Glossary*

A

Access Control: Access control is a system of privileges and permissions that restricts access to a directory or file. In content management systems permissions are assigned to authors, copy editors, graphics editors, and other contributors according to their roles.

Aggregation: Aggregation is the merging of several licensed or subscription services into a single area, thereby making otherwise scattered data, information, and services accessible at a single location, usually via a Web page.

Apache Server: Apache Server is a cross-platform Web server freely available under an open source license. The current version runs on most UNIX-based operating systems, on UNIX/POSIX-derived, and on Windows 2000. Debuted in 1995, the Apache's continued development is primarily among a set of volunteer programmers known as the Apache Group. The source code can be modified or adapted by developers.

Application Programming Interface (API): API is a set of software functions used to initiate contact with network services and the mainframe. One of the primary purposes of an API is to provide a set of commonly used functions—for example, to draw windows or icons on the screen. By making use of its functionality, programmers can save the task of programming everything from scratch.

Application Server: The application server is a server program that hosts transaction and interaction logic for an application. An application server runs on both the development and the production servers. Application servers execute the operations necessary to complete transactions and other interactions between end users, databases, and applications. For example, when a Web page is requested by a Web browser, the Web server will "hand off" the page request to the application server for files with particular file extensions. The server provides functionality such as database access classes, transaction processing, and messaging. It is a key publishing engine in the CM environment (see also Development Server, Production Server, Staging Server).

B

Blog (Weblog): Blog is short for Weblog, a combination of Web and log. Blog is a Web site that contains periodic, reverse chronologically ordered posts. Each post has an anchor that allows a hyperlink to the post from anywhere on the Web. A person who maintains a blog is called a blogger. Blogger (www.blogger.com) is a free Web-based tool that facilitates the posting of blogs. Users FTP a template of their Web page to the blogger site and are able to submit content via a Web-based form. Submissions are posted instantaneously to the Web, and the content maintains the look and feel dictated by the template. In short, it is a simple, Web-based CM system.

C

Chunk: In a content management environment, a chunk is an information content element that can carry a meaningful semantic tag. It is the result of an effort to divide information into small enough sections so that they can be managed easily.

ColdFusion: Created in 1995 by Allaire and merged with Macromedia in 2001, ColdFusion integrates browser, server, and database technologies into Web applications. ColdFusion Web pages include application tags written in ColdFusion Markup Language (CFML) that simplifies integration with databases and avoids the use of more complex language like

C++ to create translating programs. Using ColdFusion, a developer can combine a content database with a set of templates to create a site that builds and serves pages "on-the-fly." The suite of applications is composed of two pieces: ColdFusion Studio—the development interface; and the ColdFusion server, a cross-platform Web application, which deploys pages to the user. ColdFusion uses its own markup language, CFML, which incorporates elements of HTML and XML.

Content Management System (CMS): CMS is application software that manages unstructured content for Web-based access. Once the content has been developed, it can be stored in a relational database, which resides on a content server. The content is then published typically via a template. The publishing module may be capable of handling dynamic content "on the fly." The CMS often includes a security mechanism that grants or denies access to content. An embedded search function might support searches by subject, full-text, author, or other terms. The CMS may also automatically reformat content for other Web-enabled access.

D

Database-Driven Web Pages: "Database-driven" describes Web sites that store and maintain content in a database. The database can be built using a relational database, XML structure, or a structure constructed by other scripting language, such as JavaScript. The database-driven Web pages support information transactions with the user and produce Web pages "on the fly" as users request them.

Development Server: The development server is the server computer on which Web site development is performed and tested. This server sits behind the organization's firewall. Ideally, all content development and testing occur on the development server. Final content is published from the development server to the production server (see also Production Server and Staging Server).

Document Management (DM): Document management is an application or system designed to organize information and make files accessible to all users regardless of origin or format. A document management system comprises file storage (the physical location of each file), file categorization (file types and groups based on the criteria), metadata (owner, status, creation date, etc.), workflow management (files are routed from person

to person based on defined roles and standard), version control (files can be saved as historical series and can be retrieved later), and access points (user can find files through full-text searching, table of contents, and indexes).

Dublin Core: The Dublin Core is a set of standards with reference to the Resource Description Framework (RDF), a dialect of XML. It is used to ascribe library metadata to online documents or any Web page in a consistent basis to enhance retrieval. The *Dublin Core Metadata Element Set* consists of 15 metadata elements: title, creator, subject, description, publisher, contributor, date, type, format, indentifier, source, language, relation, coverage, and rights.

Dynamic Content: Dynamic content is content that can be generated on the fly from either a file system or a database using one of or all of the following: Server Side Includes, CGI scripts, Java servlets, or an application server. Dynamic content is updated frequently and can therefore be kept current and relevant for its appropriate audience. Dynamic content can include sophisticated JavaScript or Shockwave for an interactive experience or serve as a flat HTML page that is updated frequently.

E

Enterprise Content Management (ECM): Enterprise CM systems encompass the entire content creation and management for the organization and not just the Web site, as is the case with a CMS (see also CMS).

eXtensible Markup Language (XML): Put forth by the World Wide Web Consortium (W3C), XML is an offshoot of Standard Generalized Markup Language (SGML), but XML is much easier to use and apply. It was created so that richly structured documents can be described, exposed, shared, and modified over the Web. XML allows Web developers to design their own customized tags to provide functionality that is not available with HTML. The combination of XML and HTML permits powerful and content-rich Web sites. For example, a Web site using XML tags can link to multiple documents using a single HTML hyperlink. Since XML describes the underlying information and its structure, content can be separated from the display.

F

Federated Search: Federated search tools search across a variety of information sources and display results in a single interface. They can be used to consolidate data from multiple sources, while authenticating and de-duplicating results. They can simultaneously search a wide variety of online resources, such as library catalogs, electronic journals, multiple databases, Web resources, as well as locally held materials. Various products have recently been released into the market, which are designed to integrate with portal software or the OPAC (see also Metasearch).

G

GNU General Public License: The GNU General Public License (GPL) is a copyright-free software license or open source license. In contrast to proprietary licenses, the purpose of the GPL is to protect the user's freedom rather than restrict it. The Free Software Foundation (FSF), a non-profit institution, designed the GNU GPL to promote the publication of free software.

I

Information Architecture: Information architecture is the basic design of the system as it relates to the classification and organization of information content in the CMS environment. It encompasses chunking and tagging strategies, template designs, forms for content element entry, metadata collection, reuse, syndication and aggregation feeds, and database models or schemas. Schemas for classifying information, called taxonomies, are often seen as the building blocks for information architecture.

J

Java Database Connectivity (JDBC): Developed by JavaSoft, a subsidiary of Sun Microsystems, JDBC is a Java API that enables Java programs to

execute SQL statements. Since Java itself runs on most platforms, JDBC makes it possible to write a single database application that can run on different platforms and interact with different DBMSs. It is also possible to allow Java programs to interact with any SQL-compliant database. JDBC is similar to ODBC, but is designed specifically for Java programs, whereas ODBC is language independent (see also ODBC).

JavaScript: Developed by Netscape, JavaScript is a relatively simple scripting language that can be used to integrate with HTML code and add interactivity to a Web page. JavaScript is intended to provide a quicker and simpler language for enhancing Web pages and servers, whereas Java, developed by Sun Microsystems, was built as a general-purpose object language. JavaScript is an open language supported by most current Web browsers.

K

Knowledge Management (KM): Knowledge management is a term associated with the processes for the creation, dissemination, testing, integration, and utilization of knowledge. In contrast to content management systems, KM systems focus on identifying, storing, and maintaining access to key bits of information relevant to an organization's objectives.

L

Link Resolver: Link Resolver is software used to connect a link source with potential target(s). Used primarily to access a range of electronic databases, it accepts the OpenURL from the link source and determines from what target(s) the cited article is available. The link resolver has access to a database, sometimes called a knowledge base, that records what journals particular group users have access to, from which sources, and within which date range.

Link Source: Link Source is software that can recognize when a user has a link resolver available, and display a special button or link on citations (such as SFX, Search LinkSource, or more information). When a link is clicked, the link source sends an OpenURL to the user's resolver identifying the

selected resource (database, e-journal, library catalog, Web search engine, interlibrary loan system, etc.).

Link Targets: Link targets are the resources the resolver can provide links to. These targets can be electronic journals, aggregated databases, interlibrary loan systems, abstracting services, citation indexes, and library catalogs.

M

Metadata: Metadata is data about data. Metadata is commonly used to identify information that describes a Web asset, most typically an **HTML** file. Metadata is data that accompanies a piece of content or an entire document. Written in HTML, metadata describe the content and provide optional information like a caption, abstract, or keywords for search engines. It could include a creation date, publication date, and expiration date. It could include copyright information and terms of use. Document metadata might include the full list of the Dublin Core ontology properties. It is usually stored in a relational database or an object-oriented database. Metadata that describes an HTML file might include the name of the author, the language the file is written in, the source of the file, the keywords that describe the file, and the audience the content is targeted for. Metadata is typically included in the HTML code of a given Web page.

Metasearch: Metasearch is a search of other searches. Often found as a feature in some search tools, it can be used to consolidate data from multiple sources. It can simultaneously search a wide variety of online resources, such as library catalogs, electronic journals, multiple databases, Web resources, and locally held materials. Metasearch engines often cluster the results around found terms in order to assist in narrowing the result list (see also Federated Search).

MySQL: MySQL is a computer software product, a multi-threaded, multi-user, SQL (Structured Query Language) relational database server. MySQL, as free software, utilizes the GNU General Public License. It is often used in conjunction with freely available programming languages like PHP to provide a platform for developing in-house applications. The PHP-MySQL combination is also cross-platform compatible.

O

Open DataBase Connectivity (ODBC): Developed by Microsoft Corporation, ODBC is a standard database access method that provides a set of functions to enable access to databases. ODBC supports access to both relational and non-relational databases. The goal of ODBC is to make it possible to access any data from any application, regardless of which database management system (DBMS) is handling the data. For this to work, both the application and the DBMS must be ODBC compliant, that is, the application must be capable of issuing ODBC commands, and the DBMS must be capable of responding to them.

Open Source: Open source means that the source code is openly available, and refers generally to any computer software whose source code is either in the public domain or, more commonly, is copyrighted by one or more persons/entities and distributed under an open source license such as the GNU General Public License (GPL). Freely distributable means that anyone may duplicate, modify, redistribute, and use freely under a license. Some examples of open source tools are: Apache, *BSD*, Emacs, GIMP, *GNU GNOME*, KDE, *Linux*, *Moodle*, *Mozilla*, *MySQL*, *OpenOffice.org*, *PHP*, *phpBB*, *Postnuke*, *TeX*, *VIM*, *XFree86*, and *Zope*.

OpenURL: OpenURL is a uniform format for passing metadata among link sources, link resolvers, and other applications involved in context-sensitive linking. Its primary application currently is to provide persistent linking to journal articles independent of the database housing them.

P

Parsing: Parsing is the process of checking an SGML-formatted document to ensure it has met all the rules of both SGML (Standard Generalized Markup Language) and the DTD (Document Type Definition). Technically, a document is not considered to be SGML until it has been successfully parsed, as defined by the ISO Standard for SGML.

Personalization: Personalization is the process of matching categorized content to users' needs. The personalization process occurs upon page request to a Web server, and content is presented based on user request.

Preferences can be maintained on the server and linked to either a cookie on the user's machine or tied to a login.

PHP: Initially known as Personal Home Page Tools, PHP is an open source script language that was designed specifically to generate dynamic Web pages. PHP code is embedded within HTML of a Web document. The PHP script runs on the server side, and performs any programs and operations specified in the script when a user requests the document; a dynamically generated HTML page is then delivered to the user. PHP runs as an optional module within the Apache Web server. It is an open source alternative to Microsoft's Active Server Pages (ASP), Sun's Java Server Pages (JSP), and the like.

Portal: In strict technical terms, a portal is a Web site containing links to other sites. Commercial portals typically include a search engine, free e-mail, instant messaging, chat, personalized Web page, and Web hosting, and can often be personalized or tailored to the individual user's needs. Specialized library portals may include channels or a variety of information sources dealing with library news, subject guides for example, along with links to the catalog and databases. There are corporate portals, business portals, and enterprise portals. Portal technology is widely used in Web content management. A portal template, like any CMS template, contains blocks where content elements or content objects will be placed. These elements may be HTML fragments, XML converted to HTML by XSLT, news feeds including JavaScript or XML RDF, images, applets, Flash files, etc. Collectively, these content objects are sometime called portlets.

Portlet: A portlet is any content object that appears in some block in a portal. It comes from applets and servlets, which might be providing the portlet content via a Web service.

PostScript: PostScript is Adobe's original page description language. PostScript files contain images of full pages (mixed text and graphics). The black-and-white image pages are scaleable, but not searchable. PostScript files are more compressed than bitmaps, so they take up much less storage space. PostScript files retain high resolutions for printing.

Practical Extraction and Report Language (PERL): PERL is a script programming language designed specifically for processing text. PERL combines syntax from several UNIX utilities and languages. It is widely used to write common gateway interface (CGI) programs—one method through which developers can provide dynamic interaction between users and Web sites, such as Web forms. PERL is a popular choice for

programming server-side tasks such as automatically updating user accounts and newsgroup postings, processing removal requests, synchronizing databases, and generating reports. Its text-processing prowess makes it a frequent choice for building homegrown content management systems.

Production Server: The production server is the active server computer that sits outside the organization firewall. Production servers are used to serve external audiences via a Web server application running on a production server. Production servers are often designed to be specialized servers serving different types of content. For example, an organization may have one server to serve all media assets and another to run all CGI scripts, and set up multiple production servers to serve different geographic regions (see also Development Server and Application Server).

R

Really Simple Syndication (RSS): RSS is part of a family of XML-based communication standards along with Rich Site Summary and RDF Site Summary. RSS can be understood as a Web syndication protocol that is primarily used by news Web sites and Weblogs. RSS allows a Web developer to publish his or her Web site in a format that a computer program can easily understand and digest. This allows users to easily repackage the content on their own Web sites or blogs, or privately on their own computers. RSS simply repackages a Web site as a list of data items, such as the date of a post, a description of the post, and a link to it. A program known as an RSS aggregator or feed reader can then check the RSS-enabled Web page for the user and display any updated articles that it finds. This is more convenient than having users repeatedly visit their favorite news Web sites, because it makes sure that readers only see material that they haven't seen before. Web-based RSS aggregators are also available, sparing users the inconvenience of downloading an application to their computers. Such readers can make a user's feeds available on any computer with an Internet connection.

Resource Description Framework (RDF): RDF is the tool for adding semantics (meaning) to content objects in Web pages. RDF is most often sent between computers using XML as its syntax. It allows parsers and inference engines to discover the meaning. RDF is usually used for library

catalogs, large directories, media asset repositories, and personal media collections like books, CDs, and digital images.

Reuse: Reuse describes a content element that can be propagated to more than one place in the site. Maintaining an element in one place insures uniformity and accuracy. Because style and layout are controlled by the associated template, the same information can be wholly integrated into each use. Note that a content element can be a complete templated object with many sub-elements, an RSS news feed, or simply an HTML or XML fragment from a database. SCORM (Shareable Content Object Reference Model) is a proposed standard for reusing content objects. RSS Feed

S

Sandbox: A sandbox is a special Web site running a CMS that allows anyone to sign in and to test drive the application. A sandbox user normally has all the same privileges as the administrator. A sandbox does, however, restrict a program to a set of privileges and commands that make it difficult or impossible for the program to cause damage to the host's data.

Script: Script is a type of computer code that can be directly executed by a program that understands the language in which the script is written. Scripts can run on both the development and production servers. Scripts are written in interpreted languages that can contain English statements for commands like PERL and JavaScript. Interpreted languages execute more slowly than compiled programs written in C or C++. Scripting is usually used for simple, lightweight applications and is typically much easier to write than a standard program written in C or C++.

Server Side Includes (SSI): Content that sits on the server itself is dynamically included in Web pages upon page request. Server Side Includes are used to make content modular so that a content component such as page header, footer, or navigation bar can be included on multiple Web pages, and content components can be changed in one place and automatically be reflected in all Web pages.

Staging Server: A staging server or development server is used in content management development or Web development to test, develop, or compile content or applications before they are deployed to a production or delivery server. The staging server should have the same characteristics

as the production server (see also Development Server and Production Server).

Standard Generalized Markup Language (SGML): SGML is an international standard. SGML is a text-based language, describing the content and structure of digital documents. It is a meta-language included in a file's document type definition. It specifies the rules for the tagging elements of a markup language, which in turn determines the formatting of the text. It tags documents as a series of data objects rather than storing them as huge files, thus it allows organizations to structure and manage information in a cross-platform, application-independent way.

Structured Query Language (SQL): SQL is a specialized programming language for defining, maintaining, and viewing information in a relational database. First commercialized in the early 1990s, SQL has evolved into a complete language for the management of complex data objects. It can be used for most industrial-strength and many smaller database applications. Each specific application will have its own version of SQL implementing features unique to that application, but all SQL-capable databases support a common subset of SQL.

T

Taxonomy: Taxonomy, as it relates to content management, refers to either a hierarchical classification of things or the principles underlying the classification. The structure, typically highly regimented, impacts the data model, directory structure, and file naming conventions for a given implementation of a content management system.

Template: A template identifies the presentation or layout of information content, including style and positioning to be used by the browser when displaying specified elements. The content elements or content objects could be HTML or XML fragments, RSS news feeds and other remote server-driven applications (servlets), or client-side application programs (applets). Pages on a site using the same template present a common look and feel.

V

Version Control: Version control in content management assists the administrator and users of the system in coordinating production and tracking older versions of content or code. In order to prevent files from accidentally being replaced by another user's changes, version control is typically done by requiring users to "check in" and "check out" files that they are working on. Version tracking archives track old versions of content, the content management system's source code, and other files, which can be retrieved and re-published to the live site. This is done using time and date stamping, and other data about a file to maintain it.

W

Wiki: "Wiki wiki" means "super fast" in the Hawaiian language. A wiki Web site enables documents to be authored collectively in a simple markup language using a Web browser. A wiki is an organic or growing collection of Web pages with novel linking structures. Each link is the name of a page in the wiki, with a special capital letter plus medial-cap syntax that makes it a "WikiWord." The wiki builds itself when pages are added that include a WikiWord. Standard wiki pages contain no HTML. They are simple text fields in the database. All HTML is generated, turning WikiWords into hyperlinks, adding unique wiki. A wiki feels like a glossary, but it is not, since the entries are special WikiWords. It is a CMS. It manages the content in pages with its simple but powerful tools.

Workflow: Workflow describes a set of interdependent tasks that occur in a specific sequence. Workflow as it applies to content management describes the process of breaking down a high-level task into a series of orderly, pre-defined steps that answer these questions: What needs to be done? Who needs to do it? Who needs to approve it? And when must it be completed? *Approval-centered* and *task-oriented* constitute two types of workflow management. An approval-centered workflow tracks the approval chain of command on a piece of content, no matter what the form—from a press release to a video clip of a speech. A task-oriented workflow centers on the tasks that must be done by different resources, or people, to complete the task.

Z

Zope: Zope is an object-oriented Web application server written in the programming language Python. It can be almost fully managed with a Web-based user interface. A Zope Web site is composed of objects, as opposed to files, as is usual with many other Web server systems. Zope maps URLs to objects using the containment hierarchy of such objects, and methods are considered to be contained in their objects as well.

Endnote

*   The glossary of terms included in this list is primarily drawn from the disciplines of content management and, by extension, Web content management. It is not intended to be a comprehensive list but a brief guide to some of the more commonly used technical terms found in the literature and in the case studies published here. Entries are listed alphabetically. Acronyms are used for entries wherever possible with the full term placed alongside in parentheses.

Many sources were used in the preparation of this glossary. Online dictionaries and encyclopedia referenced include CMSWatch (*http:// www.cmswatch.com/*), Edit-x (*http://www.edit-x.com/glossary.php*), and Wikipedia (*http://en.wikipedia.org/*).