Managing and Presenting Digital Content in the ARHiNET System

Vlatka Lemić
Hrvatski državni arhiv
Marulićev trg 21, 10 000 Zagreb
vlemic@arhiv.hr

Hrvoje Čabrajić
Avicena Software d.o.o.
Put Supavla 30, 21 00 Split
hrvoje.cabrajic@st.t-com.hr

Summary

ARHiNET is a network information system for describing, processing and managing archival records created in 2006 by the Croatian State Archives and Avicena Software Company. It is a national archival system in Croatia, recognized by the Ministry of Culture as national project, as well as part of the e-Croatia program, the operational plan of the Government of the Republic of Croatia. Development of the archival information and institutions network is a long-term strategic archival service project and ARHiNET implementation enhanced the standardization of the archival institutions work, and enabled establishment of a unique system of processing and description of archival material, as well as data integration and exchange between the institutions that keep archival records. All archives in Croatia are included in the implementation of this unique archival information system that comprises all business processes in archival institutions, together with some other records holders under the state archives supervision. Currently, there are about 700 registered users from more than 150 institutions. Designing, realization, introduction, use, maintenance and development of such a complex program solution enclose permanent activities on system improvement, finding new functionalities and solutions, as well as upgrading of the present ones. During the three years of the system operating, more than 300 versions of program solutions have been developed and put in production, and experiences gained from work and user education led to the development of the version 2.0 that was released in February 2009. In this article authors present solutions and functionalities concerning managing, indexing and presentation of digital content developed and implemented within the ARHiNET program solution.

Key words: archival information system, archives and Internet, digital content, digitization of archives, digital records management

ARHiNET system

ARHiNET is a network information system for describing, processing and managing archival records created in 2006 by the Croatian State Archives (CSA) and Avicena Software Company. It is implemented as the national archival system in Croatia, with specific objectives:

- establishment of the unique system based on international standards,
- providing efficient and user oriented system of collecting, processing and presenting archival material,
- inclusion of all important elements of archival records management and management of business processes in the archival institutions into one comprehensive system,
- facilitating work of archival professionals,
- standardizing and assuring quality of services and products provided by archives,
- ensuring preservation and presentation of data by using information-communication technologies,
- introducing new technologies and technological solutions in the archival institutions.

ARHiNET is created on modular basis which enables design and implementation of particular modules as separate projects in a relatively short period of time and their continuous integration into the unique information system. Advantages of such a solution are the creation of an integrated base and a unique system of data protection with minimal costs.

ARHiNET system structure comprises of two parts: the open one is intended for external users who want to search databases and catalogues and use other offered services, and the protected part, intended for the employees in archives and other institutions, in which all professional-business processes that define processing and management of archival material are taking place. The program solution consists of several databases organized according to the logic of records type and user type/roles that define access to particular records:

ARHiNET Public – database containing records for access by external users. This database is read-only, that is, records are not added or changed, but only retrieved. Database is optimized for faster searching and records retrieval. Records in ARHiNET Public are daily automatically replicated from ARHiNET Master database according to the authorization criteria, i.e. only records available for searching are imported. Main purpose of defining ARHiNET Public database is database search optimization and acceleration, protection of access to the Master database and possibility of simultaneous work on the records so the units of data can be available to public while being edited.

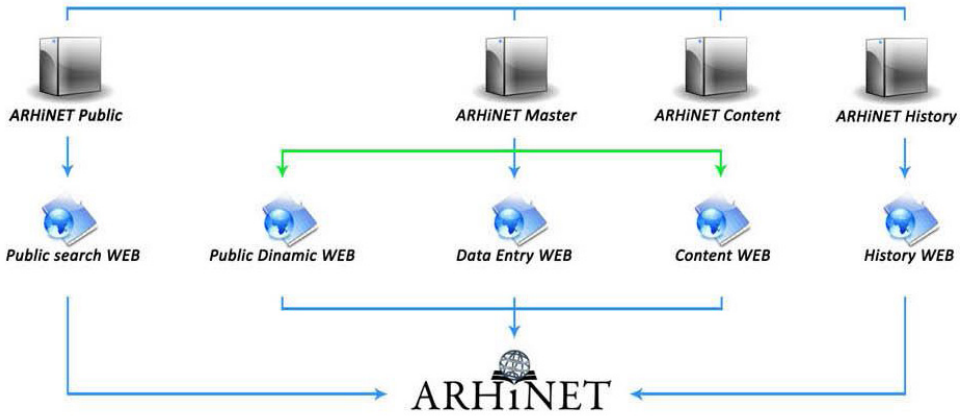
ARHiNET Master – database containing records that originate from descriptions of archival and current records, their creators and holders, special types of

archival material, archival documentation, lists and registers and all other data related to business processes management in archives (holding, processing, preservation and use).

ARHiNET Content – database containing records that are not directly related to archival records, but are used for administration and facilitating work within the ARHiNET system. It contains elements that can be found on external web pages, professional topics, archival forum and advices, help files, etc. The idea behind it is to allow users dynamic management of all modules and functionalities without programmer's additional help. It also enables system localization (translating user interface) and data import and export.

ARHiNET History – database containing all changes made in the units of data (editing, updating, deleting etc). Records life cycle management is defined with MoReq specification and implemented in ARHiNET with purpose of tracking the changes of every single unit recorded within databases, which is very important system functionality. Administrator has exclusive access to this part of the system, and every change of data is recorded in this database: who changed the record, when was the record changed and what data content has been changed.

Scheme 1: ARHiNET Structure



Digital records inside the ARHiNET system

An important segment of the new information system inside the Internet environment was to provide accessibility of digital content and professional description of digital material. Administration, processing and presentation of digital content are defined into the separate module inside the ARHiNET, and its design was accompanied with adoption of firm rules of managing digital content in the archival institutions, both based on detailed analysis of current policies and practices.

Possibilities of new technologies extended and improved ways of protection and access to archival material, and because digitization of archival records considerably facilitates their availability, archives are faced with mass production of digital records. Besides large quantities of documents and technical problems, issues concerning digitization that archives are faced with mostly refer on selecting and preparing records for digitization, their organization and presentation, as well as availability. Such condition is a reflection of complex nature of archival material and differences in provenances, arrangement and types of records kept in archives.

The ARHiNET enables description of archival material of any type and content (textual, graphic, cartographic, audiovisual, electronic, objects, photographs etc.) according to the international standards for the archival description as well as other relevant specifications. Every record is described with set of general data elements, special data depending on type of records and related tables of additional data which are available in the form of special lists. They are defined in several basic, mutually linked data sets:

- fonds and collections,
- records creators
- records holders

that all together provide data integration and saving the content and context of all records.

Since digitalization for majority of Croatian record creators and holders imply mass digitization, while digital preservation is currently in the professional background, a first step inside the ARHiNET was to provide support for digitizing archival material, and second to implement procedures for preservation and accessibility of “born digital” records. Concerning considerable efforts and resources being invested in digitization, basic principle of ARHiNET program solution – integration of data and reducing of costs – was useful in relation to problems of facilitating large-scale digitization and its cost-effectiveness.

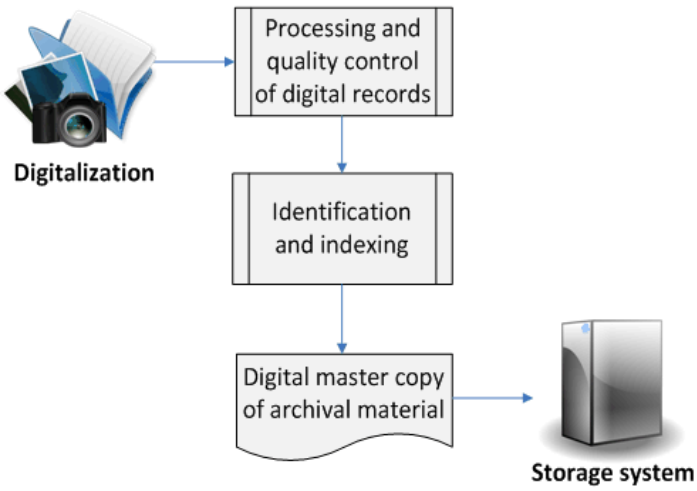
Several business processes were encompassed by digitization procedure in the ARHiNET framework:

- digitization of archival records,
- processing digitized records,
- saving master copies in the storage system,
- automatic creation of web copies in JPEG (or other) format,
- linking of archival units and digital records, and
- presentation of digital content within program solution.

With such a defined and adopted concept, the intact quality and protection of the master copy of digitized records is ensured. External users are granted an access and review of digitized content through web copies, with the possibility of their downloading and printing, while availability of master copies should be granted by the institutions which hold those records.

Basic operational unit for working with digital records inside the ARHiNET system is the digital master copy. Since activities of describing, indexing and managing user copies are depending on quality of master copies system of creation, indexing, and storage of master copies is defined by statutory procedure.

Scheme 2: Procedure of making digital master copies

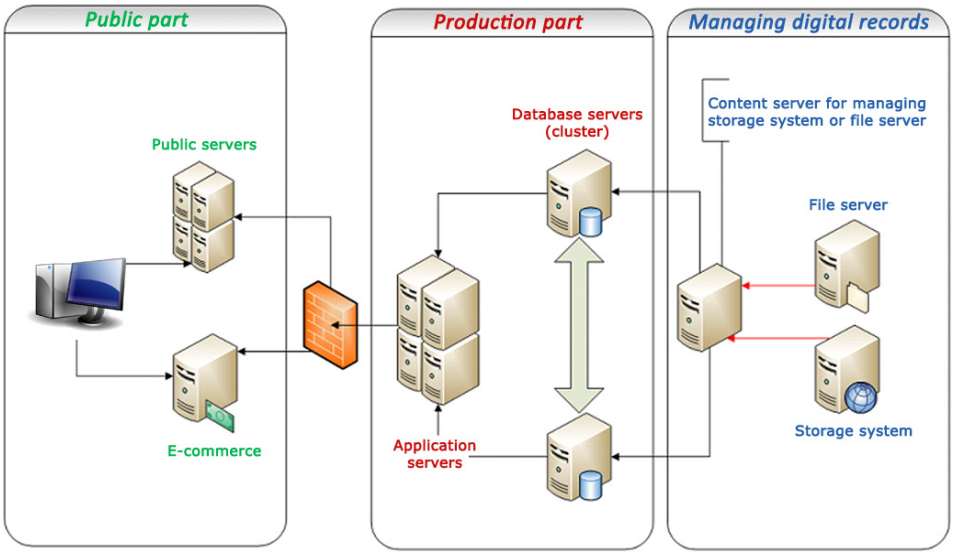


Term master copy, i.e. original digital reproduction of archival unit/record, in the framework of working inside the ARHiNET system is applied for digital copy of a single archival unit which is completely analogous with original record. Technical characteristics of master copy represent optimum of resolution and quality of digital record, depending of type and material of the original record.
Master copy represents digital material from which all user and other types of copies are made. Master copy is kept inside the storage system marked with a unique identifier, and cannot be subsequently altered after the processing, description, controlling and authorization have been finished. Access to master copies is allowed only to operators who are authorised for periodical quality control, migration and making copies.
Every archival and other institution which actively participates in national archival information system must adopt one's own set of rules for all these activities in correspondence to CSA standard procedures for ARHiNET operational work.
Managing digital content is functionally integrated on various levels with other modules inside the ARHiNET system:

- browsing and searching data,
- browsing and searching digital content,
- managing digital content on the storage system,

- integration and administration of ARHiNET databases and modules,
- managing archival material,
- managing records holders.

Scheme 3: ARHiNET system support



Protection of digital records

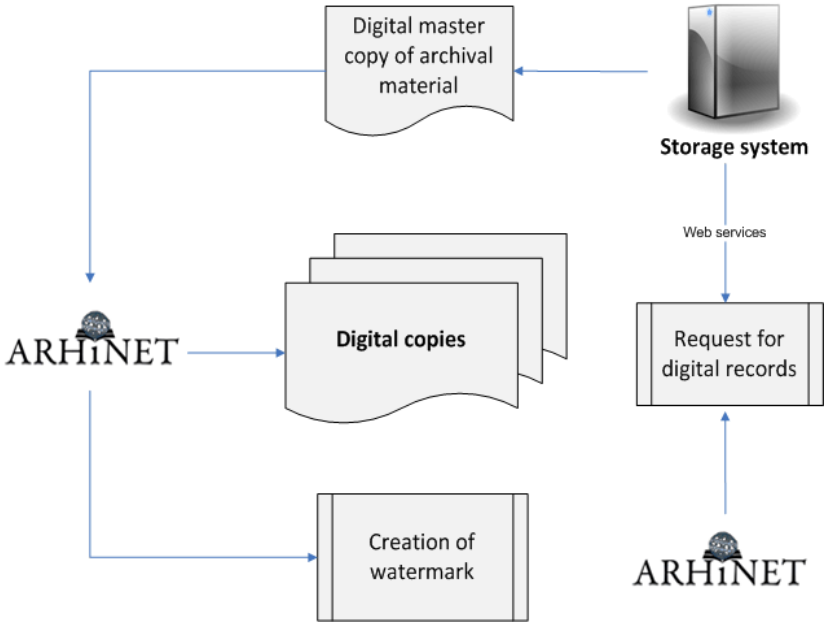
Protection of digital copies of archival material is a significant segment of ARHiNET system because using of digital records cannot be implemented without providing a system of supervision and protection of digital copies from unauthorised copying, multiplication and distribution.

By analyzing present solutions the ARHiNET project team decided to develop its own method for protection and management of digital copies in order to create unique and efficient system of controlling distribution of digital copies and to establish a mechanism of copyright protection for the material, of which originals are held in archives or other institutions/record holders. Functionalities of such a system are:

- providing implementation of visible watermark sign on every single record,
- providing implementation of invisible unique identifier for every single record,
- providing a system of control and tracking of eventual frauds of digital copies,
- recording all changes into database.

After the user has searched the ARHiNET data bases on archival material and has ordered digital copies ARHiNET automatically starts the procedure of digital master copy retrieval from storage system by using web service. After the receipt of virtual master copy, ARHiNET will automatically create digital copy with watermark consisted of name of the record holder and date of the creation.

Scheme 4: Procedure of creation of the watermark



After the creation of watermark, the next sequential part of automatic procession will add a unique identification mark on each copy of the digital record. Each digital copy is defined by pixel scale and each pixel is uniquely defined by its place and colour, while colours are defined by custom palette for colours. Having in mind these settings, it is understandable that in the cases of changes of a single pixel in relation to the master copy ($\pm$ 1) digital copy will be uniquely changed comparing to original. Regarding possible number of combinations (number of pixels × changing colour shades × possibilities of simultaneous changes of one or more pixels) it is done on unlimited number of combinations which allows unique identification and indexing of every single copy, without affected their quality.

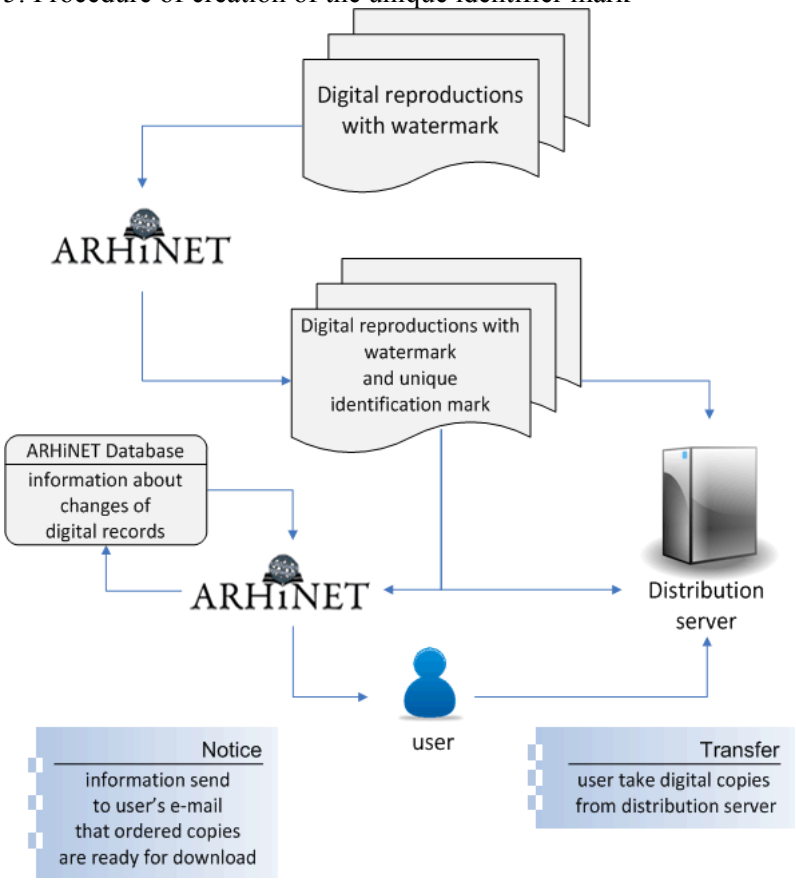
Adding a unique identification mark on each digital copy is recorded inside data base with folowing parametres:

- master copy ID,
- location of altered pixel,

- colour, and
- date of change.

Those data are linked with data about users and order forms of digital copies which all together represent base for documenting use of archival material and tracking changes.

Scheme 5: Procedure of creation of the unique identifier mark

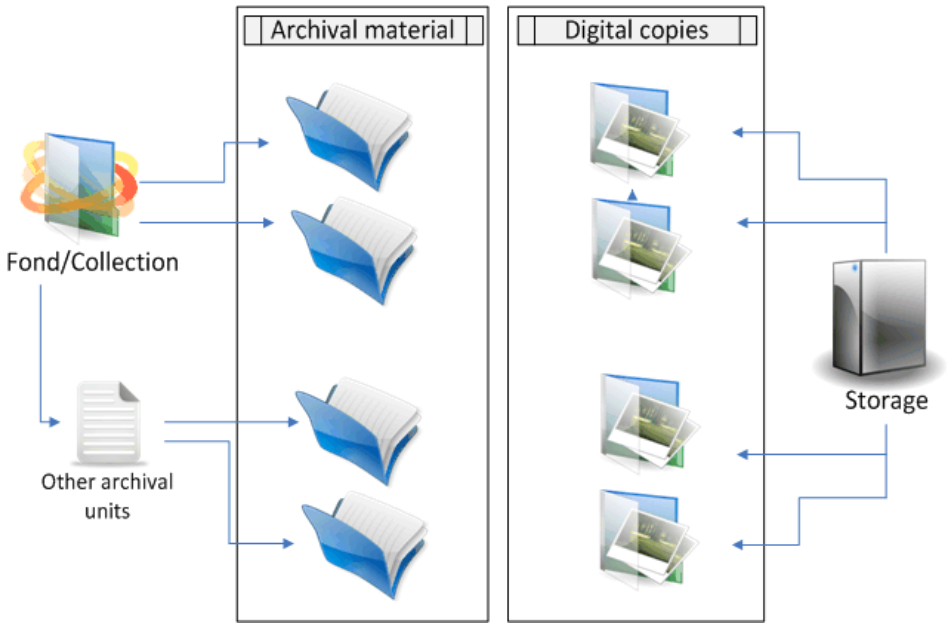


Providing context for digital material

Data on archival material are inevitably changed in time because of accessions and transfers of new and old records (new fonds and collections are created or existing are supplemented), processing of records and other procedures in the process of archival arrangement. Descriptive elements, following general standards and rules for description of archival material, elaborated and used within ARHiNET are defined not just for archival units, but also for digitized records. Data and metadata made by processing digitized records are connected with data of archival unit's descriptions which enables accuracy in efficiency, as well

as, facilitate every day work. Once those relations are made, every change of data in unit of description will automatically be reflected on digitized records so there is no need for multiple editing of same change. This is realized by providing inside the system a list connecting ID of archival units in ARHiNET with ID of master copies stored on the storage system.

Scheme 6: Connection of archival unit's descriptions with digital copies



This connection requires implementation of content management system for managing storage system, and maintenance of digital master copies data base comprising of two basic tables. One lists elements important for managing digital copies:

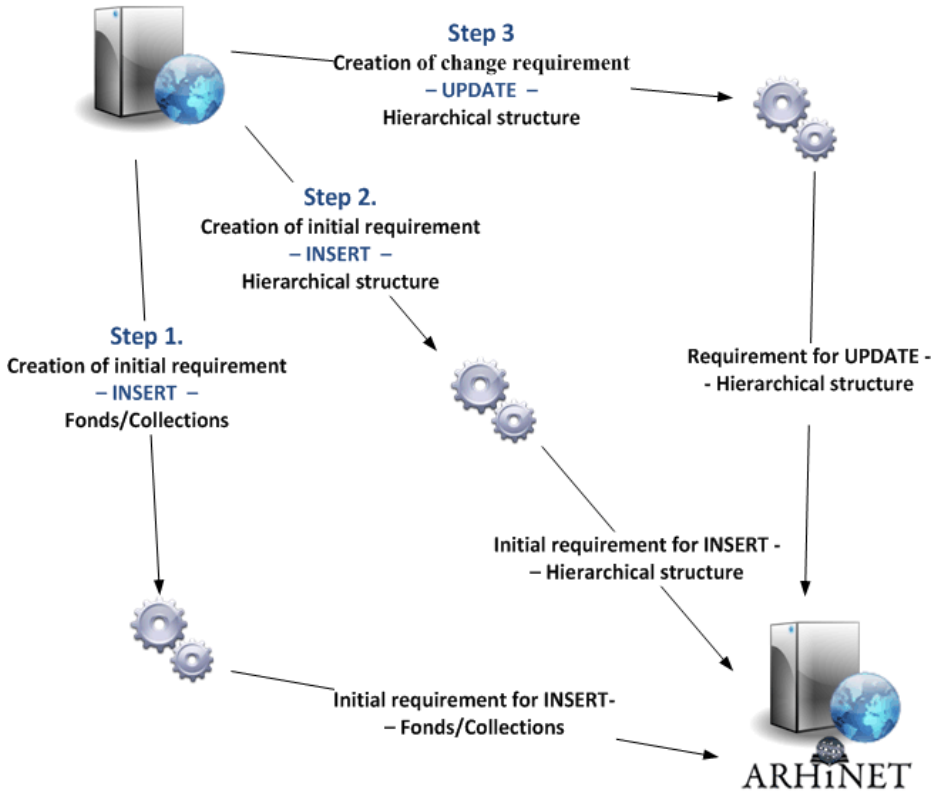
- master copy ID,
- path – position where master copy is placed on the storage system
- order – enables defining sequence of presentation of pictures inside archival unit
- accessibility – information on restriction of access and use ,
- ID of related structure,

while other administrates elements important for defining structure:

- holder and fond/collection structure,
- hierarchical structure of unit of description – level of description
- export/import of data in XML format.

Operational work of all mentioned elements insures ARHiNET and storage system communication through web service.

Scheme 7: XML scheme for connection of archival unit's descriptions with digital copies



Digital preservation – future development

Production, dissemination and filing of documents in electronic form present some of the biggest problems for modern archives. Although ARHiNET system already supports description and integration of all type of archival material, current development is directed toward upgrading present functionalities with options of online access to digital documents and their search and retrieval. This will be realized through ARHiNET Central Data Poll Model (CDP) which is designed for digital preservation and access to digital data such as databases and multimedia records. It defines XML structure, datasets and files list which enables preservation of structure, content and context of digital record and their management and use in one unique system. Implementation of such system will provide integration of traditional and digital archives, as well as, bring archives closer to their major goal - to ensure authentic, reliable and preservable records, regardless of the form and physical medium they have been created and preserved on.