

Taxonomy & Metadata

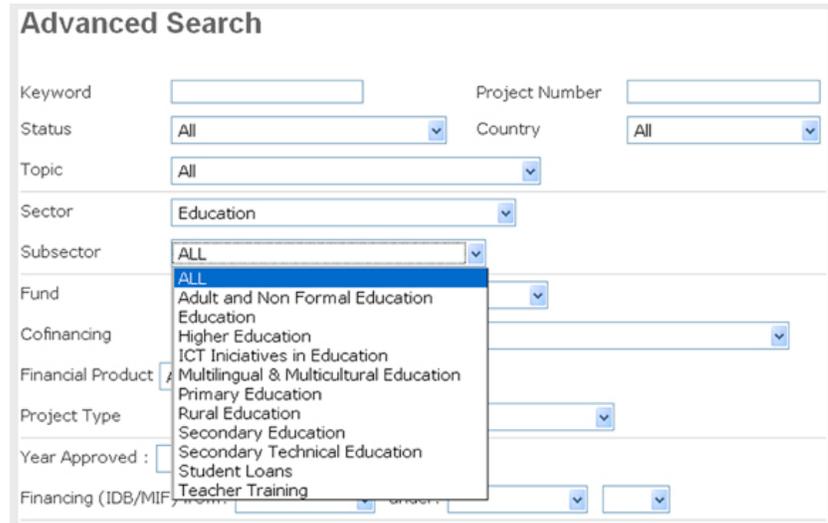
What is it?

One government leader suggests, “**Taxonomy** is the classification of an information domain, where terms are arranged into a hierarchy. It allows related terms to be grouped together and categorized in ways that make it easier to find the correct term to use - whether for navigating or searching a website or to describe an object.”¹

According to National Information Standards Organization, **Metadata** is “structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information source.”²

Figure 1 illustrates an advanced search utility for a hierarchical taxonomy of projects by “Sector” supported by two metadata fields (“Sector” and “Subsector”).

Once executed, media tagged with the sector “Education” will appear in the list. When designing metadata for digital assets, it is important that controlled vocabularies, keywords or tags follow consistent guidelines and practice, as uncontrolled keyword tagging tends to yield inconsistent and inaccurate retrieval results.



The image shows a screenshot of an 'Advanced Search' form. The form includes several fields: 'Keyword' (text input), 'Project Number' (text input), 'Status' (dropdown menu set to 'All'), 'Country' (dropdown menu set to 'All'), 'Topic' (dropdown menu set to 'All'), 'Sector' (dropdown menu set to 'Education'), and 'Subsector' (dropdown menu with a list of options including 'ALL', 'Adult and Non Formal Education', 'Education', 'Higher Education', 'ICT Initiatives in Education', 'Multilingual & Multicultural Education', 'Primary Education', 'Rural Education', 'Secondary Education', 'Secondary Technical Education', 'Student Loans', and 'Teacher Training'). Other fields include 'Fund', 'Cofinancing', 'Financial Product', 'Project Type', 'Year Approved', and 'Financing (IDB/MIF)'. The 'Subsector' dropdown is currently open, showing the list of options.

Figure 1 - Source: Inter-American Development Bank

When creating a taxonomy and metadata strategy for web search and findability: “Taxonomies (controlled vocabularies) enable consistent, accurate, and rapid indexing and retrieval of content. The fact that both indexers and end-users benefit from them, make controlled vocabularies very desirable.”³ Digital assets such as videos and pictures, which may not have text associated with them, are more easily discovered when tagged with controlled vocabularies. Thus, metadata fields and vocabulary are crucial components in a metadata strategy.

How do I start?

When designing taxonomies and metadata fields, Hedden suggests the following should be considered:⁴

- How large the subject-descriptive controlled vocabulary is:
 - If large, consider breaking it up into different vocabularies and metadata fields
 - If small, keep as one
- The ratio of names to topical subjects:
 - If names are few, more easily integrated within the topical subjects
 - If there are many names, their own metadata field may be justified
- Whether advanced search permits users to select more than one term at once from within a single metadata field:
 - If yes, a combination of term types within a metadata fields is acceptable
- How users are most likely to look up names and topics

¹ Victorian Government CIO Counsel <http://www.digital.vic.gov.au/wp-content/uploads/2014/07/WEB-STD-08-WMF-Information-Architecture-v2.1.pdf>

² NISO Press <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

^{3,4} Hedden, H. <http://www.palgrave-journals.com/dam/journal/v6/n5/full/dam201029a.html>

Taxonomy & Metadata

Important Tips⁵

- Keep it simple: While having a greater number of descriptive metadata fields can support more sophisticated searching, too many fields can be confusing
- Be sure to work with a wide set of people across the organization when developing taxonomies
- Make the determination as to whether a subject category should stand on its own as a separate controlled vocabulary and metadata field. In order to make this determination, two key questions to ask are:
 - Will users want to search and limit by this particular type of subject
 - Can the majority of digital assets in the collection be described by this type of subject
- The number and type of controlled vocabularies to create should be tailored to the particular digital asset collection and users

How can I learn more?

- **National Information Standards Organization**
<http://www.niso.org/publications/>
- **AIIM Taxonomy and Metadata**
<http://www.aiim.org/Resource-Centers/Taxonomy-and-Metadata>
- **Journal of Digital Asset Management**
<http://www.palgrave-journals.com/dam/journal/v6/n5/full/dam201029a.html#note1>
- **A Taxonomy Primer**
<http://www.lwmtechnology.com/publish/Why%20do%20You%20Need%20a%20Taxonomy%20Anyway-062003.pdf>
- **Metadata and Taxonomy Programs**
<http://www.dashboardinsight.com/articles/new-concepts-in-business-intelligence/unleashing-unstructured-data-value.aspx>
- **NASA Knowledge Map**
<http://km.nasa.gov/knowledge-map/>
- **NASA Johnson Space Flight Center CKO Corner: Irene Kaye Interview:**
<http://km.nasa.gov/cko-corner-irene-kaye-interview/>

⁵ Hedden, H. <http://www.palgrave-journals.com/dam/journal/v6/n5/full/dam201029a.html>