A middle-grained ontology should bridge between coarse & fine-grained structures

http://esw.w3.org/HCLSIG/SWANSIOC/Actions/RhetoricalStructure/alignment/mediumgrain

Jodi Schneider 2011-04-11

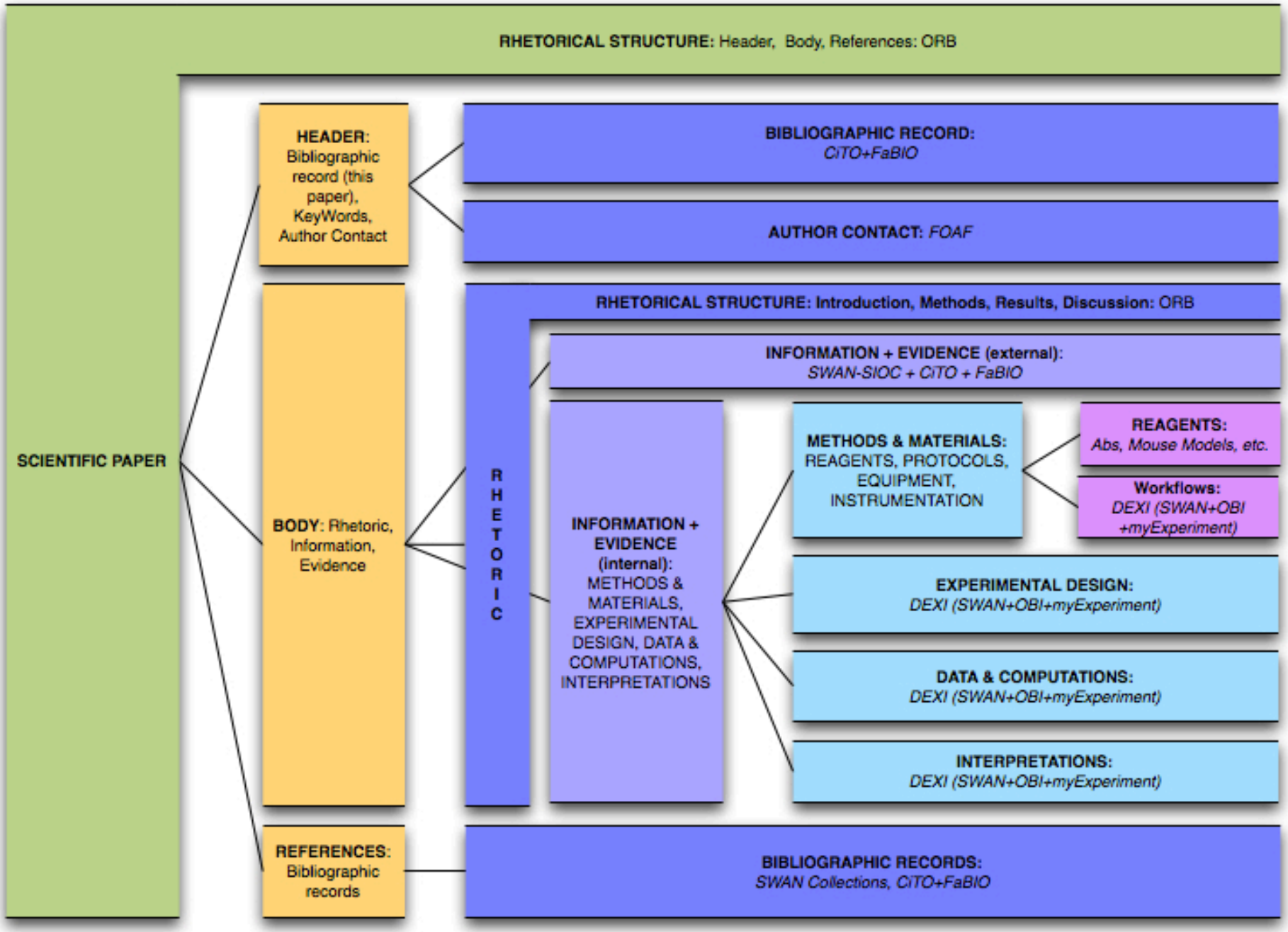
Aligning multiple perspectives

STRUCTURAL

- Block Structure (Entities worth Annotating?)
- Annotation (Relationships between Blocks?)

SEMANTIC FOCUS

- Document
- Process
- Purpose



Slide credit: Tim Clark

Coarse-grained ontology (ORB)

W3C Editor's Draft

3.1. Header

Class	Definition
orb:Header	The part of the publication that models, or captures, meta-information about the publication, including fields such as title, authors, affiliations, publishing venue or abstract.

3.2. Body

Class	Definition
orb:Introduction	The section describing why was the study in the publication undertaken, what was the tested hypothesis or what was the purpose of the research. It lays down the rationale behind the existence of the publication.
orb:Methods	The section describing when, where, and how was the study done. It includes the materials used as part of the study and who was included in the study groups (patients, etc.).
orb:Results	The part synthesizing the results of the study presented in the paper.
orb:Discussion	The section analyzing whether the tested hypothesis was confirmed. It also interprets the results to understand their consequences and importance. And finally, it shows how the approach and results fit to what other researchers in the field have discovered, including possible perspectives of future research.

3.3. Tail

Class	Definition
orb:Acknowledgments	List pointing to funding bodies or individuals that contributed in a way or the other to enabling or supporting the work presented in the publication.
orb:References	External references to other works (e.g., scientific publications, websites, software) that are relevant for the content of the current publication.

http://www.w3.org/2001/sw/hcls/notes/orb/

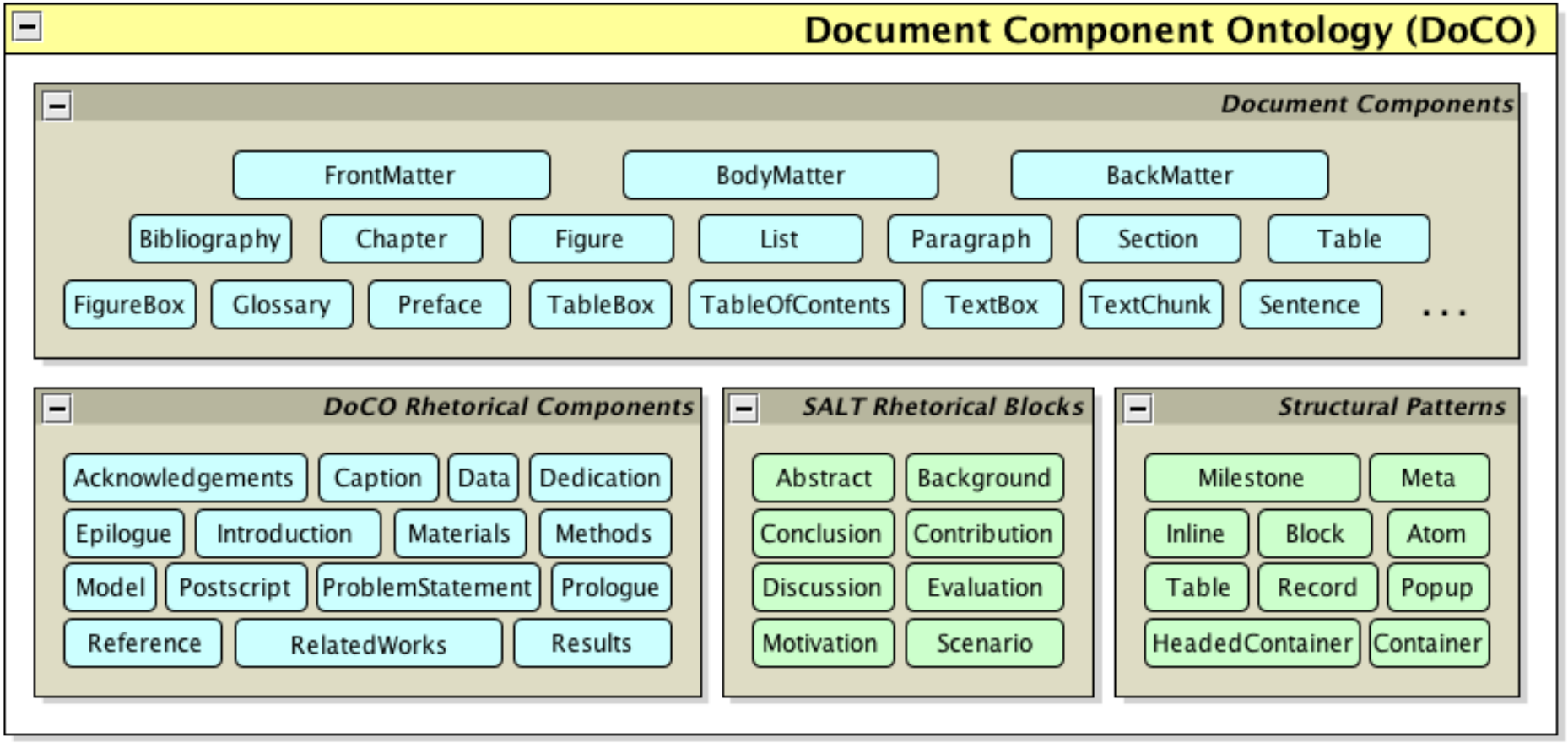
Fine-grained ontologies & approaches

- Document-focused
  - DoCO
  - Anita's block model
- Process-focused
  - DExI
  - myExperiment
- Purpose-focused
  - KeFED

Fine-grained: Document-focused

- Block level
  - Document order
  - Narrative structure (beginning, middle, end)
  - Examples:
    - DoCO
    - Anita's block structure models ("syntagmatic")
- Phrase level
  - Interpretation of claims, hedges
  - Purpose of each clause in the document
  - Examples:
    - Anita: "paradigmatic", Automatic from verb tense
    - Sandor: Automatic detection of novelty from metadiscourse

Rhetorical Blocks



DoCO overview, from
David Shotton
(slightly out-of-date)

Block structure

Medium-grained rhetorical blocks		
Medium-grained (Bio-oriented)		Medium-grained (general)
Block	**Sub-block**	**Block**
Introduction	Positioning	Context
	Central problem	Motivation
	Hypothesis	
	Summary of results	-
-		Background (Related work)
-		Contribution
Experiments	Purpose	Evaluation
	Objects of study	
	Methods	
	Results	Results
	Interpretation	
Discussion	Evaluation of experiments	Discussion
	Comparison with other work	
	Implications	
	Further research	FutureWork

Slide credit: Anita deWaard (Syntagmatic)

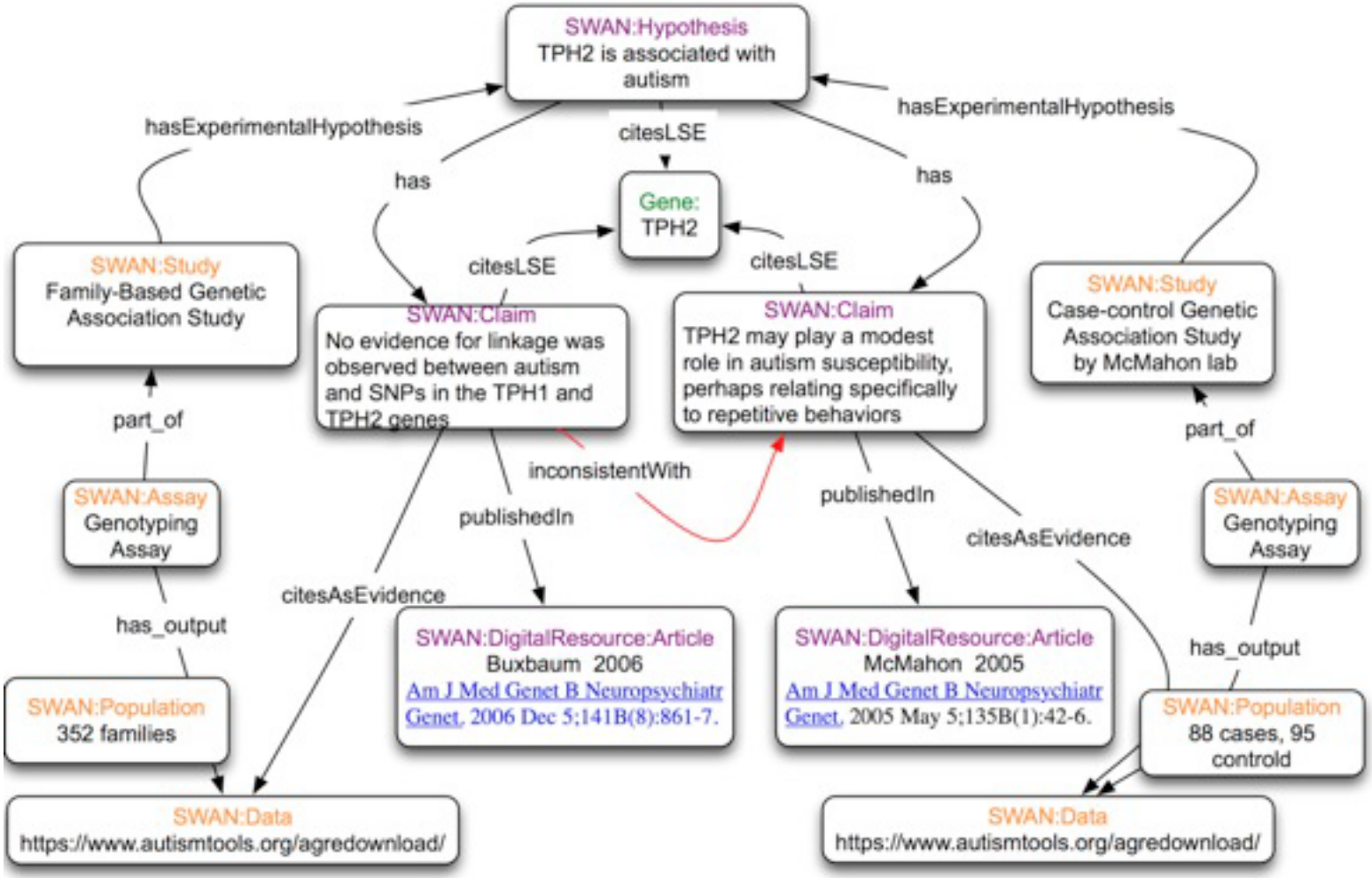
Clause structure

Both seminomas and the EC component of nonseminomas share features with ES cells.	Fact
To exclude that	
the detection of miR-371-3 merely reflects its expression pattern in ES cells,	Hypothesis
we tested by RPA miR-302a-d, another ES cells-specific miRNA cluster (Suh et al, 2004).	Method
In many of the miR-371-3 expressing seminomas and nonseminomas, miR-302a-d was undetectable (Figs S7 and S8),	Result
suggesting that	
miR-371-3 expression is a selective event during tumorigenesis.	Implication

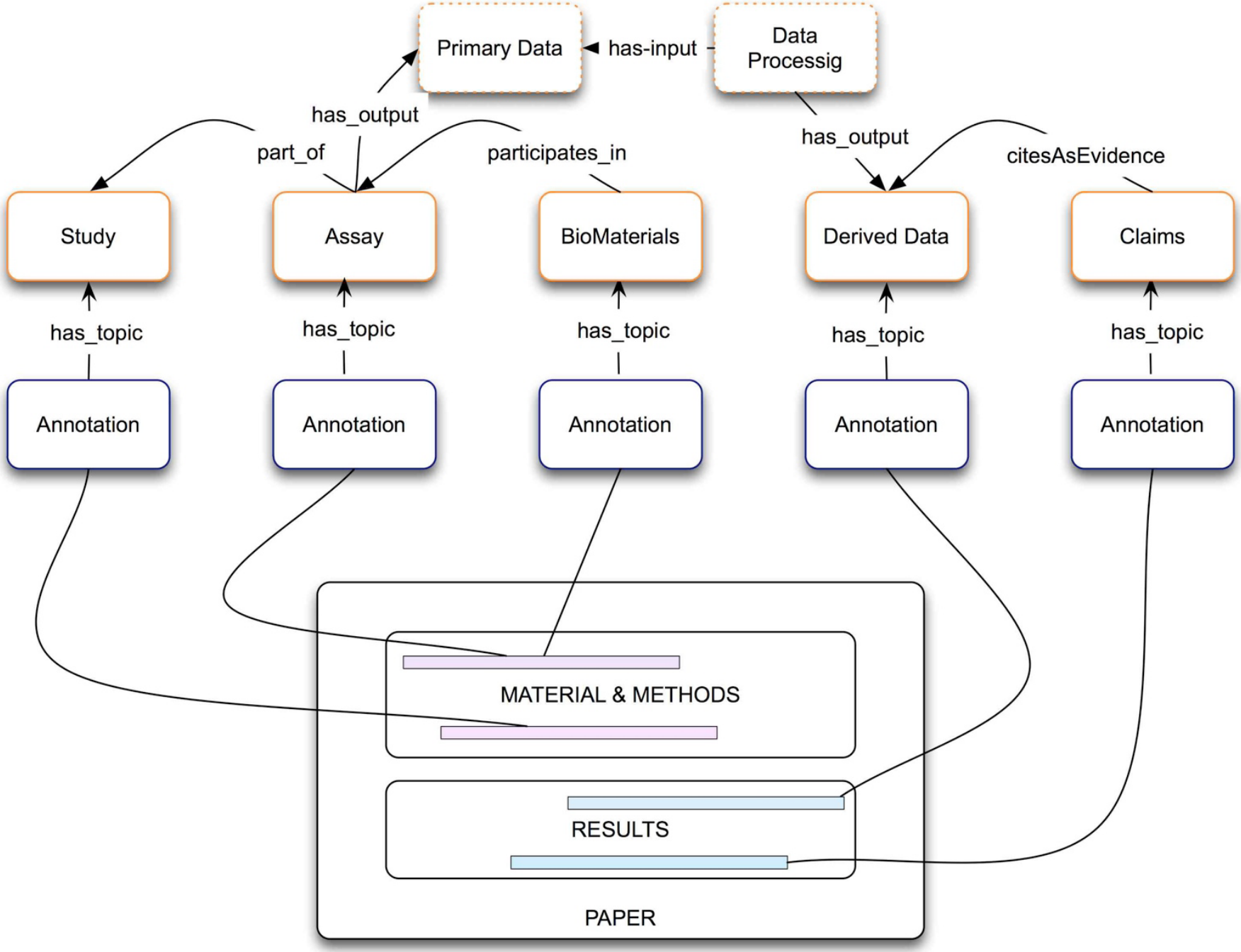
Slide credit: Anita deWaard,
Paradigmatic: verb tense & sentence structure analysis

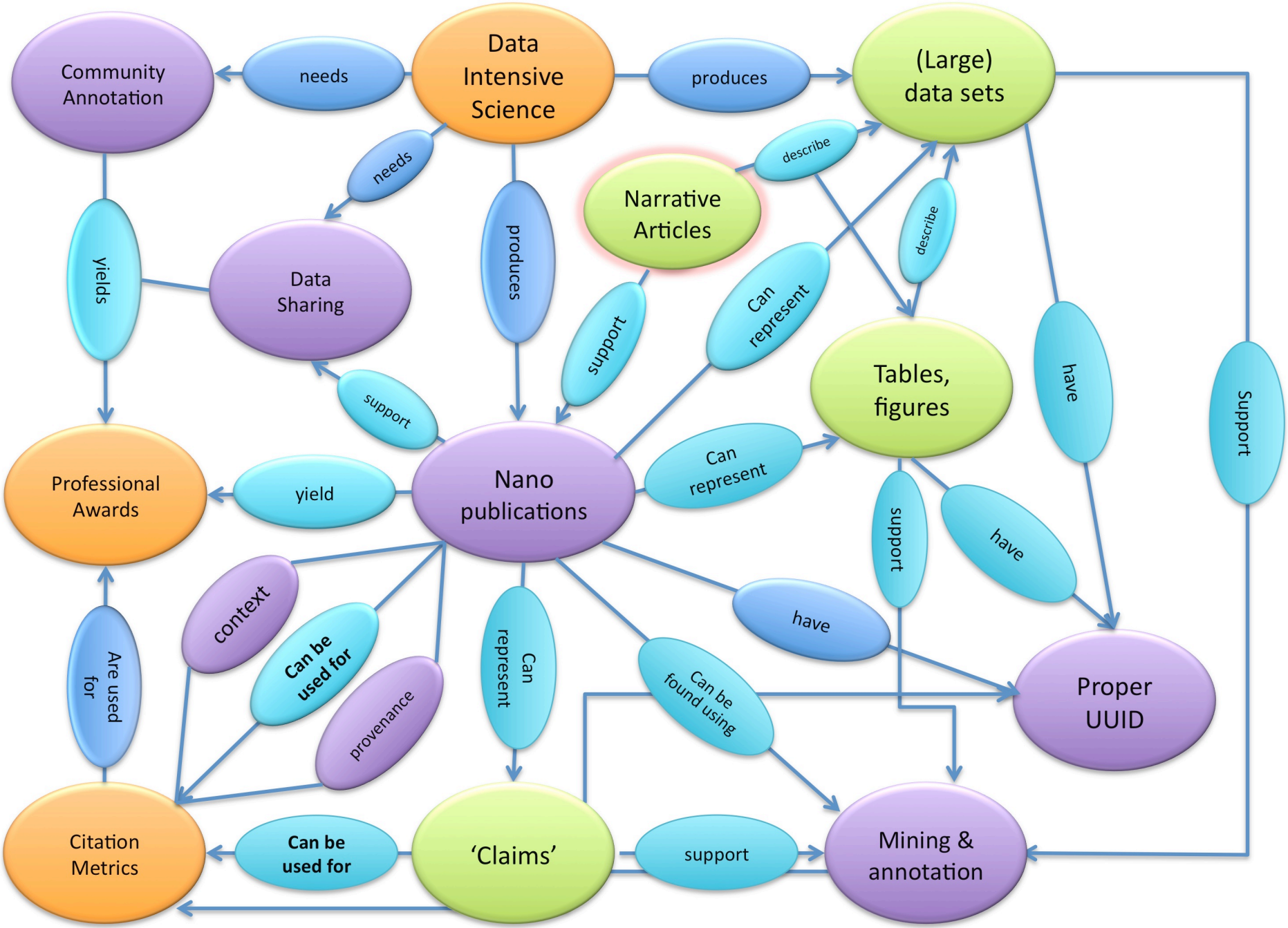
Fine-grained: Process-focused

- "Doing science" perspective
  - Focus on methods, materials, protocols, equipment
  - Examples:
    - DExI
    - myExperiment
  - Focus on domain-level overview
    - e.g. Barend Mons' view of nanopublications


SWAN:Hypothesis
TPH2 is associated with autism
hasExperimentalHypothesis
citesLSE
has
has
hasExperimentalHypothesis
Gene:
TPH2
citesLSE
citesLSE
SWAN:Study
Family-Based Genetic Association Study
SWAN:Claim
No evidence for linkage was observed between autism and SNPs in the TPH1 and TPH2 genes
SWAN:Claim
TPH2 may play a modest role in autism susceptibility, perhaps relating specifically to repetitive behaviors
SWAN:Study
Case-control Genetic Association Study by McMahon lab
part_of
part_of
inconsistentWith
SWAN:Assay
Genotyping Assay
publishedIn
publishedIn
citesAsEvidence
SWAN:Assay
Genotyping Assay
citesAsEvidence
has_output
has_output
SWAN:DigitalResource:Article
Buxbaum 2006
Am J Med Genet B Neuropsychiatr Genet. 2006 Dec 5;141B(8):861-7.
SWAN:DigitalResource:Article
McMahon 2005
Am J Med Genet B Neuropsychiatr Genet. 2005 May 5;135B(1):42-6.
SWAN:Population
352 families
SWAN:Population
88 cases, 95 controld
SWAN:Data
https://www.autismtools.org/agredownload/
SWAN:Data
https://www.autismtools.org/agredownload/

SWAN/DExI


Primary Data
has-input
Data Processig
has_output
part_of
participates_in
has_output
citesAsEvidence
Study
Assay
BioMaterials
Derived Data
Claims
has_topic
has_topic
has_topic
has_topic
has_topic
Annotation
Annotation
Annotation
Annotation
Annotation
MATERIAL & METHODS
RESULTS
PAPER



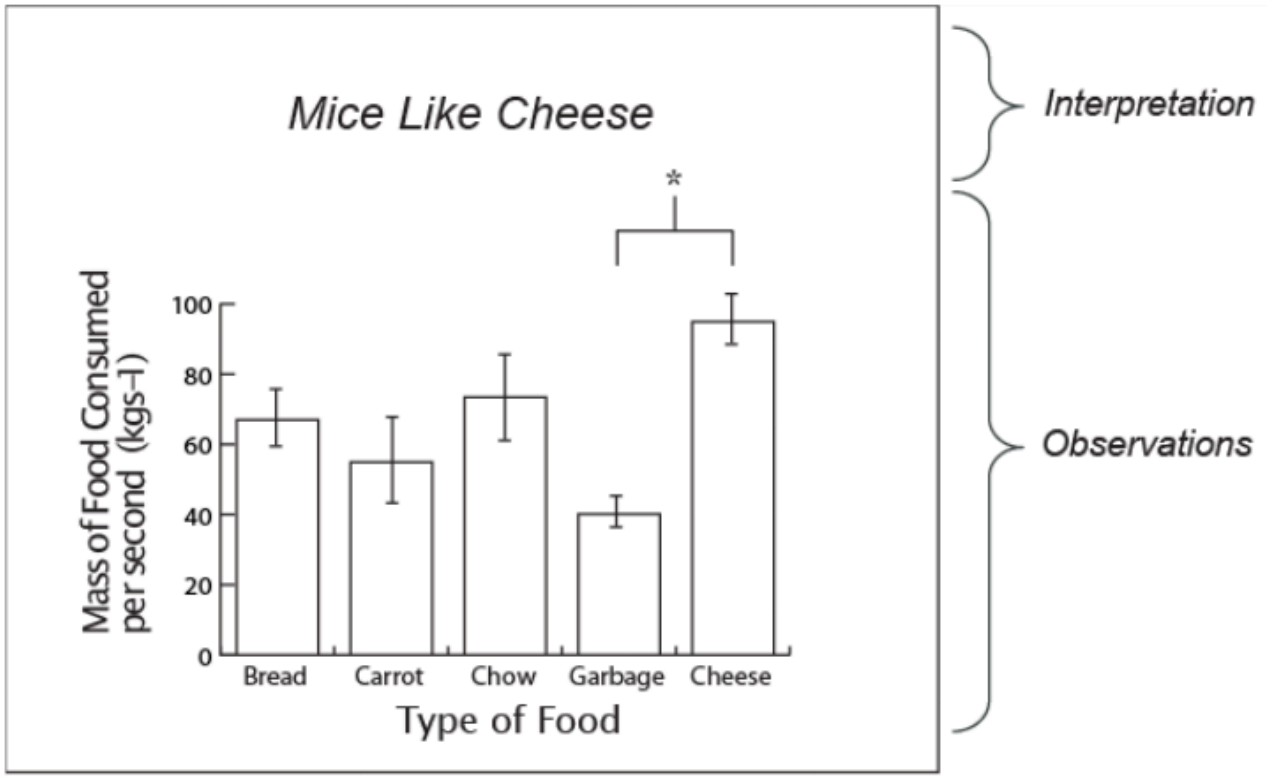
Slide credit: Barend Mons on Nanopublications

Fine-grained: purpose-focused

- Motivational perspective
  - WHY
  - Connects hypothesis, design, interpretation
  - Examples:
    - KeFED

KEFeD **connects** interpretation & observation

For example...



Slide credit: Gully Burns

Questions

- HOW to connect the document discourse to the process of science
- What fine-grained model(s) are we aiming at?
- WHAT are we modeling?
  - the document?
  - the discourse?
  - the process?
  - the underlying reasoning?