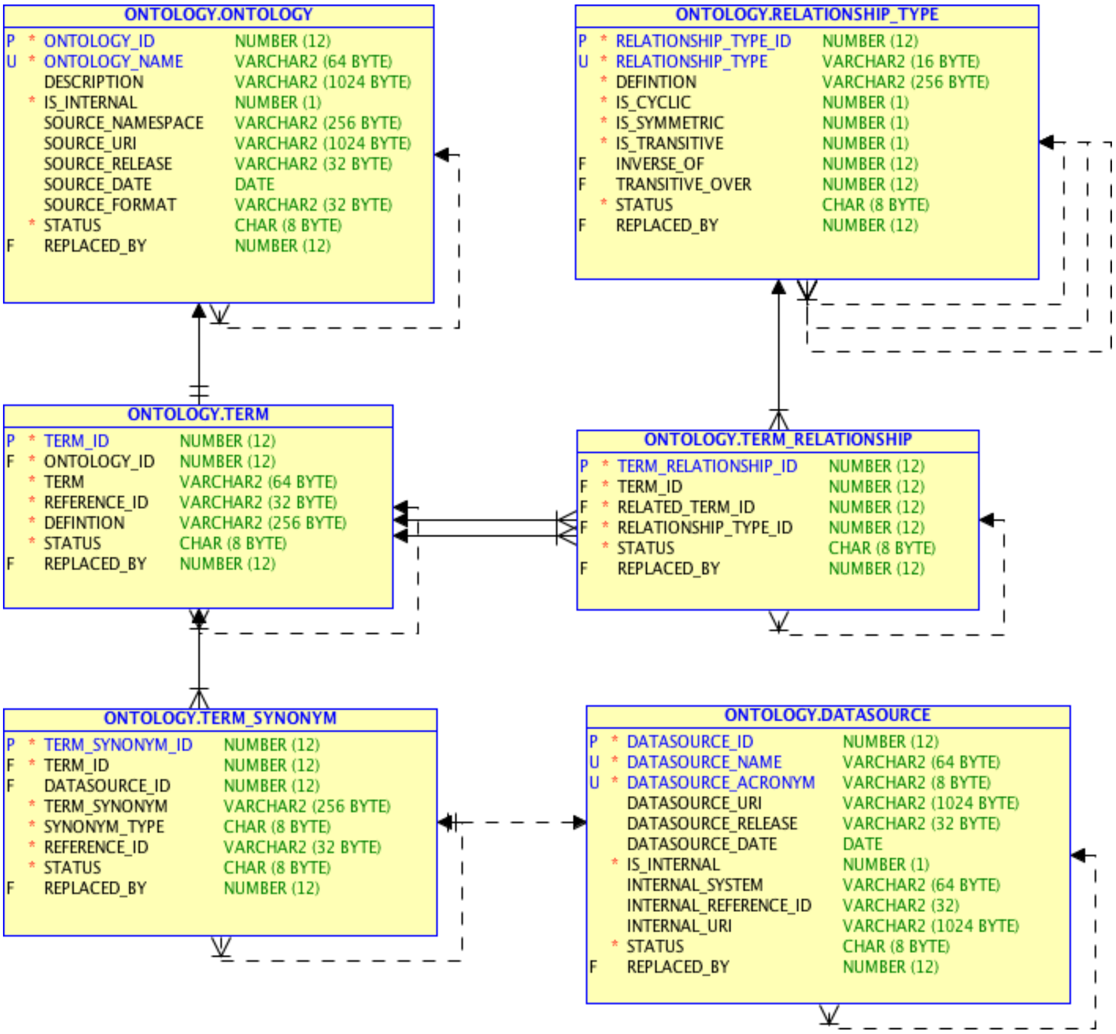
OntoBrowser Database Design

The OntoBrowser database design is based on the Open Biological and Biomedical Ontologies (OBO) file format. The scope of this document is limited to the database schema design and not intended to provide an understanding of ontology concepts.

The following Entity Relationship Diagram (ERD) represents a simplified[1] representation of the underlying database schema.



ONTOLOGY.ONTOLOGY			
P	*	ONTOLOGY_ID	NUMBER (12)
U	*	ONTOLOGY_NAME	VARCHAR2 (64 BYTE)
		DESCRIPTION	VARCHAR2 (1024 BYTE)
	*	IS_INTERNAL	NUMBER (1)
		SOURCE_NAMESPACE	VARCHAR2 (256 BYTE)
		SOURCE_URI	VARCHAR2 (1024 BYTE)
		SOURCE_RELEASE	VARCHAR2 (32 BYTE)
		SOURCE_DATE	DATE
		SOURCE_FORMAT	VARCHAR2 (32 BYTE)
	*	STATUS	CHAR (8 BYTE)
F		REPLACED_BY	NUMBER (12)

ONTOLOGY.RELATIONSHIP_TYPE			
P	*	RELATIONSHIP_TYPE_ID	NUMBER (12)
U	*	RELATIONSHIP_TYPE	VARCHAR2 (16 BYTE)
	*	DEFINTION	VARCHAR2 (256 BYTE)
	*	IS_CYCLIC	NUMBER (1)
	*	IS_SYMMETRIC	NUMBER (1)
	*	IS_TRANSITIVE	NUMBER (1)
F		INVERSE_OF	NUMBER (12)
F		TRANSITIVE_OVER	NUMBER (12)
	*	STATUS	CHAR (8 BYTE)
F		REPLACED_BY	NUMBER (12)

ONTOLOGY.TERM			
P	*	TERM_ID	NUMBER (12)
F	*	ONTOLOGY_ID	NUMBER (12)
	*	TERM	VARCHAR2 (64 BYTE)
	*	REFERENCE_ID	VARCHAR2 (32 BYTE)
	*	DEFINTION	VARCHAR2 (256 BYTE)
	*	STATUS	CHAR (8 BYTE)
F		REPLACED_BY	NUMBER (12)

ONTOLOGY.TERM_RELATIONSHIP			
P	*	TERM_RELATIONSHIP_ID	NUMBER (12)
F	*	TERM_ID	NUMBER (12)
F	*	RELATED_TERM_ID	NUMBER (12)
F	*	RELATIONSHIP_TYPE_ID	NUMBER (12)
	*	STATUS	CHAR (8 BYTE)
F		REPLACED_BY	NUMBER (12)

ONTOLOGY.TERM_SYNONYM			
P	*	TERM_SYNONYM_ID	NUMBER (12)
F	*	TERM_ID	NUMBER (12)
F		DATASOURCE_ID	NUMBER (12)
	*	TERM_SYNONYM	VARCHAR2 (256 BYTE)
	*	SYNONYM_TYPE	CHAR (8 BYTE)
	*	REFERENCE_ID	VARCHAR2 (32 BYTE)
	*	STATUS	CHAR (8 BYTE)
F		REPLACED_BY	NUMBER (12)

ONTOLOGY.DATASOURCE			
P	*	DATASOURCE_ID	NUMBER (12)
U	*	DATASOURCE_NAME	VARCHAR2 (64 BYTE)
U	*	DATASOURCE_ACRONYM	VARCHAR2 (8 BYTE)
		DATASOURCE_URI	VARCHAR2 (1024 BYTE)
		DATASOURCE_RELEASE	VARCHAR2 (32 BYTE)
		DATASOURCE_DATE	DATE
	*	IS_INTERNAL	NUMBER (1)
		INTERNAL_SYSTEM	VARCHAR2 (64 BYTE)
		INTERNAL_REFERENCE_ID	VARCHAR2 (32)
		INTERNAL_URI	VARCHAR2 (1024 BYTE)
	*	STATUS	CHAR (8 BYTE)
F		REPLACED_BY	NUMBER (12)

The following sections of this document describe the core entities and their relationships (depicted in the ERD above). See Appendix I – Curators, Data Approval and Versions for details regarding the *Curator* and *Version* entities and their relationships with the core entities.

[1] The *Curator* and *Version* entities and their relationships (i.e. foreign key columns) have been excluded for clarity

Ontology Table

An ontology represents a set of domain specific terms and their relationships (e.g. anatomical terms, histopathology finding terms, clinical pathology analyte terms etc…). Multiple ontologies can be defined in the database. Each ontology must have a unique name (as defined within the context of database). The core columns of the Ontology table are described below:

COLUMN	DATA TYPE	NULL ABLE	COMMENTS
ONTOLOGY_ID	NUMBER(12)	No	Primary key (generated)
ONTOLOGY_NAME	VARCHAR2(64)	No	Unique name/domain of the ontology e.g. anatomy
DESCRIPTION	VARCHAR2(1024)	Yes	Description of domain/ontology
IS_INTERNAL	NUMBER(1)	No	Flag indicating if the ontology is internally maintained
SOURCE_NAMESPACE	VARCHAR2(256)	Yes	Namespace as defined in the originating source data.
SOURCE_URI	VARCHAR2(1024)	Yes	Link to originating source data e.g. OBO file in SVN
SOURCE_RELEASE	VARCHAR2(32)	Yes	Version or release of originating source data
SOURCE_DATE	DATE	Yes	Date/Timestamp of originating source data
SOURCE_FORMAT	VARCHAR2(32)	Yes	Format of originating source data e.g. OBO 1.2
STATUS	CHAR(8)	No	One of four possible values: • PENDING • APPROVED • REJECTED • OBSOLETE See appendix for more details
REPLACED_BY	NUMBER(12)	Yes	Foreign key - Id of the ontology that replaces this obsolete ontology.

Term Table

An ontology term must be defined within the context of an ontology i.e. a term cannot exist without being linked to ontology. A term must also be unique with respect to all other terms defined within the associated ontology. A unique reference identifier must also be assigned to each term. The core columns of the Term table are described below:

COLUMN	DATA TYPE	NULL ABLE	COMMENTS
TERM_ID	NUMBER(12)	No	Primary key (generated)
ONTOLOGY_ID	NUMBER(12)	No	Foreign key to Ontology table
TERM_NAME	VARCHAR2(64)	No	Unique name of the term within the ontology
REFERENCE_ID	VARCHAR2(32)	No	Unique external reference identifier of the term
DEFINITION	VARCHAR2(1024)	Yes	Definition of the term

COMMENTS	VARCHAR2(1024)	Yes	Comments
STATUS	CHAR(8)	No	One of four possible values: • PENDING • APPROVED • REJECTED • OBSOLETE See appendix for more details
IS_ROOT	NUMBER(1)	No	Flag to indicate a root term
REPLACED_BY	NUMBER(12)	Yes	Foreign key - Id of term that replaces this obsolete term

Relationship_Type Table

Relationship types model the relationships between ontology terms. For example, consider the relationship: *finger* **part_of** *hand*. This means that for any particular finger in the real world there must be a particular hand in the real world that the finger is part of.

The following built-in (and/or ubiquitous) OBO relationship types will be predefined in the Relationship_Type table:

- is_a
- disjoint_from
- part_of
- union_of

Curators can define addition relationship types as required. Note that relationship types have global scope and hence are not tied to a specific ontology. The core columns of the Relationship_Type table are described below:

COLUMN	DATA TYPE	NULL ABLE	COMMENTS
RELATIONSHIP_TYPE_ID	NUMBER(12)	No	Primary key (generated)
RELATIONSHIP_TYPE	VARCHAR2(16)	No	Unique name of relationship type e.g. is_a
DEFINTION	VARCHAR2(256)	No	Definition of the relationship type
IS_CYCLIC	NUMBER(1)	No	Whether or not a cycle can be made from this relationship type. If a relationship type is non-cyclic, it is illegal for an ontology to contain a cycle made from user-defined **or** implied relationships of this type
IS_SYMMETRIC	NUMBER(1)	No	Whether this relationship is symmetric. All symmetric relationships are also cyclic.
IS_TRANSITIVE	NUMBER(1)	No	Whether this relationship is transitive. The **is_a** relationship type is transitive by definition.
INVERSE_OF	NUMBER(12)	Yes	Foreign key - the id of another relationship type that is the

			inverse of this relationship type. This symmetric relation indicates that a relation is the inverse of another relation e.g. **part_of** is the **inverse_of has_part**
TRANSITIVE_OVER	NUMBER(12)	Yes	Foreign key - The id of another relationship type that this relationship type is transitive over. If P is transitive over Q, and the ontology has X P Y and Y Q Z then it follows that X P Z
STATUS	CHAR(8)	No	One of four possible values: • PENDING • APPROVED • REJECTED • OBSOLETE See appendix for more details
REPLACED_BY	NUMBER(12)	Yes	Foreign key - Id of the relationship type that replaces this obsolete relationship type

Term_Relationship Table

Multiple relationships can be defined between ontology terms. Relationships can also exist between terms from different ontologies. Relationships between ontology terms are defined in the *Term_Relationship* table. The core columns of the Term_Relationship table are described below:

COLUMN	DATA TYPE	NULLABLE	COMMENTS
TERM_RELATIONSHIP_ID	NUMBER(12)	No	Primary key (generated)
TERM_ID	NUMBER(12)	No	Foreign key – id of the term
RELATED_TERM_ID	NUMBER(12)	No	Foreign key – id of the related term
RELATIONSHIP_TYPE_ID	NUMBER(12)	No	Foreign key – id of the relationship type
DIFFERENTIA_RELSHIP_TYPE_ID	NUMBER(12)	Yes	Foreign key – id of the differentia relationship type
STATUS	CHAR(8)	No	One of four possible values: • PENDING • APPROVED • REJECTED • OBSOLETE See appendix for more details
REPLACED_BY	NUMBER(12)	Yes	Foreign key - Id of the relationship that replaces this obsolete relationship type

Datasource Table

An ontology term can have any number of synonyms. Some synonyms are sourced from controlled vocabularies from other system (internal or external).

A datasource represents the source of a set of synonyms e.g. SEND or an external CRO. The core columns of the Datasource table are described below:

COLUMN	DATA TYPE	NULL ABLE	COMMENTS
DATASOURCE_ID	NUMBER(12)	No	Primary key (generated)
DATASOURCE_NAME	VARCHAR2(64)	No	Unique name of the datasource. This could be an external CRO or internal database system
DATASOURCE_ACRONYM	VARCHAR2(8)	No	Unique acronym of the datasource e.g. STS
DATASOURCE_URI	VARCHAR2(1024)	Yes	Link to datasource e.g URL
DATASOURCE_RELEASE	VARCHAR2(32)	Yes	Version/Release of the datasource
DATASOURCE_DATE	DATE	Yes	Release date of the datasource
IS_INTERNAL	NUMBER(1)	No	Flag to indicate if the datasource is internal
STATUS	CHAR(8)	No	One of four possible values: • PENDING • APPROVED • REJECTED • OBSOLETE See appendix for more details
REPLACED_BY	NUMBER(12)	Yes	Foreign key – Id of the datasource that replaces this obsolete datasource

Term_Synonym Table

A synonym for an ontology term represents an alternative name. Terms can have multiple synonyms potentially from multiple sources. Synonyms provide a way to map between the controlled vocabularies of various source systems to the prefer ontology terms. The core columns of the Term_Synonym table are described below:

COLUMN	DATA TYPE	NULL ABLE	COMMENTS
TERM_SYNONYM_ID	NUMBER(12)	No	Primary key (generated)
TERM_ID	NUMBER(12)	No	Foreign key – id of the term
DATASOURCE_ID	NUMBER(12)	Yes	Foreign key – id of the datasource
TERM_SYNONYM	VARCHAR2(256)	No	Synonym for the term as defined by the datasource
SYNONYM_TYPE	CHAR(8)	No	One of four possible values: • BROAD • EXACT • NARROW • RELATED
REFERENCE_ID	VARCHAR2(32)	Yes	Optional external reference id for synonym
STATUS	CHAR(8)	No	One of four possible values: • PENDING • APPROVED • REJECTED

			• OBSOLETE See appendix for more details
REPLACED_BY	NUMBER(12,0)	Yes	Foreign key – Id of the synonym that replaces this obsolete synonym

Appendix I – Curators, Data Approval and Versions

Curator Roles

A curator represents a user responsible for maintaining ontology data within the OntoBrowser Database.

There are two (superseding) curator roles:

1. Data Entry – all curators will be able to perform data entry with respect to the core entities (as described in the main part of this document). The data entry operations can be summarized as follows:
    a. Define new ontology
    b. Define new ontology term
    c. Define new relationship types
    d. Define new relationships between ontology terms
    e. Define new synonym datasources
    f. Define new synonyms
2. Approver – a restricted subgroup of curators will be responsible for:
    a. Approving (or rejecting) newly created entries
    b. Rendering previously approved entries obsolete
    c. Creating new versions/snapshots of the data (see Versioning)

See sections below for details regarding data approval and versioning.

Curator Table

Curators will need to be added to the *Curator* table as part of the registration process. The columns of the Curator table are described below:

COLUMN	DATA TYPE	NULL ABLE	COMMENTS
CURATOR_ID	NUMBER(12)	No	Primary key (generated)
USERNAME	VARCHAR2(8)	No	User's 5-2-1 i.e. ENGINE username or LDAP uid
EMAIL_ADDRESS	VARCHAR2(64)	No	User's email address
IS_ACTIVE	NUMBER(1)	No	Flag indicating if the curator is active
CREATED_DATE	DATE	No	Date/Time this curator entry was created (generated)
CREATED_BY	NUMBER(12)	No	Foreign key - Id of the curator that created this entry
MODIFIED_DATE	DATE	Yes	Date/Time this curator entry was last modified (generated)

MODIFIED_BY	NUMBER(12)	Yes	Foreign key - Id of the curator that last modified this entry

Curator Foreign Key Columns

All the core entity tables contain the following columns (in addition to the columns described in the main part of this document):

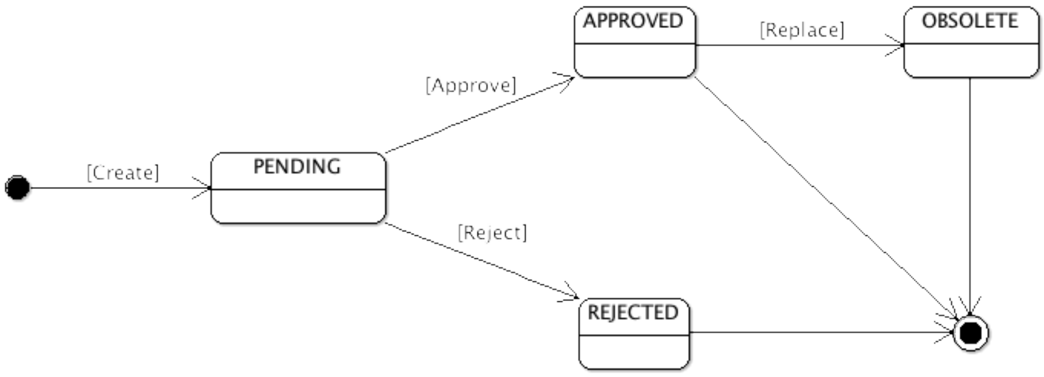
COLUMN	DATA TYPE	NULL ABLE	COMMENTS
CREATED_DATE	DATE	No	Date/Time when this entry was created (generated)
CREATED_BY	NUMBER(12)	No	Foreign key – Id of the curator that created this entry

Entity Status/State

All the core entity tables (as described in the main part of this document) include a STATUS column:

COLUMN	DATA TYPE	NULL ABLE	COMMENTS
STATUS	CHAR(8)	No	One of four possible values: • PENDING • APPROVED • REJECTED • OBSOLETE

This column indicates the current status/state of the entry. Newly created entries will always have a status of PENDING i.e. pending approval. After an entry has been approved, its status will be updated to APPROVED. If an entry has been rejected, its status will be REJECTED. Previously approved entries may be rendered obsolete (potentially being replaced by a new entry). The UML state diagram below depicts these states and transitions:



Note: only entries with status/state of APPROVED will be present in the public materialized views.

Data Approval (or Rejection)

In order for an entry to transition from the PENDING to APPROVED (or REJECTED) state, it must be approved (or rejected) by one or more curators. The number of curators required depends on the *approval weight* of each curator for the associated entity.

A curator's *approval weight* is a real number between zero and one (inclusive) i.e. 0 <= APPROVAL_WEIGHT <= 1. An entry is considered approved when the sum of the *approval weight* is equal to (or greater than) one. For example, if Bob and Jane both have an *approval weight* of 0.5 then both Bob and Jane would need to approve (or reject) an entry in order for it to transition from PENDING to APPROVED (or REJECTED).

Note: A curator can only approve or reject an entry once.

Curator_Approval_Weight Table

The Curator_Approval_Weight table stores the *approval weight* for each curator per core entity:

COLUMN	DATA TYPE	NULL ABLE	COMMENTS
CURATOR_APPROVAL_ID	NUMBER(12)	No	Primary key (generated)
CURATOR_ID	NUMBER(12)	No	Foreign key – Id of the curator
TABLE_NAME	VARCHAR2(32)	No	One of six possible values: • RELATIONSHIP_TYPE • DATASOURCE • ONTOLOGY • TERM • TERM_RELATIONSHIP • TERM_SYNONYM
APPROVAL_WEIGHT	NUMBER(9,8)	No	Approval weight of curator for table
CREATED_DATE	DATE	No	Date/Time this approval weight was created (generated)
CREATED_BY	NUMBER(12)	No	Foreign key – Id of the curator that created this approval weight
MODIFIED_DATE	DATE	Yes	Date/Time this approval weight was last modified (generated)
MODIFIED_BY	NUMBER(12)	Yes	Foreign key – Id of the curator that last modified this approval weight

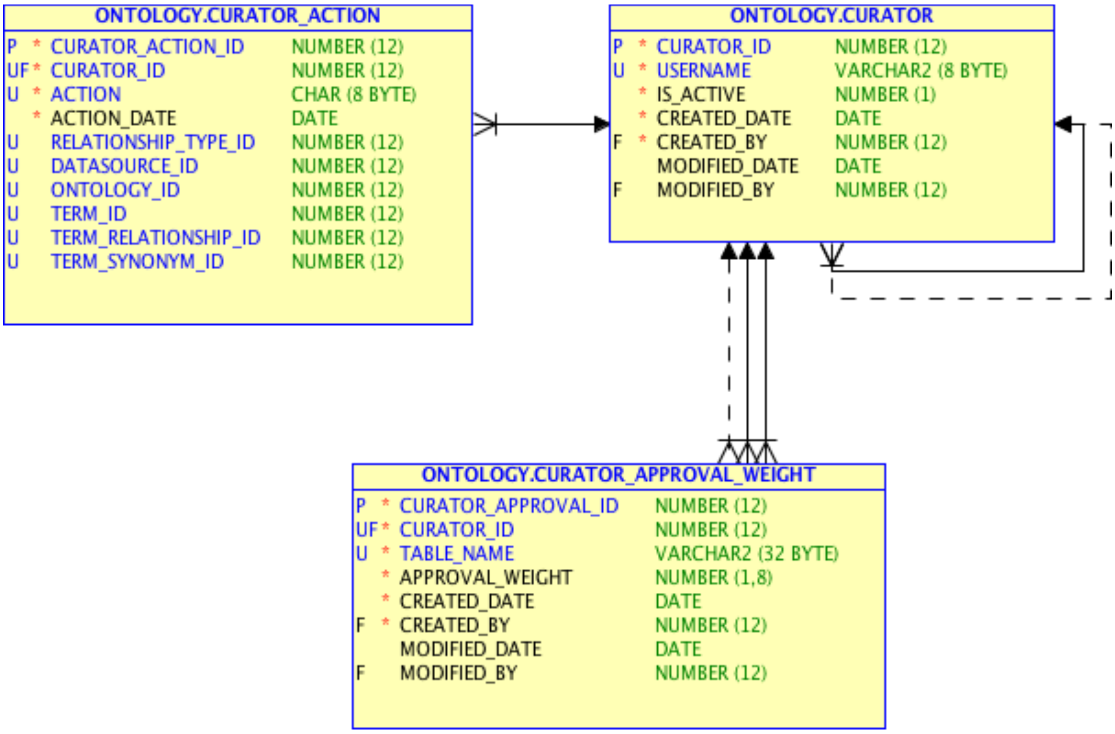
Curator_Action Table

Curator actions (i.e. approvals, rejections and replacements) are recorded in the Curator_Action table. The columns of the Curator_Action table are described below:

COLUMN	DATA	NULL	COMMENTS

	TYPE	ABLE	
CURATOR_ACTION_ID	NUMBER(12)	No	Primary key (generated)
CURATOR_ID	NUMBER(12)	No	Foreign key – Id of the curator that performed this action
ACTION	CHAR(8)	No	Action performed. One of three possible values: • APPROVE • REJECT • REPLACE
ACTION_DATE	DATE	No	Date/Time the action was performed
ONTOLOGY_ID	NUMBER(12)	Yes	Foreign key – Id of the ontology approved, rejected or replaced
TERM_ID	NUMBER(12)	Yes	Foreign key – Id of the term approved, rejected or replaced
RELATIONSHIP_TYPE_ID	NUMBER(12)	Yes	Foreign key – Id of the relationship type approved, rejected or replaced
TERM_RELATIONSHIP_ID	NUMBER(12)	Yes	Foreign key – Id of the term relationship approved, rejected or replaced
DATASOURCE_ID	NUMBER(12)	Yes	Foreign key – Id of the synonym datasource approved, rejected or replaced
TERM_SYNONYM_ID	NUMBER(12)	Yes	Foreign key – Id of the term synonym approved, rejected or replaced

The following ERD depicts the Curator_Action, Curator and Curator_Approval_Weight tables and their relationships. Note: the foreign key relationships between the Curator_Action table and the core entity tables (as described in the main part of the this document) have been excluded from the ERD for clarity.



ONTOLOGY.CURATOR_ACTION

P	*	CURATOR_ACTION_ID	NUMBER (12)
UF	*	CURATOR_ID	NUMBER (12)
U	*	ACTION	CHAR (8 BYTE)
	*	ACTION_DATE	DATE
U		RELATIONSHIP_TYPE_ID	NUMBER (12)
U		DATASOURCE_ID	NUMBER (12)
U		ONTOLOGY_ID	NUMBER (12)
U		TERM_ID	NUMBER (12)
U		TERM_RELATIONSHIP_ID	NUMBER (12)
U		TERM_SYNONYM_ID	NUMBER (12)

ONTOLOGY.CURATOR

P	*	CURATOR_ID	NUMBER (12)
U	*	USERNAME	VARCHAR2 (8 BYTE)
	*	IS_ACTIVE	NUMBER (1)
	*	CREATED_DATE	DATE
F	*	CREATED_BY	NUMBER (12)
		MODIFIED_DATE	DATE
F		MODIFIED_BY	NUMBER (12)

ONTOLOGY.CURATOR_APPROVAL_WEIGHT

P	*	CURATOR_APPROVAL_ID	NUMBER (12)
UF	*	CURATOR_ID	NUMBER (12)
U	*	TABLE_NAME	VARCHAR2 (32 BYTE)
	*	APPROVAL_WEIGHT	NUMBER (1,8)
	*	CREATED_DATE	DATE
F	*	CREATED_BY	NUMBER (12)
		MODIFIED_DATE	DATE
F		MODIFIED_BY	NUMBER (12)

Versioning

The ontology database includes a versioning system. A version represents a snapshot of the core entity data (as described in the main part of this document). Curators (with the *Approver* role) are responsible for creating new versions/snapshots. The materialized views (that provide an interface for external system to access the ontology data) are refreshed when a new version is created.

Version Table

The Version table records each new version created. The columns of the Version table are described below:

COLUMN	DATA TYPE	NULL ABLE	COMMENTS
VERSION_ID	NUMBER(12)	No	Primary key (generated)
CREATED_DATE	DATE	No	Date/Time this version was created (generated)
CREATED_BY	NUMBER(12)	No	Foreign key - Id of the curator that created this version
PUBLISHED_DATE	DATE	Yes	Date this version was published
PUBLISHED_BY	NUMBER(12)	No	Foreign key - Id of the curator that published this version

Version Foreign Key Columns

All core entity tables include the following columns (in addition to those described in the main part of this document):

COLUMN	DATA TYPE	NULL ABLE	COMMENTS
CREATED_VERSION_ID	NUMBER(12)	No	Foreign key – Id of Version the entry was created
APPROVED_VERSION_ID	NUMBER(12)	Yes	Foreign key – Id of Version the entry was approved
OBSOLETE_VERSION_ID	NUMBER(12)	Yes	Foreign key – Id of Version the entry was replaced or rendered obsolete