

Jan 10, 2024 |

## 📅 Monarch prototype for Translator - discuss Richard's prototype

Attendees: Matt Brush Richard Bruskiewich Chris Mungall Sierra Moxon Justin Reese  
Justin Reese Kevin Schaper

Attached files: 📎 Prototype Presentation Template 📄 February 2024 Prototype Signup

### Notes

#### 1. Demo of the initial MVP...

- used the PLATER code base to (successfully) bootstrap this coding effort; some quirky plater code logic tweaked; removed the 'reasoner-transpiler' module references
- initial phenotype->disease use case is 'hard-coded' (would likely wish to generalize to other SemSimian search use cases later?)
- excluded the SRI Testing data metadata operation (feels a bit deprecated within the community)
- some PLATER inspired 'common' API endpoints - i.e. `get_node()` and `get_single_hops()` are not yet implemented. Do we need one or both of them?
- would implement the common API 'overlay' (workflow) operation?
- would we implement the common API 'simple\_spec' operation
- TODO: the Open Telemetry code needs some attention - the latest OT code modules appear to have removed the `"trace.get_tracer_provider().add_span_processor"` method

#### 2. How should a QueryGraph best represent a multi-CURIE query (the MVP assumed use of the 'ids' field with a list of HPO CURIEs to get the MVP implemented)

- (Vlado et al) UUID concept to aggregate sets of phenotypes in a specific profile
  - Q: How is this represented in the QueryGraph (or do we care?). Is the UUID idea mainly used to organize the Knowledge Graph and Answers (using 'support graphs and related EPC)

#### 3. Does there need to be a new specific 'multi-curie' workflow operation for this use case?

#### 4. How to interpret the 'is\_set' flag in the context of the multi-CURIE use case:

<https://github.com/NCATSTranslator/ReasonerAPI/blob/v1.4.2/TranslatorReasonerAPI.yaml#L836>

Is the initial TRAPI JSON output compliant with the spirit of this? How does one represent the n0 QNode mapping - with 'id' string a comma-delimited list of input ids(?) - relative to the Response results?

5. How should we map the SemSimian similarity scores onto the TRAPI output? From the prototype worksheet, is the proposed metadata about the validity/score of the match (e.g. - subject\_best\_matches  
- object\_best\_matches  
- average\_score  
- best\_score  
- metric)
6. What other (EPC) metadata do we need to extract out of the raw SemSimian output, to return in the TRAPI Response JSON?
7. Matt, should the 'knowledge level' (?) and "agent" attributes be added in (Tyler Beck call for same...)? Can I be pointed to the most recent decision doc for this?
8. Comments & feedback on the initial MVP?

#### Action items

- ☐ Identify or create a more specific predicate for the SemSimian output knowledge edges (for now use **biolink:similar\_to**)
- ☐ Use **infores:semsimian** as the primary\_knowledge\_source
  - ☐ RMB Q: how do we account for the original knowledge assertions, e.g. from HPO or MONDO or ???)
  - ☐ Is SemSimian more of a method of Monarch than its own knowledge source(?)
- ☐ Review and incorporate (Vlado et al) UUID identification of phenotypic profile
- ☐ Return “the top 5” most similar pairwise comparisons in the Aux Graph to limit the size of the auxgraph
- ☐ Fill out template for presentation:  
[https://docs.google.com/presentation/d/1my3VkT9A1ZgANMatHGCgjhH1kD0UeYwamDmvSz5M69Q/edit#slide=id.g2a216b5860c\\_0\\_20](https://docs.google.com/presentation/d/1my3VkT9A1ZgANMatHGCgjhH1kD0UeYwamDmvSz5M69Q/edit#slide=id.g2a216b5860c_0_20)
- ☐

#### Minutes:

<https://github.com/NCATSTranslator/ReasonerAPI/issues/373> – for description of the UUID generation for subject or object of an edge/association based on a set of terms that contribute to the UUID

And the predicate would be “similar to” (phenotype\_of) in the result set\_uuid\_of\_phenos similar\_to MONDO:id

Associated with is probably the parent though....