

E-Archiving, the Future and Our Challenges

ENDRIT MUSAJ, PH.D. CANDIDATE

Head of Evidence and IT Unit, General Directorate of Albanian Archives, Rr. Jordan Misja, Tirana, Albania
e-mail: emusaj@albarchive.gov.al

ARDIANA TOPI, PH.D. CANDIDATE

Head of Audio-Visual Unit, General Directorate of Albanian Archives, Rr. Jordan Misja, Tirana, Albania
e-mail: dtopi@albarchive.gov.al

E-Archiving, the Future and Our Challenges

ABSTRACT

Nowadays the archives have solved the problems related with traditional archiving such as classification, preservation, storing, accessing of paper documents. They have successfully store our national heritage, our national history from one generation to another saving the documents in their original form as they are created before hundreds of years ago. Today we are facing another reality in all fields. This reality named "Information Technology" is changing our ways of working, and our ways of living. In archival field is the same situation and in order to be successfully in our mission during coming years, we should answer a lot of questions that are related with implementing and using information technology in archival procedures and in preserving of either digitized or 'born digital' content. Those are the main topics treated in our following study in Albania.

Archiviazione elettronica: il futuro e le nostre sfide

SINTESI

Al giorno d'oggi gli archivi hanno risolto i problemi relativi all'archiviazione tradizionale, quali classificazione, conservazione, archiviazione, ed accesso dei documenti cartacei. Essi sono riusciti a conservare il nostro patrimonio nazionale, la nostra storia nazionale, da una generazione all'altra salvando i documenti nella loro forma originale così come sono creati centinaia di anni fa. Oggi ci troviamo di fronte un'altra realtà in tutti i campi. Questa realtà, chiamata "Information Technology", sta cambiando il nostro modo di lavorare ed il nostro modo di vivere. In campo archivistico la situazione è la stessa e per ottenere successo nella nostra missione negli anni a venire si dovrebbe rispondere a molte domande collegate con l'attuazione e l'utilizzo di tecnologie informatiche nelle procedure di archiviazione e nella conservazione di contenuti digitali o 'nati digitali'. Questi sono in Albania i principali temi trattati nel nostro studio che segue.

E-arhiviranje, prihodnost ter naši izzivi

IZVLEČEK

Dandanašnji so arhivi rešili probleme povezane s tradicionalnim arhiviranjem, kot je klasifikacija, hramba, skladiščenje, dostop do dokumentov. Uspešno hranijo našo nacionalno dediščino, našo nacionalno zgodovino od generacije do generacije, ohranjajoč dokumente v njihovi izvorni obliki, v kakršni so bili ustvarjeni pred več sto leti. Danes se na vseh področjih soočamo z novo realnostjo. Realnost se imenuje informacijska tehnologija, ki spreminja naš način življenja in dela. S tem se sooča tudi arhivistika in, če želimo biti uspešni v naši prihodnji nalogi, moramo odgovoriti na mnoga vprašanja povezana z uvajanjem ter uporabo informacijske tehnologije v arhivskih postopkih ter pri hrambi digitaliziranih ali digitalnih vsebin. To so glavne teme našega prispevka, temelječega na albanskih študijah.

Arkivimi elektronik, e ardhmja dhe sfidat tona

ABSTRAKT

Në ditët e sotme arkivat kanë zgjedhur problemet që kanë të bëjnë me arkivimin tradicional sic janë klasifikimi, ruajtja, aksesimi i dokumente në letër. Ato kanë ruajtur në mënyrë të sukseshme trashigiminë tonë kombëtare, historinë tonë kombëtare nga njëri brez tek tjetri duke i ruajtur dokumentet në formë origjinale ashtu sic janë

krijuar qindra vitesh përpara. Sot ne përballemi me një tjetër realitet në të gjitha fushat e jetës. Ky realitet i quajtur “Teknologji Informacioni”, po ndryshon mënyrën tonë të të punuarit dhe të të jetuarit. Edhe në fushën e arkivave është e njëjta situatë dhe në mënyrë që ne të jemi të sukseshmën në misionin tonë në vitet e ardhshme, ne duhet t’u përgjigjemi një mori pyetjesh që kanë të bëjnë me implementimin dhe përdorimin e teknologjisë së informacionit në procedurat arkivore si dhe ato të ruajtjes së pasurisë arkivore dixhitale; qoftë kjo e lindur në fomë të tillë apo e krijuar si rezultat i proceseve të dixhitalizimit. Këto janë edhe qështjet kryesore të trajtuar në studimin tonë të mëposhtëm, në Shqipëri.

Introduction

Moving the complete official records from Governmental Agencies to the National Archives will be one of the biggest challenges in the coming years, not only in Albania but even in the EU countries. In order to be prepared for those challenges, we must immediately start to think and work for creating the rules and development of our electronic systems based on actual standards and best practices in the world. Our system will be digital repository of our national memory, so we need to plan carefully all steps from creation to archiving of our documents that itself are the main mean to prove the accountability and evidence of governmental activities.

Initial Situation

Albanian National Archives is the only institution in Albania responsible for ensuring and supervising the life cycle of all documents that belong to national memory regardless of their form, state or time where they are created (Law “On Archives” date 06.11.2003) including their find aids.

Until the end of 1980ies, all documents in our country were created only in paper format. After this period the use of electronic equipments (computers) to create and deliver documents is continuously increased. Despite that, the paper is still the most used method due to our actual legal situation. The digital world has not yet replaces the traditional documents (paper documents) but it is taking more and more the guiding role because of flexibility, facilities in the communication, storage capacities.

On February 2008, Albanian Government has approved the law “For digital signature” and later the law “For electronic documents”. In order to keep up this development, the General Directorate of Albania’s Archives (GDAA) is elaborating within the scope of e-government program. This article introduces the conceptual and technical aspects of this solution.

Tasks and challenges of Digital Archiving

The safe creation, circulation, storing, description and delivery of digital records, and also archiving of all other records serve as legal certainty, meaning the traceability and/or verifiability of operational activities.

In order to fulfill those business purposes all archived digital records must comply with following requirements for long-term, or for unlimited period of time:

- Integrity
- Originality
- Authenticity
- Accessibility

Our records must be saved in the unlimited period of time, but the Information Technology has a limited life cycle. Specific documents and data formats require the specific application to be able to interpret and process these formats and specific operating systems and hardware to run these programs. However, the access to the content of digital records must be ensured far beyond the life time of any individual IT component. In order to manage the challenges of digital archiving, the archives firstly have to deal with history and the prospect of the information technology and secondly, of course, to plan the long-term preservation strategy of digital records. There are two main possibilities for this:

- Emulation: keep in life of any hardware and software needed for retention and use of digital records.

- Migration: converting all digital record in new formats and/or migrating to the new media.

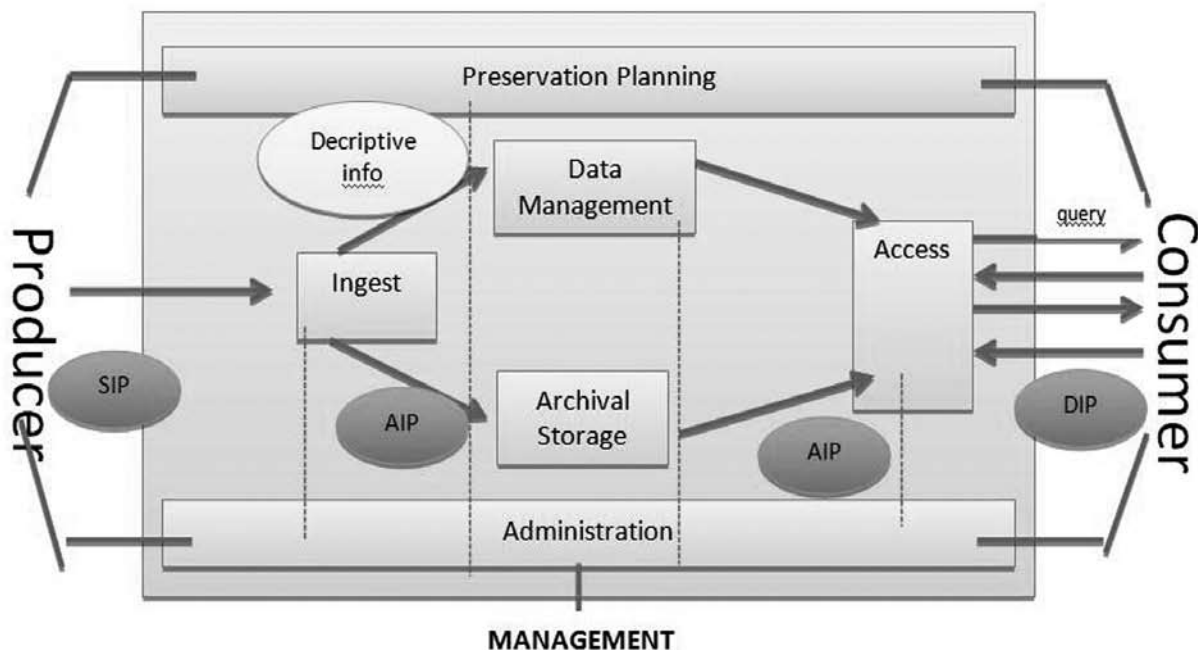
The first solution requires large costs for the maintenance of older programs and hardware and training the new generation on how to use these programs and hardware, but it ensures the preservation of all digital records without losing their properties over the time. Recently there have been many efforts to use the artificial intelligence for the implementations of emulation scenarios.

The second solution, migration strategy, consists the migration of archival data from an obsolete file format to a new one. The migration itself is a central and operational function of archiving solution. The quantity and the extent of migration should be reduced as far as possible and should occur only if authenticity, identity, integrity, accessibility and interoperability of archival data are guaranteed.

We need to take into account some considerations about possible damages from unwanted side effects during the migration process.

OAIS (Open Archival Information System) Reference Model

The digital archiving must be based on the ISO OAIS Open Archival Information System Reference Model (ISO 14721:2003). This reference model describes the archives as an organization where people and system collaborate among them in order to preserve information and make it accessible for a designed community of users. The model describes in detail how information from suppliers (Producers) is received into the archival system (Ingest), which processing steps have to be carried out under the responsibility of the archival management (Management) and how the preserved information can be made accessible to the users (Consumer).



The Governmental Agencies (Producer) submit their records to the Archives in the form of SIP objects (Submission Information Package). These objects are integrated into the archives' system through Ingest functions; they are converted into AIP (Archival Information Package) and forwarded to the archival area. Descriptive and administrative information to those objects are administered and made accessible via the process of data management. With access functions the archived objects and their metadata are made accessible as dissemination objects DIP (Dissemination Information Packages) to users (Consumer) subject of legal requirements. The area called Preservation Planning covers the tasks needed for planning and monitoring the necessary preservation actions. The essentials administrative tasks of an archive are being consolidated in the administration area.

OAIS reference model does not specify any specific organizational and technical implementation. These conceptual basics have to be elaborated in each archiving organization according to its specific needs.

Digital Record Types

There are three main types of digital records that are actually being produced in Albania from the governmental agencies, but yet not centrally organized.

- Digital files that are created from different specific electronic systems part of e-government system.
- Relational databases: data that are generated and maintained within the frame of application in specific scope.
- File collections: files generated and maintained from office applications.

For the moment we are focusing on those three record categories because this will reduce the complexity of digital archiving. Actually the GDAA observes the organizational and technological developments and takes into account the emerging requirements of submitting agencies.

All categories have to be produced in accordance with rules and must be made in packages containing both primary data and metadata. For example, a submission package must contain not only hundreds of dossiers with thousands of documents (primary data) but also the pre-archival classification system, the descriptive information of each unit of description, attributes of each object as well as information of submission agency.

The limitation of file formats also will help us to reduce the complexity of digital archiving. For the moment we recommend the following file formats to our producer (governmental agencies):

- PDF/A or Text for office documents
- TIFF for images
- WAVE for audio files
- MPEG 2 for video

Metadata and technical issues of Digital Archiving

The term 'metadata' in archiving was coined by Information Technology. The preservation actions cannot be performed without the presence of this very important component. The metadata describes the information about content and context of the records. There are so many metadata standards in the field of the archives such as: PREMIS, METS, Dublin Core and EAD, while in the pre-archival area one important role plays the standard ISO 15489.

A submission package of digital record consists of primary data in a suitable format of archiving, as well as in metadata belonging them including the administrative, descriptive, structural and technical metadata that guarantee the research, retrievability and help us to plan the preservation actions of records. This may include:

- Data on documents (reference number, title, author etc.) and their classification and their origin (records managing agency, system in use etc.). This information together with the digital records to be archived has to be submitted from governmental agencies at the Central State Archive.
- Information about the file properties (file name, format, size, checksum and other attributes).
- Files that document the management of archival records that can be simple notes added by archivist or other responsible persons.

At the moment we are closely collaborating with National Agency of Information Technology to create our standards in metadata, not only the archival metadata but in the same time the metadata standards for Electronic Record and Document Management Systems (ERDMS).

Based in the above issues the following technical requirements are necessary for digital archiving:

- Treatment of archival data as packages with multiply files and their corresponding metadata
- Creation of our system compliant with OAIS and in accordance with our legal framework
- Plan the strategy for long- term preservation of digital records

The Central State Archive of Albania has about 10 years that uses scopeArchiv application in order to facilitate the archival processes. The new version of this application offers the functionalities of digital archiving and is developed based in the international standards and of course is compliant with OAIS reference model.

In order to control and check the quality of SIP packages we are planning to create an extended function called package handler. The package handler will provide access to employees of governmental agencies in order to be able to compile independently their own submission according to archival requirements.

The functionality of e-archive

With the term functionality of e-archive we mean:

1. It is responsibility of the governmental agencies to prepare submission packages in a suitable manner for archiving. The governmental agencies extract the closed files that are ready to archive including their metadata from their ERDMS. After this, they process the packages (SIP) according to their classification schema and based in the retention period as it is specified in the Albanian Archival System.
2. Each governmental agency transfers the packages (SIP) to the Central State Archive (CSA) based on the confidentiality of in the total size of packages.
3. The CSA registers in arrival all SIP packages and after this starts with checking process of legal and archival properties. In the same time there is some very important verification like virus scan, file format check, verification of checksum etc.
4. If any of packages does not meet the required specification then a notification is send to the submitting agency.
5. After completion of the ingest phase, the system will generate the AIP according the OAIS reference model. The system must save those packages in two different geographical locations in order to be sure that the data will be safe in case of any natural catastrophe.
6. The information and metadata necessary for the description in the Archival Information System will be extracted from the packages as it is described in the "Technical and Professional Norms of Archival Service in the Republic of Albania".
7. Users can use intranet or internet for query in the find aids that are provided by AIS as DIP packages.

Conclusions

The solution presented in this paper is a summary of our work delivered in a couple of years between the computer scientists and archivists. At the moment we are working with implementing ERDMS in creating agencies and testing the delivery of the data to the final destination, that mean Central State Archive. There are three institutions that are selecting in the first phase: Council of Ministries, Ministry of Integration, and Ministry of Transport and Telecommunication. The new system will be interoperable with actual archival information system.

The web based-access to finding aids and digital records will open new possibilities for different users of archival resources like students, researchers or citizens. The right of access for official documents is guaranteed with restrictions to protect the personal and confidential information.

Bibliography

Law "On Archives". www.albarchive.gov.al. (Accessed on 23/06/2012).

ISO: *Open Archival Information System-Reference Model (ISO 14721:2003)*, International Standards Organization 2003.

ISO: *Document management-Electronic Document File Format for Long-term Preservation-Part 1: Use of PDF 1.4 (PDF/A) (ISO/CD 19005-1)*, International Standards Organization 2004.

S. STRODL – A. RAUBER, *Preservation planning in the OAIS model*, "New Technol. Libr. Inf. Serv", **1(2008)**, pp. 61–68.

J. VAN DER HOEVEN – H. VAN WIJNGAARDEN, *Modular emulation as a long-term preservation strategy for digital objects*, In *5th International Web Archiving Workshop (IWA05)*, Vienna 2005.

SUMMARY

Nowadays the archives have solved the problems related with traditional archiving such as classification, preservation, storing, accessing of documents. They have successfully store our national heritage, our national history from one generation to another saving the documents in their original form as they were created thousands years ago. Today we are facing a new reality in all the fields. This reality named Information Technology is changing our ways of thinking and working, and our ways of living. We have the same situation in archival field and in order to be successful in our mission during the present time and the future, we should answer a lot of questions that are related with implementing and using information technology in archival procedures and in preserving of either digitized or 'born digital' content. Some of these questions are:

How to create, use, store, preserve and access the digital records in order to save their authenticity, integrity and of course in accordance with archival principals?

Which are the main international standards and how to implement these standards?

Which are the best practices over the world in the field of e-archiving?

Which are the main strategies for long-term preservation that we need to plan for the future?

All these questions and their analysis will be part of our work focusing the issues regarding the long-term viability and the access to digital collections as well as finding solutions to e-archiving by adapting most important international standards such as OAIS Model, Metadata standards (PREMIS, METS), MoReq 2010, ISO 15489 standard etc, to the Electronic Archival Information systems.

Original scientific article

Submitting date: 25.06.2012

Acceptance date: 30.06.2012