DISK: Hypothesis description and requirements

Authors: Daniel Garijo, Varun Ratnakar.

# **Summary**

After analyzing the slides, story and demo, we have collected a definition and set of requirements for representing lines of inquiry and hypothesis, their contents, confidence and provenance. This documents aims to describe both the definition and requirements

# **Definition**

A **line of inquiry** represents potential analyses that can be pursued to test a type of hypothesis. They are *containers* of:

· A **hypothesis pattern** (a type of **pattern**): statements with unbound variables that will be bound to a set of assertions.

· A **data pattern** (a type of **pattern**): statements with unbound variables that will be bound to a set of assertions. The results will be associated to the input of the workflows.

· Workflows: a definition of the set of steps (and their dependencies) that are necessary to execute a scientific experiment. A workflow is a particular **type of prov:Plan**.

· MetaWorkflows: a **type of workflow** that analyzes the outputs produced by one or more workflows.

A line of inquiry can be described with other metadata like creator, date, license, description, etc. A line of inquiry is triggered when a **hypothesis** matches against its hypothesis pattern and the data pattern is found among the knowledge base.

A **hypothesis** is composed of a set of statements that aim to find a relationships among different entities. For example, "a protein is expressed in a sample from a certain patient". A hypothesis can be decomposed in one or several **hypothesis** **assertions.** Each assertion has its **provenance** (i.e., it may have been created from a particular line of inquiry, or it may have been added to the hypothesis after a revision) and a **confidence value** (assigned by a metaworkflow). A hypothesis may also have a its confidence value (e.g., by averaging the total confidence value in their independent statements). A hypothesis assertion or a hypothesis itself may be supported by data or metaworkflows.

Binding value: string of text with a pattern value, data id or a variable from the patterns. Inferred of what they are by looking at text. Example: ?data2, ?data1, hg19.fa (data id from Wings).

Binding: a container that has something and a binding value.

**Variable Binding**: contains a Variable and binding value

**WorkflowBinding**: contains a workflow and variable binding

**MetaworkflowBinding**:

**TriggeredLineOfInquiry**:

**This report focuses on hypothesis, not on the contents modeled in the assertions of the hypothesis. Those will use other vocabularies.**

# **Requirements**

What kind of questions do we want our system to answer?

Disk Ontology: <http://disk-project.org/ontology/disk/disk-1.0.0.owl>

Omics Ontology: <http://disk-project.org/ontology/omics/omics-1.0.0.owl>

Hypothesis Ontology: <http://disk-project.org/ontology/hypothesis/hypothesis-1.0.0.owl>

|  |  |
| --- | --- |
| **Competency question** | **Terms introduced to address the question** |
| 1 What are the hypotheses listed in DISK? | disk:Hypothesis (named Graph)  <http://disk-project.org/disk-server/ravali/genomics/hypotheses> |
| 2 What are the contents of a particular line of inquiry? | disk:LineOfInquiry, hasWfBindings, hasMetawfBindings, hasDataQuery, hasHypothesisQuery  <http://disk-project.org/disk-server/ravali/genomics/lois/LOI-Qa6nuyLL8BBv> |
| 2a What is the hypothesis pattern of a particular line of inquiry? | disk:LineOfInquiry,hasHypothesisQuery |
| 2b What is the data pattern of a particular line of inquiry? | disk:LineOfInquiry,hasDataQuery, |
| 2c What are the workflows or metaworkflows contained in a line of inquiry? | disk:LineOfInquiry, hasWfBindings, hasWorkflow, hasMetawfBindings |
| 2d What are the bindings contained in a Line of Inquiry | disk:LineOfInquiry, hasWfBindings, hasMetawfBindings |
| 3 Which assertions are contained in a hypothesis? | disk:Hypothesis, (named Graph)  <http://disk-project.org/disk-server/ravali/genomics/hypotheses>/[name]  In “Hypothesis Index graph” :  <http://disk-project.org/disk-server/ravali/genomics/hypotheses>  [hypId] disk:hasParentHypothesis [parentHypId]  Example:  <http://disk-project.org/disk-server/ravali/genomics/hypotheses/Hypothesis-buksqjktAwoX>  Revised hypothesis Example:  <http://disk-project.org/disk-server/ravali/genomics/hypotheses/RevisedHypothesis-52czkpkn64ucsxuzvvty3sewr> |
| 4 What is the support for assertion X in hypothesis Y? | rdf:Statement, hasTriggeredLOI.  <http://disk-project.org/disk-server/ravali/genomics/tlois/TriggeredLOI-VLRmpJb1OBjh> |
| 5 What is the confidence of hypothesis H? | **disk:Hypothesis, hasConfidenceValue** |
| 6 What is the support for hypothesis X? | **disk:Hypothesis, hasSupport (currentlyhasTriggeredLOI)** |
| 7 What is the confidence for assertion X in Hypothesis H? | rdf:Statement, hasConfidenceValue |
| 8 List all the assertions that have a confidence value above a threshold (assuming proteins are part of the assertions in a hypothesis). | **rdf:Statment, hasConfidenceValue** |
| 9 Where does hypothesis H come from? (provenance) | **disk:Hypothesis, hasParentHypothesis (prov:wasDerivedFrom), hasSupport (prov:hasProvenance)** |
| 10 Is hypothesis H a revision of another hypothesis? | **disk:Hypothesis, hasParentHypothesis.** Example of a revised hypothesis:  <http://disk-project.org/disk-server/ravali/genomics/hypotheses/RevisedHypothesis-52czkpkn64ucsxuzvvty3sewr> |
| What is the name of a hypothesis? | **name** |
| What is the description of the hypothesis | **description** |
| How is the confidence value calculated? | **???** |
| Which workflows are associated to a metawf? | **Not needed** |
| What is the hypotheses resultant from a line of inquiry ? | **TriggeredLineOfInquiry, parentHypotheis, hasLOI, resultingHypotheiss** |
| What is the status of the run? |  |
| Where is the link to the run? |  |
| What are the confidence values with their respective triggered LOIs for a given statement or hypothesis? | ConfidenceReport, hasConfidenceValue, hasTriggeredLOI |
| Where does these data come from? | Data, createdFrom |
| Add reqs for hasRevisedHypothesisVariable and hasHypothesisVariable |  |