

# **Zwischen Client und Server / technische Aspekte des Datentransfers bei der Webarchivierung**

Sebastian Vesper, FES  
(sebastian.vesper@fes.de)

„Webarchivierung“  
Gemeinsamer Workshop von nestor und dem AWW Arbeitskreis 6.2 zur  
Webarchivierung

20. März 2012, Deutsche Nationalbibliothek, Frankfurt

# Protokolle für Datenübertragungen

- HTTP (Hypertext Transfer Protocol)
- FTP/SFTP (File Transfer Protocol / Secure File Transfer Protocol)

Eigenschaften u.a.:

- Implementierung auf Anwendungsebene
- TCP/IP (Transmission Control Protocol / Internet Protocol) basiert
- Request/Response-Paradigma folgend

# Divergenz von Daten und ihrer Präsentation

FTP: Zugriff auf serverseitig gespeicherte Daten.

Problematiken:

- systemspezifische Daten
- Quelltexte (z.B. von Skriptsprachen)
- ungenutzte, nicht publizierte Daten
- Unvollständigkeit durch unzureichenden Zugriff auf Daten aus Drittsystemen (Datenbanken, Stagesysteme, CMS, verteilte Systeme)

Die Bereitstellung von Inhalten durch derart transferierte Daten mit hohem Aufwand verbunden (unterschiedliche Plattformen/Systeme, Laufzeitumgebungen)

# Datentransfer per HTTP

- Identisch mit der Perspektive üblicher Clients (Webbrowser)
- Transfer der publizierten und referenzierten Inhalte und ihrer Struktur
- Gebräuchliche Methode in vorhandener Software und Funktionsbibliotheken
- Zusätzliche Metainformationen werden übermittelt

# Struktur des HTTP Datentransfers

HTTP-Anfragen und Antworten bestehen aus einem Header (Nachrichtenkopf) und einem Body (Nachrichtenkörper)

```
GET / HTTP/1.0
```

```
HTTP/1.1 200 OK
Date: Tue, 13 Mar 2012 06:23:23 GMT
Server: Apache/2.2.9
Last-Modified: Fri, 12 Feb 2010 18:20:50 GMT
ETag: "8b116-3286-47f6b53bd8480"
Accept-Ranges: bytes
Content-Length: 12934
Vary: Accept-Encoding
Connection: close
Content-Type: text/html
```

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
 "http://www.w3.org/TR/html4/loose.dtd">
<html lang="de">
```

# Auswahl der im Header enthaltenen Informationen

Interessant insbesondere die serverseitigen Header.  
Zusätzliche Informationen bieten

- der numerische Statuscode
  - Im 200er und 300er Bereich im Erfolgsfall
  - 400er und 500er Statuscodes im Fehlerfall
- der "Last-Modified"-Header über das letzte Änderungsdatum des Inhalts in Verbindung mit dem
- "Date"-Header (abweichende Systemzeiten und Zeitpunkt des Transfers)
- der "Content-Length"-Header zwecks Vergleichsmöglichkeit
- "Content-Type"-Header für die Art der Daten

# Header Beispiele

## Weiterleitung der Anfrage mittels 302 Statuscode und "Location"-Header

```
GET / HTTP/1.1
Host: www.dnb.de

HTTP/1.1 302 Moved Temporarily
Date: Tue, 13 Mar 2012 09:45:21 GMT
Server: Apache
Vary: Host
Cache-Control: no-cache
Location: http://www.dnb.de/DE/Home/home_node.html
Content-Language: de
Content-Length: 0
Content-Type: text/html;charset=utf-8
```

# Header Beispiele

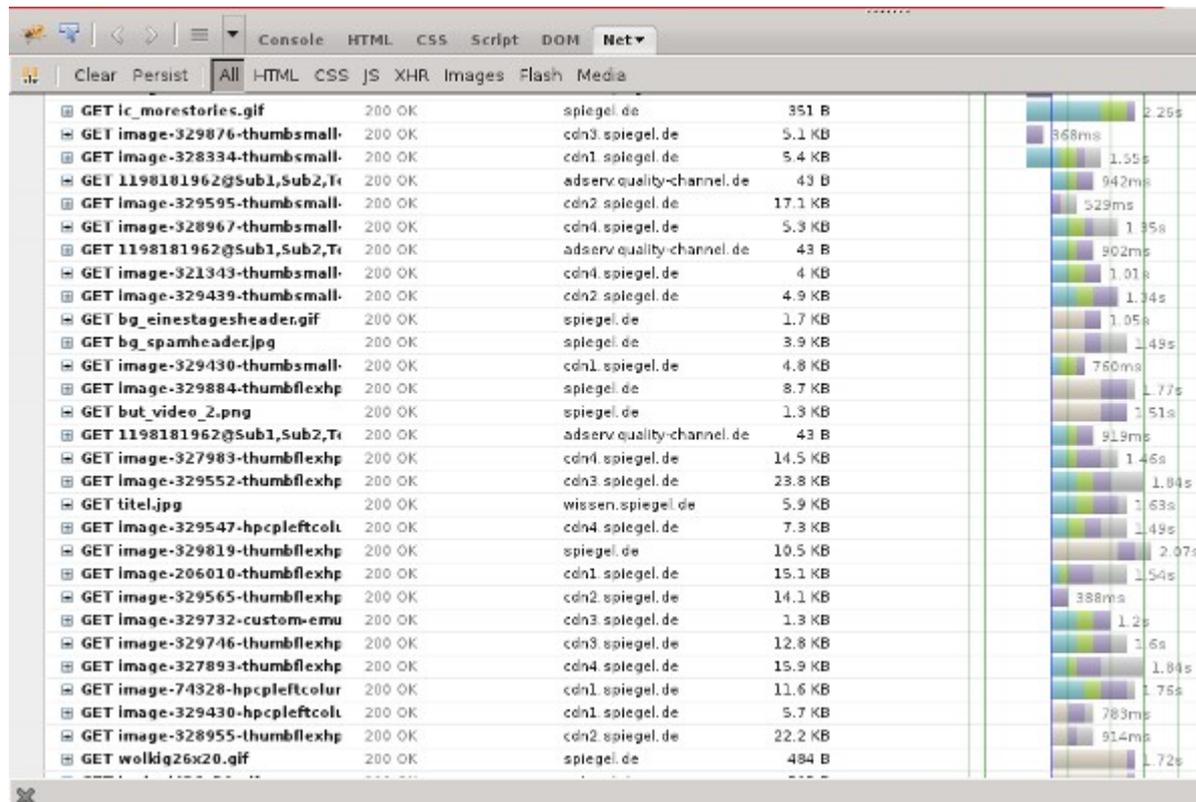
## Date, Last-Modified, Content-Length und Content-Type

```
HTTP/1.1 200 OK
Date: Tue, 13 Mar 2012 06:23:23 GMT
Server: Apache/2.2.9
Last-Modified: Fri, 12 Feb 2010 18:20:50 GMT
ETag: "8b116-3286-47f6b53bd8480"
Accept-Ranges: bytes
Content-Length: 12934
Vary: Accept-Encoding
Connection: close
Content-Type: text/html

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
<html lang="de">
```

# Visualisierungsmöglichkeiten des Datentransfers

- Firebug (getfirebug.com) Plugin für den Browser Firefox
- Google Chrome



Vielen Dank.