

**Working Glossary for NIH Workshop:
"Informatics for Data and Resource Discovery in Addiction Research"**

Gwen A. Frishkoff, Ph.D.

Attribute: A column (e.g., within a relational database)

BFO: Acronym for "Basic Formal Ontology"; An upper ontology that is widely used in bio- and neuro-informatics

Bottom-up: A process that derives information or knowledge from data

BrainMap: An online database of published functional neuroimaging (fMRI and PET) experiments with coordinate-based (x,y,z) activation locations in Talairach space.

CARMEN: Acronym for "Code Analysis, Repository & Modeling for E-Neuroscience"; A project that aims to develop and deploy a "virtual laboratory for neurophysiology, which will enable the sharing and collaborative exploitation of data, analysis code and expertise that are not physically co-located" (CARMEN eScience Longitudinal Study No. 1, p. 1)

CNARI: Acronym for "Computational Neuroscience Applications Research Infrastructure" ; A framework for research and analysis that aims to manage neuroscience tools and the heterogeneous compute resources (such as Grids and clusters) on which they can be used, to enable large-scale computational projects in the neuroscience community" (<http://www.ci.uchicago.edu/wiki/bin/view/CNARI>)

CogPO: Acronym for "Cognitive Paradigm Ontology"; A framework for the formal representation of the cognitive paradigm; project is driven by cognitive neuroscience and neuroimaging examples.

Controlled vocabulary: Preferred terms or phrases that are designated for use in a catalog, database, or other retrieval tool to be used for all the various phrases and words used to describe a concept. May also be called "descriptors" or "subject headings", and be listed in a "thesaurus".

Data annotation: The process of associating ("tagging") data with meaningful symbols ("tags"); tags may come from a controlled vocabulary or ontology

Data mining: Bottom-up process for extracting patterns (information) from data

Data model: A schema that specifies the structure of data; cf. "structured data"

Data-driven: A synonym for "bottom-up"

Database schema: Information that specifies the structure of a database

Description logic: A language that is similar to propositional logic and is used to model concepts (aka "classes"), individuals (aka "instances"), and their relationships.

DL: Acronym for "description logic"

EEG: Electroencephalogram; brainwaves

ERP: Event-related potentials (averaged EEG data time-locked to an event of interest)

Extensible mark-up language: A set of rules for encoding documents in machine-readable form; see also XML

Graph: A set of nodes (~entities) and edges (~relations) that connect the nodes

HeadIT: Acronym for "A Human Electrophysiology, Associated Anatomic Data and Integrated Tool resource"; A project that aims to provide database support for large-scale EEG data storage, management, and analysis

IAO: Acronym for "Information Artifact Ontology"; An ontology of information entities, originally driven by work by the OBI digital entity and realizable information entity branch.

Inference: A process that applies a set rules (or *axioms*) to set of assertions (i.e., *assumptions*) in order to draw a new set of assertions (i.e., *conclusions*).

Knowledge engineering: A top-down process for capturing knowledge in a domain

Knowledge-driven: A synonym for "top-down"

MATLAB: A high-level technical computing language and interactive environment for algorithm development, data visualization, data analysis, and numeric computation

MEG: Magnetoencephalogram; magnetic field potentials

Metadata: "data about data"; Statements or tags that identify the contents, structure, or provenance of data

MI checklist: "Minimal information checklist"

MIBBI: Acronym for "Minimum Information for Biological and Biomedical Investigations"; MIBBI provides a resource for exploring the range of extant minimum information checklists and fosters coordinated development of such checklists.

NEMO: Acronym for "Neural ElectroMagnetic Ontologies"; An NIH funded project researching the design and implementation of ontologies to address a critical need for tools to support representation, storage, and sharing of brain electromagnetic data.

Neurolex: An open-source lexicon (dictionary) that provides concise information about common concepts in neuroscience in a form that is easy for automated agents to understand. The lexicon is intended to help improve the way that neuroscientists communicate about their data, so that information systems like the NIF can find data more easily and support distributed data sharing and integration.

NIF: Acronym for "Neuroscience Information Framework"; A dynamic inventory of Web-based neuroscience resources: data, materials, and tools accessible via any computer connected to the Internet.

NIFSTD: Acronym for "NIF Standard ontology"; An OWL ontology consisting of multiple modules that represent various domains of biomedical reality.

OBI: Acronym for "Ontology for Biomedical Investigations"; An integrated ontology for the description of life-science and clinical investigations.

OBO: Acronym for "Open Biological and Biomedical Ontologies"

OBO Foundry: "A collaborative experiment involving developers of science-based ontologies who are establishing a set of principles for ontology development with the goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain" (<http://www.obofoundry.org/>)

Ontology: A formal representation of knowledge consisting of a set of concepts within a domain and the relationships among those concepts

Ontology mining: A bottom-up process for "discovering knowledge" (i.e., ontologies) from data

OWL: Acronym for "web ontology language"; A computer language commonly used to encode ontologies

Protégé: A free, open source ontology editor

Provenance: Information that describes the history (or lineage) of a dataset (where it came from, how it was processed, etc.)

Query: A process (request) that is designed to retrieve information from a database

RDF: Acronym for "resource description framework"; RDF is a directed, labeled graph (data model) for representing information on the Web; RDF is a generic data model that can represent just about any data

RDF graph: A set of RDF triples, which can be visualized as a graph

RDF triple: An RDF expression of the form (Subject -- Predicate -- Object)

RDFS: Acronym for "RDF schema"; An RDF data model used to describe classes, properties, and a few very basic relationships among them, such as subClassOf. (*Note that OWL extends RDFS with more complex relationships*)

Relational database: A collection of tables (aka "relations") that can be used to store structured data using the relational model

Relational model: A formal data model developed by EF Codd based on logic, where logical predicates are represented as tables (or relations) and data contained in those tables comprise the logical model (the grounding of variables) for the system

Schema: A collection of tables (within a database); Specification of the logical structure of data

SPARQL: Acronym for "SPARQL Protocol and RDF Query Language"; RDF query language; The standard query language for databases using the RDF model; Acronym for "SPARQL Protocol and RDF Query Language"

SQL: Acronym for "structured query language"; A database computer language designed for managing data in relational database management systems; The standard query language for databases using a relational model

Structured data: Data that are annotated or tagged with machine-readable symbols

Taxonomy: A hierarchical representation of terms (~ "subsumption hierarchy")

Top-down: A process that uses prior knowledge to guide data processing

Tuple: A record (e.g. a row in a table in a relational database)

Uniform Resource Identifier: A compact sequence of characters that identifies an abstract or physical resource.

URI: Acronym for "Uniform Resource Identifier"

Virtuoso: An open-source, scalable, cross-platform server that combines SQL/RDF/XML Data Management with Web Application Server and Web Services Platform functionality. (<http://sourceforge.net/projects/virtuoso>); A brand of database management system which stores RDF data and answers SPARQL queries, often called a triplestore.

XML: Acronym for "extensible mark-up language"; A set of rules for encoding documents in machine-readable form.