

Le Secteur Archives de la recherche - Phonothèque de la Médiathèque SHS de la Maison méditerranéenne des sciences de l'homme s'appuie sur les principes FAIR pour des données faciles à trouver, accessibles, interopérables et réutilisables.

Parcourir les terrains des chercheur·se·s en sciences humaines et sociales à travers leurs archives

Le Secteur Archives de la recherche – Phonothèque de la Médiathèque SHS de la Maison méditerranéenne des sciences de l'homme (MMSH)¹

réunit, conserve et met à disposition de tous les publics 600 mètres linéaires d'archives, plus de 100 000 photographies et 8 000 heures d'enregistrements sonores et audiovisuels issus de la recherche en sciences humaines et sociales. Carnets de terrain, manuscrits, dessins, documentation papier, plaques de verre, diapositives, tirages, films, cassettes, bandes, données du web ou fichiers numériques natifs : autant de documents qui témoignent du travail des chercheur·se·s et rendent compte de leurs terrains. Pour les archiver, l'équipe du secteur Archives de la recherche s'appuie sur les outils proposés par les infrastructures académiques nationales : l'Abes (Agence bibliographique de l'enseignement supérieur et de la recherche), l'IR* Huma-Num, le CCSD (Centre pour la communication scientifique directe), le Cines (Centre informatique national de l'enseignement supérieur) et OpenEdition Center. Comment interagissent entre eux et se complètent ces outils ? Pour notre méthode d'archivage, nous avons choisi de suivre les 4 lettres des principes FAIR sur lesquels nous nous appuyons pour le traitement et l'organisation de nos données et métadonnées : faciles à trouver, accessibles, interopérables et réutilisables.

FACILES À TROUVER

La plateforme Calames de l'Abes est au cœur de nos outils². Nous décrivons nos instruments de recherche dans le standard international des archives, l'*Encoded Archival Description* (EAD), et Calames fournit un identifiant stable. Lorsque les questions juridiques et éthiques pour la diffusion des données sont réglées, nous utilisons des plateformes spécifiques pour les disséminer. Dès 2011, nous avons placé nos photographies sur la plateforme mediHAL³ développée par le CCSD, tandis qu'à partir de 2015 les carnets de terrain, les archives papier numérisées, les fichiers texte ou tableurs nativement numériques ont été déposés sur Nakala⁴ géré par Huma-Num. Sur ces deux plateformes,

les données reçoivent un identifiant unique (soit un numéro HAL, soit un DOI⁵) que nous précisons dans les métadonnées et les dispositifs de mediHAL ou de Nakala permettent d'indexer les données et de les retrouver.

ACCESSIBLES

Dès 2007, alors que la numérisation des archives s'amplifiait, nécessitant des espaces de stockage toujours plus importants, nous avons d'abord bénéficié des tous premiers services du TGE Adonis⁶, devenu TGIR Huma-Num⁷ puis en 2022 l'IR Huma-Num. Dès 2011, le secteur archives de la recherche de la MMSH a également été impliqué dans le consortium « Archives des ethnologues »⁸ à travers lequel nous avons pu échanger sur nos pratiques archivistiques dans les laboratoires de recherche, nous former et comprendre comment faciliter l'accessibilité à nos données dans un monde numérique aux infrastructures en train de se construire. Aujourd'hui, nos données sont préservées à la fois sur des serveurs sur site supervisés par le service informatique de la MMSH, et sur l'espace sécurisé distant nommé « Huma-Num Box » géré par l'IR* en miroir (Paris et Villeurbanne). Enfin, engagées dans la problématique de l'accessibilité même lorsque les données ne peuvent pas être ouvertes à toutes et tous, nous travaillons à leur partage en nous appuyant sur des conventions de collaborations scientifiques. Nos métadonnées sont accessibles à travers Calames, l'accès ou l'absence d'accès aux données sont toujours documentés et nous partageons les fichiers avec celles et ceux qui bénéficient de droits spécifiques. Ainsi cette année, une convention avec le LASDEL (Laboratoire de recherche en sciences sociales)⁹ à Niamey, a été signée afin de préciser les modalités de partage des données de l'anthropologue Jean-Pierre Olivier de Sardan¹⁰ avec des chercheurs du terrain nigérien ayant participé à ces enquêtes. Un export au format CSV¹¹ de nos métadonnées à partir de Calames est partagé avec le laboratoire et l'accès distant protégé se fait à travers un espace Sharedocs, géré par l'IR* Huma-Num. Un autre exemple de facilitation de l'accessibilité

[1] En 1997, lors de la création de la Médiathèque, celle-ci s'organisait en trois secteurs, ceux de la bibliothèque, de la phonothèque et de l'iconothèque. Un nouveau projet a vu le jour sous la direction de Jean-Christophe Peyssard en 2020 qui a donné naissance à ce nouveau secteur des archives de la recherche – Phonothèque, en septembre 2021. La phonothèque a commencé à cataloguer ses fonds dans Calames dès 2013.

[2] www.calames.abes.fr/pub/ms/FileId-1361

[3] <https://media.hal.science/MMSHIMAGES>

[4] Nakala est une plateforme développée et maintenue par l'IR* Huma-Num.

[5] Digital Object Identifier <https://www.doi.org>

[6] Le très grand équipement pour l'Accès unifié aux Données et documents Numériques des Sciences humaines et sociales a vu le jour en 2009 (<https://leo.hypotheses.org/2456>) mais dès 2007 nous avons été soutenus pour sauvegarder nos données sur les serveurs distants dédiés aux SHS par le CNRS.

[7] Huma-Num est issue de la fusion du très grand équipement (TGE) Adonis et de Corpus-IR, créé en 2011.

[8] Ce consortium créé au sein de Corpus IR en 2011 a intégré ensuite ceux portés par Huma-Num, il est désormais clos. Ses activités sont présentées sur le carnet Ethnologia : <https://ethnologia.hypotheses.org>

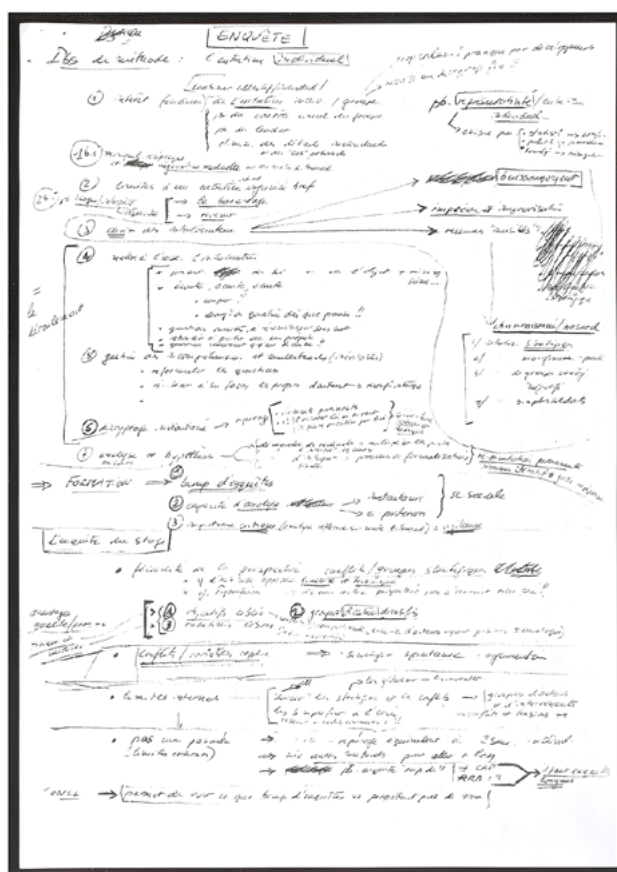
est de proposer l'enrichissement des données numérisées en mode image vers le mode texte. Ainsi une partie des carnets de terrains d'anthropologues de la MMSH (Marceau Gast, Annie-Hélène Dufour) sont accessibles sur la plateforme *Transcrire*¹² où les internautes fournissent une transcription textuelle collaborative.

INTEROPÉRABLES

L'interopérabilité des métadonnées est assurée par le format EAD de Calames et l'indexation par la plateforme IdRef. Les données elles-mêmes, placées dans Nakala et mediHAL sont documentées au format *Dublin Core* qualifié (Nakala) ou *XMP Extensible Metadata Platform* (mediHAL) et rattachées dans la zone <dao> de l'EAD de Calames par leur identifiant (DOI ou n° HAL). Métadonnées et données sont moissonnées par différents moteurs de recherche thématiques nationaux comme celui de FranceArchives¹³, d'Isidore¹⁴, une partie des photographies est moissonnée par Navigae¹⁵ et les archives sonores le sont sur le Portail du patrimoine oral¹⁶ et le serveur européen Clarin¹⁷. Pour accroître leur visibilité dans les catalogues communs, nous les référençons dans un second temps sur le Catalogue collectif de France¹⁸.

RÉUTILISABLES

Au service de tous les publics, le réusage des archives est le défi qui nous anime avant tout. Puisque dès que c'est possible, nos données sont placées sous un régime de licences libres, nous ne sommes pas toujours informées des réutilisations. Aussi c'est toujours un grand plaisir pour nous d'apprendre la création d'une installation, d'un podcast, d'une pièce de théâtre ou d'une publication qui s'appuie sur les métadonnées et les données diffusées dans nos instruments de recherche. Dans la dynamique de la devise des principes FAIR « Aussi ouvert que possible, aussi fermé que nécessaire », notre vigilance est extrême sur les droits de réusage associés aux données. Ainsi, nous précisons tous les éléments nécessaires aux futur-e-s utilisateur-ice-s des données en indiquant systématiquement des licences explicites et accessibles. La résolution de ces questions est en réflexion constante, abordée de façon collective dans un carnet de recherche spécifique : *Questions juridiques et éthiques pour la diffusion des données en SHS*¹⁹. Un autre carnet sur la plateforme d'OpenEdition²⁰, propre à notre travail quotidien, partage ces expériences de réusage et plus largement décrit nos fonds, informe sur nos



← Fonds Jean-Pierre Olivier de Sardan - Notes manuscrites sur des questions de méthodologie, sur des exposés de pairs, projet de livre, extrait de cours, chemise n°7 (cote : MMSH_ODS_B3_C-19_001-075_018), page 18, 1990.

séminaires, nos formations et notre méthodologie²¹. Signalons, pour conclure, que pour chacun des fonds, lorsque les données et l'instrument de recherche sont prêts pour un dépôt sur le long terme sur les serveurs du Cines, nous rédigeons un plan de gestion de données rétrospectif²². Ce dernier, outre les informations d'usage, décrit les pratiques, les outils, les acteurs et leur évolution tout au long du processus de traitement : notre objectif est que l'utilisateur-trice de demain puisse relire, revoir, réentendre ces archives avec tous les éléments de contexte pour les appréhender dans la durée.

VÉRONIQUE GINOUVÈS

Archiviste, Secteur archives de la recherche -
Phonothèque de la MMSH,
Aix-Marseille Université/CNRS
Veronique.ginouves@univ-amu.fr

ÉMILIE GROSHENS

Archiviste, Secteur archives de la recherche -
Phonothèque de la MMSH,
Aix-Marseille Université/CNRS
Emilie.groshens@univ-amu.fr

[9] <https://lasdel.net>

[10] L'instrument de recherche du fonds JP de Sardan est en ligne <http://www.calames.abes.fr/pub/ms/Fileid-4739>

[11] L'export dans un tableau de type CSV est plus simple pour accéder aux métadonnées lorsque le réseau est restreint.

[12] <https://transcrire.humanum.fr>

[13] <https://francearchives.gouv.fr>

[14] <https://isidore.science>

[15] <https://www.navigae.fr>

[16] <https://vu.fr/elRu>

[17] <https://vlo.clarin.eu>

[18] <https://ccfr.bnf.fr/portailccfr/ark:/16871/00210469>

[19] <https://ethiquedroit.hypotheses.org>

[20] Développée par OpenEdition Center, la plateforme Hypothèses propose des carnets de recherche scientifique, celui du Secteur des archives de la recherche - Phonothèque de la Médiathèque SHS de la MMSH est : <https://phonothèque.hypotheses.org>

[22] <https://phonothèque.hypotheses.org/40780>