METADATA: AUTOMATIC GENERATION AND EXTRACTION

Noufal P.P
Indian Institute of Technology Kanpur
(noufal@iitk.ac.in)

ABSTRACT

Today, library technical services departments have a new imperative: to carry forward their role to organize the world's information and apply it in the digital age. At the heart of the necessary transition is envisioning the future of technical services as a future with millions and millions of digital objects and e-resources in it. It is therefore a future with metadata in it. The world of information technology is awash in talk of metadata. Metadata can be thought of as a lubricant for information flow- easing the difficulty of discovery and organization of resources. It can also serve as a nozzle -- directing, channeling, and focusing information flow to make it more manageable and effective. Producing metadata has long been the task of professionals such as librarians and indexers. Nowadays metadata is also being produced by non-specialists (e.g., ordinary users), publishers, commercial agents, and even software systems. Creating metadata is often labour-intensive. Though automated procedures and new types of tools for creating and managing metadata are evolving and will continue to become more sophisticated. Some word processors, editors, and format filters generate embedded metadata tags when a document is first created. Other tools automatically extract metadata values from the documents themselves. This paper explores the features of Dublin core metadata and tools for automatic metadata extraction from electronic resources.

Metadata defined

First, let us define metadata. The common definition of metadata is ''data about data.'' The notion of metadata is relative. It is only really meaningful in a context that makes clear what the data itself is. Metadata describes the attributes of an information-bearing object (IBO) - document, data set, database, image, artifact, collection, etc.; metadata acts as a surrogate representation of the IBO. A metadata record can include representations of the content, context, structure, quality, provenance, condition, and other characteristics of an IBO for the purposes of representing the IBO to a potential user - for discovery, evaluation for fitness for use, access, transfer, and citation. According to W3C, metadata is data that provides information about, or documentation of, other data managed within an application or environment. Metadata is also machine understandable information for the web. In the digital environment, the surrogate role of metadata is key because many resources are not easily browsable and others do not carry clear data about themselves. The rise of interest in metadata is part of the effort to organize messy world of digital resources and to provide access and services for users at the right time. It is also a way to exchange data between disparate stores of resources and to allow searching across digital warehouses.

Metadata: - Types

All of these perspectives on metadata become important in the development of Digital Information System. But they lead to a very broad conception of metadata. Different kinds of metadata are designed for different purposes: - *Administrative metadata* - this is to use to manage and administer information resources, as acquisition information, documentation of legal access information etc. *Descriptive metadata* - this is to describe or identify information resources as cataloguing records, annotation by users, finding aids and specialized index etc.

Preservation metadata –This is related to the preservation management of information resources as documentation of physical condition of resources, documentation of actions taken to preserve physical and digital versions of resources, e.g., data refreshing and migration. *Technical metadata* - Metadata related to how a system functions or behave, hardware and software documentation, Digitization information, authentication and security data, e.g., encryption keys, passwords.

Metadata – Why?

Metadata serves many important purposes like data description, data browsing and data transfer. It provides additional context that makes a document or record more meaningful, accessible, and useful. Discussions of metadata in the library community have largely centered on issues of resource description and discovery. There is, however, a growing awareness that metadata has an important role in digital resource management, including preservation. Regardless of whether emulation-based or migration-based preservation strategies are adopted, the long-term preservation of digital information will involve the creation and maintenance of metadata. Properly used metadata can identify the name of the resource, the creator, who reformatted it, and other descriptive information. We can discuss the various roles of metadata in digital information system as follows: -

***Increased accessibility*:** Effectiveness of searching can be significantly enhanced through the existence of rich and consistent metadata. Metadata can also makes possible to search across multiple collections or to create virtual collections from materials that are distributed across several repositories, but only if the descriptive metadata are the same or can be mapped across each site. Digital information systems and emerging metadata standards developed by different professional communities such as Encoded Archival Description (EAD), Text Encoding Initiative (TEI), and the Dublin Core are making it easier for users to negotiate between descriptive surrogates of information objects and digital versions of the objects themselves, and to search at both the item and collection level within and across information systems.

***Interoperability*:** Describing a resource with metadata allows it to be understood by both humans and machines. Interoperability is the ability of multiple systems with different hardware and software platforms, data structures, and interfaces to exchange data accurately. Using defined metadata schemes, shared transfer protocols, and crosswalks between schemes, resources across the network can be searched more seamlessly. Two approaches to interoperability are cross-system search and metadata harvesting. The Z39.50 protocol is commonly used for cross-system search. Z39.50 implementers do not share metadata but map their own search capabilities to a common set of search attributes. A contrasting approach taken by the Open Archives Initiative is for all data providers to translate their native metadata to a common core set of elements and expose this for harvesting. A search service provider then gathers the metadata into a consistent central index to allow cross-repository searching regardless of the metadata formats used by participating repositories.

***Multi-versioning* :** Objects enter a digital information system by being created digitally or by being converted into digital format. Multiple versions of the same object may be created for preservation, research, dissemination, or even product development purposes. The creator may include some administrative and descriptive metadata for this purpose.

Legal issues: Metadata allows repositories to track the many layers of rights and reproduction information that exist for information objects and their multiple versions. Metadata also documents other legal or donor requirements that have been imposed on objects - for example, privacy concerns or proprietary interests.

Preservation: If digital information objects that are currently being created are to have a chance of surviving migrations through successive generations of computer hardware and software, or removal to entirely new delivery systems. They will need to have metadata that enables them to exist independently of the system that is currently being used to store and retrieve them. Technical, descriptive, and preservation metadata help to know how a digital information object was created and maintained, how it behaves, and how it relates to other information objects. It should be noted that for the information objects to remain accessible and intelligible over time, it would also be essential to preserve and migrate the metadata also.

System improvement and economics: Metadata is also helpful to evaluate and refine systems in order to make them more effective and efficient from a technical and economic standpoint. The data can also be used in planning for new systems.

Metadata Characteristics and standards

Some of the common characteristics of all metadata include semantics, syntax and structures. Semantics refer to type and content of metadata elements, syntax refer to the way in which content is structured according to a specifies grammar – metadata standard may range from simple syntax like Dublin core metadata element set to a complex coding systems such as mark up languages like SGML. While structures refer to the over all architecture that contain metadata content and syntax. Metadata may be embedded in the digital object and extracted as needed or they may reside in separately indexed databses.the data can be contained in a variety of architectural structures including Z39.50 compliant library catalogues, proprietary databases or Resource Description Frame work (RDF) standard [7].

Interoperability and extensibility are desirable characteristics of metadata standard systems. Interoperability refers to the ability to transfer metadata among different schemes and information systems. A core element set that is common to all will facilitate the exchange and use of metadata at a general level and description and retrieval. Extensibility allows expansion of the core element set to provide more precise description and retrieval. Element qualifiers are used to narrow the meaning of an element. Metadata schema or schemas should be applied in order to best meet the needs of the information creator, repository and users; At the time of planning of DL system, we should think carefully about which metadata attributes and format is going to use. If you do not, you may find yourself constrained by the absence of useful metadata, or have to add a new metadata field or convert an existing field to a different format when you already have several thousand resources in your database. Moreover, decisions about metadata will in turn affect the design of your interface especially the parts of it used for cataloguing and/or submitting new resources for consideration. Before selecting metadata standards we have to keep the following points in our mind.

- Deciding which aspects of metadata are essential for what they wish to achieve, and how granular they need each type of metadata to be - in other words, how much is enough and how much is too much. There will likely always be important tradeoffs between the costs of developing and managing metadata to meet current needs, and creating sufficient metadata that can be capitalized upon for future, often unanticipated uses;
- Ensuring that the metadata schemas being applied are the most current versions

Structuring Metadata

Metadata schemes (also called schema) are sets of metadata elements designed for a specific purpose, such as describing a particular type of information resource. The definition or meaning of of the elements themselves is known as the semantics of the scheme. The values given to metadata elements are the content. Metadata schemes generally specify names of elements and their semantics. Optionally, they may specify content rules for how content must be formulated (for example, how to identify the main title), representation rules for content (for example, capitalization rules), and allowable content values (for example, terms must be used from a specified controlled vocabulary). There may also be syntax rules for how the elements and their content should be encoded. A metadata scheme with no prescribed syntax rules is called syntax independent. Metadata can be encoded in any definable syntax. Many current metadata schemes use SGML (Standard Generalized Mark-up Language) or XML (Extensible Mark-up Language). XML, developed by the World Wide Web Consortium (W3C), is an extended form of HTML that allows for locally defined tag sets and the easy exchange of structured information. SGML is a superset of both HTML and XML and allows for the richest mark-up of a document. Useful XML tools are becoming widely available. XML plays an increasingly crucial role in the exchange of a variety of data on the Web.

Frameworks of Metadata

There are several metadata frame works. Most of web authors insert metadata through the use of "META"tags in HTML. A META tag can have a name and a content attribute with the value for the content determined by web authors. In the example below, the META tags provide a general description for this article.

<META name= "description" content= "This article describes about automatic metadata extraction">

<META name = "keywords" content = "Digital Libraries, Metadata, Metadata extraction, Dublin Core" >

Unlike a cataloguer, web authors who insert META tags into their documents often will not be conversant with controlled subject schemes. For this reason, there are tremendous interests in metadata formats that do not require special skills to create or maintain them. There are a number of metadata frameworks like Dublin Core, Warwick, EAD, IMS MODS, METS and MPEG -7

Dublin Core

Among the various metadata standards, it seems that Dublin Core Metadata Initiative (DC for short) has gained the special importance among the resource description communities. Within the diverse resource discovery activities of the mid 90s, ranging from unstructured indexing of full-text resources by search engines to richly structured data like MARC and TEI records, DC arose as a means to mediate these extremes [9]. DC developed in the *March 1995 Metadata Workshop* sponsored by the Online Computer Library Center (OCLC) and the National Center for Supercomputing Applications (NCSA), to advance the state of the art in the development of metadata records for networked information resources. One of the main goals of the workshop was to reach a consensus on a simple and core set of metadata elements to describe electronic resources. The result of the workshop was a set of 13-metadata elements, which was called Dublin Core Metadata Element Set (DCMES) for describing what called Document-Like Objects. By the third workshop the elements set was developed to 15 elements.

The Dublin Core Metadata Element Set includes: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, and Rights (Dublin Core, 1999).

The characteristics of Dublin Core that distinguish it as prominent candidate for description of electronic resources fall into several categories: simplicity, semantic interoperability, international consensus, and flexibility (Weibel, 1997). The Dublin Core is intended to be a maximum simplicity and flexibility. Simplicity of Dublin Core is its main feature and originates from this fact that the elements of Dublin Core designed to be used by the creators of the resources, who are not trained catalogers or have no knowledge of cataloging, to describe the resources. Since Dublin Core provides internationally agreed upon elements that are commonly understood among various communities and fields, it promotes some level of semantic interoperability. The interoperability of Dublin Core enables resource description records created using it to be implemented across disciplines and in different fields. On the other hand, Dublin Core can be qualified and extended to meet the requirements of a wide variety of communities. Though it concentrates on describing intrinsic properties of the information object, the extension mechanism will allow the inclusion of extrinsic data for objects that cannot be adequately described by the small and simple set of Dublin Core elements. It is possible to encode many controlled vocabularies and description standards such as LCSH, MeSH (Medical Subject Headings), DDC, and LC in Dublin Core elements through Dublin Core qualifiers (Dublin Core, 2000).

Among the current metadata standards, Dublin Core has the potential of being adapted as an international standard for resource description and discovery on the web and as a *lingua franca* for metadata, partly because of its simplicity. Its simplicity promotes general applicability but also suggests an important problem that is lack of consistency and trust. Regarding the importance of and the need to trust and provenance of data and metadata in the web (Lynch, 2001),

Why Dublin Core?

Dublin Core has received widespread acceptance amongst the resource discovery community and has become the defacto Internet metadata standard [3].

The Dublin Core metadata schema offers the following advantages:

- Its usability and its flexibility
- The semantics of these elements is designed to be clear enough to be understood by a wide range of customers, without the need for training
- The elements of Dublin Core are easily identifiable by having the work in hand, such as intellectual content and physical format
- It is not intended to supplant other resource descriptions, but rather to complement them. It is intended to describe the essential features of electronic documents that support resource discovery. Other important metadata such as accounting and archival data, were deliberately excluded to keep the schema as simple and useable as possible.
- It is mostly syntax independent, to support its use in the widest range of applications
- All elements are optional, but allows each site to define which elements are mandatory and which are optional
- All elements are repeatable
- The elements may be modified in limited and well-defined ways through the use of specific qualifiers, such as the name of the thesaurus used in the subject element
- It can be extended to meet the demands of more specialized communities. From the very beginning, the Dublin Core creators recognized that some resources could not be adequately described by a small set of elements. The Dublin Core creators came up with two solutions. Firstly, by allowing the addition of elements for site-specific purposes or specialized fields. Secondly, by designing the Dublin Core schema so that it could be mapped into more complex and tightly controlled systems, such as MARC.

Below is an example for Dublin Core metadata elements created for this article.

```
<HTML>
<HEAD>
<TITLE>Automatic Metadata Extraction  </TITLE>
<META NAME="DC.Title" CONTENT=" Automatic Metadata Extraction">
<META NAME="DC.Creator" CONTENT="Noufal">
<META NAME="DC.Type" CONTENT="text">
```

```
<META NAME="DC.Date" CONTENT="2005">
<META NAME="DC.Format" CONTENT="text/html">
</HEAD>
```

Automatic Metadata Creation and Extraction

Nowadays automatic metadata extraction is a hot research topic. Over the past several decades computer tecquniqes for text analysis have been developed that can achieve impressive result in constrained domains Structured markup languages such as XML help make as key aspects of documents accessible to computers and people alike. They encode certain kind of information in such a way that it can be extracted by parsing the document structure. Basic metadata about a document – its title, author, publisher, date of publication, keywords and abstract- is often present in the first page of publication for all to see. Moreover it is frequently presented in fairly uniform way. Automatic metadata extraction depends too much on the format and style of documents.

Metadata generation is the act of creating or producing metadata. Generating good quality metadata in an efficient manner is essential for organizing and making accessible the growing number of rich resources available on the Web. The success of digital libraries is mainly depends on interoperability. Metadata creation and management have become a very complex mix of manual and automatic processes and layers created by many different functions and individuals at different points in the life of an information object. This metadata can be contained within the same envelope as the information object - for example, in the form of header information for an image file, e.g. Dublin Core, or through some form of bundling, for example, with the Universal Preservation Format (UPF). [1] Metadata can also be attached to the information object through bi-directional pointers or hyperlinks, while the relationships between metadata and information objects, and can be documented by registering them with a metadata registry. As systems designers increasingly respond to the need to incorporate and manage metadata in information systems and to address how to move them forward through time, many additional mechanisms for associating metadata with information objects are likely to become available. Metadata registries and schema record keeping systems are also more likely to develop as it becomes increasingly necessary to document schema evolution and to alert implementers to version changes [4].

Major challenges for Digital Library projects and inititatives are the growing number of resources requiring metadata. Library cataloguers and indexers generate metadata are prohibitive due to the limited availability of both qualified persons and financial recourses. Author-generated metadata projects appear to be less popular in the organizational setting. The greatest problem of the author-generated metadata is the inability to relay on its accuracy. The rationale given for this predicament is that authors lack the metadata professional's expert skills, and they will therefore produce insufficient and poor quality metadata (Weinheimer 2000). In this context, automatic metadata extraction is becoming more popular and acceptable among Digital library communities. According to Anderson & Perez-Carballo, automatic metadata extraction is more efficient, less costly and more conversant than human oriented process. Researches indicate that automatic metadata generation can produce acceptable results for subject metadata (Liddy. Et al. 2001)[8]. Researchers have concluded that the most effective mean of metadata creation is to integrate to the human and automation methods (Schwartz, 2000).

Metadata may be generated or updated at various times in the life cycle of a resource, so workflows must be designed accordingly. Declarative rules or prescriptions can be used to explicitly describe how, when, and by whom a metadata element will be produced or updated. Research on metadata creation and management will enable systems to use document analysis to automate metadata capture, automatically extract subject classifications, trigger the update of metadata when a resource is modified, support the capture of temporal metadata, certify metadata at the time of publication, and remove metadata when a resource is obsolete. Such tools must become integrated into a variety of environments from Web site management utilities to databases, data warehouses, and legacy resources. If used on a wide scale, they could significantly improve the quality and cost-effectiveness of metadata in the networked environment. There are some commercial and open source tools are available for metadata creation and extraction.

Tools for metadata generation

This paper mainly concentrates on Dublin core metadata generation and extraction. Because Dublin core is widely accepted and popular among DL communities. It gives great flexibility, simplicity and ensures interoperability with others schemes. There are two types of tools that facilitate the creation of Dublin Core metadata.- Editors and generators

Templates are basic *crib* sheets that sketch a framework or provide an outline of schema elements without linking to supporting documentation. Templates have been predominant in metadata generation, probably because they are simple to produce and maintain. These tools guide metadata creation through the provision of a form without the bells and whistles. The MARC bibliographic form supporting cataloging in many second-generation online catalogs has functioned in much the same way, without any sort of automatic linking to authority files and content guidelines. This facility is fortunately changing as many catalogs become Web-based and hyperlink to cataloging documentation, thus functioning more like an *editor.*

Editors are similar to templates in that they require human input. They are more sophisticated in that they provide direct access to standards and documentation underlying metadata creation. These tools often assist with syntactical aspects of metadata creation via automatic means. One of most popular Dublin Core editors is the Nordic Dublin Core Metadata Template (www.lub.lu.se/cgi-bin/nmdc.pl). This editor supports the generation of metadata records with HTML META tags for embedding in the header of a Web resource. The Nordic Template has been adapted to many different Dublin Core projects*.* A partial list of them is at*:* http://dublincore.org/tools/. Another example is the Reggie Metadata Editor (http://metadata.net/dstc/), which allows for metadata to be generated within RDF. Editors can also include off-the-shelf software like Metabrowser (http://metabrowser.spirit.net.au/), which hyperlinks to documentation supporting metadata generation and automatically provides the correct syntactical encoding. People work with a wide variety of Web forms when joining an organization, posting information on an online community bulletin board or purchasing a product over the Internet. All of these forms function as metadata editors documenting transactions, activities, events and other types of objects beyond the traditional information resource*.*

***Generators* –**This support automatic metadata production. Generators first require the submission of a uniform resource locator (URL), a persistent uniform resource identifier (PURL) or another Web address in order to *locate the object*. An algorithm is then used to comb an object's content, including its source code, and automatically assign metadata. There are some interesting tools are available. some of them are DC-dot,. Describthis and Klarity

DC-dot

DC-dot (http://www.ukoln.ac.uk/metadata/dcdot/) is a generator developed by UKOLN (UK Office for Library and Information Networking) based at the University of Bath. DC.dot is open source and can be redistributed or modified under the terms of the GNU General Public License as published by the Free Software Foundation. DC-dot produces Dublin Core metadata, and can format output-according to number of different metadata schemas (e.g., USMARC, SOIF, IAFA/ROADS, TEI headers, GILS, RDF, and IMS). Metadata creation with DC-dot is initiated by submitting a URL. DC.dot, copies resource “identifier” metadata from the

Web browser's "address prompt," and harvests "title," "keywords," "description," and "type" metadata from resource META tags. If source code metadata is absent (meaning META tag are absent), DC-dot will automatically generate "keywords" by analyzing anchors (hyperlinked concepts) and presentation encoding, such as bolding and font size, but will not produce "description" metadata. DC.dot also automatically generates "type," "format" and "date" metadata, and can read source code programming that automatically tracks date. For example, "last modified" might be coded as: "Last Modified" + lm_day+' '+monthName[lm_month-1]+' ''+lm_year" for last updated date. This example illustrates DC.dot's ability to partially conform with the recommended Dublin Core Qualifiers.

Describethis

DescribeThis (http://www.describethis.com), a service designed for the automatic extraction of metadata from online resources. The site offers an easy to use interface where you can indicate the resource to analyze and how to download the results as XML, XHTML or RDF files. In the current version, the site's engine is able to find the resources to process using keywords, full URLs or more complex queries with operators, like "ISBN", used to collect the bibliographic data for published documents (see http://www.describethis.com/help/search.html). In the first case it works as a meta search engine using other search engines to locate the best sites where the resource can be found. The results returned back contains all the recognized and generated Dublin Core elements for the requested resource and can be downloaded as RDF, XML or XHTML collections.

Describethis's main fields of applications are to support and extend the application and development of the Dublin Core format as one of most appropriate metadata standards to describe or catalog resources, digitals or not, to use the site as a tool to support the cataloguing of online resources, oriented to information specialists and Internet users in general. And To deliver services of automatic metadata management, designed for managers of bibliographic and content databases and to create an efficient way for administrators and website authors to dynamically provide metadata information about their sites to page crawlers, bots, spiders, agents, worms and other automatic indexing and site classification systems, with the aim of contributing to the improvement of the whole Internet content organization.
In the describethis.com front page you can find several samples to illustrate the normal operation of the service. In this version, DCS can automatically analyze and to generate metadata registers for the following formats:

- HTML and XHTML Documents
- Dublin Core/RDF
- Dublin Core/XML
- Dublin Core/HTML (META tags)
- GIF, JPG (EXIF) and other image formats
- RSS
- bibTex

Support for other well-known formats like PDF, MARC , stream formats (MP3,MPEG, etc), OAI directories, FOAF networks and others will be added in the near future

Klarity

Klarity is a commercial service that was produced by the Australian company tSA, now named Intology (Intelligent Technology) (http://www.intology.com.au/). The Klarity archive is found at: http://archive.klarity.com.au/. Dublin Core metadata is automatically generated when an identifier (e.g., URL) is submitted, and the metadata is converted into HTML META tags or eXtensible Markup Language (XML) within the Resource Description Framework (RDF)

(Brickley and Guha, 2004). Klarity automatically generates metadata for the following five elements: Identifier, title, concepts, keywords, and description. “Identifier” metadata is copied from the Web browser’s “address prompt,” “title” metadata is harvested from the resource source code, and “keywords” and “description” metadata are extracted from resource text. Klarity’s “concepts” element is a unifying concept representing keywords and functions more like a classificatory node. An algorithm based on term frequencies is matched against an underlying vocabulary to create this element. Klarity’s editor feature, where a human can manually enter additional metadata by answering a series of questions

Conclusion

There is an interesting study of features of DC dot and Klarity by Prof. Greenberg. He concluded that generators have the potential to create useful metadata. Major challenges for Digital Library projects and initiatives are the growing number of resources requiring metadata. Library cataloguers and indexers generated metadata are prohibitive due to the limited availability of both qualified persons and financial recourses. Author-generated metadata projects appear to be less popular in the organizational setting. The rationale given for this predicament is that authors lack the metadata professional’s expert skills, and they will therefore produce insufficient and poor quality metadata. In this context, automatic metadata generation and extraction is the right option. It is more efficient and less costly than practicing involving human workforce. However the best metadata generation process is to integrate human and automatic process.

Acknowledgements

I would like to place on record my sincere gratitude to Shr. R.Mishra, Librarian, Indian Institute of Technology Kanpur for the encouragement and support received from him throughout.

References

1. Conway, Paul. Preservation in the Digital World. Washington, DC: Commission on preservation and access, 1996
2. Day, M. Metadata in nutshell. Information Europe 2001; 6(2) (Summer): 11
3. Dublin Core Metadata Element Set, Version 1.1: Reference description. (2005): http://dublincore.org/documents/2005/01/10/dcmi-terms/
4. Greenberg, J., Crystal, A., Robertson, D., & Leadem, E. (2003a). Iterative Design of Metadata Creation Tools for Resource Authors. In Sutton, S., Greenberg, J., and Tennis, J. (Eds.). 2003 Dublin Core Conference: Supporting Communities of Discourse and Practice Metadata Research & Applications. September 28 - October 2, 2003, Seattle, Washington. New York: ERIC Clearinghouse on Information and Technology: http://www.siderean.com/dc2003/202_Paper82-color-NEW.pdf.
5. Greenberg, J. (2003). Metadata and the World Wide Web. In Encyclopedia of Library and Information Science, 2nd ed. vol. 3 (New York: Marcel Dekker, 2003): 1876-88
6. Hlava, M. (2002). Automatic indexing: A matter of degree. Bulletin of ASIS&T, 29(1):
7. Ian H, Witten , David Bainbridge, 2003, How to build a digital library, Morgan
8. Liddy, E. D., Sutton, S. A., Paik, W. Allen, E., Harwell, S. Monsour, M., Turner, A, & Liddy J. (2001). Breaking the metadata generation bottleneck: preliminary findings. JCDL 2001: 464.
9. Nangao, Katashi, 2003, Digital content annotation and transcoding, Artech house
10. http://www.intology.com.au
11. http://dublincore.org
12. http://www.describethis.com),
13. http://www.ukoln.ac.uk/metadata/dcdot/