

7. Pôle des Humanités numériques

Responsables : M. Bui (DECU EPHE) et A. Tricoche (IE CNRS)

Le pôle Humanités Numériques d'AOROC a été officiellement créé en 2020, dans un contexte où de nombreux collègues liés à ce champ d'étude rejoignaient notre laboratoire : en 2019, une ingénieure en bases de données ; en 2020, quatre directeurs d'études de l'EPHE (trois en codicologie, un en informatique), dont la directrice de l'infrastructure numérique Biblissima/Biblissima+ (2012-2020 puis 2021-2029) et le titulaire de la chaire Humanités numériques et computationnelles de l'EPHE ; en 2020, deux ingénieurs en informatique, également issus de l'EPHE ; en 2021, un ingénieur en géomatique et analyse de données spatiales. Les humanités numériques, jusqu'alors essentiellement concentrées sur le travail de représentation spatiale en trois dimensions des objets et des sites archéologiques, se sont donc trouvées considérablement renforcées au sein d'AOROC ces dernières années, avec le traitement et l'analyse numérique des textes littéraires et épigraphiques, la recherche en intelligence artificielle, les systèmes d'information géographique, les bases de données ou les outils de publication.

Dirigé par une ingénieure (Agnès Tricoche) et un directeur d'études cumulant de l'EPHE (Marc Bui), le pôle Humanités Numériques se définit comme un axe transversal, œuvrant à définir une stratégie globale du laboratoire en termes de numérique et de science ouverte. Ses missions sont de porter des projets structurants et de mener une réflexion stratégique en collaboration avec la direction, en termes de formation, d'équipement, de gestion des données, de diffusion, etc. Intégrant huit ITA, il est aussi un service d'appui à la recherche, fournissant une veille sur l'actualité en humanités numériques (notamment à travers une newsletter mensuelle diffusée depuis 2021 à l'échelle du laboratoire et au-delà) et aidant concrètement les chercheurs à mener à bien leurs projets numériques en les orientant dans leurs pratiques et méthodologies (création d'un guichet au fil de l'eau, développement d'outils communs, mise en place d'un système de relais auprès de l'IR* Huma-Num...).

Par ailleurs, depuis 2020 AOROC est co-porteur de l'initiative Digit_Hum, sous la double responsabilité d'Agnès Tricoche (co-responsable du pôle Humanités Numériques) et d'une ingénieure issue d'un autre laboratoire (CAPHÉS, UMS 2267 CNRS/ENS). Cette initiative vise à promouvoir les humanités numériques et contribue à en structurer les pratiques à l'échelle de l'École normale supérieure et de l'Université PSL. Dans ce cadre, Digit-Hum organise des journées d'étude annuelles : "Sciences de l'Antiquité et humanités numériques" (en 2 parties, 2020-2021) et "Ouvrir les données de la recherche sur les sociétés contemporaines" (2022). Ces journées donnent lieu à un numéro spécial dans l'épître *Journal of Data Mining and Digital Humanities*, dont nous assurons depuis 2020 la co-édition scientifique au fil de l'eau pour la section consacrée aux Sciences de l'Antiquité. En 2022, Digit_Hum a également organisé avec la BnF une journée consacrée à la reconnaissance automatique d'écriture manuscrite, autour des plateformes Transkribus et eScriptorium (cette dernière étant développée au sein de notre laboratoire, voir infra) ; sous forme de retour d'expérience et de démonstrations, cette journée a dressé un état des lieux des nouvelles pratiques numériques dans le traitement des documents historiques et des archives.



Un ensemble de ressources, produites par l'initiative et sans cesse enrichies, sont proposées sur son site web (<https://digithum.huma-num.fr/>), développé et administré au sein du pôle Humanités Numériques d'AOROC : captations vidéo des journées, enquête, portraits, portfolio, carte interactive des masters en HN, glossaire, boîte à outils, etc.

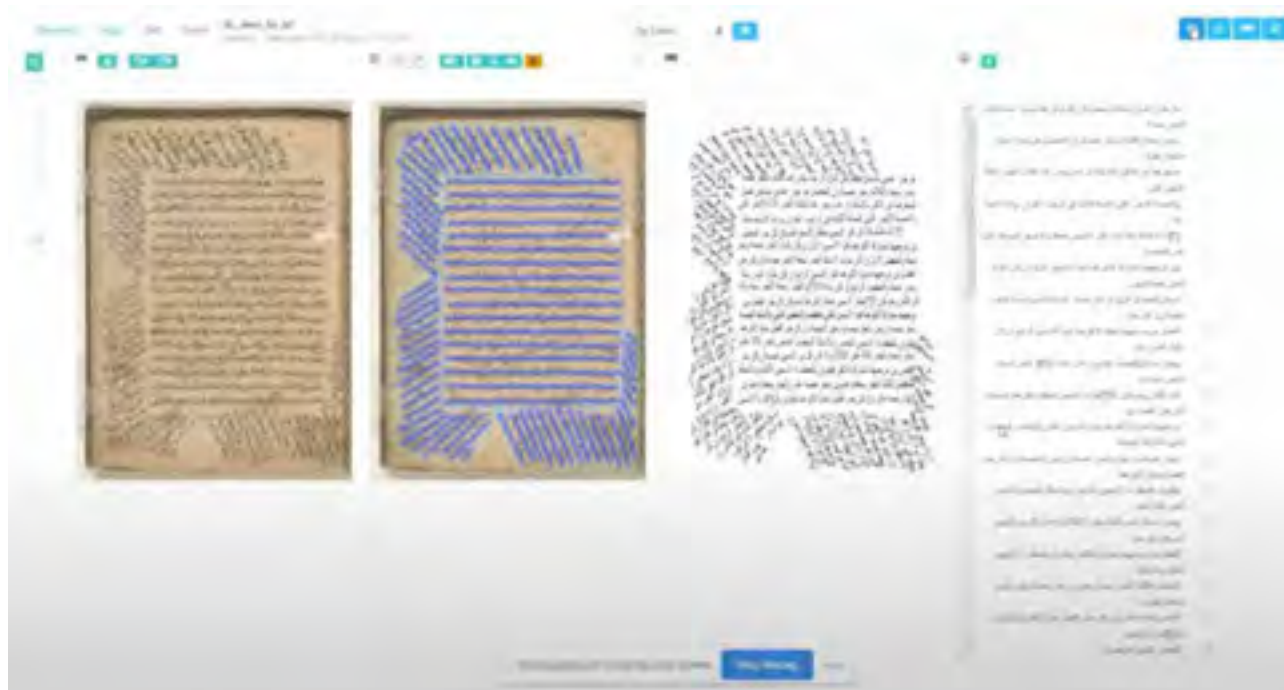
Outre les réseaux dans lesquels s'intègre le pôle Humanités Numériques, les projets présentés ci-dessous dressent un bilan des opérations les plus significatives menées dans le domaine des humanités numériques par AOROC entre 2017 et 2022, que celles-ci s'inscrivent dans le cadre de programmes de recherche ou qu'elles relèvent de services communs et d'outils collaboratifs destinés à l'ensemble de ses membres et au-delà.

1. Intelligence artificielle et logiciels

eScriptorium & Kraken. L'apprentissage profond est un domaine de recherche récent de l'intelligence artificielle visant à concevoir des modèles et des algorithmes d'apprentissage pour les réseaux de neurones profonds. Grâce à la capacité d'apprentissage à partir de données massives et aux performances supérieures de représentation et de prédiction, le Deep learning connaît de

grands succès dans diverses applications, comme la reconnaissance de caractères et d'écritures manuscrites au sein de documents, la segmentation d'images, la détection et la reconnaissance d'objets, la traduction automatique, pour n'en citer que quelques-unes.

eScriptorium : un exemple transcription de manuscrit hébreu (D. Stökl den Ezra DE EPHE/AOROC)



Le développement de la plateforme eScriptorium/Kraken est porté par des enseignants-chercheurs et des ingénieurs de l'EPHE rattachés à AOROC, pour l'étude des documents historiques. Déployée pour la transcription automatique des documents manuscrits et imprimés, elle est un projet phare pour l'IA appliquée aux écritures et langues rares. Dans ce cadre, Kraken (<https://kraken.re/>) est un système d'OCR clé en main optimisé pour les documents historiques et les textes en caractères non latins ; reposant sur Kraken, eScriptorium (<https://gitlab.com/scripta/escriptorium>) est une plateforme web opensource ayant pour but de fournir un ensemble d'outils pour transcrire, annoter et éditer numériquement des documents historiques. eScriptorium peut être déployé en de multiples instances pour des projets de différentes tailles et pour des travaux collaboratifs ; il bénéficie du soutien de Biblissima+, infrastructure numérique consacrée à l'histoire de la transmission des textes anciens et est utilisé par de nombreux collègues à l'ENC, l'INRIA, les Archives nationales et la BNF, l'infrastructure nationale DARIAH en Pologne et une vingtaine d'universités en France et à l'étranger. Parmi les projets de recherche à grande visibilité internationale ayant intégré eScriptorium, on peut citer : Hebrew Manuscripts in the Digital Age (Rothschild Foundation Hanadiv Europe), visant à apporter aux étudiants et chercheurs les méthodes de recherche actuelles utilisant le numérique ainsi que les compétences de base nécessaires pour entreprendre un projet de recherche sur les manuscrits hébraïques médiévaux ; RESILIENCE (H2020 & HEurope), infrastructure de recherche interdisciplinaire pour les études religieuses ; Manuscriptologia, projet de la région IDF pour se doter d'une infrastructure de calcul avec clusters de calcul (GPU) et de stockage pour l'analyse massive de grande quantité d'images d'objets inscrits ; Sofer Mahir (MESRI, Israeli Ministry of Science), concernant la transcription automatique des manuscrits rabbiniques. eScriptorium est également un élément central de l'ERC Synergy MIDRASH (2023-2029), qui étudie les migrations des traditions textuelles et sribales de manuscrits médiévaux en écriture hébraïque.

IA et numismatique. Récemment, la technologie du Deep learning a été également utilisée au

laboratoire dans le cadre des recherches en numismatique qui y sont menées. L'identification des monnaies issues d'un même coin est une opération visuelle longue et difficile. L'arrivée des scanners 3D à haute résolution permet de travailler sur des nuages de points et des maillages qui font abstraction des effets de surface. Le projet Coins_Riedones a pour objectif de parvenir à une restitution automatique de l'image gravée sur les coins monétaires originaux (les matrices de droit et de revers), au départ d'un corpus de référence de monnaies celtiques préalablement numérisées (musée de Bretagne de Rennes, voir infra). La mise au point du protocole a été menée dans le cadre d'une thèse de doctorat en informatique (S. Horache, 2019-2022) et est désormais fiable. La question de savoir si ce système de reconnaissance automatique de coins monétaires sera rendu accessible à la communauté scientifique, via un logiciel ou une application web, est actuellement en cours de réflexion.

Outils pour faciliter l'usage de NAKALA. Entrepôt de données de recherche pour les SHS mis en place par l'IR* Huma-Num, Nakala offre la possibilité d'archiver, préserver, publier et réutiliser les ressources déposées de façon sécurisée. Le système descriptif des données repose sur le format standard Dublin Core étendu, tandis que l'attribution individuelle d'un identifiant pérenne de type DOI assure les conditions d'accessibilité et de citabilité. AOROC exploite depuis plusieurs années les services de Nakala et a mis au point deux outils intrinsèquement liés à lui, pour accompagner les chercheurs et ingénieurs et faciliter le dépôt, la description et la consultation des données produites.

Le premier est **IOTA**, pour "Imagerie de l'Objet au Territoire Archéologique". Mené depuis 2016, le projet consiste en une application destinée au dépôt et à la consultation libre de la documentation primaire (brute) issue de fouilles menées au sein d'AOROC et depuis longtemps publiées. La version web de IOTA consiste en un moteur de recherche multicritère et inclut un système de navigation entre les divers documents numériques ou numérisés proposés à la consultation (plans, dessins, graphiques, photographies, mais aussi tableurs, diagrammes, textes ou cartes géoréférencées), avec téléchargement possible des fichiers originaux. Deux sites majeurs pour l'archéologie nationale et européenne, agglomérations artisanales de l'âge du Fer fouillées depuis plus de 30 ans par AOROC, sont actuellement accessibles sur l'interface web de IOTA : fouilles de Bourges, Port Sec Sud (Cher, Ve siècle av. J.-C.) et de Levroux-Les Arènes (Indre, Ile siècle av. J.-C.).

Le deuxième outil développé à l'initiative du pôle Humanités Numériques est plus récent et plus large. Il s'agit d'**ImportNakala**, disponible à l'adresse <http://importnakala.ens.fr>, application développée en 2022 pour publier par lots, de façon publique ou privée, des jeux de données sur NAKALA. L'outil permet également d'éditer les données publiées, une par une ou par lot, et d'envoyer de façon simultanée les modifications sur Nakala. Les types de métadonnées possibles pour chaque ressource sont illimitées (basées sur le Dublin Core qualifié). Cette application est ouverte à tous les membres d'AOROC, ainsi qu'à d'autres laboratoires SHS qui en font la demande.



Infrastructure logicielle de curation et d'exposition de données. Dans le but de développer des applications pour traiter et publier des données de recherche, simples à prendre en main, le pôle Humanités Numériques met progressivement en place depuis 2021 une infrastructure exploitant le framework Django. Elle permet d'encapsuler des programmes logiciels existants et de créer des

interfaces pour l'emploi d'applications métier. Les techniques de développement informatique récentes permettent de l'envisager sous la forme d'une "web app" (ou application web) s'exécutant dans le cloud, facilement accessible depuis un poste de travail. De premiers tests ont été réalisés au départ de bases de données de recherche (les blocs d'un monument égyptien, des inscriptions). Ce développement générique pourra permettre, à terme, d'interfacer des applications d'IA de façon commode et rapide, telles que Coins_Riedones pour la numismatique.

2. Imagerie et modélisation 3D

Les nouvelles techniques d'imagerie non destructives, comme la photogrammétrie, la numérisation 3D, la tomographie X, ou les procédés photographiques comme la RTI, transforment radicalement notre approche de l'archéologie en conservant la trace de sites, de monuments et d'objets, à un moment donné de leur histoire. AOROC documente et exploite par imagerie 2D et 3D des vestiges archéologiques mis au jour dans le cadre de ses chantiers, ainsi que des objets liés à des programmes de recherche ; il s'appuie pour cela sur un équipement performant, mobilisé selon les vestiges et objets à documenter (drones et scanners haute définition), et sur différents partenariats français (C2RMF, Centre de robotique de l'Ecole des Mines à Paris et Université de Cergy) et italiens (Istituto Nazionale di Geofisica e di Vulcanologia, Museo Nazionale Romano, Museo Archeologico di Polcoro).

Entre 2017 et 2022, plus de 150 modèles 3D ont été réalisés par notre laboratoire. Ils sont actuellement conservés sur la plateforme Sketchfab (collection AOROC dédiée : <https://sketchfab.com/AOROC/collections/edirne-museum-stone-sculptures-b5f9589652b84f32b6f61bca57ba4f4b>). La grande majorité d'entre eux constituent des documents de travail en cours (non publics), mais certains (environ 10% du total) peuvent être librement consultés, avec l'accord de nos partenaires.

Petits objets. En 2017-2022, les opérations de documentation 3D menées sur de petits objets ont principalement concerné trois types de collections.

- Les émissions monétaires celtiques. Dans le cadre d'une étude sur la reconnaissance automatique de coins (projet Coins_Riedones, voir supra), le trésor monétaire de Liffré (musée de Bretagne, Rennes) a été numérisé, représentant un lot d'un millier de monnaies riédones. De même a eu lieu la numérisation au scanner 3D d'un lot de monnaies gauloises de l'oppidum de Bibracte (environ 150 pièces), avec notamment pour objectif scientifique de comprendre les procédés de fabrication d'une partie des pièces coulées à la cire perdue.



Modélisation de monnaies par lot (2021, K. Gruel, S. Hourache EPHE/AOROC)

- Les objets d'art celtiques et étrusques. On citera d'abord la numérisation en 2018-2019 d'un ensemble de bronzes ornés provenant d'un char celtique du III^e s. av. J.-C. conservé au Musée d'archéologie nationale de Saint-Germain-en-Laye (fouilles du cimetière de La Fosse Cotheret, Roissy - Val-d'Oise, AFAN 1999) ; la numérisation des bronzes ornés a été réalisée avec un scanner adapté à la capture des objets les plus petits et des détails le plus fins. En 2021-2022, c'est avec le même souci du détail qu'a eu lieu la numérisation 3D d'un casque en fer et bronze, plaqué de feuilles d'or et de pastilles de corail, découvert en 1980 à Agris (Charente), chef-d'œuvre de l'orfèvrerie celtique du IV^e siècle av. J.-C. Associées à une couverture photographique macrographique et à des clichés en RTI, les données ont permis de réaliser un relevé à plat du décor très précis et de documenter l'ensemble des détails. Un dernier exemple notable est la numérisation 3D du mobilier des deux nécropoles (celto-)étrusques à Mon terenzio (Bologne, Italie) et de Montefortino à Arcevia (Ancône, Italie), entreprise en 2018-2019. Assortie d'une campagne de macrophotographie, cette opération a visé à documenter les principaux objets (casques, fourreaux d'épée, miroirs et strigiles) en vue de leur publication.

Modélisation 3D, dépliage et dessin du casque d'Agris (2019, B. Houal, T. Lejars, S. Verger)



- Les tablettes cunéiformes de l'Ecole Pratique des Hautes Etudes. Il s'agit de l'une des collections assyriologiques les plus importantes de France, étudiées depuis 2019 dans le cadre de l'IRIS Scripta-PSL (630 tablettes cunéiformes ou enveloppes, provenant de Mésopotamie, et une dizaine de briques inscrites provenant de Suse et de Babylonie). Du fait de la nature des supports en argile, la lecture de ces documents épigraphiques s'effectue en trois dimensions. La démarche scientifique entreprise est donc d'aborder les problématiques liées à la lecture, la conservation et la diffusion des textes par l'intermédiaire de la 3D.

La modélisation d'une tablette cunéiforme de l'EPHE (2019, J.-B. Houal)



Photogrammétrie à l'échelle de sites archéologiques. Entre 2017 et 2022, dans le cadre d'opérations archéologiques programmées ou d'urgence, a été entreprise la numérisation 3D, par photogrammétrie ou par relevé au scanner laser, de nombreux sites et structures archéologiques, en Italie et en Turquie. Faisant appel tant aux compétences des ingénieurs en interne qu'à celles de prestataires externes spécialisés, la réalisation de modèles 3D détaillés, à pied ou par drone, concerne principalement les sites suivants :

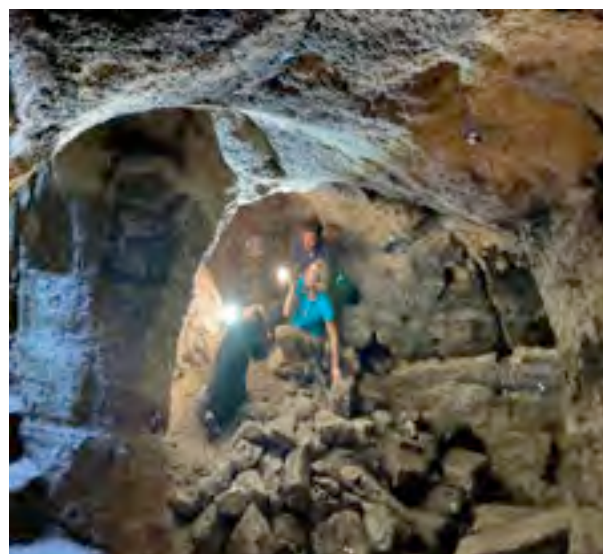
- en Italie, la tombe rupestre monumentale hellénistique de Grotte Scalina (2017-2019), la tombe des Cochons Noirs à Tuscania (2018) ; la tombe étrusque monumentale Lattanzi à Norchia et son secteur funéraire environnant (2019-2022), le sanctuaire étrusque de Piana del Lago (2021-2022) et le mithraeum de Santa Prisca à Rome (2022) ;

La modélisation 3D de l'église S.Prisca à Rome et de son mithraeum (2022, J.-B. Houal, F. Marchand-Beaulieu, C.J. Goddard)



- en Turquie, les bains Est du sanctuaire gréco-romain de Labraunda, avec ses structures souterraines et son système de chauffage à hypocauste (2017-2022), ainsi que la grotte dite de Pan à Ainos-Enez (2021), difficile d'accès et plongée dans le noir. Des modèles 3D des sculptures antiques en pierre provenant du site d'Ainos (musée archéologique d'Edirne) ont également été entrepris.

La modélisation 3D de la grotte de Pan à Ainos (2021, F. Marchand-Beaulieu, A. Dan)

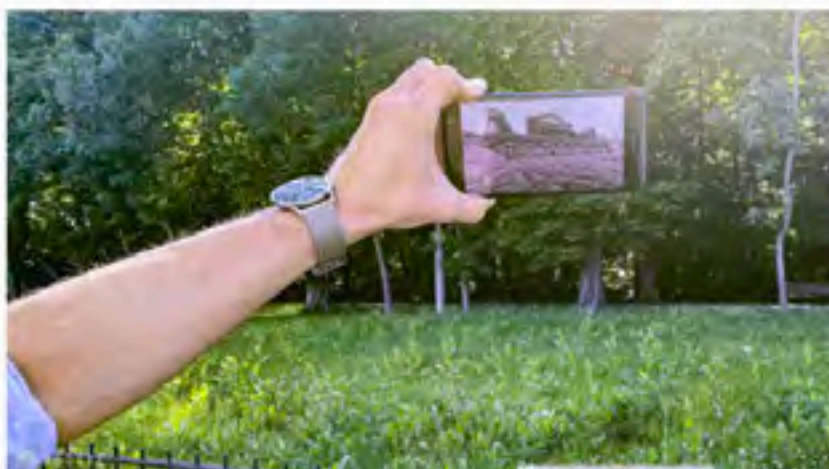


La réalisation de ces modèles 3D à l'échelle de structures archéologiques complètes, parfois condamnées à disparaître, permet d'en obtenir un plan et des coupes à l'échelle, la plupart du temps jamais réalisés jusque là, et d'en étudier l'architecture, ainsi que l'éventuel programme décoratif (par exemple pour les peintures pariétales de la grotte d'Ainos, plongée dans l'obscurité) ou les inscriptions (sanctuaire de Labraunda). Ces opérations apportent des données essentielles pour l'étude du bâti et de l'occupation de site, complémentaires au travail de fouille ou de prospection mené in situ. En outre, elles permettent la valorisation des recherches, à la fois au sein de la communauté scientifique et pour le grand public.

Archéoimmersion

Le projet ArcheoImmersion (J.B. Houal IR CNRS ; F. Marchand-Beaulieu IE CNRS) consiste en la création d'un espace numérique communautaire offrant la possibilité d'aborder le patrimoine par le biais de la réalité virtuelle du point de vue du professionnel comme de l'amateur. La collaboration du laboratoire CNRS Limsi/Venise et AOROC a permis d'obtenir en 2021 un financement pour un projet doctoral auprès de la région d'Ile de France. Il s'agira ainsi d'étudier l'utilisation de technologies immersives (allant de la réalité virtuelle à la réalité augmentée ou mixte) couplée à des approches intelligentes d'analyse de données pour aider les experts du domaine de l'archéologie à explorer en 2D et 3D, analyser et prendre des décisions informées sur les données archéologiques hétérogènes, complexes et incomplètes. Le travail consistera à développer sous Unity 3D un premier prototype d'application réalisant les fonctionnalités susmentionnées, qui devra être interopérable avec les applications métier existantes (Faro Scene, AgiSoft Photoscan, CloudCompare...), en intégrant l'utilisateur final durant tout le processus de conception, d'implémentation et d'évaluation. Existant : Ce projet de recherche s'intègre parfaitement dans les thématiques de recherche de l'équipe Imagerie 3D du laboratoire AOROC du CNRS-ENS et de l'équipe VENISE du LISN. Ce projet implique les experts dans le domaine de l'archéologie qui sont à la fois l'investigateur et aussi l'utilisateur final de tel système. AOROC dispose, comme beaucoup d'équipes en archéologie, d'acquisition de terrain 3D (photogrammétrie, scanne3D) dont l'utilisation est souvent trop limitée à ceux qui traitent l'information, en oubliant les gens de terrain, les utilisateurs finaux. Par ailleurs, les actions de recherche du projet correspondent totalement aux expertises de l'équipe VENISE (<https://www.limsi.fr/fr/recherche/venise>) dont Huyen Nguyen, Nicolas Férey et Patrick Bourdot sont membres, en particulier sur les problématiques de la supervision des interactions sensorimotrices et multimodales en réalité virtuelle et augmentée, et l'étude des systèmes de collaboration immersive colocalisée et distante. En conclusion, ce projet devrait permettre de renforcer les travaux de recherche d'AOROC et LISN et de deux équipes Imagerie 3D et VENISE, en particulier dans l'avancement des connaissances et la résolution de problèmes cruciaux qui se posent pour la collaboration en réalité mixte dans le domaine de l'archéologie et du patrimoine culturel. Impact pour la communauté scientifique SHS : Ce projet se situe dans un domaine de recherche émergent qui se trouve à l'intersection de la recherche sur l'archéologie et le patrimoine culturel, la science des données, l'interaction homme-machine, la réalité virtuelle et augmentée, et le travail coopératif assisté par ordinateur. Il a l'objectif de faire progresser les connaissances sur les méthodologies de recherche et l'application des systèmes collaboratifs utilisant la technologie immersive de réalité virtuelle et augmentée dans les domaines applicatifs pas seulement de l'archéologie et le patrimoine culturel mais aussi d'autres domaines potentiels. Le contexte immersif de réalité virtuelle et augmentée est ciblé dans ce projet afin de faciliter les interactions tridimensionnelles dans des applications scientifiques, mais aussi pour répondre à la dimension médiation scientifique pour laquelle la réalité virtuelle et augmentée est particulièrement appropriée. En outre, en faisant œuvre de pionnier dans l'étude de l'impact des "facteurs humains" au sein de ces systèmes dans des contextes de prise de décision entre des experts d'archéologie, ce projet développera les mesures comportementales implicites des utilisateurs qui ont des applications dans les domaines de la conception d'interaction, de la psychologie, de l'archéologie et de l'informatique. Pérennisation :

L'intérêt d'aborder la problématique de l'utilité de l'immersion dans le domaine du travail de terrain est de réellement modifier notre point de vue par rapport au post fouille. Il est donc important de faire correspondre les besoins par rapport aux outils. L'approche concertée entre les personnes de terrain et les informaticiens est donc importante dans une dimension prospective.



3. Cartographie et SIG

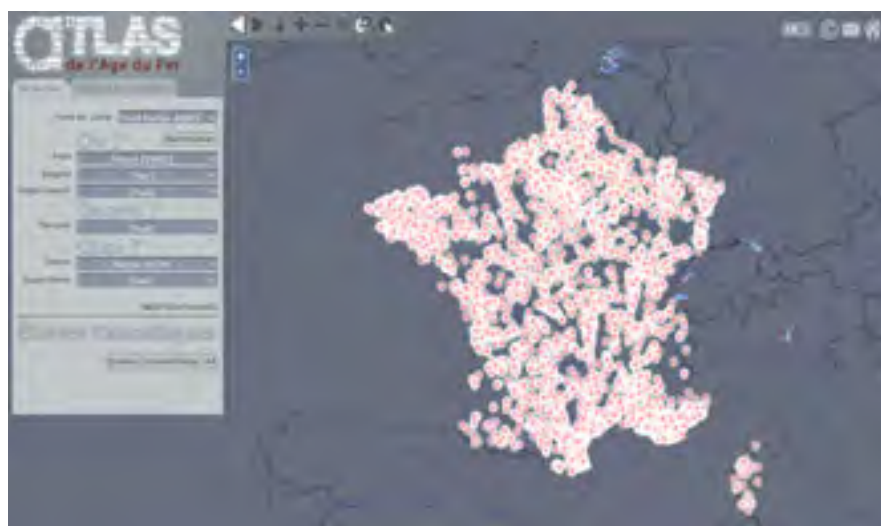
Les chantiers de fouille menés par notre UMR induisent une production systématique et exponentielle de données géoréférencées. Confronté aux besoins des chercheurs, à la fois de mettre en commun des données spatiales mais aussi d'initier un travail collaboratif sur le long terme et donc de stocker des données pour en assurer la pérennisation, le laboratoire AOROC a créé dans les années 2000 une application appelée **Chronocarto**. Il s'agit d'un Web-SIG, plateforme en ligne destinée à la manipulation de données archéologiques géoréférencées sur le Web, réalisé au départ de l'application GCServeur développée avec des outils libres par la société Géocarta. Depuis 2019, L'application Chronocarto est hébergée sur Huma-Num ; elle ne cesse d'évoluer (9ème version en avril 2023), via de nouvelles fonctionnalités émergeant au gré des besoins exprimés par les utilisateurs.



Concrètement, Chronocarto permet aux chercheurs de déposer, traiter et analyser des données archéologiques (données de terrain, prospections...) ; elle offre notamment la possibilité de superposer les données de prospections géophysiques et de fouilles, opérant de la petite à la grande échelle. Certains projets restent en accès privé (accès avec login) ; ils sont utilisés par un utilisateur ou une communauté de chercheurs désireux de travailler de manière collaborative. Une cinquantaine de projets sont en cours, qui dépassent largement le cadre de notre UMR et sont régulièrement le fruit de collaborations interinstitutionnelles. Les projets aboutis peuvent être ouverts au public, en accord avec leur(s) propriétaire(s) et font alors l'objet d'une mise en valeur en ligne, selon un classement par échelle cartographique (petite, moyenne ou grande), sur le portail de Chronocarto, également hébergé sur Huma-Num : <https://www.chronocarto.eu>.



Parmi les projets intégrés à Chronocarto figure l'atlas de l'Âge du Fer, visant à répertorier et à visualiser l'ensemble des gisements de l'âge du fer d'abord à l'échelle de la France puis depuis 2014 à l'échelle de l'Europe, en utilisant uniquement des données publiées (<https://www.chronocarto.eu/gcserver/atlas>). La base de données (BaseFer) sur laquelle s'adosse ce projet cartographique continue aujourd'hui encore d'être alimentée de façon collaborative par les chercheurs et étudiants intéressés par cette thématique, et compte à l'heure actuelle environ 20000 enregistrements.



Depuis 2022, une nouvelle rubrique ajoutée au portail de Chronocarto, intitulée "Data Papers", vise à proposer un nouveau type de publication en matière de Web-SIG, permettant de visualiser et d'interroger les données "brutes" auxquelles est associé un commentaire interprétatif. Il s'agit donc désormais d'un Web-SIG éditorial, identifié par un DOI (digital object identifier) et agrémenté de cartes interactives réalisées avec la bibliothèque JavaScript Leaflet.



Indépendamment de Chronocarto, certains programmes de recherche menés par notre UMR exploitent également cette bibliothèque libre Leaflet, pour la mise en valeur des informations géographiques produites, profitant des nouvelles possibilités et compétences offertes par le pôle

Humanités Numériques depuis 2019 (exemples de cartes dynamiques et interactives en ligne : VillaDiomedes Project, programmes ACoR et NAHAN, voir ci-dessous).

4. Sites web de mise à disposition de données de recherche

De nombreuses bases de données développées en interne constituent des outils de travail sur lesquels s'appuient les programmes de recherche que nous dirigeons ou auxquels nous contribuons. On peut ce titre citer la base de données OPUR (pour Outil Pour Unités de Réparation), conçue pour caractériser les anciennes réparations post-sismiques et socle de trois programmes d'envergure, passés ou en cours : ANR RECAP et ACCROSS, Bourse Marie-Curie Protect. La base de données OPUS permet quant à elle d'enregistrer les unités stratigraphiques construites, dans le cadre de missions archéologiques et de stages de formation à l'archéologie de la construction romaine (Pompéi, Chartres, Vieil-Evreux).

En parallèle, les années 2017-2022 ont permis de finaliser ou de lancer plusieurs gros chantiers de mise en ligne de données de recherche. Pour cela, nous bénéficions du soutien de l'Infrastructure de Recherche Huma-Num, qui héberge systématiquement les bases de données et interfaces web que nous concevons et développons en interne. Les données qui sont mises en disposition font du reste l'objet d'un dépôt sur le moteur de recherche Isidore, service en ligne d'Huma-Num, permettant l'accès à chacune des ressources numériques produites via un URI (Uniform Resource Identifier).

Villa Diomedes Project. La Villa de Diomède est l'un des tout premiers bâtiments fouillés à Pompéi, entre 1771 et 1775, et l'un des édifices les plus décrits et représentés par les voyageurs du Grand Tour. Ce programme pluridisciplinaire Villa Diomedes Project (2012-2019) piloté par AOROC en a retracé toute l'histoire jusqu'à l'éruption du Vésuve de 79. En 2019, la mise en ligne des archives complètes de la villa sur 250 ans, depuis la fouille initiale au XVIII^e siècle, jusqu'aux données numériques produites par la mission archéologique, a été réalisée par notre laboratoire : <http://villadiomede.huma-num.fr/bdd/>. Le site web trilingue comprend : 1. un corpus de près de 7600 archives (photo)graphiques ; 2. une étude du bâti, avec mise en concordance de plans de référence du bâtiment produits depuis le XVIII^e siècle, et analyse stratigraphique menée dans le cadre du le programme ; 3. un plan interactif de la villa. Associé aux opérations de restauration de la Villa de Diomède (2020-2022), AOROC en propose aussi une visite virtuelle en 3D, permettant une exploration



dans une sélection d'espaces, avec la possibilité de remonter le temps et de superposer l'état actuel de la villa à celui de l'époque romaine : <http://villadiomedehuma-num.fr/3D/app.html>.

ACoR. Depuis 2012, le programme international ACoR a pour objectif de développer un Atlas numérique des techniques de la Construction Romaine, sur un large champ géographique et chronologique. Il s'est concrétisé dans la réalisation d'une base de données, développée au sein d'AOROC et consultable en ligne depuis 2019, qui établit un corpus de techniques identifiées, classées et localisées dans un territoire, un site et un édifice (<https://acor.huma-num.fr/>). La caractérisation morphologique et technique des éléments de construction obéit à la logique des étapes d'un chantier (préparation du terrain, fondations, élévations, ouvertures, couvertures et couvertures, sols, structures de service et arcs), avec une analyse spécifique des traces de chantier. Le site web de consultation de l'atlas comprend une zone de consultation des notices, une typologie des techniques, une carte interactive, ainsi qu'un manuel interactif illustré (version papier à paraître 2023). Le thésaurus de caractérisation des techniques, trilingue (français, italien et anglais), est disponible en ligne depuis 2022 via l'instance Opentheso d'Huma-Num, outil de gestion en ligne fondé sur le web de données (schéma SKOS), ce qui lui garantit son interopérabilité et sa transposabilité à d'autres projets de recherche.

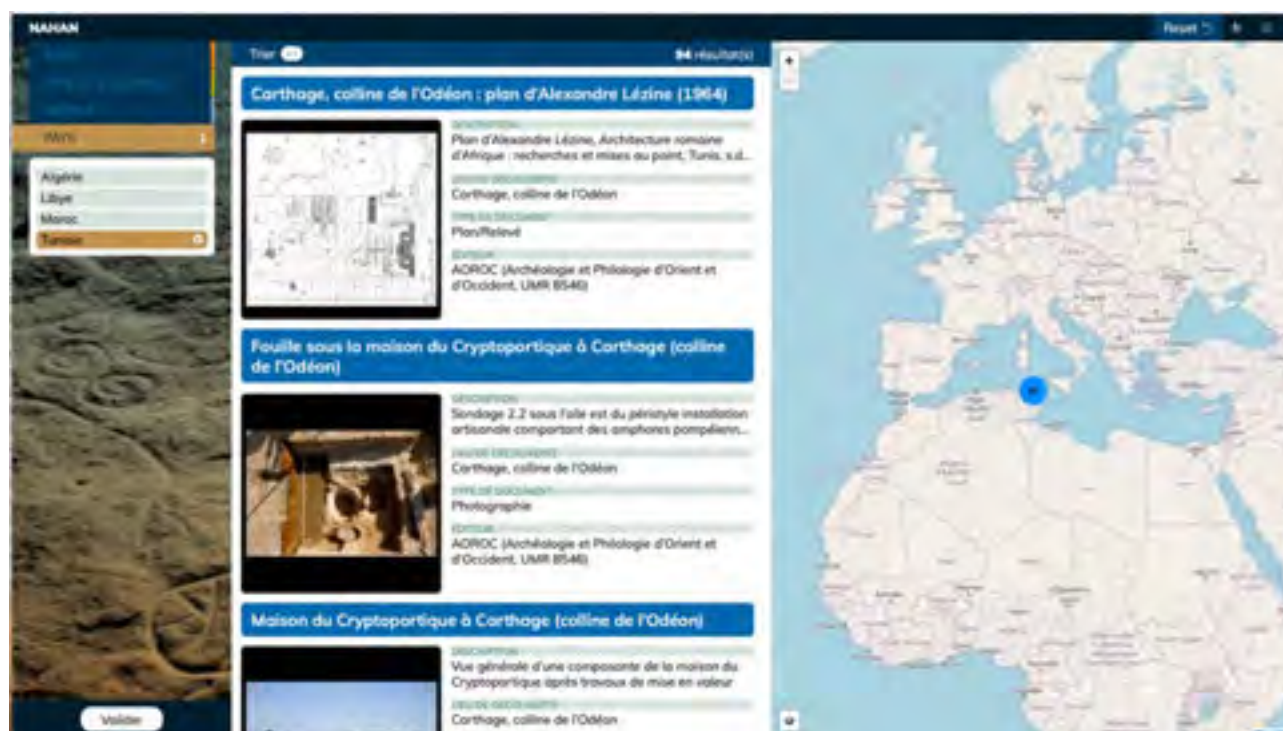
La base de données, collaborative, est aujourd'hui alimentée par des chercheurs, responsables d'opérations archéologiques ou étudiants impliqués dans un travail sur le terrain à l'intérieur du monde romain (une quarantaine d'utilisateurs inscrits actuellement, essentiellement français et italiens, étude de 208 édifices, répartis sur 35 sites).



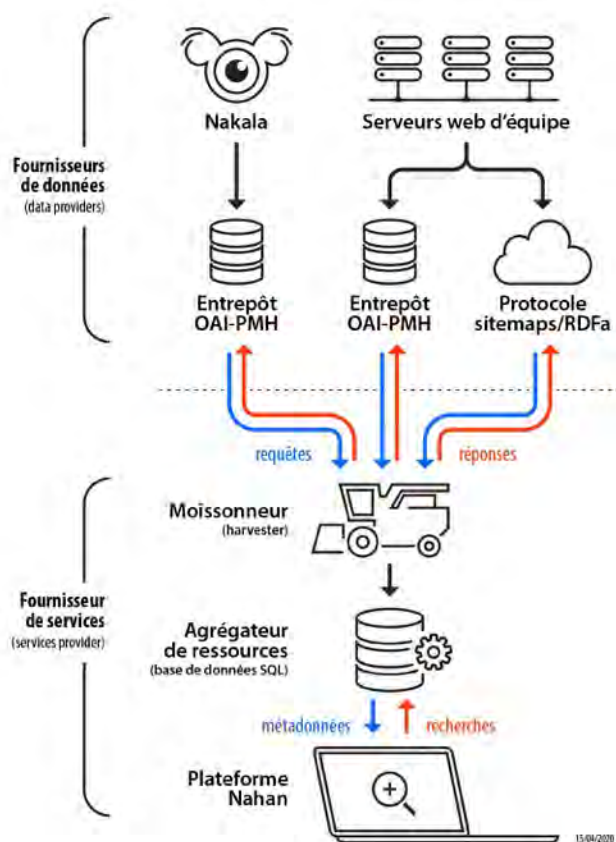
Nahan. Le programme international NAHAN (North African Heritage Archives Network : <https://nahan.huma-num.fr>) est le fruit d'un réseau d'institutions européennes et nord-africaines (19 partenaires, sous l'égide de l'ICCROM/UNESCO). Soutenu par l'infrastructure Biblissima+, son objectif est de créer une plateforme web en accès libre pour exposer, de façon centralisée, les archives numériques sur le patrimoine archéologique d'Afrique du Nord, conservées dans chaque institution.

Depuis 2020, AOROC assure la direction technique et la mise en œuvre de cette plateforme, actuellement en cours de finalisation (mise en production prévue en 2023). Elle fonctionne par moissonnage de jeux de métadonnées normalisés, afin de fournir un accès commun à des ressources stockées sur différents sites et serveurs et de proposer une redirection vers la source.

NAHAN s'adosse également à l'entrepôt NAKALA, outil de dépôt et de stockage de documents émanant de l'IR* Huma-Num, par le moissonnage de la collection NAHAN. A terme, les données pourront être géoréférencées dans le Web-SIG Chronocarto (voir supra), dans lequel le projet NAHAN aura un espace dédié.



Principes de moissonnage dans la plateforme Nahan



Facies monétaire antique. Depuis de nombreuses années, la base de données de recherche "faciès monétaire antique" permet aux spécialistes de numismatique celtique (AOROC et partenaires) de gérer leur documentation. Depuis 2021, cet outil de travail construit avec un logiciel propriétaire est en cours de refonte et de transfert vers une infrastructure répondant mieux aux objectifs de la science ouverte. Il s'agit aussi de l'adapter à l'ontologie numismatique internationale (Nomisma.org), en rassemblant plusieurs bases de chercheurs structurées de manière similaire en les intégrant dans une base unique développée dans un format opensource et référencée au format descriptif Dublin Core. L'objectif est de rendre accessibles les dizaines de milliers de données rassemblées et de permettre à terme leur moissonnage par le portail international Nomisma.



Vietnamica. Le programme ERC Vietnamica rassemble chercheurs et étudiants de l'EPHE (institution pilote) et d'institutions vietnamiennes. Ce programme vise à l'analyse des estampages qui reproduisent les inscriptions gravées sur les milliers de stèles de donation érigées dans les villages vietnamiens du XVIe au XXe siècle. Ces inscriptions, rédigées dans un chinois local émaillé d'écritures vernaculaires, ouvrent de nouvelles perspectives pour écrire une histoire des croyances religieuses, de l'économie et des structures sociales du monde rural. AOROC y contribue sur le volet des humanités numériques avec le traitement des inscriptions par les techniques d'intelligence artificielle avec l'emploi d'eScriptorium et Kraken. Le site web de Vietnamica (<https://vietnamica.online/>) rend accessible en ligne la documentation primaire numérisée, qui constitue la bibliographie de référence produite entre le XIXe et le milieu du XXe siècle. Elle comporte des milliers de livres et articles parus dans des revues savantes. La majorité de ces travaux, qui résultent d'enquêtes de terrain ou d'études fondées sur la lecture de sources primaires, est rare et d'une très haute qualité.



Notons enfin que les membres d'AOROC sont pleinement investis dans différentes actions de formation, telles le Master PSL en Humanités numériques, le projet Humanités numériques de l'EPHE ou le séminaire d'initiation aux méthodes de l'archéologie proposé à l'ENS. Les cours sont divers : ils concernent l'IA appliquée aux documents historiques, les technologies XML, la programmation Python, le traitement des images, la structuration des bases de données ou encore les systèmes d'information géographique.