

Machbarkeitsstudie

Februar 2012

zum nachhaltigen Betrieb des Dienstes



Computer- und Medienservice (CMS) der Humboldt-Universität zu Berlin (HU Berlin)
Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB Göttingen)
Saarländische Universitäts- und Landesbibliothek Saarbrücken (SULB Saarbrücken)
Universitätsbibliothek Stuttgart (UB Stuttgart)
und
Verbundzentrale des GBV (VZG)

Auf Initiative und in Abstimmung mit der
Deutschen Initiative für Netzwerkinformation e. V.



Inhaltsverzeichnis

1	Vorwort	3
2	Ausgangslage	4
3	Vom Projekt zum Dienst	6
4	Akteure des Dienstes OA-Statistik	7
5	Aspekte der Nachhaltigkeit: Anforderungen und Lösungswege	8
5.1	OAS-Data-Provider	8
5.2	OAS-Service-Provider	13
5.3	Finanzierungsmodell	15
6	Fazit und Ausblick	17
7	Anhänge	18

1 Vorwort

Seit 2008 wird im Projekt Open-Access-Statistik (OA-Statistik) ein Dienst entwickelt, der auf die Erzeugung und Bereitstellung vergleichbarer und standardisierter Nutzungsstatistiken für wissenschaftliche Open-Access-Repositorien und den darauf bereitgestellten Publikationen abzielt. Gemeinsam mit anderen von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projekten zur Verbesserung der Open-Access-Infrastruktur in Deutschland wurde dieses Vorhaben von der Arbeitsgruppe „Elektronisches Publizieren“ der Deutschen Initiative für Netzwerkinformation e. V. (DINI) initiiert und koordiniert, weil sie darin einen wesentlichen Beitrag zur Etablierung alternativer Publikationsformen im wissenschaftlichen Kommunikationsprozess sieht. Das Vorhaben steht im Einklang mit der übergreifenden Zielstellung der Arbeitsgruppe, das elektronische Publizieren an wissenschaftlichen Einrichtungen in Deutschland zu unterstützen und konzeptionell weiterzuentwickeln.

Mit dem DINI-Zertifikat für Dokumenten- und Publikationsservices¹ leistet die Arbeitsgruppe einen wichtigen Beitrag zur Harmonisierung und abgestimmten Entwicklung technischer, organisatorischer, rechtlicher und konzeptioneller Fragen des elektronischen Publizierens. In der dritten Fassung von 2010 haben unter anderem auch Ergebnisse des Projekts OA-Statistik ihren Niederschlag gefunden.

Im Bereich des Managements digitaler Publikationen ist ein standardisierter Austausch von Nutzungsdaten aus wissenschaftlichen Repositorien und kommerzieller Verlagsplattformen unabdingbar. Auf nachvollziehbaren und einheitlichen Kriterien basierende Nutzungsstatistiken auf der Ebene einzelner Dokumente bilden nicht nur einen sinnvollen Bewertungsmaßstab innerhalb der Open-Access-Welt. Sie können als Indikator auch die traditionellen auf Zitationen basierenden Metriken ergänzen und die Rezeption wissenschaftlicher Arbeiten auf einem niedrigschwelligeren Niveau kennzeichnen.

Damit dies auch nachhaltig gelingen kann, muss der konzeptionell und technisch inzwischen weit entwickelte und sich auf den Repositorienplattformen der beteiligten Projektpartner in der Erprobung befindende Dienst in eine dauerhafte Organisationsstruktur überführt werden, die den rechtlichen Erfordernissen gerecht wird. Sie muss zum einen in der Lage sein, den Dienst organisatorisch und finanziell abzusichern. Zum anderen besteht die wesentliche Herausforderung darin, die Anzahl der an OA-Statistik teilnehmenden Repositorien deutlich zu erhöhen und den Dienst technisch und konzeptionell auf dem aktuellen Stand zu halten. Dafür legen die Projektbeteiligten mit dieser Machbarkeitsstudie eine sehr gute Grundlage vor.

Die DINI-Arbeitsgruppe „Elektronisches Publizieren“ befürwortet den eingeschlagenen Weg – insbesondere vor dem Hintergrund der Diskussion um die Neuausrichtung überregionaler Informationsservices. Die Arbeitsgruppe ist bereit, die Etablierung von OA-Statistik als Dienst für wissenschaftliche Repositorien im Rahmen ihrer Möglichkeiten zu unterstützen und den Prozess hin zu einer nachhaltigen Infrastruktur weiterhin aktiv zu begleiten. Dies erfolgt durch die Vernetzung mit Institutionen und Projekten innerhalb und außerhalb Deutschlands, die für OA-Statistik, dessen Weiterentwicklung und Verbreitung relevant sind, im Rahmen von DINI-Veranstaltungen sowie durch die Erstellung von Empfehlungen und Richtlinien – wie beispielsweise das DINI-Zertifikat.

DINI-Arbeitsgruppe „Elektronisches Publizieren“, Januar 2012

¹ <http://www.dini.de/dini-zertifikat/>.

2 Ausgangslage

Ein hoher Journal Impact Factor (JIF) gilt als Zeichen für die Qualität einer Zeitschrift und der in ihr erscheinenden Artikel. Diese Art der Messung von Zitationshäufigkeiten umfasst jedoch nicht alle wissenschaftlichen Zeitschriften, einige Dokumentengattungen werden komplett ausgeschlossen. Zudem erfolgt die Berechnung nicht für den einzelnen Artikel, sondern jeweils für die gesamte Zeitschrift. Eine Alternative zum JIF sind Nutzungsstatistiken. Sie bilden das Interesse am Zugang zum Inhalt einer Veröffentlichung ab und 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. Aufbauend auf den Nutzungsmustern lassen sich zudem Empfehlungsdienste und Relevanzkriterien erstellen. 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. Als Publikationsform ermöglicht Open Access somit die transparente und kostengünstige Bewertung der Verbreitung und Nutzung wissenschaftlicher Erkenntnisse.

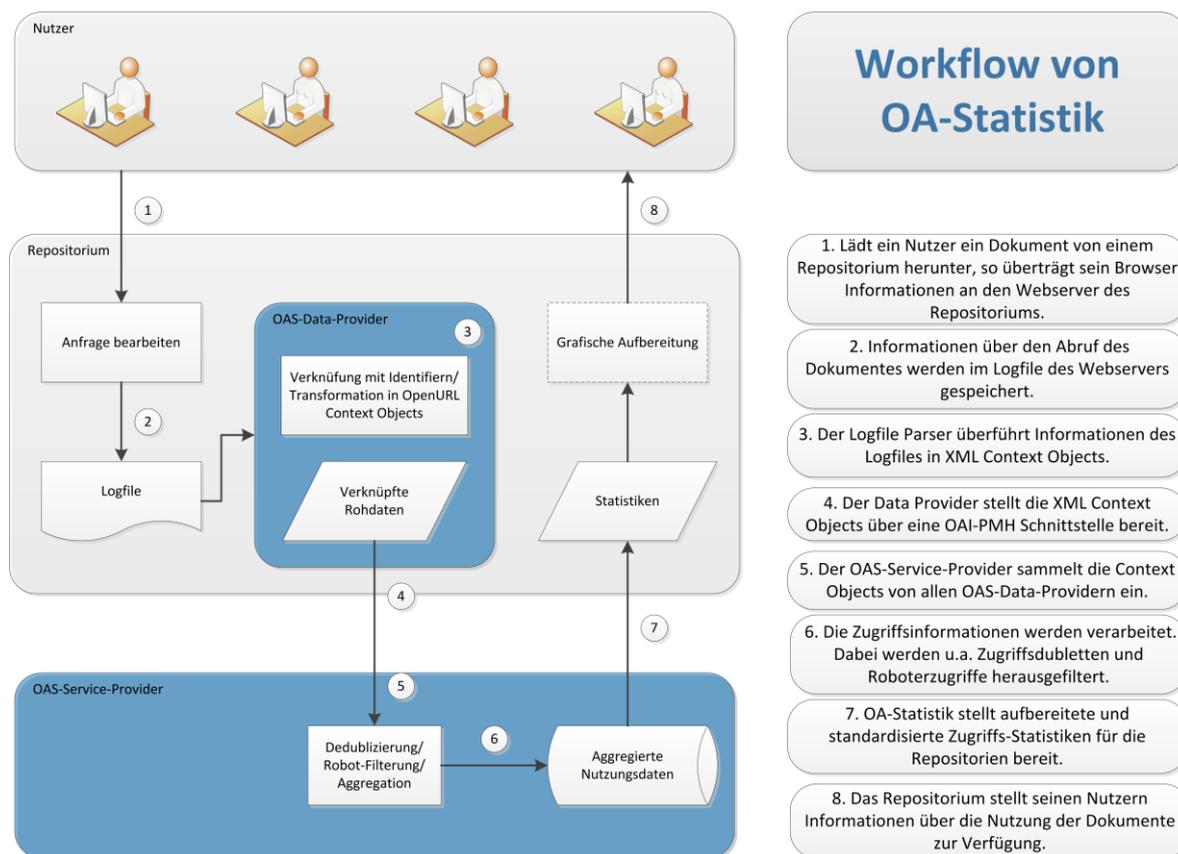
Diese Gedanken greift Open-Access-Statistik (OA-Statistik)² auf, um durch die Gewinnung international vergleichbarer Nutzungsstatistiken und durch das Angebot einer dauerhaften Infrastruktur zur Erfassung und Verarbeitung von Nutzungsdaten die Akzeptanz von Open Access bei Autoren und Rezipienten von wissenschaftlichen Publikationen zu erhöhen. Um Nutzungsstatistiken weltweit miteinander vergleichbar zu machen, hat sich OA-Statistik international abgestimmt und sich auf einheitliche Standards für die Erhebung, den Austausch und die Auswertung von Nutzungsdaten geeinigt. Gemeinsam mit der Knowledge-Exchange-Arbeitsgruppe Usage Statistics³ hat OA-Statistik Richtlinien zum standardisierten Austausch von Nutzungsdaten auf europäischer Ebene erarbeitet. Weiterhin tauscht sich OA-Statistik auf organisatorischer und technischer Ebene u. a. mit dem britischen Projekt PIRUS⁴ und dem japanischen Projekt ROAT aus, um international abgestimmt vorzugehen.

Die **Technik des entwickelten Dienstes** besteht aus zwei Komponenten. Ein OA-Statistik-Data-Provider (OAS-Data-Provider) wird lokal bei den beteiligten Repositorien betrieben, welcher die vom Webserver in Form von Logfiles protokollierten Nutzungsereignisse sammelt, aufbereitet und sie über eine standardisierte Schnittstelle bereitstellt. Alternativ besteht die Möglichkeit, dass sich Repositorien am zentralen OAS-Data-Provider beteiligen und sich so die Installation und den Betrieb eines eigenen OAS-Data-Provider ersparen. Auf der Seite des Diensteanbieters wird der OA-Statistik-Service-Provider (OAS-Service-Provider) eingesetzt, der die Daten von den OAS-Data-Providern einsammelt, aggregiert und die standardisierten Nutzungsdaten anbietet. OA-Statistik bietet bestehenden und neu hinzukommenden OAS-Data-Providern u. a. Unterstützung durch Guidelines und Workshops an. Den reibungslosen Betrieb des OAS-Service-Providers gewährleistet die stetige Anpassung der Hard- und Software an neue Anforderungen.

² <http://www.dini.de/projekte/oa-statistik/>.

³ <http://www.knowledge-exchange.info/Default.aspx?ID=365>.

⁴ <http://www.cranfieldlibrary.cranfield.ac.uk/pirus2/tiki-index.php>.



Die **OAS-Data-Provider** erfassen die Dokumentennutzung in Logfiles und pseudonymisieren die Nutzerinformationen (z. B. IP-Adressen). Sie verarbeiten diese Nutzungsinformationen, wobei eindeutige Dokumenten-IDs hinzugefügt und die Daten in OpenURL ContextObjects überführt werden. Schließlich werden diese Informationen passwortgeschützt mittels OAI-PMH zum Abruf bereitgestellt. Die Installation des OAS-Data-Providers erfolgt in wenigen Schritten über die von OA-Statistik bereitgestellten Softwarepakete, die auf die gängigsten Repositoriensysteme zugeschnitten sind.

Der Abruf und die Aufbereitung der erfassten Daten erfolgt über den **OAS-Service-Provider**. Dieser geht nach den Standards COUNTER⁵, LogEc⁶ und den Vorgaben der IFABC⁷ vor.⁸ Die Daten werden von Verzerrungen durch automatisierte nichtmenschliche Zugriffe bereinigt. Perspektivisch sollen zusätzlich Dokumentendubletten – Zugriffe auf das gleiche Dokument an unterschiedlichen Standorten – erkannt und zusammengeführt werden. Das Ergebnis, die standardisierten Nutzungszahlen, werden an die Repositorien zurückgespielt und können dort angezeigt oder z. B. zur Sortierung einer Suchtrefferliste verwendet werden.

Bei Treffen und Konferenzen in den vergangenen Jahren zeigte sich, dass die Community großes Interesse an dem Dienst von Open-Access-Statistik hat und die Entwicklungen aufmerksam verfolgt. So stießen auch beim OAS-Workshop im Oktober 2011⁹ sowohl der Dienst OA-Statistik als auch die in diesem Zusammenhang behandelten datenschutzrechtlichen Fragestellungen auf große Resonanz.

⁵ <http://www.projectcounter.org/>.

⁶ <http://logec.repec.org/>.

⁷ <http://www.ifabc.org/>.

⁸ 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.

⁹ <http://www.dini.de/projekte/oa-statistik/die-projektresultate/workshops/oa-statistik-workshop-2011/>.

3 Vom Projekt zum Dienst

In der ersten Förderphase¹⁰ wurde in dem Projekt OA-Statistik eine Infrastruktur zum Austausch und zur Aggregation von Nutzungsdaten testweise unter den Projektpartnern¹¹ aufgebaut. In der zweiten Projektphase¹² wird die Infrastruktur von OA-Statistik um weitere deutsche und internationale Repositorien und Open-Access-Verlage erweitert. Ziel ist es, das Projekt OA-Statistik nach Ablauf der zweiten Förderphase in einen langfristigen, stabilen Dienst überführt zu haben.

Ein solcher Dienst kann nur dann funktionieren, wenn sich alle Projektpartner entsprechend ihrer Fähigkeiten und Möglichkeiten langfristig engagieren. Um dieses Engagement zu dokumentieren und langfristig zu sichern, haben die Partner eine gemeinsame Absichtserklärung¹³ verfasst. Darin verpflichten sich die Partner, im Rahmen der DINI AG „Elektronisches Publizieren“ den Kontakt mit internationalen Projekten und Institutionen zu pflegen und neue Kontakte herzustellen, um die lokale und globale Vernetzung des Dienstes OA-Statistik zu gewährleisten. Sie können so die aktuellen Entwicklungen verfolgen, veränderte Anforderungen der Nutzer an den Dienst OA-Statistik erfassen und wenn nötig, Anpassungen an der Infrastruktur oder an der Software des Dienstes OA-Statistik vornehmen lassen. Die Ziele von OA-Statistik werden aus der Arbeitsgruppe heraus mittels Informationsmaterialien, Veröffentlichungen, Vorträgen und Workshops bekannt gemacht und verbreitet.

Um die Anzahl der teilnehmenden Repositorien kontinuierlich zu erhöhen, nutzen die Projektpartner neben den einschlägigen Foren vor allem Publikationen und Fachtagungen, um den Dienst OA-Statistik und dessen Ziele bekannt zu machen. Zudem werden einzelne Repositorienbetreiber über den DINI-Verteiler direkt angeschrieben und zur Teilnahme am Dienst ermuntert. Ein Registrierungsformular sowie ein Leitfaden zur Anbindung eines Repositoriums an den Dienst OA-Statistik ermöglichen eine möglichst einfache Teilnahme. Fragen zur Registrierung werden im CARPET-Forum¹⁴ gesammelt und beantwortet wie auch Fragen, Anmerkungen und Diskussionen zu weiteren, den Dienst OA-Statistik betreffenden Themen.

Der Dienst muss für die Gewährleistung eines dauerhaften Betriebs von einer vertrauenswürdigen und zuverlässigen Institution übernommen werden, welche die technischen und organisatorischen Mittel hat, diesen langfristig anbieten zu können. Eine solche Institution sollte wirtschaftlich stabil und in der Community etabliert sein sowie die notwendige technische und personelle Infrastruktur besitzen. Zudem sollte sie fundierte Kenntnisse im Betrieb ähnlicher Dienste hinsichtlich Geschäftsgang, Datenkonversion und datenschutzgerechter Datenverarbeitung vorweisen können. In der zweiten Förderphase von OA-Statistik konnte – als potentieller Dienstanbieter nach Projektende – die Verbundzentrale des Gemeinsamen Bibliotheksverbundes (VZG) als Projektpartner gewonnen werden. Bei positivem Verlauf der Testphasen des Dienstes OA-Statistik wird die VZG den Dienst in ihr Angebot übernehmen. Die VZG stellt IT-Dienstleistungen und die benötigte Infrastruktur für Bibliotheken zur Verfügung.¹⁵ Sie erfüllt die oben genannten Anforderungen und ist somit bestens geeignet, den dauerhaften und stabilen Betrieb des Dienstes OA-Statistik zu gewährleisten.

¹⁰ Laufzeit: Mai 2008 - Dezember 2010 (kostenneutrale Verlängerung).

¹¹ Projektpartner: Die Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB Göttingen), der Computer und Medien Service (CMS) der Humboldt-Universität zu Berlin (HU Berlin), die Saarländische Universitäts- und Landesbibliothek Saarbrücken (SULB Saarbrücken) und die Universitätsbibliothek Stuttgart (UB Stuttgart).

¹² Laufzeit: April 2011 - April 2013.

¹³ Siehe Anhang 3: Memorandum of Understanding.

¹⁴ http://www.carpet-project.net/forum/themen/oa_statistik/.

¹⁵ Weitere Informationen finden sich z. B. unter <http://www.gbv.de>.

4 Akteure des Dienstes OA-Statistik

Die Akteure im Dienst OA-Statistik sind grob in fünf Bereiche aufteilbar. Das im Folgenden skizzierte Rollenmodell ermöglicht einen Überblick über die Rechte und Pflichten der am Dienst OA-Statistik beteiligten Akteure.

Dienstnehmer (Repositorien)

- Stellen ihre Nutzungsdaten über eine definierte Schnittstelle in vorgegebenem Format bereit.
- Entrichten für die Teilnahme an OA-Statistik einen vertraglich geregelten Kostenbeitrag.
- Schließen einen Vertrag zwischen OAS-Service-Provider und OAS-Data-Provider.
- Beachten die datenschutzrechtlichen Vorgaben für die Beteiligung an OA-Statistik.
- Erhalten ihre Daten aggregiert und aufbereitet zurück und können diese laut Lizenzbedingungen verwenden.

Einzelne Nutzer der Repositorien

- Bekommen die Möglichkeit des Opt-Outs zur Nichterfassung ihrer Nutzungsdaten.
- Besitzen ein Informationsrecht darüber, welche ihrer Daten wie, wann, warum etc. erhoben bzw. gespeichert werden.
- Können die Nutzungsstatistiken des Repositoriums einsehen.

Die Rechte der individuellen Nutzer setzt das zuständige Repositorium um. Da der Dienstanbieter nur anonyme Daten bekommt, kann er nur beratend tätig sein.

Dienstanbieter (OA-Statistik bzw. VZG)

- Stellt die benötigte Hard- und Software bereit und pflegt sie.
- Erstellt ein Backup des Dienstes, insbesondere der aggregierten Daten.
- Schließt einen Vertrag zwischen OAS-Service-Provider und OAS-Data-Provider.
- Stellt einen Ansprechpartner für technische und organisatorische Fragen bereit.
- Hat eine Informations- und Transparenzpflicht gegenüber den beteiligten Akteuren.
- Harvestet die Nutzungsdaten der beteiligten OAS-Data-Provider.
- Stellt die aggregierten Nutzungsdaten den jeweiligen OAS-Data-Provider über eine zuvor bestimmte Schnittstelle in einem bestimmten Format zur Verfügung.
- Stellt Informationen zu Datenschutzfragen im Bezug auf OA-Statistik bereit.
- Darf die Daten im Rahmen der Lizenzen verwenden.
- Führt die Rechnungsstelle.

Externe Dienste oder Projekte

- Nutzen die Daten laut Lizenzbedingung nach.

DINI AG „Elektronisches Publizieren“

- Hält Kontakt zu internationalen Projekten und Initiativen im Bereich Nutzungsstatistiken.
- Verfolgt aktuelle Entwicklungen (Community-Watch) in der wissenschaftlichen Community.
- Bewirbt den Dienst OA-Statistik.

5 Aspekte der Nachhaltigkeit: Anforderungen und Lösungswege

Für einen nachhaltigen und dauerhaften Betrieb des Dienstes OA-Statistik sind sowohl rechtliche, technische wie auch organisatorische Anforderungen relevant, die im Folgenden aufgezeigt werden. Zu jeder Anforderung werden ein Lösungsweg sowie der aktuelle Stand der Umsetzung beschrieben.

5.1 OAS-Data-Provider

Den Repositorienbetreibern stehen zwei Möglichkeiten zur Verfügung, um sich an OA-Statistik zu beteiligen. Entweder wird der OAS-Data-Provider lokal auf dem jeweiligen Repositorium installiert und betrieben oder das Repositorium beteiligt sich an dem zentralen OAS-Data-Provider.

5.1.1 Lokaler OAS-Data-Provider

Anforderung:

Der OAS-Data-Provider wird lokal bei den beteiligten Repositorien betrieben. Dieser sammelt die Nutzungsdaten, bereitet sie auf und stellt sie über eine standardisierte Schnittstelle in einem vorgegebenen Format zum Abruf bereit. Das Bereitstellen der Nutzungsdaten zwischen OAS-Data-Provider und OAS-Service-Provider muss nach datenschutzrechtlichen Vorgaben erfolgen.

Lösungsweg:

Ruft ein Nutzer auf einem Repositorium ein Dokument auf, so überträgt sein Browser automatisch Informationen wie seinen Useragent (inkl. Browserversion und Betriebssystem), ACCEPT-Header (akzeptierte Dokumententypen und Codierungen), bevorzugte Sprachen, Protokollversion, eigene IP-Adresse und den Namen des angeforderten Dokuments. Diese Nutzungsdaten werden in Logfiles gespeichert. Aus datenschutzrechtlichen Gründen muss vom Repositorium die Möglichkeit eines Opt-Out angeboten werden. Wird diese vom Nutzer gewählt, gelangt seine IP-Adresse nicht in die Logfiles und wird dementsprechend nicht verarbeitet und ausgewertet.

Durch die Installation und die Anpassung des OAS-Data-Providers¹⁶ werden die Logfiles mit dem OAS-Logfile-Parser in OpenURL ContextObjects¹⁷ umgewandelt und im Datenformat XML abgespeichert. Den Informationen aus den Logfiles muss ein Identifier des jeweiligen Dokuments mitgegeben werden. Dies funktioniert für alle Repositorienplattformen und Softwareversionen unterschiedlich und muss bei der Installation im OAS-Logfile-Parser angepasst werden. In den meisten Fällen erfolgt das auf Grundlage der abgerufenen URL. In Deutschland ist die Speicherung von IP-Adressen aus datenschutzrechtlichen Gründen fragwürdig. Daher werden die IP-Adressen von Rechnern, die bei der Dokumentnutzung Einträge in den Logfiles hinterlassen, mit einem Salt versehen und anschließend mit dem Hashverfahren SHA-256 verschlüsselt. Die Erstellung des Hashs dient der Pseudonymisierung und damit der Identifikation der von einer IP-Adresse erfolgten Nutzungsmuster, ohne die IP-Adresse bei der Datenverarbeitung zu verwenden. Der Salt dient als weitere Schutzmaßnahme zur Verhinderung der Rückübersetzung des Hashwertes. Für eine spätere Berücksichtigung von Roboterzugriffen und zur Auswertung der Daten wird zusätzlich auch das C-Klasse-Netz der IP-Adresse (die ersten drei der vier Zahlentripel der IP-Adresse) gesalzen und gehasht. Die Abfrage der Daten erfolgt über eine standardisierte Schnittstelle. Für die Abfrage der Daten ist ein Passwort nötig.

¹⁶ Der OAS-Data-Provider basiert auf dem OAI-Standard. Ein bereits bestehender OAI-Data-Provider kann allerdings nicht genutzt, sondern muss ggf. zusätzlich installiert werden.

¹⁷ Das Datenformat OpenURL ContextObject wurde von OA-Statistik angepasst, mit anderen europäischen Projekten abgestimmt und dokumentiert. Es ist komplexer als das ursprüngliche Minimalset, ist aber vollständig kompatibel.

Stand der Umsetzung:

Die ersten sechs OAS-Data-Provider wurden installiert, konfiguriert und in die Infrastruktur von OA-Statistik eingebunden. Drei der OAS-Data-Provider gehören zu Repositorien von Projektpartnern und drei zu Repositorien von Externen¹⁸. Zwei der externen OAS-Data-Provider werden kontinuierlich geharvestet. Um die Anbindung verschiedener Repositorienplattformen an OA-Statistik zu ermöglichen, stellt der Dienst Softwarepakete sowie Dokumentationen für die verschiedenen Plattformen zum Download¹⁹ bereit. Die vorliegende Dokumentation des Workflows dient Projektpartnern und beteiligten Repositorien dazu, einen Einblick in die Verarbeitung der Logfileinformationen innerhalb des Projektes zu geben und damit Transparenz zu schaffen. In sieben Schritten wird erklärt, wie aus Zugriffen auf Artikel Statistiken generiert werden.

5.1.2 Datenschutzrechtliche Implikationen für den Dienst OA-Statistik

Anforderung:

Für den nachhaltigen Betrieb des Dienstes zur Erfassung und Aggregation von Nutzungsdaten ist es zwingend erforderlich, den aktuellen Anforderungen des Datenschutzes gerecht zu werden. In der juristischen Fachdiskussion ist man sich nicht abschließend einig, ob IP-Adressen, die für die Berechnungen in OA-Statistik von hoher Wichtigkeit sind, personenbezogene Daten darstellen. Daher muss der gesamte Workflow in OA-Statistik – vom beteiligten OAS-Data-Provider hin zum OAS-Service-Provider – datenschutzrechtlich untersucht und daraufhin angepasst werden.

Lösungsweg:

Das datenschutzrechtliche Vorgehen in OA-Statistik beruht auf einem Gutachten der Zentralen Datenschutzstelle der baden-württembergischen Universitäten²⁰ (ZENDAS). Laut einem Gutachten der ZENDAS²¹, das im Rahmen des Projekts OA-Statistik in Auftrag gegeben wurde, sind die den Nutzern zugewiesenen IP-Adressen – und somit alle in Zusammenhang mit der IP-Adresse erfassten Daten – für die OAS-Data-Provider als personenbezogen anzusehen und demgemäß datenschutzrechtlich zu bewerten.²² Aus diesem Grund müssen IP-Adressen von Rechnern, die bei der Dokumentennutzung Einträge in den Logfiles der Repositorien hinterlassen, innerhalb weniger Minuten nach Erhebung der Daten auf Seiten der OAS-Data-Provider mit einem Salt versehen und einer Hashfunktion verfremdet werden, damit kein Rückschluss auf einzelne Nutzer erfolgen kann.²³ Die OAS-Data-Provider dürfen untereinander keine Daten austauschen, da sie dasselbe Salt verwenden. Das Salt darf wiederum dem OAS-Service-Provider nicht bekannt sein und muss zwingend geheim gehalten werden, damit eine Rückübersetzung nicht erfolgen kann.²⁴ Um datenschutzkonform zu sein, ist es zudem notwendig, den Nutzern jeweils auf Seiten der OAS-Data-Provider eine Möglichkeit des Opt-Outs – der Nichterfassung ihrer Nutzungsdaten – zur Verfügung zu stellen.²⁵ Die datenschutzrechtlichen Anforderungen müssen demzufolge auf Seiten der OAS-Data-

¹⁸ Econstor (ZBW), Qucosa (SLUB Dresden) und Copernicus.

¹⁹ <http://www.dini.de/projekte/oa-statistik/die-projektergebnisse/software-download/>.

²⁰ <http://www.zendas.de>.

²¹ Das ZENDAS-Gutachten wurde der Community im Oktober 2011 im Rahmen der Open-Access-Tage in Regensburg vorgestellt und steht der Öffentlichkeit zur Konsultation auf der Webseite des Projektes zum Download unter http://www.dini.de/fileadmin/oa-statistik/projektergebnisse/ZENDAS_Gutachten_2011.pdf bereit.

²² Vgl. ZENDAS-Gutachten, S. 3.

²³ Vgl. ebenda, S. 21.

²⁴ Vgl. ebenda, S. 25.

²⁵ Vgl. ebenda, S. 18.

Provider erfüllt werden. Der OAS-Service-Provider erhält ausschließlich anonymisierte Nutzungsdaten und muss daher keine Vorschriften zum Datenschutz zu beachten.²⁶

Der Dienst OA-Statistik ist angehalten, den OAS-Data-Providern die für die Einhaltung des Datenschutzes notwendigen Schritte zu erläutern und Vorschläge für deren Umsetzung bereitzustellen. Die Berücksichtigung und Umsetzung der datenschutzrechtlichen Empfehlungen des ZENDAS-Gutachtens gewährleisten eine umfassende rechtliche Absicherung des Dienstes OA-Statistik.

Stand der Umsetzung:

Zusammen mit der ZENDAS wurde das aufwendige Verschlüsselungsverfahren der IP-Adressen entwickelt. Die Kernpunkte des Gutachtens werden von OA-Statistik erfüllt. Noch offene Anforderungen wurden so weit vorbereitet, dass sie, sobald der Dienst aus einem Testbetrieb in einen produktiven Betrieb umstellt, erfüllt werden. Dies betrifft vor allem die zeitnahe Löschung aller Logfiles. Zurzeit werden diese noch gebraucht, um die Verarbeitung der Daten zu überprüfen.

5.1.3 Salt-Lieferant

Anforderung:

Im OAS-Workflow wird auf Seiten des OAS-Data-Providers die IP-Adresse mit einem Salt versehen, bevor diese mit einer Hash-Funktion pseudonymisiert wird. Dabei müssen die Vorgaben des datenschutzrechtlichen Gutachtens der ZENDAS eingehalten werden: Das Salt muss einmal pro Monat geändert werden, die Bereitstellung des Salts darf nicht durch den OAS-Service-Provider (VZG) erfolgen.

Lösungsweg:

Der Salt-Wert muss – monatlich aktualisiert – von einer Institution allen beteiligten OAS-Data-Providern bereitgestellt werden. Dazu ist ein einfaches Skript ausreichend, das einmal im Monat einen Salt-String generiert und über eine passwortgeschützte Schnittstelle den OAS-Data-Providern zur Verfügung stellt.

Stand der Umsetzung:

Die Saarländische Universitäts- und Landesbibliothek Saarbrücken (SULB Saarbrücken) übernimmt die Generierung und Bereitstellung des Salt-Strings auch nach Ablauf der Projektlaufzeit. Das Skript muss noch erstellt und im Anschluss die Schnittstelle implementiert werden.

5.1.4 Gewinnung neuer Teilnehmer für den Dienst OA-Statistik

Anforderung:

Maßgeblich für den Erfolg von OA-Statistik ist die Zahl der teilnehmenden Repositorien. Je höher die Zahl der teilnehmenden Repositorien ist, desto mehr standardisierte Nutzungsdaten werden erfasst und verarbeitet und können so international verglichen werden.

Lösungsweg:

Die Ziele von OA-Statistik werden mittels Informationsmaterialien, Veröffentlichungen, Vorträgen und Workshops präsentiert und verbreitet. Um die Anzahl der teilnehmenden Repositorien kontinuierlich zu erhöhen, nutzen die Projektpartner neben den einschlägigen Foren vor allem Publikationen und Fachtagungen, um den Dienst OA-Statistik weiter bekannt zu machen. Zudem werden einzelne Repositorienbetreiber über den DINI-Verteiler direkt angeschrieben und zur Teilnahme am Dienst

²⁶ vgl. ebenda, S. 25.

animiert. Die Empfehlung zur Verwendung der OAS-Infrastruktur im DINI-Zertifikat 2010 führt zu einer weiteren Verbreitung des Dienstes. Informationen zum Projekt und zum Dienst OA-Statistik werden auf der Projektwebsite bereitgestellt. Über das Forum der Informationsplattform CARPET²⁷ können Fragen gestellt werden. Die Fragen und Antworten sind öffentlich sichtbar und bieten somit weiteren Interessierten die Möglichkeit, Informationen über OA-Statistik zu erhalten. Zusätzlich soll während der Projektlaufzeit ein Webinar angeboten werden, bei dem der Dienst und die Teilnahme an diesem interessierten Repositorienbetreibern näher gebracht werden kann.

Stand der Umsetzung:

Informationen zum Projekt und zum Dienst werden seit Projektbeginn gesammelt und über die Projektwebsite bereitgestellt. Dieser Datenpool wird kontinuierlich erweitert. Die FAQs im CARPET-Forum werden zurzeit erstellt und im Laufe des Frühjahrs 2012 der Öffentlichkeit zur Verfügung gestellt. Ein Webinar ist für Mitte 2012 geplant. Der Dienst OA Statistik wurde bereits auf vielen Fachtagungen präsentiert²⁸, z. B. den Open-Access-Tagen oder dem CERN Workshop on Innovations in Scholarly Communication (OAI7). Auch für 2012 sind Vorträge u. a. auf dem Deutschen Bibliothekartag, den Open-Access-Tagen und der Open Repositories geplant.

5.1.5 Registrierung neuer OAS-Data-Provider

Anforderung:

Die Integration eines OAS-Data-Providers soll so einfach wie möglich sein, bspw. durch die Bereitstellung eines möglichst automatisierten Workflows zur Registrierung neuer OAS-Data-Provider.

Lösungsweg:

Die Registrierung eines neuen OAS-Data-Providers erfolgt über ein Formular auf der Website von OA-Statistik²⁹. Dabei wird eine E-Mail an den Verantwortlichen des OAS-Service-Providers übermittelt. Das Telemediengesetz fordert nach § 13 Abs. 4 Nr. 3 TMG eine Verschlüsselung personenbezogener Daten. Das Registrierungsformular muss demnach die Anmeldeinformationen verschlüsselt übertragen, insbesondere dann, wenn die E-Mail nicht ausschließlich innerhalb eines nach außen abgeschotteten Netzbereichs verschickt wird. Nachdem der DP-Verantwortliche das Formular ausgefüllt und versendet hat, wird dessen Inhalt per E-Mail an den Verantwortlichen auf Seiten des OAS-Service-Providers geschickt. Ebenso erhält der DP-Verantwortliche eine E-Mail, in der er aufgefordert wird sich mit dem Verantwortlichen auf Seiten des OAS-Service-Providers per Telefon oder Post in Verbindung zu setzen, um mit ihm den Zugang zum OAS-Data-Provider des Repositoriums abzusprechen. Dies beinhaltet sowohl die Übermittlung des Passworts³⁰ zum OAS-Data-Provider als auch individuelle Absprachen zum Datenexport oder zu technischen Fragen. Ist diese Absprache erfolgt, muss der Verantwortliche auf Seiten des OAS-Service-Providers die Daten, die für das Hinzufügen eines neuen OAS-Data-Providers notwendig sind, manuell einspeisen. Dieses Einspeisen kann durch die automatische Verarbeitung der Formulardaten für den Verantwortlichen auf Seiten des OAS-Service-Providers vereinfacht werden.

²⁷ http://www.carpnet-project.net/forum/themen/oa_statistik/.

²⁸ <http://www.dini.de/projekte/oa-statistik/die-projektergebnisse/>.

²⁹ <http://www.dini.de/projekte/oa-statistik/die-technik/registrierung/#c2373>.

³⁰ Die Repositorien müssen ihren OAS-Data-Provider jeweils mit einem Passwort schützen, das nur dem OAS-Service-Provider bekannt ist und das die anderen OAS-Data-Provider nicht erfahren dürfen. Sie dürfen untereinander keine Nutzungsdaten austauschen, da alle OAS-Data-Provider dasselbe Salt verwenden und so ein Rückschluss auf die IP-Adressen rein theoretisch möglich wäre.

Stand der Umsetzung:

Das Formular für die Registrierung neuer OAS-Data-Provider ist erstellt und auf der OAS-Website eingebunden. In das Registrierungsformular sind noch Datenschutzhinweise – insbesondere über die Datenverarbeitung und ihren Zweck – aufzunehmen, die das für den DINI e. V. einschlägige Datenschutzgesetz vorschreibt. Auch die Verschlüsselung sowie die automatische Verarbeitung der Formulardaten muss noch erfolgen.

5.1.6 Installation des lokalen OAS-Data-Providers und Support

Anforderung:

Wie bei allen Software-Installationen ist es auch beim Dienst von OA-Statistik notwendig, dass für die Installation und auch den Betrieb der Software ein Support angeboten wird.

Lösungsweg

Nachdem das Repositorium sich am Dienst OA-Statistik registriert hat, erfolgt die Installation des OAS-Data-Providers in wenigen Schritten über die von OA-Statistik bereitgestellten Softwarepakete und Installationsanleitungen. Die VZG wird für den Betrieb des Dienstes Ansprechpartner für organisatorische sowie technische Fragen bereitstellen.

Stand der Umsetzung:

Eine Installationsanleitung steht den Verantwortlichen auf Seiten des Repositoriums auf der Website von OA-Statistik zur Verfügung.³¹ Zudem wird ein Einblick in die Verarbeitung der Logfileinformationen in sieben Schritten im OAS-Workflow³² gegeben.

Über das CARPET-Forum, per E-Mail sowie per Telefon werden Fragen zur Installation und zur Anpassung des OAS-Data-Providers gestellt. Zurzeit werden bereits beantwortete Supportanfragen ausgewertet und im CARPET-Forum als FAQs (frequently asked questions) präsentiert. So können bereits gemachte Erfahrungen als Hilfestellung bei der Installation weiterer OAS-Data-Provider genutzt werden.

5.1.7 Zentraler OAS-Data-Provider (ZDP)

Anforderung:

Die Hürde für die Teilnahme am Dienst OA-Statistik sollte so niedrig wie möglich gehalten sein. So sollten sich interessierte Repositorien auch an der OAS-Infrastruktur unter Berücksichtigung datenschutzrechtlicher Anforderungen beteiligen können, ohne einen OAS-Data-Provider lokal installieren und anpassen zu müssen.

Lösungsweg:

Die Repositorien liefern über eine sichere Verbindung die Logfiles des Repositoriums verschlüsselt an den zentralen OAS-Data-Provider. Die Einhaltung des Datenschutzes, die Konvertierung in das übertragbare Format OpenURL ContextObject, die Verknüpfung der Logfiles mit Dokumenten-IDs, und die Übermittlung der Datensätze an den OAS-Service-Provider übernimmt der zentrale OAS-Data-Provider.

³¹ <http://www.dini.de/projekte/oa-statistik/die-technik/registrierung/#c2374>.

³² http://www.dini.de/fileadmin/oa-statistik/technik/OAS-Workflow_0.9.pdf.

Stand der Umsetzung:

Das datenschutzrechtliche Vorgehen in diesem Verfahren wird zurzeit in Abstimmung mit der ZENDAS geklärt. Sollten organisatorische oder datenschutzrechtliche Gründe den Betrieb des zentralen Data-Providers verhindern, wäre der Gesamtdienst nicht gefährdet. Es handelt sich beim zentralen Data-Provider um eine Vereinfachung der Teilnahme am Dienst, nicht aber um eine Kernfunktionalität. Zurzeit wird testweise ein zentraler OAS-Data-Provider bei der VZG genutzt, der bereits zwei Repositorien bedient.

5.1.8 Zusammenfassung OAS-Data-Provider

Für die Teilnahme von Repositorien und Verlagen sind einige Anforderungen seitens der beteiligten Institutionen und auch auf Seiten des Diensteanbieters zu beachten. Es ist erforderlich, dass sich interessierte Institutionen über OA-Statistik informieren können und dass die Registrierung an den Dienst automatisiert erfolgt. Dazu sind der technische Support und die Registrierung über die Website bereits in der Projektlaufzeit vorhanden. Um die Vorgaben aus dem Datenschutz-Gutachten gerecht zu werden, wird noch in der Projektlaufzeit eine zentrale Vergabestelle für den Salt-String implementiert und nach Projektende durch den Projektpartner SULB Saarbrücken übernommen. Der aktuelle Workflow ist nach dem Gutachten der ZENDAS datenschutzrechtlich abgesichert. Im Folgenden muss noch geklärt werden, inwiefern das Vorhaben des zentralen OAS-Data-Providers umgesetzt werden kann.

5.2 OAS-Service-Provider

Der Betrieb des OAS-Service-Providers beinhaltet alle Aspekte einer verlässlichen Serviceinfrastruktur. Dazu gehören die Bereitstellung einer ausfallsicheren Hardwareumgebung, die Wartung der Software und die Überwachung des Dienstes mittels Werkzeugen zur Qualitätssicherung. Zudem bedarf es einer rechtlichen Person, die mit den beteiligten Partnern Verträge abschließen kann.

Anforderung:

Für den Dienst OA-Statistik ist der OAS-Service-Provider das technische Herzstück. Hier werden die Nutzungsrohdaten von allen beteiligten OAS-Data-Providern eingesammelt und im Anschluss aggregiert. Für diese zwei technischen Schritte ist ein ausreichend dimensionierter Server notwendig. Um die Finanzierung der gesamten Dienstleistung berechnen zu können, müssen die Kosten für diese zentrale Maschine beziffert werden.

Lösungsweg:

Die OpenURL ContextObjects werden durch die lokalen oder einen zentralen OAS-Data-Provider über eine OAI-Schnittstelle angeboten. Eine zwischen OAS-Service-Provider und OAS-Data-Provider geschlossene Lizenz regelt die Nutzung der durch OA-Statistik aggregierten und aufbereiteten Daten.³³ Der OAS-Service-Provider³⁴ holt die Daten vom OAS-Data-Provider im Fünf-Minuten-Takt oder je nach Verarbeitungsdauer ab. Im Anschluss werden die Daten auf dem OAS-Data-Provider gelöscht (Vorgabe des Datenschutzes).

Die OAS-Service-Provider-Software, welche in der ersten Förderphase entwickelt wurde, befindet sich zurzeit in einem Beta-Status. D.h. bis der Einsatz unter Produktivbedingungen möglich ist, sind noch Weiterentwicklungen und Optimierungen nötig. Entsprechend lässt sich auch zum momentanen Zeitpunkt nur eine Einschätzung hinsichtlich der benötigten Betriebsumgebung geben.

³³ Vgl. Anhang 2: Lizenzentwurf.

³⁴ Angepasster OAI-PMH Service-Provider.

5.2.1 Server

- Testumgebung: Virtuelle Maschine, 2 CPUs, 4 GB RAM, 250 GB Speicher
- Produktivumgebung: Virtuelle Maschine, 4 CPUs, 32 GB RAM, 500 GB Speicher

Beide Maschinen werden dazu standardmäßig ausfallsicher jeweils in einem VMWare-Cluster betrieben. Zusätzlich wird für beide Maschinen täglich eine Dateisystemsicherung und ein Datenbankabzug erstellt. Die verwendeten Netzwerkverbindungen sind redundant und für eine Datenübertragungsrate von 10 GBit pro Sekunde ausgelegt. Eine USV sorgt für die Möglichkeit, die Systeme auch bei einem Stromausfall sicher herunterzufahren bzw. kurze Stromunterbrechungen zu überbrücken.

5.2.2 Skalierbarkeit

Virtuelle Maschinen sind bezogen auf die Hardwareressourcen im Wesentlichen beliebig erweiterbar. Die angesetzte benötigte Hardware resultiert aus Erfahrungswerten, welche die VZG in ähnlichen Dienstleistungsprojekten gesammelt hat.

Nach bisherigem Kenntnisstand sollte die Software gemessen an den eingesetzten Algorithmen linear skalieren. Ist das im Einzelnen nicht der Fall, wird dies im Rahmen des oben erwähnten Weiterentwicklungs- und Optimierungsprozesses herzustellen sein.

5.2.3 Administration

Bei der Einrichtung der Standardaggregation pro Repositorium und Institution erscheinen einmalig 16 h Stunden Arbeitszeit als notwendig und hinreichend. Die Administration der Produktivmaschine sowie den kompletten Betrieb der Testumgebung übernimmt die VZG als Eigenleistung.

Bei Einrichtung von nicht-Standarddiensten berechnet sich die benötigte Zeit nach Aufwand (die Abrechnung erfolgt auch hier in Tagessätzen³⁵ nach einer verbindlichen Aufwandsschätzung).

5.2.4 Kosten

Die jährlichen Kosten für den Betrieb einer virtuellen Produktionsmaschine wie oben angegeben belaufen sich auf pauschal 2250 €³⁶. Eine Aufschlüsselung der Kosten findet sich im Abschnitt „Finanzierungsmodell“.

Stand der Umsetzung:

Sowohl der Testserver als auch das Produktivsystem sind in Betrieb. Der OAS-Service-Provider harvestet zurzeit zwei OAS-Data-Provider kontinuierlich und wertet diese Daten aus. Zur Zeit werden Werkzeuge zur Qualitätskontrolle entwickelt und getestet, um sicherzustellen, dass alle Daten korrekt geharvestet, erfasst, verarbeitet und ausgewertet werden. Zwei Repositorien erhalten aggregierte Daten in unterschiedlichen Formaten. Diese Daten werden sowohl von den Repositorienbetreibern wie auch von OA-Statistik einer Qualitätskontrolle unterzogen.

³⁵ Auftragsleistungen werden gemäß Benutzungs- und Entgeltordnung der VZG nach Aufwand berechnet. Die Höhe der Kosten richtet sich nach dem aktuell gültigen Tagessatz der VZG. Der Tagessatz berechnet sich aus den für den Betrieb des Verbundsystems notwendigen Kosten (Personal, Verwaltung, Mieten, Klima, Strom, Abschreibungen, etc.) geteilt durch die Anzahl der in einem Jahr erbrachten Arbeitstage. Damit sind alle Kosten abgedeckt und es werden zusätzliche keine Einzelpositionen, wie, z. B. Reisekosten gesondert in Rechnung gestellt.

Die Höhe des Tagessatzes wird jährlich auf Basis des Jahresabschlusses der VZG von der Verbundleitung des GBV überprüft. Siehe dazu auch: <http://www.gbv.de/Verbundzentrale/benutzungs-und-entgeltordnung-der-verbundzentrale>.

³⁶ Die VZG ist nicht mehrwertsteuerpflichtig.

5.2.5 Robot-Blacklist

Anforderung:

Für eine bessere Verarbeitung der nicht-menschlichen Zugriffe in den Logfiles ist eine Robot-Blacklist wünschenswert, auf der die Signaturen von Crawlern und Bots gesammelt und fortwährend ergänzt werden. Eine solche Liste sollte international abgestimmt sein, um doppelte Arbeit zu vermeiden und um eine internationale Vergleichbarkeit zu erleichtern. In der Arbeitsgruppe Usage Statistics von Knowledge Exchange wurde eine Unterstützung bei der Wartung einer solchen Liste angeboten. Ziel ist es, die Liste an einem zentralen Ort zu hosten und zu pflegen.

Lösungsweg:

Die VZG hat sich bereiterklärt eine Robot-Blacklist zu hosten, die über eine Schnittstelle für alle Interessierten zu erreichen ist. Gesammelt werden Angaben wie z. B. User-Agent und IP-Bereiche. In Abstimmung mit internationalen Institutionen und Projekten wird die Liste aktualisiert.

Stand der Umsetzung:

Zurzeit umfasst eine interne Liste über 700 Roboter. Erste konzeptionelle Überlegungen zur Robot-Blacklist bestehen, die Schnittstellen zur Ergänzung und zum Austausch der Liste wurden jedoch noch nicht realisiert.

5.2.6 Zusammenfassung OAS-Service-Provider

Der OAS-Service-Provider läuft seit Anfang der zweiten Förderphase beim Projektpartner VZG. Die in der ersten Förderphase entwickelte Software wird zur Zeit anhand der Datenbestände der Projektpartner getestet, optimiert und stellenweise erweitert, so dass diese ab der Jahresmitte 2012 als ein mandantenfähiges Softwaresystem in eine Testphase unter Produktionsbedingungen gehen kann. Zeitgleich wird die Auswertung der Zugriffsstatistiken, welche bislang den COUNTER-Standard bedient, auf die zusätzlichen Standards LogEC und IFABC ausgebaut. Die Rückgabe der aggregierten Daten wurde in verschiedenen Formaten (XML, JSON, iFrame Graphik) realisiert, wobei die Formatauswahl bei den Repositorien liegt. Weitere Rückgabeformate können leicht realisiert werden, es gibt aber bisher seitens der Repositorien bisher keine weiteren Anforderungen. Bis Ende der Projektlaufzeit wird eine international abgestimmte Robot-Blacklist initiiert, in welcher Informationen über nicht-menschliche Zugriffe gesammelt werden. Diese wird allen Interessierten zur Verfügung gestellt.

5.3 Finanzierungsmodell

Die Dienstleistungen, die über OA-Statistik erstellt werden, laufen zwar weitgehend automatisiert ab, dennoch entstehen durch Wartung, Anpassung und vor allem durch das auf ein dynamisches Wachsen des Angebots ausgerichtete System Kosten, die auch nach Ablauf der geförderten Projektlaufzeit abgedeckt werden müssen. Deswegen ist es unabdingbar, auf Basis der in den vorangegangenen Abschnitten erstellten Kostenaufwandsabschätzung ein nachhaltiges Finanzierungsmodell zu erstellen.

Anforderung:

Bis zum Ende der zweiten Projektphase muss eine zentrale Stelle eingerichtet werden, die sich um die Festlegung der jeweiligen Rechnungssummen, die Rechnungsstellung, den Versand der Rechnungen und um das Inkasso derselben kümmert. Dazu gehört auch das ordnungsgemäße

Verbuchen der Einnahmen im Rahmen einer transparenten Kontoführung. In diesem Zusammenhang muss auch die steuerliche Bearbeitung geklärt sein.

Lösungsweg

Der Projektpartner VZG hat eine Erhebung der Kosten durchgeführt. Dabei wurden einerseits die Kosten für den Betrieb des OAS-Service-Providers evaluiert, der nach Projektende bei der VZG betrieben wird. Andererseits wurde eine Kostenerrechnung für den Betrieb eines zentralen OAS-Data-Providers erhoben.³⁷

Zu den Kosten bei einer Beteiligung mit einem eigenen OAS-Data-Provider am OAS-Dienst im Einzelnen:

- Laufende Kosten pro Repositorium im ersten Jahr 1200 €. Dieser Betrag setzt sich dabei aus zwei Tagessätzen (Stand Februar 2012) zusammen (Support bei der lokalen Installation des OAS-Data-Provider, Anbindung an den OAS-Service-Provider, Betriebskosten).³⁸
- Laufende Kosten pro Repositorium und Jahr ab dem zweiten Jahr 375 €. Dieser Betrag setzt sich zusammen aus 225 € (Betriebskosten³⁹ eines Servers wie angegeben dividiert durch zehn, wobei diese Kalkulation von zehn teilnehmenden Repositorien ausgeht) für den Serverbetrieb und drei Stunden à 50 € (Stundensatz für TV-L 13) für die Softwarepflege.

Bei der Nutzung des zentralen OAS-Data-Providers fallen zusätzlich zu den zuvor genannten Beträgen folgende Kosten für die Konfiguration und Betreuung des zentralen OAS-Data-Providers an:

- Laufende Kosten pro Repositorium im ersten Jahr 600 € bei Auswertung eines Standard Apache Combined Log Formats; alle anderen auszuwertenden Formate berechnen sich nach Aufwand.
- Laufende Kosten pro Repositorium und Jahr ab dem zweiten Jahr 300 €. Dieser Betrag setzt sich zusammen aus 150 € (Betriebskosten eines Servers⁴⁰ dividiert durch zehn, wobei auch hier wieder von zehn teilnehmenden Repositorien ausgegangen wird) für den Serverbetrieb und drei Stunden à 50 € (Stundensatz für TV-L 13) für die Softwarepflege.

Die Kosten für die Teilnahme an OA-Statistik belaufen sich demnach auf:

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

Stand der Umsetzung:

Die Erhebung der Kosten ist durch die VZG erfolgt. Darauf aufbauend ist ein Finanzierungsmodell erstellt worden. Nach Inbetriebnahme des Dienstes OA-Statistik muss eine weitere Kosten-evaluierung erfolgen um das Finanzierungsmodell ggf. anzupassen.

³⁷ Alle im folgenden angegebenen Preise sind Bruttopreise.

³⁸ Für alle in der Projektlaufzeit beteiligten OAS-Data-Provider entfallen diese Kosten.

³⁹ Die Betriebskosten decken dabei ab: Hardware, Strom, Klimatisierung, Miete, Netzwerkinfrastruktur, Sicherung der Daten, ausfallsicheren Betrieb. Die Administration übernimmt die VZG als Eigenleistung.

⁴⁰ Die Kosten für den Betrieb eines Servers belaufen sich auf 1500 € pro Jahr.

5.3.1 Zusammenfassung Finanzierung

Für die Finanzierung des Dienstes wurde beim künftigen Dienstleister VZG eine Kostenevaluierung durchgeführt. Die Preise für interessierte Repositorien und Verlage stehen somit fest. Die vertragliche Regelung der Zusammenarbeit zeigen die im Anhang 1 aufgeführten Musterverträge. Die VZG wird nach Ablauf der Projektförderung für die gesamte finanzielle Abwicklung des Dienstes OA-Statistik zuständig sein.

6 Fazit und Ausblick

Der Dienst Open-Access-Statistik (OA-Statistik) stellt aggregierte und standardisierte Nutzungsdaten von in Repositorien abgelegten Open-Access-Dokumenten zur Verfügung. Diese 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 Entwicklung des Dienstes OA-Statistik wird in der Fachcommunity aufmerksam verfolgt. Bereits zu Beginn der zweiten Förderphase hatten mehrere Repositorienbetreiber ihr Interesse an einer Teilnahme an der OAS-Infrastruktur bekundet, um ihren Nutzern verlässliche und standardisierte Nutzungszahlen anbieten zu können. Die ersten sechs OAS-Data-Provider wurden installiert, konfiguriert und in die Infrastruktur von OA-Statistik eingebunden. Um diesen Dienst langfristig und performant anbieten zu können, muss das Projekt OA-Statistik nach Ablauf der zweiten Förderphase in eine stabile Struktur überführt werden. Damit dies gelingt, müssen die unter OAS-Data-Provider, OAS-Service-Provider und Finanzierung aufgeführten Anforderungen und Lösungswege bis zum Ende der zweiten Projektphase erfüllt und die Lösungen umgesetzt sein. Ab diesem Zeitpunkt kann OA-Statistik als dauerhafter und robuster Dienst durch die VZG – unterstützt durch die DINI AG „Elektronisches Publizieren“ sowie durch Eigenleistungen der Projektpartner – betrieben werden.

Durch die Teilnahme am Dienst OA-Statistik erhalten Autoren und Repositorien ein Instrumentarium zur Erstellung von Rankings und somit zur Analyse der eigenen Wettbewerbsfähigkeit. Die erhobenen standardisierten und visualisierten Nutzungsdaten über den Impact eines Dokuments können an die jeweiligen Autoren weitergegeben werden, was sicherlich ein Argument für das Einstellen des Dokuments in das Repositorium bildet. Schließlich erhalten die Repositorien die entscheidenden Daten für ein effektives Monitoring sowie für eine Messung der eigenen Reichweite. Dies ist nicht zuletzt ein werbendes Argument für die Repositorien, zumal die tagesaktuellen Auswertungen auch in die jeweiligen Websites eingebunden werden können. Auch für Verlage zeigt sich durch die Nutzung von OA-Statistik ein deutlicher Mehrwert durch ein internes Monitoring, das aufzeigt, welche Journals und Themen bei den Lesern den größten Anklang finden, woher die Leser kommen und wie lange sie auf der Seite bleiben.

Während der gesamten Projektlaufzeit fallen für Repositorienbetreiber keine Kosten für die Teilnahme an OA-Statistik an. Auch nach der Überführung in einen stabilen Dienst werden die Kosten relativ gering bleiben, so dass erwartet werden kann, dass die OAS-Infrastruktur schnell wachsen wird. Abzuwarten bleibt, ob die ermittelten Nutzungsdaten an kommerzielle Anbieter in rechtlicher Hinsicht weitergegeben werden dürfen und ob diese an einem Kauf der Daten Interesse haben werden. Hierbei wird vermutlich eine Rolle spielen, inwiefern und wie schnell die Anzahl der beteiligten OAS-Data-Provider erhöht werden kann und damit eine kritische Masse darstellt.