



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Bundesamt für Kultur BAK
Schweizerische Nationalbibliothek NB

Glossar e-Helvetica

Version:

Version 1.3

Datum:

05.02.2016

Inhaltsverzeichnis

Access	3
Administration	3
Amtliche Schriften	3
Archival Storage	3
Archivserver	3
Bretter	3
Cache	3
Data Management	3
Dateiformat	3
Datenbanken	3
Datenträger	4
Digitale Publikationen (born-digital)	4
Digitalisierte Publikationen (digitized)	4
Digitalisierung	4
DOI (Digital Object Identifier)	4
Dokumentenserver	4
Domain	4
Dynamische Websites	4
e-Helvetica	4
e-Helvetica Access	4
e-Helvetica Deposit	5
Elektronische Publikationen	5
E-Mail	5
Emulation	5
FTP (File Transfer Protocol)	5
Harvesting	5
Helvetica	5
Helveticat	5
Hochschulschriften	5
Homepage	5
HTTP (Hypertext Transfer Protocol)	5
Hyperlink	6
Ingest	6
Inkrementelles Backup	6
Internet	6
Intranet	6
JDBC (Java Database Connectivity)	6
Langzeitarchivierung	6
Langzeitverfügbarkeit	6
MARC	7
Metadaten	7
METS (Metadata Encoding & Transmission Standard)	7
Migration	7
MODS (Metadata Object Description Schema)	7
NAS (Network Attached Storage)	7
NFS (Network File System)	7
Ninive	8
OAI (Open Archives Initiative)	8
OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting)	8
OAIS (Open Archival Information System)	8
Online-Publikationen	8
PDF	8
Persistent Identifier	8
Preservation Planning	8
Proprietäres Datenformat	8
robots.txt	9
SFTP (Secure File Transfer Protocol)	9
Statische Websites	9
Tarball	9
URN (Uniform Resource Name)	9
Usenet (Newsgroups)	9
Web Harvesting	9
WebDAV	10
Weblogs (Blogs)	10
Webseite	10
Website (Site)	10
wget	10
World Wide Web (WWW)	10
XML (Extensible Markup Language)	10

Access

Prozess im OAIS-Modell: Bereitstellung der Archivdaten zur Benutzung.

Administration

Prozess im OAIS-Modell: Überwachung der Arbeitsvorgänge im Archivierungssystem.

Amtliche Schriften

Amtsschriften sind gedruckte oder auf anderen Informationsträgern gespeicherte Informationen, die von einer Verwaltungseinheit der zentralen Bundesverwaltung veröffentlicht werden.

Archival Storage

Prozess im OAIS-Modell: Speicherung der digitalen Objekte im Archivierungssystem.

Archivserver

Als Archivserver wird eine technische und organisatorische Infrastruktur bezeichnet, deren Aufgabe vorrangig die Speicherung, Verwaltung und Langzeitarchivierung digitaler Objekte ist. Systeme des Typs Archivserver werden im Kontext der Langzeitarchivierung getrennt von Systemen des Typs Dokumentenserver betrachtet, um den jeweiligen unterschiedlichen Schwerpunkten der Funktionalität gerecht zu werden.

Bretter

Bretter sind bestimmte Seiten in einer Mailbox (Bulletin Board System) oder Newsgroup, in denen die Nutzer/innen – vergleichbar mit einem Aushang am schwarzen Brett – Informationen austauschen können.

Cache

Der Cache ist ein schneller Pufferspeicher, d.h. ein Speicher, auf dessen Inhalte besonders schnell zugegriffen werden kann, und in dem mehrfach benötigte Daten zeitweilig abgelegt sind.

Data Management

Prozess im OAIS-Modell: Verwaltung und Unterhalt der Archivdaten über bibliografische, technische und administrative Metadaten.

Dateiformat

Um Daten zu einer Datei zusammenzufassen, bedarf es bestimmter Regeln. Eine Gruppe solcher Regeln bildet ein Dateiformat. Dateiformate können sehr einfach sein, indem sie etwa die bloße Aneinanderreihung der Daten vorschreiben. Sie können aber auch die Speicherung bestimmter zusätzlicher Informationen an vorgegebenen Stellen innerhalb der Datei erzwingen. Ebenso können sie eine ganz bestimmte Kodierung der Daten und der zusätzlich gespeicherten Informationen verlangen. Bis zur Schaffung eines geeigneten Archivformats ist die Wahl des Dateiformats grundlegend für die Langzeitarchivierung digitaler Daten. Sinnvoll ist die Wahl weit verbreiteter, von vielen Programmen verwendeter, Dateiformate (z.B. TIF für Bilddateien) - denn hier ist die Wahrscheinlichkeit gross, dass auch in absehbarer Zukunft Programme zur Verfügung stehen, mit denen die Informationen aus den Dateien zurückgewonnen werden können. Positiv auf die Langzeitarchivierung wirkt weiter die Wahl einfacher Dateiformate (z.B. TXT für Textdateien), die eventuell später notwendige Rekonstruktionen der Regeln, nach denen die Daten zusammengefasst sind, erleichtern. Ungeeignet sind hingegen Dateiformate, die im Besitz eines einzelnen Herstellers sind (z.B. DOC für Textdateien), der sie nach Belieben verändern oder ihre Benutzung einschränken kann. Häufig erlaubt die Dateieindung Rückschlüsse auf das Dateiformat.

Datenbanken

Datenbanken sind Systeme zur Speicherung und Verwaltung von grossen Datenmengen. Sie bestehen aus einer Datenbasis, in der die Daten nach einer bestimmten Struktur abgelegt werden, und Verwaltungsprogrammen, die Daten abspeichern, suchen oder andere Operationen mit den Daten durchführen. Meist wird auch schon ein Anwendungsprogramm zur strukturierten Organisation und Verwaltung von Daten als Datenbank bezeichnet. Die Inhalte der Datenbanken werden durch interaktive (zumeist

proprietäre) Benutzerschnittstellen erschlossen. Dabei werden von Benutzer/innen individuelle Suchanfragen übergeben und es erfolgt eine fallweise Aufbereitung individueller Ergebnismengen.

Datenträger

Maschinell oder elektronisch verarbeitbares Speichermedium, auf dem digitale Daten dauerhaft gespeichert und wieder ausgelesen werden können. Beispiele sind magnetische (Disketten, Magnetbänder) und optische Datenträger (CD-ROM, DVD).

Digitale Publikationen (born-digital)

Digitale Publikationen sind original elektronische Publikationen. Sie werden digital hergestellt und existieren nur in digitaler Form.

Digitalisierte Publikationen (digitized)

Digitalisierte Publikationen sind Publikationen, die von analoger (z.B. gedruckter Form) in digitale Form umgewandelt werden.

Digitalisierung

Für die Digitalisierung analoger Objekte werden umfangreiche Investitionen getätigt. Ziel dabei ist es vorrangig, den Nutzerzugriff auf Inhalte zu erleichtern bzw. den Inhalt der vom Zerfall bedrohten Objekte zu erhalten. Massnahmen zur Langzeitarchivierung von Digitalisaten sind erforderlich, um die getätigten Investitionen für die Zukunft zu sichern.

DOI (Digital Object Identifier)

Der Digital Object Identifier (DOI) erlaubt eine eindeutige und permanente Identifikation digitaler Objekte. Das System des DOI ist denen von ISBN und ISSN vergleichbar, geht jedoch durch seine integrierte Lokalisierungsfunktion darüber hinaus.

Dokumentenserver

Als Dokumentenserver wird ein technisches und organisatorisches System bezeichnet, dessen Schwerpunkt die endnutzerorientierte Bereitstellung von digitalen Dokumenten (bzw. dokumentähnlichen Objekten) ist. Ein Dokumentenserver kooperiert ggf. mit einem Archivserver oder Depotsystem, um die Langzeitverfügbarkeit der bereitzustellenden Objekte zu gewährleisten.

Domain

Ein Domain (eine Domäne) ist ein Bereich, der durch gemeinsame Merkmale gekennzeichnet ist; meist versteht man darunter eine Gruppe von Computern mit gemeinsamen Namensbestandteilen (Host-Name). Der kleinste gemeinsame Namensbestandteil heisst Top-Level-Domain, z.B. das Länderkürzel ".ch" oder das Kürzel ".com". Domänen können in Unterdomänen eingeteilt werden, die man Second-Level-Domain, Third-Level-Domain usw. nennt.

Dynamische Websites

Im bibliothekarischen Kontext sind dynamische Websites zum Zeitpunkt ihrer Erstveröffentlichung nicht abgeschlossen. Sie unterliegen inhaltlichen Änderungen und Ergänzungen unterschiedlicher Qualität und Häufigkeit (Integrierende Ressourcen).

e-Helvetic

e-Helvetic setzt sich zusammen aus dem "e" für elektronisch und "Helvetica" für Publikationen mit Bezug zur Schweiz.

Mit „e-Helvetic“ wird auch der Dienst der NB bezeichnet, der für den Aufbau der digitalen Sammlung sowie für deren Erschliessung, Langzeitarchivierung und -verfügbarkeit zuständig ist.

e-Helvetic Access

e-Helvetic Access ist die Schnittstelle, über die Sie die digitalen Sammlungen der Schweizerischen Nationalbibliothek (NB) abfragen können. Diese Schnittstelle ermöglicht die Volltextsuche im Inhalt der digitalen Publikationen.

In e-Helvetica Access finden Sie digitale Publikationen, die sich in den Beständen der NB befinden. Zurzeit werden in e-Helvetica Access original digitale Bücher, Zeitschriften, Hochschulschriften und Amtsschriften sowie Websites mit einem Bezug zur Schweiz verzeichnet. Die Sammlung befindet sich im Aufbau. Neben der NB sind Partnerinstitutionen daran beteiligt. Weiter finden Sie in e-Helvetica Access Publikationen, die von der NB nachträglich digitalisiert worden sind.

e-Helvetica Deposit

e-Helvetica Deposit ist eine Plattform für Verlage zur Ablieferung einzelner elektronischer Bücher.

Elektronische Publikationen

Unter elektronischen Publikationen werden sowohl Online- wie auch Offline-Publikationen verstanden. Der Begriff wird gleichbedeutend mit digitalen Publikationen verwendet.

E-Mail

Eine E-Mail (Electronic Mail) ist eine Mitteilung (Text, Dateien), die mit dem Dienst "Simple Mail Transfer Protocol" (SMTP) zwischen Absender/innen und Empfänger/innen in einem Netzwerk übertragen wird.

Emulation

Strategie zur Erhaltung der Langzeitverfügbarkeit digitaler Objekte. Die Strategie sieht vor, dass die Systemvoraussetzungen, die zur Nutzung älterer digitaler Objekte notwendig sind, durch spezialisierte Software auf aktuellen marktgängigen Systemen nachgebildet (emuliert) werden können. Die digitalen Objekte selbst werden dabei möglichst unverändert erhalten. Verschiedene z.T. konkurrierende Emulationsstrategien (Hardware- versus Software-Emulation) werden z.Zt. diskutiert.

FTP (File Transfer Protocol)

FTP ist ein Protokoll für die Übermittlung von Dateien im Internet. Auch der Internetdienst, der auf diesem Protokoll basiert, wird FTP genannt.

Harvesting

Als Harvesting, auch Webharvesting genannt, wird ein automatisches Einsammeln von Internetseiten mittels eines Harvesters (Roboter) bezeichnet.

Helvetica

Publikationen, die in der Schweiz erscheinen, sich auf die Schweiz oder auf Personen mit schweizerischem Bürgerrecht oder Wohnsitz beziehen oder von schweizerischen oder mit der Schweiz verbundenen Autor/innen geschaffen oder mitgestaltet werden. Sie entsprechen dem Sammelgebiet der Schweizerischen Nationalbibliothek.

Helveticat

Bibliothekskatalog der Schweizerischen Nationalbibliothek.

Hochschulschriften

Unter Hochschulschriften versteht man Dissertationen und Habilitationsschriften, die an einer Universität publiziert werden.

Homepage

Die Homepage ist die Eingangs- oder Startseite eines Anbieters von Webseiten im World Wide Web. Den Internetauftritt einer Privatperson durch eine eigene Website bezeichnet man ebenfalls als Homepage. Damit ist also, abweichend von der ursprünglichen Definition, nicht nur die Eingangsseite des Angebots gemeint, sondern die Gesamtheit des angebotenen Materials.

HTTP (Hypertext Transfer Protocol)

HTTP ist ein Übertragungsprotokoll für die Übertragung von Daten im WWW, das auf TCP/IP als Transportprotokoll aufbaut.

Hyperlink

Der Hyperlink, oder kurz Link, ist die Verknüpfung zwischen einem Element in einem Hypertext oder auf einer Internetseite mit einem anderen Element auf der gleichen Seite oder mit einer anderen Seite. Der Link wird dargestellt durch Unterstreichung und farbliche Hervorhebung des Begriffs, von dem der Link ausgeht, manchmal auch durch eine Grafik (ein Symbol).

Ingest

Prozess im OAIS-Modell: Übernehmen der Daten vom Produzenten, Aufbereitung für die Archivierung und Einspeisen in das Speichersystem für die Langzeitarchivierung.

Inkrementelles Backup

Ein inkrementelles Backup basiert auf einem vollständigen Backup eines Quellsystems. Das inkrementelle Backup sichert in einem zeitlich versetzten Schritt nur die neu hinzugefügten Daten. Für eine erfolgreiche Wiederherstellung wird das erste, volle Backup als Basis verwendet. Auf diese Basis werden dann die inkrementellen Backups in chronologisch aufsteigender Reihenfolge auf das System zurückgespeichert.

Internet

Das Internet ist das weltweit grösste Computernetzwerk, das aus vielen miteinander verbundenen Netzwerken besteht und eine vielfältige Kommunikationsinfrastruktur zur Verfügung stellt. Zu den Leistungen des Internets – man spricht auch von Diensten – gehören: E-Mail, die elektronische Post; World Wide Web (WWW), auf Hypertext basierendes Dokumentensystem; Usenet, ein Diskussionsforum; FTP, System zum Übertragen beliebiger Dateien; IRC, Chat; Gopher, Telnet, Wais, Archie, einige alte Dienste von nachlassender Bedeutung. Für viele Privatanwender ist das WWW der bedeutendste Beitrag, daher wird dieser Bestandteil oft mit dem Internet gleichgesetzt.

Intranet

Das Intranet ist ein privates (unternehmenseigenes) Netzwerk, das mit der Technologie des Internets arbeitet. Im Unterschied zum Internet steht das Intranet aber nur einem begrenzten und lokalen Benutzerkreis zur Verfügung.

JDBC (Java Database Connectivity)

Java Database Connectivity (JDBC) ist eine Datenbankschnittstelle der Java-Plattform, die eine einheitliche Schnittstelle zu Datenbanken verschiedener Hersteller bietet und speziell auf relationale Datenbanken ausgerichtet ist. JDBC ist in seiner Funktion als universelle Datenbankschnittstelle vergleichbar mit z.B. ODBC unter Windows oder DBI unter Perl. Zu den Aufgaben von JDBC gehört es, Datenbankverbindungen aufzubauen und zu verwalten, SQL-Anfragen an die Datenbank weiterzuleiten und die Ergebnisse in eine für Java nutzbare Form umzuwandeln und dem Programm zur Verfügung zu stellen.

Langzeitarchivierung

Unter Langzeitarchivierung versteht man die Erfassung und langfristige Aufbewahrung von Informationen. Vor allem bei der Langzeitarchivierung digital vorliegender Informationen (digital preservation) stellen sich neue Probleme.

Langzeitverfügbarkeit

Erhaltung der Langzeitverfügbarkeit digitaler Objekte umfasst alle Massnahmen, die dazu dienen, digitale Objekte für unsere Nachwelt benutzbar zu erhalten. Mit "unserer Nachwelt" ist eine unbegrenzt in die Zukunft reichende Zeitspanne gemeint, in der zu rechnen ist mit

- tiefgreifenden technologischen Veränderungen bei der Speicherung und dem Zugriff auf digitale Objekte
- damit einhergehender Entwicklung neuer und Marktverdrängung alter Datenformate und Datenträger sowie
- einschneidenden Veränderungen im Benutzerverhalten und der Art der auftretenden Informationsinhalte.

Massnahmen zur Erhaltung der Langzeitverfügbarkeit sind weiter zu untergliedern in:

- Massnahmen zur Substanzerhaltung des Datenstroms der digitalen Objekte (z.B. durch Refreshing)

- Massnahmen zur Erhaltung der Benutzbarkeit (z.B. durch Emulation oder Migration).

MARC

MARC (Machine Readable Cataloging) ist ein weit verbreiteter Standard für die Darstellung und den Austausch von Daten für bibliografische, Autoritäts-, Bestands-, Klassifikations- und Bürgerinformationsdaten in maschinenlesbarer Form.

Metadaten

Metadaten sind Daten über Daten. D.h. Metadaten ist eine Bezeichnung für Daten, die andere Daten oder Objekte beschreiben. Es geht also um Informationen über Daten, die es erlauben, auf die gewünschten Daten möglichst effizient zuzugreifen, diese auszutauschen und zu verwalten. Diese Grundinformationen beinhalten z.B. Angaben wie Autor oder Autorin des Dokumentes, Zeitpunkt der Veröffentlichung, Verweise auf Dokumente zum Thema usw. Unter Metadaten fallen also die Katalogisate von Publikationen. Im elektronischen Bereich sind in den Metadaten nicht nur bibliografische Angaben eingeschlossen, sondern auch technische und administrative Informationen (Fileformat, Filegrösse, Übernahmedatum usw.).

METS (Metadata Encoding & Transmission Standard)

METS ist ein mit XML-Schema definiertes XML-Format zur Beschreibung von digitalen Sammlungen von Objekten mit Metadaten. Das Format der Metadaten (MAB, MARC, Dublin Core etc.) kann verschieden sein und wird nicht von METS festgelegt. Stattdessen enthält METS Elemente zur Gruppierung von Objekten und ihre Verbindung mit deskriptiven und administrativen Metadaten. Mit METS lassen sich beispielsweise hierarchische und anderweitig strukturierte Werke kodieren, wobei ein Werk eine Buchreihe, ein einzelnes Buch mit Kapitel- und Seitenstruktur oder ein Film mit mehreren Szenen sein kann. Zur Verlinkung von METS-Dateien und den digitalen Objekten, über die Aussagen getroffen werden sollen, wird eine Untermenge des XLink-Schemas verwendet.

Migration

Strategie zur Erhaltung der Langzeitverfügbarkeit. Die zu erhaltenden digitalen Objekte werden, folgend den sich verändernden technischen Umgebungsbedingungen, Prozeduren unterzogen, die ihre Benutzbarkeit auch unter den neuen Bedingungen gewährleisten sollen.

MODS (Metadata Object Description Schema)

MODS steht für Metadata Object Description Schema und ist ein Metadaten-Standard der Library of Congress. MODS basiert auf einem XML-Schema, das dazu dient, bibliografische Daten zu beschreiben.

NAS (Network Attached Storage)

Network Attached Storage (NAS) bezeichnet einfach zu verwaltende Dateiserver. Allgemein wird NAS eingesetzt, um ohne hohen Aufwand unabhängige Speicherkapazität in einem Rechnernetz bereitzustellen. Ein NAS ist im Unterschied zu Direct Attached Storage ein eigenständiger Host mit eigenem Betriebssystem. Da die Funktionen sehr spezifisch auf den Einsatzzweck abgestimmt sind, werden Fehler aufgrund umfangreicher und für den speziellen Einsatzzweck unnötiger Konfigurationsmöglichkeiten von vornherein vermieden. Dateibasierende Dienste wie NFS oder SMB/CIFS stellen die Kernfunktion dar. Daher werden NAS-Systeme oft kurz als Filer bezeichnet. Eine handelsübliche Festplatte, die sich in einem externen Gehäuse befindet, das mit einer „RJ-45“-Netzwerkbuchse, entsprechender Firmware und optional weiteren Anschlussmöglichkeiten ausgestattet ist, entspricht bereits einem NAS.

NFS (Network File System)

Das Network File System – abgekürzt NFS (auch: Network File Service) – ist ein von Sun Microsystems entwickeltes Protokoll, das den Zugriff auf Dateien über ein Netzwerk ermöglicht. Dabei werden die Dateien nicht wie z. B. bei FTP übertragen, sondern die Benutzer können auf Dateien, die sich auf einem entfernten Rechner befinden, so zugreifen, als ob sie auf ihrer lokalen Festplatte abgespeichert wären.

Ninive

Ninive ist das Speichersystem der Schweizerischen Nationalbibliothek für die Langzeitarchivierung.

OAI (Open Archives Initiative)

Normierte Schnittstelle zum Austausch von Metadaten auf der Basis des 'Open Archive Initiative Protocol for Metadata Harvesting', die es ermöglicht, anhand von Metadaten in den Datenbeständen renommierter wissenschaftlicher Dokumentenarchive zu recherchieren.

OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting)

OAI-PMH ist ein Austauschprotokoll für Metadaten. Dank OAI-PMH kann ein Dokumentenserver Metadaten von anderen OAI-PMH-kompatiblen Dokumentenservern einsammeln und indizieren, so dass die mit der eigenen Suchmaschine gestarteten Recherchen auch die auf jenen anderen Servern beherbergten Dokumente miteinbeziehen können. Ein OAI-PMH-kompatibler Server kann zugleich Metadatenlieferant und –harvester sein.

OAIS (Open Archival Information System)

Das als ISO 14721 verabschiedete Referenzmodell beschreibt ein Archiv als Organisation, in dem Menschen und Systeme mit der Aufgabenstellung zusammenwirken, Informationen zu erhalten und einer definierten Nutzerschaft verfügbar zu machen. OAIS-Konformität beinhaltet die Übernahme der Verantwortung für die Einhaltung der im Referenzmodell vorgegebenen Rahmenbedingungen. Das Referenzmodell spezifiziert weder das Design noch die Implementierung eines OAIS-konformen Archivs.

Online-Publikationen

Online-Publikationen sind eine Teilmenge der elektronischen Publikationen. Online-Publikationen werden ohne die Bindung an physische Datenträger im Internet publiziert und verbreitet. Online-Publikationen treten in vielfältigen Datenformaten und Erscheinungsformen auf. Beispiele: elektronische Zeitschrift, Datenbank, E-Mail-Newsletter, Website usw. Auch verwendete Begriffe: Netzpublikationen; elektronische nicht physische Publikationen, unkörperliche Publikationen.

PDF

Von der Firma Adobe auf der Basis von Postscript geschaffenes Dateiformat zur Darstellung von Dokumenten im Originallayout (z.B. feste Seitenumbrüche, Platzierung von Abbildungen). Damit ist PDF nicht nur für kommerzielle, sondern auch für wissenschaftliche Publikationen, vor allem hinsichtlich des Zitierproblems in elektronischen Dokumenten, interessant. Die Eignung von PDF als Format zur Langzeitarchivierung ist umstritten, da es sich trotz der Offenlegung um ein proprietäres Format handelt. Für eine Untermenge des Formats (PDF/A = Archive) wurde ein Standardisierungsprozess (ISO 19005-1. Document management - Electronic document file format for long-term preservation - Part 1: Use of PDF (PDF/A)) eingeleitet, der zur grösseren Akzeptanz des Formats für Langzeitarchivierungszwecke führen soll.

Persistent Identifier

Persistent Identifiers sind eindeutige, standortunabhängige Identifikatoren für elektronische Objekte, mit denen gleichzeitig der dauerhafte Zugriff auf elektronische Ressourcen gewährleistet wird. Die Schweizerische Nationalbibliothek vergibt solche Identifikatoren auf der Basis der sogenannten National Bibliography Number (NBN).

Preservation Planning

Prozess im OAIS-Modell: Planung und Durchführung der langzeiterhaltenden Massnahmen an den digitalen Objekten im Archivierungssystem.

Proprietäres Datenformat

Bei einem proprietären Datenformat handelt es sich um eine Datenstruktur, Datei- oder Datenformat, für welche die Spezifikationen nicht offen verfügbar oder zugänglich sind. Beispiele: '.wma' oder '.doc'.

robots.txt

Nach der Übereinkunft des Robots-Exclusion-Standard-Protokolls liest ein Webcrawler (Robot) beim Auffinden einer Webseite zuerst die Datei robots.txt (kleingeschrieben) im Stammverzeichnis (Root) einer Domain. In dieser Datei kann festgelegt werden, ob und wie die Webseite von einem Webcrawler besucht werden darf. Website-Betreiber haben so die Möglichkeit, ausgesuchte Bereiche ihrer Webpräsenz für (bestimmte) Suchmaschinen zu sperren. Das Protokoll ist rein hinweisend und ist auf die Mitarbeit des Webcrawlers angewiesen. Man spricht hier auch von "freundlichen" Webcrawlern. Ein Ausgrenzen bestimmter Teile einer Webpräsenz durch das Protokoll garantiert keine Geheimhaltung, dazu sind Seiten oder Unterverzeichnisse eines Servers über eine .htaccess-Datei zu schützen. Manche Suchmaschinen zeigen die vom Webcrawler gefundenen URLs trotzdem in den Suchergebnissen an, jedoch ohne Beschreibung der Seiten.

SFTP (Secure File Transfer Protocol)

SFTP, auch Secure FTP genannt, ist ein Netzwerkprotokoll zur Übertragung von Dateien über TCP/IP-Netzwerke. Die Besonderheit hierbei ist, dass eine ansonsten ungesicherte File-Transfer-Protocol-Verbindung (FTP) teilweise über Secure Shell (SSH) getunnelt wird.

Statische Websites

Im bibliothekarischen Kontext sind statische Websites zum Zeitpunkt ihrer Erstveröffentlichung vollständig und abgeschlossen und unterliegen keinen weiteren Veränderungen (Monografische Ressource). Inhaltlich veränderte "Versionen" statischer Websites gelten als neue Online-Publikationen.

Tarball

Tar ist der Name eines im Unix-Umfeld sehr geläufigen Archivierungsprogramms und bezeichnet zudem das von diesem Programm erzeugte Dateiformat mit der Endung ".tar". Der Name wurde aus tape archiver (Bandarchivierer) gebildet, da mit dem Programm ursprünglich Daten auf Bandlaufwerken gesichert wurden. Tar bietet die Möglichkeit, viele Dateien sequenziell in eine einzige Datei zu schreiben, bzw. Dateien aus selbiger wieder herzustellen. Die entstehende Datei wird auch als Tarball bezeichnet. Der MIME-Typ für tar-Dateien ist application/x-tar.

URN (Uniform Resource Name)

Eine URN (Uniform Resource Name) ist ein Persistent Identifier. Persistent Identifier können die URLs (Uniform Resource Locator; "Link" im Internet) im Katalog oder anderen Nachweissystemen ersetzen oder als stabile Referenzen in den Dokumenten selbst angewendet werden. Dadurch werden Links stabil. Der Pflegeaufwand für die Aktualisierung der Verweise reduziert sich, da URLs nur an einer einzigen Stelle automatisiert gepflegt werden. Die Links können in mehrere Nachweisdienste integriert werden. Digitale Publikationen haben einen weltweit eindeutigen Identifier und sind damit zuverlässig zitierbar. Mit einer URN wird der dauerhafte Zugriff auf ein Objekt sichergestellt. Der dauerhafte Zugriff wird durch Langzeitarchivierung oder Archivierung der Objekte sowie durch die technische Hochverfügbarkeit des URN-Dienstes gewährleistet. Eine URN verweist auf mindestens eine URL, über welche ein Objekt adressiert wird. Eine URN kann auch mehrere Kopien desselben Objekts, also URLs verwalten sowie unterschiedliche Präsentationsformate der Objekte.

Usenet (Newsgroups)

Usenet ist ein im Prinzip vom Internet unabhängiges Netzwerk von Diskussionsforen bzw. –gruppen. Eine andere Bezeichnung für das Usenet ist Newsgroup, was allerdings zu Missverständnissen führen kann. Eine Newsgroup ist eigentlich nur ein – möglicherweise sehr kleiner – Teil des Usenet, der sich mit einem bestimmten Thema befasst. Es gibt aber auch Newsgroups ausserhalb des eigentlichen Usenet, etwa in firmeneigenen Intranets oder an Schulen bzw. Universitäten.

Web Harvesting

Siehe [Harvesting](#).

WebDAV

WebDAV (Web-based Distributed Authoring and Versioning) ist ein offener Standard zur Bereitstellung von Dateien im Internet. Dabei können Benutzer auf ihre Daten wie auf eine Online-Festplatte zugreifen. Bekannte Beispiele sind Apples virtuelle Internet-Festplatte iDisk, das GMX MediaCenter oder die aon-Online-Festplatte von Telekom Austria. Technisch gesehen ist WebDAV eine Erweiterung des Protokolls HTTP/1.1, die bestimmte Einschränkungen von HTTP aufhebt. Bisher kennt man aus Online-Formularen meist nur die Möglichkeit, einzelne Dateien hochzuladen (HTTP-POST). Mit WebDAV können ganze Verzeichnisse übertragen werden. Zudem ist eine Versionskontrolle spezifiziert.

Weblogs (Blogs)

Ein Weblog, üblicherweise einfach nur Blog genannt, ist eine Website, die periodisch neue Einträge enthält. Ein Weblog ist eine Art virtuelles Tagebuch in das der Urheber oder die Urheberin (Blogger) persönliche Gedanken und kommentierte Links zu besuchten Internetquellen festhält.

Webseite

Eine Webseite ist eine einzelne Seite eines Informationsanbieters im World Wide Web. Die Webseite ist von der Website zu unterscheiden. Eine Website besteht üblicherweise aus einer Gruppe hierarchisch angeordneter Webseiten, wobei die Hauptseite oft als Homepage bezeichnet wird.

Website (Site)

Eine Website wird üblicherweise aus mehreren Webseiten aufgebaut, die hierarchisch angeordnet sind.

wget

GNU Wget ist ein freies Kommandozeilen-Programm zum Herunterladen von Ressourcen (Dateien, Webseiten, etc) über ein Netzwerk. Zu den unterstützten Protokollen gehören ftp, http und https. Die erste Version stammt aus dem Jahr 1995 und wurde von Hrvoje Niksic geschrieben. Das Programm gibt es sowohl für UNIX und GNU/Linux als auch für OS/2, Windows und für SkyOS. Es steht unter der GNU General Public License und ist Teil des GNU-Projekts. Das Protokoll wurde 1994 von einer unabhängigen Gruppierung entwickelt, ist inzwischen jedoch allgemein anerkannt und kann als Quasi-Standard betrachtet werden. Mit ACAP 1.0 (Automated Content Access Protocol) wurde am 30.11.2007 eine mögliche Erweiterung des Robots Exclusion Standards veröffentlicht. Anfang Juni 2008 bekannten sich Google, Microsoft und Yahoo zu einigen Gemeinsamkeiten.

World Wide Web (WWW)

Das World Wide Web ist ein 1989 am europäischen Kernforschungszentrum CERN entwickeltes Hypermediasystem für das Internet. Das WWW ermöglicht den Zugriff auf beliebige digital gespeicherte Dokumente, die auf irgendeinem mit ihm verbundenen Computer in der Welt vorhanden sind. Dokumente sind sowohl normale Texte, Hypertexte aber auch Musik- und Bilddateien sowie Filme. Das World Wide Web und das Internet sind nicht identisch, sondern das World Wide Web setzt auf das Internet auf. Die meisten Internet-Surfer/innen benutzen allerdings ausschliesslich das World Wide Web für ihre Aktivitäten. Das World Wide Web nutzt HTTP als Übertragungsprotokoll, daneben gibt es im Internet noch andere Protokolle, z.B. FTP. Die Mehrzahl der Textdokumente im WWW ist heute mit der Seitenbeschreibungssprache HTML erstellt. Deren Hauptmerkmal sind sogenannte Hyperlinks, die es erlauben, aus einem Dokument heraus jedes andere, das im WWW zugänglich ist, mithilfe eines standardisierten Adressierungsverfahrens über das Protokoll HTTP aufzurufen. Neben HTML gibt es noch eine Vielzahl anderer Datenformate für Texte und andere Dokumenttypen.

XML (Extensible Markup Language)

Die Extensible Markup Language, abgekürzt XML, ist ein Standard zur Erstellung maschinen- und menschenlesbarer Dokumente. XML definiert dabei die Regeln für den Aufbau solcher Dokumente. Für einen konkreten Anwendungsfall (XML-Anwendung) müssen die Details der jeweiligen Dokumente spezifiziert werden. XML ist damit ein Standard zur Definition von beliebigen, in ihrer Grundstruktur jedoch stark verwandten Auszeichnungssprachen. XML-Strukturen haben gute Eigenschaften hinsichtlich der Langzeitverfügbarkeit, da es sich um einen offenen Standard mit umfangreichen Strukturierungsmöglichkeiten für unterschiedlichste Objekttypen handelt. XML-Strukturen sind maschinell interpretierbar, weil sie formalen Regeln folgen.