Two Traditions of Metadata Development

Bibliographic control approach

- developed before computer technology and internet were commonplace.
  mainly used in libraries and universities.
- from early on used rules and standards because of the growing need for
  local, national and international sharing of information
- well established. Now also being used in digital technologies

Examples:
- References and bibliography items
- database fields, document properties
- Dublin core, ROADS/IAFA, etc.

Data management approach

- more recent, developed for use with information technology.
- used in databases, internet webpages, software catalogues, etc.
- application is more diverse ranging from scientific, commercial, technical,
  research documents and information

Examples:
- Geospatial Metadata
- XML, Postscript


**Similarities**

- Both use metadata schemas to locate, identify, retrieve, and manipulate information.
- Both are used today in the digital environment.
- The aim of each approach is to provide a more standardised schema
- Needs of users, technology, document types may be different in the future, and already differ depending on the field, uses, etc. therefore any set of standard metadata elements and rules may become inadequate in the future.

**Differences**

- BCA places more emphasis on locating and identifying information, while DMA focuses more on the use (manipulation) of the information.
- Libraries are primarily information storage and retrieval systems, while the internet is primarily a communications medium and secondarily an info storage and retrieval system.
- There are lots of libraries, most provide a general range of information. Though there is only one internet there are lots of networks, lots of intranets, and trillions of webpages. All of these provide different information in different formats.

**Problems**

There is confusion when two types clash or crossover on the internet. Most average people would be quite familiar with the bibliographic approach and quite unfamiliar with the data management approach. So if they were to come across an example of the latter then the effective use of that information is at risk.
The data management approach must also consider security and data integrity. This is why often many of the additional metadata elements relate to administration and protection.

Many data archives are becoming very large and distributed, with diverse types of stored information formats. There can also be a need to use both approaches for different aspects of the same system.

**Bibliographic control approach is primarily concerned with metadata as a tool for information discovery.** For the people who take this approach:

Metadata is designed to enable users to locate, identify, retrieve, and manipulate information. These days there are two general ways we commonly use metadata:

- manually looking at it to give context to related info (eg. field names in an Access database)
- entering keywords into a search engine

Different search engines use metadata differently. For example, Google looks at the title, the language, the mime type, etc. but it also adds its own metadata for compiling its PageRank. Considering that, it could be said that Google adopts the Data Management Approach to metadata.

Libraries now catalogue data about all their books in a database so that people can use their computers to search for books by different criteria (eg. title, author, publisher). Those databases use a metadata schema, and most libraries will have a similar schema at least for the basic elements.
Data management approach accepts that “metadata represents information which supports the effective use of data from creation through long-term use”. For the people who take this approach:

The aim of this approach is to manage the data so that it can be used most effectively right from it’s time of creation. Therefore, it still uses any metadata schema that the bibliographic control approach may use, but also incorporates any additional data that helps make the information more useful. Its role is to support and give more context to the information.

Unlike this approach, the biblical control approach generally has a set group of data elements and does not allow for any additional data to be stored as metadata. One of the main types of additional data that is commonly stored is security and administrative information, such as authentication keys and data sequence. This is why it’s called the data management approach.

Objectives

As stated earlier, the goal of both approaches, both combined and within themselves, is to provide a standardised set of metadata elements which cover all user needs when locating, identifying, retrieving, and manipulating information.

Some people believe that this should be accomplished with a single set of elements. Others believe that half a dozen slightly different schemas are necessary, because there is a need for some specific data attributes in many cases. And finally there are some who believe that it is the way it is because that’s how it needs to be, so changing anything is unnecessary.
Because metadata is used in both traditional locations (eg. libraries) and modern digital environments (internet), there is a concerted effort to come up with a metadata standard. An example of this is the ROADS/IAFA template:

<table>
<thead>
<tr>
<th>Dublin Core element</th>
<th>IAFA template</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Title</td>
</tr>
<tr>
<td>Creator</td>
<td>Author-name (From Author (USER)* cluster)</td>
</tr>
<tr>
<td>Subject</td>
<td>Keyword</td>
</tr>
<tr>
<td></td>
<td>Subject-Descriptor-Scheme</td>
</tr>
<tr>
<td></td>
<td>Subject-Descriptor</td>
</tr>
<tr>
<td>Description</td>
<td>Description</td>
</tr>
<tr>
<td>Publisher</td>
<td>Publisher-name (From Publisher (ORGANISATION)* cluster)</td>
</tr>
<tr>
<td>Contributors</td>
<td>No direct equivalent</td>
</tr>
<tr>
<td>Date</td>
<td>Creation-date</td>
</tr>
<tr>
<td>Type</td>
<td>Category</td>
</tr>
<tr>
<td>Format</td>
<td>Format-v*</td>
</tr>
<tr>
<td></td>
<td>Requirements</td>
</tr>
<tr>
<td>Identifier</td>
<td>URI-v*</td>
</tr>
<tr>
<td></td>
<td>ISBN</td>
</tr>
<tr>
<td></td>
<td>ISSN</td>
</tr>
<tr>
<td>Source</td>
<td>Source</td>
</tr>
<tr>
<td>Language</td>
<td>Language</td>
</tr>
<tr>
<td>Relation</td>
<td>No direct equivalent</td>
</tr>
<tr>
<td>Coverage</td>
<td>No direct equivalent</td>
</tr>
<tr>
<td>Rights</td>
<td>No direct equivalent</td>
</tr>
</tbody>
</table>
For twelve out of the fifteen DC elements there is an equivalent IAFA element. The IAFA schema was designed to catalogue internet resources (articles and books online), while Dublin Core describes general webpages. Though they both have somewhat different targets, they are trying to follow a standardised set of elements.

But is it really necessary? I am of the belief that in most cases the existing schema should be left as they are, because they are created that way for a reason. For any new information or document, the appropriate metadata should be chosen by the location or related information. For example, if it’s a webpage then one of the many pre-existing schema should be chosen. However, if it’s going to be part of a new software system then a new set of tailored metadata would most likely be more suitable.

I don’t think either approach can be said to be better than the other. The chosen approach for any particular information system depends on it’s purpose, the users, and the type of information that is being stored.
References


http://www.ukoln.ac.uk/metadata/interoperability/dc_iafa.html


