

Setting up a New Project

Overview

The tool was developed and initially populated to serve specific use cases for pairs of terminologies. This page is to discuss and document what is needed to initiate a new map project.

Checklist

For the proposed project (e.g. SNOMED to ICNP map), answer the following questions

- **Q0:** Is the data for this project currently available? How is it obtained and who are the contacts?
- **Q1:** What is the format of the source and target terminologies? Do current loaders support these terminologies?
 - The tool currently supports three load formats: RF2, ClaML, and "code list and par/chd code file".
 - **Q1a:** Are there any idiosyncracies of the terminology - e.g. things like many children of a single node, or a very deep hierarchy (>11 levels).
 - **Q1b:** Is the terminology just descriptions and relationships, or are there also important other attributes (e.g. dagger/asterisk) that are relevant for mapping.
 - **Q1c:** How often is the terminology updated? What kinds of relevant changes should trigger re-editing of a map?
 - **Q1d:** Is the entire terminology to be mapped or just part of it?
 - **Q1e:** What is the style of the codes? Do they use any special characters that may adversely affect current search algorithms?
 - **Q1f:** What is the length of the longest "name" in the terminology and will that fit within bounds of the defined field (e.g. 4000 chars).
 - **Q1g:** How do data structures of the terminology align with data structures used internally in the tool - "concept", "description", "relationship", "language"
 - e.g. sometimes important "attributes" in a terminology are rendered as descriptions so they display in the right places and are searchable.
 - **Q1h:** Is there a need to create a "pseudo-hierarchy"?
- **Q2:** What is the scope definition in the source terminology?
 - e.g., SNOMED to ICD10 uses all descendants of the SNOMED concept "Clinical Finding"
 - Scope definition can use "include" concepts (with/without descendants) and "exclude" concepts (with/without descendants)
- **Q3:** What defines an "allowable code" in the target terminology?
 - This is defined by the "project specific algorithm handler" - code that must be written to support the project.
 - **Q3a:** Are there allowable codes outside the scope of codes immediately defined in the terminology (e.g. ICDO may have "valid" combinations of characters after the slash that do not per-se exist in the terminology)
- **Q4:** What is the desired release format of the map? Do current release mechanisms support this format?
 - The tool currently supports a release process based around RF2 and a terminology "map" loader to load in prior versions of maps for delta computation.
- **Q5:** Is there an existing published map that should be loaded? Do current loaders support the publication format?
 - The tool currently supports loading RF2 simple, complex, and extended map formats.
- **Q6:** What are the basic project configuration settings for this map? (this has several parts)
 - **Q6a:** Refset id (whether your map uses refsets or not, this has to be set to something non-null and unique).
 - **Q6b:** Does the map use group structure (e.g. multiple potential targets for the same)
 - **Q6c:** Does the map use rules for maps?
 - **Q6e:** What is the map publication type? - Simple, Complex, or Extended (all of these are in service of an RF2 release process)
 - Map Relation Style can be determined based on publication type. If this is a Snomed Extended map, use Map Relation Style = "Map Category Style". Otherwise use "No Relationships."
 - **Q6f:** What are the map relations?
 - e.g. "broader", "narrower", "exact", "partial", ...
 - e.g. "MAP OF SOURCE CONCEPT IS CONTEXT DEPENDENT", ..
 - **Q6g:** What are the allowable map advices?
 - **Q6h:** What are the preset age ranges? (only relevant if a mapping has rules based around age ranges)
 - **Q6i:** What are the supported "error messages" for the project? this only applies to projects with conflict workflows.
 - **Q6j:** What are the allowable map advices?
 - **Q6k:** Is there a mapping handbook document - or editorial guide? (this can be uploaded and made available to users)
- **Q7:** Who are the users for the project and what roles do they have?
 - "Specialist" - user participates in authoring the initial map
 - "lead" - user can act as a specialist, but also participates in review stages of workflows
- **Q8:** What workflow should the map use?
 - "Non-legacy" - two specialists, single lead review on conflict
 - "Legacy" - legacy map and single specialist, second specialist if conflict, single lead review on conflict
 - "Review" - single specialist, single lead review
 - "Simple" - single specialist, no review
- **Q9:** What are the needs of the project specific algorithm handler?
 - It defines valid target codes, semantic validation rules (errors/warnings), automations for adding/changing advice, etc.
- **Q10:** What reports are relevant for this project? (e.g. specialist productivity)
 - How often should reports be run? Are there important "diff" reports?
 - Do the currently defined reports provide necessary coverage or are new reports needed?
- **Q11:** What QA checks are relevant for this project? (e.g. checks that produce work to put into workflow)
 - Do the currently defined checks provide necessary coverage or are new checks needed?
- **Q12:** Does the project come with accompanying "index" data? If so, what format is it in and will current index viewer approaches suffice?
- **Q13:** Are there rules for semi-automated mapping that could apply? (e.g. this is something of a future consideration, but at present could be done outside of tool and loaded as a "legacy" map).
- **Q14:** How is this project affected by "drip feed"?

- In particular - are there dependencies for release cycles on a particular version of SNOMED? (if so - then release should be coordinated with international release)
- Also, does/should "drip feed" results affect the scope definition?

Details

Answers to the above questions will lead to decisions about the need for new features within the application (e.g. new loaders).

The strategy for initiating a new mapping project is to obtain answers to the questions above, implement necessary extensions and features, then configure and deploy the entire map project to a UAT environment for testing. This process often leads to changes in the answers to the questions above (especially as it relates to things like workflow, terminology display, and search results).

Once a new mapping project gets going in PROD, that also often triggers changes in answers to the questions, once the live effects of "drip feed" are seen, and the map is being edited with real data in the context of the previously published version. There may be need for changes in relationships or advices, qa checks, reports, and other types of things.

There are a number of "project-specific" pages on the front page of the wiki that demonstrate some of the specific ways in which those projects are maintained. It is desirable to always create a new wiki page for a new project. Answers to the checklist above should be documented there, as well as frequently used admin commands to keep the project running smoothly.