

Böhm, Peter; Rittberger, Marc
**Einsatz von Webanalyse in überregionalen
Informationsinfrastruktureinrichtungen**

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:

formally and content revised edition of the original source in:

Information - Wissenschaft & Praxis 67 (2016) 4, S. 241-251



Bitte verwenden Sie beim Zitieren folgende URN /
Please use the following URN for citation:
urn:nbn:de:0111-dipfdocs-149485

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

Deutsches Institut für
Internationale Pädagogische Forschung
Frankfurter Forschungsbibliothek
publikationen@dipf.de
www.dipfdocs.de

Einsatz von Webanalyse in überregionalen Informationsinfrastruktureinrichtungen.

Implementing web analysis in national information infrastructure institutions

Peter Böhm & Marc Rittberger

Deutsches Institut für Internationale Pädagogische Forschung, Frankfurt am Main, Berlin
Schloßstraße 29, 60486 Frankfurt am Main
{boehm,rittberger}@dipf.de

Zusammenfassung

Um die Nutzung ihrer Informationsdienste und -angebote beurteilen zu können, setzen Informationsinfrastruktureinrichtungen der Leibniz-Gemeinschaft die Webanalyse ein. Sieben Leibniz-Informationsinfrastruktureinrichtungen wurden mit einem Online-Fragebogen und einem teilstandardisierten Interview zur Nutzung der Webanalyse befragt. Es werden die verwendeten Methoden, Werkzeuge und Metriken sowie die verfügbaren Ressourcen als auch die Zukunftsperspektive der Webanalyse an den Einrichtungen beschrieben. Insgesamt wird der Stellenwert der Webanalyse an den Instituten als hoch angesehen. Die bisher wenig ausgeprägte Standardisierung und die fehlende Einheitlichkeit der Metriken und Erhebungsmethoden erschweren einen möglichen Vergleich von Nutzungsdaten allerdings erheblich.

Deskriptoren: Webanalyse, Leibniz-Gemeinschaft, Informationsinfrastrukturen, Metrik, Datenschutz, Werkzeuge, Befragung

Abstract

In order to assess the usage of information services and offers, information infrastructure institutes in the Leibniz Association have implemented web analyses. Based on online questionnaires and part-standardized interviews, seven Leibniz infrastructure institutes were assessed regarding their implementation of web analysis. The applied methods are described as well as instruments and metrics, together with available resources and the future perspective of web analysis at the Leibniz institutes. Altogether, the relevance of web analysis is rated high at the Leibniz institutes. However, a possible comparison of usage data is impeded owing to lacks of standardization, and the missing harmonization of metrics and assessment methods.

Descriptors: Web Analytics, Leibniz-Association, Information Infrastructure, Metrics, Privacy, Tools, Survey

Résumé

Descripteurs:

Motivation

2006 wurde mit der Studie zur „Neuausrichtung der öffentlich geförderten Informationseinrichtungen“ (Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung 2006) der Bund-Länder-Kommission (BLK) eine Reihe von Empfehlungen für die Weiterentwicklung der nationalen Informationseinrichtungen gegeben. An vielen Stellen des deutschen Wissenschaftssystems begann eine Diskussion zu den Veränderungen der Informationsinfrastruktur (Rahmenkonzept für die Fachinformationsstruktur in Deutschland 2009; Kommission Zukunft der Informationsinfrastruktur 2011; Wissenschaftsrat 2011; Astor et al. 2011; Lipp 2012) und es wurde intensiv über die Rolle der nationalen Informationseinrichtungen diskutiert. Die Zunahme der Sichtbarkeit der Forschungs- und Informationsinfrastrukturen zeigt sich auch bspw. an der zunehmenden Bedeutung der Frage der Forschungsinfrastrukturen im Zusammenhang mit den Geisteswissenschaften und deren Digitalisierung (siehe auch die Förderung von Digital Humanities Zentren Textgrid¹, ZenMEM², Kallimachos³, CEDIFOR⁴ oder (Burghardt et al. 2015)). Auch gewinnen Forschungsinfrastrukturen im Wissenschaftssystem selbst an Gewicht. So hat der Wissenschaftsrat im Sommer 2015 die Dokumentation von Forschungsinfrastrukturen im Kerndatensatz Forschung befürwortet (Wissenschaftsrat 2016) und die Leibniz-Gemeinschaft dokumentiert die große Fülle ihrer Forschungsinfrastrukturen und deren Qualität durch zwei Broschüren (Projektgruppe im Strategiefondsprojekt „Forschungsinfrastrukturen und Wissenstransfer in der Leibniz-Gemeinschaft – Best Practice Modelle und Strategien“ 2015; Leibniz-Gemeinschaft 2015). Allerdings steckt die Differenzierung der Qualitätsmerkmale noch in den Kinderschuhen. So wird in (Projektgruppe im Strategiefondsprojekt „Forschungsinfrastrukturen und Wissenstransfer in der Leibniz-Gemeinschaft – Best Practice Modelle und Strategien“ 2015, S. 6) zur Bestimmung von Qualität auf die Schwierigkeiten aufmerksam gemacht: „Während sich Indikatoren zur Evaluierung von Forschungsleistungen weitestgehend etabliert haben, sind diese für Forschungsinfrastrukturleistungen häufig noch nicht definiert. Zu fragen ist, wie sich für oft einzigartige, deshalb für die Forschung einerseits so wertvolle, andererseits aber auch schwer vergleichbare Forschungsinfrastrukturen Qualität bestimmen und ggf. auch mit Hilfe quantitativer Indikatoren bewerten lässt: Auf welche best-practice-Modelle können sich Akteure beziehen, wenn es um wissenschaftliche Qualität jenseits von Forschung und Forschungsevaluation geht?“ Im Verlauf des Textes werden dann sechs Qualitätskriterien identifiziert,

1 <https://textgrid.de/>

2 <http://zenmem.de/>

3 <http://kallimachos.de/>

4 <https://www.cedifor.de/de/cedifor/>

um hochwertige Forschungsinfrastrukturen auch disziplinübergreifend angemessen abzubilden. Die Infrastrukturen sollen danach zeitgemäß (1), relevant (2), qualitätsgesichert (3), nutzerorientiert und antizipativ (4), transparent zugänglich und rechtlich validiert (5) sowie nachhaltig (6) sein. Diese Kriterien können auch unterschiedlichen Sichten des Qualitätsmanagements zugeordnet werden, sei es einer kunden- (2, 4, 5), produkt- (1, 2) oder herstellungsorientierten (3, 6) Sicht (Rittberger 2004). Zentral ist dabei das Kriterium der Relevanz (2) mit dem die Bedeutung von Forschungsinfrastrukturen für die jeweilige Fachgemeinschaft und die Nutzung thematisiert werden. Insbesondere die Relevanz ist daher zentral für eine Nutzersicht („*Quality is fitness for use*“ (Juran 1951)) mit deren Erwartungen („*quality lies in the eyes of the beholder*“ (Garvin 1988)) auf Qualität. Informationsinfrastrukturen werden in aller Regel über das Internet angeboten und Nutzungsaktivitäten können daher mit einschlägigen und gängigen Werkzeugen nachverfolgt werden. Im Themenfeld der sogenannten Webanalyse sind allerdings unterschiedliche Zählweisen, Analysemethoden und Ressourceneinsätze möglich und vermutlich auch üblich.

Um der Frage der Nutzungsanalyse bei großen Informationsanbietern nachzugehen, wurde im Sommer 2015 eine Befragung durchgeführt, die vertiefende Kenntnisse über die derzeit bei sieben Informationsinfrastruktureinrichtungen der Leibniz-Gemeinschaft eingesetzten Webanalyse-Methoden und -Ressourcen erbringen sollte. Dabei sollte insbesondere geklärt werden, welche Metriken und Tools eingesetzt werden (1), welche Erkenntnisse durch die Webanalyse gewonnen werden (2), wie umfangreich Ausstattung und Ressourcen für die Webanalyse sind (3), welche Rolle der Datenschutz bei der Webanalyse spielt (4) und welche zukünftigen Entwicklungen und Herausforderungen absehbar sind (5)⁵.

Im Folgenden werden wir zunächst eine Einführung in die Webanalyse geben, anschließend die sieben Institute vorstellen, bevor wir die Methodik unseres Vorgehens erläutern und die Ergebnisse darstellen. Einige Folgerungen aus den Erkenntnissen werden zum Schluss präsentiert.

Webanalyse

Webanalyse dient der Informationsgewinnung und -auswertung bezüglich des Nutzungsverhaltens auf Websites bzw. Webauftritten. Dies dient zum besseren Verständnis des Interaktionsverhaltens von Nutzerinnen und Nutzern als auch der Optimierung der Webangebote, bspw. der Navigation oder des Layouts. Dabei hat sich jedoch noch kein einheitlicher Begriff durchgesetzt. So wird Webanalyse synonym mit Web-Analytics, Web-Controlling oder Traffic-Analyse verwendet. Zur Erhebung der Daten werden je Website eines oder mehrere Webanalyse-Tools eingesetzt.

Neben der technischen Optimierung der Website lassen sich durch die Webanalyse die Inhalte einer Seite verbessern. Die Webanalyse erlaubt es, Betreibern festzustellen, welche Inhalte besonders intensiv genutzt werden und

⁵ Die Erhebung wurde im Rahmen eines Kurses an der Hochschule Darmstadt im Studiengang Informationswissenschaft in Zusammenarbeit mit K. Mauelshagen, D. Seemann, V. Kovalenko, F. Becker, J. Browne & R. Sander durchgeführt.

welche eher wenig Nutzung erfahren. Ggf. lassen sich verschiedene Nutzergruppen identifizieren und im Idealfall können die Angebote auf die Bedarfe spezifischer Nutzergruppen hin optimiert werden (Zheng und Peltzverger 2015, S. 7675). Zur Durchführung der Webanalyse ist es notwendig, sowohl Methoden als auch Metriken zu kennen, die mit spezifischen Werkzeugen genutzt werden können. Einführend wollen wir daher im Folgenden vier gängige Methoden der Webanalyse und die wichtigsten Metriken vorstellen. Nach einem Hinweis auf datenschutzrelevante Rahmenbedingungen werden dann noch einige Werkzeuge für die Webanalyse vorgestellt.

Methoden der Webanalyse

Die vier wesentlichen Methoden der Webanalyse sind neben der Logfile-Analyse noch Page Tagging, Cookies und Application Level Logging.

- Damit eine Website über das Internet abgerufen werden kann, muss sie auf einem Webserver gehostet werden. Webserver protokollieren die verschiedenen Zugriffe auf die Website in sogenannten **Logfiles**. In diesen Logfiles, die ursprünglich für die Wartung der Webserver (z.B. die Fehleranalyse) vorgesehen waren, sind verschiedene Details über die Aufrufe der Website enthalten, etwa die Zeit des Aufrufs oder die IP-Adresse des Besuchers. Die in den Protokollen enthaltene IP-Adresse gibt nur Daten über das beim Zugriff genutzte Endgerät aus, so dass Nutzer/innen nicht über die Logfiles identifiziert werden können. Eine sitzungsübergreifende Beobachtung eines/r Nutzers/in ist nicht möglich (Zheng und Peltzverger 2015).
- Beim **Page Tagging** wird im Gegensatz zur Logfile-Analyse direkt auf das Gerät des Nutzers zugegriffen, um weitergehende Informationen zu erhalten. Dafür werden auf der genutzten Website einzelne, unsichtbare Pixel eingesetzt. Sobald der Nutzer sie mit dem Mauszeiger passiert, werden vom Browser Daten an den Analyse-Server versendet. Eine ähnliche Möglichkeit, genauere Informationen zu sammeln, sind JavaScript-Tags. Diese Tags werden in die Website integriert und übermitteln Daten über den Browser, die Systemumgebung auf dem Client und die Website-Nutzung an den Webanalyse-Server. Beide Verfahren erlauben es, Bewegungen des Mauszeigers oder das Tippen auf der Tastatur zu verfolgen (Heßler und Mosebach 2013, S. 374), womit eine tiefere Analyse des Nutzerverhaltens möglich wird.
- **Cookies** dokumentieren den Besuch einer Website und werden auf dem Gerät der Nutzerin oder des Nutzers abgespeichert. Wird erneut auf die Website zugegriffen, werden die abgespeicherten Informationen aus dem Cookie wieder an den Server übertragen. Cookies erlauben eine anonyme, aber eindeutige Identifikation der Nutzerin oder des Nutzers – und dies nicht nur auf der eigenen Website, sondern mittels Third-Party-Cookies auch auf anderen Websites, so dass das Webseiten übergreifende Surf- und Interaktionsverhalten beobachtet werden kann (Heßler und Mosebach 2013, S. 375).
- Um Vor- und Nachteile der bisher genannten Methoden zu kompensieren, werden auch häufig alle drei Methoden bei aktuellen Webanalysewerkzeugen eingesetzt, um so möglichst umfangreiche Daten über das Nutzungsverhalten zu sammeln (Heßler und Mosebach 2013, S. 376).

- Beim **Application Level Logging** sind die Funktionen für die Webanalyse bereits in einer Anwendung enthalten. Es wird somit kein weiteres Webanalysewerkzeug mit entsprechenden Methoden benötigt. Diese Lösung wird häufig bei Social Media und E-Commerce-Websites genutzt. Dort können vorgefertigte Softwarelösungen für den eigenen Blog oder den eigenen Shop genutzt werden, die Webanalysefähigkeiten innerhalb der Shop- oder Blogsoftware anbieten (Zheng und Peltzverger 2015, 2015, S. 7681).

Metriken der Webanalyse

Als Metriken werden quantitative Maße bezeichnet, die die Ereignisse auf einer Website spezifisch beschreiben und beobachtbar machen. Als übliche Metriken, die auch in unserer Untersuchung verwendet wurden, gelten Hit, Page View, Visitor, New Visitor, Returning Visitor, Sessions, Absprungrate und Downloads (Hassler 2010, S. 89; Web Analysis Association 2008; Zheng und Peltzverger 2015, S. 7681). Für eine umfangreiche Betrachtung von Metriken sei auf (Web Analysis Association 2008) verwiesen.

- Jede Anfrage bei einem Webserver wird als **Hit** erfasst. Dies umfasst alle Bestandteile einer Website, z.B. eingebettete Dateien wie CSS-Stylesheets oder Bilder. Wird eine Website mit sechs Grafiken aufgerufen sind das sieben Hits. Ein Hit für die Website, sechs für die Grafiken.
- Ein **Page View** gibt die Zahl der erfolgreichen Abrufe von normalen Internetseiten an. Dabei werden nur Seiten mit Inhalt gemessen. Weiterleitungsseiten oder Pop-Ups etc. werden nicht berücksichtigt.
- Als **Visitor** wird ein Besucher bezeichnet, der innerhalb eines bestimmten Zeitraums eine Website einmal oder in mehreren Sitzungen besucht hat. Dabei gelten automatisierte Website-Besuche, z.B. von Suchmaschinenrobotern nicht als Besucher.
- Ein **New Visitor** ist ein Besucher, der innerhalb eines Betrachtungszeitraums die Webseite erstmalig besucht.
- Als **Returning Visitor** wird ein wiederkehrender Besucher bezeichnet, der wiederholt auf der besuchten Website aktiv ist.
- **Sessions/Sitzungen** sind die zusammenhängenden Aktionen (Page Views) eines Besuchers mit der Website. Eine Sitzung kann dabei aus einem bis vielen Seitenabrufen bestehen. Wird innerhalb eines gewissen Zeitraums (meist 30 Minuten) keine weitere Aktion durchgeführt, gilt die Sitzung als beendet.
- Als **Absprungrate** wird der Anteil der Sitzungen bezeichnet, die nur eine Aktion (z.B. einen Seitenaufruf) umfassen.
- **Download** umfasst die Anzahl aller Dateien (z.B. PDF-Dateien), die von Websites heruntergeladen werden.

Datenschutz

Routinemäßig wird bei der Webanalyse die IP-Adresse jedes Besuchers protokolliert. Da die IP-Adresse als ein personenbezogenes Datum betrachtet wird, müssen die datenschutzrechtlichen Vorgaben bei der Webanalyse beachtet werden (Vollmert et al. 2012). Zur Erhebung und Speicherung personenbezogener Daten braucht es das Einverständnis der Besucher der Website. Bei Webbesuchern ist dies in der Regel nicht praktikabel und daher wird in der Regel die Erhebung personenbezogener Daten vermieden, bspw. durch Anonymisierung der IP-Adresse (Vollmert et al. 2012). Will man

Nutzungsprofile für einzelne Nutzerinnen oder Nutzer erstellen, um ihre Aktivitäten auf der Seite großflächig zu beobachten, so darf dies nur anonymisiert, bspw. unter einem Pseudonym, geschehen.

Werkzeuge

Webanalyse-Tools sind so zahlreich wie unterschiedlich in ihren Funktionen oder den Preisen. Einige Tools arbeiten logfile- und andere cookiebasiert. Manche Tools erlauben es, hohe Anforderungen an den Datenschutz zu stellen und haben entsprechende Verschlüsselungsfunktionen, wiederum andere Tools bieten einen sehr umfangreichen Funktionsumfang oder eine einfache Installation (Webanalyse Tool o.J.).

Von Belang bei der Auswahl eines Webanalyse-Tools sind insbesondere die Kriterien Datenschutz, Kosten, Installation, Support, Dashboard/Oberfläche und Datenerfassung. Die bekanntesten Werkzeuge sind Google Analytics und Piwik. Google Analytics besticht durch eine große Vielfalt an Funktionen, wohingegen Piwik, die bekannteste Open-Source-Variante ist.

Leibniz-Institute

Als Zielgruppe für die Befragung wurden nationale Informationsinfrastruktureinrichtungen der Leibniz-Gemeinschaft betrachtet, die schon aus der Logik der von ihnen zu erbringenden Kosten-Leistungsrechnung alle Webanalyse betreiben müssen, um die Zugriffszahlen zu ihren Informationsangeboten zu erheben. Allen sieben Einrichtungen ist gemein, dass sie als nationale Anbieter insbesondere im Bereich der Literaturdokumentation umfangreiche Angebote im Internet zur Verfügung stellen. Im Einzelnen handelt es sich um vier sozialwissenschaftlich orientierte Institute und drei naturwissenschaftliche Institute. Umfangreiche und detaillierte Informationen zu den Einrichtungen findet in der Webpräsenz, in den Jahresberichten der Einrichtungen und den Evaluierungsberichten der Leibniz-Gemeinschaft zu den Einrichtungen⁶.

Sozialwissenschaftliche Institute

- DIPF: <http://www.dipf.de/>, (Deutsches Institut für Internationale Pädagogische Forschung 2015)
- GESIS: <http://www.gesis.org/>, (GESIS - Leibniz-Institut für Sozialwissenschaften 2015)
- ZBW: <http://www.zbw.eu/>, (Deutsche Zentralbibliothek für Wirtschaftswissenschaften – Leibniz-Informationszentrum Wirtschaft 2015)
- ZPID: <https://www.zpid.de/>, (ZPID - Leibniz-Zentrum für Psychologische Information und Dokumentation 2015)

Naturwissenschaftliche Institute

- FIZ KA: <https://www.fiz-karlsruhe.de/>, (FIZ Karlsruhe - Leibniz Institut für Informationsinfrastruktur 2013)

⁶ Die Stellungnahmen des Senats der Leibniz-Gemeinschaft zu den sieben Instituten aus den Jahren 2011 und 2012 finden sich unter: <http://www.leibniz-gemeinschaft.de/ueberuns/evaluierung/das-evaluierungsverfahren-des-senats/senatsstimmungen/>

- TIB: <https://www.tib.eu/>, (TIB - Technische Informationsbibliothek und UB Universitätsbibliothek Hannover 2015)
- ZB MED: <http://www.zbmed.de/>, (ZBMED - Leibniz-Informationszentrum Lebenswissenschaften 2015)

Befragung

Die Befragung der sieben Institute erfolgte in zwei Stufen. Zuerst wurde im Mai 2015 eine Online-Befragung von sieben vorab von den Instituten benannten und informierten Personen durchgeführt. Der zweite Schritt war ein persönliches, halbstandardisiertes Interview am Telefon anhand eines Interviewleitfadens. Die Telefoninterviews wurden aufgenommen und transkribiert und zur Analyse kodiert.

Insgesamt enthielt der Fragebogen 39 Fragen – mit 16 Freitextantworten und 14 einfachen Antworten (z.B. Radio-Buttons oder Liste), 6 Antworten mit Mehrfachauswahl sowie 3 weiteren Antworten. Der Fragebogen wurde mithilfe von LimeSurvey⁷ realisiert und durch einen Pretest validiert.

Ergebnisse

Metriken

Die Institute nutzen fast alle genannten Metriken, Visitor und Page View werden von allen Instituten erhoben, Hits, die am wenigsten aussagekräftige Metrik findet nur bei drei Instituten Anwendung.

Welche der folgenden Metriken werden an Ihrem Institut im Rahmen einer Webanalyse erhoben?

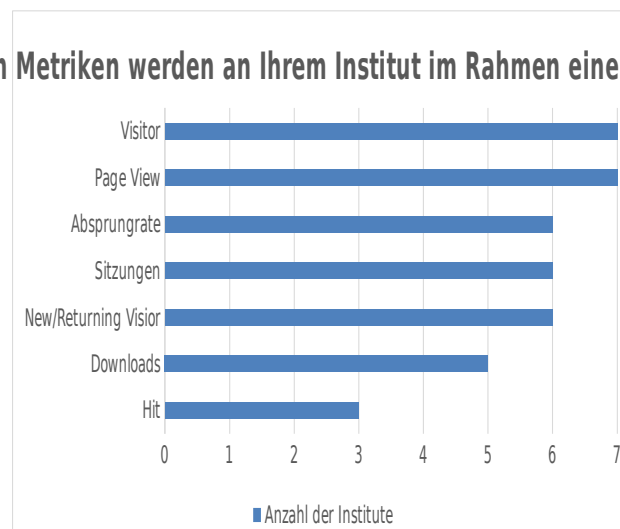


Abbildung 1: Nutzung von Metriken an den Instituten

Die von allen Instituten genutzten Metriken Visitor und Page View werden auch als sehr wichtig eingeschätzt, um die Nutzung der von den Instituten betriebenen Portale nachweisen zu können. Drei Institute verwenden sie auch als Leistungsindikatoren. Für ein Institut ist außerdem die Vergleichbarkeit mit anderen Leibniz-Instituten wichtig. Zwei Institute nutzen diese Metriken auch für den Vergleich interner Angebote. Einige Institute nutzen auch Sessions und Absprungrate, um Daten über die Bewegung der Nutzerinnen und Nutzer auf

⁷ Lime Survey ist eine frei verfügbare Applikation zur Erstellung von Online-Umfragen (<https://www.limesurvey.org/de/>)

ihren Angeboten zu erhalten. Vier der befragten Institute verfügen außerdem über Produkte zum Download, meist Volltexte, und haben daher diese Metrik mindestens als wichtig eingestuft.

Zwei Institute gaben an, dass das Land, aus dem zugegriffen wird oder die Sprache des Zugreifenden, wichtig ist. In einem Fall ist das von Bedeutung, weil der Anteil der deutschsprachigen Nutzerinnen und Nutzer zur Auftragserfüllung des Instituts erhoben wird. Bei einem anderen Institut wird diese Metrik überwacht, weil auch internationales Publikum erreicht werden soll.

Ein Großteil der befragten Institute hat angegeben, dass sowohl die Hauptseite als auch einzelne Produkte und Suchportale analysiert werden. Zwei Institute nannten in diesem Zusammenhang explizit auch Projekte. Zwei andere Institute haben ein starkes Interesse an Bewegungsprofilen und Klickpfaden ihrer Nutzerinnen und Nutzer.

Segmentierung, also die Kreuztabellierung zweier Metriken, wird nur von einem Institut zur Identifikation von mobilen Endgeräten genutzt, um die Angebote gezielt zu optimieren. Ein weiteres Institut möchte feststellen, ob der genutzte Browser der User die Website des Institutes darstellen kann. Zwei Institute können sich an eine Testsegmentierung bei bestimmten Angeboten erinnern, ein weiteres Institut nennt Datenschutzgründe für den Verzicht auf eine Segmentierung.

Verwendete Werkzeuge

Welches Werkzeuge zur Webanalyse nutzen Sie im Institut aktuell?

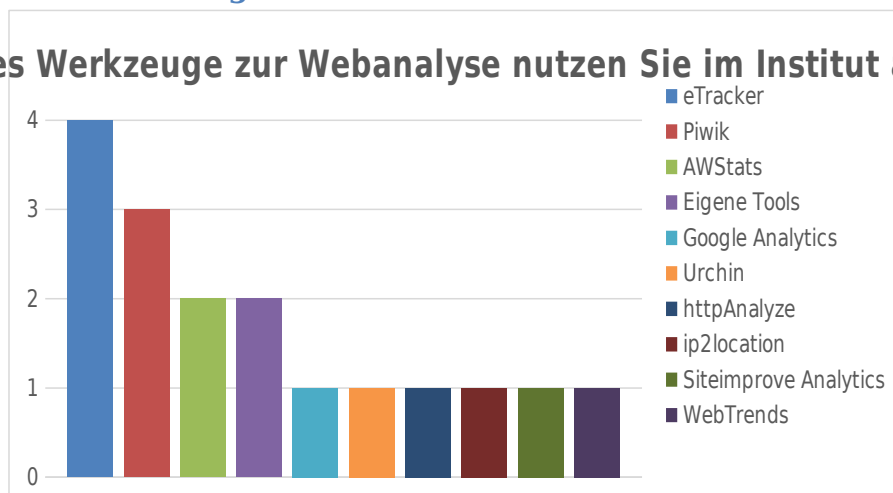


Abbildung 2: An den Instituten eingesetzte Werkzeuge

Auffallend ist, dass sechs der sieben befragten Institute angaben, mit mehreren Werkzeugen zu arbeiten. Vier der sieben befragten Institute arbeiten mit eTracker, was wohl an der Einfachheit in der Handhabung und Einbindung liegt: „[...] da ist natürlich so ein Tool mit Weboberfläche auch für weniger technisch versierte Nutzer dann einfacher auszuwerten. Und das waren somit die Beweggründe, warum man sich für e-Tracker entschieden hat [...]“. Ein weiteres Institut merkt an: „[...] also war auch die einfache Bedienung und Auswertung glaube ich ein sehr großes Kriterium.“ Drei weitere Institute nutzen Piwik, allerdings nie als einziges Werkzeug. Weitere wollen auf Piwik umsteigen oder haben es getestet und sich nicht dafür entschieden. Interessant ist hierbei

besonders, dass Google Analytics kaum eingesetzt wird, obwohl es zu den leistungsfähigsten Webanalyse-Werkzeugen zählt⁸.

Insgesamt haben drei Institute in den letzten zehn Jahren das Analyse-Werkzeug gewechselt. Das vorher genutzte Werkzeug war neben Google Analytics immer AwStats. Gründe dafür waren auch eine „erweiterte Datenerfassung [und] Standardisierung“ sowie Probleme mit dem Tool, da „nur noch Roboter erkannt“ wurden. Insgesamt sind die befragten Institute mit den von ihnen eingesetzten Tools zufrieden – zumindest vermissen nur zwei Institute Metriken in dem von ihnen eingesetzten Tool. Dabei wurde einmal die Device-Erkennung als noch „suboptimal“ benannt, ein anderes Institut vermisst „Bewegungsmuster-Korrelationen“.

Weshalb nutzt Ihr Institut das am häufigsten verwendete Webanalyse-Tool?

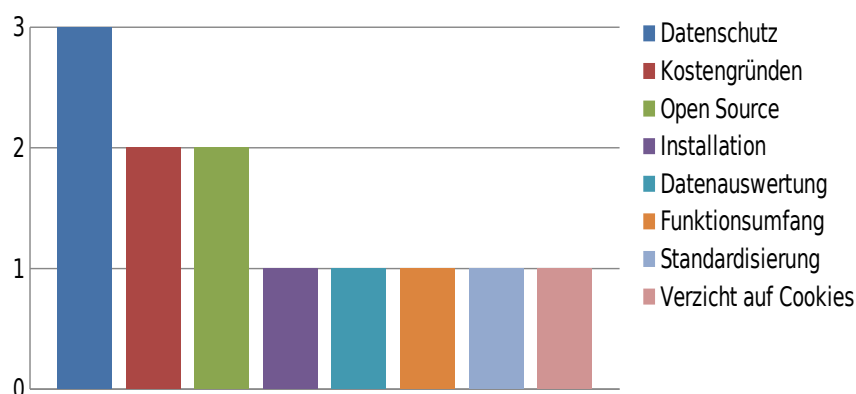


Abbildung 3: Gründe für die Nutzung des am häufigsten verwendeten Tool

Fünf der sieben Institute sehen den Datenschutz als wichtiges Auswahl-Kriterium. Als weitere besonders wichtige Gründe für die Verwendung des von ihnen präferierten Werkzeugs wurden Kosten und die Verfügbarkeit als Open-Source genannt. Außerdem wurde die Datenerfassung von vier Instituten als wichtig angesehen. Installation und Dashboard/Oberfläche wurden jeweils nur einmal genannt. Der Support als Kriterium für die Auswahl eines bestimmten Tools wurde von keinem Teilnehmer genannt.

Letztendlich wurde noch nach Maßnahmen zur Robot-Filterung gefragt. Robots sind beispielsweise Web-Crawler von Suchmaschinen. Da diese keine richtigen Besucher sind, sollen deren Seiten-Zugriffe nicht in der Webanalyse auftauchen. Die meisten Tools bieten Maßnahmen an, um dies zu ermöglichen. Darüber hinaus gaben drei Institute an, noch weitere Maßnahmen zur Robot-Filterung vorzunehmen. Zwei Institute gaben IP-Adressauflösung an. Dabei werden die IP-Adressen der Robots für die Webanalyse ignoriert. Ein Institut gab an, eine „robots.txt“ zu verwenden. Spezifische User-Agent-Filter („fragwürdige Herkunft der Zugriffe“) sowie Schwellwerte (Anzahl der Downloads) werden ebenfalls von einem Institut angewandt.

⁸ Dabei ist zu beachten, dass Google-Analytics bez. des Datenschutzes als mindestens problematisch gilt, bspw. in Hessen von öffentlichen Einrichtungen nicht genutzt werden darf.

Datenschutz

Die fünf Institute, die zu diesem Themenkomplex geantwortet haben, gaben an, die IP-Adressen der Nutzerinnen und Nutzer zu anonymisieren. Dabei wird die IP-Adresse „beschnitten“, d.h. eine gewisse Anzahl von Bits wird überschrieben, damit sie nicht mehr einer Person zugeordnet werden kann (Vollmert 2010). Drei der fünf antwortenden Institute gaben an, dass per sogenanntem Opt-Out-Cookie der Speicherung der eigenen Daten widersprochen werden kann. Dieser Widerspruch gilt auch für nachfolgende Seitenaufrufe. Nur eins der befragten Institute gab an, dass sie die No-Tracking Einstellungen der Browser respektieren. Allerdings ist es möglich, dass auch andere Institute bzw. deren Tools dies automatisch tun. Eine Pflicht, diese Einstellungen zu respektieren, gibt es in Deutschland nicht.

Zwei Institute sehen sich durch die Maßnahmen zum Datenschutz bei der Erhebung ihrer Daten eingeschränkt, zwei andere haben keine Probleme. Die Restlichen sind sich nicht sicher bzw. haben keine Antwort gegeben.

Alle befragten Institute vertreten ähnliche Prinzipien der Anonymisierung erhobener Daten unabhängig von den Tools, die sie benutzen. Ein Institut, das unter anderen Piwik benutzt, behauptet: *„[...] werden die Daten anonymisiert, indem IP-Adressen gehasht [verschlüsselt gespeichert] werden. Also die Logdaten liegen da, aber die IP-Adressen werden sozusagen versteckt.“* Genauso funktioniert auch E-Tracker, wie ein weiteres Institut sagt *„das, was wir sehen, sind alles anonymisierte Daten. Also wir sehen, wenn wir eine IP sehen, nur die ersten zwei Segmente. Und den Rest verwaltet E-Tracker in Eigenregie.“*

Drei Institute löschen die erhobenen Nutzerdaten nicht. Bei zwei Instituten werden die Daten regelmäßig gelöscht. Bei einem Institut wird dies unterschiedlich gehandhabt, da bei manchen Informationsdiensten die Daten nicht, bei anderen Diensten hingegen schon nach drei Monaten gelöscht werden. Bei einem weiteren Institut, welches ebenfalls mit Ja antwortete, werden die Rohdaten nach einem Jahr gelöscht, die aggregierten Daten allerdings nicht.

Die erhobenen Daten werden meistens nicht gelöscht, wie bei drei Instituten vermerkt wurde, obwohl ein weiteres Institut anmerkt, *„man die Daten dann nicht mehr benötigt“* und *„das Zeug liegt halt irgendwo und kein Mensch schaut es sich mehr an“*. Gründe dafür sind erstens die Notwendigkeit, die Daten manuell zu löschen und geringe Probleme, den notwendigen Speicherplatz vorzuhalten: *„Wir sind technisch glücklicherweise sehr gut ausgestattet“*. Außerdem werden die erhobenen Daten, die mit den Werkzeugen gespeichert werden, nicht weitergegeben: *„es eben keine Dritten sind, an die die Daten gehen, bleibt das quasi im Hause.“* ergänzt ein anderes Institut.

Wo weisen Sie Ihre Nutzer auf die eingesetzten Trackingverfahren hin?

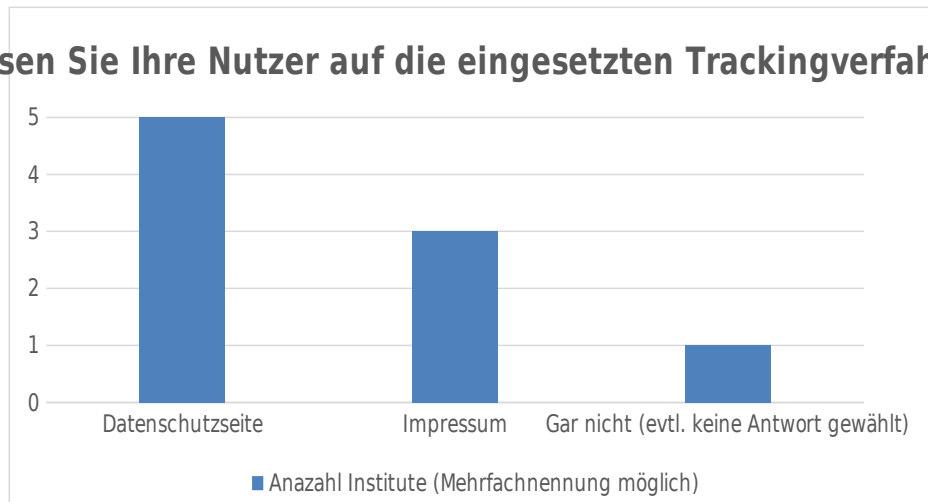


Abbildung 4: Hinweis auf Trackingverfahren

Zuletzt wurde in diesem Abschnitt gefragt, wo die Nutzer entsprechend der gesetzlichen Regelung auf die eingesetzten Trackingverfahren hingewiesen werden. Drei Institute weisen im Impressum auf die eingesetzten Trackingverfahren hin. Fünf Institute haben eine eigene, gesonderte Datenschutzseite. Zwei weitere Institute weisen auf beiden Seiten auf die Trackingverfahren hin. Ein Institut gibt an, nirgendwo auf die Trackingverfahren hinzuweisen.

Ressourcenausstattung

Zu Beginn dieses Frageblocks wurde nach der Anzahl der Beschäftigten gefragt, die sich mit der Webanalyse beschäftigen. Die Antworten fallen dabei recht unterschiedlich aus und variieren von anteilig eine Vollzeitkraft bis ca. 10 Personen in unterschiedlichen Arbeitsbereichen mit unterschiedlichen Zeitkontingenten. Dabei wird die Ausstattung als ausreichend angesehen, um bspw. die Kennzahlen für die Leistungsindikatoren zu erheben: *„was wir an Analyse fahren, was wir für Kennzahlen an die Leibniz liefern müssen, haben wir eigentlich alle Mittel und alle Mechanismen, die dazu nötig sind“*. Bei einem Institut wurde aber ein Mangel an einen Gesamtüberblick über alle Angebote und Portale bemerkt, dass *„jetzt vielleicht schon wünschenswert wäre, wenn noch mal ein Kollege das Ganze, wenn also ein Kollege noch einmal so den besseren Gesamtüberblick hätte.“*

Bei den jährlichen Sachkosten für Software gehen die Antworten recht weit auseinander. Zwei Institute haben gar keine Kosten. Diese beiden nutzen offenbar schon länger kostenfreie Tools, da sich die Kosten in den letzten zwei Jahren nicht verändert haben. Zwei weitere Institute zahlten im letzten Jahr *„knapp“* bzw. *„ca.“* 1.000 € für Lizenzen von Webanalyse-Software. Dabei haben sich die Kosten einmal erhöht. Eines der beiden Institute nutzt eTracker als kostenpflichtiges Werkzeug – das andere nutzt sowohl Siteimprove Analytics als auch ip2location. Dieses Institut hat außerdem die Nutzung eines kostenfreien Tools erwogen. Aufgrund des zu erwartenden erhöhten Arbeitsaufwands für Installation und Wartung wurde dieser Plan jedoch wieder verworfen. Ein weiteres Institut, welches auch eTracker nutzt, hat im vergangenen Jahr insgesamt 3.000 € für Lizenzen bezahlt.

Nur zwei Befragte wussten, dass ihr Institut in Zukunft mehr Ressourcen bereitstellen will. Interessant dabei ist, dass es sich nur bei Einem um Ausgaben für Lizenzen handelt. Das andere Institut plant „*Forschung zu Altmetrics*“. Dabei handelt es sich um neue, alternative Metriken (Priem et al. 2010) im Vergleich zu den herkömmlichen traditionellen. Beispielsweise soll damit der „*Einfluss*“ von z.B. (Blog-)Artikeln auf das „*akademische Ökosystem*“ gemessen werden können.

Zukunftsperspektiven

Abschließend wurde nach den zukünftigen Herausforderungen an die Webanalyse gefragt. Die Mehrheit der befragten Institute gab an, dass die Anpassung an neue digitale Geräte, wie z.B. Smartphones oder Tablets, eine zentrale Herausforderung ist. Weiterhin stimmten drei Institute zu, dass es in Zukunft nötig wird, sich an den schnell entwickelnden Markt anzupassen. Nur ein Institut sah Probleme in gegebenenfalls neuen Datenschutzbestimmungen. Auch wurde von einem Institut angegeben, dass es ein Problem sein könnte, „*ausreichend Geldmittel zur Verfügung haben, um sich kommerzielle Lösungen leisten zu können.*“

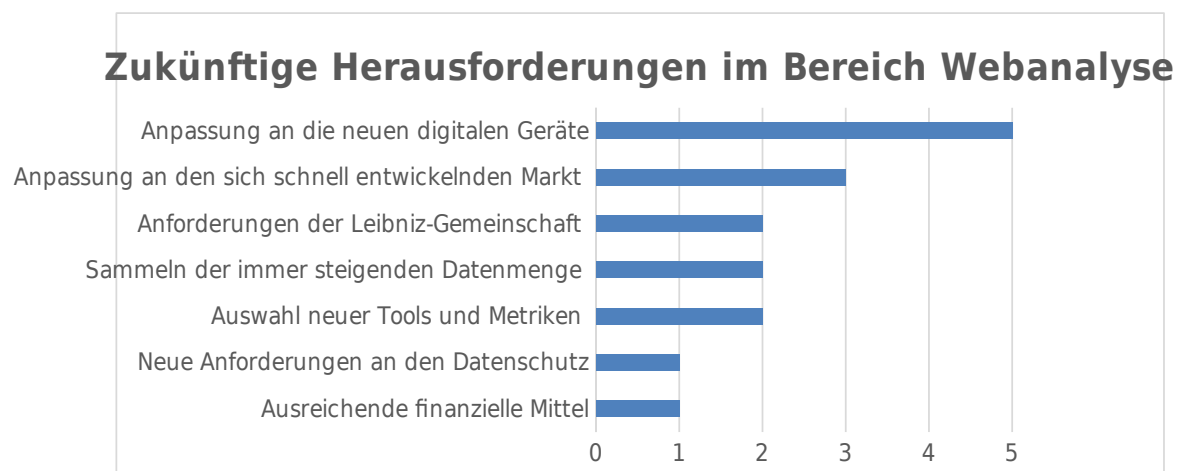


Abbildung 5: Zukünftige Herausforderungen

Übereinstimmend sehen die Befragten die Anpassung an neue digitale Geräte als Herausforderung an: „*Technik und Tools zu haben, die eben diese Kennzahlen ermitteln können*“. Ein weiteres Institut findet es auch wichtig, die Gestaltung von Webseiten zu verbessern: „*die mobile Nutzung oder dass die Nutzung mit Bildschirmen, die klassischerweise auf mobilen Endgeräten drauf sind, um diesen Umweg zu gehen, dass die Stück für Stück zunimmt. Dass ist das, was wir feststellen. Und deshalb sind wir auch responsive mit unserem Design*“. Die immer stärker steigende Anwendung von Smartphones und Tablets stellt die Institute vor die Aufgabe, Prinzipien und neue technische Verfahren für das Multi-Device-Tracking auszuarbeiten.

Zum Schluss wurde noch nach der Bedeutung der Webanalyse am jeweiligen Institut gefragt. Dabei kam heraus, dass der Stellenwert der Webanalyse bei fünf Instituten „eher hoch“ ist und nur bei einem „weder noch“. Insgesamt ist die Webanalyse also für fast alle Institute ein wichtiges Feld. Vier der Befragten sind darüber hinaus der Meinung, dass die Bedeutung der Webanalyse zunehmen wird.

Fazit

Die Befragungen zeigen, dass die verbreiteten Metriken Visitor, Page View und Sitzungen erwartungsgemäß auch bei den untersuchten Leibniz-Instituten eine wichtige Rolle spielen. Darüber hinaus wird die Absprungrate von fast allen Instituten ausgewertet.

Auffällig ist, dass der Datenschutz der mit Abstand wichtigste Aspekte bei der Auswertung eines Webanalyse-Tools ist, noch vor Kostenerwägungen. Dies zeigt die Sensibilität, mit dem diesem Thema begegnet wird.

Von Interesse war auch, welche Berufsgruppe die Antworten gab. Insbesondere bei Fragen zur Metrik, hingen die Antworten von der Berufsgruppe ab. Während Produktmanager eher Visitor, Page View und Session als wichtig ansahen, beobachteten technische Beschäftigte eher Hits, Action-Items und Absprungrate, um beurteilen zu können, ob die Infrastruktur, die Performance und Reaktion der Nutzer zueinander passen: „[...] interessante Zahlen um Performance zu messen, um natürlich festzustellen ob die Infrastruktur noch ausreichend zur Verfügung gestellt werden können. Wobei da ja eigentlich Hits eher interessant sind.“

Insgesamt ist der Stellenwert der Webanalyse hoch. Nur ein Institut gab an, dass die Webanalyse in ihrem Hause nicht so wichtig sei: „[...] für manche (.) Anbieter ist möglicherweise eine Webanalyse sehr sehr sehr sehr sehr viel wichtiger als für uns. [...] Also für uns ist die Analyse der Webseiten, das haben Sie ja bemerkt, nicht so, hat nicht einen ganz so hohen Stellenwert.“

Auch herrscht über die Vorgaben der Leibniz-Gemeinschaft noch häufig Unklarheit. Eine bessere Standardisierung für die Errechnung und Erhebung von Metriken wurde gefordert. Die fehlende Einheitlichkeit der Metriken und unklaren Vorgaben zu den Kennzahlen, die die Werkzeuge liefern, erschweren den Vergleich der Leibniz-Institute erheblich: „Und ja das muss man halt hinterfragen und generell liegt eine Herausforderung denke ich auch in dem generellen Reporting von Webanalysedaten an die Leibniz-Gemeinschaft. Ich glaube, das ist auch noch nicht so ganz ausgereift, wichtig wäre hier natürlich auch eine Einigkeit über die Metriken zu erreichen und natürlich auch noch einmal im Detail, wie diese Metriken errechnet werden, das kann ja auch sehr unterschiedlich sein.“ Ein anderes Institut merkt an: „Also es gibt keine Leibniz-übergreifende Regelung soweit ich das weiß für Webanalysetools oder Anforderungen ist da ja eher die Frage. Denn die Anforderungen sind ja von dem Online-Angebot ganz stark abhängig. Insofern wäre das wahrscheinlich gar nicht sinnvoll für 80 Institute, da irgendwelche Regelungen zu schaffen.“ Ein drittes Institut ergänzt „[...] wenn seitens der Leibniz für diese Kennzahlen, die geliefert werden müssen, neue Anforderungen, neue Kennzahlen kommen und man natürlich immer das Problem hat, kann mein Tool, was ich im Moment einsetze, das ... das leisten.“

Hier scheint sowohl von den technischen und methodischen Möglichkeiten als auch bezüglich effektiver Kriterien noch ein weiter Weg zu sein bis vergleichbare Zahlen zur Webanalyse für die beteiligten Institute verfügbar sind.

Literaturverzeichnis

- Astor, Michael; Klose, Georg; Heinzemann, Susanne; Riesenberg, Daniel (2011):
Evaluierung des DFG-geförderten Systems der Sondersammelgebiete. Online
verfügbar unter
http://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/evaluierung_ssg.pdf.
- Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (2006):
Neuausrichtung der öffentlich geförderten Informationseinrichtungen:
Abschlussbericht: Bonn : BLK, Geschäftsstelle (Materialien zur Bildungsplanung
und zur Forschungsförderung).
- Burghardt, Manuel; Wolff, Christian; Womser-Hacker, Christa (Hg.) (2015):
Themenheft Digital Humanities. *Information, Wissenschaft & Praxis* 66 (5-6).
Berlin/Boston: Walter de Gruyter GmbH.
- Deutsche Zentralbibliothek für Wirtschaftswissenschaften – Leibniz-
Informationszentrum Wirtschaft (Hg.) (2015): Jahresbericht 2014 der ZBW –
Leibniz-Informationszentrum Wirtschaft. Wandel Gestalten. Online verfügbar
unter <http://www.zbw.eu/fileadmin/pdf/ueber-uns/jb-2014.pdf>, zuletzt geprüft
am 02.05.2016.
- Deutsches Institut für Internationale Pädagogische Forschung (Hg.) (2015):
Wissen über Bildung. Bericht 2013-2014. Online verfügbar unter
<http://www.dipf.de/de/publikationen/pdf-publikationen/dipf-bericht-2013-2014>,
zuletzt geprüft am 02.05.2016.
- FIZ Karlsruhe - Leibniz Institut für Informationsinfrastruktur (Hg.) (2013):
Advancing Science. Themen und Zahlen 2012/2013. Online verfügbar unter
[https://www.fiz-
karlsruhe.de/fileadmin/redaktion/FIZ_Karlsruhe/Veroeffentlichungen/PDFs/FIZ_T
hemenuZahlen_de.pdf](https://www.fiz-karlsruhe.de/fileadmin/redaktion/FIZ_Karlsruhe/Veroeffentlichungen/PDFs/FIZ_ThemenuZahlen_de.pdf), zuletzt geprüft am 02.05.2016.
- Garvin, David A. (1988): *Managing quality. The strategic and competitive edge*.
New York: Free Press.
- GESIS - Leibniz-Institut für Sozialwissenschaften (Hg.) (2015): Jahresbericht 2014.
Online verfügbar unter
http://www.gesis.org/fileadmin/upload/institut/Jahresbericht_2014_web.pdf,
zuletzt geprüft am 02.05.2016.
- Hassler, Marco (2010): *Web analytics. Metriken auswerten, Besucherverhalten
verstehen, Website optimieren*. 2., erw. Aufl. Heidelberg u.a: mitp.
- Heßler, Armin; Mosebach, Petra (2013): *Strategie und Marketing im Web 2.0*.
Handbuch für Steuerberater und Wirtschaftsprüfer. Wiesbaden: Gabler Verlag.
- Juran, Joseph M. (1951): *Quality control handbook*. New York: McGraw-Hill.
- Kommission Zukunft der Informationsinfrastruktur (2011): *Gesamtkonzept für die
Informationsinfrastruktur in Deutschland*. Online verfügbar unter
http://www.gwk-bonn.de/fileadmin/Papers/KII_Gesamtkonzept.pdf.
- Leibniz-Gemeinschaft (2015): *Forschungsinfrastrukturen in der Leibniz-
Gemeinschaft. Research Infrastructures in the Leibniz Association*. Online

verfügbar unter http://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Presse/Publikationen/Leibniz_Infrastrukturen_2-2015_web.pdf.

Lipp, Anne (2012): "Die digitale Transformation weiter gestalten". Das Positionspapier der Deutschen Forschungsgemeinschaft zu einer innovativen Informationsinfrastruktur. In: *Zeitschrift für Bibliothekswesen und Bibliographie / Sonderhefte* 59 (2012), S. 291–300. DOI: 10.3196/186429501259614.

Priem, Jason; Taraborelli, Dario; Groth, Paul; Neylon, Cameron (2010): Altmetrics: A manifesto. Online verfügbar unter <http://altmetrics.org/manifesto>, zuletzt geprüft am 30.05.2016.

Projektgruppe im Strategiefondsprojekt „Forschungsinfrastrukturen und Wissenstransfer in der Leibniz-Gemeinschaft – Best Practice Modelle und Strategien“ (Hg.) (2015): Forschungsinfrastrukturen im Wissenschaftssystem Qualität – Reputation – Nachhaltigkeit. Online verfügbar unter http://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Presse/Publikationen/Leibniz_Broschuere_Infrastruk-Wiss-System_11-2015_final.pdf.

Rahmenkonzept für die Fachinformationsstruktur in Deutschland. Vorlage zur Sitzung des Ausschusses der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder (GWK) am 29.09.2009 (2009). Bonn. Online verfügbar unter <http://www.gwk-bonn.de/fileadmin/Papers/Rahmenkonzept-WGL.pdf>.

Rittberger, Marc (2004): Informationsqualität. In: Rainer Kuhlen, Thomas Seeger und Dietmar Strauch (Hg.): Grundlagen der praktischen Information und Dokumentation. Handbuch zur Einführung in die Informationswissenschaft und -praxis. München: Saur, S. 315–321.

TIB - Technische Informationsbibliothek und UB Universitätsbibliothek Hannover (Hg.) (2015): Jahresbericht 2014. Online verfügbar unter https://www.tib.eu/fileadmin/Daten/presse/dokumente/tibub_jahresbericht-2014.pdf, zuletzt geprüft am 02.05.2016.

Vollmert, Christian; Vollmert, Markus; Neuber, Michael (2012): BVDW-Whitepaper. Webanalyse und Datenschutz. Hg. v. Bundesverband Digitale Wirtschaft (BVDW) e.V. Online verfügbar unter <http://www.bvdw.org/mybvdw/media/download/bvdw-whitepaper-webanalyse-datenschutz-aktuell.pdf?file=3501>.

Vollmert, Markus (2010): Webanalyse datenschutzkonform betreiben: Google Analytics anonymisieren. In: *t3n Magazin* (21). Online verfügbar unter <http://t3n.de/magazin/webanalyse-datenschutzkonform-betreiben-google-analytics-225835/>, zuletzt geprüft am 30.05.2016.

Webanalyse Tool (o.J.) (OnPageWiki. Digitales Marketing Lexikon). Online verfügbar unter https://de.onpage.org/wiki/Webanalyse_Tools, zuletzt geprüft am 29.03.2016.

Web Analysis Association (Hg.) (2008): Web Analytics Definitions. Web Analysis Association. New York. Online verfügbar unter

http://www.digitalanalyticsassociation.org/Files/PDF_standards/WebAnalyticsDefinitions.pdf.

Wissenschaftsrat (2011): Übergreifende Empfehlungen zu Informationsinfrastrukturen. Berlin, Köln. Online verfügbar unter <http://www.wissenschaftsrat.de/download/archiv/10466-11.pdf>.

Wissenschaftsrat (2016): Empfehlungen zur Spezifikation des Kerndaten
Empfehlungen zur Spezifikation des Empfehlungen zur Spezifikation des
Kerndatensatz Forschung. Online verfügbar unter
<http://www.wissenschaftsrat.de/download/archiv/5066-16.pdf>.

ZBMED - Leibniz-Informationszentrum Lebenswissenschaften (Hg.) (2015):
Jahresbericht 2013/2014. Online verfügbar unter
http://www.zbmed.de/fileadmin/user_upload/Downloads/ZBMED_Jahresbericht_view.pdf, zuletzt geprüft am 02.05.2016.

Zheng, Jack G.; Peltsverger, Svetlana (2015): Web Analytics Overview. In: Mehdi Khosrow-Pour (Hg.): Encyclopedia of Information Science and Technology: IGI Global, S. 7674-7683. Online verfügbar unter
https://www.researchgate.net/publication/272815693_Web_Analytics_Overview

ZPID - Leibniz-Zentrum für Psychologische Information und Dokumentation (Hg.)
(2015): Tätigkeitsbericht 2014. Online verfügbar unter
<https://www.zpid.de/pub/profil/report2014.pdf>, zuletzt geprüft am 02.05.2016.