



**HAL**  
open science

# Cost Analysis of the Digitization, Web integration and Distribution across the Internet of Rare and Old Printed materials

Serge Lenga, Jacqueline Rey, Ahmed Silem

► **To cite this version:**

Serge Lenga, Jacqueline Rey, Ahmed Silem. Cost Analysis of the Digitization, Web integration and Distribution across the Internet of Rare and Old Printed materials. *Revue MIF (Management, information, finance)*, 2001, 1 (1), pp.143-155. hal-00821444

**HAL Id: hal-00821444**

**<https://hal.science/hal-00821444v1>**

Submitted on 13 May 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **COST - ANALYSIS OF THE DIGITIZATION, WEB INTEGRATION AND DISTRIBUTION ACROSS THE INTERNET OF RARE AND OLD PRINTED MATERIALS**

---

Serge D. LENGA, Jacqueline REY & Ahmed SILEM  
ERSICO, Université Jean-Moulin Lyon 3  
nkiama@club-internet.fr, rey@univ-lyon3.fr, silem@univ-lyon3.fr

**Résumé** : Ce travail cherche à établir aussi exactement que possible les relations de coût caractéristiques d'une numérisation d'ouvrages du XVI<sup>ème</sup> siècle, à la disposition du plus grand nombre d'utilisateurs, sur Internet. Dans un premier temps, on présente les divers projets et travaux qui ont été retenus pour faire le point sur les méthodes employées en matière d'analyse des coûts induits lors de la numérisation d'ouvrages anciens. Dans un deuxième temps, on propose une méthode originale d'analyse des coûts à base d'activités, adaptée de la chaîne de valeurs de Michael Porter.

**Mots-clés** : analyse des coûts, information numérique, modèle économique, numérisation, ouvrages rares

**Abstract** : The present paper seeks to precisely establish the cost functions characterized through the operations of digitization and distribution via Internet of rare 16th century books. Firstly a state of the art cost analysis is presented based on the documents which have been retained by the authors to make the point on the methods used in digitization cost analysis. Secondly an original activity-based cost analysis method is suggested borrowed from Michael Porter's value chain.

**Keywords** : cost analysis, digital document, digitization, economic model, rare books

## INTRODUCTION

This paper consists of two parts : in the first part are presented the economic logics which have currently course with regard to the assessment of documentary resources and techniques in general, and more precisely to the techniques and methods of appreciation of electronic library costs ; in the second part an effective model is proposed for determining typical costs involved in 16th century books digitization.

The present work seeks to exactly establish the cost functions which can be characterized through operations of rare books digitization, for which the new information and communication technologies (notably Internet) have important consequences in terms of access to knowledge for the biggest number via the production of databases and Internet interfacing.

We shall firstly present a state of the art cost analysis starting by pointing out the documents which allowed us to make the point on the methods used in digitization cost-analysis. We shall secondly suggest an original cost-analysis method inspired from the ABC method (Activity Based Costing) (R. S. Kaplan, 1997), starting from an analysis of activities adapted from Michael Porter's value chain (M. Porter, 1994).

### 1. STATE OF THE ART COST ANALYSIS

#### 1.1 Main On-line Rare Books Digitization Projects for Libraries

Several recent studies show an attempt to translate library management in economic terms. Pierre Carbone (P. Carbone, 1998) underlines the difficulties to put the ISO's latest standard – ISO 11620, published by AFNOR under the reference code Z 48-005 – into practice. The existence of a common standard for the assessment of library performances raises the question over the translation of the various performance indicators into a common language. This common interpretation of results is important, because indicators are often assessment context dependent, and limited by their economic ambiguity – a good rate of document use, for example, may not coincide with a good rate of document availability – (and so on). If the overall indicators presented in that standard are reconsidered, none allows us to find any assessment in the field of digitization, probably because it constitutes an entirely new field of research. In that sense, the results obtained by J. M. Griffiths and D. W. King (Griffiths & King, 1997) present an innovatory character, since they more particularly rely on research studies starting from 1983 to the present time. If we strictly hold to that aspect of digitization cost, a few projects seem appropriate for selection as regards their results.

We retained seven projects or studies in view of establishing a state of the art cost analysis within the DEBORA project framework :

1) the UK Public Libraries Performance Measurement Project supported by the « British Library Research and Development Department » of the Office of Arts and Libraries of the United Kingdom. From this survey resulted a taxonomy of parameters applicable to libraries as well as to their environment and the relations between them thereof ;

2) the « Bibliotheca Universalis » Project settled during the G7's top meeting devoted to « Information Highways » and which will give access to the main works of the worldwide cultural and scientific historical heritage <sup>1</sup> ;

3) the Digitization « JSTOR » (Journal Storage) Project, to digitize, preserve and provide electronic access to scholarly journals, supported by the Andrew W. Mellon<sup>2</sup> Foundation. The project represents approximately 750000 pages, « using technology developed at Michigan, high-resolution (600) bit-mapped images of each page linked to text file generated with optical character recognition (OCR) software »<sup>3</sup>.

4) the BNF Digitization Project « GALLICA » to digitize printed materials in the National Library of France (Cathaly-Pretou, 1998)<sup>4</sup>. The three-year initial program – in fact extended to four years – resulted in an overall production of 86.000 books and periodicals, that is about 25 million digitized pages ;

5) some analyses derived from many functional library surveys by specialists using such facilities and these methods( J.M. Griffiths, 1997) ;

6) the « Early Canadiana Online » Project to digitize the Canadian libraries patrimony (Kingma, 1997)<sup>5</sup>. This survey particularly emphasizes the differences between electronic document exploitation costs according to the medium. Comparisons are made between costs of microfiches, costs of printed books and costs of electronic books ;

7) the Berkeley « Museum Educational Site Licensing Project (MESL) » to allow electronic access to documentary resources for educational purposes (Besser, 1998) and more particularly designed for such structures as schools of arts or architecture using images as teaching aids.

## 1.2 Main Cost-Analysis Methods

A conceptual evolutionary framework has been developed by Griffiths (J.M. Griffiths, 1995) for the assessment of libraries. It presents five successive phases :

the aims and perspectives of the assessment, the generic assessment measure types, the specific measure types, the by-product measures, the interactions and the outside factors.

Once the structure of the various operations has been designed for the assessment and the selection of documents, the chaining of the successive operations will articulate around the eight following stages of a complete digitization process (see *Figure 1*)

---

<sup>1</sup> < <http://culture.fr/culture/bibliuni/biblio-u.htm> >

<sup>2</sup> < <http://www.mellon.org/> >

<sup>3</sup> < <http://www.jstor.org/> >

<sup>4</sup> < <http://www.culture.gouv.fr/culture/mrt/numcrisation/fr/> >

<sup>5</sup> < <http://www.canadiana.org/> >

General Digitization Process							
DOCUMENT EVALUATION AND SCREENING	DIGITIZATION PROCESS EVALUATION	« BENCH MARKING <sup>6</sup> »	OVERALL DIGITIZATION	QUALITY ASSESSMENT	POST - EDITION	METADATA IMPLEMENTATION	DELIVERY

Figure 1 : The eight usual stages of a digitization process

It is a model of a sequential production, in which cost is involved by the end user's activities. This pattern is satisfactory insofar as it brings the totality of the different necessary stages toward the complete fulfillment of the project. However splitting up the different operations in such a way, has little to do with the inherent expectations of the DEBORA project.

The taking into account of the multiple operation specificities which constitute the DEBORA project allowed us to formulate a value chain more in adequacy with the conventions and the assigned objectives. The definitive cost of the project should integrate the decisions accepted in the course of « the assessment and the selection of source documents », « the end-user satisfaction », etc in relationship with the internalization and/ or the externalization. We retained the principle of reciprocal definite services between every cost center in order to establish the cost generating connections for all the digitization processes, before grouping them into four homogeneous task sectors. In that sense, the chain below (*Figure 3*), which is an adaptation of the general model suggested by Michael Porter (M. Porter, 1994), is a general example of value sequences within libraries. We shall take it thereafter in order to illustrate and characterize the document digitization flow in the DEBORA project (*see Figure 4*).

Among the basic factors of digitization costs, it is necessary first to mention the costs linked to the project management in terms of planning, material preparation, quality of the insurance when the original documents of the collection have a certain value, and consequently their need to be under control when displaced. According to S. Tanner and J. L. Smith (Tanner & Smith, 1999), it is – and by far – the costs of book preparation which most weigh in the determination of the final cost. The diagram below gives us, in percentage, the results of a survey conducted by « MR Data Ltd » over the costs of digitization projects and their decomposition. Here are some preparation costs (as illustrated in *Figure 2*).

<sup>6</sup> One makes sure that most relevant data are captured. More formally it means an utter knowledge of the original document's main attributes ; a recognition of current and future needs ; an identification of the precise limits of the acceptance zone which will integrate every current and future remark and suggestion about digitized documents or surrogates.

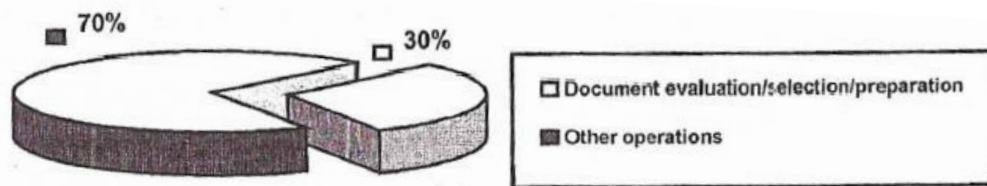


Figure 2 : Apportionment of the various operations in terms of effective working time ( Tanner & Smith, 1999)

- the displacement of the items requiring inventories and packaging for the transportation (even in the case of in house digitization) ;
- the time spent to assign unique identifiers to originals, if it had no already been done before ;
- the costs linked to operations of removing staples, cleaning transparencies, or the overall physical book processing ;
- finally, all the costs involved in checking copyright and negotiating all rights bound to the authorization of representation and reproduction.

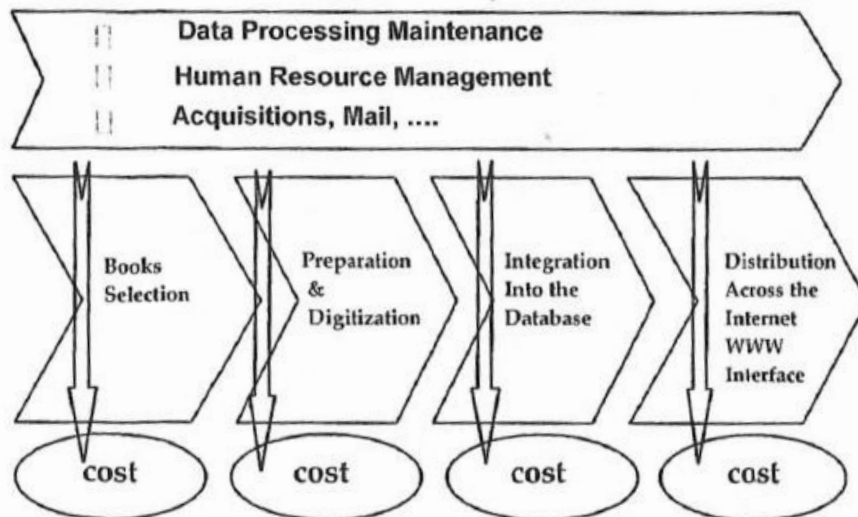
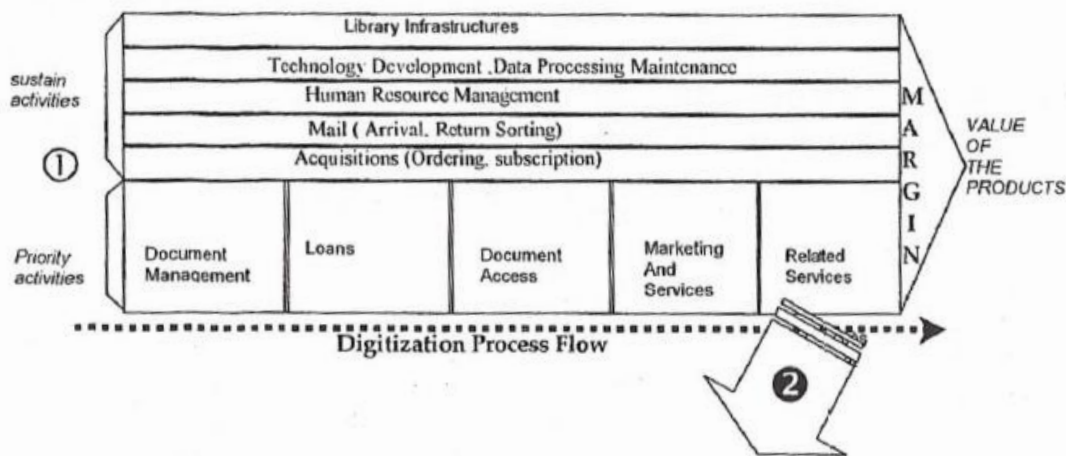
The assessment of library organization costs most of the time oscillates between several operational levels which may correspond to three great cost functions :

- user-bound functions, providing the services and the products for the library end users (access to the collections, data retrieval, users training, etc.)
- operational functions, necessary to insure services to the users (collection management and development, acquisitions, cataloguing and indexing, physical processing).
- auxiliary functions, necessary for administrative purposes and for the management of the user-bound operational functions.

The functions were split up into these three categories because the measures and the methods of assessment are different from one another in the three cases, and the costs of the auxiliary functions are often ventilated between user-bound and operational functions ; some of these operational functions being then affected to the user-bound functions, in order to appraise the total costs as well as the unit costs of individual services.

The successive costs of « book selection », of « book preparation and scanning », of « integration into the database » and lastly of « distribution across the Internet » are each calculated according to their direct cost ventilation in the operatory process. In that sense, a first development stage of the operatory mode will consist in ventilating each value function into activities. For example, the function « work selection » can cover such activities as criterion definition and choice or digitization standard specification and choice. One will at that stage speak of « cost-inductive activities » or « cost drivers

of activities » insofar as the particular activity of criterion definition and choice should affect the whole digitization process.



Figures 3 & 4 Digitization Value Chain and Scanning Activities

## 2. A METHOD FOR COST DETERMINATION

Within the DEBORA project framework, we have tried to overview all the reports written for similar projects, while searching to emphasize the specific costs linked to an Internet web distribution of 16th century books. In that sense, two issues are required to be met in our methodical choice of assessment :

- the first one is the fundamentally interactive character of the mode of distribution pinpointed. There are indeed some relationships of strong and interdependent reciprocity between costs involved and end users which make us postulate, along with C. Shapiro and H. Varian (Shapiro & Varian, 1999), some quasi infinite second rank distribution costs ;

– the second issue is the proposition of the general functional study model which will serve us as a methodical guide both in the collection, the calculation and the development of typical cost scenarios.

## 2.1 From Physical Distribution to Digital Distribution

In the cultural field, probably more than anywhere else, the offering entails the demand : the book market is thus a market of the offering and therefore a risky market since the product may not find a book purchaser. A new orientation is taken by the editorial model which was so far specially known for its high costs of design, production and distribution since the digitization process makes costs of reproduction and distribution nearly void. As argued by R. Coase (R. Coase, 1987) or more specially by O. E. Williamson's works (O. Williamson, 1975 & 1991), the exchange is known to entail two types of costs :

- the first inherent to the contract conception and follow-up as it entails the transfer of property rights of goods between distinct economic agents ;
- and the second linked to the research and acquisition time necessary to finalize the contract.

In a situation of strong uncertainty on the competitive future of the editorial « prototype » (see Figure 5 for the Analog Model), and taking into account the specific character of the investments achieved in its design and its production, institutional arrangements which will minimize the exchange and production costs should target internalization as it is supposed to reduce uncertainty (that is risk). Henceforth, economies of scale become quasi infinite insofar as goods may be sold again and again. That evolution is accompanied with an unprecedented « revolution » which redistributes the strategic deal as a whole : that is network externality effects.

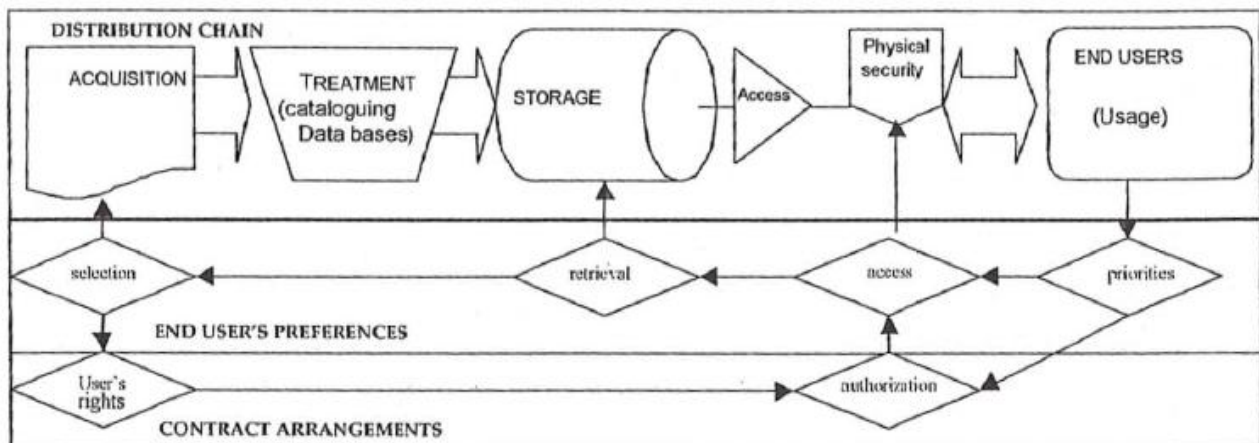


Figure 5 : « Analog Distribution Model » (Besser, 1998)

Indeed while the strategic field structure organizes itself according to a « stock of specific assets », the « errand to the size » phenomenon in library organizations should be explained by the fact that the more they dispose of « idiosyncratic » catalogs the more they should offer a considerable potential of growth in conjunction with



economies of scale effects in matter of production and of the demand. The recurrent need for the same « book » product therefore induces them to « internalize » production in order to minimize transaction costs. Inversely it increases organizational costs in terms of control, management, and supervision (etc). But it precisely is on such costs that information technology permits new economies of scale. In that sense, and according to B. R. Kingma (B.R. Kingma, 1997), « we think that » information digitization entails lower costs than usual products, for the producers as well as for the consumers or the intermediaries ( that is the libraries) ».

The digitization process is grounded on a logic of requesting supply for a product which exists in a single exemplar. Such a « demand pull » behavior decreases both the costs of information resources availability and the costs of production, properly speaking, for all those involved. One speaks of an interactive process insofar as the starting point of the production doesn't rest anymore upon a prefabricated and already available product. All the former ground-rules of product distribution are consequently modified both in organizational terms, and in terms of cost.

According to H. Besser and R. Yamashita (Besser & Yamashita, 1998), the evolution of the digital image distribution model is viewed as central in the mode of cost determination. The traditional Publisher(s) - Intermediary(ies) - End-User(s) chain provides a cost assessment on the basis of a value chain incorporating all the successive costs (see Figure 6 below).

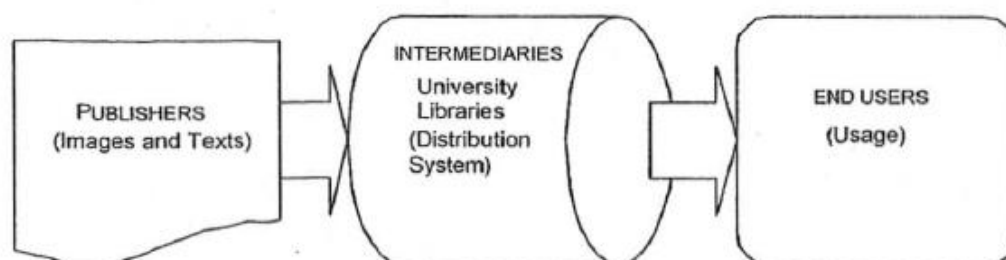


Figure 6 : « MESL Distribution Pathway » (Besser, 1998)

## 2.2 A General Cost Analysis Functional Model for Digitized Books

On the basis of the bibliography perused for this state of the art, it seems that today it doesn't exist any survey whose results could be used to evaluate the costs induced by book digitization, further integration into databases and further availability across the Internet. Our subject matter will therefore be exploratory. And it is quite obvious that if the DEBORA project makes it possible for us to make a beginning of a contribution in the field of cost analysis methods in the processes of digitizing and networking ancient documents, it is also necessary to specify that our primary objective is to establish a general method, and not to assess the specific costs of the DEBORA Project. Nevertheless, the ambition of our subject matter remains technologically limited, for it will be difficult if not impossible to mention experimental effects, or to take into account costs going lower and lower as better quality hardware and software are likely to appear on the market in the next few years.

The value chain principle developed by Michael Porter (see figure 7) is reproduced for a functional analysis of the main typical costs in the digitization process. Without deviating from the main studies on that topic, four categories of costs will be retained in our model :

- the cost of book selection and preparation built on the basis of components connected to the book preparation and selection processes. Within that function, an analysis on the typical activities will be made from an « activity cost approach ». To that end it will be necessary to search and establish the determination matrixes for main cost inductors.
- the cost of digitization must be grounded on the comparison between the externalization and the internalization. One of the main cost inductors remains the data processing material and the necessary software for the treatment and the acquisition of image collections.
- the cost of integration into the database and of constitution of the databases to be distributed across the Internet. The computer peripheral database storage and management elements are among the main central cost inductors for this functionality.
- the cost of distribution of situated documents across the Internet deserves questioning over the software distribution function. Cost inductors should be largely function of the results of the information-seeking process study, specially to identify information searching behaviors (frequency, duration, etc.).

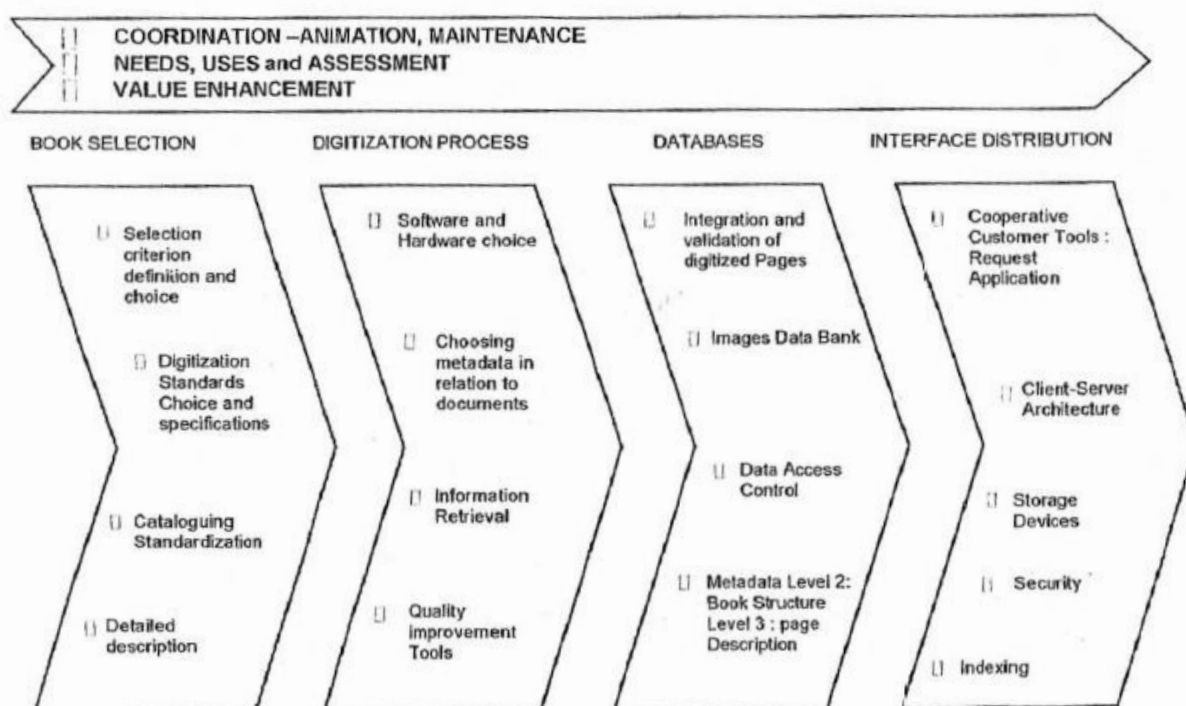


Figure 7 : Activity Chain and General Functional Model

One of the important points to underline in that model is that, when documents are to be digitized by intermediaries, a copyright authorization for copying but also a license of access to the document for authorized users must be requested. Thus costs are generated at the launching of the digitization chain although such a situation is not to be

considered in the DEBORA Project case, since all the items to be processed date back to the 16th century and there is therefore no obstacle either on the copyright or on the license of access.

We present in Figure 8 a paralleling diagram first for the analog distribution process and second for the digital distribution process. In both processes, three periods of time are to be noticed :

- the acquisition period of time which includes the various operations involved in the selection of items in the analog case, and in the acquisition and digitization operations in the digital case. Generally speaking, all the operations belonging to the activity of book acquisition are not different from the traditional editorial model. The only additional distinctive element should be the digitization activity sequence ;
- the period of unique costs here refers to all the series of processes which happen only once and are not repetitious. The production activity at the basis of the acquisition period of time is followed by the constitution of metadata, on one hand, and by the computer data processing (OCR, image compression, straightening, etc.) and database construction, on the other hand. The last one is represented by the collections owned by the library organizations in the analog hypothesis. It is represented by the electronic data base in the digital case. In both cases, even though they are physically quite different in volume, they are brought back to the same « preservation » logic from the cost process viewpoint. This process ends up by the constitution of a new product aggregating all the costs for such an activity ;
- the period of repeated or iterative costs of image distribution. Whereas the analog process comes to an end thanks to a traditional display (loan or reference), it is a different matter in the digital case : the product is available in unlimited quantity for a void reflexive cost. Moreover users have the capacity to modify priorities for retrieving precise data, and even to select complementary or contiguous works beside the constructed database. A « situated document » and a « digital image » are in conflict in terms of cost management by the fact that costs are respectively increasing for the first ones and decreasing for the second ones.

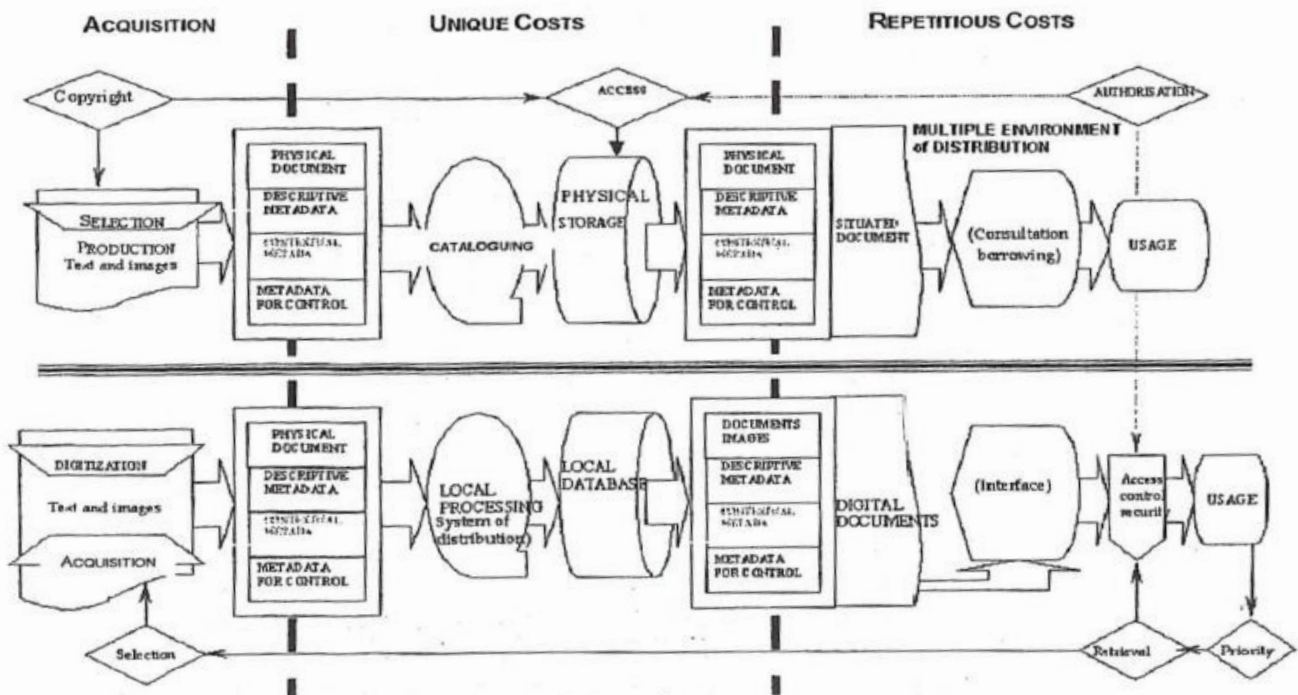


Figure 8 Analog Costs versus Digitising Costs

## CONCLUSION

The choice of the way to produce digitized documents depends on several criteria among which the following four should be worth keeping (C. Ducharme, 1997) :

- the periodicity of the digitization project : is it a repeated case of digitization (as in the case of newspaper or news files ) or a casual operation (a rare book collection or image data) ? In the case of DEBORA, it is a casual project ;
- the technical expertise skills : does the library dispose of necessary skills in-house ? These skills may largely vary, according to the items to be digitized. Regarding a casual digitization process, with the integration of old documents into a database and across the Internet network, it might seem that such tasks are far from the core skills of a librarian and therefore of the public library staff, but appropriate training can make them increasingly knowledgeable in new technologies and products ;
- the funding : does the library dispose of important financial means ? Can it invest in relatively costly equipment ? Has it the possibility to include in its annual budget a provision for digitizing documents, maintenance and updating the database ?
- the nature of the documents : some documents can be handled easily (size, medium, state), others have a patrimony value or are too fragile to be displaced. It is necessary therefore to provide for extra costs for those precious documents needing delicate handling, an insurance for loss or destruction risk.

Fundamentally and according to the theory of transaction costs, there are two possible scenarios to be developed : (1) Do it (use in house resources) or (2) have some one do it (outsource), that is make or buy.

|| In the first scenario, the library totally masters the process which goes from selecting documents in its huge, difficult of access for the public, repository of rare materials up to their networking on the Internet and their management on the web, following the digitization process. It digitizes documents in house with the help of a microcomputer device with a scanner or a digital photo equipment. Resulting images are immediately transferred to a computer to be stored after processing upon a magnetic or optical disk, or transmitted to a server across a local network. Costs are not proportional to the volume or to the size of the repository, for there are economies of scale and training effects.

|| In the second scenario, the library which has a small repository, appeals to an external service provider. The advantage of that solution is that the library does not invest in equipment and possesses a substantial safeguard for all its document collections (through CD-PHOTO). On the other hand, the current budget must have a provision for such an operation which entails transaction costs, document handling plus all the risks inherent to storing documents outside their usual repository.

Information grids over the different types of costs for both scenarios are given in appendix, knowing that it is always possible to combine the two : part of the work is processed in the library and the other is entrusted to a service provider.

## References

- BESSER, H., YAMASHITA, R., *The Cost of Digital Image Distribution*, <<http://sunsite.berkeley.edu/Imaging/Databases/1998mellon>>, 1998
- CARBONE, P., « Evaluer les performances des bibliothèques », *Bulletin des Bibliothèques de France*, n°6, 1998
- CATHALY-PRETOU, G., « La numérisation des documents imprimés à la BNF », in *Journée d'étude sur la documentation informatisée au sein du ministère de la culture et de la communication*, 16 novembre 1998, p.15–19, <[http://www.culture.gouv.fr/culture/mrt/numerisation/fr/seminaire\\_du\\_161198/projet\\_n11.htm](http://www.culture.gouv.fr/culture/mrt/numerisation/fr/seminaire_du_161198/projet_n11.htm)>
- COASE, R., « La nature de la firme », *Revue française d'économie*, II, Hiver 1987.
- DUCHARME, C. (Dir.), *Du CD-ROM à la numérisation : développer les documents numériques en Bibliothèque*. Villeurbanne : Institut de formation des bibliothécaires, Coll. La boîte à outils, 1997
- GRIFFITHS, J.-M., KING D. W., « Economic Issues Concerning Electronic Publishing and Distribution of Scholarly Articles », *Library Trends*, n°43, Spring 1995, p. 713-740.
- GRIFFITHS, J.-M., KING D. W., « Mesure de la valeur des services d'information », in SALAUN, J.-M. (Dir.), *Economie et Bibliothèques*, Editions du Cercle de la Librairie, 1997, p. 159–176.
- KAPLAN, R. S., COOPER R., « Cost & Effect : using integrated cost systems to drive profitability and performance », *Harvard Business School*, 1997
- KINGMA, R., « Economie comparée de l'achat de périodiques et de l'accès aux articles », in SALAUN, J.-M. (Dir.) *Economie et Bibliothèques*, Editions du Cercle de la Librairie, 1997, p. 187–198.
- MOULIN, E., *Les coûts en documentation : calculs, analyses et décisions*, ADBS Editions, Paris, 1995
- PORTER, M., *Choix stratégiques et concurrence*, nouvelle éd., Economica, 1994
- SHAPIRO, C., VARIAN, H. R., *Economie de l'information*, DeBoeck, 1999, (traduction de la version américaine de 1997)
- TANNER, S., SMITH J. L., *Why Digitize ?*, 1999, <<http://www.clir.org/pubs/reports/html>>
- WILLIAMSON, O. E., *Market and Hierarchies : Analysis and Antitrust Implications*, New-York, The Free Press, 1975
- WILLIAMSON, O. E., WINTER, S. G., *The Nature of the Firm : Origins, Evolution and Development*, Oxford University Press, 1991

**APPENDICES  
COST CAPTURE DOCUMENTS**

<b>Working Time Measure Chart Routine Tasks and Digitization Tasks on DEBORA</b>		
<b>FUNCTION / JOB</b> OVERALL GROSS SALARY HOURLY GROSS SALARY PERIOD FROM .....TO	<b>OVERALL JOB HOURLY VOLUME</b>	<b>PERCENTAGE FOR THE GIVEN PERIOD</b>
<b>ROUTINE TASKS :</b>  ORGANISATION & MANAGEMENT ADMINISTRATION & EXHIBITION ACQUISITIONS, MAIL, MAIL PERUSAL DOCUMENTARY FILES DATABASES QUESTIONS PRODUCTS  ETC.		
<b>subtotal I</b>		
<b>SPECIFIC TASKS ON DEBORA</b>  SELECTION PREPARATION & DIGITIZATION INTEGRATION INTO THE DATABASE DISTRIBUTION  <b>SUBTOTAL II</b>		
<b>OTHERS, SPECIFY</b>		
<b>Overall HOUR'S WORK (I + II)</b>		<b>100%</b>