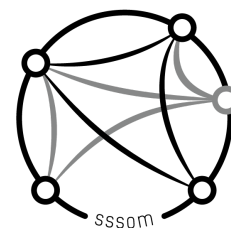


5-Star Entity Mappings - Cheatsheet

Towards more reusable and transparent mappings for Open Science.

Entity Mappings connect clinical codes, data model enums, ontology classes and terms in clinical terminologies across knowledge organization systems and databases. Entity mappings are pivotal for the integration of healthcare data, but they are expensive to produce and often use-case-dependent. Despite the cost of creating these mappings, they are rarely shared across organizations, and even if they are made available, they lack standardization and metadata. Here we outline a 5-Star mapping system (inspired by Tim Berners-Lee's 5-star system for Linked Data) to help you bring mapping production in your organization to the next level - step by step.



★ 1-Star Mappings

- **Goal:** Export mapping in a computationally accessible format, make it publicly available and record mapping precision.
- **Implementation:**
 - Record subject id, object id and mapping precision (exact, broad, narrow, close, related)
 - Use globally unique and persistent identifiers for subject id and object id (e.g. OMOP:123456)
 - Use a computable file format (JSON, XML, CSV, TSV) rather than XLSX or HTML
 - Make mappings available in a public space without access restrictions
 - **Optional:** record the subject and object labels to make it easier for humans to read the file
- **Enables:**
 - Direct integration into ETL pipelines
 - Dropping societal costs by enabling others to re-use mappings
 - Moving data between semantic spaces (rather than just “into OMOP”)

★★ 2-Star Mappings

- **Goal:** Make mappings available in a place with version control suitable for providing community feedback, make your own uncertainty explicit, add license and select semantic mapping predicate.
- **Implementation**
 - Make mapping set available in a public version control system (e.g. GitHub) with an issue tracker
 - Record the semantic predicate explicitly (e.g. owl:equivalentClass, skos:exactMatch)
 - Record a confidence value for the mapping between 0 and 1, where appropriate
 - Use a standard open license for the use of the mapping set (e.g. Creative Commons)
- **Enables:**
 - The worry-free re-use of mappings even if target or source terminologies are “closed”
 - Transparently versioned access to mappings and the opportunity to provide more direct feedback
 - Downstream users can filter for high-confidence mappings

★ ★ ★ 3- Star Mappings

- **Goal:** Export mappings in a community standard format with basic versioning and provenance information.
- **Implementation**
 - Export mappings in SSSOM (<https://w3id.org/sssom>) format (you do not have to curate using SSSOM!)
 - Record the following additional metadata
 - mapping_justification(s) (Lexical, Logical match, Human curated etc.)
 - mapping_date
 - subject_source, object_source, subject_source_version, object_source_version
 - mapping_tool (if the mapping was automatically computed using a tool), creator_id
- **Enables:**
 - Dropping costs of reusing mappings further by providing a standard format to exchange mappings
 - Enabling the decentralised production of mappings by independent expert communities
 - Basic metadata, in particular justifications, enable downstream users to assess “fitness for purpose” for a different context

★ ★ ★ ★ 4- Star Mappings

- **Goal:** Make it easier to discover mappings by registering them at public mapping registry
- **Implementation**
 - Register the mapping at a mapping commons (if none exists, create one)
 - Record the following additional metadata:
 - mapping_set_id, mapping_set_description, mapping_set_version
 - mapping_provider (if the mapping is not original, i.e. it is not derived from another source)
 - Provide an executable mapping_justification (see <https://w3id.org/sssom#minimum>)
- **Enables:**
 - Mappings can easily be made available by Open Terminology services which enable scalable data mapping services

★ ★ ★ ★ ★ 5- Star Mappings

- **Goal:** Ensure currency of mappings
- **Implementation:**
 - Mappings are up-to-date with the latest versions of the sources being mapped
 - Have no issue on their issue tracker open for more than 3 months without an interaction
 - Usually requires a lifecycle management system that integrates automated matching
- **Enables:**
 - Reduced effort dealing with mappings to deprecated codes or classes
 - Worry-free application of mappings in automated ETL processes