

Begriffe

- Hits (jeder Zugriff auf Webseite, CSS-Datei, Grafik, Skript)
- Page Impression (Seitenabruf, kann aus Vielzahl einzelner Hits bestehen)
- Visits (zusammenhängende Besuche / Sessions, kann aus Vielzahl von „Page Impressions“ bestehen)

Logfile-Analyse

- Wie viele Zugriffe (Hits, PI, Visits)
- Von wem (z.B. „reale“ Nutzer / Robots)
- Von wo (z.B. On-Campus, VPN, weltweit)
- Wo war der Nutzer vorher (z.B. Zugriffe über Suchmaschinen-Trefferlisten)
- Statuscodes (z.B. 404-Fehler)

Auszug aus einer Logfile-Datei

```
1372 129.70.22.152 - - [11/May/2011:12:14:14 +0200] "GET /de/index.php HTTP/1.1"  
    200 8938 "-" "Mozilla/5.0 (X11; U; Linux x86_64; de; rv:1.9.2.13)  
    Gecko/20101203 SUSE/3.6.13-0.1.1 Firefox/3.6.13 GTB7.1"  
1373 129.70.22.152 - - [11/May/2011:12:14:14 +0200] "GET  
    /images/menuponits_green.gif HTTP/1.1" 200 59 "http://base.ub.uni-  
    bielefeld.de/css/screen.css" "Mozilla/5.0 (X11; U; Linux x86_64; de;  
    rv:1.9.2.13) Gecko/20101203 SUSE/3.6.13-0.1.1 Firefox/3.6.13 GTB7.1"  
1374 129.70.22.152 - - [11/May/2011:12:14:14 +0200] "GET  
    /images/dreieck_orange_8p.gif HTTP/1.1" 200 844 "http://base.ub.uni-  
    bielefeld.de/css/screen.css" "Mozilla/5.0 (X11; U; Linux x86_64; de;  
    rv:1.9.2.13) Gecko/20101203 SUSE/3.6.13-0.1.1 Firefox/3.6.13 GTB7.1"
```

Auszug aus dem Apache-Logfile des Servers
base.ub.uni-bielefeld.de (Page Impression mit 3 Hits)

Logfile-Eintrag

- 129.70.22.152 - -
- [11/May/2011:12:14:14 +0200]
- „GET /de/index.php HTTP/1.1“
- 200
- 8938
- „http://www.base-search.net/“
- „Mozilla/5.0 (X11; U; Linux x86_64; de; rv:1.9.2.13) Gecko/20101203 SUSE/3.6.13-0.1.1 Firefox/3.6.13“
GTB7.1

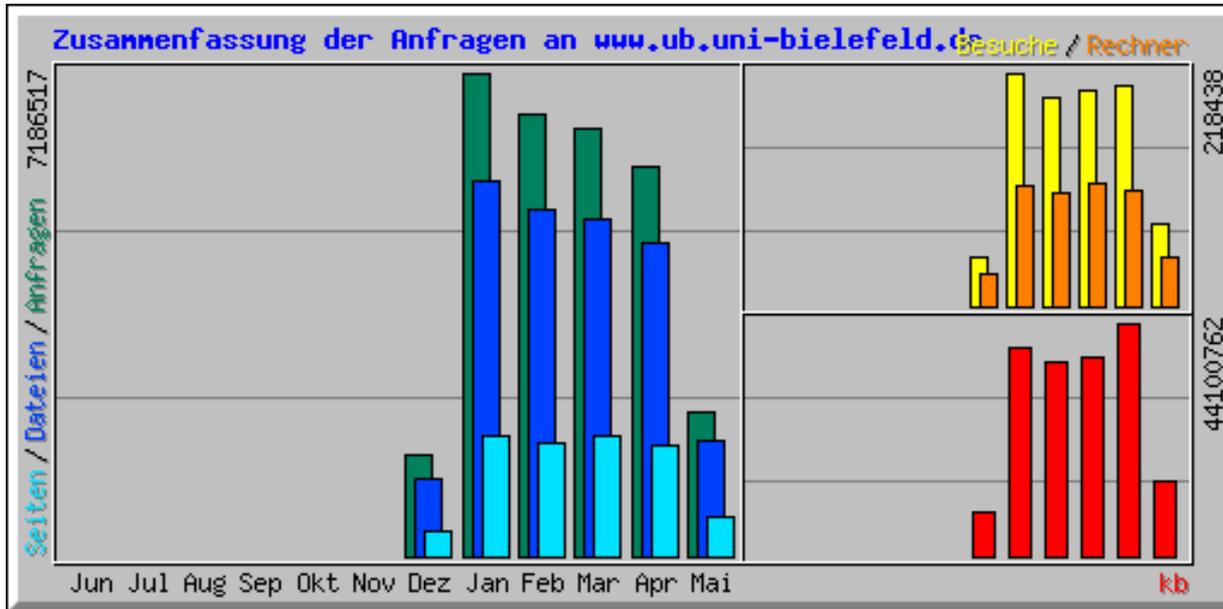
Logfile-Eintrag: Erklärung

- **Wer (IP-Adresse, Benutzername, Passwort)**
- **Wann (Datum + Uhrzeit)**
- **Was (Aktion + Protokoll)**
- **Statuscode (200 = OK, 404 = Seite nicht gefunden)**
- **Wie viel (übertragene Datenmenge in Byte)**
- **Woher (Referrer)**
- **Womit (Browser, Betriebssystem, Oberfläche)**

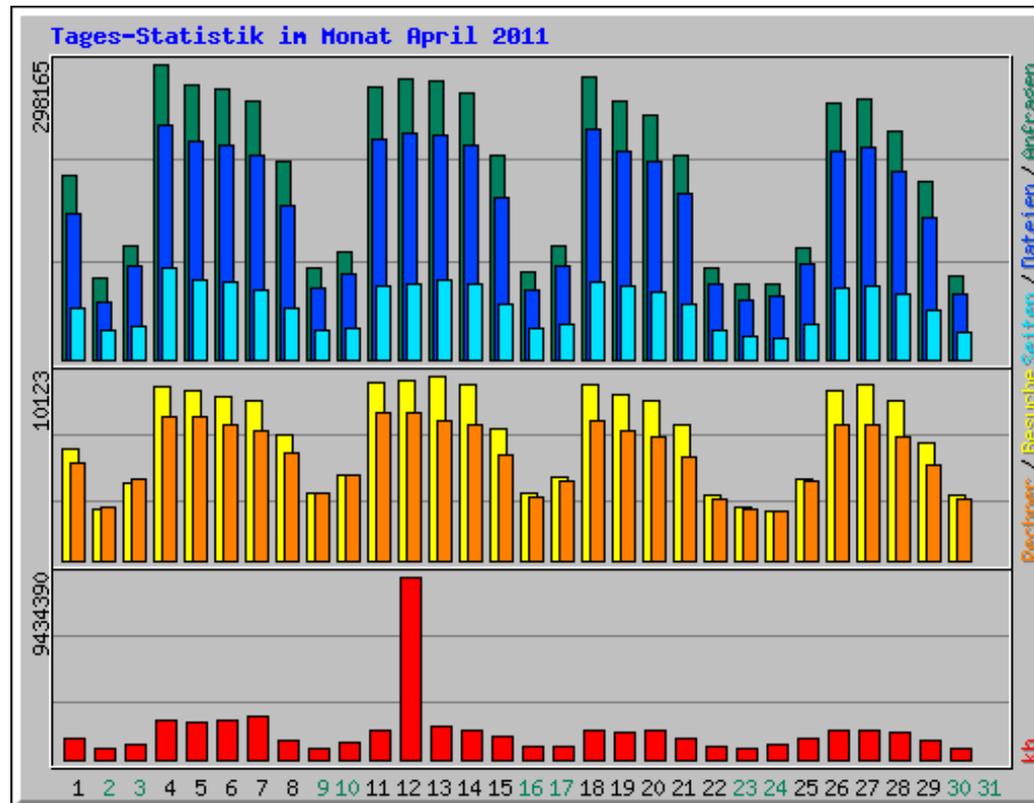
Analyse-Software: Webalizer

- Anzahl der Zugriffe (Hits, PI, Visits) und Statuscodes
- Top-Zugriffe (URLs, Eingangs- und Ausgangsseiten, Hosts, Links, Suchausdrücke, Anwenderprogramme, Länder)
- Stündliche / tägliche / monatliche Auswertung

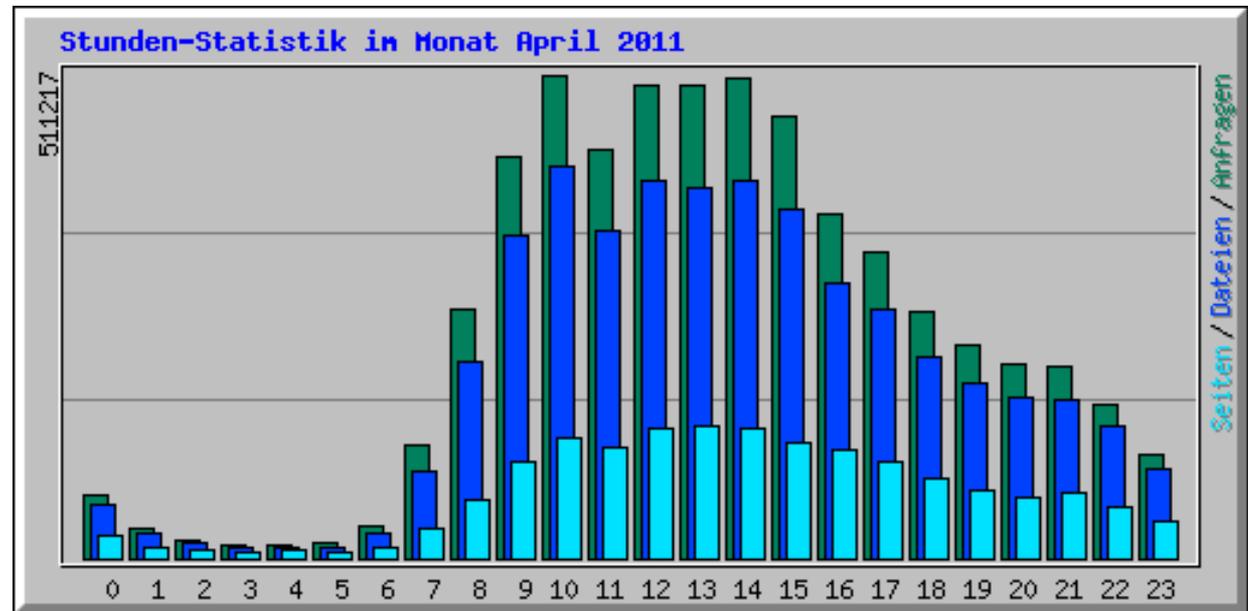
Webalizer: Monatliche Auswertung



Webalizer: Tägliche Auswertung



Webalizer: Stündliche Auswertung



YOURLS

- YOURLS ist ein Kurz-URL-Dienst, der auf unserem Server installiert ist
- Kurz-URLs sind besonders für Dienste wie Twitter
- Verwaltung der URLs inkl. einfacher Statistischer Auswertung

Auswertung YOURURLS

Hello [shorturl](#) ([Logout](#))
 Go to the [Admin Interface](#)
 Check the [Tools](#)
 Read the [Help](#)

YOURURLS: Your Own URL Shortener



Display **1** to **129** of **129** URLs.

Overall, tracking **129** links, **3,710** clicks, and counting!

Enter the URL:

Optional: [Custom short](#)

Short URL	Original URL	Date	IP	Clicks
ibase	http://itunes.apple.com/en/app/id430576014?mt=8&ign-mpt[...]	May 02, 2011 14:56	129.70.22.152	208
pubbio	http://pub.uni-bielefeld.de/pub?dept=10017&ttyp=bio4&st[...]	Jan 17, 2011 12:58	129.70.22.152	156
basefq	http://base.ub.uni-bielefeld.de/de/faq.php#chap02	Mar 16, 2011 09:54	129.70.22.152	151
basehp	http://ekvv.uni-bielefeld.de/blog/uniaktuell/entry/frei[...]	May 03, 2011 09:23	129.70.22.152	144

Auswertung YOURURLS

Information

Short URL: <http://ub.unibi.de/pubbgghs>

Long URL: <http://pub.uni-bielefeld.de/pub?dept=6747229&maxrecs=10&ttyp=bgghs&style=mla&func=search>

Traffic statistics

Traffic location

Traffic sources

Share

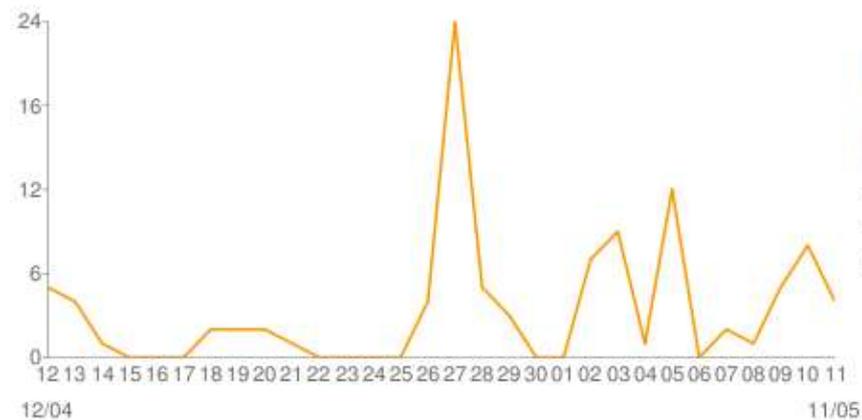
Last 24 hours

Last 7 days

Last 30 days

All time

Number of hits : Last 30 days



Historical click count

Short URL created on April 7, 2011 @ 4:00 pm (about 34 days ago)

Last 24 hours	6 hits	0.25 per hour
Last 7 days	32 hits	4.57 per day
Last 30 days	102 hits	3.4 per day
All time	116 hits	3.41 per day

Best day

24 hits on April 27, 2011. [Click for more details](#)

Auswertung YOURURLS

Information

Short URL:  <http://ub.unibi.de/pubbghs>

Long URL:  <http://pub.uni-bielefeld.de/pub?dept=6747229&maxrecs=10&ttyp=bghs&style=mla&func=search>

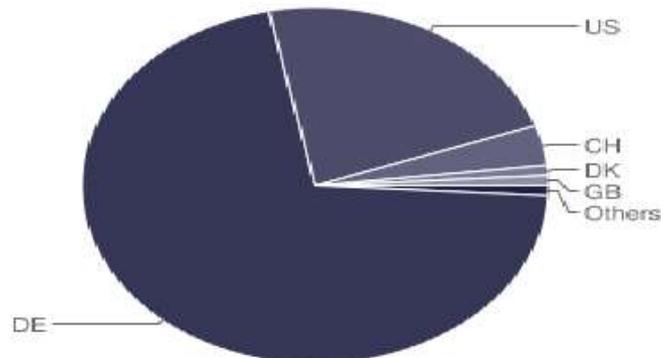
Traffic statistics

Traffic location

Traffic sources

Share

Top 5 countries



Overall traffic



Eigenes Skript

- Apache-Logfiles (Webserver) werden über ein eigenes Perlskript ausgewertet
- Auswertung nach Gruppen (On-/Off-Campus, Suchmaschinen) möglich
- Gemeinsame Auswertung für mehrere Dateien / einzelne Ordner möglich

Konfigurationsdatei für eigenes Skript

```
/GROUPS
!(38\.98\.19\.|38\.99\.|38\.114\.104\.|61\.247\.|63\.123\.238\.|64\.124\.85\.|64\.202\.165\.|64\.246\.|65\.54\.|65\.55\.|
65\.214\.44\.|65\.59\.219\.|65\.214\.45\.|66\.196\.|66\.231\.118\.|66\.231\.188\.|66\.249\.|67\.19\.114\.|68\.142\.|
68\.180\.|69\.50\.210\.|70\.42\.51\.|70\.84\.212\.|70\.85\.113\.|72\.30\.|72\.36\.|74\.6\.|74\.53\.243\.|74\.53\.249\.|
74\.86\.|75\.125\.|75\.126\.|76\.208\.|80\.86\.83\.|83\.133\.126\.|122\.152\.|129\.70\.12\.27|202\.96\.51\.|202\.160\.|
202\.165\.|208\.96\.54\.|208\.101\.44\.|209\.62\.109\.|212\.67\.202\.|213\.29\.7\.|216\.40\.222\.|217\.111\.100\.|
217\.212\.224\.)
!(dhcp11|dhcp22|uni-bielefeld\.de|129\.70\.11\.|129\.70\.22\.)
!(.*\.uni-bielefeld\.de|129\.70\.)
/GROUPS_END
/RULES
Startseite (S)= / .*HTTP
Startseite (S)= /index\.htm .*HTTP
Startseite (S)(E)= /english/ .*HTTP
Startseite (S)(E)= /english/index\.htm .*HTTP
Aktuelles (S)= /aktuell/ .*HTTP
Aktuelles (S)= /aktuell/index\.htm .*HTTP
Aktuelles (O)= /aktuell/ .*HTTP
Aktuelles (S)(E)= /english/aktuell/ .*HTTP
Aktuelles (S)(E)= /english/aktuell/index\.htm .*HTTP
Newsletter (S)= /news/ .*HTTP
Newsletter (S)= /news/index\.htm .*HTTP
Newsletter (O)= /news/ .*HTTP
RSS-Feed (S)= /rss/ .*HTTP
RSS-Feed (S)= /rss/index\.htm .*HTTP
A-Z= /index/abisz/ .*HTTP
A-Z= /index/abisz/index\.htm .*HTTP
A-Z (E)= /english/index/abisz/ .*HTTP
A-Z (E)= /english/index/abisz/index\.htm .*HTTP
Recherche (S)= /search/ .*HTTP
Recherche (S)= /search/index\.htm .*HTTP
Recherche (S)(E)= /english/search/ .*HTTP
```

Auswertung (Dezember 2008)

Content-Type: text/html

[Startseite](#) » [Intern](#) » [Statistiken](#) » UB-Webseiten

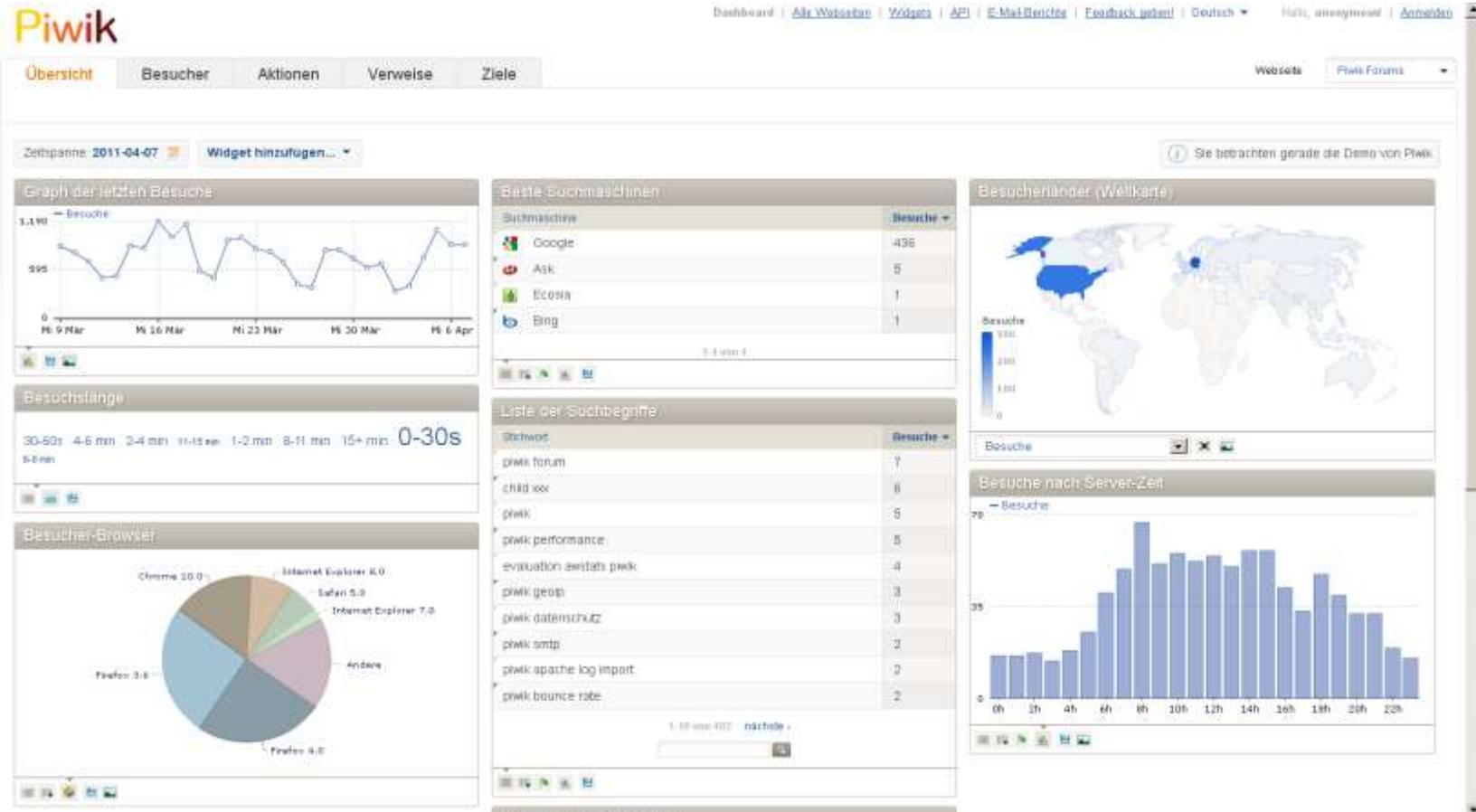
01_Jan_2009_31_Dec_2008

URLs	Gesamt	Robot	Universitätsbibliothek	Universität	extern
0-Treffer-Recherchen	23532	2	1134	11733	10663
A-Z dt.	1959	85	359	610	905
Aktuelles Newsletter (Uebersicht) dt.	354	117	8	36	193
Aktuelles Newsletter (weitere Seiten) dt.	4336	1735	81	117	2403
Aktuelles RSS-Feed (Startseite) dt.	122	47	5	2	68
Aktuelles RSS-Feed (XML-Datei) dt.	80555	119	2509	71321	6606
Aktuelles dt.	2991	115	350	751	1775
Anreise und Oeffnungszeiten dt.	1642	55	49	217	1321
Anschaffungsvorschlaege dt.	408	37	26	165	180
Ausleihkonto (Info-Seite) dt.	1253	25	23	191	1014

Piwik

- Piwik ist ein Open-Source-Programm für Webanalytik
- Freie Alternative zu Google Analytics
- Datenschutzrechtlich weniger bedenklich als GA, da die Daten lokal gespeichert werden und teil-anonymisiert werden können

Auswertung Piwik



Probleme bei der Auswertung

- Statistiken unterscheiden nicht immer zwischen menschlichen Nutzern und Robots
- Zählung von Downloadmanagern als normale Zugriffe
- Keine Bereichsauswertung (On-/Off-Campus), sondern nur nach Ländern
- Vergleichbarkeit über längeren Zeitraum nicht immer gewährleistet

Rechtliche Rahmenbedingungen

- § 12, Abs. 1 Telemediengesetz:
„Der Diensteanbieter darf **personenbezogene Daten** zur Bereitstellung von Telemedien nur erheben und verwenden, soweit dieses Gesetz oder eine andere Rechtsvorschrift, die sich ausdrücklich auf Telemedien bezieht, es erlaubt oder **der Nutzer eingewilligt hat.**“

Quelle: http://bundesrecht.juris.de/tmg/___12.html

Rechtliche Rahmenbedingungen

- § 15, Abs. 1 Telemediengesetz:
„Der Diensteanbieter darf personenbezogene Daten eines Nutzers **nur erheben und verwenden**, soweit dies erforderlich ist, um die **Inanspruchnahme von Telemedien zu ermöglichen und abzurechnen**“

Quelle: http://bundesrecht.juris.de/tmg/___15.html

Rechtliche Rahmenbedingungen

- § 15, Abs. 4 Telemediengesetz:
„Der Diensteanbieter darf Nutzungsdaten **über das Ende des Nutzungsvorgangs** hinaus verwenden, soweit sie für Zwecke der Abrechnung mit dem Nutzer erforderlich sind (Abrechnungsdaten). Zur Erfüllung bestehender gesetzlicher, satzungsmäßiger oder vertraglicher Aufbewahrungsfristen darf der Diensteanbieter die Daten sperren.“

Rechtliche Rahmenbedingungen

- IP-Adresse wird von verschiedenen Gerichten als personenbezogenes Datum angesehen, ist aber **rechtlich umstritten** (dynamische IPs, Zugriffe über Proxies)
- Bisher keine Entscheidung des Bundesgerichtshofs oder EuGH
- Protokollierung der IP-Adressen **kann** unzulässig sein, wenn Einspruch erhoben wird

Rechtliche Rahmenbedingungen

- Lösung 1: Vor dem ersten Zugriff auf eine Website muss der Nutzer seine Zustimmung geben, dass seine Daten protokolliert werden (nicht praktikabel)
- Lösung 2: IP-Adresse Teil-Anonymisieren (Analyse noch möglich, aber Zuordnung zu einem bestimmten PC nicht mehr möglich)