Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA

Swiss Federal Office of Culture FOC
Swiss National Library NL

e-Helvetica Glossary

Version:
Version 1.1

Date:
17.07.2012

Table of contents

Access
The OAIS entity that contains the services and functions which make the archival information holdings and related services visible to consumers.

Administration
The OAIS entity that contains the services and functions needed to control the operation of the archival system.

Archival Storage
The OAIS entity that contains the services and functions used for the storage and retrieval of Archival Information Packages.

Born digital publications
Born digital publications are original digital publications which are not intended to have an analogue equivalent, either as the originating source or as a result of conversion. They only exist in digital form.

Cache
Temporary storage area where frequently accessed data can be stored for rapid access.

Data Management
The OAIS entity that contains the services and functions for maintaining and accessing a wide variety of information by using bibliographic, technical and administrative metadata.

Databases
A database is a system for the storage and management of huge quantities of data. It contains data stored according to a predefined structure and management programmes that store data, search or execute other operations with the data. The term database also generally applies to the organisation and structured management of data. The contents of databases are listed by interactive user interfaces (mostly proprietary). Users transmit their individual search requests for which results are individually compiled.

Digitisation
The digital conversion of analogue objects requires interventions in various domains. The main objective is to facilitate user access to contents and preserve the content of objects threatened by degradation. Measures need to be taken for the long-term archiving of these digital objects in order to guarantee the future access to these digital copies.

Digitised publications
Digitised publications are publications which have been converted from an analogue form (i.e. paper) to a digital form.

Document server
A document server refers to a technical and organisational system whose main task is to give the final user access to digital documents (or objects similar to documents). This type of server works in relation with a repository server in order to guarantee the long-term availability of access to these objects.

DOI (Digital Object Identifier)
A digital object identifier (DOI) is for the persistent identification of content objects in the digital environment. Each DOI unequivocally and permanently identifies the object to which it is associated. The DOI can be compared to the ISBN and ISSN systems but goes further as information about a digital object may change over time, including its location, but its DOI name will not change.

Domain
A domain is characterised or defined by common characteristics; this generally means a group of computers sharing common host names. The top-level domain is the smallest common element, for

example the abbreviation of a country ".ch" or the abbreviation ".com". Domains can be subdivided into sub-domains such as second level domains, third level domains etc.

Dynamic web sites
In librarianship terms, dynamic web sites are publications which are not completed at the time of their initial publication. They can be modified; information can be added periodically (integrative resources).

e-Helvetica
e-Helvetica is the interface for consulting the digital collections of the Swiss National Library (NL). It enables full-text searches of the content of digital publications.
In e-Helvetica you may find the digital publications held in the NL's collections. These currently include born digital books, journals, university and official publications as well as websites concerning Switzerland. The collection is under construction. Partner institutions also work with the NL on e-Helvetica.
e-Helvetica also contains publications that have been digitised by the NL.
„e-Helvetica" is also the term used for the NL service responsible for the continued development of the digital collection, including its registration, cataloguing and indexing, and long-term preservation and also making it available to the public.
In e-Helvetica the „e" stands for electronic and „Helvetica" for publications relating to Switzerland.

Electronic publications
The expression electronic publications covers online as well as offline publications. The term digital can also be used.

e-mail
An e-mail is a message (text or file) which is transferred in a network of senders and receivers by the service "Simple Mail Transfer Protocol" (SMTP).

Emulation
Emulation is a strategy for long-term preservation of digital objects. Emulation addresses the original hardware and software environment of the digital object. Specialised software recreates (emulates) the digital object so it can be read on the latest systems available on the market. Today there is a debate around concurring emulation strategies: whether emulation should address software or hardware.

File format
A defined set of rules is necessary to assemble data in a file. A given set of rules constitutes the format of the file. File formats can be very simple if they only prescribe for example a simple succession of data. They can also require the storage of additional information in precise locations in the file. They can also require a precise coding of the data and the information stored in the file. Until a proper archive format is created, the choice of the right file format is critical for long-term archiving of digital data. It is prudent to choose widely used file formats (for example TIF for image files); in fact it is probable that in a relatively near future programmes will be able to retrieve information from these files. The choice of simple file formats has indeed positive consequences on long-term archiving (for example TXT for text files). If need be, these formats could facilitate the subsequent reconstruction of the rules used to assemble the data in the files. However, file formats belonging to a single producer (for example DOC for text files) are inappropriate as the producer can modify or limit their use as he wishes. Often, the suffix of the file can help in the identification of the file format.

FTP (File Transfer Protocol)
FTP is a protocol used to transfer data from one computer to another via the Internet. FTP is also the name of the Internet service on which this protocol is based.

Harvesting
Harvesting, also named webharvesting, is an automated system for the collection of web sites using a harvester (robot).

Helvetica
Helvetica are publications published in Switzerland, about Switzerland or Swiss natives living in Switzerland or abroad or publications created in association with authors who have an important link with the country. Helvetica corresponds to the holdings of the Swiss National Library.

Helveticat
Helveticat is the catalogue of the Swiss National Library.

Homepage
A homepage (often written as home page) is the main or first page of a web site. The expression is also used to designate the web site of an individual. Today, the term is not limited to its original definition but can designate the entirety of the information proposed on a web site.

HTTP (Hypertext Transfer Protocol)
HTTP is a communications protocol for the transfer of information on the World Wide Web. It is based on the transport protocol TCP/IP.

Hyperlink
A hyperlink, or simply put a link, is a reference or navigation element in a hypertext or on a web page to another element in another section of the same web page or to another web page. The term from which the link is established is usually displayed in some distinguishing way, e.g. in different colour or is represented by a graphic symbol.

Incremental backup
Incremental backup is based on a complete backup of the source system. The incremental backup backs up only the newly added data at scheduled intervals. The initial complete backup is used to successfully restore the system. The incremental backups then restore the system on top of the complete backup in ascending chronological order.

Ingest
Ingest is the OAIS entity that contains the services and functions that accept data from producers, prepares it for archiving and integration in the long-term repository.

The Internet
The Internet is the biggest worldwide, publicly accessible series of interconnected computer networks; it is a network of networks which together carry various information and services such as: e-mail, electronic mail; the World Wide Web (WWW, commonly shortened to the Web), a system of interlinked hypertext documents; Usenet, a discussion forum; FTP, File Transfer Protocol used to transfer data from one computer to another; IRC to chat, Gopher, Telnet, Wais, Archie and some other earlier services whose importance is decreasing and somewhat out-dated. For many private users the WWW is the most important contribution of Internet and this is probably why for many people the two terms are often considered equivalent.

The Intranet
The Intranet is a private network (confined to an organization) working with the Internet technology. Contrary to the Internet, the Intranet is not accessible publicly but is restricted to the use of a determined group of local users.

JDBC (Java Database Connectivity)
Java Database Connectivity (JDBC) is a database interface for the JAVA platform offering a uniform interface to databases from other manufacturers and is specially oriented to relational databases. JDBC is comparable to ODBC under Windows, for example or DBI under Perl in its function as a universal database interface. JDBC tasks include, among others, establishing and administering connections to databases, forwarding SQL queries to the database and converting the results in a form usable to Java and making it available to the program.

Long-term archiving
The act of maintaining and preserving information over the long-term. The long-term preservation of digital information (digital preservation) presents new challenges.

Long-term availability
Ensuring the long-term conservation of digital objects implies taking a series of measures that will enable future generations to have access to them. By future generations we mean an indefinite period of time in the future during which we expect:
- profound technological changes in the areas of storage and retrieval of digital objects
- development of new formats and medias and the elimination of previous models from the market-place
- decisive changes in user behaviour and the introduction of new information contents

Two types of conservation measures ensure long-term availability:
- conservation measures to preserve the substance of data flow of digital objects (for example by refreshing)
- conservation measures to preserve usability (for example through emulation or migration)

MARC
MARC (Machine Readable Cataloguing) is a standard for the representation and communication of bibliographic data, authority data, holdings, classification and community information in a machine-readable form.

Message boards
Message boards are defined pages of a mailbox (Bulletin Board System) or a newsgroup in which users can exchange information - like posting a notice on a bulletin board.

Metadata
Simply put, Metadata is data about other data. In other words, metadata is data which describes other data or objects. It is information on data that significantly facilitates access to specific data as well as the exchange and management of this data. This basic information can contain for example, indications on the author of the document, the publication date, notes on other documents pertaining to the same subject, etc. Bibliographic records of publications constitute a form of metadata. In the digital domain, metadata does not only include bibliographic information but also technical and administrative information (format, size of file, date of data retrieval etc.).

METS (Metadata Encoding & Transmission Standard)
METS is an XML format defined according to an XML schema and is a standard for encoding descriptive, administrative and structural metadata regarding objects within a digital library. The format of metadata (MAB, MARC, Dublin Core, etc.) can vary and is not defined by METS. However, METS contains elements for the grouping of objects and their association with descriptive and administrative metadata. METS allows for example the encoding of a structured work into a hierarchical structure; this work can be a collection of books, a book structured in chapters and pages, or a film made up of many scenes. It is a sub-set of the XLink schema used to link METS files and digital objects from which information has to be retrieved.

Migration
Migration is a conservation strategy for the long–term preservation of digital objects. Due to the changing technical environment, migration ensures future usability of digital objects by transferring or rewriting data from an out of date media to a current media.

MODS (Metadata Object Description Schema)
MODS (Metadata Object Description Schema) is a metadata standard developed by the Library of Congress. MODS is based on an XML schema designed to describe bibliographic data.

NAS (Network Attached System)
Network Attached Storage (NAS) refers to an easy-to-administer file server. NAS is generally used to provide independent storage capacity within a computer network at relatively low costs. In contrast to Direct Attached Storage, NAS has its own host and operating system. Functionality is carefully matched to use which prevents errors at the outset by eliminating comprehensive configurations that are not needed for the specific purpose. File-based services such as NFS or SMB/CIFS represent the core function which is why NAS systems are often simply referred to as filers. An off-the-shelf disk drive in an external housing, with a “RJ-45” plug, the appropriate software and equipped with optional connections already corresponds to NAS.

NFS (Network File System)
Network File System – or NFS (also: Network File Service) – is a protocol developed by Sun Microsystems that allows for access to files via a network. The files, for example, are not transmitted in the sense of FTP, but rather users may access the files on a remote computer as if they were stored on the local hard drive.

Ninive
Ninive ist the filing system of the Swiss National Library for long-term archiving.

OAI (The Open Archives Initiative)
The Open Archives Initiative referred to as OAI, develops and promotes interoperability standard using the Open Archive Initiative Protocol for Metadata Harvesting. The OAI seeks to enhance access to e-print archives and collections as a means of increasing the availability of scholarly communication.

OAI-PMH (The Open Archives Initiative Protocol for Metadata Harvesting)
The Open Archives Initiative Protocol for Metadata Harvesting (referred to as the OAI-PMH) is a harvesting protocol for sharing metadata between compatible OAI-PMH servers. Searches launched from a server’s own search engine can cover as well documents located on the other servers. An OAI-PMH compatible server can be a data provider and a metadata harvester.

OAIS (Open Archival Information System)
OAIS has become the recognised reference model for digital archiving in libraries and archives around the world. It is a strictly logical model that requires no implementation. The model contributes to a better understanding of digital archiving and to the creation of a common language in this area.
This reference model was approved as ISO standard 14721 in 2003. An OAIS is an archive consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available to a designated community. In order to be in conformity with OAIS, an archive must respect the set conditions defined in the reference model. The model does not specify the design or the implementation of an OAIS-type archive.

Official publications
Official publications are either printed information or information conserved on other medias. They are published by one of the departments or units of the Central Federal Administration.

Online publications
Online publications are a sub-set of electronic publications. Online publications are published and transmitted on the Internet without any physical media. They appear in all sorts of data formats and are presented in various forms. For example, there are electronic serials, databases, newsletters via e-mail etc. Other terminology used: network publications, non-physical electronic publications or non-physical publications.

PDF
The Portable Document Format (PDF) is the file format created by Adobe Systems based on the Postscript (Page description programming language) for representing documents in their original layout (for example, fixed page breaks, the positioning of illustrations). PDF is not only suitable for commercial publications but also for scholarly publications particularly in relation to the issue of quoting electronic

publications. Although PDF is an open standard it has come into question as an adequate format for long-term archiving as it is a proprietary format. For a sub-set of the format (PDF/A = Archive) a process of standardisation (ISO 19005-1. Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF (PDF/A)) was introduced to render the format more acceptable for long-term archiving.

Persistent identifier
Persistent identifiers identify unique digital objects, independently of location and at the same time, they guarantee durable access to electronic resources. The Swiss National Library attributes these identifiers on the bases of the National Bibliography Number (NBN).

Preservation planning
Preservation planning is a planning process in the OAIS model: planning and implementation of long-term conservation measures of digital objects in the archiving system.

Proprietary file format
A proprietary data format is a data structure, file or data format for which the specifications are not publicly available or accessible. Examples include: .wma or .doc.

Repository Server
The repository server refers to a technical and organisational infrastructure whose main task is to store, manage and archive digital objects for the long-term. In order to distinguish between the two essential functionality axes of long-term archiving, a separation is drawn between the systems of the repository server from the systems of the document server.

robots.txt
Upon agreement on the Robots Exclusion Standard protocol, a webcrawler (robot) first reads the file robots.txt (lower case) in the root for a domain when it finds a website. This file determines whether and how a website may be visited by a webcrawler. It provides website operators the ability to block selected areas of the web presence for (certain) search engines. The protocol is merely a reference and depends on the cooperation of the webcrawler (referred to as “friendly” webcrawlers). Restricting certain parts of a web presence does not, however, guarantee confidentiality; pages or subfolders for a server must be protected via an .htaccess file. Some search engines nevertheless display the URLs found by the webcrawlers, but without a description of the pages.

SFTP (Secure File Transfer Protocol)
SFTP, also known as Secure FTP, is a network protocol for transferring files over TCP/IP networks. The distinctive feature of SFTPs is that an otherwise insecure file transfer protocol link (FTP) can be partially tunnelled over the Secure Shell (SSH).

Static web sites
In librarianship terms, static web sites are publications which are complete at the time of their initial publication. They will not be subject to further modifications (monographic resources). Subsequent “versions” of static web sites modifying the content of the site will be considered new online publications.

Storage media
Storage media on which digital data can be stored durably and subsequently retrieved. The storage media can be processed either mechanically or electronically, for example magnetic medias such diskettes and magnetic tapes and optical medias such as CD-ROMs and DVDs.

Tarball
Tar (derived from tape archive) is both a file format ending with .tar and the name of the Unix program used to handle such files. Initially developed to be written directly to sequential I/O devices for tape backup purposes, it is now commonly used to collect many files into one larger file and to recreate files from that file. The created file is also called tarball. The MIME type for tar-file is application/x-tar.

University publications
University publications are doctorial theses and dissertations that are published at a university or university of applied sciences.

URN (Uniform Resource Name)
An URN (Uniform Resource Name) is a persistent identifier. Persistent identifiers can replace URLs (Uniform Resource Locators; links on the Internet) in the catalogue and in other listing systems or can be used as stable references in the documents themselves, and thus allow the creation of stable links. The updating of references is less time-consuming as URLS are updated automatically at the same location. Links can be integrated in numerous listing services. Digital publications have a unique identifier and can therefore be quoted in a dependable manner. The URN guarantees durable access to an object. This durable access is guaranteed through long-term archiving or object archiving as well as through the constant technical availability of the URN service. An URN refers to at least one URL which identifies an object. An URN can also generate several copies of a given object corresponding to various URLs as well as various presentation formats of these objects.

Usenet (Newsgroup)
Usenet is a network of forums or discussion groups that is, in principle, independent of the Internet. The expression Newsgroup is sometimes also used as an equivalent to Usenet and this can at times create confusion as a newsgroup is, in reality, only part and sometimes a very small part of Usenet and it deals with one particular theme. There are also other newsgroups that exist outside Usenet; for example in company, university or school intranets.

Web Harvesting
See Harvesting.

Web page
A web page or webpage is a resource of information that is suitable for the World Wide Web. The web page must be distinguished from the web site which is usually defined as a series of web pages organised hierarchically and whose main page is called a homepage.

Web site
A web site is made up of a number of web pages organised hierarchically.

WebDAV
WebDAV (Web-based Distributed Authoring and Versioning) is an open standard for preparing files on the Internet. Users may access your data as if it were an online hard disk. Known examples include Apples virtual Internet hard disk iDisk, the GMX MediaCenter or the aon online hard disk by Telekom Austria. Technically, the WebDAV represents an expansion of the HTTP/1.1 protocol that removes certain limitations of HTTP. To date, users are familiar from online forms allowing users to download individual files (HTTP-POST). WebDAV allows the user to transfer entire folders. Moreover, it specifies a version check.

Weblogs (Blogs)
A weblog, commonly abridged to “blog”, is a web site usually maintained by an individual with regular entries. Many blogs provide commentary or news on a particular subject while others function as personal online diaries in which the author or blogger offers personal views and comments on, and links to other Internet pages.

wget
AGNU Wget is a free command line program for downloading resources (files, websites, etc.) via a network. Supported protocols include ftp, http and https. The first version originated in 1995 and was written by Hrvoje Niksic. The program is available for both UNIX and GNU/Linux as well as for OS/2, Windows and SkyOS. It is subject to the GNU General Public License and is a component of the GNU project. An independent group developed the protocol in 1994 and it has become generally recognized in the interim and is considered a quasi-standard. ACAP 1.0 (Automated Content Access Protocol),

published on November 30, 2007, may represent a potential expansion to the Robots Exclusion Standards. Google, Microsoft and Yahoo acknowledged some commonalities at the start of 2008.

World Wide Web (WWW)

The World Wide Web, commonly abridged to WWW, is a hypermedia system developed for the Internet in 1989 by the European Organization for Nuclear Research (CERN). The WWW allows access to all sorts of digital documents stored in a computer, somewhere in the world, which is linked to the World Wide Web. These documents can be normal texts, hypertexts, music or images or video files. The World Wide Web and the Internet are two different entities but the World Wide Web is based on the Internet. However, most Internet surfers exclusively use the World Wide Web for their surfing activities. The World Wide Web uses HTTP, the Hypertext Transfer Protocol although today there exist other protocols for the Internet such as FTP (File Transfer Protocol). The majority of documents on the World Wide Web are created using HTML (HyperText Markup Language) mainly characterised by hyperlinks. Hyperlinks enable the linking of any document on the World Wide Web to any other document posted there using a standardised communication procedure and the HTTP protocol. Aside from HTML there is a multitude of other data formats for texts and other types of documents.

XML (extensible Markup Language)

Extensible Markup Language, commonly abridged to XML, is a standard used in the creation of documents which are legible by man and machine alike. XML defines the rules for the creation of these documents. For each concrete application (XML application), it specifies the details of the document in question. XML is a standard for the definition of any composition language; these languages being however strongly related by their basic structure. XML structures offer interesting properties for long-term availability because it is an open standard offering numerous structure possibilities for many different types of objects. XML structures are machine legible because they follow formal rules.