

LES COÛTS DE LA NUMERISATION DE DOCUMENTS ANCIENS LES LEÇONS DU PROJET DEBORA

Serge LENGA et Ahmed SILEM,
ERSICO, Lyon3

Résumé : Après un bref rappel de la nature des coûts qui accompagne la mise en œuvre de la numérisation de documents rares, en particulier des livres anciens de la Renaissance, l'article a pour objet de préciser les coûts de création d'images numériques et de leur distribution via Internet, puis de proposer des scénarii pour un passage à grande échelle de ce qui a été élaboré dans l'expérimentation du projet européen DEBORA (*Digital AccEss to Books of the RenAissance*).

Mots clés : Documents numérisés, Economie de l'information, Coûts de la numérisation, Servuction, Internet.

Abstract : After a short recall of different digitization costs of rares old Renaissance's books, we specify the digitization costs images creation and their distribution via Internet. Then we propose scenarios for a passage on a large scale of what was elaborate in experimentation of european DEBORA project (*Digital AccEss to Books of the RenAissance*).

Keywords : Digitized documents, Economics of information, Costs of digitalization, Servuction, Internet

INTRODUCTION

La numérisation permet aujourd'hui l'orientation vers un plus grand échange potentiel d'informations. Elle tend à s'imposer progressivement comme un support de documentation largement partagé. On assiste ainsi à une dématérialisation des documents qui appelle de nouveaux supports au savoir. Il s'agit en réalité :

- soit de mettre des documents rares à la disposition du public sur cédérom et DVD-ROM ou sur les réseaux ;
- soit encore de structurer des corpus afin non plus seulement de les éditer mais de les rendre disponibles et manipulables par chacun selon son propre travail et d'approfondir des études comparatives.

Ces réalisations font naître une problématique spécifique de l'utilisation de l'informatique pour le traitement des documents anciens. En effet, les projets de numérisation soulèvent des problèmes de coûts et de budget d'autant plus importants que l'évolution rapide des nouvelles technologies rend obsolète toute possibilité d'élaboration de budgets standards types. Il importe donc, pour chaque projet de banques d'images, de faire la liste et

d'évaluer les coûts. Pour cela les organisations ont besoin de réaliser des arbres de décision descriptifs de différents scénarii, c'est-à-dire d'élaborer un outil d'aide à la décision sous la forme d'une représentation graphique, d'une suite d'actions alternatives ainsi que des conséquences qu'elles suscitent. Celles-ci seront à leur tour suivies de nouvelles actions alternatives avec des conséquences chaque fois différentes en termes d'actions pouvant être envisagées avec une certaine probabilité.

Il convient de définir également à l'avance quelle sera la répartition des coûts. Ces choix sont étroitement liés aux autres choix techniques, qu'ils portent sur les aspects purement organisationnels ou encore sur les préparatifs antérieurs à la numérisation. Si le cahier de charges d'un programme de numérisation demeure difficile à se procurer – nous n'avons trouvé aucune données précises et complètes sur ce point – il est cependant possible d'identifier aujourd'hui quels sont les différents domaines clés de dépenses. Ce sont :

- le coût de la numérisation en elle-même, qui peut aller du simple au double (de 0,7 à 1,3 € la page en moyenne sur des projets courants et le quintuple sur des projets de haute qualité).
- le coût d'acquisition du matériel – les logiciels, imprimantes, système informatique si le service n'est pas encore informatisé – ce matériel va nécessiter un entretien par un contrat de maintenance.
- l'embauche de personnel qualifié ou la formation du personnel sur place, représente aussi un coût additionnel.
- le coût en termes de temps passé par les employés sur ce programme ou encore le coût occasionné par la mise à l'écart des usagers de certains documents.

Il faudra prendre en compte dans l'arbre de décision d'un programme de numérisation toutes les contraintes liées à la mise en place de ce programme. Ces contraintes impliquent des coûts, qu'ils soient en termes monétaires ou en termes de temps. Elles découlent de l'objectif même de la numérisation : mettre sous forme électronique des documents, de la façon la plus appropriée au type d'établissement concerné, mais aussi au type de public ciblé. Ainsi, le projet demande-t-il d'être adapté à une triple logique d'approche : le matériel utilisé (aspect technique), les coûts proprement dits (aspect financier) et les règles juridiques (aspect juridique).

1 LE PROJET DEBORA

Le projet européen DEBORA (*Digital AccEss to Books of the RenAissance*) a pour principal objectif la mise en commun des ressources des différents partenaires que représentent les bibliothèques à savoir : la Bibliothèque Municipale de Lyon en France ; la Bibliothèque de Casanatense à Rome en Italie ; la Bibliothèque Générale et Universitaire de Coïmbra au Portugal.

Ces trois intervenants se présentent comme des acteurs majeurs du projet de numérisation des ouvrages du 16^{ème} siècle, tant du point de vue des collections sélectionnées que de celui de la finalisation dudit projet, surtout lorsqu'il s'agit de formuler des propositions de modèles de coûts globaux. Cette dernière partie de notre travail examine les quantités de temps utiles aux bibliothèques pour remplir les missions qui leur avaient été assignées, notamment par rapport aux temps de sélection des documents, de numérisation et d'intégration dans la base de données disponible via un client Internet.

Globalement, le processus de numérisation se présente de la manière suivante : les ouvrages sont sélectionnés par chaque bibliothèque et « scannés » par leur soin respectif. En ce sens, ceux de Lyon ont été transmis à la bibliothèque Municipale de Lyon pour leur numérisation sous "DIGILIVRE". Les ouvrages des bibliothèques de Coïmbra au Portugal, d'une part, et de Casanatense à Rome, d'autre part, ont procédé de la même façon, en numérisant les collections qu'elles ont retenues, soit par leur propres moyens, soit en recourant à des intermédiaires.

Notons que quatre niveaux de métadonnées ont été définis pour l'ensemble des documents sélectionnés. Les centres de coûts suivants – caractéristiques d'une chaîne de numérisation – peuvent être retenus :

- a) Sélection, description et préparation de la collection détermination des métadonnées : on retient les étapes du processus qui conduisent au choix de la collection à travers les ouvrages faisant partie de la collection à numériser ;
- b) Numérisation et extraction de l'information : on retient le processus de numérisation proprement dit, avec l'observation des distinctions de coûts issues du matériel utilisé (DIGILIVRE et/ou l'outil logiciel mis au point par les chercheurs de l'INSA, équipe du professeur Hubert Emptoz) ;
- c) Intégration des données dans la base de données SGBI et détermination des métadonnées : on retient les quatre niveaux de métadonnées nécessaires à la structuration et à la préparation de la base de données, avec à la clé la

naissance d'un nouveau produit documentaire désignant l'ouvrage numérisé ;

d) Intégration dans le serveur DEBORA : il s'agit ici d'évaluer le processus du coût de l'intégration de la base de données au serveur ; ainsi que de celui de la distribution des ouvrages via Internet par l'interface développée sur le base du protocole ARIADNE ;

d) Les coûts additionnels : ici nous reprendrons l'ensemble des coûts relatifs à des opérations dont nous ne pouvons directement affecter les coûts aux autres centres ci-dessus définis : coordination, recherche et développement, supervision, administration,... Nous regrouperons dans la suite l'ensemble de ces coûts sous le terme de « valeur ajoutée ».

Le mécanisme d'évaluation des images numériques obtenues à partir d'une chaîne de production – fondée sur une logique industrielle – impose de recourir à des processus codifiés de détermination des coûts. Dans cette optique, le modèle d'évaluation proposé antérieurement¹ trouve ici une application concrète à travers les deux alternatives de construction d'élaboration des coûts caractéristiques rencontrés dans le projet DEBORA. L'ambition est de pouvoir fournir des éléments de « standardisation des coûts » dans une optique de passage à une plus grande échelle. Selon l'option méthodologique retenue, deux types de choix opératoires seront à effectuer pour déterminer le choix pertinent dans la structure de calcul des coûts et plus particulièrement sur la recherche de standards de coûts. Par rapport aux différentes alternatives envisagées, nous avons choisi de présenter deux méthodes parallèles : la détermination du coût internalisé pour l'ensemble du processus contrôlé par le soin des acteurs, et la détermination du coût externalisé qui fait référence à un choix de prestataire de service pour la numérisation des ouvrages, le reste de la chaîne restant identique. Il faut préalablement à la présentation des résultats définir les inducteurs de coûts.

2. LES INDUCTEURS DE COUTS

L'analyse des inducteurs de coûts est étroitement dépendante des choix technologiques (internalisation de tout ou partie des différentes opérations de la numérisation, et symétriquement externalisation) fait par chaque acteur

¹ S. D. Lenga, J. Rey et A. Silem : "Cost-analysis of digitization, web integration and distribution across the internet of rare and old printed materials", *Revue MIF*, n°1, 2001, 143-160

. Nous présenterons donc dans cette première section les acteurs, puis la matrice des centres de coûts, tandis que l'évaluation du PIDS (« Page Image Delivery System ») pour chacune des deux modalités fera l'objet des deuxième et troisième sections.

2.1. Présentation des acteurs de l'évaluation

Les trois bibliothèques participantes au projet présentent des structures économiques très disparates, aussi bien du point de vue de la taille de l'organisation (nombre d'employés, statut public national ou local, degré de modernisation, etc.) que de celle du catalogue, ou encore des missions définies (Bibliothèque Universitaire, Généraliste, Spécialisée, etc.)... Le tableau ci dessous présente de façon très succincte les éléments saillants des trois bibliothèques participantes au projet DEBORA.

Tableau 1 : Caractéristiques des Bibliothèques du Projet DEBORA

BIBLIOTHEQUES	Nombre d'ouvrages du catalogue global	Nombre d'ouvrages retenus	Nombre de personnels affectés aux tâches DEBORA
Bibliothèque Générale et Universitaire de Coïmbra	Un fonds ancien (antérieur à 1800) de 200.000 volumes, 6000 monographies et 1800 « non livres » par an	30	<ul style="list-style-type: none"> • Un Conservateur ou un Bibliothécaire, • Un agent technique principal.
Casanatense de Rome	350.000 ouvrages, 6000 manuscrits très précieux, 2200 incunables, 10000 gravures, 2000 périodes, quelques 7000 documents sur l'opéra et près de 15.000 ouvrages datant de la renaissance.	34	<ul style="list-style-type: none"> • Un Conservateur ou un Bibliothécaire, • Un Prestataire de service. • Un agent technique principal chargé de l'administration Web.
Bibliothèque Municipale de Lyon	Un fonds global de 1.700.000 documents dont 301.000 ouvrages, 11.000 manuscrits et 953 incunables.	35	<ul style="list-style-type: none"> • Un Conservateur ou un Bibliothécaire, • Un Agent technique principal.

2.5. Les matrices des centres de coûts

La chaîne de valeur porterienne montre aisément qu'un découpage des activités des bibliothèques peut être établi sur la base d'une localisation des centres de coûts repérables. Pour autant, cette analyse ne permet pas de spécifier la nature des coûts détaillés dans l'objet « livre numérisé ». En ce sens, un mode opératoire de sélection des coûts pertinents s'effectue sur la base de la définition des inducteurs de coûts. On peut ainsi recenser les inducteurs acceptables pour cette activité. La nature de l'inducteur de coût doit mettre en évidence un comportement de coût cohérent qui évolue proportionnellement au nombre d'inducteurs. L'expérience montre qu'il est possible de regrouper les activités en quatre grands types possibles :

- a) les activités liées aux volumes fabriqués : les inducteurs privilégiés sont l'heure de main d'œuvre ou l'heure machine ;
- b) les activités liées aux changements de lots ou de série : on retient comme inducteur le nombre d'ordre de fabrication ou nombre de séries fabriquées ;
- c) les activités de soutien concernant un objet de coût : les inducteurs principaux sont les nombres de composants, le nombre de modifications techniques ;
- d) les activités de soutien en général dont les charges sont indépendantes du volume et de la gamme de produit : ici l'inducteur s'apparente à une assiette de répartition comme la valeur ajoutée mais on peut préférer ne pas chercher à attribuer le coût de ces activités aux produits.

En procédant à l'analyse fine des différentes opérations imputables à chaque catégorie d'opération on débouche sur une grille qui permet d'établir des relations entre activités propres à chacune des ces catégories et les inducteurs de coûts caractéristiques. Il convient de souligner que des inducteurs de coûts multiples peuvent servir à caractériser la même activité. De la même façon, un seul inducteur peut caractériser plusieurs activités distinctes. Le tableau 2 donne les éléments de l'analyse pour chaque catégorie d'opérations qu'on désigne par WP pour workpackage.

TABLEAU 2 : ANALYSE DES ACTIVITES DES CATEGORIES D'OPERATION DU PROJET DEBORA

Désignation des opérations (WP)	Inducteurs Possibles
WP Collections	
Choix des livres	Nombres de critères, nombre de pages, Références par thèmes
Harmonisation du catalogage	Nombres de critères
Analyse des ouvrages	Nombres de pages, Contrôle qualité, ouvrages sélectionnés
Définition des métadonnées (haut et bas)	Traitements informatiques, nombre de pages
WP Numérisation, Extraction d'information	
Numérisation	Traitements informatiques, Nombre de pages, Ouvrages sélectionnés
Contrôle de la qualité	Temps de contrôle qualité
Analyses des images et pré-indexation	Traitements informatiques, nombre de pages
Adaptation et modifications logicielles	Traitements informatiques, nombre de pages
Conceptions logicielles	Temps de contrôle qualité
Intégration logicielles	Traitements informatiques, nombre de pages

Désignation des opérations (WP)	Inducteurs Possibles
WP Base de données d'images Adaptation de la structure de SGBI Gestion logicielles des accès Indexation des images	Nombre de métadonnées, Contrôle qualité Traitements informatiques, Nombre de pages Traitements informatiques, Nombres de pages
Intégration générale	Temps de développement et conception, Traitements informatiques
WP Client et accès Spécifications Interface WWW – DEBORA Implémentation fonctions de recherche & annotations Implémentation outils de formulation & requêtes Adaptation ARIADNE	Temps de développement et conception Temps de développement et conception Temps de développement et conception Temps de développement et conception

Une fois le travail de recensement effectué, il s'agit de déterminer les inducteurs pertinents à retenir. Cette analyse se présente sous la forme d'une matrice qui permet d'obtenir des centres de regroupements des coûts pour lesquels l'inducteur est le même. Le tableau 3 ci-dessous est une présentation des regroupements obtenus dans DEBORA.

Ici, les inducteurs les plus fréquents sont :

Le contrôle qualité (4 fois), le nombre de pages (3 fois), le stockage des données (3 fois) le nombre de métadonnées (2 fois), les traitements informatiques (2 fois), le temps de développement et de conception (2 fois), l'entretien et la maintenance (2 fois) les références par thèmes (1 fois)

L'activité sur le choix du nombre de critères n'appartient à aucun centre de regroupement et ce tableau ne nous permet pas de trancher sur la nature de l'inducteur à retenir. Ce travail de définition effectué, il est calculé un coût unitaire d'inducteur selon l'opération suivante :

TABLEAU 3 : DETERMINATION DES « INDUCTEURS DE COUTS » PERTINENTS DE DEBORA

<i>ACTIVITES DE DEBORA</i>	Collections	Numérisation, Extraction d'information	Base de données d'images	Client et accès	Total des liaisons
<i>INDUCTEURS DE COUTS POSSIBLES</i>					
Nombre de critères					0
Nombre de pages					3
Contrôle qualité					4
Traitements informatiques					2
Nombre de métadonnées					2
Temps de développement et conception					2
Références par thèmes					1
Stockage des données					3
Entretien et maintenance					2
Nombre total des liaisons	3	5	6	6	20

$$\text{Coût unitaire d'un inducteur} = \frac{\text{Coût des ressources attribuées à l'activité}}{\text{Nombre d'inducteurs}}$$

3. ANALYSE DES COUTS : FAIRE OU FAIRE FAIRE

L'objet de cette section est d'établir le coût de la PIDS de la manière la plus rigoureuse possible pour deux modalités : l'internalisation du travail (faire) et l'externalisation (faire faire).

3.1. Internalisation des activités : le cas des bibliothèques de Lyon et de Coïmbra

La détermination du coût internalisé concerne les bibliothèques de Lyon et de Coïmbra

La recherche d'une détermination de coûts « standards » pour les bibliothèques de Lyon et de Coïmbra équivaut à rechercher un coût de page

numérisée distribuable sur Internet qui inclut deux types d'éléments principaux :

- Les équipes techniques (Personnel) impliquées à tous les stades de la réalisation du projet ;
- Le matériel, comprenant les serveurs informatiques, le scanner et les différents dispositifs de stockage de données (DVDROM, CDROM et lecteurs enregistreurs). Ces éléments sont importants car leurs coûts peuvent littéralement « exploser » à raison de 5,3 fois le coût d'une impression normale et 1,2 fois le coût d'un Compact Disc.

Chaque fois que le contenu du projet sera enrichi par des collections plus importantes, d'autant sera rendu plus lourd le coût de sa réalisation finale.

Les séquences d'analyse des coûts se doivent toutes de commencer par la recherche d'une adéquation entre le niveau de besoin actuel anticipé (avec les coûts qui y sont rattachables) et le niveau de besoin futur (les coûts liés y compris). En l'espèce, il importe de souligner que la confrontation entre ces deux catégories de coût est extrêmement difficile car le domaine des technologies de l'information et de la communication sur lequel s'appuie ce projet est en plein essor aujourd'hui. Par suite, il devient quasiment impossible d'avoir une évaluation acceptable des coûts impliqués pour le futur. Dans cette dernière alternative, les coûts auront une qualification purement fictive en prenant le statut de « coûts cachés ». On pourra, éventuellement, prévoir par la suite une dotation forfaitaire budgétaire pour y faire face. Elle devra être déterminée en concertation avec les activités en charges des outils technologiques.

Une autre dimension des coûts cachés non pris en compte vise les pertes non évaluables lors de l'utilisation des périphériques de stockages. Les données obtenues montrent un taux moyen de remplissage des CDROM/DVDROM de l'ordre de 82%, soit une perte en zone enregistrable de près de 18 %. Ces coûts deviennent considérables lorsque les séries de travail prennent une taille relativement importante (numérisation d'une bibliothèque, etc.). Nous ferons l'hypothèse dans un premier temps de leur négligence pour les réincorporer lors des scénarii de passage à grande échelle.

L'évaluation du temps de travail s'est fait par chronométrage sur une période donnée et sur les déclarations des personnes impliquées dans le projet. (agent technique, conservateur). Sur cette base, il est alors possible d'obtenir le coût global du système de « production et de mise en ligne » de la page numérisée (« Page Image Delivery System » PIDS) pour chaque bibliothèque .

La structure des coûts est donnée par le tableau de synthèse 4 et 5 ci-dessous. Il convient de préciser qu'il existe deux fourchettes selon que l'on utilise le Robot ou pas. La variation du prix final de la page est de -6,2 % soit 5,42 € pour la fourchette haute et 5,1 € pour la fourchette basse.

TABLEAU 4: STRUCTURE DE COUT DE LA PIDS DE LA BIBLIOTHEQUE DE LYON

Eléments / Tâches	Pages	Coût unitaire		Coût des opérations	
		Valeur	%	Valeur	%
Coût de la sélection des documents	14 410	0,39 €	7,58%	6 024,33 €	7,58%
Coût de préparation et de numérisation	14 410	1,16 €	22,86%	16 793,17 €	22,86 %
Coût d'intégration dans la Base de données	14 410	2,74 €	53,70%	39 452,28 €	53,70 %
Construction de l'objet documentaire	14 410	0,62 €	12,19%	8 957,14 €	12,19 %
Coût de l'objet documentaire	14 410	4,91 €	—	70 769,58 €	—
Coût d'intégration, construction site, accès	14 410	0,19 €	3,68%	2 704,29€	3,68%
Coût "Page Image Delivery System"	14 410	5,1 €	100 %	73 473,87 €	100 %

A l'inverse des autres coûts habituellement trouvés dans des projets similaires, l'intégration dans la base de données ici représente près de 53 % du coût global. La raison incombe principalement aux différents traitements informatiques qui sont nécessaires dans la dématérialisation de la page seizième.

TABLEAU 5: STRUCTURE DE COUT DE LA PIDS DE LA BIBLIOTHEQUE DE COÏMBRA

Eléments / Tâches	Nombre de pages	Coût unitaire		Coût des opérations	
		Valeur	%	Valeur	%
Coût de la sélection des documents	14 098	0,58 €	11,97 %	8 239,41 €	11,97 %
Coût de préparation et de numérisation	14 098	0,67 €	13,80 %	9 494,68 €	13,80 %
Coût d'intégration dans la Base de données	14 098	2,8 €	57,32 %	39 441,46 €	57,32 %
Construction de l'objet documentaire	14 098	0,63 €	13,00 %	8 942,51 €	13,00 %
Coût de l'objet documentaire	14 098	4,69 €	—	66 118,05 €	—
Coût d'intégration, de construction du site, accès	14 098	0,19 €	3,91 %	2 693,01 €	3,91 %
Coût "Page Image Delivery System"	14 098	4,88 €	100 %	68 811,07 €	100 %

L'écart entre le coût de la PIDS déterminée à Coïmbra et celle de la PIDS de Lyon est de l'ordre de 4,47 % (0,22 €). En définitive, dans les contextes d'internationalisation des opérations de numérisation, on aboutit à des structures de coûts semblables.

La restriction ici est fortement tributaire de la nature du scanner utilisé et de la différence de coût inhérente au mode de répartition de cette charge fixe dans les pages délivrées. Le rapport d'un facteur de 1 à 10 du coût de scanner influence de façon radicale l'équivalence des coûts. Si l'on retenait un scanner identique pour les deux bibliothèques, le surcoût à Coïmbra atteindrait sans doute 20 à 25% sur la page livrable.

3.1. L'analyse des coûts en externalisation : le cas de la bibliothèque Casanatense (Rome)

La détermination du coût « standard » pour la bibliothèque de Casanatense à Rome pose un enjeu radicalement différent des deux approches précédentes relativement semblables. Si le coût de la PIDS (« Page Image Delivery System ») reste l'objectif à atteindre, il prendra ici la forme d'une externalisation de son activité de numérisation à proprement parler. En ce sens, la structuration des activités de numérisation recouvrira deux aspects distincts :

- Les équipes techniques (Personnel) impliquées à tous les stades de la réalisation du projet, mais dans cette alternative surtout à titre de contrôleur de la parfaite exécution industrielle ;
- Le prestataire de service informatique qui intervient dans la chaîne industrielle en lieu et place des anciennes séquences opératoires regroupant hommes et matériels (les serveurs informatiques, le scanner et les différents dispositifs de stockage de données, etc.).

La recherche d'une adéquation entre les besoins effectifs du projet et l'investissement financier nécessaire implique de partir d'une analyse fondée sur le coût d'opportunité lié à une telle réalisation. Si la solution de l'externalisation demeure envisageable, elle doit s'entendre des éléments suivants :

- a) La situation de la bibliothèque (compétence du personnel, volume des données, nature des données à manipuler, etc.) ;
- b) Le coût d'opportunité que représente le manque à gagner découlant d'une immobilisation de fonds affectés à l'éventuelle acquisition de matériels ;

- c) Le coût d'opportunité lié à l'affectation de surfaces de stockages physiques importantes chaque fois que la bibliothèque se trouve dans une situation de faiblesse de capacité de stockage physique soit en raison des surfaces disponibles, soit du coût du m² supplémentaire ;
- d) La mise en concurrence des fournisseurs de services à la fois en termes de qualité et de coût de facturation.

Cet aspect fondamental des coûts cachés n'est pas repris dans cette étude en raison de l'hypothèse « ceteris paribus » prise au départ entre les différentes structures bibliothécaires : en effet, elle aurait demandé la prise en compte de facteurs tels que les indices immobiliers de location industrielle et de construction entre les différents sites des bibliothèques, les dispositions de stockage et d'aménagement, de sécurité, etc. sans oublier la détermination des taux effectifs réels d'actualisation financière qui nous auraient conduit à une évaluation extrêmement sophistiquée.

Nous présentons ci-dessous la structure des coûts obtenus des éléments d'activité propres à l'externalisation.

TABEAU 6 : STRUCTURE DE COUT DE LA PIDS DE LA BIBLIOTHEQUE DE CASANATENSE

Eléments / Tâches	Pages	Coût unitaire		Coût des opérations	
		Valeur	%	Valeur	%
<i>Coût de la sélection des documents</i>	14 100	0,66 €	12,52 %	9 319,97 €	12,52 %
<i>Coût de préparation et de numérisation</i>	14 100	0,99 €	18,79 %	13 989,03 €	18,79 %
<i>Coût d'intégration dans la Base de données</i>	14 100	2,8 €	53,05 %	39 495,27 €	53,05 %
<i>Coût de l'objet documentaire</i>	14 100	0,63 €	12,01 %	8 942,51 €	12,01 %
<i>Coût de l'intégration, construction site, accès Internet</i>	14 100	0,18 €	3,63 %	2 704,29 €	3,63 %
Coût Page Image Delivery System	14 100	5,28 €	100 %	74 451,07	100 %

Le coût en externalisation d'une PIDS est à peine plus élevé que le coût en internalisation sur la base des calculs faits pour la bibliothèque municipale de Lyon.

4. SCENARII POUR UN PASSAGE A GRANDE ECHELLE

Les bibliothèques sont quasi exclusivement des établissements publics. Leurs ressources dépendent de leur catégorie. En France par exemple, l'État finance les projets nationaux, comme « Gallica » ou encore le programme de numérisation de la BNF (Bibliothèque Nationale de France). L'existence ou non d'une subvention ou d'un financement par un parrainage ou un mécénat est un aspect non négligeable de la décision relative à la numérisation, dans la mesure il serait alors possible de s'affranchir de la contrainte de coût grâce à ces différentes formes d'aide.

En ce sens, des scénarii de coûts sont à prévoir selon les solutions retenues. Nous proposons pour la suite des éléments de coûts comparatifs sur les plans juridiques et technique. Dans un second temps, nous présenterons des formes économiques de scénarii selon la répartition des charges retenue.

4.1. Eléments de coûts comparatifs

Les éléments comparables sont de deux ordres : technique et juridique. Si les aspects techniques se singularisent par un contexte concurrentiel important, influençant le choix définitif, les éléments juridiques pour des ouvrages seiziémistes demeurent plutôt en retrait. Nous explorons dans la suite ces deux aspects de coûts.

4.1.1. Les aspects techniques

a) Les scanners

S'il existe différentes possibilités pour numériser des documents, il convient de s'interroger sur la méthode la plus efficace en termes de coût et de qualité, mais surtout du type de document à numériser. La définition de la collection à numériser et son état est nécessaire. En ce sens, si la reliure des documents anciens est abîmée, il suffit de démembrer le livre et de le scanner à plat. Le document sera ainsi relié par un professionnel à la fin de la numérisation. A l'inverse, si la reliure est en bon état, il faudra utiliser un scanner de livre pour éviter de détériorer le document. De plus, l'état des pages peut influencer sur la qualité de la numérisation par exemple les tâches de moisissure.

La scannérisation feuille à feuille présente l'avantage de pouvoir automatiser cette étape, mais ne peut s'appliquer en réalité qu'à des documents de bureau. La « scannérisation » à plat ne permet pas de rectifier la courbure des pages tout en écrasant l'ouvrage sur une plaque de verre. Cette pratique

peut avoir des répercussions graves sur des collections prestigieuses et/ou fragiles. La scannérisation à livre ouvert est étudiée pour les ouvrages ne pouvant pas supporter les autres traitements. Elle n'en est cependant qu'à ses débuts, et est en cours d'élaboration. L'utilisation du berceau, qui maintient l'ouvrage à plat peut être préconisée autant pour les documents précieux que pour ceux qui ne peuvent s'ouvrir à plus de 90°. En général le choix du scanner revient le plus souvent à un sous-traitant spécialisé, mais en collaboration étroite avec le responsable du projet qui doit s'assurer que les contraintes réelles propres au fonds de la bibliothèque et à son usage sont bien prises en compte. Il faudra dans tous les cas arbitrer entre le coût, le débit et la préservation des documents originaux.

Les différentes solutions de scanners retenues dans chaque bibliothèque de DEBORA étaient :

TABLEAU 7 : COMPARATIFS DES OUTILS DE SCANNERISATION UTILISES

	BM-Lyon	BGUC Coïmbra	Casanatense
Scanner	DIGILIVRE 5600	MUSTEK	OmniScan 7000
Niveau des couleurs	256 niveaux de gris	256 niveaux de gris	256 niveaux de gris
Résolution	600 dpi	300 à 600 dpi	300 à 600 dpi
Image format	Tiff	Tiff	Tiff
Périphérique de stockage	DVDRAM	DVDRAM	CDROM
Nature de la transaction	Acquisition	Acquisition	Prestation
Coût approximatif en 2000 (les prix ne cessent de baisser)	26678,58 €	N/A ²	7 729,17 €

b) Les logiciels

Dans le cas du projet DEBORA, les offres logicielles existantes sur le marché sont quasi-inexistantes. En ce sens, il apparaît difficile d'établir une logique de choix économique des dits logiciels pour gérer la nature très spécifique du travail sous DEBORA. On distinguera deux catégories de logiciels selon la nature des besoins et des applications :

- Les *logiciels de scanning et de reconnaissance optique des caractères*³

² Il s'agit d'un scanner Mustek à plat que l'on peut évaluer à 380 euros environ.

souvent fournis gratuitement avec le scanner ne sont pas suffisants, car peu performants. Dans le cas du projet DEBORA, ont été développés des logiciels spécifiques par l'INSA-LRFV pour la compression de l'image et l'IST de Lisbonne pour la dématérialisation. Un bon logiciel doit permettre la numérisation en mode texte et en mode image. La numérisation en mode texte pose des problèmes car les caractères des documents seizièmistes ne sont pas reconnaissables. Le mode image permet une reproduction totale, mais qui comprend aussi toutes les dégradations, taches des documents, pour seulement des niveaux de gris (256 niveaux) mais aucune solution couleur ;

- Les *logiciels de gestion et de structuration de l'information*⁴ doivent permettre l'organisation des documents, c'est-à-dire leur structuration : pagination virtuelle, table des matières, indexation (Transvision 2.0, Zsilex, etc.). Par la suite, ils rendent possible l'optimisation de l'utilisation, c'est-à-dire la gestion : possibilité d'annotation, liens entre les fichiers... Enfin, ils permettent la normalisation des documents, qui va notamment être utile dans l'éventualité d'une diffusion sur Internet (utilisation par exemple du langage de programmation SGML) ;

- Les *logiciels de diffusion*⁵ de l'information proprement dits qui sont surtout des interfaces de

navigation Internet ou Intranet (Navigator de Netscape, Internet Explorer de Microsoft, etc.).

Ces logiciels sont des logiciels complexes. Ils nécessitent à la fois l'aide d'informaticiens professionnels et des formations appropriées pour les bibliothécaires. Il faut donc faire un choix en fonction du coût à l'achat, mais aussi du coût d'entretien et de formation. La décision devra également être prise en considérant les attentes et les besoins du public cible.

a) Les supports de diffusion, de stockage et de distribution :

Pour la diffusion, le choix se fait aujourd'hui quasiment uniquement entre deux formes de support : CDROM ou en ligne (Internet, Intranet...). Les

³ voir les livrables logicielles des WP 4 et suivants

⁴ Par exemple la base de données Transvision 2.0 proposée par SGBI couplée au serveur d'annotations Zsilex. Les domaines des logiciels étant en évolution technique rapide des comparaisons avec les outils actuels sont recommandables pour suivre le rythme technique.

⁵ Pour le cas de DEBORA, l'interface Ariadne adaptée au travail collaboratif est fournie en livrable du WP 7

autres formes de supports étant devenus obsolètes, c'est le CDROM qui est le plus utilisé (1,83 € à 2,29 € en moyenne pour 650 Mo). Le DVDROM est une solution de conservation intéressante, mais encore trop onéreuse à ce jour (36,59 € à 41,92 € l'unité en moyenne). La solution du serveur permet, en plus de la conservation et de la lecture des liens, la simplicité de gestion et d'accès au plus grand nombre. Il a également l'avantage d'être lisible par tous les ordinateurs. C'est la solution idéale de conservation, diffusion et distribution dans le cas du projet DEBORA.

Un stockage supplémentaire via des CDROM est envisageable pour répondre aux impératifs de sécurité

4.1.2. Les aspects juridiques

Les programmes de numérisations de documents anciens sont soumis à peu de contraintes juridiques. Il n'y a pas en effet à se soucier d'éventuels droits d'auteurs, pour plusieurs raisons. D'une part, les documents entrent dans le domaine public 70 ans après la mort de l'auteur. D'autre part, la majorité des documents anciens est anonyme. Il n'y a donc ni de problème de droits d'auteurs, ni de problème de droit moral. La seule obligation des bibliothèques est d'identifier leurs droits patrimoniaux, c'est-à-dire de réaliser une authentification des documents.

Ainsi, l'aspect juridique va prendre peu de place dans l'arbre de décision d'un programme de numérisation de documents anciens, contrairement à celui d'un programme de numérisation de documents contemporains.

4.2. Formes économiques des scénarii

On distinguera les perspectives pour chaque bibliothèque et les perspectives consortiales résultant de la coopération possible entre les différentes bibliothèques.

4.2.1. Les coûts hors coopération

Dans cette modalité, tous les coûts relatifs à l'utilisation de matériels par une bibliothèque sont intégralement répercutés. C'est la moins bonne solution économique pour les bibliothèques ayant un fonds d'ouvrages rares relativement limité. Si des impératifs de rentabilité économique sont attendus, ils situeront le projet dans une zone de gains particulièrement faible, qui peut aller jusqu'au résultat nul ou négatif. Deux cas de figures seront à retenir :

- Soit on affecte l'intégralité des charges liées au processus de numérisation dès la première année . elle correspond à l'hypothèse haute. N'ayant aucun avantage opératoire réel, nous ne retiendrons pas cette analyse ;
- Soit on reporte l'intégralité des charges *prorata temporis* de leur utilisation réelle. Elle constitue l'hypothèse basse. Celle-ci fera l'objet d'une étude ci-dessous.

En effectuant une simulation sur la base d'un nombre de pages réalisables uniquement par une bibliothèque et en raison du personnel attaché à la production d'un tel projet, les hypothèses de travail issues de l'enquête sont les suivantes :

Le nombre moyen de pages numérisées est 400 par livre . La durée de numérisation d'un ouvrage sur la période de l'étude a été évaluée un demi livre par jour (200 pages numérisées), soit 3 mois

pour les 35 livres dans l'hypothèse d'une internalisation de la chaîne de numérisation. On peut alors projeter une productivité de 140 livre annuels pour le seul agent technique en admettant qu'il y consacre la totalité de son temps. Le coût de travail d'un conservateur est de 37 350,01 € par an et 19 970,82 € pour un agent technique . La stockage des informations pour la conservation s'effectue sur la base d'un rapport 20/80 des fichiers sur respectivement les CDROM et DVD-RAM. La taille d'un fichier scannérisé étant évaluée à 8 Mega-octets en moyenne sous format *Tiff* ou *Jpeg* avant compression . Le coût du scanner retenu sera celui de la bibliothèque municipale de Lyon qui s'est doté d'un appareil avec berceau convenant particulièrement bien à la fragilité des ouvrages seizièmistes. L'ensemble des autres paramètres (ordinateur, logiciels divers, etc.) sont conservés depuis l'étude antérieure pour la simulation.

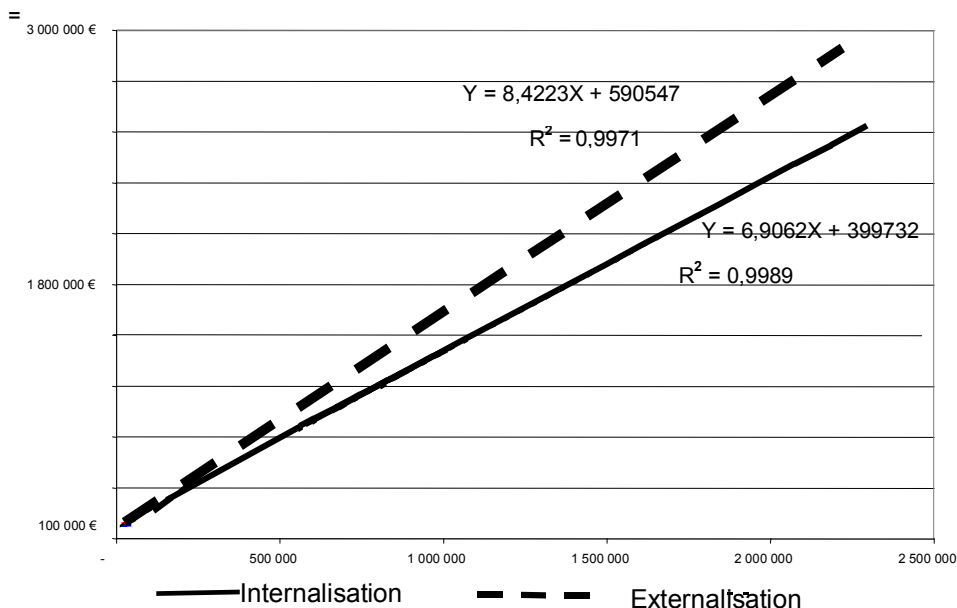
On peut admettre qu'il existe des paliers de ruptures de charges lorsque les opérations de la chaîne de numérisation se trouvent internalisées. L'une des explications envisageables se fonde sur le fait que la répartition des charges est effectuée proportionnellement à l'évolution du nombre de pages numérisées : amortissement linéaire pratiqué, masse salariale à taux constant, etc. A l'inverse, pour les données externalisées, la croissance des charges prend la forme d'une courbe exponentielle sans doute en raison de la mise en concurrence du prestataire de service (numérisation) qui fait évoluer son coût en fonction de la demande effective et de la marge attendue.

Cette structure d'évolution des charges globales a une influence relativement marginale sur le coût de la PIDS. Les simulations montrent une évolution relativement linéaire et homogène de la structure des coûts globaux

en fonction du nombre de pages numérisées. L'écart constant entre les deux modalités – correspondant sans doute à la marge bénéficiaire que le prestataire de service prélève sur le prix facturé – rend ce coût proportionnellement plus important que celui obtenu avec une internalisation des outils de production. En effet, on constate que la meilleure fonction d'ajustement possible sur les deux catégories de coûts (internes et externes) est une fonction affine – avec un degré de précision de la corrélation de 99,71% pour les coûts externes et 99,89% pour les charges internes – dont on peut ainsi déduire une structure de prévision des charges.

TABEAU 8 : ELEMENTS DE SIMULATION DES COUTS PIDS ET COUTS GLOBAUX

<i>Nbre de Pages</i>	Coût PIDS		Coût global estimé du projet		Variation % Inter/Ext
	<i>Internalisation</i>	<i>Externalisation</i>	<i>Internalisation</i>	<i>Externalisation</i>	
14 410	5,86 €	6,56 €	84 470,48 €	92 446,91 €	11,86 %
30 000	2,82 €	3,33 €	84 470,48 €	100 020,27 €	18,41 %
45 000	1,88 €	2,39 €	84 470,48 €	107 642,72 €	27,43 %
56 000	1,51 €	2,03 €	84 470,48 €	113 588,23 €	34,47 %
112 000	1,49 €	2 €	167 212,63 €	224 073,12 €	34,01 %
168 000	1,49 €	1,96 €	249 956,62 €	328 620,32 €	31,47 %
224 000	1,48 €	1,91 €	332 697,87 €	428 588,3 €	28,82 %
560 000	1,21 €	1,55 €	676 622,85 €	869 775,76 €	28,55 %
1 008 000	1,13 €	1,43 €	1 135 189,5 €	1 443 136,5 €	27,13 %
2 016 000	1,07 €	1,3 €	2 166 964,4 €	2 627 970,2 €	21,27 %

Figure 1 : Evolution des charges globales par pages numérisées et ajustement linéaire

Nous considérerons par la suite les fonctions linéaires obtenues comme des fonctions de d'offre du prix sur le marché interne des bibliothèques dans le cadre d'une recherche d'équilibre économique entre coût de fabrication et coût de cession via Internet. En ce sens, nous admettrons que chaque fois que les bibliothèques souhaiteraient mettre en œuvre un programme de numérisation, elle le font sur la base d'une fonction d'offre qui sera une dépendante de la décroissance de la PIDS.

La figure 7 qui illustre l'évolution générale des coûts de la PIDS montre qu'il existe une relation de puissance qui voit correspondre à chaque accroissement du nombre de pages dans le projet une baisse concomitante du coût unitaire de ladite page d'une fonction inverse de -0,25.

4.2.2. Les coûts pour le consortium des bibliothèques

La solution la plus avantageuse économiquement voit répartis sur plusieurs acteurs d'un unique consortium les charges liées au projet. Proportionnellement au nombre d'utilisateurs, il y a réduction du coût global et les effets de synergies et de complémentarité du catalogue commun offrent un atout commercial non négligeable. Si on retient l'idée d'un

consortium de bibliothèques composé des bibliothèques ayant participé au projet DEBORA, on peut déduire le coût final de chaque page numérisée. En fait elle sera juste fonction du nombre d'acteurs présents dans le consortium. On obtient des structures de coûts divisées par le nombre d'acteurs présents dans le consortium. Il est à noter néanmoins, dans ce dernier cas de figure, qu'il y aura des charges relatives aux déplacements des ouvrages à répercuter. Ces derniers pouvant prendre une dimension importante.

Tableau 9 : Coûts simulés dans le cas d'un Consortium de 3 acteurs

Nombre Pages	Internalisa Tion	Externalisa tion	Variations Inter/Ext
14 410	28 156,88 €	30 815,59 €	9,44%
30 000	28 156,88 €	33 340,14 €	18,41%
45 000	28 156,88 €	35 880,86 €	27,43%
56 000	28 156,88 €	37 862,7 €	34,47%
112 000	55 737,49 €	74 692,09 €	34,01%
168 000	83 318,88 €	109 540,01 €	31,47%
224 000	110 899,34 €	142 862,71 €	28,82%
560 000	225 541 €	289 925,25	28,55%
1 008 000	378 396,44 €	481 045,55 €	27,13%
2 016 000	722 321,43 €	875 990,04 €	21,27%

CONCLUSION

Si les arbres de décision (voir annexe) sont des outils précieux pour tous les organismes documentaires souhaitant mettre en place un projet de numérisation de leurs documents anciens, ils demeurent des instruments stratégiques difficiles à trouver.

S'agissant de projets similaires à DEBORA, plusieurs autres ont été mis en œuvre pour des orientations diverses. On peut citer à titre d'exemples le projet *Gallica* de la Bibliothèque Nationale de France⁶ ; le projet *Charette* du Centre d'Etudes Supérieures de Civilisation médiévale (Poitiers) ; le Département des Langues Romanes de l'université Princeton (Etats-Unis) qui représente la tradition du Chevalier de la Charrette (Lancelot, 1180)⁷ et enfin le projet *Athena*⁸ qui se propose de numériser des œuvres classiques ainsi que des textes genevois, des travaux universitaires.

Ces projets favorisent des collaborations internationales entre diverses institutions et réveillent les idées utopiques des Lumières autour du grand partage du savoir où « *les bibliothèques participent (...) à la construction de (...) l'intelligence répartie.* »⁹

Si « *le livre est le moyen le plus sûr de transmettre un héritage* »¹⁰, il faudra du temps pour que dans les esprits le document numérique remplace le support papier. De tout temps, les « bibliothécaires » et les « documentalistes »¹¹ ont eu pour mission de conserver et de distribuer l'information par tous les moyens quittent à les transférer d'un support à l'autre. La numérisation est une nouvelle étape, elle répond à de nouvelles exigences financières, techniques de diffusion et de préservation de ces documents.

ANNEXE : LES MATRICES DE DECISIONS

⁶ <http://gallica.bnf.fr/>

⁷ <http://www.mshs.univ-poitiers.fr/cescm/lancelot/>

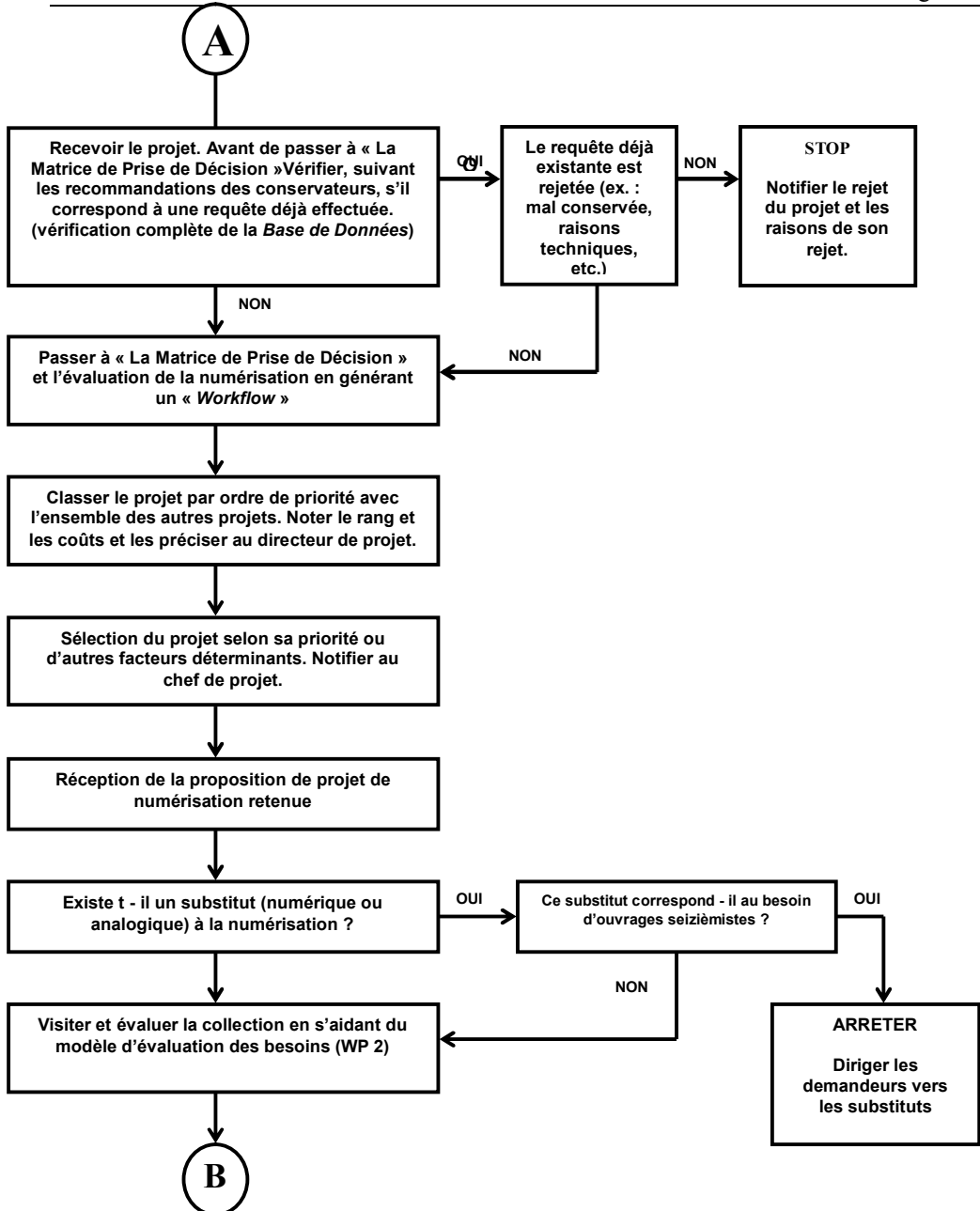
⁸ <http://un2sg4.unige.ch/athena/html/authors.html>

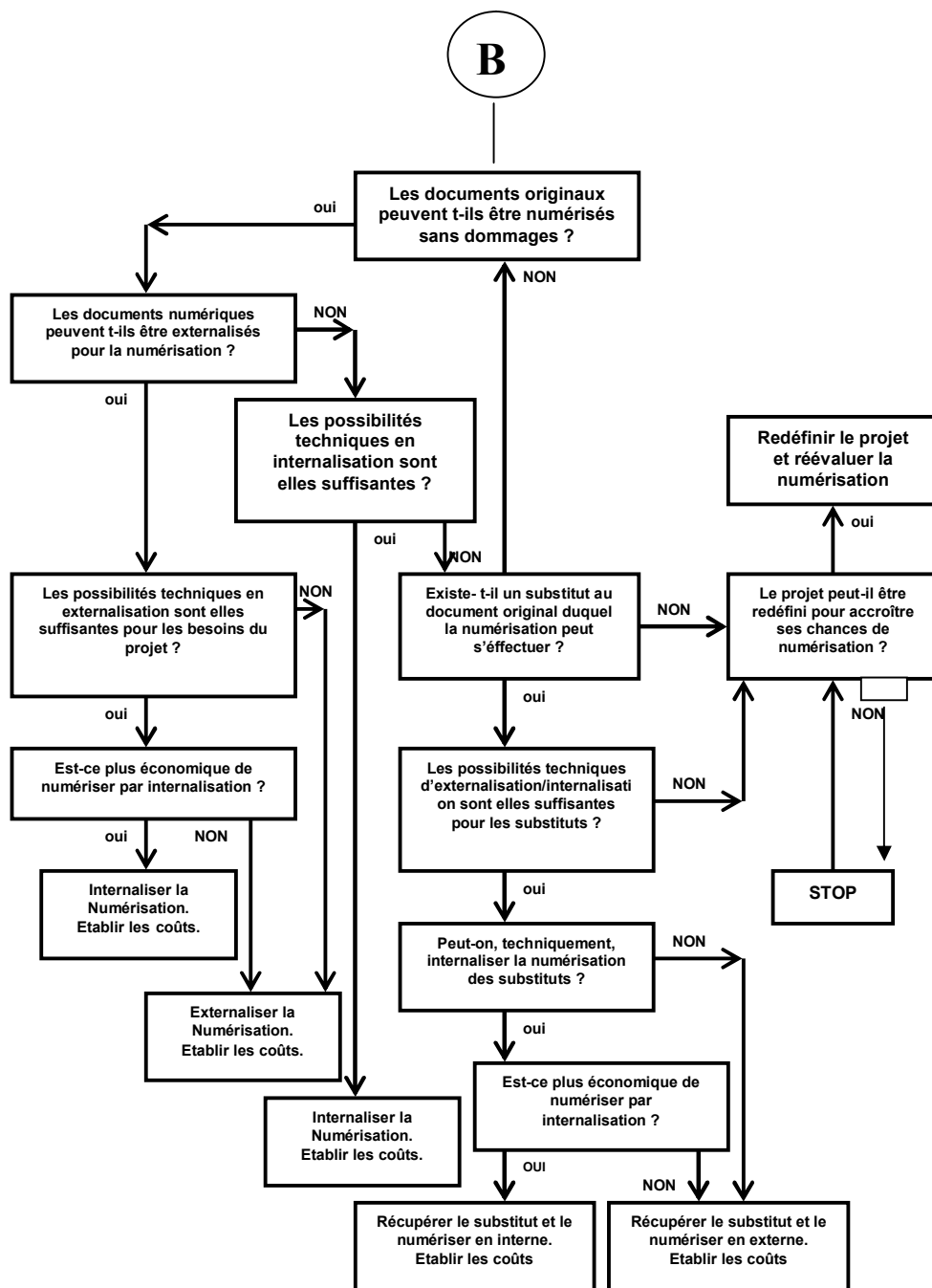
⁹ Conférence « La bibliothèque virtuelle : une antinomie ? » *National Library of Medicine*, Washington, 1998.

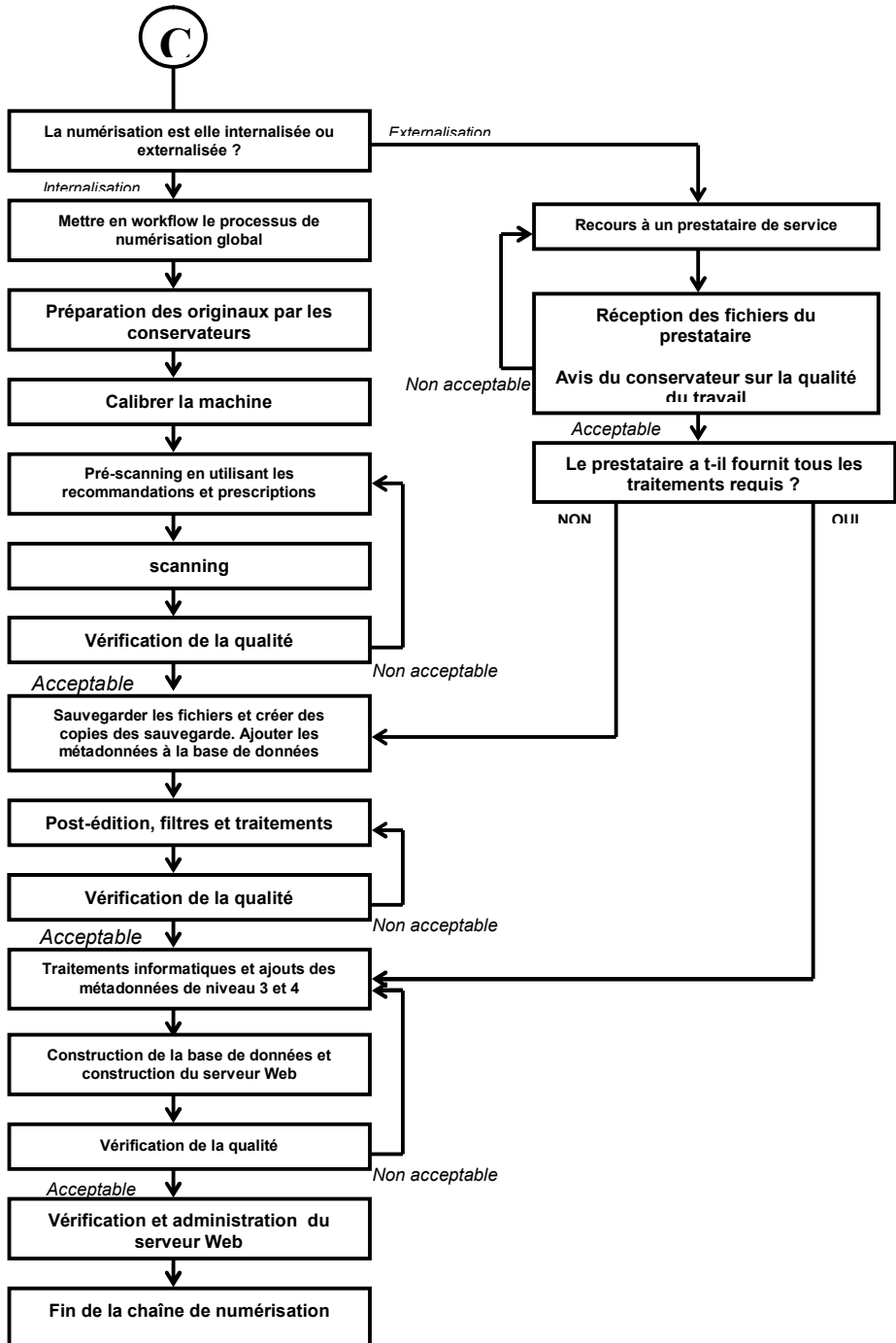
¹⁰ Direction du Livre et de la Lecture, *Protection et mise en valeur du patrimoine des bibliothèques* : Recommandations techniques. Paris : Direction du Livre et de la Lecture, 1998, p.7.

¹¹ Ce terme est utilisé de manière anachronique, il est d'usage récent et s'adresse aux personnes dont l'activité professionnelle à trait aux documents, il n'était pas employé pour parler par exemple des moines qui recopiaient à la main les manuscrits avant l'invention de l'imprimerie

Les matrices présentées ici sont des adaptations des principaux arbres de décisions rencontrés dans la littérature consacrée à la numérisation des livres. On distingue trois arbres : le premier (A) concerne l'analyse du projet de numérisation ; le deuxième (B) analyse la problématique du faire (internalisation) et du faire-faire (externalisation) ; le troisième (C) analyse les différentes étapes dans le processus de numérisation jusqu'au contrôle de la qualité.







BIBLIOGRAPHIE

- Besser H., Yamashita R., (1998), *The Cost of digital image distribution*, <http://sunsite.berkeley.edu/Imaging/Databases/1998mellon>
- Dalbéra Jean-Pierre "Culture et société de l'information. Numériser le patrimoine : un enjeu collectif". *Culture et recherche*, n°77, mars-avril 2000, pp. 3-5
- Ducharme Christian (sous la direction de-) *Du CD-ROM à la numérisation : développer les documents numériques en bibliothèque*. Villeurbanne, Institut de formation des bibliothécaires, 1997, Collection La Boîte à outils, Mémento, 172 p.
- Dupoirier Gérard (coord.), *Les bibliothèques numériques*, Paris : Editions Hermès Sciences, 1999
- [INF99], *Put it here - AIIM 1999 : Storage Summary*, *Inform*, vol. 13, Issue 6, June.
- Jacques André (coord.), Chabin Marie-Anne , *Les documents anciens*, Paris: Editions Hermès Sciences, 1999
- Jacquesson Alain, Riviers Alexis , *Bibliothèques et documents numériques : concepts, composantes, techniques et enjeux*. Paris : Editions du Cercle de la Librairie, 1999, 377 p.
- Kaplan R. S., Cooper R., (1997), *Cost & effect : using integrated cost systems to drive profitability and performance*, *Harvard Business School*.
- Kenney A. R., (), *Digital to Microfilm Conversion : A demonstration Project, 1994-1996*, <http://www.library.cornell.edu/preservation/com/comfin.html>
- Kenney A. R., Rieger O. Y., (1998), *Managing Digital Imaging Projects : an RLG Workshop*, RLG May.
- Lee Stuart D. (1999), *Decision Matrix for proposed Digitization Projects*, University of Oxford
- Library of Congress, *NDLP Project Planning Checklist*, 1997
- Lowry C., Troll D., (1996), *Virtual Library Projet*, *Serials Librarian*, NASIG Proceedings : Tradition, Technology and Transformation, Part 1, vol. 28, n° 1/2.
- Puglia Steven (1999), *The costs of Digital Imaging Projects*, *RLG DigiNews*, October 15, vol. 3, n° 5, <http://www.rlg.org/preserv/diginews/diginews3-5.html>
- Lenga S.D., Rey J. et Silem A.: "Cost-analysis of digitization, web integration and distribution across the internet of rare and old printed materials," *Revue MIF*, n°1, 2001, 143-160
- Tanner S., Lomax Smith J. (1999), *Digitisation : How much does it really cost ?*, September, <http://heds.herts.ac.uk/HEDCinfo/Papers/drh99.pdf>