

Bureau des normes pour la production et la conservation des ressources numériques (BdN-PCRN)

Ordre du jour de la réunion du 27 avril 2017, à 13 h 30
Salle 0270, Direction de la Bibliothèque

MEMBRES	ORDRE DU JOUR
<p>présents : Allard, Louise Bissonnette, Patrick Fortin, Émilie Lapierre Dominique Lasou, Pierre Léger, Sonia Raja, Aladin Ruest, Jean-François</p> <p>absents : Samson, Sarah</p> <p>secrétaire : Ruest, Jean-François</p>	<ol style="list-style-type: none"> 1. Adoption de l'ordre du jour 2. Fonctionnement du BdN 3. Dépôt numérique fiable (DNF) <ol style="list-style-type: none"> 3.1. Retour sur la présentation des consultantes 3.2. Impacts d'un DNF sur les projets de numérisation 4. Impartition de la numérisation des thèses rétrospectives <ol style="list-style-type: none"> 4.1. Résumé du travail effectué jusqu'à maintenant 4.2. Nos exigences de base pour la conservation 4.3. Présentations des avantages et inconvénient de l'ALTO et les formats JPEG 2000 vs TIFF 4.4. Autres caractéristiques du fichier de diffusion 5. Création de groupes de travail <ol style="list-style-type: none"> 5.1. Numérisation des collections de la BUL 5.2. Normes de numérisation 5.3. Politique de numérisation et de production numérique de la BUL 6. Prochaine réunion : mi-juin 2017

1. Adoption de l'ordre du jour

L'ordre du jour est adopté avec ce changement : le point 5 sera abordé à l'intérieur du point 2, fonctionnement du BdN.

2. Fonctionnement du BdN

Présence de Louise Allard

Louise Allard apporte le point de sa présence aux rencontres du Bureau des Normes. Elle n'a l'intention d'y participer que pour initier la constitution de cette nouvelle entité. Elle dresse un tableau d'ensemble des sujets qui seront développés en cours d'année, dont le projet de numérisation des thèses rétrospectives par impartition, un projet qui devrait débiter en juin.

Définir le nom de l'entité (Bureau des normes)

Le nom de l'entité étant très générique, une proposition plus spécifique a été faite au groupe : Bureau des normes pour la production et la conservation des ressources numériques (BdNPCRN)

Les membres sont en accord à la majorité. Le nouveau nom est donc adopté.

Fréquence des réunions

Un rythme de quatre réunions statutaires par année a été proposé et accepté.

Ces rencontres auront lieu à la mi-septembre, la mi-novembre, la mi-février ainsi que la mi-avril. Des réunions ad hoc pourront également être demandées par les groupes de travail dédiés.

Nomination d'un(e) président(e) et d'un(e) secrétaire ainsi que la durée de leur rôle (décision)

Louise Allard propose que la durée des rôles soit de deux ans.

Deux postes sont proposés : celui de président et celui de secrétaire. Étant donné que le premier projet du BdN est lié au projet de numérisation des thèses par impartition, un projet initié par la DATC, Louise Allard propose la nomination d'Émilie Fortin à titre de présidente du BdNPCRN. Cette nomination fait l'unanimité.

Le rôle du secrétaire, en plus des tâches habituelles, est de faire office de bras droit au président pour la préparation des dossiers d'approbation qui doivent être déposés au CDI. Jean-François Ruest se présente pour le poste et est élu sans opposition.

Louise Allard rappelle le mandat du BdNPCRN qui doit, pendant la première année, produire une politique de numérisation et de production numérique.

Création de groupes de travail dédiés

Les trois premiers groupes de travail ont été lancés lors de cette réunion. Le quatrième devra être précisé lors des prochaines rencontres. Les responsables de ces trois groupes ont été nommés et ce sont eux qui seront chargés de recruter les autres membres du groupe de travail. D'ici la prochaine réunion de la mi-juin, il faudra que les groupes soient constitués et qu'ils aient un plan de travail.

1) Priorisation de la numérisation des collections de la BUL

Mandat : identifier et prioriser les collections de la BUL à numériser ; identifier le niveau de conservation qui doit être appliqué à ces collections (conservation ou uniquement diffusion) ; proposer un calendrier qui commencerait en septembre 2017.

Membres proposés : Sonia Léger + invités.

Échéance du livrable : à la réunion de septembre 2017.

Note : la présidente fera suivre les livrables de ce groupe de travail au CDI pour approbation.

2) Normes de numérisation

Mandat : choix des normes à appliquer sur les projets de numérisation des collections qui seront à numériser en 2017-18. Ces normes devront couvrir l'image et le son, mais excluront la vidéo.

Membres proposés : Émilie Fortin + invités.
Échéance du livrable : à la réunion de septembre 2017.

3) **Logistique**

Mandat : faire l'inventaire des équipements actuels dans la BUL et assurer leur répartition en fonction des missions de chaque prestataire de numérisation dans la BUL ; voir à la mise à jour des équipements en fonction des besoins de numérisation exprimés ; s'assurer de l'espace disponible pour loger les collections numérisées ; assurer une veille sur la création de nouveaux dépôts numériques fiables et sur de nouveaux logiciels favorisant la création de dépôts numériques fiables.

Membres proposés : Aladin Raja + invités.

4) **Politique de numérisation et de production numérique de la BUL**

Dominique Lapierre informe le groupe qu'elle a commencé la francisation d'une politique de numérisation, mais qu'elle a mis le projet sur la glace, car l'avocat de l'Université l'a informée que ce ne serait pas possible de diffuser les thèses rétrospectives sans l'accord signé des auteurs. Suite au lancement du dernier groupe de travail, ce qui pourra être fait à la prochaine réunion du BdN, il faudra déterminer si le travail de Dominique pourrait servir de point de départ à la politique que le BdN doit rédiger.

3. **Dépôt numérique fiable [DNF]**

3.1. **Retour sur la présentation des consultantes**

L'ensemble des membres du groupe a assisté à la présentation des consultantes au sujet du DNF, Louise Allard ne fait donc qu'un retour rapide sur le sujet. Elle mentionne que même si la BUL attend avant de déposer ses documents numériques dans un DNF, elle doit faire en sorte que ses fichiers soient conformes dès leur création aux normes exigées par les DNF et une veille doit être assurée afin de surveiller les développements au Canada et dans les solutions logicielles (voir groupe de travail sur le volet logistique).

3.2. **Impacts d'un DNF sur les projets de numérisation**

Déposer des fichiers dans un DNF implique des normes de conservation strictes. Il faut donc savoir au début de chacun des projets de numérisation s'il sera versé dans un DNF ou pas. Émilie Fortin présente les quatre niveaux d'archivage selon la NDSA (National Digital Stewardship Alliance) et indique où la BUL se situe (voir annexe), ce qui donne une idée au groupe des améliorations que nous pouvons apporter pour nous rapprocher d'un DNF et améliorer nos pratiques en matière de conservation des fichiers numérique.

4. **Impartition de la numérisation des thèses rétrospectives**

4.1. **Résumé du travail effectué jusqu'à maintenant**

Louise Allard fait un résumé du travail effectué jusqu'à maintenant. Avec l'aide des consultantes, nous avons envoyé des critères de numérisation à trois prestataires qui ont chacun reçu 5 thèses papier qui offraient différents défis : impression des caractères, papier, couleur, etc. Le résultat de chacun des prestataires a été analysé en fonction des critères envoyés. Notre constat est que nos critères devront faire l'objet d'exigences plus précises dans le cahier des charges que nous élaborerons pour le projet-pilote à venir qui couvrira quelque 180 thèses.

4.2. **Nos exigences de base pour la conservation**

La BUL considère les thèses et mémoires comme des documents importants pour l'Université. En ce sens, des normes de conservation élevées seront appliquées. La plupart des normes ont été tirées du Recueil de règles de

numérisation rédigé en collaboration entre BAnQ, BnF et le Musée canadien de l'histoire. Voici les critères de numérisation qui seront appliqués et qui n'ont pas besoin d'être discutés :

- résolution de 300 ppp pour une feuille A4 ;
- numérisation en couleur pour les pages en couleur ;
- niveaux de gris pour les pages en noir et blanc ;
- ajout de métadonnées adéquates ;

4.3. Présentations des avantages et inconvénient de l'ALTO et les formats JPEG 2000 vs TIFF

Émilie Fortin présente le schéma ALTO (voir PowerPoint). Une discussion s'ensuit à savoir s'il est utile de demander au prestataire un fichier XML du texte des thèses et mémoires mis sous la forme du schéma ALTO. Aladin Raja suggère de demander une soumission qui ventile toutes nos exigences afin de déterminer leur coût. Ce sera fait. Nous établissons qu'il est fondamental de conserver le contenu des thèses et mémoires et que ce format permet à la fois de conserver le contenu, mais également de le situer sur une page. Il est également fait mention qu'il serait excessivement complexe de créer un fichier XML en ALTO si nous changeons d'idée et qu'il n'est pas créé au moment de la numérisation. Le groupe convient qu'un fichier texte de format ALTO devrait être demandé au prestataire.

Jean-François Ruest présente les avantages et les inconvénients des formats de fichier TIFF ainsi que JPEG2000 (voir PowerPoint). La discussion qui suit cette présentation ne permet pas de décider lequel des formats devrait être choisi, car l'impact financier en lien avec l'entreposage des fichiers est assez important. Il est entendu que Louise Allard présentera une fiche décisionnelle à ce sujet au Comité de direction du 4 mai.

4.4. Autres caractéristiques du fichier de diffusion

Émilie Fortin demande l'avis du groupe sur certaines caractéristiques du PDF de diffusion afin de déterminer ce qui est vraiment important de ce qui ne l'est pas, tout en prenant en compte que chaque demande incluse dans le cahier des charges aura une incidence sur le montant que couvrira la numérisation des thèses et mémoires. Elle présente quelques exemples reçus des prestataires.

- Pages blanches : les pages des plus vieux documents sont jaunies, ce qui résulte en des pages grises. Nous nous entendons tous pour dire que la lecture d'un texte sur fond gris n'est pas agréable, sans compter que l'impression donnerait un très mauvais résultat. Pierre Lasou mentionne qu'il n'est pas simple de blanchir un fond gris et qu'il est risqué de perdre du texte. Patrick Bissonnette indique qu'il serait possible de communiquer notre « recette » au prestataire, c'est-à-dire lui dire de quelle façon nous effectuons ces modifications avec Photoshop.
- Angle du texte : nous nous entendons que le texte doit être numérisé le plus droit possible et qu'il doit être redressé dans le PDF afin de l'aligner sur les extrémités de la page, ce qui assure une lecture agréable lors de la consultation du document.
- Pages horizontales : il y a consensus sur le fait qu'il est plus agréable d'avoir le texte dans le sens de la lecture et donc de faire pivoter les pages horizontales.
- Taille des pages : avoir des pages qui sont toutes de la même taille (ou sensiblement) une fois qu'elles ont été retaillées est une contrainte de base.
- Couleur dans le cas d'une vieille thèse : il s'agit d'un cas qui ne concerne pas le fichier de diffusion, mais plutôt le fichier de conservation, à savoir si une thèse avec son papier jauni doit être numérisée en couleur ou en niveau de gris. Une bonne discussion s'ensuit. Le groupe remarque que nous n'avons pas d'information à savoir quelle est la différence de poids d'un fichier en couleur versus en niveau de gris. Cette décision est retournée dans les mains de la DATC, responsable de la conservation des fichiers numériques.

5. Création de groupes de travail

Voir point 2.

6. Prochaine réunion : mi-juin 2017

Nous effectuerons un suivi sur les groupes de travail à savoir qui en fait partie et les prévisions de résultats d'ici la rencontre de la mi-septembre.

Nous ferons un suivi sur le projet-pilote d'impartition et sur ce qui a résulté des décisions du CDI.

LISTE DES ANNEXES :
Niveau d'archivage de la NDSA

Rédaction du compte rendu :

Jean-François Ruest

	Niveau 1 : protection des données	Niveau 2 : connaissance des données	Niveau 3 : surveillance des données	Niveau 4 : réparation des données
Entreposage et emplacement géographique	<ul style="list-style-type: none"> . Deux copies complètes qui ne sont pas sur le même serveur . Pour des données qui sont sur des supports hétérogènes (CD, disques durs), mettre les données sur des serveurs. 	<ul style="list-style-type: none"> . Au moins trois copies complètes . Au moins une copie dans une localisation géographique différente . Documenter le système de stockage (ce qu'il faut savoir pour l'utiliser) 	<ul style="list-style-type: none"> . Au moins une copie dans une localisation géographique différente qui a des risques potentiels différents . Avoir une procédure de surveillance de l'obsolescence du système de stockage. 	<ul style="list-style-type: none"> . Au moins 3 copies dans des emplacements géographiques différents avec des risques potentiels différents . Avoir un plan de conservation des fichiers et métadonnées sur des supports viables.
L'intégrité des données	<ul style="list-style-type: none"> . Vérifier l'intégrité du fichier lors de son arrivée si celle-ci a été fournie . Créer l'information sur l'intégrité si elle n'a pas été fournie 	<ul style="list-style-type: none"> . Vérifier l'intégrité sur tous les fichiers qui sont déposés . Utiliser des bloqueurs pour éviter de modifier des médias originaux s'il faut travailler avec . Vérifier les virus pour les contenus les plus à risques. 	<ul style="list-style-type: none"> . Vérifier l'intégrité des fichiers à intervalles déterminés . Garder dans des registres l'information sur l'intégrité ; les fournir sur demande lors d'audit . Avoir la capacité de détecter les données corrompues . Tester tous les contenus pour les virus 	<ul style="list-style-type: none"> . Vérifier l'intégrité de tous les fichiers en réponse à un événement ou activité spécifique . Capacité à remplacer/réparer les données corrompues . S'assurer qu'aucune personne n'a la possibilité de modifier toutes les copies.
Sécurité de l'information	<ul style="list-style-type: none"> . Identifier les autorisations de ceux qui peuvent lire, écrire, déplacer et supprimer les fichiers . Restreindre ceux qui ont des autorisations sur des fichiers individuels 	<ul style="list-style-type: none"> . Documenter la restriction d'accès aux contenus 	<ul style="list-style-type: none"> . Maintenir des registres pour savoir qui a fait quelles actions sur quels fichiers incluant les suppressions et les actions de préservation 	<ul style="list-style-type: none"> . Effectué des audits sur les registres
Métadonnées	<ul style="list-style-type: none"> . Inventorier le contenu et le lieu d'entreposage . S'assurer de sauvegarder l'inventaire dans un emplacement différent 	<ul style="list-style-type: none"> . Entreposer les MD administratives . Entreposer les MD de transformation et enregistrer les événements 	<ul style="list-style-type: none"> . Entreposer les MD techniques et descriptives selon des standards 	<ul style="list-style-type: none"> . Entreposer les MD de préservation selon des standards.
Formats de fichiers	<ul style="list-style-type: none"> . Lorsqu'il est possible de donner son avis, encourager l'utilisation d'un ensemble défini de formats ouverts et de codecs 	<ul style="list-style-type: none"> . Inventorier les formats de fichiers utilisés 	<ul style="list-style-type: none"> . Faire le suivi sur l'obsolescence des formats 	<ul style="list-style-type: none"> . Effectuer la migration des formats, l'émulation et des activités semblables si nécessaire.