



SLA 2019 Annual Conference

June 14 – June 18, 2019

Huntington Convention Center of Cleveland

Cleveland, Ohio, USA

www.sla.org

#SLA2019 #SLArocks #AlwaysSLA



“Taxonomy Tools and Tool Evaluation”

SLA Taxonomy Division

Agenda and Presenters

Overview of Taxonomy Tools

- Presented by Heather Hedden
Senior Vocabulary Editor, Gale, A Cengage Company
Author of The Accidental Taxonomist (Information Today, 2016)

Evaluating Taxonomy Tools

- Presented by Marti Heyman
Executive Director, Metadata Strategy and Operations, OCLC

Overview of Taxonomy Tools

Heather Hedden

Senior Vocabulary Editor, Gale, A Cengage Company

heather.hedden@cengage.com



What are “taxonomy tools”?

- No authoritative industry list of taxonomy software
- “Taxonomy software” can mean different things
 - ➔ • Thesaurus/ontology management software
 - Auto-categorization/auto-classification software
 - Mindmapping or concept modeling tools
 - Other software with a key taxonomy component
- Web lists are miscellaneous taxonomy tools or out-of-date
 - www.taxobank.org/content/thesauri-and-vocabulary-control-thesaurus-software
 - www.taxotips.com/resources/tools
 - www.searchtools.com/info/classifiers-tools.html
- Market is too small and specialized to be followed by industry analysts

Background

- Taxonomy, thesaurus, or ontologies - The distinctions are blurred. Most software enables the creation of a combination: taxonomy/thesaurus, thesaurus/ontology, taxonomy/thesaurus/ontology
- Excel suffices for flat term lists (such as for facets), and small hierarchical taxonomies, but not for the complexities of large taxonomies, thesauri, ontologies, or multilingual vocabularies.
- Software tools enforce/support standards, but not all the same standards: thesaurus or records management (ANSI/NISO or ISO), SKOS/RDF, OWL
- Software tools support integration with auto-classification, content management systems, SharePoint, etc.

Summary of Controlled Vocabulary Types



Term List	Synonym Ring	Authority File	Taxonomy	Thesaurus	Ontology
Ambiguity control	Synonym control	Ambiguity control Synonym control	Ambiguity control (Synonym control) Hierarchical relationships	Ambiguity control Synonym control Hierarchical relationship Associative relationships	Ambiguity control (Synonym control) Semantic relationships Classes (Linked Data)

Types of software for vocabulary management

- Spreadsheet software (Excel)
- ➔ • Dedicated thesaurus/ontology management software
- Taxonomy creation & editing module of a content management, document management, digital asset management, collaborative software (SharePoint)
- Taxonomy creation & editing module of auto-classification (automated indexing) software
- Vertical market software for creating classification structures
- Proprietary programs developed in-house in organizations with large or core taxonomy management needs

Types of software used by taxonomists

An internally developed taxonomy/thesaurus management system	25.5%	36
Commercial dedicated thesaurus/taxonomy/ontology management software	22.7%	32
Commercial software, of which taxonomy management is a feature, module or component	12.1%	17
Open-source ontology/taxonomy management software	9.2%	13
Other commercial software that is not intended for taxonomies (such as a word processor, spreadsheet, or database management)	30.5%	43

Thesaurus/ontology software basic features

- Maintain terms/concepts and their relationships
 - As reciprocals
 - When renaming, merging, subsuming, or deleting terms
- Support controlled variants/NPTs/synonyms/alternative labels
- Support notes/definitions and other attributes for terms
- Manage categories or classes for terms/concepts
- Manage candidate and approved terms; term creation and update dates
- Generate reports in various display formats
- Export data in format for importing into a content indexing/search/retrieval system: CSV, Excel, HTML, XML (ZThes, RDF, SKOS, and OWL)

Software feature of enforcing thesaurus standards

Thesaurus standards: ANSI/NISO Z39.19-2005 (R2010) or ISO 25964

- Preferred terms (preferred labels) must be unique; no duplicates
- A nonpreferred term (alternative label) can point to only one preferred term (concept)
- A pair of terms (concepts) can be either hierarchically (broader/narrower) or associatively (related) linked to each other, but not both.
- Hierarchical relationship logic extends:
 - A is narrower to B, and
 - B is narrower to C,
 - C *cannot* be narrower to Term A.

Thesaurus software points of comparison

- interface design (default view) and ease of use
- multiple taxonomy display options
- term searching
- spell-checking
- speed (limited mouse clicks) for repeated term and relationship additions
- single-step new term & relationship creation
- single-step branch (term and narrower terms) moving
- drag & drop relationship adding
- user-defined (customizable) relationships
- user-defined term notes and term attributes
- bilingual or multilingual taxonomy support
- importing and exporting formats
- connectors to content SharePoint and enterprise search systems
- auto-categorization add-on module
- support for linked data



Commercial dedicated thesaurus/ontology management software

Includes:

- MultiTes Pro
- Synaptica KMS, Synaptica Graphite
- Data Harmony Thesaurus Master
- Semaphore
- Mondeca Intelligent Topic Manager (ITM)
- PoolParty
- TopBraid Enterprise Data Governance (EDG) Vocabulary Management

MultiTes Pro

Multisystems (Miami, FL)

www.multites.com

- Single product independent vendor since 1983.
- Windows single user \$295 (multi-user and enterprise packages also available)
- Web/cloud-based option: \$4950/year per thesaurus for 20 accounts
- Thesaurus model ANSI/NISO Z39.19 based
- Supports user-defined relationships, classes, and notes; multilingual thesauri
- Imports delimited text. Outputs text, HTML, XML, SKOS/RDF, and CSV
- Add-on products: web development kit, enterprise development kit
- Free 1-month downloadable trial and online video tutorials

Search Bird's-eye view prints
 [ESC] Clear [Enter] Search

Term	Status	Flag	Rel
Billheads	A		BT
Bills of fare	A		USE
Bird's-eye view prints	A		BT
Bird's-eye views	A		BT,
Bird feeders	A		BT
Birth certificates	A		BT,
Birthday cards	A		BT
Black-and-white prints	A		USE
Black & white photographs	A		BT
Black light works	A		USE
Black line prints	A		USE
Blind embossed prints	A		USE
Blind stamps	A		BT,
Block prints	A		USE
Blocks, Printing	A		USE
Blue line prints	A		USE
Blueprints	A		BT
Boards, Game	A		USE
Boards, Instruments	A		USE
Booby traps	A		BT
Book & magazine posters	A		BT,
Book burning	A		RT

Bird's-eye views

Clipboard Edit Print Window

Record Details Multilevel Hierarchy 2-way Hierarchy Edit

Bird's-eye views

- + General
- Classification
- Notes
 - + CN: For the class of engravings, lithographs, and
 - + FUN: Formerly TGM I term & TGM II term.
 - + SN: Graphic representations of scenes portrayed as if viewed
- Relationships
 - UF: Balloon views
 - UF: Views, Bird's-eye
 - BT: Views
 - NT: Bird's-eye view prints
 - RT: Aerial views
 - RT: Maps
 - RT: Panoramic views
 - RT: Projections

Buttons: Add Rels..., +Rel Wiz..., Delete..., Personal Note, Web search, Close All, Close

Stay on top

MultiTes Pro

Synaptica

Synaptica Software LLC (Franktown, CO)

www.synaptica.com

Synaptica KMS (Knowledge Management System) – thesaurus model (since 1995)

Synaptica Graphite – SKOS ontology model on a linked data graph database (since 2018)

- Web browser-access, inside the firewall or hosted.
- Supports user-defined relationships, classes, and notes; multilingual vocabularies
- Features drag-and-drop editing, automatic term mapping
- Imports: CSV, Excel, XML (Zthes, RDF SKOS, RDF OWL). Exports also HTML, Word.
- Related add-on products: Indexing Management System (IMS), Text Analytics Platform (TAP), Image Annotation & Indexing, Linked Data Manager, SharePoint connector
- Online video tutorial for editing terms in Synaptica KMS

[Hierarchy](#) [Search](#) [Results](#)

- Training_Thes +**
 - Botany +**
 - Applied botany +**
 - Agriculture +
 - Forestry +**
 - Afforestation +
 - Reforestation +
 - Research Botany +
 - Exobiology +
 - Medicine +**
 - Dentistry +
 - Family practice +
 - Veterinary medicine +
 - Zoology +**
 - Applied zoology +**
 - Livestock farming +
 - Veterinary medicine +

[Training_Thes](#) [Forestry \(Training_Thes\)](#)

Forestry

Save Refresh Copy
Categories History Subsume
Deactivate Delete Restore
Linked Data

Attributes

Descriptor:	Forestry
Scope Note:	Forestry scope note
History Note:	

Administrative Attributes:

Approval:	Candidate ▾
Workflow:	Normal ▾
Language:	English ▾
Locked:	Unlocked ▾
BLAT:	<input type="radio"/> Yes <input checked="" type="radio"/> No
BTN:	<input type="radio"/> Yes <input checked="" type="radio"/> No

Term Metadata:

Relationships

[Flat View](#) [Add/Edit](#) [Visualize](#)

Limit to Project Show All

Forestry

Top Level Parents 1
 TT Botany (Training_Thes)

Parents 1
 BT Applied botany (Training_Thes)

Children 2 (2)
 NT Afforestation (Training_Thes)
 NT Reforestation (Training_Thes)

Associations 0
Variants 0

Synaptica KMS

Hierarchy Search Results

- Training_Thes +
 - Botany +
 - Applied botany +
 - Agriculture +
 - Forestry +
 - Afforestation +
 - Reforestation +
 - Research Botany +
 - Exobiology +
 - Medicine +
 - Dentistry +
 - Family practice +
 - Veterinary medicine +
 - Zoology +
 - Applied zoology +
 - Livestock farming +
 - Veterinary medicine +

Training_Thes Applied botany (Training_Thes)

Visualize

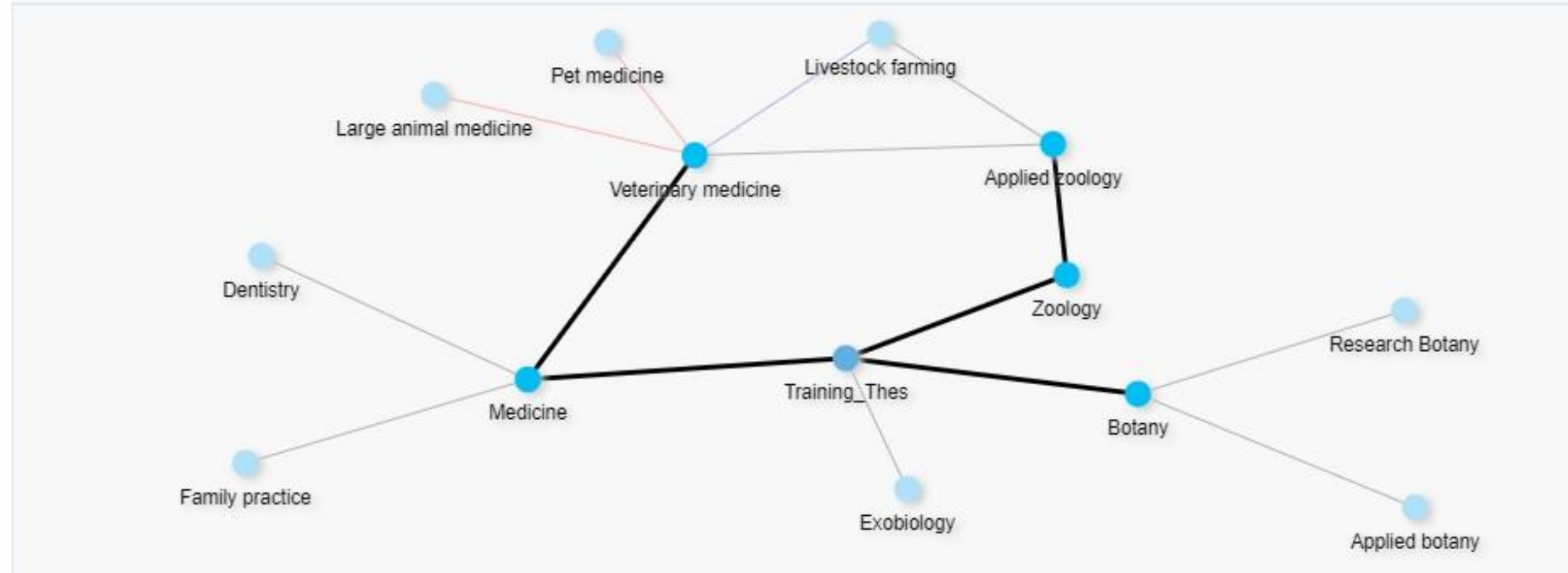
Home



Visualization Type: Radial Map

Concept Language: English

Display Options





Search bar with a magnifying glass icon and a language dropdown set to 'en'.

Hierarchy Collections Advanced Search Search Results

Buttons for 'Collapse All', 'Expand All', 'Move', 'Copy', and a refresh icon, along with a language dropdown set to 'en'.

- Test Scheme +
 - Botany +
 - Applied Botany +
 - Forestry +
 - Research Botany +
 - Exobiology +
 - Medicine +
 - Dentistry +
 - Family Practice +
 - Veterinary Medicine +
 - Zoology +

Concept Export Import Audit Trail

Forestry

Delete

Assigned class: SKOS Class Class Assignment Collection Assignment

In Scheme: Test Scheme

Broader Concepts (SKOS) 1

Applied Botany ✖

Narrower Concepts (SKOS) 2

Afforestation ✖ Reforestation ✖

Related Concepts (SKOS) 0

Preferred Labels ?

Pref Label (SKOS) +

Forestry en ✖

Alternative Labels

Alternative Label (SKOS) +

Forest management en ✖

Properties ?

notation (SKOS) +

en ✖

Note (SKOS) +

en ✖

Synaptica Graphite

Save Properties

Data Harmony Thesaurus Master

Access Innovations (Albuquerque, NM)

www.dataharmony.com

- Commercial software (originally used for indexing in-house) offered since 1998
- Multi-platform java-based (used on Windows, Mac, Solaris, Linux). Client software allows remote access. Also a web-hosted version.
- Thesaurus model ANSI/NISO Z39.19 based
- Separately or combined with M.A.I. (Machine Aided Indexer) as MAIstro.
- Related products: XIS (XML Intranet System), Inline Tagging, Search Harmony
- API connectors for SharePoint, MarkLogic, OpenText, Oracle, SAP
- Access Innovations also offers taxonomy creation services.

File Edit View Help

- Business and finance
- Community life
- Government and politics
- Health and wellness
 - Health sciences
 - Allied health sciences
 - Alternative medicine
 - Food engineering
 - Health care
 - Health care facilities**
 - Health concerns
 - Health supplies
 - Health support services
 - Medical conditions and diseases
 - Medical patients
 - Medical prostheses
 - Medical research
 - Medicine
 - Nursing
 - Patient care
 - Physical rehabilitation
 - Public health
 - Veterinary sciences
 - Wellness
- Locations
- News
- Science and technology
- Sports

Term: Health care facilities

Broader Term

+

-

V

Health sciences

Narrower Term

+

-

V

Hospices

Hospitals

Mental health facilities

Nursing and retirement homes

Rehabilitation centers

Status

Candidate

Accepted

Related Term

+

-

V

Non-Preferred Term

+

-

V

Scope Note

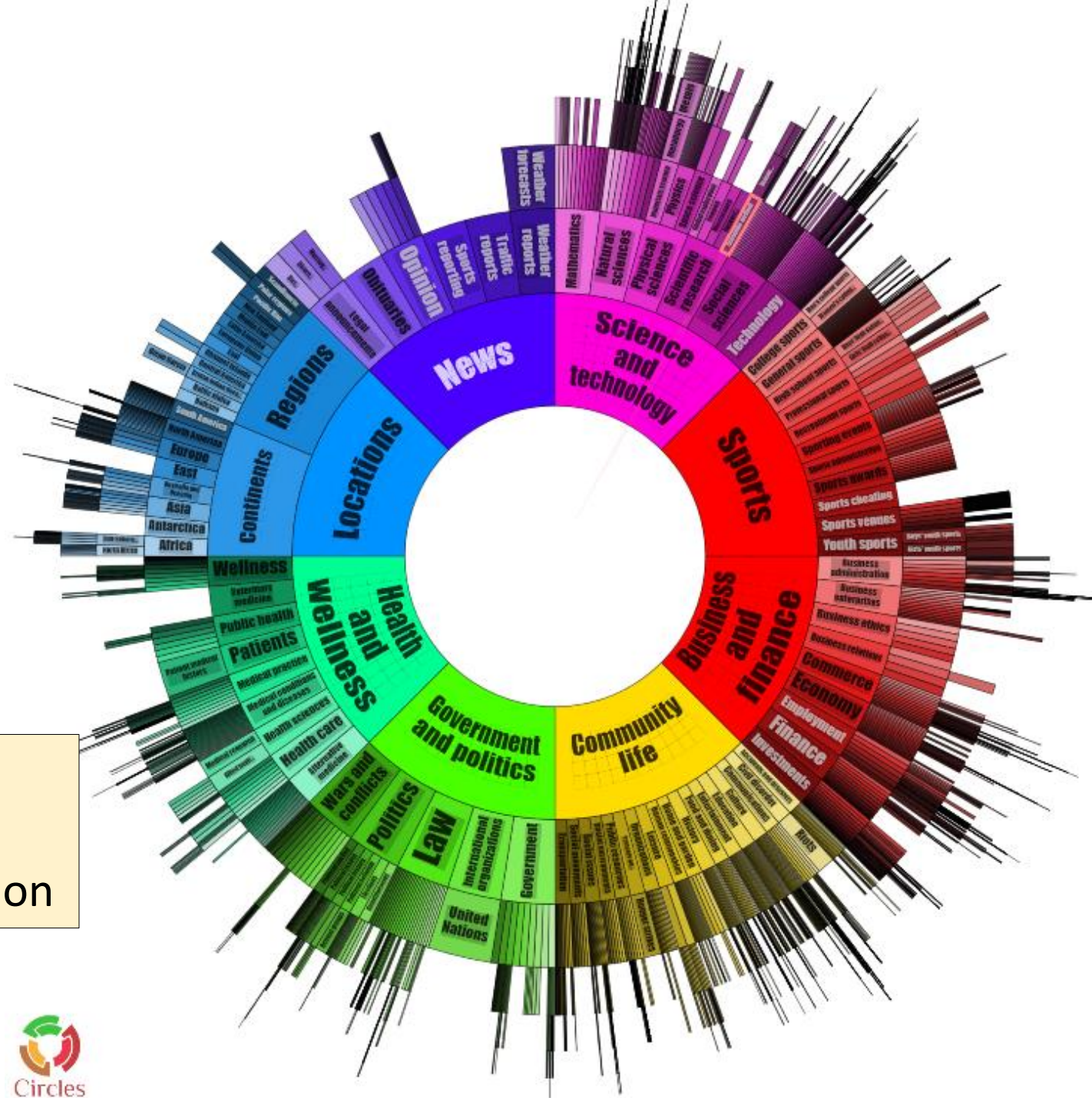
Editorial Note

Facet+

NNID

History+

Data
Harmony
Thesaurus
Master



Data Harmony
Thesaurus Master
taxonomy visualization



Semaphore

Smartlogic Semaphore Ltd. (London, UK)

www.smartlogic.com

- Introduced in 2006.
- Supports SKOS, RDF ontology standard, and ISO 25964 thesaurus standard
- Imports/export CSV, XML (RDF SKOS, Turtle, N Triple), SQL databases, and MultiTes files
- Related products: Classification Server for automated classification; Ontology Service for a navigation system
- Download free 30-day trial: <https://trial.smartlogic.com/S4Trials/>
(Sign in with LinkedIn.)

Edit Model

Edit History

Search...

- Business and industry
 - Business people
 - Business practice and regulation
 - Business advice services
 - Business development
 - Business management
 - Business studies
 - Business travel
 - e-Commerce
 - Internet shopping
 - Ethical business practices
 - Fair trading
 - Health and safety at work
 - Insolvency
 - Marketing
 - Regulation and deregulation
 - Business sectors
 - Companies
 - Consumer affairs
 - Energy and fuel
 - International trade
 - Economics and finance
 - Education and skills
 - Employment, jobs and careers

Smartlogic Semaphore

e-Commerce

Concept Class

No user class defined

Preferred Labels

Create a preferred label

e-Commerce en

Alternative Labels

Create an alternative label

alternative label > Electronic Commerce en

Metadata

Related Concepts 5

Select a related concept

has related > Communications industries

has related > Data security

has related > Information and communication technology

has related > Internet

has related > Retail trade

Broader Concepts 1

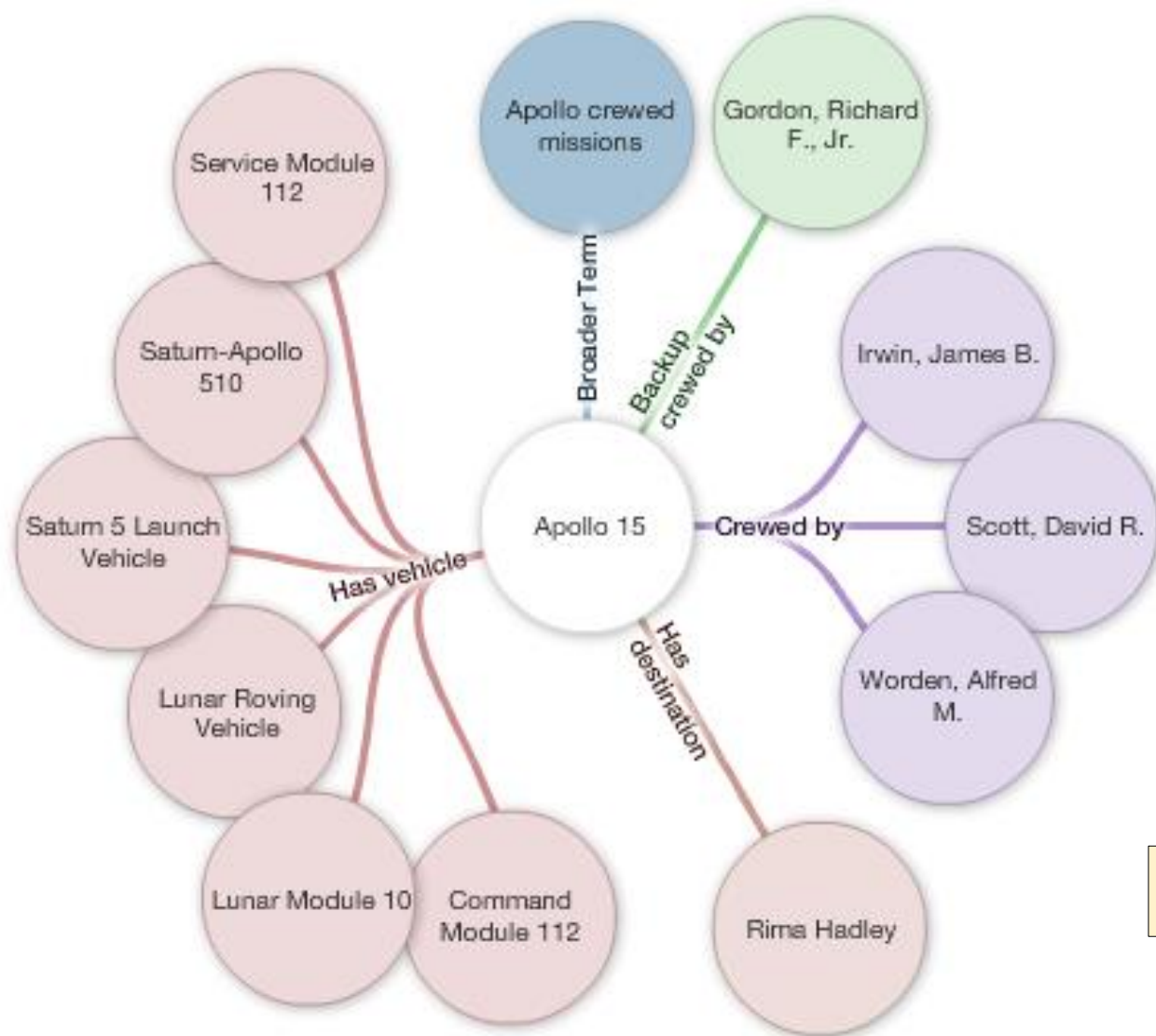
Select a broader concept

Business practice and regulation

Narrower Concepts 1

Select a narrower concept

Internet shopping

[> back to overview](#)

Term:

Apollo 15

Class:

Missions

Synonyms:

Apollo XV, Endeavour (call sign), Falcon, J mission, Lunar landing

Term information:

Search Results:

No results found

Semaphore ontology visualization

Mondeca Intelligent Topic Manager (ITM)

Mondeca S.A. (Paris, France)

<https://mondeca.com/itm>

- Introduced in 2008
- Supports SKOS vocabularies and OWL-standard ontologies
- Linked data feature
- SharePoint term store connector
- Visualization of hierarchies and relationships
- Exports to Excel, XML, RDF, SKOS, and Topic Maps



GROUPS

Home > Descriptors > UNESCO Thesaurus > Information and communication > Information management

+ Add group	+ Add group	+ Add group
Countries and country groupings M >	Communication research and pol... M >	
Culture M >	Documentary information proce... M >	
Education M >	Documentary information syste... M >	
Information and communication M >	Information industry M >	
Politics, law and economics M >	Information management M >	
Science M >	Information sciences M >	
Social and human sciences M >	Information sources M >	
	Information technology (hardwa... M >	
	Information technology (software) M >	

INFORMATION MANAG... [Settings] [Edit]

- Home
- + archive management
- information/library administration
 - + information/library management
 - + information/library legislation
 - + information systems evaluation
- + library collections
- + information/library planning
- + information/library policy
- + information/library profession
- + information users
- + library use promotion

ATTRIBUTES

Name	es información/biblioteca, gestión
	fr information/bibliothèque, gestion
	en information/library management
URI	http://skos.um.es/unescothes/C02010
in group	<ul style="list-style-type: none">▶ UNESCO Thesaurus▶ Information management▶ Information and communication
used for	fr gestion de la bibliothèque
	en library management
	es gestión de las bibliotecas

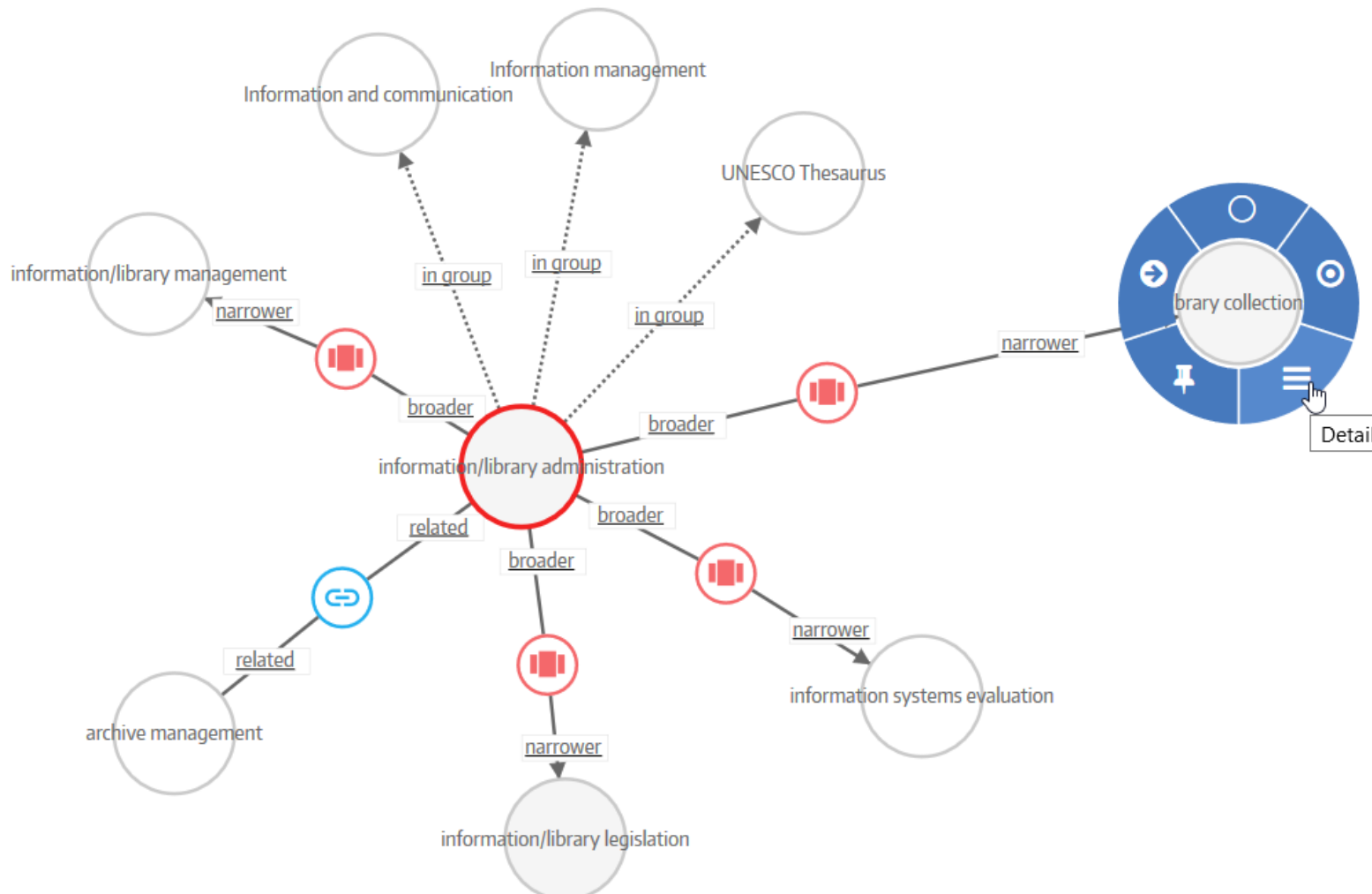
ASSOCIATIONS

UNESCO CONCEPT INFORMATION/LIBRARY ADMINISTRATION

Edit Delete Actions

information/library administration

Close



library collections

ID
162243

Name

- es colecciones de biblioteca
- en library collections
- fr collection de bibliothèque

used for

- es colección de libros
- es fondo documental
- en book collections
- fr collection de livres
- fr fonds de bibliothèque
- fr fonds de livres
- es fondo de libros
- fr fonds documentaire

Details

PoolParty

Semantic Web Company (Vienna, Austria)

www.poolparty.biz

- Introduced in 2009.
- Built on W3C Semantic Web standards: SKOS, RDF, OWL, SPARQL
- Installed server or web-hosted options
- Can link domain-specific thesauri to Linked Open Data
- Import/export formats: Excel, N3, N-Quads, Trix, Binary-RDF, MultiTes, RDF/XML, Turtle, N-Triples, RDF/JSON, Trig, JSON-LD, and Zthes
- Add on modules: Concept Tagging, Linked Data Management, Semantic Search, Text Mining & Entity Extraction, Classification, Data Analytics & Visualization
- Connectors for SharePoint, Drupal, WordPress, Confluence, Alfresco, FontoXML
- Download free 30-day trial: <http://www.poolparty.biz/test-demo/thesaurus-server-entity-extractor>



en

Last automatic snapshot at 04:58 - 28.01.2015



- Food and Recipes
 - Recipes (2)
 - Dishes (14)
 - Appetizers (1)**
 - Dips (0)
 - Baked goods and breakfast (1)
 - Desserts (0)
 - Dressings and Sauces (1)
 - Egg Dishes (1)
 - Meat and Poultry (4)
 - Pasta, Rice, Potatoes (0)
 - Pizza and Savory Baking (1)
 - Salads (0)
 - Seafood (2)
 - Smoothies and Parfaits (0)
 - Soups and Stews (0)
 - Vegetable Dishes (0)
 - Ingredients (6)
 - Lists
 - Collections

Appetizers

<http://advanced.poolparty.biz/FoodandRecipes/2>

+ Add to Collection x Delete Concept

Pool Party

- Details
- Notes
- Documents
- Linked Data
- Triples
- Visualization
- Quality Report
- History

SKOS

Relations - Add to Blacklist abc Linguistics

Broader Concepts

[Dishes](#)



Preferred Label

Appetizers

en

Narrower Concepts

[Dips](#)



Alternative Labels

Starters

en

Related Concepts



Hidden Labels



en

Scope Notes

Dishes served as appetizers

en

- Contemporary Classics (30)
- New Era Drinks (16)
- The Unforgettables (30)
 - Alexander (0)
 - Americano (0)
 - Angel Face (0)
 - Aviation (0)
 - Bacardi cocktail (0)
 - Between the Sheets (0)
 - Casino (0)
 - Clover Club Cocktail (0)
 - Daiquiri (0)
 - Derby (0)
 - Gin Fizz (0)
 - John Collins (0)
 - Manhattan (0)**
 - Martini (0)
 - Mary Pickford (0)
 - Negroni (0)
 - Old Fashioned (0)
 - Paradise (0)
 - Planter's Punch (0)
 - Porto flip (0)
 - Ramos Gin Fizz (0)
 - Rusty Nail (0)
 - Sazerac (0)
 - Screwdriver (0)
 - Sidecar (0)
 - Stinger (0)
 - The Monkey Gland (0)
 - Tuxedo (0)
 - Whiskey sour (0)
 - White Lady (0)
- Mixed drink (8)
- Sour (12)
- Wine cocktail (8)
- Garnish and other ingredients (2)
- Glassware (3)
- Sauces (2)

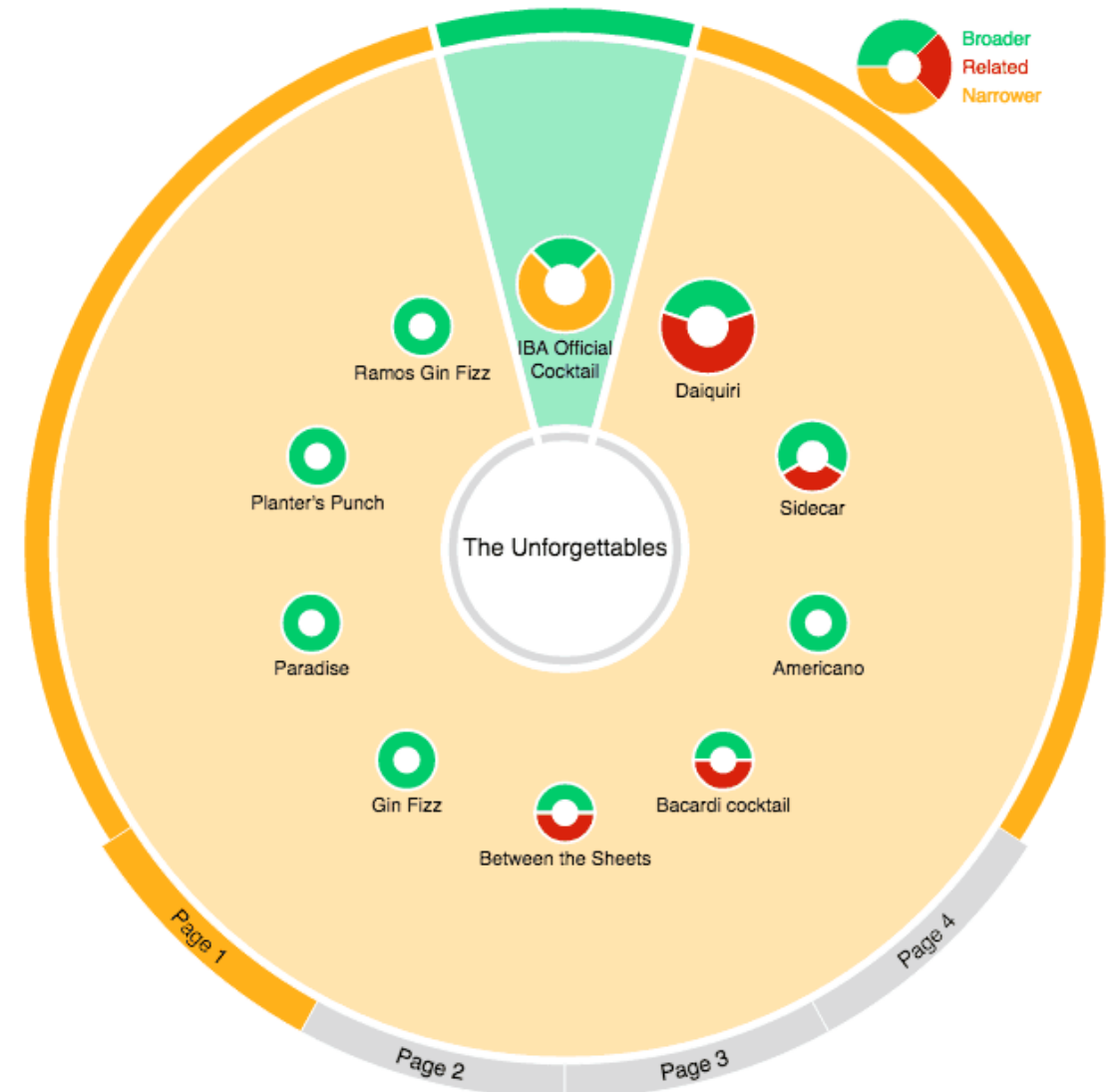
Manhattan

<http://vocabulary.semantic-web.at/cocktails/#632b16d-1490-4587-a637-fb393d7b1e7f>

+ Add to Collection X Delete Concept

Cocktail

Details Notes Documents Linked Data Triples Visualization Quality Report History



Pool Party
concept
visualization

TopBraid Enterprise Data Governance (EDG) Vocabulary Management

TopQuadrant Inc. (Raleigh, NC)

www.topquadrant.com/products/topbraid-edg-vocabulary-management/

Originally as Enterprise Vocabulary Net (2010), stand-alone vocabulary management tool (replacing Enterprise Vocabulary Net) or module of TopBraid Enterprise Data Governance, introduced in 2016

- Web-based access to a Linux server installation, with browser access
- Based on a graph database
- Taxonomies in SKOS or SKOS-XL; ontologies based on SHACL or OWL
- Import/export formats Excel/CSV, XML, RDF/OWL
- Automatic creation of crosswalks between two vocabularies
- Video demos at: www.topquadrant.com/knowledge-assets/videos

Concept Hierarchy

- Peru
- Geography
 - Africa
 - Antarctica
 - Asia
 - Atlantic Ocean
 - Europe
 - Global
 - Indian Ocean
 - Latin America
 - Central America
 - South America
 - Argentina
 - Bolivia
 - Brazil
 - Chile
 - Colombia
 - Ecuador
 - French Guiana
 - Guyana
 - Paraguay
 - Peru
 - Suriname
 - Uruguay
 - Venezuela
 - North America
 - Pacific Ocean
 - West Indies

Peru (Country)

Edit Delete Print Start Workflow (0) Tasks (0) Comments Country

Hide Properties Show History

http://topquadrant.com/ns/examples/geography#Peru

- Labels and Description
- Label Objects (SKOS-XL)
- Geopolitical Characteristics
- Location
- Standard Relationships
- Notes
- Matching Relationships
- Other Properties

preferred label: Peru (en)
Perú (es)
Pérou (fr)

alternative label:

hidden label:

notation:

type: Country

definition:

top concept of:

TopBraid EDG
Vocabulary Management

Search for Concept

Type Country

Free Text

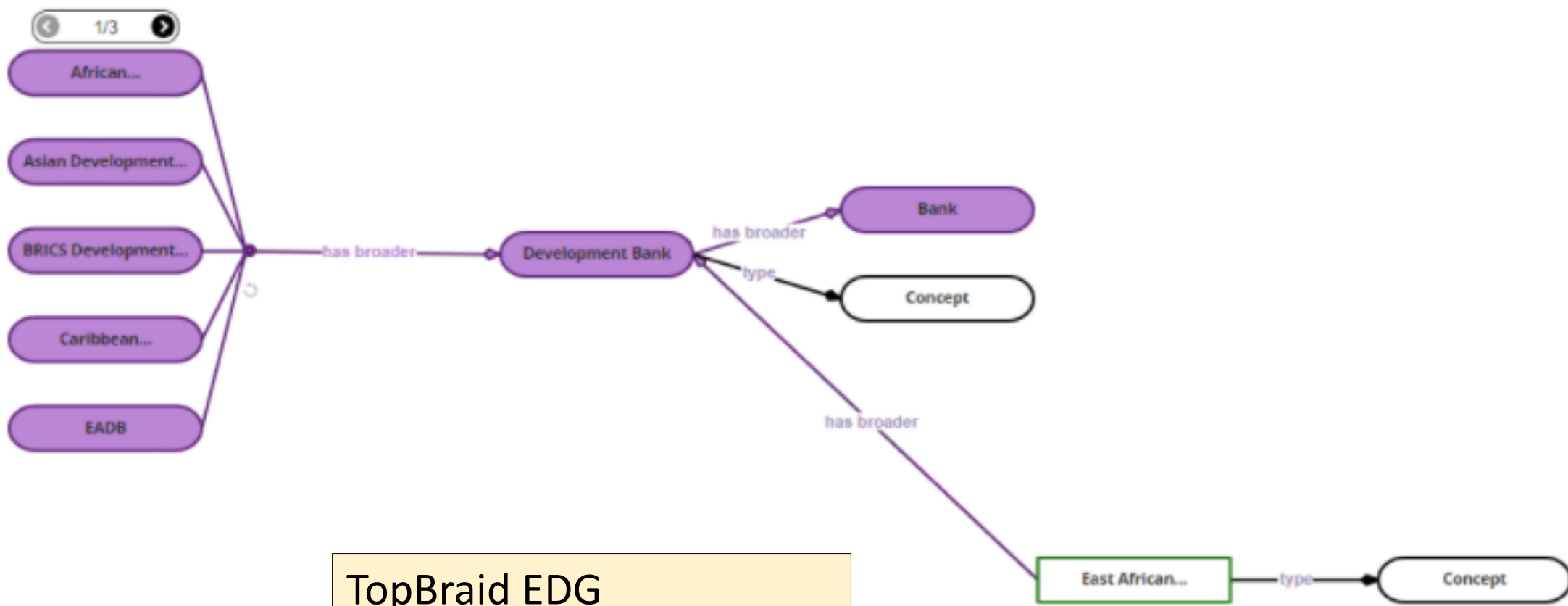
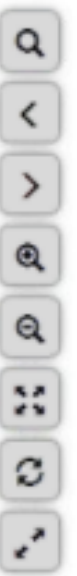
Filters Columns

Return local results only

Details Export

Show 25 entries Refine:

<input type="checkbox"/>	Country	broader concept	area (KM)
<input type="checkbox"/>	Afghanistan	Asia	647500
<input type="checkbox"/>	Albania	Europe	28748
<input type="checkbox"/>	Algeria	Africa	2381741
<input type="checkbox"/>	American Samoa	Pacific Ocean	
<input type="checkbox"/>	Andorra	Europe	468
<input type="checkbox"/>	Angola	Africa	1246700
<input type="checkbox"/>	Anguilla	West Indies	
<input type="checkbox"/>	Antarctica		
<input type="checkbox"/>	Antigua and Barbuda	West Indies	442
<input type="checkbox"/>	Argentina	South America	
<input type="checkbox"/>	Armenia	Asia	
<input type="checkbox"/>	Aruba	West Indies	193



TopBraid EDG
Vocabulary Management
visualization

Free and Open Source Software

- **Protégé** – Developed by the Center for Biomedical Informatics Research at Stanford University School of Medicine. <https://protege.stanford.edu>
 - Dedicated ontology software; not so suitable for taxonomies/thesauri
- **VocBench** – Developed by the Artificial Intelligence Research group at University of Tor Vergata, Rome <http://vocbench.uniroma2.it>
 - For OWL ontologies, SKOS(/XL) thesauri
 - Introduced in 2010 for UN Food & Agriculture Organization's AGROVOC thesaurus.
 - Now funded by the EU European Commission's ISA² program. Current version 3.
 - Can be installed on a web server or on a single desktop
- **TemaTres** – *Originally* developed by the Library and Information Science program of the University of Buenos Aires <https://www.vocabularyserver.com>
 - Available On-Premise on a web server; Software as a Service (SaaS), or On-Demand
 - Version 3.0, November 2017
 - Uses SKOS model and supports ISO thesaurus standards

Class Concept Scheme Collection Property

⊕ ⊖ ⊗ A ↻

- ▼ Car (en)
 - FamilyCar (en)
 - LuxuryCar (en)
 - SportsCar (en)**
- ▼ Veichle (en)
 - Bus (en)
- ▼ Car (en)
 - FamilyCar (en)
 - LuxuryCar (en)
 - SportsCar (en)
 - Truck (en)

SportsCar (en) http://test.project/c_74ee3f83

▼ Types: —

rdf:type

- skos:Concept

▼ Top Concept of: —

▼ Schemes: —

skos:inScheme

- main scheme (en)

▼ Broaders: —

skos:broader

- Car (en)

▼ Lexicalizations: —

<http://www.w3.org/2008/05/skos-xl#prefLabel>

- 🇬🇧 SportsCar

▼ Notes: —

▼ Properties: —

Thesaurus/Ontology Management Software - Other

- a.k.a. by Synercon – information management software with taxonomy/thesaurus/ontology component; Australian company
- Coreon – taxonomy/thesaurus + terminology management; German company
- Lucidia’s STAR/Thesaurus – part of the CuadraSTAR (2008 acquisition) suite, software marketed at libraries, archives, and museums
- Soutron Global – library management system with thesaurus component
- Unilexicon – web hosted open source, but all vocabularies are open, too.
- Wordmap – offered by a consulting company (Earley Information Science), not their main focus

Software for indexing/tagging content with controlled vocabulary terms

Different methods:

- Manual indexing
- Automated indexing/auto-categorization
- Machine-aided indexing

Different types:

- Dedicated software, if automated (but not for manual)
- Add-on module to the taxonomy management software (usually for automated)
- Component of a content management system, if manual (not for automated)
- Custom-built

Auto-Classification Software

Example dedicated tools for auto-classification/automated indexing
Taxonomy management components, if any, are limited.
(Create the taxonomy in an external management tool.)

- Attivio
- BA Insight
- Concept Searching
- Coveo
- Expert System
- Lexalytics
- Lucidworks
- OpenText
- SAS Text Miner
- Sinequa

Evaluating Taxonomy Tools

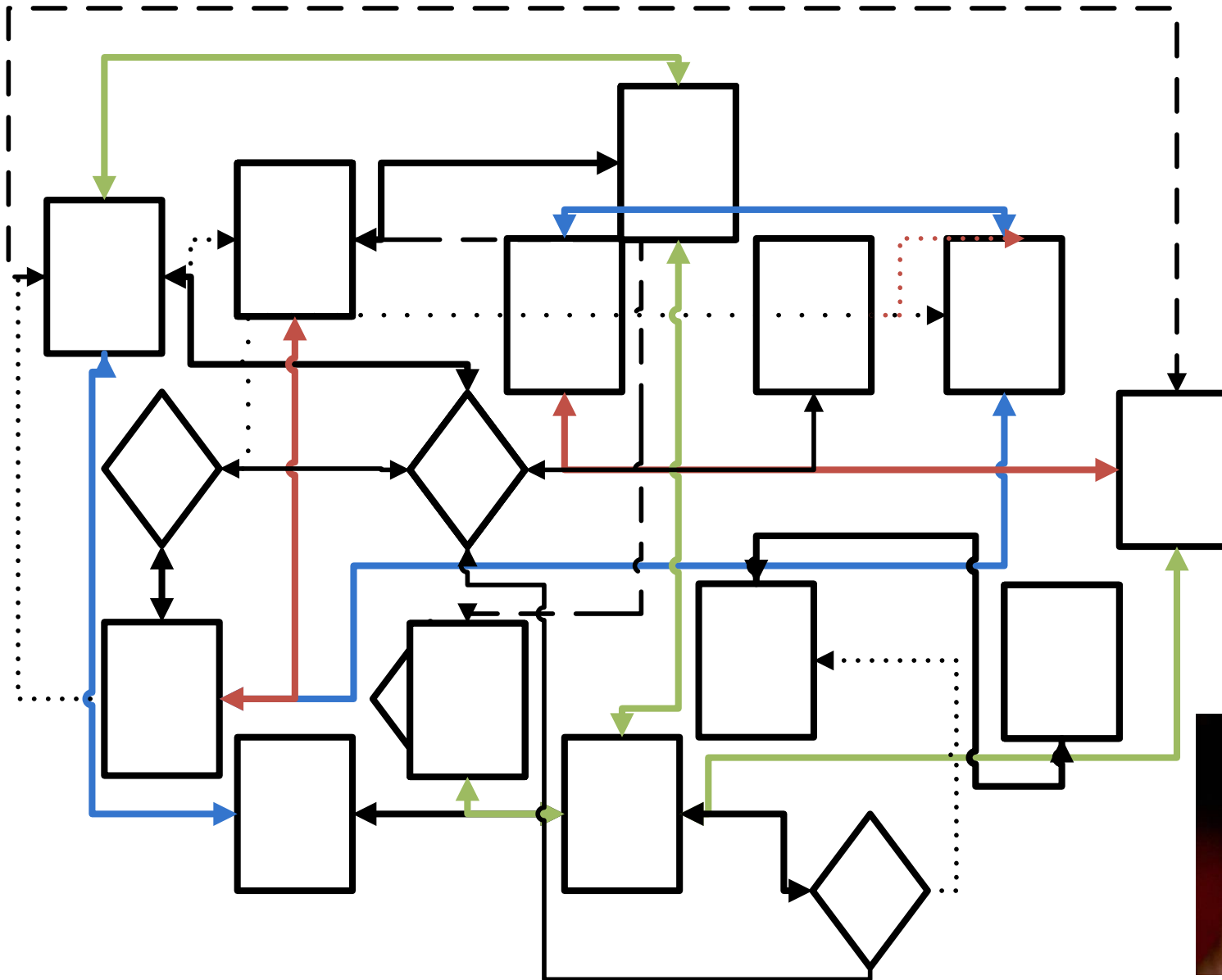
Marti Heyman

Executive Director, Metadata Strategy and Operations

OCLC

heymanm@oclc.org





Setting the Context

- My assumptions:
 - You have an enterprise-wide taxonomy program underway
 - You have sufficient corporate support to fund at least yourself
 - You have sufficient business support that you have subject matter experts assisting you
 - You've reached the point where you sense a taxonomy management package is needed

Setting the Context

- My case for the Spec & Select effort:
 - You've invested significant time and money in developing the enterprise-wide controlled vocabularies you have
 - You've invested significant time and effort gaining the confidence of management in your judgment
 - You've invested significant time and effort establishing your credibility

So, why risk all that for the sake of saving a small amount of time?

Possible Outcomes

Over kill

- Bought more than you needed

Under kill

- Didn't buy as much as you needed

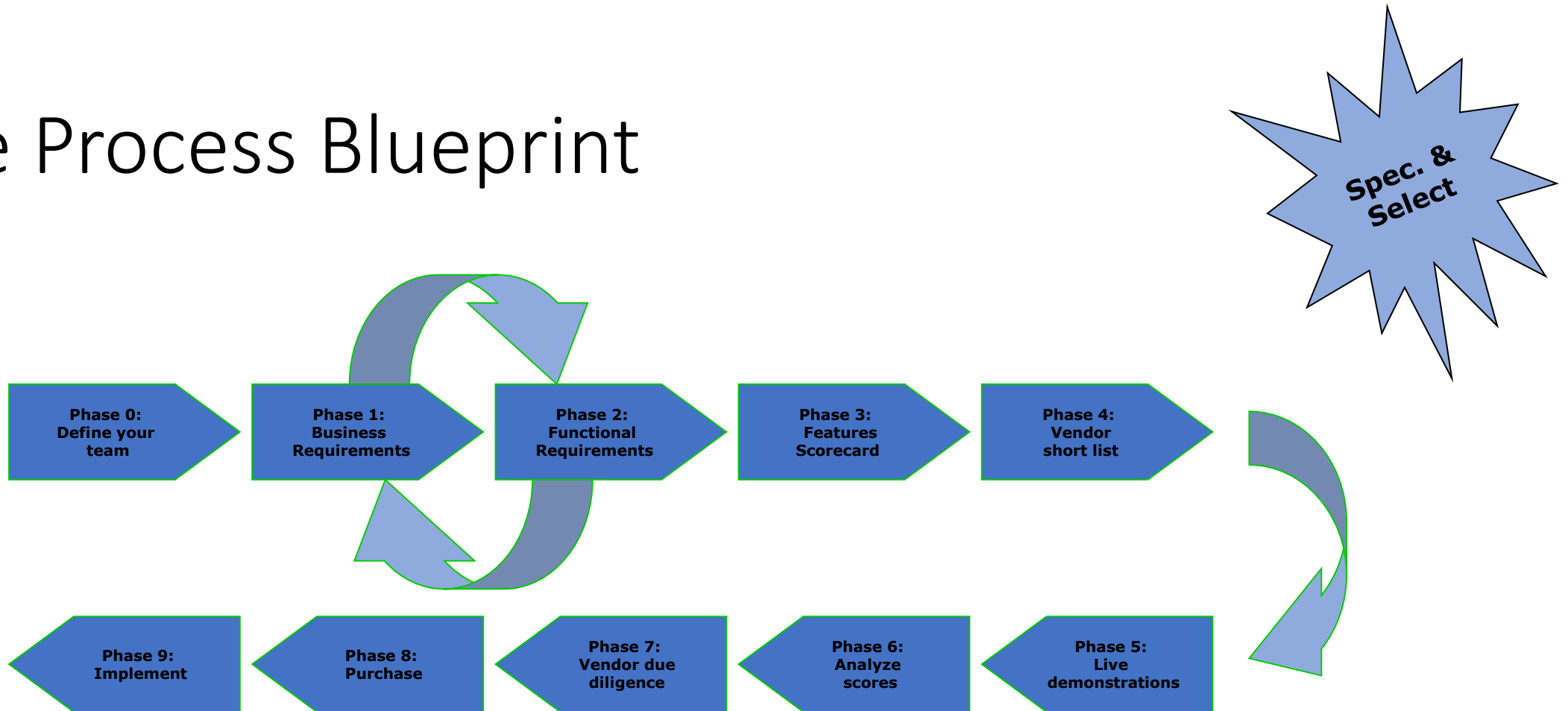
Road kill

- Conflicts with IT infrastructure and other systems

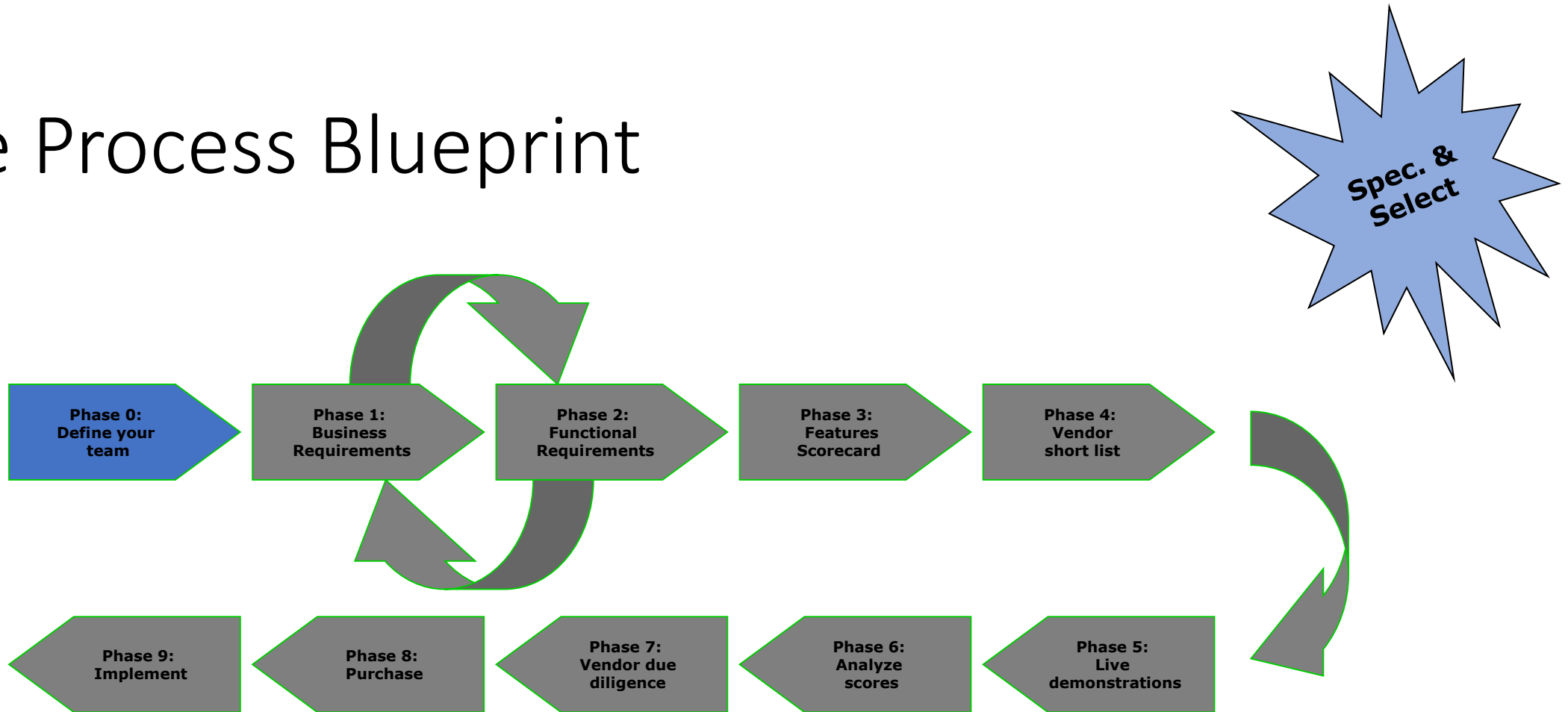
Living the Good Life

- Bought just what you needed for now with a little room to grow!

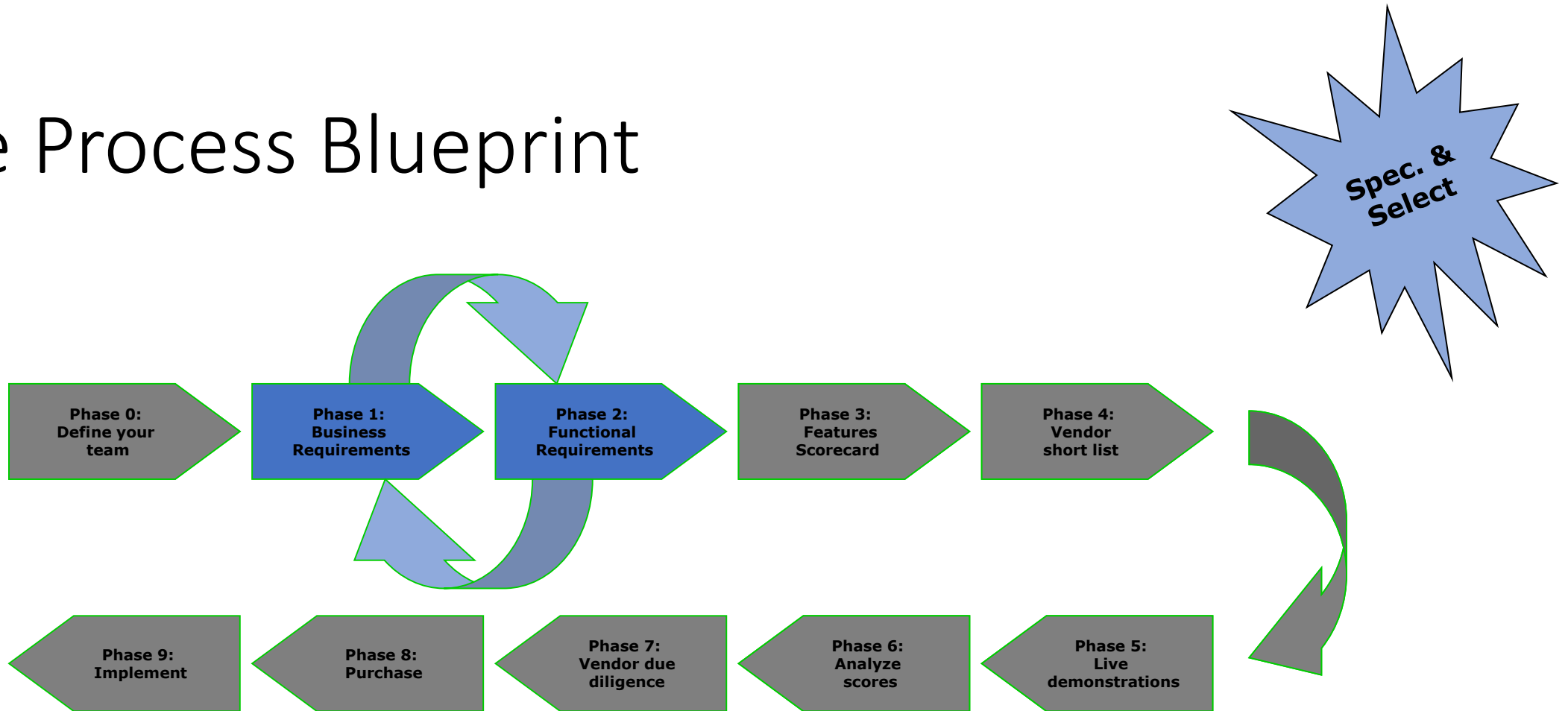
The Process Blueprint



The Process Blueprint



The Process Blueprint



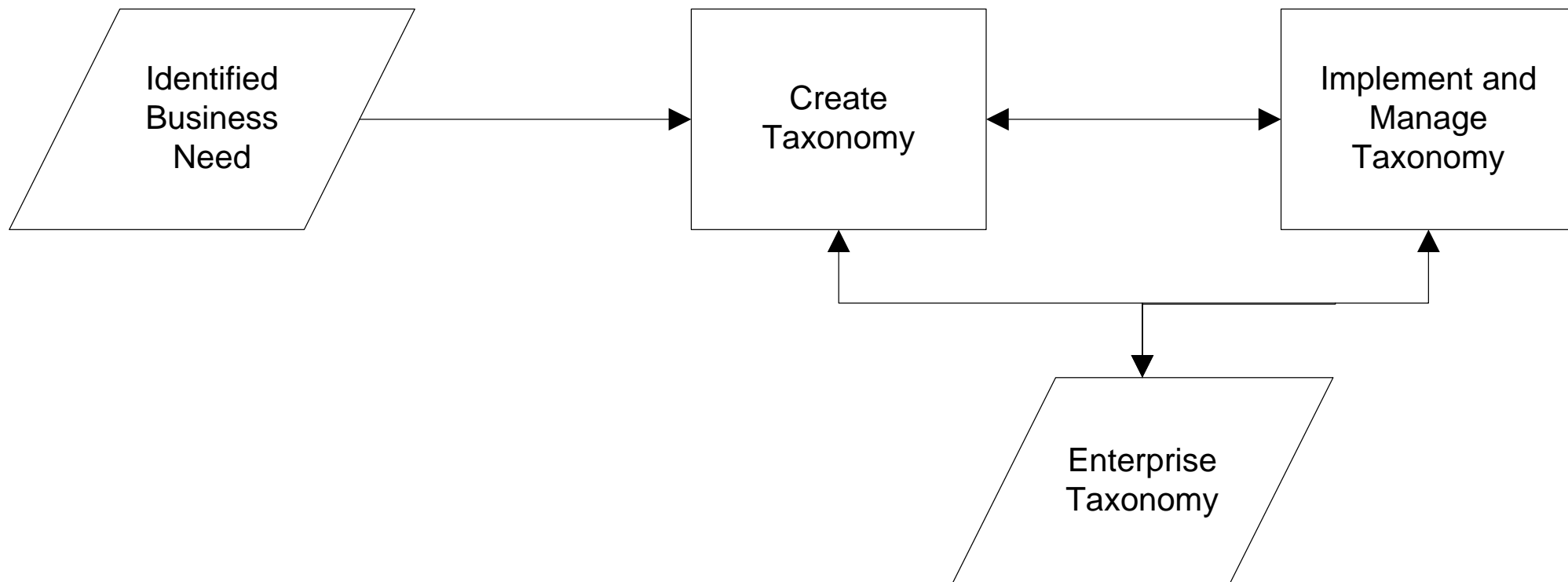
Sample business requirements

- Make the enterprise vocabulary accessible and available to a geographically dispersed set of taxonomy managers
- Make the enterprise vocabulary accessible to all applications that depend on it
- Ensure scalability of the enterprise vocabulary
- Low operating costs
- Unix environment compliant

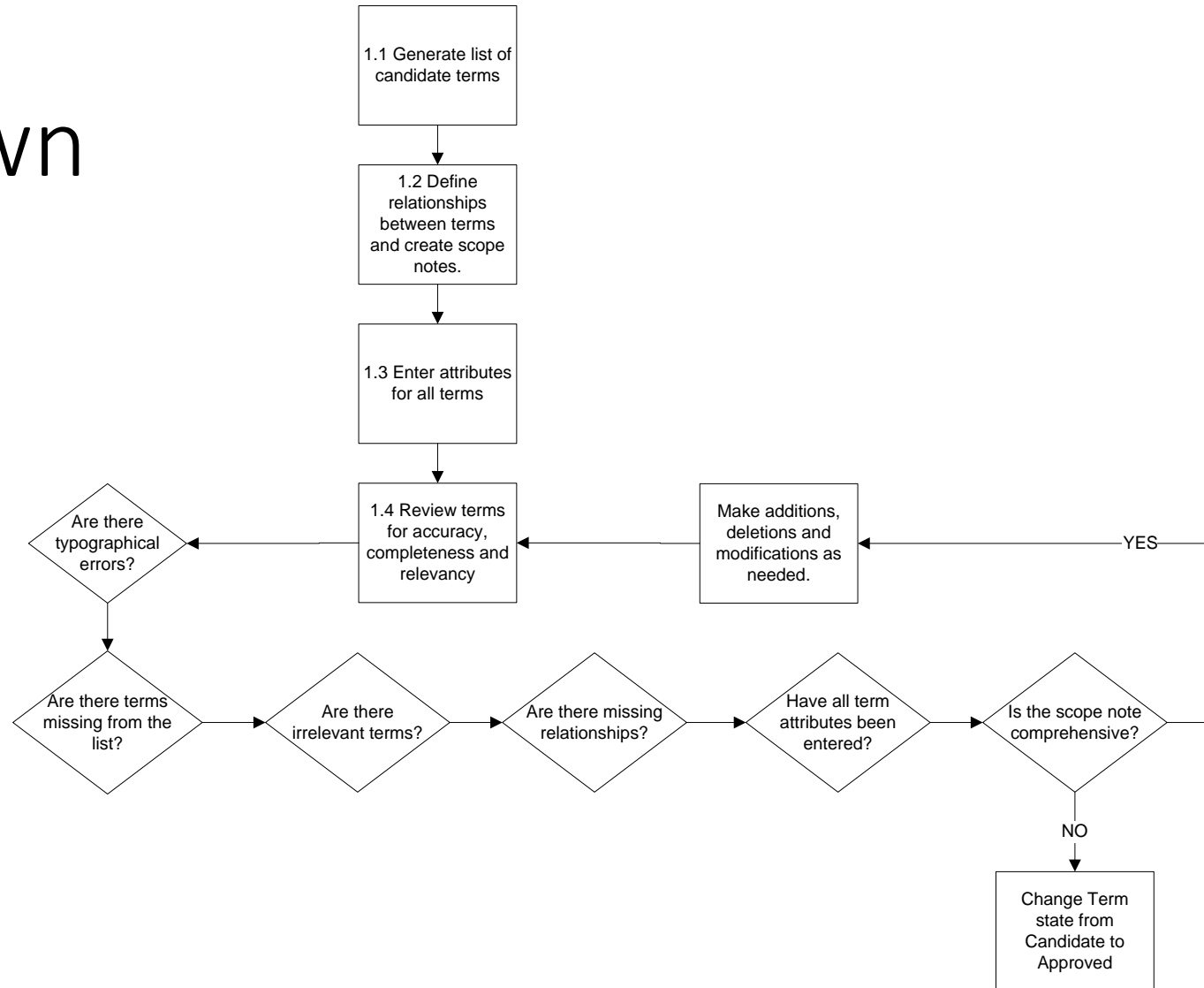
Identifying functional requirements

- This can be trickier
- Be sure you have the use case(s) and technical constraints clearly in mind
- Challenge what is a true “requirement” (i.e. necessity)

Step 1: Map the Process



Step 2: Drill Down



Capture the full context

- Design Considerations
 - Supports multiple languages including foreign characters.
 - Compliant with ANSI/NISO Z39.19.
- Business Rules
 - A single term may exist in multiple categories.
 - Duplicate terms can not be entered.
 - New terms exist in the “Candidate” state. All terms must be reviewed by a second person before being moved to the “Approved” state.

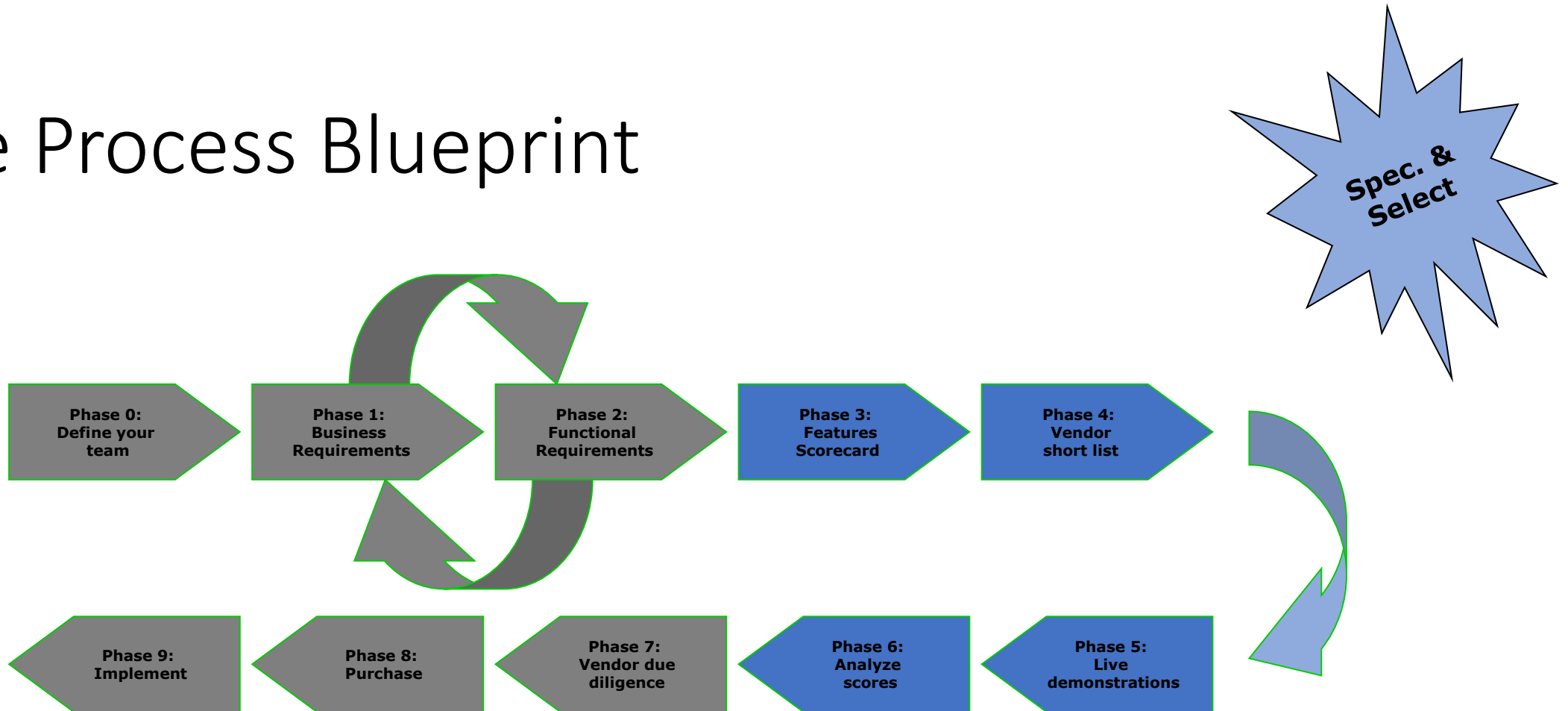
Sample functional requirements

1.0	General
1.1	Supports multiple categories.
1.2	Supports multiple languages, including foreign characters
1.3	Compliant with ANSI/NISO Z39.19

4.0	Relationship Creation/Management
4.1	Automatic reciprocal relationship management.
4.2	Ability to create and delete relationship types without developer assistance. BT/NT, RT, Use/Use for.....
4.3	Poly-hierarchies (i.e. ability for a single term to be in more than one category).

2.0	Term Creation/Editing
2.1	General editing - multi term select using ctrl and shift
2.2	General editing - right mouse menu for undo , cut, copy, paste, select all
2.4	General editing - inbuilt spell check software for global spell-check (automatically runs on "save record" action)
2.5	Prevents duplicate term entry (any level). System should warn that the term already exists.

The Process Blueprint



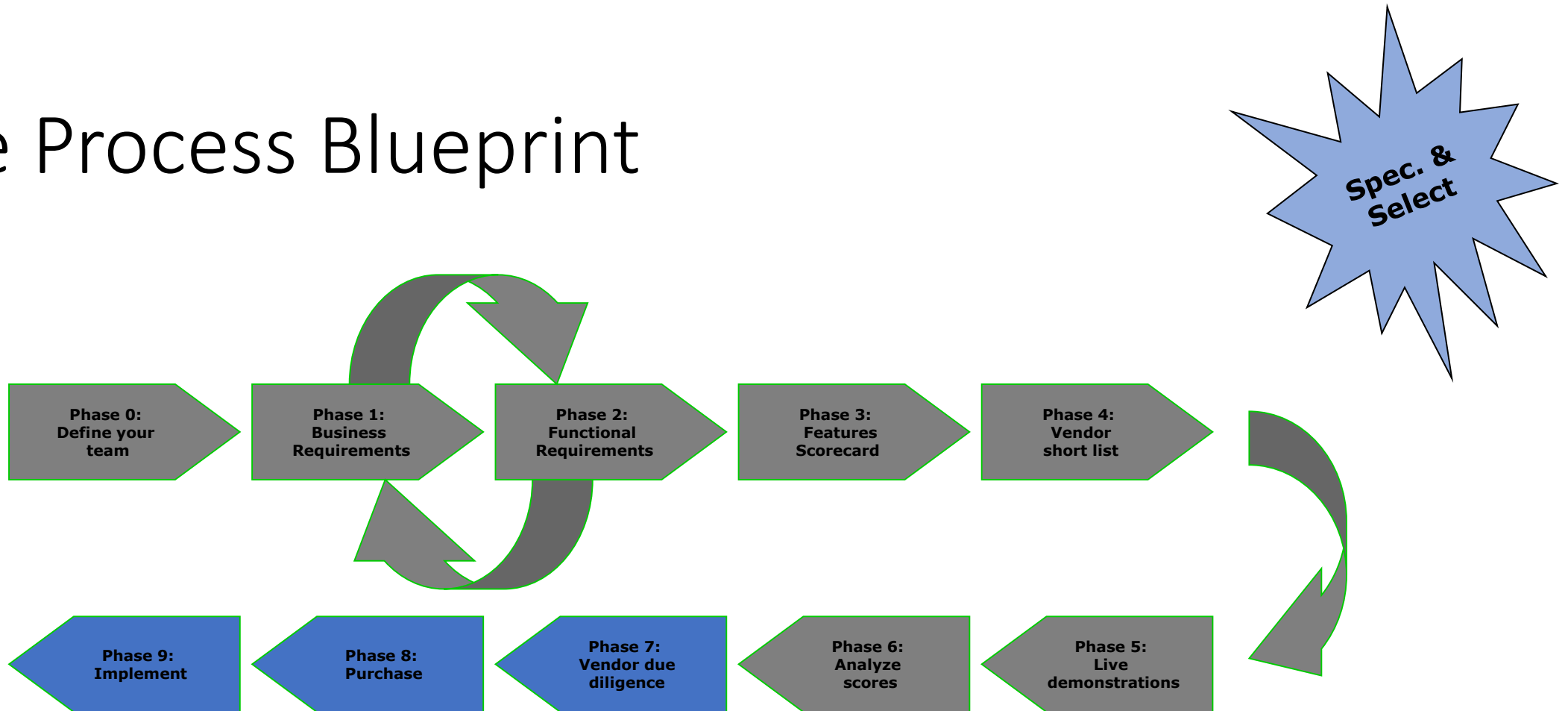
Sample Score Card

Taxonomy Features Review Sheet				
Priority 1=Must have 2=Important but don't need in first release 3=Nice to have				
Relevant User Group	Feature	Priority	Vendor 1	Comments
All (Content Mgrs, Taxonomy Mgrs, End Users)	General - Supports multiple Languages	1		
All (Content Mgrs, Taxonomy Mgrs, End Users)	General - ANSI/NISO Z39.19 compliant	1		
All (Content Mgrs, Taxonomy Mgrs, End Users)	General - Windows operating system compliant	1		
Taxonomy Managers	Editing - Ability to select multiple terms.	1		
Taxonomy Managers	Editing - Built in Spell checker	1		
Taxonomy Managers	Editing - Prevent duplicate term entry within a category. Warn when duplicate is entered.	1		
Taxonomy Managers	Administration - terms should not be deleted by taxonomy managers, just changed to an inactive status.	1		
System	Relationship Mgmt - Poly-hierarchies (i.e. Ability for a single term to be in more than one category).	1		
Taxonomy Managers	Workflow - There should be an audit trail showing changes.	1		
Taxonomy Managers	Workflow - roles should be used when defining taxonomy managers responsible for each workflow step. Ability to change the roles without developer assistance.	2		
All (Content Mgrs, Taxonomy Mgrs, End Users)	Display - Ability to see full record of a term on the screen (relationships, attributes, comments, etc.)	1		
All (Content Mgrs, Taxonomy Mgrs, End Users)	Reporting - Ability to export reports in a common format (CSV, XML, etc.)	1		
Taxonomy Managers	Import/Export - Import/Export terms, categories, taxonomies in a standard format (CSV, XML, etc.)	1		
Taxonomy Managers	Search - Search across all terms and all categories (taxonomies) and attributes.	1		
Taxonomy Managers	Security - Ability to set security at a category (taxonomy) level.	2		
All (Content Mgrs, Taxonomy Mgrs, End Users)	Help - On line help, quick reference guide	1		

Sample Completed Score Card

Taxonomy Features Review Sheet							
Priority 1=Must have 2=Important but don't need in first release 3=Nice to have							
Relevant User Group	Feature	Priority	Vendor 1	Comments	Vendor 2	Comments	
All (Content Mgrs, Taxonomy Mgrs, End Users)	General - Supports multiple Languages	1	0		1		
All (Content Mgrs, Taxonomy Mgrs, End Users)	General - ANSI/NISO Z39.19 compliant	1	1		1		
All (Content Mgrs, Taxonomy Mgrs, End Users)	General - Windows operating system compliant	1	1		1		
Taxonomy Managers	Editing - Ability to select multiple terms.	1	0		1		
Taxonomy Managers	Editing - Built in Spell checker	1	0		0		
Taxonomy Managers	Editing - Prevent duplicate term entry within a category. Warn when duplicate is entered.	1	1		1		
Taxonomy Managers	Administration - terms should not be deleted by taxonomy managers, just changed to an inactive status.	1	0.5	business process	1		
System	Relationship Mgmt - Poly-hierarchies (i.e. Ability for a single term to be in more than one category).	1	1		1		
Taxonomy Managers	Workflow - There should be an audit trail showing changes.	1	0		1		
Taxonomy Managers	Workflow - roles should be used when defining taxonomy managers responsible for each workflow step. Ability to change the roles without developer assistance.	2	1		1		
All (Content Mgrs, Taxonomy Mgrs, End Users)	Display - Ability to see full record of a term on the screen (relationships, attributes, comments, etc.)	1	1		1		
All (Content Mgrs, Taxonomy Mgrs, End Users)	Reporting - Ability to export reports in a common format (CSV, XML, etc.)	1	1		1		
Taxonomy Managers	Import/Export - Import/Export terms, categories, taxonomies in a standard format (CSV, XML, etc.)	1	0.5	not intuitive	1		
Taxonomy Managers	Search - Search across all terms and all categories (taxonomies) and attributes.	1	1		0.5	no wild cards?	
Taxonomy Managers	Security - Ability to set security at a category (taxonomy) level.	2	0		1		
All (Content Mgrs, Taxonomy Mgrs, End Users)	Help - On line help, quick reference guide	1	0.5	pdf guide	0	print manual	
			9.3	TOTAL	13.1	TOTAL	

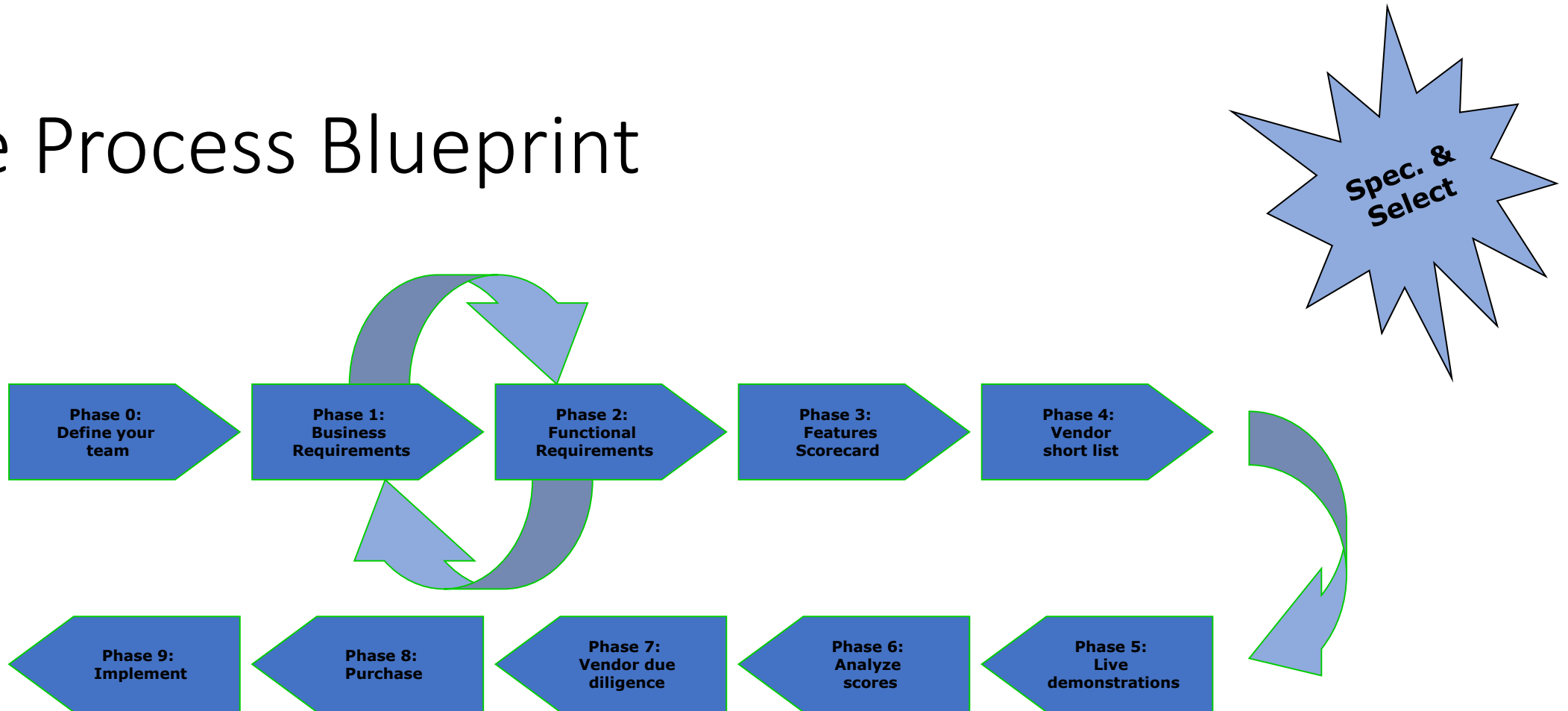
The Process Blueprint



Key Points

- Requirements: to purchase the software that meets your needs and expectations, you need a strong, clear, unambiguous definition!
- Process Blueprint: follow a tried and true path to success and ensure you have the data to explain your decisions
- Ensure stewardship of corporate funds
- Avoid being road kill, spend the time to do the “front-end analysis” work

The Process Blueprint





Taxonomy Tools and Tool Evaluation

Heather Hedden (heather.hedden@cengage.com)

Marti Heyman (heymanm@oclc.org)