Controlled vocabularies definition method for bridging formal ontologies development

Paolo Ciccarese, PhD

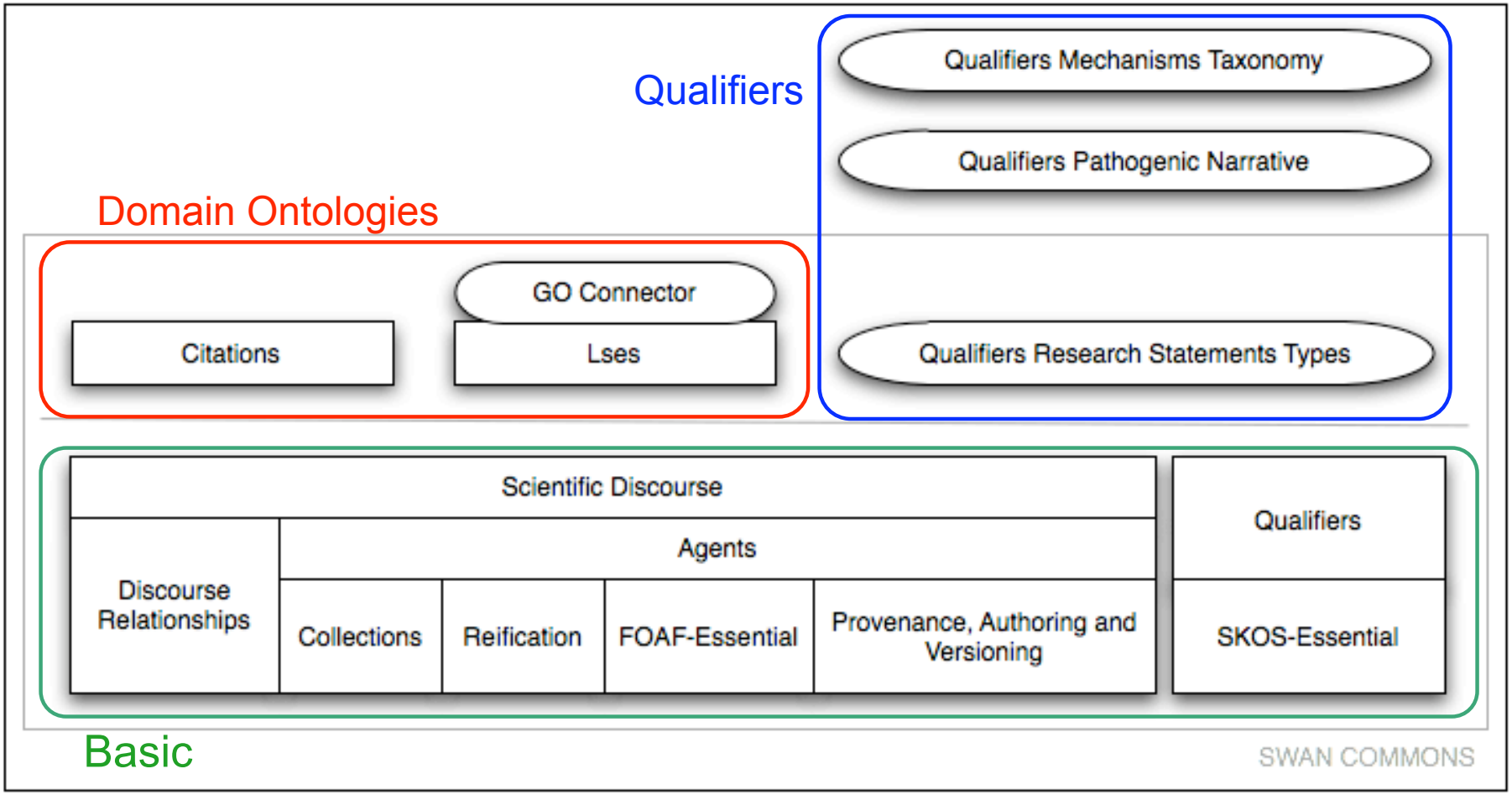
Mass General Hospital / Harvard Medical School

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

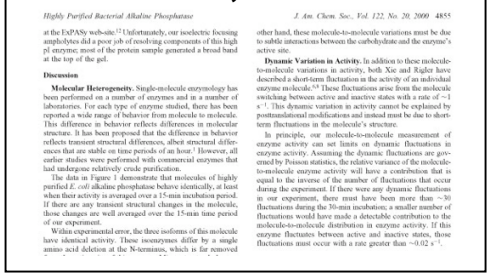


SWAN ALZHEIMER

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



Journal Article

Muresan, Zoia

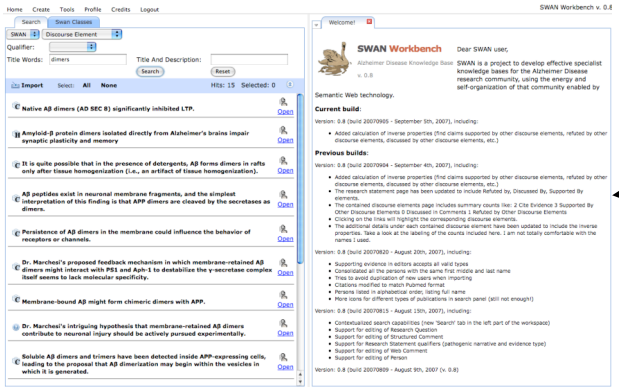
Muresan, Virgil

Comments
Qualifiers

SWAN Curator
Gwen Wong, PhD

SWAN Browser

Triple store



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

1. 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)

2. 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)

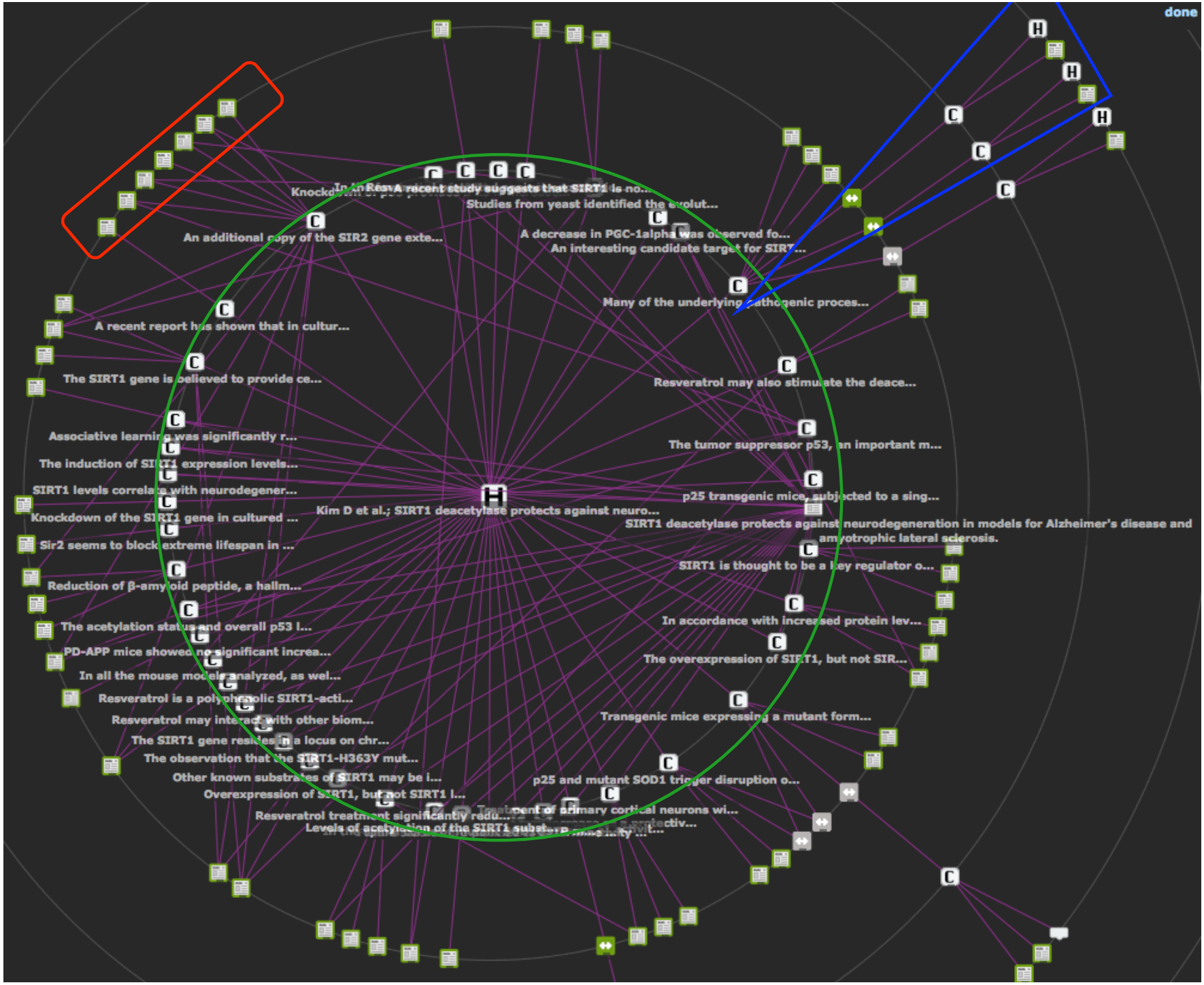
3. An additional copy of the SIR2 gene extends lifespan in yeast and metazoans by a process seemingly analogous to caloric restriction, a diet that delays diseases of aging in mammals including neurodegeneration.

   SHOW Info SHOW 12 Citations : Supporting(12)

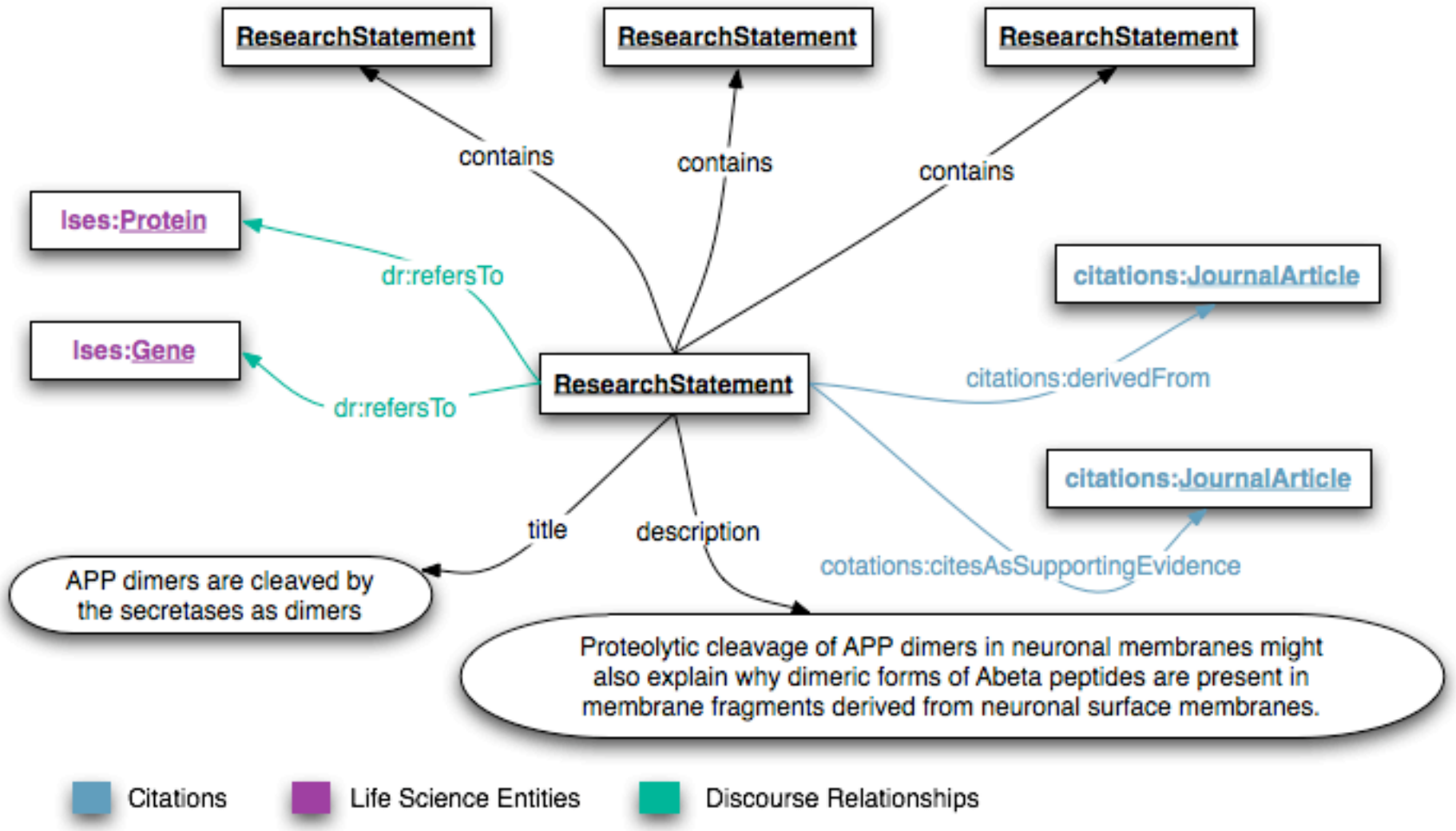
4. The SIRT1 gene is believed to provide cell protection during times of cell stress; however, the role of SIRT1 in vivo in age-dependent chronic neurodegenerative disorders remains undefined.

   SHOW Info SHOW 5 Citations : Supporting(5)

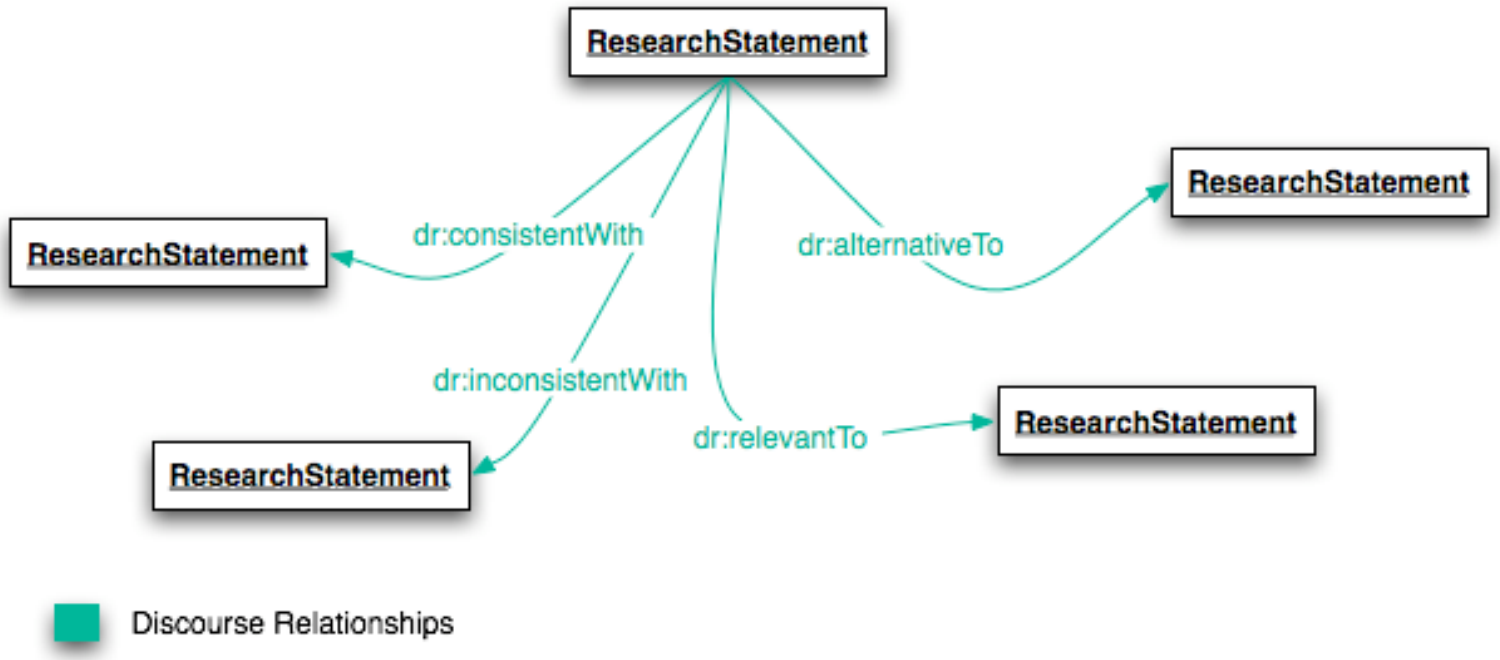
5. Knockdown of the SIRT1 gene in cultured mouse dorsal roots ganglion sensory neurons abrogates the protective effects of increased NAD+ synthesis on axonal degeneration


done
Kim D et al.; SIRT1 deacetylase protects against neuro...
SIRT1 deacetylase protects against neurodegeneration in models for Alzheimer's disease and amyotrophic lateral sclerosis.
Studies from yeast identified the evolut...
An additional copy of the SIR2 gene exte...
A decrease in PGC-1alpha was observed fo...
An interesting candidate target for SIRT...
Many of the underlying pathogenic proces...
A recent report has shown that in cultur...
The SIRT1 gene is believed to provide ce...
Resveratrol may also stimulate the deace...
Associative learning was significantly r...
The tumor suppressor p53, an important m...
The induction of SIRT1 expression levels...
SIRT1 levels correlate with neurodegener...
p25 transgenic mice, subjected to a sing...
Knockdown of the SIRT1 gene in cultured ...
Sir2 seems to block extreme lifespan in ...
SIRT1 is thought to be a key regulator o...
Reduction of β-amyloid peptide, a hallm...
In accordance with increased protein lev...
The acetylation status and overall p53 l...
PD-APP mice showed no significant increa...
The overexpression of SIRT1, but not SIR...
In all the mouse models analyzed, as wel...
Resveratrol is a polyphenolic SIRT1-acti...
Transgenic mice expressing a mutant form...
Resveratrol may interact with other biom...
The SIRT1 gene resides in a locus on chr...
The observation that the SIRT1-H363Y mut...
Other known substrates of SIRT1 may be i...
p25 and mutant SOD1 trigger disruption o...
Overexpression of SIRT1, but not SIRT1 l...
Resveratrol treatment significantly redu...
Levels of acetylation of the SIRT1 subst...

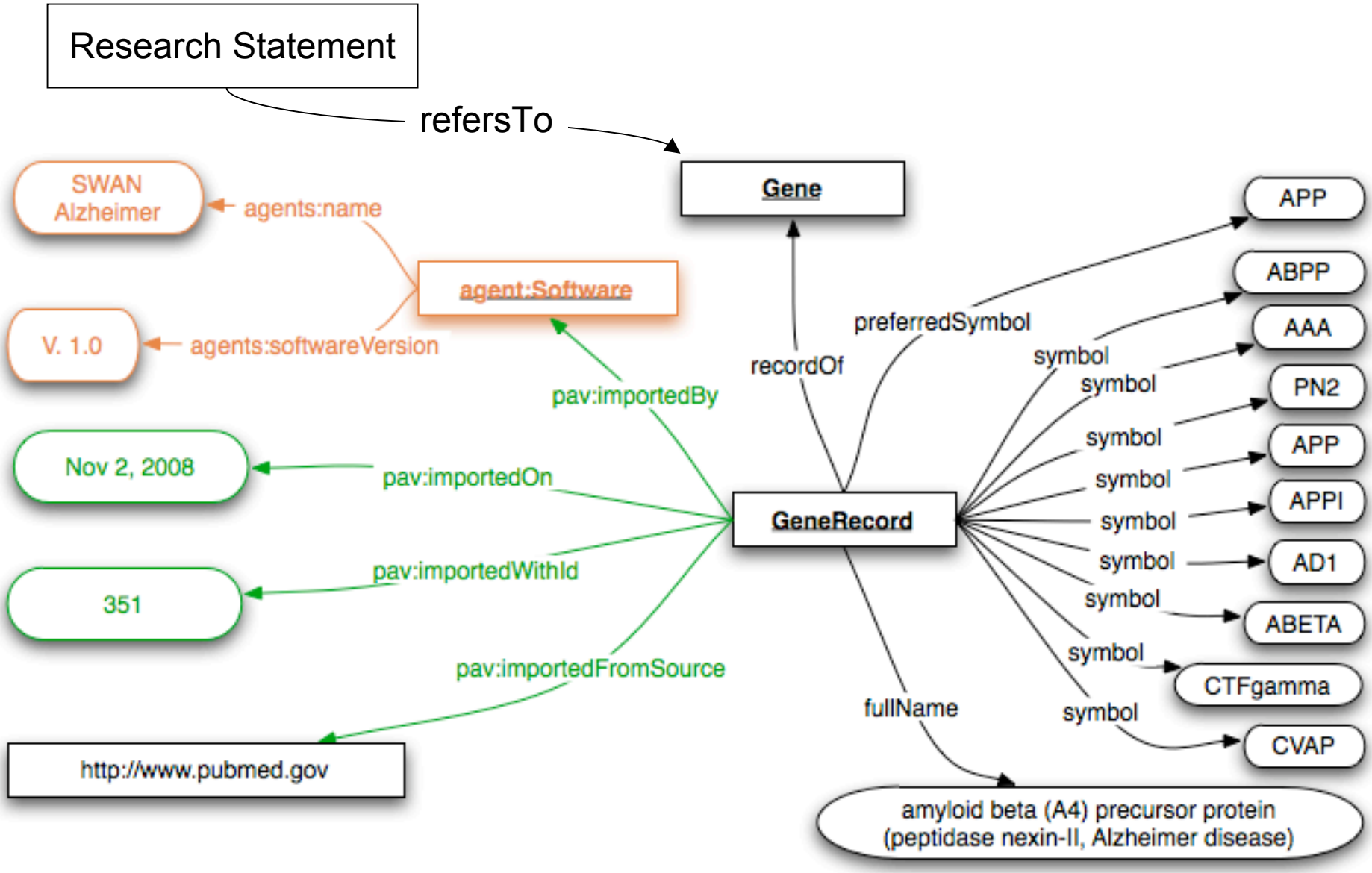
An example of Research Statement



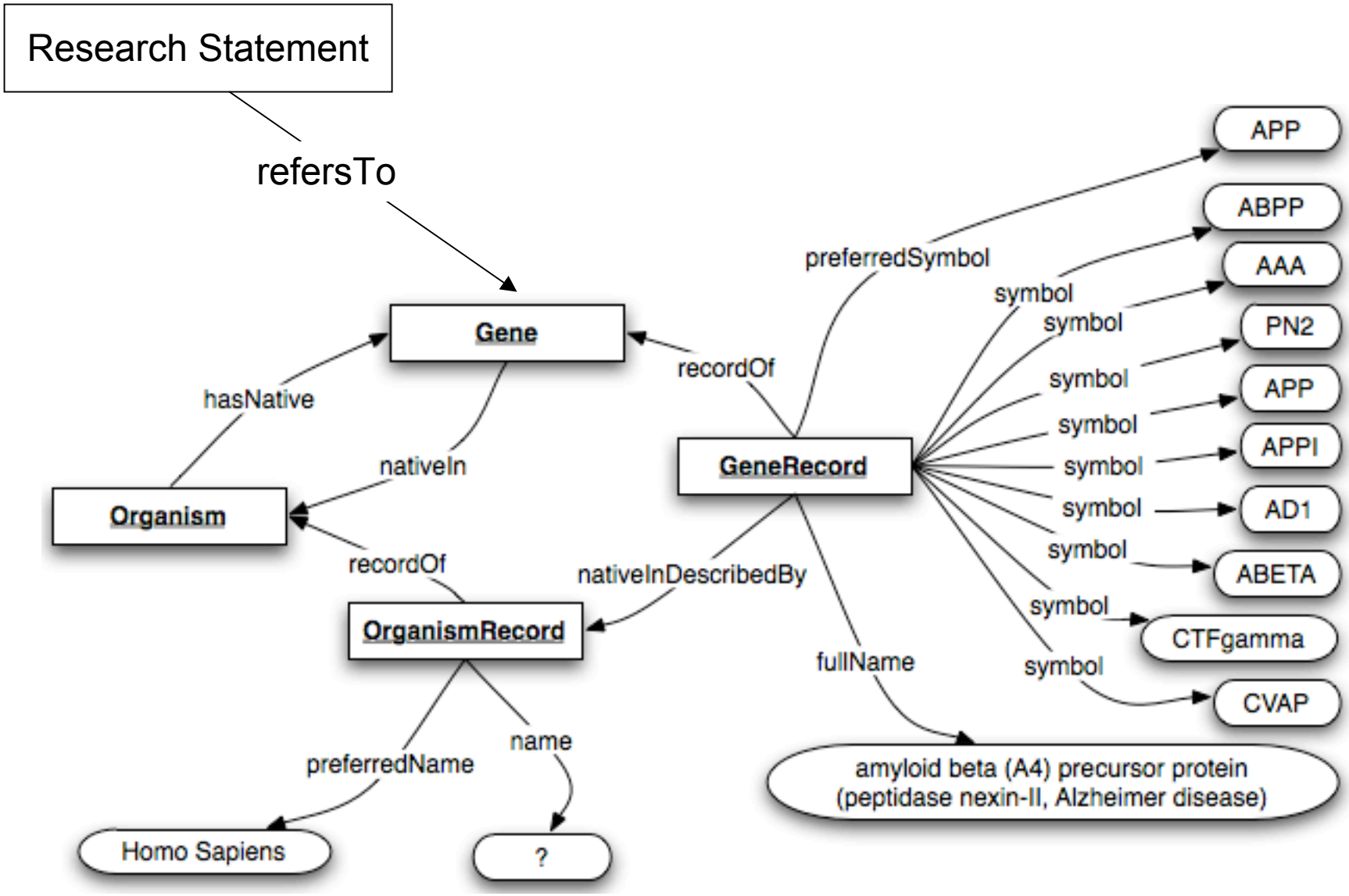
Relationships between discourse elements



Life Science Entities - 1



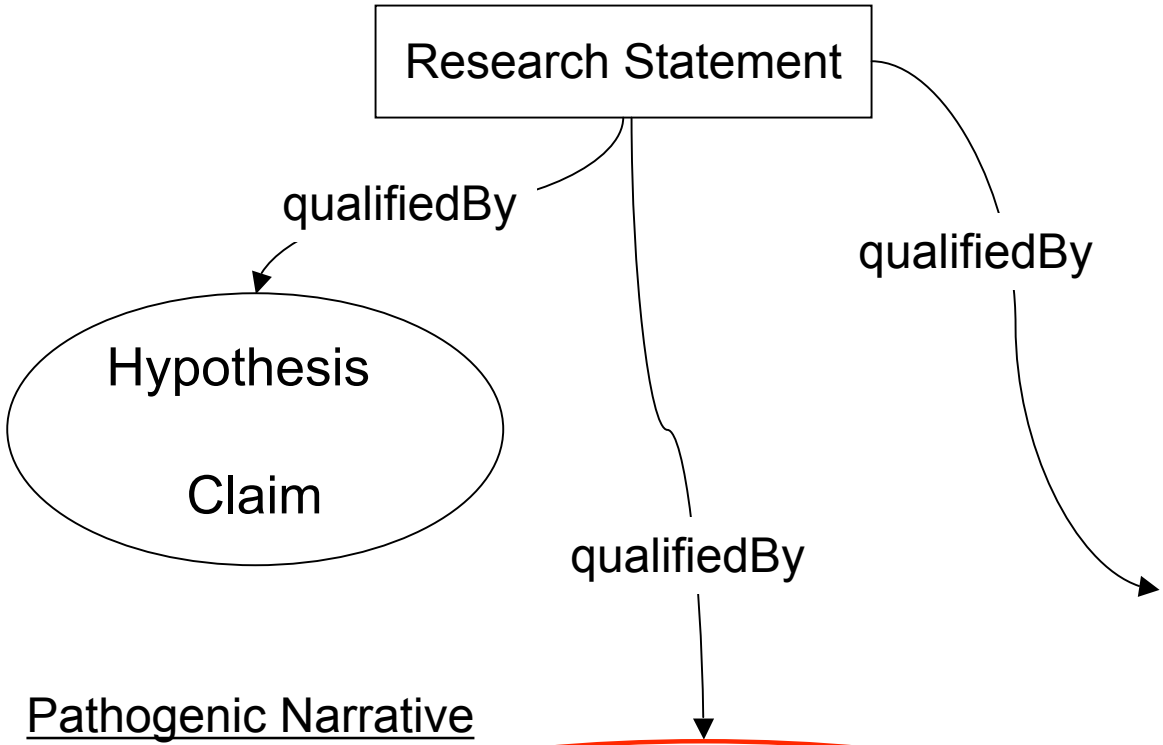
Life Science Entities - 2



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
imer disease.	9	5	36	19
cium plays a key role in the process of synaptic	9	11	47	1
laques 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 *Drosophila*, 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 Armadillo 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

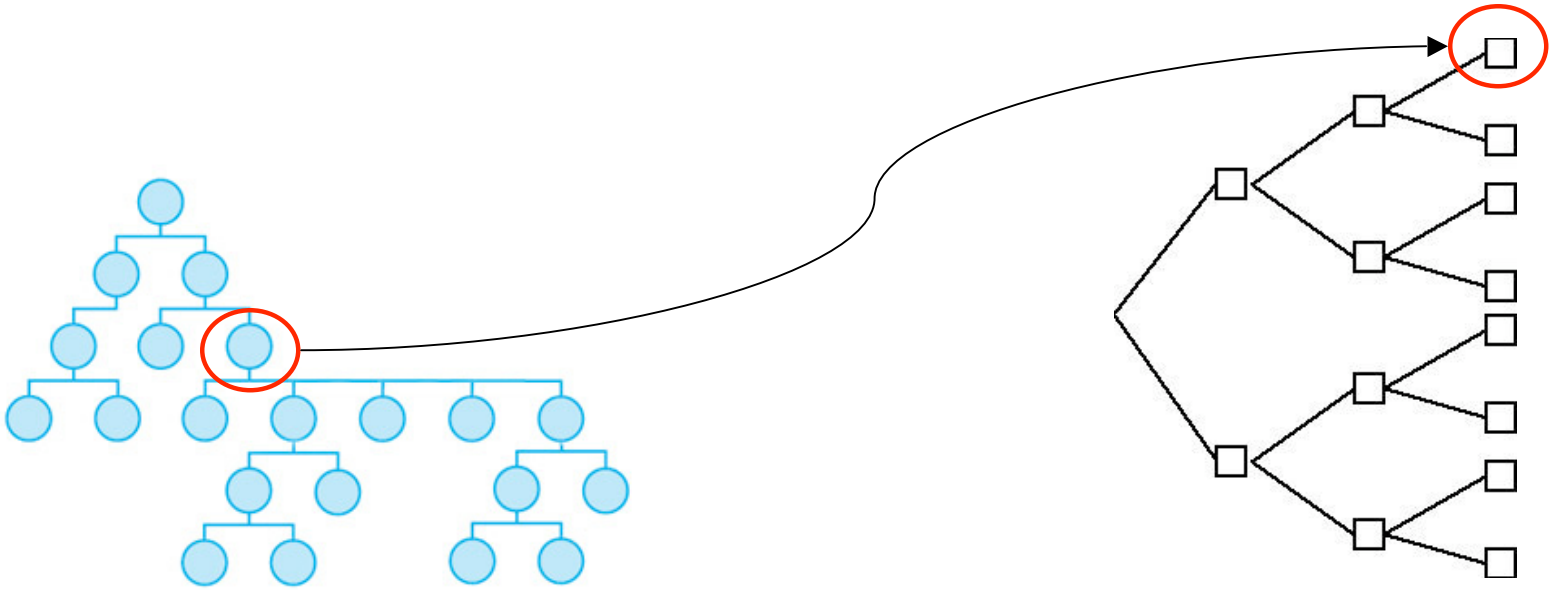
- Controlled vocabularies
- Controlled vocabularies should be easy and fast to be defined by “not ontologists”
- Link the terms to existing (or future) ontologies and keep track of provenance/authoring
- OWL-DL but without impacting the reasoning
- Share the vocabularies in RDF format

Agile definition and mapping

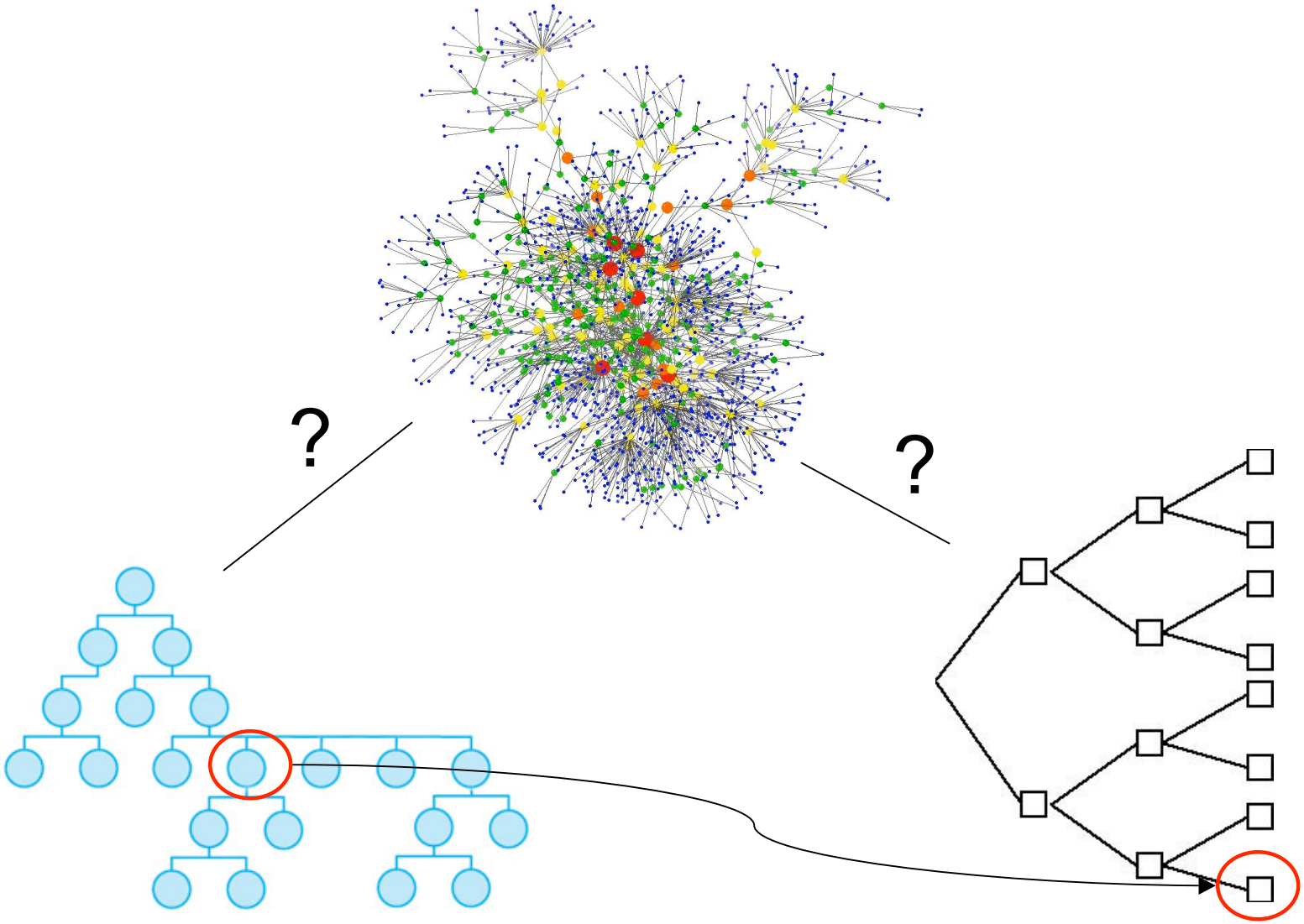
- We don’t want users “not ontologists” to deal with classes, properties and restrictions
- Taxonomies are often enough for simple classifications and can be defined easily (terms + definitions)
- We want trained ontologists to figure out how to map the terms to existing ontologies (if we have resources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

Possible solution: SKOS

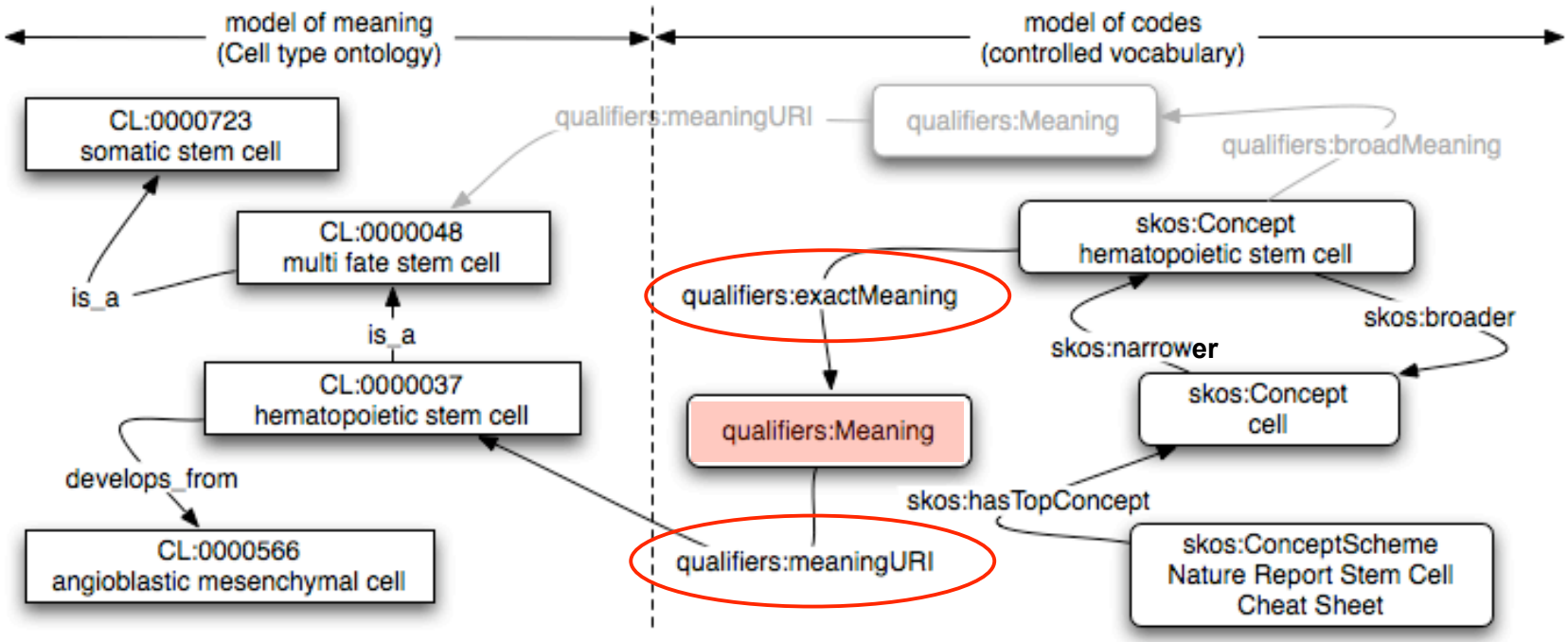
The Simple Knowledge Organization System provides a model for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, folksonomies, and other similar types of controlled vocabulary.



How to refer to formal ontologies?



An example



skos:Concept —— *qualifiers:exactMeaning*
qualifiers:broaderMeaning
qualifiers:narrowerMeaning ⟶ qualifiers:Meaning

owl:Thing ⟵ *qualifiers:meaningURI*

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>
```

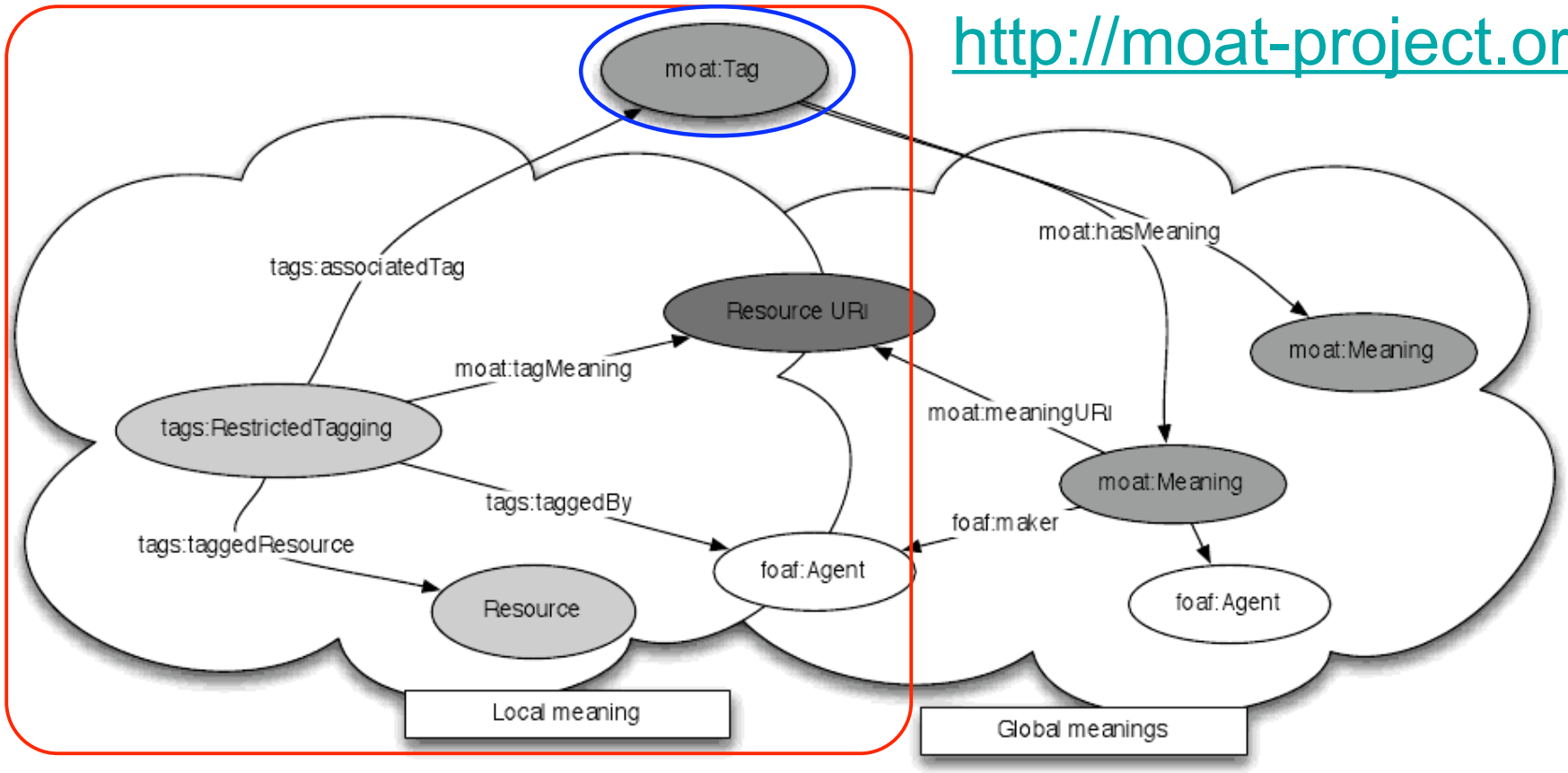
Meaning Of A Tag (MOAT) - 1

```
<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>
```

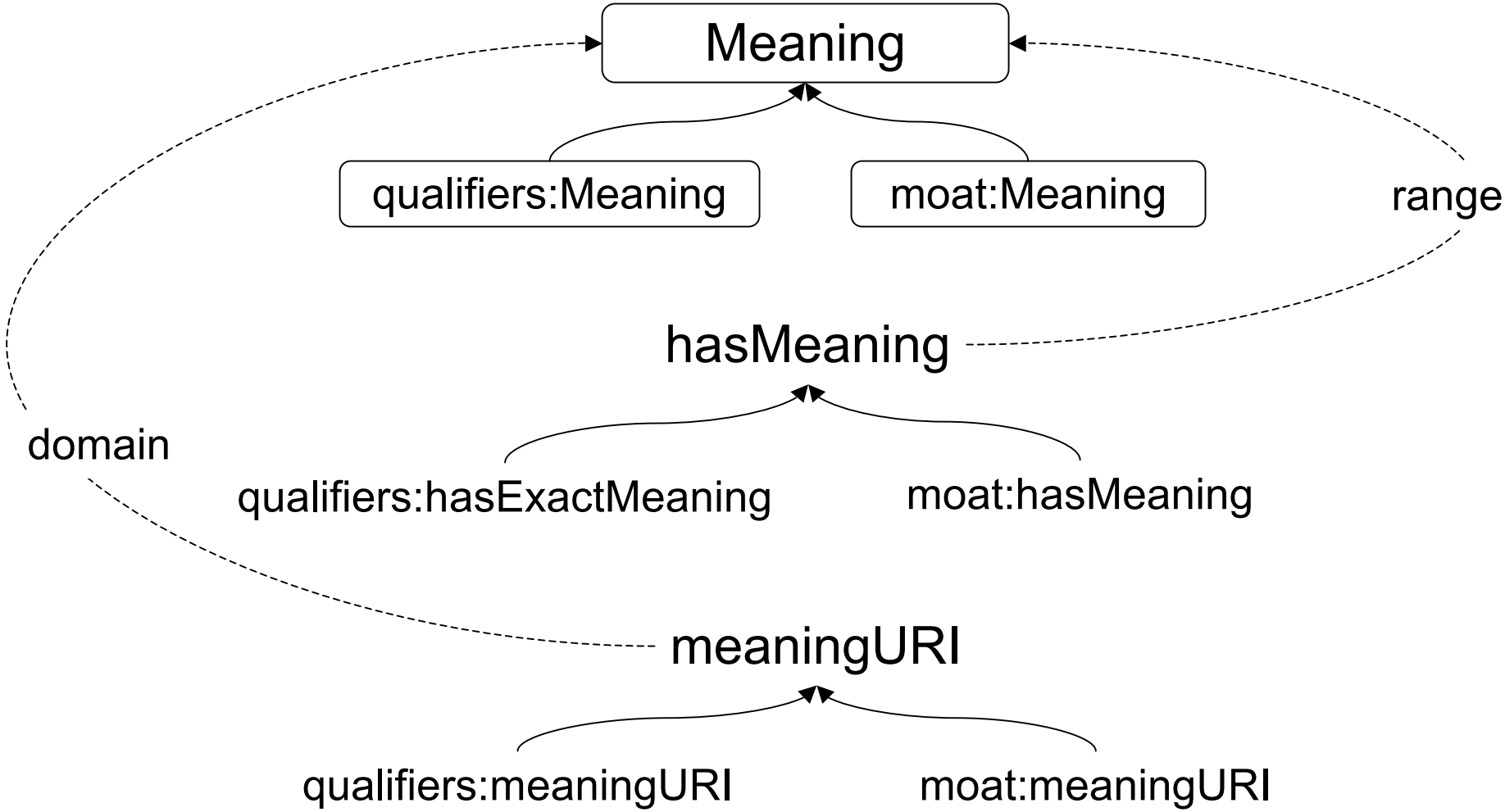
Meaning Of A Tag (MOAT) - 2

```
<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>
```

http://moat-project.org/



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)