



XML DTDs and other Alternatives: Vocabulary Markup Language (Voc-ML) Project & Friends

Joseph A. Busch
Director, Solutions Architecture

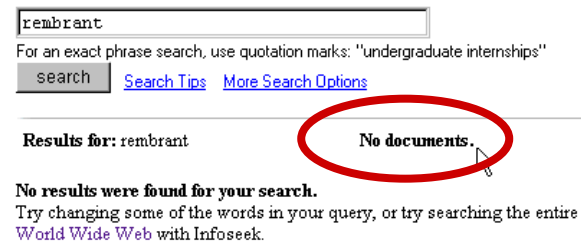
Outline

- ◆ The “real” Semantic Web
- ◆ Vocabulary Markup Language (Voc-ML)
 - Namespace registry
 - Schema
 - Services definition
- ◆ Voc-ML applications

The problem IS search!

- ◆ Data values, NOT just data structures are needed.

My News Feed



Soergel's SemWeb Proposal

- ◆ System of integrated access to data on concepts and terminology.
- ◆ Bring together variety of sources that exist largely in separate worlds, including dictionaries, thesauri, classification schemes, etc.
- ◆ Federated system with multiple collaborators.
- ◆ Common interface to all concept & terminology knowledge bases on the Internet.

Dagobert Soergel. "SemWeb: integrated access to distributed ontological resources." (April 1998) Last checked March 29, 2002.
<http://www.clis.umd.edu/faculty/soergel/soergelsemwebprop.pdf>

The Real Semantic Web

- ◆ Namespace for uniquely identifying a semantic scheme & each concept within each scheme.
- ◆ Broad template or conceptual schema for holding all types of semantic information & specifying relationships among them.
- ◆ Definitions of services for interacting with the System.

Namespace: NKOS Registry

Asset metadata— <i>The Who, Where and When</i>	Title, Alternate, Creator, Publisher, Date, Type, Format, Identifier, Language
Subject metadata— <i>The What and Why</i>	Subject, Description , Application
Relational metadata— <i>The Linkages</i>	Relation
Use metadata— <i>The How</i>	Rights, Entity Types, Relationships, Info Given

NKOS registry example

Vocabularies

ADD EDIT METASOURCE

Title	Vocabulary ID	167	Namespace	Identifier	Type
	Vocabulary	UN Standard Products and Services Classification			
	Abbreviation	UNSPSC			
	URL/ISBN	http://eccma.org/unspsc/browse/			
Description	Description	Designed to group together common goods and services. 10-digit hierarchical tree includes 2-digit category, 6-digit product identifier, and 2-digit business function identifiers.			
Creator	Authority	Dunn & Bradstreet and UN Development Programme			
Date	Pub Date	2001	Copyright	free	
	No. of Terms		Cost Info	"...can be used and distributed without restrictions or license fees."	
	Priority	high	Acquisition Info	download from web	
	Status	identified	Source	Electronic Commerce Code Management Association	
			Format		

Domains Topics **Contacts** Find Record Print Record Add New Contact

Record: 232 of 245 **Subject** **Publisher**

NKOS registry example

Vocabularies

ADD EDIT METASOURCE

Title

Vocabulary ID

Vocabulary

Abbreviation

URL/ISBN

Description

Authority

Pub Date

No. of Terms

Priority

Status

UN Stand

UNSPSC

http://eco

Designed services, 1 category, function id

Dunn & Br

2001

high

identified

Domains

Topics

Record: 1 232 of 1 (Filtered)

Vocabulary Topics

Vocabulary

Topics

UN Standard Products and Services Classification

Products, Industrial

Products, Consumer

*

Record: 1 of 2

Subject

Publisher

Schema: Vocabulary Markup Language (Voc-ML)

- ◆ XML schema for the Semantic Web.
- ◆ Broad template for structured representation of semantic schemes.
 - Z39.19-1993 and ISO 2788
 - Dublin Core metadata
 - Tags and syntax for uniquely identifying each concept
 - Typed relationships (hierarchical, associative, etc.)
- ◆ Host agency: Networked Knowledge Organization Systems

Voc-ML schema example

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!DOCTYPE MetaSource SYSTEM "Voc-ML.dtd">
<MetaSource>
<SVHeader>
```

Dublin Core

```
<dc:Title>UN Standard Product and Services Classification</dc:Title>
<dc:Creator>Dunn & Bradstreet </dc:Creator>
<dc:Subject>Products, Industrial</dc:Subject>
<dc:Subject>Products, Consumer</dc:Subject>
```

```
<UIDprefix>unspsc </UIDprefix>
</SVHeader>
```

Unique ID

```
<SVTerm UID="unspsc::501921">
```

```
<label>Snack foods</label>
```

```
<parent UREF="unspsc::5019">
```

```
<child UREF="unspsc::50192101">
```

```
<child UREF="unspsc::50192102">
```

```
<child UREF="unspsc::50192103">
```

```
<child UREF="unspsc::50192104">
```

```
</SVTerm>
```

Typed Relationships

ADL Thesaurus Protocol: XML Elements

<properties>	Overall properties of the thesaurus	
<term>	A term name and its “preferred” status	(format)
<term-description>	Full term description	(format)
<extended>	Any other thesaurus format	(format)
<list>	List of zero or more terms	<response>
<hierarchy>	A hierarchy of terms	<response>
<error>	Error code and description	<response>
<response>	Response from a thesaurus service	

ADL Thesaurus Protocol: Services

← get-properties

← query? (*operator, text, fuzzy, format*)

```
<query-operators  
  equals="true"  
  contains-all-words="true"  
  contains-any-words="true"  
  matches-regexp="false"/>
```

- text = text

- fuzzy = {true|false}

- format = <term>, <term-description>, <extended>

← get-hierarchies? (starting-term, broader-levels, narrower-levels, format)

Service definition example: ← get-properties

<http://nkosregistry.org/unspsc/get-properties>

<response>

 <properties>

 <dc.name>UN Standard Product and Services Classification</dc.name>

 <dc.Creator>Dunn & Bradstreet</dc.Creator>

 <dc.Subject>Products, Industrial</dc.Subject>

 <dc.Subject>Products, Consumer</dc.Subject>

 <query-operators

 equals="true"

 contains-all-words="true"

 contains-any-words="true"

 matches-regexp="false"/>

 <extended-schema><http://eccma.org/unspsc.dtd></extended-schema>

 </properties>

</response>

Service definition example: ← query?

<http://nkosregistry.org/unspsc/query?operator=contains-any-words&text=snack+foods&format=term>

<response>

<term-description>

<term>Snack food</term>

<scope-note>Use this category for food eaten between regular meals.</scope-note>

<broader>

<term>Prepared and preserved foods</term>

</broader>

<narrower>

<term>Pretzels</term>

<term>Corn chips</term>

<term>Potato chips</term>

<term>Popcorn</term>

</narrower>

<used-for>

<term preferred="false">Junk foods</term>

</used-for>

<related>

<term>Crackers</term>

</related>

</term-description>

<response>

Service definition example: ← get-hierarchies?

<http://nkosregistry.org/unspsc/get-hierarchies?starting-term=snack%20foods&broader-levels=-2&narrower-levels=1&format=term>

<hierarchy direction="broader" maxlevels="-2">

<node>

<term>Snack foods</term>

<node>

<term>Prepared and preserved foods</term>

<node>

<term>Food Beverage and Tobacco Products</term>

</node>

</node>

</node>

</hierarchy>

<hierarchy direction="narrower" maxlevels="1">

<node>

<term>Pretzels</term>

<term>Corn chips</term>

<term>Potato chips</term>

<term>Popcorn</term>

</node>

</hierarchy>

Application: Visual vocabulary editor

Portability.	<ul style="list-style-type: none">▪ Voc-ML input/output.▪ Utilities to convert outline, spreadsheet, directory to Voc-ML
Manage namespace unique ID's (not just a list of labels).	<ul style="list-style-type: none">▪ Enforce namespace and ID uniqueness.
Allow polyhierarchy (membership in multiple classes).	<ul style="list-style-type: none">▪ Copy & paste term to additional parent.
Allow typed equivalents.	<ul style="list-style-type: none">▪ Add/edit equivalents with pre-defined types.
Easy to use.	<ul style="list-style-type: none">▪ File manager style hierarchy display.▪ Drag & drop/cut & paste terms and their children.▪ Undo.▪ Right-click functions.

Application: Visual vocabulary editor

MetaSource Editor - C:\Documents and Settings\jbusch\Desktop\Vocabulary Stuff\unspsc.xml*

File Edit View Tools Help

Search For: snack

Search Results: 1 Nodes Found [Searched On: UID, Label] [Searched All Descendants]

Tree ID	UID	Label
18.10.7	unspsc::501921	Snack foods

Term Details

Unique Identifier: unspsc::501921

Label:

Associated Terms

Name	Type
Junk foods	Used for
Crackers	Related

Add Delete

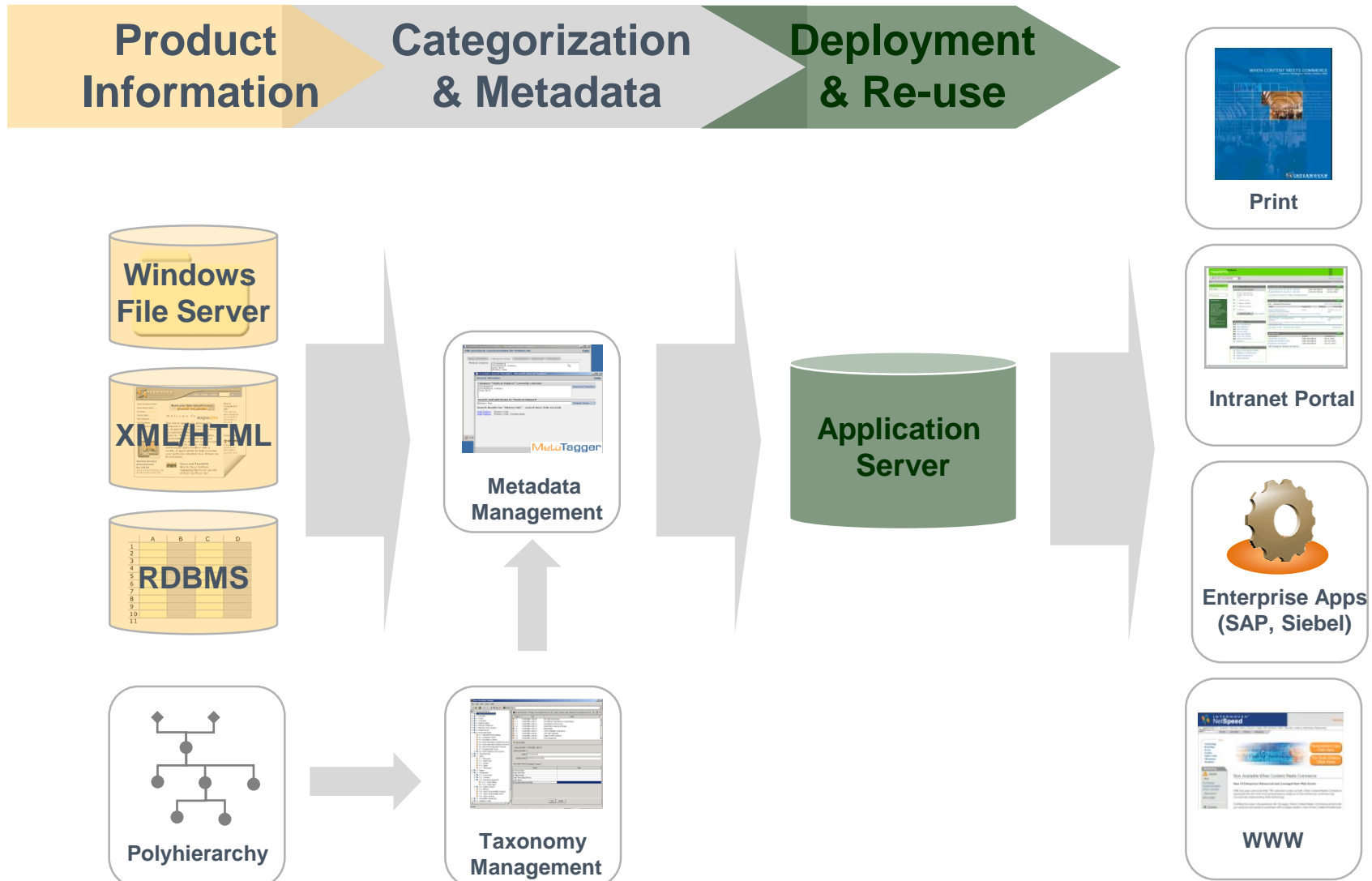
UN Standard Products and Services Classification

- 1 - Apparel, Luggage, and Personal Care Products
- 2 - Building and Construction Machinery and Equipment
- 3 - Building, Construction and Maintenance Services
- 4 - Chemicals (including Bio-Chemicals) and Allied Products
- 5 - Cleaning Equipment and Supplies
- 6 - Commercial, Military and Private Vehicles, Aircraft, and Marine Vessels
- 7 - Communications, Computer Equipment, and Office Machinery
- 8 - Defense, Law Enforcement, Security and Safety Equipment
- 9 - Distribution and Conditioning Systems, Equipment, and Supplies
- 10 - Drugs and Pharmaceutical Products
- 11 - Editorial, Design, Graphic and Fine Art Services
- 12 - Education and Training Services
- 13 - Electronic Components and Supplies
- 14 - Environmental Services
- 15 - Farming, Fishing, Forestry and Wildlife Conservation
- 16 - Farming, Fishing, Forestry and Wildlife Management
- 17 - Financial and Insurance Services
- 18 - Food Beverage and Tobacco Products
 - 18.1 - Fruits, vegetables and nuts
 - 18.2 - Meat
 - 18.3 - Seafood
 - 18.4 - Dairy products and eggs
 - 18.5 - Grains, cereals and their flours
 - 18.6 - Oils and fats
 - 18.7 - Sugar, sweeteners and confectionery
 - 18.8 - Seasonings and preservatives
 - 18.9 - Bakery products
 - 18.10 - Prepared and preserved foods
 - 18.10.1 - Soups and stews
 - 18.10.2 - Frozen foods
 - 18.10.3 - Canned foods (bottled etc.)
 - 18.10.4 - Dried foods
 - 18.10.5 - Salted or smoked foods
 - 18.10.6 - Pickled foods
 - 18.10.7 - **Snack foods**
 - 18.10.7.1 - Pretzels
 - 18.10.7.2 - Corn chips
 - 18.10.7.3 - Potato chips
 - 18.10.7.4 - Popcorn

Application: Manage product taxonomies

Organize (and reorganize) product classes for diverse purposes	<ul style="list-style-type: none">▪ Drag and drop editing; Preserve unique ID within namespace
Allow products to have many aliases	<ul style="list-style-type: none">▪ Alternates associated with unique ID
Allow products to exist in more than one class	<ul style="list-style-type: none">▪ Polyhierarchy (to allow multiple parents)
Map products across multiple taxonomies	<ul style="list-style-type: none">▪ Relationships across different namespaces (e.g. linked parallel hierarchies in different languages)
Generate and maintain linkages to associated documentation	<ul style="list-style-type: none">▪ Associate metadata labels as well as namespace with unique ID

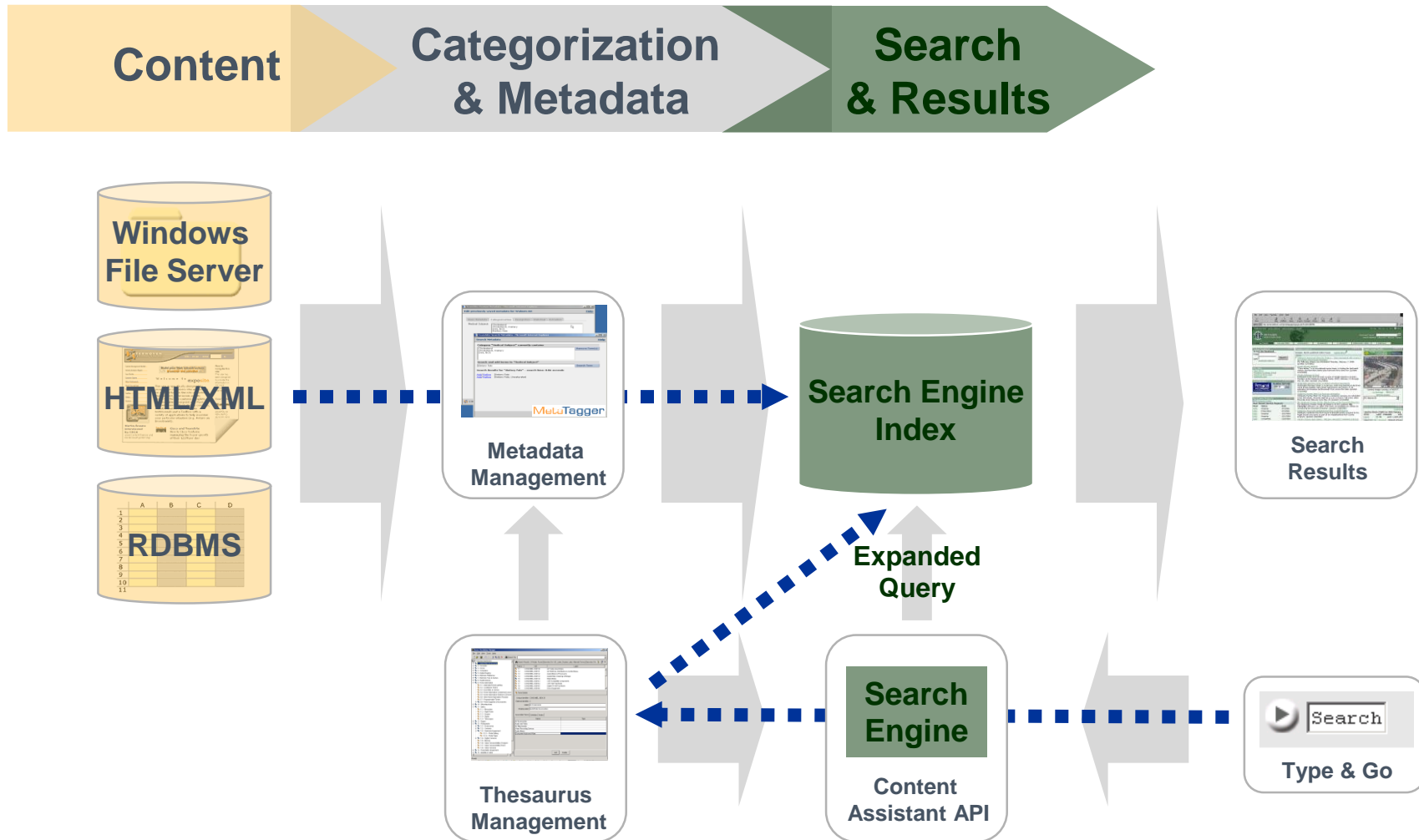
Application: Manage product taxonomies



Application: Search query intermediation

Map content to controlled vocabulary (thesaurus, etc.)	<ul style="list-style-type: none">▪ Associate metadata labels.
Control exactly what is indexed (metadata not just full text) and when.	<ul style="list-style-type: none">▪ Deploy metadata direct to search engine without spidering.
Deploy controlled vocabulary (thesaurus, etc.) to search engine.	<ul style="list-style-type: none">▪ Re-direct user queries to thesaurus, and return expanded query based on rules.

Application: Search query intermediation



Contact Information



Joseph A. Busch
Director, Solutions Architecture
Interwoven
803 11th Avenue
Sunnyvale, CA 94089
(408) 220-6974
jbusch@interwoven.com

Visit www.interwoven.com
Enterprise Content Management