

# Criteria Definition for Implementing Digital Archives in Distributed Organizations - A Case Study

Katarina Tomičić-Pupek, Martina Tomičić Furjan, Željko Dobrović  
Faculty of Organization and Informatics  
University of Zagreb

[ktomicic@foi.hr](mailto:ktomicic@foi.hr); [mtomicic@foi.hr](mailto:mtomicic@foi.hr); [zdobrovic@foi.hr](mailto:zdobrovic@foi.hr)

**Abstract.** *While executing business operations organizations generate and use documents. Management systems for handling these documents depend on specific demands of every organization and the way the organization communicate with other business partners. Under specific demands we may consider availability of documents in their original or other appearances, time needed for the document retrieval, document handling costs and other criteria that effect on the document management systems. The aim of lowering the costs of document handling organizations is implementing various document management systems, especially when documents are considered as working assets. Contemporary information and communication technologies enable “paperless” work, by implementing digital archiving systems that store documents temporarily or permanently.*

*Before implementing such digital archiving systems it is necessary to develop the most suitable configuration which consists of available technical and technological solutions that could meet the specific demands of organizational needs for managing documents. This paper will show some suitable configurations for digital archives, define criteria for the selection of suitable technology for digital documents storage systems and evaluate the suitability of configurations considering the criteria for various organizational demands.*

**Keywords.** Electronic documents, configurations of digital archives, criteria for implementation technology

## 1 Introduction

The term electronic communication includes the exchange and the storage of documents in a digital form, and for the purposes of reducing the costs of document filing and managing. There are different technologies available for implementing a digital archive, but one of the problems inherent is the difficulty users have in understanding the scope and the nature of the archived materials [1]. The problems

lie also in medium preservation and technology preservation, in immense volume, instable storage media and obsolete hardware and software [3].

The basic benefits of a digital archive, or a digital document system, are that a process is done more easily, more quickly, with better quality and it is easier to find out about it afterwards [2].

To assess the cost of introduction and benefits associated with the new DMS is more complex than the literature of those companies that manufacture hardware and software components for the DMS would have us believe [5].

This paper defines the configurations and the criteria for selecting the appropriate technology for the storage of digital documents in spatially distributed organizations with the high level of decentralization of operative work. An example is given for an organization that comprises the central service that is responsible for direction and supervision over operative work execution in the field of organization activity, and for the associated branch departments where the professional tasks from the field of activity are done. By executing the operative work the greatest part of business documentation is generated (e.g. in public service administration). Business documentation is the subject of the book of rules about archives and about managing filed documents and it often stands as a burden instead of work resources.

In order to manage the business documents in an efficient way, the implementation of a document digital archive is being offered as a solution. A digital archive, as a part of an information system, has to be made according to the demands that arise from the optimized work execution of the organization. The quantity of digitalized information, which is used every day, is causing a need to collect and store information [4] in order to meet dynamic, changing and complex requests for data. For this reason, this paper shows the analysis of adequacy of four digital archive configurations depending on the organization needs that are shown through the criteria of the individual configurations selection.

## 2 Description of storage configurations in implementing digital archives

The organization for which the digital archive has been designed operates distributive with the high level of decentralization in operative work execution. The organization does the work in the public administration area by using the central service and branch departments. The central service defines the work guidelines, directs and regulates the work procedures and supervises the work of the branch departments. To execute these tasks the central service uses the documentation that is in a large part being generated by the workers of twenty branch departments in the execution of the operative work from the field of the organization activity.

In order to reduce the costs of document filing, by taking a long-term view, the organization shall introduce the digital archive of documents. The procedure of document digitalization has been defined, as well as all the procedures connected with the usage of digital archive documents. However, it is necessary to make a decision on the selection of information technology that shall be used to store the digital documents.

For this purpose, four configurations have been defined that are based on two possible ways of distributing two technologies:

1) of disc systems united in the depository of large capacity and

2) of WORM media (write once read many, optical medium) with the appropriate equipment.

Technology distribution implies the creation of digital archives in organizational units:

1) Digital archive installed in the central service, where documents will be sent to from the branch departments via the internet connection and that shall be used in the operative work in the same way.

2) Digital archive installed locally in branch services, which means that each branch department would have its own digital archive.

The configurations are individually described in details in the following text.

### 2.1 Configuration 1: The depository of large capacity in the central location

All the branch department entries are stored on a computer of large capacity (e.g. Disc Storage System IBM Total Storage DS8000 Series of 192TB capacity) in the central location.

The advantages of such storage are:

- Documents are recorded according to the principles of grouping that have been defined in advance and according to the folder structure.

- The number of accesses to the medium (of reading and writing) is not limited.
- The document recording medium is a physically inseparable entity (disc system) upon which the processing is done.
- Making a security copy means copying the contents of one medium.
- The capacity is sufficient for the central digital archive: with an average data quantity of 100KB per document, it is possible to record about 10.000.000 documents on 1TB, which means that it is possible to record about 2.000.000.000 documents on 192 TB.
- It is possible to get an on-line access to all documents, because the whole content of the depository is available through the communication server.
- There is no redundancy: all the entries are physically situated in only one place (except for the security copy).

The disadvantages of the central digital archive creation by using the disc systems are:

- All users compete for one resource: All the organization users are served by one central depository; so a larger number of servers are required, as well as a high performance communication server.
- Jammed communication paths are probable, which demands communication infrastructure strengthening.
- In the event of the central depository failure, not a single branch department can proceed with the work.
- Relatively high communication costs between the users and the central digital archive due to the transfer of relatively large data quantities.
- By taking a long-term view, the depository capacity is limited.
- A relatively high price at the beginning of implementation.

### 2.2 Configuration 2: Local depository of large capacity

Computers of large capacity (e.g. Disc Storage System IBM Total Storage DS8000 Series of capacity 192 TB) are installed in all locations of the branch departments. Each branch department has its own digital archive where all digital documents, associated with the activities for which the branch office is authorized, are stored.

The advantages of the local depository of a large capacity are:

- Documents are recorded according to the principles of grouping that have been defined in advance and according to the folder structure.
- The number of accesses to the medium (of reading and writing) is not limited.

- The document recording medium is a physically inseparable entity upon which the processing is done.
- Making a security copy means copying the contents of a limited number of media, totally 20 media from 20 branch departments.
- The depository capacity can be different for each branch department, depending on the needs of the branch department and on the estimated number of documents that shall be stored; with the average data quantity of 100KB per document it is possible to record about 10.000.000 documents on 1TB.
- Relatively low costs of communication that can be achieved by creating the local network within the branch department.
- Department workers can have an on-line access to all documents of the local archive, because the depository content is available through the communication server of the local network.
- A small number of users are served by the local digital archive so the demands on the communication server are significantly low.
- In the event of the local archive failure, work interruption will only happen in the branch department where the failure occurred, while the other branches continue to work independently to this failure.

These are the disadvantages:

- There is no unique digital archive in one place, but there are a larger number of smaller archives distributed in branch departments.
- Possible redundancy due to the spatial change of competence for clients (e.g. caused by moving).
- By taking a long-term view, the depository capacity is limited.
- A relatively high price at the beginning of implementation.

### 2.3 Configuration 3: The central depository system of optical discs

The depository system of WORM media includes Jukebox (a device that allows writing and reading of digital documents on more optical media of the same kind at the same time) and a cupboard for putting down the media.

The advantages of the central depository system of optical discs are:

- By taking a long-term view, the depository capacity is unlimited.
- Contrary to the disc systems, the costs are significantly lower at the beginning of the system implementation because optical discs are obtained according to the real and not according to the estimated needs.

- The depository capacity can be increased permanently depending on the real needs; with the average data quantity of 100KB per document, it is possible to record about 90.000 documents on 1 optical WORM medium of 9GB capacity.
- It is possible to have an on-line access to those documents that are in the JUKEBOX, while a near-line access is applicable for the others.
- There is no redundancy: all the entries are physically situated in one place (except for the security copy).
- Documents get recorded on a medium as they come to the storage system, which means that before storing the documents it is not necessary to create a folder structure.
- Each optical medium is independent of other optical media, so, in the event of a failure of any kind (fire, viruses or similar), the entire digital document fund does not have to be lost except for the media system that is caught by the failure.

The disadvantages of this configuration are:

- The document recording media are physically separated entities (2200 WORM media are needed for 200.000.000 documents stored on 9GB WORM media).
- Making a security copy means copying the contents of all WORM media that have to be inserted in the JUKEBOX.
- The guarantee for the medium quality is limited to 5 years.
- Medium indexation is decisive for the utilization of such document storing system.
- All users compete for one resource: All the organization users are served by one central depository; so a bigger number of servers are required (JUKEBOX is limited to a very small number of servers).
- A relatively small number of documents can be available on-line, e.g. with the possibility of managing 256 WORM media with the capacity of 9 GB, JUKEBOX enables an on-line access to about 200.000.000 documents.
- Jammed communication paths are probable, which demands communication infrastructure strengthening.
- Relatively high communication costs between the users and the central digital archive due to the transfer of relatively large data quantities.
- In the event of the central depository failure, not a single branch department can proceed with the work.
- There always have to be some empty media in the JUKEBOX on which new documents can be recorded. As the number of optical media on which documents are stored gets bigger than the number of WORM media that can be in JUKEBOX at the same time, it is necessary

to ensure the space for keeping the WORM media outside the JUKEBOX.

- There always have to be some empty media in the JUKEBOX on which new documents can be recorded. As the number of optical media on which documents are stored gets bigger than the number of WORM media that can be in the JUKEBOX at the same time, it is necessary to provide a worker who will take out the written media and put the new ones in the JUKEBOX.
- In case the required document on the medium is outside the JUKEBOX, the JUKEBOX worker will have to insert the required medium in the JUKEBOX.
- The equipment costs should include the costs of the JUKEBOX worker position.

## 2.4 Configuration 4: Local depository systems of optical discs.

The depository system of WORM media includes Jukebox (a device that allows writing and reading of digital documents on more optical media of the same kind at the same time) and a cupboard for putting down the media. The equipment should be installed in all locations of the branch departments. Each branch department has its own digital archive where all digital documents that are associated with the activities for which the branch office is authorized are stored.

The advantages of this configuration are:

- By taking a long-term view, the depository capacity is unlimited.
  - Contrary to the disc systems, the costs are significantly lower at the beginning of the system implementation because optical discs are obtained according to the real and not according to the estimated needs.
  - The depository capacity can be increased permanently, depending on the needs of a single branch department; with the average data quantity of 100KB per document it is possible to record about 90.000 documents on 1 optical WORM medium of the capacity of 9GB.
  - The number of documents that can be available on-line in the branch service, e.g. JUKEBOX with the possibility of managing 256 WORM media of the 9GB capacity, enables an on-line access to about 200.000.000 documents.
  - Documents get recorded on a medium as they come to the storage system, which means that before storing the documents it is not necessary to create a folder structure.
  - Each optical medium is independent of other optical media, so, in the event of a failure of any kind (fire, viruses or similar), the entire digital document fund does not have to be lost
- except for the media system that is caught by the failure.
  - Relatively low costs of communication that can be achieved by creating the local network within the branch department.
  - A small number of users are served by the local digital archive so the demands on the communication server are significantly low.
  - In the event of the local archive failure, a work interruption will only happen in the branch department where the failure occurred, while other branches continue to work independently to this failure.

The disadvantages of the local depository system of optical discs are:

- There is no unique digital archive in one place, but there are a bigger number of smaller archives distributed in branch departments.
- Possible redundancy due to the change of competence for the clients business (e.g. caused by moving).
- A relatively high price at the beginning of implementation due to the JUKEBOX price and the management applications.
- The document recording media are physically separated entities (2200 WORM media are needed for 200.000.000 documents stored on 9GB WORM media).
- Making a security copy means copying the contents of all WORM media that have to be inserted in the JUKEBOX.
- The guarantee for the medium quality is limited to 5 years.
- Medium indexation (so far it has been the one of the MF coils) is decisive for the utilization of such document storing system.
- There always have to be some empty media in the JUKEBOX on which new documents can be recorded. As the number of optical media on which documents are stored gets bigger than the number of WORM media that can be in the JUKEBOX at the same time, it is necessary to ensure the space for keeping the WORM media outside the JUKEBOX (the number of times as there are the local archives).
- There always have to be some empty media in the JUKEBOX on which new documents can be recorded. As the number of optical media on which documents are stored gets bigger than the number of WORM media that can be in the JUKEBOX at the same time, it is necessary to provide a worker who will take out the written media and put the new ones in the JUKEBOX (at least the number of workers has to be the same as the number of local archives).
- In case the required document on the medium is outside the JUKEBOX, the JUKEBOX

worker will have to insert the required medium in the JUKEBOX.

- The equipment costs should include the costs of the JUKEBOX worker position.

### **3 Criteria for selecting the storage configurations**

A detailed analysis of the above described configurations has established the possibility for the implementation of digital archives by the individual application of individual configurations or by their combination. For this purpose, a parallel analysis has been made, specifically the one in terms of criteria that result from the business needs of the organization from practice. The analysis results are shown in tables 1. and 2. In the first column, the criteria according to which the compatibility between technology and the needs that result from doing business are outlined, while in other columns, descriptive grades according to an individual technology application and according to their combinations are stated. While doing so, the abbreviations K1 to K4 are used for individual configurations (table 1 in the order as they are described) and the abbreviations K1+K2, K1+K4, K3+K2 and K3+K4 for meaningful configuration combinations of two components that always comprise one configuration from the group K1 and K2 and one configuration from the groups K3 and K4(table 2).

Table 1. A detailed analysis of the individual storage configurations

Criteria/Configuration	K1	K2	K3	K4
The possibility of work in the event of one minimal component failure	It does not exist	In the event of the failure in one branch service (BS), the others can proceed with the work.	It does not exist	In the event of the failure in one BS, others can proceed with the work.
The speed of accessing the document	Technology (T) of accessing the distant depository + T of data transfer from the distant location	T of accessing the local depository + T of data transfer from the local depository	T of accessing the distant depository + T of inserting the WORM media in the JUKEBOX (many times as there are WORM media outside the JUKEBOX) + T of data transfer from the distant location	T of accessing the local depository + T of inserting the WORM media in the JUKEBOX (many times as there are WORM media outside the JUKEBOX) + T of data transfer from the local depository
Equipment costs	High at the beginning for the resource purchase for the whole period of utilization	High at the beginning for the resource purchase for the whole period of using for 20 BS	Medium at the beginning, and during the system utilization they are proportional to the needs	Relatively high at the beginning due to a large number of JUKEBOXS, and during the system utilization they are proportional to the needs
Communication costs	High due to external communication (of BS with the distant location)	Almost irrelevant due to a possibility of establishing a local network	High due to external communication (of BS with the distant location)	Almost irrelevant due to a possibility of establishing a local network
The probability of the system jam regarding the number of servers	High	Relatively low	High	Relatively low
The system utilization costs	Relatively irrelevant	Relatively irrelevant	The costs of the JUKEBOX worker position, of the place for keeping the WORM media and copying due to a limited medium shelf life	The costs of the JUKEBOX worker position, of the place for keeping the WORM media and copying due to a limited medium shelf life (20 times)
Storage structure	Defined in advance	Defined in advance	Determined by the occurrence of the documents with a very low possibility of grouping that can be achieved by the physical exchange of the WORM media	Determined by the occurrence of the documents with a very low possibility of grouping that can be achieved by the physical exchange of the WORM media
The number of servers for workers	1 for all workers	20 in 20 branch services	1 for all workers	20 in 20 branch services
The limited quality of the storage capacity	Limited in advance	Limited in advance	Unlimited	Unlimited
The quantity of contents available on-line	100%	100%	Limited due to a limited number of WORM media that can be in the JUKEBOX at the same time	Limited due to a limited number of WORM media that can be in the JUKEBOX at the same time

Table 2. A detailed analysis of the combined storage configurations

Criteria/Configuration	K1+K2	K1+K4	K3+K2	K3+K4
The possibility of work in the event of one minimal component failure	If the K1 fails, BS can work without interruption; If one of the 20 K2 fails, it is possible to use the K1	If the K1 fails, BS can work without interruption; If one of the 20 K4 fails, it is possible to use the K1	If the K3 fails, BS can work without interruption; If one of the 20 K2 fails, it is possible to use the K3	If the K3 fails, BS can work without interruption; If one of the 20 K4 fails, it is possible to use the K3
The speed of accessing the document	T of accessing the local depository + T of data transfer from the local depository; <i>if the K2 fails: T of accessing the distant depository + T of data transfer from the distant location</i>	T of accessing the local depository + T of inserting the WORM media in the JUKEBOX (many times as there are WORM media outside the JUKEBOX) + T of data transfer from the local depository; in case the K2 fails: T of accessing the distant depository + T of data transfer from the distant location	T of accessing the local depository + T of data transfer from the local depository; in case the K2 fails: T of accessing the distant depository + T of inserting the WORM media in the JUKEBOX (many times as there are WORM media outside the JUKEBOX) + T of data transfer from the distant location	T of accessing the local depository + T of inserting the WORM media in the JUKEBOX (many times as there are WORM media outside the JUKEBOX) + T of data transfer from the local depository; in case the K2 fails: T of accessing the distant depository + T of inserting the WORM media in the JUKEBOX (many times as there are WORM media outside the JUKEBOX) + T of data transfer from the distant location
Equipment costs	High at the beginning, but without additional costs	High at the beginning due to 20 K4 components, and during their utilization they are proportional to the needs	Relatively high at the beginning due to 20 K2 components, and during their utilization they are proportional to the needs	High at the beginning due to 20 K4 components, and during their utilization they are proportional to the needs
Communication costs	In communication almost irrelevant due to reaching the documents; for copying the local archives into the central one, medium costs are expected	In communication almost irrelevant due to reaching the documents; for copying the local archives into the central one, medium costs are expected	In communication almost irrelevant due to reaching the documents; for copying the local archives into the central one, medium costs are expected	In communication almost irrelevant due to reaching the documents; for copying the local archives into the central one, medium costs are expected
The probability of the system jam regarding the number of servers	Relatively high	Low	Medium	Low
The system utilization costs	Relatively irrelevant	The costs of the JUKEBOX worker position, of the place for keeping the WORM media and copying due to a limited medium shelf life (20 times)	The costs of the JUKEBOX worker position, place for keeping the WORM media and copying due to a limited medium shelf life	The costs of the JUKEBOX worker position, place for keeping the WORM media and copying due to a limited medium shelf life (21 times)
Storage structure	Defined in advance in both cases	Defined in advance in the central location; locally determined by the occurrence of the documents with a very low possibility of grouping that can be achieved by the physical exchange of	In central location and locally determined by the occurrence of the documents with a very low possibility of grouping that can be achieved by the physical exchange of the WORM media;	In the central location and locally determined by the occurrence of the documents with a very low possibility of grouping that can be achieved by the physical exchange of

		the WORM media	locally defined in advance	the WORM media
The number of servers for workers	20 in 20 branch services, and 1 for all in the event of the K2 failure	20 in 20 branch services, and 1 for all in the event of the K4 failure	20 in 20 branch services, and 1 for all in the event of the K2 failure	20 in 20 branch services, and 1 for all in the event of the K4 failure
The limited quality of the storage capacity	Limited in advance both locally and in the central location	Limited in advance in the central location, and locally unlimited	In the central location unlimited, and locally limited in advance	Unlimited both in the central location and locally
The quantity of contents available on-line	100%	It is locally limited due to a limited number of WORM media that can be in the JUKEBOX at the same time, but in the event of the K4 failure it is 100% in the central location	Locally it is 100%, but in the event of the failure it is limited due to a limited number of the WORM media that can be in the JUKEBOX at the same time	Limited due to a limited number of the WORM media that can be in the JUKEBOX at the same time

#### 4 Conclusion

The digital archive is a part of an information system, and as such it should help, in the best possible way, the realization of organization everyday activities. The problem that appears in practice is the true selection of an appropriate configuration of the digital archive implementation for managing business documents.

Different organizations need different solutions, depending on several criteria for choosing the right configuration.

This paper gives four possible digital archive configurations: The depository of high capacity in the central location, The local depository of high capacity, The central depository system of optical discs and The local depository systems of optical discs; and four combinations of two components of these configurations.

An organization can choose between one of these suggested configuration.

Criteria that are listed in this paper can be adjusted or changed according to the type of organization that uses the decision model. In order to choose the appropriate configuration, an analysis of business demands should be done and it should be determined which concept suits these demands the best.

While doing this, it is necessary to choose carefully (and also to weigh) the criteria according to which the analysis of the compatibility between the configurations and the business demands. Some of these criteria are being offered in this paper: The speed of accessing the document, The probability of the system jam regarding the number of servers, Digital archive costs, The quantity of contents available on-line and The limited quality of the storage capacity.

#### References

- [1] Falk, H.: Digital archive developments, The electronic library, 21(4), 2003, pp. 375-279.
- [2] Johnston, G., P., Bowen, D., V.: The benefits of electronic record management system, Records management Journal, 15(3), 2005, pp.131-140.
- [3] Lin, L., S., Romaiah, C., K., Wal, P. K.: Problems in the preservation of electronic records, Library Review, 52(3), 2003., pp117-125.
- [4] Mateljan, V., Juričić, V., Šimović, V.: Content management systems functionality, Proceedings of the 19th International Conference on Information and Intelligent Systems, Varaždin, 2008.
- [5] \*\*\*: *CIO Survey-Content and Document Management Trends 2004*, Rethink Research Associates, 2004.