

enssib

école nationale des sciences
de l'information et des bibliothèques

Diplôme de conservateur de bibliothèque

Rapport de Stage

**Création, management et développement des
collections numériques**
Stage du 4 septembre au 24 novembre 2000 à la
Digital Unit, The Research Libraries, New York
Public Library

Olivier Bosc

Sous la direction de Michael J. Alexander
Head of Digital Unit, Research Libraries, New York Public Library

Décembre 2000



M 2000 DCB ST 04

enssib

école nationale des sciences
de l'information et des bibliothèques

Diplôme de conservateur de bibliothèque

Rapport de Stage

**Création, management et développement des
collections numériques**

**Stage du 4 septembre au 24 novembre 2000 à la
*Digital Unit, The Research Libraries, New York
Public Library***

Olivier Bosc

Sous la direction de Michael J. Alexander
Head of Digital Unit, Research Libraries, New York Public Library

Décembre 2000

Mes remerciements à Sharon Frost et à Heike Kordish qui m'ont ouvert les portes de NYPL, ainsi qu'à Bob, Terry, Nicole et au reste de la brillante équipe de la *Digital Unit* avec qui j'ai eu un immense plaisir à travailler.

Introduction

1. Présentation de l'institution d'accueil : The Research Libraries of New York Public Library

La *New York Public Library*¹ est l'une des premières bibliothèques au monde². Les chiffres – 51 millions de documents, un réseau de 85 bibliothèques de quartier, 4 centres de recherche, un budget pour 1999 de 223 millions de dollars, 3500 employés – ainsi que l'histoire – fusion en 1895 de l'*Astor and Lenox libraries* et du *Tilden Trust*, lancement en 1901 à l'instigation d'Andrew Carnegie d'un réseau de bibliothèques de prêt, construction en 1902 d'un bâtiment au centre de New York – font de cette bibliothèque une institution de premier ordre dans le paysage mondial des bibliothèques. Cependant le gigantisme des chiffres et la symbolique des événements ne font pas de NYPL une institution monolithique, bien au contraire. C'est ce que notre stage au sein des *Research Libraries* – Bibliothèques de recherche – à la *Digital Unit* – bibliothèque numérique – nous a, entre autres choses, permis de découvrir.

1.1. NYPL, une bibliothèque à l'image d'une ville et à l'échelle du monde

Généralement présentée, trop unilatéralement peut-être, comme la bibliothèque municipale de New York, NYPL est aussi une bibliothèque patrimoniale et de recherche d'envergure internationale. Ces deux missions distinctes expliquent sa partition en deux structures, les *Branch Libraries* d'une part, qui remplissent des missions de lecture publique, et les *Research Libraries* d'autre part, orientées vers la recherche et la

¹ Désormais désignée dans ce texte par l'acronyme *NYPL*.

conservation. De carrure internationale les collections de ce second ensemble étaient estimées en 1999 à 41 millions de documents³, dont 14 d'imprimés, alors que le nombre de ses usagers avoisinait les 2 millions. Les bibliothèques de recherche regroupent 5 entités distinctes : *Humanities and Social Science Library* [Bibliothèque des « humanités » et des sciences sociales], *Library for the Performing Arts* [Bibliothèque des Arts du spectacle], *Schomburg Center for Research in Black Culture* [Centre de recherche *Schomburg* sur la culture noire], *Science, Industry and Business Library* [Bibliothèque de science, d'industrie et des affaires] et depuis mai 2000 la *Digital Unit* [Bibliothèque numérique]⁴.

NYPL est à l'image de la ville qu'elle sert. Ses collections reflètent les divers groupes sociaux, nationaux ou religieux qui par le passé et aujourd'hui encore ont façonné New York. Mais ce sont aussi des individus qui par le don de collections ou par des aides financières à la bibliothèque marquent et ont marqué leur attachement profond à la ville. Cependant si les collections des bibliothèques de recherche sont de par leur histoire typiquement « new-yorkaises » leur composition et leur richesse font de la bibliothèque une institution culturelle de premier plan au niveau national et international. Par opposition à la Bibliothèque du Congrès qui a les prérogatives et les obligations d'une bibliothèque nationale, *NYPL*, essentiellement à travers le rôle moteur de ses *Research Libraries*, se montre plus ouverte sur le monde⁵.

1.2. NYPL, une institution privée au service du public

Les grandes étapes de la création et du développement de *NYPL* ont essentiellement reposées sur la générosité de grands et de petits mécènes. Ce système, par-delà les interventions croissantes de l'état fédéral au fil du XX^e siècle, a perduré en se renforçant jusqu'à nos jours. La

² Voir sur ce point l'annexe 1 où figure une présentation factuelle de l'institution.

³ Contre 11 millions d'imprimés pour les *Branch Libraries*.

⁴ Pour plus de détails voir l'organigramme reproduit dans l'annexe 2.

⁵ Cette division du travail est clairement visible dans les programmes de numérisation. La bibliothèque du congrès concentre sur ce plan ses efforts sur des collections ayant trait à l'histoire américaine – programme *American Memory* – *NYPL* s'occupe pour sa part de thèmes beaucoup plus larges et variés.

bibliothèque, comme l'immense majorité des institutions culturelles américaines, est une institution privée. Fondation elle doit rechercher l'équilibre financier et rendre des comptes précis de l'utilisation des moyens mis à son service par l'Etat fédéral, la ville ou de généreux donateurs. Cependant ce souci de bonne gestion est commun à l'ensemble des bibliothèques, qu'elles soient privées ou publiques, et c'est peut-être dans l'environnement économique plus global, la loi du milieu, à laquelle l'institution se trouve très exposée, qu'il faut rechercher l'origine d'une posture confinant au paradoxe. A l'heure où l'information devient une marchandise *NYPL* se trouve en concurrence avec des entreprises privées, délivrant elles-aussi du contenu. Par opposition aux universités qui peuvent supporter les coûts élevés de diffusion de l'information vers un public délimité, la bibliothèque fait face à une mission difficile : s'adresser au public sans limitation et de manière gratuite. Ce problème se pose essentiellement pour les *Research Libraries* qui concentrent dans leurs collections le contenu patrimonial et sont tenues de rester accessibles à tous ceux qui le souhaitent. Ce principe fondateur se trouve à l'heure numérique simultanément renforcé par les nouvelles possibilités techniques et menacé par leurs coûts et les aléas liés à leur utilisation et à leur développement.

1.3. The Research Libraries, une institution de recherche au sein d'une bibliothèque publique

La particularité de *NYPL* est d'offrir ses collections à une communauté de chercheurs potentiellement illimitée. Des infrastructures d'accueil en perpétuelle augmentation, des outils mis à la mesure et à la disposition du public et une politique d'accréditation des lecteurs relativement souple pour les collections spécialisées permettent de servir un vaste public. A ce titre *NYPL*, cumulant des collections axées « recherche » et une mission de service public apparaît comme une institution unique, une sorte de « bibliothèque publique de recherche », sans grand équivalent dans le monde.

Mais les *Research Libraries* apparaissent surtout comme la vitrine prestigieuse de *NYPL*. Leur image composite mêlant haute culture et accès au plus grand nombre permet d'affirmer l'identité et de rendre plus visible la mission véritable de la bibliothèque. Les donateurs sont sensibles et attachés à cette image. Elle devient un enjeu vital pour la survie et le développement de l'institution. A l'intérieur de *NYPL* ce rôle moteur des *Research Libraries* se confirme sur un autre plan. En effet, elles constituent les secteurs les plus dynamiques de l'institution chargés d'impulser les réformes et d'introduire les nouvelles technologies. A terme les *Branch libraries* bénéficient pleinement des expériences menées dans ce domaine comme plus largement du prestige qui en découle pour *NYPL* dans son ensemble.

2. Présentation du service : The Digital Unit

Le stage s'est effectué au sein de la *Digital Unit*⁶. Entre la demande de stage en mai et son début effectif en septembre de profonds bouleversements ont affecté cette structure. Par une décision récente des instances dirigeantes de la bibliothèque la *Digital Unit* a été érigée en un département à part entière, à égalité avec les autres composantes des *Research Libraries*. Cette décision traduit l'importance de l'expérience acquise par ce service dans le domaine de la numérisation. Mais surtout elle illustre la politique volontariste mise en place dans ce domaine par la direction de la bibliothèque. Sous l'instigation de William Walker, *director*, mais surtout de Heike Kordish, *deputy director*, sous l'autorité de laquelle la *Digital Unit* est directement placée, la numérisation est devenue une priorité stratégique au sein des *Research Libraries* et delà pour *NYPL*⁷.

⁶ Ce service correspond à celui de la Bibliothèque numérique à la Bibliothèque Nationale de France.

⁷ Les enjeux de cette montée en puissance sont nombreux : dynamiser l'activité des *Research Libraries*, redonner à *NYPL* un seul visage, affirmer l'image de haute culture à laquelle les donateurs sont sensibles, bâtir un programme à la page sur le plan technologique correspondant au standing – culturel et symbolique – de l'institution, enfin inclure la bibliothèque dans la compétition actuelle pour la mise en ligne de contenu.

Les premières initiatives concernant la constitution de collections numériques remontent aux années 1990. Autour d'une équipe restreinte, Anthony Troncale pour les prises de vue et Robert De Candido pour le traitement des données, la *Digital Imaging Unit* a accumulé un savoir solide dans le domaine de la numérisation et du traitement des images fixes. Plusieurs programmes de numérisation ont ainsi été menés à bien. Les succès remportés dans ce domaine ont fait de *NYPL* une structure pionnière dans ce domaine, ce qui a motivé ma demande pour effectuer en stage sous la responsabilité d'Anthony Troncale.

L'obtention d'une bourse de 5 millions de dollars, émanant d'un donateur individuel et anonyme, pour la numérisation en 5 ans de 600 000 images (programme des *Visual Treasures*) en avril 2000 a bouleversé la donne. Immense succès pour l'équipe en place ce nouveau programme représente aussi un défi remettant en cause son organisation et son fonctionnement. En effet, la structure doit désormais grandir pour adapter ses capacités de production. C'est dans cette optique qu'a été créée, au milieu de cette année, la *Digital Unit*, ayant rang de département au sein des *Research libraries*. Elle est nouvellement dirigée par Michael J. Alexander, venu de la *British Library* en septembre 2000 afin de superviser le travail des structures déjà en place (*metadata unit* et *digital imaging unit*). Les effectifs du service sont passés en quelques mois de 3 personnes à plus d'une quinzaine.

Placé sous l'autorité de Michael J. Alexander et sous l'égide d'Anthony Troncale et de Robert De Candido, j'ai intégré l'équipe de la *Digital Unit* pour participer au programme de numérisation des *Visual Treasures*, les Trésors visuels de la *New York Public Library*.

3. Présentation de la collection : Visual Treasures

La bibliothèque n'en est pas à son coup d'essai dans le domaine de la numérisation⁸. Cependant le nouveau programme concernant la numérisation de 600 000 images tirées des collections de la bibliothèque sur la période 2000-2004 est inédit de par son ampleur. Ambitieux, il vise à numériser des photographies, des gravures, des ouvrages illustrés en mettant l'accent sur la qualité visuelle des matériaux sélectionnés et en cherchant à mettre en lumière l'infinie variété des collections détenues par *NYPL*.

Bien que d'autres programmes sont en cours⁹ l'ampleur des *Visual Treasures* mobilise l'essentiel des énergies à la *Digital Unit*. Aussi à mon arrivée il a été décidé d'un commun accord que je travaillerais à la numérisation d'une collection déjà sélectionnée dans le cadre de ce projet : *The Journal of Saints*. Aux yeux des responsables du stage sa durée relativement courte et l'importance présumée du travail nécessaire à la numérisation de la collection ne permettait pas un traitement complet de l'ensemble des matériaux. Aussi la mission a été définie de manière à concentrer le travail sur une fraction de ceux-ci, pour permettre un passage par toutes les étapes nécessaires à la construction d'une collection numérique : description, rassemblement et/ou création des métadonnées, numérisation et construction d'une infrastructure *Web* pour la diffusion. J'ai ainsi travaillé à la construction d'un modèle permettant d'achever ensuite dans un délai plus contenu la numérisation de l'ensemble de la collection. Les bénéfices d'une telle procédure sont clairs pour les deux parties prenantes. La bibliothèque ouvre un chantier nouveau, portant certes sur une fraction seulement d'une collection, mais en s'assurant de disposer ensuite d'un modèle éprouvé pour son achèvement. Pour sa part, le stagiaire fait un apprentissage concret de

⁸ Voir l'annexe 3 qui décrit les différentes collections numériques déjà constituées à ce jour.

⁹ Ils concernent des documents venus de la *Map division* [Département des cartes et plans], du *Schomburg center*, et de la *Performing Arts library* [Bibliothèque des Arts du spectacle].

toutes les étapes nécessaires à la création et au management des collections numériques. En ayant un point de vue global sur le processus en cours, j'ai pu appréhender concrètement l'enchaînement des différentes opérations. Le déroulement du stage s'est fait selon ce programme, mais la rapidité dans son exécution a décidé les responsables du stage à me confier une autre mission concernant le développement futur des collections numériques de la *NYPL* dans le domaine du son et de l'image et animée. Ce faisant ils m'ont permis de connaître à fond et de faire l'expérience de toutes les étapes nécessaires à la constitution d'une collection numérique : création, management et développement.

Créer une collection numérique : The Journal of Saints

Après une visite complète de l'établissement les différentes étapes du processus de création d'une collection numérique m'ont été présentée de manière générale. L'expérience acquise par la *Digital Unit* par le passé l'a conduit à bâtir une véritable chaîne de production de document numérique (annexe 3 bis). J'ai donc réalisé pour chaque étape les opérations nécessaires à la numérisation de la collection qui m'a été confiée. Elle a donc traversé, et moi avec elle, les différents stades nécessaires à la création d'une collection numérique.

1. Sélection des matériaux

La sélection des matériaux est l'étape première dans la constitution d'une collection numérique (cf. annexe 4). Elle doit s'effectuer en liaison avec les responsables des collections. C'est un long travail de consultation et de projection. Au début du stage le choix des responsables du stage s'est arrêté sur *The Journal of Saints* conservé à la *Print division*. La taille contenue, l'indéniable intérêt graphique de cet ensemble, mais surtout l'existence d'un fichier papier, élément essentiel dans la construction des métadonnées, ont motivé ce choix. Cohérent avec le vaste

programme des *Visual Treasures* (cf. annexe 4 bis) cette décision s'est faite en accord avec le conservateur de la *Print division*.

1.1. Les différentes collections en cours de numérisation

A l'heure actuelle la *Digital Unit* travaille à la numérisation des collections suivantes dans les cadres des *Visual Treasures*¹⁰ :

The Avery collection, gravures et eaux-fortes françaises réunis par le collectionneur Avery et datant de la seconde moitié du XIX^e siècle. Cette collection est conservée à la *Print division*.

The Russian photos albums, albums photographiques représentant des églises russes à la fin du XIX^e siècle. Cette série d'albums est conservée à la *Slavic and Baltic division*.

Les ouvrages d'égyptologie illustrés, allemands, français et italiens du XIX^e siècle conservés à l'*Oriental division*.

Ces collections ont été choisies pour leur intérêt historique et leurs qualités visuelles.

1.2. The Journal of Saints

The Journal of Saints est selon la terminologie américaine un *scrap book* (album). Il se présente en douze volumes reliés en cuir rouge de format *in quarto*. Chacun des douze albums représente un mois de l'année dont le nom est porté sur la tranche en latin. Chaque volume contient un sommaire manuscrit pour chaque jour du mois énumérant, toujours en latin, une série de Saints ou d'événements en rapport avec la vie des Saints. Ces Saints ou ces événements sont matérialisés par des images collées. Ces vignettes, des gravures découpées, sont de toutes les tailles, de toutes provenances, de toutes les époques et d'auteurs différents, le plus souvent anonymes. Chacun des douze volumes comporte environ 200 pages et rassemble autour de 500 images. Pour l'ensemble de la collection c'est plus de 1500 Saints qui sont ainsi représentés.

¹⁰ Cf. également l'annexe 4 ter pour plus de détails.

Quoique populaire dans les collections de la *Print division*, les informations concernant l'origine et la provenance de cet ouvrage sont rares. De toute évidence il s'apparente à un martyrologe romain illustré. Selon notre propre enquête, il pourrait s'agir d'une production d'origine bollandiste. Les images les plus anciennes remontent au XIV^e siècle et la plus récente à 1803. Le martyrologe romain ayant peut-être servi de référence pour sa rédaction pourrait être celui de Benoît XIV rédigé en 1745 et dans lequel apparaissent pour la première fois les Saints fêtés en Angleterre. Nous n'avons pas poussé plus avant les travaux d'identification puisqu'ils n'étaient pas nécessaires à la numérisation et échappaient très clairement à notre compétence.

La collection étant imposante il a été décidé que je ne travaillerai dans le cadre du stage que sur le premier volume, janvier, en vue d'apprécier la faisabilité de l'opération et de faire le point sur les problèmes rencontrés. Mon travail a donc visé à numériser partiellement la collection, tout en travaillant à la fabrication d'un modèle permettant de la compléter ultérieurement. Cette méthode a été appliquée avec succès lors des projets précédents par la *Digital Unit*.

2. Récupération des données et création des métadonnées

La première étape a consisté à rassembler ou à créer des informations décrivant la collection. Ces informations indispensables à la numérisation sont appelées les métadonnées¹¹. C'est l'étape initiale qui permet ensuite d'organiser la numérisation en même temps que la collection numérique elle-même. Les fichiers images doivent être rattachés de manière permanente à leur description (notice catalographique, source, description factuelle) et être distribués selon une architecture préexistante à leur création. Numériser en mode image toutes les pages d'un livre revient à rendre autonomes les éléments d'un tout qui

ne l'était à l'origine pas. Les metadonnées servent à rattacher entre eux des fichiers numériques physiquement différents mais issus d'un même document analogique.

Ce raisonnement s'applique également à un niveau plus global. En effet, la cohérence intellectuelle d'une collection doit être préservée. La création des metadonnées permet de rendre permanente à travers la base de données cette continuité. Mais les metadonnées servent principalement à la fabrication d'index et de répertoires permettant une recherche à l'intérieur de la collection. C'est une étape stratégique dans la création d'une collection numérique.

Pour *The Journal of Saint*, ce travail s'est déroulé au sein de l'équipe de Robert De Candido spécialiste des metadonnées à la *Digital Unit*. La base de données actuellement constituée et qui contient les collections déjà numérisées repose sur le logiciel Microsoft *Access*. C'est un outil peu sophistiqué et limité à deux niveaux d'arborescence : objet, et sous-objet. Mais il est d'usage simple – le système en place a pu être construit sans aide extérieure – et permettra dans un proche avenir une migration aisée vers un système plus élaboré de type Oracle.

L'existence d'un index des noms sur fichier papier regroupant environ 1500 fiches semblait pouvoir simplifier la tâche de description de la collection. La transcription de cet index s'est effectuée sur un traitement de texte en saisissant – séparé d'un signe de tabulation – les informations figurant sur chaque fiche : nom du saint, date auquel il apparaît dans la volume, numéro du volume, numéro de page. Ce fichier texte est destiné à être converti afin d'alimenter la base de données.

Au fil de cette transcription – quelque peu fastidieuse – la piètre qualité du fichier papier s'est révélée : fautes de frappe, inexactitudes dans les transcriptions, manque de rigueur (certains noms étaient traduits en anglais - Peter pour *Petrus* – alors que d'autres non)¹². En conséquence j'ai suggéré pour le mois de janvier une comparaison avec le document

¹¹ Pour avoir un aperçu des méthodes en place dans ce domaine à la *Digital Unit* cf. les annexes 6 et 6 bis.

¹² Ces erreurs doivent apparemment être imputés aux volontaires longtemps employés par la bibliothèque dans ce genre de tâches peu gratifiantes.

original afin de rectifier les erreurs et contrôler pour chaque page quel était le Saint présent sur les illustrations. La vérification complète a conduit à de nombreuses rectifications. Il a fallu à l'aide d'outils bibliographiques identifier chaque Saint ou événement et, en cas de doute vérifier la cohérence entre date et Saints. Les noms en latin apparaissant sur l'ouvrage ont été systématiquement délaissés pour leurs équivalents en anglais¹³, lorsqu'ils existaient, en construisant en parallèle un fichier de conversion. Ce travail sur le document s'est poursuivi par une investigation, partielle, sur la base d'autorité *Rlin*. À terme ce recoupement doit être effectué pour l'ensemble de la collection en vue d'assurer la création et l'utilisation de données fiables et de qualité pour la construction des index.

L'étape suivante a consisté à réfléchir sur l'architecture de la base données. Le logiciel *Access* limite le nombre des niveaux de hiérarchisation des données à deux : objet et sous-objet. L'objet est toujours la collection elle-même et dans le cas présent il s'est agi des douze volumes intitulés : *The Journal of Saints*. Il restait à définir le sous-objet qui devait donner sa physionomie définitive à la base de données. Deux idées, apparaissant de prime abord évidentes, ont été écartées (une définition par Saint¹⁴ ou par date¹⁵) et c'est finalement la page qui a été définie comme sous-objet¹⁶. Chaque sous-objet est rattaché à un nom de fichier et à un numéro de séquence qui correspondent au futur fichier image numérisé.

Une fois cette architecture définie il a été possible de faire entrer les données dans la base générale en leur attribuant un titre et un numéro de collection. Cette étape permet l'impression d'une *carrier sheet* (feuille de travail, cf. annexe 6 ter) où sont résumés ces informations et qui sert de

¹³ Ce choix a été fait car la collection est en priorité destinée à un public anglophone.

¹⁴ Puisqu'il n'y a pas que des Saints représentés mais aussi des événements et que plusieurs Saints peuvent apparaître sur une même page, voire sur une même image...

¹⁵ Car les dates peuvent se répéter d'une page à l'autre puisqu'il y a plus d'une page par date.

¹⁶ La page est liée au numéro de volume et le titre de chaque sous-objet fait référence à la date [jan. 1] ou pour les pièces liminaires à leur nature [frontispice].

référence aux opérateurs chargés de la numérisation proprement dite du document.

3. Processus de numérisation des images

Nous avons effectué, après une rapide initiation aux techniques de la photographie numérique, la numérisation complète du volume de janvier du *Journal of Saints* dans le laboratoire de la *Digital Imaging Unit* dirigé par Anthony Troncale¹⁷. Composé de 4 personnes ce service dispose de 4 appareils photos numériques et de 2 scanners à plat pour les prises de vue et de deux postes de travail dit de *post-processing* où sont effectuées les opérations d'étiquetage et de contrôle de la qualité. L'expérience accumulée dans ce domaine par la *Digital Unit* est énorme. En ayant depuis plusieurs années son travail sur la qualité des documents, elle a mis au point des méthodes efficaces et fiables.

3.1. Manipulation et conservation

Les exigences techniques considérables que demandent les méthodes actuelles de numérisation ne doivent pas faire oublier la fragilité des collections et des matériaux. L'objectif de la numérisation des collections anciennes est bien souvent d'éliminer les risques liés à leur communication traditionnelle. Il ne faut pas que ce traitement leur soit plus dommageable sur ce plan. En plus des questions habituelles liées à la sécurité de collection précieuses, l'exposition à la lumière et des manipulations répétées doivent rendre ce travail efficace et rapide.

A la *Digital Imaging Unit* le travail *in situ* permet de limiter les risques pour les collections les plus fragiles et les plus précieuses. La numérisation des matériaux les plus transportables se fait, elle, à l'extérieur.

¹⁷ Cf. la présentation de ce service dans l'annexe 7.

3.2. La photographie numérique

Les techniques de numérisation des images se sont stabilisées ces dernières années¹⁸. Avec un parc de machines modernes important et des procédures très établies, fruit d'une longue expérience, la *Digital Imaging Unit* est une structure remarquablement efficace.

Après une rapide initiation technique j'ai entamé la numérisation du *Journal of Saints* sur un appareil *Kontron Progress 3012* couplé au logiciel de traitement des images *Photoshop*. Le banc de numérisation est éclairé par deux lampes au tungstène spécifiques, car la régularité dans l'alimentation lumineuse est essentielle pour la qualité. Les images ont été numérisées en couleur avec une définition de 350 *dpi* au format *Tif*. La procédure est simple : mise en route du système, éclairage, étalonnage, installation du document sur un support, cadrage, contrôle de la luminosité, mise au point, photographie, enregistrement. Les fichiers créés sont nommés selon la *carrier sheet* et enregistrés sur un disque de stockage. J'ai ainsi numérisé 200 images d'une taille de 20 mb chacune en l'espace d'une dizaine de jours.

Manager les collections numériques : un impératif à l'heure de la production de masse d'images numérisées

On l'aura compris à la lecture des chiffres la numérisation d'une petite collection d'images nécessite déjà des structures conséquentes. Le programme des *Visual Treasures* portant sur 600 000 images dépasse très clairement les capacités actuelles. A ce titre le management des collections numériques dans l'optique d'une production de masse d'images devient une priorité. La prise de conscience d'une telle nécessité

¹⁸ La période expérimentale est terminée alors qu'une littérature de plus en plus abondante est disponible.

nous est apparue au fil de la numérisation de notre collection mais elle s'est encore plus clairement révélée lors de la visite d'un groupe de conservateurs français dont l'organisation nous a été confiée par Heike Kordish¹⁹. La rédaction du programme et la conduite de la visite m'a permis d'envisager de manière concrète le fonctionnement global de la *Digital Unit* et ses transformations actuelles. Il s'agit par une politique à long terme d'envisager la production des collections numériques en terme de management.

1. Le *post processing* : repenser les méthodes actuelles

A l'issue de la création des fichiers images un certain nombre d'opérations sont menées. Elles visent à assurer la qualité et l'utilisation des fichiers images.

1.1. Contrôle qualité

La qualité des clichés est essentielle. Le fichier *Tif*, conservé dans son format original, sans retouches ni compression, constitue un document précieux qui est archivé. C'est vers lui qu'on pourra s'appuyer dans l'avenir en cas de perte d'information ou de destruction de l'original analogique. Aussi le contrôle qualité le concernant doit être établi selon des standards pré-définis visant à détecter tous les défauts de l'image. Pixels morts, effets de moiré, problèmes de cadrage, de luminosité sont détectés et donnent lieu à de nouveaux tirages. A la *Digital Imaging Unit* ce contrôle est toujours réalisé par un opérateur différent de celui qui a réalisé la numérisation. J'ai donc été formé à ce travail sur une autre collection.

¹⁹ Visite le 25 octobre 2000 de 8 conservateurs français (musées et bibliothèques) sous l'égide de la *Franco-american Foundation*.

Ce contrôle s'effectue aussi pour les collections numérisées par un prestataire de service extérieur. Ceci permet d'évaluer sur pièce la qualité du travail et, sur le long terme, de prévenir toute dérive éventuelle.

Les opérateurs de la *Digital Imaging Unit* sont formés à l'ensemble des tâches, de la prise de vue au contrôle qualité. Deux postes de travail sont dévolus à celui-ci. Dans un proche avenir ces capacités doivent augmenter en liaison avec le recours à des sociétés extérieures pour la numérisation des collections les moins fragiles. C'est ainsi qu'un ensemble de 500 vues stéréoscopiques ont été envoyés à une société extérieure à des fins de numérisation.

1.2. Duplication des fichiers, étiquetage, gravage des CD

L'étape consécutive consiste à reproduire les fichiers en les modifiant selon les besoins. Ces derniers doivent être définis lors de la mise en place de la chaîne de production, pour satisfaire différents besoins. Il faut, en plus du fichier archive, créer un fichier de service, ainsi qu'un fichier dit de référence et une imagerie (*thumbnail*) pour le *Web*. Ces trois types de fichiers sont étiquetés avec le logiciel utilitaire *TiffTag* pour permettre une identification ultérieure sans ouverture du fichier lui-même.

Le fichier archive est ensuite enregistré sur un CD, sans modification, au format *Tif*. Sa taille est de 20 Mo.

Le fichier de service est une copie du fichier archive toujours au format *Tif* dont les bords sont recoupés (afin d'éliminer la barre de code couleur utilisée lors de la prise de vue) et retravaillée à l'aide de *Photoshop* pour améliorer sa qualité visuelle. Il sert aux éventuelles demandes de reproduction ou pour des travaux d'édition, ou d'exposition. Sa taille est identique au fichier archive.

Les fichiers *Web* sont deux copies du fichier de service, réduites et converties au format *Jpeg*. L'une, image de référence, mesure autour de 100 Ko l'autre, imagerie (*thumbnail*), mesure 10 Ko.

The Journal of Saints représente un ensemble de 800 fichiers image, occupant 900 Mo d'espace disque. La collection a nécessité le gravage de 7 CD nommés et étiquetés selon un code précis (cf. annexe 7 bis).

2. Archivage et conservation : l'occasion d'un nouveau départ

NYPL n'a jusqu'alors pas considéré la question de l'archivage des données numériques comme prioritaire. Jusqu'ici le CD a été le seul outil de stockage des collections numérisées²⁰. Le déclin programmé de ce support et l'augmentation des cadences de production offre l'occasion de réfléchir à l'utilisation de nouvelles techniques.

2.1. Les méthodes actuelles

Le gravage sur CD est une méthode simple et souple. Les collections numérisées sont stockées sur ce support. Les procédures actuelles d'étiquetage des fichiers ont été conçues pour permettre en cas d'obsolescence du format d'enregistrement une identification du fichier sans ouverture. En plus les CD sont identifiés par des étiquettes en papier collées sur le disque et sur le boîtier. Ils sont ensuite stockés sur place. Cependant ces supports ne sont pas catalogués de manière traditionnelle et leur inventaire actuel nous a semblé très approximatif.

2.2. Des perspectives nouvelles

Récemment *NYPL* a lancé un programme de création d'une unité de stockage pour ses données numériques. La perspective d'une production de masse de fichiers images a provoqué cette réflexion en même temps que l'afflux programmé dans les collections de documents « nés numériques » (*born digital*). Les fichiers actuellement stockés sur CD vont être transférés sur disque dur. Sans entrer dans les détails techniques il s'agit d'un stockage en ligne sur un ensemble de disques durs d'une

²⁰ La politique suivie par le passé est décrite dans l'annexe 7 ter.

capacité de plusieurs millions de Mo. Un serveur doit permettre le transfert automatique des fichiers dès leur création. L'intégrité des archives, leur conservation, est assurée par une procédure robotisée de contrôle des données. Ce faisant la bibliothèque espère combler son retard dans le domaine de l'archivage des données numériques. L'adoption de cette procédure est récente et sa mise en place vient à peine de commencer avec l'acquisition d'un serveur.

3. Mise à disposition et diffusion

Là encore, le programme des *Visual Treasures* nécessite la remise à plat des méthodes actuelles. Par le passé les collections numérisées ont été mises en ligne sur www.nypl.org dans les pages de la *Digital Library* en construisant des infrastructures spécifiques à chaque collection. Aujourd'hui ce système, malgré son succès dont son audience témoigne (cf. annexe 5), est inadapté car il demande un travail lourd et complexe.

3.1. Adapter les méthodes à la politique de diffusion

Le programme des *Visual Treasures* vise un public illimité d'utilisateurs. A la fois sur place et en ligne ces collections vont être accessibles. Le système de présentation doit être orienté vers l'utilisateur tout en respectant la collection sur le plan graphique mais surtout intellectuel.

En premier lieu, qu'il soit érudit ou novice le lecteur/spectateur doit être également satisfait lors de la consultation proprement dit de la collection (esthétique signifiante, ergonomie soignée et intuitive, niveaux de recherche distincts, respect de l'intégrité intellectuelle des œuvres et des collections).

Ensuite le système doit établir un bon équilibre entre image et texte (création d'index, association étroite entre image et information relative à l'image, qualité intellectuelle de cette information)

Enfin la qualité de la consultation, c'est-à-dire l'équivalence entre la consultation du document original et son double numérique doit être

optimale (rapidité d'accès au document, possibilité de zoom, sélection d'images particulières).

Le *Web*, par la création de pages Html où sont intégrées les images, offre toutes ces possibilités et c'est cet outil qui a été jusqu'ici privilégié par *NYPL*. Cependant la taille du projet actuel incite là encore à innover en adoptant un logiciel spécifique permettant le stockage et la présentation des images : le logiciel *Insight* de la société *Luna*.

3.2. Rédaction d'une proposition Web pour le Livres des Saints

L'étape suivante du stage a consisté à travailler dans le service du *Web Development* pour réfléchir à la construction d'une infrastructure permettant de rendre accessible *The Journal of Saints*. Il s'est agi d'un exercice car dans l'avenir c'est le logiciel *Insight* qui assurera la présentation des images de cette collection comme de l'ensemble des supports numériques image de *NYPL*. En accord avec M. Misner, du *Web Development*, une architecture générale a été définie. J'ai ensuite réfléchi sur la mise en œuvre et l'amélioration de cette structure en produisant une nouvelle proposition (annexe 9). Ensuite à l'aide du logiciel *Dreamwaver* j'ai mis au point une maquette du site *Web* en construisant plusieurs pages (annexe 10).

L'accueil réservé à ma proposition a été très positif, la direction insistant notamment sur ses qualités graphiques.

3.3. Une solution pour la gestion des grands nombres : Le logiciel Insight de la société Luna

L'adoption du logiciel *Insight* par *NYPL* est la conséquence directe de la taille du projet en cours et son orientation vers l'image. Déjà utilisée par de grandes institutions muséales américaines, le logiciel *Insight* permet la présentation, la manipulation et l'étude d'images numérisées sur une très grande échelle. Il est possible d'utiliser sa version pour le *Web* sur le site du célèbre collectionneur américain de cartes David Rumsey à l'adresse suivante : www.davidrumsey.com. Ses fonctionnalités sont par ailleurs

abondamment décrites par son fabricant (www.luna.com). Une démonstration par une équipe de cette société m'a permis d'apprécier les capacités de cet outil.

Facile à personnaliser, un point important pour conserver l'identité de la bibliothèque, il permet des recherches élaborées et offre, par une puissance de zoom remarquable, un confort de consultation étonnant. Pleinement sécurisé, au dire de ses développeurs, il empêche toute exportation frauduleuse des images. Cependant il faut remarquer qu'il laisse peu de place aux notices catalographiques et au texte de manière générale. De plus il paraît très dépendant du système d'exploitation *Windows*²¹.

A ce jour la *Digital Unit* a envoyé une première série de 5000 images avec leur métadonnées à la société *Luna*. Elles servent à la mise au point de la version client d'*Insight* destinée à *NYPL*. Une procédure de chargement automatique permettra d'entrer les nouvelles collections dans ce système dès leur création. C'est, là encore, une modification radicale des méthodes en place dictée par l'ampleur du projet en cours.

Développer les collections numériques : au-delà des Visual Treasures la numérisation de l'image animée et du son des collections d'art du spectacle de NYPL

Avec la rédaction d'une proposition pour le *Web* concernant *The Journal of Saints* le programme touchait à son point final. Le fragment de cette collection était passé par toutes les étapes du circuit de production d'une

collection numérique. Il aurait été possible de poursuivre la numérisation en passant au volume de février. Mais le responsable du stage, Michael Alexander, a jugé préférable de me confier une autre mission concernant l'accroissement des collections numériques de *NYPL*.

1.1. Nouveau projet, nouvelle mission pour le stagiaire

A la demande du directeur de *NYPL*, Paul LeClerc, une étude concernant la numérisation des collections de son et d'image animée de la *Performing Art Library*, structure rattachée aux Bibliothèques de Recherche, a été lancée. En effet l'un des donateurs fortunés de *NYPL* a fait part de son vœux de financer un projet dans le domaine des Arts du spectacle. Prenant la forme d'un document succinct, la proposition devait donner une vision globale de la pertinence d'une telle opération, des possibilités techniques actuelles au coût de cette entreprise. La mise au point et la rédaction de ce document m'a donc été confiée par M. Alexander qui a insisté sur le nombre et le type de documents concernés et surtout sur le chiffrage précis de son coût.

1.2. Veille technologique

La première étape a consisté à faire le point des projets actuellement en cours dans le domaine de la numérisation de l'image animée et du son à travers le monde. Il a fallu également s'informer sur les techniques actuelles et sur leur coût²². Ce travail de veille technologique s'est essentiellement déroulé sur le *Web*.

Il est impossible de rendre en quelques lignes l'ampleur de cette recherche cumulée sur plusieurs jours. Précisons simplement que le site du *Performing Arts Data Service* (www.pads.ahds.ac.uk) constitue une ressource essentielle et concernant les projets en cours dans ce domaine (son répertoire est très précieux). Plus largement le répertoire *Padi* de la

²¹ Il fonctionne avec les formats propriétaires favoris de *Microsoft*. S'en servir à distance avec une machine tournant avec *Linux* plutôt que *Windows* et en utilisant *Netscape* plutôt qu'*Internet Explorer* ne donne pas de très bon résultats.

²² A ce jour *NYPL* n'a que ponctuellement mis en œuvre ce genre de technologie. Le projet portant sur les *PFA* n'est pas destiné au *Web* ce qui règle la douloureuse question des droits.

Bibliothèque Nationale d'Australie (www.nla.gov.au/padi/topics) fait utilement le point sur les projets de numérisation les plus avancés et propose des ressources multiples sur tous leurs aspects.

La technologie actuelle rend possible la numérisation de l'image animée et du son et leur diffusion par les réseaux à haut débit. Seules les incertitudes concernant l'archivage de ces matériaux et la permanence des formats en l'absence, à l'heure actuelle, de normes établies et reconnues, constituent des obstacles de nature finalement plus économique que technique.

1.3. Evaluation des collections de la Performing Arts Library

The New York Public Library for the Performing Arts rassemble la plus grande collection au monde consacrée aux arts du spectacle. Pour prendre la mesure de cette richesse on pourra consulter leur description à l'adresse suivante : www.nypl.org/research/lpa/lpa.html.

La seconde étape a consisté à rencontrer les responsables de ces collections sous la houlette du sous-directeur M. Tolleson afin de connaître les priorités en termes de numérisation. J'ai donc rencontré les responsables du *Rodgers and Hammerstein sound archive* et de la *Jerome Robbins Dance Collection* et évalué avec eux les objectifs d'un éventuel programme à grande échelle. Il est clairement apparu qu'une telle opération ne devait pas uniquement concerner les collections anciennes mais offrir les moyens de récolter les documents numériques qui échappent aujourd'hui à la collecte, faute d'infrastructures techniques adaptées.

1.4. Chiffrage du coût et rédaction de la proposition

L'ultime étape a consisté à rédiger un document en deux parties (annexes 10 et 11). La première partie, un texte d'une page, a visé à mettre en perspective l'intérêt d'un tel programme à l'échelle de la *Library for the Performing Arts* mais aussi de *NYPL* en insistant sur les deux aspects de la numérisation (collections patrimoniales et documents « nés »

numériques). Le second document rassemble les chiffres et un calendrier prévisionnel pour l'exécution du projet. Supervisés par Michael Alexander ces documents ont ensuite été remis à Paul LeClerc le mardi 21 novembre.

Conclusion

Le bilan du stage a clairement mis en lumière la cohérence du travail que j'ai effectué au sein de la *Digital Unit*. Le passage en grandeur nature par les différentes étapes de la construction d'une collection numérique pour *The Journal of Saints* m'a permis d'en connaître, en temps réel, les différents aspects. Dans le même temps, j'ai été associé aux différentes réunions de travail concernant l'organisation et le fonctionnement du service. J'ai été pleinement intégré à l'équipe en travaillant à un projet clairement délimité dont les résultats seront dans l'avenir, je l'espère sincèrement, pleinement exploitables.

Plus largement le stage m'a permis d'envisager la délicate transition que traverse, avec le passage à la production de masse, la *Digital Unit*. Ce qu'une terminologie en vogue appelle le *Digital Asset Management*, ce qui pourrait se traduire par Gestion des actifs numériques, nécessite une transformation radicale des méthodes jusqu'ici utilisées. Il s'agit d'un saut qualitatif comparable au passage d'une production artisanale à une production industrielle. La numérisation du premier volume du *Journal of Saints* s'est effectuée selon l'ancienne méthode, mais l'imminence de la grande transformation a été clairement perçue. L'évolution radicale des procédures touche tous les secteurs de la *Digital Unit*. Le système de la base de données est sur le point d'être remplacé par un logiciel Oracle plus puissant, le recours aux prestataires extérieurs va être accru, l'archivage réorganisé. La généralisation du format Xml, déjà utilisé pour certaines collections, ainsi que la construction d'un système d'information sur mesure est en gestation.

Enfin la mission de prospection qui a couronné le stage m'a permis de comprendre les enjeux technologiques futurs dans le domaine, plus que

jamais mouvant, du numérique et son rôle stratégique dans la vie et le développement des bibliothèques.

Annexes

**Annexe 1 : " The New York Public Library. The
Research Libraries. The Branch Libraries.
General Fact Sheet 2000 "**²³

²³ Document interne.

General Fact Sheet 2000

Background The New York Public Library, Astor, Lenox and Tilden Foundations, is a private corporation responsible for the operation of 4 research centers and 85 branch libraries. It is guided by a 42-member Board of Trustees. The Mayor, the Comptroller, and the Speaker of the City Council are ex-officio members. Samuel Butler serves as Chairman of the Library's Board of Trustees. Dr. Paul LeClerc serves as the Library's President.

The Research Libraries depend heavily on private contributions. These noncirculating collections constitute one of the world's major research libraries and the greatest of all privately supported libraries open to the public. Four centers comprise The Research Libraries: the Humanities and Social Sciences Library; The New York Public Library for the Performing Arts; the Schomburg Center for Research in Black Culture; and the Science, Industry and Business Library, which opened in the former B. Altman building in 1996.

The 85 Branch Libraries, located in the Bronx, Manhattan, and Staten Island, are primarily supported by the City and State of New York. They offer circulating and reference collections and a wide variety of services to more than six million people of all ages who live and work in the New York area. The branches also provide literacy tutoring, English-language instruction, and special services for people with disabilities and the unemployed. Brooklyn and Queens have separate library systems.

History The New York Public Library was created in 1895 by the consolidation of the privately financed reference libraries of John Jacob Astor and James Lenox with the Samuel Jones Tilden Trust. The City of New York agreed to erect and maintain the Library at Fifth Avenue and 42nd Street, a national landmark designed by John Mervyn Carrère and Thomas Hastings. In 1901, the existing New York Free Circulating Library merged with The New York Public Library; Andrew Carnegie offered to build new circulating libraries if the City would provide the sites and make adequate provisions in its annual budget for their continued maintenance. The New York Public Library was then contracted by the City to operate these libraries.

Facts and Figures

as of June 1999

The Research Libraries

Locations: 4 – Humanities and Social Sciences Library; The New York Public Library for the Performing Arts; Schomburg Center for Research in Black Culture; and the Science, Industry and Business Library

Collections: 41,709,454 items
(13,991,994 books)

Users in FY99: 1,974,992

Number of items consulted: 2,162,459

Reference inquiries (telephone): 257,017

Materials preserved: Through restoration, preservation, microfilming, and reprint: 110,811

Other items: 17,134

Hours of recorded sound: 637

Feet of motion picture film: 36,000

Hours of videotape media: 407

Volumes cleaned: 52,807

Employees: 1,141 (346 are hourly)

The Branch Libraries

Locations: 85 – in Manhattan, the Bronx, and Staten Island

Collections: 10,421,691 items (4,341,485 books)

Users in FY99: 13,145,709

Library card holders: 1.86 million

Reference inquiries: 6,739,512
(90,363 telephone reference service)

Directional inquiries: 5,746,416

Items circulated: 12,652,723

Distribution of free material: 202,177

Internet user sessions: 1,622,318

Number of free programs: 21,898

Free program attendance: 419,539

Number of CLASP (Connecting Libraries and Schools Project) programs: 4,036

CLASP program attendance: 99,339

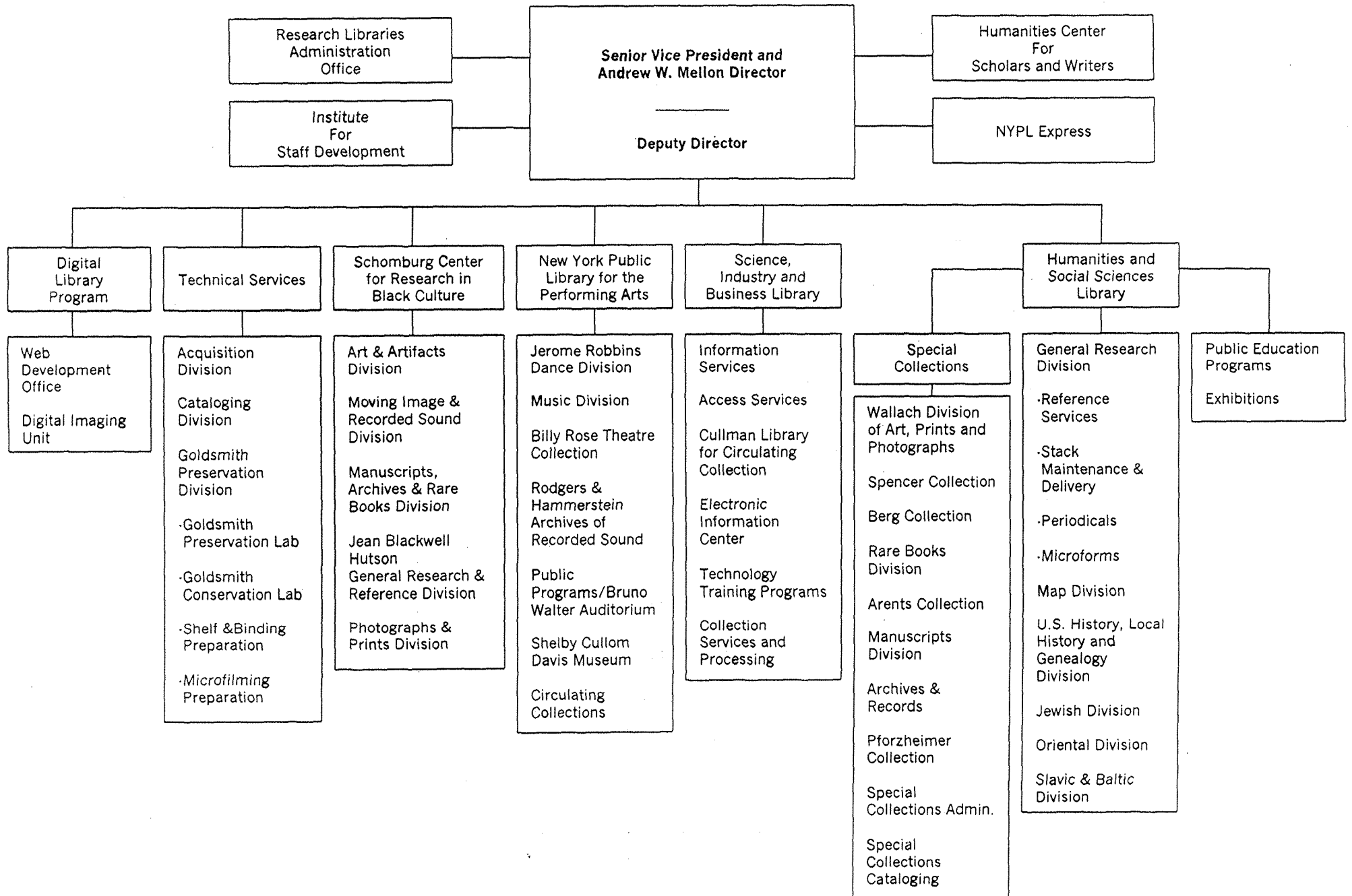
Number of branch exhibits: 2,336

Employees: 2,448 (719 are hourly)

Annexe 2 : " The New York Public Library. The Research Libraries. Organization Chart 2000 "²⁴

²⁴ *Ibid.*

August 2000



Annexe 3 : " Digital library collections factsheet "²⁵

²⁵ Extrait des pages de l'Intranet local de NYPL, *Lair*, consacrées à la numérisation.

digital library collections factsheet

FY 1998	FY 1999 (projected) Collections	FY 2000 (projected)
<ul style="list-style-type: none"> >Images of African Americans >Nineteenth-Century Women Writers >Dance Heritage Coalition >Finding Aids >Abbott and Javitz 	<ul style="list-style-type: none"> >Small-Town America (Dennis) >Images of Dance >Black New York (WPA) >Ortiz Bibliography >Schomburg Finding Aids >Travels along the Hudson >OCLC Schomburg >Marriage, Women, and the Law 	<ul style="list-style-type: none"> >Millennium >Delmas Finding Aids >Making of America II >IMAGINE >Digital Scriptorium II >Digital Archives
Digital Objects		
<ul style="list-style-type: none"> >1,500 images >50 page images >5,000 pages of encoded text >10 video clips 	<ul style="list-style-type: none"> >15,500 images >109,000 page images >17,000 pages of encoded text >1 VR file 	<ul style="list-style-type: none"> >56,000 images >121,000 page images >26,500 pages of encoded text >10 audio clips >10 video clips >10 VR files
Functionality		
<ul style="list-style-type: none"> >searching text >paging through documents >browsing topics >viewing images 	<ul style="list-style-type: none"> >cross-collection searching >setting aside images >streaming audio/video >ordering online >integrating bib record data into Web DBMS 	<ul style="list-style-type: none"> >selecting audio/video frames >zooming in on details >webcasts of events >payment online >connectivity between OPAC and Web DBMS
Web Statistics (for digital.nypl.org)		
<ul style="list-style-type: none"> >300,000 hits >20,000 user sessions >80 countries >1,500 subscribers to newsgroup 	<ul style="list-style-type: none"> >1,000,000 hits >50,000 user sessions >80 countries >5,000 subscribers to newsgroup 	<ul style="list-style-type: none"> >2,000,000 hits >100,000 user sessions >80 countries >7,500 subscribers to newsgroup
Staff:		
<ul style="list-style-type: none"> >3 RL staff (FTE) >1 ITG staff (FTE) 	<ul style="list-style-type: none"> >4 RL staff (FTE) >1.5 ITG staff (FTE) >10 interns 	<ul style="list-style-type: none"> >9 RL staff (FTE) >3 ITG staff (FTE) >10 interns
Software/Hardware		
<ul style="list-style-type: none"> >Relational database (Allaire's Cold Fusion) >SGML-to-HTML (Inso's DynaWeb) <hr/> <ul style="list-style-type: none"> >UNIX server (Sun Solaris) >NT server >CD-ROM storage 	<ul style="list-style-type: none"> >Object-relational database (Informix's Web Datablade and Universal Server) >Text Searching (Excalibur's Text Datablade) >Image Handling(Excalibur's Image Datablade) >On-the fly compression (Lizardtech's Mr. Sid) <hr/> <ul style="list-style-type: none"> >UNIX server (Sun Enterprise) 	<ul style="list-style-type: none"> >Audio Handling (Musclefish's Audio Retrieval Datablade) >Video Handling(Video Foundation's Video Datablade) <hr/> <ul style="list-style-type: none"> >Handle server (name resolution) >Storage server (lossless files) >DVD storage

**Annexe 3 bis : " Digital library collections.
Forthcoming collections "²⁶**

²⁶ *Ibid.*

digital library collections

forthcoming collections

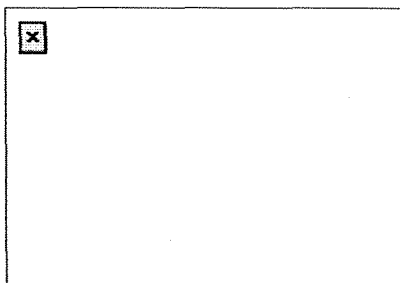
Images of Dance

- *Content:* Photographs documenting the history of dance. Project also includes non-digital components: video documentation and oral history.
- *Budget:* \$50,000.
- *Budget Notes:* \$8,200 for expenses related to digitizing (contractual scanning, storage, shipping); \$9,900 for staff devoted to all phases of project.
- *Funding Source:* New York State Council on the Arts.
- *Timeline:* March 1997 - December 1998.
- *Status:* Scanning and cataloging complete. Web site scheduled for Summer 1999.
- *Product:* Web access to 2,000 photographs.

Black New York

- *Content:* Featured are biographical sketches, studies relating to cultural achievements, history, slavery, economics, sports, theater, and churches. Authors of the studies include Ralph Ellison, Abram Hill, and Ellen Tarry.
- *Budget:* \$50,000.
- *Funding Source:* In-house.
- *Timeline:* August 1998 - August 1999.
- *Status:* Contractual scanning of page images complete. Text conversion and SGML encoding completed to begin in October 1998.
- *Product:* 9,000 page images and searchable texts.

Travels Along the Hudson



- *Content:* Text, images, prints, photographs, and maps documenting railroad and river travel along the Hudson River in the 19th Century. A cooperative project with 9 institutions in New York State.
- *Budget:* \$61,000.
- *Funding Source:* New York State
- *Timeline:* January 1998 - December 1998.
- *Status:* Selection of materials and cataloging complete for demo site. Materials being scanned for demo Web site. Web site to be completed by Spring 1999.
- *Product:* Web access to thematic treatment of 19th Century social and economic history of the New York State Hudson River Valley.

Performing Arts Archival Finding Aids

- *Content:* Finding aids to manuscripts collections in the performing arts at the Library for the Performing Arts, Schomburg Center, and Center for the Humanities.
- *Budget:* \$210,000.
- *Funding Source:* Delmas foundation.

- *Timeline:* January 1998 - January 2000.
- *Status:* First group of finding aids sent to vendor for encoding. Project staff hired. Finding aids will be available on the Web by Summer 1999.
- *Product:* Web access to finding aids for 200 archival collections.

Digital Library for the Performing Arts



- *Content:* Selections from performing arts collections in a variety of media.
- *Budget:* \$300,000.
- *Funding Source:* NEA Millennium program.
- *Timeline:* January 1998 - December 1999.
- *Status:* Selection complete, cataloging ongoing. Scanning to begin in Fall 1998. Web site will be available in Winter 2000.
- *Product:* Web access for 50,000 items, including sound and film.

Making of America II

- *Content:* Primary source material documenting the history of transportation in the U.S., 1869-1900. From collections at UC Berkeley, Cornell, Penn State, Stanford and NYPL, a subset of the membership of the Digital Library Federation (DLF).
- *Budget:* \$350,000.
- *Budget Notes:* \$20,000 to NYPL for digitizing costs.
- *Funding Source:* DLF, National Endowment for the Humanities.
- *Timeline:* July 1998 - July 1999.
- *Status:* Selections completed. Finding aids being encoded in EAD. Scanning and text encoding scheduled for Fall 1998. Web site to be completed in Summer 1999.
- *Product:* Project emphasizes developing best practices for technical architecture of digital library projects. Web access to virtual collection contributed by project participants. NYPL contribution: diaries and photos of William Henry Jackson and Robert Brewster Stanton, and New York Central Railroad archives.

Digital Scriptorium II

- *Content:* Medieval illuminated manuscripts from NYPL, Columbia, UC Berkeley, the Huntington Library and other manuscript repositories.
- *Funding Source:* National Endowment for the Humanities.
- *Timeline:* May 1999-April 2001.
- *Status:* Award notification in late Fall 1998. An SGML-based model for encoding the descriptions of medieval manuscripts has been designed and is being tested.
- *Product:* Phase II of a project will develop a prototype Web-based world union catalog of medieval manuscripts.

Digital Library Projects
Last updated 2/15/99, pe

Annexe 4 : " Selection criteria for Digitization projects. Content "27

²⁷ *Ibid.*

SELECTION CRITERIA FOR DIGITIZATION PROJECTS - CONTENT

1. COPYRIGHT

Digital projects must be undertaken with a full understanding of ownership rights, difficult as they often are to ascertain, and with full recognition that permissions are essential to convert materials that are not in the public domain.

If a proposed project involves materials in the public domain, the work can proceed. If the source materials are protected by copyright, and appropriate permissions can be obtained, the work can move ahead.

If permissions are not forthcoming for copyrighted sources, the materials cannot be reproduced and the focus of the project must change.

2. INTELLECTUAL NATURE OF THE SOURCE MATERIAL

a. Does the intellectual quality of the source material warrant the level of access made possible by digitizing?

Materials of marginal scholarly value are best left in their original form or made accessible in a less costly manner.

Most users would opt for electronic access to original source material rather than to derivative works; or to the papers of a prominent person over the administrative records of a department. This is a subjective assessment and judgment must apply.

b. Will digitization enhance the intellectual value of the material?

Texts can be encoded and made fully searchable to allow comparisons between successive drafts or editions, or translations; can be moved into a personal computer and into a word processing program, shared with other users and manipulated for multiple purposes. Digitized prints, drawings, and other visual resources can be viewed in groups at low resolution or inspected individually for very fine detail.

Digital charts and tables can be loaded easily into statistical software for additional analysis, etc.

c. Will electronic access to a body of information add significantly to its potential to enlighten, or are the original books, manuscripts, photographs, or papers sufficient to the task?

A collection of thousands of images, however promising a resource, might be nearly unapproachable because of its size and the conditions and dimensions of individual items. Well-indexed and in digitized form, however, the collection could be searched with relative ease for images of a particular person, location, or some indexed characteristic.

d. To what extent will the combination or aggregation of original sources increase their value?

Digitizing related material, like building a coherent collection of paper copies, can strengthen the context within which each title is approached. Ephemera are most often useful when studied in the aggregate, as are posters, broadsides, and popular literature. One can digitize scattered holdings to facilitate the combinations, and comparisons that are otherwise precluded by the fragility, value, and dispersion of the original images.

3. CURRENT AND POTENTIAL USERS

a. Are users now consulting the proposed source materials? Are the materials being used as much as they might be?

This is a complicated question. Ephemera produced by a community political organization may be of great interest to local scholars, and of limited value to a worldwide audience. But, if the use is heavy and widespread, digitizing may guarantee convenient and reliable access. JSTOR is an example of a project focussing on high-use material only.

b. Is current access to the proposed materials so difficult that digitization will create a new audience?

Low use may signal that a collection has marginal intellectual value, but there are many other reasons for valuable materials to generate little interest - its remote location, or highly restrictive access policy; bibliographic access may be poor. Broader access, as it creates new

users, can facilitate more active interest.

c. Does the physical condition of the original materials limit their use?

Some resources are too fragile to be consulted. Aging newspapers, manuscripts that are too fragile to be browsed. In such cases, a digital copy might be provided to improve access, and to protect originals from handling.

d. Are related materials so widely dispersed that they cannot be studied in context?

Cooperative efforts to digitize disparate pieces of a greater whole can create or restore a more usable collection. Papyrus fragments, far-flung correspondence, scattered photographs, broken runs of a title are some of the materials whose coherence, accessibility, and utility can be enhanced through digitization.

e. Will the proposed digital files be of manageable size and format?

Digital resources need to match user's technical capabilities and equipment. Will images delivered require special software or viewers or printers. Are files so large, that connections are far too slow to support browsing at satisfactory speeds?

These limitations are not reasons to rule out digitization, since the trend is toward greater capabilities. The more important the resources available electronically, the greater the incentive to acquire the technology to use these resources.

f. Will digitization address the needs of students and scholars?

Immediate demand can inject a measure of practical reality into decisions to create electronic resources. Ready access to shared electronic files can transform use, and projects in support of such demand can have greater support and financial backing.

4. ACTUAL AND ANTICIPATED NATURE OF USE

An object encounters few barriers to use. One might

have to handle an object carefully, or use a magnifying glass to read fine print, but in general, the work is immediately approachable. The same resource, when digitized, should be equally accessible and approachable. Ideally, the electronic version will also permit new kinds of uses and more sophisticated types of analysis.

a. How do readers use the existing source material? What approach to digitization will facilitate their work?

Different digitizing techniques result in electronic files with different characteristics. This in turn, can respond well or poorly with different readers needs. E.g. if the goal is to provide an image-based finding aid to help users identify original materials of interest, then mounting slow-loading high-resolution images would be counter-productive. On the other hand, if the intention is to reduce or eliminate handling of original materials, an image that fails to convey all critical information will fail to serve its intended purpose.

ASCII texts do not readily replicate the structure and format of an original document; but these capabilities are possible in marked-up texts (SGML). Different approaches carry very different costs, benefits, and resource requirements.

b. Will digitization increase the utility of the source material? Will it enable new kinds of research use, or other innovative uses?

Thumbnail images of visual resources can be browsed to discover patterns, trends, and relationships among individual items; and specific images can then be scrutinized at higher resolution. Electronic transcriptions of texts in marked-up files can be linked to bitmap images of original documents. Comparisons of different versions of a text are likewise simplified. Related texts and images can be assembled together within a single, unified corpus.

Projects intended primarily for research tend to be less aggressive in ordaining relationships among sources, since their creators assume that researchers will build their own structures of meaning.

c. Are there other scholars, librarians, and experts who can collaborate to create a useful product?

Colleagues and potential users can clarify ideas, help select meaningful materials for conversion, improve project design, and stimulate early interest. "User demand" reflects both the intrinsic utility of specific source materials as well as a social context of participation and promotion.

5. THE FORMAT AND NATURE OF THE DIGITAL PRODUCT

a. What critical features of the source material must be captured in the digital product? Are very high resolution copies, accurate rendition of colors, a seamless combination of images and text, or other qualities considered essential?

The cost and nature of digitizing hardware and software continue to evolve and preferred solutions are likely to shift as well. Sometimes it makes sense to defer certain projects until technology catches up to needs. E.g. a project to digitize oversized maps depended on the ability to read detail closely. As a result, the maps are scanned at very high resolution, thereby creating challenges for digital image delivery and presentation. File sizes can out run the library's computers and network.

b. If the original sources are to be retained, can they withstand the digitization process?

Automatic sheet feeders are fast and efficient, but they may destroy brittle paper. Digital cameras can minimize manipulation of source materials, but subjecting certain media - watercolors, for example, to prolonged lighting is problematic.

c. What type of hardware should be used for conversion?

Color slides, for instance, cannot be fully represented by scanners that only produce black and white images. Even a color scanner with limited capacity to reproduce tonalities will be inadequate when high-quality images are important. Digitizing equipment can be very expensive and costs can be difficult to justify when use is sporadic. Outsourcing to external vendors, to prevent expensive equipment and staff costs, nevertheless, requires intimate knowledge of the process in order to specify technical specifications, technical conditions, performance expectations, and safety/security

precautions.

d. Will a digitized sample meet user's needs? If so, how should the sample be constructed?

Many collections are too large to convert in their entirety. Subsets of large collections can be defined in many ways and for many purposes.

e. Will the information resources upon which the project is based continue to grow?

Ongoing commitments and extended arrangements for copyrights may be required when collections are still expanding, as is the case for current journals and annual reports or the papers of a living individual.

f. How will users navigate within and among digital collections?

Printed sources orient readers by means of tables of content, chapters and sections, pagination, indexing, and formatting cues. Manuscript materials often rely on finding aids linked to the organization of file folders. Photographs may be mounted in albums. At a minimum, electronic products need to provide the same kind of functionality. This process may require several steps. The files for critical pages of a work - e.g. the title page, table of contents, and the first page of every new chapter - can be linked to electronic navigational tools so that they can be easily located.

6. DESCRIBING, DELIVERING, AND RETAINING THE DIGITAL PRODUCT

While libraries can point to their collective achievements in organizing and describing an enormous number and variety of collections and material types, some perplexing issues have not yet been resolved. There is still no consensus on how to handle ephemera that cannot realistically be cataloged by the piece and that are too insubstantial to shelve like most books and journals.

Decisions as to what resources should be digitized must be informed by an understanding of how the product will be described for users, delivered to them, and managed over time.

a. How will users know that the digital file exists?

One of the principal challenges is to determine what information is essential in describing an electronic product. The "Dublin Core" and other special initiatives for structuring and standardizing descriptive data propose to combine information about the technical characteristics of digital files (how they were created), their location, and a summary of their contents. The resulting records, are known as "metadata". Their function is to provide users with a standardized means for intellectual access to digitized materials.

b. How can the digital product best be delivered to users?

Alternative modes of storage and delivery must be considered from the outset of a project. CDRoms are distributed and used differently than information made accessible over the Internet. How immediate must access be? Files can be instantaneously available online, or they can be stored and loaded on demand. Expected demand, file sizes, fee structures, and available staffing and equipment must all be considered.

c. Who will be authorized to use the digital resource and under what circumstances?

Access issues impinge upon selection decisions in a number of ways. A library may mount high-resolution images of unique holdings for local use only and not allow unauthorized printing of those images. Or, be unwilling to provide the infrastructure needed to support universal access. A cost-recovery enterprise cannot by definition make a file available without restriction.

Access, when it is not universal, must be managed. Current alternatives include passwords, direct user fees, and limitations according to organizational affiliation.

Different capabilities for viewing, downloading, and printing may be offered to different sets of users, or at different prices.

d. How will the integrity of the digitized data be ensured?

Digitized files must be embedded with detailed information concerning the methods used to create them. The same information should be included in external bibliographic or descriptive records. Users who

are consulting or copying the sources must also be able to confirm that the files they see or receive match the originals. Means to authenticate and protect digital products are only now being developed for use in the scholarly world.

e. Particularly for digital products created to meet local demand, is the existing technology infrastructure adequate?

Robust computer systems and an appropriate number of workstations are perhaps more easily provided than such ancillary features as network printing capabilities.

f. What are the long-term intentions for the digital file?

In some cases, images must be legible but need not be perfect replicas; copyright constraints, indexing complexities, and storage economies make it simpler to rescan on demand than to organize and retain random files. In other cases, the file may be kept for a longer, but still limited period and then discarded - images of a text or art work scanned to support a limited exhibition or temporary use.

g. Is the long-term preservation of deteriorated materials a project goal?

Digital preservation also requires a supporting organization and infrastructure dedicated to storing the electronic files and to migrating them to new formats and/or new media as technologies change. Unless these capacities are all in place, digital files cannot be regarded as permanent. Creating an enduring digital preservation master file is a multidimensional task with long-term implications.

Digital processes meet preservation objectives without pretending to permanence.

7. RELATIONSHIPS TO OTHER DIGITAL EFFORTS

a. Have the materials proposed for digitization already been converted to electronic form?

b. Can cooperative digitization efforts bring together a cohesive body of material that would otherwise remain disassociated?

Standardized descriptors and a common approach to indexing and storage can allow dispersed materials to be combined in an amalgamated digital resource. The process involves technological conventions and institutional alliances.

Successful projects to combine digital resources through a common system for organization and delivery suggest a new kind of model for collection building. The workflow of digital collection development can remain radically decentralized provided a robust infrastructure for collaboration is in place.

8. COSTS AND BENEFITS

Cost-benefit analysis assesses the relationship between functionality, demand, and expense. Limited resources and competing demands on organizational time and energy mean that the analysis must be rigorous and complete. The costs of creating electronic resources vary considerably. File size and processing requirements account for part of the differences, though labor requirements are even more important.

Materials converted by machine must be painstakingly proofread, or the source documents must be rekeyed in combination with careful attention to the detection and correction of errors. Costs rise for marked-up text, which entails yet another level of analysis and intervention.

All other things being equal, for example, it is less expensive to scan from single sheets than from bound volumes. Small sheets are less expensive to scan than oversized ones. Items in good condition are less costly to process than those that are deteriorated and thus require special handling.

Incomplete cost analyses can impute benefits that are difficult to represent on a project cost sheet. Functions such as preparation of materials for scanning, indexing, bibliographic description, post-scan processing, and long-term file management may not be factored into cost equations.

QUESTIONS TO ASK

- a. Who will benefit from the proposed digital product?

It may allow librarians to manage collections and provide services more effectively, or to provide traditional services at a lower cost or at less risk to collections.

b. Is the intellectual value of the proposed product commensurate with the expense?

The resources available for digitization might have greater impact if they were directed at another project, or directed toward an entirely different approach to providing access - through more detailed indexing, or microfilming, or some other type of reformatting that would prove in the end more useful to users.

c. Could an acceptable product be created a lower cost?

d. How will the proposed project address the long-term costs associated with digital files?

The accumulated body of digital products may enable savings elsewhere in the institution e.g. by lowering the costs of storage, circulation, and preservation. It is essential to realize that the costs of digitization are just beginning at the time of capture. The programmatic capacity to distribute and maintain electronic resources and to migrate them to new forms as original digital platforms fail and formats and software are superseded, is fundamental to long term efforts.

DIGITAL RESOURCES DEPEND ON THE NATURE AND IMPORTANCE OF THE ORIGINAL SOURCE MATERIALS, BUT ALSO ON THE NATURE AND QUALITY OF THE DIGITIZING PROCESS ITSELF AND ON HOW WELL RELEVANT INFORMATION IS CAPTURED FROM THE ORIGINAL AND THEN ON HOW THE DIGITAL DATA ARE ORGANIZED, INDEXED, AND DELIVERED TO USERS, AND MAINTAINED OVER TIME. DISCIPLINED EFFORTS TO ADDRESS THE QUESTIONS OUTLINED HERE HELP ENSURE THAT NEW PROJECTS WILL FULFILL THE EXPECTATIONS OF LIBRARIES, USERS, AND FUTURE MANAGERS.

RLWEB/HK 06/02/99

Annexe 4 bis : " Candidate Collections for Visual Treasures Project "

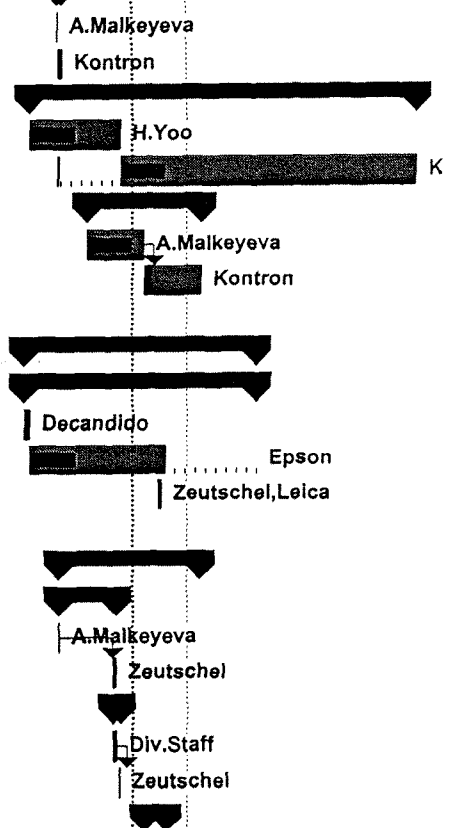
Candidate Collections for Visual Treasures Project

517 de dollars in San
600.000 images in San

<i>NAME</i>	<i>COLLECTION SIZE</i>	<i>IMAGES</i>	<i>NOTE</i>
Spalding Baseball		400	names index
ARM			
Photographs of construction and plans for Central Building			
Lewis Hine Collection of Immigration, Labor and Social Conditions in the U.S.		431	index to the images in an electronic format
ART			
Deco Art and Design, American, British, W.Europe, Tibetan, Scandanavian	multiple volumes		All cataloged
BRG			
Portraits of authors			cataloged
Drawings by artists "Pen and Brush" exhibit			exhibit catalog
Photographs, by authors		5,000	cataloged
GRD			
Menus		19,500	Database
Scrapbooks containing examples of book jackets produced in the United States, England, Germany	22 scrapbooks		
Vinkhuizen Collection		32,236	Images sourced.
JWS			
Oral Histories	2,500 interviews estimated at 75 pp/interview	190,000	Requires transcription

Annexe 4 ter : " Visual Treasures Production Schedule "

ID	No.Item	DIV	Task Name	% Co	Duration	Start	Finish	Resource Nam	a	u	u	u	e	c	o	e	a	e	a	p	a	u	u	u	e	c	o
1	0		PRIORITY STANDARD SIZE - Kontron	26%	342 days	Thu 6/22/00	Fri 10/12/01		[Gantt bar]																		
2									[Gantt bar]																		
3	50	ORT	Champollion, Pantheon of Egypt	25%	4 days	Wed 7/26/00	Mon 7/31/00		[Gantt bar]																		
4	50		Metadata Collection	100%	1 day	Wed 7/26/00	Wed 7/26/00	A.Malkeyeva	[Gantt bar]																		
5	50		Digitization	0%	3 days	Thu 7/27/00	Mon 7/31/00	Kontron	[Gantt bar]																		
6	4000	SLV	"Original photographs of Russia,	23%	342 days	Thu 6/22/00	Fri 10/12/01		[Gantt bar]																		
7	4000		Metadata Collection	50%	80 days	Thu 6/22/00	Wed 10/11/00	H.Yoo	[Gantt bar]																		
8	4000		Digitization	15%	264.85 days	Wed 7/26/00	Fri 10/12/01	Kontron	[Gantt bar]																		
9	1200	SPN	Islamic Manuscripts	38%	100 days	Fri 9/1/00	Thu 1/18/01		[Gantt bar]																		
10	1200		Metadata Collection	75%	50 days	Fri 9/1/00	Thu 11/9/00	A.Malkeyeva	[Gantt bar]																		
11	1200		Digitization	0%	50 days	Fri 11/10/00	Thu 1/18/01	Kontron	[Gantt bar]																		
12									[Gantt bar]																		
13	0		OUTSOURCE AND EPSON	22%	204 days	Thu 6/15/00	Tue 3/27/01		[Gantt bar]																		
14	15000	PRN	Emmett Collection Hist Prints	22%	204 days	Thu 6/15/00	Tue 3/27/01		[Gantt bar]																		
15	5000		Metadata Collection	100%	5 days	Thu 6/15/00	Wed 6/21/00	Decandido	[Gantt bar]																		
16	4800		Digitization	20%	200 days	Wed 6/21/00	Tue 3/27/01	Epson	[Gantt bar]																		
17	200		Digitization	50%	4 days	Mon 11/27/00	Thu 11/30/00	Zeutschel,Leica	[Gantt bar]																		
18									[Gantt bar]																		
19	0		PRIORITY OVERSIZE - ZEUTSCHEL	40%	124 days	Thu 7/27/00	Tue 1/16/01		[Gantt bar]																		
20	50	ORT	Sundhya Daily Prayers Brahmin	20%	51 days	Thu 7/27/00	Thu 10/5/00		[Gantt bar]																		
21	50		Metadata Collection	100%	1 day	Thu 7/27/00	Thu 7/27/00	A.Malkeyeva	[Gantt bar]																		
22	50		Digitization	0%	4 days	Mon 10/2/00	Thu 10/5/00	Zeutschel	[Gantt bar]																		
23	24	JWS	Kettubot (Marriage Contracts)	0%	7 days	Mon 10/2/00	Tue 10/10/00		[Gantt bar]																		
24	24		Metadata Collection	0%	5 days	Mon 10/2/00	Fri 10/6/00	Div.Staff	[Gantt bar]																		
25	22		Digitization	0%	2 days	Mon 10/9/00	Tue 10/10/00	Zeutschel	[Gantt bar]																		
26	362	PRN	Russian Popular Prints, 19th Ce	4%	19 days	Fri 11/10/00	Wed 12/6/00		[Gantt bar]																		



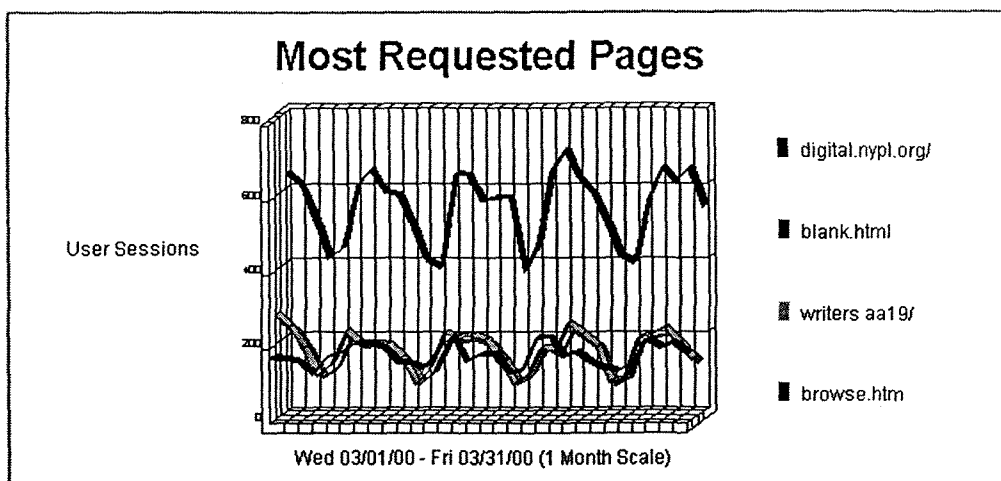
Project: VisTreas
Date: Wed 10/25/00

Task	[Solid Bar]	Rolled Up Task	[Solid Bar]	Project Summary	[Solid Bar]
Split	[Dotted Line]	Rolled Up Split	[Dotted Line]	External Milestone	[Diamond]
Progress	[Solid Bar]	Rolled Up Milestone	[Diamond]	Deadline	[Down Arrow]
Milestone	[Diamond]	Rolled Up Progress	[Solid Bar]		
Summary	[Solid Bar]	External Tasks	[Solid Bar]		

Annexe 5 : “ Most requested Pages ”²⁸

²⁸ Extrait des pages de l’Intranet local.

This section identifies the most popular web site pages and how often they were accessed. The average time a user spends viewing a page is also indicated in the table.

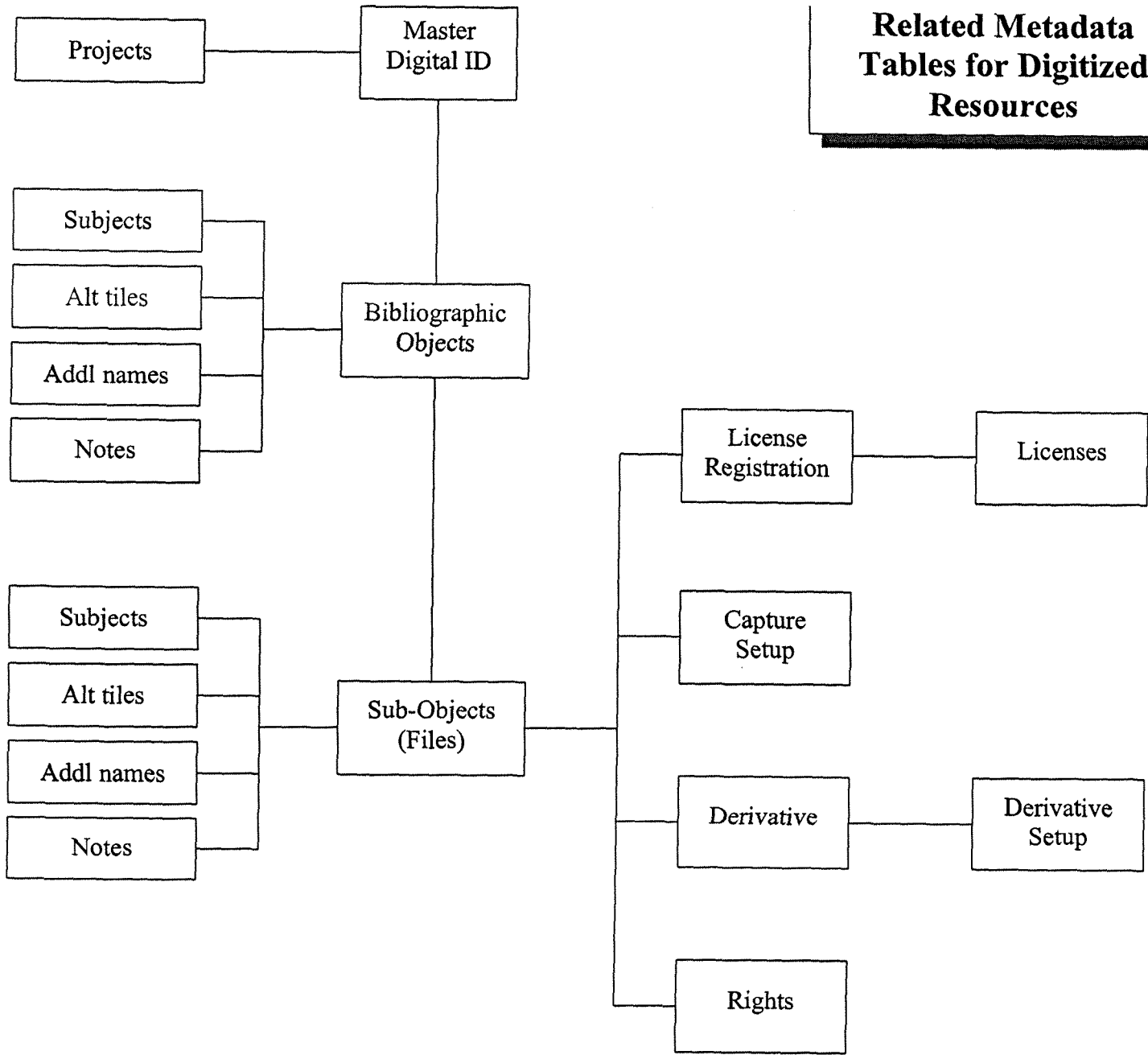


Most Requested Pages				
	Pages	Views	% of Total Views	Use Sessi
1	Digital Library -The New York Public Library http://digital.nypl.org/	19,503	12.39%	16,97
2	NYPL Digital Schomburg Blank Frame http://digital.nypl.org/schomburg/writers_aa19/blank.html	7,035	4.47%	5,09
3	NYPL Digital Schomburg African American Women Writers the 19th Century http://digital.nypl.org/schomburg/writers_aa19/	6,399	4.06%	5,80
4	Digital Library: Finding Aids http://digital.nypl.org/browse.htm	5,840	3.71%	5,61
5	African American Women Writers of the 19th Century http://digital.nypl.org/schomburg/writers_aa19/toc.html	5,457	3.46%	5,02
6	NYPL Digital Schomburg African American Women Writers of the 19th Century http://digital.nypl.org/schomburg/writers_aa19/main.html	5,361	3.4%	5,09
7	NYPL Digital Schomburg Images Fine Print http://digital.nypl.org/schomburg/writers_aa19/fineprint.html	5,353	3.4%	5,08
8	NYPL Digital Schomburg African American Women Writers of the 19th Century http://digital.nypl.org/schomburg/writers_aa19/inmain.html	5,344	3.39%	5,08
9	Error Occurred While Processing Request http://digital.nypl.org/schomburg/writers_aa19/gothere.cfm	4,992	3.17%	3,63
10	NYPL Digital Schomburg Images of African Americans from the 19th Century http://149.123.1.8/cgi-shl/vsc30b.exe/schomburg/images_aa19/toc.html?E+nyplbeta	4,647	2.95%	3,33
11	NYPL Digital Schomburg Blank Frame http://digital.nypl.org/schomburg/images_aa19/blank.html	4,196	2.66%	3,00
12	NYPL Digital Schomburg Images of African Americans from the 19th Century http://digital.nypl.org/schomburg/images_aa19/	3,708	2.35%	3,30
13	Digital Library: Finding Aids http://digital.nypl.org/full.htm	3,592	2.28%	3,49
14	NYPL Digital Schomburg Images of African Americans from the 19th Century http://digital.nypl.org/schomburg/images_aa19/main.html	3,576	2.27%	3,38
15	Digital Library: Finding Aids http://digital.nypl.org/photo.htm	3,226	2.05%	3,10

Annexe 6 : " Related Metadata Tables for Digitized Resources "²⁹

²⁹ Extrait des pages de l'Intranet local de NYPL consacrées à la numérisation.

**Related Metadata
Tables for Digitized
Resources**



Annexe 6 bis : " Digital library collections metadata registry "³⁰

³⁰ *Ibid.*

digital library collections metadata registry

1 - 2 - 3

MARC		NYPL					DC
ELEMENT NAME	MARC	CATNYP	HUDSON	MILLENIUM	DENNIS	TEI LITE	DUBLIN CORE
Main Entry- Personal Name	100\$a	100\$a	CREATOR	CREATOR	PHOTOGRAPHER / STUDIO	<docAuthor>	DC.CREAT
Author's dates	100\$d	100\$d	CREATOR	CREATOR	PHOTOGRAPHER / STUDIO	.	.
Author's Role- Code	100\$r	100\$r	CREATOR	CREATOR	PHOTOGRAPHER / STUDIO	.	.
Author's Role- Text	100\$e	100\$e	CREATOR	CREATOR	PHOTOGRAPHER / STUDIO	.	.
Affiliation	100\$u	100\$u	CREATOR	CREATOR	PHOTOGRAPHER / STUDIO	.	.
Main Entry- Corporate Name	110	110	CREATOR	CREATOR	PHOTOGRAPHER / STUDIO	.	DC.CREAT
Main Entry- Meeting Name	111	111	CREATOR	CREATOR	PHOTOGRAPHER / STUDIO	.	DC.CREAT
Uniform Title- Main Entry	130	130	.	OTHER TITLE	.	.	.
Uniform Title	240	240	.	OTHER TITLE	.	.	.
Title Statement	245\$a	245	TITLE	TITLE	TITLE	<docTitle>	DC.TITLE
Title, Subtitle	245\$b	245\$b	TITLE	TITLE	TITLE	.	.
Statement of Responsibility	245\$c	245\$c
Medium	245\$h	245\$h	.	GENRE	.	.	.
Title, Number	245\$n	245\$n	.	TITLE	.	.	.
Title, Part	245\$p	245\$p	.	TITLE	.	.	.
Varying Form of Title	246	246	SUPPLIED TITLE	OTHER TITLE	.	.	.
Former Title	247	247	.	OTHER TITLE	.	.	.
Edition Statement	250	250	.	.	.	<edition>	.
Place of Publication	260\$a	260\$a	PLACE OF PUBLICATION	PLACE OF PUBLICATION	PLACE OF PUBLICATION	<pubPlace>	.
Publisher	260\$b	260\$b	PUBLISHER	PUBLISHER	PUBLISHER	<publisher>	.
Publication Date	260\$c	260\$c	DATE	DATE	DATE	<docDate>	.
Physical Description	3xx	3xx	.	PHYSICAL DESCRIPTION	PHYSICAL DESCRIPTION	.	.

**digital library collections
metadata registry**

1 - 2 - 3

MARC		NYPL					DC
ELEMENT NAME	MARC	CATNYP	HUDSON	MILLENIUM	DENNIS	TEI LITE	DUBLIN CORE
General Note, Coverage	5xx	5xx	NOTES	NOTES	NOTES	<note>	DC.COVERAG
Provenance	541	541	PROVENANCE	PROVENANCE	NOTES	.	.
Subject, Personal Name	600	600	PERSONAL NAME	SUBJECT NAME	SUBJECT	<name type="person">	DC.SUBJECT
Subject, Corporate Name	610	610	CORPORATE BODY	SUBJECT NAME	SUBJECT	<name type="institution">	DC.SUBJECT
Subject, Meeting Name	611	611	.	SUBJECT NAME	SUBJECT		DC.SUBJECT
Subject, Topical	650	650	SUBJECT	SUBJECT	SUBJECT	.	DC.SUBJECT
Subject, Geographic Name	651	651	COUNTY	SUBJECT NAME	PLACE	<name type="place">	DC.SUBJECT
Index Term-Uncontrolled	653	653	.	.	.	<keywords>	DC.SUBJECT
Index Term-Genre	655	655	GENRE	GENRE	GENRE	.	DC.TYPE
Local Index Term	690	690	PLACE	.	SUBJECT	.	.

Digital Library Projects

Last updated 11/15/98, pe

**digital library collections
metadata registry**

1 - 2 - 3

MARC		NYPL				
ELEMENT NAME	MARC	CATNYP	HUDSON	MILLENIUM	DENNIS	TEI LITE
Added Entry- Personal Name	700	700 except \$t	.	CREATOR	PHOTOGRAPHER/STUDIO	<name type="person">
Added Entry- Corporate Name	710	710 except \$t	.	CREATOR	PHOTOGRAPHER/STUDIO	<name type="institutio
Added Entry- Meeting Name	711	711	.	CREATOR	.	.
Added Entry- Uniform Title	730	730	.	OTHER TITLE	.	.
Added Entry- Uncontrolled Title	740	740	.	OTHER TITLE	TITLE	.
Added Entry- Place Name	752	752	.	.	.	<name type="place">
Genre	755	755	.	.	PROCESSES	.
Host Item Entry (bibliographic relation)	773	773	PARENT DOCUMENT	PARENT DOCUMENT	.	.
Constituent Text-Main Entry (physical relation)	774\$a
Relationship Entry	787n
Donor/Gift	799	799	.	FUNDER	.	<funder>
Library Location (division + collection)	852\$a	852\$a	INSTITUTION	INSTITUTION	.	.
Library Location Code	852\$b	852\$b	INSTITUTIONAL DIVISION	DIVISION	.	.
Electronic Location (URL)	856\$u	856\$u
Classmark	950	950	LOCAL ID	CALL NO	CLASSMARK	.
Bibliographic Record Number	001	001	BIB RECORD NO	BIB RECORD NO	.	.
Country Code	008\$15	008\$15

**Annexe 6 ter : " Digital Library Programs.
Digitizing Carrier Sheet VT PRN Journal of
Saints "31**

³¹ Document de travail ayant servi à la numérisation du *Journal of Saints*.

VT PRN Journal of Saints

Journal of Saints

WWM0040 / 31097

Image Count: 200

	Sub Object Title	FILE ID:	Sequence	Page	
X	Frontispiece	001	1	v.I, p.1	/
X	Frontispiece	002	2	v.I, p.2	/
X	Frontispiece	003	3	v.I, p.3	/
X	Frontispiece	004	4	v.I, p.4	/
X	Frontispiece	005	5	v.I, p.5	/
X	Frontispiece	006	6	v.I, p.6	/
X	Frontispiece	007	7	v.I, p.7	/
X	Frontispiece	008	8	v.I, p.8	/
X	Frontispiece	009	9	v.I, p.9	/
X	Jan. 1	010	10	v.I, p.11	/
X	Jan. 1	011	11	v.I, p.12	/
X	Jan. 1	013	13	v.I, p.13	/
X	Jan. 1	014	14	v.I, p.14	/
X	Jan. 1	015	15	v.I, p.15	/
X	Jan. 1	016	16	v.I, p.16	/
X	Jan. 2	017	17	v.I, p.17	/
X	Jan. 2	018	18	v.I, p.18	/
X	Jan. 2	019	19	v.I, p.19	/
X	Jan. 2	021	21	v.I, p.20	/
X	Jan. 3	022	22	v.I, p.21	/
X	Jan. 3	023	23	v.I, p.22	/
X	Jan. 3	024	24	v.I, p.23	/
X	Jan. 3	025	25	v.I, p.24	/
X	Jan. 3	026	26	v.I, p.25	/
X	Jan. 4	027	27	v.I, p.26	/
X	Jan. 4	029	29	v.I, p.27	/
X	Jan. 4	030	30	v.I, p.28	/
X	Jan. 4	031	31	v.I, p.29	/
X	Jan. 5	032	32	v.I, p.30	/
X	Jan. 5	033	33	v.I, p.31	/
X	Jan. 5	034	34	v.I, p.32	/

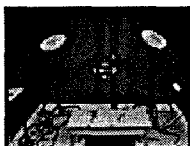
Annexe 7 : " Digital imaging unit homepage " ³²

³² Extrait des pages de l'Intranet local de NYPL consacrées à la numérisation.

**Digital Imaging Unit - The New York Public Library
Room 68R
(212) 930-0821**



Epson flatbed scanner



Leica S1 Pro digital camera



Zeutschel M-6000 digital camera



Kontron ProgRes 3012 digital camera



Post-processing/Quality control center

Anthony Troncale, Head
(212) 930-0821 atroncale@nypl.org

Robert DeCandido, Automation Specialist,
(212) 930-0622 rdecandido@nypl.org

Robyn Mericle, Scanning Technician
rmericle@nypl.org

Timothy Russell, Scanning Technician
trussell@nypl.org

- [Selection Guidelines for Scanning Library Materials](#)
- [Production Goals for Scanning, Visual Treasures Project](#)
- [NYPL Digital Conversion Specifications and Long-term Retention Policies](#)
- [Related Metadata Tables for Digital Resources\(.pdf file only\)](#)
- [Guidelines for Conservation Assessment\(.pdf file only\)](#)

Annexe 7 bis : " Digital library collections file naming ³³

³³ *ibid.*

digital library collections file naming

The following file naming conventions helps to ensure the longevity of a file by enabling us to track files through cataloging, scanning, conversion, quality review, mounting on the server, and archiving. For most projects, several files will be associated with one bibliographic identifier, that is, with one cataloging record representing several versions of the object-- thumbnail image, reference image, pdf file, and so on. Since the location, pathname, and formats of digital objects are certain to change, reference the elements of the filename that are uniquely identify the item.

NYPL File Naming Scheme

Folder Name	
RLIN # OR OCLC # OR CALL #	1-letter identifier for file type (see chart below)

File Name		
3 or 4 digit sequential number	Page number	1-letter identifier for location within work (see chart below)

Identifiers for File Type	
o	audio
r	reference image
s	service copy
t	thumbnail image
u	uncompressed image
v	video

Identifiers for Location within Work	
a	attachment
b	back of item
c	cover
d	detail
e	advertisement
f	illustration (figure)
g	title page
i	inscription
l	list of illustrations
m	map
n	table of contents
p	frontispiece
x	index
y	irregularity target

Annexe 7 ter : " The Research Libraries. The New York Public Library. Digital Library program-Digital imaging unit. Draft 2.1 NYPL Digital Conversion specifications and Long-term Retention Policies "³⁴

³⁴ *Ibid.*

The Research Libraries-The New York Public Library

Digital Library Program-Digital Imaging Unit

DRAFT 2.1

NYPL Digital Conversion Specifications and Long-term Retention Policies

1. Definition

1. The Research Libraries of The New York Public Library [RL-NYPL] digitally converts selected holdings of its materials either through direct capture, or through intermediary film, using *high-resolution* digital cameras and/or flatbed scanners for textual and/or visual materials, and analog to digital binary encoding for its audio/visual materials. RL-NYPL may also contract out digital conversion and post-processing tasks to third party contractors who agree to follow RL-NYPL's specifications, stated below.
2. The term '*high-resolution*' is defined in this context as the point at which a digital file reproduces or resolves to a level of quality or definition that allows for effective surrogate representation within a printed or electronic format.

2. Objectives

1. Digital imaging technology, and the digital files created, constitute a means towards achieving RL-NYPL's goal of making its holdings and catalog records widely and publicly accessible via networked telecommunications to a global audience. The specific objectives of this retention policy are to ensure the long-term viability of digital files that are secure, discoverable, documented, recoverable, and maintain their integrity. A third goal is to utilize digital technologies to create an online research environment to serve students, researchers and scholars. The use of digital technologies also assists RL-NYPL in its commitment to the preservation of its collections. By providing digital surrogates, RL-NYPL reduces the physical handling of fragile or rare items.
2. Digital files created for RL-NYPL designated to be held for long-term retention must meet specifications and policy goals spelled out in this memorandum. These specifications are designed to insure that all master digital collections maintain their physical and intellectual integrity, are environmentally secure, and their methods of creation are documented for future refreshing and migration purposes.

3. Scope

1. Digital imaging is capable of capturing textual and visual materials, audio/moving image materials, three-dimensional objects, computer graphics and multimedia constructs and

composites to levels that constitute an effective surrogate for research. The digital files created by RL-NYPL, and the metadata associated with them, are designed to represent but not replace materials held in its collections. The RL-NYPL digital archives are converted into an uncompressed, unaltered format, usually .TIF 5.0 or 6.0 for print collections, and .WAV for video/moving image materials. Sound recordings are currently being archived in analog on open-reel tape.

4. Metadata

1. RL-NYPL is committed to the long-term retention of digital files for use in pursuing the stated goals and mission of the RL-NYPL. To fulfill that commitment, best practices are followed to ensure that the collections captured digitally are managed through the use of *metadata*. The three types of metadata defined below are used for access to intellectual content (intellectual metadata), definition of a particular class of materials or objects (structural metadata), and the description of the digital files themselves (administrative metadata).[] [Note:definitions taken from MOA II Testbed Project: Univ. of California, Berkeley 1997, seen at: <http://sunsite.berkeley.edu/MOA2/wp-v2.html>]
2. Administrative metadata is currently being archived in both an MS-Access master metadata database and/or as a text file that accompanies the Archival files stored on CD-R disks.
3. Descriptive metadata is stored onto a Master MS-Access database which is mapped into an Informix database for web access. A Metadata crosswalk is being devised to allow integration of selected data elements exported from NYPL's OPAC, CatNYP, into the MS-Access metadatabase for eventual cross-collection searching.
4. All forms of metadata, including structural, are to be stored on the ITG near-line disk array where the Archival image files are stored and continually backed up and/or refreshed with DLT tape.

5. Conversion

1. The Digital Imaging Unit of the Preservation Reformatting Division is responsible for all digital conversion and is the repository of all RL-NYPL converted materials to be held for long term retention.
2. File types. The Digital Unit uses specific file types for most of its digital files. *Master files* are those intended to be held for long-term retention. *Service files* are to be used by the Copy Services Division for hard copy delivery of image files for requests from the public. *Derivative files*, used primarily for Internet access, may encompass several file types, depending on the need for resolution and navigation of the respective materials.

1. **Master files** for textual and visual materials are saved as uncompressed .TIF version 5.0 or 6.0 file format. Master files for audio materials are saved in the uncompressed .WAV file format. [Master files for video and film are currently being stored on Betacam analog tape.] While attempts are made to render the item as close to the fidelity of the original as possible during the capture process, they not altered or manipulated once the initial scanning process is complete. These are the files in which all others are derived. Once all derivatives are generated, the Master files for the TIF file format are tagged in the TIF header with *Administrative metadata* such as color look up tables, scanner profiles, pixel resolutions, file path, and copyright or ownership citation. See Appendix A for full tag set example. Additionally, a .TXT file is written to contain the Structural, Intellectual and Administrative metadata of the respective item digitized to accompany the Master. Master files are currently being written to the ISO 9660 standard using CD-R gold-coated disks.
2. **Service files** are designed for eventual output to hardcopy prints [currently Fujix Pictography 3000 system]. Generated from the Master files, the Service files are saved in either uncompressed TIF or TIF with LZW compression. They have been optimized for color fidelity, sharpness, briteness and contrast. [Note: The availability of audio and moving image files has not been considered to date.]
3. **Derivative files** are generated for use on the Internet through the RL-NYPL's World Wide Web site. They consist of several file types employing compression algorithms designed to create small file sizes while still retaining an adequate visual representation of the original. Current file types used are .JPG, and .GIF, with compression ratios for .JPG ranging from 5:1 to 15:1 depending on content resolution requirements.

5.2.3a **Reference files** are created to provide a visual representation of the digitized object on the full screen of a computer monitor. As computer monitors continue to improve in video resolution, and as network bandwidth expands, the pixel array for Reference files may increase. Currently the standard used is 800 x 600 @ 72 dpi but may be decreased to 640 x 480 or increased to 1024 x 768 as deemed appropriate. To determine whether the standard reference resolution is sufficient, a significant detail of an item is identified as a target. In the case of textual materials, the pixel resolution of the file may need to be large enough to render the text sufficiently readable. For visual materials, such as a photograph, file sizes may vary according to the need for color or to render text or other details in the image, such as street signs. In some instances, Interactive derivative (see 5.3) files may supercede or compliment Reference files.

5.2.3b **Inspection files** are created when there is a need to provide more detail in an item on the monitor than can be rendered in a Reference file. This may include textual materials with small fonts or engravings with minute details. Generally, Inspection files are in grayscale to keep the file size low enough to be used on the Internet. Current standard is 1024 x 768 @ 72 dpi in 8-bit grayscale.

5.2.3c **Thumbnail files** are low resolution files designed to load onto a computer screen in groups to allow for quick visual browsing. Usually these files are hyper-linked to the Reference or Inspection files for more detailed viewing. Thumbnail files are not always required since their small file sizes and rendering capability negate its use for material groups such as texts. Current standard pixel dimensions are 150 x 150 @ 72 dpi.

1. Control

Management and control of digital image files effects all aspects of the process, from file naming to cataloging. Administrative metadata is the means which image files can be tracked, shelved, loaded to a server, refreshed and migrated as needed. There are several aspects of administrative metadata designed to document creation and

purposing of the image files. Control for Master image files are digital call numbers, and implementation of a PURL server is underway for on-line files.

1. The Digital Imaging Unit tracks all aspects of the scanning process and any customized applications applied to the item being scanned. This includes make and model of capture device, lighting used to illuminate the item, name of scanner operator and time/date scanned, software used to operate scanning device, lens type and focal length if a camera is used, monitor type and settings used to match image files with the item, and an explanation of the file naming structure. Most of this data is placed into an ASCII text file (.txt) which accompanies the image files when written to a CD-R disk. An MS-Access database aggregates all metadata types generated for digital collections.
2. TIF header tags are fields at the beginning of an image file that informs the computer trying to open the file what file type it is and what its parameters are. Most administrative metadata needed here is written either automatically by the software used to scan the items, currently Adobe Photoshop 5.x, or by post-processing. See Appendix A for full TIF tag set example. Usually the tags 269 (file path) and 315 (copyright holder, artist or owner) need to be inserted through post-processing using a software program such as Infothek Tiffkit.
3. File naming schemes are devised on a project-by-project basis. This is because most collections have varying or unique structures and contents. Ideally, it is best to try and incorporate a truncated version of the catalog record number into the file path or file name. However, this is often difficult due to the 8.3 character limit imposed with the ISO 9660 standard for writing to CD-R. It is hoped that once we adopt a new storage medium we will not have to be constrained to the 8.3 character limit. A file-naming scheme has been devised for paginated books and suggests other schemes for unpaginated materials. See Appendix B.
4. Digital class marks are assigned by the RL-NYPL Cataloging Department for all internally generated digital collections. These class marks are placed in the item-level record that is attached to the bibliographic record, for the item, if one exists, to indicate it has a digital surrogate. The class mark consists of a star, (*) followed by the letters *WWM* or *WXM* followed by the year in which the collection was digitized and a collection number. The first letter, 'W', is a designation for an electronic resource. The second letter, 'W' indicates that this is an RL-NYPL assigned classification for specific items or collections, the other second letter, 'X', indicates that this is a group of *assembled items or collections for a particular purpose*, for example, an exhibition.

Examples: *WWM 99-001 (indicates the first collection digitized for the year 1999).

*WXM 99-005 (indicates fifth assembled collection digitized for the year 1999).

1. Digital class marks are also used as the upper most file directory when digitized images are written to CD-R disks. Master files are stored on labeled CD-R disks and shelved in a cabinet in the Digital Unit by digital class mark. See Appendix B for further explanation and examples.
5. The New York Public Library creates, captures and maintains descriptive, structural and administrative metadata on all digital objects it creates. The descriptive metadata is of two kinds: descriptions of the intellectual and perceptible content of the object including information such as bibliographic, visual and auditory data; and descriptions of the technical characteristics of the digital objects and the method of its capture such as pixel dimensions, pixel depth, file size, encoding format, etc. Since digital objects are often parts of some greater whole--e.g. pages of a book, leaves of a letter in an archive--the hierarchical relationships between objects is documented in the structural metadata. The permissions and rights associated with digital objects, is recorded to assist in the administration of them. All this data is recorded in an MS-Access meta-database registry which can also be used to create web products, assist in the maintenance and migration of files and provide for the discovery of resources. This metadatabase will be archived in much the same way as the data in the Library's catalog, with daily backups and constant migration to hardware and new platforms.

1. Storage formats and labeling

1. Currently, all Master files for still digital images are stored on gold-coated CD-R disks following the ISO 9660 standard for writing CD-R disks. Digitized video/film files are currently stored on DAT tape and/or CD-R disks. Sound files are currently stored in analog format on open reel tapes.
2. CD-R disks are labeled using Avery #5824 white CD-R disk labels for laser printers. Each disk label has "The

New York Public Library, Astor, Lenox and Tilden Foundation" printed at the top, followed by the project name. The bottom half of the label displays the digital class mark(s), followed by a folder number (usually a truncated RLIN number), disk volume number, followed by the range of file names contained within each class mark. If there are too many class marks included in one disk to list on the CD-R label, it is placed on the jacket label. Finally the file type, such as Archival (Master) and its file format (.TIF) is included. See example below:

The New York Public Library,
Astor, Lenox and Tilden Foundation
Travels Along the Hudson
 *
***WWM 98-003/G97F123U Disk 03 of 03**
175136.TIF -204165.TIF
Archival (TIF uncompressed)

CD-R jacket labels are custom printed in the Digital Unit following a uniform format. The front and back covers are exactly like the CD label above except that "Digital Library

Collections" is included below the project name. The full list of digital class marks and range of file names are listed as well. See example below:

The New York Public Library
Astor, Lenox and Tilden Foundation
 ~
Travels Along the Hudson
 ~
Digital Library Collections
***WWM 98-003/G97F123U Disk 03 of 03**
175136.TIF -204165.TIF
Archival (TIF uncompressed)

CD-R jacket spines are labeled first with the digital class mark, then folder number, then the disk volume number.

Example: *Travels Along the Hudson* *WWM 98-003/G97F123U Disk 03 of 03

3. Service file disks are labeled accordingly and eventually are delivered to the Photographic Services Division for public service orders. Reference files are also labeled in a similar fashion and are cut to CD-R disks and delivered to RL-NYPL's Information Technology Group for mounting on the web server. Copies of Reference files are maintained in the Digital Unit for on hand referral.

2. Long-term storage

1. Currently, all Master CD-R disks are stored with the Digital Unit for ready access. Eventually all Archival files stored onto CD-R disks are to be uploaded to a high-capacity disk array located at the Central Computing facility of the Information Technology Group located at 20 West 53rd Street.

9. Refreshing and migration

1. The first digital collections captured at RL-NYPL were created in 1996 on CD-R format in .TIF 5.0 uncompressed format. Other collections were digitized from microfilm and captured in .TIF Group IV bi-tonal and stored onto CD-R disks by a vendor. While periodic (every 6 months) random checks are made of these files to see if they open with current CD players, no refreshing and/or migrations are planned at this time. RL-NYPL is concerned about the proliferation of CD-R disks and the control issues of an ever-growing collection of them. This is especially true when new digital cameras are capable of capturing 48-bit files at 5000 pixels or more, resulting in single image files of 100 MB or more.

10. Archival storage and long-term retention strategies

1. 650 GB of archival digital storage will have been generated by RL-NYPL's Digital Collections program by the end of 1999. Projections for the next three years call for from 400 GB to 600 GB of additional data storage for FY 1999/00, 600 GB - 800 GB for FY 00/01, and 800 GB - 1 TB for FY 01/02. The NYPL Information Technology Group has established a near-line disk array storage and off-line (DLT tape) storage of this data and to prepare for future growth over the next three years. This storage will replace the Master CD disks currently being retained in the Digital Unit.

APPENDIX A**TIFF header tags (bold tag lines require manual insertion)**

Tag, Description, and Comment

254 NewSubfileType

256 ImageWidth (LC uses actual pixel count)

257 ImageLength (LC uses actual pixel count)

258 BitsPerSample

259 Compression

262 PhotometricInterpretation

269 DocumentName (NYPL Digital Call Number Example: WWM98001\NYPG98f123U\001000C.TIF)

273 StripOffsets

277 SamplesPerPixel

278 RowsPerStrip

279 StripByteCounts

282 XResolution (LC uses dots per inch)

283 YResolution (LC uses dots per inch)

296 ResolutionUnit (LC uses "2" [inch])

306 DateTime (LC uses date and time scanned)

315 Artist (The New York Public Library, Astor, Lenox and Tilden Foundation)

282 XResolution (LC uses actual pixel count or dpi)

283 YResolution (LC uses actual pixel count or dpi)

296 ResolutionUnit (LC uses "1" [no unit specified] or "2" [inch])

* Each collection identifier is a designation that shall also be used as part of the CD-ROM volume names.

** At least two options exist for tags 282, 283, and 296, either one of which is acceptable:

Option 1 (often used for full-size uncompressed images)

Xresolution 282 actual pixel count

YResolution 283 actual pixel count

ResolutionUnit 296 1 (no unit specified)

Option 2 (often used for thumbnail images)

Xresolution 282 dots per inch

YResolution 283 dots per inch

ResolutionUnit 296 2 (inch)

In order for the digital images to open in all software packages, the TIFF header tags shall be sorted into **ascending numerical order**.

APPENDIX B

File naming schema for digitized collections

(For the Digital Archiving Disk Array, these directories would be strung together to create unique file names.)

Directories

Upper-level file folders incorporate the assigned master digital class mark. Secondary file folders may be used to incorporate catalog record numbers into the file path, followed by an alpha indicator which designates the file type status. *Always use capital letters for all alpha indicators.*

Sample Master file number: **WWM99015**

The second-level directory should be the RLIN or OCLC or Call number (in that order) specific to the collection at hand. [Note: CATNYP numbers are derived from either RLIN or OCLC record numbers.] If the record number exceeds 8 characters, one may truncate it by eliminating, for example, the "NYP" preceding an RLIN number. Truncating the record

number is not always possible. Many record numbers exceed the 8 character limit and cannot be truncated without erasing the ability to reconstruct it. In this case one may have to substitute a name, such as author or artist for the second-level directory. Following the this number is the alpha indicator which designates the file type status. See examples below:

Example #1: Full RLIN record number: NYPG96-f123

Truncated for directory: G96F123

Example of full file path: WWM99015/G96F123U/008003G.TIF

Example #2: Full RLIN number exceeding 8 characters: NYPG82-B21494

Example of full file path: WWM99015/FRITHU/008003G.TIF

Example #2: Full OCLC record number: 3699872

Example of full file path: WWM98012/3699872U/000001C.TIF

See above exception if OCLC number exceeds 8 character limit.

Individual file names for paginated books:

The first three digits of the individual file names shall be a running sequential number, (**001 to 999**); the second three digits will reflect the page number of the item, if a page number exists. If **NO** page number exists, then the second three digits will be "000".

File Name: 000000X.ext

First three digits: 3 digit sequential number: 000---X.ext

Second three digits: Page number, if present, if no page number

then "000" is used: ---000X.ext

A final digit, "X", will be a capitalized alpha character identifier which represents the type of item represented. See chart below:

Identifiers for Location within Work:

A attachment

B back of item

C cover

D detail

E advertisement

F illustration (figure)

G title page

I inscription

L list of illustrations

M map

N table of contents

P frontispiece

X index

Y irregularity target

Example file names: 001000C.TIF

002000P.TIF

003000G.TIF

004001.TIF

005002F.TIF

The extensions **.TIF**, and **.JPG and/or .GIF** will be used for the uncompressed, and service, (.TIF), reference, detail and thumbnail files(.JPG/.GIF), respectively.

Individual file names without pagination or other schema that identifies each page, plate or leaf in a collection or publication: Use above file directory structure for upper and secondary directories, followed by a sequential number of *no less* than four digits, followed by an alpha indicator, if necessary.

Examples: WWM9920/G97F123U/0001C.TIF

WWM9920/G97F123Y/0002P.TIF

WWM9920/G97F123U/0003G.TIF

WWM9920/G97F123U/0004F.TIF

Annexe 8 : " Digital Imaging Unit. In-house production schedules for digital capture, Visual Treasures Project "³⁵

³⁵ Extrait de l'Intranet local de Nypl.

Digital Imaging Unit

In-house production schedules for digital capture, Visual Treasures Project

- Kontron - 3072 x 2320 pixel array (limit: 11 x 14 inches for visual materials)
- Leica - 5000 x 5000 pixel array (all sizes for visual and textual materials)
- Epson flatbed - 800 dpi (8000 pixels) (limit: 12 x 17 in.. textual, graphic or line art materials, or negatives and transparencies.
- Zeutschel - 5000 pixels (up to 22 x 30 in. for maps, graphic or line art, books)

Yearly Output Goal: 68,500 items/yr.

Total actual work hours per year: 1570 (5 hours for scanning, 1 hour setup) Based on full time scanning technicians performing capture **ONLY**.

	Capture Rate	
	Capacity - Required	
Estimated output per hour on Kontron camera:	25 items	[13]
Estimated output per hour on Leica camera:	10 items	[7]
Estimated output per hour on Epson flatbed:	25 items	[13]
Estimated output per hour on Zeutschel:	45 items	[33]
Daily output goals for Kontron:	125	[67]
Daily output goals for Leica:	50	[35]
Daily output goals for Epson:	125	[67]
Daily output goals for Zeutschel:	225	[166]
Four (4) day output goals for Kontron:	500	[333]

Four (4) day output goals for Leica:	200	[177]
Four (4) day output goals for Epson:	500	[333]
Three (3) day output goal for Zeutschel:	675	[583]
* 1-2 days/week dedicated to post-processing, quality control, depending on output.		
Monthly goals for Kontron (16 days):	2000	[1333]
Monthly goals for Leica (16 days):	800	[708]
Monthly goals for Epson (16 days):	2000	[1333]
Monthly goals for Zeutschel (12 days):	2700	[2333]
Yearly output goals for Kontron:	24,000	[16,000]
Yearly output goals for Leica:	9600	[8500]
Yearly output goals for Epson:	24,000	[16,000]
Yearly output goals for Zeutschel:	32,400	[28,000]
Total yearly output:	90,000	[68,500]

Post-Processing

Quality Control-Stage 1

Checks file integrity and file name accuracy

Kontron [20 MB files] 150 images/hr. = 750/day (LTA I position working 5 hrs.)

Leica [70 MB files] 100 images/hr. = 500/day (LTA I position working 5 hrs.)

Epson [30-50 MB files] 135 images/hr. = 725/day (LTA I position working 5 hrs.)

Epson - 12-20 images = one CD-R

Production: 40 images/hr. = 200 images/day, 1000 images/wk, 4000 images/mo.

Zeutschel - 9-12 images = one CD-R

Production: 18-20 images/hr. = 90-100 images/day, 450-500 images/wk,

1800-2000 images/mo.

Quality Control-Stage 2

Checks CD-R integrity and labeling accuracy

Kontron [20 MB files] 200 images/hr. = 1000/day (LTA I position working 5 hrs.)

[approx. 8 days worth of scanning]

Leica [70 MB files] 150 images/hr. = 750/day (LTA I position working 5 hrs.)

[approx. 8 days worth of scanning]

Epson [30-50 MB files] 195 images/hr. = 975/day (LTA I position working 5 hrs.)

[approx. 8 days worth of scanning]

Zeutschel [50-70 MB files] 150-200 images/hr. = 750-1000/day (LTA I position working 5 hrs.) [approx. 8 days worth of scanning]

Out-sourced production schedules for digital capture, Visual Treasures Project

(based on Dennis Stereo Collection production rate from JJT, Inc.)

Single-sheet, uniformed sized materials, up to 9 x 12 in.: 2500 items/month

Single-sheet, uniformed sized materials, 9 x 12 in.-16 x 20 in.: 1750 items/month

Single-sheet, *non-uniform* materials, up to - 11 x 14 in.: 1500 items/month Single-sheet, *non-uniform* materials, 11 x 14 in.-16 x 20 in.: 1100 items/month

Uniform, non-fragile, bound materials, such as

(pamphlets, programs, journals) up to 9 x 12 in.: 2000 pages/month

Quality Control for Out-Sourced Digital Files

QC rate will likely be similar to Stage 2 above. File sizes will vary according to material types and subsequent resolution specifications provided to vendor.

For example:

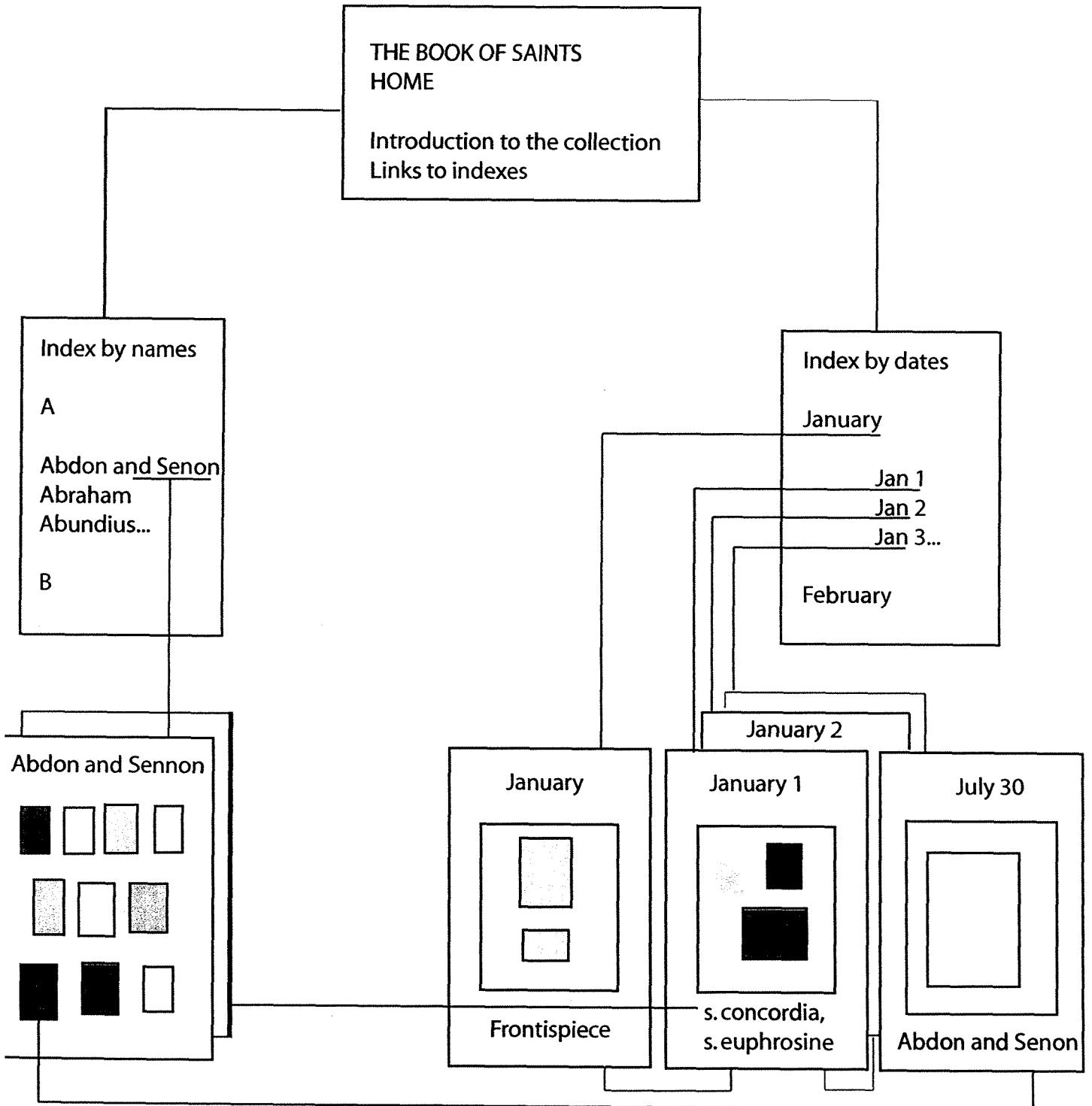
8 x 10 in. silver gelatin photographs: Scanned at 3000 pixels= +/-20 MB

Theater Key sheet: approx. 11 x 17 in.: Scanned at 7000 pixels= +/-70 MB

Restaurant menu, 11 x 14 in. graphic art: Scanned at 3000 pixels= +/-30 MB

Annexe 9 : " The Book of Saints. Website architecture proposal "

The Book of Saints - Website architecture proposal



Annexe 10 : " The Journal of Saints. Website pages "³⁶

³⁶ Contient : page d'accueil, index des noms, index des dates et une image.

THE JOURNAL OF SAINTS

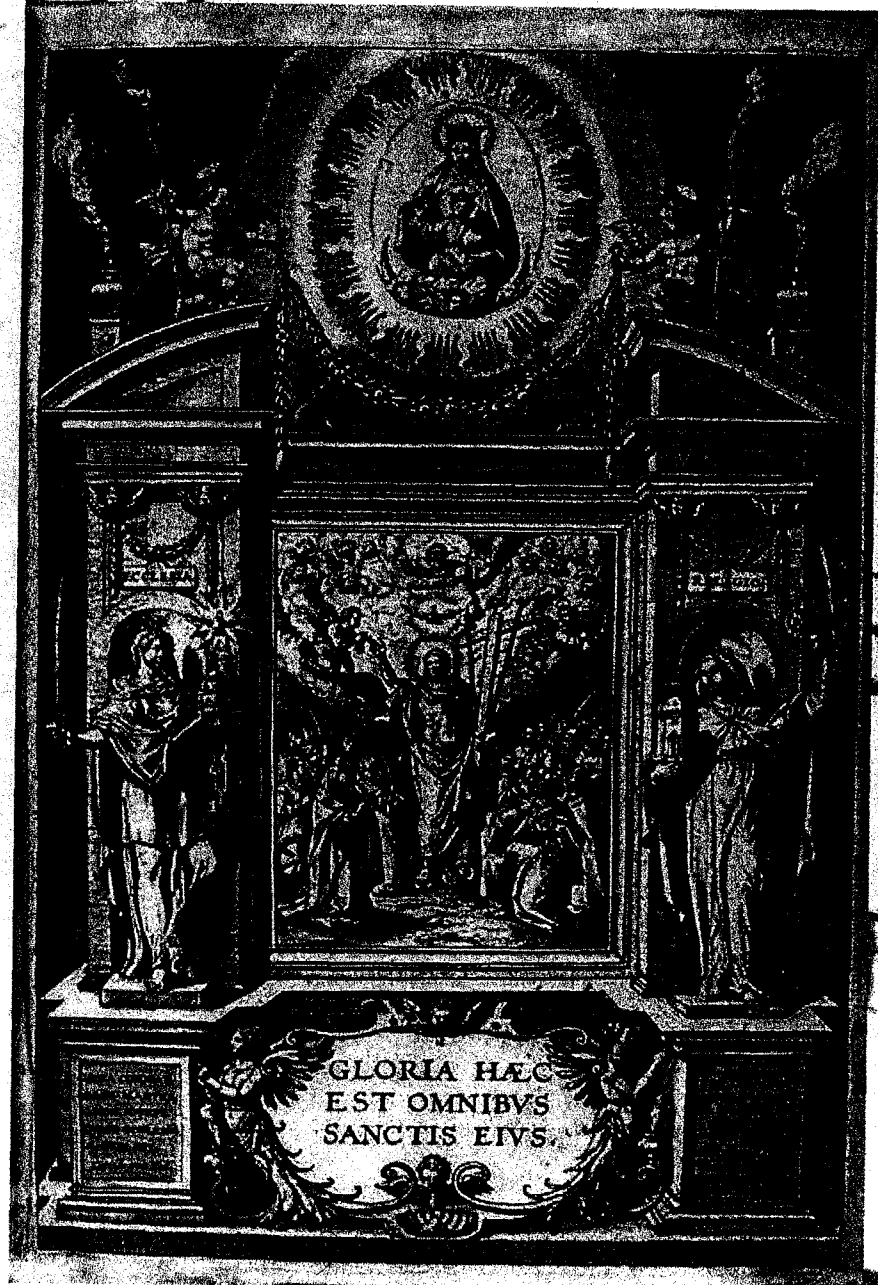
Miriam & Ira D. Wallach Division of Art, Prints and Photographs

Belnydenis

Introduction to the collection

Browse the collection

Index	Index
by	by
dates	names



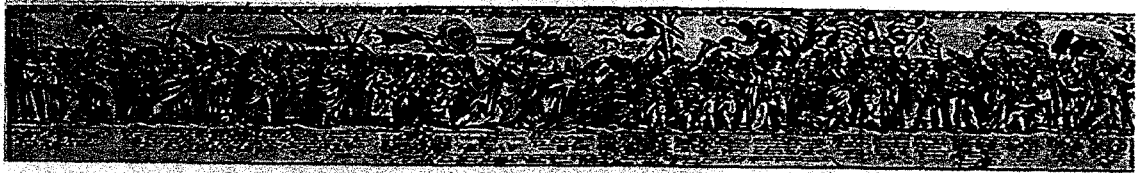
Mantua

M

IANUARIUS .

The
Journal
of
Saints

Visual Treasures of The New York Public Library



Print collection, Miriam & Ira D. Wallach Division of Art and, Print and photographs



Index by dates

index
by
names

Befnydenis

January	February	March	April	May	June
	16		16		16
1	17	1	17	1	17
2	18	2	18	2	18
3	19	3	19	3	19
4	20	4	20	4	20
5	21	5	21	5	21
6	22	6	22	6	22
7	23	7	23	7	23
8	24	8	24	8	24
9	25	9	25	9	25
10	26	10	26	10	26
11	27	11	27	11	27
12	28	12	28	12	28
13	29	13	29	13	29
14	30	14	30	14	30
15	31	15	31	15	31

Fulgentius

Marcellus

Odilo

July	August	September	October	November	December
	16		16		16
1	17	1	17	1	17
2	18	2	18	2	18
3	19	3	19	3	19
4	20	4	20	4	20
5	21	5	21	5	21
6	22	6	22	6	22
7	23	7	23	7	23
8	24	8	24	8	24
9	25	9	25	9	25
10	26	10	26	10	26
11	27	11	27	11	27
12	28	12	28	12	28
13	29	13	29	13	29
14	30	14	30	14	30
15	31	15		15	31

Concordia

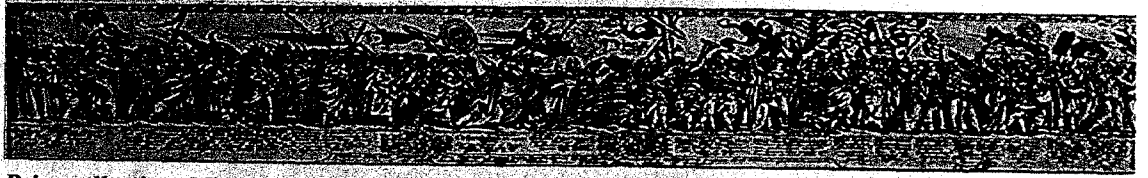
Europa

Carolus Thomatius

IANUARIUS.

The
Journal
of
Saints

Visual Treasures of The New York Public Library



Print collection, Miriam & Ira D. Wallach Division of Art and, Print and photographs



Index by names

Index
by
dates

Befnydenis

A-B-C-D-E-F-G-H-I-K-L-M-N-O-P-Q-R-S-T-U-V-W-X-Y-Z

Fulgentia

Abdon and Sennon Agathon

Aleta

Abraham

Agericus

Alexander

Abraham

Agilus

Alexander

Carbonatus

Martina

Abundus

Agnellus

Alexander Oliva

Acasius

Agnes of

Montepulciano

Alexandra

Achas

Agnes

Alexis

Achilla

<u>Achilles</u>	<u>Agricola</u>	<u>Alix Le Clerc (d. 1622)</u>
<u>Adalbert</u>	<u>Agrippina</u>	<u>Alipius</u>
<u>Adalbert of Prague</u>	<u>Aibertus</u>	<u>All Saints Day</u>
<u>Adalhard</u>	<u>Aichardus</u>	<u>Aloysius Gonzaga</u>
<u>Adelheid</u>	<u>Aidanus</u>	<u>Altamanus</u>
<u>Adelhindis</u>	<u>Aigulphus</u>	<u>Alto</u>
<u>Adelricus</u>	<u>Alatus</u>	<u>Amabilis of Riom</u>
<u>Adjutor</u>	<u>Alardus</u>	<u>Amadeus IX of Savoy</u>
<u>Adrian</u>	<u>Alban</u>	<u>Amalberga</u>
<u>Aegidius</u>	<u>Albertus Siculus9</u>	<u>Amandus</u>
<u>Aegil</u>	<u>Albert</u>	<u>Amatus</u>
<u>Aelfred</u>	<u>Albert of Oгна</u>	<u>AmbroseAlferius</u>
<u>Afra</u>	<u>Albert the Great</u>	
<u>Agape</u>	<u>Albert of Brabant</u>	
<u>Agapetus</u>	<u>Albert</u>	...
<u>Agapitus</u>	<u>Albinus</u>	...
<u>Agatha</u>	<u>Albricus</u>	...
<u>Agathengelus</u>	<u>Albantara</u>	...
<u>Agatho</u>	<u>Allegandis</u>	...
	<u>Aleidis</u>	...
	<u>Alena</u>	...
		...

IA NUAKIUS

The Visual Treasures of The New York Public Library
Journal of Saints



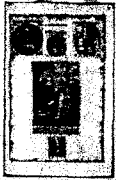
Print collection, Miriam & Ira D. Wallach Division of Art and, Print and photographs

January 17th



Bel

Fw



Previous

Ma



Next

Odilo

Sanctus
s. Antonius [s. Anthony]

Annexe 11 : " Creating Digitized Collections in the NYPL for the Performing Arts. Pilot project global cost estimation "³⁷

³⁷ Fiche détaillant les calculs pour l'estimation du coût de la numérisation des différentes collections.

Creating Digitized Collections in the NYPL for the Performing Arts

Rodgers and Hammerstein Archives of Recorded Sound

	Whole collection	Eligible items ¹	Item average running time	Selected items	Estimated cost ²
Acetate disc	60 000	12 000	15 mn	120 items 1 800 mn	\$90 000
Open reel tape	50 000	20 000	60 mn	200 items 12 000 mn	\$600 000
Video tape	8 000	3 000	60 mn	-	-
Audio cassette	25 000	10 000	60 mn	-	-
Audio cd	30 000	-	-	-	-
78	150 000	7 500	-	-	
LP	150 000	7 500	-	-	
45 single disc	50 000	-	-	-	
Total		60 000		320 items 13 800 mn	\$690 000

¹ Already catalogued and needing preservation

² Digitization in reel media format ; estimated cost : 50 \$ per minute.

Jerome Robbins Dance collection

	Whole collection	Eligible items ¹	Item average running time	Selected items ²	Estimated cost ³
Video	30 578	23 873	30 mn	575 items 17 500 mn	\$1 435 000
Reel	5 873				
Total	36 451				

¹ Already catalogued and needing preservation

² Around 2.5 % of the eligible items

³ Digitization in mpeg3 format ; estimated cost : 82 \$ per minute.

Theatre on Film and Tape archive

	Whole collection	Eligible items ¹	Item average running time	Selected items ²	Estimated cost ³
Total	20 000	20 000	60 mn	250 items 15 000 mn	\$1 230 000

¹ Already catalogued and needing preservation

² Around 1.5 % of the eligible items

³ Digitization in mpeg3 format ; estimated cost : 82 \$ per minute.

Pilot project global cost estimation

	# of items in the collection	# of eligible items for digitization	Pilot Project			
			# of items selected	Size of digitized collection		Estimated cost
				in minute & hours	in storage space	
Rodgers and Hammerstein Archives of Recorded Sound	138 000	60 000	320	13 800 mn 230 h	138 GB	\$ 690 000
Jerome Robbins Dance Collection	38 631	23 873	575	17 500 mn 290 h	525 GB	\$ 1 435 000
Theatre on Film and Tape Archive	20 000	20 000	250	15 000 mn 250 h	450 GB	\$ 1 230 000
Total :			1145	46 300 mn 770 h	1 113 TB	\$ 3 355 000

Annexe 12 : " Creating Digitized Collections in the NYPL for the Performing Arts " ³⁸

³⁸ Document final remis à Michael J. Alexander le 20 novembre 2000.

Creating Digitized Collections in the NYPL for the Performing Arts

Why now ?

If the digital revolution challenges the traditional activities of some libraries, it confirms the leadership of the NYPL for the Performing Arts in its field. Its immense and unique historical collection, its international reputation, and the skills and up to date technical competences it has developed since its creation are major strengths which should be used to promote the Performing Arts through new technologies worldwide. It is now the decisive moment to determine a strategy for the Performing Arts Library's future development in this area.

The question of the preservation of archival material is becoming more and more crucial. The Library for the Performing Arts - and hence the NYPL - must become a leader at an international level for the preservation of the sound and moving image documents of the performing arts.

As the Library for the Performing Arts will be soon housed in a technology intensive building, its collections and its services should also benefit from the newest network and digital technologies.

Securing the past: launching a preservation program to protect the historical collections of the Library for the Performing Arts

Priceless audio and video collections are threatened by media degradation and format obsolescence. The digitization process offers the possibility of protecting the documents by transferring them to a more reliable and durable media. This fulfills the preservation task of a major international repository. The creation of digital surrogates and their corresponding descriptions will also improve document access.

- Digitization of selected material from Jerome Robbins Dance division
- Digitization of selected material from Rodgers and Hammerstein Archives of Recorded sound
- Digitization of selected material from Toft Theatre on Film and Tape Archive

Enriching the future: collecting born digital documents in the domain of the performing arts

The digital revolution is creating new tools, new recording methods, new artistic forms in the domain of the performing arts. Unique on-line dance performances, sound broadcast on the web, new software to design sets, costumes or light, computer music, and new multimedia creations are today's Avant-garde expressions. The library must be able to collect and record these new forms of artistic expression. In this way it will become a pioneer in the archiving of digital performing arts.

- Collect and record the own production of the library - theatrical recordings, interviews, conferences - in a digital format
- Develop methods to collect new audio and video digital content

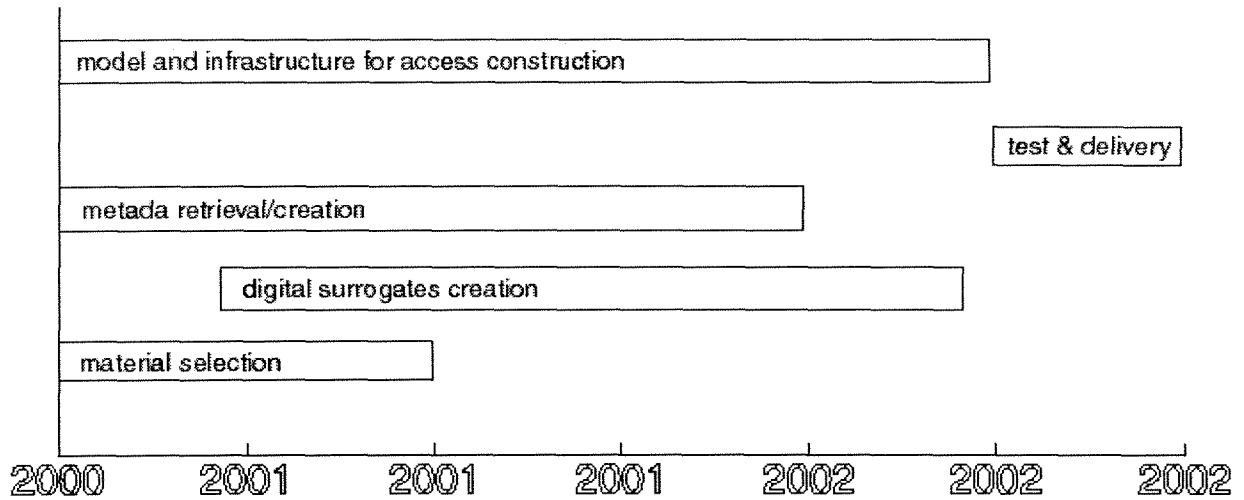
Developing accessibility for present and future generations

For the current and future digital collections, the Library for the Performing Arts must provide ready-access procedures to all forms of media.

- Develop methods and large facilities to store this content : durable and/or migratable sound and moving image formats ; durable and reliable medias
- Develop tools to give an easy access to this new digital content : finding aids ; local and/or distant streaming access to digital documents

NYPL for the Performing Arts digitization project

Pilot project timeline



Pilot project cost estimation

	# of items in the collection	# of eligible items for digitization	Pilot Project		
			# of items selected	Size of digitized collection	Estimated cost
Rodgers and Hammerstein Archives of Recorded Sound	138 000	60 000	320	138 GB	\$ 690 000
Jerome Robbins Dance Collection	38 631	23 873	575	525 GB	\$ 1 435 000
Theatre on Film and Tape Archive	20 000	20 000	250	450 GB	\$ 1 230 000
Total :			1145	1.113 TB	\$ 3 355 000

Illustrations

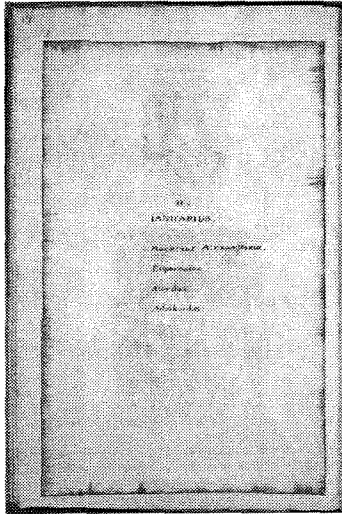


Figure 1: page de titre



Figure 2: St Fulgentius



Figure 3 : St François Xavier



Figure 4 : Ste Geneviève



Figure 5 : St Sebastien



Figure 6 : Ste Martina

TABLE DES MATIERES

<u>INTRODUCTION</u>	3
<u>1. Présentation de l'institution d'accueil : The Research Libraries of New York Public Library</u>	3
1.1. <u>NYPL, une bibliothèque à l'image d'une ville et à l'échelle du monde</u> 3	
1.2. <u>NYPL, une institution privée au service du public</u>	4
1.3. <u>The Research Libraries, une institution de recherche au sein d'une bibliothèque publique</u>	5
<u>2. Présentation du service : The Digital Unit</u>	6
<u>3. Présentation de la collection : Visual Treasures</u>	8
<u>CRÉER UNE COLLECTION NUMÉRIQUE : THE JOURNAL OF SAINTS</u>	9
<u>1. Sélection des matériaux</u>	9
1.1. <u>Les différentes collections en cours de numérisation</u>	10
1.2. <u>The Journal of Saints</u>	10
<u>2. Récupération des données et création des metadonnées</u>	11
<u>3. Processus de numérisation des images</u>	14
3.1. <u>Manipulation et conservation</u>	14
3.2. <u>La photographie numérique</u>	15
<u>MANAGER LES COLLECTIONS NUMÉRIQUES : UN IMPÉRATIF À L'HEURE DE LA PRODUCTION DE MASSE D'IMAGES NUMÉRISÉES</u>	15
<u>1. Le <i>post processing</i> : repenser les méthodes actuelles</u>	16
1.1. <u>Contrôle qualité</u>	16
1.2. <u>Duplication des fichiers, étiquetage, gravage des CD</u>	17

<u>2.</u>	<u>Archivage et conservation : l'occasion d'un nouveau départ</u>	18
2.1.	<u>Les méthodes actuelles</u>	18
2.2.	<u>Des perspectives nouvelles</u>	18
<u>3.</u>	<u>Mise à disposition et diffusion</u>	19
3.1.	<u>Adapter les méthodes à la politique de diffusion</u>	19
3.2.	<u>Rédaction d'une proposition <i>Web</i> pour le Livres des Saints</u>	20
3.3.	<u>Une solution pour la gestion des grands nombres : Le logiciel Insight de la société Luna</u>	20
<u>DÉVELOPPER LES COLLECTIONS NUMÉRIQUES : AU-DELÀ DES VISUAL TREASURES LA NUMÉRISATION DE L'IMAGE ANIMÉE ET DU SON DES COLLECTIONS D'ART DU SPECTACLE DE NYPL</u> ..		
1.1.	<u>Nouveau projet, nouvelle mission pour le stagiaire</u>	22
1.2.	<u>Veille technologique</u>	22
1.3.	<u>Evaluation des collections de la Performing Arts Library</u>	23
1.4.	<u>Chiffrage du coût et rédaction de la proposition</u>	23
<u>CONCLUSION</u>		24
<u>ANNEXES</u>		26
<u>ILLUSTRATIONS</u>		46