**DRAFT Translator Reasoning Tools API Output Format Specification**

DRAFT version 0.7.0 2018-05-15

Goal: Facilitate comparison and co-development of the Reasoning Tools by providing a suggested common API output format. Compliance to this standard is presumably not required, but it would make comparison of results a lot easier!

Comments:

- Target output format is JSON-LD

Related:

- See information about earlier versions of this specification at: <https://github.com/NCATS-Tangerine/NCATS-ReasonerStdAPI>

- Emerging KG Standard: [https://docs.google.com/document/d/1TrvqJPe\_HwmRJ5HCwZ7fsi9\_mwGcwLOZ53Fnjo8Sh4E/edit#](https://docs.google.com/document/d/1TrvqJPe_HwmRJ5HCwZ7fsi9_mwGcwLOZ53Fnjo8Sh4E/edit)

# Specification

### Top level (Response)

* **@context** - URI - URI of the JSON-LD context file of this document.
* **id** - URI - URI of this Reasoning Tools response if it is persisted somewhere.
* **type** - text - Type definition of this response object.
* **schema\_version** - text - The API standard will likely evolve over time. This encodes the schema version used in this response.
* **tool\_version** - text - The version string of the reasoning tool that provided this response.
* **datetime** - datetime - The datetime stamp when this response was provided to a user.
* **original\_question\_text** - text - The exact string that the original user provided to the reasoning tool
* **restated\_question\_text** - text - A restatement of the question that the reasoning tool understood and is answering with this response. This may not match the intent of the original\_question\_text.
* **query\_template** - object - Some reasoners may not work from an English text question, but may begin with a series of notes or node types. This section is intended to encode such a beginning. It is still not completely specified. There is some initial work on this in the QuerySpecification doc in the NCATS Hackathon 44/51 folder. To be fleshed out later with appropriate input from groups who want this functionality.
* **response\_code** - text - A terse code indicating success or error message for the query overall. OK is normal completion. Available error codes are not yet defined. These probably should be mapped to HTTP error codes in the YAML or entirely replaced by YAML-defined error codes.
* **message** - text - A detailed message from the Reasoning Tool to the user about degree of success of answering the query. If there are no results returned, then this message should detail why there are no results. If there are results returned, the Reasoning Tool may still provide some commentary to the user about how act of addressing the query result went. This is NOT intended to describe and answer/result, but rather just for the Reasoning Tool to provide information external to any specific result to the user.
* **result\_list** - array - A response from a tool may contain multiple results, where a result is an independent potential answer to the query.

### result (each object within result\_list)

* **id** - URI - URI of this specific result if it is persisted somewhere.
* **text** - text - A free text field describing this result (answer to the query).
* **confidence** - float - A numerical confidence score for this result, where 1.0 denotes the highest confidence and 0.0 denote no confidence.
* **result\_graph** - object - A serialization of the thought pattern or graph path for this result (answer to the query).

### result\_graph (a container for nodes and edges)

* **node\_list** - array - An array container for multiple node objects in arbitrary order
* **edge\_list** - array - an array container for multiple edge objects in arbitrary order

### node (each object within node\_list)

* **id** - text - CURIE corresponding to the bioentity
* **type** - text - bioentity type of the node, defined by the KG standard property “category”
* **name** - text - bioentity name of the node
* **uri** - URI - Full URI corresponding to the bioentity
* **description** - Full 1+ sentence description/definition of the bioentity
* **symbol** - text - Equivalent symbol for this bioentity. This is most common with the protein or gene bioentity types, but other types may also have symbols or abbreviations
* **node\_attributes** - array - container for a series of node\_attribute objects

### node\_attribute (each object within node\_attributes)

* **type** - text - controlled type of the property
* **name** - text - name of the node property
* **value** - any - value associate with the name and type
* **url** - URL - potential URL associated with this node property

### edge (each object within edge\_list)

* **type** - text - controlled edge type / predicate from the KG standard
* **source\_id** - text - id of the subject node
* **target\_id** - text - id of the object node
* **is\_defined\_by** - text - Name of the Translator group that instantiated the edge in the KG used to achieve this result
* **provided\_by** - text - Name of the original knowledge source that is credited with this edge/relationship/assertion that was inserted into the KG
* **confidence** - float - Confidence metric for this relationship/assertion/edge. 1.0 indicates the highest confidence. 0.0 indicates no confidence. The confidence may come directly from a knowledge source, or may come from come from the KG builder or even Reasoning Tool based on other contextual information. (NOTE: This is not in the KG standard, but is being proposed as an addition there)
* **edge\_attributes** - array - container for a series of edge\_attribute objects

### edge\_attribute (each object within edge\_attributes)

* **type** - text - controlled type of the property
* **name** - text - name of the edge property
* **value** - any - value associate with the name and type
* **url** - URL - URL associated with this edge property

# Example 1

{

"@context": "https://raw.githubusercontent.com/biolink/biolink-model/master/context.jsonld",

"id": "http://rtx.ncats.io/api/rtx/v1/response/1234",

"type": "medical\_translator\_query\_response",

"schema\_version": "0.7.0",

"tool\_version": "RTX 0.5.1",

"datetime": "2018-05-15 12:04:45",

"original\_question\_text": "what proteins are affected by sickle cell anemia",

"restated\_question\_text": "Which proteins are affected by sickle cell anemia?",

"query\_template": {

( Still to be fully defined. See QuerySpecification doc)

},

"response\_code": "OK",

"message": "1 result found",

"result\_list": [

{ "id": "http://rtx.ncats.io/api/rtx/v1/result/2345",

"text": "A free text description of this result",

"confidence": 0.932,

"result\_graph": {

"node\_list": [

{ "id": "OMIM:603903",

"category": "disease",

"name": "sickle cell anemia",

"uri": "http://omim.org/entry/603903",

"description": "A disease characterized by chronic hemolytic anemia...",

"symbol": "",

},

{ "id": "UniProt:P00738",

"category": "protein",

"name": "Haptoglobin",

"uri": "https://www.uniprot.org/uniprot/P00738",

"description": "Haptoglobin captures and combines with hemoglobin...",

"symbol": "HP",

"node\_attributes": [

{ "type": "n\_citations",

"name": "Number of citations",

"value": 329,

"url": "http://www.uniprot.org/uniprot/P00738/publications",

} ],

} ],

"edge\_list": [

{ "type": "affects",

"source\_id": "OMIM:603903",

"target\_id": "UniProt:P00738",

"is\_defined\_by": "RTX",

"provided\_by": "OMIM",

"confidence": 1.0,

"edge\_properties": [

{ "type": "PubMed\_article",

"name": "Orthopaedic Manifestations of Sickle Cell Disease.",

"value": "Pubmed:29309293",

"url": "https://www.ncbi.nlm.nih.gov/pubmed/29309293",

} ],

}

]

}

]

}

}

}