synaptica®

Jump-start Your Taxonomy Project

Dave Clarke
CEO Synaptica

@synaptica

Taxonomy Bootcamp London 2017 #tbcl17

© 2017 Synaptica, LLC www.synaptica.com

Jump-start Topics

25 Minute Session

3 things to help jump-start a taxonomy project

1. **Know How**
   - Management & stakeholder buy-in
   - Industry standards
   - Take an audit of your existing content & systems
   - Whiteboard your knowledge domain
   - Adopt & adapt ontologies
   - Design your taxonomy scheme(s)

2. **Starter Taxonomies**
   - Leveraging Linked Open Data resources
   - Licensing third-party taxonomies

3. **Taxonomy Management Tools**
   - Use a taxonomy management tool
   - Share work-in-progress with stakeholders often

© 2017 Synaptica, LLC

synaptica®

Number 1 - KNOW HOW

- Management & stakeholder buy-in
- Industry standards
- Audit existing content & systems
- Whiteboard your knowledge domain
- Adopt & adapt ontologies
- Design your taxonomy scheme(s)

Management Buy-In

Information Today, Inc. **REACH OUT** TO THE LIBRARY AND INFO PRO MARKET

ONLINE SEARCHER EDITORIAL CALENDAR 2017

ONLINE SEARCHER

Information Discovery, Technology, Strategies

Jan/Feb
- International Policies
- Digital Humanities
- Talking Among Ourselves
- **Bonus conference distribution:** American Library Association Midwinter (ALA)

Mar/Apr
- Open Source Solutions
- Scholarly Digital Social Sciences
- Financing the Stone
- **Bonus conference distribution:** Computers in Libraries '17 (CIL), Association of College & Research Libraries (ACRL)

May/Jun
- Big Data and Library Research
- Public Health
- Toy Story
- **Bonus conference distribution:** Data Summit, Special Libraries Association (SLA), American Library Association Annual (ALA), Medical Library Association (MLA), Association of Independent Information Professionals (AIIP)

Jul/Aug
- Getting GLAMerous
- Legal Limits
- Text and Data Mining
- **Bonus conference distribution:** American Association of Law Libraries (AALL)

Sep/Oct
- Artistically Searching
- Taking an Analytical Approach
- Dealing Libraries a Data Hand
- **Bonus conference distribution:** WebSearch University, Internet Librarian International, Internet Librarian, and Charleston Conference on Collection and Development

Nov/Dec
- Search Goes Back to the Future
- Taming Taxonomy Terrors
- Global Perspectives
- **Bonus Conference Distribution:** KMWorld, Enterprise Search and

Editorial content focuses on subjects of topical interest to librarians and other information professionals:
- Science & Technology
- Business & Finance
- Medical & Pharmaceutical
- Social Sciences & Humanities
- News & Current Awareness
- Legal, Tax, Regulatory, & Intellectual Property
- Competitive Intelligence
- Usability

In every issue:
- Information Literacy
- Internet Technologies
- Open Access, Open Source
- Business Research
- Conference Coverage
- Information Industry News
- Library Website Design
- Book Reviews

Nov/Dec
- Search Goes Back to the Future
- Taming Taxonomy Terrors
- Global Perspectives

Ahren Lehnert

Senior Manager Text Analytics, Synaptica

"What is the plot of my story?" You need a good story to sell taxonomy to upper management, communicate the value and the everyday use to your end users, and to ***present your project in a nice, compact, elevator pitch clear to all levels and roles in the organization****."*

© 2017 Synaptica, LLC

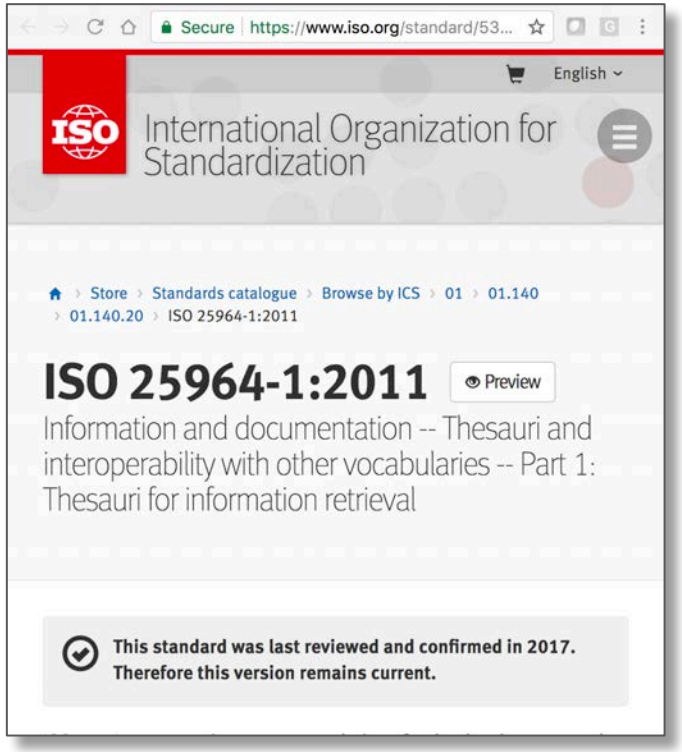
Stakeholder Buy-In

- Segment all potential user groups and talk to them
- Describe their needs and how they will use the taxonomy
- Document the systems they will be using and the user experience they will need

© 2017 Synaptica, LLC

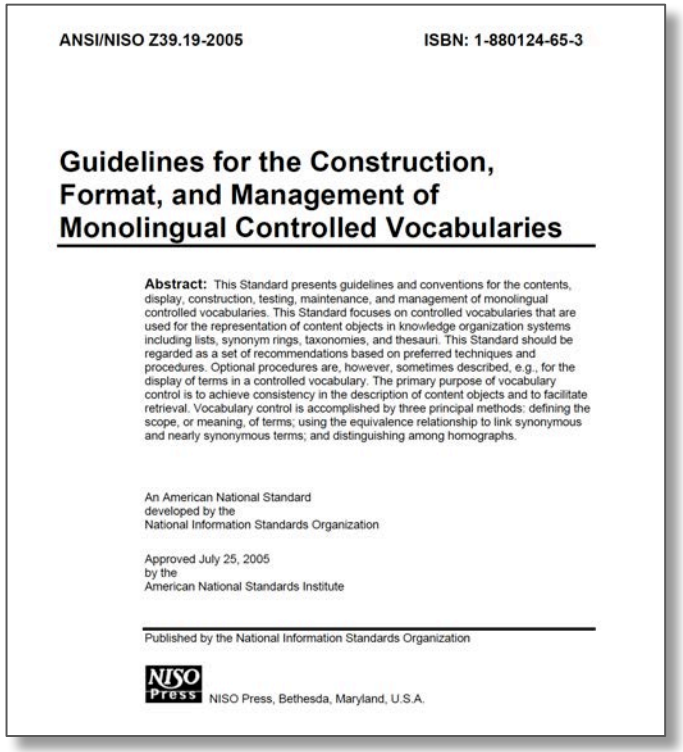
Industry Standards

The Bible



https://www.iso.org/standard/53657.html

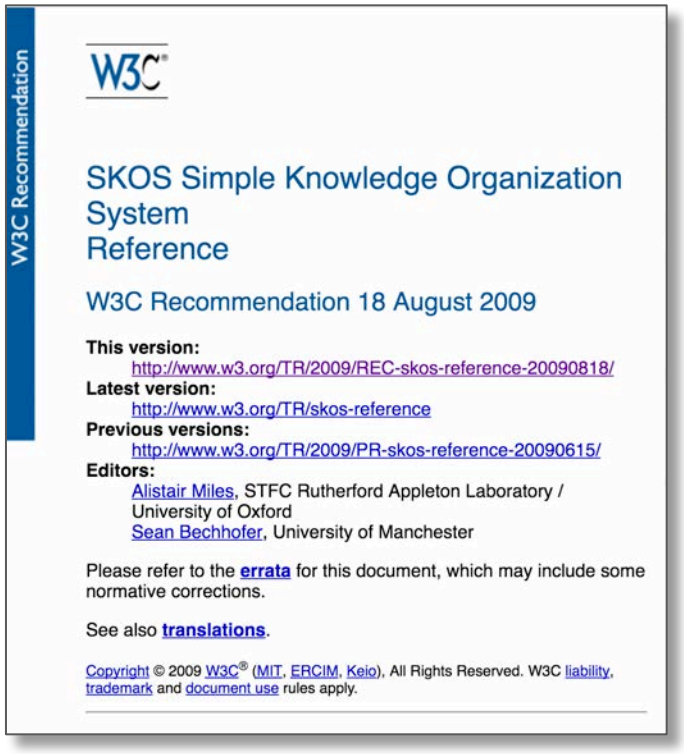
An Older Testament



www.niso.org/standards/resources/Z39-19.html

Lexicographic standards

SKOS



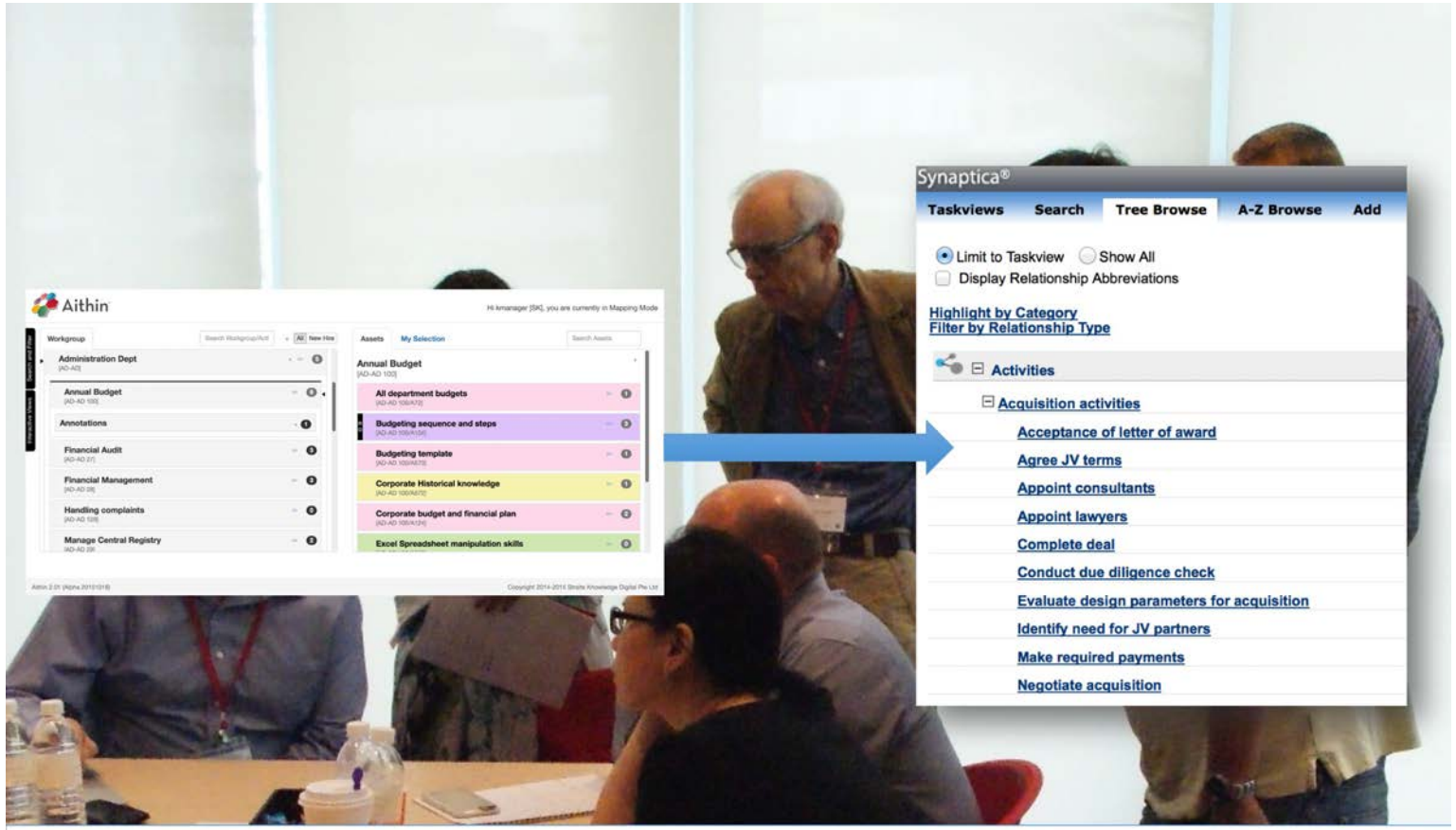
https://www.w3.org/TR/2009/REC-skos-reference-20090818/

Semantic web data standards

© 2017 Synaptica, LLC

What comes before taxonomy – knowledge audits

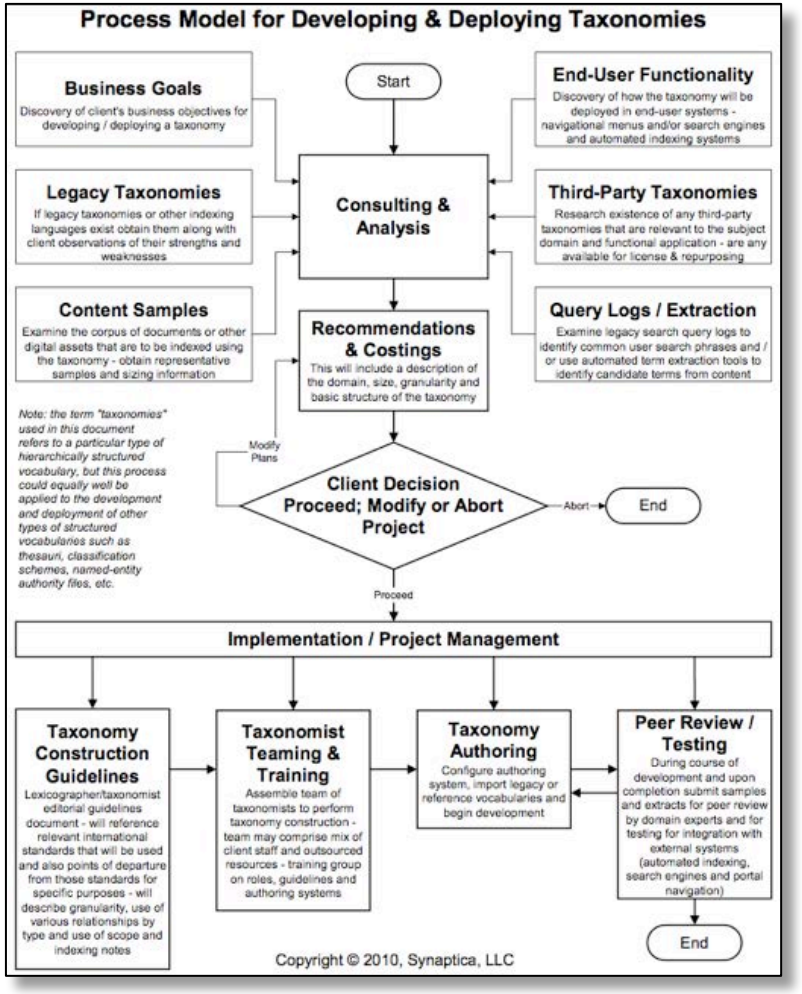
KM Knowledge Audits & Mapping



Aithin™

Activity and Asset descriptions are exported into facets in Synaptica®

Systems audits



Why Audit Existing Systems? Because…

"When you automatically or manually select metadata to classify your content, it comes from the taxonomy. When you type into your search box and there are type-ahead suggestions, they come from your taxonomy. When you look at the global navigation, the values come from the taxonomy. When you apply text analytics to content to determine what it is about or to cluster items by topic, classification terms come from the taxonomy and extracted concepts go back into the taxonomy."

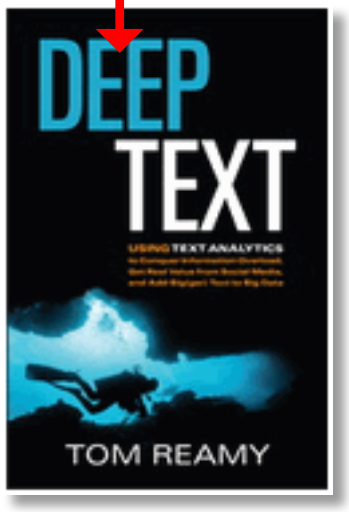
Ahren Lehnert, from taming Taxonomy Terrors in Info Today's Online Searcher

email info@synaptica.com our discovery questionnaire

© 2017 Synaptica, LLC

What comes after taxonomy – indexing and categorization

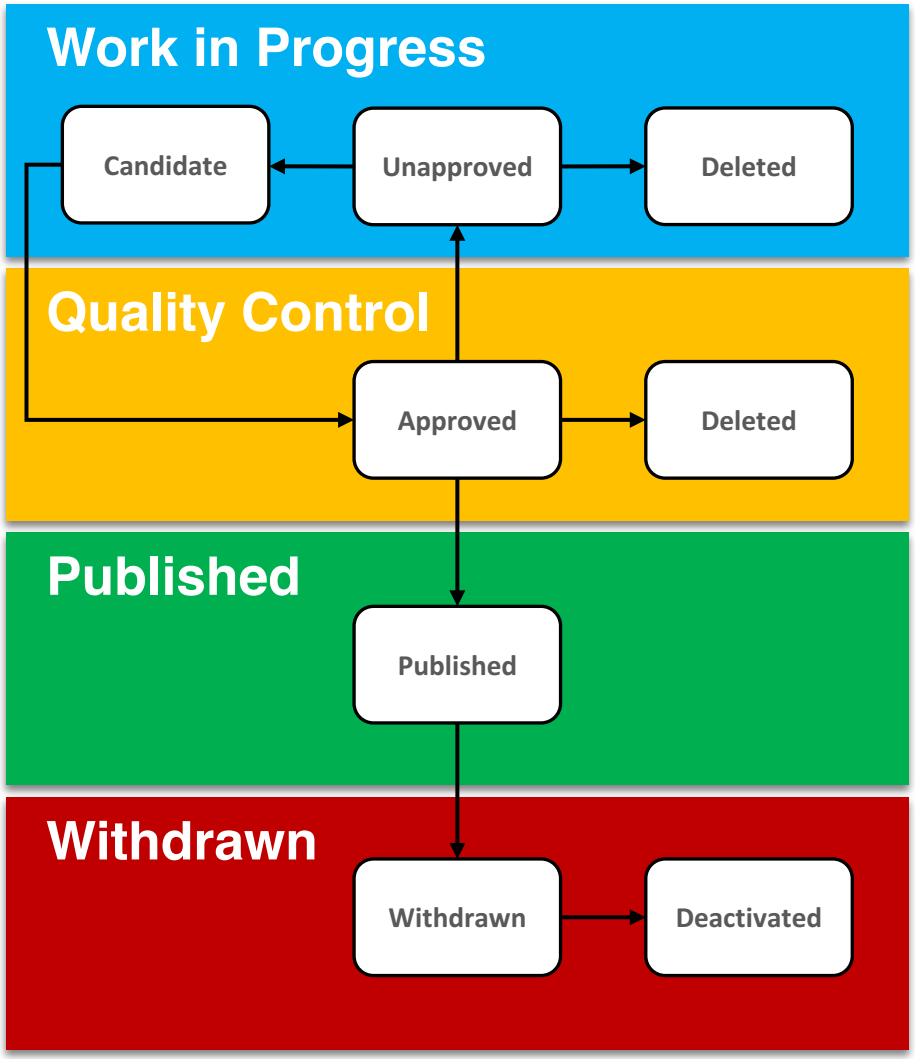
- ❑ Human tagging by content authors
- ❑ Human indexing by professional indexers
- ❑ Machine-aided indexing systems
- ❑ Catonomy developers
- ❑ Faceted end-user navigation interface
- ❑ Search & query-refinement



User Groups	Needs
Taxonomy curators	Import & adopt Build & maintain Human indexing guidelines Machine indexing rules Workflow & governance Publication cycles & version control
Content curators	SME & stakeholder review Content navigation
Human & machine-aided indexers	Search & browse Access guidelines & rules Submit candidates
Search & discovery	Resolve ambiguous concepts Map user query language to indexing language & Submit candidates

© 2017 Synaptica, LLC

Workflow and governance



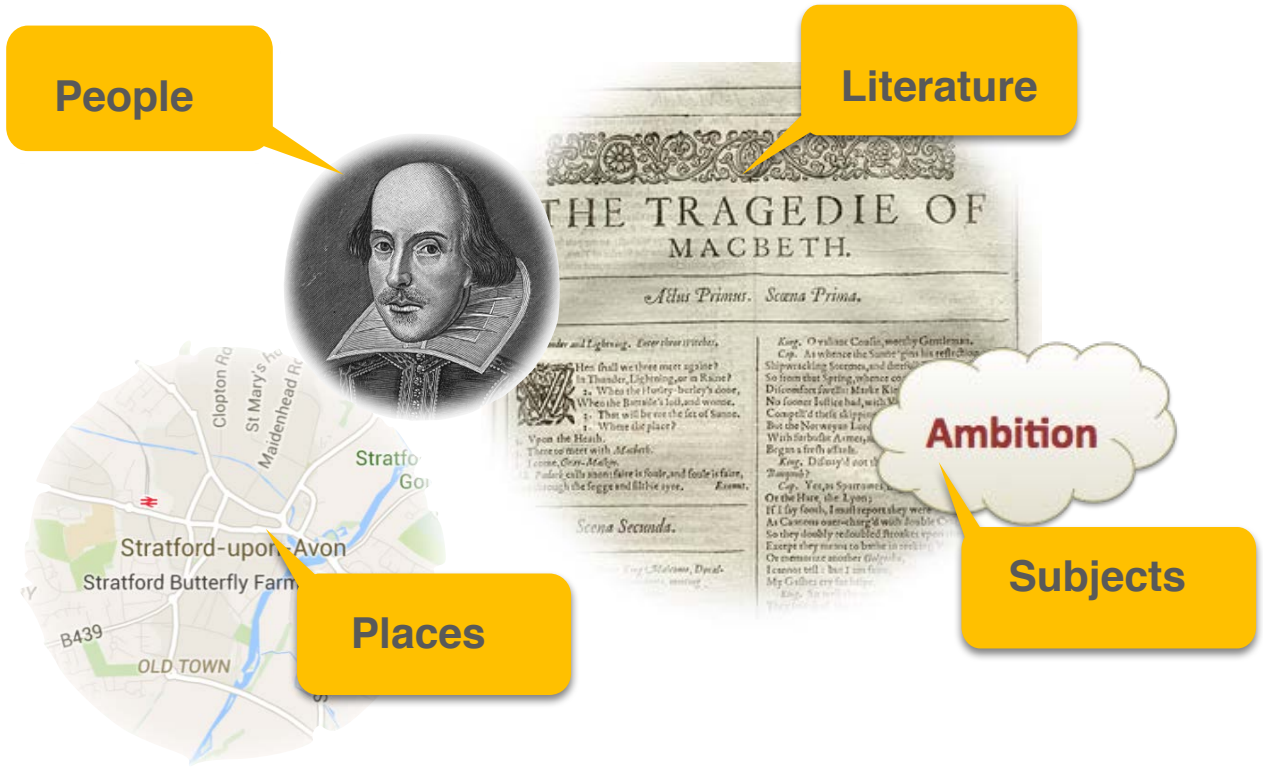
"Taxonomy governance comprises the policies, procedures, and documentation for the ongoing management and use of taxonomy. …the existence of taxonomy is itself a form of governance."

Heather Hedden – *Accidental Taxonomist*
http://accidental-taxonomist.blogspot.co.uk/2013/12/taxonomy-governance.html

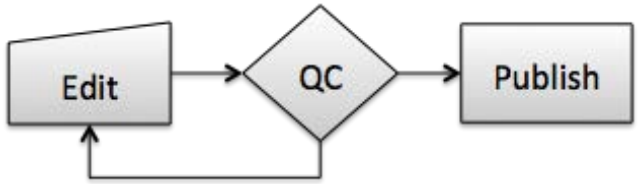
© 2017 Synaptica

Getting Started with Domain Models and Process Models

1. Describe the *types* of real-world objects and **abstract ideas** that an information system needs to reference



2. Describe the business processes and functions that an information system needs to support



3. Choose the formal data models suited to the information, processes, databases and IT systems to be supported

ISO International Organization for Standardization ISO25964

Dublin Core® Metadata Initiative
Making it easier to find information.

W3C SKOS

© 2017 Synaptica, LLC

Diagram the Entities, Attributes and Relationships

Descriptive information about real-world objects and abstract ideas comprises three generic components:

Entities

(ovals)

Unique things and ideas

Attributes

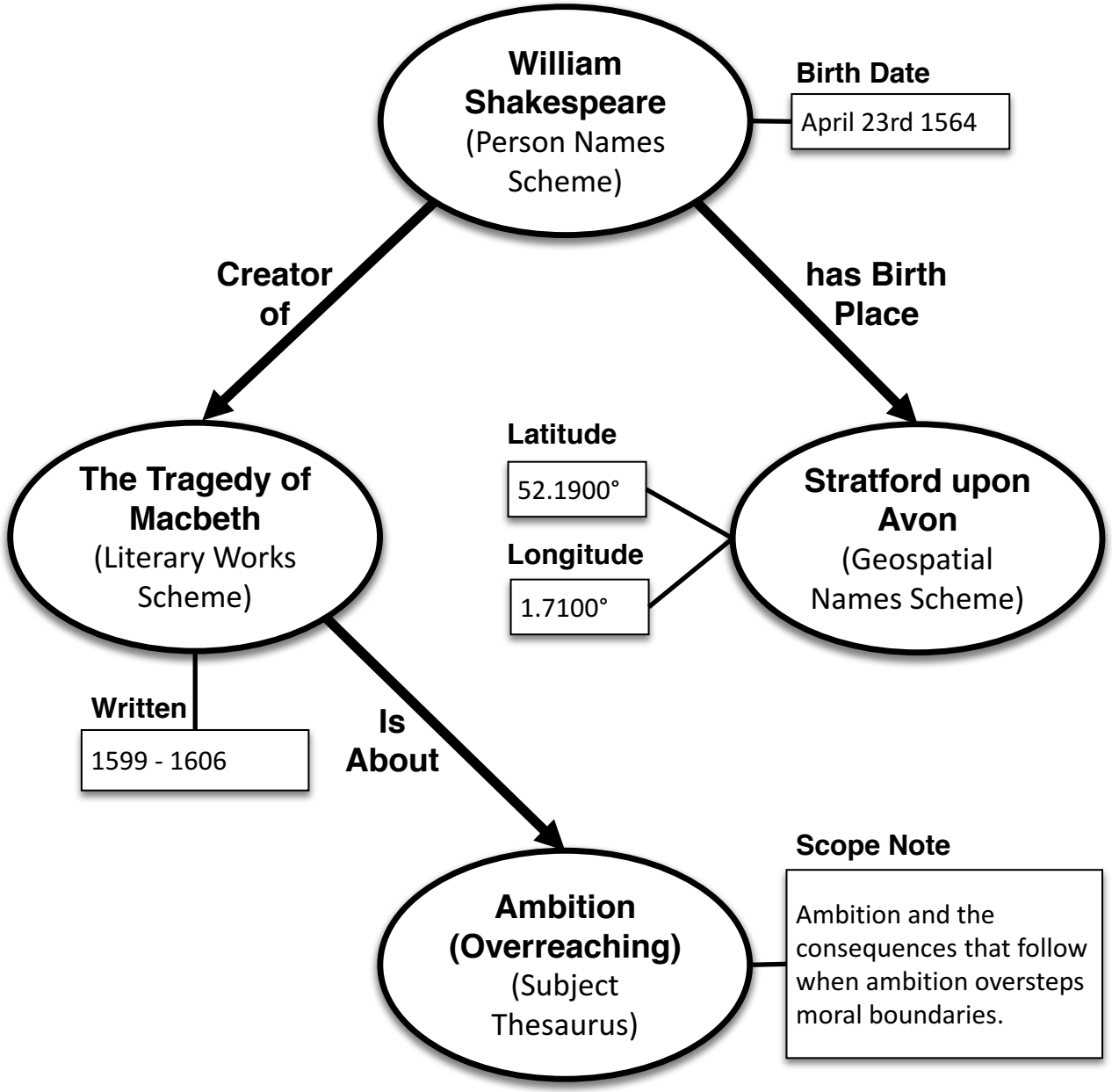
(boxes)

Descriptive properties of entities

Relationships

(arrows)

Connections between entities



© 2017 Synaptica, LLC

Multiple KOS Schemes

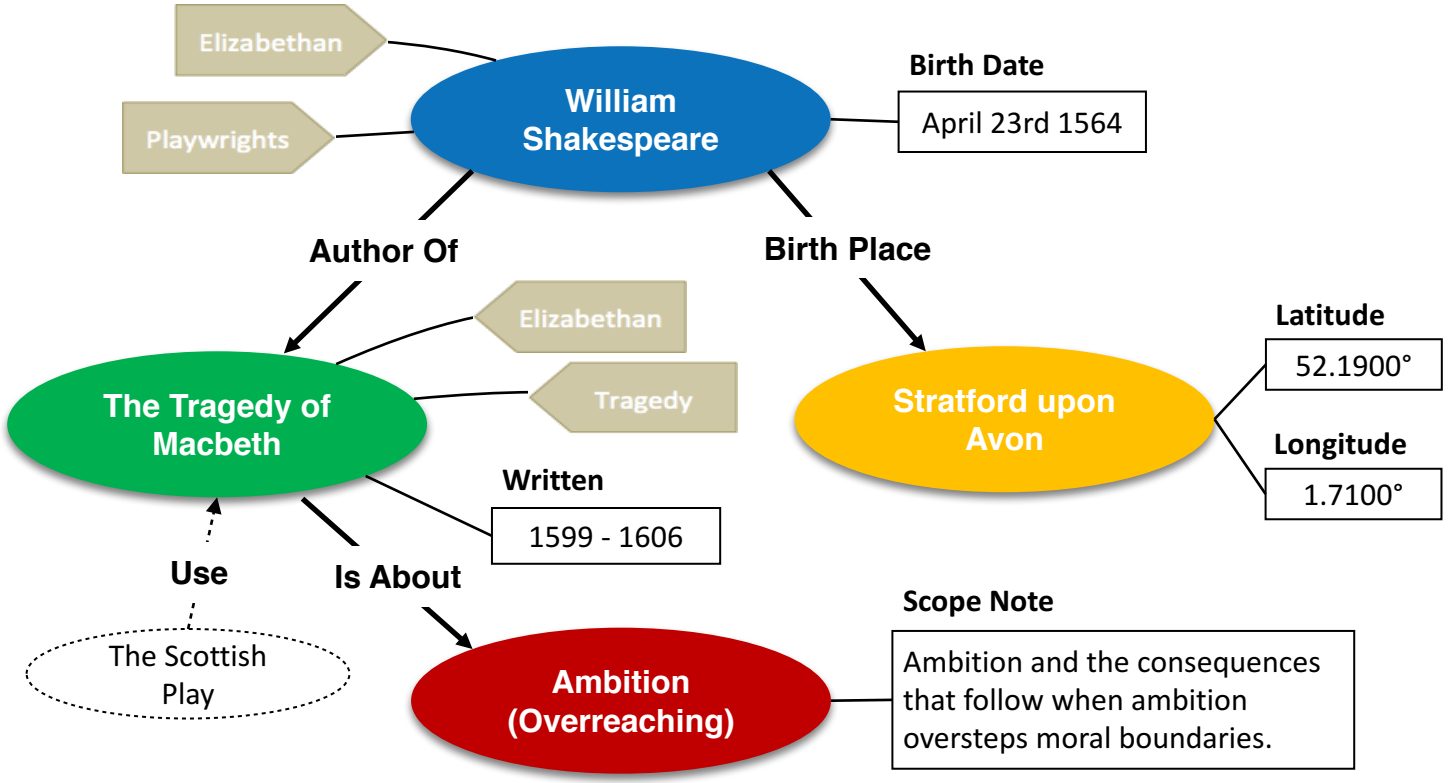
Concept Schemes
collections of like things
(i.e. with common set of attributes)

Entities [ovals]
uniquely named things and concepts
(non-preferred terms redirect to preferred terms)

Attributes [rectangles]
properties of things with literal values

Relationships [arrows]
named directed relationships between entities

separate concept schemes are used to represent collections of like-things



Subjects Scheme

Person Names Scheme

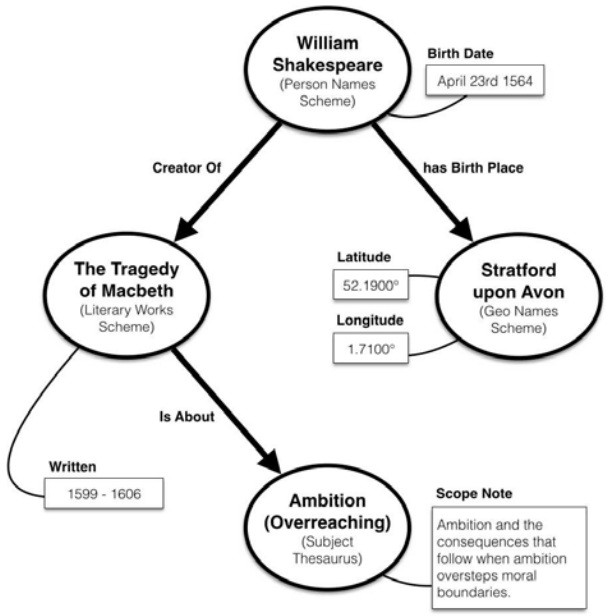
Literary Works Scheme

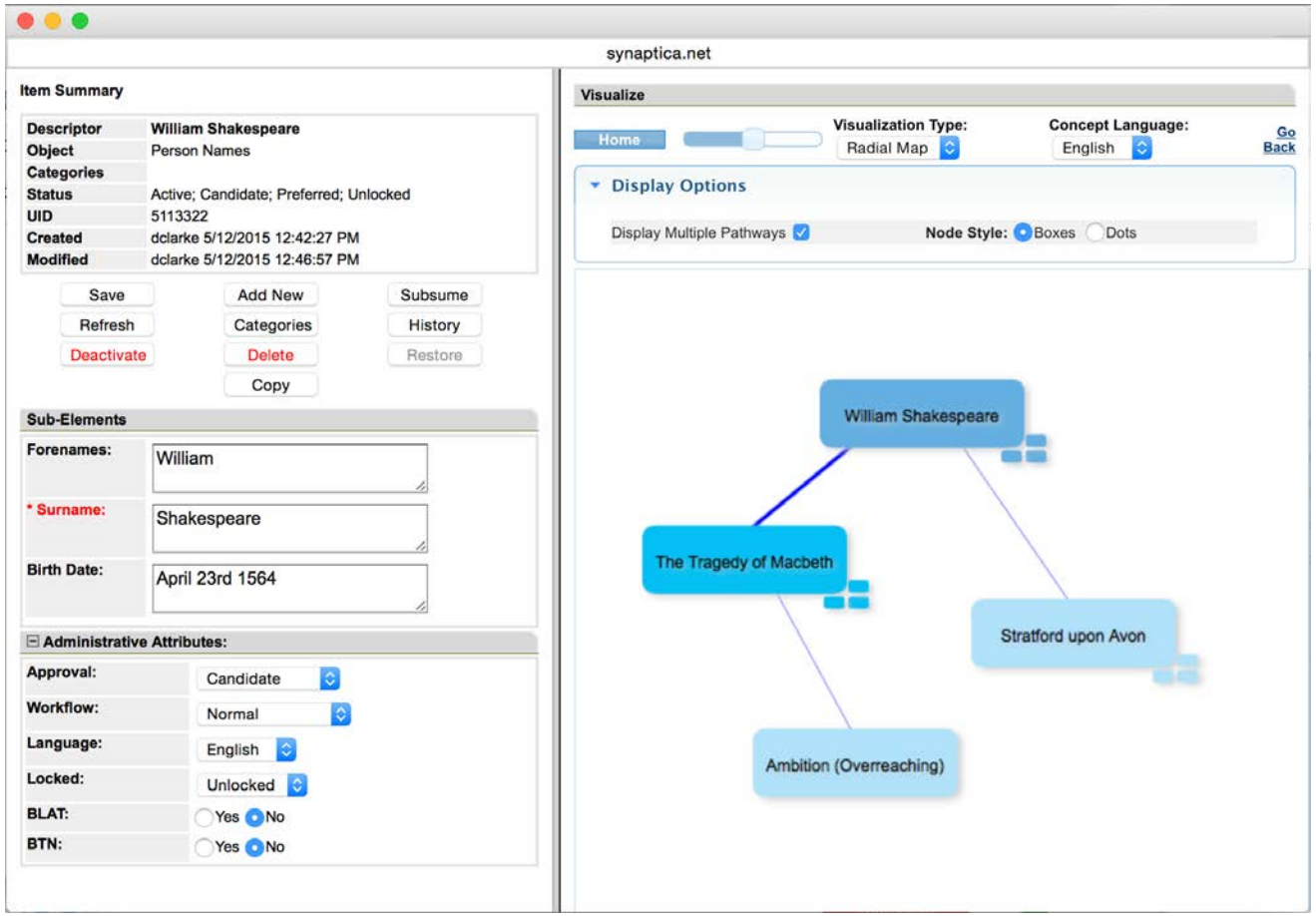
Place Names Scheme

in addition to intra-scheme relationships a multi-scheme KOS may support inter-scheme relationships that make factual assertions about the properties of an entity in relation to entities in other schemes – if formally defined this becomes an ontology

© 2017 Synaptica, LLC

Worked Example – From Data Model to KOS in 30 Minutes





- Configure four KOS schemes for storing the four fundamentally different types of thing (People, Places, Works and Subjects)
- For each scheme configure *property* fields to store the attributes about each thing (Birth Date, Latitude, Longitude, etc.)
- Configure *semantic relationship types* to support the relationships between things (Creator Of, has Birth Place, etc.)

Time to configure entire data model with multiple schemes, properties and attributes - less than 30 minutes…

© 2017 Synaptica, LLC

synaptica®

Number 2 - STARTER TAXONOMIES

- Leveraging Linked Open Data resources
- Licensing third-party taxonomies

synaptica®

Build

Buy

New mantra for taxonomy projects:

ADOPT first

ADAPT second

CREATE third

© 2017 Synaptica, LLC

When do third-party taxonomies work well

Corporate & Enterprise Taxonomies

Products & Services
Commodities
Finance
Legal & Regulatory

STEMs: Science, Technology, Engineering & Mathematics
HCLS: Health Care & Life Sciences
Cultural Heritage
News Media
Geospatial
Person Names

© 2017 Synaptica, LLC

synaptica®

Linked Open Data Taxonomy Sources

Trusted authorities

Many different subject domains

Millions of concepts

Many sources in the public domain

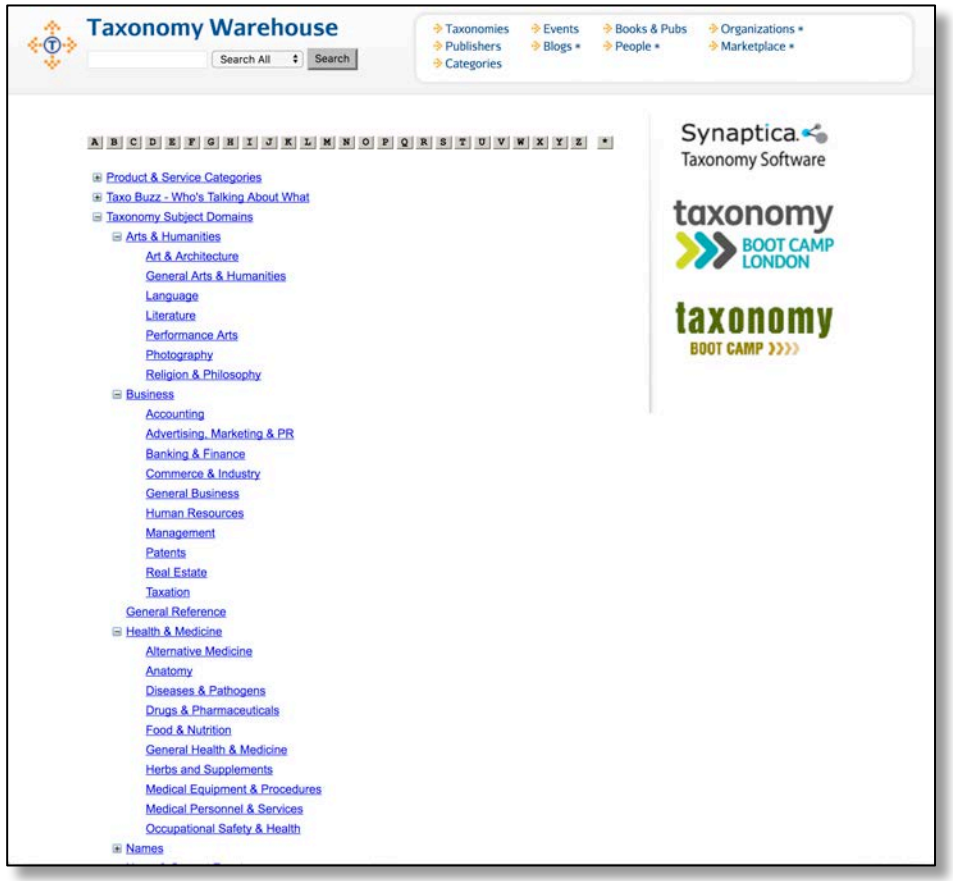
Standard electronic format

Live query and/or download

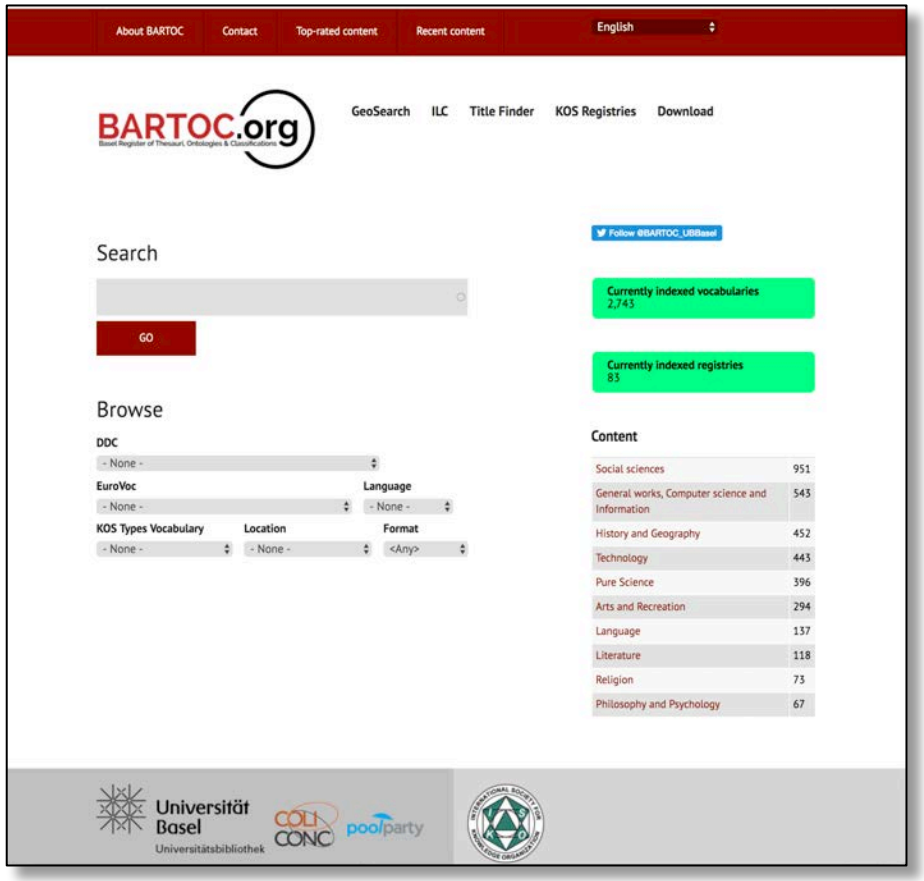
© 2017 Synaptica, LLC

Online Directories of Taxonomies

http://www.taxonomywarehouse.com

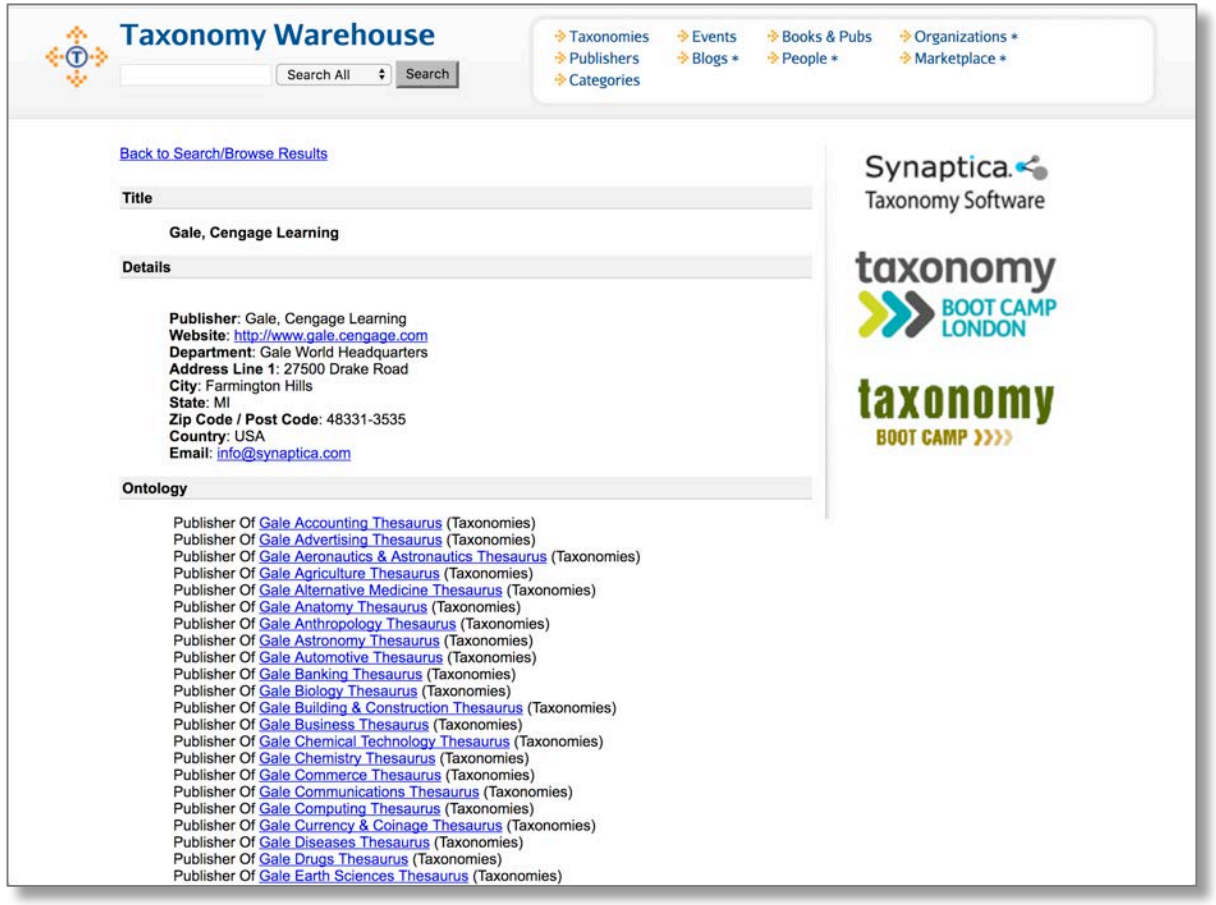


http://www.bartoc.org



© 2017 Synaptica, LLC

Commercial Example 1: Gale-Cengage Taxonomies

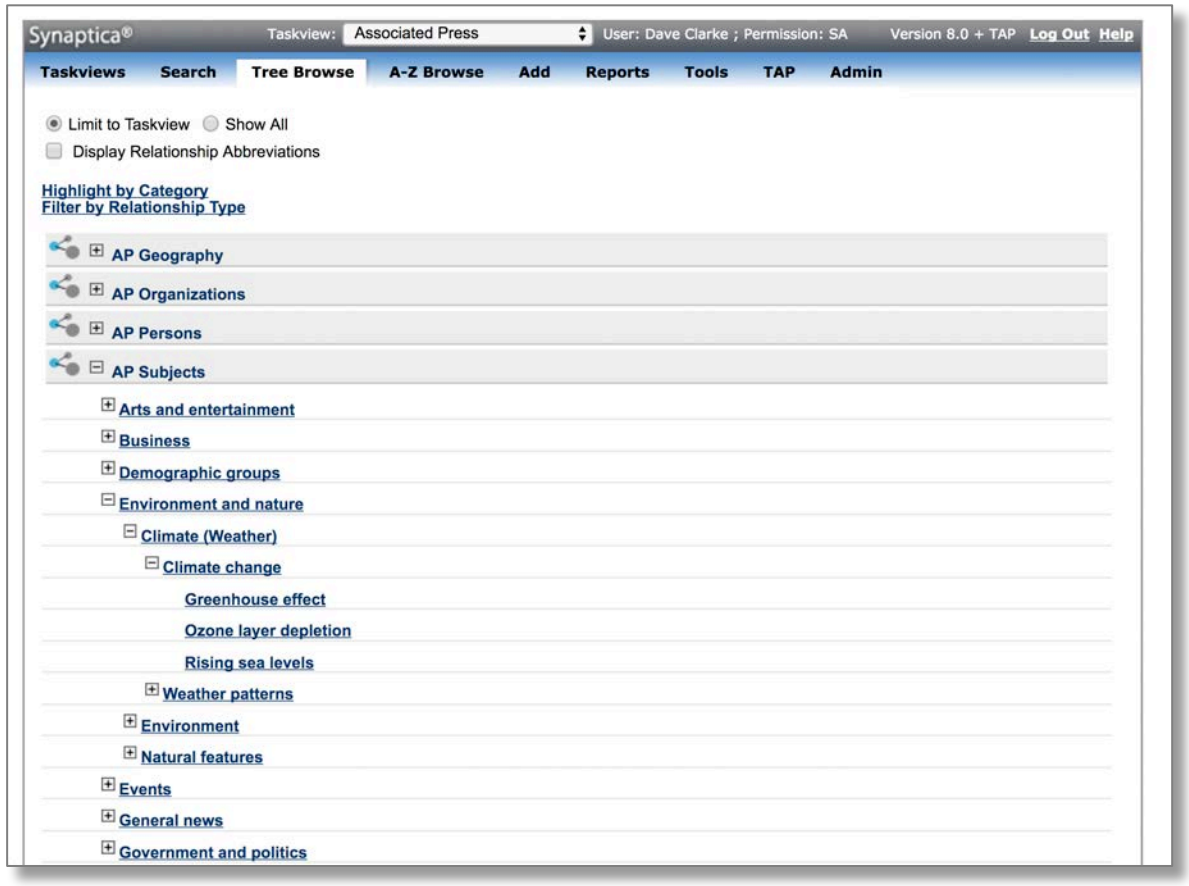


- 115,000 concepts
- Professional standards-compliant thesaurus
- Arts, Sciences & Business domains
- Available as master thesaurus
- Or in 69 topical subsets
- Demos and licensing available from Synaptica info@synaptica.com

http://www.taxonomywarehouse.com/details.aspx?vunid=89115

© 2017 Synaptica, LLC

Commercial Example 2: Associated Press Taxonomies



- 200,000 concepts
- Professional standards-compliant thesaurus
- News-related topics, People and Organization and Geospatial Names
- Demos available from Synaptica
- Licensing available from AP at the link below

https://www.ap.org/en-us/services/planning-and-media-tools/metadata-services

© 2017 Synaptica, LLC

synaptica®

Number 3 - TOOLS

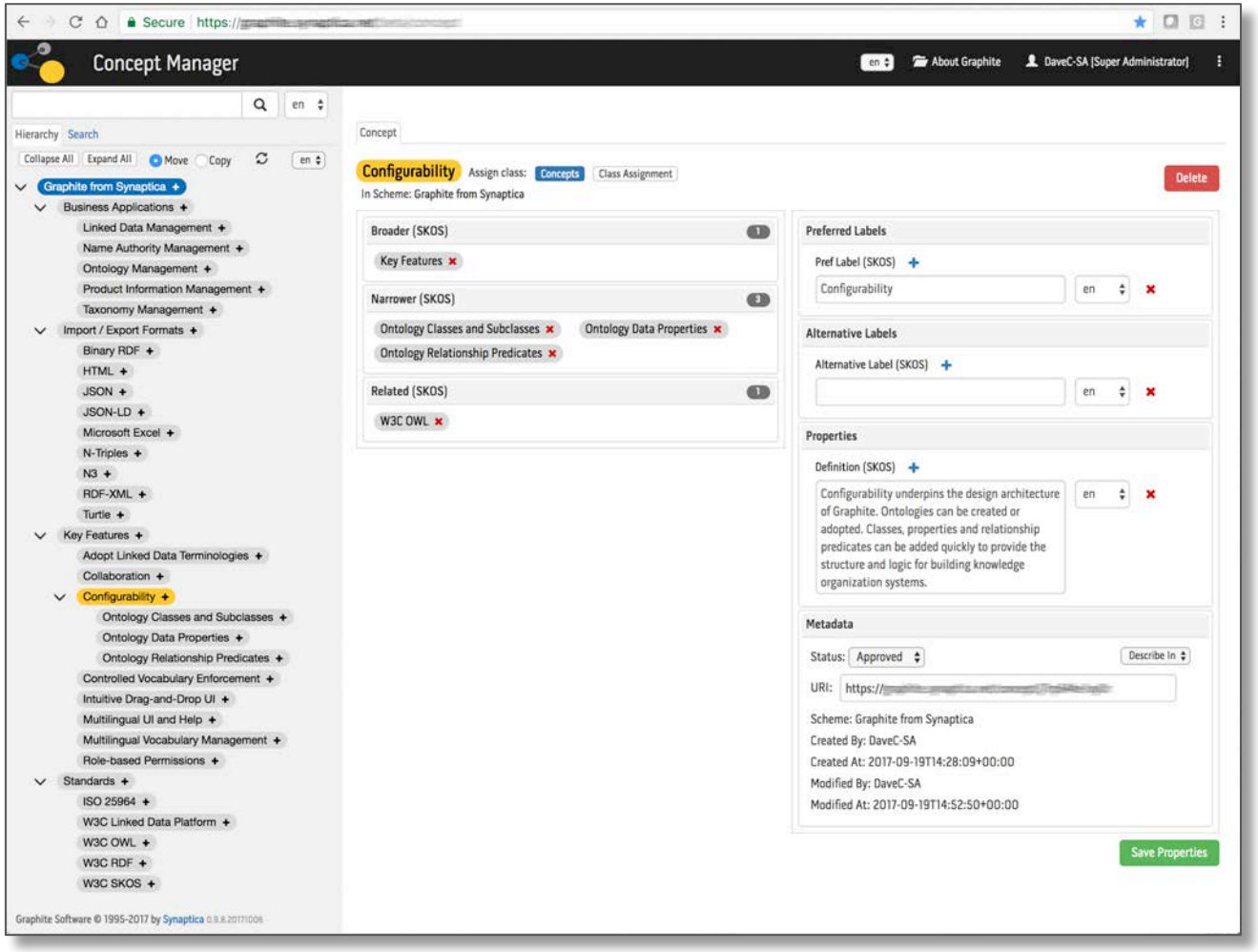
- Use a taxonomy management tool
- Share work in progress with stakeholders often

Why use a taxonomy management tool

- ✔ They enforce lexicographic standards preventing many common editorial errors and ensuring semantic integrity
- ✔ They support data exchange standards that facilitate data exchange and systems interoperability
- ✔ They allow taxonomy teams to collaborate together through role-based permissions, governance and workflow
- ✔ They enable taxonomists to share work-in-progress with other stakeholders securing buy-in throughout the development process

© 2017 Synaptica, LLC

Example tool: Graphite – you saw it first at TBCL17

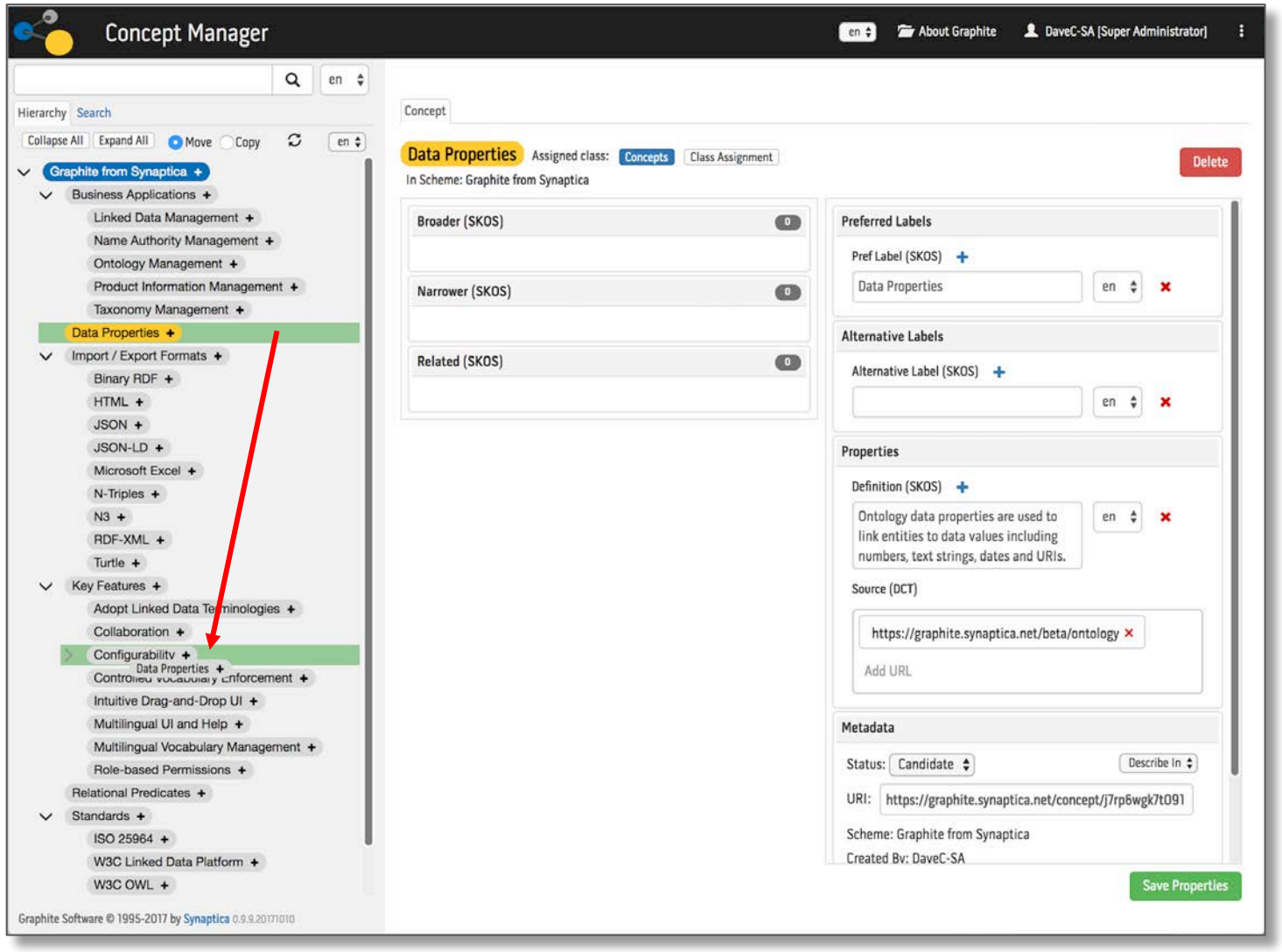


Jim Sweeney
Senior Manager Taxonomy & Ontology, Synaptica

"Graphite presents an all new graphical user experience that makes taxonomy editing child's-play, while also packing in powerful ontology management tools that can support the most sophisticated knowledge organisation systems."

© 2017 Synaptica, LLC

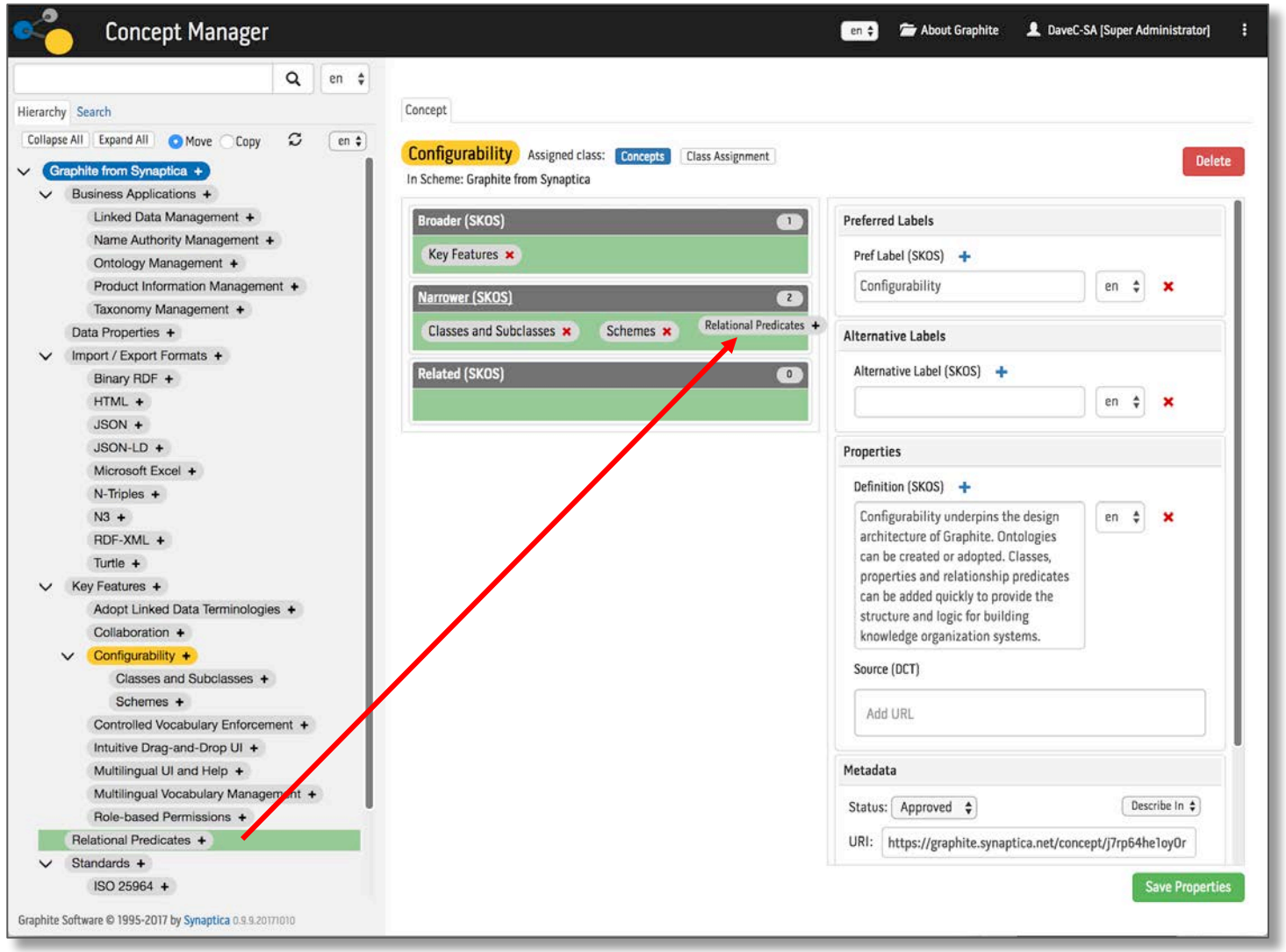
synaptica®



Drag-and-drop **within the tree structure** to create and adjust hierarchical structures

© 2017 Synaptica, LLC

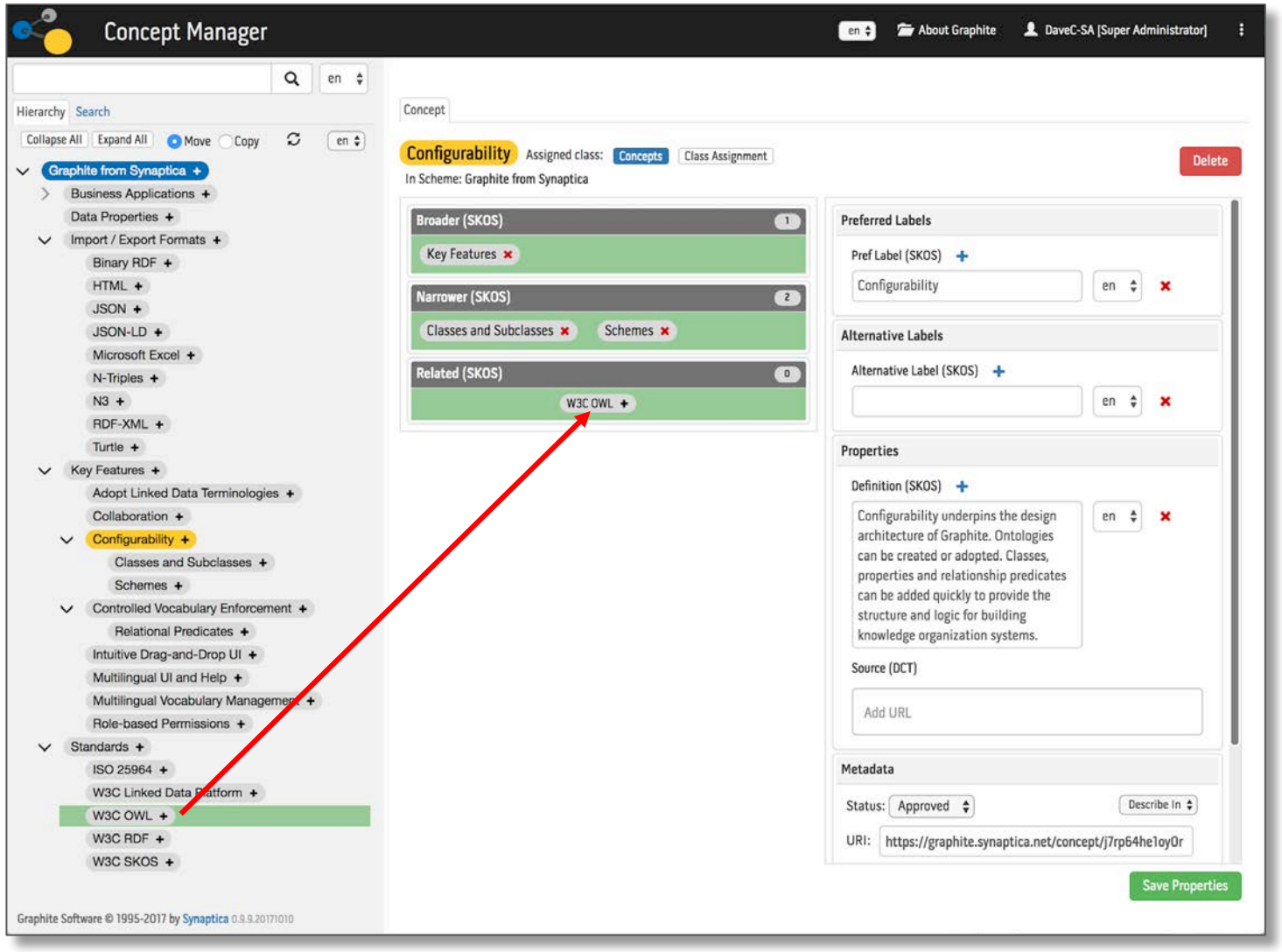
synaptica®



Drag-and-drop **into the SKOS Narrower panel** to create a new hierarchical relationship

© 2017 Synaptica, LLC

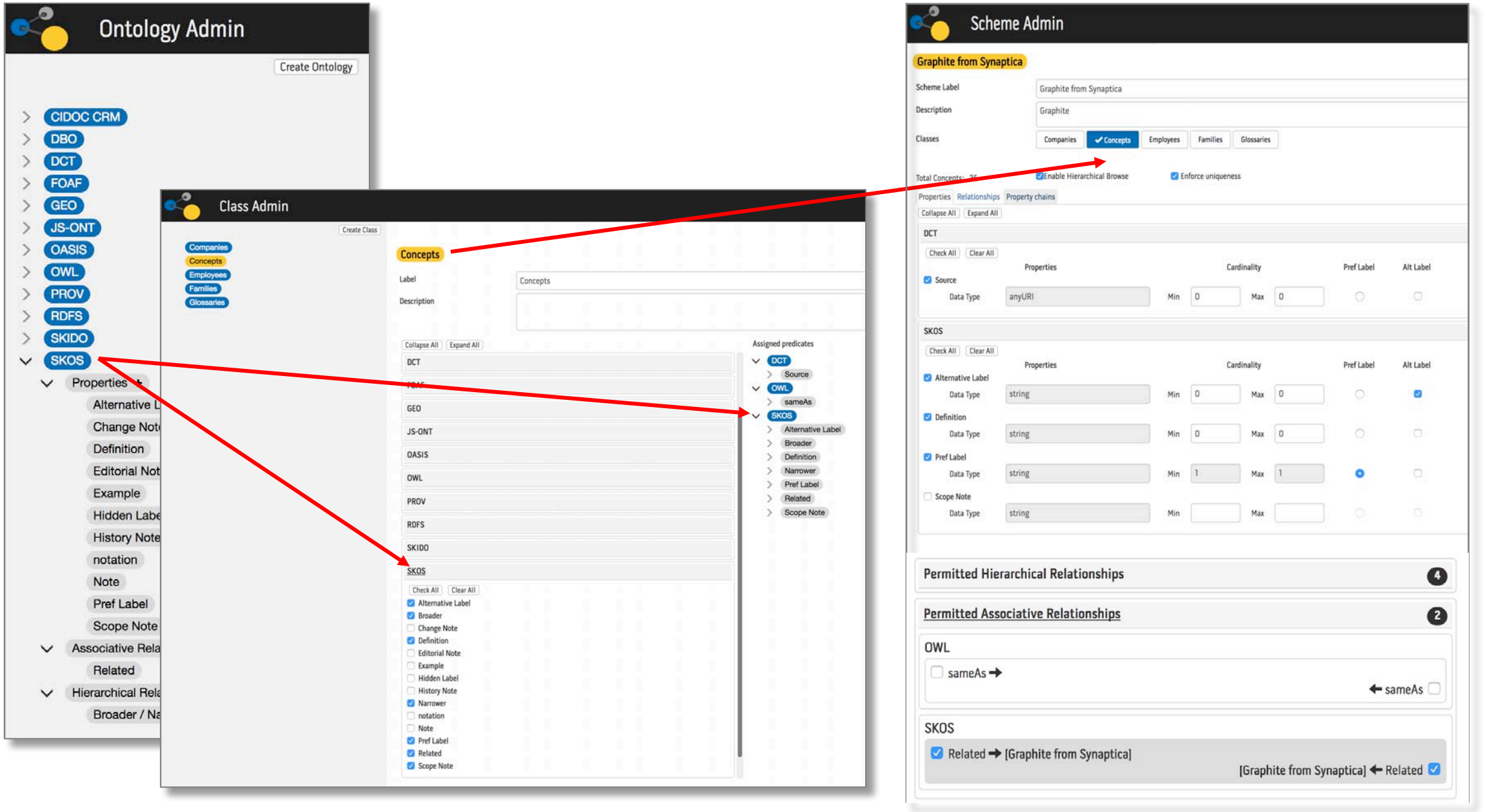
synaptica®



Drag-and-drop into the SKOS **Related** panel to create a new associative relationship

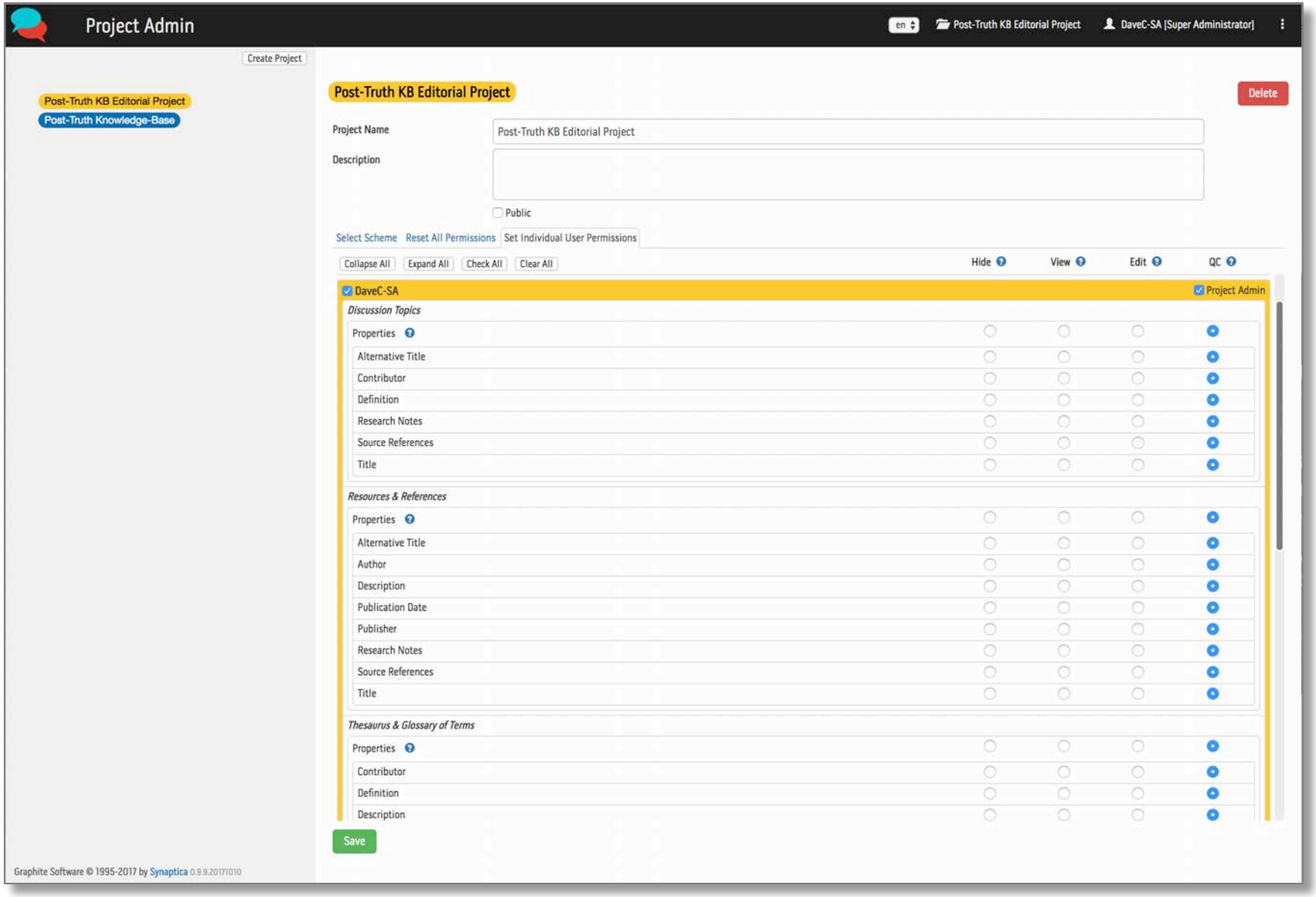
© 2017 Synaptica, LLC

Beyond SimpleKOS – working with Ontologies and classes



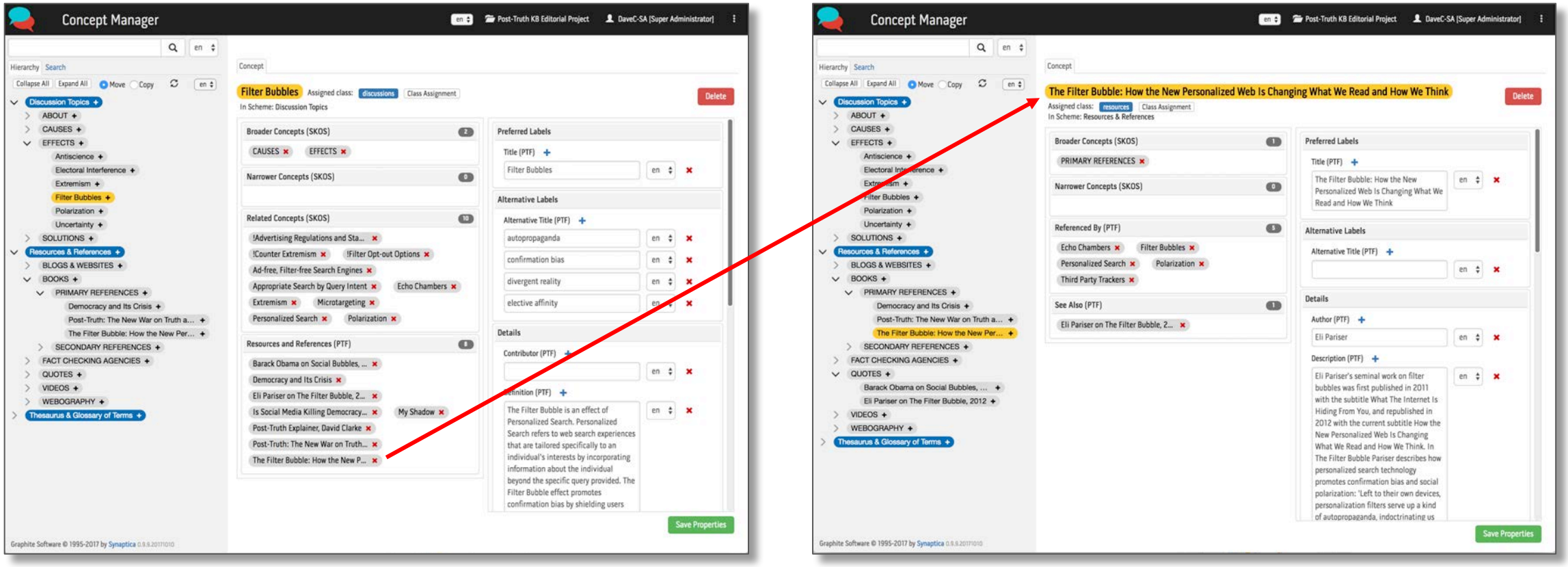
© 2017 Synaptica, LLC

Involving Stakeholders with Granualar Project Views



© 2017 Synaptica, LLC

Beyond SimpleKOS – building Ontologies and classes



Ontology *Classes* support *Knowledge Organizations Systems* that contain resource types with different *Properties* such as taxonomy **Concepts** and published **Resources**

© 2017 Synaptica, LLC

#tbcl17 #SynapticaGraphite Competition

To enter just reply to or re-tweet this post and tell us why you love taxonomy

All prizes are completely free-of-charge cloud-hosted Graphite systems with full tech support:

1st Prize: One year system
2nd Prize: Six month system
3rd Prize: Three month system

Competition starts now and closes at the end of TBCL - 17:00 on October 18th.

Prizes will be awarded by the Synaptica team and announced on Thursday October 19th they will be based on our judgment of most **profound**, **enthusiastic** or **witty** answers.