Metadata Interoperability
Business consultants who specialize in applying taxonomies, metadata, automatic classification, and other information retrieval technologies to the needs of business and government.

Leadership in enterprise content management, knowledge management e-commerce, e-learning and web publishing.

Spin-off from Metacode Technologies, developer of XML metadata repository, automated categorization methods and taxonomy editor acquired by Interwoven in 2000 (now part of Autonomy).

More than 30 years experience in digital text and image management.

Metadata and taxonomy community leadership.
- President, American Society for Information Science & Technology
- Dublin Core Metadata Initiative Board Member
- American Library Association Committee on Accreditation External Reviewer

Recent taxonomy projects

Public Sector

For Profit Sector

Not-for-Profit Sector

Taxonomy Strategies LLC  The business of organized information
Agenda

- Interoperability
- Dublin Core
- Dates, Roles and Topics
- Types of semantic schemes
- Tagging assets
- Tools for automating tagging
Interoperability

- The ability of diverse systems and organizations to work together by exchanging information.
- Semantic interoperability is the ability to automatically interpret the information exchanged meaningfully and accurately.
Interoperability ROI

- Assets are expensive to create so it’s critical that they can be found, so they can be used and re-used.
- Every re-use decreases the asset creation cost and increases the asset value.
Interoperability (2)

- If assets are so important, why can’t they be found?
  - They contain no searchable text.
  - They exist in different applications, file shares and/or desktops.
  - … Other reasons?

- When they are found why can’t assets be reused?
  - When there are multiple versions, it’s difficult to choose which one to use.
  - The usage rights may not be clear.
  - … Other reasons?
Interoperability (3)

- Digital assets are sourced from multiple applications and locations
  - Product lifecycle management (PLM) application
  - Product information management (PIM) application
  - Third party contractors’ systems
  - In-house graphic design department
  - Marketing and Communications servers
  - …Other applications and locations?
Interoperability vision

- I want to easily find any assets in a particular format that can be used for a specific purpose regardless of where they are located.
- Challenges:
  - How to align different metadata properties
    - E.g., Title and Caption; Location and Setting; etc.
  - How to align different vocabularies
    - E.g., CA and California; RiM and Research in Motion; etc.
What is metadata?

- Metadata provides enough information for any user, tool, or program to find and use any piece of content.

<table>
<thead>
<tr>
<th>Complexity</th>
<th>Enabled Functionality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Subject metadata – *What, Where & Why:* Subject, Type, Coverage
- Use metadata – *When & How:* Date, Language, Rights
- Asset metadata – *Who:* Identifier, Creator, Title, Description, Publisher, Format, Contributor
- Relational metadata – *Links between and to:* Source, Relation

http://dublincore.org/documents/dces/
What is metadata

- Metadata provides enough information for any user, tool, or program to find and use any piece of content.

**Metadata Types**

- **Asset metadata** – Who:
  - Identifier, Creator, Title, Description, Publisher, Format, Contributor

- **Subject metadata** – What, Where & Why:
  - Subject, Type, Coverage

- **Relational metadata** – Between and to:
  - Source, Relation

- **Use metadata** – When & How:
  - Date, Language, Rights

---

**Enabled Functionality**

- More efficient editorial process
- Better navigation & discovery

---

http://dublincore.org/documents/dces/
But Dublin Core is a little more complicated

<table>
<thead>
<tr>
<th>Elements</th>
<th>Refinements</th>
<th>Encodings</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Identifier</td>
<td>Abstract</td>
<td>Box</td>
<td>Collection</td>
</tr>
<tr>
<td>2. Title</td>
<td>Access rights</td>
<td>DCMIType</td>
<td>Dataset</td>
</tr>
<tr>
<td>3. Creator</td>
<td>Alternative</td>
<td>DDC</td>
<td>Event</td>
</tr>
<tr>
<td>4. Contributor</td>
<td>Audience</td>
<td>IMT</td>
<td>Image</td>
</tr>
<tr>
<td>5. Publisher</td>
<td>Available</td>
<td>ISO3166</td>
<td>Interactive</td>
</tr>
<tr>
<td>6. Subject</td>
<td>Bibliographic citation</td>
<td>ISO639-2</td>
<td>Resource</td>
</tr>
<tr>
<td>7. Description</td>
<td>Conforms to</td>
<td>LCC</td>
<td>Moving Image</td>
</tr>
<tr>
<td>8. Coverage</td>
<td>Created</td>
<td>LCSH</td>
<td>Physical Object</td>
</tr>
<tr>
<td>9. Format</td>
<td>Date accepted</td>
<td>MESH</td>
<td>Service</td>
</tr>
<tr>
<td>10. Type</td>
<td>Date copyrighted</td>
<td>Period</td>
<td>Software</td>
</tr>
<tr>
<td>11. Date</td>
<td>Date submitted</td>
<td>Point</td>
<td>Sound</td>
</tr>
<tr>
<td>12. Relation</td>
<td>Education level</td>
<td>RFC1766</td>
<td>Sound</td>
</tr>
<tr>
<td>13. Source</td>
<td>Extent</td>
<td>RFC3066</td>
<td>Still Image</td>
</tr>
<tr>
<td>14. Rights</td>
<td>Has format</td>
<td>TGN</td>
<td>Text</td>
</tr>
<tr>
<td>15. Language</td>
<td>Has part</td>
<td>UDC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Has version</td>
<td>URI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Is format of</td>
<td>W3CTDF</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Is part of</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Is referenced by</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Is replaced by</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Is required by</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Issued</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Is version of</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>License</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mediator</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modified</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Provenance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>References</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Replaces</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Requires</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rights holder</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spatial</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Table of contents</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Temporal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Valid</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
... and Dublin Core has gotten more abstract
So why Dublin Core?

- Dublin Core is a de-facto standard across many other systems and standards
  - RSS (1.0), OAI (Open Archives Initiative)
  - Inside organizations – DAMS, ECMS, SharePoint, etc.
- Mapping to DC elements from most existing schemes is simple
- Metadata already exists in enterprise applications
  - avid, OpenText, MarkLogic, SAP, Documentum, MS Office, SharePoint, etc.

Source: Todd Stephens, BellSouth

Per-Source Data Types, Access Controls, etc.

Taxonomies, Vocabularies, Ontologies

Enterprise Business Intelligence
- Portals
- Business Intelligence
- Search Engines
- Topic Maps

Semantic Metadata
- Integration Metadata

Repository Collection

Structural Metadata

Asset Collection
- Systems
- Applications
- Interfaces
- Databases
- XML
- Web Services
- Metrics
- Schema, DTD
- email
- Web Pages
- Documents
- Postals
- People
- Transformation
- Messaging
- Models

Source: Todd Stephens, BellSouth
Dublin Core is the top vocabulary in the linked data cloud

<table>
<thead>
<tr>
<th>Vocabulary prefix</th>
<th>Vocabulary link</th>
<th>Number of usages in data sets</th>
<th>Data sets that use the vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc</td>
<td><a href="http://purl.org/dc/elements/1.1/">http://purl.org/dc/elements/1.1/</a></td>
<td>92 (31.19 %)</td>
<td>Data sets that use dc</td>
</tr>
<tr>
<td>foaf</td>
<td><a href="http://xmlns.com/foaf/0.1/">http://xmlns.com/foaf/0.1/</a></td>
<td>81 (27.46 %)</td>
<td>Data sets that use foaf</td>
</tr>
<tr>
<td>skos</td>
<td><a href="http://www.w3.org/2004/02/skos/core#">http://www.w3.org/2004/02/skos/core#</a></td>
<td>58 (19.66 %)</td>
<td>Data sets that use skos</td>
</tr>
<tr>
<td>geo</td>
<td><a href="http://www.w3.org/2003/01/geo/wgs84_pos#">http://www.w3.org/2003/01/geo/wgs84_pos#</a></td>
<td>25 (8.47 %)</td>
<td>Data sets that use geo</td>
</tr>
<tr>
<td>xhtml</td>
<td><a href="http://www.w3.org/1999/xhtml/vocab#">http://www.w3.org/1999/xhtml/vocab#</a></td>
<td>19 (6.44 %)</td>
<td>Data sets that use xhtml</td>
</tr>
<tr>
<td>akt</td>
<td><a href="http://www.aktors.org/ontology/portal#">http://www.aktors.org/ontology/portal#</a></td>
<td>17 (5.76 %)</td>
<td>Data sets that use akt</td>
</tr>
<tr>
<td>bibo</td>
<td><a href="http://purl.org/ontology/bibo/">http://purl.org/ontology/bibo/</a></td>
<td>14 (4.75 %)</td>
<td>Data sets that use bibo</td>
</tr>
<tr>
<td>mo</td>
<td><a href="http://purl.org/ontology/mo/">http://purl.org/ontology/mo/</a></td>
<td>13 (4.41 %)</td>
<td>Data sets that use mo</td>
</tr>
<tr>
<td>vcard</td>
<td><a href="http://www.w3.org/2006/vcard/ns#">http://www.w3.org/2006/vcard/ns#</a></td>
<td>10 (3.39 %)</td>
<td>Data sets that use vcard</td>
</tr>
<tr>
<td>sioc</td>
<td><a href="http://rdfs.org/sioc/ns#">http://rdfs.org/sioc/ns#</a></td>
<td>10 (3.39 %)</td>
<td>Data sets that use sioc</td>
</tr>
<tr>
<td>cc</td>
<td><a href="http://creativecommons.org/ns#">http://creativecommons.org/ns#</a></td>
<td>8 (2.71 %)</td>
<td>Data sets that use cc</td>
</tr>
<tr>
<td>geonames</td>
<td><a href="http://www.geonames.org/ontology#">http://www.geonames.org/ontology#</a></td>
<td>6 (2.03 %)</td>
<td>Data sets that use geonames</td>
</tr>
<tr>
<td>frbr</td>
<td><a href="http://purl.org/vocab/frbr/core#">http://purl.org/vocab/frbr/core#</a></td>
<td>6 (2.03 %)</td>
<td>Data sets that use frbr</td>
</tr>
<tr>
<td>xsd</td>
<td><a href="http://www.w3.org/2001/XMLSchema#">http://www.w3.org/2001/XMLSchema#</a></td>
<td>6 (2.03 %)</td>
<td>Data sets that use xsd</td>
</tr>
<tr>
<td>time</td>
<td><a href="http://www.w3.org/2006/time#">http://www.w3.org/2006/time#</a></td>
<td>5 (1.69 %)</td>
<td>Data sets that use time</td>
</tr>
<tr>
<td>event</td>
<td><a href="http://purl.org/NET/c4dm/event.owl#">http://purl.org/NET/c4dm/event.owl#</a></td>
<td>5 (1.69 %)</td>
<td>Data sets that use event</td>
</tr>
<tr>
<td>dbpedia</td>
<td><a href="http://dbpedia.org/resource/">http://dbpedia.org/resource/</a></td>
<td>5 (1.69 %)</td>
<td>Data sets that use dbpedia</td>
</tr>
<tr>
<td>gr</td>
<td><a href="http://purl.org/goodrelations/v1#">http://purl.org/goodrelations/v1#</a></td>
<td>4 (1.36 %)</td>
<td>Data sets that use gr</td>
</tr>
</tbody>
</table>

http://www4.wiwiss.fu-berlin.de/lodcloud/state/#structure
## Dates, roles and topics

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
<th>Set By</th>
</tr>
</thead>
<tbody>
<tr>
<td>date.added</td>
<td>Date the asset was first added to the DAM.</td>
<td>DAM</td>
</tr>
<tr>
<td>date.lastModified</td>
<td>Date the asset was last reviewed for accuracy and relevance. Used for provenance and to validate content or rights.</td>
<td>DAM</td>
</tr>
<tr>
<td>date.reviewed</td>
<td>Date the content was last reviewed for accuracy and relevance. Used for provenance, and to compute a future date to recheck the content.</td>
<td>DAM</td>
</tr>
<tr>
<td>date.nextReviewed</td>
<td>Date of next scheduled review for accuracy and relevance.</td>
<td>Rule</td>
</tr>
<tr>
<td>date.embargoed</td>
<td>Date and time that content is scheduled to become available on the site. Content can be prepared in advance and system will push it out once the embargo date is reached.</td>
<td>Manual</td>
</tr>
<tr>
<td>date.subject</td>
<td>Date of the event, data, or other information depicted in the asset. Used for search and recall purposes. (This is not the date the asset was uploaded or last updated).</td>
<td>Manual</td>
</tr>
</tbody>
</table>
Dublin Core dates

- “A date associated with an event in the life cycle of the resource”
- Woefully underspecified.
- Typically the publication or last modification date.
- Best practice: YYYY-MM-DD

### Refinements
- Created
- Valid
- Available
- Issued
- Modified
- Date Accepted
- Date Copyrighted
- Date Submitted

### Encodings
- DCMI Period
- W3C DTF (Profile of ISO 8601)
## Dates, roles and topics

<table>
<thead>
<tr>
<th>Role</th>
<th>Description</th>
<th>Admin</th>
<th>Add</th>
<th>Edit</th>
<th>Delete</th>
<th>Approve</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrator</td>
<td>Technical administration of the DAM. Generally allowed to do anything, to keep the system running and up-to-date.</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Approver</td>
<td>Senior DAM staff with the authority to approve assets for publication. In small shops Contributors may also be Approvers. Larger shops, and those using outsider contractors will have many Contributors but just a few Approvers.</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Contributor</td>
<td>Editorial staff with authority to contribute new assets to the DAM. Their work must be approved by an Approver before it can be published. Administrators have the authority to approve content for publication, but only as an exception not the rule.</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>
Dates, roles and topics

Locations
- Setting
  - Classroom & Seminar Room
  - Common Area
  - Campus Exteriors
  - Housing
  - Laboratory
  - Office
  - Clinical
  - Community
  - Nature
  - Community Pharmacy
  - Culture
- Campuses & Locations
  - Bay Area
  - San Francisco
  - National
  - International
  - Laurel Heights
  - Mission Bay Campus
  - Mission Center
  - Mount Zion Campus
  - Parnassus Campus
  - ...

Concepts
- Caring for Patients
- Collaboration
- Concentration
- Conducting Science
- Contemplation
- Diversity
- Growth and Progress
- Happiness
- Innovation
- Leadership
- Learning
- Passion
- Questioning
- Recreation
- Service
- Socializing
- Systems & Organizations
- Teaching/Presenting
- Unhappiness

Expertise
- Basic and Applied Research
- Health Policy Research
- Clinical Research
- Pharmacy Practice Research

Events
- Awards Ceremonies
- Community Outreach
- Conferences & Courses
- Graduations & Courses
- Professional Program
- Graduations, Graduate Programs
- Homecomings & Reunions
- Orientations & Registrations
- Parties & Receptions
- Recruitment
- Students Organizations & Extracurricular Activities
- White Coat Ceremonies

Objects
- Lab Equipment
- Research Core Equipment
- Computing, Networking & IT Equipment
- Medicines, Medicine Containers, & Delivery Devices
- Medical Devices
- Transportation Vehicles
- Lab coats

Organization
- Departments / Units
- ... Research Centers
- ... Labs
- ...

Titles
- Alumnus
- Associate / Assistant Dean
- Board of Advisors Chair
- Dean
- Donor
- Faculty
- Friend
- Graduate Students
- PharmD Students
- Postdocs, professionalPostdocs, science
- Staff / Administrator
- Visitors
- Other UC

Other People
- Infants
- Children
- Youth
- Families
- Elderly
- Patients
- Researchers
- Clinicians
- Teachers
- University Students

Taxonomy Strategies LLC The business of organized information
### UCSF topics (1)

<table>
<thead>
<tr>
<th>Concepts</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caring for Patients</td>
<td>Diversity</td>
</tr>
<tr>
<td>Collaboration</td>
<td>Growth &amp; Progress</td>
</tr>
<tr>
<td>Concentration</td>
<td>Joy</td>
</tr>
<tr>
<td>Conducting Science</td>
<td>Innovation</td>
</tr>
<tr>
<td>Contemplation</td>
<td>Leadership</td>
</tr>
<tr>
<td>Teaching &amp; Presenting</td>
<td>Public Service</td>
</tr>
<tr>
<td>Sadness</td>
<td>Socializing</td>
</tr>
<tr>
<td>Learning</td>
<td>Systems &amp; Organizations</td>
</tr>
<tr>
<td>Questioning</td>
<td>Surprised</td>
</tr>
</tbody>
</table>

**UCSF Titles**
- Alumnus
- Associate / Assistant Dean
- Board of Advisors
- Chair
- Dean
- Donor
- Faculty
- Postdocs - Professional
- Postdocs - Science
- Staff / Administrator
- Visiting Faculty or Student
- Other UC Campus Persons

**Other People**
- Infants & Toddlers
- Families
- Researchers
- University Students
- Children
- Elderly
- Clinicians
- VIPs & Dignitaries
- Youth
- Patients
- Teachers

**Technique (people)**
- Candid
- Portrait
- Mugshot
- Other Posed

---

**Taxonomy Strategies LLC**  The business of organized information
## UCSF topics (2)

<table>
<thead>
<tr>
<th>Technique (people)</th>
<th>Candid</th>
<th>Portrait</th>
<th>Mugshot</th>
<th>Other Posed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of People</td>
<td>Unspecified</td>
<td>One Person</td>
<td>Large Group (6-15)</td>
<td>Crowd (15+)</td>
</tr>
<tr>
<td></td>
<td>No People</td>
<td>Small Group (2-5)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Events</th>
<th>Graduations - Graduate Programs</th>
<th>Students Organizations &amp; Extracurricular Activities</th>
<th>White Coat Ceremonies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Awards Ceremonies</td>
<td>Homecomings &amp; Reunions</td>
<td>Meetings</td>
</tr>
<tr>
<td></td>
<td>Community Outreach</td>
<td>Orientations &amp; Registrations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Conferences &amp; Courses</td>
<td>Portraits &amp; Receptions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Graduations - Professional Programs</td>
<td>Recruitment</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Objects</th>
<th>Lab Equipment</th>
<th>Medicines / Medicine Containers / Delivery Devices</th>
<th>Lab coats</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Research Core Equipment</td>
<td>Medical Devices</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Computing / Networking / IT Equipment</td>
<td>Transportation Vehicles</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Settings</th>
<th>Campus Exteriors</th>
<th>Office</th>
<th>Nature</th>
<th>Aerial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Classroom &amp; Seminar Room</td>
<td>Housing</td>
<td>Clinical</td>
<td>Pharmacy</td>
</tr>
<tr>
<td></td>
<td>Common Area</td>
<td>Laboratory</td>
<td>Community</td>
<td>Culture</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Campuses &amp; Locations</th>
<th>Mission Center</th>
<th>VAMC SF</th>
<th>Bay Area</th>
<th>Mount Zion Campus</th>
<th>Los Angeles-Orange County Pharmacy Education Program</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sea Francis</td>
<td>Panama Campus</td>
<td>North Bay Pharmacy Education Program</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>National</td>
<td>Peking University</td>
<td>Santa Clara-South Bay Pharmacy Education Program</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>International</td>
<td>SF General Hospital</td>
<td>UC Davis-Sacramento Pharmacy Education Program</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Laurel Heights</td>
<td>UC Washington Center</td>
<td>UC San Diego Pharmacy Education Program</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mission Bay Campus</td>
<td>UCSF Fresno</td>
<td>UCSF Fresno Pharmacy Education Program</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Departments &amp; Units</th>
<th>Dean's Office and Units</th>
<th>Department of Clinical Pharmacy</th>
<th>Other UC Departments and Units</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Department of Pharmaceutical Chemistry</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Department of Biopharmaceutical Sciences</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Department of Pharmacology and Toxicology</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Department of Orthodontics and Dental Public Health</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
UCSF topics (3)
UCSF asset and use taxonomies

### Asset Types
- Templates
- Documents
- Logos
- Images
- Art & Illustrations
- Data & Diagrams
- Photograph
- Rich Media
- Audio
- Video

### Usage / Rights
- Use
  - Project Only
  - School of Pharmacy Only
  - UC Regents
  - Unrestricted
- Privacy
  - Model-Released
  - Not Released

### Characteristics
- Color
  - Black & White
  - Color
- Orientation
  - Horizontal
  - Vertical
  - Panorama
  - Aerial
- Source Type
  - Amateur
  - Commissioned
  - Historic
  - Stock

### People
- Technique
  - Candid
  - Portrait
  - Mug shot
  - Other
- Posed Number
  - No People
  - One Person
  - Small Group
  - Medium Group
  - Large Group
- Types of People
  - Infants
  - Children
  - Youth
  - Families
  - Elderly
  - Patients
  - Researchers
  - Clinicians
  - Teachers
  - University Students
# Types of semantic schemes

<table>
<thead>
<tr>
<th>Type</th>
<th>Definition</th>
</tr>
</thead>
</table>
| Synonym Ring          | - A set of words/phrases that can be used interchangeably for searching.  
                        | - Example: Hypertension, High blood pressure                                                                                             |
| Controlled Vocabulary | - A list of preferred and variant terms, with defined hierarchical and associative relationships. A taxonomy is a type of controlled vocabulary.  
                        | - Typically used for names of countries, individuals, organizations                                                                     |
| Taxonomy              | - Identifying and naming things, and arranging them into a classification according to a set of rules.  
                        | - Often understood to mean biological classification, but can be applied to naming and classifying anything.                              |
| Classification Scheme | - An arrangement of knowledge that does not follow taxonomy rules.  
                        | - Usually enumerated; e.g., Dewey Decimal Classification                                                                                 |
| Thesaurus             | - A tool that controls synonyms and identifies the semantic relationships among terms.                                                  |
| Ontology              | - Resembles faceted taxonomy but uses richer semantic relationships among terms and attributes and strict specification rules.          |
Types of semantic schemes – simple to complex

**Semantic Schemes**

- Synonym Ring
- Controlled Vocabulary
- Taxonomy
- Classification Scheme
- Thesaurus
- Ontology

**Relationships**

Equivalence → Hierarchy → Associative

After: Amy Warner. *Metadata and Taxonomies for a More Flexible Information Architecture*
# Semantic relationships

<table>
<thead>
<tr>
<th>Concept</th>
<th>A unit of thought, an idea, meaning, or category of objects or events. A Concept is independent of the terms used to label it.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preferred Label</td>
<td>A preferred lexical label for the resource such as a term used in a digital asset management system.</td>
</tr>
<tr>
<td>Alternate Label</td>
<td>An alternative label for the resource such as a synonym or quasi-synonym.</td>
</tr>
<tr>
<td>Broader Concept</td>
<td>Hierarchical link between two Concepts where one Concept is more general than the other.</td>
</tr>
<tr>
<td>Narrower Concept</td>
<td>Hierarchical link between two Concepts where one Concept is more specific than the other.</td>
</tr>
<tr>
<td>Related Concept</td>
<td>Link between two Concepts where the two are inherently &quot;related&quot;, but that one is not in any way more general than the other.</td>
</tr>
</tbody>
</table>
Some semantic relationships for IBM

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM</td>
<td>prefLabel</td>
<td>IBM</td>
</tr>
<tr>
<td>International Business Machines</td>
<td>altLabel</td>
<td>IBM</td>
</tr>
<tr>
<td>I.B.M.</td>
<td>altLabel</td>
<td>IBM</td>
</tr>
</tbody>
</table>
Some semantic relationships for Mitt Romney

Lexical Relationships

- W. Mitt Romney
- Willard Mitt Romney
- Willard M. Romney
- W. M. Romney
- Governor Mitt Romney
- Romney, Mitt
- Romney, W. Mitt (Willard Mitt)

Subject | Predicate | Object
--- | --- | ---
Mitt Romney | prefLabel | Mitt Romney
W. Mitt Romney | altLabel | Mitt Romney
Willard Mitt Romney | altLabel | Mitt Romney
Willard M. Romney | altLabel | Mitt Romney
W.M. Romney | altLabel | Mitt Romney
Governor Mitt Romney | altLabel | Mitt Romney
Romney, Mitt | altLabel | Mitt Romney
Romney, W. Mitt (Willard Mitt) | altLabel | Mitt Romney
Negotiations With Iran Over Nuclear Program May Resume

By STEVEN LEE MYERS and RICK GLADSTONE
Published: February 17, 2012

WASHINGTON — The United States and the European Union signaled on Friday that negotiations with Iran over its nuclear program could soon resume for the first time in more than a year, even as a telecommunications network vital to the global banking industry prepared to expel Iranian banks.

While senior American and European officials stopped short of declaring a diplomatic breakthrough, Iran dropped previously unacceptable preconditions for talks in a letter this week from its senior nuclear negotiator, Saeed Jalili, who declared his country’s “readiness for dialogue” at “the earliest possibility.”

After weeks of official bluster, ominous threats of military action and a spectacular assassination attempt, both sides appeared to have reached a brief pause in their months-long僵持

More news and information about Iran.
The Tagging Problem

- How are we going to populate metadata elements with complete and consistent values?
- What can we expect to get from automatic classifiers?
Cheap and Easy Metadata

- Some fields will be constant across a collection
  - e.g., format, color, photographer or location
- In the context of a single collection those kinds of elements may add little value, but they add tremendous value when many collections are brought together into one place, and they are cheap to create and validate.
## 4 Indexing rules: How to use the taxonomy to tag content

<table>
<thead>
<tr>
<th>Rule</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use specific terms</td>
<td>Apply the most specific terms when tagging content. Specific terms can always be generalized, but generic terms cannot be specialized.</td>
</tr>
<tr>
<td>Use multiple terms</td>
<td>Use as many terms as necessary to describe <em>What the content is about</em> &amp; <em>Why it is important</em>.</td>
</tr>
<tr>
<td>Use appropriate terms</td>
<td>Only fill-in the facets &amp; values that make sense. Not all facets apply to all content.</td>
</tr>
<tr>
<td>Consider how content will be used</td>
<td>Anticipate <em>how the content will be searched for</em> in the future, &amp; <em>how to make it easy to find it</em>. Remember that search engines can only operate on explicit information.</td>
</tr>
</tbody>
</table>
Methods used to create & maintain metadata

- Paper or web-based forms widely used:
  - Distributed resource origination metadata tagging
  - Centralized clean-up and metadata entry.

**Source:** CEN/ISSS Workshop on Dublin Core.
Tagging considerations

- Who should tag assets? Producers or editors?
- Taxonomy is often highly granular to meet task and re-use needs, but with detailed taxonomy it’s difficult to get complete and consistent tags.
- The more tags there are (and the more values for each tag), the more hooks to the content, but the more difficult it is to get completeness and consistency.
- If there are too many tags or tags are too detailed, producers will resist and use “general” tags (if available).
- Vocabulary is often dependent on originating department, but the lingo may not be readily understood by people outside the department (who are often the users).
Tagging considerations (2)

- Automatic classification tools exist, and are valuable, but results are not as good as people can do.
  - “Semi-automated” is best.
  - Degree of human involvement is a cost/benefit tradeoff.
## Tools for tagging

<table>
<thead>
<tr>
<th>Vendor</th>
<th>Taxonomy Editing Tools</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autonomy Collaborative Classifier</td>
<td><a href="http://www.autonomy.com/content/Functionality/idol-functionality-categorization/index.en.html">www.autonomy.com/content/Functionality/idol-functionality-categorization/index.en.html</a></td>
<td></td>
</tr>
<tr>
<td>ConceptSearching</td>
<td><a href="http://www.conceptsearching.com">www.conceptsearching.com</a></td>
<td></td>
</tr>
<tr>
<td>Intelligent Topic Manager</td>
<td><a href="http://www.mondeca.com/Products/ITM">www.mondeca.com/Products/ITM</a></td>
<td></td>
</tr>
<tr>
<td>Temis Luxid® for Content Enrichment</td>
<td><a href="http://www.temis.com/?id=201&amp;selt=1">www.temis.com/?id=201&amp;selt=1</a></td>
<td></td>
</tr>
</tbody>
</table>
Thank You

QUESTIONS