

Glossary and Bibliography for Controlled Vocabularies

Glossary

AACR2 (Anglo-American Cataloging Rules, 2nd edition)

A content standard published by the American Library Association (ALA), Canadian Library Association (CLA), and Chartered Institute of Library and Information Professionals (CILIP). It is being replaced by **RDA**.

abbreviation

A shortened form of a name or term (for example, *Mr.* for *Mister*). *See also* **acronym** and **initialism**.

access point

An entry point to a systematic arrangement of information, specifically an indexed field or heading in a work record, vocabulary record, or other content object that is formatted and indexed in order to provide access to the information in the record.

acronym

An abbreviation or word formed from the initial letters of a compound term or phrase (for example, *MoMA*, for Museum of Modern Art). *See also* **initialism** and **abbreviation**.

ad hoc query

Also called a **direct query**. A query or report that is constructed when required and that directly accesses data files and fields that are selected only when the query is created. It differs from a **predefined report** or querying a database through a **user interface**.

administrative entity

In the context of a geographic vocabulary, a political or other administrative body defined by administrative boundaries and conditions, including inhabited places, nations, empires, nations, states, districts, and townships.

administrative data

In the context of cataloging art, information having to do with the administrative history and care of the work and the history of the catalog record (for example, insurance value, conservation history, and revision history of the catalog record). *See also* **descriptive data**.

algorithm

A formula or procedure for solving a problem or carrying out a task. In the context of this book, an algorithm is a procedure, formula, or rules in a computer program or set of programs, often expressed in algebraic notation, that follows a logical unambiguous step-by-step process to retrieve a set of results, solve a problem, make a decision, manipulate or alter data, or achieve some other result or state. Although a computer program may be considered one large algorithm, in common usage in computer science, the term typically refers to a small procedure applied recurrently. *See also* **program**.

alphanumeric classification scheme

A set of controlled codes (letters or numbers, or both letters and numbers) that represent concepts or headings and generally have an implied taxonomy that can be surmised from

the codes (for example, the Dewey Decimal System number 735.942). *See also chain indexing.*

alternate descriptor (ALT)

A variant form of a descriptor available for use; usually a singular form or a different part of speech than the descriptor (for example, *lithograph* is an alternate descriptor for the plural descriptor, *lithographs*). The relationship indicator for this type of term is ALT.

ancestor

In a hierarchy, any record that is a broader context for the record at hand, including parents, grandparents, and all other broader contexts at higher levels; any node in the succession of parent nodes on a path all the way up to the root. *See also descendent.*

antonym

A term that is the opposite in meaning or antithesis of another term (for example, *roughness* is an antonym for *smoothness*).

API (Application Programming Interface)

A software intermediary that allows two applications to talk to each other. For example, for the Getty vocabularies, APIs provide one way in which the vocabulary data is made available to end users.

application

Also called an **application program**. A software program designed to accomplish a task for an end-user (for example, word processing or project management), as distinguished from the **operating system** program that runs the computer itself.

application profile

A set of metadata elements, policies, and guidelines defined for a particular application or community. The elements may be from one or more element sets, thus allowing a given application to meet its functional requirements by using metadata from several element sets, including locally defined elements.

application programming interface (API)

In the context of this book, an online system, source code, and interface that a provider of data (for example, a vocabulary provider or library) employs to allow users to have access to the data. It may be language-dependent (designed for a specific programming language) or language-independent (working with multiple programming languages).

architect

A person or firm involved in the design or creation of structures or parts of structures that are the result of conscious construction, are of practical use, are relatively stable and permanent, and are of a size and scale appropriate for—but not limited to—habitable buildings.

architectural works

See built works.

architecture

Refers to the built environment that is typically classified as fine art, meaning it is generally considered to have aesthetic value, was designed by an architect, and was constructed with skilled labor. *See also built work.*

archival group

See **group**.

art

In the context of this book, refers to the *visual arts* such as painting, sculpture, drawing, printmaking, photography, ceramics, textiles, and decorative arts of the type and caliber generally collected by museums. Performance art is also included, but the performing arts are not. Note that these are works of visual art of the *type* collected by art museums. The objects themselves may actually be held by an ethnographic, anthropological, or other museum, or owned by a private collector.

artist

Any person or group of people involved in the design or production of visual arts that are of the type collected by art museums.

ascending order

In the context of a string of hierarchical parents, refers to the display of parents from narrowest to broadest (for example, *Columbus (Bartholomew County, Indiana, United States)*). See also **descending order**.

associative relationship

In a thesaurus, the relationship between concepts that are closely related conceptually, but the relationship is not hierarchical because it is not whole/part or genus/species. The relationship indicator for this relationship is RT (for *related term*). See also **equivalence relationship and hierarchical relationship**.

asymmetric relationship

In the context of a thesaurus, refers to a reciprocal relationship that is different in one direction than it is in the reverse direction, for example BT/NT. See also **symmetric relationship**.

authentication

A human or machine process that verifies that an individual, computer, or information object is who or what it purports to be.

authoritative source

A published source that is based on reliable documentary evidence that is accepted as true by most experts, and used as a standard source in a given discipline. The source may be online. Crowd-sourced resources are typically not considered authoritative, unless they were produced by experts only. For purposes of access, terminology from non-authoritative sources may be included in a vocabulary.

authority file

Also called simply an **authority**. A file, typically electronic, that serves as a source of standardized forms of names, terms, titles, etc. Authority files should include references or links from variant forms to preferred forms. The main purpose of an authority is to enforce usage, often requiring users to use only the preferred term for a given concept. Any type of vocabulary can be used as an authority. See also **controlled vocabulary and local authority**.

authority heading

A preferred, authorized heading used in a vocabulary, particularly a bibliographic authority file, typically including a string of names or terms with additional information as necessary

to allow disambiguation between identical headings (for example, *United States—History—Civil War, 1861-1865—Battlefields* and *United States—History—Civil War, 1861-1865—Campaigns*). The types of authority headings used by the U.S. Library of Congress are subject authority headings, name authority headings, title authority headings, name/title authority headings, and keyword authority headings. *See also* **heading**.

authorization

In the context of vocabularies, the process by which the creators of a vocabulary or an oversight group regulate the selection of terms and establishment of relationships in a controlled vocabulary. *See also* **warrant**.

automatic indexing

In the context of online retrieval, indexing by the analysis of text or other content using computer algorithms. The focus is on automatic methods used behind the scenes with little or no input from individual searchers, with the exception of relevance feedback. The results tend to be broad and imprecise, as contrasted to human **indexing**.

autoposting

See **up-posting**.

back-end database

A database that contains and manages data for an information system, distinct from the presentation or interface components of that system.

batch load

In the context of populating or contributing to vocabulary systems or other databases, refers to moving or manipulating a group of records as a single unit for the purpose of data processing, typically accomplished by the computer without user interaction, as contrasted to entering records manually, one at a time. *See also* **load** and **processing**.

best match

Also called a **weighted term ranking**. Refers to a variety of electronic term-matching and ranking methods that attempt to predict the potential relevance of query results by assigning relevance scores and ranking based on comparing search terms to the indexing terms of the target database.

blind reference

In the context of a vocabulary that is being used for indexing or retrieval on a defined data set, refers to a term in the vocabulary that is not linked to any content in the data set. End users should typically not receive blind references in a retrieval situation because they will result in a failed search; but these terms should be retained in structured vocabularies that are used for indexing because they may be needed in the future or in another context.

Boolean operators

Logical operators used as modifiers to refine the relationship between terms in a search. The four Boolean operators most commonly used are AND, OR, NOT, and ADJ (adjacent). They may be used with parentheses and other punctuation to form logical groupings of criteria in queries (for example, *(Castillo OR Rancho) AND Diego*).

bound term

A compound term representing a single concept, characterized by the fact that the words almost always occur together and the meaning is lost or altered if the term is split into its component words. *See also* **compound term** and **unbound compound term**.

brand name

A trade or proprietary name for a thing or process (for example, *Super Glue*).

broadcast searching

See **federated searching**.

broaden results

To adjust criteria in a search in order to retrieve a larger number of results, typically because the searcher did not find what he or she wanted in an initial narrower search. See *also* **narrow results**.

broader term (BT)

Also called a **broader context**. A vocabulary record to which another record or multiple records are subordinate in a hierarchy. In thesauri, the relationship indicator for this type of term is BT. Variations on the notation include BTG, (broader term generic), BTP (broader term partitive), BTI (broader term instance), BT1 (broader term level 1), BT2 (broader term level 2), etc.

browsing

The process whereby a user of a system or Web site visually scans and maneuvers through navigation lists, results lists, hierarchical displays, or other content in order to make a selection, as contrasted to the user entering a search term in a search box.

built work

An instance of architecture, which includes structures or parts of structures that are the result of conscious construction, are of practical use, are relatively stable and permanent, and are of a size and scale appropriate for—but not limited to—habitable buildings. Built works in the context of art information are manifestations of the built environment that is typically classified as fine art, meaning it is generally considered to have aesthetic value, was designed by an architect (whether or not his or her name is known), and constructed with skilled labor. See *also* **movable work**.

candidate term

Also known as a **provisional term**. A term under consideration for admission into a controlled vocabulary because of its potential usefulness.

cataloger

In the context of this book, the person who records information in records for works. See *also* **indexer and end user**.

cataloging

In the context of this book, the process of describing and indexing a work or image, particularly in a collections management system or other automated system. Cataloging involves the use of prescribed fields of information and rules (for example, the rules described in CCO and CDWA).

cataloging rules

See **editorial rules**.

cataloging tool

A system that focuses on content description and labeling output (for example, wall labels or slide labels), often part of a more complex **collection management system**.

CCO (Creative Commons zero)

A no rights reserved open license, intended for scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.

<https://creativecommons.org/share-your-work/public-domain/cc0/>

CCO (Cataloging Cultural Objects)

A data content standard for describing works of art, architecture, and material culture.

http://cco.vrafoundation.org/index.php/toolkit/cc0_pdf_version/

CDWA (Categories for the Description of Works of Art)

A set of metadata categories and recommendations that may be used to design information systems and to do cataloging for art, architecture, objects of material culture, and archaeological and archival materials.

http://www.getty.edu/research/publications/electronic_publications/cdwa/

CGI script

A computer program, most frequently written in C, Perl, or a shell script, that uses the Common Gateway Interface (CGI) standard and provides an interactive interface between a user or an external computer application and a World Wide Web server. CGI scripts are most commonly used to develop forms that allow users to submit information to a Web server.

chain indexing

Also called **chain procedure**. A technique for indexing that uses a numeric or alphanumeric classification system where the entries have meaning beyond simple numeric sequencing, such as the Dewey Decimal System (for example, in Dewey number 735.942, 735 means *sculpture after the year 1400 CE*, 9 means *geographic area*, 4 means *Europe*, and 2 means *England*).

Chenhall's Nomenclature

System for classifying man-made objects published by Robert G. Chenhall in 1978.

child

See **narrower term**.

CIDOC CRM (CIDOC Conceptual Reference Model)

An object-oriented ontology for the mediation and interchange of heterogeneous cultural heritage information. It was published in 2006 as an ISO standard (ISO 21127:2006).

<http://network.icom.museum/cidoc/standards/cidoc-standards-guidelines/L/12/>

classification

In the context of this book, the process of arranging works or other content objects systematically in groups or categories of shared similarity according to established criteria and using terms to identify the classes.

classification notation

In a vocabulary, a numeric, alphabetic, or alphanumeric code in a system of codes used to classify or categorize entries; may be used in a hierarchical arrangement to impose a

display or sorting order on the lines or levels in the hierarchy (for example, *V*, *V.PC*, *V.PE*).
See also notation.

classified display

See hierarchical display.

client

An application that retrieves and/or renders resources or resource manifestations. Often used to denote a computer or other kinds of devices connected to a network, equipped with software that enables users to access resources available on another computer connected to the same network, called a server. *See also server.*

clustering

In the context of automated data, usually refers to the process of grouping or classifying through automatic or algorithmic means rather than incorporating human judgment.

code

See computer code.

coined term

A new term invented for the purpose of translation, generally by literally translating the word or words of the term from the source to the target language, but without authoritative warrant in the target language for the usage of the term as a loan term.

collection management system (CMS)

A type of database system that allows an institution to control various aspects of its collections, including description (artist, title, measurements, media, style, subject, etc.) as well as administrative information regarding acquisitions, loans, and conservation information.

collection

In the context of cataloging art, refers to multiple works that are physically or conceptually arranged together, including the entire set of objects curated by a given museum or other repository.

complex term

A single phrase denoting more than two distinct concepts, which could be broken out and used independently, as defined by the Library of Congress. *See also bound term, compound term, and heading.*

component

In the context of cataloging art and architecture, a part of a larger item. A component differs from an *item* in that the item can stand alone as an independent work but the component typically cannot or does not stand alone (for example, a panel of a polyptych, a façade of a basilica). *See also item and group.*

compound term

A term consisting of two or more words. In the context of this book, mention of compound terms generally refers to bound terms, which are compound terms that represent a single concept (for example, *flying buttresses*). *See also bound term.*

computer code

Also called **code**. The machine-readable form, arrangement of data, and instructions of a computer program that are created when the computer program, which was written by a human programmer, is converted into binary code that can be read by the computer.

computer program

Also called a **program**. A specific set of instructions for ordered operations that will result in the completion of a task by the computer; a computer program consists of **computer code**. While the program is technically a type of data, computer programs are generally considered as separate from the data to which the programs refer (for example, data would be the terms, scope notes, etc. in a vocabulary record). A program is *interactive* if it acts when prompted by an action or information supplied by a user, or *batch* if it automatically runs at a certain time or under certain conditions, and then stops after the task is completed. A program is written in a **programming language**.

concept

In the context of the AAT and other thesauri comprising generic terms, the subject of the vocabulary record (i.e., the concept to which the terms refer), including abstract concepts, physical attributes such as shape, pattern, and color, style or period, activities, terms for performers of activities, materials, objects, and visual and verbal communication forms.

concept record

See **record**.

conceptual data model

An abstract model or representation of data for a particular domain, business enterprise, field of study, etc., independent of any specific software or information system; usually expressed in terms of entities and relationships. See *also* **logical data model**.

content object

In the context of a database, any entity that contains data. A content object can itself be made up of content objects. For example, a journal is a content object made up of individual journal articles, which are themselves content objects. See *also* **information object**.

contribution

In the context of controlled vocabularies, a term or record that is submitted for admission into a thesaurus or other vocabulary by an agency or individual outside the group responsible for maintaining the vocabulary; contributions are typically made by users of the vocabulary. See *also* **candidate term**.

controlled field

In the context of this book, a field in a record that is not free-text, meaning it is specially formatted and often linked to controlled vocabularies (authorities) or controlled lists to allow for successful retrieval. See *also* **free-text field**.

controlled format

Rules applied to the field regarding the types of values that may be included (for example, a controlled measurements' value field would allow only numbers). Fields may have controlled format in addition to being linked to controlled vocabulary, or the controlled format may exist in the absence of any finite controlled list of valid values.

controlled list

A simple list of terms used to control terminology. In a well-constructed controlled list, the following should be true: each term must be unique; terms should all be members of the same class; terms should not be overlapping in meaning; terms should be equal in granularity/specificity; and terms should be arranged alphabetically or in another logical order. A type of controlled vocabulary.

controlled vocabulary

An organized arrangement of words and phrases used to index content and/or to retrieve content through browsing or searching. A controlled vocabulary typically includes preferred and variant terms and has a limited scope or describes a specific domain.

co-occurrence mapping

Also called **co-occurrence clustering**. An automated method of compiling groups of terms that tend to occur together in certain contexts and are therefore presumed to be related in some way; the resulting groups of terms are considered to be loosely related and may be used to automatically broaden a user's search or to suggest alternative search terms to users in order to improve search results. *See also* **automatic indexing**.

core fields

Also called **core elements**. In the context of this book, the set of fields representing the fundamental or most important information required for a minimal record, whether the record is a work record or a vocabulary record. *See also* **required fields**.

corporate body

In the context of vocabularies discussed in this book, an organized, identifiable group of individuals working together in a particular place and within a defined period of time, whether or not they are legally incorporated (for example, architectural firms, artist studios, and art repositories).

criteria

In the context of this book, a specific set of limiting conditions used to create a query or select a subset of entries (for example, a *WHERE statement* in SQL). *See also* **variable**.

cross-database searching

See **federated searching**.

cross-reference links

See **syndetic structure**.

cross-language equivalence

When terms are in different languages in a multilingual thesaurus, each language may have descriptors, alternate descriptors, and used-for terms in its own language. The equivalence relationship of terms across languages is known as cross-language equivalence. *See also* **multilingual equivalency work**.

crosswalk

A chart or table (visual or virtual) that represents the semantic or technical mapping of fields or data elements in one database or a metadata framework, standard, or schema to fields or data elements that have a similar function or meaning in one or more other databases, frameworks, standards, or schemas (for example, the *artist* element in one standard may map to the *creator* element in another). *See also* **mapping**.

crowdsourcing

The act of soliciting ideas and labor from a group of people outside one's organization or institution. This primarily online exchange draws on collective knowledge to accomplish a task, and the work is typically undertaken for free or for minimal payment. *See also expert crowd sourcing and social tagging.*

cultural heritage

The total corpus of activities and the artifacts of activities that provide a record of the life of a culture.

cultural works

In the context of this book, art and architectural works, and other artifacts of cultural significance, including both physical objects and performance art. In related disciplines, the scope could be broader, also including the performing arts.

DACS (Describing Archives: A Content Standard)

A data content standard for describing archival collections.

<https://www2.archivists.org/standards/DACS>

data

In common usage in computer science, used as a singular noun to refer to information that exists in a form that may be used by a computer, excluding the program code. In other uses, *datum* is the singular and *data* is the plural, referring to facts or numbers in a general sense.

database

A structured set of data held in computer storage, especially one that incorporates software to make it accessible in a variety of ways. A database is used to store, query, and retrieve information, typically comprising a logical collection of interrelated information that is managed as a unit, stored in machine-readable form, and organized and structured as records that are presented in a standardized format in order to allow rapid search and retrieval by a computer.

database field

Also called a **data field**. A placeholder for a set of one or more adjacent characters comprising a unit of information in a database, forming one of the searchable items in that database. It is a portion of a structured record, especially a machine-readable record, containing a particular category of information (for example, *term* and *scope note* would be fields included in a vocabulary record). *See also field.*

database index

Also called a **data index**. A particular type of data structure that improves the speed of operations in a table by allowing the quick location of particular records based on key column values. Indexes are essential for good database performance. The concept is distinguished from **indexing** (human indexing) and **automatic indexing**.

database normalization

See normalization.

data content

The organization and formatting of the words or terms that form data values.

data content standard

Rules that determine the vocabulary, syntax, or format of content entered into data fields or metadata elements, for example, Anglo-American Cataloguing Rules (AACR), ISO 8601 (rules for recording date and time), Describing Archives: A Content Standard (DACS), Cataloging Cultural Objects (CCO).

data elements

The specific categories or types of information that are collected and aggregated in a database.

data index

See **database index**.

data pre-processing

See **preprocessing**.

data processing

See **processing**.

data provider

In OAI nomenclature, a resource, person, or organization that exposes metadata records in one or more repositories (specially configured servers) for harvesting by service providers.

data structure

A given organization of data, particularly the data elements, the logical relationships between data elements, and the storage allocations for the data.

data table

Sets of data that are organized in a grid or matrix comprising rows and columns.

data values

In the context of this book, the terms, words, or numbers used to populate fields in a work or vocabulary record. See *also* **data content**.

decoordination

In the context of a thesaurus, the splitting of a compound term into its component words to stand as individual terms. This typically would happen if a **compound term** had been added to the thesaurus, but was later determined not to be a **bound term**.

deep Web

See **hidden Web**.

default values

Values that are assumed or supplied automatically, for example, by a computer system, if a value is not specified.

derivation

Also called **modeling**. In the context of this book, the process of building a new vocabulary based on an existing vocabulary. In this approach, an appropriate controlled vocabulary is selected as a model for developing controlled terminology for local use, so that the local terms will be interoperable with the larger original vocabulary. See *also* **microcontrolled vocabulary and local authority**.

descendant

In a hierarchy, any record that is a narrower context for the record at hand, including children, grandchildren, and all other narrower contexts at all lower levels; any node in the succession of parent nodes on a path all the way down to the tips (leaves) of the hierarchies. *See also ancestor.*

descending order

In the context of a string of hierarchical parents, the display of parents from broadest to narrowest (for example, *Columbus (United States, Indiana, Bartholomew County)*). *See also ascending order.*

descriptive data

In the context of cataloging art, data intended to describe and identify a work, as contrasted to information necessary for administrative, technical, or accounting purposes. *See also administrative data.*

descriptor

In a thesaurus, the term recommended to represent the concept in displays and indexing. Also called the **main term**, **postable term**, or **preferred term** in a monolingual thesaurus. A multilingual thesaurus may have multiple descriptors (one in each language represented), but may possibly have only one preferred term for use as default in displays. *See also preferred term.*

diacritics

Also called **diacritical marks**. Signs or accent marks found over, under, or through alphabetic letters in many languages (for example, the umlaut in German, *München*) and used to indicate emphasis or pronunciation, often to distinguish different sounds or values of the same letter or character without the diacritical mark.

digital asset management system (DAMS)

A type of system for organizing digital media assets, such as digital images or video clips, for storage and retrieval. Digital asset management systems sometimes incorporate a descriptive data cataloging component, but they tend to focus on managing workflow for creating digital assets and managing asset rights, requests, and permissions.

digital surrogate

A digital "copy" of an original work or item, for example, a JPEG or TIFF image of a painting or sculpture or a PDF file of an article or book. In OAI nomenclature, digital surrogates are often referred to as "resources."

direct mapping

In the context of interoperability of vocabularies, refers to the matching of terms one-to-one in two controlled vocabularies. While the vocabularies need not be the same size (one may be smaller or larger) or cover exactly the same content, where overlap exists, there should be the same meaning and level of specificity between the two terms in each controlled vocabulary. *See also switching.*

direct query

See ad hoc query.

direct translation

See literal translation.

disagreement among sources

Know the level of currency and reliability of your sources. When two sources disagree, prefer the information obtained from the most scholarly, authoritative, recent source.

disambiguation

In the context of creating and displaying a vocabulary, the use of qualifiers, headings, or other methods to clarify and remove ambiguity between homographs (for example, *Smith, John (English printmaker, 1654-1742)* and *Smith, John (English architect, 1781-1852)*). See also **word sense disambiguation**.

discrete concept

In the context of a generic concept vocabulary, a discrete thing or idea as opposed to a subject heading, which often concatenates multiple terms or concepts together in a string. See also **concept**.

display field

A field intended for viewing by the end-user, typically showing data in natural language that is easily read and understood, and that can convey nuance and ambiguity. Display information may in some cases be concatenated from controlled fields; in other cases, this information is best recorded in **free-text fields**. See also **indexing**.

displayed index

An index that is visible and available to end-users for browsing. See also **non-displayed index**.

document

In the context of search and retrieval, the combination of a defined, primarily self-contained, machine-readable text or other information and the format in which it is housed.

dominant language

In the context of multilingual vocabularies, the more prominent or original language to which terms in other languages are mapped and in which other fields in the record (for example, scope note or date notes) are written. In a purely multilingual vocabulary, no language is dominant; but in a rich and complex vocabulary (for example, the AAT), a dominant language may be required for practical purposes.

download

See **load**.

Dublin Core Metadata Element Set (DCMES)

A set of 15 metadata elements that can be assigned to information resources, optimized for resource discovery on the World Wide Web. Also often used as a "lowest common denominator" in metadata mapping. <http://www.dublincore.org/documents/dces/>

dynamically generated

Refers to a Web page, metadata record, or other information object that is generated on demand, typically from content stored in a database, and usually either in response to a user's input or from dynamic data sources that are refreshed periodically. The expression "on the fly" is often used in relation to dynamically generated content.

EAD (Encoded Archival Description)

A data structure standard for encoding archival finding aids in SGML or XML according to the

EAD DTD or EAD XML schema, making it possible for the semantic contents of a hierarchically structured finding aid to be machine processed. <https://www.loc.gov/ead/>

editorial rules

In the context of this book, written rules and guidelines for creators or editors of vocabulary records that dictate how to populate fields and choose or interpret data. They should include which fields are required, how to choose appropriate values for various fields (for example, how to choose a preferred term), how to choose hierarchical positions, the format and syntax for each field, authorized sources, etc. Analogous rules for catalogers of works are called **cataloging rules**.

end user

In the context of this book, usually the searcher, client, or patron who retrieves, views, and uses the data in a vocabulary or work record, as distinguished from the editors or catalogers of the vocabulary or work record. In the context of systems design, the term refers to any client for whom a database system is designed and used; from that perspective, it could include the editors or catalogers for whom an editorial or cataloging system has been designed.

end-user thesaurus

A thesaurus designed for direct access by searchers rather than for use by indexers. Instead of controlling the terminology, the purpose of an end-user thesaurus is to help searchers find useful terminology for improving, narrowing, and broadening their queries. *See also* **indexer thesaurus**.

entity

In the context of computer science, a self-contained piece of data that can be referenced as a unit. In a more general sense, the term is used in this book to refer to a distinct person, place, or concept in a vocabulary.

entity-relationship model

A type of conceptual data model that represents structured data in terms of entities and relationships. An **entity-relationship diagram** can be used to visually represent information objects and their relationships. Because the constructs used in the entity-relationship model can easily be transformed into relational tables, this type of model is often used in database design.

entry array

A type of display, often used for headings, in which any two or more entries that have the same broader heading (for example, *Religious art—Ancient Egyptian*, *Religious art—Christian*, *Religious art—Hindu*, etc.) are grouped together vertically under the broader heading. While this is not a true hierarchical display, it may resemble a hierarchical display through use of indentation.

equivalence relationship

In a thesaurus, the relationship between synonymous terms or names for the same concept, typically distinguishing preferred terms (*descriptors*) and non-preferred terms (*variants* or *UFs*). *See also* **associative relationship and hierarchical relationship**.

equivalent term

A term that is considered an equivalent in search-and-retrieval, including not only true synonyms, but possibly also near-synonyms and any other terms that are considered

closely enough related to be useful in broadening a query; to narrow a query, exact equivalents could be used instead.

exact equivalence

The relationship between synonyms in one language or terms in different languages that have the same usage and meaning. *See also inexact equivalence.*

exact match

Electronic term-matching that produces a result that precisely matches the user's query term and does not implement automatic Boolean operators, truncation, proximity ranges, or stemming. In a strictly applied exact match, normalization will not be used, so that differences in punctuation, spacing, and diacritics will be maintained in the match. *See also best match.*

exhaustivity

In the context of cataloging and indexing, the degree of depth and breadth that the cataloger uses in assigning indexing terms or writing a description. Measures of greater exhaustivity include the use of a greater number of optional fields and the assignment of a greater number of indexing terms for each field. *See also specificity.*

expansion

See query expansion.

expert crowd sourcing

The act of soliciting ideas and labor from a group of people, limited only to those people having expertise in the content of the materials.

explode a hierarchy

To retrieve and display all the descendants of any given node, typically in a graphic display.

extension vocabulary

A thesaurus that is created with the intention of, or is later adapted for, linking to another vocabulary that is larger, broader, or more generic; it is typically linked through node linking, rather than being integrated at many points in the original vocabulary. *See also node linking, satellite vocabulary, and microcontrolled vocabulary.*

external node

See leaf node.

facet

Also called a **faceted display**. A fundamental, homogeneous, and mutually exclusive category of information in a thesaurus (for example, the AAT has seven facets: *Associated Concepts, Physical Attributes, Styles and Periods, Agents, Materials, Activities, Objects*).

facet indicator

See guide term.

false hit

Also called a **false drop**. In search and retrieval, an entry in a list of results that does not comply with the user's intended results.

federated searching

Also called **metasearching**, **broadcast searching**, **cross-database searching**, and **parallel searching**. Performing queries simultaneously across resources that are in different domains and created by different communities. Federated searching may involve searching across multiple databases, different platforms, and varying protocols, thus requiring the application of **interoperability** between resources and vocabularies.

field

In the context of this book, an area (often mapping to a metadata element in a metadata element set) in the user interface of a system where a discrete unit of information is displayed or the cataloger can enter information. Note: In this context, *field* is not necessarily equivalent to a **database field**.

field mapping

See **crosswalk**.

filing rules

A set of guidelines that determine how letters, numbers, spaces, and special characters will be processed when assembling an alphabetical or other listing. See *also* **sorting**.

finding aid

A descriptive tool widely used in archives. Finding aids typically take the form of hierarchical narrative descriptions of cohesive groups of archival records or collections of manuscript materials. Finding aids traditionally were paper documents; **EAD** is a structured way of expressing finding aids as machine-readable data.

first name

Also called a **given name**. In Western tradition, the name of a person that identifies that individual, typically unique in the immediate family and used with a last name (for example, *Richard* in *Richard Meier*). See *also* **last name** and **middle name**.

flat format

In the context of a thesaurus, an alphabetical display in which only one level of broader contexts and one level of narrower contexts are displayed for each focus record. See *also* **generic structure**.

flat-file database

A database with a data model designed around a single table, often a single file containing many records that all have exactly the same fields. It is a simpler model than the more highly structured relational and object-oriented models.

focus

Also known as **head noun** for terms, and a **trunk name** for proper names. In the context of a compound term, the noun component that identifies the class of concepts to which the term as a whole refers (for example, *buttresses* in the term *flying buttresses*). In the context of a modified name such as a place name, the part of the name that is not a modifier (for example, *Etna* in *Mount Etna*). See *also* **modifier**.

folksonomy

A neologism referring to an assemblage of concepts, which are represented by terms and names (called *tags*) that are compiled through **social tagging**, generally on the Web. A folksonomy differs from a **taxonomy** in that it is not structured hierarchically, and the authors of the folksonomy are typically the casual users of the content rather than

professional indexers following standard protocols and using standardized controlled vocabularies.

format

Used in two senses in this book: In the context of cataloging art, the configuration of a work, including technical formats, or the conventional designation for the dimensions or proportion of a work (for example, *cabinet photograph* or *IMAX*). In the context of computer science, the physical layout of a data storage device or the logical structure or composition of a file.

format control

See **controlled format**.

FRBR (Functional Requirements for Bibliographic Records)

A set of requirements and a conceptual entity-relationship model developed by the International Federation of Library Associations and Institutions (IFLA) to support bibliographic access and control. <https://www.oclc.org/research/activities/frbr.html>

free-text field

A field that may contain data entered without any vocabulary control or system-defined structure. It may be used to express ambiguity, uncertainty, and nuance in a note. See *also* **controlled field**.

free translation

A translation that conveys the sense of the source-language text in the target language, emphasizing the content rather than order of words or the phrasing. Free translation is the recommended method of translating most scope notes. However, the translator must be careful not to distort the meaning of the original text; in a free translation, the target-language translation should ideally not contain additions, deletions, or modifications of the source-language text. See *also* **literal translation** and **pragmatic translation**.

FRSAD (Functional Requirements for Subject Authority Data)

Previously known as *FRSAR (Functional Requirements for Subject Authority Records)*, a conceptual entity-relationship model developed by IFLA (International Federation of Library Associations and Institutions) and published in 2010.

FTP (File Transfer Protocol)

A TCP/IP protocol that allows data files to be copied directly from one computer to another over the Internet.

generic concept

In the context of this book, a concept in a vocabulary that is described by terms other than proper nouns or names (for example, the type of art work, such as *amphora*, or a material, such as *terracotta*). Generic concepts do not include proper names of persons, organizations, geographic places, named subjects, or named events.

generic posting

In controlled vocabularies, the use of narrower terms as *used for* terms for a descriptor that is really a broader term in the same vocabulary record. A generic posting is typically used as a time-saving strategy (or for other purposes) rather than making separate records for all the terms and linking them hierarchically. See *also* **up-posting**.

generic structure

A display format for a thesaurus in which all hierarchical levels are displayed by using indentation, codes, or punctuation marks. *See also flat format.*

genus/species relationship

Also called a **generic relationship**. A hierarchical relationship in which all children must be a *kind of, type of, or manifestation of* the parent. The genus/species relationship is the most common hierarchical relationship in thesauri and taxonomies, because it is applicable to a wide range of topics. *See also whole/part relationship and instance relationship.*

given name

See first name.

gloss

See qualifier.

grandparent

In a thesaurus, the level immediately above the parent of the focus record (for example, in *Columbus, Bartholomew county, Indiana, United States, Indiana* is the grandparent of *Columbus*).

granularity

See specificity.

group

Also called an **archival group** or **record group**. In the context of cataloging works, refers to an aggregate of items that share a common provenance.

group-level cataloging

Describing and assigning indexing terms for a group of works as a whole, typically focusing on the most important or most frequently occurring characteristics in the items of the group. *See also item-level cataloging.*

guide term

Also called a **node label**. A record represented by a term or phrase that is created as a hierarchical level to provide order and structure to thesauri by grouping narrower terms according to a given logic. Guide terms are not used for indexing and are often enclosed in angled brackets or otherwise distinguished from other terms in displays (for example, *<photographs by form>*).

harvester

In OAI nomenclature, a computer system that sends OAI-PMH requests to OAI data providers' repositories and harvests metadata records from them.

hardware

The physical components of a computer system, including mechanical, electronic, magnetic, and electrical components such as disks, disk drives, chips, electronic circuitry, keyboard, monitor, modem, and printer. *See also software.*

harmonization

In the context of vocabularies and standards, the process of preventing, minimizing, or eliminating technical and content differences and contradictions between standards or

vocabularies that have the same or similar scope or that must work interchangeably or in concert.

head noun

See **focus**.

heading

Also called a **label**. A string of words comprising a term combined with other information that serves to modify, disambiguate, amplify, or create a context for the main term in displays. Examples include the listing of qualifiers and/or broader contexts for terms (for example, *rhyta* (<*vessels for serving and consuming food*>, *containers*)), place types and administrative broader contexts for place names (for example, *Dayr al-Bahri* (*deserted settlement*) (*Qinā governorate, Egypt*)), or biographical information for people's names (for example, *Francesco Aliunno* (*Italian calligrapher, active 15th century*)). See also **subject heading list**, **name authority**, and **authority heading**.

hidden Web

Also called the **deep Web** or **invisible Web**. The sum of the Web pages that are not accessible to Web crawlers or robots, usually because they are either dynamically generated by a user querying a database, or are password-protected or subscription-based.

hierarchical display

Also called a **systematic display** or **classified display**. In a thesaurus, a graphic arrangement of terms showing broader/narrower relationships through the use of indentation, codes, or another method.

hierarchical relationship

The broader and narrower (*parent/child*) relationship between two entities in a thesaurus, namely whole/part (for example, *Montréal* is part of *Québec*), genus/species (for example, *bronze* is a type of *metal*), or instance relationships (for example, *Montréal* is an instance of a *city*). It is the basic structure that creates a **hierarchy**.

hierarchy

An organization of records related by levels of superordination and subordination. Each record in the hierarchy, except the root, is a narrower context of the record above it. See also **sub-facet**, **polyhierarchy**, and **monohierarchy**.

historical term

Also called a **historical name**. In the context of the vocabularies discussed in this book, a name or term that was used to refer to a person, place, subject, or concept in the past, but in current usage has been replaced with a different term or name (for example, historical names for St. Petersburg, Russia, are *Leningrad* and *Petrograd*).

hits

See **results**.

homograph

A term that is spelled the same as another term, but the meanings of the terms are different (for example, *drums* can have at least three meanings: components of columns, membranophones, or walls that support a dome). Homographs exist whether or not the terms are pronounced alike. Terms are generally considered homographs despite differences in capitalization, punctuation, or diacritics. A homograph may be a **homonym** or a **polyseme**.

homonym

Homonyms are terms that are spelled the same but have different meanings and unrelated origin. *Contrast polyseme.*

homophone

A term or word that is pronounced like another term or word but spelled differently (for example, *bows* and *boughs*). Traditional controlled vocabularies generally need not concern themselves with labeling homophones.

HTML (HyperText Markup Language)

An SGML-derived markup language used to create documents for World Wide Web applications. HTML has evolved to emphasize design and appearance rather than the representation of document structure and metadata elements.

human indexing

See **indexing**.

hyperlink

Also called a **hypertext link**. In the context of online information, an embedded link that connects different parts of an online document or data set to other parts of the document or to other documents. It is usually indicated by color or other emphasis applied to a word, phrase, icon, or symbol.

hypertext database

A dataset that resides as a collection of online documents with links joining various parts to each other, and access provided via an interactive browser.

Iconclass

A classification system designed for art and iconography. <http://www.iconclass.nl/home>

IIIF (International Image Interoperability Framework)

Framework that defines several application programming interfaces that provide a standardised method of describing and delivering images over the web, as well as presentation based metadata. <http://iiif.io/>

image

In the context of cataloging art, a visual representation of a *work*, typically existing in photomechanical, photographic, or digital format. In a typical visual resources collection, an *image* is a slide, photograph, or digital file.

indentation

Also called **indention**. In the context of printing or other displays of typed words or texts, refers to the white or blank space of a fixed width on a row or rows along the right or left margin of a display, as commonly used to indicate the first line in a new paragraph of text. Right-hand graduated indentation is used to indicate relationships between parents and their descendents in hierarchical displays of thesauri.

indexer

A person who assigns indexing terms for a work or image, typically the same person as the cataloger. See *also* **cataloger**.

indexer thesaurus

A thesaurus designed to control terminology and guide indexers in the choice of terms. *See also end-user thesaurus.*

indexing

Also called **human indexing**. In the context of this book, the process of evaluating information and designating indexing terms by using controlled vocabulary that will aid in finding and accessing the cultural work record. Refers to indexing done by human labor, not to the automatic parsing of data into a *database index*, which is used by a system to speed up search and retrieval.

indexing important information

The scope note is not an access point for retrieval. Therefore, if you mention a variant term, a date for usage of a term, or other important information in the scope note, index it in the appropriate fields elsewhere in the vocabulary record; for example add a variant term as a used-for term, or add the date of usage in the date fields for the descriptor.

inexact and partial equivalences

The relationship between synonyms in one language or terms in different languages that have similar or overlapping meaning and usage, but are not true synonyms (for example, *floating* and *flying*). *See also exact equivalence.* In some vocabularies, in cases where a suitable preferred term with the exact meaning and usage of the original term is not available in the second language, terms are sometimes considered equivalents when they have only inexact or partial matches in scope and meaning. For example, the English *science* and the German *Wissenschaft* have overlapping but not identical meanings. In the AAT, the equivalent terms must have identical meanings.

information object

A digital unit or group of units, regardless of type or format, that a computer can address or manipulate as a single discrete object. *See also content object.*

initialism

A set of initials that stand for the full form of a name (for example, *MFA*, for Museum of Fine Arts). *See also acronym* and *abbreviation*.

instance relationship

A hierarchical relationship in which all children must be an example of a broader context, most commonly seen in vocabularies where proper names are organized by general categories of things or events (for example, if the proper names of mountains and rivers were organized under the general categories *mountains* and *rivers*).

interoperability

In the context of controlled vocabularies, the ability of two or more vocabularies and their systems or components of their systems to map to each other's data, with the goal of exchanging information or enhancing discovery.

Internet

A global collection of computer networks that exchange information by the TCP/IP suite of networking protocols.

inverse document frequency (IDF)

An automatic ranking method often used in a formula with **term frequency** in information retrieval and text mining to estimate how important a term is to a set of data and how useful it will be in retrieval.

inverted form

In the context of a controlled vocabulary, the indexing form of a multiple-word name or term, where the last name or trunk portion of the term is listed first, followed by a comma and the descriptive word (for example, *Wren, Christopher*, or *buttresses, flying*). *See also natural order form.*

inverted index

See permuted index.

invisible Web

See hidden Web.

information retrieval database

Also called an **IR database**. Any database designed primarily for discovering and retrieving information. The systems that work with IR databases provide a search interface to permit users to compose queries, methods for searching through the target data, viewable or behind-the-scenes indexes, and results displays.

ISO (International Organization for Standardization)

A worldwide voluntary, non-treaty network of national standards institutes of approximately 160 countries, one member per country. The standards bodies work in partnership with international organizations, governments, industry, business, and consumer representatives to reach consensus, set standards, and promote their use with the goal of facilitating trade and meeting the broader needs of society.

item

In the context of cataloging art, an individual object or work. *See also group.*

item-level cataloging

Describing and assigning indexing terms for individual items in a collection of works. *See also group-level cataloging.*

jargon

A characteristic terminology of a particular group or discipline that is typically not understood by a more general audience.

JSON (JavaScript Object Notation)

A lightweight data-interchange format. <https://www.json.org/>

keyword

In the context of vocabularies, a verbal unit or word of a term that may be used in a search expression (for example, for the place name *Sena Julia*, *Sena* is one keyword and *Julia* is another). In the broader context of online retrieval, any significant word or phrase in the title, subject headings, or text associated with an information object.

keyword index

An index based on individual words (keywords) found in a vocabulary term, text, or other content object.

KeyWord In Context (KWIC)

A type of automatic indexing in which each word in a text, title, subject heading, string of words, or term becomes an entry word in the index, with the exception of stop-words. Variations on KWIC are KWOC (KeyWord Out of Context) and KWAC (KeyWord Alongside Context).

knowable vs. unknowable information

Editors should not express information that is not included in authoritative sources. Do not express information based upon your own opinion or observations. On the other hand, take care not to imply that a fact is unknown simply because you happen not to know it (generally because time and editorial priorities do not allow you to do the research required to resolve the issue). Write from the sources' point of view. If a fact is knowable, but just not known by you, omit it entirely rather than stating it with qualifying phrases such as probably, because using probably implies that the sources have stated it thus.

KOS (Knowledge Organization Systems)

A generic term used in knowledge organization to refer to authority files, classification schemes, thesauri, topic maps, and ontologies. KOS are used to organize materials for the purpose of retrieval and to manage a collection.

label

See **heading**.

language model

A type of automatic indexing based on term weighting and relevance prediction that attempts to predict probable query search terms based on term frequencies within documents and the inverse document frequency of terms across the target data. It is similar to the **probabilistic model**.

last name

Also called a **surname**. In Western tradition, the family name used with a first name to identify a person (for example, *Meier* in *Richard Meier*). See also **first name** and **middle name**.

latent semantic indexing (LSI)

A form of automatic indexing based on co-occurrence clustering of terms in combination with content that is associated with these clusters of terms; it attempts to partially address the problem of the variety of terms that can be used to express similar concepts.

latinization

See **romanization**.

lead-in term

See *used for term*.

leaf linking

See **node linking**.

leaf node

Also called an **external node**. In a thesaurus, a node that has no children, as with the ends or tips of hierarchical trees.

legacy data

Data that has been inherited from an old system or old data set, typically in reference to the data now being repurposed or transferred.

legacy system

An information system that has been developed and modified over a period of time and has become outdated and difficult and costly to maintain but that holds important information and involves processes that are deeply ingrained in an organization. Legacy systems usually are eventually replaced by a new hardware and software configuration.

lexeme

A fundamental unit of the words of a language, around which may be clustered a set of words that are different forms of the same word (for example, *paint*, *paints*, *painted*).

lexical unit

Also called a **lexical item**. One or more words that refer to a single concept (for example, *flying buttresses* or *bills of sale*). *See also compound term and bound term.*

lexical variant

A term that is a different word form for another term, such as spelling differences, grammatical variation, and abbreviations (for example, *watercolor* and *water-colour*). Lexical variants are considered as and grouped with synonyms in a vocabulary record, but they technically differ from synonyms in that synonyms are different terms for the same concept. *See also synonym.*

link

In the context of this book, any relationship between two vocabulary records, two works, a work and image, or a work or image and an authority. *Compare to hyperlink.*

Linked.Art

A new initiative, spearheaded by the Getty with other institutions; it is a “community working together to create a shared Model based on Linked Open Data to describe Art.” The model will be implemented in software and used to provide valuable content. As of this writing, the project is under active development and additional partners and collaborators are welcome. <https://linked.art/>

linked data

A method of publishing structured data so that it can be interlinked and become more useful through semantic queries.

Linked Open Data (LOD)

Data that is *linked* by using a method of publishing structured data so that it can be interlinked and useful in semantic queries. LOD is also *open*, meaning it is freely available to everyone to use and republish as they wish, with no or few restrictions from copyright or other mechanisms of control.

literal translation

Also known as a *direct translation*, is in its purest form a word-for-word translation of a text from the source language to the target language. It may be used where the basic structure and syntax of the source language is the same as the target language. Literal translation is generally not recommended for AAT scope notes, because the result may be stiff, and may convey idioms, grammar, and even meaning incorrectly in the target-language text. *See also free translation and pragmatic translation.*

literary sources

If certain information in a scope note or other text ultimately comes from a literary or mythological source, as opposed to a modern authoritative source, editors should avoid stating it as if it were a proven fact (e.g., references from the Bible, legends, etc.). At the same time, refer to religious texts in a neutral and respectful tone. Distinguished from

literary warrant.**literary warrant**

Justification for the inclusion of a term in a vocabulary based on published evidence that is sufficient to prove that the form, spelling, usage, and meaning of the term are widely agreed upon in authoritative sources. *See also* **source**, **organizational warrant**, and **user warrant**.

load

The process of moving or transferring files or software from one disk, computer, or server to another disk, computer, or server. To **upload** means to transfer files from a local computer to a remote computer; to **download** means to transfer from a remote computer to a local one.

loan term

Also called *loan word*. In the context of a given language, a word or words that are taken directly from another language (for example, *sotto in su*, an Italian phrase used in English to mean painted in correct perspective as if viewed from below).

local authority

An authority developed for local use, often compiled from one or more standard authoritative published vocabularies, but enforcing preferences and usage pertinent for the local setting. *See also* **authority**.

locator

In a bibliographic index, the part of an index entry that indicates the location of the book, page, or other resource. In an online index, it may be a hyperlink to the source.

logical data model

A data model that includes all entities and the relationships among them based on the structures identified in a conceptual data model, and that specifies all attributes for each entity. The data is described in as much detail as possible, without regard to how it will be implemented in a specific database. *See also* **conceptual data model**.

logical record

See **record**.

main term

See **descriptor**.

manual indexing

See **indexing**.

mapping

A set of correspondences between terms, fields, or element names, used for translating data from one standard or vocabulary into another, or as a means of combining terms or data for search and retrieval.

MARC (Machine-Readable Cataloging format)

A set of standardized data structures for describing bibliographic materials that facilitates cooperative cataloging and data exchange in bibliographic information systems.

<https://www.loc.gov/marc/>

markup language

A formal way of annotating a document or collection of digital data using embedded encoding tags to indicate the structure of the document or data file and the contents of its data elements. This markup also provides a computer with information about how to process and display marked-up documents (for example, HTML, XML, and SGML are examples of standardized markup languages).

material culture

A term referring to art together with the broad realm of physical objects and edifices produced by a culture. *See also* **cultural heritage**.

memory institution

A generic term used to describe an institution that has a responsibility to collect, care for, and provide access to the human record—for example, museums, libraries, and archives.

metadata

A structured set of descriptive elements used to describe a definable entity. This data may include one or more pieces of information, which can exist as separate physical forms. In the context of art information, metadata could include includes data associated with information about the creation, physical characteristics, history, location, administration, or preservation of the work.

metadata mapping

A formal identification of equivalent or nearly equivalent metadata elements or groups of metadata elements within different metadata schemas, carried out in order to facilitate semantic interoperability.

metadata mining

The automated extraction of metadata from electronic documents.

Metaphone

A phonetic algorithm for matching terms and names by sound, as pronounced in English, by translating words into a standard code or representation. It was developed by Lawrence Philips to address the perceived deficiencies in the Soundex algorithm. Metaphone and its later improvements are available as a built-in operators in a number of systems. *See also* **Soundex**.

metasearching

See **federated searching**.

METS (Metadata Encoding Transmission Schema)

A standard for encoding descriptive, administrative, and structural metadata relating to objects in a digital library, expressed in XML. METS enables the "packaging" of complex digital objects that include a range of metadata as well as related digital surrogates.

<http://www.loc.gov/standards/mets/>

microcontrolled vocabulary

Also called a **microthesaurus**. A controlled vocabulary that is limited in the range of topics covered, but fits within the domain of a larger, broader, or more generic controlled vocabulary. It typically contains highly specialized terms that are not necessarily in the broader controlled vocabulary, but that map to the hierarchical structure of the broader controlled vocabulary. *See also* **satellite vocabulary** and **extension vocabulary**.

middle name

In Western tradition, any name for a person placed before the last name (surname) but after the first name (for example, *Alan* in *Richard Alan Meier*). *See also* **first name** and **last name**.

minimal description

In the context of cataloging art, a record containing the minimum amount of information in the minimum number of fields or metadata elements.

modeling

See **derivation**.

modifier

In a compound term or name, the adjectival component that modifies the noun (for example, *flying* in *flying buttresses*; *Mount* in *Mount Etna*). *See also* **focus**.

MODS (Metadata Object Description Schema)

An XML schema for bibliographic records, developed and maintained by the Library of Congress. <http://www.loc.gov/standards/mods/>

monohierarchy

A hierarchy in which each child has only one immediate parent. Distinguished from a **polyhierarchy**.

monolingual

Expressed in a single language, as distinguished from **multilingual**. In a monolingual thesaurus, the terms and names are expressed in only one language.

movable work

In the context of cataloging art, any tangible object capable of being moved or conveyed from one place to another, as opposed to real estate or other buildings. Distinguished from **built work**.

multilingual

Expressed in more than one language, as distinguished from **monolingual**. In a multilingual thesaurus, terms and other information may be expressed in more than one language.

multilingual equivalency work

In translating terms, the process of finding terms in a target language that are equivalent in meaning and usage to terms in the source language. *See also* **cross-language equivalence**.

N3 network

The IP network that provides the NHS with high availability, high speed broadband network services.

name authority

An authority containing proper names, most often personal names. *See also* **subject authority list**.

namespace

The set of unique names used to identify objects within a well-defined domain, particularly relevant for XML applications. An XML Namespace is a W3C recommendation for providing uniquely named elements and attributes in an XML instance. A namespace is declared using the reserved XML attribute `xmlns`, the value of which must be a URI (Uniform Resource Identifier) reference. For example, the Dublin Core Metadata Element Set, Version 1.1 (original 15 elements) has the approved DCMI namespace URI as <http://www.dublincore.org/documents/dcmi-namespace/>

narrow results

To adjust criteria in a search in order to retrieve a smaller number of more precise results that better match the intention of the searcher.

narrower term

Also called **narrower context**. A record to which another record or multiple records are superordinate in a hierarchy (for example, *Brewster chair* is a narrower term to *armchair*). In thesauri, the relationship indicator for this type of term is NT. Variations on the notation include NTG, (narrower term generic), NTP (narrower term partitive), NTI (narrower term instance), NT1 (narrower term level 1), NT2 (narrower term level 2), etc.

natural language

Spoken or written texts, as distinguished from fielded data and controlled vocabulary.

natural order form

In the context of a controlled vocabulary, the form of a multiple-word name or term, where the name or term is not inverted (as may be appropriate for an index) but appears in the form that would be used in speech or a written text (for example, *Christopher Wren* or *flying buttresses*). *See also* **inverted form**.

navigation

In the context of search and retrieval, the facility that allows users to move through a controlled vocabulary or other content object by using pre-established links or relationships.

near synonymy

Also called **quasi-synonymy**. The characteristic of a term with meaning that is regarded as different from another term, but the terms are treated as equivalents for the purposes of broadening retrieval. *See also* **true synonymy**.

neologism

A term that has been newly invented or an existing term to which a new meaning is applied, often arising in the professional literature of a discipline.

nesting

The way in which subelements may be contained within larger elements, resulting in multiple levels of metadata.

nickname

A familiar, affectionate, derogatory, or humorous name that is used to refer to a person, place, or corporate body as a replacement for, or in addition to, the real or official name (for

example, *Masaccio*, meaning "big Tom," is a nickname for the painter *Tommaso Guidi*). (In the case of *Masaccio*, in ULAN it is the preferred name based on literary warrant.) *See also pseudonym*.

NISO (National Information Standards Organization)

A non-profit association that is accredited by the American National Standards Institute (ANSI), and identifies, develops, maintains, and publishes technical standards to manage information.

node

In the context of a thesaurus, any point or record in the hierarchy that is a location at which a branch or individual record (leaf) is attached; thus, the basic conceptual unit used to build hierarchies.

node label

See guide term.

node linking

Also called **leaf linking**. In the context of combining multiple vocabularies, a method that uses various nodes in the hierarchical structure of a source controlled vocabulary to link to more detailed controlled vocabularies that are applicable to a single node of the parent hierarchy. The vocabulary linked to a broader vocabulary in this way is often called an **extension vocabulary**.

nomen

As defined by FRSAD, *nomen* is an entity at the same logical level as Work, Expression, etc., and Person. All are subclasses of **Res**. Nomen is defined as Whatever appellation is used to refer to any entity found in the bibliographic universe, depending on the context of use, the same sequence of symbols can be assigned as a nomen of different entities in the real world even within the same language (polysemy and homonymy). Conversely, the same entity can be referred to by any number of nomens (synonymy). The association of nomens to entities is in general many-to-many.

non-displayed index

An machine-readable index that is not displayed for browsing or other direct access of end-users, but is used behind the scenes to improve accuracy or speed in search and retrieval. Such indexes may be created beforehand or on the fly at the time of the query. *See also displayed index*.

non-leaf node

Also called an **internal node**. In a hierarchy, nodes that link to one or more narrower contexts. *See also leaf node*.

non-equivalence

In mapping one vocabulary to another, the situation where there is no exact match, no term in the second language has partial or inexact equivalence, and there is no combination of descriptors in the second language that would approximate a match. *See also exact equivalence and inexact equivalence*.

non-preferred parent

In a polyhierarchical thesaurus, any parent that is not the parent chosen as preferred for the default for displays. *See also preferred parent*.

non-preferred term

Also called a **non-preferred name**. Any term in a vocabulary record that is not the **preferred term**, which is the term flagged as preferred for use as default in displays.

normalization

In the context of vocabulary retrieval, normalizing terms through a process of converting a term to its simplest form by removing case sensitivity, spaces, punctuation, and diacritics. It differs from **database normalization**, which is the process of reducing a complex data structure into its simplest structure, a technique to eliminate data redundancy, converting Unicode text into a standardized form, among other things.

notation

For a thesaurus, the alphabetic code used to express term types (D, ALT, UF), associative relationship (RT), hierarchical relationships (BT, NT, BTG, NTG, BTP, NTP, BTI, NTI, BT1, BT2, NT1, NT2), and scope notes (SN), among others. *See also classification notation.*

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

A protocol used to harvest or collect metadata records from data providers.

<http://www.openarchives.org/pmh/>

N-triples

N-Triples is an easy to parse line-based subset of **Turtle**. *See also triples.*

object

See work.

object-oriented

A programming or data modeling methodology that utilizes the notion of classes and their properties. Members (or instances) of a class share the same properties—for example, color or weight (however, note that although members of a class all share the same properties, the values of those properties do not need to be the same). Classes can contain subclasses, members of which inherit the properties of the parent or "superclass."

ODC-by (Open Data Commons Attribution License v1.0)

A license agreement intended to allow users to freely share, modify, and use a given database, subject only to specified attribution requirements.

<https://opendatacommons.org/licenses/by/1-0/>

online catalog

In the context of art information, a type of system used by end users to search for and view data and images.

ontology

A formal, machine-readable specification of a conceptual model, in which concepts, properties, relationships, functions, constraints, and axioms are all explicitly defined. While an ontology is not technically a controlled vocabulary, it uses one or more controlled vocabularies for a defined domain and expresses the vocabulary in a representative language that has a grammar for using vocabulary terms in an automated way to express something meaningful.

open licensing

Licensing methods in which data is made freely available to everyone to use and republish as they wish, with no or few restrictions from copyright or other mechanisms of control. Examples are **ODC-by** and **CCO**.

operating system

Also called an **operating system program**. A software program that runs the computer itself, as distinguished from an **application** program, which is designed to accomplish a task for an end-user (for example, word processing).

operational specificity

An automated method that attempts to predict the specificity of terms in a domain based on the number of postings or links to that term in a content object (for example, a term that is linked to very few content objects is predicted to be highly specific).

optical coincidence system

See **peek-a-boo system**.

organizational warrant

Justification for the inclusion of a term in a vocabulary based on the specialized requirements or jargon of the group or organization that is creating or sponsoring the vocabulary. See *also* **literary warrant** and **user warrant**.

orphan term

In a thesaurus, a record that has no associative or hierarchical relationship to any other term in the thesaurus. Orphan terms should be avoided in best practice thesaurus construction.

orthography

Correct or proper spelling and form of a word or words, including capitalization, diacritics, and punctuation, based on standard usage or convention.

paradigmatic relationship

Also called a **semantic relationship**. A relationship between terms or concepts that is permanent and based on a known definition.

parallel searching

See **federated searching**.

parent

See **broader term**.

parent string

The display of hierarchical parents in a horizontal string, as distinguished from vertical indented displays or displays using notation.

parenthetical qualifier

See **qualifier**.

parsing

In processing data, a process where data is broken or filtered into smaller, more distinct units.

partial equivalence

See **inexact equivalence**.

partitive relationship

See **whole/part relationship**.

patronymic

Also called a **patronym**. A word or words used with a given name to identify the person, common in early Western personal names when last names were uncommon (for example, *Bartolo di Fredi* means "Bartolo, son of Fredi"); may also refer to a surname derived from a paternal ancestor (for example, *Robinson* means "son of Robin").

permuted index

A type of index where individual words of a term are rotated to bring each word of the term into alphabetical order in the term list. See *also* **inverted form**.

phonetic matching

a process by which terms are matched to other terms that are presumed to sound like the original term, in an attempt to compensate for users' misspellings or general variation in spelling of names or terms (for example, *Meier* and *Meyer* are pronounced alike). A *phonetic algorithm* is the algorithm used for indexing of words by their pronunciation, such as Soundex, Metaphone, and others.

physical feature

In the context of geographic information, a characteristic of the Earth's surface that has been shaped by natural forces, including continents, mountains, forests, rivers, and oceans. See *also* **administrative entity**.

pick list

A user interface feature that allows the user to select from a pre-set list of terms, typically used to control vocabulary for indexing or to provide options in a query. A pick list is generally populated with a **controlled list**.

polyhierarchy

A thesaurus in which any record may be linked to multiple parent records.

polyseme

Polysemes are terms that are spelled the same and have related etymology and related, but not identical, meaning. In a dictionary, polyemes are typically listed under a single heading, with several definitions. In the AAT, polysemes would comprise separate records. *Contrast homonym*.

postable term

See **descriptor**.

postcoordination

The process of combining two or more terms at the time of retrieval rather than at the indexing stage; usually uses the Boolean operators "and," "or," or "not" (*Baroque and cathedral*) in formulating a query. See *also* **precoordination**.

postings specificity

See **operational specificity**.

posting

In the context of indexing, any instance of a given indexing term having been assigned to records, documents, or other content objects. Formulas used for predicting the usefulness of terms or methods of retrieval may count the number of postings relative to the target content objects or use the numbers of postings in other statistics.

pragmatic translation

A translation that allows certain additions or explanations in order to convey the meaning efficiently to the readers of the target language. Pragmatic translation may be used for scope notes as necessary.

Pragmatic translation is particularly important where the culture and tradition of the target-language readers is different from those of the source-language readers.

precision

A measure of a search system's effectiveness in terms of retrieving only relevant results, expressed as the ratio of relevant records or documents retrieved from a database to the total number retrieved in response to the query. A high-precision search means that most of the results retrieved will be relevant; however, a high-precision search will not necessarily retrieve all relevant results. *Recall* and precision are inverse ratios. When one goes up, the other goes down. *See also recall.*

precoordination

The formulation of a compound term or multiword heading at the time of indexing, rather than at the time of retrieval. An example of a precoordinated term is *Baroque cathedrals*; an example of a precoordinated heading is *United States—History—Civil War, 1861-1865*. *See also postcoordination.*

predefined report

A report for which the query and the output have been written and made available for repeated use by users; users may be allowed to enter variables that are plugged into the report. *See also ad hoc query.*

preferred term

Also called a **preferred name**. The term designated among all synonyms or lexical variants for a concept to be used as the default term to represent the concept in displays and other situations. In a monolingual thesaurus, the preferred term is also the only *descriptor* in the record. In a multilingual thesaurus, there may be a descriptor for every language, but there is often only one preferred term for the record as a whole. *See also descriptor.*

preferred flag

A designation indicating that a term or other data instance is preferred over others of the same type in a record. In addition to a *preferred term* for the record overall, there may be a *preferred-indexing name* flag for the inverted order version of the term, a *preferred-display name* for the natural order form of the name, a *preferred role* or *preferred place type* flagged among a list of roles or place types, and so on.

preferred parent

In a polyhierarchical thesaurus, the broader context that is chosen as conceptually preferred or to serve as the default in hierarchical displays. *See also non-preferred parent.*

preprocessing

Also called data preprocessing. Preliminary processing or transformation of data in order to facilitate further processing, parsing, etc.

probabilistic model

An automatic relevance and weighting method in which terms in a text or other content object are modeled as random variables, so that term frequency and distribution are used to predict the probability of relevance. *See also language model.*

procedure

Also called a **subroutine** or **subprogram**. A relatively independent portion of computer code within a larger computer program that performs a specific task in a series of steps.

processing

Also called **data processing** or **information processing**. The manipulation or transformation of data through a series of operations. In *batch processing* the operations are grouped together in batches and performed automatically; in *interactive processing*, the operations are prompted by input from a human programmer or user. *See also computer program.*

program

See computer program.

programming language

A formal language, defined by syntactic and semantic rules; used to write instructions that can be translated into machine language and then executed by a computer (for example, SQL, C++, C#, Java, Perl).

protocol

A specification—often a standard—that describes how computers communicate with each other, for example, the TCP/IP suite of communication protocols or the OAI-PMH.

provisional term

See candidate term.

pseudonym

A false or fictitious name, especially one assumed by an artist, author, or other person to maintain anonymity, to designate an identity for a particular activity, or for another reason (for example, *Le Corbusier* is a pseudonym assumed by the architect, *Charles Édouard Jeanneret*). *See also nickname.*

punctuation

In the context of vocabulary terms, the marks from standard written communication used to clarify, organize, or indicate how a word or words should be read (for example, hyphen, comma, period, quotation mark, parentheses).

qualifier

A word or phrase used to distinguish a term in a vocabulary from otherwise identical terms that have different meanings. A qualifier is separated from the term, generally displayed within parentheses. It is also called a **gloss**, although strictly speaking a qualifier should be used only with homographs, and a gloss has a more general meaning in the field of linguistics. *See also homograph.*

quasi-synonymy

See near synonymy.

query

Also called a **search**. In the context of retrieval, a command to look in a database and find records or other information that meet a specified set of criteria (for example, *select subject_id from term where normalized_term like 'A%' and historic_flag = 'H';*). The most precise queries are those that return the fewest false drops.

query expansion (QE)

Reformulating a query in order to return a broader or more comprehensive set of results (for example, adding synonyms to the user's search term expands a query).

RDA (Resource Description & Access)

Bibliographic cataloging standard that will replace **AACR2**. <http://www.loc.gov/aba/rda/>

RDF (Resource Description Framework)

An application of XML that enables the creation of rich, structured, machine-readable resource descriptions. It is a standard model for data interchange on the Web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed. RDF represents information using **triples**, each of which consists of a subject, a predicate, and an object. Each of those items is expressed as a Web URI.

<https://www.w3.org/2001/sw/RDFCore/>

RDF schema

A set of semantics within a defined namespace for use with specific applications of RDF.

recall

A measure of a search system's effectiveness in terms of retrieving all results that are possibly relevant, expressed as the ratio of the number of relevant records or documents retrieved over all the relevant records or documents. A high recall search retrieves a comprehensive set of relevant results; however, it also increases the likelihood that marginally relevant content objects will also be retrieved. Recall and precision are inverse ratios. *See also* **precision**.

reciprocity

In reference to vocabulary records, the characteristic of a two-way relationship in which both entities have mutual dependence, action, or influence on each other. Semantic relationships in controlled vocabularies must be reciprocal, meaning each relationship from one record to another must also be represented by a reciprocal relationship in the other direction. Reciprocal relationships may be *symmetric* (for example RT / RT) or *asymmetric* (for example BT / NT).

record

Also called a **logical record**. In the context of this book, a conceptual arrangement of fields referring to a vocabulary concept or a work; not the same thing as a **database record**, which is one row in a database table or another set of related, contiguous data.

record group

See **group**.

related term (RT)

A concept that is associatively (not hierarchically) linked to another concept in a thesaurus. In thesauri, the relationship indicator for this type of term is RT. *See* **associative relationship**.

relational table database

Also called a *relational database*. A database in which data is organized into columns and rows according to specific defined relationships (for example, in a vocabulary database, a table of terms may be linked to a table for languages).

relationship

In the context of this book, a link between two types of data, two records, two files, or any two entities of the same or different types in a system or network. *See also link*.

relationship indicator

A word, code, or other device used in thesauri to identify the semantic relationship between terms (for example, UF), other fields (for example, SN), or records (for example, BT).

relevance

The extent to which information retrieved in a search is judged by the user to meet the criteria of the query.

relevance ranking

Ranking and sorting of query results, typically estimated by an algorithm that calculates the number and weight of occurrences of the search term in the targeted data.

report

An organized set of data presented in a format suitable for viewing or printing, typically produced by a pre-established query that may or may not have variables that are manipulated by the user.

repository

In the context of art and related disciplines, refers to an institution, agency, or individual that has physical or administrative responsibility for an art object, work of architecture, or other thing.

required fields

Fields or data elements that are required to meet a standard or the requirements of a system's operations. *See also core fields*.

Res

As defined in FRSAD, *Res* includes both material or physical things and conceptual objects. Everything considered relevant to the bibliographic universe, the universe of discourse in this case, is included. *Res* is a superclass of all the other entities that are explicitly defined, as well as of any other entities not specifically labeled. *See also nomen*.

results list

The records or other data retrieved in response to a query and presented on line or in a system in an organized display.

reserved characters

Letters, numbers, or symbols that have special uses or meanings in a programming or querying language.

resource discovery

The process of searching for specific information objects on the Web.

retrieval

In the context of this book, the activity of using a search or other method to find records or other data in a database. *See also* **query**.

robot

See **Web crawler**.

romanization

Also called **latinization**. The conversion of a character or word expressed in a non-Roman alphabet or writing system (for example, *Cyrillic* or *Korean*) into the Roman alphabet by means of transcription, transliteration, or a combination of the two methods.

root

Also called **root node** or **top term**. The highest level of the hierarchy, from which all branches descend. In thesauri, the relationship indicator for this type of term is TT, for *top term*.

rotated listing

See **permuted display**.

satellite vocabulary

A thesaurus that is created with the intention of, or is later adapted for, linking to another vocabulary that is larger, broader, or more generic; it may be integrated at many points in the original vocabulary. *See also* **node linking**, **extension vocabulary**, and **microcontrolled vocabulary**.

schema

Also called a **scheme**. In the context of this book, the organization, structure, and rules for a set of data (for example, the set of tables, views, indices, and descriptions for columns in a database, or the organization and description of an XML document).

scope note (SN)

A note explaining coverage, specialized usage, and meaning of terms. In thesauri, the relationship indicator for this note is SN.

screen scraping

A technique in which display data (usually unstructured) is automatically retrieved and extracted, for example, from a Web page.

search engine

A computer program that allows users to search electronic resources. In the context of the World Wide Web, the term usually refers to a program that searches a large index of Web pages generated by an automated Web crawler. *See also* **Web search engine**.

see reference

A type of cross-reference, usually in a printed index, directing the reader from a non-preferred term or subject heading to the preferred term or subject heading for the same concept. The term or subject heading at the *see* reference is a synonym for the preferred term or heading.

see also reference

A type of cross-reference, usually in a printed index, directing the reader to a related term or entry. A *see also* reference differs from a *see* reference in that the *see also* reference is

not made between synonyms, but between terms or headings that are more peripherally related.

semantic linking

A method of linking terms in a vocabulary or larger database according to the meaning of the terms and relationships between terms.

semantic relationship

See **paradigmatic relationship**.

semantic triples

See **triples**.

Semantic Web

As coined by Tim Berners-Lee, the inventor of the World Wide Web and director of the W3C, a web of data that can be processed directly and indirectly by machines. It would be an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.

server

An application that supplies resources or resource manifestations. Often used to refer to a networked computer that acts as a source of data and/or applications used by multiple client computers or devices. See also **client**.

service provider

In OAI nomenclature, an institution or organization that harvests metadata from data providers and uses the aggregated metadata as a basis for building value-added services.

SGML (Standard Generalized Markup Language)

International Standards Organization standard ISO/IEC 8879:1986; a markup language first used by the publishing industry, for defining, specifying, and creating digital documents that can be delivered, displayed, linked, and manipulated in a system-independent manner. XML and HTML are derived from SGML.

sibling

A concept that shares the same immediate broader context (one level higher) as other concepts. Siblings are subordinate to the same broader concept and that are at the same hierarchical level.

single-to-multiple term equivalence

In the context of mapping terms from different vocabularies to each other, the situation that occurs when a term in one vocabulary has no direct match in the second vocabulary, but instead must be mapped to a combination of terms.

SKOS (Simple Knowledge Organization System)

An area of work developing specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading lists and taxonomies within the framework of the Semantic Web.

<https://www.w3.org/2004/02/skos/>

social tagging

Also referred to as *collaborative tagging*, *social classification*, *social indexing*, *mob indexing*, *folk categorization*. The decentralized practice and method by which individuals and groups

create, manage, and share tags (terms, names, etc.) to annotate and categorize digital resources in an online "social" environment. *See also* **crowd sourcing** and **expert crowd sourcing**.

software

The components of a computer system that are not physical, including programs, procedures, algorithms, and documentation pertaining to the operation of a system and the performance of specific tasks, such as word processing, Web browsers, photo editing, and art cataloging or vocabulary editing.

sorting

In the context of this book, the automated process of organizing a results list, data elements in a record, or other data, in a particular sequence based on established criteria or attributes of the data, for example alphabetically, by parent string, or by an associated date. There may be *primary sort criteria* and *secondary sort criteria* (for example, first sort place names in a results list alphabetically, and then — for homographs in the list — sort by parent string).

Soundex

A phonetic algorithm for matching terms and names by sound, as pronounced in English, by translating words into a standard code or representation. It was developed by Robert Russell and Margaret Odell and patented in 1918 and 1922. The National Archives and Records Administration (NARA) maintains the current rule set for the official implementation of Soundex used by the U.S. Government. *See also* **Metaphone**.

source

In the context of building vocabularies, a citable reference to a term in the literature that helps establish its form, spelling, usage, and meaning. *See also* **literary warrant**.

source authority

In the context of this book, a bibliographic authority file used to control the citations providing warrant for terms in a vocabulary or information in a work record.

source language

In the context of translating or mapping one vocabulary to a vocabulary in another language, the language of the original vocabulary. *See also* **target language**.

SPARQL end point

A conformant SPARQL (Semantic Search Query Language) protocol service as defined in the SPROT specification, that enables users (human or other) to query a knowledge base via the SPARQL language. Results are typically returned in one or more machine-processable formats.

specialized vocabulary

See **microcontrolled vocabulary**.

specifications

In the context of designing an information system, the formal, detailed description of user and technical requirements, including specific descriptions of procedures, functions, screens, reports, materials, other features, and hardware. *See also* **user requirements**.

specificity

In the context of indexing, the degree of precision or *granularity* used in assigning terms. Measures of greater specificity include the use of the narrowest applicable indexing term rather than a broader, more generic term. *See also exhaustivity.*

spider

See **Web crawler**.

SQL (Structured Query Language)

A standard command language used with relational databases to perform queries and other tasks.

standard

A vocabulary, set of rules, code of practice, or description of characteristics and parameters, that is documented, established by experts or approved by an authoritative body, and widely recognized or employed as an authoritative exemplar of correctness or best practice; used within a discipline or domain in order to promote interoperability and efficiency.

statistical specificity

See **operational specificity**.

stemming

In the context of mapping terms for search and retrieval, the alteration of a term by automatically truncating or removing common suffixes, word endings, or prefixes in order to find a match, usually applied to sets of related words that are derived from a common root and appear in a variety of grammatical forms (for example, *paint*, *painting*, *painted*).

stop list

In the context of search and retrieval, words in a vocabulary or target data that are ignored in searching or matching because they occur too frequently or are otherwise of little value in retrieval for a given domain. Common stop lists for a text contain articles, conjunctions, and prepositions, although these words are typically not included in a stop list for a vocabulary.

string syntax

Also called **string indexing**. The creation of headings by computer algorithm, characterized by headings that are more consistent than the typically idiosyncratic headings created by hand (for example, the automatic concatenation of a parent string in a heading for a geographic place, *San Gimignano (Siena province, Tuscany, Italy)*).

structure

See **data structure**.

sub-facet

A major conceptual division of a thesaurus that is located near the top of the tree but under a facet. Also called a **hierarchy** in the AAT, although *hierarchy* has a more general meaning as well.

subject

In the context of this book, the focus concept of a vocabulary record (for example, the subject of a ULAN record is a person). Also used to refer to the subject matter (often iconographical content) of what is depicted in or by a work of art or the content of a text.

subject heading list

An alphabetical list of words or phrases used to indicate the content of a text or other thing; characterized by *precoordination* of terminology, meaning that several unique concepts are combined in a string (for example, *Archaeology and art—China—History—20th century*). A type of controlled vocabulary. *See also* **authority heading**.

subject indexing

A term typically used in the context of bibliographic cataloging but also applicable to cataloging art, referring to the application of indexing terms to the content of the document, as contrasted to a description of its physical characteristics.

surface Web

See **visible Web**.

surname

See **last name**.

surrogate

See **digital surrogate** and **visual surrogate**.

switching

In the context of mapping one vocabulary to another, refers to the use of a third vocabulary (a *switching vocabulary*) that itself can link to terms in each of the two original controlled vocabularies, useful when the original two vocabularies do not map well directly to each other. *See also* **direct mapping**.

symmetric relationship

In the context of a thesaurus, a reciprocal relationship that is the same in both directions, e.g., RT/RT. *See also* **asymmetric relationship and reciprocity**.

syndetic structure

Also called **cross-reference links**. In the context of a vocabulary, refers to the linking of equivalent, broader, narrower, and other related terms so that they can be used as cross references to each other and to related headings for the purpose of access.

synonym

A term having a different form, but exactly or very nearly the same meaning as another term. *See also* **true synonymy** and **near-synonymy**. *Compare* **lexical variant**.

synonym ring list

A type of controlled vocabulary containing terms that are considered equivalent for the purposes of retrieval, but not necessarily having true synonymy.

synonymy

A type of semantic relation in which two words or terms have the same or very similar meaning. *See also* **true synonymy and near synonymy**.

syntax

In the context of this book, the structure of elements in a compound term or name (for example, *last name first, comma, first name, middle initial*) or heading; also used to refer to the structure of elements in a search query (for example, rules for the placement of Boolean operators OR, AND, or NOT between terms), analogous to the linguistic structure of elements in a sentence.

synthesis note

A brief preliminary finding, example, or recommendation; this expression was used in the original print publication of the AAT to refer to bottom-of-page notes throughout each sub-facet (or hierarchy) that suggested ways in which descriptors from that sub-facet could be combined in postcoordination with other descriptors (these recommendations are now found in the AAT Editorial Manual).

system

Also called a **computer system**. A number of interrelated hardware and software components that work together to store and convert data into information by using electronic processing. In the context of this book, a system for building and maintaining vocabularies, cataloging art, or performing search and retrieval. *See also* **database**.

systematic display

See **hierarchical display**.

table

See **data table**.

tagging

In the context of the Web, the act of associating terms (called tags) with an information object (e.g., a Web page, an image, a streaming video clip), thus describing the item and enabling keyword-based classification and retrieval. Tags—a form of user-generated metadata—from communities of users can be aggregated and analyzed, providing useful information about the collection of objects with which the tags have been associated. *See also* **social tagging**.

target language

In the context of translating or mapping one vocabulary to a vocabulary in another language, the language into which the original vocabulary is being translated. *See also* **source language**.

taxonomy

A classification organized into a hierarchical structure and applicable for a defined domain. Often used to refer to the classification of living organisms according to physical characteristics, but the term and principles can be applied to classification in any discipline. Unlike thesauri, taxonomies typically do not include synonyms and associative relationships.

term

A word or group of words representing a single concept; a vocabulary *record* comprises terms and other information, including relationships, scope note, sources, etc. Additionally, in the jargon of thesaurus construction, the word *term* is often used as shorthand to actually refer to the concept that is represented by that term (for example, BT and NT actually refer to the relationships between concepts). The distinction between a *term* in the strict sense and *term* meaning a *record* must often be inferred from the context of the discussion.

term frequency (TF)

An automatic ranking method often used in a formula with *inverse document frequency* in information retrieval and text mining to measure how important a term is to a set of data and how useful it will be in retrieval.

term record

In the jargon of thesaurus construction, the collection of information associated with a descriptor, including the history of the term, its relationships to other terms and records, etc. In this book, it is referred to as a **record** (or a *concept record*), in order to distinguish it from the information that is actually associated only with the term table in a relational database model (for example, language of the term, contributor of the term).

text

In the context of this book, data that is not vocabulary controlled and generally unstructured beyond the common structure of standard language expressions of characters, words, sentences, or paragraphs. *See also free-text field.*

thesaurus

A controlled vocabulary arranged in a specific order and characterized by three relationships: equivalence, hierarchical, and associative. Thesauri may be monolingual or multilingual. Their purposes are to promote consistency in the indexing of content and to facilitate searching and browsing.

top term

See root. In thesauri, the relationship indicator for this type of term is TT.

translation

The process of changing a term or text from one language into another by interpreting the meaning of the original (source) term and expressing it as an equivalent in the second (target) term (for example, *copper mines* in English is translated as *mines de cuivre* in French).

transcription

In the context of cataloging art, the process of recording a term or text word for word, letter for letter, accurately copying capitalization, punctuation, spacing, line breaks, illegible passages, and all other possible aspects of the original (for example, to accurately express the nuances of an artist's signature or an ancient architectural inscription). Transcriptions in this context are typically *semi-diplomatic* or *semi-normalized* transcriptions, meaning both substantive and accidental features of the original are retained, but abbreviations are spelled out using brackets or other punctuation to distinguish the original from the editorial content.

transliteration

The process of rendering of the letters or characters of one alphabet or writing system into the corresponding letters or characters of another alphabet or writing system, generally based on phonetic equivalencies. While a common noun will often be translated, a proper name in a non-Roman alphabet is more often transliterated. There are often multiple standards for transliterating from one writing system to another, thus producing multiple variant names.

tree structure

A controlled vocabulary display format in which the complete hierarchy of records is shown or accessible by clicking. The tree structure may be constructed by assigning a tree number or line number to each record, or by another method. *See also systematic display.*

triples

Also called **semantic triples**. A triple is the atomic data entity in the **RDF** data model. Each triple is a set of three entities that codifies a statement about semantic data in the form of

subject–predicate–object expressions (e.g. "Leonardo da Vinci had as patron Francis I, King of France" or "Leonardo da Vinci painted the Mona Lisa"), each of which consists of a subject, a predicate, and an object.

true synonymy

The characteristic of terms or names that have meanings that are identical or as nearly identical as is possible with language. The purpose of enforcing true synonymy in a vocabulary is to increase precision in indexing and retrieval. *See also near synonymy.*

truncation

In searching and matching, the action of cutting off characters in a search term in order to find all terms with a certain common string of characters; it typically involves the user employing a wildcard symbol to search for a string of characters no matter what other characters follow (or sometimes, precede) that string (for example, searching for *arch** will retrieve *arch*, *arches*, *architrave*, *architecture*, *architectural history*, etc.).

trunk name

See focus.

Turtle (Terse RDF Triple Language)

A format for expressing data in the **RDF (Resource Description Framework)** data model with a syntax similar to **SPARQL**. Turtle provides a way to group three URIs to make a triple, and provides ways to abbreviate such information, for example by factoring out common portions of URIs.

typography

The font style and size, arrangement, appearance, and layout of words and texts on a page; in the context of this book, one of the critical elements in designing an end-user display of vocabulary records.

unbound compound term

Multi-word terms in which the words are not necessarily bound together in a given vocabulary, because the terms do not represent a single concept. For example, in the AAT, *Baroque cathedral* is not included because joining a style plus a work type in the thesaurus will result in many thousands of potential terms that will clutter the resource. Instead, *Baroque* plus *cathedral* may be combined on the documentation and retrieval sides. *See also bound term, compound term, postcoordination, and precoordination.*

uncertainty and ambiguity

In writing scope notes, editors should explain any controversies or ambiguous issues. If an issue is in dispute, be careful not to express it as a certain fact.

Unicode

A 16-bit character encoding scheme and standard for representing letters, characters, and diacritical marks in most of the world's modern scripts.

unique identifier

A number or other string that is associated with a record or piece of data, exists only once in a database, and is used to uniquely identify and disambiguate that record or piece of data from all others in the database.

upload

See **load**.

up-posting

Also known as **autoposting**. The automatic generation of search terms or indexing terms by adding broader terms to the specific term requested by a searcher or used by the indexer. See *also* **generic posting**.

URI (Uniform Resource Identifier)

A short string that uniquely identifies a resource such as an HTML document, an image, a downloadable file, or a service. URLs and URNs are types of URIs. Examples of URIs for AAT are the following:

AAT: rhyta

Human-readable full record: <http://vocab.getty.edu/page/aat/300198841>

Human-readable hierarchy view: <http://vocab.getty.edu/hier/aat/300198841>

Semantic RDF concept: <http://vocab.getty.edu/aat/300198841>

URL (Uniform Resource Locator)

A type of URI consisting of an Internet address that tells users how and where to locate a specific file on the World Wide Web. A URL includes not only the name of a file but also the name of the host computer, the directory path to get to that file, and the protocol needed in order to use it (e.g.,

<http://www.getty.edu/research/tools/vocabularies/guidelines/index.html#aat>)

URN (Uniform Resource Name)

A type of URI consisting of a unique, location-independent identifier of a file available on the Internet. The file remains accessible by its URN regardless of changes that might occur in its host and directory path. For example, *urn:issn:0167-6423* is the URN for the journal *Science of Computer Programming*.

used for term

Also called a **UF**. In thesaurus jargon, a term that is not a descriptor and not an alternate descriptor. If the thesaurus is being used as an authority, a *used for* term is not authorized for indexing. *Used for* terms typically comprise spelling or grammatical variants of the descriptor or have true synonymy with the descriptor.

user

See **end user**.

user interface (UI)

The portion of the design and functionality of a cataloging, editorial, search and retrieval, or other system or Web site with which end-users interact, including the arrangement of displays, menus, clickable text or images, pagination, etc. A user interface that is easy for users to use is called *user friendly*.

user requirements

In system design, the initial formal explanation of functionalities, displays, and reports expressed from the point of view of the users' needs and expectations. See *also* **specifications**.

user warrant

Justification for a term in a controlled vocabulary based on the frequency of user queries that employ the term. User warrant may be used for terms intended for retrieval, but is

typically not sufficient warrant for posting a term in a thesaurus used for indexing. *See also literary warrant and organizational warrant.*

variable

In a query, criteria or factors that may be changed to produce different results (for example, as may be expressed in a *where* clause, as the relationship type code in this query: *select distinct subjecta_id from associative_rels where rel_type_code = '2110';*). *See also criteria.*

variant term

In a vocabulary, a term that is not the preferred term but refers to the same concept, including *used for* terms and *alternate descriptors*.

vector-space model

A method of automatic weighting in retrieval where an algebraic model is used for term frequency and distribution, creating representative vectors in multiple dimensional space; when compared to the vectors of an incoming query, the relevance of results may be predicted.

visible Web

The subset of the World Wide Web that is visible to Web browsers and indexable by search engines' Web crawlers. To be accessible to Web crawlers, the pages must be accessible simply by following links (i.e., not generated dynamically in response to user input) and not protected by a password).

visual arts

See art.

visual surrogate

A representation of an object as opposed to the tangible object itself, as when distinguishing between a photograph or digital image, and the art object depicted in the photograph or digital image.

vocabulary control

The process of enforcing the use of certain terminology with the goal of providing consistency and improving retrieval.

vocabulary

See controlled vocabulary.

VRA Core 4.0

An XML schema for describing works of art and architecture and their visual surrogates. <http://www.vraweb.org/projects/vracore4/index.html>

W3C (World Wide Web Consortium)

The main international standards organization for the World Wide Web.

warrant

In the context of vocabularies, sources that provide justification for the spelling and usage of a term to refer to a particular for a concept, including warrant of publications, common usage by experts of a discipline, or other sources.

Web 2.0

The second stage of development of the World Wide Web, characterized especially by the change from static web pages to dynamic or user-generated content and the growth of social media.

Web browser

A software application that enables users to view and interact with information and media files on the Web (for example, Internet Explorer, Mozilla Firefox, and Netscape Navigator).

Web crawler

Also called *robot* or *spider*. A software program that systematically traverses the Web, either for the purpose of generating a searchable index of Web content or to gather statistics.

Web search engine

Also called *Internet search engine*. A software program that collects data taken from the content of files available on the Web and puts them in an index or database that Web users can search in a variety of ways. The search results provide links back to the pages matching the user's search in their original location

Web site

A collection of related electronic pages (*Web pages*), generally formatted in HTML and found at a single address where the server computer is identified by a given hostname.

weighted term

See **best match**.

wiki

A collaborative Web site that contains pages that any authorized user can edit. Wikis typically retain all former versions of each page, allowing the revision history of a page to be tracked and for unwanted revisions to be reversed.

wildcard

Also called a **wildcard character** or **wildcard symbol**. In searching, a character or symbol, such as an asterisk or percent sign, that is used to represent any other character or characters in a Boolean query or other string (for example, the asterisk in *Buonar**).

whole/part relationship

Also called a **partitive relationship**. A hierarchical relationship between a *larger entity* and a part or *component*. In the context of cataloging art, it typically refers to a relationship between two work records or two records in a thesaurus (for example, *Florence* is part of *Tuscany*). See also **genus/species relationship** and **instance relationship**.

word sense disambiguation (WSD)

In automatic search and retrieval, the problem of determining in which sense a homograph is intended in a given data set or text.

work

In the context of this book, a creative product, including architecture, art works such as paintings, drawings, graphic arts, sculpture, decorative arts, photographs that are considered to be art, and other cultural artifacts. A work may be a single item or it may be made up of many physical parts.

World Wide Web

A vast distributed wide-area client-server architecture for retrieving hypermedia documents over the Internet.

XML (eXtensible Markup Language)

A simple, flexible markup language derived from SGML. Originally designed for large-scale electronic publishing, but now playing an increasingly important role in the publication and exchange of a wide variety of data on the Web.

Z39.50

An ISO 23950 and ANSI/NISO Z39.50 standard information retrieval protocol. Z39.50 is a client/server-based protocol for searching and retrieving information from remote databases.

Selected Bibliography for Vocabularies

- Agirre, Eneko and Philip Edmonds, eds. *Word Sense Disambiguation: Algorithms and Applications*. New York: Springer Publishers, 2007.
- Aitchison, Jean, Alan Gilchrist, and David Bawden. *Thesaurus Construction and Use: A Practical Manual*. 4th edition. New York: Fitzroy Dearborn Publishers, 2002.
- Alexiev, Vladimir. *GVP LOD: Ontologies and Semantic Representation*. 2014.
http://www.getty.edu/research/tools/vocabularies/cidoc_alexiev_getty_vocab_lo.pdf
(8 June 2018).
- ANSI/NISO Z39.19-2005 (R2010): Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Bethesda (MD): National Information Standards Organization, 2010.
- ANSI/NISO Z39.50-2003 (S2014), Information Retrieval: Application Service Definition & Protocol Specification*. Bethesda: National Information Standards Organization, 2015.
- ANSI/NISO Z39.85-2001, Dublin Core Metadata Element Set*. Bethesda: National Information Standards Organization, 2001.
- Baca, Murtha. "Controlled Vocabularies for Art, Architecture, and Material Culture." in *Encyclopedia of Library and Information Sciences* [electronic resource]. New York: Taylor & Francis, 2015.
- Baca, Murtha. *Digitization Does Not Equal Access: Challenges in Creating and Disseminating Cultural Information in the Age of the Internet*. 2014.
http://www.getty.edu/research/tools/vocabularies/cidoc_baca_keynote.pdf
(8 June 2018).
- Baca, Murtha, ed. *Introduction to Art Image Access: Issues, Tools, Standards, Strategies*. Los Angeles: Getty Research Institute, 2002. [available online].
http://www.getty.edu/research/publications/electronic_publications/intro_aia/
(7 June 2018).
- Baca, Murtha, ed. *Introduction to Metadata: Pathways to Digital Information*. 3rd ed. Los Angeles: Getty Information Institute, 2016. [2nd ed. Online]
http://www.getty.edu/research/publications/electronic_publications/intrometadata/
(7 June 2018).
- Baca, Murtha. "On Language." *Visual Resources: An International Journal of Documentation* 30, no. 2 (2014): 121-124.
- Baca, Murtha. "Practical Issues in Applying Metadata Schemas and Controlled Vocabularies to Cultural Heritage Information." *Electronic Cataloging: AACR2 and Metadata for Serials and Monographs, Cataloging & Classification Quarterly* 36/3-4 (2003): 47-55.
- Baca, Murtha and Melissa Gill. "Encoding Multilingual Knowledge Systems in the Digital Age: the Getty Vocabularies." in *ISKO C-US: NASKO 2015: Proceedings*.
<http://www.iskocus.org/nasko2015-proceedings.php> (8 June 2018).

- Baca, Murtha, Patricia Harpring, Elisa Lanzi, Linda McRae, and Ann Baird Whiteside. *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images*. Chicago: American Library Association, 2006. [available online]. http://cco.vrafoundation.org/index.php/toolkit/cco_pdf_version/ (7 June 2018).
- Baca, Murtha and Patricia Harpring. "The Getty Vocabularies and Standards: Describing, Cataloging, and Accessing Information about Architecture and Architectural Documents." *COMMA: International Journal on Archives*. 2009.
- Baker, Thomas. "A Grammar of Dublin Core." *DLIB* 6/10 (2000). [online]. www.dlib.org/dlib/october00/baker/10baker.html (1 June 2018).
- Barry, Randall K., ed. *ALA-LC Romanization Tables: Transliteration Schemes for Non-Roman Scripts*. Washington, DC: Library of Congress, 2011-. [online] <https://www.loc.gov/catdir/cpsd/roman.html> (7 June 2018).
- Bates, Marcia J. "Indexing and Access for Digital Libraries and the Internet: Human, Database and Domain Factors." *Journal of the American Society for Information Science* 49, no. 13 (November 1998): 1185-1205.
- Beebe, Caroline, "Image Indexing for Multiple Needs." *Art Documentation* 19/2 (2000): 16–21.
- Beecroft, Antonio. *Describing a Digital World: New Media Descriptors in the AAT*. 2017. http://www.getty.edu/research/tools/vocabularies/beecroft_itwg_2017.pdf (8 June 2018).
- Bell, Lesley Ann. "Gaining Access to Visual Information: Theory, Analysis and Practice of Determining Subjects—A Review of the Literature with Descriptive Abstracts." *Art Documentation* 13/2 (Summer 1994): 89–94.
- Benedetti, J. M. "Words, Words, Words: Folk Art Terminology--Why It (Still) Matters." *Art Documentation* 19/1 (Spring 2000): 14–21.
- Bermes, Emmanuelle. "Convergence and Interoperability: a Linked Data perspective." in *IFLA World Library and Information Congress, 77th Puerto Rico*, 2011.
- Berners-Lee, Tim, James Hendler, and Ora Lassila. "The Semantic Web." *Scientific American* 284, no. 5 (2001): 28-37.
- Besser, Howard. *Introduction to Imaging*. Revised edition, edited by Sally Hubbard with Deborah Lenert. Los Angeles: Getty Research Institute, 2003. [online]. http://www.getty.edu/research/publications/electronic_publications/introimages/ (1 June 2018).
- Borgman, Christine L. *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World*. Cambridge, Massachusetts: MIT Press, 2000.

- Bourcier, Paul, Heather Dunn, and the Nomenclature Task Force, eds. *Nomenclature 4.0 for Museum Cataloging. [Chenhall's Nomenclature]*. Lanham, Maryland: Rowman & Littlefield/AASLH, 2015.
- BS 1749:1985: British Standard Alphabetical Arrangement Filing Order of Numerals and Symbols*. London: BSI Group, 1985.
- BS 8723-1:2005: BS 8723-2:2005, BS 8723-3:2007, BS 8723-4:2007: Structured Vocabularies for Information Retrieval*. London: BSI Group, 2005–2007.
- Canow, Joanne, David Kerr, and Patricia Whittaker. *Faceted Classification, A Group Perspective: History, Current and Future Applications*. Vancouver: University of British Columbia, School of Library, Archival, and Information Studies, 2002.
- CDWA Lite: XML Schema Content for Contributing Records via the OAI Harvesting Protocol*. J. Paul Getty Trust, 2005.
http://www.getty.edu/research/publications/electronic_publications/cdwa/cdwalite.html
 (1 June 2018).
- Charles, Valentine, Nuno Freire, and Antoine Isaac. "Links, languages and semantics: linked data approaches in The European Library and Europeana." in *IFLA World Library and Information Congress, 80th Lyon, 2014*.
- Chen, Shu-Jiun, Marcia Lei Zeng, and Hsueh-Hua Chen. "Alignment of conceptual structures in controlled vocabularies in the domain of Chinese art – A discussion of issues and patterns." *International Journal on Digital Libraries* 17(1) (2016): 23-38.
- Clarke, Stella G. Dextre, and Marcia Lei Zeng. "From ISO 2788 to ISO 25964: the evolution of thesaurus standards towards interoperability and data modeling." *Information Standards Quarterly* 24(1) (2012): 20-26.
- Cobb, Joan. *Getty Vocabularies: Why LOD? Why Now?* 2014.
http://www.getty.edu/research/tools/vocabularies/cidoc_cobb_getty_vocab_lod.pdf
 (8 June 2018).
- Contributors to the Getty Vocabularies.
<http://www.getty.edu/research/tools/vocabularies/contributors.html>
 (8 June 2018).
- Craven, Tim. *Thesaurus Construction: An Introductory Tutorial*. London, Ontario: University of Western Ontario, 2002.
- Dewey, Melvil. *Dewey Decimal Classification and Relative Index*. Joan S. Mitchell, Julianne Beall, Winton E. Matthews, Jr., and Gregory R. New, eds. Albany, New York: Forest Press, 1996.
- Garcia, Gregg. *Getty Vocabularies as LOD: Process Discovery and Evolution*. 2016.
http://www.getty.edu/research/tools/vocabularies/garcia_itwg_tech_update_2014.pdf
 (8 June 2016).
- Garcia, Gregg. *Getty Vocabularies LOD: Querying, Dumping, Re-Using and Serving*. 2017.

http://www.getty.edu/research/tools/vocabularies/garcia_vocab_loq_querying.pdf
(8 June 2018)

Golub, Koraljka, Douglas Tudhope, Marcia Lei Zeng, and Maja Zumer. "Terminology registries for knowledge organization systems." *Journal of the American Society for Information Science and Technology (JASIS&T)* 65(9) (2014): 1901-1916.

Greenberg, Jane, Stuart Sutton, and D. Grant Campbell. "Metadata: A Fundamental Component of the Semantic Web." *Bulletin of the American Society for Information Science and Technology* 29, no. 4 (April/May 2003): 16-18.

Harpring, Patricia. *AAT: Introduction and Overview*. 2016.
http://www.getty.edu/research/tools/vocabularies/aat_in_depth.pdf
(8 June 2018).

Harpring, Patricia. "Cataloging Architectural Drawings: The Architectural Subject Authority of the Foundation for Documents of Architecture." *Visual Resources* 7 (1990).

Harpring, Patricia. "Can Flexibility and Consistency Coexist? Issues in Indexing, Mapping, and Displaying Museum Information." *Spectra* 26/1 (Spring 1999): 33-35.

Harpring, Patricia. *Cataloging Architecture and Related Documentation*.
http://www.getty.edu/research/tools/vocabularies/cco_cdwa_for_architecture.pdf
(8 June 2018).

Harpring, Patricia. *Cataloging for Museum Objects*. 2018.
http://www.getty.edu/research/tools/vocabularies/cco_cdwa_for_museums.pdf
(8 June 2018).

Harpring, Patricia. *Cataloging Manuscripts, Prints, Drawings, and Rare Books*. 2018.
http://www.getty.edu/research/tools/vocabularies/cco_cdwa_for_ms_prints.pdf
(8 June 2018).

Harpring, Patricia. *CDWA and CCO: Introduction and Application*. 2018.
http://www.getty.edu/research/tools/vocabularies/intro_to_cco_cdwa.pdf
(8 June 2018).

Harpring, Patricia. *CONA: Introduction and Overview*. 2018.
http://www.getty.edu/research/tools/vocabularies/cona_intro.pdf
(8 June 2018).

Harpring, Patricia. "Contributing to the Getty Vocabularies." *VRA Bulletin* (2009).

Harpring, Patricia. *Contributing to the Getty Vocabularies*. 2018.
http://www.getty.edu/research/tools/vocabularies/intro_to_contributing.pdf
(8 June 2018).

Harpring, Patricia. "The Getty Cultural Objects Name Authority (CONA)." *Art Documentation* (Spring 2010).

- Harpring, Patricia. *Guidelines for Multilingual Equivalency Work* [online].
http://www.getty.edu/research/tools/vocabularies/guidelines/aat_4_4_appendix_d_contributors.html#4_4_2
 (7 June 2018).
- Harpring, Patricia. "How forcible are right words!: Overview of Applications and Interfaces Incorporating the Getty Vocabularies." *Museums and the Web 1999: Selected Papers* (1999).
- Harpring, Patricia. *IA: Iconography Authority: Introduction and Overview*. 2016.
http://www.getty.edu/research/tools/vocabularies/cona_ia_in_depth.pdf
 (8 June 2018).
- Harpring, Patricia. *Indexing Using the Getty Vocabularies*. 2016.
http://www.getty.edu/research/tools/vocabularies/vocab_brief_intro_and_indexing.pdf
 (8 June 2018).
- Harpring, Patricia. *Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works*. Revised version. Edited by Murtha Baca. Los Angeles: Getty Publications, 2013. [also online]
<http://www.getty.edu/publications/virtuallibrary/160606018X.html>
 (7 June 2018).
- Harpring, Patricia. "The Language of Images: Accessing Images by Applying the Categories for the Description of Works of Art and Structured Vocabularies," in *Introduction to Art Image Access*. edited by Murtha Baca. Los Angeles: Getty Publications, 2002.
- Harpring, Patricia. "The Limits of the World: Theoretical and Practical Issues in the Construction of the Getty Thesaurus of Geographic Names." *Proceedings: ICHIM 97* (1997): 237–251.
- Harpring, Patricia. *Managing an Art Documentation or Vocabulary Project*. 2017.
http://www.getty.edu/research/tools/vocabularies/managing_project.pdf
 (8 June 2018).
- Harpring, Patricia. *Overview of the Getty Vocabularies*. 2016.
http://www.getty.edu/research/tools/vocabularies/overview_getty_vocababs.pdf
 (8 June 2018).
- Harpring, Patricia. "Proper Words in Proper Places: The Thesaurus of Geographic Names." *MDA Information* 2/3 (March 1997): 5–12.
- Harpring, Patricia. "Resistance is Futile: Inaccessible Networked Information Made Accessible Using the Getty Vocabularies." *ASIS (American Society for Information Science) Annual Conference Proceedings*, 1999.
- Harpring, Patricia. "The Role of Metadata Standards in Mapping Art Information: The Visual Resources Perspective." *VRA Bulletin* 27/4 (Winter 2000): 71–76.
- Harpring, Patricia. *Subject Access to Art Works*. 2017.
http://www.getty.edu/research/tools/vocabularies/subject_access_for_art.pdf
 (8 June 2018).

- Harpring, Patricia. *TGN: Introduction and Overview*. 2016.
http://www.getty.edu/research/tools/vocabularies/tgn_in_depth.pdf
(8 June 2018).
- Harpring, Patricia. *ULAN: Introduction and Overview*. 2016.
http://www.getty.edu/research/tools/vocabularies/ulan_in_depth.pdf
(8 June 2018).
- Harpring, Patricia. "When We Mean to Build: Using Standards and Vocabularies to Lay the Foundations for Accessing Information on Art and Architecture." *Society of Architectural Historians (SAH) Annual Conference*, 2001.
- Harpring, Patricia and Murtha Baca, ed. *Categories for the Description of Works of Art*. Revised 2016. Los Angeles: J. Paul Getty Trust and the College Art Association, 2016.
http://www.getty.edu/research/publications/electronic_publications/cdwa/
(7 June 2018).
- Harpring, Patricia and Getty Vocabulary Program. *Editorial Guidelines*. Los Angeles: J. Paul Getty Trust, Vocabulary Program, 2003–.
<http://www.getty.edu/research/tools/vocabularies/guidelines/index.html>
(1 June 2018).
- Haynes, David. *Metadata for Information Management and Retrieval*. 2nd ed. London: Facet Publishing, 2018.
- Hooland, Seth van, and Ruben Verborgh. *Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish your Metadata*. Chicago: Neal-Shuman, 2014.
- Hourihane, Colum. "Subject Classification for Visual Collections: An Inventory of Some of the Principal Systems Applied to Content Description in Images." *VRA Special Bulletin* 12 (1999).
- Hyvönen, Eero. 2010. *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*. San Rafael, California: Morgan & Claypool Publishers, 2010.
- International Terminology Working Group. Presentations given at ITWG Meetings.
<http://www.getty.edu/research/tools/vocabularies/training.html#itwg>
(8 June 2018).
- ISO 2788:1986: Guidelines for the Establishment and Development of Monolingual Thesauri*. Geneva: International Organization for Standardization, 1986.
- ISO 5964:1985: Guidelines for the Establishment and Development of Multilingual Thesauri*. Geneva: International Organization for Standardization, 1985.
- ISO/CD 25964-1: 2011: Information and documentation -- Thesauri and interoperability with other vocabularies*. Geneva: International Organization for Standardization, 2011.

- ISO/CD 25964-2: 2013: *Information and documentation -- Thesauri and interoperability with other vocabularies*. Geneva: International Organization for Standardization, 2013.
- Johnson, Robin. *Building Connections: Some Observations*. 2017. http://www.getty.edu/research/tools/vocabularies/johnson_itwg_2017.pdf (8 June 2018).
- Lanzi, Elisa. "The Linguistic Challenge of a Multilingual AAT." *Art Documentation* 14/2 (Summer 1995): 19.
- Lee-Smeltzer, Kuang-Hwei. "Finding the Needle: Controlled Vocabularies, Resource Discovery, and Dublin Core." *Library Collections, Acquisitions & Technical Services* 24 (2000): 205–215.
- Library of Congress. *ALA Filing Rules*. Chicago: American Library Association, 1980.
- Library of Congress. *Library of Congress Subject Headings: Principles of Structure and Policies for Application*. Washington, DC: US Library of Congress, 1990–2001.
- Long, Matthew, and Roger C. Schonfeld. *Supporting the Changing Research Practices of Art Historians*. New York: Ithaca S+R, 2014.
- McCulloch, Emma and George Macgregor. "Analysis of Equivalence Mapping for Terminology Services." *Journal of Information Science* 34/1 (2008): 70–92.
- Miller, Uri. "Thesaurus Construction: Problems and Their Roots." *Information Processing & Management* 33/4 (July 1997): 481–493.
- Mitchell, Joan S., Marcia Lei Zeng, and Maja Zumer. "Modeling classification systems in multicultural and multilingual contexts." *Cataloging & Classification Quarterly*. 52(1) (2014): 90-101.
- Molholt, Patricia and Toni Petersen. "The Role of the *Art & Architecture Thesaurus* in Communicating about Visual Art." *Knowledge Organization* 20/1 (1993): 30–34.
- Nagel, Lina. "The Spanish-Language Version of the Art & Architecture Thesaurus: History, Current Status, Implementation, and Dissemination." *Visual Resources Association Bulletin* 32/2 (Fall 2005): 30–34.
- National Information Standards Organization. *ANSI/NISO Z39.19-2005: Guidelines for the Construction, Format, and Management of Multilingual Controlled Vocabularies*. Baltimore: National Information Standards Organization, 2005.
- Olson, Tony and Gary Strawn. "Mapping the MeSH and LCSH Systems." *Information Technology and Libraries* 16/1 (March 1997): 5–19.
- Panofsky, Erwin. *Meaning in the Visual Arts: Papers In and On Art History*. Garden City, New York: Doubleday Books, 1955.
- Petersen, Toni. "Developing a New Thesaurus for Art and Architecture." *Library Trends* 38/4 (1990): 644–658.

- Porter, Vicki and Robin Thornes. *Guide to the Description of Architectural Drawings*. Los Angeles: J. Paul Getty Trust, updated 2005 by Patricia Harpring. http://www.getty.edu/research/publications/electronic_publications/fda/ (1 June 2018).
- RBMS Bibliographic Standards Committee. *Thesaurus Construction and Maintenance Guidelines*. Chicago: Association of College and Research Libraries, Rare Books and Manuscripts Section, 1998-. https://rbms.info/rbms_manual/thesaurus_construction/ (7 June 2018).
- RDA Joint Steering Committee for Development of RDA. *RDA: Resource Description and Access: Information and Resources*, 2009-. <https://www.loc.gov/aba/rda/> (7 June 2018).
- Roberts, Helene E. "A Picture is Worth a Thousand Words: Art Indexing in Electronic Databases." *Journal of the American Society for Information Science and Technology* 52/11 (September 2001): 911–916.
- Roberts, Helene E. "Do You Have Any Pictures of--? : Subject Access to Works of Art in Visual Collections and Book Reproductions." *Art Documentation* 7/3 (Fall 1988): 87–90.
- Rowley, Jennifer and John Farrow. *Organizing Knowledge: An Introduction to Managing Access to Information*. 3rd ed. Burlington, Vermont: Gower, 2000.
- Sanderson, Rob. *IIIF, Linked Data, and Vocabularies*. 2016. http://www.getty.edu/research/tools/vocabularies/sanderson_iif_intro_itwg_pnc_2016.pdf (8 June 2018).
- Sanderson, Rob and David Newbury. *Linked.Art & Vocabularies: Linked Open Usable Data*. http://www.getty.edu/research/tools/vocabularies/newbury_sanderson_itwg_2017.pdf (8 June 2018).
- Soergel, Dagobert. *Functions of a Thesaurus/Classification/Ontological Knowledge Base*. College Park, Maryland: College of Library and Information Services, University of Maryland, 1997.
- Stanley, Janet L. "African Art and the Art & Architecture Thesaurus." *Museum Studies Journal* 2 (1986): 42–52.
- Straten, Roelof van. *Introduction to Iconography*. Revised English ed. Translated from the German by Patricia de Man. Yverdon, Switzerland: Gordon and Breach, 1994.
- Svenonius, Elaine. *The Intellectual Foundation of Information Organization*. Cambridge, Massachusetts: MIT Press, 2000.
- Taylor, Arlene G. and Daniel N. Joudrey. *The Organization of Information*. 3rd ed. Westport, Connecticut: Libraries Unlimited, 2009.

- Taylor, B. L. "Chenhall's Nomenclature, the Art & Architecture Thesaurus, and Issues of Access in America's Artifact Collections." *Art Documentation* 15/2 (1996): 17–23.
- Thornes, Robin, with Peter Dorrell and Henry Lie. *Introduction to Object ID: Guidelines for Making Records that Describe Art, Antiques, and Antiquities*. Los Angeles: J. Paul Getty Trust, 1999.
- Tillett, Barbara. *FRBR (Functional Requirements for Bibliographic Records): A Conceptual Model for the Bibliographic Universe*. Washington, DC: Library of Congress Cataloging Distribution Service, 2004.
- Tudhope, Douglas, et al. "Query Expansion Via Conceptual Distance in Thesaurus Indexed Collections." *Journal of Documentation* 62/4 (2006): 509–533.
- Turner, James M. "Subject Access to Pictures: Considerations in the Surrogation and Indexing of Visual Documents for Storage and Retrieval." *Visual Resources* 9/3 (1993): 241–271.
- U.S. National Center for Biotechnology Information. *NCBI Taxonomy*. 2000-. <https://www.ncbi.nlm.nih.gov/taxonomy> (1 June 2018).
- Vellucci, Sherry L. "Metadata and Authority Control." *Library Resources & Technical Services* 44/1 (January 2000): 33–43.
- Vizine-Goetz, Diane, Carol Hickey, Andrew Houghton, and Roger Thomson. "Vocabulary Mapping for Terminology Services." *Journal of Digital Information* 4/4 (March 2004) no. 272.
- Ward, Jon. *Historical Journeys with the Getty Thesaurus of Geographic Names (TGN)*. 2017. http://www.getty.edu/research/tools/vocabularies/ward_historical_journeys.pdf (8 June 2018).
- Ward, Jon. *Kingdoms, Empires, and Sites: Historic Concepts in TGN*. 2017. http://www.getty.edu/research/tools/vocabularies/ward_itwg_2017.pdf (8 June 2018).
- Warner, Amy J. *A Taxonomy Primer*. Ann Arbor: Lexonomy, 2002.
- Warren, Susanne. "Workshop: Using the AAT Art & Architecture Thesaurus: Practical Applications." *Art Documentation* 11/2 (Summer 1992): 63.
- Whitehead, Cathleen. "Faceted Classification in the Art & Architecture Thesaurus." *Art Documentation* 8/4 (Winter 1989): 175–177.
- Wielinga, B. J., A. T. Schreiber, et al. "From Thesaurus to Ontology." *International Conference on Knowledge Capture Archive; Proceedings*. New York: ACM Press, 2001, 194–201.
- Zeng, Marcia. *Create Microthesauri and Other Datasets from the Getty LOD Vocabularies*. 2017.

http://www.getty.edu/research/tools/vocabularies/zeng_microthesauri_getty_lod.pdf
(8 June 2018).

Zeng, Marcia Lei. *The Getty Vocabularies and the Significance of Five-Star LOD Datasets*. 2016.
http://www.getty.edu/research/tools/vocabularies/zeng_vocab_and_five_star_lod.pdf
(8 June 2018).

Zeng, Marcia Lei and Lois Mai Chan. "Trends and Issues in Establishing Interoperability among Knowledge Organization Systems." *Journal of the American Society for Information Science and Technology* 55/5 (2004): 377–395.

Zeng, Marcia Lei, Karen Gracy, and Maja Zumer. "Using a semantic analysis tool to generate subject access points: A study using Panofsky's theory and two research samples." *Knowledge Organization* 41(6) (2014): 440-451.

Zeng, Marcia Lei and Tao Hu. *Extending Exhibitions to Historical Journeys through Data [in the Semantic Web]*. 2017.
http://www.getty.edu/research/tools/vocabularies/zeng_silk_road_tgn.pdf
(8 June 2018).

Zeng, Marcia Lei and Philipp Mayr. "Knowledge Organization Systems (KOS) in the Semantic Web: A Multi-Dimensional Review." *International Journal on Digital Libraries* 17(1) (2016): 23-38.

Zeng, Marcia Lei and Jian Qin. *Metadata, Second Edition*. London: Facet Publishing, 2016.

Zhang, Jin and Marcia Lei Zeng. "A new similarity measure for subject hierarchical structures." *Journal of Documentation*. 70(3) (2014): 364 – 391.