

Réunion Groupe Numérisation
28 avril 2020
Synthèse des échanges

=====

Présents :

AD09 : Marigeorges Allabert

AD11 : Anne Robert, Claude Robion

AM Narbonne : Cécilia Gazel

BM Nîmes : Didier Travier

AM Toulouse : Audrey Joly, Pierre Gastou

BEP Toulouse : Sophie Grabielle

AD34 : Céline Dehondt, Kristina Caron

AD48 : Fabienne Salles

AD65 : Cécile Ricard

AD81 : Valérie Gautrand

Arkhenum : Amandine D'Hubert, Damien Tekla

Occitanie Livre & Lecture : Mélanie Marchand

Excusés :

AD12 : Anne-Lise Delouvrie

=====

1. Bilan des contrôles – phase 2019

Retours contrôles et reprises :

Anne Robert (AD11) a relevé un souci d'images légèrement penchées, surtout pour *le Républicain de Narbonne*, à hauteur de 18% du titre)

> Amandine D'Hubert (Arkhenum) propose un redressement automatique pour l'ensemble du titre.

Didier Travier (BM Nîmes) revient sur l'OCR. Lors des contrôles, certaines pages n'étaient absolument pas reconnues.

> Amandine D'Hubert (Arkhenum) précise qu'il s'agit d'OCR brut (traitement logiciel sans intervention directe de leur part). Il peut advenir des problèmes de passage d'OCR de manière aléatoire. Dans ce cas précis, un nouvel essai sera effectué.

Marigeorges Allabert (AD09) a noté l'existence de fichiers nommés « _bis ».

Ces fichiers « _bis » sont présents pour l'ensemble des collections.

S'il s'agit dans certains cas d'une erreur d'indexation (mauvaise date), il peut s'agir également de numéros différents parus à la même date.

Plusieurs solutions de nommage :

Mélanie Marchand a proposé de renommer les fichiers bis, en continuation de _05 à _08, afin d'avoir 1 seule entrée par date.

Didier Travier remarque qu'il ne s'agit pas de la solution choisie par la BnF (qui différencie par exemple les suppléments). Il suggère de renommer plutôt avec _1 et _2.

Exemple : CODE ETS_COTE_AAAAMMJJ_0001 et CODE ETS_COTE_AAAAMMJJ_0001_bis

 > CODE ETS_COTE_AAAAMMJJ_1_0001 et CODE ETS_COTE_AAAAMMJJ_2_0001

> Alice Martin-Vallas vérifie la possibilité technique

> Mélanie vérifie les préconisations de la BnF (également pour les suppléments sportifs numérisés en 2019 pour la BEP de Toulouse)

Améliorations en vue de la phase 2020 :

Anne Robert (AD11) souhaiterait revoir la structuration des fichiers dans le DD de contrôle :
Afin de faciliter les contrôles, il serait souhaitable de ne pas avoir un sous-dossier par fascicule. Cela permettrait une vision d'ensemble des vignettes et un gain de temps.

> Arkhenum génère directement les sous-dossiers par fascicule. Amandine D'Hubert propose de rajouter un répertoire jpeg en plus de la structuration habituelle.

[Pour mémoire, Dossier Titre > sous-dossier fascicule > sous-dossiers par extension (jpeg, tiff, pdf, xml...) > fichiers]

Didier Travier (BM Nîmes) s'interroge sur le contrôle d'exhaustivité. Le prestataire a-t-il besoin d'un certain format ?

> Amandine D'Hubert (Arkhenum) confirme le croisement des données issues du fichier de recolement (si bulletiné) et des données indexées issues de la numérisation.

Le fichier de recolement avec 1 ligne par fascicule mentionnant les colonnes année/mois/jour (AAAA/MM/JJ) est suffisant.

Cécile Ricard (AD65) s'interroge sur la pertinence du contrôle de l'OCR.

> Il est décidé de l'alléger en vérifiant avec les fichiers pdf qu'il est bien placé (= xml-alto) et qu'il est bien « passé ». Pour le taux de reconnaissance, il est proposé de le vérifier ponctuellement.

Marigeorges Allabert (AD09) demande à ce que les métadonnées (IPTC, EXIF) à vérifier soient précisées.

Calendrier et mise en ligne :

Suite aux réunions du sous-groupe METS, Alice Martin-Vallas a finalisé le schéma METS.

Les métadonnées METS pour la phase 2018 (est et ouest) devraient être générées courant mai.

Les métadonnées METS pour la phase 2018 (est et ouest) devraient être générées courant juin.

La mise en ligne des phases 2018 et 2019 est envisagée pour juillet 2020.

2. Phase 2020

Structure		Vol.	Enlèvement	Tests	Num	Contrôle
AD09	Convention ok recolement ok	13000	Mi mai	Fin mai		
AD34	Convention ok recolement ok	18000	15 mai	Fin mai		
AD48	Convention ok recolement ok transport des collections aux AD34	14000	15 mai	Fin mai		
AD11	Convention ok Récolement en cours (fini fin mai)	72000	Début juin	Mi juin		
BEP Toulouse	Convention ok Récolement en cours (fini fin mai)	49000	Début juin	Mi juin		
AD81	Convention ??? Récolement en cours	11000	Début juin	Mi juin		
BM Nîmes	Convention ??? Recolement en cours (fini fin mai ?)	110 000	Fin juin	juillet		
AD 12	Convention juillet ou septembre Recolement ok	27 000	septembre	septembre		

Question du nommage des fichiers

Il avait été envisagé de renommer les fichiers avec l'ISSN.

Le retour d'expérience des AM de Toulouse n'est pas concluant, en termes d'ergonomie.

Pour mémoire, l'annuaire de Haute-Garonne a été publié sous de multiples ISSN successifs.

> Il est décidé de garder l'ancien nommage [CODE ETS_COTE_AAAAMMJ_0001.ext].

L'ISSN est intégré aux métadonnées METS, ainsi que l'ISSN précédant et suivant en cas de filiations.

Facturation :

il sera envoyé en début d'opération un **devis** basé sur le nombre estimatif de pages à numériser (soit de mai à juin en fonction des établissements).

Le coût total sera calculé après validation du nombre effectif de vues numérisées. Sauf avis contraire, une seule facture sera émise et ce, avant la fin de l'année budgétaire 2020.

3. Divers

Phase 2021

Il est proposé de réfléchir aux propositions de titres pour la phase 2021 :

A ce jour, j'ai noté :

- AD65 : presse thermale (corpus à définir)
- AD81 : Le Patriote Albigeois (36 000 pages)
- AM Narbonne : Le Midi vinicole (avec les AD34 ?)
- BEP Toulouse : Le Télégramme (44 000 pages)

Il est demandé aux établissements qui ne pensent pas pouvoir participer à la phase 2020 en raison d'une baisse du budget alloué à la numérisation de l'indiquer dès que possible à la chargée de mission.

Pour mémoire, la proposition de corpus est envoyée à la BnF pour avis en septembre.

Point sur le projet de valorisation thermalisme

La Direction Culture et Patrimoine de la Région développe des « récits multimédias » s'attachant à valoriser les données de l'Inventaire régional.

Sur la thématique du thermalisme, la DCP et OLL avons souhaité croiser ces données avec les fonds patrimoniaux des bibliothèques et archives.

Une représentativité (et non pas une exhaustivité) des différentes stations thermales sur le territoire est bien évidemment recherchée.

Suite aux réponses des établissements, plusieurs corpus ont retenu l'attention du comité de pilotage (entre autres, le fonds Garrigou de la BU de Toulouse, le fonds des archives de la station thermale de Ginoles des AD de l'Aude, la presse thermale de Bagnères-de-Bigorre des AD de Hautes-Pyrénées, un album souvenir manuscrit d'un voyage à Amélie-Les Bains de la médiathèque de Perpignan) Certains sont numérisés et d'autres pourront être amenés à l'être. Le corpus est en cours de définition.

Afin de proposer un regard contemporain sur ces fonds, 8 auteurs en région ont été invités à s'inspirer de ces fonds pour créer une narration, en se fondant sur des profils sociologiques (les médecins, les personnels des bains, les mondains, les habitants locaux, les militaires...)

Prochaine réunion du groupe à l'automne 2021

> Point sur la phase 2020

> Point sur la phase 2021

> La question des éditions locales

> La question de la conservation pérenne des données (création d'un sous-groupe ?)

Certains membres du groupe évoquent la possibilité de réitérer les réunions en visioconférence.

Une limite évoquée est la difficulté de télécharger un logiciel adéquat (pour certaines collectivités, la DSI pourrait anticiper l'installation mais d'autres craignent un refus catégorique).

Il existe aussi la possibilité de faire une visioconférence dans les locaux d'Oll à Toulouse et Montpellier.