

## Introduction

This report explores issues encountered while creating metadata records using two different metadata schemas for an assigned set of five web-based resources.

The following five resources were used for this exercise. Each resource is followed by a brief identifier in bold which is used to reference the corresponding resource for the remainder of the document.

- Introduction to Metadata: Pathways to Digital Information. Murtha Baca. 1998. [http://www.getty.edu/research/conducting\\_research/standards/intrometadata/](http://www.getty.edu/research/conducting_research/standards/intrometadata/) (**Intro to Metadata**)
- Application Profiles: Mixing and Matching Metadata Schemas. Rachel Heery and Manjula Patel. 2000. <http://www.ariadne.ac.uk/issue25/app-profiles/> (**Application Profiles**)
- Indecs Framework. [Website]. <http://www.indecs.org/> (**Indecs**)
- CARL Institutional Repository Pilot Project: Online Resource Portal. CARL. 2003. [http://www.carl-abrc.ca/frames\\_index.htm](http://www.carl-abrc.ca/frames_index.htm) (**CARL**)
- Iterative Design of Metadata Creation Tools for Resource Authors. D. Robertson, E. Leadem, A. Crystal, and J. Greenberg. 2003. [http://www.siderean.com/dc2003/202\\_Paper82-color-NEW.pdf](http://www.siderean.com/dc2003/202_Paper82-color-NEW.pdf) (**Iterative Design**)

Metadata records for each resource were created with NoteTab Lite using supplied templates. One metadata record was created for each resource using the Electronic Resource Citation (ERC) metadata schema. Then one metadata record was created for each resource using the Dublin Core (DC) metadata schema.

## Electronic Resource Citation Records

The following records represent the five resources using the ERC metadata schema. The records are listed in the order in which the associated resources are presented in the Introduction of this document.

```
who: ,Baca, Murtha|, Gill, Tony|, Gilliland-Swetland, Anne J.|, Woodley,  
Mary  
what: ,Introduction to Metadata: Pathways to Digital Information  
when: 2000  
where: http://www.getty.edu/research/conducting\_research/standards/intrometadata/index.html
```

```
who: ,Heery, Rachel | ,Patel, Manjula  
what: ,Application profiles: mixing and matching metadata schemas  
when: 2000 09 24  
# Publication date used rather than date file modified  
where: http://www.ariadne.ac.uk/issue25/app-profiles/
```

**who:** ,indecs Framework Ltd. | ,Dorner, Jane  
**what:** ,Indecs Home Page  
 # Home page for indecs Framework Ltd.  
**when:** 2001  
 # Latest copyright date  
**where:** <http://www.indecs.org/>

**who:** ,CARL | ,Canadian Association of Research Libraries | ,McColgan, Katherine  
**what:** ,CARL Homepage (English language version)  
**when:** 2005 01 24  
**where:** [http://www.carl-abrc.ca/frames\\_index.htm](http://www.carl-abrc.ca/frames_index.htm)

**who:** ,Greenberg, Jane | ,Crystal, Abe | ,Robertson, W. Davenport | , Leadem, Ellen  
**what:** ,Iterative Design of Metadata Creation Tools for Resource Authors  
**when:** 2003  
 # No date is provided anywhere in the document, however links to this page from other sites identify it as being part of the proceedings from the 2003 Dublin Core Conference.  
**where:** [http://www.siderean.com/dc2003/202\\_Paper82-color-NEW.pdf](http://www.siderean.com/dc2003/202_Paper82-color-NEW.pdf)

## Dublin Core Records

The following records represent the five resources using the DC metadata schema. The records are listed in the order in which the associated resources are presented in the Introduction of this document.

```
<META NAME="DC.Title" LANG="en" CONTENT="Introduction to Metadata: Pathways
to Digital Information">
<META NAME="DC.Title" LANG="en" CONTENT="Introduction to Metadata (Getty
Research Institute)">
<!-- Above title came from page source -->
<META NAME="DC.Creator" LANG="en" CONTENT="Baca, Murtha">
<META NAME="DC.Creator" LANG="en" CONTENT="Gill, Tony">
<META NAME="DC.Creator" LANG="en" CONTENT="Gilliland-Swetland, Anne J.">
<META NAME="DC.Creator" LANG="en" CONTENT="Woodley, Mary">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata standards">
<META NAME="DC.Subject" LANG="en" CONTENT="crosswalks">
<META NAME="DC.Description" LANG="en" CONTENT="This web site presents three
articles. The first introduces and explains the importance of metadata.
The second discusses metadata and standards in relation to the World
Wide Web. The third discusses the costs and benefits to mapping
different metadata standards in order to achieve interoperability. The
site also describes several crosswalks and provides a glossary and list
of acronyms and links to related materials.">
<META NAME="DC.Publisher" LANG="en" CONTENT="J. Paul Getty Trust. Getty
Research Institute.">
<META NAME="DC.Date" LANG="en" CONTENT="2000">
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Type" LANG="en" CONTENT="collection">
<META NAME="DC.Format" LANG="en" CONTENT="text/html">
```

```
<META NAME="DC.Identifier" LANG="en"
  CONTENT="http://www.getty.edu/research/conducting_research/standards/in
  trometadata/index.html">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Rights" LANG="en"
  CONTENT="http://www.getty.edu/legal/copyright.html">

<META NAME="DC.Title" LANG="en" CONTENT="Application profiles: mixing and
  matching metadata schemas">
<META NAME="DC.Title" LANG="en" CONTENT="Main Articles: 'Application
  profiles: mixing and matching metadata schemas', Ariadne Issue 25">
<!-- Above title came from page source -->
<META NAME="DC.Creator" LANG="en" CONTENT="Heery, Rachel">
<META NAME="DC.Creator" LANG="en" CONTENT="Patel, Manjula">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata schemes">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata interoperability">
<META NAME="DC.Description" LANG="en" CONTENT="The authors of this article
  push for a common approach to the way metadata schemas are constructed
  in order to facilitate sharing of metadata.">
<META NAME="DC.Publisher" LANG="en" CONTENT="UKOLN">
<META NAME="DC.Date" LANG="en" CONTENT="2000-09-24">
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Format" LANG="en" CONTENT="text/html">
<META NAME="DC.Identifier" LANG="en"
  CONTENT="http://www.ariadne.ac.uk/issue25/app-profiles/">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Relation" LANG="en" CONTENT="IsPartOf Ariadne Issue 25 ">
<META NAME="DC.Rights" LANG="en"
  CONTENT="http://www.ariadne.ac.uk/about/copyright-new.html">

<META NAME="DC.Title" LANG="en" CONTENT="Indecs Home Page">
<META NAME="DC.Creator" LANG="en" CONTENT="Dorner, Jane">
<!-- Above creator information from source. -->
<META NAME="DC.Subject" LANG="en" CONTENT="Indecs Framework Ltd.">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata">
<META NAME="DC.Subject" LANG="en" CONTENT="e-commerce">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata interoperability">
<META NAME="DC.Description" LANG="en" CONTENT="Provides documentation and a
  framework within which interoperability of metadata for IP in a
  networked environment can be achieved.">
<META NAME="DC.Publisher" LANG="en" CONTENT="indecs Framework Ltd.">
<META NAME="DC.Contributor" LANG="en" CONTENT="Kopiosto">
<META NAME="DC.Contributor" LANG="en" CONTENT="Copyright Agency Limited">
<META NAME="DC.Contributor" LANG="en" CONTENT="Editeur">
<META NAME="DC.Contributor" LANG="en" CONTENT="IFPI">
<META NAME="DC.Contributor" LANG="en" CONTENT="Muze Inc.">
<META NAME="DC.Contributor" LANG="en" CONTENT="The International DOI
  Foundation">
<META NAME="DC.Date" LANG="en" CONTENT="2001">
<!-- Latest copyright date -->
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Type" LANG="en" CONTENT="collection">
<META NAME="DC.Format" LANG="en" CONTENT="text/html">
<META NAME="DC.Identifier" LANG="en" CONTENT="http://www.indecs.org/">
```

```

<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Rights" LANG="en"
  CONTENT="http://www.indecs.org/help/credits.htm">

<META NAME="DC.Title" LANG="en" CONTENT="CARL Home Page (English version)">
<!-- There is also a French version -->
<META NAME="DC.Title" LANG="en" CONTENT="Canadian Association of Research
  Libraries">
<META NAME="DC.Subject" LANG="en" CONTENT="CARL">
<META NAME="DC.Subject" LANG="en" CONTENT="Canadian Association of Research
  Libraries">
<META NAME="DC.Subject" LANG="en" CONTENT="research libraries">
<META NAME="DC.Subject" LANG="en" CONTENT="information policy">
<META NAME="DC.Subject" LANG="en" CONTENT="resource sharing">
<META NAME="DC.Subject" LANG="en" CONTENT="scholarly communication">
<META NAME="DC.Description" LANG="en" CONTENT='Home page of the Canadian
  Association of Research Libraries (est. 1976) to provide leadership for
  the Canadian research library community. "Areas of interest to the
  Association include automation, collections inventory projects,
  copyright, joint purchasing consortia, preservation and resource
  sharing."'>
<!-- May not have the syntax right for embedded quotes. -->
<META NAME="DC.Publisher" LANG="en" CONTENT="Canadian Association of Research
  Libraries">
<!-- Felt more confident that CARL was publisher than creator, so it is here.
  -->
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Type" LANG="en" CONTENT="collection">
<META NAME="DC.Format" LANG="en" CONTENT="text/html">
<META NAME="DC.Identifier" LANG="en" CONTENT="http://www.carl-
  abrc.ca/frames_index.htm">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Relation" LANG="en" CONTENT="IsVersionOf CARL homepage">
<META NAME="DC.Coverage" LANG="en" CONTENT="Canada">

<META NAME="DC.Title" LANG="en" CONTENT="Iterative Design of Metadata
  Creation Tools for Resource Authors">
<META NAME="DC.Title" LANG="en" CONTENT="Microsoft Word - 202 Paper82-color-
  NEW.doc">
<!-- Above title came from page source -->
<META NAME="DC.Title" LANG="en" CONTENT="202_Paper82-color-NEW.pdf
  (application/pdf Object)">
<!-- Above title came from browser's title bar -->
<META NAME="DC.Creator" LANG="en" CONTENT="Greenberg, Jane">
<META NAME="DC.Creator" LANG="en" CONTENT="Crystal, Abe">
<META NAME="DC.Creator" LANG="en" CONTENT="Robertson, W. Davenport">
<META NAME="DC.Creator" LANG="en" CONTENT="Leadem, Ellen">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata, metadata generation
  tools, metadata applications, author-generated metadata, Dublin Core">
<META NAME="DC.Description" LANG="en" CONTENT="Compares v1.0 and v2.0 of the
  NIEHS Dublin Core-based metadata generation tool designed to assist
  authors with metadata creation. Summarizes key findings and identifies
  further research needs regarding these types of tools.">
<META NAME="DC.Date" LANG="en" CONTENT="2003">
<!-- Date is from 2003 Dublin Core Conference site. -->
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Format" LANG="en" CONTENT="application/pdf">

```

```
<META NAME="DC.Identifier" LANG="en"
  CONTENT="http://www.siderean.com/dc2003/202_Paper82-color-NEW.pdf">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Relation" LANG="en" CONTENT="IsPartOf Online text of full
  papers of 2003 Dublin Core Conference">
```

## Discussion

The ERC schema specifies a minimum set of record elements. At least one value must be specified for each of the four basic elements in the first segment of the record (also known as the ‘Anchoring Story’). The only semantics specified are that the selected values be deemed by the record’s creator to be those ‘most appropriate’ for the audience. The basic elements are ‘who:’, ‘what:’, ‘when:’, and ‘where:’. If a value cannot be provided for an element in spite of the creator’s best effort, the element must still appear and the appropriate ‘unknown’ specifier must be included in place of the value.

DC consists of two forms: ‘simple DC’ and ‘qualified DC’. Simple DC is made up of fifteen elements. Each element is given broad semantics to guide the record creator in deciding which values are most appropriate to use. Each element is optional (i.e. no elements are required).

In general, I felt compelled to enter more data in the DC records than I did with the ERC records. I believe this is due to the fact that more basic elements are defined for DC. Likewise, I found it difficult to omit elements in DC even when I felt that they were really not applicable.

It seems that an organization adopting either schema should have a metadata policy. This policy is necessary to establish the ground rules for the metadata records, including which metadata schema to use. I encountered issues with creating records for the assigned resources in all of the areas described below. Many of these issues would likely be addressed by a thorough metadata policy.

## ***Metadata Policy***

### Record Purpose and Audience

A metadata policy should identify the purpose of the metadata records, as well as the audience for the records. This part of the metadata policy answers key questions such as: Is the record intended to accompany a resource in an archive? Is it intended to be a tool for discovering what resources are available? Will the users of the record be the general public finding and selecting resources in a public library? Will the users of the record be a group of professionals selecting resources to solve a problem related to their profession? Will the users of the record be maintainers of a digital archive doing digital data migration to ensure continued access to the resources?

Establishing the purpose of the metadata records and identifying the audience can help the record creator make decisions about which elements to include, which segments/elements/qualifiers to add, which values are most appropriate for each element, etc.

For the assigned resources I considered adding additional segments to the ERC records, but was unable to come to a decision about which segments were most appropriate to add. Having a well defined purpose and audience would provide a basis for making the proper decisions.

In the case of **Iterative Design**, for example, an ERC story to identify that the document was produced as a result of the 2003 Dublin Core Conference might be helpful to users of the record. However, not knowing who those users were or what they would be using the record for prevented me from deciding to include that story.

### Element Semantics

A metadata policy should also refine the semantics of the elements for the given situation. Even in DC, many of the semantics are vague. Take the DC 'Date' and ERC 'when:' elements, for example. Both schemas offer semantics such as: 'a significant date in the object's lifecycle.' Obviously there are possibly many significant dates in an object's lifecycle: date of creation, date of publication, date of copyright, date of last update, etc. Which date is best? A metadata policy can provide a guide to answering that question.

#### *DC Date and ERC when:*

The semantics of the DC 'Date' and ERC 'when:' elements caused me a lot of trouble when creating records for the assigned resources. For example, the **Application Profiles** resource contains two dates 'Publication Date' and 'File Last Modified' date. These dates have no description, and the 'File Last Modified' date is not accompanied by any data describing what was modified on that date, or even what the 'file' is. Is the file just the contents of the article? Is it the entire html page? I chose to use the publication date on the instinct that the 'File Last Modified' date probably indicated updates to supporting materials (fixes to broken links, etc.) If the 'File Last Modified' date is used, and indicates only that a broken link was fixed, it seems that this information would be misleading to the user of the record.

These elements were even more troublesome for the **Iterative Design** resource. There are no dates included in the document. I was unable to backtrack on the Sidrean site to locate a date, and my attempts to locate the document from the Sidrean homepage failed miserably. Finally I searched for the document's title using Google. I found a link to the document from the 'DC-2003' web site: <http://dc2003.ischool.washington.edu/index.html> (follow links to 'Online Text of Full Papers & Posters' and then to 'Metadata Generation'.) As a result, I chose to use the value '2003' for 'when:', but I'm still not confident that it was the best decision.

The only date I could find for the **CARL** resource was the 'date last updated' and it did not seem like a very appropriate date to include in the metadata. This is a significant date in the document's lifecycle, but is really only significant relative to the time the page is viewed. Since the date cannot be updated dynamically in the metadata, it quickly becomes erroneous. For this particular resource I chose different courses of action in regards to the metadata record. For the ERC record, I chose to include the date last updated as the value for the 'when:' element. I'm still not comfortable with this decision, but the ERC documentation seems to imply that you should use whatever date you can find that is significant in the lifecycle of the object and only resort to the unknown values when there are no other options. I omitted the 'Date' element from the DC record for the **CARL** resource because that was an acceptable option in DC.

#### *ERC who:*

In the ERC records, the 'who:' element gave me some trouble. On the **Indecs** resource, I specified both the company name and the page creator (Jane Dorner). For discovery purposes, I'm not sure anyone would care to know that Jane Dorner created the page. However, for archive purposes, it seems like a potentially valuable piece of information. The **CARL** resource gave me

a similar problem as it has the webmaster specified as Katherine McColgan. I arrived at the same conclusion, this is not necessarily an important value for discovery, but it is potentially important for archiving. A metadata policy clarifying the purpose of the records and the audience would guide decisions such as these.

### *DC Creator, DC Publisher, DC Contributor*

The DC elements of ‘Creator’, ‘Publisher’ and ‘Contributor’ caused me some trouble too. The semantics seem clear enough when I read them, but when I go back to try to apply them, I do not feel very confident. For example, I had difficulty determining whether ‘CARL’ was more appropriate as the Creator or as the Publisher for the **CARL** resource. I finally felt that CARL was more responsible for making the site available for access (‘Publisher’) than for creating the content (‘Creator’). This decision was additionally supported by the guidelines which say that when the responsibility is ambiguous to use ‘Publisher’ for organizations and ‘Creator’ for individuals.

I also was not comfortable putting Sidrean as the ‘Publisher’ for the **Iterative Design** resource. The guidelines say that a ‘Publisher’ is “the entity responsible for making the resource available.” I didn’t feel that Sidrean was necessarily “making the page available” simply by storing it on their server and opening it to public access. “Making the page available” seems to imply more. (Perhaps making it findable?) The University of Washington Information School provides a public link to it, but I could not find a similar link from the Sidrean home page. In this case, I was unable to achieve a comfort level with any value for ‘Publisher’, so I omitted the element.

The ‘Contributor’ element only seemed clearly useful in the case of the **Indecs** resource. It seemed very useful to have the names of the companies that compose Indecs Frameworks Ltd. in the record, and “an entity responsible for making contributions to the content of the resource” seemed the perfect description for the role of each of these companies. On the other hand, the **Intro to Metadata** resource listed the authors of its individual articles as ‘contributors’, but I felt that these people and the editor shared equal weight in the creation of the content, so I listed them as ‘Creators’ in my record.

### *DC Type, DC Source, DC Relation*

The DC ‘Type’ element description contains a value of ‘collection’ in the usage guide. This value was not available in the NoteTab template, but it seemed to describe three of my resources well. **Intro to Metadata**, **Indecs**, and **CARL** are all comprised of parts that can stand on their own, so I used ‘Type=collection’ in addition to ‘Type=text/html’ for these resources.

I chose to omit the ‘Source’ element from the **Intro to Metadata** resource because it was unclear how the web site related to the book that was available for purchase from the site. I also chose not to list French as a ‘Language’ for the **CARL** resource because the assigned URL took me directly to the English translation of their page.

DC ‘Relation’ semantics are not entirely clear. For example can ‘IsVersionOf’ represent a translation? Should I have included Relation=IsVersionOf CARL Homepage (English) or something to that effect for the **CARL** resource?

## Controlled Vocabularies

A metadata policy should direct record creators to controlled vocabularies for fields like the DC 'Subject' field and the 'erc-about: what:' field. In spite of following the recommendation to write the DC 'Description' values first, I had difficulties determining the 'correct' keywords to use for DC 'Subject'. A controlled vocabulary and some guidance in what specificity and exhaustivity to target when choosing keyword values are necessary guidelines to achieve the highest level of consistency possible in the metadata records.

## Input Rules

A metadata policy should specify input rules including rules for capitalization and other details. Input rules could also clarify when to use one DC 'Subject' element versus when to repeat the element, as there are circumstances under which the DC schema leaves this decision to the creator of the record. The 'empty record' problem that is inherent with the idea that all DC elements are optional could also be addressed with input rules. A metadata policy could specify which DC elements are required for the given situation.

## **User Tasks**

Metadata records should assist users in four basic tasks: *finding* resources relevant to their needs, *identifying* more specifically what those resources are, *selecting* amongst multiple available resources and *acquiring* the selected resources. Both the ERC and DC metadata schemas can produce records which can be used for all four tasks, but both require clarification and refinement in order to achieve this.

The first user task is finding. I believe that the ERC schema is actually better suited to this task since it has mandatory elements. The optional nature of the DC elements makes it difficult to predict which elements will exist and, thus, which data to search for. In addition, the theoretical 'empty' DC record cannot be found at all! The problem with the ERC schema in regard to the finding task is the lack of semantics for the elements. Qualifiers may help to mitigate this problem for systems that can understand them.

It seems that, if used to their full extent, both schemas can provide adequate information for identification and selection. The ERC schema segments and provisional and local qualifiers definitely enhance the ability to relay information helpful for identification and selection of the represented resources. The issue remains that the information cannot be used if the record creator chooses not to provide it. The DC metadata schema might have an advantage in that all 15 simple elements are given equal weight. Since any element may be used or omitted, it seems that all 15 should be reviewed by the record creator which might result in more extensive identification information. Again, the difficulty is in the ability to omit any of the elements. If the data is not there, it cannot be used to identify and select.

The acquisition task can be accomplished by either schema equally well. One caveat being the optional nature of the DC 'Identifier' element. An additional consideration is the close relationship between the ERC metadata schema and the development of the Archival Resource Key (ARK) persistent identifier scheme. If the ARK scheme becomes widespread, the ERC metadata schema will be an easy metadata solution for networked resources.

### ***Ease of Use and the ERC Metadata Schema***

An advantage to ERC is the intuitive nature of the basic elements. We are taught who, what, when, where (and how) from a young age. However, the ‘what:’ and ‘where:’ elements in this exercise seemed more straightforward than the ‘who:’ and ‘when:’.

If more than one ‘who:’ is used, it cancels the opportunity to create a minimal ERC as described by Kunze. A minimal ERC is one where the labels are removed and the values are separated only by a pipe (‘|’). Since multiple values for a single element are also separated by a pipe, ambiguity would be introduced if multiple values were used for a single element in a minimal ERC. So, the problem arises: Which ‘who:’ is the best ‘who:’? However, this is only an issue if the minimal ERC is desired.

The issues with the ‘when:’ element are discussed in detail in a prior section of this document.

### ***Ease of Use and the DC Metadata Scheme***

I found the ERC schema easier to use than the DC schema primarily due to the fact that DC provides semantics. I found myself rereading the semantics repeatedly to ensure I was entering the appropriate values. This is an interesting problem. In my case, as a record creator, I was more likely to include additional data about a resource because there were more basic elements presented for consideration in DC. On the other hand, I found it more difficult to use and more time consuming for exactly the same reasons.

## **Summary and Conclusion**

Many of the issues identified in the process of creating metadata records revolve around the vague semantics provided by the schema definitions. This might seem to indicate a need for more detailed semantics for the schemas. On the contrary, I believe that the vague nature of the semantics is, overall, a positive aspect of these schemas. The enormous variety of information resources that must be defined using metadata necessitates some imprecision. A clearly documented metadata policy can solve many of these issues.

On the other hand, this very imprecision makes it difficult for an outsider to interpret the metadata very precisely. Is the value specified for ERC ‘when:’ or DC ‘Date’ the date of publication? Perhaps it is the date of the last modification? Maybe it is the date the document was originally created? Possibly all that is needed is skilled application of qualifiers and a method of identifying which policy was used to create a given record.

Both ERC and DC are very powerful. I favor ERC over DC for two reasons: ERC is simple, and I believe that the mandatory ‘Anchoring Story’ segment is important. We have been taught about who, what, when, and where from a young age. They are easy to remember and applicable in some way to almost everything. This simplicity makes it easy to create a very basic core record for any object. In addition, a mandatory core set of elements will help facilitate discovery of resources without previous knowledge of their existence.