HAL open science

# PANDOR

Laurent Gautier, Céline Alazard, Agnès Viola, Hédi Maazaoui, Arnaud Millereux

**To cite this version:**

Laurent Gautier, Céline Alazard, Agnès Viola, Hédi Maazaoui, Arnaud Millereux. PANDOR. La Lettre de l'InSHS, 2015, pp.4-6. hal-01368198

**HAL Id: hal-01368198**

**https://hal.science/hal-01368198v1**

Submitted on 14 Jun 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

## PANDOR : Portail Archives Numériques et Données de la Recherche

*Le portail PANDOR est un puissant outil d'interrogation et de valorisation des ressources numériques gérées par la Maison des Sciences de l'Homme (MSH) de Dijon, la faisant entrer dans le vaste mouvement des Humanités Numériques. Lancé avec succès le 6 octobre 2014, PANDOR permet non seulement aux chercheurs, aux collectivités et institutions, ainsi qu'aux entreprises, mais aussi au grand public de localiser et d'accéder à un ensemble de données, le plus souvent inédites, issues de programmes de recherche pluridisciplinaires. Il couvre tous les champs thématiques des Sciences Humaines et Sociales représentés à la MSH de Dijon et intègre tous les types de données (textes, images fixes et mobiles, son, multimédia), qu'elles soient natives ou le fruit d'une numérisation. Répondant aux standards internationaux en matière de traitement de données, PANDOR permet, grâce à une fine description des contenus, de repérer des documents difficiles d'accès. Témoin et acteur de la recherche développée à la MSH de Dijon, PANDOR inclut aussi des archives et des productions de chercheurs.*

### Histoire de PANDOR : une nécessité d'unifier, simplifier et faciliter l'accès aux ressources archivistiques et documentaires

Le recensement des pratiques en matière de production de ressources documentaires et le recensement des outils utilisés par la MSH de Dijon, les laboratoires SHS du Grand Campus et les partenaires de la MSH ont permis de prendre conscience de la nécessité d'unifier, de simplifier et de faciliter l'accès aux données. En effet, les premières réflexions, amorcées par la MSH de Dijon, ont révélé une profusion d'outils de diffusion qui engendrait des problèmes de visibilité, de maintenance, de sécurité des serveurs et de coût ; une trop grande hétérogénéité des métadonnées et une absence totale d'interopérabilité nuisaient en outre à l'accès et à la valorisation des données disponibles.

Les méthodes et préconisations techniques mises en œuvre par les grands acteurs du numérique en France (le Très Grand Équipement ADONIS — aujourd'hui Très Grande Infrastructure de Recherche Huma-Num —, le Service Interministériel des Archives de France, la Bibliothèque Nationale de France, le Centre Informatique National de l'Enseignement Supérieur) se sont imposées alors dans la création d'un nouvel outil de diffusion de ressources archivistiques et documentaires facile à utiliser, centralisant les données disponibles au sein d'un portail unique : PANDOR.

PANDOR a été développé par la société bordelaise AJLSM avec le soutien financier du Fonds Européen de Développement Régional (FEDER) et de la Région Bourgogne dans le cadre du Plan d'Action Régional et de l'Innovation (PARI).

Par ailleurs, le portail participe à l'alimentation du portail européen *Europeana* dans le cadre du programme HOPE (*Heritage Of People's Europe*). Ce programme vise, quant à lui, à rassembler des collections numériques d'histoire sociale et ouvrière depuis la fin du XVIIIe siècle jusqu'au début du XXIe siècle. La MSH de Dijon est ainsi l'une des 13 institutions des 10 pays participants ayant

rendu accessible une partie de ses collections numériques *via* le portail multilingue *Europeana*.

## Des fonds variés, reflet des thématiques de recherche portées par la MSH de Dijon

Reflétant les thématiques de recherche scientifique portées par l'unité, les fonds d'archives ou d'imprimés actuellement publiés sur le portail mettent en avant des documents constitués et/ou (ré)utilisés dans le cadre de programmes de recherche. Ces fonds d'archives scientifiques de chercheurs en sciences humaines et sociales sont décrits, indexés et numérisés, rendant accessibles des données primaires qui pourront être à nouveau interrogées par d'autres chercheurs.

A titre d'exemples, les archives des sociologues Marcel Jollivet, Hugues Lamarche, Henri Mendras ou de l'historien Gaston Roupnel ont été déposées et traitées à la MSH de Dijon et les inventaires sont publiés sur PANDOR.

Le portail donne également accès à des documents d'archives produits par l'entreprise Schneider et Compagnie du Creusot (documents conservés à l'Académie François Bourdon) au moment de la Grande Guerre : on y découvre non seulement la participation de cette grande entreprise à l'effort de guerre avec la fabrication massive d'armement, mais aussi la vie des salariés.

Outre des documents d'archives, PANDOR propose des collections d'imprimés telles que les revues *Cahier d'Histoire* et *Société française*, dont les numéros papier ont été numérisés et catalogués.

La MSH de Dijon a également collaboré avec le Théâtre Dijon Bourgogne au traitement et à la mise en ligne d'un fonds d'affiches de spectacles pour la période 1955-2007.

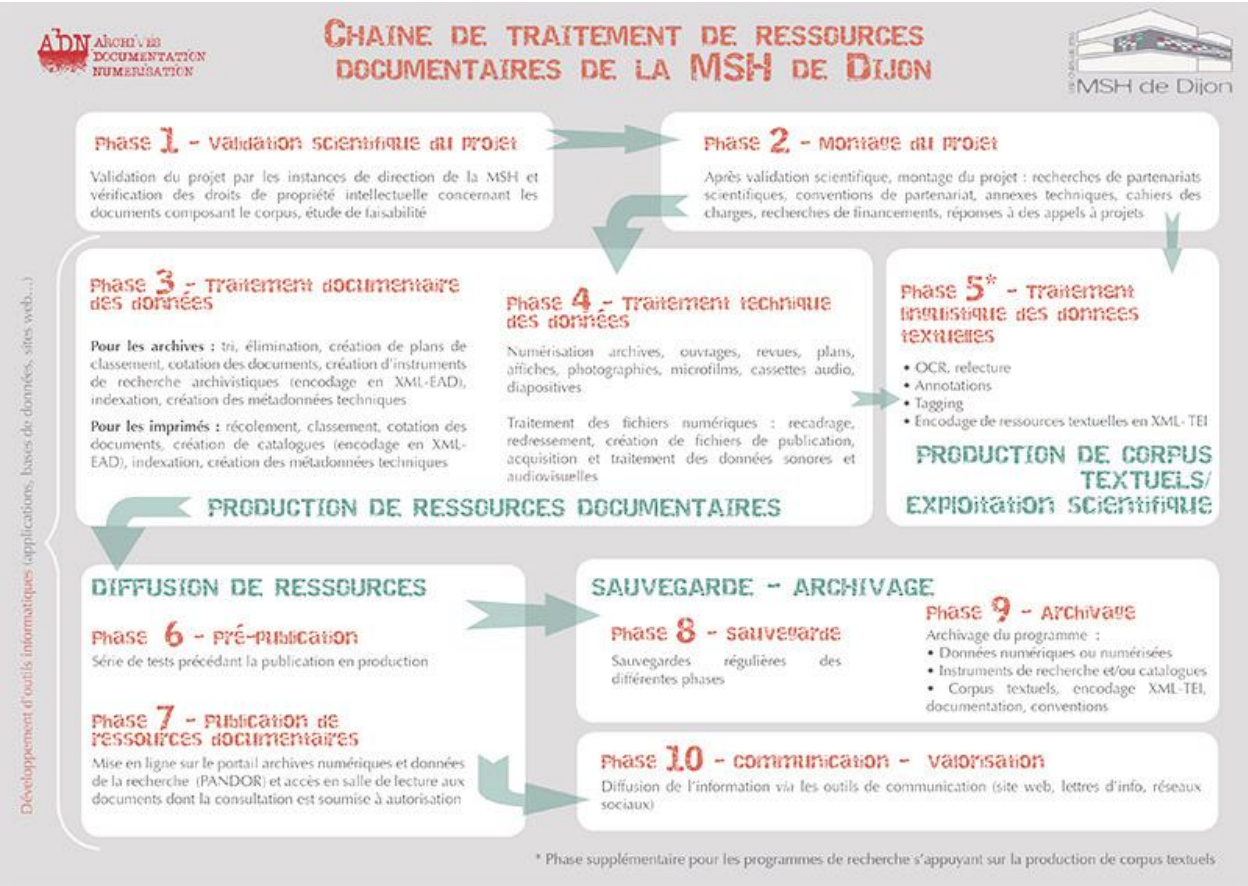
Cette liste, non exhaustive, illustre les différentes thématiques scientifiques et typologies documentaires actuellement disponibles sur PANDOR.

Si tous les instruments de recherche électroniques correspondant à la description des ressources sont publiés sur le portail *via* Internet, certains des documents numérisés sont accessibles uniquement en salle de lecture. Cette restriction concerne les archives contenant des données personnelles et nominatives (c'est le cas des enquêtes sociologiques, par exemple) et pour lesquelles la consultation ne peut se faire que dans les locaux de la MSH de Dijon, à des fins de recherche scientifique.

### Des modes de recherche variés

▶ Moteur de recherche simple : recherche plein texte dans les notices, les ressources associées ainsi que les contenus importés d'autres services web ;
▶ Recherche thématique dans les notices ;
▶ Recherche par géolocalisation ;
▶ Recherche spécifique à l'aide de formulaires spécialement conçus et adaptés aux termes et mots-clés d'un corpus ;
▶ Possibilité de recherche de termes (dans les images) directement depuis la console de visualisation d'images ;
▶ Nombreux outils d'aide à la recherche (tables d'index).

Une aide en ligne fournit des conseils et astuces pour formuler ses recherches (utilisation des opérateurs booléens notamment).



## Des types de contenu et des fonctionnalités

▶ PANDOR permet la mise en ligne d'instruments de recherche et de catalogues créés par les archivistes et documentalistes ainsi que la mise en ligne d'objets numériques numérisés par les techniciens (ouvrages, périodiques, affiches…) et nativement numériques (sons, vidéos…)

▶ PANDOR propose de nombreux outils de recherche dans les notices, les ressources numériques ou les données moissonnées. Il propose notamment la recherche en texte intégral dans les ressources numériques de type documents imprimés (ouvrages, périodiques, etc.), ce qui est rendu possible par la production en amont et la publication de fichiers XML METS/ALTO pour ce type de contenus.

## Quelques données techniques : la technologie du portail

PANDOR est propulsé par Pleade, logiciel *Open Source* développé et maintenu par la société bordelaise AJLSM. L'application peut être installée et utilisée sur tous les systèmes d'exploitation du marché compatible JAVA (Linux, Windows, Mac) puisqu'elle s'appuie sur ce langage de programmation. En tant qu'application web, cette dernière nécessite l'utilisation d'un serveur de type Apache Tomcat. Les données peuvent également être stockées dans une base de données MySQL pour étendre les capacités de l'application. Pleade s'appuie sur les standards du web pour le traitement et la diffusion des données : XML, HTML, CSS et JS.

L'échange de données est assuré par la présence d'un entrepôt de données utilisant le protocole OAI-PMH, un format d'échange de données au format XML. Ainsi, PANDOR peut, d'une part, mettre ses données à disposition de moteurs comme Isidore ou *Europeana* ou, d'autre part, moissonner des données présentes sur d'autres serveurs compatibles avec ce protocole d'échange.

## L'avenir de PANDOR

La plateforme ADN (Archives Documentation Numérisation) de la MSH de Dijon travaille actuellement à plusieurs projets qui seront l'occasion :

▶ d'étendre les compétences de l'équipe ;

▶ de poursuivre sa participation aux échanges nationaux visant à la définition de bonnes pratiques dans le domaine des humanités numériques ;

▶ d'imaginer un transfert du domaine de la recherche au monde socio-économique.

Pour exemple, le domaine de la vigne et du vin, déjà visible sur PANDOR puisqu'on y trouve les archives du secteur Vins et Eaux-de-vie de l'Institut National des Appellations d'Origine (INAO), sera renforcé dans les mois à venir. Lauréate de l'appel à projets BSN5 en 2014, la MSH de Dijon va œuvrer dès le début de l'année 2015 à la mise en ligne du *Bulletin de l'Organisation Internationale de la Vigne et du Vin*. Ce nouveau corpus fera l'objet d'une exploitation inédite : aux notes de dégustation en français, qui constituent une rubrique récurrente de la revue, seront appliqués les prérequis de l'analyse de sentiment pour parvenir à l'extraction automatique du profil, positif ou négatif, des évaluations de vins. Nous étudions pour ce faire l'opportunité d'appliquer la TEI (*Text Encoding Initiative*) au corpus et projetons de nous rapprocher des institutions pilotes dans le domaine : Consortium international TEI, Consortium Ecrits, MSH Val de Loire… Nous développerons également des compétences dans le domaine de l'extraction automatique de données linguistiques, par l'étiquetage morphosyntaxique et sémantique d'un certain nombre des corpus existants. Ces compétences seront diffusées à la communauté scientifique, et notamment aux doctorants, par le biais de formations. Le but est la création de prototypes qui pourront être présentés et adoptés après adaptation par les filières professionnelles concernées.

L'objectif pour la plateforme ADN est de passer de données traitées comme corpus statiques à des données dynamiques et intelligentes dans le sens de la fouille de données, un incontournable pour qui veut inscrire son travail dans le mouvement des humanités numériques.

### L'équipe du projet PANDOR

Les contenus mis en ligne sur PANDOR sont produits par l'équipe de la plateforme Archives Documentation Numérisation (ADN) de la MSH.

▶ Laurent GAUTIER, Professeur de linguistique appliquée, responsable de la plateforme ADN et du projet PANDOR
▶ Céline ALAZARD, archiviste
▶ Agnès VIOLA, documentaliste
▶ Hédi MAAZAOUI, responsable de fabrication
▶ Arnaud MILLEREUX, développeur informatique
▶ Personnels contractuels intervenant aux différentes étapes de production des données

contact&info
▶ Laurent Gautier,
Laurent.gautier@u-bourgogne.fr
pandor@u-bourgogne.fr
▶ Pour en savoir plus
https://pandor.u-bourgogne.fr