Managing e-Content : Ways and Issues

Veena A Prakashe

Abstract

With the continuous downpour of e-content and web content in today's highly demanding world, it has become inevitable to design and choose an appropriate yet exhaustive e-content management system that would allow us to effectively create our own e-content which is viable and sustainable. For this purpose, an attempt is made to enumerate the most important features that one must look for in an e-content management system. Also, several issues of digitisation have been discussed that one often comes across while converting their documents into e-content.

Keywords: XML, Digital Archive, Digitization, E-content Management, E-content Empowerment.

1. Introduction

In today's knowledge society, information centres differentiate themselves by their capabilities for managing information. The effectiveness today is predicated on a library or information centre's ability to deliver the right information to the right person at the right time, and by its ability to leverage information to improve services, user satisfaction, and overall efficiency.

The future of empowerment would lie on how dominating you are in terms of creating content and putting them on the digital media, especially the content pertaining to commerce and culture. It is now fair to assert that technology is as good as it delivers and it enables. It is extremely important for information and communication technologies to turn around in terms of their impact, usability, and result in empowerment at a large scale.

In the age of digital revolution and ICT boom, the most obvious issues are whether ICT is accountable for whatever it is doing, and whether ICT is enabling its users to create content or not. Whether ICT is empowering people to create digital content, content which converts into economy and commerce, content which empowers the mass with information and knowledge and the content which enables the mass to know their rights and wrongs and finally the content which not only pose the typical Indian mass into consumer but also providers through global communication media. Realising this need well in time, many companies have come up with e-Content management systems and are competing with each other to woo the clients.

2. e-Content Management System

e-Content management services are primarily focussed towards the digital preservation and access of newspapers, books, journals and a variety of historic handwritten documents. The digitisation and bibliographic conversion services help libraries, museums, repositories of special collections and other organisations to digitise their manuscripts, catalogues, finding aids, theses and

other valuable collections. With a robust data security policy & business continuity plan in place, one can fully safeguard customer data at all times.

In enable one to choose the most appropriate e-Content management systemas also to ascertain and test the claims as put forth by the suppliers, the information professionals need to have some yardsticks or guidelines to compare and evaluate these systems. Therefore, some most important features or facilities that are expected in these e-Content management systems is discussed below which one can look for in any such proposed system.

2.1 XML/SGML

Extensible markup language (XML) is the universal format for structured as well as unstructured content on the Web and is a subset of SGML . We can convert media rich, unstructured paper, microfilm or scanned documents to powerful XML files. This involves the following activities:

- Analyzing Your data.
- Developing/modifying DTD/Schema
- Writing conversion specifications
- Pre - Migration data clean-up.
- OCR / Data Capture
- SGML/XML encoding
- Proofing,
- Validation and QC Audits
- XML FO transformation

2.2 Bibliographic Services

It should provide for converting card catalogues to MARC21, MARCXML or any other format. The scanning of also has expertise to perform EAD (Encoding of Archival Descriptions). Since most card catalogues are in constant use in Libraries, AEL can organize the scanning of the catalogue cards should be organised in such a manner that it does not cause disruption of use of the cards in the library

2.3 Newspaper Digitisation

An end-to-end solution for the digitization of archival newspapers include:

1. Scanning of paper copies / microfilm
2. Image Enhancement
3. Article zoning
4. OCR
5. Headline keying
6. XML encoding
7. Web Delivery of digitised newspapers.

2.4 Microfilm/ Microfiche Imaging & Conversion

This being highly technology intensive task, it calls for state of the art high high speed microfilm & microfiche scanning and printing equipment to scan black white or grayscale images from microfilm or microfiche, index & convert the text and images to a full-text searchable (XML /SGML/HTML) format or to PDF format.

In addition, microfilm, microfiche positive or negative film can be processed and printed as well as written to CD-ROM, DVD-ROM or Optical Disc for instant On-Demand service.

2.5 Digital Archiving

Digitizing full-text searchable newspapers, magazines, legal, scientific, technical & medical journals, government papers and other archives is required to be done for on-line publication. Data is captured by OCR / manual data entry and encoded as per archive standards. One must comply with archiving standards such as ISAD(G) published by the International council on Archives.

2.6 Yellow and White page conversion

The e-Content management system should perform Yellow, White pages, and other directory conversion services in many languages. We have the software and latest address databases to validate the street address usinf the post code for most countries.

2.7 Forms Processing

Many types of Handwritten forms do not easily lend themselves to automatic capture using ICR/OCR tools. Data capture is from images or paper forms. Efforts should be made to ensure high accuracy output. Following are certain essential management functions of an e-Content manangement system for form processing:

1. Document Preparation
2. Forms Scanning
3. Automatic Forms
4. Data Capture
5. Manual Forms Data Capture
6. Document Warehouse
7. Document Retrieval
8. Database creation/ maintenance :

To create and maintain small and large databases of various kinds, e.g. Technical, Scientific and business database / Indexes, directories, mailing addresses, customer lists, library cards (Marc 21) & indexes and catalogues.

9. Extraction/Maintenance / Validation from the Internet :

To extract data from the web, create a database to your requirement and periodically maintain your database by regular updates.

10. Maintenance / Validation by calling the customer

11. Data Security & Privacy :

Implementing an Information Security Management System to ISO 27001:2005 standard to improve its internal security processes and enhance customer confidence for the Organisation.

3. Digitisation

Given below is a discussion on the fundamental issues associated with the digitisation process .

3.1 Originals

Having a good knowledge of the contents of the collections that are intended to be digitised will make it much easier to decide on processes and techniques for converting the originals to digital form.The physical processes required to create a digitised version of an original item depend on many factors, including:

- The format of the original - is it printed text, photographic material, video, audio etc.?
- The condition of the original – will it stand up to automated procedures (if used), will conservation be required before scanning?
- The size of the original
- The colour content of the original and whether that colour is important.

For paper and photographic originals, issues to consider include the following:

Photographic media (transparencies, prints, negatives)

- What size are the originals, are they all the same size? It makes for a smoother workflow if items of a similar size are grouped together.
- What proportion of the items have colour content? Is it important to capture the colour?
- What condition are they in, for example, are they dirty from heavy use? If they are dirty a better scan will be achieved if the items can be cleaned first.

- What format are they in? Slides in sleeves or strips will take longer to prepare for scanning and may cost more if a bureau is scanning them.
- Glass negatives are prone to breakage and require careful handling.
- Are the photographs flat or have they bowed? Bowed originals cause difficulties with focus and may need weighting down.
- What is the quality of the original? A bad original (i.e. out of focus) will not be improved by scanning.

Paper media

- What size are the pages, are all items the same size?
- What general condition is the material in? Pristine pages will produce a better result and the scanning process may be able to be automated. Any damage in an original may be exacerbated by the scanning procedure.
- Can books that are bound be stripped to loose pages for scanning? Scanning from bound volumes is more complex and therefore expensive than from loose pages.
- Is there any artwork? – is it black and white or colour photographs or line art? Colour scanning is generally more complex and resource intensive.
- Is the text size particularly small or large? Very small text may need a higher resolution to extract the information.

Objects require a different approach. Artifacts, art works and sculptures cannot generally be successfully scanned using the techniques available for 'flat' media such as photographs. It will therefore be necessary to use photography, either traditional or digital, to get an image of the original.

Cost Remember that real costs and prices are bound to vary from those given in any guidance documents and it is essential that such guidance is used purely as a starting point in accurate costing.

3.2 Creating a Digital Master

It is recommended that the digital preservation issues be observed when producing digital content. A good baseline to creating a digital file that will be long-lasting would be Scan Once for All Purposes – this means that all the complex and expensive preparation work will only need to be done once.

The resultant would be a fully documented high-quality 'digital master' from which all other versions (e.g. compressed versions for accessing via the Web) can be derived. This 'digital master' file should be created at the highest suitable resolution and bit depth that is both affordable and practical. This master file then becomes the source for every other version of that item that the project will require, such as Web surrogates, versions for high quality printing and so on.

The 'digital master' file will become an archive version of the data – it remains as pure a representation of the original as possible. Ideally more than one copy should be stored on more than one media type and in more than one geographical location, thus providing a degree of protection against data corruption, media failure and physical damage to equipment.

'Surrogate' or 'access' versions of the digitised item can be created from the 'digital master' using image manipulation software such as Adobe Photoshop or Paintshop Pro.

3.3 Choosing Scanning Equipment

Digitisation equipment can be separated into 'contact' and 'no-contact'. 'Contact' equipment, i.e.

flatbed scanners, requires that the original be flat against the scanbed to get a scanned image. This approach will only work if your original is flat or can be pressed flat without damage to it.

No-contact equipment includes overhead scanners or book scanners and digital cameras that are able to obtain a digital image with the bare minimum of contact with the original.

Choosing the equipment for scanning your originals will depend largely on the characteristics of the collection: in general terms, photographic materials are usually scanned on a flatbed or a transparency scanner while bound volumes and oversized flat materials such as maps and plans require a digital camera or an overhead scanner.

If you have a mixed media collection then it may not be possible to use one scanner for everything. A flatbed that is ideal for high speed, high volume paper scanning may not be capable of the resolution required for high quality scans of transparencies. A digital camera studio set-up will be overkill for loose leaf paper scanning and for most general photographic materials.

Generally, make sure that your requirements match the capability of the scanner(s) that you buy. Look carefully at the resolution that the scanner is capable of, the scanner will often be listed with a maximum optical resolution and an interpolated or software resolution. The optical resolution is the figure to look for.

The software that runs the scanner is also important. It should be straightforward to use and an ability to run batch scans will save time as the scan bed can be loaded with originals and more or less left to get on with it.

Colour management software is essential to ensure that the digital representation is as accurate as possible. This can often be purchased with the scanner.

Digital cameras : 'Home use' cameras are aimed at non-professional users for taking general casual photography.

There are two kinds of professional digital camera; the first has developed from medical and industrial uses and is a complete unit. The second is where the film from a traditional camera is replaced with computer sensors which transmit the image to a computer rather than to film; this is known as a digital scanning back.

3.4 Set up an in-house scanning unit or use a bureau?

The conversion of the materials can be done either in-house on specially purchased or existing equipment or sent to an external agency or commercial bureau.

Setting up a digitisation unit gives the institution the value of equipment and trained staff for future projects and the movement and treatment of the materials can be closely controlled. Using an external supplier to do the scanning means that the equipment and expertise of a third party can be exploited while the project team concentrates on their specialist area of the project. Using a bureau also means that the cost of buying and maintaining specialist and expensive equipment is not borne by the project.

Both approaches have their merits but there are certain situations where the choices are more clear cut.

4. Concluding Remarks

This document is intended to be used as a means of focusing attention upon the key issues associated with the e-Content management process. The advice that it contains is intended as guidance rather than as the only solution to these issues. There may be valid institutional or curatorial reasons for following or discarding different aspects of this guidance, especially in relation to the handling of original materials that may make certain processes unsuitable for that class of material.

References

1. **Francis, A. T. and Sushama Devi, C. K. and Abdul Razak,** C. Content Management for Digital Delivery of Agricultural Information: Redefining need of libraries in the context of digitization of theses and research reports., 2007 . In International Conference on Semantic Web and Digital Libraries, ICSD 2007, Bangalore (India), 21-23 February, 2007.

2. **Holley, Rose** The Australian Newspapers Digitisation Program: Development of the Newspapers Content Management System., 2008 . In ANPlan - ANDP

   Workshop, Canberra (Australia), 28 November 2008.

3. **Patel, Yatrik and Vijayakumar, J. K. and Murthy, T. A. V.** Institutional digital repositories/E-Archives : INFLIBNET's initiative in India., 2006 . In 7th MANLIBNET Annual National Convention, Kozhikode (India), 5-7 May, 2005.

4. **Peset, Fernanda and Ferrer-Sapena, Antonia** E-LIS : Central repository on Library and Information Science., 2008 . In Online Information 2008, London (UK), 2th December 2008.

5. **Rajasekharan, K. and Nafala, K. M.** Creating Digital Document Archives with Winisis & Genisisweb Software., 2007 [cited 1/19/2009]. Available from WWW: http://eprints.rclis.org/9027/. [Guide/Manual]

6. **Raman Nair, R.** . Digitization of indigenous materials : problems and solutions in the Context of Kerala University., 2006 In: Library and Information Studies in the Digital Age. Ess Ess Publications, New Delhi (India), pp. 222-247.

7. **Sridhar, M. S.** Information Management in Digital Environment: A Librarian's Perspectives. DESIDOC Bulletin of Information Technology, 2007, vol. 27, n. 4, pp. 45-50. [cited 1/19/2009].

About Author

Ms. Veena A Prakashe, Information Scientist,
RTM Nagpur University, Nagpur.
E-mail: sh_veena@hotmail.com