

## Glossary of Terms, Aquifer context

---

Alison Parks-Whitfield  
v.7  
9/5/07

**Aggregate:** Gathering of files or metadata, etc. in a repository. Aggregating collections implies gathering digital objects in addition to metadata, from multiple domains.

**Aggregator:** A service that gathers information published by different sources and organizes it together under a common search interface. The aggregator may also license access to a collection of journals from many different publishers. [British Library]

**Aquifer:** A Digital Library Federation program that models and develops network-based, scalable solutions to make digital library content easy to find and use across institutional boundaries. DLF Aquifer draws upon digital content pertaining to American culture and life to develop a distributed open digital library that can be used in a variety of environments. For program details see: <http://www.diglib.org/aquifer/>

**Asset Action Package:** An asset action package is an XML-defined set of actionable URIs for a digital resource that delivers named, typed actions for that resource. Packages are made up of action groups, which are sets of actions, with an optional set of parameters which are specific to the group. Every asset action package contains at least a "default" action group which provides a basic high-level set of actions. It is optimal but not required that it also contain at least one basic medium-specific action group which provides a set of basic functions for a known type of resource. An asset action package may contain any number of additional action groups. [TWG]

**Browse:** A guided search. To constrain search by displaying multiple logically grouped categories on a website. Users can select from a category/ sub-category to receive targeted results.

**Business process:** A process, workflow, or set of operations used in support of the (business) requirements of a domain, delivering results for the user. [JISC e-Framework glossary]

**Collection aggregation:** *See Aggregate.*

**Collection Registry:** The database in which descriptions of collections are stored. [UIUC DLI Glossary]

**Constraint:** A restriction on behaviour or interaction. Constraints may be imposed for technical, policy, or other reasons, e.g., accessibility, scope, ease of evolution, efficiency, extensibility. [WWW Arch]

**CMS integration:** Providing DLF Aquifer collections and services through a Course Management System such as Sakai or Blackboard.

**Container:** Containers are places in OAI-PMH responses where XML complying with any external schema may be supplied. Containers are provided for extensibility and for community specific enhancements. The OAI Implementation Guidelines lists the existing optional containers and provides links to existing schemas. [Open Archives Forum Glossary]

**Design:** A description (typically represented in a collection of documents in different forms) used by developers as the basis to convert a concept into a software product or service. [JISC e-Framework glossary]

**Document:** Content represented in a defined, known format. [WS Gloss]

**Dublin Core:** *Dublin Core Metadata Element Set* (Dublin Core Metadata Initiative): <http://dublincore.org/>. A standard set of 15 elements (title, creator, subject, etc.), with optional qualifiers and community-specific extensions. All elements are optional and repeatable within an application profile used to structure data elements into records customized for specific audiences. Dublin Core is used to structure descriptive information about a resource but also to map readily to other descriptive schema, to facilitate sharing information across different metadata schemas and user communities. First developed in the mid-1990s, and originally intended for use in describing web sites and web pages, Dublin Core is now used also for describing physical and digital collections in museums, libraries, archives and other repositories. [Rutgers MIC glossary].

**Enhancement workflow:** Processes that support automated or semi-automated improvement of the metadata associated with a given digital object.

**Federated Search:** Federated searching consists of transforming a [query](#) and broadcasting it to a group of disparate [databases](#) with the appropriate syntax, merging the results collected from the databases, presenting them in a succinct and unified format with minimal duplication, and allowing the library patron to sort the merged result set by various criteria (definition by [Peter Jacso](#), 2004). [Wikipedia].

**Federated Search Tools:** Cross-database search tools that can search multiple catalogs, online databases, search engines, or commercial databases. They can often merge and de-duplicate results and provide unified access to a variety of information resources. [dmoz open directory project]

**Finding service:** Support for the FRBR <http://www.ifla.org/VII/s13/frbr/frbr3.htm#6> find function: "to **find** entities that correspond to the user's stated search criteria" (i.e., to locate either a single entity or a set of entities in a file or database as the result of a search using an attribute or relationship of the entity). May also support the ability to browse collections for a given object or set of objects or to locate a given collection.

**Framework:** A collection of cooperating *service genres* within a domain. [GoF]  
A vocabulary of components (e.g., service genre, service expression and service usage model) used to model concepts and behaviours within a domain or to create a particular type of infrastructure (system or environment). Associated with the vocabulary is the grammar or language used to describe how the elements are combined or constrained to create systems and environments. [JISC e-Framework glossary]

**Focused crawling:** Similar to topical crawling. A crawling service that will attempt to locate and download pages on a given topic.

**Gap analysis:** To compare the actual with the potential to accurately visualize areas for improvement.

**Identifier:** An unambiguous name for a resource. [WWW Arch]

**MARC:** The acronym for MACHine-Readable Cataloging. It defines a data format that emerged from a Library of Congress-led initiative that began thirty years ago. It provides the mechanism by which computers exchange, use, and interpret bibliographic information, and its data elements make up the foundation of most library catalogs used today. MARC became USMARC in the 1980s and MARC 21 in the late 1990s... MARC 21 has been mapped to the following metadata standards: MODS, Dublin Core, MARC Character Sets to UCS/Unicode, Digital Geospatial Metadata ... The following metadata standards have been mapped to MARC 21: MODS, Dublin Core, UNIMARC to MARC21, ONIX, Digital Geospatial Metadata to MARC. [MARC Standards home page: <http://www.loc.gov/marc/>]

**Metadata:** Data about data. ... "the sum total of what one can say about any *information object*<sup>1</sup> at any level of aggregation." [Introduction to Metadata (Research at the Getty)].

The word **metadata** can yield varying interpretations depending upon one's vantage point.

**Metadata Enhancement:** *coming soon.*

**Metadata Remediation:** Correcting or improving the existing metadata. *See Metadata Enhancement.*

**Metadata Schema:** A labeling, tagging or coding system used for recording cataloging information or structuring descriptive records. A metadata schema establishes and defines

data elements and the rules governing the use of data elements to describe a resource. [Rutgers MIC glossary].

**Middleware:** Communications software used to link requestors to providers or to link support software in order to integrate service implementations into an application. [JISC e-Framework glossary]

**MODS:** (Metadata Object Description Schema). MODS is intended to be able to carry selected data from existing MARC 21 records as well as to enable the creation of original resource description records. It includes a subset of MARC fields and uses language-based tags rather than numeric ones, in some cases regrouping elements from the MARC 21 bibliographic format. MODS is expressed using the XML schema language of the World Wide Web Consortium. The standard is maintained by the Network Development and MARC Standards Office of the Library of Congress with input from users. [ MARC Standards home page: <http://www.loc.gov/marc/>]

**OAI Harvesting:** Open Archives Initiatives allows users to gather metadata. OAI-PMH (Open Archives Initiatives Protocol for Metadata Harvesting) defines a mechanism for data providers to expose their metadata.

**OAIster:** A service which harvests collection data from a large variety of institutions. Data is made available to any interested end user for searching. A product of the University of Michigan Digital Library Production Service. [OAI workshop].

**Open standard:** A standard (*de jure* or *de facto*) or specification, developed collaboratively through an open, consensus-building process, that is platform independent, vendor neutral, extensible, reusable, publicly accessible, and not encumbered by royalties, licenses or constraints on use and implementation. [e-Framework Overview]

**Persistent identifiers:** An Identifier which is associated with the same resource for all time. [IPTC Standards Draft, NewsML 2 Architecture, v.1.0]

**Personas:** The persona technique defines archetypical users, defined by actual user data. While a persona has a fictional name, his or her roles are based on needs, current behavior and goals of an aggregate of actual users. A persona is described in a narrative format that can include name, a photo, goals, attitudes, frustrations, work flow, environment and behavior. A person may have multiple personas.

**Portal:** Tool or set of tools for organized knowledge discovery that assists identification and selection of appropriate target resources; provides federated searching and information retrieval of descriptive metadata from multiple, diverse target resources, including but not limited to commercial or licensed electronic resources, databases, Web pages, and library catalogs; manages access to target resources and portal functionalities for authenticated user communities based on various user classes and roles. [ Excerpted

from: Library of Congress Portals Applications IssuesGroup  
<http://www.loc.gov/catdir/lcpaig/introduction.html>]

**Protocol:** A set of formal rules describing how to transmit data, especially across a network. [WS Gloss]

**Recommender systems:** A set of services that supports finding entities that are similar in some way to an entity the user has found or used. Amazon's "recommendations for you" is the most frequently cited example of a recommender system.

**Repository:** A **repository** is a central place where [data](#) is stored and maintained. A repository can be a place where multiple [databases](#) or [files](#) are located for distribution over a network, or a repository can be a location that is directly accessible to the user without having to travel across a network. [Wikipedia]

**Results:** Logic generated matches obtained from a search or browse query performed within a website's databases, then displayed to the user.

**Rights expression language:** A syntax that provides information about how an object can be used, who owns the copyright, if it is in the public domain, etc.

**Rights use cases:** Use cases that demonstrate various scenarios for digital object use. For example, educational use of archival material in the public domain or commercial use of images or audio files that are under copyright protection.

**Search:** To enter a query in an online textbox, submit and receive targeted results from a website's databases.

**Service Genre:** A core e-Framework term referring to a generic or abstract capability expressed in terms of behaviors; the overall concept of a service without regard to how it is converted into an operational capability. A service genre is independent of specific interfaces, technologies and standards. [JISC e-Framework glossary]

**Service toolkit:** A reusable collection of service implementations designed to be used in any application or with any service usage model. A service toolkit is typically distributed with auxiliary software (sample applications, data).  
[GoF]

**(Software) architecture:** The overall structure of a software system, including software components, the externally visible properties of those components, their behaviours, the relationships among them and the constraints on their use. [Ref Arch, OPF]

Alternate definition: The software architecture of a program or computing system is the structure or structures of the system, which comprise software elements, the externally visible properties of those elements, and the relationships among them.

"Externally visible" properties refers to those assumptions other elements can make of an element, such as its provided services, performance characteristics, fault handling, shared resource usage, and so on. [From the book [Software Architecture in Practice \(2nd edition\)](#), Bass, Clements, Kazman; Addison-Wesley 2003]

**SRU:** SRU (Search/Retrieve via URL) is a standard search protocol for Internet search queries, utilizing CQL (Common Query Language), a standard query syntax for representing queries. SRW (Search Retrieve Web Service) is a companion protocol to SRU. The Library of Congress serves as the maintenance agency for these standards. [Library of Congress]

**Static Repository:** A Static Repository provides a simple approach for exposing relatively static and small collections of metadata records through the OAI-PMH. The Static Repository approach is targeted at organizations that:

- Have metadata collections ranging in size between 1 and 5000 records;
- Can make static content available through a network-accessible Web server;
- Need a technically simpler implementation strategy compared to acting as an [OAI-PMH Repository](#), which requires processing [OAI-PMH requests](#);

A Static Repository is an XML file that is made accessible at a persistent HTTP URL. The XML file contains metadata records and repository information. [www.openarchives.org].

**Tools:** Software utilities to facilitate development and testing of software products and services. Tools can also mean software that enables content consumers, and content providers, to perform specific activities, (i.e. the UVA "Collector tool").

**Use cases:** [OPTIONAL, MANY] SERVICE EXPRESSION CASE. Brief informal descriptions of how the service usage model is used. An illustration of the service usage model within a workflow or overall set of business processes SHOULD be included.

[OPTIONAL, MANY] REFERENCE MODEL CASE. Brief, informal presentations of how the reference model is used. An illustration of the reference model showing an overall set of business processes involving workflow and service genres XOR service expressions, or an illustration of the use within an application or collection of interacting service genres XOR service expressions SHOULD be included. [JISC e-Framework glossary]

**Web service:** A software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processible format (specifically WSDL). Other systems interact with the web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialisation in conjunction with other Web-related standards. [WS Arch, WS Gloss]

**Workflow:** The paths through a networked structure of tasks and activities; the steps in a business process. [JISC e-Framework glossary]