BFP, Vol. 36, pp. 312-318, Dezember 2012 • Copyright © by Walter de Gruyter • Berlin • Boston. DOI10.1515/bfp-2012-0041

Open-Access-Statistik – Vom Projekt zum Dienst

Daniel Beucke
Niedersächsische Staats- und Universitätsbibliothek Göttingen
Platz der Göttinger Sieben 1
D-37073 Göttingen
E-Mail:
beucke@sub.uni-goettingen.de

Justine Haeberli
Universitätsbibliothek Stuttgart
Holzgartenstraße 16
D-70174 Stuttgart
E-Mail:
justine.haeberli@ub.uni-stuttgart.de

Julika Mimkes
Niedersächsische Staats- und Universitätsbibliothek Göttingen
Platz der Göttinger Sieben 1
D-37073 Göttingen
E-Mail:
mimkes@sub.uni-goettingen.de

Im Rahmen des DFG-geförderten Projekts Open-Access-Statistik werden international vergleichbare Nutzungsstatistiken erzeugt sowie eine dauerhafte Infrastruktur zur Erfassung und Verarbeitung von Nutzungsdaten aufgebaut. Erklärtes Ziel ist es, so die Akzeptanz von Open Access bei Autoren und Rezipienten von wissenschaftlichen Publikationen zu erhöhen. Im Gegensatz zu zitationsbasierten Metriken sind die Nutzungsstatistiken eines digitalen Dokuments eine zeitnahe Messung seiner Relevanz und bilden so dynamische Trends des World Wide Web ab.
Schlüsselwörter: Open Access; Nutzungsstatistiken; Alternative Metriken

Open Access Statistics – from a Project to a Service

The German project Open Access Statistics aims to increase the acceptance of Open Access among authors and readers of scholarly publications by gathering internationally comparable usage statistics, and by providing a lasting infrastructure for the collection and processing of usage data. Usage statistics reflect the level of interest in accessing the content of an individual article and thereby enable item-level, usage-based assessment.
Keywords: Open Access; usage data; altmetrics

1 Motivation

Der Einfluss und die Reputation einer Wissenschaftlerin oder eines Wissenschaftlers im akademischen Diskurs werden durch die Bedeutung seiner Arbeiten und seiner wissenschaftlichen Präsenz gemessen. Die gängige „Währung“, in der sich die Sichtbarkeit wissenschaftlicher Veröffentlichungen bemisst, ist dabei hauptsächlich der auf Zitationen basierende Impact Factor (IF). Je höher der IF einer Zeitschrift ist, in der Autoren publizieren, umso höher ist deren Ansehen in der Wissenschaft. Oft hängen Entscheidungen bei Berufungen oder Evaluationen im wissenschaftlichen Kontext davon ab. Der IF bezieht sich auf die durchschnittliche Anzahl Zitationen einer gesamten Zeitschrift. Dabei wird nicht betrachtet, wie oft einzelne Artikel, sondern wie oft alle Artikel einer Zeitschrift in einer bestimmten Zeitspanne nach Erscheinen der Zeitschrift im Durchschnitt zitiert wurden. Als Grundlage dienen Datenbanken, die von kommerziellen Anbietern gegen Lizenzgebühren zur Verfügung gestellt werden. Der bekannteste aber auch der umstrittenste[1] IF ist der – ehemals vom Institute for Scientific Information (ISI) und heute von Thomson Reuters publizierte – Journal Impact Factor (JIF).

Im Jahr 2005 entwickelte der Argentinier Jorge E. Hirsch ein bibliografisches Maß für das Ansehen von Wissenschaftlerinnen und Wissenschaftlern[2], das auf Grundlage von Zitationen berechnet wird. Bei dem sogenannten Hirschfaktor (oder h-Index) erhält, im Gegensatz zum IF, die Autorin oder der Autor eine Messzahl[3]. Um den h-Index zu berechnen, wird die Anzahl der Publikationen einer Person mit der jeweiligen Anzahl der Zitierungen in Relation gesetzt. So hat eine Autorin oder ein Autor den h-Index von 5, wenn fünf Publikationen jeweils mindestens fünfmal zitiert wurden.

Eine alternative und ergänzende Bemessung des wissenschaftlichen Einflusses, die nicht auf der Auswertung von Zitationen beruht, ist die Auswertung der Nutzerzugriffe auf elektronische Dokumente. Diese Statistiken bilden das Interesse an einer Veröffentlichung ab, sie ermöglichen so eine nutzungsbasierte Bewertung. Nutzungsstatistiken digitaler Objekte können in Echtzeit und mit geringem Aufwand erstellt und unabhängig von ihren Objekteigenschaften wie Publikationsplattform, Materialart, Herkunftsland, Sprache und Themengebiet verglichen werden.

1 Dong, Peng; Loh, Marie; Mondry, Adrian: The „impact factor“ revisited. In: Biomed Digit Libr 2 (1), S. 7. 2005. doi:10.1186/1742-5581-2-7 sowie Seglen, Per O: Why the impact factor of journals should not be used for evaluating research. BMJ. 1997; 314 (7079):498-502.

2 Sidiropoulos, Antonis; Katsaros, Dimitrios; Manolopoulos, Yannis: Generalized h-index for Disclosing Latent Facts in Citation Networks. arXiv:cs/0607066v1.

3 Hirsch, Jorge E.: An index to quantify an individual's scientific research output. In: Proceedings of the National Academy of Sciences of the United States of America. Bd. 102, Nr. 46, 2005, S. 16569–16572. doi:10.1073/pnas.0507655102.

Open-Access-Dokumente unterliegen keinen Zugangsbeschränkungen, so dass sie eine sehr gute Basis für die Erhebung von Nutzungsstatistiken und die Etablierung eines international vergleichbaren und standardisierten Verfahrens bilden. Aus diesem Grund hat die DINI[4]-Arbeitsgruppe „Elektronisches Publizieren“[5] das von der Deutschen Forschungsgemeinschaft (DFG) gefördert Projekt „Open-Access-Statistik“ (OA-Statistik) mit den folgenden Hauptzielen initiiert, auf die in den folgenden Kapiteln ausführlich eingegangen wird:

– Erhöhung der Akzeptanz von Open Access
– International standardisierte Nutzungsstatistiken
– Nachhaltige Service-Infrastruktur

2 Projektdarstellung

Seit 2008 wird im Projekt OA-Statistik ein Dienst entwickelt, der auf die Berechnung und Bereitstellung international vergleichbarer und standardisierter Nutzungsstatistiken für wissenschaftliche Open-Access-Repositorien und die darauf bereitgestellten Publikationen abzielt.

Die erste Projektphase endete im Dezember 2010. Während dieser Laufzeit wurde die Grundlage für eine dauerhafte Infrastruktur zur Erfassung und Verarbeitung von Nutzungsdaten geschaffen. Sowohl die Entwicklung und Etablierung eines einheitlichen Standards zur Ermittlung von Zugriffszahlen als auch das Anbieten von Mehrwertdiensten für Repositorien standen im Fokus der ersten Projektphase. Ergebnisse aus OA-Statistik sind durch die Beteiligung von Projektmitarbeitenden in der DINI-AG „Elektronisches Publizieren“ in das DINI-Zertifikat 2010[6] eingeflossen. Dort wird die Infrastruktur von OA-Statistik explizit als Empfehlung für zertifizierte Repositorien genannt.

Im April 2011 startete die zweite Projektphase, die voraussichtlich im Mai 2013 mit der Überführung des Projekts in einen stabilen Dienst ihren Abschluss finden wird. Grundlage der Überführung des Projekts in einen Dienst stellte die in der Mitte der Projektlaufzeit erstellte Machbarkeitsstudie[7] dar. In der zweiten Projektlaufzeit liegt der Fokus auf der Erweiterung der OAS-Infrastruktur um weitere deutsche Repositorien sowie der performanten Bereitstellung von standardisierten Nutzungsstatistiken.

Projektpartner seit der ersten Förderphase sind die Niedersächsische Staats- und Universitätsbibliothek Göttingen, der Computer und Medien Service der Humboldt-Universität zu Berlin, die Saarländische Universitäts- und Landesbibliothek Saarbrücken sowie die Universitätsbibliothek Stuttgart. In der zweiten Projektphase kam die Verbundzentrale des GBV als weiterer Projektpartner hinzu. Um international abgestimmte Standards für den Austausch und die Berechnung von Nutzungsdaten zu garantieren, kooperieren die Projektpartner mit nationalen und internationalen Partnern.[8]

3 Erhöhung der Akzeptanz von Open Access

Open Access wird als ein Weg aus der Zeitschriftenkrise gesehen. Wie der aktuelle Boykott[9] von Elsevier zeigt, ist der Leidensdruck in der Wissenschaft mittlerweile so groß, dass ganze Zeitschriftenpakete großer Wissenschaftsverlage von Bibliotheken abbestellt [10] werden. Alternative oder ergänzende Publikationswege sind daher sehr gefragt. Während sich der „Goldene Weg“ des Open Access immer mehr durch die Gründung von Open-Access-Verlagen wie PLoS[11] oder Copernicus Publications[12] etabliert, muss der „Grüne Weg“, also die Zweitpublikation von Artikeln auf Repositorien, gestärkt werden. Da die Wissenschaftskommunikation durch das World Wide Web neue Facetten erhalten hat, sollte eine zeitgemäße Metrik auch die dadurch entstandenen Publikationswege berücksichtigen.

Open-Access-Veröffentlichungen ermöglichen durch ihren freien Zugang eine sehr gute Möglichkeit der nutzungsbasierten Impact-Messung. Nachfolgend werden die wesentlichen Unterschiede (siehe Tabelle 1) zwischen nutzerbasierten Metriken und klassischen zitationsbasierten Metriken einander gegenübergestellt und erläutert.

Zitationsbasierte Metriken		**Nutzungsbasierte Metriken**
Autor	Wer?	Leser
zeitversetzt	Wann?	sofort
indexierte Journale	Was?	alle digitalen Objekte
Auswertung auf Journal- oder Autorenebene	Wie?	Auswertung auf Objektebene
etabliert	Warum?	alternative Metriken

Tabelle 1: Unterschiede zwischen zitationsbasierten und nutzungsbasierten Metriken

Bei zitationsbasierten Metriken sind es die Autorinnen und Autoren, die im Rahmen der Wissenschaftskommunikation durch ihr Referenzieren von Publikationen die Metrik bestimmen. Bei den nutzerbasierten Metriken sind es die Leserinnen und Leser, die mit ihren Zugriffen auf Dokumente zeigen, dass diese für sie von Interesse sind. Die Messung des Impacts bei zitationsbasierten Metriken kann nur zeitversetzt stattfinden, da die Zitationen rückwirkend – nach Erscheinen des zitieren-

4 Deutsche Initiative für Netzwerkinformation e.V., http://www.dini.de.
5 http://www.dini.de/ag/e-pub/.
6 http://www.dini.de/dini-zertifikat/.
7 OA-Statistik: Machbarkeitsstudie zum nachhaltigen Betrieb des Dienstes Open-Access-Statistik. 2012. http://www.dini.de/fileadmin/oa-statistik/Machbarkeitsstudie/OAS-Machbarkeitsstudie.pdf.
8 u.a. mit OA-Netzwerk (Deutschland), COUNTER (Großbritannien), ROAT (Japan) und OpenAIRE (Europa).
9 Herb, Ulrich: Mobilmachung gegen Elsevier? Wissenschaftler drohen mit Boykott des Wissenschaftsverlages. Telepolis 2012. http://www.heise.de/tp/blogs/10/151310 und http://thecostofknowledge.com/.
10 z.B. Fakultät für Mathematik der TU München: http://www.ma.tum.de/Mathematik/BibliothekElsevier.
11 http://www.plos.org/.
12 http://publications.copernicus.org/.

den Artikels – gezählt werden und der Impact Factor berechnet wird. Die Messung bei nutzerbasierten Metriken kann hingegen unmittelbar erfolgen. Sobald ein Zugriff stattgefunden hat, wird ein Zähler um eins nach oben gesetzt.

Ein häufiger Kritikpunkt bei den zitationsbasierten Metriken ist, dass nur im Index erfasste Journale berücksichtigt werden. Dies ist bei den nutzerbasierten Metriken nicht gegeben, da hier jedes Dokument, auf das zugegriffen wird, gezählt wird. Die Auswertung erfolgt zudem auf Objektebene. So kann jeweils pro Artikel ein Impact-Wert ermittelt werden. Die zitationsbasierten Metriken sind in ihrer Anwendung bereits bei den Wissenschaftlern etabliert und gehören zum akademischen Alltag. Die nutzerbasierten Metriken befinden sich hingegen im Aufbau und bedürfen noch der Anerkennung in der wissenschaftlichen Community.

3.1 International standardisierte Nutzungsstatistiken

Um den internationalen Vergleich der aggregierten Nutzungsdaten zu ermöglichen, fanden im Rahmen der Projektarbeit in OA-Statistik Abstimmungen mit verschiedenen internationalen Institutionen und Projekten statt. Das Austauschformat und der Datentransfer wurden von Bollen et al.[13] für den Austausch von Nutzungsdaten spezifiziert. Darauf aufbauend hat OA-Statistik als Teil der Knowledge-Exchange-Arbeitsgruppe Usage Statistics[14] Richtlinien[15] zum standardisierten Austausch von Nutzungsdaten auf europäischer Ebene erarbeitet.

Neben dem Austausch muss auch die Aggregierung der Nutzungsdaten anhand international abgestimmter Prozesse erfolgen. Das Projekt Counting Online Usage of Networked Electronic Resources (COUNTER)[16] definiert einheitliche Standards, wie die Nutzung von Informationen gemessen werden kann. Hierzu wurden die COUNTER Codes of Practice[17] entwickelt, die von OA-Statistik als Standard zur Aggregierung der Nutzungsdaten und zur Berechnung der Nutzungsstatistiken eingesetzt werden. Die COUNTER Codes of Practice werden von Bibliothekarinnen und Bibliothekaren eingesetzt, die damit ermitteln können, wie die Informationen, die von Verlagen eingekauft werden, in der Bibliothek genutzt werden. Auch die Anbietenden dieser Informationen profitieren von diesem Standard, indem sie auf konsistente Nutzungsstatistiken zugreifen und diese als Verkaufsargument einsetzen können.

Um die nicht-menschlichen Zugriffe in den Logfiles herauszufiltern, ist eine Robot-Blacklist notwendig. Hier ist es von besonderer Bedeutung, dass diese Liste international abgestimmt bzw. erweitert wird. Nur wenn alle dieselbe Liste verwenden, ist gewährleistet, dass die Ergebnisse vergleichbar bleiben. Ein Ziel von OA-Statistik ist es, am Ende der Projektlaufzeit eine solche Liste zu etablieren.

4 Nachhaltige Service-Infrastruktur

Um einen Dienst wie OA-Statistik langfristig anbieten zu können, sind eine robuste Infrastruktur sowie dessen rechtliche Absicherung essentiell. Auch klar definierte Rollen mit vertraglich geregelten Rechten und Pflichten sind nötig, um einen reibungslosen Betrieb zu gewährleisten. Aus diesem Grund wurde eine Machbarkeitsstudie für den laufenden Betrieb des Dienstes erstellt und eine datenschutzrechtliche Bewertung des Projekts OA-Statistik in Auftrag gegeben. Die wichtigsten Aspekte werden im Folgenden vorgestellt.

4.1 Datenschutz

Die Verarbeitung von Zugriffen auf Repositorien basiert auf der Auswertung von Logfiles, in denen Informationen über das Abfrageereignis gespeichert werden. Neben dem angeforderten Dokument und dem Zugriffserfolg werden auch Informationen über den Zugreifenden wie beispielsweise seine IP-Adresse erfasst. Zugreifender kann eine Suchmaschine sein oder aber eine reale Person, die die abgerufene Publikation lesen möchte. In diesem Fall muss der Datenschutz beachtet werden, der bei der Verarbeitung von personenbezogenen Daten klare Vorschriften macht. Laut einem Gutachten[18] der Zentralen Datenschutzstelle der baden-württembergischen Universitäten (ZENDAS)[19], das vom Projekt OA-Statistik in Auftrag gegeben wurde, sind sowohl die IP-Adressen wie auch die im Zusammenhang mit ihnen erfassten Daten bei der Verarbeitung als personenbezogen anzusehen und daher datenschutzrechtlich zu bewerten[20]. Eine Anonymisierung der IP-Adressen durch ein Verschlüsselungsverfahren wird daher erforderlich. An die IP-Adressen wird ein monatlich wechselnder Wert („Salt“) angehängt. Diese Zeichenkette wird anschließend durch das sogenannte Hashverfahren verschlüsselt. Wenn alle beteiligten Repositorien das gleiche Salt und das gleiche Hashverfahren benutzen, können Zugriffe auf Dokumente mit der gleichen Dokumenten-ID im Anschluss repositorienübergreifend ausgewertet werden. Der zentrale Dienst von OA-Statistik kennt den Salt nicht und darf ihn auch nicht kennen, damit er keine Möglichkeit hat, die Verschlüsselung rückgängig zu machen. Während die Daten bei den Datenlieferanten – den Repositorien – pseudonymisiert sind, können die verschlüsselten IP-Adressen somit nach der Übertragung zum zentralen Dienst von OA-Statistik als anonymisiert angesehen werden. Datenschutzrechtliche Aspekte müssen bei anonymisierten Daten nicht weiter beachtet werden.[21]

13 Bollen, Johan; Van de Sompel, Herbert: An Architecture for the Aggregation and Analysis of Scholarly Usage Data. To be published in Proceedings of the Joint Conference on Digital Libraries, 2006. http://arxiv.org/abs/cs.DL/0605113.

14 http://www.knowledge-exchange.info/Default.aspx?ID=365.

15 Verhaar, Peter et al.: Guidelines for the aggregation and exchange of Usage Data. Surf Wiki 2010. http://wiki.surf.nl/display/standards/KE+Usage+Statistics+Guidelines.

16 http://www.projectcounter.org/.

17 http://www.projectcounter.org/code_practice.html.

18 Stemmer, Bastian: Datenschutzrechtliche Bewertung des Projekts Open-Access-Statistik“. Zentrale Datenschutzstelle des baden-württembergischen Universität (ZENDAS) 2011. http://www.dini.de/fileadmin/oa-statistik/gutachten/ZENDAS_Gutachten_2011.pdf.

19 http://www.zendas.de/.

20 Vgl. Stemmer (Anm. 18) S. 3.

21 Vgl. Stemmer (Anm. 18) S. 25.

4.2 OAS-Infrastruktur

Die Infrastruktur von OA-Statistik besteht aus zwei Komponenten: den OAS-Data-Providern, die bei den teilnehmenden Repositorien installiert sind, sowie dem OAS-Service-Provider als zentralem Dienst von OA-Statistik.

Die Zugriffe von Nutzerinnen und Nutzern auf Dokumente, die auf Repositorien liegen, werden vom Webserver in Form von Logfiles erfasst. Diese Daten werden vom OAS-Data-Provider des Repositoriums aufbereitet. Der OAS-Service-Provider holt die Nutzungsereignisse von den beteiligten Partnern ab, wertet diese standardisiert aus und gibt die berechneten Zugriffsstatistiken an die Repositorien zurück.

Für jedes Repositorium muss ein eigener OAS-Data-Provider installiert werden. Dieser filtert die Zugriffe (Logfiles) vor und pseudonymisiert die IP-Adressen der Nutzerinnen und Nutzer. Aus dem Speicherort des besuchten Dokuments, der im Logfile abgespeichert wird, ermittelt der OAS-Data-Provider zusätzlich eindeutige Dokumenten-IDs. Diese aufbereiteten Daten werden in ein international abgestimmtes XML-Format (OpenURL ContextObjects) überführt. Über eine passwortgeschützte, spezifizierte OAI-PMH-Schnittstelle werden die Daten anschließend für den OAS-Service-Provider bereitgestellt. Die Funktionsweise des OAS-Data-Providers wird in Abbildung 1 veranschaulicht.

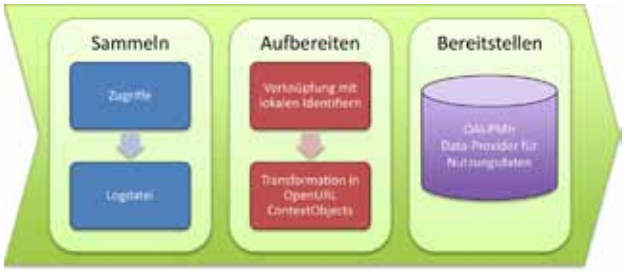

Abb. 1: OAS-Data-Provider

Der OAS-Data-Provider kann entweder im eigenen Repositorium installiert werden oder ein Dienstanbieter übernimmt diese Aufgabe.

Der OAS-Service-Provider holt die aufbereiteten Nutzungsdaten bei den Repositorien ab. Die eingesammelten Daten werden in eine zentrale Datenbank eingelesen und nach den internationalen Standards COUNTER[22], LogEc[23] und den Vorgaben der IFABC[24] ausgewertet.[25] Dabei werden die Zugriffe von Personen von den automatisierten nichtmenschlichen Zugriffen getrennt und ungültige Zugriffe herausgefiltert. Zum Schluss werden die aggregierten Daten den Repositorien im gewünschten Format über zuvor gemeinsam abgestimmte Schnittstellen angeboten. Abbildung 2 fasst die Funktionen des OAS-Service-Providers zusammen.

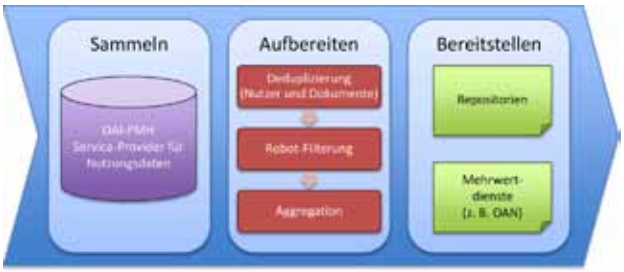

Abb. 2: OAS-Service-Provider

Die OAS-Infrastruktur wurde in der ersten Projektlaufzeit aufgebaut. In der zweiten Projektphase wurden die einzelnen Komponenten optimiert und mittels Monitoringinstrumenten überwacht. In enger Zusammenarbeit mit den Repositorien EconStor[26] und Qucosa[27], die an der Optimierungsphase von OA-Statistik beteiligt sind, wurde die Belastbarkeit und Zuverlässigkeit des Dienstes überprüft. So konnten einige Schwachstellen entdeckt und behoben sowie der Dienst auch in Hinblick auf seine Betriebsgeschwindigkeit optimiert werden. Zudem wurde OA-Statistik auf eine große Anzahl von Repositorien vorbereitet.

4.3 Rückführung der aggregierten Nutzungszahlen in die Repositorien

Die aggregierten Nutzungszahlen der beteiligten Repositorien werden vom OAS-Service-Provider über verschiedene Schnittstellen angeboten. Unterschiedliche Austauschformate werden in Absprache mit den beteiligten Repositorien implementiert, wie beispielsweise die Ausgabe als Textdatei im JSON-Format auf Wunsch eines Repositorien-Betreibers erstellt wurde. Zeilenweise werden in diesen Dateien die Zugriffe pro Dokumenten-ID aufgelistet. Für jeden Tag gibt eine Zeile Auskunft über die standardisierten Zugriffe von Menschen sowie von Suchmaschinen auf ein Dokument. Als Exportformate für die Rückführung der Statistiken liegen zum jetzigen Zeitpunkt JSON, XML und eine grafische Darstellung (siehe Abbildung 3) zur Einbindung als iFrame vor. Eine Sushi-Schnittstelle wird bis zum Ende der Projektlaufzeit ebenfalls implementiert werden.

4.4 Teilnahme an der Infrastruktur

Repositorien, die ihren Nutzerinnen und Nutzern international standardisierte Zugriffszahlen anbieten möchten, können sich als OAS-Data-Provider an der OAS-Infrastruktur beteiligen. Auf der Website des Projekts OA-Statistik sind alle dafür notwendigen Informationen zusammengestellt. [28] Dazu gehören eine Beschrei-

22 http://www.projectcounter.org/.
23 http://logec.repec.org/.
24 http://www.ifabc.org/.
25 Die von anderen Projekten wie COUNTER, LogEc und IFABC genutzten Standards werden von OA-Statistik auf ihre Anwendung hin überprüft mit dem Ziel, aussagekräftige Ergebnisse über die funktionale Anwendbarkeit dieser Standards zu erhalten und daraus ggf. nachhaltige und international vergleichbare Standards zu erzeugen.

26 http://www.econstor.eu/.
27 http://www.qucosa.de/.
28 http://www.dini.de/projekte/oa-statistik/die-technik/registrierung/.

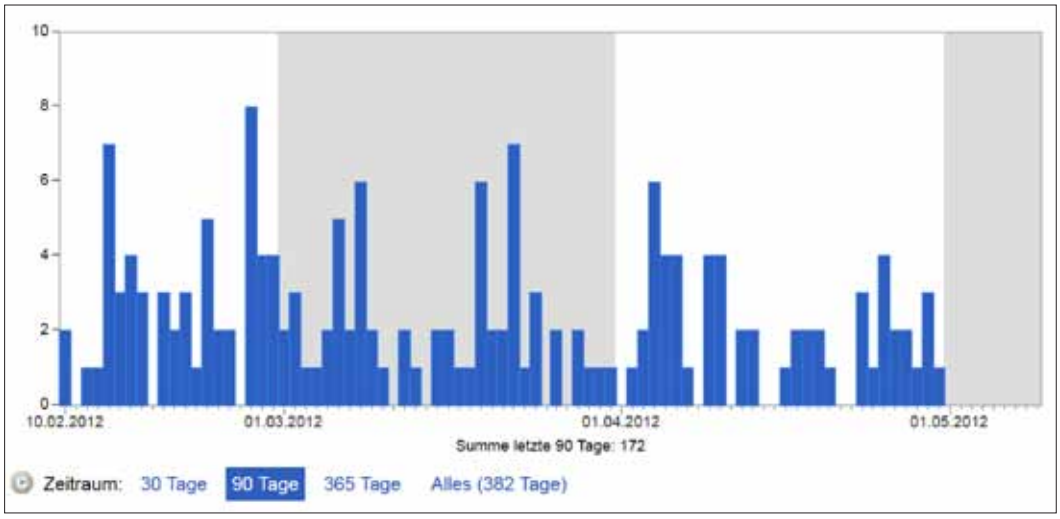

Abb. 3: Grafisch aufbereitete Nutzungsstatistiken

bung des Workflows, die Software des OAS-Data-Providers, der Installationsleitfaden, der XML-Validator sowie ein Registrierungsformular für neue OAS-Data-Provider.

Der OAS-Workflow beschreibt die Infrastruktur von OA-Statistik Schritt für Schritt, wobei das Zusammenspiel von OAS-Service-Provider und OAS-Data-Provider detailliert erklärt wird. Die Software des OAS-Data-Providers steht in der neuesten Version für verschiedene Repositorien-Plattformen zur Verfügung. Sie muss auf die lokalen Gegebenheiten angepasst und konfiguriert werden. In acht Schritten beleuchtet der OAS-Leitfaden alle Teilaspekte der Installation. Als Referenz fungiert eine Demoinstallation des OAS-Data-Providers. Auch werden Beispiele für die Rückführung der aggregierten Nutzungsstatistiken in den Formaten XML und JSON sowie ein Bespiel für die grafische Einbindung mittels iFrame bereitgestellt.

Nach der Installation eines neuen OAS-Data-Providers kann das von der Schnittstelle ausgegebene Datenformat durch den XML-Validator überprüft werden. Ist das XML valide, kann der OAS-Data-Provider über das Registrierungsformular angemeldet werden. Daraufhin kontaktiert OA-Statistik das jeweilige Repositorium, um das Rückgabeformat der aggregierten Daten und den Austausch von Passwörtern zu klären.

Eine Zusammenstellung von häufig gestellten Fragen und Antworten sind in den FAQs zu finden. Sofern Fragen nicht durch die FAQs beantwortet werden können, steht der Support von OA-Statistik zur Klärung der offenen Punkte zur Verfügung.

4.5 Kosten

Nach Ablauf der Projektförderung wird die VZG den Dienst übernehmen und dauerhaft anbieten. Damit verbunden sind zum einen der technische Betrieb des Dienstes auf Servern der VZG und zum anderen die gesamte finanzielle und vertragliche Abwicklung.

Durch Wartung, Anpassung und das dynamische Wachsen des Dienstes entstehen Kosten, die nach Ablauf der Projektlaufzeit von den beteiligten Repositorien getragen werden müssen. Eine Evaluation der zu erwartenden Kosten durch die VZG hat folgendes Kostenmodell ergeben:[29]

- Die laufenden Kosten pro Repositorium werden im ersten Jahr 1.200 Euro betragen. Dieser Betrag setzt sich dabei aus dem Support bei der lokalen Installation des OAS-Data-Providers, der Anbindung an den OAS-Service-Provider und den anfallenden Betriebskosten zusammen.[30]
- Die laufenden Kosten pro Repositorium und Jahr werden sich ab dem zweiten Jahr auf 375 Euro summieren. Dieser Betrag setzt sich aus den Betriebskosten[31] des Servers sowie der Pflege der Software zusammen.

Nutzt ein Repositorium statt einer lokalen Installation eines OAS-Data-Providers die Möglichkeit, den OAS-Data-Provider von einem Dritten Anbieter hosten zu lassen, so fallen zusätzlich zu den zuvor genannten Beträgen weitere Kosten für die Konfiguration und die Betreuung des OAS-Data-Providers an:[32]

- Die laufenden Kosten betragen pro Repositorium im ersten Jahr 600 Euro, wenn es sich um die Auswertung eines Standard-Webservers handelt. Alle anderen auszuwertenden Formate berechnen sich nach Aufwand.
- Die laufenden Kosten pro Repositorium und Jahr addieren sich ab dem zweiten Jahr auf 300 Euro. Dieser Betrag setzt sich zusammen aus den Betriebskosten eines Servers und der Softwarepflege.

29 Die Kostenberechnung bezieht sich auf zehn beteiligte Repositorien. Die exakte Höhe der Kosten kann erst bei der Übernahme des Dienstes durch die VZG genannt werden.

30 Für alle in der Projektlaufzeit beteiligten OAS-Data-Provider entfallen diese Kosten.

31 Die Betriebskosten decken dabei ab: Hardware, Strom, Klimatisierung, Miete, Netzwerkinfrastruktur, Sicherung der Daten, ausfallsicheren Betrieb. Die Administration übernimmt die VZG als Eigenleistung.

32 Diese Berechnungen basieren darauf, dass die VZG als Hosting-Anbieter fungiert.

Übersicht über die Kosten für die Teilnahme an OA-Statistik:

	Lokaler OAS-Data-Provider	Extern betreibender OAS-Data-Provider
Erstes Jahr	1200,00 €	1800,00 €
Folgende Jahre	375,00 €	675,00 €

Tabelle 2: Kosten für die Teilnahme an OA-Statistik

Eine neue Kostenevaluation wird nach Projektende stattfinden, um das Finanzierungsmodell auf den dann produktiven Dienst anzupassen.

5 Ausblick

Bis Projektende im Frühjahr 2013 wird die OAS-Infrastruktur sukzessive um weitere Repositorien erweitert. Im Anschluss an die Förderphase wird der Dienst an die VZG übergeben, die OA-Statistik in ihr Serviceangebot mit aufnehmen wird. Dabei wird die VZG durch die DINI-AG „Elektronisches Publizieren“ sowie durch Eigenleistungen der Projektpartner unterstützt.

Abb. 4: Impact-Faktoren wissenschaftlicher Publikationen[33]

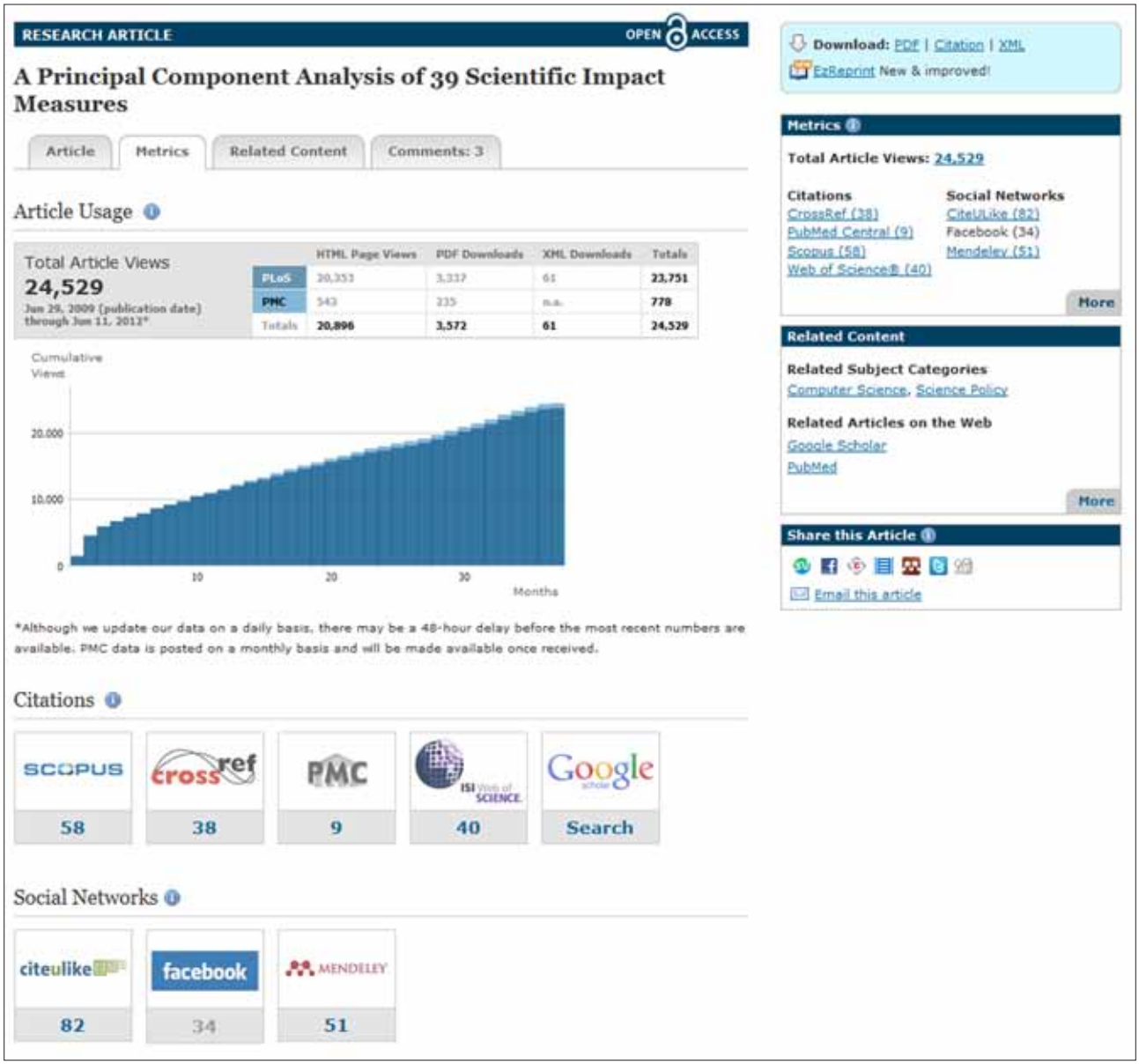

Abb. 5: Metrics bei einer Artikelanzeige in PLoS ONE

33 Nach: Priem et al. (Anm. 33).

Die gesamten Projektergebnisse werden zum Ende der Projektlaufzeit in einer DINI-Schrift veröffentlicht. In dieser werden auch neueste Entwicklungen der bibliometrischen Community beleuchtet. Diese Entwicklungen werden beispielsweise bei Priem et al.[34] beschrieben, die die alternativen Metriken (auch: altmetrics) als eine weitere Säule zur Bemessung des wissenschaftlichen Impacts (siehe Abbildung 4) benennen.

Die Wissenschaftlerinnen und Wissenschaftler verlegen ihre Arbeit immer weiter ins World Wide Web, arbeiten kollaborativ an Publikationen und veröffentlichen ihre Ergebnisse in Weblogs. Dadurch bedingt ändern sich auch die Werkzeuge, mit denen sie arbeiten; bei Literaturverwaltungsprogrammen wie Zotero und Mendeley steigen die Nutzerzahlen, sie verzeichnen eine Vielzahl von Artikeln. Viele Wissenschaftlerinnen und Wissenschaftler teilen ihre Erkenntnisse und Ergebnisse über Twitter mit.[35] Wie können diese Social-Media-Erwähnungen mit in die Bemessung des wissenschaftlichen Impacts aufgenommen werden?[36] Viele Social-Media-Dienste bieten Schnittstellen, über die Bookmarks, Links oder Diskussionen abgefragt werden können, andere Anbieter aggregieren diese Daten und stellen sie aufbereitet zur Verfügung. Der Open-Access-Verlag PLoS ONE[37] nutzt diese Möglichkeit schon länger und gibt für seine Artikel verschiedenen alternative Metriken an (siehe Abbildung 5).

6 Zusammenfassung

Der Dienst OA-Statistik stellt aggregierte und standardisierte Nutzungsdaten von in Repositorien abgelegten Open-Access-Dokumenten zur Verfügung. Auf Grundlage dieser nutzungsbasierten Metriken bietet OA-Statistik ein weiteres Maß zur Impact-Messung an. Dieses ergänzt die zitationsbasierten Metriken, da deren Nachteile bei nutzungsbasierten Metriken nicht zum Tragen kommen. Die nutzungsbasierten Metriken sind unabhängig von Publikationsplattform, Materialart, Herkunftsland, Sprache und Themengebiet weltweit miteinander vergleichbar und ermöglichen somit die transparente und kostengünstige Bewertung der Verbreitung und Nutzung wissenschaftlicher Erkenntnisse. Die Messung kann unmittelbar erfolgen und beschränkt sich nicht auf indexierte Journale.

Nach Ende der Förderphase wird OA-Statistik als dauerhafter und robuster Dienst durch die VZG – unterstützt durch die DINI AG „Elektronisches Publizieren“ sowie durch Eigenleistungen der Projektpartner – betrieben werden. Alle interessierten Repositorienbetreiber können die Projektpartner von OA-Statistik kontaktieren und erhalten nach der Bereitstellung der Nutzungsdaten für geringe Kosten die aggregierten, nach internationalen Standards berechneten Nutzungsstatistiken zurück.

34 Priem, Jason et al.: Altmetrics: A manifesto. V. 1.0 2010. http://altmetrics.org/manifesto.

35 Vgl. dazu: Barr, Kelli: The Role of altmetrics and Peer Review in the Democratization of Knowledge. 2012. http://altmetrics.org/altmetrics12/barr/.

36 Bar-Ilan, Judit et al.: Beyond citations: Scholars' visibility on the social Web. 2012. arXiv:1205.5611v1.

37 http://www.plosone.org/.