Jump-start Your Taxonomy Project

Taxonomy Bootcamp London 2017  #tbcl17
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
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)
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.”
• 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
The Bible

An Older Testament

SKOS

Lexicographic standards

Semantic web data standards

© 2017 Synaptica, LLC  www.synaptica.com
What comes before taxonomy – knowledge audits

Activity and Asset descriptions are exported into facets in Synaptica®
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
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

<table>
<thead>
<tr>
<th>User Groups</th>
<th>Needs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxonomy curators</td>
<td>Import &amp; adopt</td>
</tr>
<tr>
<td></td>
<td>Build &amp; maintain</td>
</tr>
<tr>
<td></td>
<td>Human indexing guidelines</td>
</tr>
<tr>
<td></td>
<td>Machine indexing rules</td>
</tr>
<tr>
<td></td>
<td>Workflow &amp; governance</td>
</tr>
<tr>
<td></td>
<td>Publication cycles &amp; version control</td>
</tr>
<tr>
<td>Content curators</td>
<td>SME &amp; stakeholder review</td>
</tr>
<tr>
<td></td>
<td>Content navigation</td>
</tr>
<tr>
<td>Human &amp; machine-aided indexers</td>
<td>Search &amp; browse</td>
</tr>
<tr>
<td></td>
<td>Access guidelines &amp; rules</td>
</tr>
<tr>
<td></td>
<td>Submit candidates</td>
</tr>
<tr>
<td>Search &amp; discovery</td>
<td>Resolve ambiguous concepts</td>
</tr>
<tr>
<td></td>
<td>Map user query language to indexing language &amp; Submit candidates</td>
</tr>
</tbody>
</table>
“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
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

   - People
   - Literature
   - Ambition
   - Subjects
   - Places

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 25964
   - Dublin Core Metadata Initiative
   - W3C SKOS
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

- **William Shakespeare** (Person Names Scheme)
  - Creator of
  - Has Birth Place
  - Birth Date: April 23rd 1564
  - Latitude: 52.1900°
  - Longitude: 1.7100°

- **The Tragedy of Macbeth** (Literary Works Scheme)
  - Written: 1599 - 1606
  - Is About
  - 1599 - 1606

- **Stratford upon Avon** (Geospatial Names Scheme)
  - Is About
  - Latitude: 52.1900°
  - Longitude: 1.7100°

- **Ambition** (Overreaching)
  - Is About
  - (Subject Thesaurus)
    - Scope Note
      - Ambition and the consequences that follow when ambition oversteps moral boundaries.

© 2017 Synaptica, LLC  www.synaptica.com
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

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
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...
Number 2 - STARTER TAXONOMIES

- Leveraging Linked Open Data resources
- Licensing third-party taxonomies
New mantra for taxonomy projects:

ADOPT first
ADAPT second
CREATE third
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  www.synaptica.com
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

Jump-start your taxonomy project
Online Directories of Taxonomies

http://www.taxonomywarehouse.com

http://www.bartoc.org
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

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

Number 3 - TOOLS

- Use a taxonomy management tool
- Share work in progress with stakeholders often
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.
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.”
Drag-and-drop **within the tree structure** to create and adjust hierarchical structures.
Drag-and-drop into the SKOS Narrower panel to create a new hierarchical relationship.
Drag-and-drop into the SKOS Related panel to create a new associative relationship
Beyond SimpleKOS – working with Ontologies and classes
Involving Stakeholders with Granular Project Views
Ontology Classes support Knowledge Organizations Systems that contain resource types with different Properties such as taxonomy Concepts and published Resources
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.