

**MINISTRY OF DEVELOPMENT OF INFORMATION  
TECHNOLOGIES AND COMMUNICATIONS**

**TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES**

*Manuscript Copyright*

*UDC:002.55:004.4*

Mardonova Dilnoza Ismoil qizi

**Methods of Information exchange between archives**

Specialty 5A350602 – Electronic library and archive

A DISSERTATION

Submitted for the Master's degree

Supervisor: Dr. A.Kh.Abdullayev

Tashkent 2016

## **Introduction**

With the development of information activities of the human wealth of information created, processed and stored only grows. There are estimates that currently only about 30% of all corporate information is stored in electronic form, structured (in databases) and unstructured.

The share of the information in paper form currently accounts for about 70%. Storage of information in paper form creates difficulties when searching and processing information, and require physical space for storage, and does not allow management personnel to accelerate the process.

Every year in the authorities and organizations are becoming more and more urgent the establishment, organization and functioning of archives of electronic documents. It is expected that in the coming years the archives of electronic documents will become a major source of information to make critical decisions at all levels of government. In this regard, the problem of storage of electronic information resources and ways to overcome these problems become crucial in the methodology of organizing such archives.

The introduction of electronic document management system in the archival institutions enables companies to simplify the problems related to search, access and storage of documents, and therefore avoid many of the problems encountered in the process of document. This statement applies not only to companies producing any goods, but also companies that provide services, including the state archival institutions such as the Agency "Uzarhiv"

**Research actuality.** Nowadays archive organizations is providing of information, special materials in a diferrent field to various strata of society, and it is a call of the times. Because providing of general and personal resource in time to organizations and citizens may cause a better efficiency to their work. Using of ICT to enhancing archive work give a efficiency. For example, electronic archives (EA) gives a fastest searching of electronic papers in circulation of electronic documents or supporting a circulation of electronic

documents. Analyzing of leading countries experience in field of circulation of electronic documents, state registration, state preservation is a main field of research of archive organizations of Republic of Uzbekistan.

**Research tasks.** - Studying and analyzing of foreign archive organizations;

- Comparing studied experience of foreign archive organizations with our experience;
- Choosing optimal of compared results and creating deployment algorithm to our archive;
- Studying current status of archive organizations and implementation of modern technologies;
- Analyzing of problems which cause in information circulating between archives and problem-solving.

**Object of research.** Purpose of research work: increase of effectiveness of archive organizations, creating improvements in circulation of documents, increase of effectiveness of service of archive organizations of Uzbekistan.

**Subject of research.** Cinema archives, photo archives, Central photo archive Republic of Uzbekistan, museums of Republic of Uzbekistan, Shahrizabz state archive. Implementation of faster searching support, electronic catalogue creating technology for users facility, programs for purpose of automation in archive organizations and creating methodology of information circulating in archive organizations is a subject of research.

**Methods of research.** In this research such methods as the bases of system and comparative analysis and catalogization in MARC21 format were used.

**Scientific degree of novelty of research results:** Analyzing of information circulation technologies and organization information circulation technologies of developed countries of the world; implementing of new types of rendering of service to users also creating electronic catalogue for purpose of supporting of information circulation in field of archive organization and methods of enhancing of corporate archive networks is shown.

**Practical value.** research proposals and sugGESTions is useful for implementation in archive development field of Republic of Uzbekistan, and can be used in information circulation in archive organization of Republic of Uzbekistan.

**Structure and volume of dissertation.** Dissertation is contain, three chapters, conclusion, literature and applications. Consists \_\_\_ pages.

# **CHAPTER 1. RESEARCH SUBJECT DOMAIN AND GENERAL PRINCIPLES OF SOFTWARE COMPLEX ELECTRONIC DOCUMENT EXCHANGE ARCHIVAL INSTITUTIONS**

## **1.1. Technology of exchange archival documents of the developed countries and their analysis**

In order to preserve human cultural heritage and to keep continuity of access to content that has already disappeared from the web, or soon will disappear, many libraries and archives have started to archive web resources. Much has been reported about how libraries and archives select, acquire, store, organize and describe web resources for preservation. As libraries and archives become more settled with these collection-building practices, they need to pay more attention to the users, usability, and functionality of web archives, since the ultimate purpose of archiving web resources is to use them.

There have been a limited number of studies on the uses and functionalities of web archives. The International Internet Preservation Consortium (IIPC) Access Group has defined a variety of use cases that web archives are expected to support (IIPC Access Working Group, 2006). Each of these use cases requires a number of functionalities. Many of these functionalities, such as "search by URL", are basic to search and retrieval in web archives. A number of functionalities, such as "data mining," require more advanced means of information discovery and analysis. The IIPC use cases can be used as best practice guidelines for the functionalities of web archives.

After the publication of the IIPC use cases, two user studies identified user expected or preferred functionalities. Ras & Bussel (2007) studied user needs and identified potential types of users of web archives of the National Library of the Netherlands (KB). Based on their user study, Ras & Bussel identified user-preferred search interface features of web archives, which include both usability and functionality features.

The most recent user study on web archives was conducted by Costa & Silva (2010) on the Portuguese Web Archive (PWA). They studied the needs

and behaviors of people searching the PWA. They found that searching for a known page or site was more frequent than collecting information about a subject written about in the past, which was more frequent than performing web-based transactions. Another finding is that users prefer the oldest documents over the newest. This study also identified user preferred or expected functionalities. For example, although users prefer full-text search over URL searches, URL queries are common and should be supported. Specialized search engines for imaGES, videos and old news, and the ability to demonstrate the evolution of a page or site, are also expected.

To expand upon the findings of a comprehensive literature review of web archiving methods, the author conducted an overview of the functionalities supported by current web archives, by creating a checklist of functionalities based on the IIPC use cases and the two user studies mentioned above. Although the sample size is small due to language barriers, and the access restrictions of some web archives, the findings did reveal the current functionalities provided by some of the most established web archives in the English speaking world (or for which there is an English language interface). The findings also highlight the gap between the current functionalities, and the expected functionalities identified by the IIPC use cases and recent user studies. The checklist and current findings may help to inform the functionality design of future web archives and assist future archivists. The checklist can also be used as an evaluation or self-assessment tool for existing web archives.

## **Methodology**

### ***Web archives evaluated***

This project evaluated the functionalities of ten member archives of IIPC. IIPC has a web archive registry, [netpreserve.org](http://netpreserve.org) (International Internet Preservation Consortium), that includes 24 web archives. The archive of one of the 24, [Bibliotheca Alexandrina](http://Bibliotheca.Alexandrina), is a mirror site of the Internet Archive (IA), and provides identical functionalities to the IA Wayback Machine. Since IA is also on the registry, Bibliotheca Alexandrina's web archive was excluded from the

study. Eight of the archives are either dark archives or allow only onsite access, making them inaccessible.

Fourteen web archives are publicly accessible online; however, only nine of them, listed below, have an English interface:

- IA Wayback Machine (started in 1996)
- Preserving and Accessing Networked Documentary Resources of Australia (PANDORA) Web Archive created by the National Library of Australia (started in 1996)
- UK Government Web Archive created by the UK National Archives (started in 1997)
- New Zealand Web Archive of the National Library of New Zealand (started in 1999)
- Library of Congress Web Archives (started in 2000)
- Web Archiving Service (WAS) of the California Digital Library (started in 2003)
- Government of Canada Web Archive created by Library and Archives Canada (started in 2005)
- Web Archive Collection Service (WAX) created by Harvard University Library (launched in 2009, piloted in 2006)
- UK Web Archive provided by the British Library in partnership with the National Library of Wales, Joint Information Systems Committee and the Wellcome Library (started in 2005)

The Archive-It service (started in 2005) is provided by IA and is not listed as a separate member on the IIPC registry. However, Archive-It and the IA Wayback Machine provide very different functionalities. They were included as two separate web archives in this study, bringing the total to ten.

These ten web archives vary in their collection scope, size and age. At the time of the study, IA Wayback Machine was 15 years old, but WAX was only two years old. The collection scope of IA Wayback Machine is the most comprehensive, followed by the web archives of several national libraries. The

UK Government Web Archive and Government of Canada Web Archive preserve only government websites. WAX, due to its very young age and the fact that it is provided by a university library, only had four collections at the time of this study.

Unlike the other eight web archives, WAS and Archive-It are service providers that provide technical infrastructure, data storage and training for other organizations. Subscribers of WAS and Archive-It can focus on resource management issues such as organization, description and access control, and do not need to worry about the technical concerns of web archiving. These two service providers also represent the web archiving practices of a large number of organizations in the United States. As of April 2011, WAS had 16 subscribers and Archive-It had over 160 subscribers, including national libraries, state libraries and archives, academic libraries and government agencies. Many subscribers provide a link on their websites to Archive-It and do not have their own web archive interface, such as the Arizona State Library, Archives and Public Records and the Michigan Government Web Collection.

Although ten web archives sounds like a small sample size, their status as member archives of IIPC shows that they are among the most established web archives in the world. In addition, since many subscribers rely on WAS and Archive-It to provide an access interface, the functionalities of these two web archives also represent the functionalities of the web archive collections created by many other organizations.

### ***Creation of the functionality checklist***

The IIPC use cases were divided into five categories. The "Use Cases for Archive Internal Use" in the last category were excluded from this study because they were created for people who build the web archives. Content analysis was conducted on the first four categories of the use cases. Functionalities required to support each use case were identified. For example, one use case states:

"Jane Jones learns that a competitor of her company copied the appearance, including trademarks, of her business website, and engaged in a mass mailing to customers directing them to the spoofed website, where some were tricked into supplying proprietary information. By the time legal action is contemplated, the competitor has removed the offending material from the website, but copies exist in the ArcSys. Jane Jones keys in the URL and relevant dates of the offending website, and wishes to receive evidentiary-quality printouts of the relevant paGES, which may require signed declarations of ArcSys personnel."(IIPC Access Working Group, 2006, p. 8)

This use case requires the following functionalities: search by URL, record the dates on which the URL was harvested, allow people to print out authentic copies and certify the authenticity of a printed copy.

Some functionalities, such as search by URL, were identified from more than one use case. The duplicates were removed and all the remaining functionalities identified from the use cases were gathered together, and combined with the user-preferred or expected functionalities extracted from the two users studies (Ras & Bussel, 2007; Costa & Silva, 2010). There was also overlap in the functionalities identified from the two user studies and those from the use cases. Again, duplicates were discarded. In the end, all the remaining functionalities were re-organized into groups: search parameters, search results, browsing, policy-related functionalities, personalized services, data mining, and reconstruction of lost websites.

### ***Method of investigation***

The ten archives were evaluated using the checklist in April 2011. The author mined the interfaces of targeted web archives thoroughly to find out whether each functionality was supported by each web archive. This included reading information about the archives (Frequently Asked Questions, technical information, statistics reports, etc.), conducting searches, observing the ranking and presentation of search results and the way un-archived web paGES were dealt with, trying to print the web paGES, and testing whether the forms and

search boxes work in some web archives. Based on this exploration, the author judged whether each functionality was supported, and then recorded the findings in a table. The results were either "Yes," "No," or "No information found", occasionally supplemented by notes about how the functionality was supported. The author also recorded some functionalities provided by the web archives but not covered by the checklist, and some evident usability issues discovered along the way. One item on the checklist, "Archived websites are the same as the original version of this site" was not used in the evaluation because the original versions of many archived websites are no longer available on the Internet.

*Search by URL and Keyword:* URL is the mostly widely supported search parameter followed by keyword. Six web archives support search by exact URL. The other four support search by URL as a keyword. Searching by exact URL and searching using a URL as a keyword sometimes return different search results. Searching the URL "www.umich.edu" at IA Wayback Machine will only return web paGES at that URL, whereas searching using that URL as a keyword at Archive-It will return results that contain the URL anywhere in the archived web content. The Library of Congress Web Archives and the New Zealand Web Archive support keyword search on bibliographic records. The other web archives support keyword search on the full text of web paGES.

*Domain-based search:* Five web archives support restricting a search by domain, offering users the option of searching content within a single website. PANDORA allows a search to be limited to a given Top Level Domain, e.g., .gov or .edu.

*Narrow the search by date:* Six web archives support narrowing searches by date, and the granularity of the dates varies. The Government of Canada Web Archive, IA Wayback Machine and UK Web Archive allow the user to limit the search to a time period defined by days (YY/MM/DD - YY/MM/DD). Archive-It can limit the search to a time period defined by months (YY/MM - YY/MM). PANDORA allows its users to limit the search to a time period defined by years (YY-YY). The Library of Congress allows the user to restrict the search to a

particular year. Theoretically, larger web archives need more detailed granularity to avoid very long lists of search results. It is not known how these web archives decide the granularity nor whether the size of the web archives plays a role in their decision.

For example, IA Wayback Machine uses "text" as the media type, while Government of Canada lists specific types of text: 'DOC' and 'HTML'. The UK Web Archive lists 'MS Word', 'MS Excel' and 'PPT', whereas WAS uses 'Office Documents', which covers all three types in the UK Web Archive. In most web archives, the search by media type feature is provided on the search interface as a drop-down box and serves as a pre-search filter, as shown in Figure 1 below.



*Figure 1: Media type as pre-search filter.*

The UK Web Archive presents this feature as icons on the search results page and uses them as a post-search filter. When a user clicks on an icon, the search results are filtered by the media type represented by that icon, as shown in Figure 2.



Figure 2: Media type as post-search filter.

*Integrated search:* All web archives that are affiliated with a traditional library are accessible through the library catalog (the New Zealand Web Archive, PANDORA, WAX and the Library of Congress Web Archives). Archive-It provides an integrated search interface for the web archive collections of all its subscribers. WAS provides separate search interfaces for each collection but does not support an integrated search on the whole web archive. In other words, the search interface shown in Figure 1 is provided for each collection in WAS, and there is no such interface for the whole web archive of WAS. IA's Wayback Machine archive is accessible through the integrated search portal for all the IA collections. The UK Government Web Archive is integrated with the live government websites. When a requested URL is no longer available from the UK government live website, the request is automatically redirected to the UK Government Web Archive.

*Unsupported search parameters:* None of the web archives support searching for identical copies using MD5 fingerprint, or searching based on genre and update frequency. Safe search to filter out adult content was not found in any of the ten web archives. The scope and highly selective nature of some of these web archives may make such a safe search feature unnecessary. For example, the Government of Canada Web Archive and UK Government Web Archive only archive government websites; it is thus highly unlikely that those archives include adult content. Web archives that do bulk harvesting, such as the

IA Wayback Machine, are likely to include adult content. The author was able to find adult websites using the IA Wayback Machine. For example, according to the search results page of IA, the porn website <http://www.pornhub.com/> has been crawled 429 times since February 1, 2001. There are no access restrictions on these archived versions.

*Usability issues on the search interface:* Although not intended, some usability issues affecting functionality were identified.

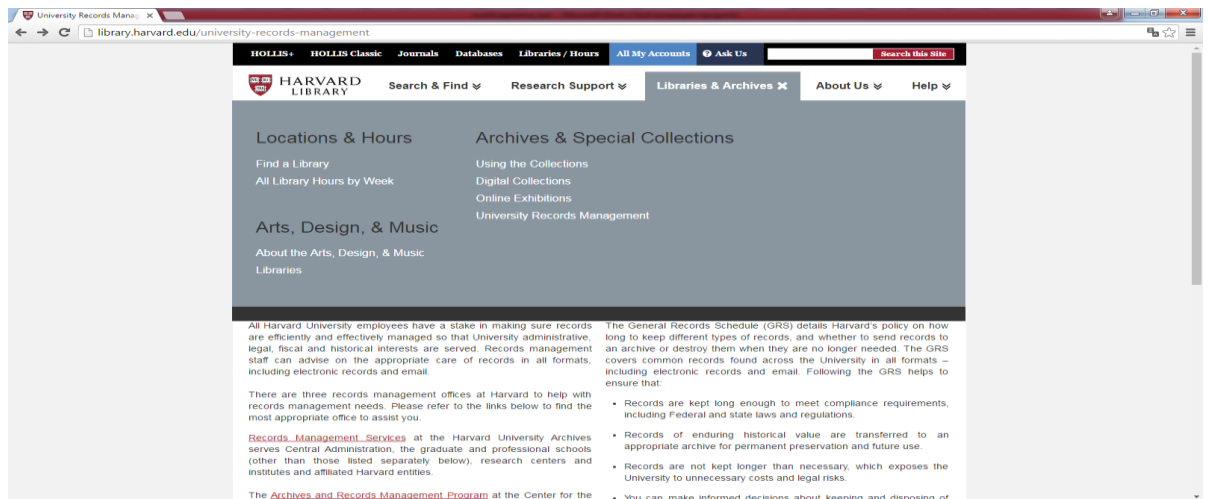


Figure 3: Interface of Harvard Library

WAX provides a very simple Google-like search box, and there is no advanced search option on the search interface. The "help" link, which provides instructions for conducting advanced searches, is at the bottom of the page, far away from the search box itself, as shown in Figure 3.1.



Figure 3.1: The search interface of WAX.

The instructions on the help page of WAX require users to use a certain syntax to conduct advanced searches. For example, to limit a search to files of a particular PDF format, the user needs to type in the search box: "type: application/pdf". In this example, users also need to memorize and type in the MIME type instead of choosing one from a list. At a time when search interfaces are becoming more user-centered and easier to use, this interface looks primitive. The very short history of this web archive may explain this obvious usability issue. However, similar problems also exist in web archives that have a longer history, although to a lesser degree. PANDORA and Library of Congress also provide only one simple search interface, and leave instructions for advanced search only in the "help" text. See Figure 4.

The PANDORA "help" page also provides a link to the advanced search interface. This link should be provided on the same web page as the single search box, since according to Ras & Bussel (2007), few users will use the "help" page.



Figure 4: The search and browse interface of PANDORA.

## Search results

All the web archives present all archived versions of a URL, the date and time when each version was captured, and group all archived versions of a URL together. Five web archives provide a content summary for each returned URL that is automatically extracted from the archived web page. See Figure 5.

**Government - Utah.gov**  
 Government - Utah.gov Skip to content Registrants 24/7 Help Font Size: Decrease Font Size Normalize Font Size Increase Font Size Utah.gov The Official Web Site for the State of Utah Government Residents Visiting Business Employment Education About Service Government State Agencies A complete list of state government agencies Office of the Governor Links to information on the Utah Governor's... branches City & County Government Links to local government information Federal Government Links to federal government resources Election & Voting Links to information on voting and elections Utah Laws A... government employees Connect with Government Mobile and social networking channels Data Find online government data Residents Citizen & Community An alphabetical list of Utah counties, cities, and towns...  
 text/html - 45.2 KB - crawled once Mar 22, 2010  
<http://www.utah.gov/government/agencylist.html> - more results from utah.gov

**Montana State Government - mt.gov**  
 Montana State Government - mt.gov . State of Montana Access Your Government Tourism Working & Living Online Services Business Government Education Home State Agency Listings & Employee Directory... and Subjects Indexes can help you find information using a database of subjects or common government... three branches of government and describes the basic rights belonging to the people of the State. The... Counties, Montana League of Cities and Towns Montana Association of Counties Tribal Government... Affairs , or Indian Nations . Federal Government Federal Government resources, including the Montana Congressional delegation, government agencies, USA.gov and branches of the military. Congressional... Legislative Branch Montana Laws (MCA) 2011 Biennium Budget 2009  
 Legislative Roster State Government State...  
 text/html - 15.4 KB - crawled once Feb 19, 2010  
<http://mt.gov/government.asp> - more results from mt.gov

Figure 5: Content summary in the search results of Archive-It.

The Library of Congress and the New Zealand Web Archive provide a bibliographic record for each returned result instead of a content summary. See Figure 6.

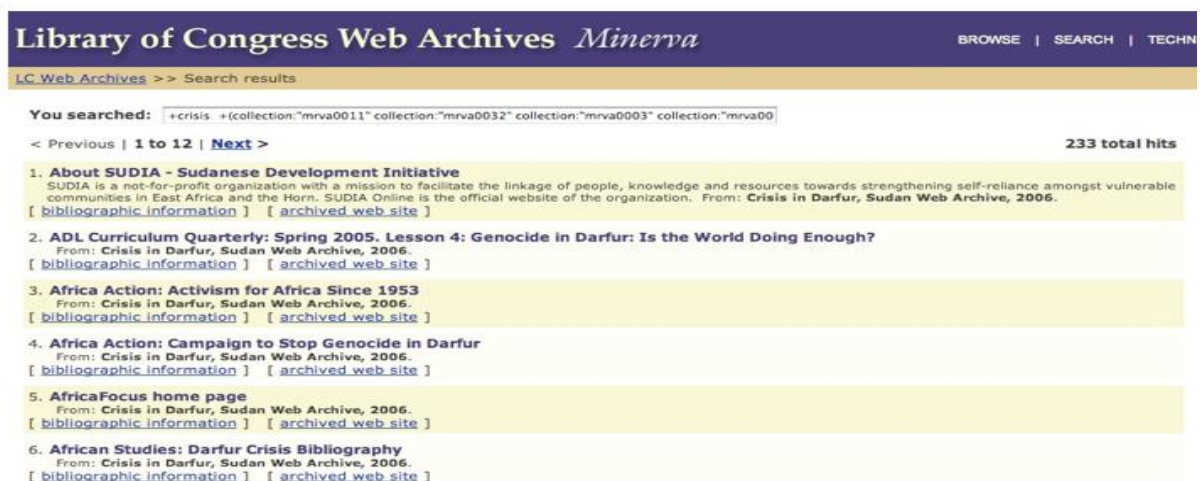
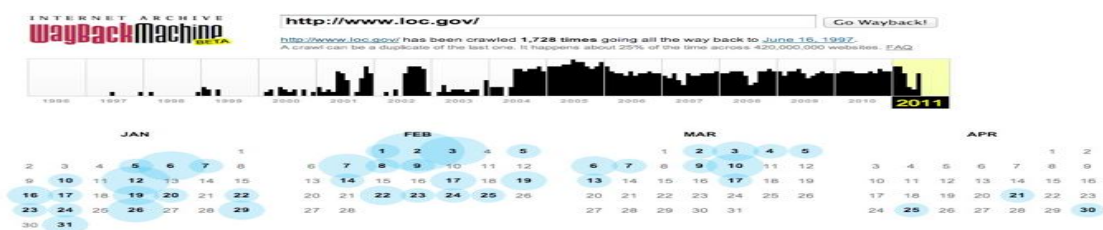


Figure 6: Search results page from the Library of Congress Web Archives.

Other web archives return only matching URLs with their capture dates. IA Wayback Machine presents the resulting URLs in the format of a calendar. See Figure 7. The UK Web Archive presents the resulting URLs as a list as shown previously in Figure 2. All the web archives support navigation through hyperlinks, just like surfing on the Web.



Fig

Figure 7: The calendar view of search results at IA Wayback Machine.

**Group search results by domain:** Five web archives show the hierarchical relationships between websites and the paGES that belong to those websites. In other words, all of the web paGES from the same web site that match the search criteria are grouped together in the search results. (See the example from Archive-It in [Figure 5](#).) The UK Web Archive and UK Government Web Archive return matching web paGES from the same website as different search results. Fortunately, paGES from the same website are listed adjacent to each other. Thus, users are still able to see all web paGES from the same site, although the results are not hierarchically presented. (See the earlier example from the UK Web Archive in [Figure 2](#).)

**Persistent identifiers:** Four web archives provide persistent identifiers for archived web paGES. PANDORA provides persistent identifiers for citing a title (a journal, a website, etc.) and the component parts of a title (e.g. an article within an issue of an electronic journal, or an image or a table within a web site). The persistent identifier for a title is ready-made. The persistent identifiers for a part of a title can be automatically generated upon request by using the citation service. The Library of Congress Web Archives provides a citation ID in its metadata records for websites and sugGESTs that users include this citation ID when citing the archived web page. WAX offers ready-made citations in three styles — APA, Chicago and MLA — for three levels of resources: a web archive collection, a website and an individual archived version of a website. A URL is included in the ready-made citation format. Since this URL is used for the citation, it should be a persistent identifier. WAS displays persistent identifiers (archival URL) in the banner on top of the browser window for each archived web page.

**Indicating archived content:** All web archives construct URLs for archived web content based on the URL of the web archive and the original URL of the archived web page.

All web archives except PANDORA use a banner to indicate that the version being viewed is an archived web page. IA Wayback Machine and Archive-It provide the option to hide or close the banner. While the option to close the banner is useful, users who later need the information in the banner cannot re-open it. The Library of Congress Web Archives and the UK Web Archive provide better designs. They allow users to hide the banner partially and recover the whole banner when needed. See Figure 8.



Figure 8: The banner from the Library of Congress Web Archives.

**Printing:** Three web archives do not allow users to print the banner at the top of the browser window. Two web archives restrict certain images from printing, and the layout of the prints is also different from what is presented in the browser. Figure 9 compares the two sites layouts. The difference between the two is so dramatic that without careful examination a user may not recognize that the two printed pages are actually the same web page.



Figure 9: Left image: the view of an archived page in the browser from WAX. Right image: the print view of the same page.

***Certification of reliability:*** None of the web archives mention that they certify the reliability of records archived in their web archives. On the contrary, several web archives (IA Wayback Machine and WAX) provide a disclaimer saying that they are not responsible for the accuracy or reliability of content in their web archives. PANDORA says it takes various measures to preserve the authenticity and integrity of archived resources, but does not mention that the archive can certify the authenticity, integrity and reliability of the archived web content. The Government of Canada Web Archive preserves government websites, which are more likely to be used for government accountability. However, even that web archive provided a disclaimer for the reliability and authenticity of its archived web resources.

Three items on the functionality checklist are about the comparison of different versions of an archived URL. IA Wayback Machine is the only web archive that provides the functionality of comparing two versions of a web page. However, this functionality did not work during the time this study was conducted, so it is not known to what extent, and how, IA Wayback Machine supports the comparison of different versions of web paGES.

### **Browsing**

Eight of the ten web archives provide browsing functionality. Seven of them organize content into collections or categories based on subject or genre, as shown in [Figure 4](#). The Government of Canada Web Archive organizes websites by government departments. The collection of IA Wayback Machine is very broad and there is no human generated metadata that helps categorize its collection. However, in the future, its collection might be organized into categories and sub-categories based on automatically extracted metadata. For example, the collection could be categorized first by country domains, then the web of a country could be categorized by other Top Level Domains (.gov, .edu, etc.), each sub-category could be further divided by genre (blogs, newspapers, virtual worlds), then categorized by media types (PDF, HTML, video, etc.), and

so on. This browsing structure would allow users to use the web archive when they do not know a specific URL.

### **Policy-related functionalities**

Policies for handling users' requests to block public access to content already in an archive, to remove content from an archive, or to add content to an archive, are not consistent across the ten archives. For example, blocking public access is only mentioned explicitly by WAS; the Library of Congress proactively blocks online access to websites that they do not have permission from copyright owners to show; and PANDORA proactively restricts access to a very small portion of titles for commercial reasons, or because their content is sensitive.

Regarding users' requests to add or remove specific URLs, WAS and six other web archives state that they have a "takedown policy" for removing specific web pages from the archive upon users' requests. The Library of Congress explicitly states that it does not accept user requests to add specific URLs to their archive. The UK Government Web Archive and WAX do not mention whether they allow users to request specific URLs to be archived, but some of the other web archives do accept user-recommended URLs for archiving.

### **Personalized services**

The National Library of New Zealand catalog provides personalized service for users, such as saving searches and records for later user. Since the New Zealand Web Archive is accessible through the library catalog, the personalized service is also available for the web archive users. However, it was not created specifically for the web archive. IA Wayback Machine allows users to create an account, but this personal account is used for uploading and sharing information to other databases of IA Wayback Machine and seems unrelated to the Wayback archive. No other web archive was found to provide personalized services. PANDORA is the only web archive that publishes monthly reports about new titles added to the web archive. This public notification is not the

same as the personalized alert service mentioned in the IIPC use cases. Personalized services are a common functionality in many traditional library OPACs and digital libraries. They are also provided by some online archives with blended web resources and digitized content, such as the [UCLA Online Campaign Literature Archive](#) of the University of California, Los Angeles. By adding this functionality, web archives are more likely to be capable of attracting and keeping repeat users.

### **Data mining**

None of the web archives show or mention that they support any of the data mining functionalities on the checklist. The UK Web Archive offers two visualization interfaces based on the mining of its archived content: tag clouds and a 3D wall. However, it does not seem to provide data mining functionalities for researchers. None of the web archives were found to preserve website log files that include information about operating systems, web servers, versions, and network connection speeds. Thus, these web archives are unlikely to be able to support the kind of research about the evolution of web technology that is mentioned in IIPC use cases, although the archived web paGES themselves demonstrate the evolution of web technology. Website log files are more likely to be preserved for an organization's internal records management purposes or for specific research purposes. All of the web archives in this study rely mainly on online harvesting for external archiving, and this at least partially explains why they do not archive log files.

### **Functionalities discovered in the study that are not on the checklist**

***Duplicates management:*** IA Wayback Machine provides the option of showing duplicates in search results.

***Discoverability by search engines:*** Several web archives explicitly state that they block search engine crawlers from indexing archived web paGES. The Government of Canada Web Archive only allows Google to index its main page. The UK Web Archive and PANDORA allow commercial search engines to index the information page of each archived title. Titles archived by the UK

Web Archive and PANDORA will show up in Google search results if the right search query is used. For example, the query "Australian Industry Group/American Express cash management — managing cash flow in troubled times" will return the archived version at PANDORA. The search result will not lead users directly to the archived page. Instead, it brings the user to the information page about this archived document on the PANDORA website. See Figure 10. The web page clearly shows the requested document is archived by PANDORA and keeps users from confusing it with the live website. A similar mechanism is used by the UK Web Archive.



Figure 10: A Google search result for a web page archived by PANDORA.

**Indicating non-archived content:** When a requested URL is not archived, all of the web archives return a page showing the requested URL, why the page is not archived, and/or what users should do. IA Wayback Machine also automatically captures the requested URL, and notifies users about this capture in the banner on top of browser window.

### **Limitation of the study**

This study relied solely on data gathered from the websites of ten publicly accessible web archives with an English interface. This poses concerns about the completeness of the findings. It is possible that some web archives that are accessible on-site only, and non-English web archives, provide more advanced functionalities that were not discovered in this study. It is also possible that some functionalities, such as extracting a subset of the web archive for outside

analysis and recovering portions of lost websites, are supported but not visible on the websites of the web archives. A more thorough study should include gathering information from the people who build and maintain the web archives, in addition to examining the archives themselves.

## **1.2. Principles of exchange documents**

To fully understand the operation of the electronic exchange documents management system, we introduce a summary of the main components.

The document - is fixed on a material bearer with requisites allowing to identify it.

An electronic document (ED) - the information recorded in electronic form, certified electronic digital signature and having other details of the electronic document, allowing him to be identified.

Exchange document - the movement of documents in the organization since its creation or receipt to complete the execution or administration.

Exchange Electronic Document Management (EEDM) - Electronic document is a collection of sending and receiving electronic documents through the information system. <sup>1</sup>

Automation System workflow, exchange electronic document management system (EEDMS) - an automated multi-user system that accompanies the work of the management of hierarchical organization to ensure compliance with the organization of its functions. It is assumed that the control process is based on the human-readable documents containing instructions for employees required to perform.

Consider two of the electronic document management system, which was introduced earlier in the archival institutions in some of the CIS (Commonwealth of Independent States) countries:

---

<sup>1</sup> Закон республики Узбекистан «Об электронном документообороте»

✓ The free application "BOSS-Referent 4J» - exchange electronic document and interaction management system, aimed at improving the efficiency of work across the organization in different areas of their joint activities.

AdvantaGES:

- Manage the organizational structure
- Manage external recipients
- Processing of incoming and outgoing correspondence

DisadvantaGES:

• The lack of easy access to all the documents on the subject: categorization and search of documents

- Can not support the full route management processing document
- Enough complex installation, configuration and restart the program
- Low performance and scalability

✓ The software for a fee, the MOTIVE system - the first system of management of the company in real time. Control of execution of orders, a collective work on the project, coordination of documents, electronic archive and much more.

AdvantaGES:

- Save time and reduce the routine work;
- risk control (the ability to keep control of the most important tasks, not forgetting the less important);
- Reducing the time to search for the information you need;

DisadvantaGES:

- Lack of remote access and efficient interaction with branches; accumulation of intellectual experience of the company.
- The main disadvantage is the limited system of work with a lot of information - each user of the system may not be thousands of tasks, but only two or three.

After reviewing the above exchange electronic document management system, the management of archival institutions, it was decided to create a "Program complex exchange electronic document archival institutions", which will include materials and documents relating to the archival institutions.

When designing the system already in the initial staGES, it is necessary to lay the basic principles of rational construction of the structure, which will significantly affect the course of the design, development and implementation of the system, ensure the correct course of action leaders and experts in solving various issues. These principles specify that you must first be provided, and then perform to achieve this goal the most effective way.

Collection of information and research activities of the program complex exchange electronic document archival institutions should be carried out in several staGES:

- study of functional areas in order to determine the required input and output information;
- building the organizational and functional charts and analysis of information flows, reflecting the specifics of the structure and activities;
- definition of information objects and details of appropriate composition.

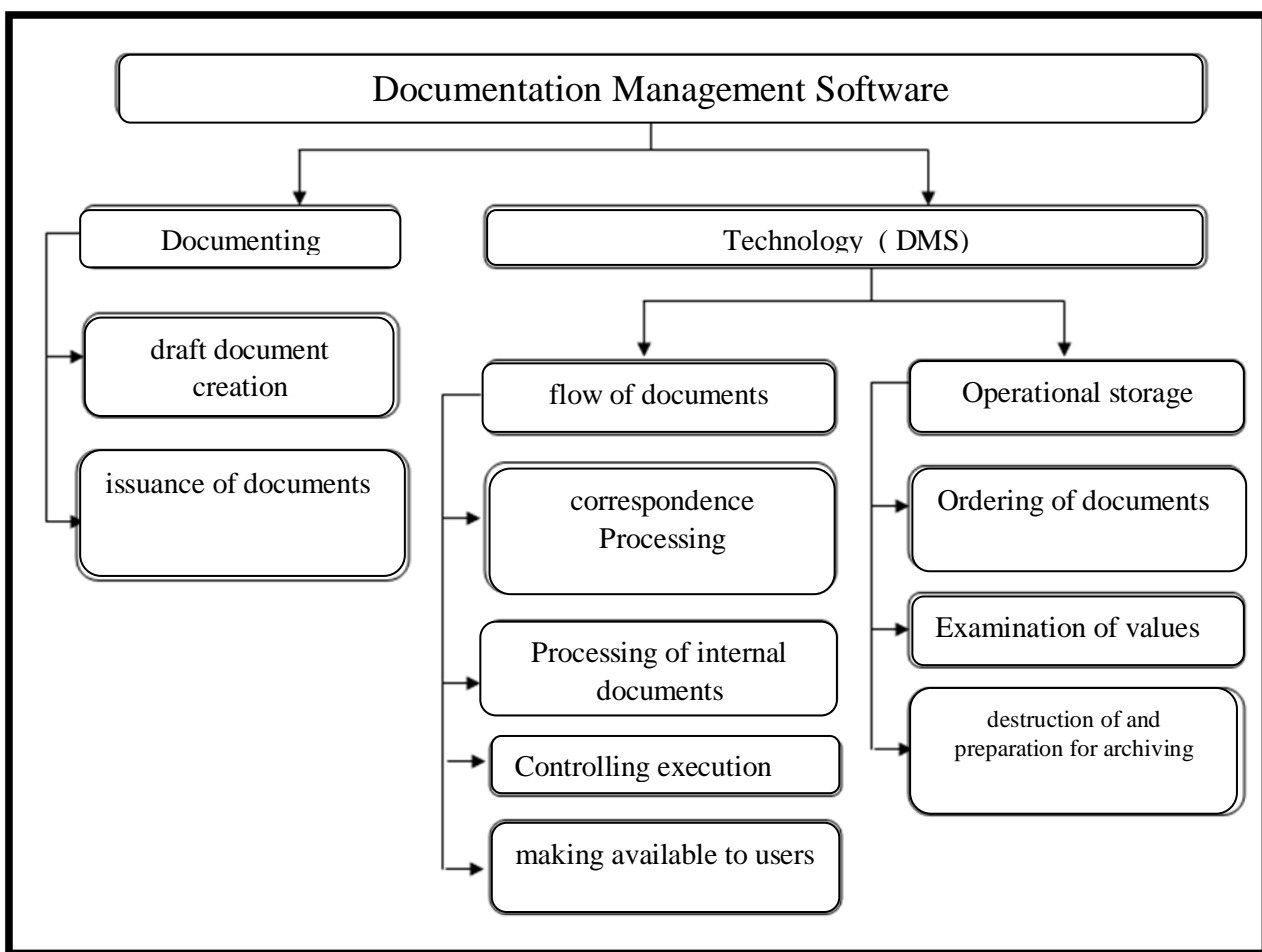
In order to achieve the efficiency of the automated system is required complete information. In addition to access to current information, it is desirable to have access to the archive data in accordance with the principles of the document, which set mandatory storage of archival documents. Thus, the system should have the function of information storage.

Tradition and workflow rules were formulated in 1811 by M.M.Speransky, it was then for the first time feature the requirements for working with documents in public institutions. There was a records management - the science that studies the laws of records and exchange document management and how these patterns affect the management of organizations and enterprises in the framework of documentary maintenance of management of the enterprises.

Fundamentals and challenges facing Documentation, since then has not changed significantly. Many of the requirements and rules fixed in O`zDST 1157-2008 and a number of other regulations.

It is known that the exchange document is part of a records management. Documentary management software

- A branch of activity, providing documentation and organization of the official documents, which has its own technology (see Figure 1.). Technology document management - a document the organization, storage and use of documents in the current activities of the institution. <sup>2</sup>



*Fig. 1. Documentary management software*

Next, you need to identify the main components involved in the process workflow. The basic concept - a document. The document is considered to be a set of three components:

<sup>2</sup>Documentary management software -

[http://www.docsup.ru/ver1/content/articles/1\\_dou\\_principles/dou\\_principles.shtml](http://www.docsup.ru/ver1/content/articles/1_dou_principles/dou_principles.shtml)

1. The physical recording of information;
2. The format of the information;
3. activation of certain activities.

That some activity and converts the information into a document. But the document ceases to exist in the future if it is not intended processing procedure. It should be noted that the form of the document is closely linked to the nature of the further activities of the officials, it generates the necessary documents. It follows a simple principle: the more activity - more documents.

It can be argued that the document - the basis of the scientific organization. Such base management subsystem as accounting, planning and decision-making is impossible to imagine without the use of documents. The document has a definite place in the normal course of business at the border separating the execution of functions and acts as a link between the individual employees of the company. It should be noted that with the help of documents implemented some rules of interaction within the organization.

By defining the document as distribution functions of the tool between employees, look at a few types of such functional chains. The simplest model: one - to one. From the title it is clear that the document has one sender and one receiver. An example is, the initial stage of the application for leave, when it is supplied to the personnel department and is not causing global activities. In this case, the document has the same sender (one who submits an application) and a recipient (personnel officer, the receiving application).

Next model: one - to many. This model is used when a task should be entrusted to several persons, used just such a model. At the same time one document is sent to the set of employees.

You can complicate the model, increasing the number of instruments and performers. However, as the most important characteristics of the instrument performs its versatility, it is possible to use the same document in multiple functions. This feature allows you to not to multiply the number of types of documents in the enterprise. At the same time it is this property is used in the

main control functions of the system - control. But this is a function of office, rather than exchange document.

Exchange electronic document management systems and electronic archive systems have a number of common features. Firstly, the processing unit in both classes is an electronic document systems. Secondly, the two-class systems are designed to solve similar tasks, ie. Structuring and organizing the flow of documents within the organization. Third, as noted above, both classes of systems have similar interfaces and include approximately the same basic set of functions.

The differences between the two classes of systems are primarily in the approach to solving problems. you must include the following significant differences include:

- Exchange electronic document management system captures all current records of the organization; archives (and electronic files included) intended to preserve important documents; significance determination can not be carried out automatically and is a purely human task;

- Exchange electronic document management systems are designed for everyday work with documents (amendments, approval, distribution to various employees of the organization, and so on.). Electronic archive system may include a means to support the daily work, but their main goal is to secure (including by modification) storage of important documents.

- Exchange electronic document management systems are designed to work with small amounts of information; electronic archive system was originally designed to work with large sets of documents;

- Exchange electronic document management systems are designed to structure the flow of documents, which the organization has to deal in the present; electronic archive system is designed for organizing and structuring the last document; In addition, they allow you to define a framework for organizing and storing documents in the future;

- Exchange electronic Document Management System provides only short-term storage of documents, while the electronic archive system ensures the long-term (as highlighted in many publications on the subject - almost eternal) storage;

- The electronic archive systems are enhanced to search for documents.

When you create a software system of exchange electronic document archival institutions need to define the basic principles of exchange electronic document:

- Single sign a document allowing to identify the document uniquely.

- Possibility of parallel operations, allowing to reduce the movement of documents and improve their performance efficiency

- The continuity of the movement document allowing to identify the person responsible for the execution of the document (tasks) in each moment of life instrument (the process).

- One (or coordinated distributed) base document information, to eliminate the possibility of duplication of documents.

- Efficiently organized document retrieval system that allows you to find the document, having a minimum of information about him.

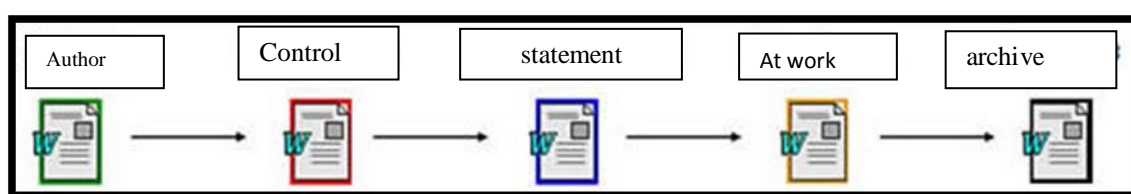
- Advanced reporting system on the various statuses and attributes of documents, allowing to control the movement of documents and workflow processes to make management decisions based on the data from the reports.

Having considered the concept of workflow and basic principles of electronic document, proceed to the analysis of this process in enterprises. Firstly, it is necessary to analyze the challenges that will be faced in implementing the system.

### 1.3. Problems in the implementation of electronic document archival institutions

In general, enterprise management can be divided into two components: the actual management and its information support. The first group includes the activities involved in the "decision-making system", and the second - the processes that ensure the accumulation, processing, analysis of the information needed to guide the company. Any large organization, such as archival institutions, operates the huge flow of information.

Consider the circuit of the electronic document life cycle. (Fig.2)



*Fig. 2. The life cycle of an electronic document*

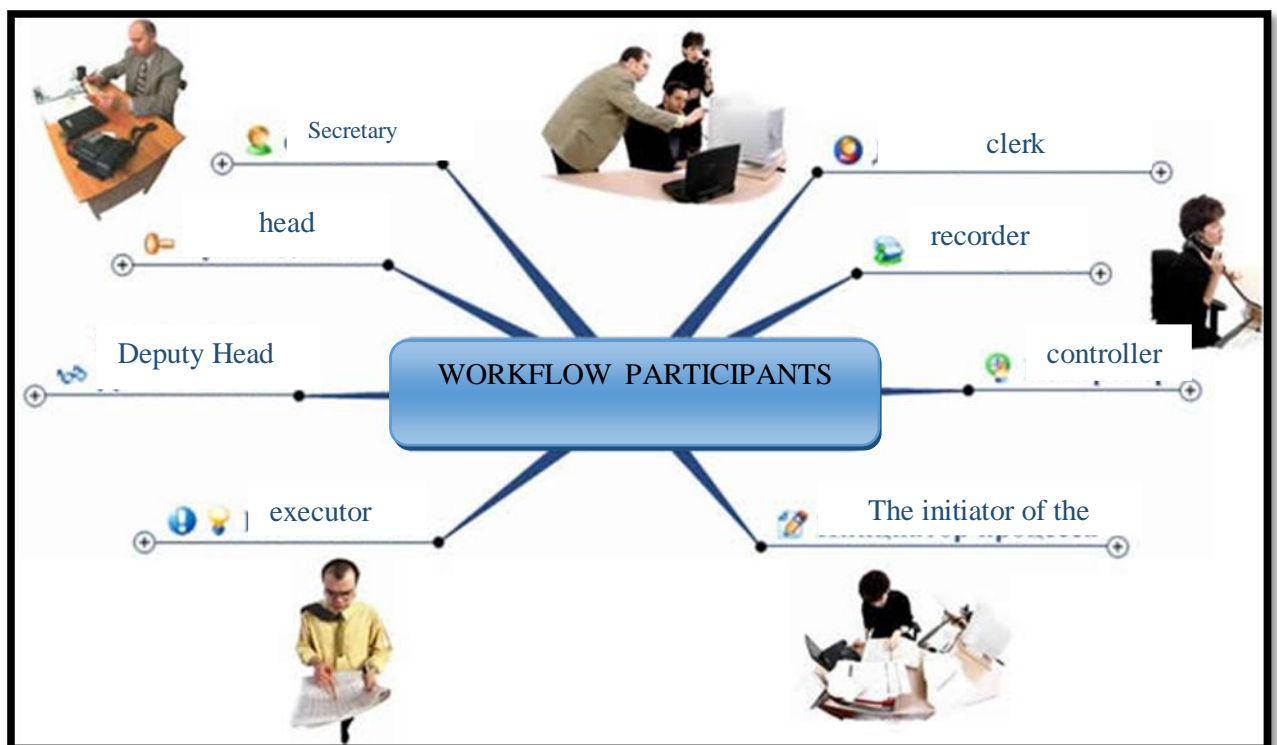
It should be noted that there are specialized software for implementing document management process, in principle, take into account all that is necessary. However, in this case, we have to deal with some range of issues.

First of all, note that when you buy the software, third-party manufacturers in our region there is practically no support, and not only technical. No support for staff working with the software. Let us try to explain the problem.

Electronic document management systems have a certain feature: a system should be implemented in all workplaces, related to the creation, editing and storage of information (documents), or the efficiency of its use will be minimal. This approach immediately identifies one of the key implementation problems: conservatism of personnel, due to the reluctance of trained and retrained, and possibly, low education in the field of computer technology. This problem can to stall the whole process of implementation. This is especially true of organizations in which the personnel policy is conservative, and nobody, not even the head is not free to move or upgrade training.

Another problem for any organization under the current electronic document are permanent structural changes. This is required even if small, but all the restructuring of the system structure. In the absence of a part of the staff of specialists in the field of information technology, this process can take a lot of time at all, which in turn will lead to a simple production process and as a result, financial losses.

Try to imagine the work process of exchange electronic document archival institutions. (Fig. 3)



*Fig. 3. The staff involved in the document are divided into different categories*

The third problem may lie in the complete absence of exchange document (even paper). In such a situation has its advantages. First and foremost is the lack of need someone to retrain. In addition, there are objective prerequisites for to convince the management to implement exchange document management system. If the leadership given a choice: to introduce archival paper or modern electronic, most likely that the latter will be chosen.

A serious problem is that if a large organization does not do any formal document, then it is constantly confronted with many challenges, and

management does not always understand the reason for the lack of a formalized scheme of doing things. As a result, managers are constantly falls heap of problems, and they do not have time to understand the issues of document management systems. In this case it is necessary to carry out a gradual introduction of the system, to run, so to speak, a pilot project, the work of which will be invisible to the authorities, but the effectiveness of it still appears. This process can take several months before you see results. Support for the pilot project, as well as continued full implementation of the enterprise can result in a considerable amount.

Consider the main features and workflow issues in an organization or enterprise:

- The document management documents must be legally significant
- The documents can be divided into original and copy. Quite often, only to original documents received consideration. What would a copy of the document was the legal value it is necessary to assure a notary.
- Documents are the primary (cash receipt, invoice, certificate, report, etc.) and secondary. The primary document is the first evidence of the facts occurred and confirms validity of produced economic operation. It sets individual performers responsible for the performance of their business operations.
- If your organization has a large document (large volume of documents), then sooner or later the question arises about the correct storage of documents and the organization of the archive documents (document repository).
- When moving offices, or other difficult situations some of the documents may be lost or corrupted. When the competent organization of workflow, these risks can be minimized.
- If the wrong workflow or storage of documents is often difficult and time consuming to find the right document in a document repository or to make a thematic selection from the archive documents.

If you go to exchange electronic document, it is possible to minimize the risk of loss or damage of documents to facilitate the signing and transmission of documents, significantly reduce document storage costs.<sup>3</sup>

These problems apply to all businesses regardless of their specificity. We now consider the specific archival institutions. These problems apply equally to them, that does not speak in favor of third parties. The archival institutions are quite extensive parks of computer technology, which have the potential to significantly improve the quality and efficiency of work with data. Nevertheless, the formation of a local area network, as well as the creation of a local network, only slightly impact on increasing the performance efficiency of the archival institution units.

In addition there are now many units of archival institutions use to exchange information rather inefficient methods described previously. There are no common standards on the size of incoming and outgoing data streams: one used drives, flash drives and other shared resources - which ultimately leads to disruption of the single information space of archival institutions and, in general, to the inability to find the necessary document. In many cases, departments keep up not only their own documents but also included, resulting from the work of other departments. The consequence of such an approach to the process workflow is a redundancy and low efficiency display the chanGES.

It follows from the foregoing that the archival institutions need to focus efforts not only on the development of the hardware platform, but also for the development of software that would allow the most efficient use possible of personal computers. As already mentioned the purchase of third-party software does not meet the archival institutions, not only in terms of financial costs, but also in terms of support staff and the introduction into the structure. In addition, the software market there are also systems developed by various organizations for their needs. The price of such programs are much lower, but they are

---

<sup>3</sup>Основные особенности и проблемы документооборота в организации или на предприятии - <http://www.esign-pro.ru/glossary/a-6-dokumentooborot.html>

designed for specific universities, and their use in archival institutions require serious revision.

The best solution for archival institutions is the creation of the necessary software on their own. The benefit of this solution is the fact that currently the archival institutions have sufficient intellectual capacity to create an electronic workflow system.

#### Description of the subject area

In general, the subject area covers all processes, actions and interactions that take place within the system, as well as between the system and the environment. Thus, this section is devoted to the description of the archival institution's structure, namely, the agency "Uzarhiv" its individual units and links between them.

#### Main unit "Uzarhiv" agency

Agency "Uzarhiv" has a fairly simple structure, which can be identified koordinatoraskuyu group and information - resource center, the main sub-sections, which are based on an automated workflow system. The first group includes the following structures:

- Department of organizational and methodical work on archives of the Agency "Uzarhiv". The main tasks of the Department:

- development of guidelines for the development of archives in the country, together with other departments of the Agency and organizational units, as well as the improvement of the organizational and methodological work and work to ensure the preservation of documents; the organization of long-term planning of archives and records management, development and planning of archival institutions of the republic for the current year and improvement;

- increase the economic efficiency of archives and records management, preservation of archival documents, maintaining the definition of public accounting principles and guidelines for the development of archives and records management;

- organization and planning of the board of the central apparatus of the Agency, to coordinate the activities of departments and divisions;
  - organization of state registration documents National Archival Fund of the Republic of Uzbekistan and centralized accounting documents of National Archival Fund(NAF) of the Republic of Uzbekistan;
  - maintain control and provide organizational and methodological assistance to archival institutions for recruitment issues, placement, training and skills development;
  - control over the activities of security and fire safety of State archives;
  - study, generalization and introduction of advanced domestic and foreign experience in the field of archives and records management.
- Department of Electronic Archives and Records Management Agency "Uzarhiv". The main tasks of the Department:
    - control over the right conduct of documents on archives and records management in Uzbekistan and make appropriate recommendations;
    - organization to streamline and systematize the creation of electronic documents, providing storage of information resources related to the documents of the National Archival Fund of the Republic of Uzbekistan (NAF RUz);
    - To provide guidance to the Central Office of the agency, structural divisions, regional offices to work in the fields and the intended use of Internet resources, including system information resources, consultation on improving the registration of documents, analysis of their turnover, ensuring control on the basis of the introduction of electronic document management and document management system;
    - participation in the implementation of the agency's structure and its departments of information and communication technologies and ensure that the necessary programs;
    - Organization preservation of electronic documents and information resources NAF RUz;

- maintenance management and control in the delivery of NAF RUz goskhranenie documents in electronic form on the institutions, which are the sources of acquisition of public archives.

- Department of the use of documents and information service of the Agency "Uzarhiv". The main tasks of the Department:

- Control over the execution of orders, instructions and regulations of the Agency "Uzarhiv" under the Cabinet of Ministers of the Republic of Uzbekistan;
- realization of methodical management and control of the republic's state archives on the use of documents and provision of information services;
- Execution of orders and instructions of the Agency "Uzarhiv" distribution of working hours in the form of instructions and assignments territorial archives and public records offices in the field;
- organization of work for the timely and quality execution of socio-legal and case requests received by the Agency "Uzarhiv" from citizens, institutions and organizations;
- Prepare information on scientific and practical significance of the documents NAF Uzbekistan, stored in the archives, together with the state archives of the Republic

Human Resource Management. management staff is made employment of citizens in the State Archive of the structure of Uzbekistan and their dismissal. Ensure the timely preparation of reports, regular forms, conducting competitive selection coordinators and contracts. Human Resources Management provides timely and regular registration of employees for retirement, the organization and conduct of medical preventive measures; document preparation and issuance of health insurance policies; registration forms for the opening of a personalized account in the Pension Fund each employee and the issuance of insurance certificates; preparation of documents and transfer them to storage to the

national archive; organization of reception on personal issues and control over the execution.

The structure of personnel management include:

- Department of Employment;
- Accounting department personnel;
- General Division;
- Archive;
- Staffing service regional centers.

### **Conclusions**

1. formulated and set out the general principles of the software system for electronic document archival institutions, which significantly affected the course of the design, development and implementation of the system.

2. Using a relational approach to building Infological model allows to identify and eliminate duplication of data objects for further database design.

3. Since examination of the structure of information flows observed in other organizations, it can be concluded about the universality of the structural data model. Consequently, the program is designed for the construction of such a system in any organization.

## CHAPTER 2. INFORMATION MODELS AND ALGORITHMS FOR PROCESSING OF ARCHIVAL DOCUMENTS

### 2.1. Structure metadata of archives.

Metadata is commonly defined as "data about data." Metadata is frequently used to locate or manage information resources by abstracting or classifying those resources or by capturing information not inherent in the resource. Typically metadata is organized into distinct categories and relies on conventions to establish the values for each category. For example, administrative metadata may include the date and source of acquisition, disposal date, and disposal method. Descriptive metadata may include information about the content and form of the materials. Preservation metadata may record activities to protect or extend the life of the resource, such as reformatting. Structural metadata may indicate the interrelationships between discrete information resources, such as page numbers.

In terms of archives, MARC format and EAD are standards for structuring descriptive metadata about collections. Dublin Core is a standard for structuring metadata that is intended for describing web resources.

In terms of information technology, metadata includes the documentation of data architecture, properties, and methods necessary to store, retrieve, and use the data in a meaningful manner. To the extent that data is a record, it may also include administrative, descriptive, preservation, and structural information.

**Digital Identification.** Most metadata schemes include elements such as standard numbers to uniquely identify the work or object to which the metadata refers. The location of a digital object may also be given using a file name, URL (Uniform Resource Locator), or some more persistent identifier such as a PURL (Persistent URL) or DOI (Digital Object Identifier). Persistent identifiers are preferred because object locations often change, making the standard URL (and therefore the metadata record) invalid. In addition to the actual elements that point to the object, the metadata can be combined to act as a set of identifying data, differentiating one object from another for validation purposes.

**Archiving and Preservation.** Most current metadata efforts center around the discovery of recently created resources. However, there is a growing concern that digital resources will not survive in usable form into the future. Digital information is fragile; it can be corrupted or altered, intentionally or unintentionally. It may become unusable as storage media and hardware and software technologies change. Format migration and perhaps emulation of current hardware and software behavior in future hardware and software platforms are strategies for overcoming these challenges.

Metadata is key to ensuring that resources will survive and continue to be accessible into the future. Archiving and preservation require special elements to track the lineage of a digital object (where it came from and how it has changed over time), to detail its physical characteristics, and to document its behavior in order to emulate it on future technologies. Many organizations internationally have worked on defining metadata schemes for digital preservation, including the National Library of Australia, the British Cedars Project (CURL Exemplars in Digital Archives), and a joint Working Group of OCLC and the Research Libraries Group (RLG). The latter group developed a framework outlining types of presentation metadata. A follow-up group, PREMIS (PREservation Metadata: Implementation Strategies)—also sponsored by OCLC and RLG—is developing a set of core elements and strategies for the encoding, storage, and management of preservation metadata within a digital preservation system. Many of these initiatives are based on or compatible with the ISO Reference Model for an Open Archival Information System (OAIS).

**Structuring Metadata.** Metadata schemes (also called schema) are sets of metadata elements designed for a specific purpose, such as describing a particular type of information resource. The definition or meaning of the elements themselves is known as the semantics of the scheme. The values given to metadata elements are the content. Metadata schemes generally specify names of elements and their semantics. Optionally, they may specify content rules for how content must be formulated (for example, how to identify the main title),

representation rules for content (for example, capitalization rules), and allowable content values (for example, terms must be used from a specified controlled vocabulary). There may also be syntax rules for how the elements and their content should be encoded. A metadata scheme with no prescribed syntax rules is called syntax independent. Metadata can be encoded in any definable syntax. Many current metadata schemes use SGML (Standard Generalized Mark-up Language) or XML (Extensible Mark-up Language). XML, developed by the World Wide Web Consortium (W3C), is an extended form of HTML that allows for locally defined tag sets and the easy exchange of structured information. SGML is a superset of both HTML and XML and allows for the richest mark-up of a document. Useful XML tools are becoming widely available as XML plays an increasingly crucial role in the exchange of a variety of data on the Web.

### **Metadata Schemes and Element Sets**

Many different metadata schemes are being developed in a variety of user environments and disciplines. Some of the most common ones are discussed in this section.

**Dublin Core.** The Dublin Core Metadata Element Set arose from discussions at a 1995 workshop sponsored by OCLC and the National Center for Supercomputing Applications (NCSA). As the workshop was held in Dublin, Ohio, the element set was named the Dublin Core. The continuing development of the Dublin Core and related specifications is managed by the Dublin Core Metadata Initiative (DCMI).

The original objective of the Dublin Core was to define a set of elements that could be used by authors to describe their own Web resources. Faced with a proliferation of electronic resources and the inability of the library profession to catalog all these resources, the goal was to define a few elements and some simple rules that could be applied by noncatalogers. The original 13 core elements were later increased to 15: Title, Creator, Subject, Description,

Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, and Rights.

The Dublin Core was developed to be simple and concise, and to describe Web-based documents. However, Dublin Core has been used with other types of materials and in applications demanding some complexity. There has historically been some tension between supporters of a minimalist view, who emphasize the need to keep the elements to a minimum and the semantics and syntax simple, and supporters of a structuralist view who argue for finer semantic distinctions and more extensibility for particular communities.

These discussions have led to a distinction between qualified and unqualified (or simple) Dublin Core. Qualifiers can be used to refine (narrow the scope of) an element, or to identify the encoding scheme used in representing an element value. The element Date, for example, can be used with the refinement qualifier created to narrow the meaning of the element to the date the object was created. Date can also be used with an encoding scheme qualifier to identify the format in which the date is recorded, for example, following the ISO 8601 standard for representing date and time.

All Dublin Core elements are optional and all are repeatable. The elements may be presented in any order. While the Dublin Core description recommends the use of controlled values for fields where they are appropriate (for example, controlled vocabularies for the Subject field), this is not required. However, working groups have been established to discuss authoritative lists for certain elements such as Resource Type. While Dublin Core leaves content rules to the particular implementation, the DCMI encourages the adoption of application profiles (domain-specific rules) for particular domains such as education and government. An application profile for libraries is being developed by the Libraries Working Group.

Because of its simplicity, the Dublin Core element set is now used by many outside the library community—researchers, museum curators, and music collectors to name only a few. There are hundreds of projects worldwide that use

the Dublin Core either for cataloging or to collect data from the Internet; more than 50 of these havelinks on the DCMI website. The subjects range from cultural heritage and art to math and physics. Meanwhile the Dublin Core Metadata Initiative has expanded beyond simply maintaining the Dublin Core Metadata Element Set into an organization that describes itself as “dedicated to promoting the widespread adoption of interoperable metadata standards and developing specialized metadata vocabularies for discovery systems.”

**The Text Encoding Initiative (TEI).** The Text Encoding Initiative is an international project to develop guidelines for marking up electronic texts such as novels, plays, and poetry, primarily to support research in the humanities. In addition to specifying how to encode the text of a work, the TEI Guidelines for Electronic Text Encoding and Interchange also specify a header portion, embedded in the resource, that consists of metadata about the work. The TEI header, like the rest of the TEI, is defined as an SGML DTD (Document Type Definition)— a set of tags and rules defined in SGML syntax that describe the structure and elements of a document. This SGML mark-up becomes part of the electronic resource itself. Since the TEI DTD is rather large and complicated in order to apply to a vast range of texts and uses, a simpler subset of the DTD, known as TEI Lite, is commonly used in libraries.

It is assumed that TEI-encoded texts are electronic versions of printed texts. Therefore the TEI Header can be used to record bibliographic information about both the electronic version of the text and about the non-electronic sourceversion. The basic bibliographic information is similar to that recorded in library cataloging and can be mapped to and from MARC. However, there are also elements defined to record details about how the text was transcribed and edited, how mark-up was performed, what revisions were made, and other non-bibliographic facts. Libraries tend to use TEI headers when they have collections of SGML-encoded full text. Some libraries use TEI headers to derive MARC records for their catalogs, while others use MARC records as the basis for creating TEI header descriptions for the source texts.

**Metadata Encoding and Transmission Standard (METS).** The Metadata Encoding and Transmission Standard (METS) was developed to fill the need for a standard data structure for describing complex digital library objects. METS is an XML Schema for creating XML document instances that express the structure of digital library objects, the associated descriptive and administrative metadata, and the names and locations of the files that comprise the digital object. The metadata necessary for successful management and use of digital objects is both more extensive than and different from the metadata used for managing collections of printed works and other physical materials. Structural metadata is needed to ensure that separately digitized files (for example, different pages of a digitized book) are structured appropriately. Technical metadata is needed for information about the digitization process so that scholars may determine how accurate a reflection of the original the digital version provides. Other technical metadata is required for internal purposes in order to periodically refresh and migrate the data, ensuring the durability of valuable resources.

METS was originally an outgrowth of the Making of America II project, a digitization project of major research libraries that attempted to address these metadata issues, in part by providing an encoding format for metadata for textual and image-based works. The Digital Library Federation (DLF) built on that earlier work to create METS, a standard schema for providing a method for expressing and packaging together descriptive, administrative, and structural metadata for objects within a digital library. Expressed using the XML schema language, METS provides a document format for encoding the metadata necessary for management of digital library objects within a repository and for exchange between repositories.

METS document contains seven major sections:

- METS Header – Contains metadata describing the METS document itself, including such information as creator, editor, etc.

- Descriptive Metadata – Points to descriptive metadata external to the METS document (for example, a MARC record in an OPAC or an Encoded Archival Description finding aid maintained on a webserver), or to internally embedded descriptive metadata, or both.

- Administrative Metadata – Provides information regarding how the files are created and stored, intellectual property rights, the original source object from which the digital library object derives, and the provenance of the files comprising the digital library object.

- File Section – Lists all files containing content that comprise the electronic versions of the digital object.

- Structural Map – Outlines a hierarchical structure for the digital library object and links the elements of that structure to content files and metadata that pertain to each element.

- Structural Links – Allows METS creators to record the nodes in the hierarchy outlined in the Structural Map.

- Behavior – Associates executable behaviors with content in the METS object.

The METS header, file section, structural map, structural links, and behavior sections are defined within the METS schema. METS is less prescriptive about descriptive and administrative metadata, relying on extension schemas—externally developed metadata schemes—to provide specific elements. The METS Editorial Board has endorsed three descriptive metadata schemes: simple Dublin Core, MARCXML, and MODS (discussed below). For technical metadata the METS website makes available schemas for text and digital still imaGES. The latter standard is called MIX, Metadata for ImaGES in XML Schema, and is based on a proposed NISO standard, Z39.87, Data Dictionary: Technical Metadata for Digital Still ImaGES. Further work is in process on extension schemas for audio, video, and websites. Another current area of concentration for the METS development community is the creation of METS application profiles to give guidance regarding the creation of METS

documents for particular object types. Use of the METS schema is widespread. A list of implementation registries using METS, a tutorial, and other important information can be found on the METS website.

**Metadata Object Description Schema (MODS)** The Metadata Object Description Schema (MODS) is a descriptive metadata schema that is a derivative of MARC 21 and intended to either carry selected data from existing MARC 21 records or enable the creation of original resource description records. It includes a subset of MARC fields and uses language-based tags rather than the numeric ones used in MARC 21 records. In some cases, it regroups elements from the MARC 21 bibliographic format. Like METS, MODS is expressed using the XML schema language.

Although the MODS standard can stand on its own, it may also complement other metadata formats. Because of its flexibility and use of XML, MODS may potentially be used as a Z39.50 Next Generation specified format, an extension schema to METS, a metadata set for harvesting, and for creating original resource metadata records in an XML syntax. Rich description of electronic resources is a particular focus of MODS, which provides some advantages over other metadata schemes. MODS elements are richer than the Dublin Core; its elements are more compatible with library data than the ONIX or Dublin Core standards; and it is simpler to apply than the full MARC 21 bibliographic format. With its use of XML Schema language, MODS offers enhancements over MARC 21, such as the use of an optional ID attribute to facilitate linking at the element level; the ability to specify language, script, and transliteration scheme at the element level; and the ability to embed a rich description of components in the related Item element.

The ability in MODS to give granular descriptions of constituent parts of an object works particularly well with the METS structural map for complex digital library objects.

**The Encoded Archival Description (EAD).** The Encoded Archival Description (EAD) was developed as a way of marking up the data contained in

finding aids so that they can be searched and displayed online. In archives and special collections, the finding aid is an important tool for resource description. Finding aids differ from catalog records by being much longer, more narrative and explanatory, and highly structured in a hierarchical fashion. They generally start with a description of the collection as a whole, indicating what types of materials it contains and why they are important. If the collection consists of the personal papers of an individual there can be a lengthy biography of that person. The finding aid describes the series into which the collection is organized—such as correspondence, business records, personal papers, and campaign speeches—and ends with an itemization of the contents of the physical boxes and folders comprising the collection.

Like the TEI Header, the EAD is defined as an SGML DTD. It begins with a header section that describes the finding aid itself (for example, who wrote it) and then goes on to the description of the collection as a whole and successively more detailed information about the records or series within the collection. If individual items being described exist in digital form, the EAD can include pointers to the digital objects. The 2002 version of the EAD DTD provides support for both SGML and XML through the use of defined “switches” for turning off features used only in SGML and turning on features used only in XML. The EAD standard is maintained jointly by the Library of Congress and the Society of American Archivists.

**Learning Object Metadata** The IEEE Learning Technology Standards Committee (LTSC) developed the Learning Object Metadata (LOM) standard (IEEE 1484.12.1-2002) to enable the use and re-use of technology-supported learning resources such as computer-based training and distance learning. The LOM defines the minimal set of attributes to manage, locate, and evaluate learning objects. The attributes are grouped into eight categories:

- General, containing information about the object as a whole;
- Lifecycle, containing metadata about the objects evolution;

- Technical, with descriptions of the technical characteristics and requirements;
- Educational, containing the educational / pedagogical attributes;
- Rights, describing the intellectual property rights and use conditions;
- Relation, identifying related objects;
- Annotation, containing comments and the date and author of the comments; and
- Classification, which identifies other classification system identifiers for the object.

Within each category is a hierarchy of data elements to which the metadata values are assigned. Examples of learning-related metadata elements found in the Education category are Typical Age Range (of the intended user), Difficulty, Typical Learning Time, and Interactivity Level.

The IMS Global Learning Consortium has developed a suite of specifications to enable interoperability in a learning environment. Their MetaData Information Model specification is based on the IEEE LOM scheme with only minor modifications.

**MPEG Multimedia Metadata.** The ISO/IEC Moving Picture Experts Group (MPEG) has developed a suite of standards for coded representation of digital audio and video. Two of the standards address metadata: MPEG-7, Multimedia Content Description Interface (ISO/IEC 15938), and MPEG-21, Multimedia Framework (ISO/IEC 21000).

MPEG-7 defines the metadata elements, structure, and relationships that are used to describe audiovisual objects including still pictures, graphics, 3D models, music, audio, speech, video, or multimedia collections. It is a multipart standard that addresses:

- *Description Tools* including Descriptors that define the syntax and the semantics of each metadata element and Description Schemes that specify the structure and semantics of the relationships between the elements.

- *A Description Definition Language* to define the syntax of the Description Tools, allow the creation of new Description Schemes, and allow the extension and modification of existing Description Schemes.

- *System tools*, to support storage and transmission, synchronization of descriptions with content, and management and protection of intellectual property.

*Descriptors* for visual and audio are defined separately using a hierarchy of elements and sub elements. For visual objects there are descriptors for Basic Structure, Color, Texture, Shape, Motion, Localization, and Face Recognition. Audio descriptors are divided into two categories: low-level descriptors that are common to audio objects across most applications, and high-level descriptors that are specific to particular applications of audio. The cross-application low-level descriptors cover Structures and Features (temporal and spectral). The domain-specific high-level descriptors include such elements as Musical Instrument Timbre, Melody Description, and Spoken Content Description.

*The Description Schemes* are based on XML, and can be expressed in textual form suitable for editing, searching, filtering, and human readability; or in a binary form for storage, transmission, and streaming delivery. Since the full description of a multimedia object can be quite complex, the standard provides for a Summary Description Scheme geared to browsing and navigation.

The standard envisions that search engines could use MPEG-7 metadata descriptions to identify audiovisual objects in entirely new ways, such as digitizing a musical phrase played on a keyboard and then retrieving a list of musical pieces that contain the sequence of notes; drawing some lines on an electronic drawing tablet and retrieving images with similar graphics; or using a voice excerpt to retrieve related speech files, photographs, video clips, and biographical information of the speaker. These retrieval mechanisms are outside the scope of MPEG-7, but the standards developers wanted to accommodate these futuristic capabilities and have included many interoperability requirements beyond the typical metadata elements.

MPEG-21 was developed to address the need for an overarching framework to ensure interoperability of digital multimedia objects. The multi-part standard is not yet fully completed but is intended to include the following:

- *Part 1: Vision, Technologies and Strategy* provides the overview of the complete vision and plan for the framework. It was issued as an ISO technical report (ISO/ IEC TR 21000:1-2001) and is available as a free download from ISO's publicly available standards website. A second edition of the vision document is underway to address comments and suggestions received from other organizations following the initial publication.

- *Part 2: Digital Item Declaration*, issued in 2003, describes a model for defining Digital Items. It includes a description of the syntax and semantics of each of the Digital Item Declaration elements and a corresponding XML schema.

- *Part 3: Digital Item Identification*, also issued in 2003, describes how to uniquely identify Digital Items and how to link Digital Items with related information such as descriptive metadata.

- *Part 4: Intellectual Property Management and Protection* is still in development. It is intended to define the framework for ensuring interoperability of intellectual property management tools, including authentication, and accommodates the Rights information defined in the following two parts.

- *Part 5: Rights Expression Language*, issued in 2004, is a machine-readable language that can declare rights and permissions.

- *Part 6: Rights Data Dictionary is still in development*. It will define a standard set of terms to be used with the Rights Expression Language. It is also expected to include specifications for mapping and transforming rights metadata terminology. The Rights Data Dictionary and Expression Language are being viewed as models for the handling of intellectual property metadata for applications beyond audiovisual.

*Part 7: Digital Item Adaptation*, also in development, is intended to standardize networking and interoperability description tools. Included in this

part will be User Characteristic description tools that specify user preferences. There are some seven additional parts identified and in various staGES of development that deal with technical interoperability issues of less specific relevance to metadata. All of the published parts are available from ISO as ISO/IEC 21000-[part#].

**Metadata Crosswalks** The interoperability and exchange of metadata is further facilitated by metadata crosswalks. A crosswalk is a mapping of the elements, semantics, and syntax from one metadata scheme to those of another. A crosswalk allows metadata created by one community to be used by another group that employs a different metadata standard. The degree to which these crosswalks are successful at the individual record level depends on the similarity of the two schemes, the granularity of the elements in the target scheme compared to that of the source, and the compatibility of the content rules used to fill the elements of each scheme.

Crosswalks are important for virtual collections where resources are drawn from a variety of sources and are expected to act as a whole, perhaps with a single search engine applied. While these crosswalks are key, they are also labor intensive to develop and maintain. The mapping of schemes with fewer elements (less granularity) to those with more elements (more granularity) is problematic.

Table 1 shows a crosswalk between Dublin Core, MARC 21, and VRA Core for selected elements. In this case, there is no attempt to map at the content level.

<b>Table 1. Example of Metadata Crosswalk Mapping</b>			
	<b>Dublin Core</b>	<b>EAD</b>	<b>MARC 21</b>
<b>Title Element</b>	Title	<titleproper>	245 00\$a (Title Statement/Title proper)
<b>Author Element</b>	Creator	<author>	700 1#\$a (Added Entry--Personal Name) (with \$e=author) 720\$a (Added Entry--Uncontrolled Name/Name) (with \$e=author)
<b>Date Created Element</b>	Date.Created	<unitdate>	260 ##\$c (Date of publication, distribution, etc.)

## **2.2. Information model for archives**

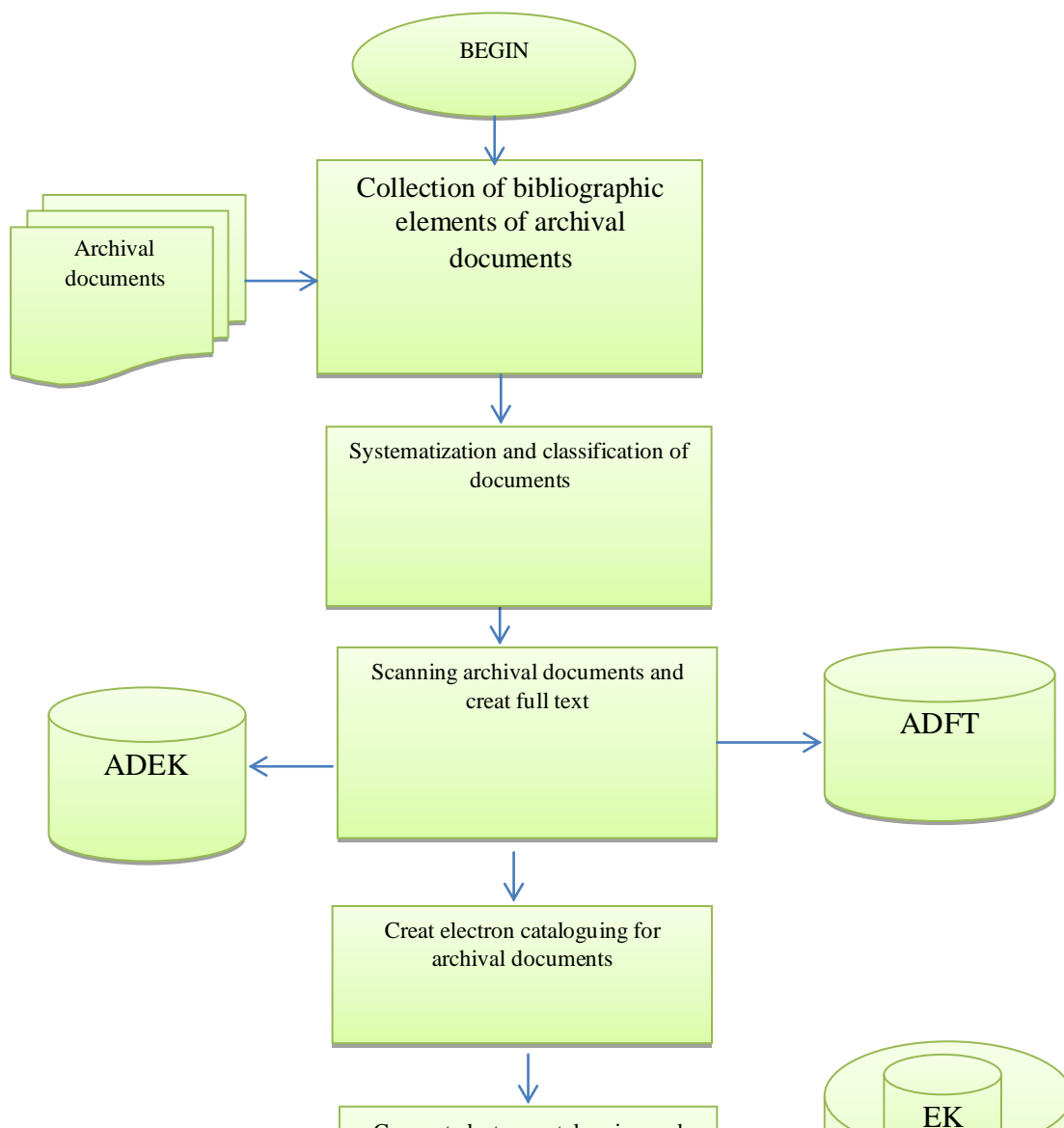
In order to present the information models and algorithms for processing of archival documents, we introduce some basic concepts.

Information model - object model presented in the form of information that describes essential for the consideration of options and variables of the object, the connection between them, the inputs and outputs of the object and allows by applying a model of information about changes of the input variables to simulate the possible states of an object.

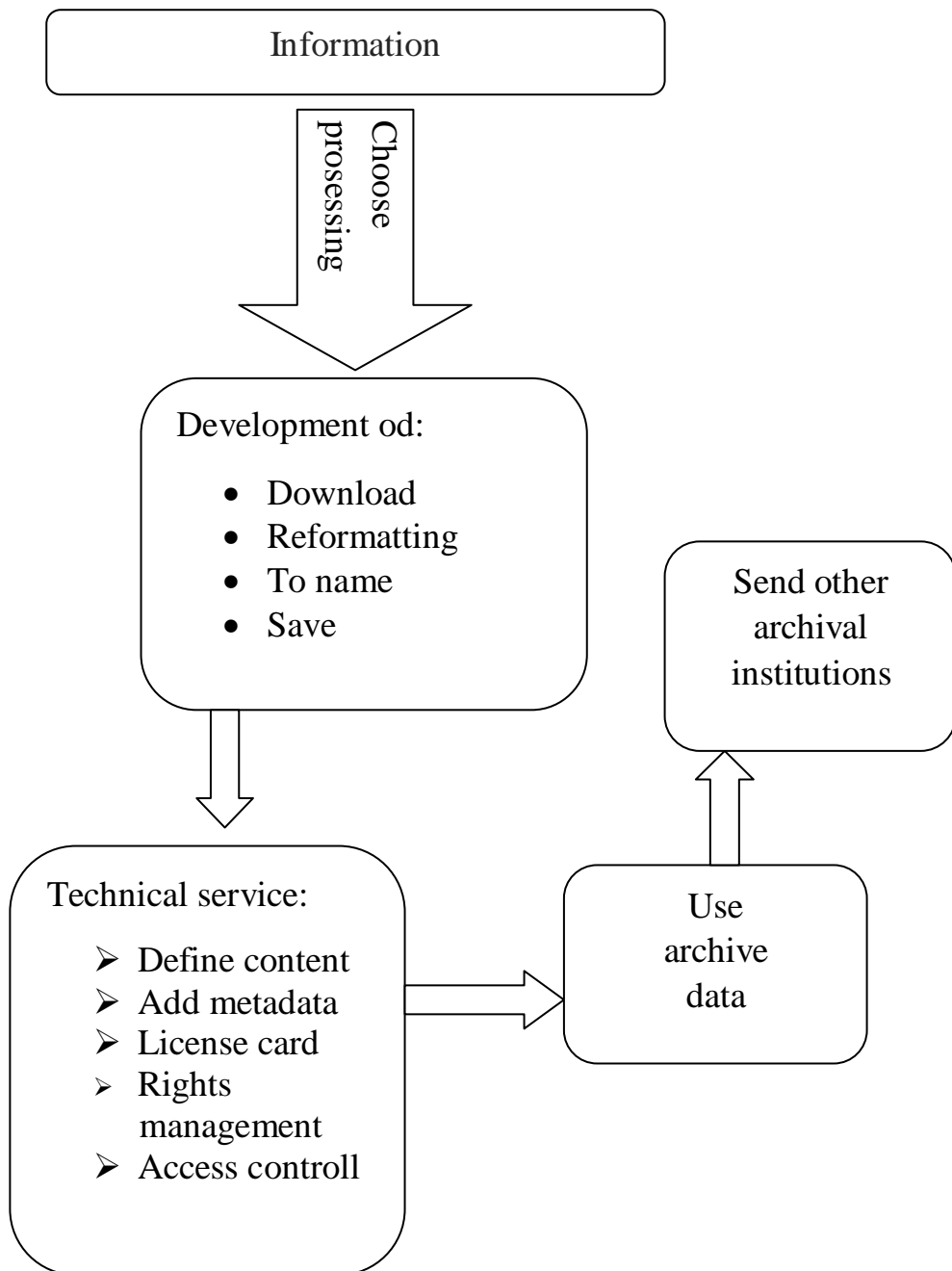
Algorithm - a set of instructions that describe procedures artist to achieve the result of solving the problem in a finite number of steps.

Archival documents - material carrier with fixed on it information, which has details that enable him to be identified, and shall be kept because of the importance of these media and information for citizens, society and state.

### **Information model of creat electron archive documents in archival institutions**



2.1. Block-scheme. Information model of creat electron archive documents in archival institutions.



2.2 Block-scheme. Information medel of exchange, collect and transmission of data in archival institutions.

The concepts of infological and datalogical data models were introduced by Langefors in a series of publications starting in 1963 <sup>4</sup>. Infological data models represent information in ways that are supposed to be similar to how people perceive the information (infological realm), without considering their final computer-related representations (data- logical realm). The ideal situation for an information system designer is to have a powerful infological data model that can be easily communicated between humans, and that there is a way to perform a non-loss translation of this infological data model into the datalogical realm.

### **Infological data model**

In the early theoretical work on infological data models, the concepts of object, property, relationship and time were identified as basic. An elementary fact is in this framework represented as a triple (a collection of objects + a property or relationship + time), called an elementary constellation.

Structured textual descriptions (natural language), formal logic (specification in for instance the logic-based programming language Prolog <sup>5</sup>) and other structural techniques (with visualisation through diagrams) have been proposed as infological data models.

Structured textual descriptions can express things in a human readable format, but have severe limitations when it comes to data structuring and formalisation for translation into the datalogical realm. Data modelling <sup>7</sup>

Logic has the advantage of being a formal description, having its roots in mathematics. It is therefore more easily translated into the datalogical realm. A problem is that logic lacks mechanisms for efficient communication of structure.

Diagrams have the advantage that they can show structure (relationships) in a human readable way (usually as two-dimensional maps), and diagrams have therefore become very popular for “semantic” data modelling. A problem with

---

4"Data Models" Dionysios C. Tschritzis, Frederick H. Lochovsky Prentice Hall, Inc., 1982

5""Programming in Prolog" W.F. Clocksin, C.S. Mellish Springer-Verlag, 1984

diagrams is that they can be difficult to translate into the datalogical realm, and there is a limit to the amount of information that can be put into a diagram without making it difficult to comprehend.

Semantic data models<sup>6</sup> introduce many useful methods for data structuring and abstraction, and constitute the most interesting branch of infological data models for database modelling. In this chapter, the ER model and an EER model are described to give a background in high level data modelling.

The entity-relationship (ER) approach (or ER diagrams, initially proposed by Chen has been the most popular diagrammatic representation for data modelling in the last decade. The expressiveness of the original ER model has been extended in many directions to capture more real-world semantics in the diagrams.

The latest direction in real world modelling for computer representation is the object-oriented approach. Object-oriented methods add encapsulation and behaviour to the traditional structuring mechanisms of semantic data models.

Infological carry out data analysis and define objects. The purpose infological simulation<sup>7</sup> - providing the most natural for the human ways of collecting and presenting the information that is supposed to be stored in the created database. Therefore infological need to build a data model similar to natural language (the latter can not be used alone because of the complexity of computer word processing, and any ambiguity of natural language). Define the main structural elements Infological model: Essentially, the connection between them and their properties (attributes).

The analysis of documents flow in the organization of the following basic essence of the model have been identified: the document and the employee.

---

<sup>6</sup> "Semantic Database Modeling: Survey, Applications, and Research Issues" Richard Hull, Roger King ACM Computing Surveys, vol.19, no.3, Sept. 1987, pp. 201-260

<sup>7</sup> Инфологическая модель базы данных

[http://www.lapshin.biz/6778/infologicheskaja\\_model\\_bazy\\_dannyh.php](http://www.lapshin.biz/6778/infologicheskaja_model_bazy_dannyh.php)

Special attention should be paid to the essence of "worker". Since every employee can have different attitudes to the document, it is necessary to distinguish two subentity: the sender and the recipient.

Consider the attributes of certain entities:

1. Document. It is the core essence of the system and has the following attributes:

- author (sender, recipient) of the document;
- the date of creation / sign the document;
- period of execution;
- importance;
- Status of implementation;
- a type;
- document file;
- Document the parent (this attribute is optional, is used to track the responses of the tree).

2. Worker. It is also the core essence.

- Full Name;
- unit;
- status within the division.

The relationship between the identified entities are displayed in the form of the circuit in Fig. 4. It clearly shows the difference between subentity "receiver" and "sender".

### **Datalogical model**

After building the model Infological exchange electronic document management system is considered prospective database structure.

Under datalogical understood model<sup>8</sup> that reflects the logical relationships between data elements, regardless of their content and the physical organization.

---

<sup>8</sup>Даталогическая модель данных (базы данных)

[http://ru.wikipedia.org/wiki/%D0%94%D0%B0%D1%82%D0%B0%D0%BB%D0%BE%D0%B3%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F\\_%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D](http://ru.wikipedia.org/wiki/%D0%94%D0%B0%D1%82%D0%B0%D0%BB%D0%BE%D0%B3%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D)

This datalogical model is developed taking into account the specific implementation of the database, also taking into account the specifics of a particular domain on the basis of its Infological model.

For the purposes of the content of the information highlighted in the following table:

- Document - the table contains all the attributes of the essence of the "Document", as well as additional parameters, the use of which was required in the development process;



All the tables included in the database are given to the third normal form, which avoids excessive storage of information and access to data is not in their values, and on the identifier.




Many different lower level data models (more closely tied to the datalogical realm) have been used through the years. They are by definition computer oriented, but the evolutionary trend of these data models is that they are approaching infological data models in expressiveness. The first low level data models from the 1950s and 1960s were based on simple file and record structures. Beginning in the late 1960s, there has been an “evolution” of the datalogical models, starting with the hierarchical data models and continuing with network data models and relational data models. In the last decade the object-oriented data models have been proposed. Reaching object-oriented data models, the distinction between the datalogical and the infological realm is getting fuzzy. Object-oriented models are claimed to cover both the infological and the datalogical realm, being directly implementable through object-oriented database systems.

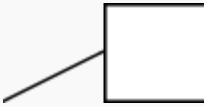



As datalogical data models are approaching infological data models in expressive power and sophistication, their implementation is becoming more and more complicated.


### 2.3. Algorithm of exchange information in archival institutions

In mathematics and computer science, an **algorithm** ( i/ˈælgərɪðəm/ AL-gə-ri-dhəm) is a self-contained step-by-step set of operations to be performed. Algorithms perform calculation, data processing, and/or automated reasoning tasks. An algorithm is an effective method that can be expressed within a finite amount of space and time and in a well-defined formal language for calculating a function. Starting from an initial state and initial input (perhaps empty), the instructions describe a computation that, when executed, proceeds through a finite number of well-defined successive states, eventually producing "output" and terminating at a final ending state. The following are some of the commonly used shapes used in flowcharts. Generally, flowcharts flow from top to bottom and left to right.

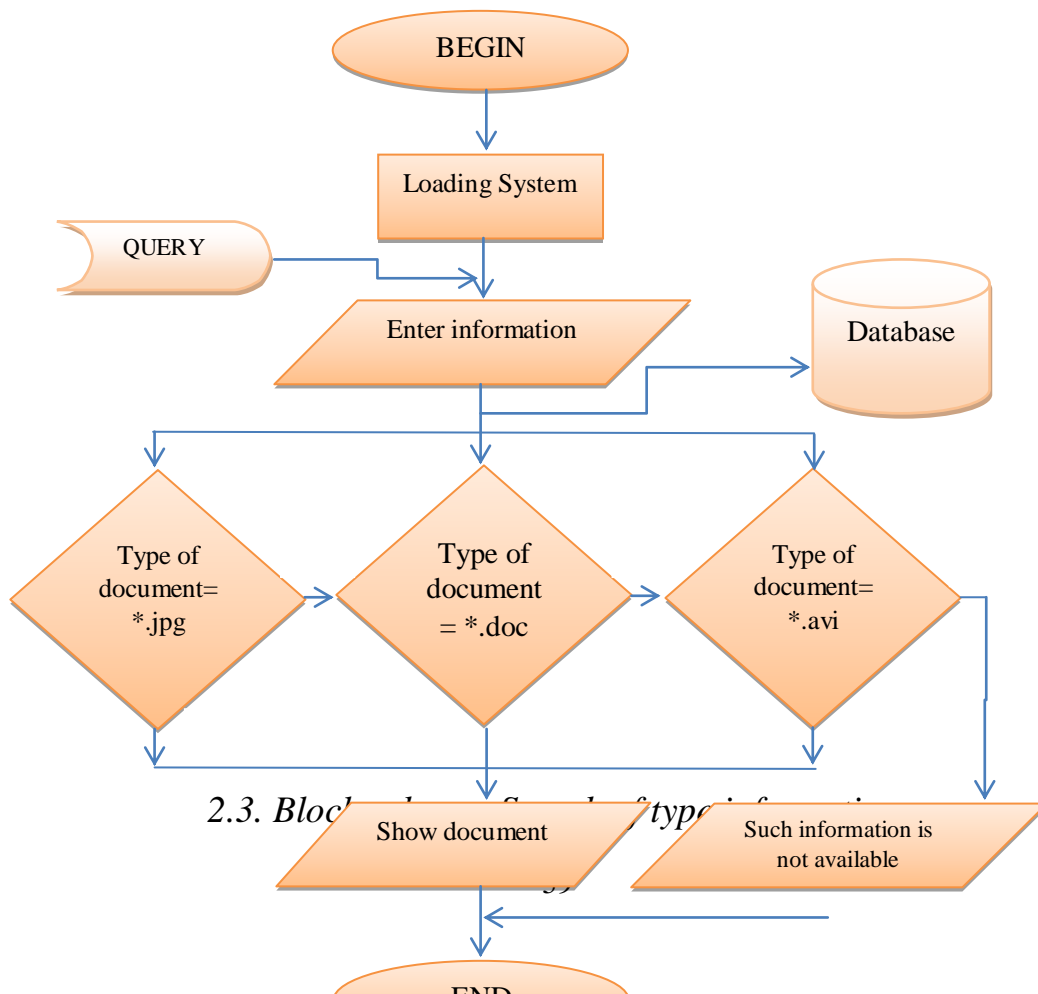
Shape	Name	Description
	Flow Line	An arrow coming from one symbol and ending at another symbol represents that control passes to the symbol the arrow points to. The line for the arrow can be solid or dashed. The meaning of the arrow with dashed line may differ from one flowchart to another and can be defined in the legend.
	Terminal	Represented as circles, ovals, <u>stadiums</u> or rounded (fillet) rectangles. They usually contain the word "Start" or "End", or another phrase signaling the start or end of a process, such as

		"submit inquiry" or "receive product".
	Process	Represented as <u>rectangles</u> . This shape is used to show that something is performed. Examples: "Add 1 to X", "replace identified part", "save chanGES", etc....
	Decision	Represented as a diamond ( <u>rhombus</u> ) showing where a decision is necessary, commonly a Yes/No question or True/False test. The conditional symbol is peculiar in that it has two arrows coming out of it, usually from the bottom point and right point, one corresponding to Yes or True, and one corresponding to No or False. (The arrows should always be labeled.) More than two arrows can be used, but this is normally a clear indicator that a complex decision is being taken, in which case it may need to be broken-down further or replaced with the "predefined process" symbol. Decision can also help in the filtering of data.
	Input/Output	Represented as a <u>parallelogram</u> . Involves receiving data and displaying processed data. Can only move from input to output and not vice versa. Examples: Get X from the user; display X.

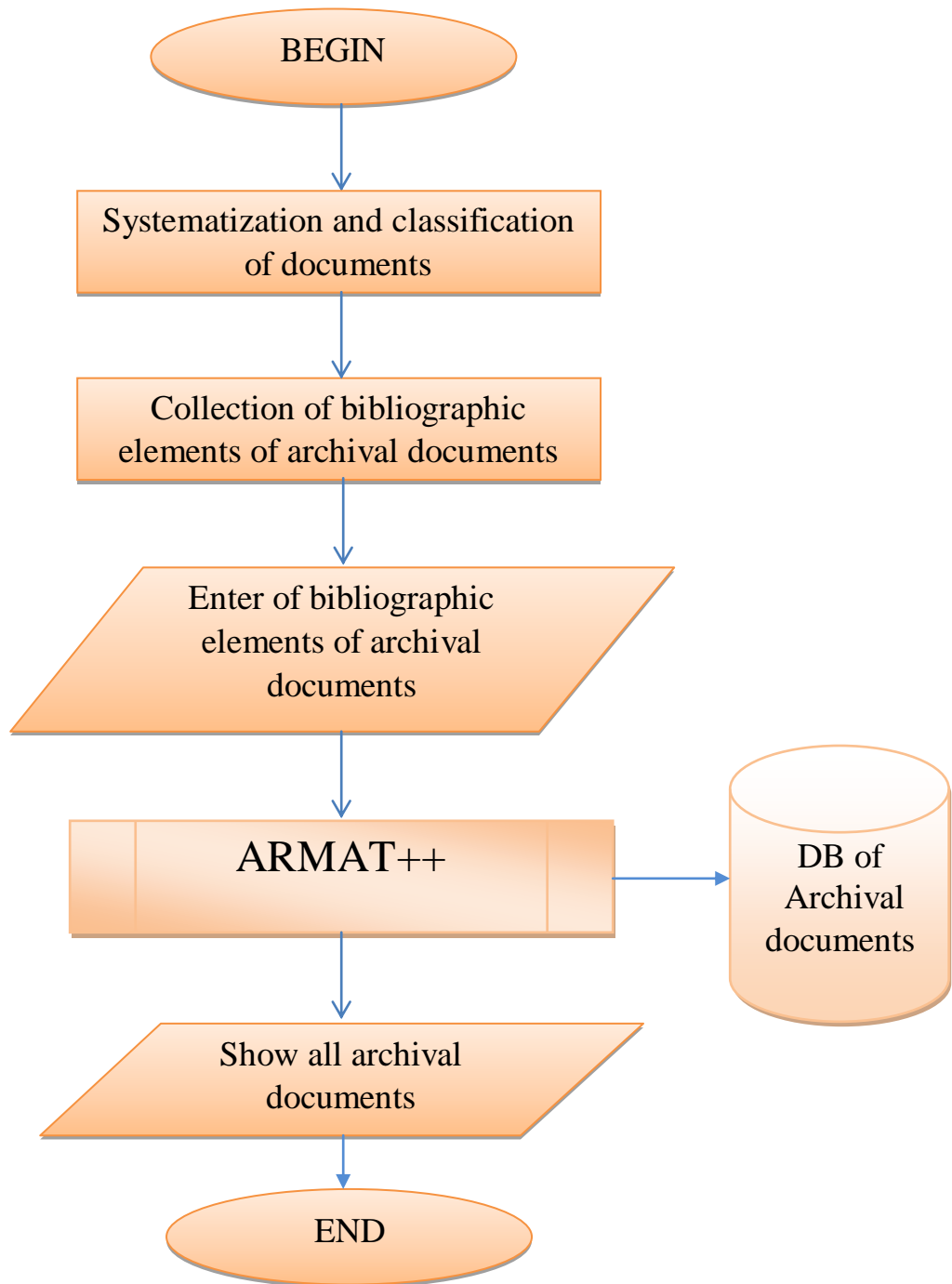
	<p>Annotation</p>	<p>Annotations represent comments or remarks about the flowchart. Like comments found in high-level programming languages, they have no effect on the interpretation or behavior of the flowchart. Sometimes, the shape consists of a box with dashed (or dotted) lines.</p>
	<p>Predefined Process</p>	<p>Represented as rectangles with double-struck vertical edges; these are used to show complex processing steps which may be detailed in a separate flowchart. Example: PROCESS-FILES. One subroutine may have multiple distinct entry points or exit flows (see <u>coroutine</u>). If so, these are shown as labeled 'wells' in the rectangle, and control arrows connect to these 'wells'.</p>
	<p>Preparation</p>	<p>Represented as a <u>hexagon</u>. May also be called initialization. Shows operations which have no effect other than preparing a value for a subsequent conditional or decision step. Alternatively, this shape is used to replace the Decision Shape in the case of conditional looping.</p>
	<p>On-Page Connector</p>	<p>Generally represented with a circle, showing where multiple control flows converge in a single exit flow. It will have more than one arrow coming into it, but only one going out. In simple cases, one may simply have an arrow point to</p>

		<p>another arrow instead. These are useful to represent an <u>iterative</u> process (what in Computer Science is called a <u>loop</u>). A loop may, for example, consist of a connector where control first enters, processing steps, a conditional with one arrow exiting the loop, and one going back to the connector. For additional clarity, wherever two lines accidentally cross in the drawing, one of them may be drawn with a small semicircle over the other, showing that no connection is intended.</p>
	<p>Off-Page Connector</p>	<p>Represented as a <u>home plate-shaped pentagon</u>. Similar to the on-page connector except allows for placing a connector that connects to another page.</p>

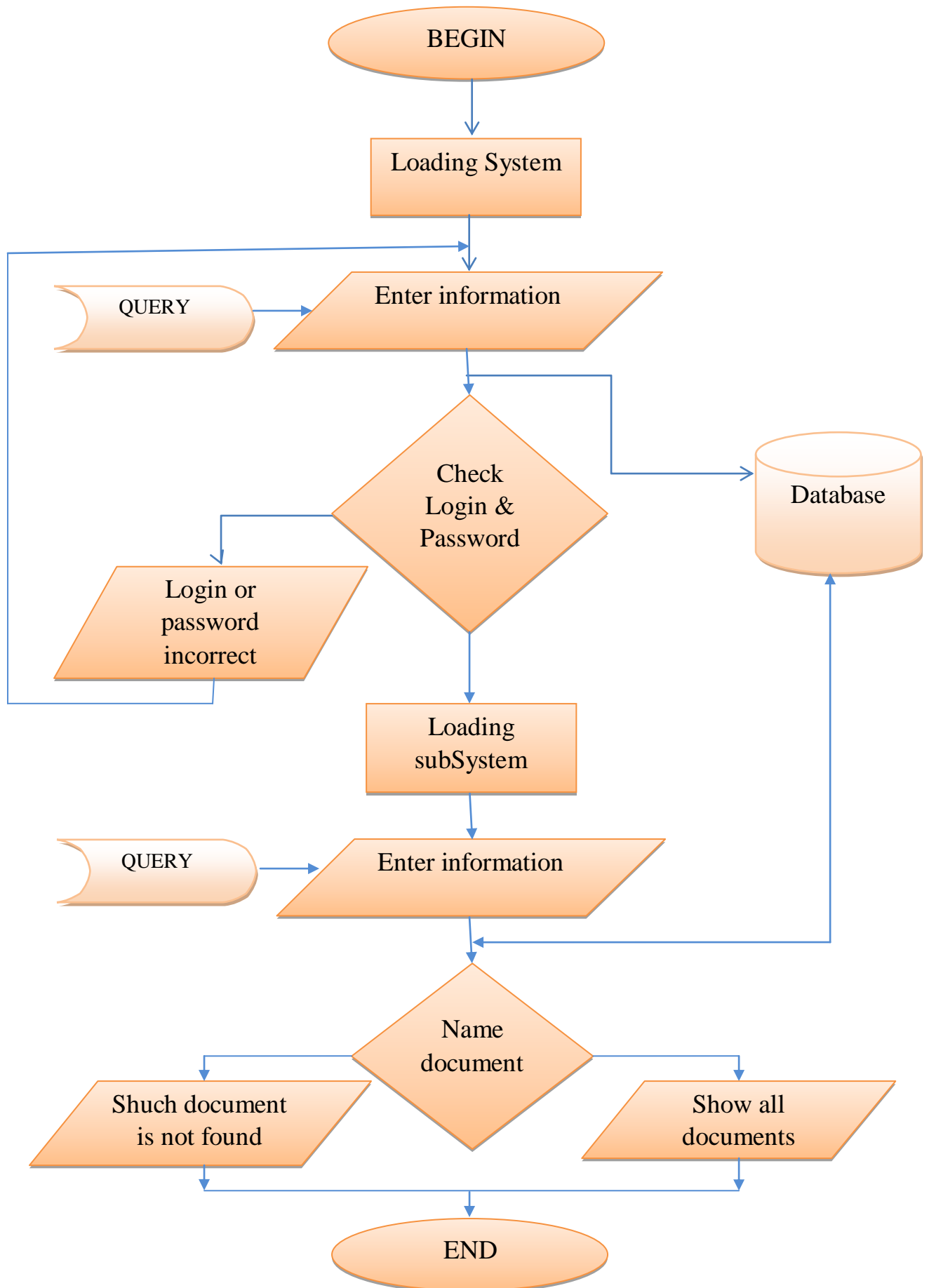
**SEARCH OF TYPE INFORMATION.**



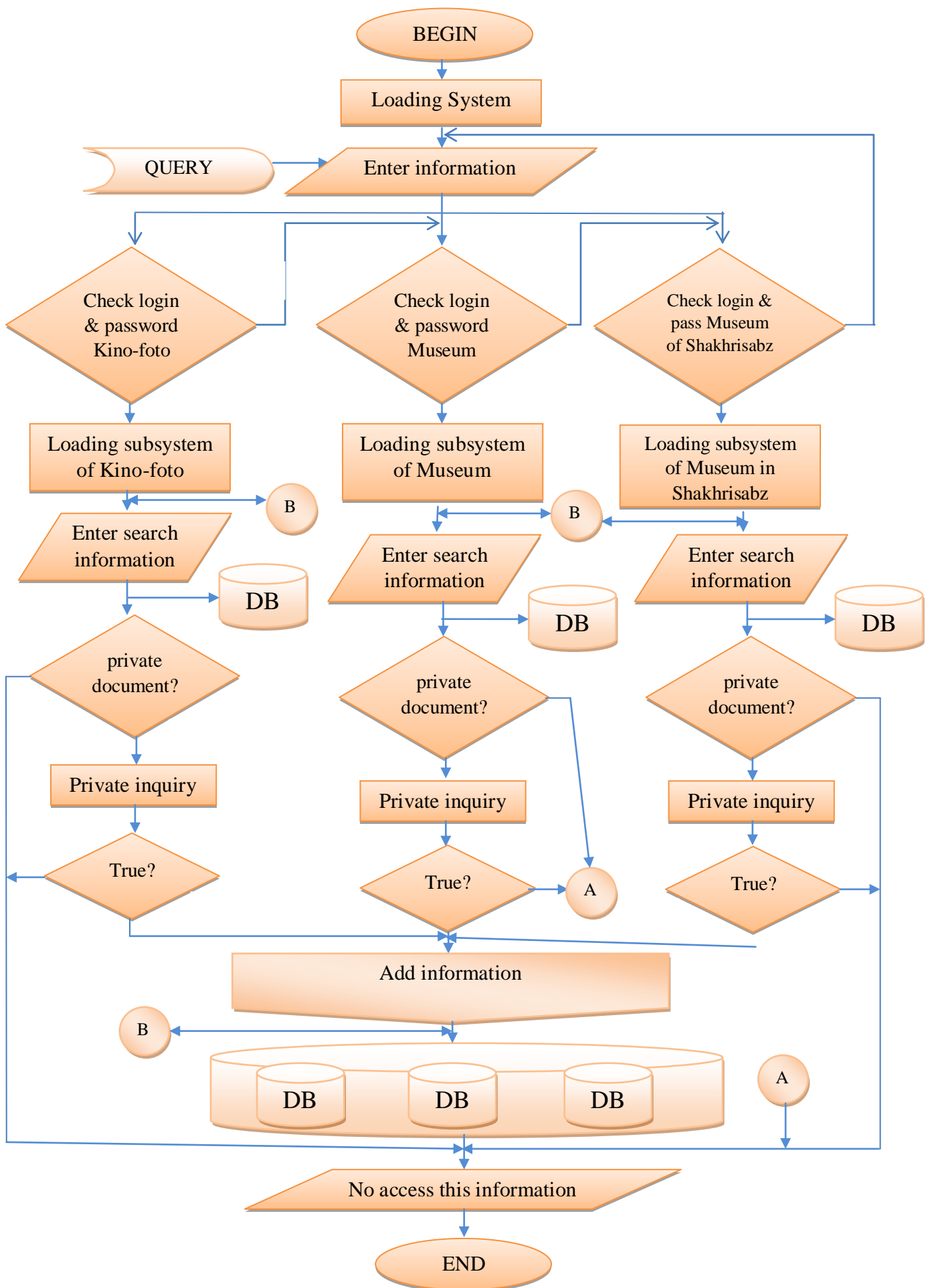
## **Creata Database for archival documents.**



2.4. Block-scheme. Creat Database for archival documents.



2.5. Block-scheme. Authorization and search document.



## *2.6. Block-scheme.Exchange information in archival institutions*

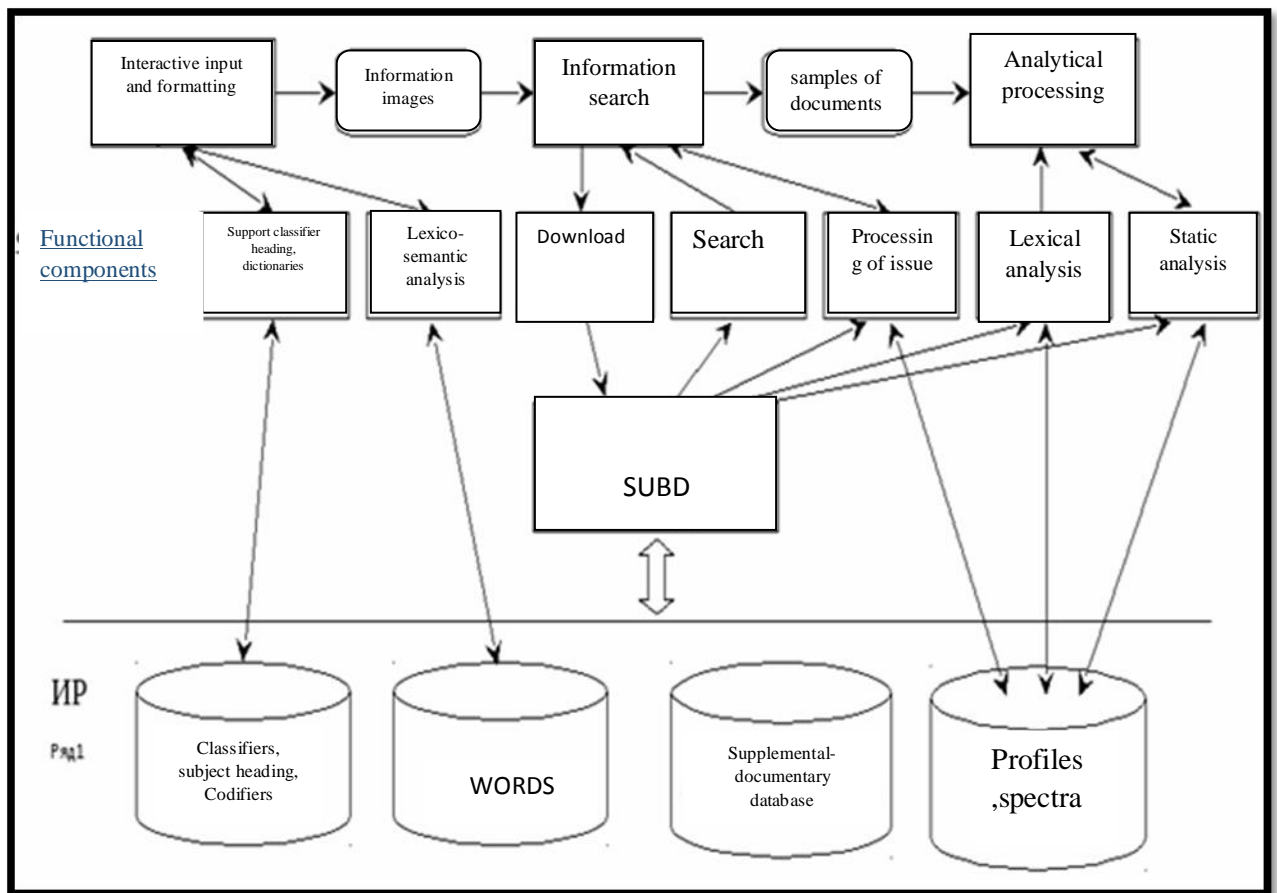
### CONCLUSIONS

1. Define general concepts about decision-making model. Also, an analysis of methods and tools for modeling complex objects leads to the conclusion that the expansion of opportunities for implementation of models is greatly extended and the range of problems solved by simulation methods, which is inherent in the decision-making problem.
2. The functional and organizational structure of the agency "Uzarhiv".  
Painted duties of each of the employees.
3. A scheme of tasks subsystems, electronic document archiving system institutions. The detailed description of the construction of each of the subsystems of the agency "Uzarhiv".

# CHAPTER 3. PROGRAMMING AND ORGANIZATIONAL SUPPLY OF EXCHANGE INFORMATION IN ARCHIVAL INSTITUTION

## 3.1 The functional structure of the exchange electronic document management system for archival institutions

The functional structure of the system determines the composition of the sub-systems and data communication between them, and also serves to improve the efficiency of the research object and the relationship between its components. (Block diagram 1)



*1. Block diagram of functional structure of electronic document management system for archival institutions.*

### Description of the subsystems

Electronic document archival institution supports three levels of cognition: the external, responsible for communicating with the external environment; interior in communication between subsystems and including

processing and storage of all kinds of documents; administrative - the level of support and decision-making.

In accordance with the functions performed, the system consists of the following nine subsystems.

#### Registration and calculation of results

State registration and inventory consists of several steps:

- identification of the organization - the manufacturer of the document and enter the incoming number and date of registration in the e-zine. In the absence of information about the organization - the manufacturer of the document in the electronic journal it entered information about the new organization, and then the data of the document,
- State registration and inventory taken of documents received in the course of which cards are assigned numbers that are fixed to the respective documents,
- output registration, during which the electronic magazine of the registration number is entered and the inventory numbers corresponding to the incoming document number. Thereafter, the secondary documents are transmitted to the semantic processing.

As part of the subsystem you want to automate the following operations:

- preparation and reception of input documents,
- the registration and management of documents,
- prepare and send letters of notification about the registration carried out correction or the need to refine and clarify the information materials,
- Accounting registration status,
- control of the registration status.

#### **Processing of documents received**

Document processing comprises the following operations:

- **rubritcirovanie documents.** Rubritcirovanie involves assigning document the full list of headings used standard heading corresponding to the content of the document;

- indexing of documents. Indexing consists in attributing to the document a limited list of key words (up to 10), more fully and accurately reflect the content of the document;

- edit text documents and details of individual documents from the standpoint of control of spelling and stylistic literacy;

- create digital forms of documents if the documents received on paper. Full keyboard input secondary documents to your computer. The result is an array of documents in digital format and print it on the printer for further processing;

- **readability proofreading** documents print after keyboard input documents;

- Input proof in the documents submitted in a digital format.

As a result, the semantic processing operations formed an array of current receipts (ICC) secondary documents, ready for transfer to the electronic archive and upload in a retrospective database.

Because the subsystem "Registration and calculation of results" to the input subsystem receives the following documents:

- a registration form
- information card (IC).

The "secondary documents Processing" subsystem they are unified processing and transferred to the subsystem "Archiving documents", "publication" and "Management".

### **Archiving documents**

The subsystem "Archiving documents" is designed to create and support of scientific and technical information assets received by the organization.

Documents in the subsystem "Archiving documents" comes from the sub-systems:

- Registration and stocktaking of the documents;
- Processing of secondary documents;

The subsystem "Archiving documents" performs the following functions:

- verification of the completeness of the documents received,
- Maintaining the current array of documents received,
- conducting annual arrays of documents received from the recording on CD,
- Maintaining the registry and fund organizations that submitted their documents

Because the subsystem "Archiving documents" documents arrive in the subsystem "Keeping database" and "User Maintenance."

#### Maintaining databases

The subsystem "Database Management" is intended for conducting and supporting historical database of received documents and information is the main component of the system under development. The documents and information to the subsystem "Keeping abstract database" comes from the subsystems:

- Registration and stocktaking of the documents,
- Processing of secondary documents;
- Document Archiving.

Storage is carried out in a medium document-oriented database that will, along with the construction of quantitative characteristics, to conduct a meaningful analysis based on content analysis of accounting documents.

Document submission tools - search results, allow the use of a variety of output formats, sort and Rubritsirovanny set of found documents and perform a spot (for groups of records - previously obtained search results on the chosen theme) analytical and statistical processing for the construction of direct or indirect estimates. Because the subsystem "Keeping abstract database" documents arrive in the subsystem "User service" and "management".

## User Service

The subsystem "User Service" is designed to provide access to archival collections stored in the agency "Uzarhiv". The documents and information to the subsystem "User Service" comes from the sub-systems:

- Maintain databases,
- Maintaining linguistic support,
- Document Archiving.

The subsystem "User Service" supports two modes of operation with resource users:

- Remote, implemented through the Internet,
- Local supporting the work of the reading room users.

subsystem performs the following functions for the implementation of remote access:

- creation and support of Web-site
- Providing abstracts database users,
- accepting orders for thematic search and their implementation specialists

organizations

- Support for e-mail and conduct correspondence.

To implement local mode subsystem performs the following functions:

- organization of the reading room of the library in the classical form,
- Access to the abstract databases and electronic stores,
- accepting orders for the execution of thematic research specialists

organizations

- Search for a thematic organization of customized,
- transfer of the executed orders to users.

Because the subsystem "User Service" information is delivered to the subsystem "Management".

### Information-analytical subsystem

Implementation of information and analytic functions must be based on the use of the concepts of data warehousing. Analytical data processing based on

the thematic classification of texts, extracting characteristic concepts, allocating connections in the text. The selected objects and relationships are compared with the corresponding classifications; a certain class of objects communicate the appropriate type of relationship.

The information-analytical system is implemented on the basis of modern computer and communication technologies registration, accounting, analytical and statistical processing and storage of accounting documents through the use of database technology with advanced retrieval apparatus, a network of electronic document management and intelligent content analysis provides the following tasks:

- Provision of statistical references (current and for the reporting period) on the performed tasks;
- Formation of the various thematic and statistical spectra and time series, characterizing the direction, effectiveness and level of scientific research;
- Identify scientific relationships in the domain of a particular direction.

Information-analytical subsystem allows operatively and to change the conditions of selection and aggregation of data, the data management system meets the following requirements:

- Support for high-speed data acquisition;
- support internal consistency of the data;
- the possibility of obtaining and comparing the data slices;
- the availability of convenient application for viewing data;
- completeness and accuracy of the data stored
- Support data replenishment process.

Basic requirements for the automated preparation and processing of data:

- adaptability to training;
- versatility of use.

The documents and information received from the subsystem subsystems:

- Registration and accounting of results,
- Processing of secondary documents;

- Archiving documents
- Maintain databases

Because the subsystem "Information-analytical" documents arrive in the subsystem "Management".

#### Subsystem conduct linguistic support

The subsystem reference linguistic software implements the construction and maintenance of terminology systems and rubricators used, including, for problem-thematic structuring of the domain based on the automatic categorization of documents. For ordering can be used as a conventional imported or domestic industry classification system and specially created.

Because the subsystem "Keeping linguistic support" information received in the subsystem "Database Management" and "User Service", "Information and analytical."

#### Management

The subsystem "Management" is designed for the collection, storage and processing of all incoming documents, staGES of processing and the requested. Information subsystem "management" comes from the sub-systems:

- Registration and profitability analysis;
- Processing of secondary documents;
- Maintain databases;
- Service users;
- Information-analytical
- Maintaining linguistic support,
- Document Archiving.

The subsystem "Control" performs the following functions:

- collection of information and support of the database of statistical information on incoming documents, staGES of processing and the requested,
- Preparation of weekly progress reports for receiving and processing incoming documents,

- collect and store information about the user orders the stage of their implementation, and the funds received by,
- preparation of orders and user reports.

The subsystem "Management" prepares and provides information to the management company and the parent bodies.

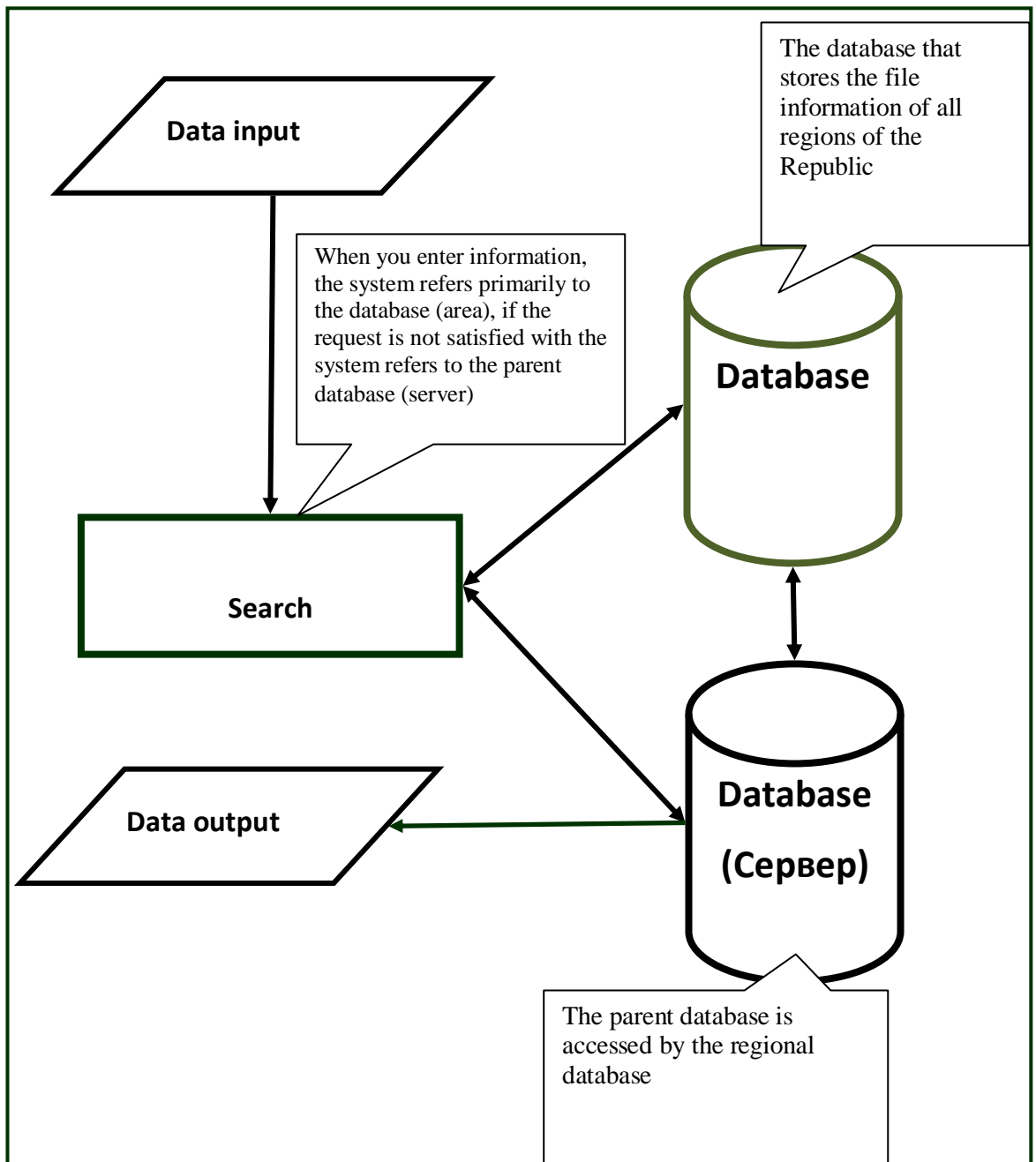
#### Administration

Subsystem "Administration" is used for the support and development of corporate network infrastructure. The subsystem must provide transportation, hardware and system components of the electronic document management system. The purpose of this subsystem is:

- Develop and support a cable network;
- securing funds network environment and different user groups;
- Use load control the operating status of hardware, etc.

Subsystem "Administration" is a component which provides for any of the subsystems of the electronic document management system.

The functional structure of the electronic document, implemented by the system shown in Fig. 5.



*Fig. 5. The functional structure of the electronic document management subsystem*

Subsystems - is an integral part of the system, which has independent functional significance in the framework of the objective system.

Divide the electronic document archival institutions in the sub system:

- Data input
- Search
- The database (server)
- Database (Province)

- Conclusion

#### Data input

After logging in to the Block "input data", you must enter the information for data retrieval or to obtain information for further work.

Entering data such as features, such as messaging, categories of users, including some mandatory fields, as well as the procedures performed on documents.

After logging into the system unit "Messaging", if you can check received from someone new message.

This software includes two major subtypes documents:

- Draft document - handwritten or typewritten document reflecting the work of the author or editor of the text.
- The saved document - a document stored in the database.

Developers provided a method for rapidly adding a large number of users. If you think that in your case it is necessary, you must create a text file of a certain format: the first line contains a list of field names that are introduced or changed, the user data are located in other lines.

The field names and data are separated by commas, or other homogeneous symbols. The use of all possible fields is unnecessary. There are required fields, optional and custom.

- For each operation the following fields must be used:
- To add users: name, surname (firstname, lastname)
- ChanGES to the login name (username)

#### Role

- determines the status of the user in a certain context. For example, the chairman, coordinator, etc.

#### Opportunity

- a description of one characteristic feature of a particular element of the system. For example, Task: Viewing Forum: Document management. For each element of the system identified a lot of opportunities.

## Resolution

- value set by a certain potential for a specific role. For example, to allow or deny.

Searching documents stored in the database. Use the search bar, you can select a phrase to search for, select the necessary documents and customize the search mode and return to the previous search assignments. Perhaps the only restriction of the use for the administrator and the director of the archival institution. A prerequisite for the implementation of the document search, it is necessary to documents have an explanatory note, and filled key concepts.

## Database

The database is presented in the form of an objective set of independent materials, systematized so that these materials can be found and processed with the help of an electronic computer (PC)

Database (Server) - keeps the full-text information on the status of documentation of all archival institutions, including the regional divisions, all employee data archival institutions, as well as information on the status of documentation and resource center resources.

Database (Province) - stores information about the subdivisions of archival institutions in each region and the state archive fund.

## data output

After completion of the system, get all the information about the projects of orders, Focal Point, on the status of the document library, and other elements and resources, you can display certain information on printing or form in any form, such as a report, table or in the form of a diagram.

### **3.2 Organizational structure of the electronic document management system for archival institutions**

Organizational Structure<sup>9</sup> - A set of ways in which the labor process is first divided into individual work tasks, and then achieved the coordination of action to address the problems (Henry Mintzberg, "The structure of the fist"). In fact, the organizational structure determines the allocation of responsibility and authority within the organization. As a rule, it appears in the form of organizational charts (English organigram.) - Graphical design, elements of which are hierarchically ordered organizational units (departments, job position).

Organizational structure - a scheme consisting of units and individual officers of the company, located at the level of significance and responsibility of containing the relationship between them and the chain of command.

A well-built organizational structure makes it possible to optimize the number of employees and the number of units, to simplify the interaction between departments, to evenly distribute the load on the staff, to avoid duplication of functions and their "slack", to eliminate the double and triple subordination, to delineate the scope of activity of heads, determine their powers and areas of responsibility, improve productivity. The organizational structure is the basis for building an effective management system.

Building organizational structure comprises the following steps:

Company operation analysis (type of activity, stage of development, the main business processes, strategic and tactical objectives, profitability, management system);

Assessment of the existing functions of the distribution system:

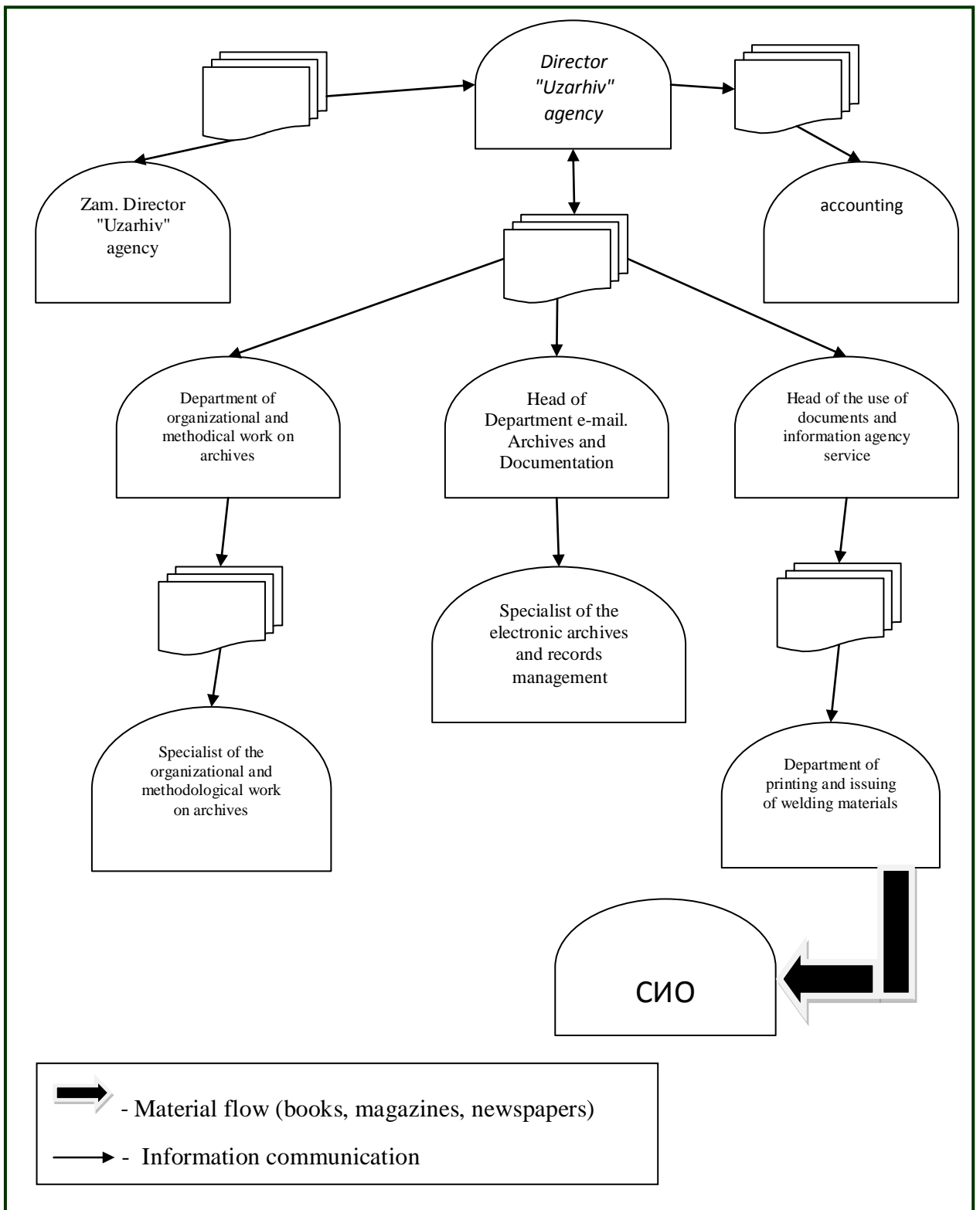
- analysis of staffing and the principle of separation of structural units;

---

<sup>9</sup>Организационная структура

[http://ru.wikipedia.org/wiki/%D0%9E%D1%80%D0%B3%D0%B0%D0%BD%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F\\_%D1%81%D1%82%D1%80%D1%83%D0%BA%D1%82%D1%83%D1%80%D0%B0](http://ru.wikipedia.org/wiki/%D0%9E%D1%80%D0%B3%D0%B0%D0%BD%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D1%81%D1%82%D1%80%D1%83%D0%BA%D1%82%D1%83%D1%80%D0%B0)

- analysis of the basic functions, powers and performance indicators for top and middle management;
- analysis of the basic functions, objectives and performance indicators of structural units and performers (*ex officio*);
- detection of deviations in the distribution of functions and subordination
- Making recommendations on the optimal number of staff;
- Construction of the organizational structure;
- Develop an implementation plan of the organizational structure and chanGES in the work of the staff;
- Conducting information events for staff.



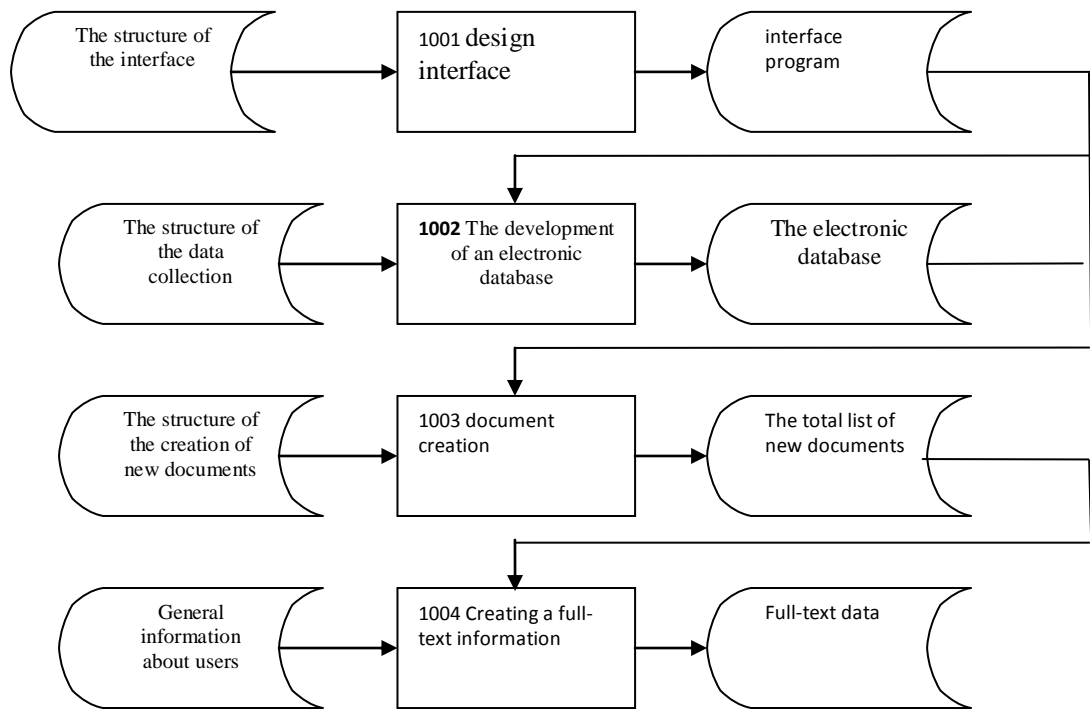
*Block - diagram 2. The organizational structure of the agency electronic document system "Uzarhiv"*

### 2.3 Schemes tasks subsystems, systems document archival institutions

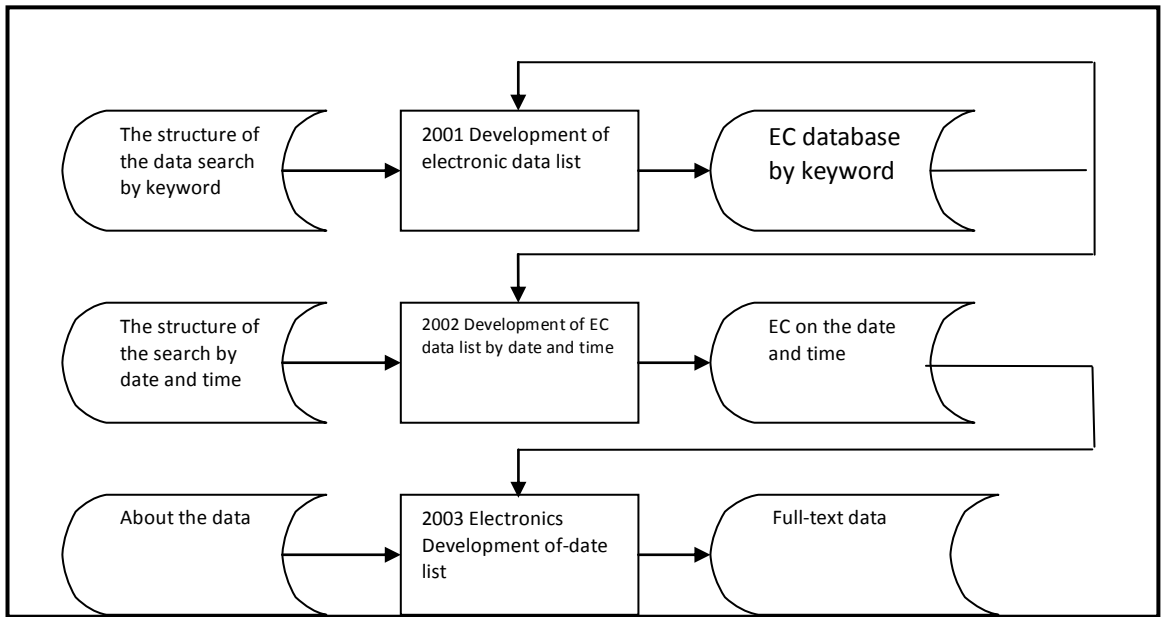
Driving tasks document archival institutions subsystems is a logical structure interaction problems subsystem Research (functions), and also reflects the (track) stream input and output data.

The problem is the lowest stage of decomposition system

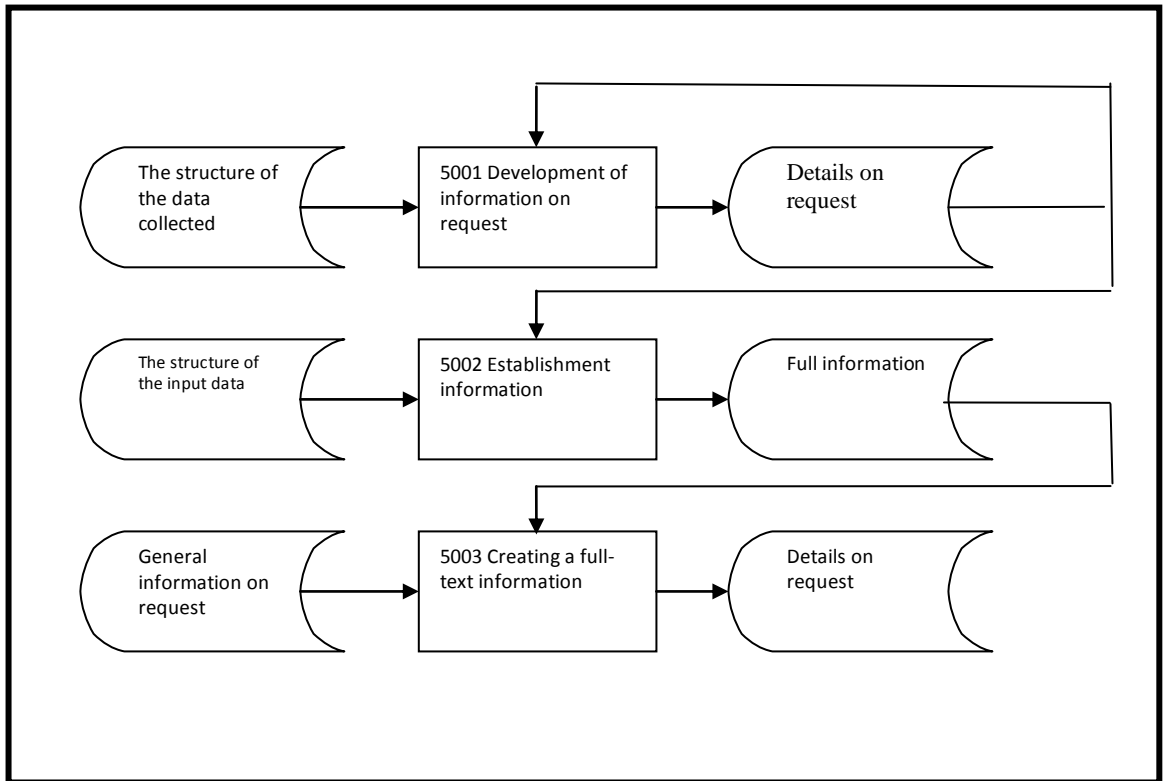
#### 1000. Driving subsystem tasks "Data Entry"



#### 2000. Driving tasks subsystem "Search"



### 3000. Driving subsystem tasks "Data Output"



### 3.3 User guide for use programming for exchange between archive

We will use ARMAT++ (library management software) for catalogization of film and photo. Because this system is compatible for these kind of documents. Also its database structure is based on MARC21 metadata standard. In this guide step-by-step catalogization process is explained.

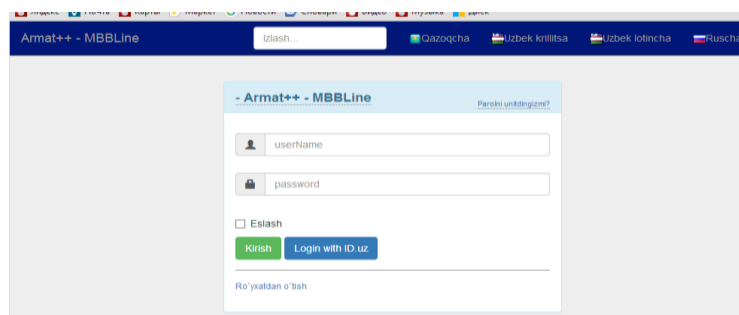


Figure 3.6. Main page of software interface

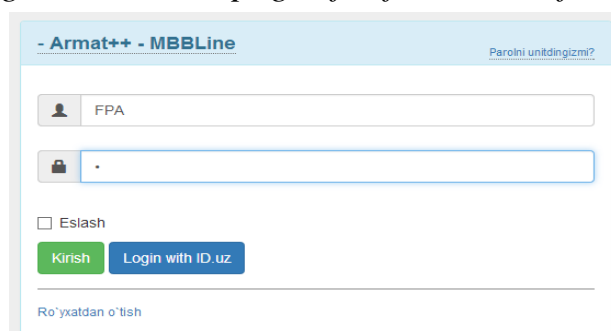


Figure 3.8. Sign into catalogizator workspace

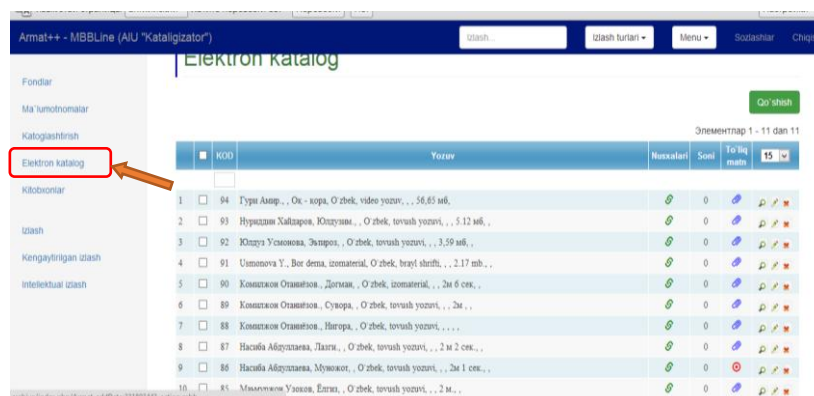


Figure 3.9. Open en electronic catalog page and click on 'Add new record' link.

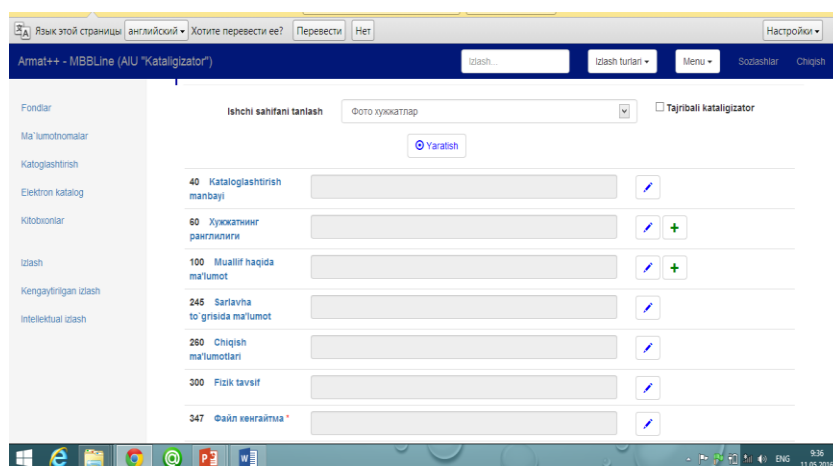


Figure 3.10. Choose the template is 'Catalogization' and start describing the item.

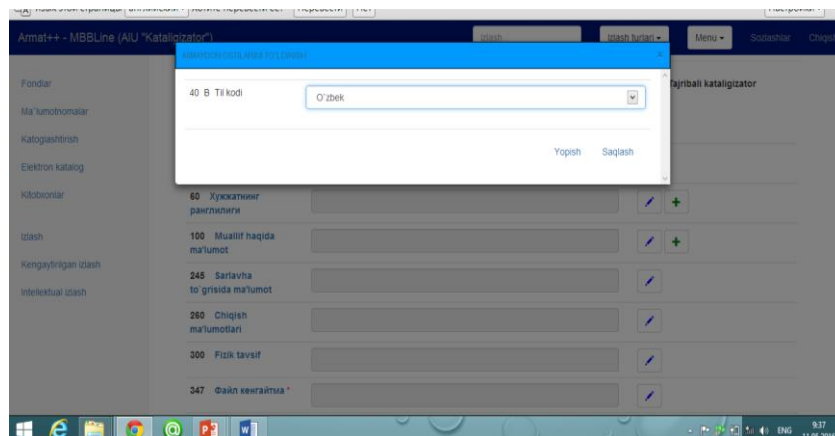


Figure 3.11. Field 040: catalogization source and language.

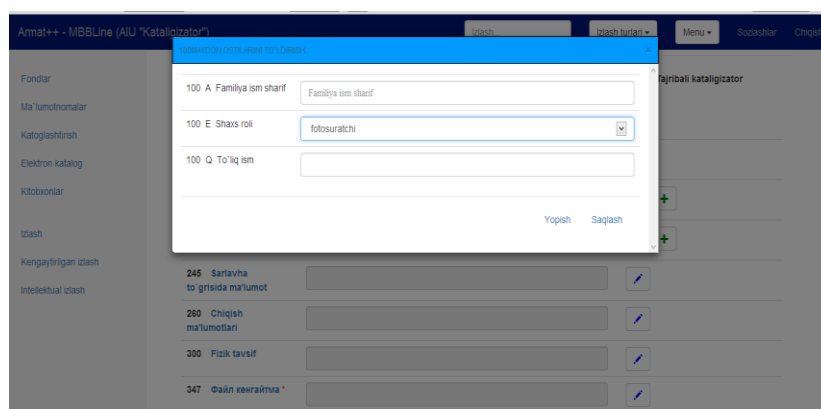


Figure 3.12. Field 100: Main name of creator - photograph.



YUVAN RAQIDA YOKI RAQIM	MAYDON	OSTI	MA'LUMOTI
40	B	O'zbek	
160	A	On - on	
245	A	G'uri, Asim	
245	H	video ushuv	
300	A	56,65 so	
300	E	2 cha	
420	A	G'uri, Asim	
653	A	G'uri, Asim	
927	A	Horayona	
3472	A	avi	

Figure 3.24. Record on catalog.

Archive facilities in one of the documents included in the electronic catalog, and then they receive data.

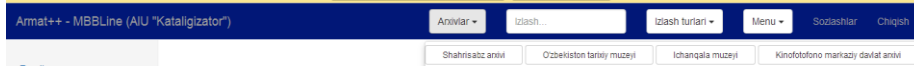


Figure 3.24. exchange between for archive.

## **The conclusion of the third chapter**

In this chapter the following works were done:

1. define general concepts about decision-making model. Also, an analysis of methods and tools for modeling complex objects leads to the conclusion that the expansion of opportunities for implementation of models is greatly extended and the range of problems solved by simulation methods, which is inherent in the decision-making problem.
2. The functional and organizational structure of the agency "Uzarhiv". Painted duties of each of the employees.
3. A scheme of tasks subsystems, electronic document archiving system institutions. The detailed description of the construction of each of the subsystems of the agency "Uzarhiv".
4. User guide of catalogization process was created and exchange between for archive instutions.