

Proposition de communication pour le n° 151 d'En Direct (partie chantiers scientifiques)

Titre

« Retour vers le futur, saison 2, épisode 4 »

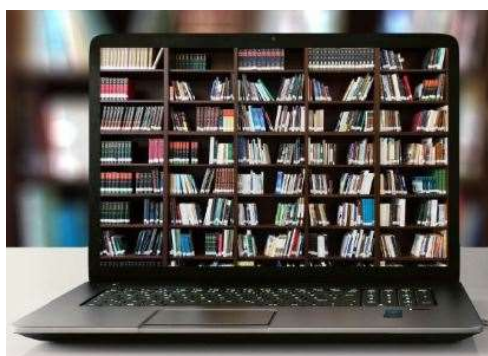
Sous-titre

Le chantier de reprise des données numériques archivées (1983-2018) : le traitement des fichiers des synthèses journalières de la Direction Centrale de la Sécurité Publique

Accroche diffusée dans la partie mail

Retrouvez dans *En direct* la deuxième saison de « Retour vers le futur » et suivez les orientations prises par cette nouvelle saison. [lire la suite]

Illustration



©Pixabay

Texte de l'article [lien vers sémaphore]

Six versements numériques constitués des fiches de synthèses journalières produites par la Direction Centrale de la Sécurité Publique (DCSP) du ministère de l'Intérieur (1996-2011) seront bientôt pérennisés dans les modules numériques du SIA (Système d'Information Archivistique) dans le cadre de la reprise des données.

Ces synthèses, rédigées par le bureau de l'information et transmises à la Direction Générale de la Police Nationale (DGPN), rapportent les événements quotidiens qui se sont déroulés sur le territoire national et qui ont nécessité le recours aux forces de l'ordre. Des statistiques ou des dossiers thématiques ont aussi pu être réalisés par la police dans le cadre de ces synthèses, ce qui accroît l'intérêt de ce fonds.

Le chantier de reprise est mené par le département de l'administration des données (DAD) en étroite collaboration avec le département Justice et Intérieur (DJI).

Au cours de l'étude sur les données bureautiques, cet ensemble a été repéré parce qu'il couvre une même typologie et forme une suite chronologique d'un même fonds ouvert. Les métadonnées des instruments de recherche qui les décrivent dans la salle des inventaires virtuelle (SIV) ont été extraites, permettant une observation plus fine des fichiers et un enrichissement des métadonnées de contexte.

L'héritage de CONSTANCE

Les 9 163 fichiers que représentent les 6 versements avaient été pérennisés selon la méthode CONSTANCE, utilisée depuis les années 1980 par les Archives nationales pour conserver et traiter les archives nativement numériques. De ce fait, ils sont mis et stockés « à plat » c'est-à-dire tous au même niveau du répertoire racine.

Cette méthode prévoyant un nommage spécifique (numéro du producteur + numéro de versement + numéro de l'article + nom du fichier donné par le producteur), le rangement à plat des fichiers n'interdit pas du tout la compréhension du classement d'origine, ces métadonnées permettant de reconstituer l'arborescence du versement. Mieux encore, dans le cas des synthèses journalières et grâce à ce nommage rigoureux, il a été possible d'extraire nombre d'informations sur l'état des fichiers et sur leur contexte de production.

Audit de l'état des fichiers grâce à leur nommage signifiant

Les synthèses journalières sont divisées en deux types : celles rapportant les faits survenus de 17h à 5h, diffusées le matin, et celles relatant les faits de 5h à 17h et diffusées le soir. Le nommage permet de faire ressortir la typologie de chaque fiche mais surtout le quantième de l'année, c'est-à-dire un numéro entre 1 et 365 (ou 366 les années bissextiles), en fonction de leur date de diffusion.

L'analyse effectuée sur les noms relègue la permanence d'un chiffre attendu de synthèse par année, à savoir 365 (ou 366), par type de synthèse ainsi que la date de diffusion d'un document.

Un programme informatique, créé à cet effet, vérifie alors de manière automatisée la liste des fichiers, permettant à la fois de gagner en temps de traitement et d'éviter les erreurs dues à la masse de fichiers à traiter, en repérant notamment les lacunes et les erreurs de nommage.

Une meilleure connaissance du fonds

Les résultats obtenus ont, par exemple, souligné que les lacunes étaient plus nombreuses dans les synthèses du soir. Cette vue macroscopique sur les fichiers a révélé que ces lacunes n'en étaient pas et permis de conclure que ces rapports du soir n'étaient pas produits les week-ends et jours fériés.

Cette vue plongeante dans les fichiers a également permis d'observer une évolution dans la pratique des tris et éliminations et dans le nommage des fichiers à partir de 2009. Il est intéressant de noter que cette évolution, relevée directement à partir des fichiers, se trouve confirmée par le changement du bureau en charge des synthèses l'année précédente, d'où cette différence dans les méthodes de travail.

Ces éléments, et d'autres, ont nourri un rapport que le DJI et le DAD, ont décidé d'ajouter au paquet d'informations à verser. Cela permettra aux chercheurs et archivistes travaillant sur les synthèses journalières de comprendre la nature des anomalies de ces fichiers et une connaissance accrue de ce fonds.

Vers la constitution du SIP (*Submission Information Package*)

Les travaux du DJI et du DAD ont permis d'aborder d'autres questions soulevées par ces travaux.

Cette analyse n'en est que la première étape et la constitution du SIP, c'est-à-dire de boîtes numériques contenant les archives et leurs métadonnées, dans le SIA numérique en est la prochaine. Avec elle se pose la question du classement de ces fichiers, de l'enrichissement des métadonnées de description (existantes ou non), du choix des métadonnées de gestion et enfin la place de l'analyse, que nous avons présentée, dans ce SIP.

Actuellement, des tests de constitution du SIP sont en cours, nous ne manquerons pas de vous tenir informé-e-s de la suite de ce petit chantier !

Contact

Émeline Levasseur – Cheffe de projet archivage électronique – DAEAA/DAD

emeline.levasseur@culture.gouv.fr

Jeanine Gaillard – Experte fonctionnelle archivage électronique – DAEAA/DAD

jeanine.gaillard@culture.gouv.fr

Julien Fenech – Stagiaire sur le projet reprise des données – DAEAA/DAD

julien.fenech@culture.gouv.fr