



## INTERNATIONAL JOURNAL OF APPLIED TECHNOLOGIES IN LIBRARY AND INFORMATION MANAGEMENT

International Journal of Applied Technologies in Library and Information Management 3 (1) 1 - 01 - 09  
ISSN: (online) 2467 - 8120  
© 2017 CREW - Colleagues of Researchers, Educators & Writers  
Manuscript Number: JATLIM - 2017-03.01/01-01

### Cataloguing Digital Materials: Review of Literature and The Nigerian Experience

#### Abstract

*Libraries today have moved from the arena of being mere repositories of books, journals and other print materials to collecting and disseminating online resources. The authors discuss cataloguing digital materials in libraries. Librarians in Nigeria are gradually taking up the challenge of cataloguing online materials. The problem often encountered with the situation is 'how to catalogue these online materials so that they can find a suitable Uniform Resource Locator (URL) for them within the databases?' Another challenge is 'how to render this catalogues searchable to users through the Online Public Access Catalogue (OPAC)?' The authors argued that digitized materials need to and must be catalogued online and discussed some online cataloguing equipments that provide the solution. These equipments include: Dublin Core (DC), MARC 21, Text Encoding Initiative (TEI) and Electronic Archive Description (EAD). The author looked at how Standard Generalized Mark-up Language (SGML) can be used as a language of metadata for Text Encoding Initiative (TEI). The author compared MARC 21 with TEI, pointing out that these are the most suitable equipments that can easily accept metadata for online cataloguing. The authors finally explained how Sears List of Subject Headings (SLSH) and Library of Congress List of Subject Headings (LCLSH) can be used to provide alternative access points for catalogued digital materials.*

**Keywords:** *Cataloguing, Digital Materials, Electronic Archive Description, MARC, OPAC*

**Chorun, Matthew Terfa (CLN)**  
matthewchorun@yahoo.com.  
ICT Division,  
University of Agriculture,  
Library, Makurdi, Nigeria.

**Esther Wurrola Oyeniyi**  
ICT Division,  
University of Agriculture,  
Library, Makurdi, Nigeria.

**Jirbee, Benjamin**  
Technical Service Division,  
University of Agriculture,  
Library, Makurdi, Nigeria

#### 1.1 Introduction

Monographic materials and online books, journals, reference materials, audio visual materials are acquired in thousands by libraries in Nigeria. More often than not online journals from AGORA, OARE, JSTOR, TEEAL, EBSCO-HOST and others are donated free of charge to some agricultural

university libraries and databases are created to host them. In a bid to beat down costs, Nigerian universities acquire online materials by forming consortium as gateway to purchasing them; and virtually all Federal and State Universities are currently members of the consortium in Nigeria. But the question is: how are these digital materials stored in

databases being utilized? Do they ever receive online cataloguing? On the other hand, hard copied of thousands of text books and non-book materials are received in the library to be catalogued, classified and assigned subject-headings.

Besides, materials made available in the library in digital form are dumped in the databases and access to them is not as popular as expected. Manual technical processing and their attendant manipulations remain the order of the day. In Nigeria, a good number of universities have received one database or the other, they have digitized or automated their collections but these innovations have not replaced the traditional card catalogue. Those university libraries that have automated their collections have so far succeeded in making a name rather than a mark on the sand of digital libraries. What then is responsible for these?

Librarians must remember that materials downloaded into the database from the Internet and other digital sources need to be catalogued online. A few titles stored on the laptop and desktop computers are automatically assigned locations either in the “my documents”, desktop, my photos, music, video, etc. But the vast numbers of titles downloaded in our databases have no searchable locations. The desktop or laptop computers have some limitations which must be supplemented by some external equipment to assist them find appropriate locations known as Uniform Resource Locators (URL). Cataloguing library materials into digital form before putting them online is therefore of uttermost importance.

## **2.1 Literature Review**

### **2.2 The Need to Catalogue Electronic Resources**

It has been observed that most Nigerian libraries are contempt with downloading electronic journals and other online materials into databases. Some libraries have even automated their

collections with great enthusiasm but then stopped at the point of data capturing often propagated and taught by information and communication technology (ICT) experts.

Another reason is that in a library situation, cataloguing equipments for digital purposes must be procured and put in place. These equipments standardize online cataloguing in addition to sending signals to the Online Public Access Catalogue (OPAC). These equipments include Dublin Core (DC), MARC 21, Text Encoding Initiative (TEI) and Electronic Archive Description (EAD). These equipments have bibliographic elements which have been re-examined, selected and applied to suit the purpose of digitization. Each and every document made available for the purpose of library collection, in electronic form or otherwise, must be in line with the requirements of stability, source reliability, and bibliographic information in order to be useful to all needful users, not just to the author of the document. Librarians have endeavoured to organize digital documents and scholars in the humanities have long been interested in the possibilities of storing and retrieving textual materials in electronic format but many of the electronic documents originally encoded did not satisfy digital requirements for easy retrieval.

Of prime importance is the expectation that the electronic materials be accurate in the transcriptions, provide retrieval guides and contain detailed and accurate bibliographic information. Wayne, Ahronhiem and Crawford (2002) note that the difficulties encountered were that the bibliographic information required by scholars of ancient texts also needs to be more descriptive than a bibliographic citation for library purposes. In addition to traditional bibliographic information such as that provided by proper cataloging of an electronic text, humanities scholars need more detailed description of the physical book or manuscript they study.

For example, According to Wayne, Ahronhiem and Crawford (2002) “information such as where a page break occurs in the original texts and manuscripts of the same work, the appearance of the title page, and how the lines of verse or prose are arranged on the page in a play or poem, are only a few items indispensable to a scholar who tries to compile a new edition of an ancient work. While this type of meta-information is not in the scope of cataloging per se, it is required, for instance, for an electronic version of a first quarto of a Shakespearean play to be useful. Electronic texts for the humanities thus create problems as to how much and what type of meta-information they must provide.”

Because of this specific need for accurate and detailed meta-information to accompany electronic texts of this sort, the Text Encoding Initiative (TEI) was established in 1987 as a combined volunteer effort in humanities computing, an effort driven by a joint-project of the Association for Computers and the Humanities, the Association for Computational Linguistic, and the Association for Literary and Linguistic Computing. The principal objective of the TEI project is to define a set of generic guidelines for the representation of textual materials in electronic form.

## **2.2 Standard Generalized Mark-up Language (SGML) and Text Encoding Initiative (TEI)**

A major breakthrough in what type or how much of meta-information was required for cataloguing electronic textual materials was the establishment of TEI guidelines. The guidelines specified a number of required bibliographic elements for each text, in particular a header made of four distinctly outlined elements which contain the bibliographic information and a body which can contain text, images, and other objects. In

order to provide rules for the encoding and interchange of electronic texts, the TEI scheme relies upon the use of Standard Generalized Mark-up Language (SGML) and its sets of mark-up tags for the encoding of textual materials. SGML is an international standard language for encoding electronic information which defines device- and system-independent methods for representing text and other objects and is concerned with content and arrangement rather than format or appearance. SGML is made of various sub-sets of tags, called Document-Type Definitions (DTD) which specifies content requirements and sets of tags for each type of document. Hypertext Mark-up Language (HTML), which is one type of a SGML. DTD is the best-known sub-set of SGML tags. The TEI guidelines are another.

Like other SGML applications, TEI is independent of platforms, systems application and devices and one advantage of TEI is that it conforms to network protocols for the exchange of information. Any electronic document encoded in SGML includes three parts:

1. the SGML declaration (i.e. a statement declaring that this is an SGML document),
2. the DTD (i.e., the sets of SGML tags used for this particular document), and
3. the document instance (i.e., the encoded document according to the rules declared by the DTD).

With SGML, and therefore with TEI, each electronic document carries along its own meta-information or metadata. The TEI scheme defines how to write a specific class of SGML document-type-definitions that specify how each electronic document is encoded in SGML. According to the TEI guidelines, the content requirements specify minimum mark-up requirements for low level of encoding, but may also provide for very

complex encoding appropriate to the detailed marking necessary to humanities research.

Each electronic document encoded according to TEI also contains a TEI header in addition to the body of the document. TEI headers carry meta-information about the electronic document; besides bibliographic information that is of direct use to libraries, because it has been designed in consultation with librarians. Although the TEI guidelines were originally concerned with printed texts such as those studied in literature, linguistics, and history, and designed as “a common encoding scheme for complex textual structure”, electronic texts must be understood in a broad way: TEI is interested in textual and non-textual resources such as those contained in a research database or components of non-proper publications.

Perhaps in a less obvious manner, catalogers may also apply experience gained from their exposure to the demands of electronic texts in the humanities: the efforts of cataloging and encoding meta-information in electronic texts for the humanities provide direction for cataloging all electronic materials, particularly resources on the internet. According to Marco (2011), head of the monograph cataloging division at the university of Michigan library, catalogers may face a situation similar to that of the ice industry at the advent of refrigeration unless cataloging practice adapt to the changing environment brought about by electronics formats.

The library of congress recommends that catalogers prepare for the future of organizing for access to digital libraries by finding ways to expend the use of metadata that forms part of the digital project..., include it on digital resources and develop mechanisms for integrating different forms of metadata. Although metadata efforts are more advanced for digital text materials, other digitized resources could also benefit from metadata schemes.

### **3.1 The Text Encoding and Initiative (TEI) Header**

According to Wayne, Ahronhiem, and Crawford (2002) the TEI header as attached to electronic documents, is a label containing directions about the document's logical structure. It may also function independently from the electronic text it pertains to as a vehicle for meta-information about the electronics materials. It is this capacity of producing independent records of meta-information following prescribed rules that make the TEI header of interest to catalogers: in effect, the rules designed by the TEI serves as guidelines for using SGML for purpose of describing electronic documents in a manner which is pertinent both to the requirement of electronic forms and to those of more tradition cataloging.

The TEI header includes four parts. All the four parts pertain to issue specific to cataloging electronic forms but only one is mandatory.

- file description and source from which the text was derived (mandatory) print, electronic.
- Encoding system which has been applied (optional)
- Profile description (optional) class, keywords, subjects heading.
- Revision history (optimal)

Although only the file description is mandatory, the other parts are highly recommended, especially in the case of independent TEI headers because they contain the possibility of including information that is difficult to describe using AACR2 rules for computer files.

### **3.2 Analysis of a Header**

- (a) the files description (>file des<),
- (b) the encoding description (<encoding des>) show as non-application
- (c) the profile description (<profile Desc>),
- (d) the revision history (<revision Desc>)

also known as non-applicable.

The file description element and the profile-description element are the most familiar to a traditional cataloger since they contain the metadata that form the basis of a catalog record. The file-description contains descriptive information, and other information. It also contains the source-description element which allows encoding information about the physical text or source from which the electronic text has been derived. The profile description also contains some assigned subject headings and classification numbers and accordingly, the electronic text may be assigned a call number or accession number. The profile-description element may also contain information necessary to humanistic studies ranging from the human language in which the text has been written to information about the social context.

### 3.3 Subject Headings and Classification Numbers

The TEI header used for encoding online documents contains Library of Congress Subject Headings (LCSH) and the library of congress classification numbers. Not all TEI header encoded documents do have these because the profile description element is not mandatory, but only highly recommended where applicable, as prescribed by the TEI scheme. The profile Desc tag uses the following LCSH keywords: internet computer network keywords, keywords used for cataloging of computer files, those of information networks, library of congress subject-heading communication systems, information storage and retrieval system, and library information networks.

The inclusion or exclusion of Library of Congress Subject Headings (LCSH) within a TEI-encoded document should not, however, constitute a criterion for assessing the appropriateness of the TEI scheme. The

profile description makes room for including subject headings, and the inclusion depends on the purposes for which the electronic document was created. Although the use of LCSH is recommended in certain instances, it may not always be appropriate. A TEI header allows for the possibility of declaring any scheme of subject headings or using standardized schemes such as the Sears List of Subject Heading and any other developed Subject Heading. The following classification numbers are typical of the numbers found in the profile Desc element: Dewey Decimal Classification 025.28 for Educational Psychology and library of congress classification QD 251.2 for Organic Chemistry. However, examples from CETH and ETC provide no space for such a classification.

### 3.4 Practical description of TEI header

The profile description element of the TEI-header includes the following fields:

- Title statement
- Edition statement
- Extent
- Publication statement
- Series statement
- Notes statement
- Source description

It should be noted that each of these fields may be translated to the fields of a USMARC records as well as MARC 21. The Online Computer Library Center (OCLC) has announced a prototype program (Spectrum) which is capable of allowing automatic translation. The fields "extent" and "series statement" are sometimes empty because they are unknown or not applicable to some particular electronic text to be cataloged.

Gaynor's (1994) example of a TEI-header for Electronic Text Center (ETC) includes the size of the Electronic File in Kilobytes in extent, the size of the Electronic

File in Note Statement. The source description describes the physical item various sources from which the computer file is derived. It is often the case that electronic texts such as those available from CETH and ETC are digitized from one or several printed books or manuscripts, and it is very important for the research use of these texts that the user knows exactly which edition he/she is studying.

In addition, some electronic archives and repositories of ancient texts share their records: the source description of some texts from Electronic Text Center (ETC) indicates that the source is another computer file from the electronic repository at the Oxford Text Archive. The source description often gives the full bibliographic record of the printed sources, and duplicates the format of the first six fields of the profile description. Regardless of how much the first six fields of the profile description resemble the source description fields in content, the user of an electronic text must not forget that the source description contains meta-information about the physical sources, whereas the profile description contains meta-information for the electronic record itself.

In the case of electronic document that may never have appeared in print, such as world wide web pages, the source description field need only contain the indication "original" and nothing else. The source description field may concern with requirements of intellectual property for digitization of print sources. It should also be noted that the information that may be translated into fields 245 (Title Statement), 260 (Publication, Distribution, etc.) and 650 (subject-added entry) of a MARC record for an electronic document must be taken from the first six fields of the file description, as well as from the profile description, but not from the source description, as this would duplicate the record for the printed text and not generate one for the electronic text.

### **3.5 Authority Control (Main Entry)**

It is noteworthy that many catalogers have observed that early electronic texts and those created without regard for cataloging rules do not fall within the descriptive field of authority control. The rules for encoding Author/Title information into a TEI header do not prescribe the use of AACR2 for writing information into the fields. Therefore the title and author fields in the title statement element of profile description may be spelled and capitalized according to the whims and caprices of the encoder or to the policies in effect in his or her organization. This may render access to information from these fields difficult (Anglo-American Cataloguing Rule 2 (1988)).

Fortunately, efforts to create texts on a large scale, such as those at Center for Electronic Text in the Humanities (CETH) and Electronic Text Centre (ETC), have been working in partnership with catalogers. Gaynor (1994) described how the ETC developed asset of local guideline for entering information concerning the author, publication and edition statement in the TEI headers used at ETC, with the purpose of making the TEI header as congruent to a MARC record as possible. A particular effort was made for the completeness and accuracy of publication information, often a volatile area for electronic information.

According to Caplan, (1993) the indexing and retrieval software provided flexibility for the author/title related fields, with the result that TEI headers at ETC do not have to conform to AACR2. Thomas Jefferson as an author's name may be entered in the TEI header and searched as Jefferson; Thomas; Jefferson, Thomas; Thomas Jefferson and T. Jefferson. Similarly, Chorun, Matthew as an author's name may be entered in the TEI header and searched as: Chorun, Matthew; Matthew, Chorun; and M. Chorun; according to how it is spelled in the print source. Even a TEI-header created by OCLC,

does not carry the same format for the author in the title statement as that of the AACR2. The author's name are listed as "Martin Dillon, Erik Jul" and not "Dillon, Martin; Jul, Erik".

#### **4.1 Similarity between TEI header and MARC guidelines**

Two attempts were made: one by OCLC and another at ETC for translating a TEI header into a MARC record but these attempts revealed that a number of issues needed further discussion and elaboration before satisfactory answers may be found. One is that the granularity of a TEI header must be refined in order to allow effective flow of data to a MARC record. Most TEI headers only specify a general author field, which may contain several authors, and distinguish between author as person, corporate author, and conference proceeding only occasionally.

This renders the author assignment of MARC tags difficult in the case of a corporate body as main entry for texts encoded with TEI. This lack of granularity does not render TEI headings invalid, because the TEI guidelines are flexible enough to accommodate such refinements. But this will depend on each individual creator and distributor of electronic texts; whether by a corporate body, a library, an educational institution or organization, to ensure that the necessary granularity is achieved.

Besides, a part of the TEI header has space in the current design fields and subfields for the revision history of an electronic file. MARC field 856 allows for a wide range of information regarding access to an electronic file, but does not appear to include an indicator for revision history. More investigation is necessary to ascertain if the header field 005 conveys similar kin information as the revision history field in a TEI header.

Another issue that needs clarification is how to present the original version or what

constitutes an original version of an edition statement, and a publication status for an electric document. To what extent does the digitization process for the creation of an electronic document constitutes a new intelligent work, a new edition of an existing work, or simply a new format of the same work? Questions which arise with the cataloging of electronic document call for an agreement between MARC catalogers and TEI encoders, because they affect how information is represented in the publication statement, series statement, and notes statement of the TEI header and of a MARC record. The variations between them have to be harmonized.

This is not a problem of harmonizing one scheme in favor of the other MARC and TEI, but a common one that appears in both schemes and therefore must be resolved in accord. Project now underway at the ETC and CETH have showed that the use of a TEI header for the cataloging of electronic documents in general and electronic texts in particular is a fruitful endeavor. TEI has introduced minimum standards which may be refined and adapted to suit the specific requirements of electronic-texts creation and distribution. Both partners in the design of TEI headers have found that a TEI header may accommodate descriptive information which is difficult to encode into a MARC record, such as the source description and the revision history of an electronic document.

It has also been discovered that a standard measure of authority control is necessary for the profile description field of the TEI header although TEI guidelines do not prescribe it. The TEI header and a MARC record differ significantly because they were created for different purposes: TEI was for encoding online materials while MARC was for digitizing hard copies of monographic materials, but when it comes to actual usage, the convergence that exist between the two may be exploited with numerous advantages.

When adapted to allow better congruence, TEI headers and the MARC structure offer possibilities for cataloguing all sorts of electronic documents and electronic textual materials.

#### **4.2 Subject Heading and Classification**

Subject headings are words, group of words, phrases or acronyms that are used to describe the subjects of a work. They provide alternative access points to a work with several possible subjects. As access points in the catalogue, these headings are extremely important in communicating the holding of a library. This explains why a single article or book is easily retrieved online through several access points.

To add subject heading to the T-slip to provide access to the intellectual content of the material means the cataloguer for the book must select the most appropriate subject headings. The chosen heading are numbered with Arabic numerals and are recorded. It is important to use consistently the same terminology to designate a particular subject. For this reason some libraries have developed standard lists in Canadian school libraries are the sears list of subject headings and the Canadian companion to the sears list. Most Nigerian university libraries use Library of Congress List of Subject Headings.

A supplemental choice would be the most recent edition of the library of congress subject headings and the national library of Canada's Canadian subject headings. The latter two titles supply an expanded number of subject possibilities. Before assigning subject heading, the main subject or subjects of the book must be determined. This is done by examining the title, table of contents, description on the dust jacket, preface, text, and illustrations.

The next step is to write down the subject and check to see if that choice is permitted in the standardize list. If the subject selected is not in the list, the related subjects

must be consulted. It is possible, although not advisable, to add a local subject headings. This should be done only when all other avenues have been exhausted. The standard subject list represents a controlled vocabulary that has been carefully constructed over a number of years. Local subject creativity if taken to excess results in a garbled catalogue. If local heading are added to the catalogue, write them in your copy of sears List of subject headings or LCSH for future reference and to expand the scope of subject heading.

Typically, two subject headings per item is sufficient for most school libraries, although there are occasions when three or more may be required to cover the scope of the item especially the case with academic libraries. There is also the need to select as specific a heading as possible. A book about bears should be given the subject heading BEARS rather than general heading ANIMALS. For a book about several different animals, the more inclusive heading ANIMALS should be used rather than separate headings for each animal mentioned in the text. Determining whether to use a specific or a general subject entry can pose difficulties. The introduction to sears describes how to designate correctly.

The subject heading is listed at the bottom of T-slip preceded by an Arabic numeral. Such a notion is called a tracing and tells the cataloguer what other cards exist. School and academic libraries often find it valuable to included subject or theme. In general, add the word to indicate that the item is a work of fiction.

#### **Conclusion and Recommendation**

Cataloguing online resources in the digital age play significant role in shaping library services. The basic terms and intricacies associated with cataloguing digital materials and the currently accepted practice for cataloguing electronic-based library resources and the processes for standards-



based data entry have been discussed. Any library collection catalogued in accordance with international standard rules like AACR2, MARC 21, Dublin Core, etc, and practices allow the library users to retrieve information effectively, thereby maximizing the use of the library's collection.

Skills for cataloguing library and multimedia resources are a necessary requirement in the modern electronic library environment. It is therefore recommended that all library professionals (especially those who work in the technical services sections of the library), should acquire these skills so as to be able to manage the electronic resources effectively. By so doing library users will have unlimited access to the multimedia resources available in the libraries.

## References

Anglo-American Cataloguing Rule, 2<sup>nd</sup> ed. (1988). *Cataloguing Electronic Resources*. Chicago: American Library Association.

Caplan, P. (1993). Cataloging Internet Resources. *The Public-Access Computer Systems (PACS) Review*, 4 (2), 61-102

Gaynor, E. (1994). Cataloging electronic texts: The University of Virginia Library experience. *Library Resources and Technical Services*, 38(4), 403-412.

Marco, G. A. (2011). *The American Public Library Handbook*. Michigan: Michigan Library Association

Online Computer Library Centre (2009). WorldCat Statistics. Retrieved on 8th August, 2016. from <http://www.oclc.org/ca/en/worldcat/statistics/default.htm>.

Wayne, J., Ahronhiem, J. R. & Crawford, J. (2002). *Cataloguing the Web: Metadata, AACR, and MARC 21*. Lanham, Md: Scarecrow Press; (Published in corporation with the Association for Library Collection & Technical Services).