Scientific Discourse Ontologies
Context, Use Cases, Strategy

Tim Clark

Harvard Medical School &
Massachusetts General Hospital

CSHALS Tutorial February 24, 2010

Outline

- Task Description
- Background
- Complex Medical Disorders
- Ontologies of Discourse
- Use Cases
- Strategy
- Conclusion

Task Description

Create an open, shared, widely interoperable Semantic Web vehicle for biomedical discourse

- which can be evolved over time,
- is linked to key biological categories
- is fully specified by ontologies
- meets a clear set of practical scientific needs &
- is taken up into existing ecosystems of software developers, ontologists and scientists.

Discourse categories

- research questions,
- scientific claims,
- hypotheses,
- comments and discussion,
- experiments,
- data,
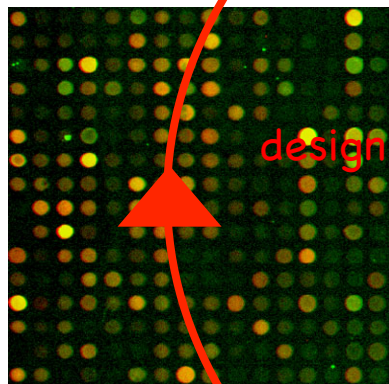- analysis,
- and evidence.

Biological categories

- Organisms, anatomical structures, genes, proteins, biological & chemical reagents, laboratory protocols, biological processes, disease classifications, user-generated taxonomies, and bibliographic references.
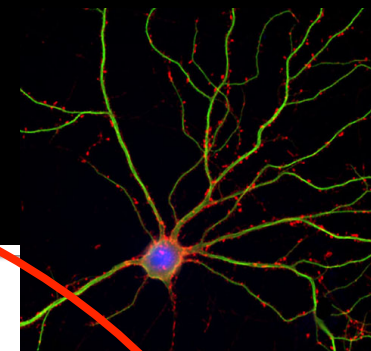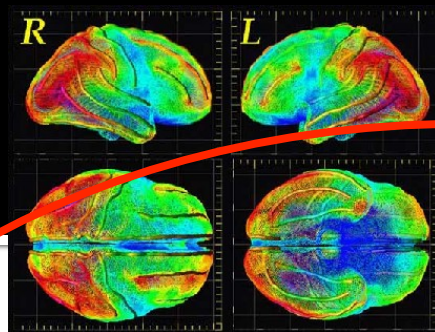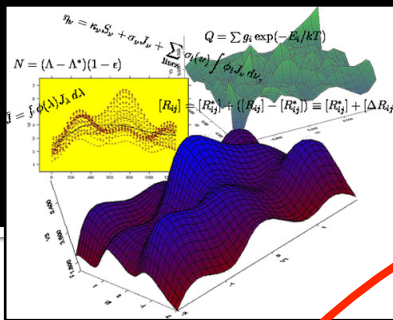
run experiment

collect data

design experiment
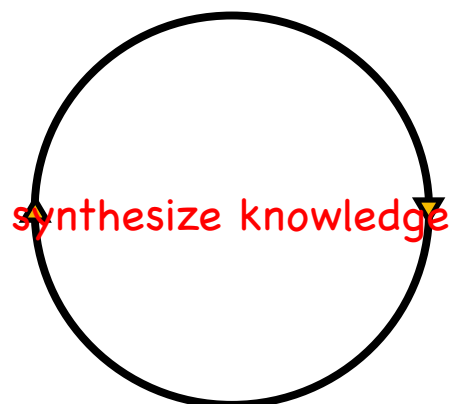
synthesize knowledge

interpret data
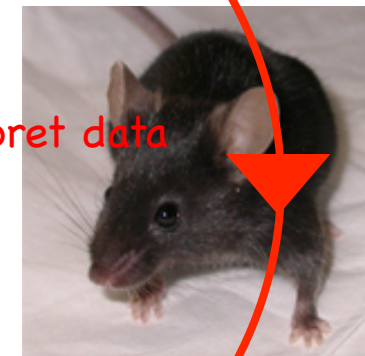
create hypothesis

share interpretations
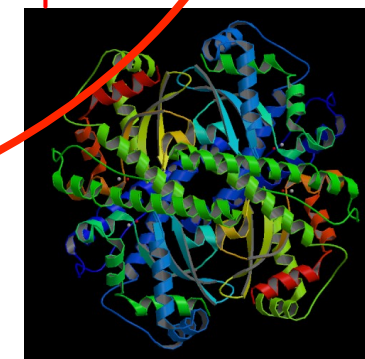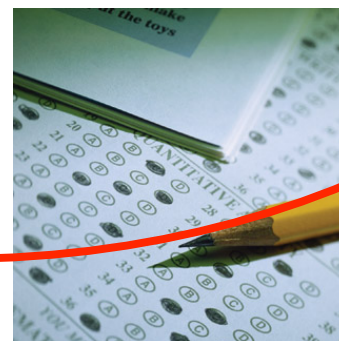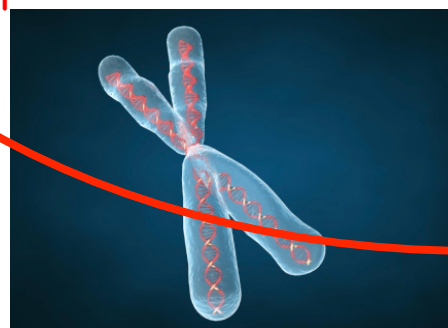
Background

- Modern scientific discourse dates to the 17th century
- Scientific societies
  - Royal Society of London (1665),
  - Académie des Sciences (1666),
  - Kurfürstlich-Brandenburgische Societät der Wissenschaften (1700), etc.
- Founded as alternatives to the Universities.
- Concerned with "the new Philosophy", i.e. experimental science.
- Published journals of their discoveries, met at the local coffee house, collaborated on experiments.

Information Explosion

- 16th c: A scientist might follow <u>all of science</u> by reading three or four journals.

- Today: Not even possible to follow all the literature in one's own (niche) speciality.

- > 671,000 scientific articles / yr & > 20,000 books / yr published in biomedicine alone.

Technological Explosion

- Access to scientific papers primarily via web.

- NLM PubMed searches/yr > 775.5 million.

- Massive use of websites & databases.

- Open Access to publications, data, software.

Terminologies & Ontologies

- Complex terminology systems in biomedicine since (at least) Linnaeus (1735)
- Today: MeSH, SNOMED, GO, PRO, HUGO, NCBO, OBO, CheBI, etc., etc.
- Massive investment in digitized terminology systems in the biomedical community
- <u>We can leverage these much better than we have done</u>.

18th century

21st century

Discoordination

tech tools & methods

Technologist

objective

Δ Product | Useful Tools

"Productive" Tension

"perfecting the ontology"

"perfecting the science"

Scientist

objective

Δ Science
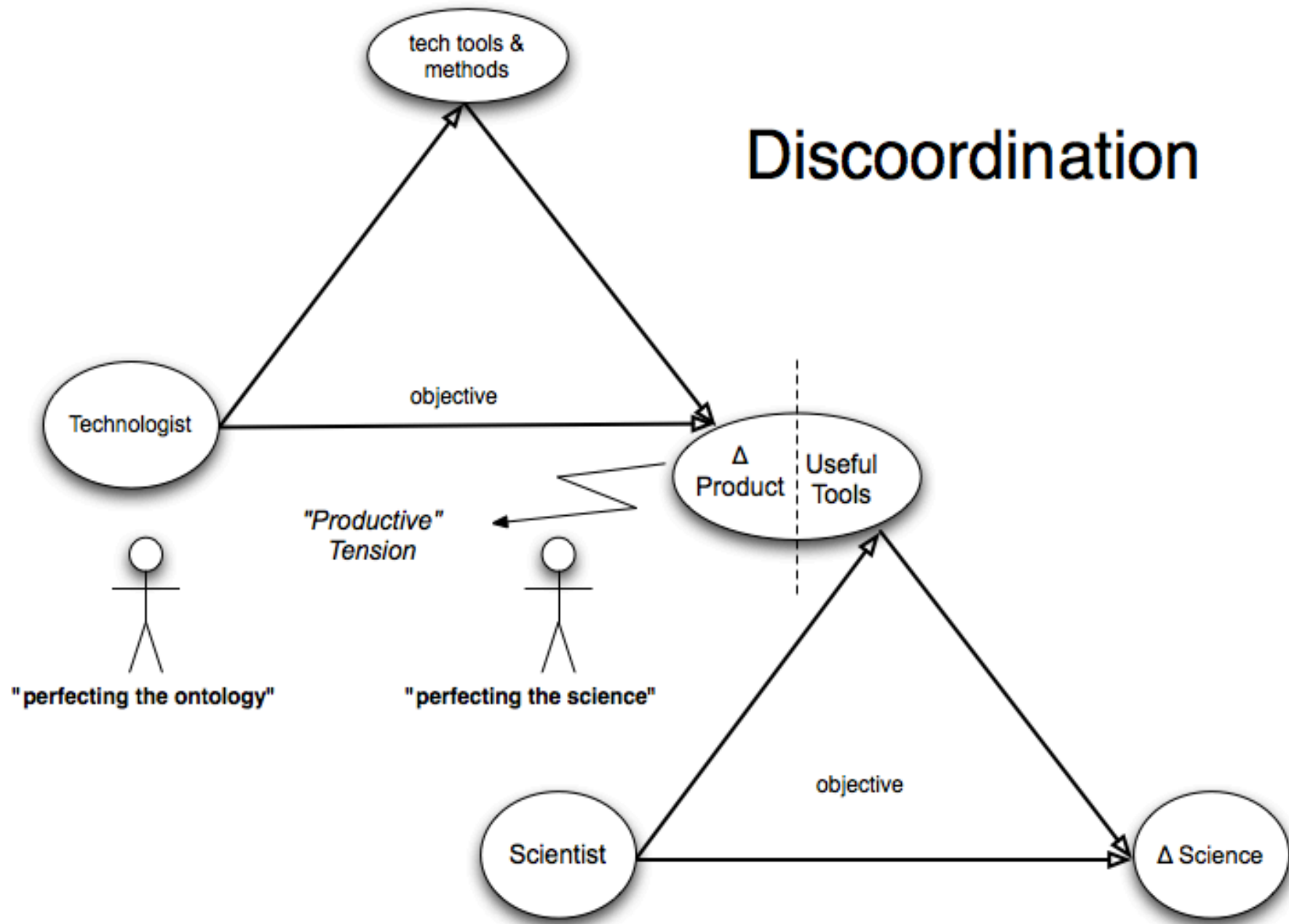
The Poster Child

- Complex medical disorders are the poster child for semantic integration

- <u>Especially: Cancer, neurology & neuropsych</u>

- Alzheimer's, ALS, Autism, Bipolar, Cancer, Drug Addiction, Huntington's, Major Depressive Disorder, Multiple Sclerosis, Parkinson's, Schizophrenia…
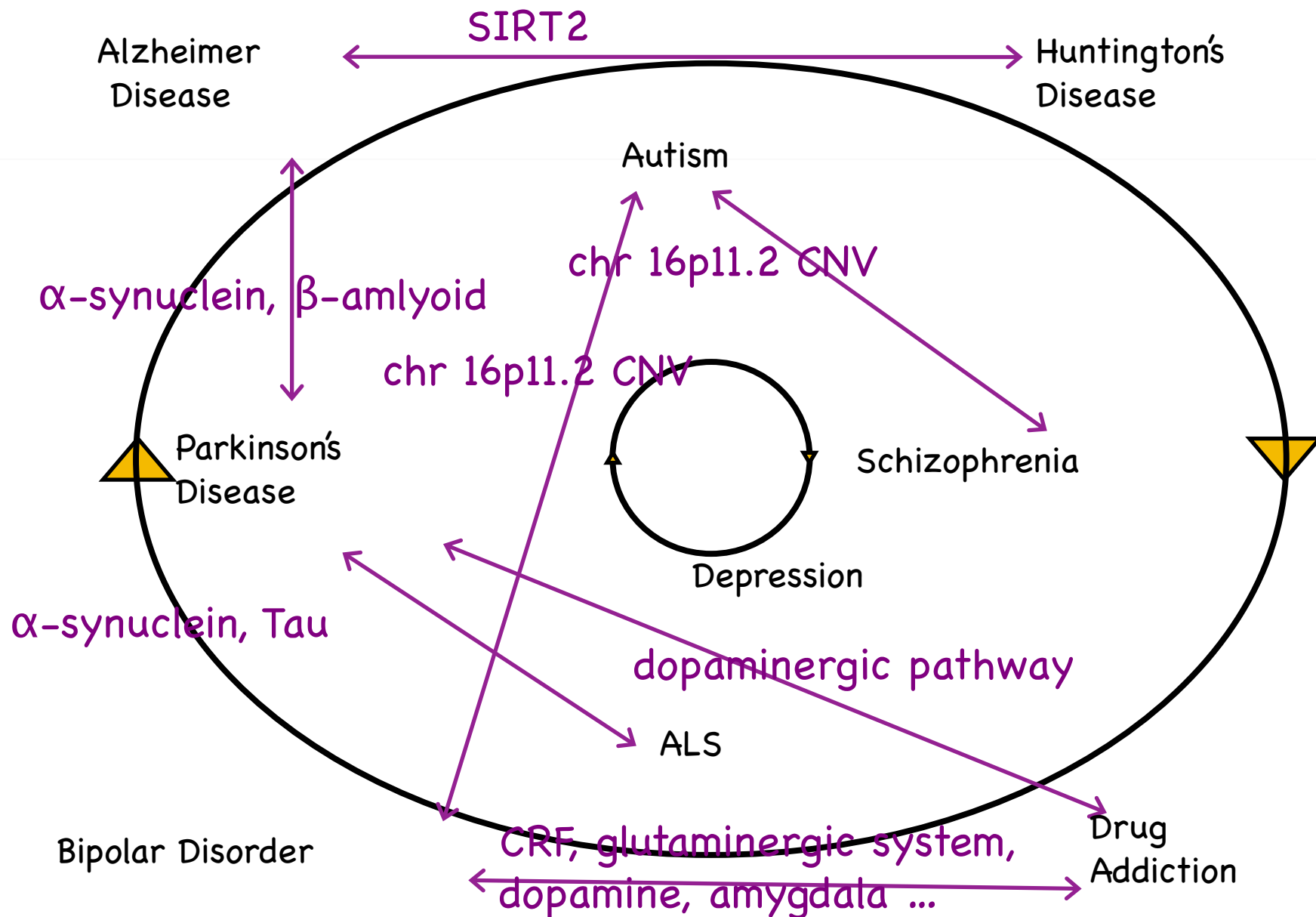
Linking biomedical communities

- Common software toolkits
- Shared terminology systems
- Open data & metadata

- Track the connections as found
- Track challenges to interpretation

We want it all

- Computer-interpretable language is powerful.

- Human readable language is rich and flexible.

- <u>We want it all</u>: rich, flexible and powerful.

Ontologies of discourse

- Reasoning about discourse provenance, evidence, claims, and connections of terms.
- In leading edge science, people are fighting about "what the elephant is".
- Our goal is to HELP THEM bring more materials to the fight and fight efficiently.
- We want them NOT TO MISS ANYTHING.
- Outcome ultimately decided - <u>by experiment</u>.

Nfkb1 ← hasOfficialName — Gene — involvedInProcess → GO:006594

Nf-kappaB ← hasName — Gene

hasName → NF-kB

Gene — encodes → Protein — hasAntibody → Antibody

Protein — hasName → nuclear factor kappa-B, subunit 1

Gene — discussedIn → TextRef

GO:006594 — hasName → inflammatory response

Neuroinflammation

By Chris Scientist, Clinician, Washington University

refTo → NF-kB

The brain has generally been considered to be immune privileged and thus largely protected from immune factors. However, it is now recognized in can initiatory responses including neuroinflammation as a means to reduce injury and clean up injured brain tissue. Responders to brain injury are microglia, which produce a number of factors to regulate the injury response including cytokines and protective factors. Although neuroinflammation may have initial protective effects, prolonged and chronic neuroinflammation may lead to prolonged neurodegeneration such as that seen in PD. Whether neuroinflammation acts as an initial trigger of PD pathogenesis or is a downstream result of another triggering pathogenic event is unclear. Neuroinflammation may lead to increased oxidative stress. Evidence for neuroinflammation in PD includes presence of activated microglia, increased expression of cytokines and pro-inflammatory signaling cascades (e.g., NF-kB). Furthermore, epidemiological data supports a reduce risk of PD in users of anti-inflammatory drugs (e.g., NSAIDS).

textmining: genes

atLocation — refTo → Supplier

hasSupplier → Supplier

TextRef — atLocation

Annotates → Claim

Claim — AlternativeTo

Claim — discussedIn → TextRef

Comment — discussedIn

GO:0050728 — hasName → anti-inflammatory response

user contribution

textmining: GO

References

Tansey MG, Frank-Cannon TC, McCoy MK, Lee JK, Martinez TN, McAlpine FE, Ruhn KA, Tran TA. **Neuroinflammation in Parkinson's disease: is there sufficient evidence for mechanism-based interventional therapy?** Front Biosci. 2008 Jan 1;13:709-17. PubMed

Wilms H, Zecca L, Rosenstiel P, Sievers J, Deuschl G, Lucius R. **Inflammation in Parkinson's diseases and other neurodegenerative diseases: cause and therapeutic implications.** Curr Pharm Des. 2007;13(18):1925-8. PubMed

Contributions

Cannabinoids and neuroprotection in basal ganglia disorders   6 May 2008
   J. Fernandez-Ruiz, Ciudad Universitaria, Madrid

Existing anti-inflammatory agents are ineffective or dangerous   8 Mar 2008
   J. McGeer, USC

Comment — annotates → TextRef

TextRef — refTo

atLocation

Existing anti-inflammatory agents are ineffective or dangerous 8 Mar 2008
J. McGeer, USC

Epidemiological evidence indicates that when there is extended use of nonsteroidal anti-inflammatory drugs (NSAIDs) results in a reduced risk of PD. However, NSAIDs and other existing similar agents are either ineffective or dangerous.

Ref. Bartels & Leenders 2007

Integrate across specilizations

Stem Cell Technology

Parkinson's Disease

SWAN Sample Extract

```
<http://hypothesis.alzforum.org/researchstatement/100> a swande:ResearchStatement ;
    swande:title "Aside from its well-established role in promoting the stabilization
        of microtubules (MTs), tau may have additional functions as a result of its interactions with
        other structures and enzymes"@en;
    swande:description "Poorly defined interactions and functions of tau contribute
        to the difficulty of understanding how pathologically altered tau mediates neurodegeneration.
        For example, tau interacts with the plasma membrane, the actin cytoskeleton and with src
        tyrosine kinases such as FYN."@en;

    swanco:citesAsSupportiveEvidence <http://hypothesis.alzforum.org/citation/321>,
        <http://hypothesis.alzforum.org/citation/322>,
        <http://hypothesis.alzforum.org/citation/323>,
        <http://hypothesis.alzforum.org/citation/324>,
        <http://hypothesis.alzforum.org/citation/325>;
    swanco:researchStatementQualifiedAs
       <http://swan.mindinformatics.org/ontologies/1.2/rsqualifiers/claim>;

    swanqs:qualifiedBy
       <http://swan.mindinformatics.org/ontologies/1.2/pathogenic-narrative/initial_condition>;

    swandr:refersTo
       <http://hypothesis.alzforum.org/protein/201>;
```
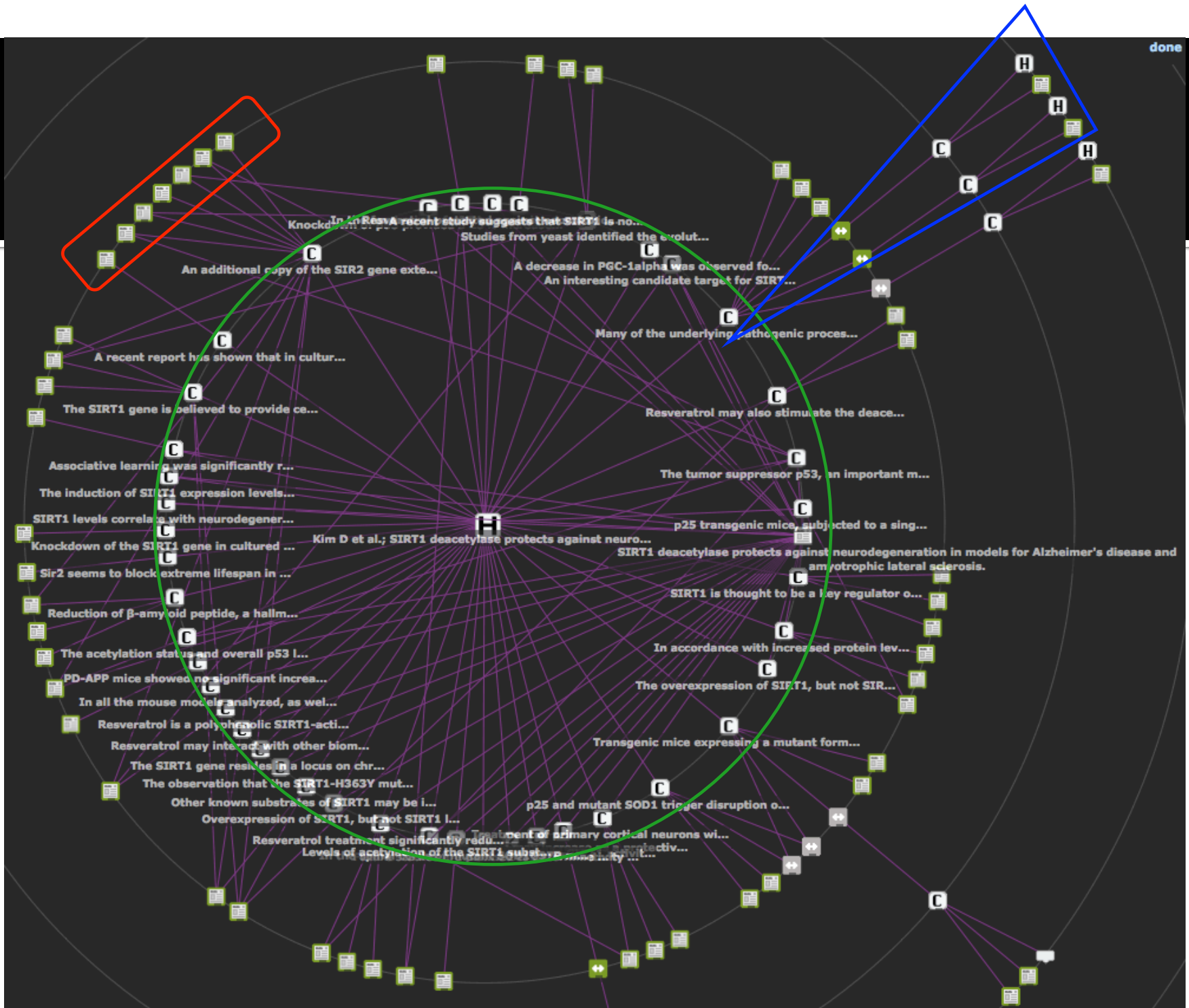
done

Kim D et al.; SIRT1 deacetylase protects against neuro...

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

A recent study suggests that SIRT1 is no...
Studies from yeast identified the evolut...
A decrease in PGC-1alpha was observed fo...
An interesting candidate target for SIRT...
Many of the underlying pathogenic proces...
Resveratrol may also stimulate the deace...
The tumor suppressor p53, an important m...
p25 transgenic mice, subjected to a sing...
SIRT1 is thought to be a key regulator o...
In accordance with increased protein lev...
The overexpression of SIRT1, but not SIR...
Transgenic mice expressing a mutant form...
p25 and mutant SOD1 trigger disruption o...

Knockdown of p25 provides...
An additional copy of the SIR2 gene exte...
A recent report has shown that in cultur...
The SIRT1 gene is believed to provide ce...
Associative learning was significantly r...
The induction of SIRT1 expression levels...
SIRT1 levels correlate with neurodegener...
Knockdown of the SIRT1 gene in cultured ...
Sir2 seems to block extreme lifespan in ...
Reduction of β-amyloid peptide, a hallm...
The acetylation status and overall p53 l...
PD-APP mice showed no significant increa...
In all the mouse models analyzed, as wel...
Resveratrol is a polyphenolic SIRT1-acti...
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...
Overexpression of SIRT1, but not SIRT1 l...
Resveratrol treatment significantly redu...
Levels of acetylation of the SIRT1 subst...

Use Cases in Neurodegeneration

- Semantic integration of collaboratories: Alzforum, SWAN, PD Online, …
- with blogs, wikis & discussion groups …
- with biological databases (genes, proteins, bibliographic, reagents…)
- with publisher websites & databases.

Strategy – Stepwise Integration

- ## SWAN formalization ✔
    - http://www.w3.org/TR/hcls-swan/
- ## SIOC+SWAN ✔
    - http://www.w3.org/TR/hcls-swansioc/
- ## + CITO (and other bibliographic models)
    - Shotton, Ciccarese, Gibson et al.
- ## + Rhetorical Structure (SALT, ABCDE, etc.)
    - DeWaard, Groza, Guttfreund, et al.
- ## + Annotation Model
    - Ciccarese et al.
- ## + Data +Experiment
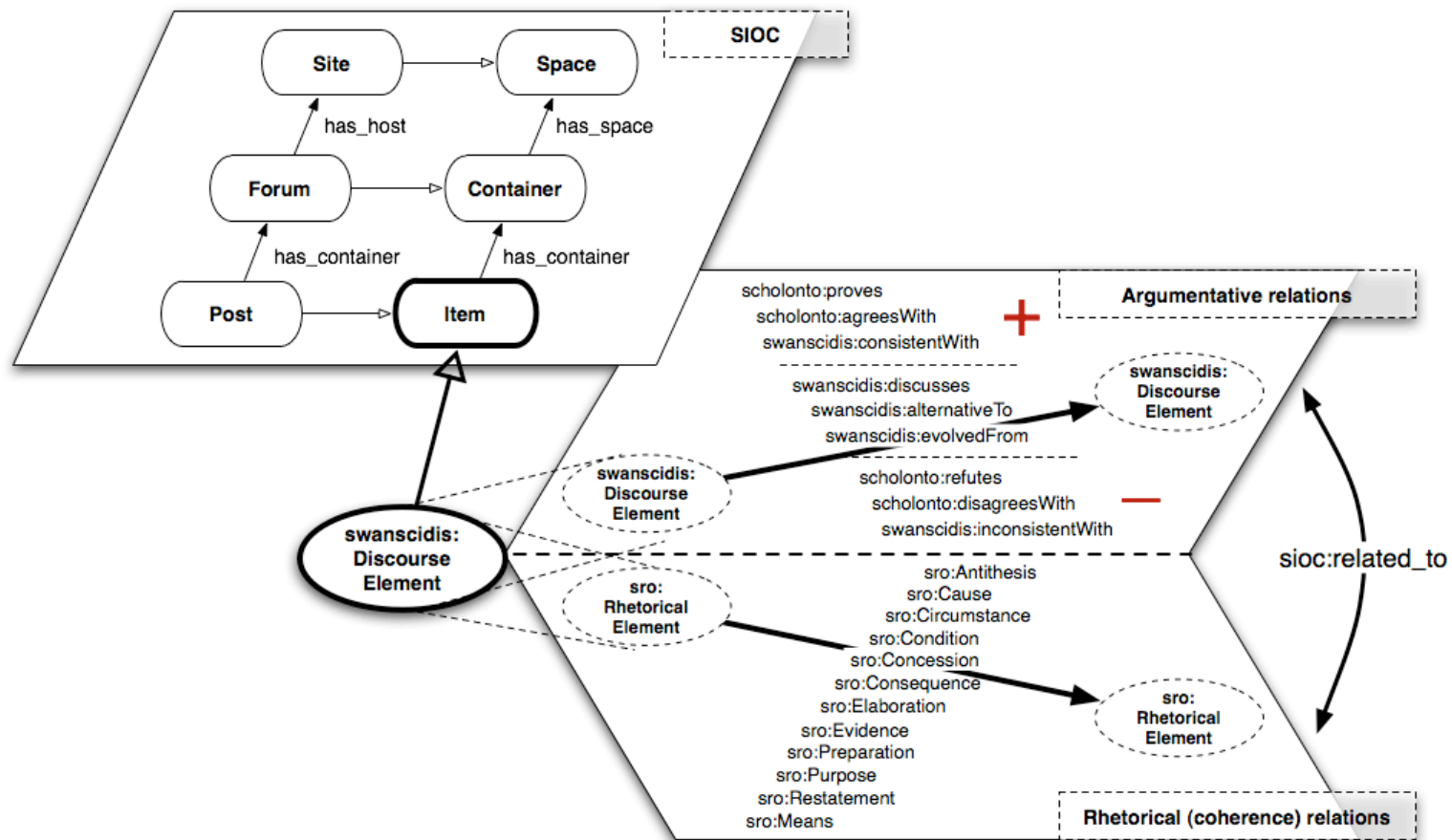   - Gamble et al.

Architecture

- Developed in Groza et al. 2009
  - General structure
- Layered – e.g. SWAN, SALT
  - Coarse-grained structure
  - Rhetorical blocks – e.g. ABCDE, SALT, "Zones"
- Fine-grained structure
  - Discourse elements
- Relations
  - Argumentative + cognitive coherent
  - Rhetorical relations

Abstract Layering
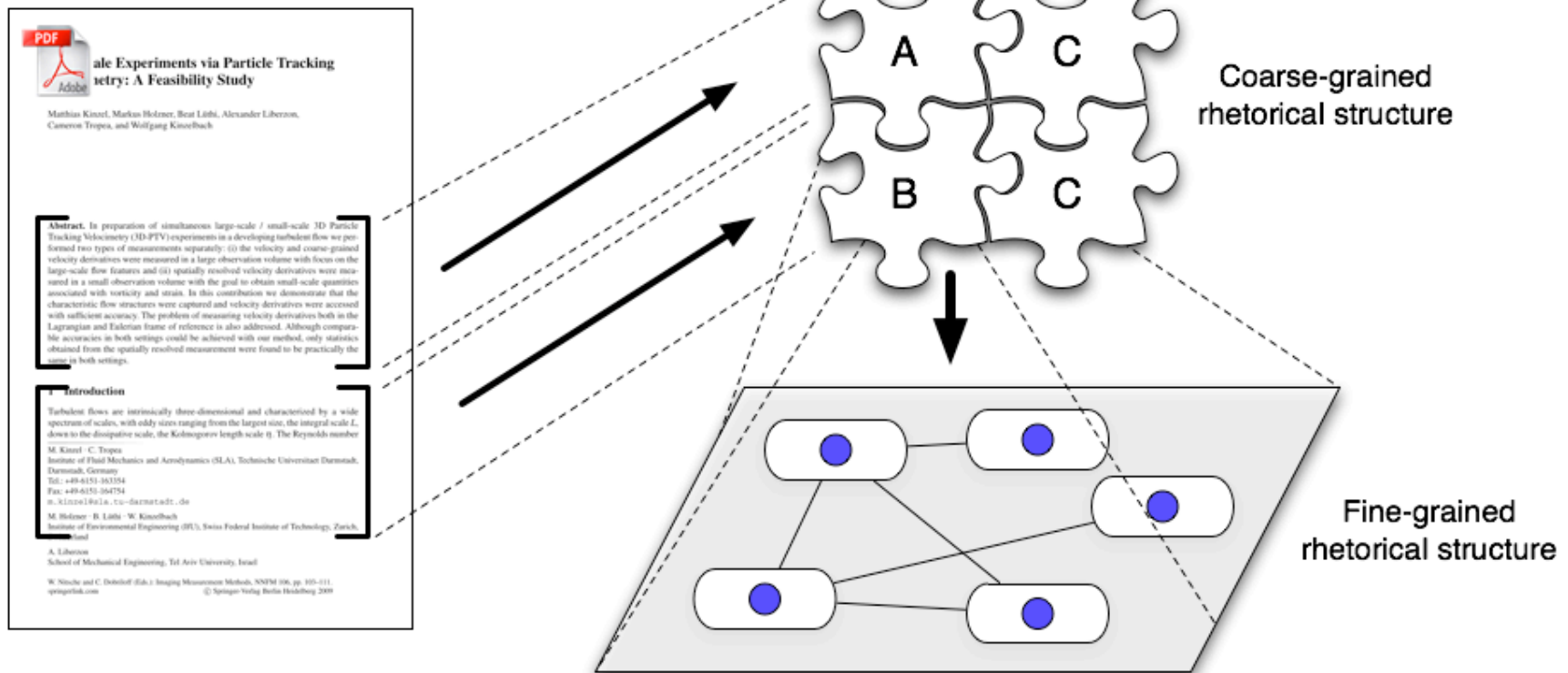
adapted from Groza et al. 2009

Concrete Layering*

*adapted from Groza et al 2009

Summary

- Scientific Discourse requires specialized ontologies to capture discourse relationships in the real "evolving truth" of science.
- These can leverage all the other ontologies and terminological systems built to date.
- Our task group is integrating and refining existing ontologies in this area.
- Use cases are to support multi-disciplinary biomedical research in neurodegeneration.