

## RESUMES DES PROJETS PRESENTES

### Session 1

#### **DATTACC'**

DATTACC' est un dispositif complet d'accompagnement des chercheurs en physique et en chimie pour la gestion des données de recherche tenant compte des fortes spécificités disciplinaires. Le dispositif comprend :

- une plateforme "boite à outils" en ligne pour faciliter le dépôt de jeux de données, la rédaction de PGD, de « data papers », etc.
- l'accès à un outil de cahier de laboratoire électronique répondant aux besoins spécifiques des deux disciplines,
- l'accès à un programme de « tutorat-données de recherche » par les pairs,

Le dispositif est entièrement co-construit avec les représentants des deux disciplines associés à la conception du projet. Les besoins des jeunes chercheurs et doctorants sont particulièrement pris en compte. Un focus group composé de chercheurs, de techniciens et de doctorants teste et valide chacune des solutions proposées. Des laboratoires partenaires "embarquent" des professionnels de l'IST tout au long de la mise en place du dispositif afin de s'assurer que les solutions proposées restent en phase avec les pratiques de terrain.

#### **ARCHIVES PARLEMENTAIRES DE LA REVOLUTION FRANCAISE**

Le projet ARCPA vise à poursuivre la numérisation, le balisage et l'indexation de l'une des sources les plus utilisées pour l'histoire de la Révolution française et de la vie parlementaire de la France moderne, les *Archives parlementaires*.

Il s'agit d'un ensemble de 102 volumes qui restituent, journée par journée, les débats parlementaires du printemps 1789 à décembre 1794. Ces 80 000 pages sont une source de premier ordre pour les chercheurs de différentes disciplines. Lancée par le Corps législatif en 1862, l'édition de ce corpus sous forme papier est toujours en cours, sous la responsabilité scientifique de l'Institut d'histoire de la Révolution française (IHRF), au sein de l'Institut d'histoire moderne et contemporaine (IHMC), qui possède les droits sur les volumes récemment publiés. La numérisation et le traitement des volumes publiés offrent des perspectives vastes et enthousiasmantes de recherche.

Le but de ce projet est de poursuivre le travail [déjà en ligne sur Persée](#) en accélérant significativement le calendrier et en améliorant la présentation et les services associés. 15 nouveaux volumes seront traités via cet appel à projet et l'ensemble sera mis en ligne sur un portail dédié de type [Perséide](#).

#### **BIBLIOTHEQUE HISTORIQUE DE L'EDUCATION**

Le projet de Bibliothèque Historique de l'Education porte sur la numérisation, le traitement documentaire et la valorisation de collections patrimoniales en histoire de l'éducation, choisies pour leur pertinence au regard des besoins de la communauté scientifique.

Le corpus documentaire choisi est composé de trois revues :

- la *Revue pédagogique* (1878-1926)
- la *Revue internationale de l'enseignement* (1878-1939)
- le *Manuel général de l'instruction primaire* (1844-1964)

Dans un second volet, le projet élargira ses perspectives en procédant à l'enrichissement des collections numérisées et à la production d'outils d'interrogation et de traitement. Les premières réalisations entreprises au titre de la BHE sont disponibles sur une Persée dédiée: <https://education.persee.fr/>

## DIGITAL ALFIERI

Le projet *Digital Alfieri* a pour objectif l'édition numérique des archives de l'écrivain italien Vittorio Alfieri (1749-1803) afin de réunir virtuellement des manuscrits et des livres éclatés entre France et Italie à travers un site web conçu comme un collecteur de matériaux, rassemblant des documents déjà numérisés et d'autres qui seront numérisés pour la première fois. Le soutien du CollEx Persée permettra notamment de poursuivre la numérisation des livres annotés ayant appartenu à la bibliothèque personnelle de l'auteur et conservés dans des établissements français, afin d'enrichir le dossier génétique des ouvrages de l'écrivain à travers la mise en relation de ses brouillons et des livres annotés. Une édition génétique numérique des archives d'Alfieri est en cours de réalisation à l'ITEM, sur une plateforme d'édition modélisée sur OMEKA, ce qui permettra une double diffusion aisée de ces matériaux à travers la plateforme EMAN et le portail du consortium franco-italien « Fonte Gaia », qui bénéficie d'une large visibilité en France et à l'étranger.

## RESEAUX PLURIDISCIPLINAIRES ET PRATIQUES DOCUMENTAIRES : LES TRANSITIONS POLITIQUES EN AMERIQUE LATINE DANS LES ANNEES 1960/2000

La problématique des dictatures et des transitions politiques en Amérique latine dans les dernières décennies du 20<sup>ème</sup> siècle mobilise nombre de chercheurs dans une approche pluridisciplinaire, faisant se croiser américanistes, juristes, politistes et historiens. Elle fait aussi ressortir la nécessité d'un travail en réseau pour collecter, constituer et sauvegarder des archives de cette période. En lien avec les programmes de recherche [Transatlantique](#) et [Archives des juristes internationales](#), le projet se centrera plus particulièrement sur les archives orales. Celles-ci constituent un élément central des sources documentaires à produire / collecter et à transmettre dans le cadre de partenariats associant chercheurs et établissements documentaires. La consultation de ces entretiens, qui sont des archives numériques natives, soulève des enjeux techniques de coordination et peut créer entre chercheurs et bibliothécaires une dynamique de sensibilisation favorable à la collecte d'autres sources documentaires. L'objectif de ce projet est de mettre en commun de bonnes pratiques de traitement, de mise à disposition et de conservation, de la part des chercheurs et avec les bibliothèques : ateliers méthodologiques, enquête sur les pratiques documentaires des chercheurs et production d'un guide en ligne sur les archives orales. De nouveaux matériaux pour la recherche seront aussi produits et traités, permettant de formaliser et mettre en œuvre les règles définies au cours du projet.

## Session 2

### ARCHELEC4

L'objectif du projet Archelec4 est de permettre aux chercheurs et aux communautés intéressées d'accéder à de nouvelles formes d'exploitation des métadonnées et des documents du corpus de sources primaires d'Archives électorales françaises (Archelec).

Archelec est un corpus en accès ouvert comptant 40 000 documents de nature très diverse (professions de foi, bulletins de vote, tracts, affiches, etc.), couvrant les élections françaises de 1958 à 2012, disponible sur [la plateforme Internet Archive](#).

Archelec4 se focalise sur les 10 élections législatives de 1958 à 1993, soit 33 000 professions de foi concernant 18 400 candidats. Il permettra :

- de bénéficier d'un accès à la pièce aux professions de foi et aux bulletins de vote, permettant la constitution de bases de données et la recherche plein texte combinée avec un ou plusieurs critères ;
- de mener des recherches sur les documents mis en ligne grâce à de nouvelles métadonnées provenant de sources distinctes (activité professionnelle des candidats, étiquette politique notamment) ;
- d'apporter une nouvelle dimension d'analyse de contenu du corpus à travers l'évolution des logos des partis et organisations politiques pendant la période couverte ;
- d'offrir un accompagnement à l'utilisation de la plateforme, à la compréhension de la nature et du périmètre des données du corpus à travers un tutoriel, un atelier d'utilisation de la plateforme et un atelier d'utilisation d'outils d'extractions de termes (fouille de texte).

### DATALAB à la BNU

Placé au sein d'une structure plus large de type « Lab » et dédiée à l'innovation, le service du DataLab a vocation à coordonner les services numériques à la recherche développés par la Bnu en étroite collaboration avec l'Université de Strasbourg et ses autres partenaires. Il est destiné à favoriser l'usage, en particulier par les chercheurs, des collections numériques de la BNU ainsi que d'autres réservoirs de données existants. Cet équipement sera au centre de l'offre de services que la BNU prévoit de déployer en soutien à la science numérique, et plus particulièrement dans le domaine des sciences humaines et sociales, dans le cadre de son projet d'établissement 2018-2022.

Ce nouveau service constitue un outil d'appui pour les projets menés au niveau national inscrits dans le cadre de CollEx. Plusieurs projets pilotes permettront d'expérimenter des outils développés par les opérateurs dans le domaine des humanités numériques, en partenariat avec les chercheurs, et d'en assurer la diffusion dans le cadre des réseaux CollEx.

## **PLATEFORME DES ETUDES NORDIQUES**

Le Réseau « CollEx Etudes nordiques » est engagé dans un projet de plateforme documentaire en ligne consacrée à la recherche :

- dans les domaines des pays de langues scandinaves et finno-ougriennes,
- sur l'Europe du Nord et sur les pays baltes,
- sur les concepts de Nord et de boréalisme.

Cet outil prendra la forme d'un site web dédié au référencement des ressources utiles aux chercheurs francophones dans ces domaines, site qui leur permettra également de promouvoir leurs travaux.

Il s'agit d'un outil complet d'information bibliographique, avec, dans la mesure du possible, accès au texte intégral des documents. L'animation de la plateforme sera assurée de façon collaborative par les membres du réseau.

Cette plateforme constituera également une vitrine de l'activité des chercheurs francophones sur les études nordiques et un outil de coopération nationale et internationale dans ce domaine, avec un rapprochement vers les réseaux documentaires et de recherche similaires. Le financement demandé dans le cadre de cet appel à projet sera consacré à la constitution d'un site-test permettant de valider les orientations initiales par le public-cible des chercheurs en études nordiques, mais permettra également de planifier les phases suivantes tout en initiant d'éventuelles réorientations ou développements supplémentaires du projet.

## **NUMERISER DES DONNEES ASTRONOMIQUES CONTEMPORAINES : UN TRAVAIL CROISE DE CHERCHEUR, BIBLIOTHECAIRE, ARCHIVISTE ET INFORMATICIEN**

Depuis 1970, des observations des émissions radio basse fréquence du Soleil et de Jupiter sont quotidiennement réalisées avec des radiotélescopes de la station de radioastronomie de Nançay. Avant 1990, les données étaient enregistrées de manière analogique sur des films de 35 mm à défilement continu (environ 1500 films, soit 45km de pellicule).

Ces films sont d'une grande valeur scientifique et patrimoniale. La base de données radio basse fréquence de Nançay est unique au monde par sa taille, qui couvre aujourd'hui 47 ans, et a joué le rôle important de « support sol » à de nombreuses missions spatiales internationales d'étude du Soleil et de Jupiter. La numérisation des archives analogiques de la période 1970-1990 est indispensable pour réhabiliter l'ensemble de la base de données et permettre aux équipes du LESIA et de l'USN d'en faciliter l'exploitation scientifique sur toute sa durée, d'en assurer la diffusion (mise en ligne à destination de la communauté scientifique) et de les conserver (archivage électronique et physique).

## Session 3

### **DATAPOC.MNHN.FR**

Le projet [datapoc.mnhn.fr](http://datapoc.mnhn.fr) vise à évaluer la faisabilité et l'opportunité de construire un référentiel « personnes » commun à l'ensemble des services et types de collections et de données produits et gérés au Muséum. Il s'agit à la fois de consolider les données et les bases de données existantes, mais aussi de permettre aux chercheurs de croiser, lier et exploiter des données qu'il leur est difficile d'apparier aujourd'hui compte tenu de la dispersion des applications. Ce référentiel devra respecter des standards d'interopérabilité afin de pouvoir être réutilisé par les grands réservoirs de données nationaux et internationaux, s'y insérer et être partagé avec d'autres institutions et communautés dans une logique de données ouvertes.

Le projet prend la forme d'une preuve de concept établie à partir d'un corpus limité de noms de personnes (environ 500). Il consiste à :

- tester la performance de technologies et de méthodes de traitement et de rapprochement de données à partir des données disponibles,
- réaliser un prototype d'interface orienté chercheur pour la visualisation humaine et la réutilisation par des machines des résultats et des données exposées,
- émettre des préconisations pour un passage à l'échelle et une industrialisation (à la fois en volume et en type de données).

### **MISTARA : améliorer la qualité des métadonnées relatives aux personnes physiques dans les langues à écriture arabe**

Les métadonnées décrivant des ressources en alphabet arabe (langues arabe, persane, ourdou et turque-ottomane principalement) présentent de nombreuses anomalies liées à la complexité du système onomastique, aux divergences phonétiques ou orthographiques entre langues ou à des erreurs de codage des caractères.

Le projet vise l'amélioration des entités décrivant des personnes physique. Il entraînera les outils d'alignement automatique élaborés dans le cadre de la transition bibliographique pour traiter ce type de métadonnées et des écritures non-latines. Ces métadonnées seront enrichies en assurant l'alignement des référentiels catalographiques (Idref au premier chef) avec des référentiels onomastiques issus de la recherche : base *AlKindi* (IDEO), *Onomasticon Arabicum* (IRHT).

### **CAIRMOD : Géo-visualisation de contenus de la Perséide Athar : le cas du Caire moderne**

Le projet CAIRMOD vise à structurer finement des albums iconographiques, déjà numérisés par Gallica, documentant la ville moderne du Caire pour les diffuser sur la Perséide Athar. Cette structuration, enrichie par un référentiel des toponymes (avec OpenTheso), sera effectuée grâce à la plateforme technique de Persée, et alimentera un module de géo-visualisation. L'outil mis en œuvre permettra de connecter les données disponibles sur la Perséide à un système d'information géographique (SIG) développé sur un logiciel libre (QGis) et accessible en ligne. La spatialisation cartographique intègrera une dimension historique pour rendre compte du renouvellement urbain dans le temps. Le cas du Caire moderne où cette dimension est particulièrement importante est à ce titre exemplaire.

L'expérimentation se fonde sur un corpus visuel concernant Le Caire, mais l'outil a vocation à être déployé, à terme, sur un volume plus large de documents visuels ou sur d'autres lieux. Toutes les données seront structurées de manière à être réutilisables.