
DIM PAMIR

Patrimoines matériels – innovation, expérimentation et résilience
Appel à projet 2024-10, Référentiels franciliens

PROJET



Corpus, Référentiels
et Dictionnaires Ouverts

Descriptif

Document coordonné
par Agnès Tricoche

avec Michel Dabas, Katherine Gruel, Eneko Hiriart et Florence Weber

1. L'infrastructure CReDO 2

2. Référentiels spécialistes

Basefer : sites archéologiques de l'Âge du Fer	3
CELTclic : monnaies en Gaule	4
Riedones3D : coins monétaires celtiques	5
ACoR : techniques de la construction romaine	5

3. Référentiels transversaux

ArcheoLocalis : référentiel géographique européen	7
OCthe : référentiel typologique sur les monnaies celtiques	7
Kelticon : référentiel iconographique sur les symboles celtiques	8
ACoR : référentiel terminologique sur les techniques de construction	8

4. Méta-référentiels

Chronocarto : méta-référentiel cartographique	9
ArchEthno : méta-référentiel en ethnologie	10

1. L'infrastructure CReDO – Corpus, Référentiels et Dictionnaires Ouverts

L'infrastructure CReDO est un projet collectif et pluridisciplinaire visant à finaliser, à rendre accessible et interopérable une série de référentiels numériques en SHS, structurés depuis des années par le laboratoire AOROC et ses partenaires. Développé autour de deux disciplines principales, l'archéologie et l'ethnographie, le projet implique la mise à la disposition d'outils d'enregistrement, de stockage, de gestion et de partage de ces référentiels, Via l'acquisition d'une importante infrastructure de stockage et de calcul, pour permettre à des spécialistes de les consulter, de les manipuler et bien sûr de les réutiliser, depuis une plateforme en ligne.

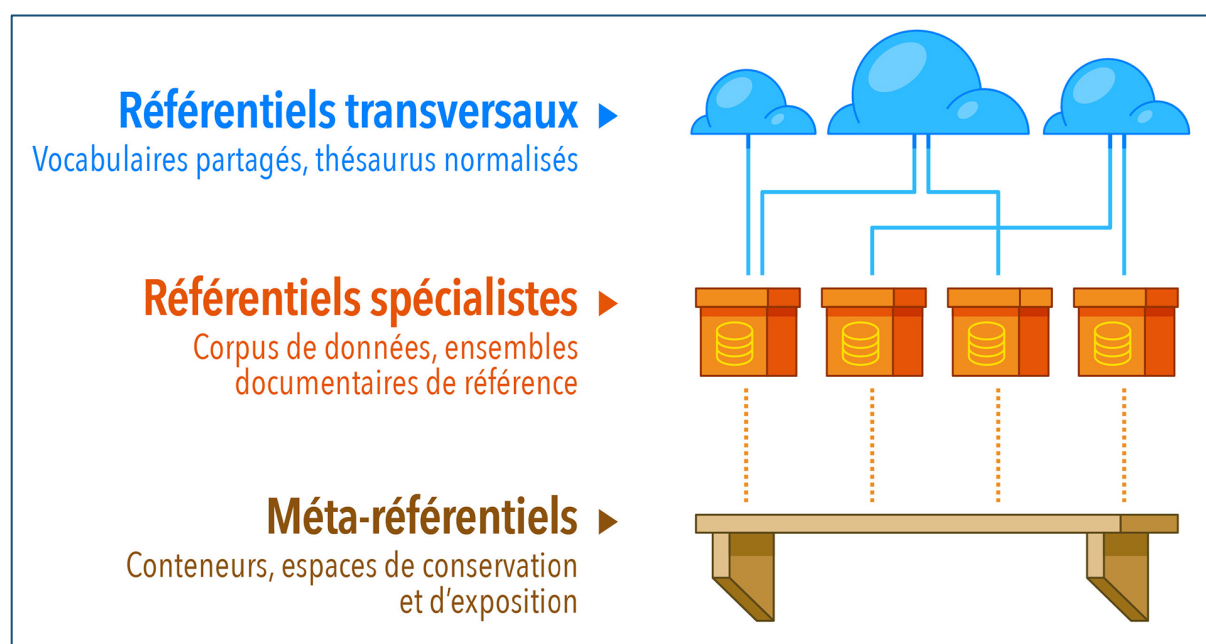
Les référentiels CReDO sont déjà en partie labellisés par des structures de recherche collaboratives au niveau de l'Université PSL (IRIS Scripta, GPR Faire collection, EquipEx+ Biblissima+) et du ministère de la Culture (programme Siamois – Système d'Information Archéologique Mutualisé et Ouvert reposant sur l'Intelligence Sémantique – porté par l'EPCC Bibracte). Le projet s'intègre en outre dans différents réseaux internationaux (International Research Network NEMESIS relatif aux territoires et sociétés européennes à l'âge du Fer, consortium ACoR sur les techniques de construction romaine, projet international Nomisma.org sur la numismatique, etc.).

L'écosystème des référentiels CReDO s'articule en trois catégories :

1- Les référentiels spécialistes : corpus de données de référence, portant sur des domaines spécifiques (objets, techniques anciennes, gisements archéologiques, enquêtes, etc.).

2- Les référentiels transversaux : vocabulaires partagés, thésaurus normalisés permettant d'adopter un langage commun pour décrire et faire dialoguer des jeux de données distincts, dans un réseau de données liées. Ils sont de nature géographique, typologique, iconographique ou terminologique.

3- Les méta-référentiels : espaces pour conserver et exposer les référentiels spécialistes dans un cadre conforme aux principes FAIR. CReDO comprend deux méta-référentiels : Chronocarto pour les données cartographiques et ArchEthno pour les données ethnologiques.



Ecosystème des référentiels CReDO

2. Référentiels spécialistes

■ BaseFer : sites archéologiques de l'Âge du Fer

<https://www.chronocarto.eu/projet/base-fer/>

La BaseFer est un référentiel collaboratif international qui répertorie à ce jour 33548 gisements archéologiques de l'âge du Fer géolocalisés par commune en Europe grâce à Archeolocalis (voir *infra*, p. 7). Il est disponible en 5 langues.

Son objectif est de cartographier les données actuellement disponibles pour caractériser l'originalité des régions occupées par les Celtes de l'âge du Fer. La description des gisements décline le type d'occupation (habitat, sanctuaire, funéraire, atelier, dépôts, objets isolés...) et les catégories de mobilier récoltés (armes, outillage, importations, parures...). C'est une base qualitative, de présence-absence de données publiées. La description des sites se développe sur 6 questions. Le code d'analyse compte 284 variables. La BaseFer vise à donner une image globale des connaissances sur le 1er millénaire avant notre ère en Europe, sans prétendre à l'exhaustivité. L'analyse chronologique simplifiée en 4 phases (50 % des sites sont seulement datés « âge du Fer » dans nos sources !), propose déjà un premier tri. Son but est de proposer à l'échelle régionale un arrière-plan chronologique.

Depuis février 2026, ces données sont ouvertes à tout public en téléchargement sous forme de tableur et de cartes de points. Nous souhaitons à présent développer un système d'alimentation automatisé de la BaseFer "l'Extractor" à partir des catalogues de thèse, en faisant appel aux outils de l'IA (annotation, recherche d'entités nommées, alignement sur nos listes de valeur, extraction sous forme de tableurs), en collaboration avec le consortium Huma-Num Pictoria. Un module bibliographique, actuellement inexistant, sera également ajouté. L'atlas de de l'âge du Fer doit évoluer vers une visualisation des données par cartes thématiques et un archivage des cartes thématiques déjà existantes et publiées. L'application offre plusieurs fonctionnalités pour interroger la base BaseFer, récupérer des données, obtenir des statistiques, identifier des communes, mettre à jour les fiches descriptives de site et obtenir une documentation sur la base. Un manuel en présente l'essentiel. Il est disponible dans la partie Documentation de l'application. Le manuel de saisie plus détaillé permet de rentrer plus avant dans la philosophie de BaseFer et de maintenir une certaine cohérence des données disponibles dans chaque pays au 1er millénaire av. J.-C.. Il est disponible en plusieurs langues comme la Base.

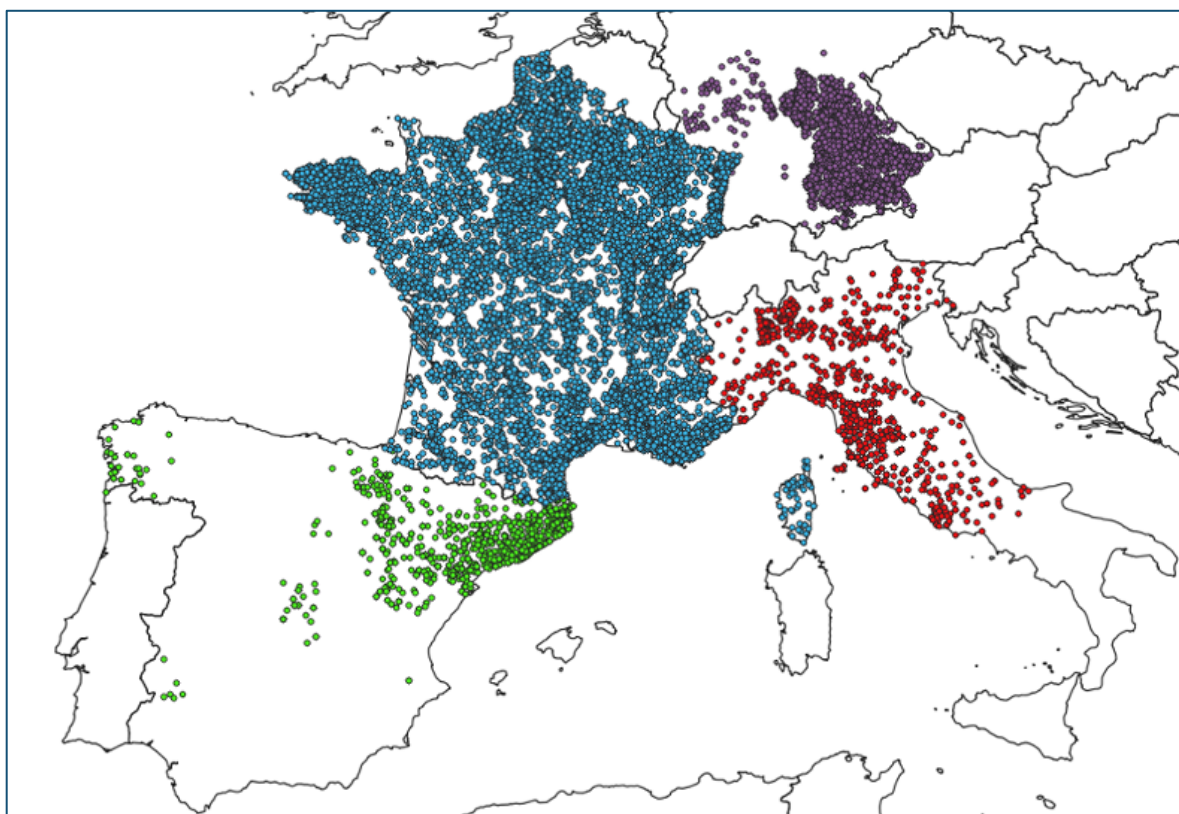
Calculs d'effectifs disponibles :

- totaux : nombre de sites avec au-moins une valeur, pour chaque élément caractéristique,
- sous-niveau géographique : répartition des sites selon ce sous-niveau (pays pour toute la base, super-régions et régions pour pays, départements pour région),
- élément caractéristique de site : nombre de sites pour chaque valeur de cet élément (typochronologie, structure principale et de niveau 1, 2, 3, écofacts et matériaux, catégorie et type de mobilier),

Les effectifs concernent les seuls sites répondant à la question en cours.

Un onglet Documentation extrêmement complet est disponible sur le site :

- Manuel d'utilisation de BaseFer en ligne (2015)
- Guide pratique pour la saisie des fiches de site (v3, 2020)
- Formulaire de saisie manuelle des données (v4, 2020)
- Documentation sur la datation des sites de BaseFer (2026)
- Documentation sur les communes dans BaseFer (2024)
- Documentation sur le lexique de BaseFer (2024)
- Documentation sur l'administration de BaseFer (2026)
- Documentation générale sur la structure de BaseFer (2026)
- Documentation technique sur la structure de la base pour l'atlas (2026)



▪ **CELTclic : monnaies en Gaule**

Porté par deux UMR, AOROC et Archéosciences Bordeaux, CELTclic est un vaste référentiel en construction, unique en Europe, devant s'intégrer dans le projet européen [Nomisma.org](https://nomisma.org) déjà opérationnel pour les monnayages grecs et romains. Il comprend une base de données typochronologique géoréférencée contenant plus de 50 000 monnaies en cours de vérification ; elle s'adosse à un thésaurus sous Opentheso, actuellement développé en collaboration avec la MOM dans le cadre de l'écosystème SIAMOIS, qui propose une typologie des monnaies en séries et classes monétaires (3000 classes répertoriées, réparties sur 1230 séries).

L'objectif est à la fois de mieux structurer les descriptions monétaires du monde celtique mais aussi de les appréhender dans leur contexte archéologique en association avec les autres marqueurs chronologiques. Le classement par région et par site permet enfin de dresser automatiquement des cartes de diffusion des séries monétaires présentes dans la base de données. Une application web dédiée est en cours de construction, pour permettre la consultation en libre accès de ce référentiel. Sa mise en réseau est assurée par des liens vers les collections monétaires de référence, dont la BNF (Gallica), les symboles dessinés dans KELTICON (ArchéoGrid), les sites référencés dans la BaseFer (Chronocarto), tous déposés sur Huma-Num. Les images sont stockées sur Huma-Num par l'outil ArcheoGrid développé par Ausonius. Le même traitement par l'IA que pour la BaseFer sera également testé (océrisation de catalogues et annotation par les chercheurs pour l'enrichissement du thésaurus et du corpus de monnaies via les outils de reconnaissance automatique).

▪ **Riedones3D / DIANet : coins monétaires celtiques**

La fabrication en série des pièces facilite le recours à l'intelligence artificielle pour l'identification du coin d'origine de deux pièces bien conservées mais aussi des images partielles ou dégradées par le décentrage, l'usure, la corrosion. Partant de l'expérience acquise dans le classement manuel des monnaies par coins, nous avons mis au point un protocole d'analyse des images intégrant des techniques d'apprentissage profond (« deep learning ») et d'algorithmes de regroupement vérifiés par des diagrammes de similarité entre les pièces. Après validation des résultats par l'expert, ceux-ci pourront être intégrés dans le processus pour affiner encore les capacités d'identification (logiciel auto-apprenant).

A la base de ce travail est l'exploitation d'images 3D des pièces scannées au 1/10^e de microns qui fournissent la précision nécessaire et un maillage géométrique sur lequel appuyer l'apprentissage profond. « DIANet » vise donc à intégrer l'IA dans nos pratiques régulières de travail par la création d'une interface utilisateur qui permet à partir des microtopographies de surface de chaque pièce d'en analyser les ressemblances et les différences et de les classer. Le trésor de monnaies gauloises de Liffre (FR_35) attribuées aux Riedones a servi de base à la mise en place de DIANet et Riedones3D constitue le fichier de référence nécessaire à l'entraînement du système maintenant opérationnel.

▪ **ACoR : techniques de la construction romaine**

<https://acor.huma-num.fr/>

Le référentiel des techniques de construction dans le monde romain est issu du programme ACoR (Atlas des techniques de la Construction Romaine), initié en 2012 autour d'un consortium international réunissant différents partenaires en France, Italie et Espagne. Son principal objectif est d'établir une grille d'observation des techniques conforme au processus de construction, en recourant à des catégories de description objectives, sans utiliser les appellations latines inventées par la tradition savante moderne (opus vittatum par exemple). Cette classification prend en compte et suit rigoureusement la logique des phases du chantier en proposant une lecture de l'édifice selon les

interventions de construction : préparation du terrain, fondations, élévations, ouvertures, couvremets/couvertures, sols, structures de service et arcs (qui ont une place à part, à la fois comme intervention, et comme des éléments constitutifs de techniques).

La base de données est en ligne et librement consultable depuis 2020. Trilingue, elle est alimentée au fil de l'eau, de façon collaborative et comprend en 2026 près de 1300 techniques de construction, concernant 430 édifices répartis sur plus de 100 sites du monde romain. L'Atlas des techniques de la Construction Romaine est développé sur un large champ géographique et chronologique et propose quatre modes de consultation des techniques sur son site web : un formulaire de recherche multicritères, une consultation par types techniques (regroupements de techniques selon des caractères communs), une exploitation cartographique et un manuel illustré interactif présentant, pour chaque intervention de construction, les principaux éléments de caractérisation. Le thésaurus terminologique des techniques est disponible sur l'instance Huma-Num d'OpenTheso (<https://opentheso.huma-num.fr/?idt=th455>).

3. Référentiels transversaux

- **Archeolocalis : référentiel géographique européen**

<https://www.chronocarto.eu/projet/archeolocalis/>

Archéolocalis est un référentiel ouvert à tous, qui rassemble toutes les coordonnées de géoréférencement des divisions administratives européennes jusqu'à l'échelle de la commune (découpage NUTS) ; il inclut les références aux nomenclatures internationales (ISO, Geonames, OpenStreetMap, Wikidata) et contient non seulement les communes actuelles, mais aussi celles ayant disparu par séparation ou intégration dans une autre commune. 17 pays européens sont actuellement intégrés (Allemagne, Autriche, Belgique, Danemark, Espagne, France, Hongrie, Irlande, Italie, Liechtenstein, Luxembourg, Pays-Bas, Pologne, Portugal, Principauté d'Andorre, République Tchèque, Suisse), les autres pays restant à compléter (récupération des codes NUTS, collecte des informations fines sur l'IGN de chaque pays).

Chaque commune est repérée dans un découpage administratif à 2 ou 3 niveaux au sein de son pays. Elle est localisée par un point central (son centre géographique ou sinon le centre administratif) en WGS 84 et selon un référentiel géographique local, avec une précision approximative de l'ordre du kilomètre. Archéolocalis sera à terme intégrable aux différents référentiels pour cartographier les données à *petite ou moyenne échelle*. Une api est déjà utilisable pour les bases de données hébergées sur le même serveur.

La documentation disponible sur le site comprend une présentation générale de la structure de ArcheoLocalis et une documentation sur l'interface de commandes (API) avec un [formulaire pour essai de l'interface de commandes \(API\)](#).

- **OCthe (Open Celtic Thesaurus) : référentiel typologique sur les monnaies celtiques**

Les Celtes de l'âge du Fer ont laissé une importante documentation iconographique dont les pièces de monnaies demeurent le support le plus éloquent (tant au niveau du nombre que de la variété des représentations). Néanmoins, l'iconographie monétaire n'a été que très partiellement analysée. Les approches se fondent le plus souvent sur des études stylistiques, mais aucune recherche globale n'a encore été menée sur des critères quantitatifs, ou sur des analyses spatiales et statistiques de vaste ampleur. Un examen plus approfondi du corpus d'images montre un usage de la plénitude de l'espace pictural et l'introduction d'un vocabulaire stylistique complexe, mais visiblement codifié que l'on retrouve sur tous les objets décorés de la période (fourreaux, torques, monnaies...). Les images centrales des droits (côté face) et des revers (côté pile), ainsi que les différents symboles monétaires présents au pourtour de ces éléments centraux semblent très codifiés. Leur position sur la pièce et leur traitement doivent être exploités d'un point de vue statistique et cartographique afin de mieux appréhender les phénomènes perçus et leur sens. Les inscriptions monétaires, bien que limitées à certains types, associent parfois plusieurs alphabets ou caractères.

L'objectif du thésaurus iconographique des monnaies celtiques OThe est la standardisation de l'ontologie et sa mise à disposition de tous dans le but d'une appropriation par la communauté scientifique et d'un enrichissement collectif. Le thésaurus défini dans Opentheso a pris en compte dès la phase de conception et de modélisation les référentiels de Biblissima afin de faciliter les opérations de mise en correspondance et d'alignement des données dans Biblissima+, Pactols, et, au niveau européen, Nomisma.org.

Le thésaurus est structuré en 4 niveaux. Le top-concept pour classer les monnaies celtes est géographique, générant des groupes monétaires obéissant à des règles de diffusion communes. Les niveaux suivants sont typologiques, distinguant des séries monétaires divisées en classes, pouvant elles-mêmes engendrer des variantes. Pour chaque niveau de concepts, on peut intégrer des alignements entre plusieurs typologies et intégrer des équivalences linguistiques. Chaque fiche est dotée d'un identifiant permanent, les illustrations sont appelées par des liens ARK dans les collections de références comme celle de la BnF ou des dessins stockés sur ArchéoGrid. La Base CELTclik (voir *supra*, p. 4-7) récupère automatiquement les données du thésaurus.

▪ **Kelticon** : référentiel iconographique sur les symboles celtiques

Le thésaurus Kelticon est un outil de classification et d'indexation de l'iconographie des monnaies gauloises, développé sous l'environnement Opentheso. Il vise à établir un vocabulaire normalisé permettant de décrire de manière homogène les motifs figurés, en réponse à l'instabilité terminologique de l'analyse iconographique. Organisé en trois niveaux hiérarchiques, il distingue d'abord de grandes catégories (anthropomorphe, zoomorphe, géométrique, etc.), puis des sous-catégories, et enfin des motifs précis correspondant aux formes observées. Ce système permet de couvrir l'ensemble des expressions figurées présentes sur les monnaies, des représentations humaines et animales aux motifs abstraits. Chaque image peut ainsi être indexée de manière structurée, facilitant la comparaison et l'analyse.

Ce thésaurus constitue un instrument de travail central pour coder les informations iconographiques. Il rend possible un traitement systématique, notamment statistique et spatial, des données. Avec plus de 800 concepts, il offre une base exhaustive et évolutive. Il s'inscrit dans une démarche visant à structurer un corpus iconographique complexe. Enfin, il constitue un socle méthodologique pour une approche renouvelée de l'étude des symboles protohistoriques.

▪ **ACoR** : référentiel terminologique sur les techniques de construction

Voir *supra*, p. 5-6.

4. Méta-référentiels

■ Chronocarto : méta-référentiel cartographique

<https://www.chronocarto.eu/>

Chronocarto est un web-SIG développé depuis 2006 à AOROC (ENS-PSL) pour la représentation de données spatiales en archéologie sous forme de cartes. Il a été conçu pour permettre un travail de réflexion en réseau entre chercheurs, non spécialistes de SIG, depuis la phase d'acquisition jusqu'à celle de l'enrichissement de documents existants en ligne et leur publication sous forme d'articles interactifs (avec DOI et contrôle par un comité scientifique international) où chaque figure est une instance de Chronocarto.

Chronocarto a été développé comme une solution client/serveur riche utilisant plusieurs bibliothèques simultanément en fonction du type de rendu côté client (bibliothèques javascript OpenLayers ou Leaflet) et de la mise en forme des données côté serveur (Mapserver plus GeoServer). Développé à partir de briques open-source (GeoServer, GeoTools, Apache Tomcat, MySQL, PostgreSQL, etc.), il est en développement permanent en fonction des demandes des utilisateurs (V9 en 2026).

Chronocarto est hébergé sur l'IR* Huma-Num et supporte la notion de projet associée à une communauté d'utilisateurs dont les droits sont modulables par un administrateur. La notion de copyright a été introduite aussi bien au niveau de la donnée que pour les cartes. Chaque projet est constitué d'un certain nombre de cartes éventuellement regroupées dans des dossiers différents. Les données peuvent être spatiales (fichiers images géoréférencées, fichiers vecteurs mais aussi base de donnée ou simple tableur) ou non-spatiales (fichiers de traitement de texte, images, vidéos, liens Internet). L'archivage est journalier et incrémentiel.

La création des cartes suit un workflow en 5 étapes : Définition du Projet >> Introduction des Données >> Style à appliquer aux données >> Définition des données à utiliser pour les Cartes >> Visualisation des Cartes.

Cette plateforme permet entre autres : un accès direct à des BD avec possibilité de modifications en ligne de tables attributaires ; la possibilité de créations en ligne d'entités spatiales (Points, Lignes, polygones, textes, dépôts documents, etc...) ; communiquer avec des systèmes externes via des API, facilitant l'échange automatisé d'informations avec des applications métiers, des catalogues patrimoniaux ou des plateformes institutionnelles ; requêter des données ; suivre à long terme des projets ; permettre la publication de données complexes géolocalisées ; croiser des informations spatiales (WMS, WFS, Données « communes ») ; utiliser des exemples pour l'enseignement de l'archéologie ; Mettre en valeur un site pour le public.

Le statut des projets peut évoluer dans le temps : 'en projet', 'en cours', 'achevé', 'publiable en ligne (public)' et 'archivé' (non visible mais stocké sur le long terme). Il est possible de ne rendre public qu'une partie des cartes d'un projet.

La plateforme gère des centaines de projets et plusieurs milliers de cartes (2026). Chronocarto continue d'être développé en interne et son accès est libre.

- **ArchEthno : méta-référentiel en ethnologie**

<https://archweb.ens.fr/>

En ethnologie et dans d'autres disciplines des SHS, ArchEthno est un dispositif et une méthode œuvrant en faveur de la mutualisation des données qualitatives. Modulaire, ArchEthno permet de documenter, gérer et diffuser des données de recherche polymorphes. Chaque projet repose sur la construction ad hoc d'un dictionnaire (Dictionary Editor) organisé en concepts hiérarchisés selon les principes SKOS, permettant de définir la structure d'une collection de données (corpus, enquête ou de tout autre objet scientifique). L'outil local de saisie (ArchEditor) permet de documenter ces données et de les analyser collectivement en toute sécurité. L'interface web de consultation (ArchEthno-Web) est disponible pour des enquêtes déposées sur un serveur sécurisé.