ $$0StatID:(DE-HGF)0150$$2StatID$$aDBCcoverage$$bWeb of Science Core Collection
915_ $$0StatID:(DE-HGF)0111$$2StatID$$aWoS$$bScience Citation Index Expanded
```

Abbildung 2: Beispiel der JOIN² Statistikschlüssel

Ein Beispiel dieser Schlüssel ist in Abbildung 2 zu sehen. Bei JOIN² werden alle Daten im MARC21-Standard abgespeichert. Die Qualität der Daten konnte zusätzlich durch den bei JOIN² angebotenen Digital-Object-Identifier-Import, kurz DOI-Import, erhöht werden, so dass in den Daten deutlich weniger Fehler zu finden sind, als noch vor ein paar Jahren. Zusätzlich wird die Eingabe eines Datensatzes einfacher und geht damit deutlich schneller, da viele notwendige Felder, wie Autor, Titel, Journal, etc. schon befüllt wurden.

In Abbildung 3 ist ein Teil unserer Masken direkt nach einem DOI-Import dargestellt, die Pflichtfelder sind jeweils rot mit Stern gekennzeichnet. Wie man sieht, werden nahezu alle Pflichtfelder befüllt. Die Autoren müssen nach dem Import noch bestätigt werden, damit IDs zugeordnet werden können, nach denen man später auch suchen kann.



Author(s) / Contributor(s) * ⓘ

Mendis, C. L. [Extern] Author	✓	✕
Tolnai, D. -> Tolnai, Domonkos (>Extern / HASYLAB HAS-User) Author	✓	✕
Stark, A. -> Stark, Andreas (>Extern / HASYLAB HAS-User) Author	✓	✕
Schell, N. -> Schell, Norbert (>Extern / HZG) Author	✓	✕
Kainer, K. U. -> Kainer, Karl U. (>Extern / HASYLAB HAS-User) Author	✓	✕
Hort, N. -> Hort, Norbert (>Extern / HASYLAB HAS-User) Author	✓	✕

Start typing lastname and select...

Title * ⓘ

In Situ Investigation of Microstructure Evolution during Solidification of Mg₁₀CaxGd (x=5, 10, 20) Alloys

Title preview:

In Situ Investigation of Microstructure Evolution during Solidification of Mg₁₀CaxGd (x=5, 10, 20) Alloys

Journal * ⓘ Acta physica Polonica / A

DOI ⓘ 10.12693/APhysPolA.128.606

Volume * ⓘ 128 Issue ⓘ 4 **Pages *** ⓘ 606 - 611

Title of Special Issue ⓘ ISBN of Special Issue ⓘ

Publication Year * ⓘ 2015 Language ⓘ Click to select...

Abbildung 3: Typische Eingabemasken bei JOIN²

Nachdem eine Publikation durch den JOIN²-Workflow (USER-EDITOR-STAFF) gegangen ist und von der Bibliothek abschließend bestätigt wurde, ist er freigegeben, hat den Stempel VDB (Veröffentlichungsdatenbank) und erscheint auf unserer dynamischen Statistikseite.

Abbildung 4 zeigt einen Ausschnitt aus dem Gesamtüberblick der Statistik-Webseite. Die Webseite besteht aus einem Kopf mit allen Zahlen der Einrichtung, sortiert nach Statistikschlüssel, sowie nach den Publikationstypen. Danach kommen die POF-Zahlen, wiederum aufgeschlüsselt nach Statistikschlüssel sowie den Publikationstypen. Als drittes können die gleichen Tabellen auch auf Gruppen- bzw. Institutsebene erhoben werden.

865	records	ISI + Scopus (POF3)	(POF3)	EndNote BibTeX RIS
845	records	WOS listed journal OR entry (POF2)	(WOS_UT)	EndNote BibTeX RIS

Abbildung 4: Auszug aus HGFStatistic

Alle Zahlen sind anklickbar und man bekommt sofort die Suche in der Datenbank aufgelöst. Wenn die Treffermenge leer ist und eine Null in der Tabelle steht, wird nur die Zahl angezeigt, ohne Suche. Zusätzlich können die Suchen gegen die Datenbank in eines der drei Ausgabeformate (Endnote, BibTeX, RIS) gebracht werden, so dass man damit gängige Literaturverwaltungsprogramme füttern kann.

Year

At least one internal author?

At least one external author?

POF Period:

Additional Parameters:

Collection:

POF Category:

Abbildung 5: Web Frontend zur Auswahl der gewünschten Daten auf der Statistik-Webseite

Mit einem einfachen Web Frontend, wie in Abbildung 5 zu sehen, kann man die Parameter für die Berechnung und Anzeige einstellen. Hier kann man das Jahr, ob alle Autoren von der Anstalt kommen sollen oder nur externe betrachtet werden sollen, sowie die POF-Periode und die Kategorie eingestellt werden. Danach kann die Berechnung gestartet werden. Somit sind auch Fragen nach älteren Jahren, aber in der aktuellen POF möglich und werden beantwortet, solange die Datensätze die dafür benötigten Informationen tragen.

Differenzen und Probleme beim Übergang des bestehenden Evaluierungstools HGStatistic zum Kerndatensatz Forschung

Nachdem gerade aufgezeigt wurde, welche Kennzahlen wir bereits regelmäßig ermitteln, gehen wir nun im zweiten Teil auf mögliche Realisierungen und Probleme im Hinblick auf den Kerndatensatz Forschung ein. Die „Empfehlungen zur Spezifikation des Kerndatensatz Forschung“ (Wissenschaftsrat, 2016), die auch die Spezifikationen des Kerndatensatzes in der Version 1.0 enthalten, hat der Wissenschaftsrat Anfang dieses Jahres zur Standardisierung der Forschungsberichtserstattung verabschiedet.

Bei unseren Überlegungen beschränken wir uns dabei vor allem auf den Objektbereich Publikationen, der aber auch zahlreiche Verknüpfungen zu allen anderen Objektbereichen aufweist (wie den Objekten Beschäftigte, Nachwuchsförderung, Drittmittel & Finanzen, Patente & Ausgründungen oder den Forschungsinfrastrukturen). Diese Verknüpfungen spiegeln bereits die ersten Schwierigkeiten, denen wir als Bibliotheken gegenüberstehen: Um die funktionierende und nicht gebrochene Verknüpfungen in den Daten vorhalten zu können, müssen wir Zentrums-intern an Standardisierungen und der Etablierung von Identifikatoren arbeiten sowie Mappings und Ontologien aufbauen und pflegen, die wir in der für den Kerndatensatz Forschung gebrauchten Weise noch nicht in hinreichender Form haben. Aus Datenschutzgründen müssen dabei auch Zwischen-Identifikatoren in Betracht gezogen werden, so enthalten die bibliothekarischen Personennorm-datensätze beispielsweise keinesfalls Personalnummern. Die Verknüpfung von internen IDs zu Personenidentifikatoren wie der ORCID, der Scopus-ID, der GND-ID oder auch Fachcommunity-spezifischen Identifikatoren wie der PMID-ID, der JaCOW-ID oder der INSPIRE-ID muss ebenfalls geleistet werden, was nicht ohne manuelle Nacharbeiten geschehen kann.

Auch müssen zur Integrität der Verknüpfungen neue Automatismen aufgebaut werden, um die Verknüpfung der Identifikatoren zu gewährleisten. Neben der Pflege der Verknüpfungen wird für uns als Bibliothek auch die Synchronisierung mit anderen internen Abteilungen deutlich mehr Zeit als bisher in Anspruch nehmen.

Wie Sie bereits gesehen haben, können wir mit unseren bisherigen Evaluierungsabfragen auch Publikationen ermitteln, an denen keine WissenschaftlerInnen unserer Einrichtungen beteiligt waren, die aber an unseren Forschungsinfrastrukturen entstanden sind. Damit kommen wir der Definition, was eine Publikation, Fi14 nach dem Kerndatensatz Forschung ist, bereits nach. Bei Strahlzeitbewilligungen werden die externen ForscherInnen aufgefordert, auf die genutzten Forschungsinfrastrukturen zu verweisen. DOIs für unsere Forschungsinfrastrukturen im Journal of large-scale research facilities (JLSRF) sind in Vorbereitung.

Neben den Publikationen in Listenform werden im Objektbereich von Publikationen die folgenden Ausdifferenzierungen gefordert, die wir nun im Hinblick auf unsere bisherigen JOIN²-Evaluisierungsroutinen darstellen und problematisieren wollen:

- Schöpfer Pu2
- Titel des Werkes Pu5
- Veröffentlichungsjahr Pu84
- Verlag Pu67
- Quelle Pu143 [meint u.a. Zeitschriftentitel, Buchtitel]
- Identifier Pu132
- Format Pu87
- Zugangsrechte Pu19

Bei etwa 90% unserer Veröffentlichungseinträge sind diese Ausdifferenzierungen in der Regel automatisiert über den bereits vorgestellten DOI-Import aus den Verlagsangaben oder anderen ID-Importen (wie PMID oder INSPIRE-ID) aus Datenbanken gefüllt. Bei etwa weiteren 8% der Einträge können wir die von den WissenschaftlerInnen eingetragenen Ausdifferenzierungen im Rahmen unserer Bibliotheksfreigabe recherchieren und ggfs. ändern. Bei den restlichen 2% müssen wir auf die Nutzerangaben vertrauen.

Selbst bei diesen „einfachen“ Angaben, die im Prinzip vollständig abdeckt sind, zeigen sich bei den verschiedenen Ausprägungen erste Probleme in der Praxis. Zum Beispiel beim Schöpfer: Die Eingabe von Herausgebern, von Körperschaften mit Herausgeberfunktionen oder von Gruppen von Herausgebern wird bei Eintragenden vermutlich auf wenig Resonanz stoßen. Und eine Nachrecherche solcher Angaben, insbesondere im Zeitschriftenbereich, werden wir als Bibliothek nicht leisten können. Bei Veröffentlichungen von mehr als tausend AutorInnen ist es nicht möglich, die Vornamen aller AutorInnen nachzuerfassen; über die Normdatenverknüpfungen erhalten wir nur die Vornamen bei unseren eigenen AutorInnen.

Auch scheinbar einfache Identifikatoren wie ORCID oder das Publikationsjahr haben in der Praxis ihre Tücken. Immer wieder führen wir im Alltag mit den WissenschaftlerInnen Diskussionen, wann etwas als „erschieden“ gilt, weisen online-first aus und pflegen das Publikationsjahr nach. Die Verbreitung von ORCID ist in unseren Zentren gänzlich unterschiedlich und zeigt sich damit in der Praxis als unterschiedlich sinnvoll. Nur das Forschungszentrum Jülich strebt zurzeit von unseren Zentren an, für jeder Wissenschaftler/jede Wissenschaftlerin eine ORCID zu vergeben und unterstützende Prozesse dazu anzubieten. Auch basieren die ORCID-ID-Zuordnungen in der Regel auf den Personenzuordnungen unserer WissenschaftlerInnen. Wenn dort Fehler bei der Zuordnung von Einzelpersonen gemacht werden, was bei mehreren Tausend AutorInnen für eine einzige Veröffentlichung in der Praxis durchaus vorkommt (z.B. bei Veröffentlichungen der ALICE Collaboration <<https://repository.gsi.de/record/50756>>), sind damit auch die Abbildungen zu den ORCIDs falsch. Bei den Autorenzusordnungen erfassen wir nur die eigenen WissenschaftlerInnen; niemand wird uns die Affiliationen anderer Einrichtungen eintragen wollen. Die Fach-ID, die sich im DFG Research Explorer zu unseren Einrichtungen findet, gibt unsere interdisziplinären Forschungsfeldern nicht wieder, so dass wir vermutlich, wir bisher, die Helmholtz-Systematik der POF als Forschungsfeld Pu141 weiterhin melden werden. Der Import in die Veröffentlichungsdatenbank erfolgt, wie bereits beschrieben, in der Regel über den DOI-Import und auch der Doublettencheck erfolgt bei JOIN² über die DOI, sowie andere über weiterer Identifikatoren, wie zum Beispiel die arXiv-ID. Daher haben wir in fast allen Journal-Datensätzen eine DOI als Identifikator.

Veröffentlichungen ohne DOI werden dabei nicht nur ungern per Hand eingetragen, sie werden auch oftmals von den Eintragenden nicht als „richtige“ Veröffentlichung betrachtet, da sie in der Regel nicht den Evaluierungskriterien von Helmholtz-reviewed entsprechen. Unsere eigenen Instituts-/Abteilungs-ID ist in allen Datensätzen enthalten. Wir erfassen keine fremden Institute und die Institut-ID unserer Einrichtungen beim DFG-Research Explorers spiegelt i.d.R. nur die Institution wider. In unseren Zentren und Instituten gibt es unterschiedliche Kulturen, was beteiligtes Institut meint und welchen Fällen diese eingetragen werden. Wie oben bereits dargestellt wurde, haben wir ein eigenes gemeinsames Normdatenset für Förderprojekte und –Programme. Einen Umstieg nach Fundref <<http://www.crossref.org/fundingdata/>> ziehen wir bei unseren eigenen Datenanforderungen zurzeit nicht in Betracht, da wir u.a. eine tiefere Erfassung benötigen (z.B. auf Workpackage-Ebene bei Sonderforschungsbereichen). Unter Format versteht der Kerndatensatz Daten, die wir in der Regel bereits automatisiert erhalten können wie Band, Heft oder Seitenbereich. Andere Attribute wie Name der Konferenz werden von unserer WissenschaftlerInnen eingetragen oder auch zwecks Zeitersparnis gerne ausgelassen. Zugangsrechte, sofern sie Open Access, Creative Commons oder die Nationallizenzen widergeben, kommen in der Regel durch den Import mit. Creative Commons können für eigene Verlagsveröffentlichungen von den Bibliotheken nachgetragen werden.

- Sprachcode Pu95

Da unsere Zentren in der Regel in Englisch veröffentlichen, ist dieses Feld für die Eingebenden freiwillig. Die MitarbeiterInnen der Bibliotheken versuchen, fehlende Sprachcodes nachzutragen, bzw. diese automatisiert nachzuerfassen.

- Peer-Reviewed Pu104

Aus dem Bibliothekswesen ist uns keine handhabbare Definition von peer-reviewed mit den Ausprägungen ja-nein bekannt, die auch automatisiert abrufbar ist. Wir können hier, wie bereits vorgestellt, allerdings formelle Ersatz-Definitionen liefern, wie wir sie im Rahmen der Programm-orientierten Förderung über die Journal-Masterlisten von kommerziellen Produkten definieren. Die Definition in der Spezifikation des Kerndatensatzes Forschung ist in der Praxis nicht erklärbar, sowohl für die Bibliothek als auch für die eintragenden WissenschaftlerInnen.

- Qualifikationsschrift Pu146
- Dokumenttyp Pu101
- Publikationstyp Pu6/Pu22-Pu51
- Ressource Pu102
- Förderer Pu86
- Förderkennzeichen Pu90
- Forschungsinfrastrukturen Fi0

Bei den Qualifikationsschriften haben wir ein breiteres Spektrum als die Ausprägungen des Kerndatensatzes (Dissertation und Habilitation). Die Angaben der WissenschaftlerInnen zur Qualifikationsschrift werden von der Bibliothek um URNs ergänzt und im Detail geprüft. Eine gewissenhafte Prüfung des Publikationstyps und des Dokumenttyps, den der Eintragende bei der Eingabe auswählt, erfolgt ebenfalls durch die Bibliothek. Allerdings können wir bestimmte Ausprägungen wie Bibliografie, Editorial, Arbeitspapier, Quellenedition, Wissenschaftliche Vortragsfolien, „Letter to the Editor“ oder Sonderheft einer Zeitschrift nicht erfassen. Da unsere Eintragungen von WissenschaftlerInnen und Verwaltungspersonal gemacht werden, können wir nicht alle Ausprägungen des Kerndatensatzes Forschung als Eingabemaske anbieten oder abfragen. Wir beschränken uns auf die in unseren Einrichtungen typische Ausprägungen des Publikationstyps, die auch für die Eintragenden verständlich sind. Ausprägungen wie ePaper oder Sammelbandbeitrag erfassen wir gar nicht. Über die Wahl der Eingabemasken können wir bestimmte Ressourcen als Bilder, Daten oder Multimedia (Kerndatensatz: Audio, bewegte Bilder) ausweisen. Schwierig wird es bei Journalveröffentlichungen, die auch Videos oder ähnliches enthalten, diese werden wir automatisiert weiterhin nur als Text kennzeichnen können. Förderinformationen und Förderkennzeichen werden von den WissenschaftlerInnen eingetragen – oder je nach Institutskultur auch nicht. Informationen über Forschungsinfrastrukturen können wir vermutlich aus anderen Eintragungen wie Beamlines oder Experimenten automatisiert ableiten. Dies wird allerdings von der Definition der Forschungsinfrastrukturen und derer Detailtiefe abhängen.

Die vom Kerndatensatz geforderten Aggregationsmöglichkeiten nach Fach, Organisationseinheit, Publikationstyp, Dokumenttyp, Schöpfer, Peer-reviewed und Veröffentlichungsjahr werden mit den oben ausgeführten Einschränkungen mit JOIN² möglich sein.

Ob JOIN² als Forschungsinformationssystem die XML-Exporte für den Kerndatensatz abliefern wird oder einen Zentrums-eigenen Forschungsinformationssystem die entsprechenden Daten automatisiert zuliefern wird, wird in unseren Zentren vermutlich unterschiedlich realisiert werden.

Zusammenfassend lässt sich sagen, dass wir durch unsere JOIN²-Projektarchitektur und die konsequente Nutzung von Normdatensätzen in der Lage sind, viele Ausdifferenzierungen des Objektes Publikationen in Kerndatensatz Forschung im Prinzip unter den ausgeführten Einschränkungen erfüllen zu können. Nur wenige Ausdifferenzierungen wie peer-reviewed können wir gar nicht liefern. Mit den JOIN²-Instanzen können unsere Zentren daher der Empfehlung des Wissenschaftsrates nachkommen, über die an unseren Zentren entstandenen Veröffentlichungen auskunftsfähig zu sein (Wissenschaftsrat, 2016, S. 41).

Bei der Umsetzung des Kerndatensatzes Forschung dürfen wir unsere WissenschaftlerInnen nicht aus den Augen verlieren, die die Eingaben und vor allem die Verknüpfungen zu Projekten, Abteilungen und Einzelpersonen vornehmen, und deren Aufwand in vertretbarem Rahmen belieben muss. Kennzahlen sind bei unseren JOIN²-Instanzen lediglich ein Nebenprodukt. Aus Sicht unserer WissenschaftlerInnen und unseren Zentren ist die Nachnutzbarkeit für eigene Listen im Web oder für das Schreiben von Veröffentlichungen die zentrale Aufgabe von JOIN². Diese Nutzungsmöglichkeiten für die WissenschaftlerInnen sollten von einer deutlich stärkeren Differenzierung für den Kerndatensatz Forschung nicht eingeschränkt werden.

Referenz

Wissenschaftsrat : *Empfehlungen zur Spezifikation des Kerndatensatz Forschung*. Berlin, 2016. <<http://www.wissenschaftsrat.de/download/archiv/5066-16.pdf>>

Liste der Autoren

Autorenliste

Franziska Ackermann Kommunikations- und Informationszentrum (kiz) der Universität Ulm Abt. Informationsmedien 89069 Ulm franziska.ackermann@uni-ulm.de	
Dr. Ursula Arning ZB MED - Leibniz-Informationszentrum Lebenswissenschaften Gleueler Str. 60 50931 Koeln arning@zbmed.de	
Dr. Martin Blenkle Staats- und Universitätsbibliothek Bremen Informationstechnik und Digitale Dienste Bibliothekstraße 28359 Bremen blenkle@suub.uni-bremen.de	
Dr. Thomas Dierkes Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV) Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB) Takustr. 7 14195 Berlin-Dahlem dierkes@zib.de http://www.kobv.de	
Michael Ehrismann ETH Zürich, ETH-Bibliothek DigiCenter, IT-Services Rämistrasse 101 8092 Zürich michael.ehrismann@library.ethz.ch	
Rachel Ellis Staats- und Universitätsbibliothek Bremen Informationstechnik und Digitale Dienste Bibliothekstraße 28359 Bremen ellis@suub.uni-bremen.de	

Ralf Flohr

Abteilung Publikationsdienste / EconStor
ZBW – Deutsche Zentralbibliothek für Wirtschaftswissenschaften
Leibniz-Informationszentrum Wirtschaft
Neuer Jungfernstieg 21
20354 Hamburg
r.flohr@zbw.eu

Sandra Geisler

RWTH Aachen University - Informatik 5
Ahornstr. 55
52056 Aachen

Kai Geschuhn

Max Planck Digital Library
Amalienstr. 33
80799 München
geschuhn@mpdl.mpg.de

Katrin Große

GSI Helmholtzzentrum für Schwerionenforschung GmbH
Planckstraße 1
64291 Darmstadt
k.grosse@gsi.de

Dr. Elmar Haake

Staats- und Universitätsbibliothek Bremen
Informationstechnik und Digitale Dienste
Bibliothekstraße
28359 Bremen
haake@suub.uni-bremen.de

Rihan Hai

RWTH Aachen University - Informatik 5
Ahornstr. 55
52056 Aachen

Birte Lindstädt

ZB MED Leibniz-Informationszentrum Lebenswissenschaften
Gleueler Str. 60
50931 Köln
lindstaedt@zbmed.de

Estelle Lorraine

Director COUNTER
 COUNTER, 25 Egbert Road, Winchester, Hampshire SO23 7EB,
 United Kingdom
 lorraine.estelle@counterusage.org

Dr. Bernhard Mittermaier

Forschungszentrum Jülich GmbH – Zentralbibliothek
 Leo-Brandt-Straße
 52428 Jülich
 b.mittermaier@fz-juelich.de

Dr. Thomas Mutschler

Thüringer Universitäts- und Landesbibliothek Jena (ThULB)
 Bibliotheksplatz 2
 07743 Jena
 thomas.mutschler@thulb.uni-jena.de

Dr. Manfred Nölte

Staats- und Universitätsbibliothek Bremen
 Bibliothekstraße
 28359 Bremen
 noelte@suub.uni-bremen.de

Dirk Pieper

Universität Bielefeld
 Postfach 10 01 31
 33501 Bielefeld
 dirk.pieper@uni-bielefeld.de

Annette Polly

Max Rubner-Institut
 Bibliothek Standort Karlsruhe
 Haid-und-Neu-Str. 9
 76131 Karlsruhe
 annette.polly@mri.bund.de

Markus Putnings

Universitätsbibliothek Erlangen-Nürnberg
 Technisch-naturwissenschaftliche Zweigbibliothek
 Erwin-Rommel-Str. 60
 91058 Erlangen

Dr. Christoph Quix

RWTH Aachen University - Informatik 5
Ahornstr. 55
52056 Aachen
quix@dbis.rwth-aachen.de

Dagmar Sitek

Deutsches Krebsforschungszentrum (DKFZ)
in der Helmholtz Gemeinschaft
Stiftung des öffentlichen Rechts
Zentralbibliothek
Im Neuenheimer Feld 280
69120 Heidelberg
www.dkfz.de

Ariane Streicher

Universitätsbibliothek Johann Christian Senckenberg
HeBIS / Verbundzentrale - CIB-Projekt
Bockenheimer Landstr. 134-138
60325 Frankfurt am Main
a.streicher@ub.uni-frankfurt.de

Dr. Robert Thiele

Deutsches Elektronen-Synchrotron DESY
Notkestraße 85
22607 Hamburg

Marco Tullney

Technische Informationsbibliothek (TIB) –
Leibniz-Informationszentrum Technik und Naturwissenschaften
und Universitätsbibliothek
30167 Hannover
marco.tullneytibeu

Regina Wanger

ETH Zürich, ETH-Bibliothek
DigiCenter, Leitung
Rämistrasse 101
8092 Zürich
regina.wanger@library.ethz.ch

Dr Alexander Wagner

Deutsches Elektronen-Synchrotron DESY-Library and Documentation
Notkestr. 85
22607 Hamburg
alexander.wagner@desy.de

Dennis Wehrle

Albert-Ludwigs-Universität Freiburg
Hermann-Herder-Str. 10
79104 Freiburg
dennis.wehrle@rz.uni-freiburg.de

Jan B. Weiland

Abteilung Publikationsdienste / EconStor
ZBW – Deutsche Zentralbibliothek für Wirtschaftswissenschaften
Leibniz-Informationszentrum Wirtschaft
Düsternbrooker Weg 120
24105 Kiel
j.weiland@zbw.eu

**Sponsoren
und
Aussteller**

Wir danken folgenden Sponsoren für Ihre freundliche Unterstützung:

Sponsoren:

Dietmar Dreier Wissenschaftliche Versandbuchhandlung GmbH

EBSCO Information Services GmbH

Elsevier B.V

ImageAccess GmbH

OECD Berlin Centre

SAM - Standards And More -> im Impressum

Wiley

Ausstellende Firmen:

EBSCO Information Services GmbH

Elsevier B.V

ImageAccess GmbH

Lehmanns Media GmbH

Schweitzer Fachinformationen / Goethe + Schweitzer GmbH

Springer Nature

Walter DeGruyter GmbH

Wolters Kluwer / Ovid Technologies GmbH

Band / Volume 18

WissKom 2007: Wissenschaftskommunikation der Zukunft

4. Konferenz der Zentralbibliothek, Forschungszentrum Jülich,
6. – 8. November 2007, Beiträge und Poster
hrsg. von R. Ball (2007), 300 pp
ISBN: 978-3-89336-459-6

Band / Volume 19

**Bibliometrische Verfahren und Methoden als Beitrag zu
Trendbeobachtung und –erkennung in den Naturwissenschaften**

von D. Tunger (2009), 311 pp
ISBN: 978-3-89336-550-0

Band / Volume 20

WissKom 2010: eLibrary – den Wandel gestalten

5. Konferenz der Zentralbibliothek, Forschungszentrum Jülich,
8. – 10. November 2010, Proceedingsband
hrsg. von B. Mittermaier (2010), 389 pp
ISBN: 978-3-89336-668-2

Band / Volume 21

WissKom 2012: Vernetztes Wissen – Daten, Menschen, Systeme

6. Konferenz der Zentralbibliothek, Forschungszentrum Jülich,
5. – 7. November 2012, Proceedingsband
hrsg. von B. Mittermaier (2012), 379 pp
ISBN: 978-3-89336-821-1

Band / Volume 22

**WissKom 2016: Der Schritt zurück als Schritt nach vorn –
Macht der Siegeszug des Open Access Bibliotheken arbeitslos?**

7. Konferenz der Zentralbibliothek, Forschungszentrum Jülich,
14. – 16. Juni 2016, Proceedingsband
hrsg. von B. Mittermaier (2016), ca. 270 pp
ISBN: 978-3-95806-146-0

Der Schritt zurück als Schritt nach vorn – Macht der Siegeszug des Open Access Bibliotheken arbeitslos?

7. Konferenz der Zentralbibliothek, Forschungszentrum Jülich, 14. – 16. Juni 2016

Bernhard Mittermaier (Hrsg.)

Proceedingsband

WissKom
2016

