

Controlled vocabularies definition method
for bridging formal ontologies
development

Paolo Ciccarese, PhD

Mass General Hospital / Harvard Medical School

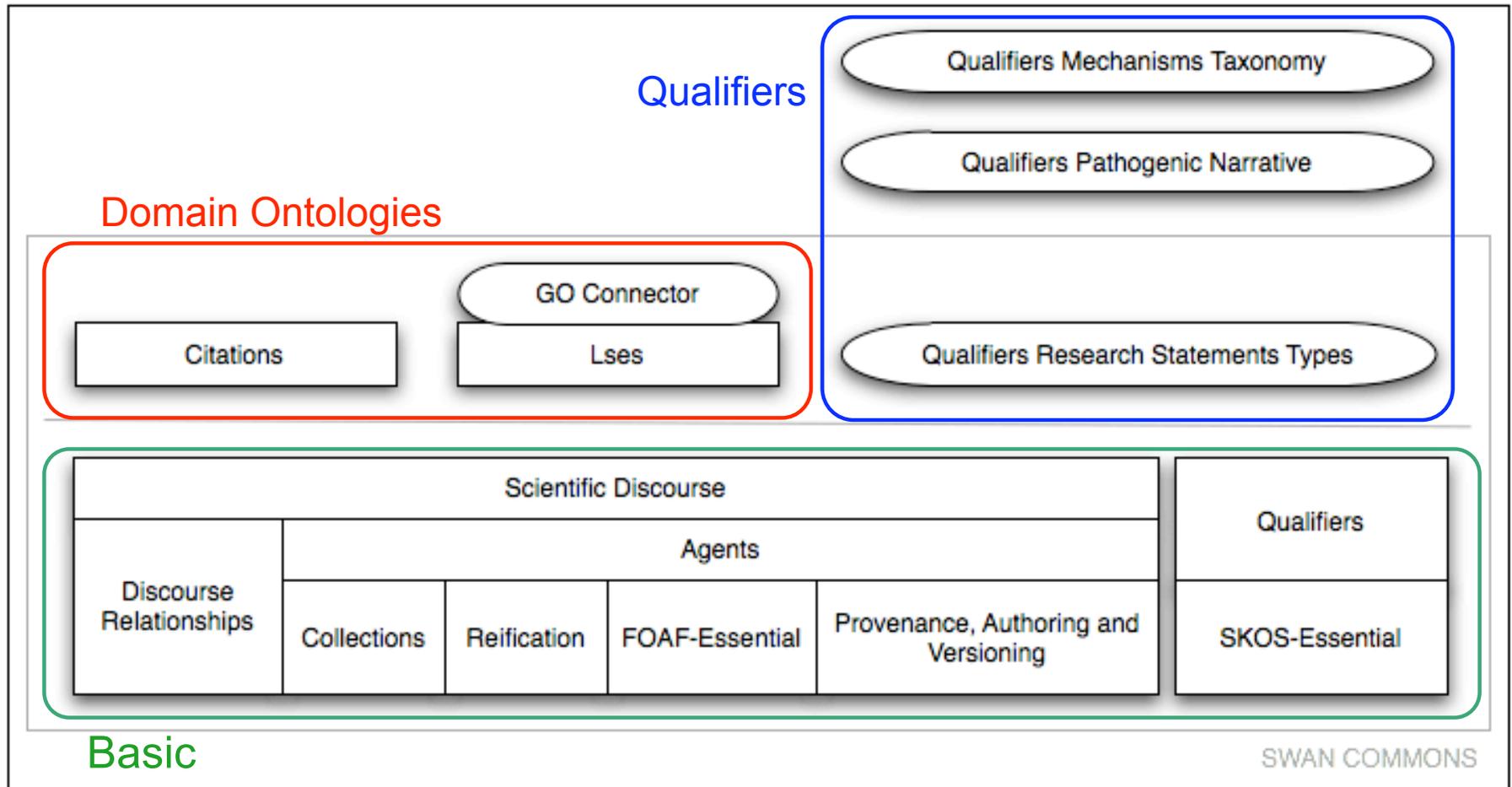
Background

time



- AlzForum <http://www.alzforum.org/>
- Semantic Web Applications in Neuromedicine
SWAN: <http://swan.mindinformatics.org/>
SWAN Alzheimer: <http://hypothesis.alzforum.org/>
- Science Collaboration Framework
SCF: <http://www.sciencecollaboration.org/>
StemBook: <http://www.stembook.org/>
- HCLSIG Scientific Discourse Task Force
<http://esw.w3.org/topic/HCLSIG/SWANSIOC>

SWAN Ontology Ecosystem



Basic

SWAN COMMONS

SWAN ALZHEIMER

An example of SWAN content curation

Seeding neuritic plaques from the distance: a possible role for brainstem neurons in the development of Alzheimer's disease pathology.
Muresan Z, Muresan V

Highly Purified Bovine Brain Phosphatase

at the EPANF website.¹² Unfortunately, our radioactive focusing antibodies did a poor job of resolving components of this high pI enzyme; most of the protein sample generated a broad band at the top of the gel.

Discussion

Molecular Heterogeneity. Single-molecule enzymology has been performed on a number of enzymes and in a number of laboratories. For each type of enzyme studied, there has been reported a wide range of behavior from molecule to molecule. This difference in behavior reflects differences in molecular structure. It has been proposed that the difference in behavior reflects transient structural differences, albeit structural differences that are stable on time periods of an hour.¹³ However, all earlier studies were performed with commercial enzymes that had undergone relatively crude purification.

The data in Figure 1 demonstrate that molecules of highly purified *L. collidolabris* phosphatase behave identically, at least when their activity is averaged over a 15-min incubation period. If there are any transient structural changes in the molecule, those changes are well averaged over the 15-min time period of our experiment.

Within experimental error, the three isoforms of this molecule have identical activity. These isoforms differ by a single amino acid deletion at the N-terminus, which is far removed

other hand, these molecule-to-molecule variations must be due to subtle interactions between the carbohydrate and the enzyme's active site.

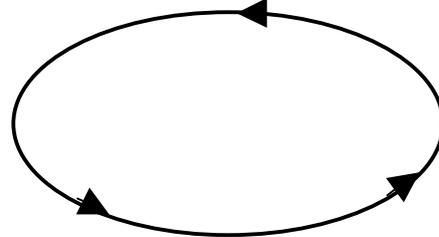
Dynamic Variation in Activity. In addition to these molecule-to-molecule variations in activity, both Xie and Rigler have described a short-time fluctuation in the activity of an individual enzyme molecule.^{14,15} These fluctuations arise from the molecule switching between active and inactive states with a rate of $\sim 1 \times 10^3 \text{ s}^{-1}$. This dynamic variation in activity cannot be explained by posttranslational modifications and instead must be due to short-time fluctuations in the molecule's structure.

In principle, our molecule-to-molecule measurement of enzyme activity can set limits on dynamic fluctuations in enzyme activity. Assuming the dynamic fluctuations are governed by Poisson statistics, the relative variance of the molecule-to-molecule enzyme activity will have a contribution that is equal to the inverse of the number of fluctuations that occur during the experiment. If there were any dynamic fluctuations in our experiment, there must have been more than ~ 30 fluctuations during the 30-min incubation; a smaller number of fluctuations would have made a detectable contribution to the molecule-to-molecule distribution in enzyme activity. If this enzyme fluctuates between active and inactive states, these fluctuations must occur with a rate greater than $\sim 0.02 \text{ s}^{-1}$.



Muresan, Zoia

Muresan, Virgil



SWAN Curator
 Gwen Wong, PhD

Comments
 Qualifiers

SWAN Browser

Brainstem neurons are initiators of neuritic plaques.

© Muresan Z, Muresan V

Description:
 This hypothesis provides a mechanism for triggering the formation of the neuritic plaques in the brains of Alzheimer Disease (AD) patients, aiming to explain why and how these plaques form preferentially in specific brain regions, such as the cerebral cortex and the hippocampus. According to this hypothesis, plaques are triggered by oligomeric A β seeds that form at the terminals of axons of brainstem neurons, which extend into lower prone regions, such as the hippocampus and the cerebral cortex. Brainstem neurons, rather than neurons in the cerebral cortex or hippocampus, are prone to accumulation of oligomerized A β at the terminals of their processes.

Authors: Muresan Z, Muresan V

Derived from:

Contains 28 Statements:

- Neuritic plaques contain at their core extracellular deposits of A β peptide. [View on SWAN Browser](#) [Supporting\(1\)](#) [View Related Statements](#) [Consistent\(1\)](#)
- The polymerization of soluble A β , which leads to plaque formation, is nucleated by "seeds" of oligomeric A β , but it remains unclear where seeds of polymerized A β originate or how they are formed. [View on SWAN Browser](#) [Supporting\(1\)](#) [View Related Statements](#) [Consistent\(1\)](#)
- Neuritic plaques are relevant to AD. [View on SWAN Browser](#) [Supporting\(1\)](#) [View Related Statements](#) [Consistent\(1\)](#)
- Plaques form preferentially in specific brain regions, such as the cerebral cortex and the hippocampus. [View on SWAN Browser](#) [Supporting\(1\)](#) [View Related Statements](#) [Consistent\(1\)](#)
- Cultured neuronal cells originating from the brainstem - but not neurons derived from the cerebral cortex or hippocampus - accumulate oligomerized A β at the terminals of their processes. [View on SWAN Browser](#) [Supporting\(2\)](#)
- SWAN only uses a useful model for studies of APP cell biology and AD pathology. [View on SWAN Browser](#)

Journal Article

SWAN Workbench v. 0.8

Search:

SWAN Workbench

Dear SWAN user,
 SWAN is a project to develop effective specialist knowledge bases for the Alzheimer Disease research community, using the energy and self-organization of that community enabled by Semantic Web technology.

Current build:

- Added description of inverse properties that claims supported by other discourse elements, related by other discourse elements, displayed by other discourse elements, etc.

Previous builds:

- Added citation of inverse properties that claims supported by other discourse elements, related by other discourse elements, displayed by other discourse elements, etc.
- The research statement page has been updated to include related by, Occurrence By, Supported By elements.
- The combined discourse elements page includes summary counts like: 2 Cite Evidence 3 Supported By Other Discourse Elements 2 Occurrence In Comments 1 Related by Other Discourse Elements
- Clicking on the line will highlight the corresponding discourse element.
- The additional links under each individual discourse element have been updated to include the inverse properties. Take a look at the labeling of the source included here. I am not totally comfortable with the syntax used.

Version: 0.8 (June 20070920 - September 30th, 2007), including:

- Supporting evidence in editors accepts all valid types
- Consistent of the editors with the same that includes and list name
- Check for errors in the editor
- Changes needed to match human format
- Private user is experimental only. Items for name
- More icons for different types of indications in search panel (still not enough)

Version: 0.8 (June 20070913 - August 15th, 2007), including:

- Conceptual search capabilities (now "header" tag in the left part of the workspace)
- Support for editing of Research Question
- Support for editing of Structural Comment
- Support for editing of Research Statement
- Support for editing of Research Comment
- Support for editing of Person

Version: 0.8 (June 20070809 - August 20th, 2007), including:



SWAN Workbench

Kim hypothesis

SIRT1 deacetylase protects against neurodegeneration in models for Alzheimer's disease and amyotrophic lateral sclerosis. 0 Comment(s)

 [Show graph](#)  [Submit a comment](#)

Description:

This hypothesis suggests that SIRT1 is a unique link between aging and human neurodegenerative diseases. SIRT1 expression confers significant neuroprotection, and resveratrol, a SIRT1 activating agent, promotes neuronal survival.

Authors: Kim D Nguyen M Dobbin M Fischer A Sananbenesi F Rodgers J Delalle I Baur J Sui G Armour S Puigserver P Sinclair D Tsai L

Derived from:



Tsai, Li-Huei

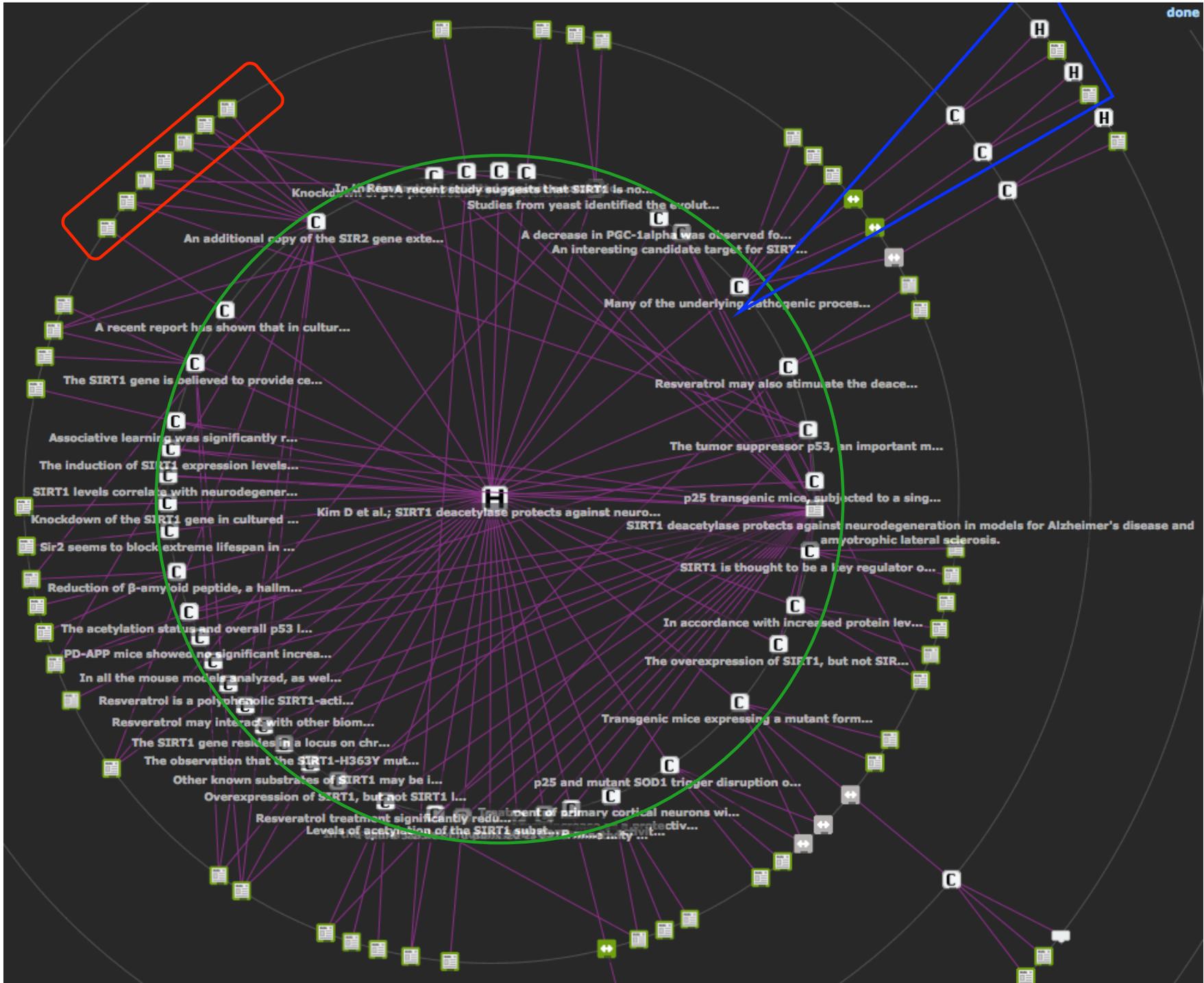
Kim D, Nguyen M, Dobbin M, Fischer A, Sananbenesi F, Rodgers J, Delalle I, Baur J, Sui G, Armour S, Puigserver P, Sinclair D, Tsai L
 **SIRT1 deacetylase protects against neurodegeneration in models for Alzheimer's disease and amyotrophic lateral sclerosis.**
The EMBO journal. 2007 Jul 11;26(13):3169-79

Contains 40 Statements:

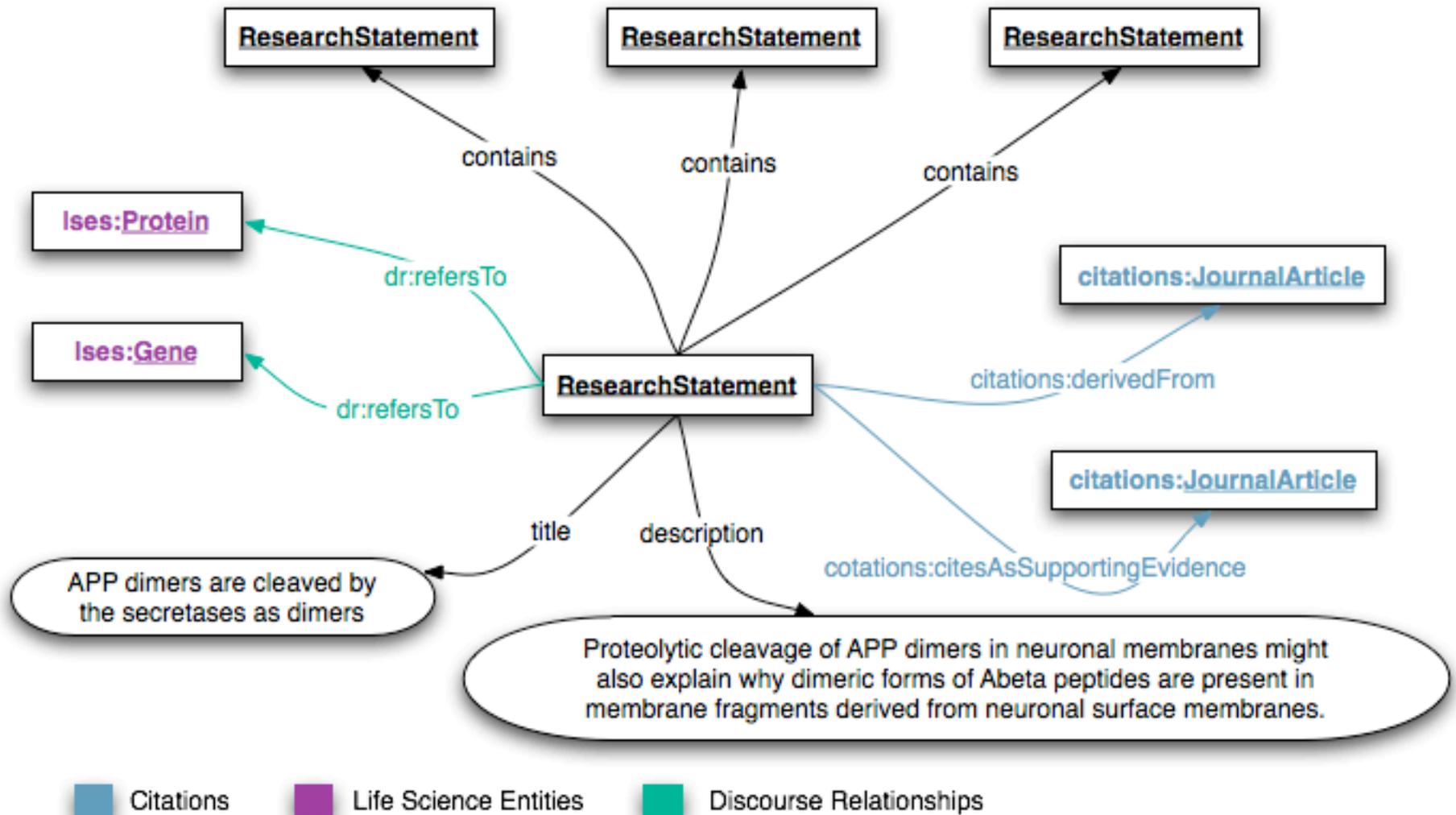
 40 with evidence  0 without evidence and a total of:  96 citations ; Related to external statements:  2 consistent  0 inconsistent  3 discussed  0 with alternatives

 Expand Info  Collapse Info

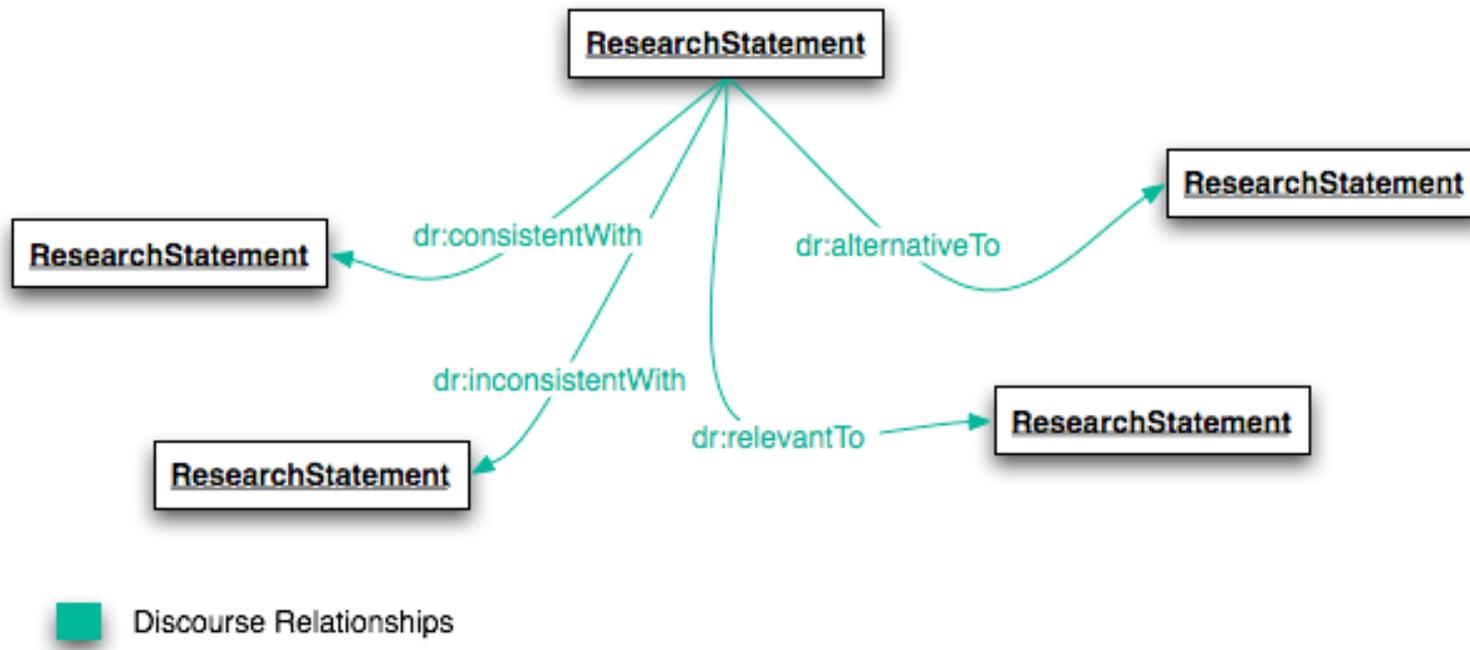
-  **Many of the underlying pathogenic processes of neurodegeneration are similar, including protein misfolding, oxidative stress, cytoskeletal abnormalities, disruption of calcium homeostasis, and inflammation, all of which increase during aging.**
[SHOW Info](#) [SHOW 3 Citations](#) :  Supporting(3) [SHOW Related statements](#) :  Consistent(2)  Discussing: (1)
-  **Studies from yeast identified the evolutionarily conserved NAD⁺-dependent deacetylase Sir2 as a critical regulator of the aging process.**
[SHOW Info](#) [SHOW 6 Citations](#) :  Supporting(6)
-  **An additional copy of the SIR2 gene extends lifespan in yeast and metazoans by a process seemingly analogous to caloric restriction, a diet that delays diseases of aging in mammals including neurodegeneration.**
[SHOW Info](#) [SHOW 12 Citations](#) :  Supporting(12)
-  **The SIRT1 gene is believed to provide cell protection during times of cell stress; however, the role of SIRT1 in vivo in age-dependent chronic neurodegenerative disorders remains undefined.**
[SHOW Info](#) [SHOW 5 Citations](#) :  Supporting(5)
-  **Knockdown of the SIRT1 gene in cultured mouse dorsal roots ganglion sensory neurons abrogates the protective effects of increased NAD⁺ synthesis on axonal degeneration**



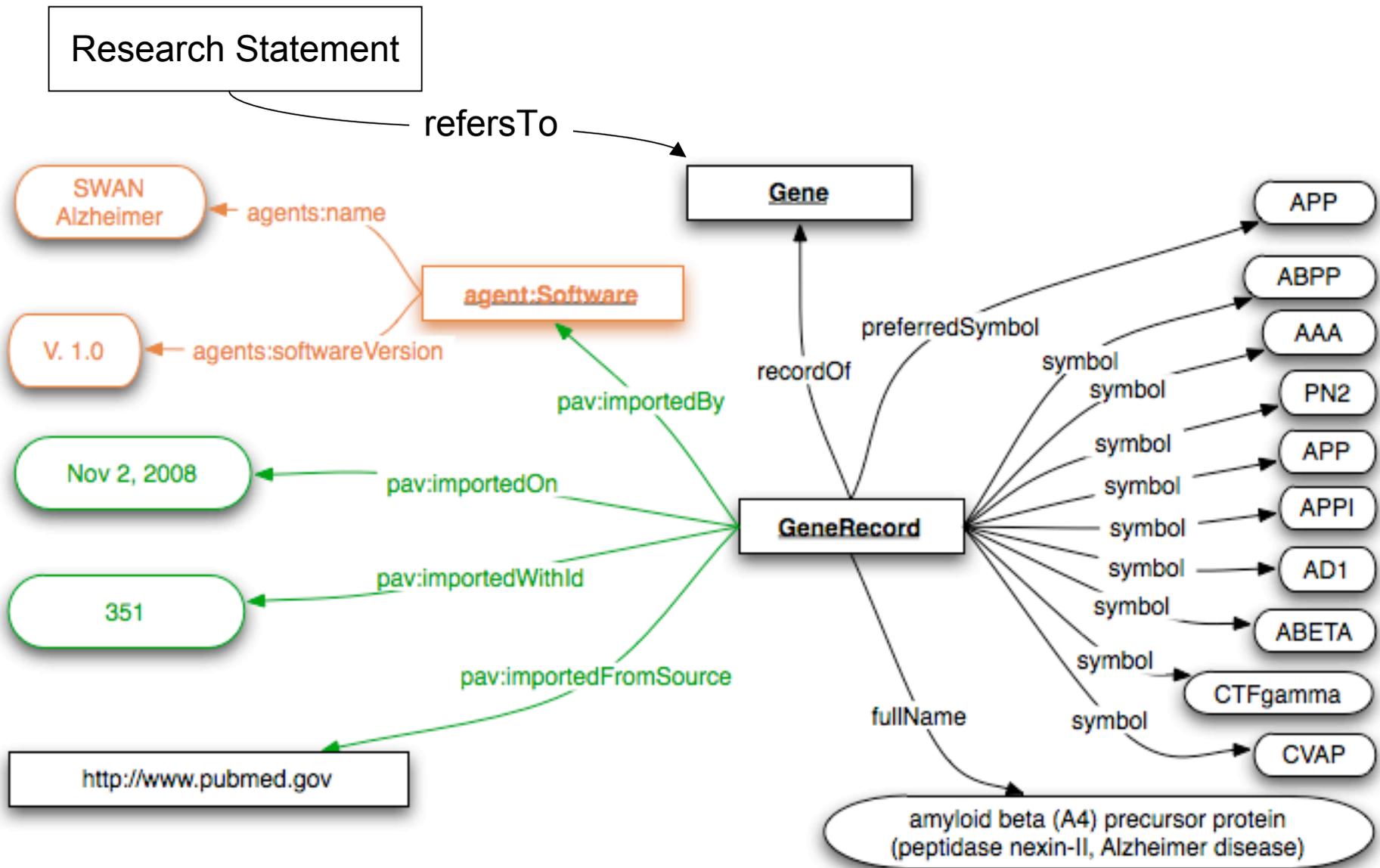
An example of Research Statement



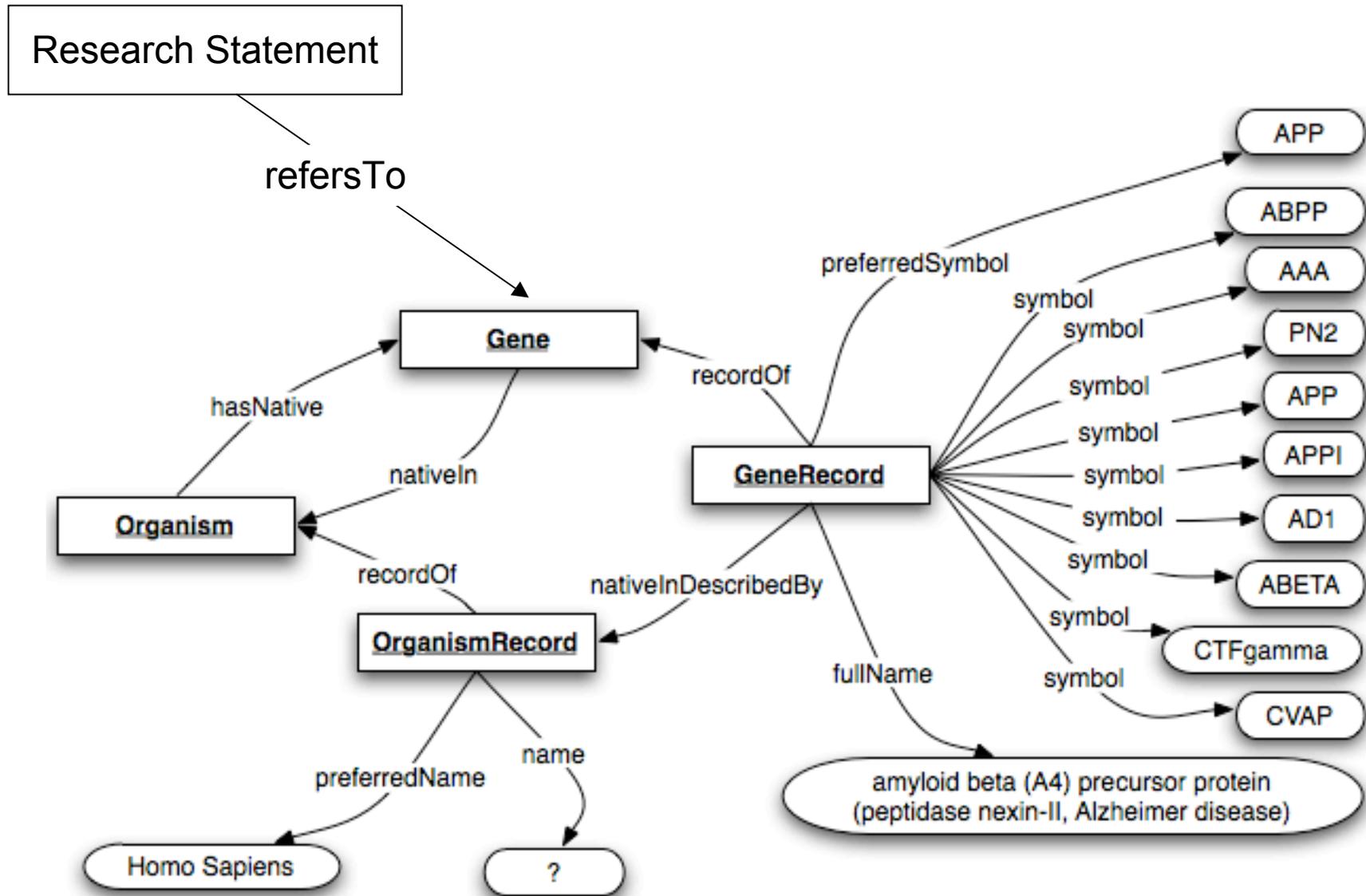
Relationships between discourse elements



Life Science Entities - 1



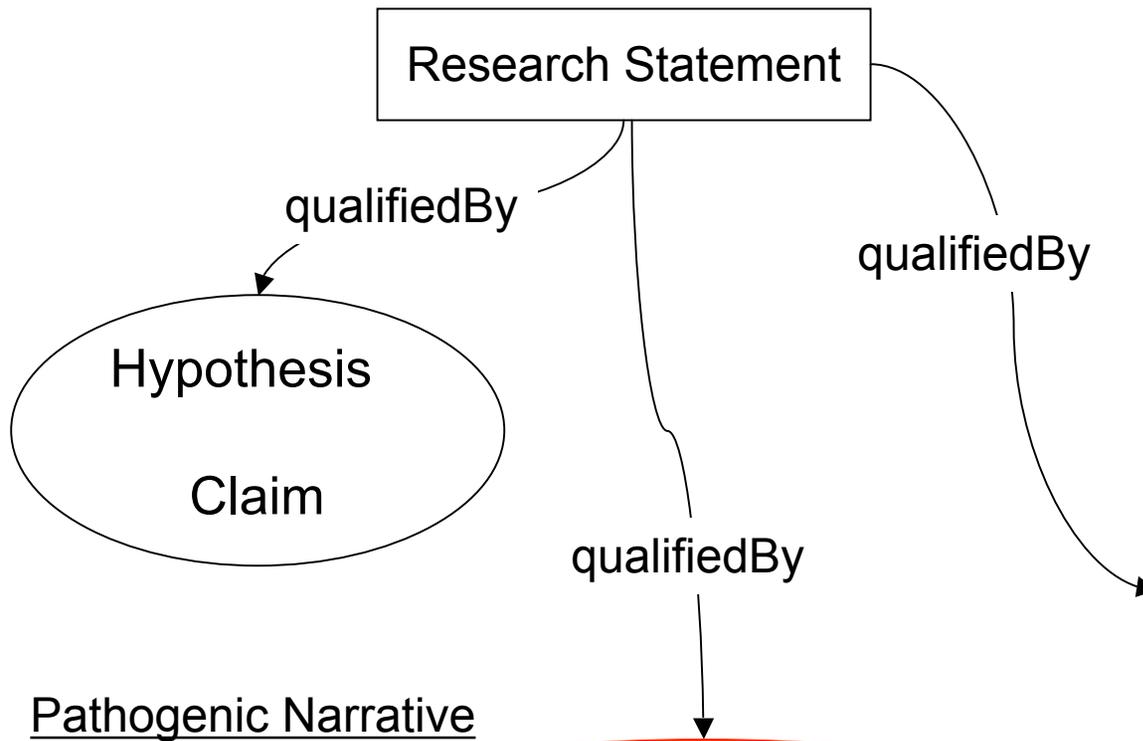
Life Science Entities - 2



But...

- Creating ontologies ~~takes~~ time long time
- Creating ontologies requires skilled knowledge engineers and skilled domain experts
- The time needed for developing ontologies is influencing the application development
- Is creating ontologies always the optimal solution?

SWAN Additional Annotation - 1



Mechanisms Taxonomy

- 21 Energetics ▾
 - 3 Circulation
 - 8 Metabolism
 - 10 Oxidation
- 32 Functional Changes of Proteins ▾
 - 11 ApoE
 - 1 Apoptosis
 - 4 Hormones
 - 6 Inflammation
 - 1 Protease Activation
 - 8 Trophic Factors
- 52 Structural Changes of Proteins ▾
 - 32 Amyloid ▾
 - 13 Protein Aggregation and Deposition
 - 5 Protein Degradation
 - 9 Protein Processing
 - 1 Protein Synthesis
 - 4 Cytoskeletal Proteins
 - 14 Synapses

Pathogenic Narrative

	Initial condition	Perturbation	Pathogenic event	Pathologic change
Alzheimer disease.	9	5	36	19
Calcium plays a key role in the process of synaptic	9	11	47	1
Amyloid plaques in a mouse model of Alzheimer's disease.	1	0	34	27

Other Additional Annotation - 2

Term	Type	What
AKT	kinase; intracellular signalling protein	A key kinase in the PI3-kinase signalling pathway. Plays pivotal roles in cell growth and survival. The pathway is activated in many cancers
APC1, APC2	intracellular signalling protein	Most studied in <i>Drosophila</i> , these tumour suppressors make gigantic protein complexes that inhibit Wnt signalling by binding to beta catenin
beta-catenin	intracellular signalling protein	A dual-purpose protein. In one role it forms protein complexes with cadherin at cell junctions. It can also act as a transcription factor: if a cell is not stimulated by Wnt, beta-catenin is degraded. In response to Wnt signalling, it enters the nucleus to spur transcription. Called <i>Armadillo</i> in flies
Bmi-1	Polycomb protein	Part of the PRC complex, Bmi-1 is a Polycomb protein that remodels chromatin so that transcription factors do not bind; it has been implicated in blood and neural cancers and stem cell maintenance
BMP/ BMPR	secreted signal protein	The 20 or so bone morphogenetic proteins interact with cell-surface receptors (BMPRs) and help set body plans in embryonic development. Among other functions, BMP4 regulates bone production from mesoderm
BrdU	marker	A modified nucleotide used as a DNA label cells incorporate the label when they make DNA
CBF1	intracellular signalling protein	A protein target of Notch that activates transcription; among other functions, it seems important to maintain neural stem cells
ChIP	technique	A way to know where proteins are binding on DNA; it stands for chromatin immunoprecipitation
C-kit, also KIT, CD117	cell-surface receptor	This cytokine receptor triggers hematopoietic stem cells to move into the bloodstream; when mutated, it can cause leukemia and other cancer
c-Myc/Myc	transcription factor	A Yamanaka factor (see later) that affects thousands of genes; when overactive, it leads to cancer
Cre/Lox	genetic tool	A tool for tissue-specific gene expression; Cre systems snip a desired gene (the floxed gene) from the genome; depending on the construct, this can be controlled by small molecules or another gene's expression

* Nature Reports - Stem Cells Cheat Sheet
Nature Publishing Group

Need for controlled vocabularies

- Using terms coming from controlled vocabularies is a good compromise in the case of high quality human curated scientific content
- Maybe we can use controlled vocabularies as an easy and incremental method to get to formal ontologies

Requirements

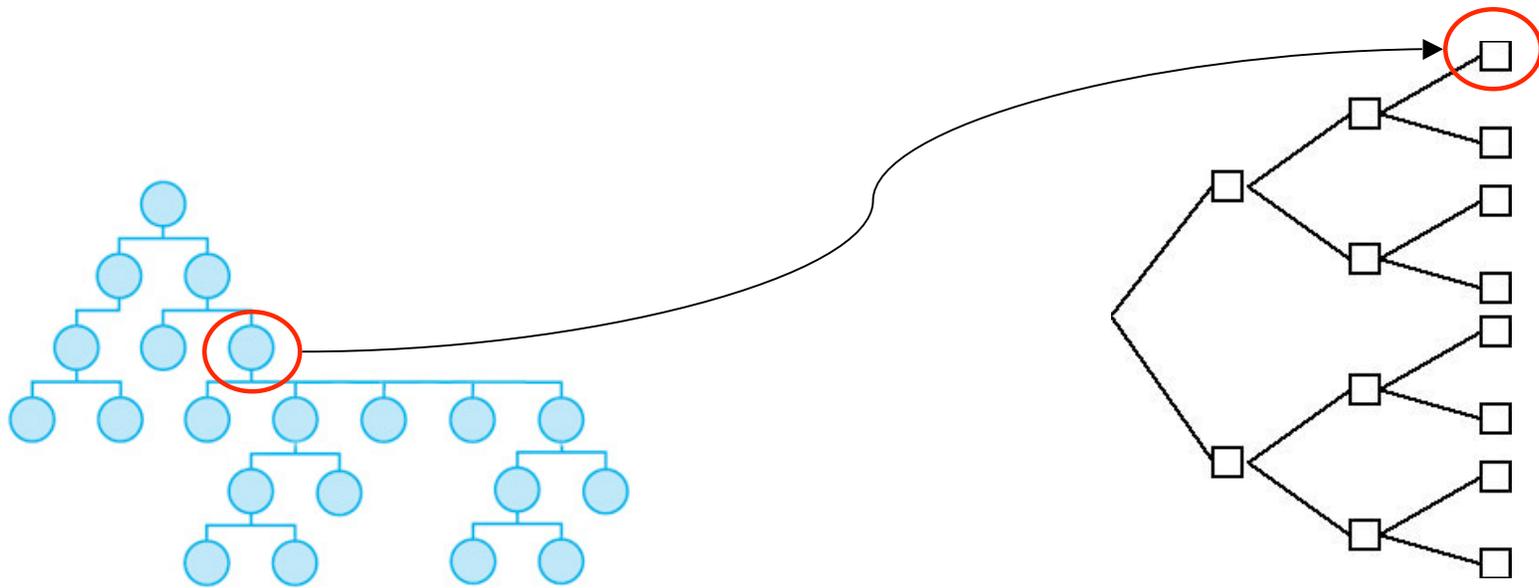
- Controlled vocabularies
- Controlled vocabularies should be easy and fast to be defined by “not ontologists”
- Link the terms to existing (or future) ontologies and keep track of provenance/authoring
- OWL-DL but without impacting the reasoning
- Share the vocabularies in RDF format

Agile definition and mapping

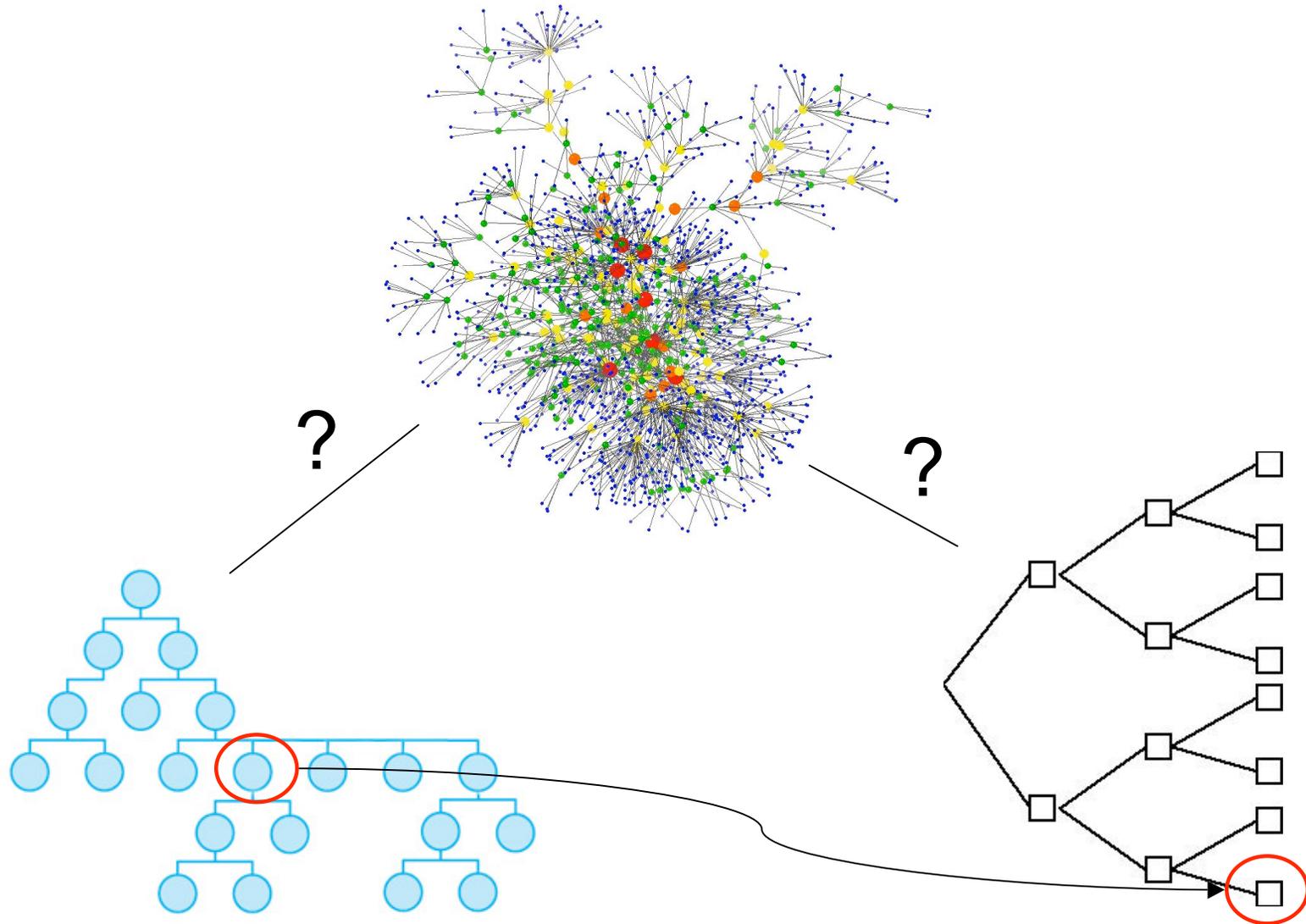
- We don't want users "not ontologists" to deal with classes, properties and restrictions
- Taxonomies are often enough for simple classifications and can be defined easily (terms + definitions)
- We want trained ontologists to figure out how to map the terms to existing ontologies (if we have resources to do so)

Possible solution: SKOS

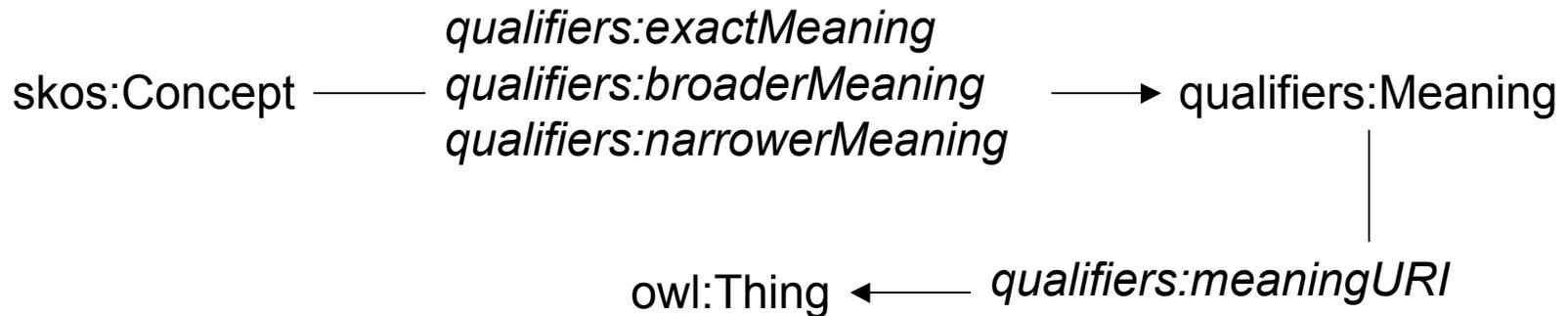
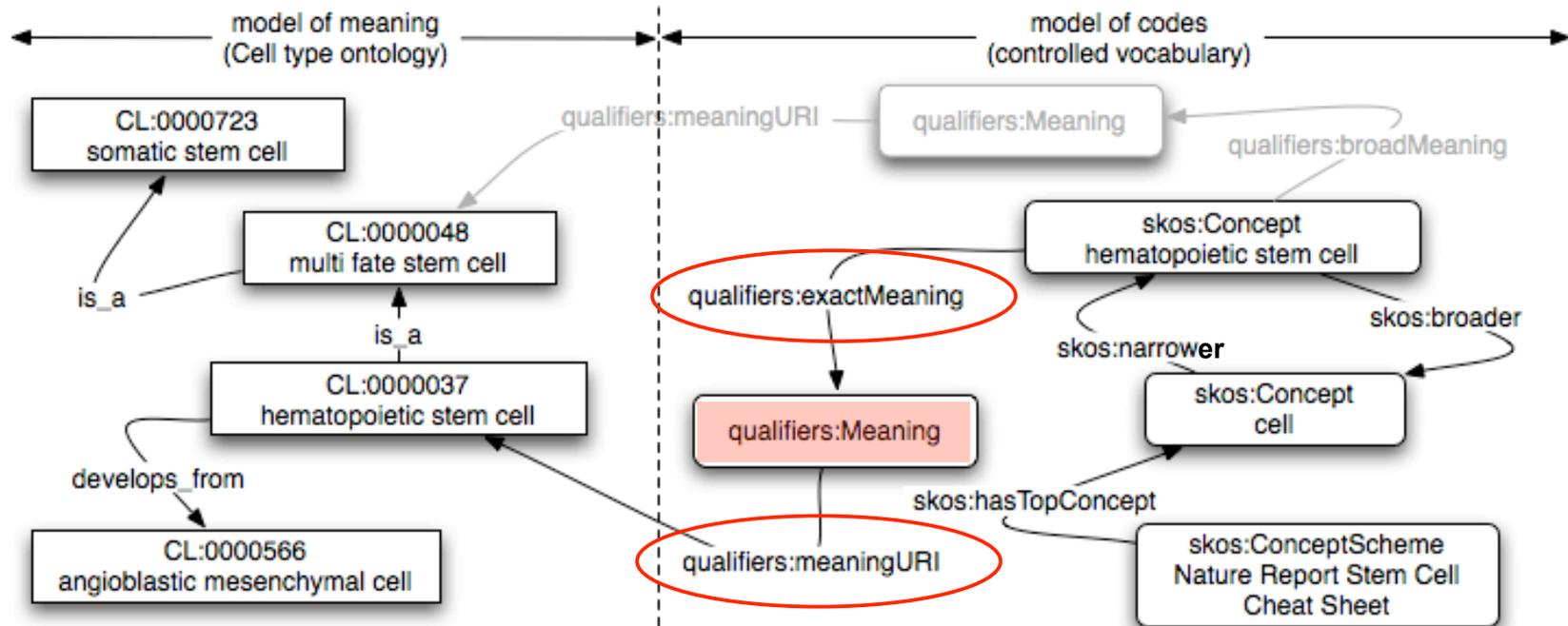
The Simple Knowledge Organization System provides a model for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, folksonomies, and other similar types of controlled vocabulary.



How to refer to formal ontologies?



An example



Mapping to formal ontology

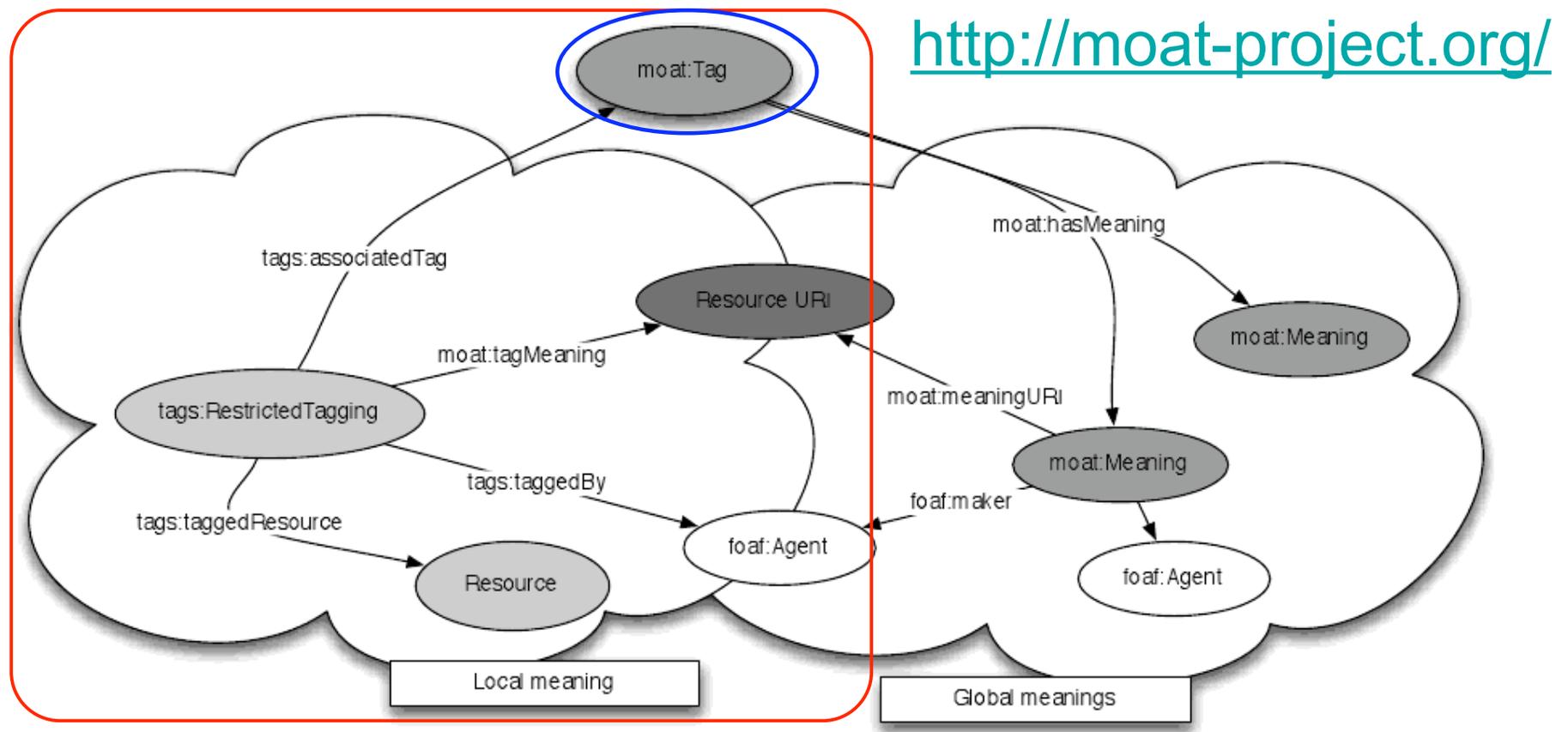
```
<skos:Concept rdf:about="&stemcellcheatsheet;hsc">
  <skos:prefLabel>HSC</skos:prefLabel>
  <skos:definition>Haematopoietic stem cells (blood-forming stem cells that reside in bone
    marrow)</skos:definition>
  <skos:broader rdf:resource="&stemcellcheatsheet;cell"/>
  <!-- Mapping -->
  <qualifiers:exactMeaning>
    <qualifiers:Meaning>
      <qualifiers:meaningURI rdf:resource="http://purl.org/obo/owl/CL#CL_0000037"/> <!-- hema
      <dcterms:publisher rdf:resource="http://swan.mindinformatics.org"/>
      <dcterms:created rdf:datatype="&xsd;dateTime">2009-01-28T00:00:00+05:00</dcterms:cr
      <dcterms:issued>2009-01-29T10:00:00+05:00</dcterms:issued>
      <pav:curatedBy rdf:resource="http://swan.mindinformatics.org/people/tim-clark"/>
      <dcterms:creator>
        <foaf:Person rdf:about="http://www.hcklab.org/people/pc/">
          <foaf:name>Paolo Ciccarese</foaf:name>
        </foaf:Person>
      </dcterms:creator>
    </qualifiers:Meaning>
  </qualifiers:exactMeaning>
</skos:Concept>
```

Meaning Of A Tag (MOAT) - 1

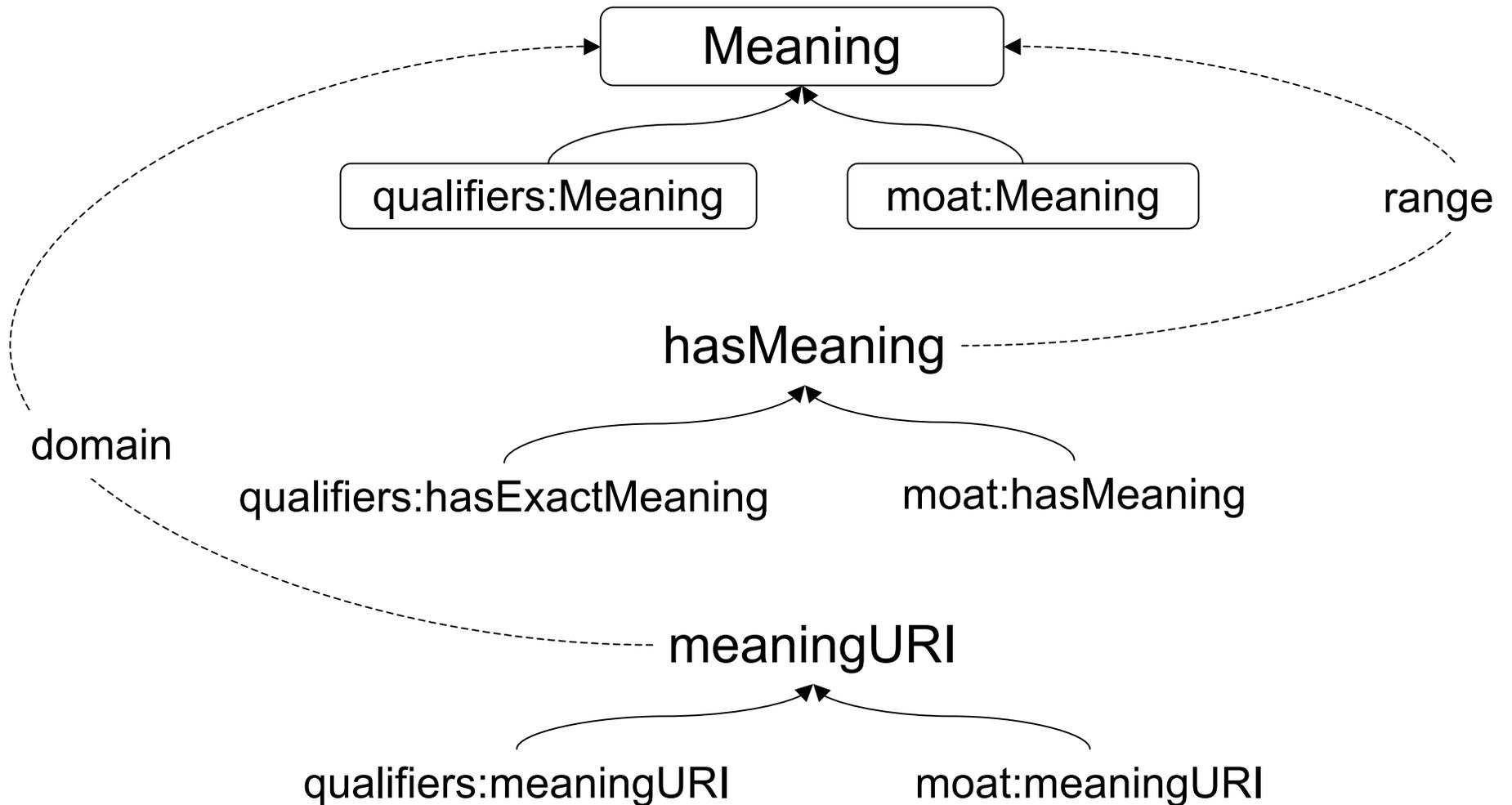
```
<moat:Tag rdf:about="http://tags.moat-project.org/tag/apple">  
  <moat:name><![CDATA[apple]]></moat:name>  
  <moat:hasMeaning>  
    <moat:Meaning>  
      <moat:meaningURI rdf:resource="http://dbpedia.org/resource/Apple_Records"/>  
      <foaf:maker rdf:resource="http://apassant.net/alex"/>  
      <foaf:maker rdf:resource="http://example.org/user/foaf/1"/>  
    </moat:Meaning>  
  </moat:hasMeaning>  
  <moat:hasMeaning>  
    <moat:Meaning>  
      <moat:meaningURI rdf:resource="http://dbpedia.org/resource/Apple"/>  
      <foaf:maker rdf:resource="http://example.org/user/foaf/1"/>  
    </moat:Meaning>  
  </moat:hasMeaning>  
  <moat:hasMeaning>  
    <moat:Meaning>  
      <moat:meaningURI rdf:resource="http://dbpedia.org/resource/Apple_Inc."/>  
      <foaf:maker rdf:resource="http://apassant.net/alex"/>  
    </moat:Meaning>  
  </moat:hasMeaning>  
</moat:Tag>
```

Meaning Of A Tag (MOAT) - 2

```
<tag:RestrictedTagging>  
  <tag:taggedResource rdf:resource="http://example.org/post/1"/>  
  <foaf:maker rdf:resource="http://apassant.net/alex"/>  
  <tag:associatedTag rdf:resource="http://tags.moat-project.org/tag/apple"/>  
  <moat:tagMeaning rdf:resource="http://dbpedia.org/resource/Apple_Records"/>  
</tag:RestrictedTagging>
```



Alignment with MOAT



Conclusions

- ⊕ Easy to implement (also at application level)
- ⊕ OWL-DL (with OWL2 - punning)
- ⊕ Provenance
- ⊕ Mapping to existing ontologies
- ⊕ Doesn't change reasoning (unless we post-process the annotation)
- ⊕ Aligned with MOAT
- ⊖ Introduces a level of complexity (but we can always post-process the annotation)