

## Resolving the confusion over metadata-creation in digital archives

Jiban K. Pal

Scientific & Technical Professional, Library, Documentation & Information Science Division,  
Indian Statistical Institute, 203 B. T. Road, Kolkata - 700108, India, Email: jiban@isical.ac.in

*Received: 13 October 2015; revised 20 May 2016; accepted 21 May 2016*

Metadata is a growing concern in every sphere of information management, especially for libraries worldwide. Indeed the Internet has rigorously prompted the revitalization of metadata with more sustainable framework to obtain the efficient discovery of resources for reuse. This paper aims to present a bird's-eye view on metadata creation, thus provide glimpses of diverse opinions in producing structured metadata consistently across the digital repositories. It recognizes the obvious inquiries viz. how to create metadata; whether it can be created manually (using human-intellectual efforts) or by automatic means; who is responsible to create it; are they resource-authors or metadata-professionals; who can produce better quality metadata. Discussion brings out a few paradoxical views on the process of metadata creation and responsibilities by the concerned players. The paper dwells on a variety of practices having their own merits and arguments, and draws attention to the responsibility of metadata creation, a fuzzy area. The paper argues that best results could be achieved by integrating both automatic and manual efforts. Finally, better possible ways of metadata creation are suggested.

**Keywords:** Information management; Metadata engineering; Metadata creation; Strategic planning; Metadata quality; Automated extraction; Author-generated metadata; Metadata professional; Integrated system-design; Digital transformation; Digital repository.

### Introduction

Approaches to organizing information as well as library cataloguing practice can be traced back to 1830s, when Sir Anthony Panizzi's 91 rules became implemented in catalog entries for better consistency. Gradually, card catalog came into existence at the time of industrial revolution due to prevalence in production of printed materials. However the changes over time had transformed the library catalogs into many forms, mostly by related evolution of libraries. But a radical change happened when computer technology made catalog cards into bibliographic records, and subsequently allowed the creation of Online Public Access Catalog in 1980s.

Gradual changes in technology have offered the opportunities to change in the workflow of cataloguing practices. Eventually the bibliographic records were available through distributed network environment and libraries became integrated into the virtual information world. However in twenty-first century, an unprecedented change has been occurred in electronic cataloguing – metadata (refers to a set of descriptors) is here to stay and evolve. Indeed

metadata is perceived to be essential for the librarians in pursuing long-term management and preservation of digital objects.

### Metadata explosion

There has been an explosion of metadata activities; especially on standards, schemas, identifiers, vocabularies, ontologies, registries, repositories, crosswalks, harvesting, etc. These activities are being reached increasingly often elsewhere. No doubt, metadata plays a pivotal role in organizing, managing, preserving, discovering and sharing access to digital information assets world over. This is equally important for system managers, web-designers, service providers, digital archivists, resource authors and seekers of electronic information. The metadata is essential for resource discovery and also life-blood of e-commerce. Thus everyone in the modern information society realizes the absolute potential of it.

Over two decades, several metadata initiatives have been taken across the globe. Dozens of metadata standards have greatly expanded the access to digital

information. Many of those standards are still persistent to meet a variety of needs with the hierarchy of complexity. Increasingly we are having conferences and round-tables on metadata, even online web-forums of metadata have become very active. More often the working cataloguers are called upon to contribute as metadata librarian; thus creating metadata for digital repositories, selecting standards, identifying harvesting tools, assigning local application guidelines, etc. Nevertheless, metadata is being considered an essential phenomenon for electronic-cataloguing, federated-searching, and open-URL's.

Apart from its manifold activities, metadata offers tremendous versatility that enables better relevance in discovering resources from the noisy world of information. Metadata behaves just like a secret sauce. When it works, no one can trace it (i.e. quite invisible), but it is primarily the inherent force of resource discovery. Thus metadata itself is a service, which will have far-reaching impact on virtual information world. Hence it needs to be customized and managed carefully in a regular basis towards realizing the maximum benefits. Therefore, operative guidelines have become obvious in extracting metadata with quality and consistency that can be accomplished by standard metadata schema. Such accomplishments ensure compatibility and facilitate interchangeability of sources across the global information system.

There is no doubt, metadata endeavors are gaining momentum with varied standards. Technological capability now allows multiple schemas for producing metadata in complex digital environments. However, semantic web technologies could enable the integrated use of specialized metadata in order to improve the precision by standardizing structure and content of cataloguing (or indexing) information. Further, linked-data technology has become operative to enhance sharing-of-information on the Web, which really expanded the context of metadata creation in the realm of metadata explosion.

### Rationale of the study

Metadata is amazingly a growing concern in every sphere of information management, especially for libraries worldwide. By and large, the use of metadata is desired to make an efficient way of characterizing the digital objects for discovery. Thus metadata

initiates to obtain the digital resources available on the Internet for reuse. In reverse, Internet rigorously prompted the re-vitalization of metadata with more sustainable framework. However, increasing volumes of increasingly valuable metadata substantially demanded for quality to ascertain the discovery of more relevant objects in a timely manner.

Worthy to mention, a few ambitious projects on metadata unification and extraction have '*gone the way of all flesh*' or become spoiled. Most of the researchers have conceptualized the metadata creation issues without overall coordination. Even a number of studies have reported with less confidence. Indeed they raised much controversy on the creation of metadata. Therefore, my enthusiasm has been pursued for a discussion with regard to a few paradoxical views on metadata creation process, and the responsibilities of concerned players in creating metadata for digital archives.

Current metadata research has been continuing the tradition of cataloguing research. It emphasizes on the process of metadata creation, tools for automated extraction, and commitment of the creators in producing adequate quality metadata to enable interoperability and standardization. Although a number of studies have concentrated on this issue, but no such uniform standard and coherent practice of metadata creation has been found in the global scenario. Rather many of them have created severe ambiguities, instead of making a standpoint on how to create the best quality metadata or what could be the best possible way of metadata creation in digital archives.

Therefore, it becomes imperative to deal with the obvious inquiries; viz. How to create metadata in digital archives? Whether it can be generated through automatic or traditional means? Who is really responsible to create metadata for digital repositories? Who can create a better quality metadata? Notably the aforesaid queries will have widespread implications as far as metadata is concerned, particularly to facilitate better strategies of metadata creation in a large digital archiving initiative like National Digital Library (NDL) of India.

### **Ambiguities in metadata creation (*Automated vs. Traditional*)**

Since earlier time, metadata creation was obligatory for the libraries and traditionally performed

by the cataloguers or indexers (currently known as metadata professionals) through catalogue entries. Gradually library custodians have experienced that *traditional means* are highly labor intensive and time consuming. Otherwise they realized the limitations of metadata creation in a large decentralized system where dynamic resources are involved. Crystal and Land (2003)<sup>1</sup> opined that, "it would take about 60 employee-years to create metadata for one million documents". Eventually the problems of traditional technique demanded for generating metadata by *automatic means*, which pose a challenge over the manual entries being made traditionally.

Keeping in view of aforesaid discussion, metadata can be created either automatically (using metadata extraction tools), or by traditional means (using human-intellectual efforts). Both the processes have their own merits and arguments. While manual efforts will not sustain in the long-run; then automatic means will not work equally on disparate sets, thus would be error-prone and unreliable. So an inherent controversy primarily exists upon the provisions of metadata creation. However the human intellectual efforts can be pursued either by resource-authors or metadata-professionals (i.e. cataloguers, indexers, etc.); which seems to be another cause of contradiction.

Anderson and Pe'rez-Carballo (2001)<sup>2</sup> viewed that automated extraction of metadata from digital objects is less costly than manual entries. Even they assumed that automated extraction will become more efficient and consistent over the time. The Directorate for Cataloguing of the US Library of Congress has also recognized the overwhelming cost of entering metadata manually (Adams, 2009)<sup>3</sup>. Subsequently they sponsored the Automatic Metadata Generation Applications (AMEGA) project (Greenberg et al., 2006)<sup>4</sup>. In fact a number of devices like search engine spiders, web crawlers, and XML editors produce numerous types of metadata through automatic means.

Practically such devices can generate fairly accurate metadata for some elements viz. date, language, etc. But they fail to produce metadata appropriately for some other elements like creator, subject, geographic code, etc.; especially when it is more intellectually demanded. Besides that, automatic method does not have consistent filtering practice to ensure the quality and credibility of extracted metadata. Obviously some structural factors in

metadata generating software's and search engine spiders bring displeasure in producing optimum quality metadata. Therefore, many systems prefer traditional processing, so as to generate schema-specific metadata using human intellectual efforts.

In spite of the aforesaid limitation, a number of studies have been reported that automatically generated metadata provides acceptable performance<sup>5-8</sup>. Although, most often researchers concluded that the best results are achieved by integrating automated and manual methods<sup>9</sup>. Mayernik (2011)<sup>10</sup> opined that metadata can be created both automated and manual processes. He also explained that automated techniques better perform in text-based documents. But these techniques will abandon at the time of creating metadata from scientific papers those having significant proportion of non-textual data (i.e. mathematical signs and symbols, chemical formulas, engineering designs, etc.).

Progressively an improvised technique of metadata extraction had developed by Kovacevic et al (2011)<sup>11</sup>. This automated system is highly capable of extracting eight types of metadata from scientific papers (formatted in PDF only), based on the SVM classifier and open source tools. Finally he realized that, automatically extracted metadata elements cannot be captured into the database records directly. So an automated system requires substantial human efforts through enormous control of the archivist for possible entries and/or corrections in the repository.

Indeed automated techniques essentially require customization for each new type of data creation instrument and processes, so as to demand staff-expertise on critical components of metadata description. Most of the metadata creation processes thus depends on manual efforts.

#### **Responsibility goes to whom? (*Metadata professional vs. Resource author*)**

Once it is presumed that traditional methods (hereafter referred to manual process vis-à-vis human intellectual efforts) of metadata creation are quite common in real practice; then obvious question arises with regard to the responsibility of metadata creation. Subsequently it implies, who is actually responsible to create metadata for digital repositories; are they resource-authors or metadata-professionals; otherwise, who can produce better quality of metadata.

It can be primarily understood that *metadata professionals* (so called catalogers, indexers, data-entry operators) and *resource authors* (those responsible for intellectual content of digital objects) represent two main classes of metadata creators. Though such responsibility of metadata creation goes to numerous individuals depends on the organizational capacity and settings. The National Science Foundation (2005)<sup>12</sup> has reported the four main actors who play important roles in metadata endeavors.

They are *data creators* (i.e. scientists, scholars, students, and others involved in research); *data managers* (responsible for database operation and maintenance); *data scientists* (information scientists, software engineers, domain experts, curators, librarians, archivists and others involved in mentoring digital data and management of data archives); *data users* (academic and professional communities, government, NGO's, etc.). Additionally Swan and Brown (2008)<sup>13</sup> described the role of *data librarians*; which is primarily confined to librarians dealing with their competency and skill in archiving, preserving, and curation of digital datasets.

Let me consider the first group (i.e. metadata professional) among two main classes of metadata creators. Metadata professionals (viz. cataloguers, indexers, etc.) have their intellectual ability (achieved through training and experiences) and proficiency in the use of content value and descriptive standards. Regularly they are being involved in creating structured metadata by exerting their professional knowledge and technical skills. So they are acquiring better experiences and reinventing newer skills for generating optimum quality metadata consistently.

Although a few researchers have noted problems with inter indexer consistency<sup>14</sup>. Often they produce metadata inconsistently for the same digital object. Ideally professional metadata creators could ensure the efficiency in resource-discovery, and obviously they can produce acceptable quality of structured metadata<sup>15</sup>. But they are limited in their availability, demand high incentives, and too costly; so as to violate the law of parsimony.

On the other hand, resource authors (hereafter referred as authors) make them viable with sole responsibility to create the intellectual content of digital objects. In fact, resource-authors (viz. writers, scholars, painters, artists, etc.) regularly create

metadata for their technical or artistic works in the form of abstract, keyword, etc. to make their objects more visible. As such, metadata representations to be created by resource authors are likewise products of their working epistemologies, which can be performed in different ways in different situations. Eventually they might also be involved in creating metadata with their exciting motivation and dynamic performance.

However, in agreement with author-generated metadata, Greenberg et al (2001)<sup>16</sup> reported that resource authors have an ability to create adequate quality metadata "as they are more intimate with their work, they want their work to be discovered and consulted, they know their audience and can thus describe their resources appropriately". Even he vowed that, in some cases authors may be able to create schema-specific metadata (using Dublin Core), which is of better quality than what a metadata professional could produce.

"Yet there is a perception that author-generated metadata will be of poor quality and may actually hamper rather than aid to resource discovery"<sup>17</sup>. Crystal and Land (2003)<sup>1</sup> viewed that ideally metadata should be entered by the resource authors, but practically authors rarely do that, even when they are provided with appropriate tools. Mayernik (2011)<sup>10</sup> also outlined a study of metadata creation by the resource authors in a digital repository. He observed that resource authors faced a number of typical problems in creating useful metadata. Thereby he suggested for investigating new methods of group-oriented (community-wise) metadata creation by the resource authors.

Typically authors do not have good understanding on metadata creation workflow and metadata harvesting tools. Truly they are lacking relevant experience, otherwise having limited skill for producing structured metadata when compared to professionals. More specifically, author-generated metadata system may have diminishing support as the authors find a minimum time for creating metadata elements; rather they prefer to invest more time for their academic works. Resource authors therefore are not necessarily a good player in creating metadata than professionals. However they feel more comfort, when facilitated by professionals through metadata generation tools having appropriate interfaces to understand.

Worthwhile Greenberg et al (2003)<sup>18</sup> provided more detailed views by extending their previous analyses of author-generated metadata quality<sup>16</sup>, collaborative metadata generation<sup>19</sup>, and interface design for metadata creation<sup>20</sup> from the resource author's perspective. Even they posed an iterative design approach, supported by cognitive walkthroughs for identifying further research needs pertaining to author generated metadata tools. Park and Tosaka (2010)<sup>21</sup> conducted a survey on the state of metadata creation practices (use of schema and interoperability) across digital repositories. They found that despite the proliferation of newer metadata schema; metadata selection is a collection specific consideration, where technological infrastructure and staff-expertise remains a major issue.

Crystal and Greenberg (2005)<sup>22</sup> prescribed that effective information system design can mitigate some of the difficulties sought by resource authors while creating metadata. Many systems provide access to an array of metadata generation tools; include *web-templates* (simple forms), *web-editors* (forms enhanced with documentation), and *metadata-generators* (merely require the submission of DOI or URL for metadata generation). Expectedly these tools facilitate the resource authors in such a manner that dramatically simplify the process of metadata creation in a structured format.

Several agencies (FGDC, EPA, etc.) and open-archives (NDLTD, NEEDS, etc.) have taken a dominating role in developing web-based metadata generation tools (via entry forms or relevant interfaces), as they prefer author generated metadata. In such an orientation resource authors usually create metadata (either by him or under his supervision) at the time of object creation. Certainly this practice makes sense to produce huge amount of consistent and quality metadata in consideration with the economics of hiring professional metadata creators for digital archiving.

Similar endeavors in many institutional repositories are truly indicative for large-scale metadata production and to ensure prospective future of digital transformation. Resource authors need to have good understanding on metadata creation process, its' workflow, and actual use of metadata for intelligent resource discovery in a complex information environment. Notably, it has to be conducted in different organizational contexts with different classes

of authors in order to identify institutional factors that influence metadata creation.

So the metadata creation may be performed by the resource authors partially, eventually endorsed by the professionals too. Otherwise, considering the above discussions, information organizers may have to presume and draw their own conclusions.

### **Summing up the ideas**

Libraries were among the earliest of social institutions to understand the function and value of metadata. Library metadata began as the library catalog, a finding aid for librarians and library users. Over a passage of time, catalog entries are changed its format into bibliographic records, thereafter OPAC, and ultimately interoperable bit-streams driven by something else that can seem mysterious – is the metadata, an emerging toolkit for digital libraries to survive.

So, gradual changes in cataloging have been occurring with time; rather changes were inevitable, and still obvious. Anyone needs to change something does not mean that what he has done was wrong, but it means he is intended to improve something else. Metadata also necessitated the required changes. It is primarily constructed with certain purpose, but increasingly valuable metadata gradually has become versatile and demands to produce it in a structured way. However the Internet has rigorously prompted the re-vitalization of metadata with more sustainable framework towards an intelligent discovery and reuse.

While metadata creation is perceived to be essential as a basis of relevance in retrieval, then integrated use of specialized metadata has become obligatory to improve the precision by standardizing the structure and content of cataloguing information. Indeed metadata creation is now a complex issue as it is demanding far more intellectually than earlier, often combining with semantic-web as well as linked-data technologies. So the creation of metadata is amazingly a growing concern in every sphere of information activities. Indeed all digital-archiving initiatives intend to realize the absolute potential of it.

Therefore, the process of metadata creation for dramatically huge resources is obviously a valid agenda, and the responsibility of metadata creation is also noteworthy. It can be created automatically using metadata extraction tools or by manual process using

human intellectual efforts. However, such responsibility of human-efforts for metadata creation can be pursued either by metadata professionals or may be performed by resource authors. While the manual efforts will not sustain in a long-run, then automatic extraction techniques will not work equally on disparate sets of resources with the hierarchy of complexity.

Automatic generation is although fairly-good process for producing large amount of metadata consuming less-time and less-costly – thus efficient one. But, it is often criticized because of its inconsistent filtering practice, error-prone, unreliable, and fails to produce optimum quality metadata consistently; especially when it is more intellectually demanded. In reverse, automatically extracted metadata; despite its lesser quality than the manually entered one, is a good alternative for dramatically speeding up the metadata creation of extensive digital objects.

Nevertheless, metadata creation primarily became obligatory for the libraries and handled by the professionals (i.e. cataloguers & indexers), initially through manual process and increasingly by automatic means. They have an intellectual ability achieved through training and experiences, and are well-conversant in the use of content value and descriptive standards. Even though professional metadata creators can produce acceptable quality of structured metadata and obviously holds an immense commitment to ensure the efficiency in resource-discovery; but they are limited in their availability, labor-intensive, and too costly.

On the other hand, resource authors have an ability to create optimum quality metadata as they are more intimate with their work and can thus describe their resources appropriately. As such, metadata representations to be created by resource authors are likewise products of their working epistemologies. Though authors have deep concern to their works (bit-streams), thereby highly deserves for creating metadata (refers to a set of descriptors). But authors rarely do that, even when they are provided with appropriate tools. Probably authors find a minimum time for creating metadata elements; rather they prefer to invest more time for their academic works. Authors are therefore not necessarily a good player in creating metadata than professionals.

Typically authors do not have good understanding on metadata creation workflow and metadata harvesting tools. Truly they are lacking professional knowledge and relevant experience on actual use of metadata for intelligent resource discovery. Otherwise they are having limited skill for producing structured metadata when compared to professionals. However they feel comfort to create schema-specific metadata, especially when facilitated by metadata generation tools with appropriate interfaces to understand. Therefore the discussion brings out a few paradoxical views on the process of metadata creation and responsibilities of the concerned players.

### **Conclusion**

Finally it appears that, the responsibility of metadata creation can be equally as fuzzy. Author generated metadata may have increasing support, but typically metadata professionals (include librarians) are primarily engaged with the task of creating metadata for digital repositories. Indeed both the practices have their own merits and arguments. Perhaps the best results could be achieved by integrating automated and manual efforts. Author sincerely believes that human endeavors will continue to play, and no automated system is likely to be able to describe the metadata elements perfectly without frequent assistance of human beings.

Author can plead to beg for an integrated system design that will facilitate automated metadata creation to be performed primarily by resource authors, eventually supervised and endorsed by the metadata professionals through manual efforts. So an optimum quality to be ensured consistently within a structured format, even to produce more intellectually demanded metadata. But truly it is disappointing in real practice.

Obviously, there is a need for efficient tools and techniques with appropriate interfaces to understand that could enhance automatic metadata extraction process further. Besides integrated system design, reengineering the workflow of metadata creation, more generic handler system and development of intelligent metadata harvesting tools would be a great frontier of research in the foreseeable future. Indeed the vocabulary control by means Faceted Application of Standard Terminologies (FAST) could bring a reasonable solution to a greater extent.

If such integrated systems are to succeed, both information generators and organizers or their

successors, could be able to realize a prospective future of digital transformation. Otherwise, such motivations of metadata creation will continue to be an unattainable solution.

## References

1. Crystal A and Land P, *Metadata and Search*, In summary report of the Global Corporate Circle DCMI 2003 Workshop, Seattle, Washington, USA, 28 September 2003. Available at: [www.dublincore.org/groups/corporate/Seattle/](http://www.dublincore.org/groups/corporate/Seattle/) (Accessed on 6 June 2015).
2. Anderson J D and Pe'rez-Carballo J, The nature of indexing: how humans and machines analyze messages and texts for retrieval - Part II: machine indexing, and the allocation of human versus machine effort, *Information Processing & Management*, 37 (2) (2001) 255-277.
3. Adams J, The use of bibliometrics to measure research quality in UK higher education institutions, *Archivum Immunologiae et Therapiae Experimentalis*, 57 (1) (2009) 19-32.
4. Greenberg J, Spurgin K and Crystal A, Functionalities for automatic metadata generation applications: a survey of metadata experts' opinions, *International Journal of Metadata, Semantics and Ontologies*, 1 (1) (2006) 3-20.
5. Liddy E D, Allen E, Harwell S, Corieri S, Yilmazel O, Ozgencil N E, Diekema A R, McCracken N and Silverstein J, *Automatic metadata generation and evaluation*, In Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Tampere, 11-15 August, 2002, p.401-402.
6. Han H, Giles C L, Manavoglu E, Zha H, Zhang Z and Fox E A, *Automatic document metadata extraction using support vector machines*, In Proceedings of the Third ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL-2003, Houston, TX, USA, 27-31 May, 2003, IEEE Computer Society, p.37-48.
7. Takasu A, *Bibliographic attribute extraction from erroneous references based on a statistical model*, In Proceedings of the Third ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL-2003, Houston, TX, USA, 27-31 May, 2003, IEEE Computer Society, p.49-60.
8. Peng F and McCallum A, *Accurate information extraction from research papers using conditional random fields*, In Proceedings of Human Language Technology Conference and North American Chapter of the Association for Computational Linguistics, HLTNAACL, Boston, MA, 2-7 May, 2004, p.329-36.
9. Schwartz C, *Sorting out the Web: Approaches to Subject Access*, (Ablex Pub.; Westport, CT), 2001, 169p.
10. Mayernik M S, *Metadata realities for cyberinfrastructure: Data authors as metadata creators* (2011). Available from Social Science Research Network (SSRN), record no. 2042653, at: <http://dx.doi.org/10.2139/ssrn.2042653> (Accessed on 14 June 2015).
11. Kovacevic A, Ivanovic D, Milosavljevic B, Konjovic Z and Surla D, Automatic extraction of metadata from scientific publications for CRIS systems, *Program*, 45 (4) (2011) 376-396.
12. National Science Foundation, *Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century*, (National Science Board; Washington, D.C.), September 2005, p.25-28. Available at: <http://www.nsf.gov/pubs/2005/nsb0540/> (Accessed on 16 June 2015).
13. Swan A and Brown S, *The skills, role and career structure of data scientists and curators: An assessment of current practice and future needs*. In Report to the JISC, July 2008, p.8. Available at: <http://www.jisc.ac.uk/media/documents/-programmes/digitalrepositories/dataskills-careersfinalreport.pdf> (Accessed on 25 June 2015).
14. Chan L M, Inter-indexer consistency in subject cataloging, *Information Technology and Libraries*, 8 (4) (1989) 349-358.
15. Weinheimer J L, How to keep the practice of librarianship relevant in the age of the Internet, *Vine*, 29 (3) (1999) 14-37.
16. Greenberg J, Pattuelli M C, Parsia B and Robertson W D, *Author-generated Dublin Core metadata for web resources: a baseline study in an organization*. In Proceedings of the International Conference on Dublin Core and Metadata Applications, National Institute of Informetrics [NII], Tokyo, Japan, 24-26 October, 2001, p.38-46. [Also in *Journal of Digital Information*, 2 (2) (2002). Available at: <https://journals.tdl.org/jodi/index.php/jodi/article/view/42>].
17. Thomas C F and Griffin L S, Who will create the Metadata for the Internet? *First Monday*, 3 (12) (1998). Available online at: <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/633> (Accessed on 25 June 2015)
18. Greenberg J, Crystal A, Robertson W D and Leqadem E, *Iterative design of metadata creation tools for resource authors*. In International Conference on Dublin Core and Metadata Applications, North America, September 2003. Available at: <http://dc.papers.dublincore.org/pubs/article-/view/733> (Accessed on 22 June 2015).
19. Greenberg J and Robertson W D, *Semantic Web Construction: An Inquiry of Authors' Views on Collaborative Metadata Generation* (DC-2002: Metadata for e-Communities: Supporting Diversity and Convergence). In Proceedings for the International Conference on Dublin Core and Metadata for e-Communities, Florence, Italy, 13-17 October, 2002, Firenze University Press, p.45-52.
20. Crystal A, *Interface Design for Metadata Creation*. In Extended Abstracts of CHI 2003 (Ft. Lauderdale, 5-10 April, 2003). Available at: [http://ils.unc.edu/~acrystal/chi03\\_crystal.pdf](http://ils.unc.edu/~acrystal/chi03_crystal.pdf) (Accessed on 21 June 2015).
21. Park J R and Tosaka Y, Metadata creation practices in digital repositories and collections: Schemata, selection criteria, and interoperability, *Information Technology and Libraries*, 29 (3) (2010) 104-116.
22. Crystal A and Greenberg J, Usability of a metadata creation application for resource authors, *Library and Information Science Research*, 27 (2) (2005) 177-189.