

LIVRE DES RÉSUMÉS

Journée d'étude du 26 mai 2021

**Documents et données :
l'université à l'ère
de la transition numérique**

Table des matières

Edito	3
Organisation	4
Comité scientifique	4
Comité d'organisation	4
Partenaires	4
Programme	5
Résumés des communications	8
Références bibliographiques	23
Index de mots-clés	25
Table des sigles et abréviations	26
Bloc-notes	28

Edito

La transition numérique bouleverse en profondeur nos pratiques professionnelles et pose plusieurs défis aux institutions qui peinent à s'y adapter, faute de connaissances et de stratégies appropriées. L'enseignement supérieur et la recherche ne sont pas exempts des bouleversements qu'elle induit.

Dans un monde numérique en perpétuelle mutation, la maîtrise de la production massive de documents nativement numériques ou numérisés et de données, la gestion de leur cycle de vie, leur conservation constituent un enjeu majeur de compétitivité. L'accès à l'information est certes facilité, mais il soulève néanmoins des questions quant au cadre juridique adapté, notamment la protection des données à caractère personnel ou encore la fiabilité et la sécurité des sources. La conception et la mise en œuvre de dispositifs de gestion des documents et données nécessite une approche interdisciplinaire, ce qui signifie une fédération d'acteurs d'horizons divers : DSI, bibliothécaires, archivistes, délégués à la protection des données, enseignants-chercheurs pour une culture commune et une vision intégrée. Sans oublier l'utilisateur qui doit être placé au cœur des dispositifs.

Dans ce contexte, il semble impératif d'accompagner la transition numérique de nos institutions en garantissant la préservation de son patrimoine informationnel, mais également les services rendus aux usagers.

Cette journée d'étude organisée par le Service des archives, en partenariat avec la Direction de la Recherche et de la Valorisation, de l'université Paris 1 Panthéon-Sorbonne, l'Amue, la DINUM et le SIAF sera l'occasion de faire le point de manière transversale sur les méthodes de traitement, de conservation à long terme, de valorisation des données (données pédagogiques, administratives, de recherche, ressources des bibliothèques et centres de documentation...) et l'apport de ces nouvelles sources aux chercheurs.

Organisation

Comité scientifique

Lorène BECHARD, Archiviste et Responsable fonctionnel plateforme PAC, Département Archivage et Diffusion, CINES

Marie-Anne CHABIN, Professeure associée, Université Paris 8 - Vincennes - Saint-Denis

Cyprien HENRY, Chef de la Mission des archives et du patrimoine culturel MEN-MESRI, Ministère de l'Éducation nationale, de la Jeunesse et des Sports

Marie-Caroline LUCE, Cheffe du Service des archives, Université Paris 1 Panthéon-Sorbonne

Marion MASSOL, PMO, Experte en archivage électronique, Département Stratégie et Programmation SI, AMUE

Magalie MOYSAN, Maîtresse de conférences en archivistique, Université d'Angers

Éric ZYLA, Directeur de la Recherche et de la Valorisation, Université Paris 1 Panthéon-Sorbonne.

Comité d'organisation

Marie-Caroline LUCE, Cheffe du Service des archives, Université Paris 1 Panthéon-Sorbonne

Marion MASSOL, PMO, Experte en archivage électronique, Département Stratégie et Programmation SI, AMUE

Partenaires

Agence de Mutualisation des Universités (Amue)

Direction interministérielle du numérique (DINUM)

Direction de la Recherche et de la Valorisation, Université Paris 1 Panthéon-Sorbonne (DIREVAL)

Service interministériel des archives de France (SIAF)

Programme

- 9h00** Accueil des participants
- 9h15** Discours introductifs
Christine **NEAU LEDUC**, *Présidente, Université Paris 1 Panthéon-Sorbonne*
Stéphane **ATHANASE**, *Directeur, Agence de mutualisation des universités*
Patrick **RUESTCHMANN**, *Chef de la mission Transformation numérique de l'État, DINUM*
- SESSION 1 : LA TRANSITION NUMERIQUE, UN SUJET REGLEMENTAIRE ET TRANSVERSE**
Présidence : Magalie **MOYSAN**, *MCF en archivistique, Université d'Angers*
- 9h45** Manager la donnée dans l'enseignement supérieur et la recherche : une approche par les métiers et compétences
Antoine **BLANCHARD**, *Consultant, Dataactivist*
Romain **PIERRONNET**, *Chercheur associé, Institut de Recherche en Gestion, Université Gustave Eiffel – Université Paris-Est Créteil et consultant, cabinet Adoc Mètis*
- 10h05** Les mails engageants à l'université : aspects juridique, informatique et organisationnel
Marie-Anne **CHABIN**, *Professeure associée, UFR Sciences et Technologies du Numérique, Master GSI, Université Paris 8*
- 10h25** Comment s'applique le RGPD dans un ensemble de données à caractère personnel et non personnel ?
Kin **MATCHEU-MADJEU**, *DPO, ENSAM*
- 10h45** Echanges avec le public
- 11h15** Pause
- SESSION 2 : LA NECESSAIRE MUTUALISATION**
Présidence : Cyprien **HENRY**, *Chef de la Mission des archives et du patrimoine culturel, MESRI*
- 11h30** Orientations stratégiques SI de l'Amue : zoom sur un futur dispositif co-construit avec la communauté ESR pour la réutilisation des données de gestion et des autres informations de l'établissement
Marion **MASSOL**, *PMO, Département Stratégie et Programmation SI, Amue*
- 11h50** Mettre en œuvre l'archivage électronique en université : le choix de la mutualisation pour les universités bordelaises et de Pau et des pays de l'Adour
Béatrice **CASTETBON**, *Chargée de mission Dématérialisation et archivage électronique et responsable du pôle SI à la Direction du numérique, Université de Pau et Pays de l'Adour*
Lorène **BECHARD**, *Archiviste et Responsable fonctionnel plateforme PAC, Département Archivage et Diffusion, CINES*
- 12h10** Echanges avec le public
- 12h30** Déjeuner

SESSION 3 : **ANALYSER, GERER ET DIFFUSER LES DONNEES DE RECHERCHE**

Présidence : Marie-Laure **BACHELERIE-GOUVERNEUR**, *Responsable du Pôle national de conservation des données et documents, Direction des affaires juridiques, CNRS*

14h00 Cycle de vie des données d'enquêtes à l'Institut national d'études démographiques (Ined)
Lindsay **FUSFELD**, *Archiviste, DPO, Ined*

Arianna **CAPORALI**, *Ingénieure de recherche, Service des enquêtes et sondages, Responsable du DataLab, Ined*

14h20 dat@UBFC, un service de gestion des données de la recherche pour la COMUE UBFC
Sylvie **DAMY**, *MCF en Informatique, Université de Bourgogne Franche-Comté, UMR 6249 Laboratoire Chrono-environnement, UFR Sciences et Techniques, Besançon*

Rachel **PROST**, *Documentaliste du projet dat@UBFC, OSU THETA, Professeure documentaliste, Université de Franche-Comté, Académie de Besançon*

14h40 Vers un entrepôt de données scientifiques pour l'université : retours d'expérience du projet Data Paris 8

Goran **SEKULOVSKI**, *Chargé de mission Données de la recherche, Université Paris 8*

Johann **HOLLAND**, *Chargé de mission MSH, TGIR Huma-Num*

15h00 Echanges avec le public

15h30 Pause

TABLE RONDE : **ENRICHIR POUR EXPLOITER**

Présidence : Sébastien **OLIVEAU**, *Directeur, TGIR PROGEDO, USR 2006 CNRS-EHESS*

15h45 L'archivage du web à la BnF : le cadre légal, les conditions d'exploitation et l'accompagnement des chercheurs

Dorothee **BENHAMOU-SUESSER**, *Chargée de collections numériques, Responsable de l'accès aux archives de l'internet et de la préservation, BnF*

Alexandre **FAYE**, *Chargé de collections numériques, Responsable projets de recherche et action internationale, BnF*

16h05 L'archive augmentée : cohérence et continuité des collections à l'INA

Jérôme **THIEVRE**, *Responsable R&D au Dépôt légal web, INA*

16h25 Echanges avec le public

16h45 Conclusion générale

Françoise **BANAT-BERGER**, *Directrice, Service Interministériel des Archives de France*

Résumés des communications

Session 1 :

La transition numérique, un sujet réglementaire et transverse

Présidente de séance : Magalie MOYSAN, *MCF en archivistique, Université d'Angers*

Cette première session interroge la transition numérique au prisme de la réglementation et de l'organisation des établissements d'enseignement supérieur. Il s'agit d'aborder ici la transformation des organisations en lien avec la dématérialisation et le rôle croissant des données. Trois aspects seront plus précisément étudiés : le management de la donnée, la gestion des emails engageants et l'application du Règlement européen sur la protection des données (RGPD).

Manager la donnée dans l'enseignement supérieur et la recherche : une approche par les métiers et compétences

Résumé : Au-delà de l'informatisation désormais bien ancrée dans les nombreux métiers présents dans les universités, la transition numérique constitue un enjeu managérial central qui questionne les compétences de l'ensemble des parties prenantes du monde universitaire : DSI, chercheurs, bibliothèques et archives, fonctions support, tutelles... En particulier, la maîtrise de la donnée est devenue, à tous les niveaux et dans tous les métiers, un enjeu croissant dont la vocation est désormais d'être moins une affaire de spécialistes que l'affaire de toutes et tous.

Dans cette communication, nous rapprocherons la transition numérique et les transformations managériales du secteur public en général, universitaire en particulier, afin de proposer une discussion prospective et systémique autour de la question suivante : en quoi l'enjeu de développement d'une culture partagée de la donnée interroge-t-il plus globalement l'ensemble du management des établissements de recherche et d'enseignement supérieur ?

Pour ce faire, nous avons réalisé une analyse quantitative de fiches métiers du secteur académique, que nous complétons et mettons en perspective avec les enseignements issus de 12 entretiens exploratoires réalisés dans deux universités, où nous avons interrogé des acteurs du niveau « central » et du niveau « périphérique ». Nous mettons ainsi en évidence les transformations et les progrès réalisés, les acteurs impliqués, ainsi que les freins à l'appropriation de l'enjeu par la communauté académique dans son ensemble. Nous présenterons alors des orientations et des pistes d'actions susceptibles d'accroître l'adoption de nouvelles pratiques au plus près des personnels.

Antoine BLANCHARD est ingénieur et titulaire d'un Master en études sociales des sciences et techniques. Il est consultant pour la société coopérative Dataactivist où il accompagne des institutions scientifiques vers la science ouverte et l'ouverture des données.

Romain PIERRONNET est docteur en sciences de gestion et chercheur associé à l'Institut de Recherche en Gestion (Université Gustave Eiffel – Université Paris-Est Créteil). Il est également consultant pour le cabinet Adoc Mètis.

Mots-Clés : Données, management, recherche, université, compétences, métiers

Les mails engageants à l'université : aspects juridique, informatique et organisationnel

Résumé : La messagerie électronique, encore marginale à la fin du 20^e, s'est imposée dans les années 2010, à l'université comme ailleurs, comme vecteur de communication entre deux personnes ou entre un émetteur et un groupe de destinataires (directeur vers collaborateurs, enseignant vers étudiants, membre d'un groupe vers l'ensemble du groupe). Parallèlement à "l'épanouissement" des réseaux sociaux pour diffuser et commenter l'actualité, les idées ou les projets, le mail reste la forme principale pour les échanges professionnels interpersonnels et, au sein de l'institution, comme le successeur du courrier classique, administratif ou pédagogique.

Mais force est de constater que l'usage du mail est davantage le fruit de tâtonnements plus ou moins heureux que l'application d'une méthode pensée et accompagnée par une quelconque autorité. Ceci serait sans conséquence si le contenu des messages était futile mais ce n'est pas le cas. Une partie des messages échangés entre les acteurs de l'université engageant l'émetteur au plan administratif ou pédagogique, et une partie (parfois les mêmes) relèvent de la réglementation européenne sur la protection des données (RGPD). Dès lors, pour minimiser les risques d'incidents ou de non-conformité de cette masse de messages échangés, il importe de définir et de mettre en œuvre des règles de conservation, destruction, sécurité des messages.

Cette communication vise à dresser un tableau des enjeux et des solutions pour une gestion optimale des mails engageants à l'université.

Marie-Anne CHABIN est *Professeure associée, à l'Université Paris 8, UFR Sciences et Technologies du Numérique, Master GSI. Consultante indépendante depuis une vingtaine d'années, elle a créé la méthode ArcategTM pour structurer la gouvernance des données à risque et des documents engageants.*

Mots-Clés : messagerie, mail engageant, courrier, risque, RGPD

Comment s'applique le RGPD dans un ensemble de données à caractère personnel et non personnel ?

Résumé : L'entrée en application du règlement général sur la protection des données (RGPD) le 25 mai 2018 a causé un séisme chez les acteurs publics et privés anxieux à l'idée de revoir leurs pratiques de protection des données à caractère personnel, particulièrement chez les établissements d'enseignement supérieurs et de recherche qui peinent encore à attraper le train en marche.

Pour parer à ce mal-être, le règlement (UE) 2018/1807 du Parlement européen et du Conseil du 14 novembre 2018 établissant un cadre applicable au libre flux des données à caractère non personnel dans l'Union européenne a été voté quelque mois après le RGPD. Il délimite le champ d'application du RGPD par l'apport de la définition de la donnée à caractère non personnel qui continue d'exister aux cotés des données à caractère personnel.

Pour autant, la définition ne vise que des catégories de données précises. De plus, le règlement instaure une supériorité de la donnée à caractère personnel sur la donnée à caractère non personnel dans les ensembles de données mixtes inextricablement liées pour les faire basculer dans le périmètre du RGPD. D'une base de données de santé anonymisées, à la réalisation d'un algorithme d'apprentissage, les ensembles de données mixtes sont légions dans la recherche. Les opposants au RGPD seront déçus d'apprendre qu'une donnée anonymisée utilisée aux côtés de données à caractère personnel peut former un ensemble de données mixtes qui va replacer la donnée anonymisée dans le champ du RGPD malgré sa nature non identifiante.

Les partisans du RGPD seront ravis de savoir que les apports du règlement ne devraient pas bousculer les actions de mise en conformité débutées par les établissements d'enseignement supérieur de de recherche. La donnée à caractère non personnel relève de l'exception et de ce fait les programmes de recherche autour du big data et de l'intelligence artificielle devront être menés dans le respect du RGPD.

Kin MATCHEU-MADJEU est la déléguée à la protection des données d'Arts et Métiers Sciences et Technologies (connu également sous le nom d'« Ecole Nationale Supérieure d'Arts et Métiers » (ENSAM). Elle est également membre du réseau SupDpo.

Mots-Clés : RGPD, DPO, CNIL

Session 2 :

La nécessaire mutualisation

Président de séance : Cyprien HENRY, *Chef de la Mission des archives et du patrimoine culturel, MESRI*

Orientations stratégiques SI de l'Amue : zoom sur un futur dispositif co-construit avec la communauté ESR pour la réutilisation des données de gestion et des autres informations de l'établissement

Résumé : Depuis près de 30 ans, l'Amue propose aux établissements de l'ESR des solutions SI, des formations et des dispositifs d'accompagnement pour améliorer la qualité de leur gestion.

En qualité d'acteur de mutualisation, l'agence a lancé, en 2019, la co-construction, avec la communauté, d'un dispositif pour l'archivage intermédiaire des données numériques produites par les applications de son offre. Ce projet, cadré en partenariat avec la section Aurore (Association des archivistes français), le CINES (établissement public agréé par le SIAF et unique opérateur sous tutelle du MESRI dont les missions statutaires comportent l'archivage numérique pérenne), le réseau SupDPO, la MipNes et l'Association des DSI vise à doter prochainement les établissements :

- + d'un outillage pour l'extraction et l'archivage pérenne (dans un système d'archivage électronique agréé) des données et documents issus de logiciels de gestion ;
- + et d'actions d'accompagnement : sensibilisation des gouvernances des établissements à la problématique de l'archivage numérique, dispositif d'accompagnement à la mise en œuvre des solutions développées par le projet, formation complémentaire en établissements.

Ce dispositif à l'ambition multiple a été conçu pour être utilisé au-delà des produits de l'Amue et du périmètre des données de gestion.

De par son adéquation avec le cadre stratégique commun de modernisation des archives 2020-2024, ce projet est soutenu par le service interministériel des archives de France.

*Ancienne responsable du département « archivage pérenne » du CINES, **Marion MASSOL** est ingénieure de recherche et PMO à l'Amue au sein du département « stratégie et programmation SI » ; elle est notamment en charge du schéma directeur SI de l'agence.*

Mots-Clés : système d'information, ESR, mutualisation, archivage pérenne, schéma directeur SI

Mettre en œuvre l'archivage électronique en université : le choix de la mutualisation pour les universités bordelaises et de Pau et des pays de l'Adour

Résumé : L'université de Pau et des pays de l'Adour, l'université de Bordeaux, l'université Bordeaux Montaigne, Bordeaux INP, Sciences Po Bordeaux et Bordeaux Sciences Agro ont finalisé en 2018 une étude sur les chantiers à conduire collectivement sur la dématérialisation et l'archivage électronique. L'archivage électronique étant un préalable notamment à l'adoption de la signature électronique, les établissements aquitains ont examiné alors les différents systèmes d'archivage électronique du marché et ont retenu la solution VITAM. D'un commun accord, il a été décidé d'expérimenter le SAE VITAM en s'appuyant sur l'expertise et les moyens techniques proposés par le CINES. Le périmètre du POC a porté sur le versement dans VITAM des pièces issues des séances du conseil d'administration.

Le CINES a accompagné ces travaux sur le plan fonctionnel, autour de la construction du modèle de données en SEDA 2.1 et la rédaction du profil d'archivage, mais également, sur le plan technique, en concourant au développement du connecteur de génération des paquets d'archives.

Ce projet s'est étalé sur une période de 11 mois et a permis pleinement de démontrer la faisabilité de cet archivage. Le passage en mode production de versements de flux vers l'instance VITAM hébergée au CINES est en cours de préparation. Parapheur électronique, signature électronique, orchestrateur de flux, GED, SAE, le socle technologique est désormais constitué pour entrer pleinement et de manière sécurisée dans l'ère du numérique.

***Béatrice Castetbon** est chargée de mission Dématérialisation et archivage électronique et responsable du pôle Système d'Information à la Direction du numérique de l'Université de Pau et des pays de l'Adour. Elle est également coordinatrice des actions menées collectivement en matière de Dématérialisation et d'archivage électronique pour les établissements d'Aquitaine ; l'université de Pau et des pays de l'Adour étant porteuse du projet. Elle est enfin correspondante Informatique et Libertés de 2008 à 2018 et membre du groupe des animateurs du réseau national SUPCIL durant 6 années.*

***Lorène Béchard** est archiviste experte en archivage numérique au Centre Informatique National de l'Enseignement Supérieur (CINES). Membre de la CN171 de l'AFNOR, de la Commission Archives Electroniques de l'AAF et du comité de pilotage du SEDA, elle est aussi co-pilote de la cellule nationale de veille sur les formats.*

Mots-Clés : VITAM, SEDA, SAE, archivage pérenne

Session 3 :

Analyser, gérer et diffuser les données de recherche

Présidente de séance : Marie-Laure BACHELERIE-GOUVERNEUR, *Responsable du Pôle national de conservation des données et documents, Direction des affaires juridiques, CNRS*

Les trois communications de cette session offrent une vision transversale des problématiques associées aux données de recherche, de leur analyse à leur diffusion. Les aspects documentaires liés à la gestion et la diffusion des données de recherche s'articulent ici avec la réponse aux besoins des usagers, les choix techniques et la prise en compte du cadre réglementaire.

Cycle de vie des données d'enquêtes à l'Institut national d'études démographiques (Ined)

Résumé : L'Institut national d'études démographiques (Ined), établissement public à caractère scientifique et technologique, a pour mission de mener des travaux de recherche ayant pour objet l'étude des populations, notamment au moyen d'enquêtes. Celles-ci, en particulier les enquêtes quantitatives, génèrent une quantité importante de données et nécessitent une gestion tout au long de leur cycle de vie. Bien que chaque enquête ait des particularités (selon la nature des données collectées, le type de collecte, la population enquêtée, etc.), il est possible d'identifier les grandes étapes communes ainsi que des catégories de fichiers qui en résultent. Entre les données brutes de collecte et les fichiers pseudonymisés mis à la disposition de la communauté scientifique (dits « Fichiers de production et de recherche »), il existe un nombre d'autres traitements intermédiaires : les fichiers produits aux différentes étapes posent des questions différentes en termes d'accès, de stockage, d'archivage notamment. L'ensemble des étapes doit également répondre aux exigences en matière de protection des données à caractère personnel (particulièrement importantes vu la nature sensible de nombreuses données collectées), d'ouverture des données scientifiques ainsi qu'aux dispositions du Code du patrimoine. Les enquêtes de l'Ined fournissent donc un éventail assez large des problèmes posés par la vie des données dans le monde numérique. Mise à disposition, conservation ou archivage impliquent en effet des approches autant scientifiques que juridiques et patrimoniales, soulevant le problème du difficile équilibre entre protection des données personnelles et libre diffusion de la recherche, conservation limitée et besoins de la recherche sur le long terme, ou encore réutilisations locales et nécessité d'une conservation pérenne.

Diplômée d'un Master 2 Archives de l'Université Paris 8 et d'un DU DPO de l'Université Paris II, Lindsay FUSFELD est archiviste à l'Ined depuis 2017. Depuis 2019, elle y exerce également la fonction de déléguée à la protection des données.

Arianna CAPORALI a obtenu un doctorat en démographie à l'Université de Rome (Italie) « La Sapienza ». Elle est ingénieure de recherche à l'Ined depuis 2011 et dirige le DataLab, une équipe du Service des enquêtes et sondages créé en 2019 et chargée de la documentation et de la mise à disposition des enquêtes et des bases de données agrégées de l'Ined.

Mots-Clés : données de recherche, RGPD, enquêtes, cycle de vie, données structurées, données statistiques

dat@UBFC, un service de gestion des données de la recherche pour la COMUE UBFC

Résumé : Proposer un service commun d'accompagnement à la communauté scientifique d'une université afin de soutenir la gestion et la valorisation des données de la recherche est aujourd'hui un enjeu majeur et fait partie des services à mutualiser. Le projet dat@UBFC propose la mise en place d'un tel service pour l'Université Bourgogne Franche-Comté, en s'appuyant naturellement sur des expériences/projets locaux tels que le projet dat@OSU de l'OSU THETA pour la création d'un portail de valorisation des données de la recherche et le projet de datacenter régional (DCR) UBFC pour le traitement et le stockage des données. Ce service permettra de guider les chercheurs dans la gestion de leurs données dès leur création, afin d'obtenir, comme cela est préconisé dans le plan national science ouverte, des données FAIR grâce à une description riche des données.

Ce projet lancé en 2019 à la demande de la COMUE bénéficie d'un financement régional, a permis le recrutement de deux ingénieurs, informaticien et documentaliste. Afin de s'assurer la collaboration des différents établissements, le CoPil du projet est constitué de représentants de chaque établissement et service (SCD et le DCR). De plus un réseau de référents a été mis en place. Leur rôle est de participer au projet sur des points techniques et de faire la sensibilisation à la science ouverte dans leur équipe.

La présentation permettra de faire le point sur le travail réalisé dans le cadre du projet dat@UBFC et montrera comment ce travail s'intègre et participe à la politique Science Ouverte de la COMUE UBFC, tout en répondant aux préconisations du Plan Science Ouverte. Ce projet institutionnel concerne l'ensemble des établissements d'UBFC (Agrosup, BSB, ENSAM, ENSMM, uB, UFC, UTBM) et par conséquent l'ensemble de leurs disciplines.

Sylvie DAMY, MCF en Informatique, Porteuse du projet dat@UBFC, Co-Responsable de l'axe Science Ouverte du laboratoire Chrono-environnement, Université de Bourgogne Franche-Comté, UMR 6249, Laboratoire Chrono-environnement, UFR Sciences et Techniques, Besançon

Rachel PROST, Documentaliste du projet dat@UBFC, OSU THETA, Professeure documentaliste, Université de Franche-Comté, Académie de Besançon

L'équipe "données de la recherche" de l'OSU THETA a lancé en 2019, à la demande de la COMUE UBFC, le projet dat@UBFC qui a pour but la création d'un service de gestion des données de recherche pour l'ensemble de la COMUE et qui fait suite au projet dat@OSU.

Auteurs : Sylvie DAMY (Chrono-environnement, UFC), Bernard DEBRAY (UTINAM – CNRS), Raphaël MELIOR (OSU THETA – CNRS), Rachel PROST (OSU THETA, UFC), Nadir TABBOU (OSU THETA, UFC), Hélène TISSERAND (OSU THETA – CNRS)

Mots-Clés : données de recherche, valorisation, science ouverte, mutualisation

Vers un entrepôt de données scientifiques pour l'université : retours d'expérience du projet Data Paris 8

Résumé : Cette intervention proposera un retour d'expérience sur la mise en place du projet Data Paris 8 à l'université Paris 8 dans l'objectif de dynamiser la recherche grâce à un entrepôt permettant le dépôt, la diffusion et le stockage à long terme des données collectées par les unités de recherche de Paris 8, en partenariat avec la TGIR Huma-Num. Le projet en cours de réalisation vise à moyen terme à généraliser l'usage des outils d'Huma-Num (ShareDocs et Nakala) pour l'ensemble des laboratoires en SHS de l'université dont six la première année, avec un objectif transversal de préservation des données déposées dans Nakala.

La présentation exposera le travail accompli ainsi que le processus nécessaire pour que les données soient documentées, diffusées en fonction de règles éthiques et juridiques et archivées sur le long terme. Il sera également question des solutions pratiques qu'il est possible de mettre en œuvre lorsqu'il s'agit d'affronter les problématiques liées au traitement des données et à leur mise à disposition en s'appuyant sur des exemples issus des données déposées dans le cadre du projet Data Paris 8.

Goran SEKULOVSKI est actuellement chef de projet Gestion des données de la recherche à l'Université Paris 8. Docteur en géographie (Université Paris 1 Panthéon-Sorbonne), il est chercheur associé au laboratoire Géographie-cités (CNRS/Paris 1/Paris 7/EHESS).

Johann HOLLAND est ingénieur spécialisé en stratégies d'innovation et en politiques numériques. Après huit années passées au service de l'établissement public Campus Condorcet où il dirigeait le pôle numérique, il est désormais chargé de mission auprès des maisons des sciences de l'homme (MSH) pour la TGIR Huma-Num. Il est familier des problématiques d'usages et d'accompagnement au changement, d'innovation permanente, d'ingénierie numérique et de transfert de technologies.

Mots-Clés : données de recherche, infrastructure, entrepôt de données, science ouverte

Table ronde :

Enrichir pour exploiter

Président de séance : Sébastien OLIVEAU, *Directeur, TGIR PROGEDO, USR 2006 CNRS-EHESS*

La 4^e et dernière séance de cette journée prend la forme d'une table ronde qui réunit deux acteurs majeurs de l'archivage des médias : la Bibliothèque nationale de France (BnF) et l'Institut national de l'audiovisuel (INA). Elle portera sur les conditions de conservation et de description de ces objets hybrides pour répondre aux besoins d'exploitation et de fouille de données de la recherche scientifique.

L'archivage du web à la BnF : le cadre légal, les conditions d'exploitation et l'accompagnement des chercheurs

Résumé : Cette communication se propose d'explorer les enjeux techniques, organisationnels et scientifiques de l'archivage du web tel que mis en œuvre à la BnF. Expérimenté depuis 2002, celui-ci se voit doté d'un cadre légal par la loi DADVSI de 2006. Destiné à "préserver le patrimoine culturel français" et fondé sur un principe de représentativité, le dépôt légal du web français s'effectue selon deux modalités principales et complémentaires : la collecte large une fois par an, de l'ensemble des sites du domaine français, et des collectes ciblées mises en place sur des axes disciplinaires ou des sujets d'actualité. A travers quelques exemples, nous reviendrons sur le processus de production de l'archive web, de la définition d'un périmètre de collecte à l'archivage pérenne des données collectées.

Cette communication s'attachera ce faisant à mettre en lumière les coopérations institutionnelles et scientifiques déployées à l'échelle nationale et internationale, tant pour développer et maintenir les outils qui permettent l'archivage technique du web que pour mettre en œuvre la démarche de sélection qui préside aux collectes ciblées.

Enfin, si les outils de collecte et de préservation du web ont atteint une certaine maturité, la question des outils permettant d'analyser et d'exploiter les 1,40 pétaoctets que constituent les collections et de leur valorisation restent encore largement à explorer. Nous dresserons un bref panorama des travaux de recherche ayant exploité ces collections et reviendrons sur quelques exemples de collaboration avec le milieu académique qui permettent à la BnF d'étoffer les services proposés autour des archives du web. Parmi celles-ci, on citera les collectes menées en partenariat avec des équipes de recherche dans le cadre des projets LIFRANUM et BodyCapital, ou encore le projet ResPaDon, qui associe la BnF, l'Université de Lille, le Campus Condorcet et Sciences-Po, et dont l'objectif est d'étendre et de diversifier les usages des archives du web par les chercheurs.

Dorothee BENHAMOU-SUESSER, *Chargée de collections numériques, Responsable de l'accès aux archives de l'internet et de la préservation, BnF*

Alexandre FAYE, *Chargé de collections numériques, Responsable projets de recherche et action internationale, BnF*

Bibliographie : Valérie Beaudouin, Philippe Chevallier, Lionel Maurel. Le web français de la Grande Guerre. Réseaux amateurs et institutionnels. Presses universitaires de Paris Nanterre, 2018. [hal-02103837]

Sophie Gebeil. Les mémoires de l'immigration maghrébine sur le web français de 1999 à 2014. Les Cahiers du numérique, Lavoisier, 2016, La médiation des mémoires en ligne, 12 (3). [hal-01470905]

Schafer Valérie, « Sous les pavés, l'archive ! Luttés sociales et archives du Web », Le Temps des médias, 2020/2 (n° 35), p. 121-138. DOI : 10.3917/tdm.035.0121. URL : <https://www.cairn.info/revue-le-temps-des-medias-2020-2-page-121.htm>

Mots-Clés : archives du web, dépôt légal, collecte du web, données de la recherche, patrimoine numérique

L'archive augmentée : cohérence et continuité des collections à l'INA

Résumé : La télévision s'enrichit désormais de contenus délinéarisés : les émissions possèdent ainsi leur chaîne Youtube, leur compte Twitter et leur page Facebook, autant de déclinaisons d'un même contenu poussé vers des publics différents. À l'INA, les pratiques d'archivage se sont adaptées à ces évolutions et le périmètre de l'archive web media s'est progressivement enrichi de ces « objets » qui documentent et complètent les programmes archivés de 170 chaînes de radio et télévision. Dans le cadre de sa mission de Dépôt Légal du Web initiée en 2009, l'INA collecte en continue 15 000 sites Web, 13 000 comptes et 1 200 hashtags Twitter, ainsi que près de 20 000 chaînes de vidéos ou de podcasts.

Du fait de la nature hétéroclite de ces objets, le concept d'archive augmentée se dessine et traduit à la fois l'évolution du processus d'archivage et de l'archive ainsi constituée. Cette manne est mise en consultation à des fins de recherche via des interfaces innovantes qui s'appliquent aux pratiques nouvelles de "second screen" ou "télévision sociale", ainsi qu'à la fouille et analyse de données.

L'objectif de cette présentation ne sera pas seulement de présenter les nouveaux outils de recherche d'une archive dite augmentée, mais également d'étudier les étapes de leur développement. En effet, la mise en place d'ateliers ou "labs" a permis aux chercheurs de formuler des demandes en phase avec leurs besoins et d'éprouver l'ergonomie ou la pertinence des interfaces. Outre qu'elle est une garantie de l'efficacité des outils de recherche, cette collaboration étroite favorise le renouvellement des méthodologies ainsi que le partage des savoirs.

Ces collections sont à disposition des chercheurs dans les nombreux lieux de consultation répartis dans toute la France, et l'INA est partenaire de projets de recherche dans lesquels nous apportons notre expertise dans la collecte et l'exploitation des données issues du Web.

Jérôme THIEVRE, *Docteur en informatique, Responsable R&D au Dépôt légal web, INA*

Projets de recherche en cours : BodyCapital : Individus, sociétés de marché et politiques corporelles dans une Europe audio-visuelle du 20e siècle – Projet ERC mené par l'Université de Strasbourg - <https://bodycapital.unistra.fr>

RwandaMAP2020 - International research network on the (archival, heritage and memorial) traces of the Genocide against the Tutsi in Rwanda – Université de Versailles Saint-Quentin-en-Yvelines - <https://rwandamap.hypotheses.org>

Carnet de Recherche de l'Inathèque : <https://inatheque.hypotheses.org/>

Mots-Clés : réseaux sociaux, archive augmentée, méthodologie, dépôt légal, fouille de données

Références bibliographiques

BANAT-BERGER Françoise, DUPLOUY Laurent et HUC Claude, L'archivage numérique à long terme : les débuts de la maturité ? 2009.

BOUNTOURI Lina, Archives in the Digital Age : Standards, Policies and Tools, 2017

BRÉGIER Frédéric et al., L'archivage des messageries électroniques, Preuve de concept VITAM, Version 1.1, 2013

CHABIN Marie-Anne. « L'expertise Archivage managérial/Records Management : un atout pour la mise en œuvre du RGPD », I2D - Information, données & documents, vol. 1, no. 1, 2019, pp. 78-81.

CHARTRON Ghislaine, « Open Access : origine, enjeux, actualités ». 11 fév. 2013. url : <http://www.openaccess-shs.info/wp-content/uploads/2012/12/Pr%C3%A9sentation-de-Ghislaine-Chartron-11-f%C3%A9vrier.pdf> (visité le 19/04/2018)(cf. p. 25).

LAROUSSE Nicolas, MASSOL Marion, La préservation à long terme des données de la recherche en Sciences Humaines et Sociales : un retour d'expérience. JRES 2015, Dec 2015, Montpellier, France. (hal-02151185) : <https://hal.archives-ouvertes.fr/hal-02151185/>

LEDOUX Thomas, HUMBERT Marion, VASSEUR Edouard, BAUCOM Erin, KUSSMANN Carol, et al.. Utiliser les Niveaux de Préservation Numérique comme un outil d'évaluation, traduction française. 2020. (hal-02552208) : <https://hal-bnf.archives-ouvertes.fr/hal-02552208/>

MAUREL Lionel, Données personnelles et recherche scientifique : quelle articulation dans le RGPD ? 18 juillet 2018.

MASSOU Luc, et LAVIELLE-GUTNIL Nathalie, Enseigner à l'université avec le numérique. Savoirs, ressources, médiations. De Boeck Supérieur, 2017

PINTARIC Pierre, « Sécuriser ses archives numériques », I2D - Information, données & documents, 2017/3 (Volume 54), p. 40-41. DOI : 10.3917/i2d.173.0040. URL : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2017-3-page-40.htm>

PROST Hélène, SCHÖPFEL Joachim, Les données de la recherche en SHS. Une enquête à l'Université de Lille 3. : Rapport final. [Rapport de recherche] Lille 3. 2015. <hal-01198379>

SERRES Alexandre, MALINGRE Marie-Laure, MIGNON Morgane, PIERRE Cécile, COLLET Didier, Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2 : Rapport ; Annexe 1 : Résultats de l'enquête statistique ; Annexe 2 : Croisements statistiques ; Annexe 3 : Extraits des entretiens ; Synthèse des résultats... [Rapport de recherche] Université Rennes 2. 2017, 159 p., 47 p., 114 p., 26 p., 23 p. (hal-01635186v2)

Big Data - Open Data : Quelles valeurs ? Quels enjeux ? Actes du colloque « Document numérique et société », Rabat, 2015. De Boeck Supérieur, 2015

La Gazette des archives, n°243, 2016-3. Quel accès, quel traitement pour les documents et données de l'enseignement et de la recherche ? Actes des journées d'études de la section Aurore - Archivistes des universités, rectorats, organismes de recherche et mouvements étudiants - de l'Association des archivistes français des 28 novembre 2014 et 6 novembre 2015.

Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, Stratégie nationale des infrastructures de recherche, Edition 2018 n° 2.

Modèle de référence pour un Système ouvert d'archivage d'information (OAIS) [= ISO 14 721], CCSDS, 2012 (version française) : <https://public.ccsds.org/Pubs/650x0m2%28F%29.pdf>

Pour aller plus loin

ARISTOTE est une société savante qui regroupe des organismes de recherche, des grandes écoles, des entreprises, des PME et des associations pour favoriser l'échange d'expériences et l'innovation dans le domaine des technologies numériques : <https://www.association-aristote.fr/cellule-format/>

Le groupe de travail "**Atelier Données**", actuellement composé de représentants de plusieurs réseaux s'attache à construire une réflexion sur la gestion des données porteuse d'une vision « métiers » et « réseaux » : <https://mi-gt-donnees.pages.math.unistra.fr/guide/06-archiver.html>

Le **Blues du JPEG**, Une histoire haute en couleur d'Alix Bruys, Bertrand Caron, Yannick Grandcolas et Thomas Ledoux de la Bibliothèque nationale de France (BnF) : <https://openpreservation.org/blogs/le-blues-du-jpeg/>

Cairn.info : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-1-page-62.htm>

Cat OPIDoR, wiki des services dédiés aux données de la recherche.
https://cat.opidor.fr/index.php/Cat_OPIDoR_wiki_des_services_d%C3%A9di%C3%A9s_aux_donn%C3%A9es_de_la_recherche (catalogue de services ESR)

Le **Cines** est l'opérateur pour l'archivage des données et documents numériques produits par la communauté Enseignement Supérieur et Recherche française (ESR). Il propose des solutions d'archivage numérique sur le moyen et long terme, mutualisées, économiques et personnalisable : <https://www.cines.fr/archivage/>

CNRS Le Journal. Donner du sens à la science
<https://lejournel.cnrs.fr/articles/preserver-les-donnees-de-la-recherche-a-lerc-du-big-data>

DoRANum (Données de la Recherche Apprentissage Numérique) propose un dispositif de formation à distance d'accès coordonné, intégrant différentes ressources d'auto-formation sur la thématique de la gestion et du partage des données de la recherche. L'internaute y trouvera un tutoriel sur la thématique « Stockage et archivage. Quelles données conserver à long terme et comment ? <https://doranum.fr/stockage-archivage/>

France archives : <https://francearchives.fr/fr/article/88482504>

Siaf hypothèse : <https://siaf.hypotheses.org/tag/formats>

Index de mots-clés

A

Archivage pérenne, 15
Archive augmentée, 22
Archives du web, 21

C

CNIL, 12
Collecte du web, 21
Compétences, 10
Courrier, 11
Cycle de vie, 17

D

Dépôt légal, 21, 22
Données, 10
Données de la recherche, 21
Données de recherche, 17, 18, 19
Données statistiques, 17
Données structurées, 17
DPO, 12

E

Enquêtes, 17
Entrepôt de données, 19
ESR, 14

F

Fouille de données, 22

I

Infrastructure, 19

M

Mail engageant, 11
Management, 10
Messagerie, 11
Méthodologie, 22
Métiers, 10
Mutualisation, 14, 18

P

Patrimoine numérique, 21

R

Recherche, 10
Réseaux sociaux, 22
RGPD, 11, 12, 17
Risque, 11

S

SAE, 15
Schéma directeur SI, 14
Science ouverte, 18, 19
SEDA, 15
Système d'information, 14

U

Université, 10

V

Valorisation, 18
VITAM, 15

Table des sigles et abréviations

AAF : Association des Archivistes Français.

AFNOR : Association Française de Normalisation

Amue : Agence de mutualisation des universités et établissements

BnF : Bibliothèque nationale de France

BSB : Burgundy School of Business

CINES : Centre Informatique National de l'Enseignement Supérieur

CN171 de l'AFNOR : Commission de Normalisation CN171 de l'AFNOR en charge de la norme sur l'archivage électronique

CNIL : Commission Nationale de l'Informatique et des Libertés

CNRS : Centre National de la Recherche Scientifique

COMUE : Communautés d'universités et établissements

CoPil : Comité de Pilotage

DCR : Data Center Régional

DINUM : Direction Interministérielle du Numérique

DPO : DPO, Délégué à la Protection des Données (Data Protection Officer)

DSI : Direction des Systèmes d'Information

EHESS : École des Hautes Etudes en Sciences Sociales

ENSAM : École Nationale Supérieure d'Arts et Métiers

ENSMM : École Nationale Supérieure de Mécanique et des Microtechniques

FAIR : Findable, Accessible, Interoperable, Reusable

GED : Gestion Électronique des Documents

INA : Institut National de l'Audiovisuel

LIFRANUM : Littératures FRAncophones NUMériques

MCF : Maître de conférences

MESRI : Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation

MipNes : Mission de la Pédagogie du Numérique pour l'Enseignement Supérieur

MSH : Maison des Sciences de l'Homme

OSU THETA : Observatoire des Sciences de l'Univers Terre-Homme-Environnement-Temps-Astronomie

PMO : Project Management Office

POC : Preuve de concept (Proof of Concept)

RESPADON : Réseau de Partenaires pour l'Exploration et l'Analyse de Données Numériques

RGPD : Règlement Général sur la Protection des Données

SAE : Système d'Archivage Electronique

SCD : Service Commun de la Documentation

SEDA : Standard d'Echange de Données pour l'Archivage

SHS : Sciences Humaines et Sociales

SI : Système d'Information

SIAF : Service Interministériel des Archives de France

SUPCIL : Correspondant Informatique et Libertés de l'enseignement supérieur et de recherche

SupDPO : Réseau des Délégués à la Protection des Données de l'enseignement supérieur et de recherche

TGIR : Très Grandes Infrastructures de Recherche

TGIR Huma-Num : Très Grandes Infrastructures de Recherche des humanités numériques

TGIR PROGEDO : Très Grandes Infrastructures de Recherche pour Production et Gestion des Données

uB : université de Bourgogne

UBFC : Université Bourgogne Franche-Comté

UE : Union européenne

UFC : Université de Franche-Comté

UFR : Unité de Formation et de Recherche

UTBM : Université de Technologie de Belfort-Montbéliard

UTINAM : Univers, Temps-fréquence, Interfaces, Nanostructures, Atmosphère et environnement, Molécules

VITAM : Valeurs Immatérielles Transmises aux Archives pour Mémoire

Bloc-notes

