

Glossary

A

- AAT (Art and architecture thesaurus)** Reference work for standard terms used in art and architecture.
- Administrative metadata** Provide information about dates of creation, rights, file format, and other information for management of resources.
- AI (Artificial intelligence)** Automated systems that appear to perform some activities that mimic human capacities.
- Analog** Information with continuous values, such as a conventional clock with hands and a dial, that are not divided into discrete units.
- Analytics** Techniques for processing data, generally using statistical methods.
- API (Applied Programming Interface)** Interface designed to support data export from a project, repository, or its metadata.
- Artificial neural networks (ANN)** Automated systems that imitate the learning capacities of neurons through reinforcement and training.
- AR (Augmented reality)** Result of computer-generated images or text overlaid onto an actual scene, object, or situation.
- Association table** Table in a database used to cross-reference information in one-to-one or many-to-many relationships.

B

- Base-map** Map used in geospatial information systems as the foundation for data display provides coordinates and uses a single projection.
- Between-ness (in networks)** Measure of centrality based on shortest point between two nodes or vertices.
- BIBFRAME (Bibliographical framework)** Standard used to make metadata machine-readable.
- Big data** Very large data sets that pose complex problems for analysis and processing.
- Born digital** Materials that have no instantiation except in digital form (such as iPhone photographs or text messages or web-based artifacts).
- Boundary conditions** The constraints that limit or define the domain or range of a problem or topic.
- Browser** An application that supports the use of files online, it retrieves web pages from a server and displays them.

C

- Cascading style sheets (CSS)** Style sheets that take effect based on <tags> that separate presentation format from intellectual content.
- CCO (Cataloguing Cultural Objects)** A standard for describing cultural objects that makes metadata shareable across institutions like museums, collections, and archives.
- Centrality (in networks)** A metric used to assess significance by the number of links to any node within a network.
- CERIF (Common European Research Information Format)** Standard developed in Europe to make research data interoperable across projects and repositories.
- ChatGPT** Launched by the company OpenAI in 2022, it is a generative AI chatbot that produces responses to prompts and queries that resemble those created by humans.
- Choropleth** Maps in which entire areas are given a color value to show information aggregated by region (rather than population).
- CLIP training** Web-based platform for employee training in the managed service provider sector.
- Closeness (in networks)** A metric assessing the proximity of one node to another based on the shortest distance between nodes.
- Collections management system** Used by institutions to manage the materials and artifacts in their collections.

- Command line interface** Text-based method of issuing computer program commands.
- Content management system** A set of pre-packaged tools designed for the use of digital assets.
- Content model** A conceptual scheme to formalize the description of intellectual (or artistic) content to turn it into structured data.
- Content types** The range of different kinds of web-presentation formats such as blogs, e-books, infographics, and other modes of information sharing.
- Context windows** Data or information stored in an AI system to produce more accurate and “natural” seeming conversational text.
- Continuous data** Data that has an unbroken gradient of values, such as changes in temperature or tonal values in sound or color.
- Controlled vocabulary** A set of terms defined for a project, field, or discipline to standardize metadata creation and facilitate search and retrieval.
- Conversational analytics** System in which queries and exchanges are used to generate search results as well as for training through reinforcement.
- Coordinates** Points that can be located in a defined metric space, such as a grid or ruler.
- Crosswalk** Software used to take data from one file format or structure to another.
- CSV** Comma-separated values, a common mode of exporting data for re-use.
- Cultural analytics** A method developed for the analysis of images, sound, and web-based information at a very large scale.

D

- DALL-E** An artificial intelligence system that generates images from text prompts.
- Dashboard** Interface in which the parts of an application or project are displayed to support user experience.
- Data mining** Techniques for extracting meaningful patterns of information from data in any format.
- Data parallelism** Processing of data on multiple processors or computing environments to optimize efficiency.
- Deepfake** AI-manipulated video, audio, or photo that uses someone’s or something’s likeness to make it appear as if something happened that did not.
- Descriptive metadata** Metadata that describes the intellectual content or other features of the information in a file.
- Dependent variables** Values in an equation that change in connection with each other such that altering one will alter the other. (If I change the independent variable of the time of day for an event it will probably affect how many people show up, which is a dependent variable).
- Digital asset** Any information stored in a digital file or format, generally refers to materials in a collection or repository that are considered valuable.
- Digital asset management system (DMS)** Application or software designed to manage substantial quantities of assets.
- Discrete value** Any value that is defined in discrete units, such as integers, rather than in continuous values.
- Distant reading** Technique developed for doing large-scale analysis of texts and summarizing them through machine generated statistical methods.
- Distant viewing** Technique developed for doing large-scale analysis of images and summarizing them through machine generated statistical methods.
- Directed graphs** Graphs or diagrams constructed with vectors that connect the nodes in a specific direction.
- Dynamic content** Display of information generated on the fly from data content.

E

- EAD (Encoded Archival Description)** XML standard developed specifically for archival description.
- Edge (in networks)** The line that connects nodes and indicates a relationship exists between them.

F

- FAIR** “Findable, Accessible, Interoperable, and Repurposable” Definition of a set of guidelines meant to ensure that research data can be successfully shared across projects.
- Fair use** Set of legal standards that define the ways intellectual property can be used without permission in specific cases.
- Feature set** Elements in the design of a digital project that describe its functionality and format features.
- Fine-tuning** In AI, the process of refining a pre-trained model to improve performance.
- Foreign key** Data attribute used to create a connection between two tables in which the other contains a primary key.
- Form (data form)** A graphic format designed to make it easier to enter data into a table through use of a form with defined fields (e.g. name, address, and zip code).
- Freeware** Software that can be used without payment.

G

- GANS (Generative adversarial networks)** A machine learning framework that pits one system against another in a contest to achieve an optimized outcome.
- Geocoding** Standard for assigning geographical coordinates to information so that it can be put onto a map.
- Geographic Information Systems** Applications and platforms designed to store, retrieve, analyze and present data about geographic features and support mapping.
- Georectification** The process of cleaning or correcting data to conform to geographical standards and coordinate systems for the purpose of mapping.
- Georeferencing** The process of associating data with a geographical coordinate system so that it can be visualized in a spatial format.
- Getty vocabularies** A thesaurus of standards for describing artworks, architecture, and cultural objects developed at the Getty Research Institute to provide information on production techniques, historical styles, and other features.
- GIS** Acronym for Geographic Information Systems, platforms and applications for storing, processing, and representing geographical data.
- GOF AI** Acronym for Good-Old-Fashioned-Artificial-Intelligence which was based on logical rules and symbolic protocols that supported well-defined tasks and other operations, but were not capable of learning.
- GraphML** Graph Markup Language is an XML format for storing network data to generate a wide range of graph structures.
- Graphic variables** The elements of visualizations that can be used to designate specific values through the systematic use of color, shape, pattern, tonal value, position, orientation, and size.
- GUI (Graphical user interface)** The visual display in which applications, devices, and browsers support access and use of content and features.

H

- Hallucinations (in AI)** Inaccurate information generated by AI systems.
- HCI (Human-computer interaction)** Term used to describe the range of ways in which users can interact with digital platforms through touch, keyboard, voice, motion, or other means.
- Href (Hypertext reference)** An HTML attribute that contains information about the destination of a hyperlink.
- HTML (HyperText Markup Language)** A standard markup used to define the content and structure of content so it can be displayed in a web browser.
- HTTP (HyperText Transfer Protocol)** The set of rules that supports the transfer of information on the World Wide Web, it is the basis of all networked data communication.

I

- Information architecture** The design of the structure of information in digital environments and computational platforms includes decisions about data formats and workflows.
- Independent variables** Factors that can be changed to assess their effect in an experiment, equation, or data protocol (e.g. if you change the number of chickens (independent variable) will it change the total of eggs produced [dependent variable]).
- Information visualization** Graphic representation of data that makes quantitative information into a legible format as a chart, graphic, map, or other display.
- Intellectual property** Any expression of human intelligence or imagination that is protected by standards of ownership includes texts, music, patents, and artworks.
- Interface** The space that supports interaction between computational devices and users.
- Interoperable** Refers to making it possible to share data, formats, or software across platforms, projects, and applications.
- IP address (Internet Protocol address)** A unique label that identifies any and every device that is connected to the internet, it is essential for making the device detectable for routing information.

J

- JavaScript** A programming language used to create dynamic and interactive content online.
- JSON (JavaScript Object Notation)** A standard for representing structured data in a text-based format so it can be transmitted between servers and browsers.

K

- Keyboarded** Data that is created using a keyboard, can include text, numerical information, and special characters but usually refers to print or manuscript materials that have been turned into a text file.
- KML (Keyhole Markup Language)** A standard XML format for expressing geospatial information to be displayed on two-dimensional maps or three-dimensional globes.

L

LAION family of data sets Referred to as a “dataset of datasets,” it is a collection managed by a not-for-profit organization to support research in machine learning.

LAMP (Linux, Apache, MySQL, and PHP) A software stack containing an operating system (Linux), web server (Apache), database management system (MySQL), and programming language (PHP) to host materials for web presentation.

LLM (Large Language Models) A machine learning model that performs natural language processing using huge collections of text.

Legacy data Data that is inherited by a project, collection, or repository that was created in an earlier iteration and may or may not conform to updated standards.

Legends On a map, the set of symbols used to define features including natural and human-made elements (e.g. roads, rivers, schools, churches, mountains).

Licensing agreements Terms governing the re-use of intellectual property, including software code.

LIDAR (Light/laser detection and ranging) A method of creating a scan of a surface or area by timing the rate at which reflected light returns to the signal source.

Linked data Data that is connected to other datasets to enhance search and retrieval and increase the usefulness of information, it requires standardization across datasets.

Lossless Data compression format that preserves original information in a way that allows it to be reconstructed perfectly depends on redundancy in the data.

Lossy A data compression format that only allows information to be reconstructed in an approximation of the original data.

LVM (Large vision model) AI systems that are designed to analyze large amounts of visual information without making use of language.

M

Macro-analysis Analysis of phenomena at a large scale to perceive broad patterns and trends.

MAMP (Mac, Apache, MySQL, and PHP format) Web application stack of software used to develop, store, and transmit online content, consists of an operating system (Mac), web server (Apache), data management system (MySQL), and programming language (PHP).

MARC (Machine-Readable Cataloging) Set of standards for making library catalogue records (describing books, recordings, manuscripts, and other library collections) machine-readable.

Manual modeling When a three-dimensional model is built by drawing features directly rather than by using automated techniques.

Markdown A simple markup language designed to make it easy to format text for browser display.

Mark-up languages Text-encoding systems used to specify the structure and format of documents can be limited to design features or developed to include content analysis and interpretation of intellectual content.

Mediation The process of exchanging or transmitting information from one platform, system, or project to another.

Metadata Information used to describe data, generally following a standard format that is discipline specific.

Metadata schemes Standards developed for specialized description of collections, data, artifacts, archives, or other records.

ML (Machine learning) Artificial intelligence systems that can learn to generalize from data and advance their analytic capacities through a combination of statistical and reinforcement processes.

Microanalysis Analysis of phenomena at a small, detailed, scale to perceive underlying causes of events or specific entities.

Multi-modal analysis The description and analysis of data that includes multiple forms of information, including text, image, gesture, time-based recording, and sound.

N

Nested hierarchies Any data structure in which one level is nested within or below the next, usually from general to increasingly specific classification. (Contrast to a flat data structure where all elements are at one level.)

Networks An environment where one or more computers are connected for the purpose of sharing information.

NN (Neural networks) Artificial intelligence systems built on a model that simulates neurons to “learn” through reinforcement and signal exchange to perform increasingly complex tasks.

NLI (Natural Language Interface) A computer environment where the user communicates to the system using text commands to create, analyze, retrieve, and use data.

NLP (Natural Language Processing) The use of computers to understand and produce natural language in a manner that simulates that of human beings.

Node (in networks) Any vortex/point that is a defined entity within a network, connected by edges/lines to other nodes.

Non-representational geography Term coined by geographer Nigel Thrift to emphasize the study of processes and spatial experience rather than representing thinking about space in conventional forms such as maps.

O

Object-oriented database A format in which data and functions are stored in the same database so that many operations can be realized more efficiently than when they are stored separately.

Ontology Scheme of knowledge organization that references a field or discipline and attempts to represent its entities through description, naming, and definition of basic categories.

OAI (Open Archives Initiative) An informal organization that designed standards meant to support interoperability standards for archives, particularly focused on sharing metadata in catalogue records.

OpenRefine An open-source application used in cleaning data or transforming it from one format to another.

Open source Code that is made available for use and modification by others.

P

Paradata Information to describe the processes by which data were produced.

Parallel processing Capacity of a computer system to run processing routines simultaneously on multiple tasks to achieve optimal performance.

Parameterization The creation of defined categories to transform unstructured data into structured data by establishing workflow by fixing the parameters on which values can be defined (e.g. capturing words as data by using spaces between or capturing sentences by identifying full stops/periods).

Parse The ability of a program to extract specific information from raw data through various methods of identification and selection.

Performative metadata Instructions attached to a file that contain instructions on how it should be used, made active, or displayed, often by communicating with a browser, program, or application.

Pick list In an interface menu, a list of terms to assist users in search and retrieval by supplying vocabulary relevant to a specific domain, project, or field.

Pipeline parallelism A data workflow in which the processing takes place in a series of operations which may be carried out simultaneously. Used in neural networks to describe its multiple layers and the connections among them across distributed systems and operations.

Pivot table The manipulation of a database to transform single or multiple rows into columns in order to extract information effectively from the data for the purpose of analysis.

Pixel-based images Images whose values are stored in individual units of display, like a mosaic or tapestry. These images do not store visual information about shapes or vectors.

Plagiarism The act of using someone else's intellectual property without attribution which can be subject to legal action.

Plug-ins Additions to a program or application that add functionality.

Predictive modeling Ability of an automated system to predict what comes next based on statistical analysis.

Primary key A unique identifier for records in a table that ensures data integrity by avoiding confusion or ambiguity.

Procedural modeling Models created using text-based instructions or specifications.

Processing Any activity of a program or algorithm that carries out instructions to create, analyze, manipulate, retrieve, or transform digital information from one state to another.

Projections In cartography these are the conventions used to generate a representation of spatial features on a map or globe.

Prompt engineering (or prompt craft) Skill in designing queries for artificial intelligence systems that produce desired results.

Prototyping Initial creation of designs for a project to test features of interface and functionality, usually done iteratively to fold the results back into the design process.

Python A common programming language, considered a general-purpose language, used to develop applications from simple calculation to data analysis and automation.

Q

Qualitative data Information in a non-numerical form, such as natural language, sound, and other phenomena, frequently identified with the arts and humanities.

Quantitative data Information that is structured numerically, such as time stamps, financial data, population statistics, and temperature, tends to be found in natural and social sciences.

Query A structured inquiry designed to extract specific results from the analysis of a database or digital files.

R

Raster Pixel maps with individual values for each point.

Reality-based models Models created using data captured from actual scenes or objects using photogrammetry or other techniques.

Records Any written, inscribed, recorded, or otherwise made and instantiated information in permanent or semi-permanent form.

RDF (Resource Description Framework) Standards designed for the exchange of graph data that contains triples in the form of subject/predicate/object (Jane/loves/kitten).

RDBMS Relational Database Management Systems are software designed to manage the functions of relational databases.

Relational database A database in which information is broken into tables that must be connected to each other by specified relationships (rather than a flat database of a single table of rows and columns).

Remediation The act of transforming information or phenomena from one format or material instantiation to another, such as the work of scanning an image into a digital file.

Report An output generated from a query or process performed on a database.

S

Shape files Carry geospatial vector data that can be scaled up or down without losing its form.

Simulation An image or model that attempts to pass for what it represents.

SGML (Standard Generalized Markup Language) A standard for defining markup languages, it contains the rules for proper syntax and is the foundation for HTML, XML, KML, and other markup for specific domains.

SORA dynamics Used in generating and producing video footage, often in combination with drones, through a process of noise elimination using artificial intelligence, resulting in high-fidelity simulations.

SPARQL A query language for working with data stored in RDF (Resource Description Framework) formats.

Spreadsheet A flat database organized in rows and columns that can be used to hold information for further analysis.

SQL Structured query language (often referred to as Sequel) used to extract information from a database in response to specific filters or procedures.

Stable diffusion A technique used in generative artificial intelligence as part of neural information processing that generates images through improved sampling and training.

Static content Information displayed in a manner that is the same for every viewer (rather than generated on the fly).

Statistical processing Techniques for consolidating or analyzing data to discern patterns and trends using statistical methods such as median, mean, averages, predictive, and descriptive analyses.

Structured data Information that has unambiguous form at the level at which it can be processed, such as ASCII (keyboarded) files or numerical and statistical data.

Style sheets A set of encoded instructions for telling a browser how to format a file.

Sustainability The ability of a system, file, process, or activity to be maintained over time in a manner that is cost-effective in terms of ecological, human, and financial resources.

Syntax A set of rules governing the way a computer program's statements can be formulated.

T

Table A structured form for data consisting of rows and columns with defined values.

Tags Structured elements between angle brackets used to define parts of XML, HTML, SGML, and other markup languages in order to define the content, form, and other elements of the file.

Task parallelism In a computer environment or system, the ability to perform part of tasks or multiple tasks at the same time.

Taxonomy A classification system of names and terms that organizes a domain or field.

Text analysis Any process using computational methods to extract information and analyze patterns in linguistic material (e.g. word frequency, themes, topics, and sentence length). Often used in stylistic analysis and/or authorship attribution.

Text markup Method of inserting tags into a text to define or identify elements of form and/or content (e.g. <speaker>, <heroine>, <nature>, <excitement>, <title>, or <author>).

Tokenization Process by which information or phenomena can be broken into discrete units that lend themselves to automated processing.

Topic modeling An approach to text analysis that extracts thematic information from an individual or collection of text files at a speed and scale impossible for human readers.

TUI (Tactile User Interface) Any feature of interface that makes use of touch, such as the swiping, hitting, and scrolling actions on a screen.

U

- UDCS (Universal Decimal Classification System)** A standard for organizing bibliographical and library materials by subject area, author, and other information so that they can be ordered and retrieved.
- URL (Uniform Resource Location)** A unique web address that consists of information about the location of files on the internet and consists of access protocol (http), domain name, and path to the file.
- Unique ID** A string of numbers and/or characters that identifies a digital resource (file, folder, or entry in a database) so that it cannot be confused with any other resource in that system.
- Unstructured data** Information that has no predefined format and is not organized in discrete units, which makes it difficult to process (such as color along a tonal gradient or sound values in a continuum).
- UX (User experience)** The full set of activities, behaviors, and features essential for supporting interaction between a user and a device, platform, or project.
- UI (User interface)** The interaction of a person with the functions and features of a digital platform or environment.

V

- Vector-based images** Images created using point-to-point lines to describe shapes that can be scaled up or down without losing their form.
- Vision Language Models (VLMs)** Combine NLP (natural language processing) with LVMs (large vision models) to create textual descriptions for images, answer questions about visual data or create visual data based on text inputs.

W

- WAMP** Windows, Apache, MySQL, and PHP format for web presentation in Windows platforms (see also LAMP and MAMP).
- W3C (World Wide Web Consortium)** The consortium of organizations that oversee the standards for all activity and protocols on the Web.
- WCAG (Web Content Accessibility Guidelines)** Recommendations developed by the W3C to make web content accessible to persons with various disabilities.
- Well-formedness** The characteristic of a file that has been created in conformance with the standards of the program or platform in which it is made so that it will be able to be processed effectively.

X

- XML (eXtensible Markup Language)** A standard markup language that creates structured data that is both human and machine readable and designed to facilitate data exchange, processing, and display.
- XR (Extended reality)** Umbrella term used to describe projects using three-dimensional and/or immersive techniques to create projects in augmented, mixed, and virtual reality that may mirror reality or create realistic but imagined environments.
- XSLT (eXtensible Stylesheet Language Transformations)** A language used to transform XML files into other formats, such as HTML for web display.