

HyQue: A Semantic Web tool for evaluating scientific hypotheses

Alison Callahan¹ & Michel Dumontier^{1,2,3}

¹Department of Biology, ²Institute of Biochemistry, ³School of Computer Science, Carleton University

```

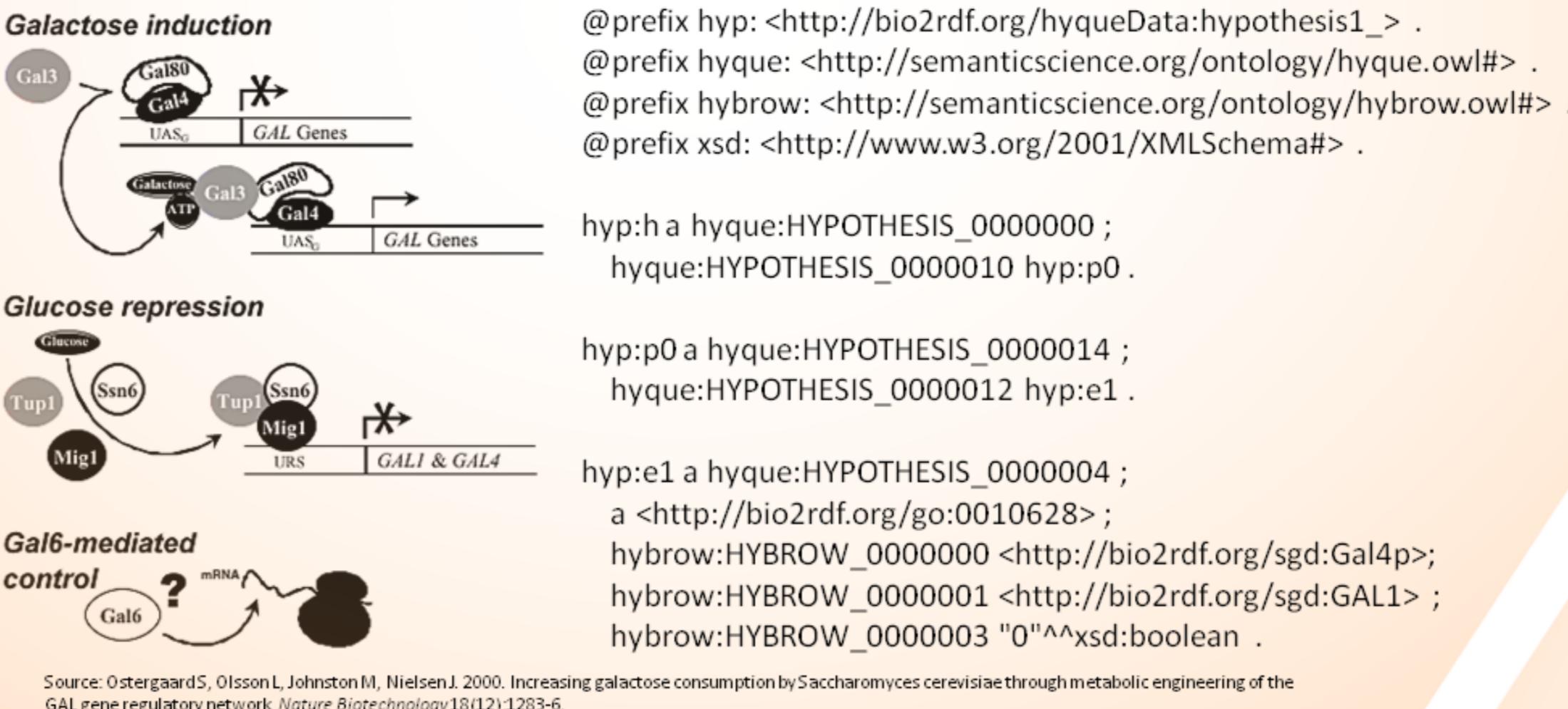
<spin:Function rdf:ID="calculateInduceAgentTypeScore">
<spin:body>
<sp:Select>
<sp:resultVariables rdf:parseType="Collection">
<rdf:Description rdf:nodeID="A126">
<sp:varName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
>score</sp:varName>
</rdf:Description>
<sp:where rdf:parseType="Collection">
<rdf:Description>
<sp:object>
<rdf:Description rdf:nodeID="A127">
<sp:varName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
>agent</sp:varName>
</rdf:Description>
<sp:predicate rdf:resource="http://semanticscience.org/ontology/hybrown.owl#HYBROW_00000000"/>
<sp:subject rdf:resource="http://pspnrdf.org/spin#_arg1"/>
</rdf:Description>
<sp:Bind>
<sp:expression>
<retrieveEntityTypeFromBio2RDFSGD>
<arg1 rdf:nodeID="A127"/>
</retrieveEntityTypeFromBio2RDFSGD>
</sp:expression>
<sp:variable>
<rdf:Description rdf:nodeID="A128">
<sp:varName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
>agent_type</sp:varName>
</rdf:Description>
</sp:variable>
</sp:Bind>

```

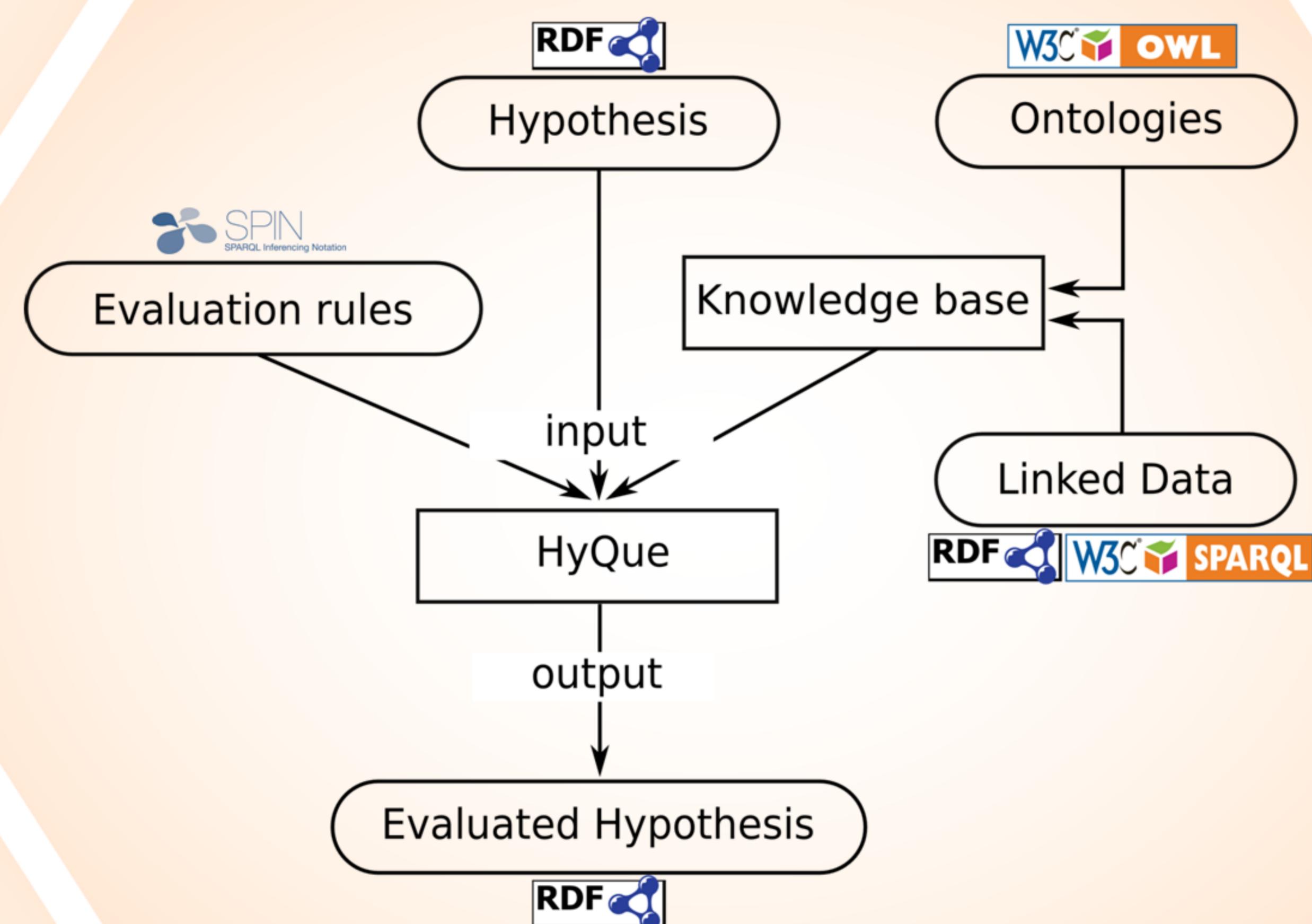
HyQue data retrieval and evaluation rules are written in the SPARQL Inferencing Notation (SPIN). SPIN rules can be executed using a reasoner, and also published as linked data and queried using SPARQL.

The screenshot shows the HyQue interface. At the top, there are links for Home, Learn more, Build & evaluate a hypothesis, Publications, Download, and Contact. Below this, the "Build a HyQue hypothesis" section has a "Hypothesis details" tab selected. It contains fields for Title (Galactose metabolism hypothesis), Description (Gal4p binds to and induces expression of GAL1 in the presence of galactose), Author (Alison Callahan), and Source (Ostergaard et al. 2000 [PMID: 11101808]). Under "Group events into propositions", there is a table with columns: Event identifier, Event label, Event type, Negated?, Agent, Target, Event location, Perturbation context. One row is shown: e1, Gal4p induces expression of GAL1, Induction, false, http://bio2rdf.org/sgd:Gal4p, http://bio2rdf.org/sgd:GAL1. Below the table are dropdowns for "Select operator" (AND) and "Group propositions". The bottom of the screen shows buttons for "<< Back" and "Confirm hypothesis details >>".

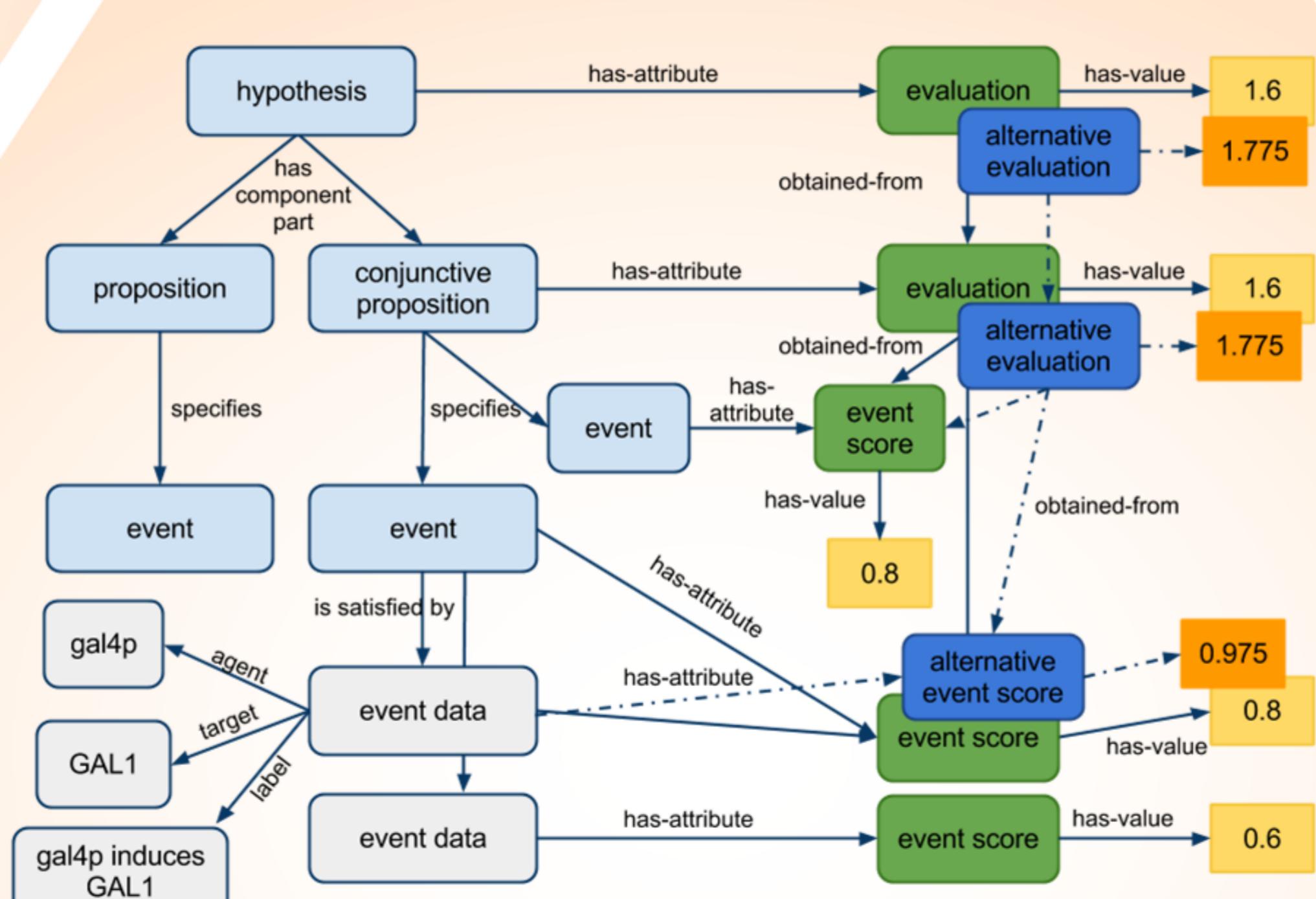
HyQue users can formulate hypotheses online and submit them for evaluation. If users are familiar with RDF they can also download the RDF serialization of their hypothesis and its evaluation.



HyQue can evaluate hypotheses about biological events. The hypothesis shown above is that the protein Gal4p induces expression of the gene GAL1.



HyQue uses SPIN rules to evaluate a hypothesis over Bio2RDF linked data and bio-ontologies. Hypotheses, data, ontologies, evaluation rules, and evaluation are all serialized using Semantic Web standards (RDF/OWL).

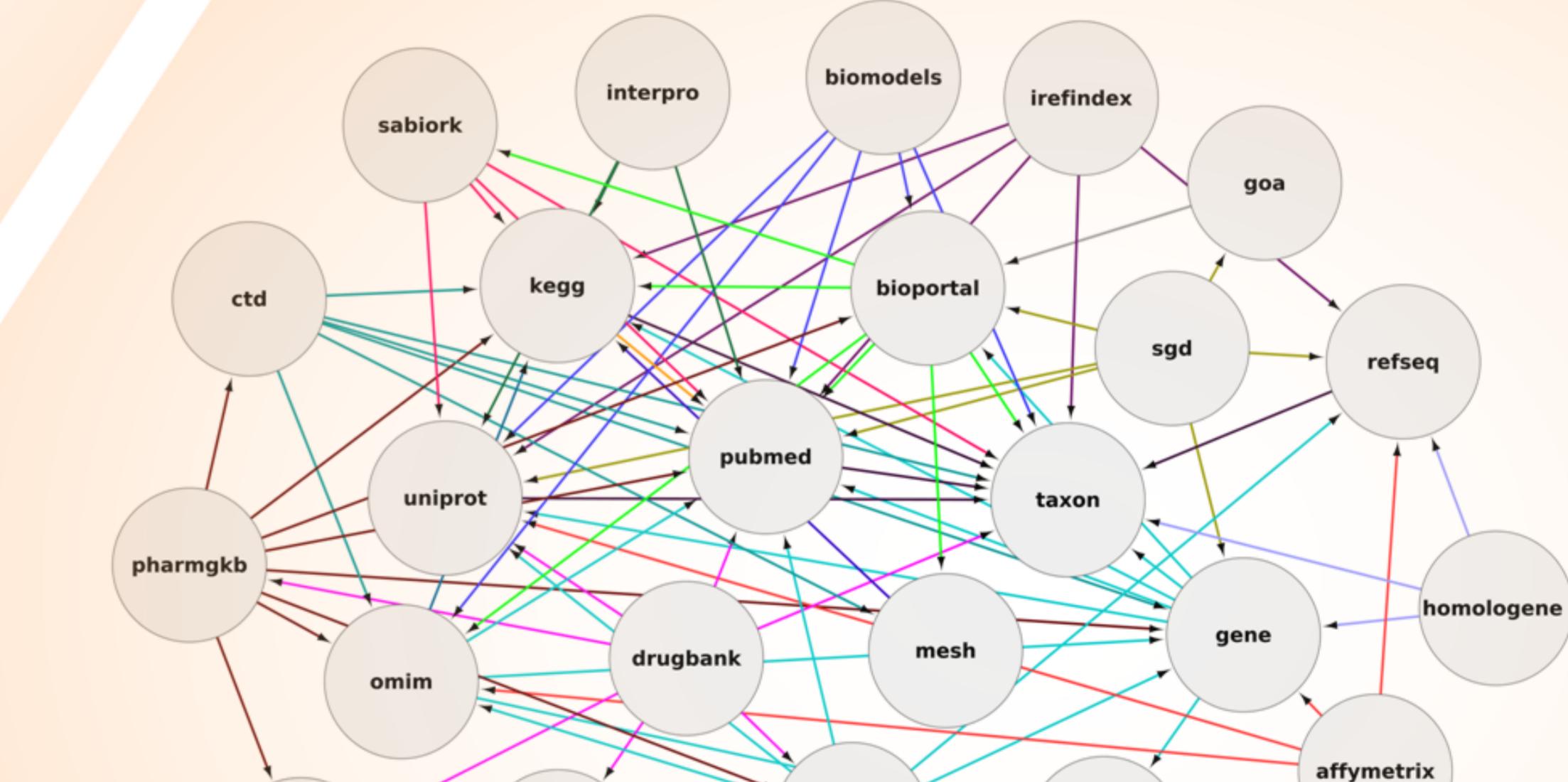


HyQue evaluations are automatically linked to both the input hypothesis and the data used to generate the evaluation. As a result it is possible to follow the path from a hypothesis to its supporting data and vice versa.

The screenshot shows the BioPortal interface. On the left, there is a "Class hierarchy" tree for "biochemical reaction". On the right, there is a "Browse" section with a table of ontologies and a "Search" section with a query for "biochemical reaction".

ONTOLOGY NAME	VISIBILITY	TERMS
BioAssay Ontology (BAO)	Public	1,298
BioModels Ontology (BioModels)	Public	187,519
BioPAX (BP)	Public	98
Breast Cancer Grading Ontology (BCGO)	Public	129
Cell Behavior Ontology (CBO)	Public	241
Cell Culture Ontology (CCONT)	Public	5,516
Cell Cycle Ontology (CCO)	Public	106,397
Cell Line Ontology (CLO)	Public	35,893
Enzyme Mechanism Ontology (EMO)	Public	259
Experimental Factor Ontology (EFO)	Public	13,289

HyQue uses bio-ontologies collected and maintained by the National Center for Biomedical Ontology (NCBO) to describe hypothesis details including event type, location and the biological function of event participants.



BIO2RDF

HyQue leverages Bio2RDF, a growing network of linked data for the life sciences, to gather data that supports or refutes a hypothesis.

