

Lignes de conduite sur la numérisation des documents d'archives

du BSG, P.O. Box 459, Grand Central Station, New York, NY 10163

INTRODUCTION

Nous vivons dans un monde de technologie en évolution. De plus en plus d'archives et de bibliothèques utilisent des dépôts numériques et testent des technologies qui ne produisent pas toujours les résultats escomptés. La création d'un dépôt numérique peut poser des défis, en partie à cause de la variété de sujets et de problèmes divers. Vous voudrez peut-être évaluer la méthodologie de la création de dépôts numériques, vous faire une idée des coûts que cela engendre et évaluer votre capacité à décider ce que vous voulez numériser avant de songer à créer une bibliothèque numérique.

De plus, vous devrez bien connaître les critères techniques des différents documents de votre collection, comme les documents manuscrits, les photos, les magazines, les livres, les documents sonores, les films et autres. D'autres aspects à considérer sont la sécurité de votre dépôt, les questions d'accès et d'utilisation ainsi que les questions concernant les serveurs et les systèmes informatiques de sauvegarde.

Comme les Archives du Bureau des Services généraux reçoit souvent des questions des Archives locales des AA sur la numérisation des collections, nous avons décidé de parler de certains de ces sujets. Ces lignes de conduites se veulent de nature informative ; elles ne traitent pas de tous les sujets concernant la gestion des dépôts numériques, mais c'est un bon début. Aucun dépôt, peu importe sa taille, ne pourra numériser la totalité de sa collection. La numérisation peut être onéreuse, demander beaucoup de temps et certains documents ne se prêtent pas à la numérisation, ou n'en valent peut-être pas la peine. Ce ne sont pas toutes les archives des régions ou des districts qui auront les moyens de numériser leurs collections, et cela est normal.

PROTÉGER LA VIE PRIVÉE ET L'ANONYMAT

La protection de la vie privée et de l'anonymat des membres des AA est une question importante qui doit être considérée quand on produit des dossiers numériques. L'élaboration des procédures semble être un sujet qui relève du comité local des archives, mais la nécessité

de protéger la confidentialité de la correspondance et l'anonymat des correspondants est, de toute évidence, une question importante et une responsabilité qui incombe à tous les archivistes et comité des archives des AA. Vous voudrez peut-être formuler une politique pour l'utilisation des documents numérisés et y adjoindre des lignes de conduite sur l'affichage de documents sur l'Internet. Pour plus d'informations sur l'anonymat en ligne, consultez les Lignes de Conduite des AA sur l'Internet, émises par le BSG (FMG-18).

QU'EST-CE QUE LA NUMÉRISATION?

La numérisation est la conversion d'informations, comme des textes, des photos, des documents sonores et vidéo, en format numérique. La numérisation permet de préserver le contenu des documents en créant un substitut accessible, ce qui évite de manipuler l'original. Les objets numérisés sont alors conservés dans un dépôt numérique qui offre une façon commode d'entreposer, de gérer, d'accéder et de préserver ces substituts.

Un dépôt numérique répond à trois besoins fondamentaux. Le premier, acquérir ou capter le contenu numérique. Le deuxième est l'entreposage et la gestion du contenu numérique et le troisième concerne l'accès au contenu numérique et la création de documents livrables (ou ce qu'on peut faire avec un contenu numérique). Pour répondre à ces besoins, il faut du matériel et des logiciels particuliers. Leur coût peut varier de zéro à très cher, selon le choix de matériel et de logiciels.

CONSIDÉRATIONS FONDAMENTALES DANS LA PLANIFICATION DE VOTRE PROJET

Avant de créer un dépôt numérique, vous devrez répondre aux questions suivantes: Pourquoi voulez-vous établir un dépôt numérique? Quelles raisons sont les plus importantes en rapport avec la mission de vos archives? Par exemple, vos raisons pour numériser pourraient inclure un meilleur accès et un meilleur usage, la sécurité, la conservation, la gestion et l'intégrité. Aux Archives du Bureau des Services généraux, nos efforts de numérisation ont contribué à améliorer l'accès ainsi qu'une meilleur

leure préservation des documents en réduisant la manipulation des originaux. Vous voudrez prendre le temps de décider quel équipement de numérisation convient le mieux à votre dépôt. Vous devriez faire une recherche approfondie avant d'acheter du matériel et des logiciels. Analysez les menaces potentielles à votre dépôt, incluant une défaillance des médias, du matériel, des logiciels ou une panne de courant. Voici quelques questions qui pourraient vous permettre d'éviter des conséquences désastreuses :

Vos archives peuvent-elles répondre aux besoins financiers à long terme que demandent les équipements et les logiciels requis?

- Quel médium utilisez-vous pour stocker les données?
- Comment allez-vous sauvegarder les données de vos périphériques de stockage?
- Avez-vous pensé à l'obsolescence du matériel et des logiciels?
- Comment allez-vous contrôler l'accès aux données?
- Comme allez-vous protéger les données des logiciels électroniques malveillants, des virus et autres?

Exigences des logiciels:

Il vous faudra d'abord déterminer le type de dispositif de numérisation qui répondra le mieux à vos besoins. Pour les petites collections locales, des applications sur Windows ou Mac pourront suffire. Pour les collections plus importantes (5000 images ou plus), il est recommandé d'utiliser des logiciels de Gestion du contenu [Content Management] et des dossiers.

Il est important d'avoir de bons outils et logiciels pour numériser, éditer les fichiers et travailler les dossiers en format PDF. Plusieurs vendeurs de numériseurs incluent des logiciels adaptés à leurs appareils. Il existe aussi plusieurs logiciels libres [open-source] que vous pourrez télécharger de l'Internet. Plusieurs vendeurs d'équipements et de logiciels accordent des rabais aux organismes sans but lucratif.

Les logiciels libres offrent gratuitement leur code-source à leurs utilisateurs. Ces logiciels ne comportent aucun frais de licence ni autres limitations. Ainsi, un utilisateur peut accéder au code-source, le modifier et le redistribuer. Les logiciels libres offrent rarement de soutien technique et les utilisateurs comptent sur la communauté en ligne pour des conseils et du soutien. L'utilisation d'un logiciel libre demande une certaine compétence

technique. Par contre, les logiciels propriétaires ou fermés sont des programmes distribués pour vente par des fournisseurs en qui on peut avoir confiance. Les ententes de licence des logiciels propriétaires contiennent des restrictions qui empêchent les usagers de les modifier et de les redistribuer. Les logiciels propriétaires offrent un soutien aux utilisateurs ce qui est particulièrement intéressant pour les personnes ne possédant de connaissances techniques avancées. Peu importe votre choix, vous devrez le modifier et vous aurez besoin d'un professionnel en TI pour installer et faire fonctionner votre système. Il est important de bien lire les termes de votre licence de logiciel avant de signer un accord. Une information importante que vous devrez connaître est le niveau de contrôle vous conserverez sur l'accès et l'édition de vos données à l'expiration de l'accord de licence.

Besoins en matériel:

Le matériel de base dont vous aurez besoin consistera en un ordinateur personnel (préférentiellement avec un processeur à haute vitesse), un moniteur de qualité supérieure, un lecteur de CD-ROM, un disque dur de 40GB (ou plus), une imprimante au laser couleur et noir et blanc, un numériseur qui peut traiter les photos et, au besoin, un scanner portable. La variété de documents à numériser (texte, photos, graphiques), vous dictera la décision à prendre à propos de l'équipement et de la taille du scanner dont vous aurez besoin. Les numériseurs à plat sont recommandés pour la plupart des projets de numérisation alors que les scanners portables seront utiles pour les documents super grands, les documents fragiles ou les livres reliés.

La meilleure façon d'entreposer les fichiers maîtres du dépôt est probablement sur un disque dur. Il est important de savoir que l'entreposage des fichiers pour accès futur sur des CD-R, DVD-R, disques Blu-Ray et disques flash ne garantit pas la longévité de vos fichiers maîtres. Si vous décidez d'entreposer vos documents sur des CD-R ou des DVD-R, nous vous recommandons d'utiliser des CD-R et des DVD-R de haute qualité ou de qualité «archives» (tels les CD-R Mitsui Gold Archive) et d'entreposer ces disques dans un environnement à température et humidité contrôlées.

Les technologies de logiciels et de périphériques demandent une attention soutenue car elles changent continuellement et rapidement. Quand une nouvelle technologie apparaît, elle remplace habituellement l'an-

cienne version assez rapidement. Quand une technologie logicielle est abandonnée ou qu'un périphérique n'est plus fabriqué, les dossiers numériques produits par de telles technologies pourraient être perdus. C'est ce qu'on appelle l'obsolescence numérique. Étrangement, nous pouvons toujours lire des documents originaux vieux de plusieurs siècles dans leur langue originale. Par contre, dans un espace de temps relativement court, les moyens de technologie numérique ont progressé des disques souples aux diskettes, aux CD, aux DVD, aux disques Blu-Ray et aux supports de mémoire flash. N'oubliez pas que les supports d'entreposage peuvent tomber en panne en tout temps. Archivez vos données sur plus d'un support et faites régulièrement des vérifications de vos archives.

Informez vous sur les «Sociétés de nuage» qui peuvent sembler offrir une option idéale de sauvegarde de vos données. La sauvegarde sur le nuage signifie que vous devez envoyer une copie de vos données par Internet vers un serveur hors site. Ces serveurs sont habituellement des tiers fournisseurs de services qui vous facturent selon l'importance de la bande passante, la capacité ou le nombre d'utilisateurs. La plupart de ces sociétés prennent le contrôle absolu sur les données que vous leur confiez. Avant d'adopter le nuage, lisez attentivement les conditions de service et examinez ces aspects importants:

- Questions de sécurité. Vos données sont-elles bien protégées?
- Que se produit-il si vous décidez de mettre fin au service?
- La société de nuage peut-elle conserver une copie de vos données?
- Qu'arrive-t-il si la société cesse ses activités et vend vos données?
- Rigidité. Assurez-vous que vous n'avez pas enfermé vos données dans une application ou en format propriétaire.

CHOIX DES ORIGINAUX À NUMÉRISER

Plusieurs archivistes sont confrontés à la question déroutante du choix des documents à numériser. Certains prennent connaissance des efforts de numérisation d'autres dépôts et concluent qu'ils ne procèdent pas de la bonne manière. Sachez que vos efforts pour préserver l'histoire des AA sont un bon point de départ. À cause du caractère unique de votre collection, il nous est impos-

sible de vous dire quels documents il vous faut numériser.

Soyez sélectif dans le choix des documents que vous planifiez de numériser en vous concentrant sur la qualité et non la quantité. Nous vous suggérons aussi de vous concentrer sur la numérisation des documents qui traite spécifiquement de votre région ou district, comme les procès verbaux, les affiches, la correspondance et autres articles produits par votre structure locale des AA. En numérisant ces articles, vous créez une copie numérique qui sera facile à accéder et utiliser, tout en protégeant l'article original d'une trop grande manipulation.

Certains documents sont numériques dès leur création, c'est-à-dire qu'ils ont été créés en format numérique (par ex. publication assistée par ordinateur, appareils photo numériques, et autres). Ils devraient néanmoins être importés dans votre dépôt numérique et sauvegardés dans un format qui respecte le plan que vous avez choisi pour les objets numérisés. Les textes peuvent être numérisés tant que la procédure ne risque pas d'endommager le document.

Les documents qui ne peuvent être numérisés par un scanner à plat sont appelés documents hors normes. Ces documents, tout comme les livres reliés, devraient être numérisés en utilisant un scanner portable, un appareil photo à scanner ou un appareil photo normal. Il n'est pas recommandé de numériser les livres reliés sur un scanner à plat car cela pourrait causer des dommages irréparables à l'épine ou à la reliure du livre.

CRÉER UNE STRUCTURE POUR VOS DONNÉES

Après avoir répondu adéquatement à ces importantes questions, vous devrez maintenant définir votre plan pour organiser vos données dans un ordre logique. Vous devrez aussi déterminer une convention de nomenclature pour vos dossiers et fichiers. Le système de nomenclature de dossiers devrait être choisi avant la numérisation pour respecter une logique. Les noms descriptifs de fichiers devrait respecter l'article numérisé et pourraient comprendre le nom de l'objet ou un numéro de dépôt. Vous pourriez envisager de fonder votre système de nomenclature de vos dossiers numériques sur la même base que celui que vous utilisez pour les fichiers en papier de votre dépôt.

Le Manuel des archives contient une page intitulée «Liste historique» des archives du BSG. C'est une structure à haut niveau de l'organisation et de la classification

de la collection des Archives du BSG. Vous devriez utiliser le même genre de structure simplifiée pour organiser le contenu des dépôts numériques.

Voici un exemple de la structure du dossier dans la base de données électronique des Archives du BSG :

Bulleting (Dossier mère)

➔ *Information sur les A.A. (dossier accessoire)*

- *Information sur les AA, Été 1972 (nom du fichier)*

➔ *Box 4-5-9 (dossier accessoire)*

- *Box 4-5-9, Automne 2010 (nom du fichier)*

MÉTADONNÉES

Toute personne qui s'adonne à un projet de numérisation rencontrera le terme «métadonnées». En langage simple, les métadonnées sont des «informations sur l'information». C'est une information utilisée pour décrire, localiser et récupérer les fichiers entreposés dans une bibliothèque numérique. C'est la clé pour s'assurer que les ressources électroniques sont toujours facilement accessibles à l'avenir. En Microsoft Windows et autres logiciels, les métadonnées sont aussi appelées «propriétés». Les deux mots sont synonymes.

La plupart des métadonnées, comme la taille du fichier et sa date, sont automatiquement générées lors de la création d'un fichier. Les logiciels d'application génèrent aussi des métadonnées lors que la création d'un fichier, tel le nom et le genre de fichier.

Les noms et chiffres devraient être simples et cohérents. Une bonne façon d'y arriver est d'établir un voca-

bulaire contrôlé. Il s'agit d'une simple liste structurée de mots et de phrases utilisées pour baliser le contenu et le récupérer en faisant une recherche. Utilisez des noms de dossiers et de fichiers et des mots clés faciles à comprendre. Un nom de fichier est une métadonnée et devrait décrire son contenu.

En bibliothéconomie, il existe plusieurs normes et schémas standards pour décrire les fichiers numériques de divers types de documents (livres, photos, audio, et autres) et chacun possède sa propre structure unique. Néanmoins, tant que vous resterez cohérents, vous pourrez facilement créer votre propre schéma de métadonnées qui respecteront les ressources et besoins de votre dépôt.

Le logiciel que vous choisirez pour votre dépôt utilisera son propre mécanisme pour ajouter les métadonnées. Par exemple, le dépôt des archives électroniques des Archives du BSG contient des milliers de documents de groupes. Pour retracer les informations d'un groupe en particulier, nous avons créé une structure de métadonnées qui sert uniquement aux Dossiers des groupes. Les champs que nous utilisons sont: «Nom du Groupe», «Numéro du Groupe», «Lieu de réunion», «Région», «District», «Ville», «État ou province» et «Pays». Ceci permet au logiciel de retracer efficacement un groupe quand nous mettons des données dans un des champs. Créer des métadonnées est plus complexe que ce que nous venons de décrire, mais n'oubliez pas de garder ça simple et cohérent.

Formats de fichiers	Caractéristiques techniques	Utilisation recommandée
TIFF (Tagged Image File Format)	<ul style="list-style-type: none"> • Compatible avec la plupart des plateformes. • Sans perte (aucune perte de qualité suite à la compression). • Accepte les grands fichiers. • Format souple. • Format préféré pour la préservation d'images. 	Recommandé pour l'archivage. Le plus universel et répandu.
JPEG (Joint Photographic Expert Group)	<ul style="list-style-type: none"> • Acceptable pour les images photographiques les moins compressées. • Perte de qualité chaque fois qu'un fichier JPEG est compressé et sauvegardé. 	Pour fichiers d'accès seulement. Non recommandé pour la production de fichiers maîtres.
PDF (Portable Document Format)	<ul style="list-style-type: none"> • Format complexe de fichier. 	Pour fichiers d'accès seulement. Non recommandé pour la production de fichiers maîtres ou l'archivage. Peut être traité pour la reconnaissance de texte OCR pour rendre les fichiers interrogeable.

SPÉCIFICATIONS DE NUMÉRISATION

Plusieurs facteurs peuvent affecter les images, qu'elles soient affichées ou imprimées. En conséquence, quand vous numérisez, il existe différentes résolutions et exigences de types de fichiers selon le type, le format et la nature de l'original. Cependant, il y a beaucoup trop de caractéristiques particulières à un document et chacun exige des paramètres spécifiques.

- Un document en caractères d'imprimerie (par ex. un document imprimé au laser ou typographié) il est recommandé de choisir des réglages de 300 ppp et une échelle de gris de 8 bits au minimum.
- Quand vous scannez un document difficile à lire, un manuscrit, des copies carbone ou des photos, les paramètres recommandés sont de 400 ppp et une échelle de gris de 8 bits.
- Si vous désirez une représentation exacte de la couleur originale d'un document, réglez le scanner à 300-400 ppp en mode RGB 24 bits.
- Vous devez utiliser les formats de fichiers ouverts car ils sont disponibles au grand public et reconnus par plusieurs logiciels.

Le tableau qui suit couvre les genres de fichiers ouverts les plus courants qu'on utilise pour les projets de numérisation.

L'Open Office XML est aussi un format de fichier ouvert utilisé par Microsoft Office 2007, 2010 et 2013. Ces fichiers incluent les DOCX (pour le traitement de texte), XLSX (tableurs et PPTX (exposés). Les versions plus anciennes de Microsoft Office pourraient ne pas être en format de fichier ouvert et en conséquence ne peuvent être ouverts que par Microsoft Office.

CONTRÔLE DE LA QUALITÉ

L'examen de la qualité des images numérisées est une partie importante du processus de numérisation. Il est aussi recommandé de faire une inspection régulière du scanner pour s'assurer que des débris comme la poussière ou des morceaux de papiers n'obstruent pas le plateau.

N'oubliez pas de toujours nettoyer votre surface de travail, y inclus le nettoyage quotidien de votre équipement de numérisation. Plusieurs documents ou livres anciens peuvent être poussiéreux ou moisissés et peuvent laisser des traces nuisibles sur le plateau. Vous devriez

régulièrement faire une vérification visuelle des images numérisées pour vous assurer de la qualité dans les domaines suivants :

- Utiliser le bon format (TIFF, JPEG, ou autre) et la bonne résolution.
- Bien orienter l'image numérisée.
- Dans le cas de documents de plusieurs pages, vous assurer d'avoir toutes les pages dans le bon ordre.
- Vous assurer que l'image numérisée est entière et non découpée.
- Vous assurer que l'image soit bien droite.
- Que l'image soit libre d'artéfacts numériques (poussière, morceaux de papier et autres).
- Vous assurer de l'uniformité de la couleur, du ton, de la netteté et du contraste de l'image.
- Les métadonnées — bien nommer les fichiers numériques.

PRÉSERVATION DES ORIGINAUX APRÈS LA NUMÉRISATION

La numérisation produit un substitut numérique de l'original. Il est important de bien prendre soin des originaux. Même si vous avez une copie numérique, il est vital de protéger les documents historiques de la détérioration en utilisant du matériel de qualité d'archive et le meilleur environnement possible pour l'entreposage.

Consultez les Lignes de conduite pour la Préservation des Archives du BSG pour plus d'informations et une approche plus détaillée de la préservation des documents d'archives.

VEUILLEZ NOTER...

Si vous avez des questions particulières et pour une liste de ressources additionnelles, communiquez avec les Archives du BSG à archives@aa.org ou au 212-870-3400. Vous trouverez d'autres informations importantes sur le site Web des AA du BSG, www.aa.org.

AUTRES SOURCES D'INFORMATION :

Veillez noter que les Archives du BSG ne cautionnent ni n'endossent ces sites Web, et ne vous les fournit qu'à titre de ressources externes.

National Archives and Records Administration (NARA)
<http://www.archives.gov/preservation/technical/guidelines.html>

Digital Preservation Coalition
<http://www.dpconline.org/advice/preservationhandbook>