RHYTHM OR MIX: DIGITIZATION AN END TO GLOBAL VISIBILITY OF INFORMATION RESOURCES?

Kasa, M. Gabriel
Agricultural Library
Institute for Agricultural Research
Ahmadu Bello University, Zaria

Abstract

This paper provides the stepping stone to successful digitization of information resources from the scratch. It gave benefits, characteristics of digitized resources, stages of digitization, different file format and digitization challenges. The paper displays how digitization is used to solve problems of access and global visibility of research, academic and the leisure information needs for effective development of the society. The study proposes learning from other initiatives, and gave basic equipment for all the processes. Digitized resources can be made available on-line through the internet and/or off-line on storage devices like CD-ROM, external/detachable hard disk drives to mention few. Information on the various document formats were given to facilitate choice and long-term consideration, the tool for digitization recommended are those very familiar to ease digitizing and correction which are two of the major constraints to effective implementation of digitization.

Keywords: Digitization, information resources, global visibility, access, tools

Introduction

Numerous advantages and benefits to institutions and researchers on the effect of their information resources and research output to global development have been discussed and expressed (Kalogeressis and Labrianidis, 2009). According to Bhattacharya (2004) global visibility of these resources must therefore be encouraged through digitization. Digitization can remedy and ensure the global visibility of information resources in Nigeria. Though various arguments emanate to express the beclouding constraints that include lack of funds, technical-know-how, poor infrastructural facilities, awareness, duplication of research, scholarship cheating and collaborations. Whatever form a resource is born (paper or electronic), digitization has comfortably serve as scholarship publisher and author. These roles enable it to ensure quality through editorial works, text formatting, peer reviewing and suitable presentation of research for public consumption.

Characteristics of Digitized Resources

Numerous reports on the characteristics of digitized resources have been made, which also hinges on diverse definitions. Gbaje (2007) defined digitization as a process of converting non digital born documents to digital formats. The study underpins the characteristics of digitized resources to the ones itemized by Gbaje (2007). Others who described digitized resources linked them to this (Smith, 1999;

UNESCO, 2002; Ochai, 2007). This includes:

1. Digital resources are not localized with ICTs. A digital record can be used from anywhere without any geographical restriction.
2. Several people can use the same records at the same time
3. Digital records can be stored electronically, arranged, indexed and retrieved easily.
4. Easily copied, flexible and less bulky.

Digitization and Access to Information Resources

Poor access to information is pointed to affect drastically the advancement and development of the human society (Gbaje, 2009). Ways to ameliorate the inadequacies as proposed since the turn of the 20th Century, to implement the universally tested and verified precepts now on the scale among developed and developing countries on scientific knowledge and experiences is digitization and supporting institutional repositories. Utulu and Akadri (2010) advocated relying on evidences and procedures available globally, develop new ones while expanding existing ones. It is also evident that lack and ineffective use of Information and Communication Technologies affect access to information resources. Training, re-training, upgrading, migration and follow-up were suggested as appropriate measures to salvage the changing roles of information professionals, researchers, scientist and the general public (Ezeani and Ekere, 2009).

To achieve the aims and objectives of digitization, professional skills are required, these skills enable the ability to convert, locate, evaluate and use effectively needed information. Reinforcing these are the description of contents and attributes of items digitized (Singh and Pinki, 2009).

Challenges of Digitization and Global Visibility of Information Resources

Digitization must be standardized as a process and as a system, linked to others to facilitate access on-line and remotely. According to Calhoun (2006), this process, referred to as "switching layer" (help) to leverage information delivery services, thereby positing that libraries in the future will be evaluated based on capability to provide users with technologies that allow applications to communicate across platforms and programming languages.

Experience has revealed that the major challenge to digitization lies with information centres which try to re-invent the wheels. Digitization is not expected to be embarked upon in isolation. In other word, it supports sharing and copying of what has been done successfully elsewhere so that dialogue to resolve identified constraints in the process of actualizing digitization from a beneficiary on a benefactor is readily obtainable. This will also resolve concerns expressed about poor access on the global level.

As technological development that preserves resource materials against deterioration and conserve space among the many benefits already mentioned, it will therefore be inappropriate to write it off on grounds of the economy, poor infrastructure, erratic power supplies or lack of knowledge. The challenges to be overcome are those constraining us from adding our contents and making them available for others to use. These may not possibly be available on the wider networks, but could be made available to others on CD-ROMS, higher institutional web sites and others whose support can facilitate the dissemination of our resources. It is also noteworthy that contents must cover vast areas. Today, common contents are those of theses and abstracts. We should be ready to have our institutional publications in public and private domains, academic research works, and undergraduate and postgraduate works and many more we take for granted.

The Challenge to digitization is the legal implications, constantly referred to as the copyright issue. To overcome this, the principle of fair use and enacting of clauses and gaining authorization from copyright holder can be overcome at personal and institutional levels. Agreements, bargaining and collaboration can hasten these processes with concise but deliberate persuasions. Sensitization and mobilization of immediate community will ultimately help to overcome some salient constraints, they as stakeholders can provide the support for growth. The challenge of capacity building can be overcome with human resource development, but, the need for the necessary equipment remains the tentative to achieve desired results.

Embarking on Digitization

The priority of any information centre must form the basis for digitization. This will help to define the purpose, process, beneficiaries, equipments, contents, standard, funding and targets. Getting this straight, the need to underscore the intellectual nature of the source material become imperative. This will define potential users, anticipated nature of use, the best format for the product and generating befitting metadata to facilitate access and retrieval. Digitization can commence with two network computers, a CD writer and a scanner with good optical character resolution.

Stages of Digitization

Data entry stage: This could be through typing, scanning and the utilization of Optical Character Recognition (OCR). Whatever form utilized, it is important to at the end, regularized the character encoding. There are incidences that typing becomes inevitable, as in the case of handwritten notes and when the product has poor clarity and degraded results.

Scanning stage: This could be achieved using inexpensive flatbed scanner even though it is slow. It is recommended that the scanning resolution should not be less than 300 dpi (dots per inch) and note that the higher the resolution, the better the results from an OCR program. It is really worth it to spend some time experimenting with the different scanning settings, this gives greatly improved OCR results. TIFF is the most commonly used file format for scanning and OCR. OCR program allows you save the files in the exact format you want.

Renaming files stage: Always make sure you name your files with cognizance because this will save you many headaches later.

Post-processing the images stage: This may not be necessary on all documents, but it is also recommended. Some suggestions include leaving ample margin when setting the crop area, since each page is always positioned a little differently when you scan. This allows further processing, like un-tilt the text and to crop away the dark margins, it helps to make the files smaller, which is a consideration if you're going to be publicly posting the page images.

It is recommended that you don't throw away your image files when you're done with scanning. You can fall back to them to see whether an apparent error is in the original or not. You may need to repeat the OCR at some point, it demands patience and care because often the output text may be so degraded that one probably gives up and enter the data by hand instead. However, there are usually some regularity to the errors; the same errors tend to happen over and over. We can take advantage of this partial regularity to greatly reduce the number of hand-corrections which must be made and change to the corrections.

Choosing File Formats

There are two important considerations

to keep in mind when considering what file formats to use:

1. Your digitized data will still be useful 10 years from now. However, it's very unlikely that very many people will still be using today's word processor formats 10 years from now. It might be difficult to find software which can even open the files.
2. It's often useful for data to be presented in more than one format. For example, one might want both an HTML and a PDF version of the same document. Conversions between presentation formats can be problematic; for example, it would be very difficult to convert a PDF file to an HTML file. Also, exporting from proprietary word processors is often not as straightforward as it seems.

It is because of these considerations that many professionals who have to manage data have adopted a model in which an abstract base document (often encoded in XML) can automatically be converted to any desired presentation format. XML has a simple and transparent structure which a programmer can easily work with; therefore data in XML will remain usable indefinitely.

Reasons for Considering File Formats

PDF and PostScript: These file formats allow you near complete control over the way the data appear when the user views them. However, the data are in a frozen form, meaning that it is basically impossible to extract the data from the file and use it for some other purpose. The user might want to reprocess and reuse the data in all kinds of useful ways which you might not have envisioned. So it's OK to use the formats as a supplementary presentation format, but it is strongly recommend against making them the only way to publish your data.

Microsoft Word and other proprietary word processors: Word has the advantage of being widely used by Windows and Macintosh users, at least at present.

Some significant disadvantages

The files are in a binary-barf format, which makes it very difficult for a programmer to manipulate the data. You can export the data to another format, but word processors rarely export the data neatly; there are usually anomalies in the output format which the programmer has to fix by hand. (If you are using Word for your base document, this means that this clean up will need to happen every time you make further corrections and re-export the data, which is obviously not a good solution).

Special characters, and text attributes such as bold and italic, might look good when you open the file with the word processor you created it in. However, since each word processor has its own quirky ways of storing this kind of information, these characters and attributes tend to get mingled or lost when trying to do other things with the data.

Most importantly, word processor formats tend to go obsolete fairly quickly. Microsoft changes the Word file format every few years. Think ten years down the road: your data will probably still be of valuable, but they will be useless if they are locked up in a word processor format which nobody uses any more.

The Issue of Tools for Effective Digitization

Always consider the tools available to you and are familiar with. The biggest thing is digitizing and correcting the text. Therefore go ahead and digitize even if you don't act on all of the recommendations made here.

Conclusion

Access to knowledge remains the only solution to sustainable growth and development. Digitization of information resources can facilitate access either on-line and/or off-line. Digitization do not require huge expenses to commence effectively, tools familiar with should be exploited to enhance competences. Digitization promotes knowledge creation and sharing, it provides the support necessary to overcome weaknesses on endeavours, promote collaboration and securing of funds for cutting edge findings.

References

Calhoun, K. (2006). *The Changing nature of the catalog and its integration with other discovery tools* (prepared for the Library of Congress) Available at http://www.loc.gov/catdir/calhoun-report.final.pdf.

Ezeani, C. and Ekere, J. (2009). Use of ICTs by Library Practitioners in Nigeria: Implications for the Library and Information Curriculum. *Paper presented at the NALISE Conference*, June 2 – 5, University of Nigeria, Nsukka.

Gbaje, S.E. (2007). Digitization and its challenges: Paper presented in a Workshop on Digital Records and Archival Management for Members' Federal Capital Territory, Abuja, Archives and History Bureau Committee and Liaison Officers, Held in Arewa House, Kaduna.

Gbaje, S.E. (2009). Global Visibility of Scholarly Journals Published in Nigeria. Nigerian Libraries: Journal of the Nigerian Library Association Vol. 42: 32 – 44.

Kalogeressis, A. and Labrianidis, L. (2009). De-Localization. In: *International Encyclopedia of Human Geography*, pages 60 – 69.

Ochai, A. (2007). Modern trends in the provision and use of library resources through digitization. Being a keynote address at the National Workshop on Digitization of Library Materials: Process and Tools at the Computer Center, University of Jos, July 16, 6p.

Bhattacharya, P. (2004). Advances in Digital Library Initiatives: A Developing Country Perspective. *The International Information & Library Review* 36(3): 165 – 175.

Singh, S.P. and Pinki (2009). *New skill requirements for LIS professionals in technology-intensive environment*. In: *International Conference of Academic Libraries* (2008), held at University of Delhi, 3-5 October. Pp. 331-336. New Delhi: Mittal Publications.

Smith, A. (1999). "Why digitize?". Council on Library and Information Resources. Available at: http"//www.clir.org/pubs/reportspub80.html.

UNESCO (2002). Guidelines for digitization projects for collection and holdings in the public domain, particularly those held by libraries and archives. UNESCO, Paris pp. 1 – 15.

Utulu, S.C.A. and Akadri, A.A. (2010). Institutional Repositories: The untapped academic goldmine. In: *University of Nigeria, Nsukka Proceedings of Second Professional Summit on Information Science and Technology (PSIST)*. Pp. 16 – 23, Nnamdi Azikiwe Library, University of Nigeria, Nsukka. Available at www.unn.edu.ng/library, 3- 7 May