DSPACE

https://dspace.org/

DIGITIZATION OF ARCHIVAL RECORDS: THE KENYA NATIONAL ARCHIVES AND DOCUMENTATION SERVICE EXPERIENCE.

Namande, Ben Wekalao

2012-06-14
http://hdl.handle.net/10539/11522

Downloaded from a DSpace repository. https://dspace.org/

DIGITIZATION OF ARCHIVAL RECORDS: THE KENYA NATIONAL ARCHIVES AND DOCUMENTATION SERVICE EXPERIENCE.

Ben Wekalao Namande
Principal Librarian, Kenya National Archives and Documentation Service.
Ag. Deputy Director, Department of Government Library Services,
Ministry of National Heritage and Culture, Kenya.

Abstract

The Kenya National Archives was established in 1965 by an Act of Parliament to provide custody, among other functions, to public and private records of enduring value. It is worth noting that some information resources at the Kenya National Archives were created as early as the 17th century. Some of these old materials have since deteriorated as a result of both environmental and biological dangers and become fragile, brittle and discoloured. This situation highlights the need for long-term availability of such resources.

This paper is a sojourn in the digitization endeavours at the Kenya National Archives. It shows what has been done and the challenges that have been experienced. Despite the four-phase attempt to digitize over 680 million pages of records, the process has experienced myriad challenges those of particularly inadequate human capital and financial resources.

Digital information centres: an overview

A digital information centre is an information facility in which collections are managed in digital formats (as opposed to print, microform or any other media) and are accessible by computer. The digital content may be stored locally or remotely and accessed remotely via computer networks. A digital library is a library in which collections are stored in digital formats (as opposed to print, microform, or other media) and accessible by computers. The digital content may be stored locally, or accessed remotely via computer networks. In a nutshell, a digital information centre is one in which a significant proportion of resources is available in machine readable formats as opposed to print or microform, and is accessible by means of computers.

Background

The concepts "digital" or "electronic" libraries are not really new. It may be recalled that Lancaster wrote extensively on the paperless society and on electronic library developments in the late 1970s and early 1980s. The development of digital libraries could be said to have been started by the National Library of Medicine at Bethesda, in the United States, with its use of computer controlled photo composition or computer typesetting in the production of Index Medicus in 1964. This was a transitional phase in evolution from print to paper to electronics. Digital resources have greatly changed the way in which libraries function as information centres. With the transformation from traditional print collections into versatile electronic

resources, information including full text, can now be accessed from laboratories, offices and home 24 hours a day.

Requirements for a digital information service

The development of a digital information service involves: digitization of existing library materials; integration with networking; connectivity to the user in the world online and offline; and availability on the World Wide Web. To achieve these, the following are important:

Technical architecture

The first requirement that underlies any digital information system is that of the technical architecture. According to Chepsesuik (1997), libraries will need to enhance and upgrade current technical architectures to accommodate digital materials. The architecture comprises components such as:

- High-speed local networks and fast connections to the internet;
- Relational databases that support a variety of digital formats;
- Full text search engines to index and provide access to resources;
- A variety of servers, such as Web and FTP servers;
- Electronic document management functions that will aid in the overall management of digital resources;
- Hardware: desk tops, hard discs, UPS, printers, scanners.

Trained staff

With the introduction of the digital age, the library and information profession has assumed new trends and therefore library and information professionals have to keep themselves abreast with current technology. According to Sharma (2001) librarians must prepare themselves for the transformation from an era of scientific management to systems and structural management. Tennant (2006) asserts

> … we need an agile, imaginative, and engaged staff. We need people who are not afraid of jettisoning traditional activities in favor of new ones. We must have people who can learn constantly, foster change, and create new kinds of collections and services. Some of these people are in our buildings; others will need to be hired. We must train, support and encourage them.

Digital software

Once one has an idea of how a digital information centre should work, it is important to select some software. If a particular piece of software can meet the requirements, this is good; if it cannot there are three options: customize the software to meet your needs, find a different piece of software, or adjust your requirements. Some level of customization is to be expected in any digital project. Someone with reasonable experience with Hyper Text Markup Language (HTML) and style sheets should be able to achieve this. More complex customization requires more time, effort, and skill. Weigh those against the value added and the resources available.

Building collections in digital format

One of the most important requirements for creating digital information resources is the building of digital collections. For any digital information centre to be viable, it must eventually have a digital collection with the critical mass to make it truly useful. There are essentially three methods of building digital collections:

- digitization, converting paper and other media in existing collections to digital form;
- acquisition of original digital works created by publishers and scholars (born digital), (example items would be electronic books, journals, and data sets);
- access to external materials not held in-house by providing pointers to web sites, other library collections or publishers' servers.

Why digital information centres?

In the ideal digitized information environment, the following features of added value may be realized:

- **Multiple access to information.** The same resources can be used simultaneously by a number of institutions and patrons. Digital libraries provide a faster method of accessing and exchanging information in all sectors such as research, scholarship, medicine, government services and business.
- **Resource sharing.** Digital information can easily be shared and therefore is available to everybody, as opposed to conventional information materials which require expensive duplication of material in order to meet the needs of many users.
- **Remote access**. Digital libraries can make it possible to provide everybody, even those in the remotest parts of the Africa, with information in its most up-to date form. With a link to the information superhighway, libraries in the region, even the smallest and the remotest, can serve as a gateway to local as well as global information.
- **Increased accessibility.** Digital libraries are available on a 24 hour basis from anywhere in the world, offering flexible arrangements for students, researchers, scholars and the community. Digital libraries can be widespread and accessed as full text from any location or workstation.
- **Information retrieval.** The user is able to use any search term (word, phrase, title, name, and subject) to search the entire collection. Digital libraries can provide very user-friendly interfaces, giving clickable access to resources.
- **Preservation and conservation.** Digitization is not a long term preservation solution for physical collections, but does succeed in providing access to copies for materials that would otherwise fall into degradation from repeated use.
- **Added value.** Certain characteristics of objects, primarily the quality of images, may be improved. Digitization can enhance legibility and remove visible flaws such as stains and discoloration.
- **Space.** Whereas traditional libraries are limited by storage space, digital libraries have the potential to store much more information simply because digital libraries require very little physical space to contain them and media storage technologies are more affordable than ever before. The cost of maintaining a digital library is also much lower

than that of traditional library, since a traditional library must spend large sums of money paying for staff, book maintenance, rent, and additional books. Digital libraries do away with these costs.

It is important to note here that until the project is fully realized, KNA&ADS cannot be measured against these advantages.

Digitizing records at Kenya National Archives and Documentation services (KNA&DS)

Goal: To digitally preserve content and context of archival records and other endangered publications for purposes of access and dissemination.

Objectives:

- Develop institutional digital information infrastructure to ensure preservation and longevity of archival information resources;
- Develop online *Record Search* databases where researchers are able to submit requests for records;
- Offer multiple, faster and cheaper access to resources, reference processes and resource sharing.

Justification for the project:

The core function of any archival institution is to ensure long-term preservation of valuable records for posterity. Consequently, the major objectives of the KNA&DS include ensuring that the records in its collection are accessible to all those who need to use them, irrespective of where they live. This is a challenge because the collection is large (several shelf kilometres) and not all potential users have access to our Reading/Search Room. In addition, most researchers use the Archives online catalogue, accessed on the KNA&DS website, to identify records that are most useful for their research. They must then:

- visit the Search Room at the Archives; or
- hire a Research Assistant to do research on their behalf;
- request and pay for reprographic services and postage; and
- be patient until the information is retrieved and delivered.

Statistics show a steady increase in the number of materials utilized for academic research from 4600 in 2001 to 5900 in 2004, over 7000 in 2008 and about 10,000 currently.

The aim of this digitization project is therefore primarily to offer researchers faster and cheaper access, probably when the material has been uploaded on to the online *Record Search*. This would mean that the records could be read on-screen or printed at home, thus saving researchers a visit to the Search Room and the expense of paying for reprographic services. Additionally this would go a long way to enhancing the exploitation of archival resources by the global research fraternity.

Thus digitization itself not only lends credence to the production of images of the originals but also provides all sorts of added features including the creation of searchable text files, indexing and remote use and consequently it becomes open to reciprocal hyperlinking with other online

resources. The potential of remote usage, the ease of accessibility and the digital add-ons have tended to make digitization the method preferred by knowledge professionals as well as researchers for reformatting and repackaging.

Further to the above, Kenya National Archives does not operate the equivalent of an inter-library loan service. Most records in the collection exist in single copies. This makes them unique and irreplaceable. To transfer them between the storage areas and Search Room exposes them to too great a risk of loss or damage. Digitization reduces the physical handling of the original records since the captured images can be made available online to the users while the originals are permanently preserved for posterity.

Other opportunities to be provided by this project are those of establishing backup copies of our archival collection that could then be stored offsite. At the moment, the Archives does not have either backup or offsite storage for its unique and valuable documents, hence the likelihood of unfathomable losses in case of a disaster.

Methodology

The process starts with pre-digitization, then digitization, metadata capture, post-digitization and finally monitoring and evaluation. At KNA&DS the detailed overview of its digitization process can be outlined as follows:

Baseline survey (pre-digitization stage)

According to the Quarterly work report (2008), the baseline survey cum assessment was carried out in 2007 by the KNA&DS to establish the nature of about 1.7 million Coast Province records. This was phase one of the project. The records comprised manuscripts, handwritten and typed documents of varying sizes. It was established that about 30% of the records were fragile and brittle and hence required special equipment and handling when scanning in order to avoid damage. Due to repeated physical handling by users over the years, the materials had suffered wear and tear. It therefore became necessary to create surrogate files of the documents for use while original documents are preserved permanently unless under extreme circumstances.

It was established that about 10% of the resources to be digitized required wide format scanners since they were not in standard sizes. Filtered copies from digital masters such as PDFs were to be done with metadata and full text search capabilities for retrieval. Due to yellowing and fading, extra manual editing was to be done before and after optical character recognition (OCR) for full text searchable PDFs. A summary of the state of records shows:

Fragile pages	30%
Normal	70%
A4 pages	90%
Non A4	10%
Legibility	Good
Yellowing	97%

Selection of materials

This was one of the outputs from the survey. In the first phase, 1.7 million document pages were selected for digitization. In carrying out the selection exercise, some of the key factors which were considered included physical conditions, age, uniqueness and demand, over and above the fact that such documents exist in the public domain and are not tied to copyright restrictions. The selected materials were prepared for the further steps through cleaning, demagnetizing, unclipping and batching.

Equipment for digitization and other resources

Digitization is an equipment-intensive venture unless the institution already has in place some degree of computerization. Some of the equipment needs that KNA&DS had to contend with included:

- **Computers hardware and software:** KNA&DS had to purchase 50 computer workstations for the exercise**.** In addition to an upper range of RAM, disk storage, and processor speed, the computers had to be fast enough to handle scanning and other digitization tasks. CD-ROM drives were important for creating and copying CDs. Adobe Photoshop application software was adopted in the department because it is the de facto standard in use. The quality of its images is also higher than those produced by other software.
- **Scanners:** All the computers were fitted with quality flatbed scanners to do adequate scanning. It was realized that a large bed scanner was also needed and therefore was acquired for oversize special collections.
- **Storage:** The department adopted server-based storage of the digitized information with an offline backup in CDs.

Staffing

It is important to note that because of inadequate staff capacity (both in quality and quantity) the initial digitization services were outsourced from commercial professional service providers. To supplement the staff, the consultant hired casual employees. The challenge arising from this arrangement was that the temporary work force was more driven by profit motives and survival than by professionalism. Most of the casual workers recruited had elementary knowledge in IT and minimal knowledge of information management. This called for regular professionals to constantly check and correct work done by the casual employees to ensure that digitization standards were adhered to.

Scanning

The process for scanning documents entailed:

1. Creation of master image, saved as a TIFF (Tagged Image File Format) file;
2. Generation of derivative or access images from the master image;
3. Transferring the master images onto server and backup CD-R disks.

This process is associated mostly with paper-based records; the department does however also digitize audio resources from their analogue forms to digital, simply using a digital converter, the analogue tape player and a computer.

Quality Control

For the success of the programme, it was realized that quality control was essential. Among the key control strategies used in the KNA&DS were round-the-clock supervision by senior staff, quantitative summaries of digitized records, and periodic reports which factor in achievements and challenges. The department came up with a standing committee to monitor digitization activities: quality control is very important in error rectification since even the most conscientious person doing digitization will make some errors.

To date, the above processes have been implemented to some extent. Due to human, financial and technical hurdles, however, there are other steps which have so far not been addressed. These include:

- **Text encoding:** eventually the department will have to transcribe the digitized images into text format as part of the digitization process. This can be done through Optical Character Recognition (OCR) technology. When this is done, the resulting text may then be presented on a web page or encoded as SGML or XML for inclusion in a searchable database.
- **Metadata:** this describes the content and attributes of an item in a digital format. It is a concept familiar to librarians because it is one of the primary things that librarians do - they create cataloguing records that describe documents. (Cleveland, 1998).Thus simpler schemes for metadata need to be introduced to organize the digital information that KNA&DS has created. It is envisaged that metadata will be derived from finding aids already in the department.
- **Online Delivery:** online delivery of content is probably the ultimate goal in any digitization project. Unfortunately KNA&DS is still far from realizing this. Basic HTML web pages are the simplest way to put digitized materials online. More advanced methods include web-accessible databases, and XML or SGML solutions. These options have to be addressed for the successful implementation of the programme.

Cooperative digitization venture

In its endeavour to convert information resources into digital format, the KNA&DS has partnered with other government organizations in undertaking the exercise. Key in this venture are the National Council for Law Reporting and the National Assembly. The National Council for Law Reporting is a state corporation whose mandate is the publication of law reports containing judicial opinion of the Court of Appeal and the High Court and other related publications (Kenya Law Reports, 2008).

A consensus was reached on the need to work jointly in the venture of creating and managing digitally created documents in workflow processes across organizations to reduce duplication of

effort and costs of reformatting documents. It was agreed that a common format and integrated workflow processes at each step of the different phases of document lifecycle in different organizations be adopted. Together with the National Assembly, the Council for Law Reporting partnered with KNA&DS to enhance public access to information by digitizing the Laws of Kenya, Kenya (Official Gazette) and Hansard reports to include content which captures our country's legislative history. All legislative enactments since 1900 have been digitized. The KNA&DS role in this was the provision of hardcopies of Legal notices, Acts and Hansards for purposes of digital conversion. The Council for Law Reporting thus undertakes the digitization process and gives digital copies (banned CDs) to the KNA&DS AV library for preservation. The same is shared with the National assembly.

Challenges of digitization at KNA&DS

In view of the above experience, a number of challenges must be addressed for KNA&DS to reap the full benefits of digitization and to be at the cutting edge of the new information order. Notable among these challenges include:

- **Low internet connectivity.** Bandwidth is generally low in developing countries. This affects data transfer rates and access speed for digital resources. In Kenya, the broadband capacity has systematically improved in recent years; the effect has not however been felt across the board. E-resources require high bandwidth to be accessed effectively and because of the low connection speed witnessed in most libraries, users tend to grow impatient as it takes a long time to download information. This may become a thing of the past in eastern Africa once the Eastern Africa Submarine Cable System (EASSy) project is completed. The EASSy is an initiative to connect countries of eastern Africa via a high bandwidth fibre optic cable system to the rest of the world. Until this is done, institutions like KNA&DS will continue to operate below the necessary broadband threshold for effective service delivery.

- **Low levels of ICT literacy/e-resource use among users.** Whereas full exploitation of digital resources and services call for detailed knowledge of ICT tools and information searching skills, most internal and external customers of KNA&DS only have basic ICT literacy skills that scarcely extend beyond Microsoft Office applications. The institution faces a glaring challenge of investing in serious staff development initiatives in readiness for digital service delivery environment.

- **Frequent power cuts.** Power surges are a common occurrence in most of the developing countries; this is due to unreliable power supplies. Coupled with this are the high power tariffs being witnessed especially in Kenya; such tariffs make power very expensive. Information technologies require power for them to function and therefore with the frequent power cuts, e-dependent users remain frustrated.

- **Information security.** Information security is the means and method of protecting data from unauthorized access, theft, alteration or deletion, and ensuring continued

ability to access data whenever required (Kumar, 2004). Maintaining information security in a digital environment is one of the most demanding challenges for the KNA&DS. In the electronic environment, information is prone to many threats such as viruses, hackers and masquerades, among others. The average employee of KNA&DS is less empowered to confront this challenge.

Digital divide

Access to digital repositories and their collections is dependent upon a stable information technology infrastructure. Hence, despite the egalitarian potential of the digital information service, many of those who could most benefit from its global reach (for instance in the third world) are not able to do so. Smaller libraries and repositories in developed countries may also have limited resources in dealing with long term digitization projects.

Copyright management

Copyright has been called the single most vexing barrier to digital library development. Digital age is posing a big challenge for libraries, and the challenge is to protect intellectual property rights in a networked world. The issue of copyright also includes information ethics; the use of library networks for access, copying, and printing of copyrighted material without permission can have serious repercussions for the library. In the digital age it has become difficult to protect intellectual property rights, because of the lack of understanding among users that copying of a stream of bits without necessarily reducing the availability of that stream to other users can potentially be considered as a breach of intellectual property rights (Chepesuik, 1997:49). In order to develop common grounds for the enforceability of law, it is necessary to educate users as well as information professionals about the importance of ethical and moral behaviour.

Preservation

Digital materials are very fragile and their availability depends on technologies that are rapidly changing. In fear of technological obsolescence and other factors, institutions like KNA&DS have to maintain records in both paper and digital formats. This almost amounts to double work. Technical obsolescence in the digital age is synonymous with the deterioration of paper in the paper age. Libraries in the pre-digital era had to worry about climate control and the de-acidification of books, but the preservation of digital information will mean constantly coming up with new technical solutions.

Technological ripples and obsolescence

Technological standards change over time and forward migration must be a constant consideration of every library. Migration is a means of transferring of unstable digital objects to other more stable formats, operating systems, or programming languages. Migration allows the ability to retrieve and display digital objects that are in danger of becoming extinct.
Migration is also flawed in the sense that when the digital files are being transferred, the new platform may not be able to capture the full integrity of the original object. There are countless

resources in libraries all over the world that are essentially useless because the technology required to access the source is obsolete. In addition to obsolescence, there are rising costs that result from continually replacing the older technologies.

Metadata

The revolution to metadata from traditional methods of collection management presents a challenge to information mangers globally. Simpler schemes for metadata need to be introduced to organize digital information as opposed to complicated ones which also pose difficulties to end-user searching digital information. These can be listed as follows.

- Selecting source: end-users have difficulty in choosing from the many Internet search engines and in keeping up with the features they offer.
- Formulating queries: people have difficulty in expressing their information needs in the syntax appropriate to the search system they are using. They use only a few words in their search strategies, often even misspelling them. They are reluctant to explore the more sophisticated features many search systems offer, such as phrase and name searching.
- Examining results: they accept that a significant number of the hits they do get will be irrelevant, because their strategies are matched against the whole of the Internet and the ambiguities in their search strategies produce noisy search results

Way forward

It is encouraging to note that KNA&DS has made tremendous strides in digitization of its information holdings and is indeed a case study for most information centres in the country. However, the fact that the materials so far digitized have not been taken to the higher stages in the digitization process (especially online delivery) negates the essence of the whole process. It is therefore the recommendation of this paper that the relevant adjustments, website maintenance, software, skills and other resources are deployed to realize the fundamental objective on which the project was grounded.

References

Chepsuik. 1997. The future is here: America's libraries go digital. *American Librarie*s.

Greenstein, D. and S. E. Thorin. 2002. *The digital library: a biography* [sic]. Digital Library Federation, September.

Gary Cleveland. 1998. Digital libraries, definitions, issues and challenges. UDT Occasional paper no. 8. IFLA

Graham, P.S. 1995. Requirements for the digital research library. Available at: http://aultnis.rutgers.edu/texts/DRC.html

Kenya National Archives. 2008; 2009; 2010. Departmental quarterly Reports (1st, 2nd, 3rd and 4th quarters)

Kenya National Archives. 2008; 2009; 2010. Departmental annual reports. Nairobi.

Kenya Law Reporting. 2008. Document imaging partnership proposal. Nairobi.

Kumar. 2004. Information security and disaster recovery. India. Unpublished.
Tennant, R.2004. *Managing the digital library*. New York: Reed.