Indexing and Cataloging in Digital Environment

Wan Elija Usna Mohd Radzuan, Saiful Farik Mat Yatin, Nurul Aina Sakinah R. Junaide, Nurhidayah Mazlan

Indexing and Cataloging in Digital Environment

Wan Elija Usna Mohd Radzuan, Saiful Farik Mat Yatin, Nurul Aina Sakinah R. Junaide, Nurhidayah Mazlan

Faculty of Information Management, Puncak Perdana Campus,
Universiti Teknologi MARA (UiTM), UiTM Selangor, Malaysia

Abstract
This paper explain what is cataloging and indexing. It also touch about digitization in cataloging and indexing as digitally. In cataloging and indexing there is many process pertaining to it as cataloging and indexing is an important part in libraries and archival institution. This paper also explain digitization in cataloging and the different between indexing and cataloging. This paper explain about metadata as metadata plays important components in digitizing the indexing and cataloging. Metadata help organize electronic records, facilitate interoperability across systems, provide digital identification, and support both archiving and preservation. This paper also mention the challenges faced in cataloging into digitization and the current issues happened in indexing and cataloging.
Keywords: Cataloging, Indexing, Metadata, Digital information

Definition of Indexing and Cataloging

Indexing is one of the most important process in the Retrieval Maklumat (IR) system. It forms the main function in the Information Retrieval process as it is the first step in helping to retrieve the information accurately and efficiently. The indexing can reduce the documentation activity to the mandatory term contained therein. Mapping from information to the respective documents contained therein. Among the effective indexes created from document collection, the material recovery process is simplified. Indexing continues to follow four levels of process namely content specification, tokenization of documents, processing of document terms, and index building. Indexes are stored in the form of different data structures such as direct indexes, index documents, lexicon and reverse indexes. Indexes can be built using different algorithms or schemes, such as indexing such as single-pass in-memory indexing, blocked-indexing, etc. This indexing process helps the public in finding things faster in just where they are looking for terms, so users do not have to hunt for information by flipping through the page sheets to find out the terms or to find the information needed in information retrieval.

Meanwhile cataloging is a process that creates metadata and represents information sources such as books, sound recordings, moving images and so forth. The catalog provides information such as the name of the creator, the title and the subject term that describes the source, usually through the creation of a record bibliography. The code serves as a substitute for the information source

stored. Cataloging allows users to search for items in the use of manual or electronic methods. The catalog users can search for items easily. According to Gorman (1998), technical services are the tasks carried on in a library that are concerned with the processing of library materials in order to make them accessible to the users of the library. In general, it has two main administrative divisions, namely the acquisition and cataloging. The work of redemption work includes selecting and cataloging. The use of the cataloging system for all materials is important to make it easy for users to get information or get materials and to write reference materials. Cataloging towards electronic materials is now rapidly as many materials are produced through electronic materials. So it will make it easier for users to get material and important information about the material.

Cataloging Planning
The cataloging department is given the responsibility of selecting the most suitable cataloging catalog. It is a guideline to catalog this source and create a record. Based on Lorraine M. Nero article "Cataloging digital resources" as the experience of the University of the West India, St Augustine Campus. Cataloging test is being conducted to test several strategies to decrease the possibilities of errors.

Cross Check
Checking cataloging is requested to cataloguer edit and revise the catalog record made by cataloguer the sender given to the project.

Authority Control
The index will be used as a control tool for the authorities, to check and verify the title. When setting up corporate and personal headlines, the Index will also be used as a guide.

Staffing
The cataloging team lobbied to hire the publisher for the document. Student, who conducts research on Trinidad and Tobago's economic history, is used to transcribe. Student is proficient to read the document as it is used as a major source of research. Researchers also suggested that transcription should be done first, as this text will help in understanding the subject of the document. The team welcomes this solution, as transcriptions are also held on the site along with the image. Therefore, add value to the collection. Another source, a historian, can also find further consultation on the documents if the researchers consider it necessary.

Type of Indexing
Full text
There are several types of indexing. One of the type in indexing is full text that needed a unique words used for identified the similar documents. The full text indexing is an automatically created in a computer software that it will show all the list and location of documents that have been save in the database. Users can find the documents using any keywords that related to the materials that they want and the computer will then find any related match terms and text regarding the documents. Not only that, full text index is an easy searching because it will help users finding the materials that they want even if they are not sure of the details information.

General Indexing

This type of indexing is needed a basic information or a popular and current words to find the materials. With general indexing, we will find less reliable information of articles because it is a broad searching that include all article and materials related to the searching words. For example when the users want to find article about digitization, the results will show all articles that contain word of "digitization" even though it is not reliable with the information that the users want. The keywords also in general that contain the most words have been used in materials.

Specialized Index

The specialized index is better than general index because it is more specific to find the materials using any terms that related to the articles. This type of index will show the searching results that more reliable than general index because it include scholarly journals or articles. The specialized will only have the keywords that specific for the title and contain in materials.

Electronic Index

With electronic index, users can find the searching using any keyword, title of the materials and author at one time. This type of index give extra searching because we can locate the materials with any word in the article especially the year of article publication, the author and title or page that will give small results of searching. This is very easy to search because it only shows the materials that have the same title, year or author at the same time rather than search one by one and the results found the specific as the users want. If the result is error or not found what the users want, it means that there are no materials as the searching keywords. Sometimes the system cannot found the materials because the keyword is not correct and need to use other words or title.

Types of Cataloging
Machine-Readable Catalog (MARC)
        Based on articles "Cataloguing Digital Resources (2005) and Metadata and Cataloging Practices", Machine –Readable Catalog (MARC) is a computer that can read and interpret the data in cataloging records. It is a standard format for cataloging and any related information in machine-readable form. MACR is using a data format for cataloging in digitization that use a system of numbers, letters and symbols that stored exchange bibliographic information. MARC is a standardization of cataloging because it prevent the duplication of catalog work. Moreover, it also more better for sharable bibliographic resources and enables to acquire reliable data. Not only that, it enables the library to available the automation system in managing library operations that follows advance in technology. This system also enables cataloger to replace with other system and ensures that their available data will still compatible.
        The bibliographic records or the information should include the description of the item that have to catalog. The description includes the title of materials, statement of responsibility, the edition of the materials if there are the second and several edition. It also includes the specific details such as the summary of the materials, and what is the content of the materials to make readers

understand the whole records only by reading the description. There are also main entry needed and added entries such access points where users should be able to find by main subject of the item. The cataloger also has to show the subject headings as a proved list that are important to enable the consistent and ensure that the items in a specific subject can be found in the same place and heading in the system.

Function of Indexing

Indexing function for the users to directly go to the keyword that have listed by the person that made the records. It is important because it help to access the information and reduce time in retrieve the records. When the users want to find only the important information directly and not waste time by read thoroughly all the information, they can search using the specific words and find the pages that contain the words. Sometimes users does not sure what they exactly have to search in finding the materials that they needed and indexing help them in searching the information. Not only have that, indexing been used in printed version and now with digital format, the retrieval tools help the users more easy as it increase speed searching. In printed format, the indexing usually at the back of pages for the users to refer. It contain words that important in the materials and arranged alphabetically together with the pages which the words located. In the digital format, the index have been prepared as one of the searching types and the system will automatically find the keyword that the users insert in the searching.

Function of Digital Cataloging

Cataloging digital materials is to decide and determine access point of the main entry and added entry based on the tools that used to catalog. The cataloging helps to determine proper heading for keyword search in order to enable the users search the materials efficient effectively. The details of materials will be type in system that include the title, file number, date, keyword for easy searching and also description about the materials because it helps the users the easy access the documents or materials efficiently. All of the details for cataloging are made in digital format that researchers and also the staff of archives does not have to do manually by writing the catalogue information.

Moreover, cataloguing in digitization are combining both collections of printed and electronic materials that the cataloger will not have to do separate catalog. The facility for cataloging will increase the speed of processing cataloging and give advantages to improve the cataloger works. For example, if using the manual cataloging, the cataloger can finish their cataloging work about 100 files in a day and then have to check several times and they will have to check the spelling correctly. If they are using a system cataloger, they can do about 300 files in a day because it is automatically detect spelling error while insert the details and delete the mistake before the data is saving in the system.

Not only that, the system is available to link and to make users or researchers aware of the collection and materials especially the new collection and this technology is improving all the time. So that, the system will always have improvement to help in serving the users in searching materials that they need in a short time. Since we are in technology era, the economics are also forcing the cataloger to do the cataloging in digital format.

Digital document is a set of files that labelled by unique characters. The documents must be explained so that it is identifiable by majority of individuals. The description includes a set of metadata. It is because the digital document does not exist as it remains inaccessible without

metadata. The function of descriptive metadata is to give details of bibliographic description in standardized form which allows the access to document and the exchange of other institutions.

Different Between Index and Cataloging
Index and cataloging can be confusing because these two are similar function but there are different that we need to know. The index is a unique keyword that the users needed to find and locate the specific terms of words that they need to know. For example, in the articles, there are terms that hard to understand but indexer have given the meaning of the terms at the last page with alphabetical arrangement. It easy for the users to find the word with the front letter.

For cataloging, it is to locate the materials with any subject, title, and author or any details that regarding the materials. Cataloging is a physical description of the materials to give understanding for the users that will help to understand about the materials especially for the content of materials. The differences between index and cataloging is the function that we need to know because it is not the same. The index is to find materials and information using the keywords that easy to retrieve while the cataloging is to describe the materials.

What is Metadata
Metadata a data (information) that provides information about the data, metadata allows user to easily locate and evaluate data. For example, author, date created and date modified and file size are examples of very basic document metadata.   It can help organize electronic records, facilitate interoperability across systems, provide digital identification, and support both archiving and preservation. Based on ISO15489-1 Information and Documentation (Records Management) the first international records management standard, defines metadata as "data describing context, content, and structure of records and their management through time." The definition of metadata provided by the National Information Standards Organization states that "metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource".

There is three (3) types of metadata it is descriptive metadata, structural metadata and administrative metadata describes a resource for purposes such as discovery and identification. Based on article "Why data matters" in The Information Management Journal, September/October issues types of metadata is described as descriptive include such elements as author, title, and abstract. Elements that one would find in a typical bibliographic record are examples of metadata. For structural metadata, it indicates how compound objects are put together. They identify data format, media format or the type of data representation and file types, hardware and software needed to render the data, and the compression method and encryption algorithms used. Meanwhile, administrative data provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it. Very often, there are subsets in this category. The two most typically included are rights management and preservation metadata.

A digital project may include many different components and activities such as material selection, copyright issues, digitization, cataloging and metadata, access and delivery, funding, publicity, and assessment. Among those, metadata is a critical component. Metadata is fundamental to effective retrieval and use of digital materials, as well as the organization, preservation, and management of

those materials in the digital environment. Different types of metadata can be used to fulfill the above purposes. Descriptive metadata is for describing or identifying information resources. Structural metadata documents relationships among objects. Administrative metadata provides information to help manage a resource.

Unstructured data, which are automatically extracted from resources and indexed for use by robot-based Web services such as AltaVista. Structured formats, is simple that it can be created by non-specialist users. Usually manually created, but some data may be extracted automatically. Examples include ROADS templates and DC. Structured formats, which are rich and complex to organize complex relations between objects or collections of objects and are often based on implementations of SGML. Examples include MARC21, TEI headers and CIMI formats as studied by Magda El – Sherbini and George Klim in their articles "Metadata and cataloging practices" (2004).

Process of Cataloging
The seven processes are:
1) ***Examining the issue(s) of the serial***
    - Cataloging determines that the source is an important serial as it helps to obtain information about the information,
    - Among the examples is knowing the title, what is the main sources of the series, the system used by and how does the tittle change by number of issues.
2) ***Searching***
    - Searching is an important process for obtaining information, users need to find information using the system provided based on established institutions. Appropriate information stipulation should be made to facilitate the search process. Determining where the source is, whether the keyword for each source, any system designated is to facilitate and speed up the search time of the information. The source setting is according to the chosen keyword either by institution name or coding**.**
3) ***Describing the serial***
    - After completing the search process then need to refer to the rules section according to the predetermined serial. See where the appropriate location is to keep the records in the system but must comply with the rules according to the predetermined serial.

4) ***Deciding what to do with what is found on the issue***
    - After acquiring the source and information, the cataloger needs to know what to do and the next step in managing the information. Examine whether there is a change in the occurrence of increasing issues in the source. This is a difficult process because it is important to understand what the publisher or institution has to say to facilitate the resource being identified. Cataloging electronically is the main step involved as it is the main medium for users to find information. Therefore, the cataloger needs to know what is the appropriate measure for each resource to be classified according to the rule.

5) Determining name and title access points

- "Access points" are the entries under which the serial can be searched. The first decision is usually what will be the "main entry". For serials, this is generally a question of corporate body/conference heading vs. title. Added entries may be made for corporate bodies that have not been given as the main entry, persons responsible for the serial (in some cases), or conference names. Added entries are also given for the title when it is not the main entry and, in most cases, for other titles or variations of the title. If the serial is issued within a series, the series may also be given as an access point. Since access points are critical to the retrieval of records, many of them, such as series titles, series like phrases, and corporate names are established in authority files.

6) Assigning subject headings and a call number

- Determining what the serial is about setting the place for the information, This is the process of the name or keyword setting for each record. So the user will know where to get the information.

7) Inputting the record

- Before inputting the new record, it is advisable to search the title again if much time has elapsed since it was first searched. When there is a new record , it is necessary to find the appropriate keyword and if it is the connection from the first source it should be the keyword related to the first sum and the existing issues need to be completed more efficiently.

8) Updating related records

- When new information is updated, the cataloger is responsible for entering the data and also need to update it online. Information placed electronically should be updated more frequently to minimize the risk of missing out on setting new information. catalogers need to ensure that all information is protected from any suspicious sources.

Indexing in Cataloging

With recent technology in advance, compressing and indexing information have yielded a qualitative change in the feasibility of scale large full text retrieval, it is not only textual document but also of document containing black-white, greyscale and color image, video clips and sounds bite. To permit retrieval of textual and non-textual documents alike we assume that every document is describe by a set of representative term or simply term which use to index it. The index is capable of identifying all documents that contains combination of specified term are or in some other ways. In term of textual documents the entire text is generally regarded as its own set of representative terms thus any words contained in the message can be used as query term. Modern compression technique make it quite to index every single words in the text, so no consideration need be given to a selecting index terms when documents are filed.

This kind of indexing provides full-text retrieval (FTR) which makes information immediately accessible to the end user without requiring that any attention be paid to organize or labelling it when it is filed. No manual indexing or keyword selection is necessary because the index terms are drawn

from the text itself. Nothing is missed because every single words is being indexed. Information can be retrieved either by specifying a Boolean combination of words and seeking all documents that include that combination or alternatively by specifying a list of words and applying a heuristic to gauge the relevance of each documents to the list. No special database, information handling or library skill are required. This is end-user technology and user can learn quickly through this responsive full-retrieval system.

The Search index is generated by retrieving information from each of the applications based on a schedule defined by the administrator.
- ***Text acquisition** – **identifies*** and stores documents for indexing
- ***Text transformation*** – Trans document's into index term or features
- ***Index creation*** – Takes index terms and creates data structures (indexes)to support fast searching

Challenges Cataloging into Digitization
Resource Description and Access (RDA) have been introduced as an alternative to past cataloguing practices for the electronic environment. This new code for identifying resources has emerged from years of international collaborations, and it produces well formed, interconnected metadata for the digital environment, offering a way to keep libraries relevant on the Web.

Lack of adequate physical information
The majority of respondents described this as a barrier to cataloging electronic sources. Electronic sources often do not provide standard bibliographic information that requires AACR2. Terms like author information, title, and publication information are most often not available for cataloging to work together. Because of this issue, more detailed information about the author, and more detailed information about the publication will be difficult to find. However, electronic resources can help in saving safer information.

Copyright Challenge
This is another challenge highlighted by the respondents. Electronic resources, especially those found on the web, need to be handled with great sensitivity. Cataloging will violate copyright laws by cataloging the source and belonging to their university databases without authorization or publisher permission. Legislation in the exploitation of copyright resources must also place the security guard necessary to ensure nothing is done to violate the terms of the license agreement as it relates to licensed electronic sources (Abioye, 2010).

Insufficient Workflow
Respondents claim that inadequate workflow has influenced the cataloging of electronic sources in their libraries. This is confirmed by what Oketunji and Iyoro (2009) note that in some libraries, publishers are involved in administering administrative tasks, operations, and other services while in some other libraries there is an integrated workflow across what is traditionally called technical services, where acquisition and cataloging activities are often performed by the same staff. Cataloging electronic resources requires a large degree of mental alertness in the catalog section. Therefore, multi-tasking will not match the level of concentration required in the cataloging process.

Transferring legacy Collections of Indexed Materials into the Tools Available Comfortably.
Many experiment has been conduct to provide several level of access via the OPAC, HTML, WebPages and digitization the resources that they get. In library where MARC has been the standard for bibliographic description to however the format or the collection using MARC the information rely is more to summary than the content should have for records based content. For example, UWI St. Augustine campus library in Trinidad and Tobago all the special collection has been catalogue at the collection level and made accessible via the integrated online search system. However, several of the prized collection including the papers of the first Prime Minister have supplementary funding aids in separate database. The idea in this situation is to integrate both set of data thus it can be assist in user access while ensuring that the context of the archives is not lost to whatever data is being presented.

Current Issues
Digital Technology
Digital technology has revolutionized the information production, acquisition, storage and retrieval processes. Digital computing, application to emailing and text, video, and audio have created a new world and generation of information sources, resources and systems that operate in the electronic environment. Using digital technology save times and energy. Lighten the workload and other. How every each technology have their own way to use thus the manual is provided as a guide. Meaning user must have knowledge and skill in using the technology. The technology changes every time due to the vast and advance of technology. So do the technology related to cataloging and indexing in the libraries or archival institution. The institution have to find which technology can serve the institution in helping making great contribution to the users in preserving the information and records.

Automation
This is perhaps the first and most critical phase in the process of creating an electronic information environment. For the cataloguer, this may entail turning the conventional card catalogue into an Online Public Access Catalogue (OPAC). Creating the interface tools or templates and managing the transition process are critical challenges. Automation refers to the use of the computers to automate the typical procedure of libraries, archival, records center and so on for cataloging, indexing and circulating. It is a process using the machine for easily working and saving the human power and time. In the old days, we would use card cataloger to do cataloging and searching for the records, but now we use RFID barcode to scan the barcode attached at the back of the books or records and the in link to the system in computer.

Standardization
For indexing and cataloging for records or archival institution the standard is not standardize like the libraries. The libraries is using AACR as the standard in determine how the description or details of information should be write. ACCR is refer as "libraries bible" for the whole world when it comes to it, however for records or archival institution it is opposite. This is because there is no standardization, each institution is making their own rules according to the country rules standard. That can be said because there is to many department or subject to be categorize in each subject. So when it is being indexing and cataloging it also very hard. Doing standardization may come true but until there is the followed and formal standardization such as AACR for libraries. Perhaps there is might exist for records or archival institutions.

Conclusion

In conclusion digitizing have been a great helps to libraries and archival institution on helping in doing cataloging and indexing job becoming easier. With the advance of the technology there is many ways in reducing the burden and workload from the librarians and archivist that they faced in doing their work, however there is still challenges about using digitization on their field that still need to discover more on how digitization can contribute more in helping the libraries and archivist institution. Doing cataloging and indexing in digitizing have many area to develop as there is space for development but it is hard to decide in which way the digitization will develop. There is also current issues happening while doing digitization that need to look over. However, it will be a great loss if the digitization in indexing and cataloging is not being develop and used well. As the function of doing indexing and cataloging are important task in libraries and archival institutions.

References

Abioye, A. (2010). *Copyright: Interpreting the law for library, archives and information centers in Nigeria.* A Paper presented at the Cataloguing, Classification and Indexing Section of Nigerian Library Association (October 2010), 27-40.

El-Sherbini, M. (2001). Metadata and the future of cataloging. *Library Review,50*(1), 16-27. doi:10.1108/00242530110363217

El-Sherbini, M., & Klim, G. (2004). Metadata and cataloging practices. *The Electronic Library*, 22(3), 238-248. doi:10.1108/02640470410541633

Ma, J. (2006). Managing metadata for digital projects. *Library Collections, Acquisitions, and Technical Services,30*(1-2), 3-17. doi:10.1016/j.lcats.2006.07.001

Matusiak, K. K. (2006). Towards user-centered indexing in digital image collections. *OCLC Systems & Services: International digital library perspectives,22*(4), 283-298. doi:10.1108/10650750610706998

Nero, L. M. (2005). Cataloguing digital resources. *Library Review,54*(2), 100-107. doi:10.1108/00242530510583048

Nero, L.M. (2015). Manuscript libraries and archival description in the Caribbean. *New Library World*, *116*(5/6), 289-301. Doi:10.1108/nlw-08-2014-0098

Oketunji, S. F.; & Iyoro, A. O. (2009). *Effective workflow for cataloguing and handling problem materials in cataloguing departments.* A Paper presented at the Cataloguing, Classification and Indexing Section of Nigerian Library Association (October 2009), 80– 91.

Omekwu, C. O. (2008). Cataloguers in a global information network environment. *The Electronic Library,26*(2), 188-201. doi:10.1108/02640470810864082

Waugh, A. (1998). Specifying metadata standards for metadata tool configuration. *Computer Networks and ISDN Systems, 30*(1-7), 23-32. doi:10.1016/s0169-7552(98)00102-0

Willer, M. (1999). Formats and cataloguing rules: developments for cataloguing electronic resources. *Metadata And Cataloging Practices*, 33(1), 41-55. doi:10.1108/eum0000000006912

Why METADATA Matters - EBSCO Information Services. (n.d.). Retrieved January 2, 2018, from http://www.bing.com/cr?IG=88B3D157F9494FC89CD0651FA57E7500&CID=23C5853330076B2F139 48E5831A86A36&rd=1&h=WygJo1G5au0m1adGPuOKbh6AqfZ7ahHh01zbmADlUI&v=1&r=http%3a %2f%2fconnection.ebscohost.com%2fc%2farticles%2f23005802%2fwhymetadatamatters&p=DevEx ,508.1

Types of Indexes. Elmer E. Rasmuson Library. Retrieved January 2, 2018, from https://library.uaf.edu/ls101-types-of-indexes

Current Trends in Cataloguing and the Challenges of a Cataloguer in the Digital Age
Esse, U. %. (2013). Figure 2f from: Irimia R, Gottschling M (2016) Taxonomic revision of Rochefortia Sw. (Ehretiaceae, Boraginales). Biodiversity Data Journal 4: e7720. https://doi.org/10.3897/BDJ.4.e7720. CURRENT TRENDS IN CATALOGUING AND THE CHALLENGES OF A CATALOGUER IN THE DIGITAL AGE, 1(2). doi:10.3897/bdj.4.e7720.figure2f