

ARSOL « Archives du Sol »

Mise en ligne d'archives archéologiques numérisées

Rapport scientifique



Porteur : Olivier Marlet (olivier.marlet@univ-tours.fr)

Ingénieur de Recherche CNRS à l'UMR 7324 CITERES - Laboratoire Archéologie et Territoires, Tours

Le 11 janvier 2019, le conseil scientifique du GIS CollEx-Persée a sélectionné le projet Arsol « Archives du Sol » déposé en septembre 2018 dans le programme Numérisation.

Contexte

Ce projet visait à achever le programme de numérisation des archives d'archéologie de terrain du <u>Laboratoire</u> <u>Archéologie et Territoires</u> (LAT) de Tours (UMR 7324 CITERES, <u>CNRS</u> – <u>Université de Tours</u>).

Engagé dans le <u>consortium MASA</u> (Mémoires des Archéologies et des Sites Archéologiques) de la <u>TGIR Huma-Num</u> depuis sa création fin 2012, le LAT a pu trouvé différentes ressources pour commencer la numérisation de ses archives archéologiques en vue de leur mise en ligne.

Environnement institutionnel

Le Laboratoire Archéologie et Territoires (LAT) est une équipe de l'UMR 7324 CITERES (Université de Tours/CNRS), composée d'archéologues, d'historiens, de sociologues et de géographes (http://citeres.univ-tours.fr/). CITERES fait partie de la MSH Val de Loire qui assure la coordination du Consortium « Mémoire des Archéologues et des Sites Archéologiques », labellisé par la TGIR Huma-Num. Outre les chercheurs CNRS et les enseignants-chercheurs de l'Université de Tours, le LAT compte aussi de nombreux membres archéologues à l'INRAP. Le LAT coordonne la formation d'archéologie de l'Université de Tours au sein de l'Ecole Supérieur en Intelligence des Patrimoines (https://esipat.univ-tours.fr/) et compte une dizaine de doctorants.

Partenariat

Engagé depuis l'origine dans le fonctionnement du consortium MASA, le Laboratoire Archéologie et Territoires suit les préconisations données par le consortium en terme de bonnes pratiques de gestion et diffusion des données numériques en archéologie, et notamment le respect des principes FAIR.

Le consortium MASA réunit 10 partenaires, essentiellement des MSH (Maison des Sciences de l'Homme) plus le réseau (GDS) FRANTIQ, le Musée d'archéologie nationale (Ministère de la Culture), le Réseau des Écoles Françaises à l'Étranger et l'INRAP, autour de la thématique de l'accès à la documentation archéologique (archives et données numériques). Leurs compétences sont mises en commun autour de quatre axes de travail, afin de mettre à disposition de la communauté des archéologues des corpus, des outils et des guides de bonnes pratiques : archivage, web de données, interopérabilité, formation.

Ces enjeux sont liés à la nature même de la discipline : en fouillant, l'archéologue détruit irrémédiablement son objet d'étude sans qu'il soit possible de renouveler l'expérience. L'interrogation des données de terrain est nécessaire, dans des objectifs ré-interprétatifs et comparatifs, souvent à la lumière des progrès technologiques et scientifiques, d'où l'enjeu fondamental de traitement et de conservation des archives de fouille qui constituent les données primaires en archéologie. MASA aide les archéologues à profiter de la révolution numérique pour faciliter la consultation de cette documentation et la rendre pérenne. Avec l'augmentation croissante de l'informatisation des données, il devient indispensable de normaliser et rendre interopérables les données mises à disposition (référentiels commun, thésaurus normalisés).

Les objectifs de MASA découlent de la nécessité de mettre en réseaux les diverses compétences des acteurs scientifiques nationaux de l'archéologie afin de mettre en place des bonnes pratiques.

Projet ArSol

Description du fonds

Le projet porte sur le fonds des archives de fouilles du Laboratoire Archéologie et Territoires et est composé de deux collections. D'une part les archives de fouille de la ville de Tours initiés par Henri Galinié en 1969 et aujourd'hui coordonnées par le Laboratoire Archéologie et Territoires (une quarantaine de sites), d'autre part les archives de fouille du centre paroissial de Rigny (37) menées entre 1986 et 1999 par Elisabeth Zadora-Rio (chantier école de l'Université de Tours). Tours a joué un rôle majeur en France dans le développement de l'archéologie urbaine et le déploiement des méthodes d'enregistrement stratigraphique importées du Royaume Uni par Henri Galinié à la fin des années 1960. Quant à Rigny, il s'agit du seul centre paroissial fouillé sur une aussi grande surface en France. Ces deux fonds sont donc emblématiques de l'archéologie urbaine et de l'archéologie médiévale et on l'avantage de présenter une documentation parfaitement ordonnée selon un système d'enregistrement homogène.

Les unités d'archive sont constituées de classeurs contenant des fiches d'enregistrements de terrain (unités stratigraphiques, faits, murs, sépultures, mobilier, phasage), de cahiers ou de boîtes d'archive. Certains classeurs sont dédiés aux photos ou aux négatifs (avec planches-contacts). Les plans et relevés de terrain de formats supérieurs au format A4 sont archivés dans des armoires à plan. Pour les archives de Tours, il s'agit de plus de 100000 vues (fiches, pages, photos) et pour Rigny environ 30000 vues.

Pour les archives de Tours, environ 70000 vues avaient déjà été effectuées depuis 2015 grâce au soutien de la Ville, de la Région Centre Val de Loire et du consortium MASA. Pour Rigny, seule une petite partie des photos avait été numérisée.



Le fonds d'archives archéologiques du Laboratoire Archéologie et Territoires

Le système d'information ArSol

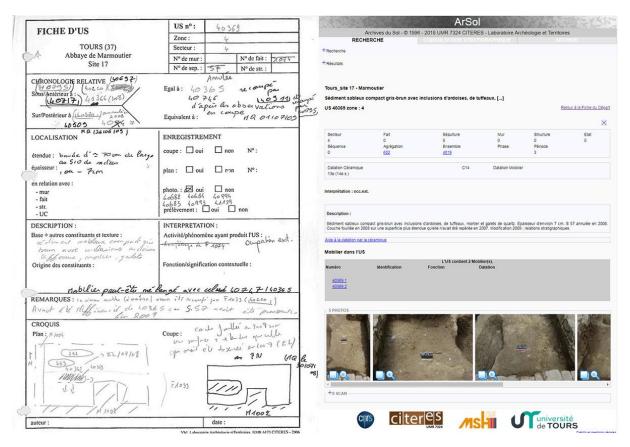
L'ensemble des données archéologiques du LAT sont enregistrées dans le système d'information ArSol (« Archives du Sol »). Ce système de base de données, dont une version est en ligne (http://arsol.univ-tours.fr) permet d'associer aux fiches d'enregistrement leur version papier pour permettre au chercheur de consulter la donnée primaire. Entre 2013 et 2018, 70000 vues du fonds de Tours ont déjà été numérisées. Le fonds de Rigny n'a pas encore été numérisé, mis à part quelques photographies pour les besoins de la publication.

Portée des données mises en ligne

A l'heure actuelle, ArSol est l'un des rares, si ce n'est le seul en France et en Europe, système d'information archéologique à la granularité de la fouille qui soit prêt à exposer les données dans le *Linked Open Data* selon l'ontologie du CIDOC-CRM pour les données patrimoniales et à produire un archivage en XML-EAD. Actuellement le fonds ArSol est accessible sous forme de triplestore via la plateforme <u>OpenArchaeo</u> mise en place par le consortium MASA. Ces données archéologiques et les archives numérisées associées alimenterons prochainement la plateforme européenne ARIADNE d'une part et seront versées au CINES d'autre part.

Conservation

L'archéologie détruit la majeure partie de son objet d'étude lors des fouilles. Cela confère un statut particulier aux archives produites. L'enregistrement de terrain constitue la seule documentation primaire à laquelle il est possible de se référer après la fouille. Si certaines fouilles récentes ont pu bénéficier d'une informatisation, ce n'est pas le cas des fouilles antérieurs aux années 90 et le seul moyen d'avoir accès à ces données est de consulter les archives. En outre, même pour les sites dont l'enregistrement a été informatisé, les enregistrements de terrain fournissent souvent davantage d'informations que celles saisies dans les champs de bases de données. Aussi, outre l'intérêt évident pour la conservation de ces données, la numérisation de ces fonds et leur mise en ligne est indispensable pour permettre aux chercheurs d'y accéder plus facilement.



Le système ArSol permet de lier l'enregistrement de terrain source aux données informatisées.

A ce titre les archives de fouilles de Tours (depuis 1968) et du centre paroissial de Rigny (1986-1999) sont des fonds emblématiques de l'archéologie urbaine et médiévale. Ils présentent une documentation structurée fondée sur un système d'enregistrement stratigraphique homogène importé du Royaume Uni à Tours dès la fin des années 1960 puis diffusé en France. Le système d'information <u>ArSol</u> développé sur cette base au début des années 1990 intègre les données nativement numériques plus récentes et permet d'associer la documentation ancienne numérisées. Il est le pivot d'une chaine de traitement de l'information archéologique allant de la fouille à la publication en assurant l'administration de la preuve.

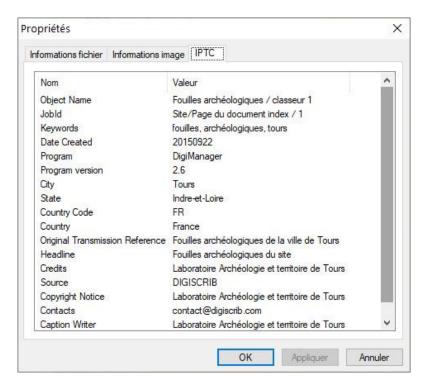
Les contraintes techniques liées au SGBD-R

Si le système d'information ArSol répond parfaitement aux attentes des chercheurs, il n'en est pas de même pour le système de gestion de bases de données relationnels (SGBD-R) avec laquelle il a été développé. Au début de l'informatisation en archéologie au début des années 80', les SGBD-R n'étaient pas légions : Microsoft Access et une application français nommée 4^e Dimension développée par la société ACI. Depuis le produit et la société se nomment « 4D ». A cette époque, les solutions web tels que mySQL ou postgreSQL n'étaient pas envisageables. La Laboratoire Archéologie et Territoires s'est donc tourné vers 4D. Jusqu'à 2014, de nombreux développements ont été réalisés successivement pour intégrer des modules spécifiques (contrôle d'intégrité des données, notamment des relations stratigraphiques, essentielles en archéologie pour établir une chronologie du site). Malheureusement, les développements à l'international de 4D ont induit des changements notables dans la syntaxe du code de développement qui a nécessité des révisions régulières des développements effectués en interne, souvent extrêmement chronophages. N'ayant pas nécessairement les budgets pour financer une prestation pour ces corrections parfois indispensables, nous avons décidé de changer d'application et de migrer tout le système ArSol sur une solution libre et full web en PHP-mySQL, nous affranchissant ainsi d'un logiciel propriétaire pour gérer nos données. Ce projet de migration a trouvé un financement conséquent grâce au Contrat-Plan État Région et pourra être lancé courant 2022.

En attendant cette migration, il nous a fallu prendre une décision concernant la mise en ligne des archives numérisées dans le cadre de l'AAP CollEx-Persée: soit intégrer nos données dans la version 4D d'ArSol (sachant que certaines fonctionnalités ne sont plus opérationnelles et que nous ne sommes pas en mesure de les corriger), soit publier directement dans une application web PHP-mySQL qui pourra être intégrée facilement à la future version PHP-mySQL d'ArSol. Ayant les compétences en interne pour le développement de la seconde solution, c'est bien sûr la solution qui a été retenue. Ce choix est d'autant plus pertinent qu'il nous permet d'abandonner une solution propriétaire et de nous orienter vers du libre, suivant ainsi les recommandations des principes FAIR.

La numérisation et l'indexation du fonds

Après avoir consulté plusieurs prestataires, nous avons confié la numérisation à un prestataire local (Digiscrib) avec lequel nous avions déjà travaillé pour les précédentes numérisations. Le prestataire a appliqué de manière automatique à chaque fichier numérique des métadonnées globales en format IPTC : date, identifiant de l'unité d'archive, titre du projet, responsables, mots clés, etc.

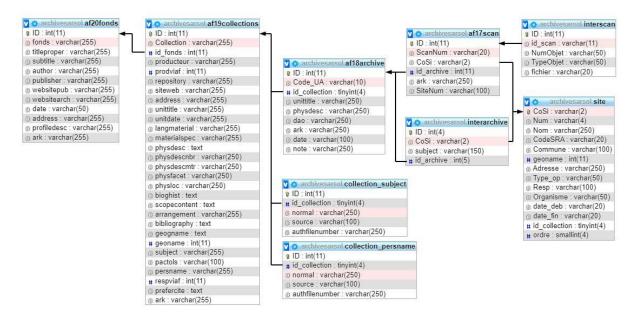


Exemple de métadonnées IPTC mises en place par Digiscrib

Pour les archives papiers, il a été décidé de numérisé en 300dpi couleur, certaines pages comportant des informations en couleurs comme par exemple des annotations *a posteriori* ou des corrections. Les plans et relevés grands formats ont également été numérisés en 300dpi couleur, qualité largement suffisante pour pouvoir détailler l'information. Les négatifs et les diapos ont en revanche été scannés à très haute résolution (2400 dpi) afin d'en permettre des agrandissements de qualité sans perte d'information.

Un contrat de 4 mois, niveau Ingénieur d'étude, a permis de traiter l'indexation de l'ensemble des fichiers (ceux numérisés depuis 2013 et ceux numérisés dans le cadre de ce projet CollEx-Persée). Cette indexation a permis de constater quelques oublis de numérisation. Celles-ci (moins d'une dizaine) ont alors été traitées par nos soins en interne dans les mêmes conditions que celles fournies au prestataire afin de ne pas perdre de temps. Ce contrat a été confié à Alexandre Viale.

L'indexation a suivi la structure globale suivante, préparée à des fins d'archivages en suivant le format EAD.



Nous avons donc un fonds (af20fonds), comportant plusieurs collections (af19collections). Chaque collection est décrite avec plusieurs sujets (collection_subject) et plusieurs responsables (collection_persname).

Chaque collection est composée de plusieurs unités d'archive (af18archive) : classeur, boîte d'archive, cahier, ouvrage, etc. Chaque unité d'archive est composée de plusieurs vues numériques (af17scans). Chaque scan est indexé (interscan) fait éventuellement référence à un ou plusieurs objets conceptuels dans la base ArSol (une sépulture, une fiche d'enregistrement, une structure, etc.). En effet, parfois une même vue peut présenter plusieurs objets conceptuels (par exemple 3 fiches d'Unité stratigraphique au sein d'une même page) et un même objet conceptuel peut renvoyer à plusieurs page (une fiche de sépulture dont la description peut se faire sur deux pages).

Une table inventorie les sites archéologiques (site). Une unité d'archive n'est pas nécessairement dédiée à un seul site archéologique et on peut retrouver les fiches de plusieurs sites au sein d'une même unité d'archive. Afin de pouvoir retrouver rapidement toutes les unités d'archives d'un site, on a mis en place un lien permettant de savoir quels sont les unités d'archives associées à un site (interarchive) et également un lien pour associer directement un scan à un site archéologique (une page est nécessairement associée à une seule site).

Volumétrie

Voici la volumétrie des données présentes dans la base de données :

Table	Quantité	Volume
Fonds	1	2,4 Ko
Collection	2	23,9 Ko
Unité d'archive	675	66,5 Ko
Vue	137840	9,4 Mo
Indexation	189713	12,8 Mo
Site	53	10,7 Ko

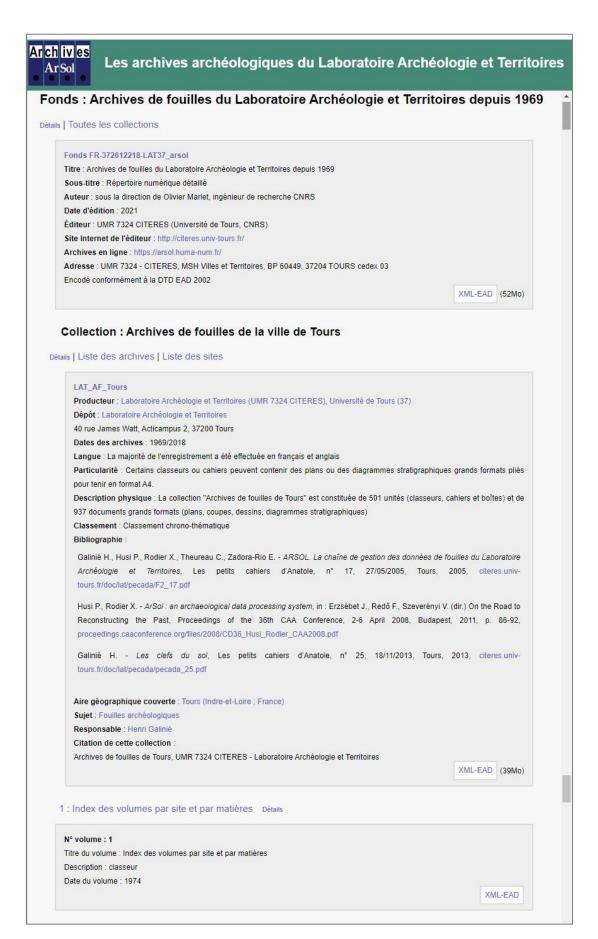
La totalité des fichiers bruts représente un poids de 683 Go (pour environ 140 000 fichiers).

Le poids de l'ensemble de fichiers en version vignette et en version web est de 30,5 Go.

Le développement du site web a nécessité environ 50h de travail pour une personne.

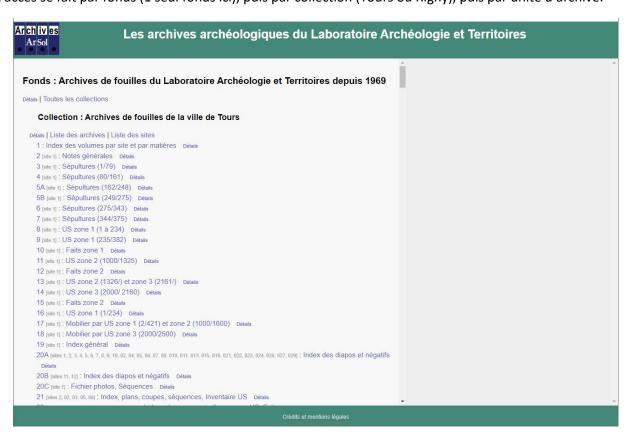
Publication des métadonnées

L'ensemble des données de la base permet de remplir automatiquement toutes les métadonnées nécessaires à la mise en place d'un fichier documentaire EAD (<u>Encoded Archival Description</u>).



La mise en ligne du fonds

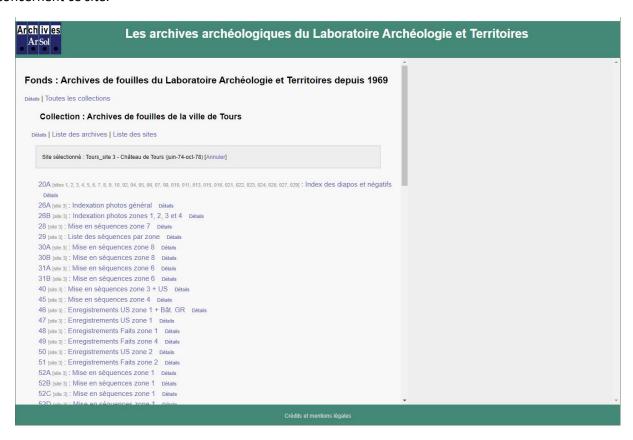
Le fonds numérisé est hébergé chez Huma-Num à l'adresse : https://arsol.huma-num.fr/
L'accès se fait par fonds (1 seul fonds ici), puis par collection (Tours ou Rigny), puis par unité d'archive.



Il est possible d'afficher la liste des sites archéologiques concernés par cette collection.



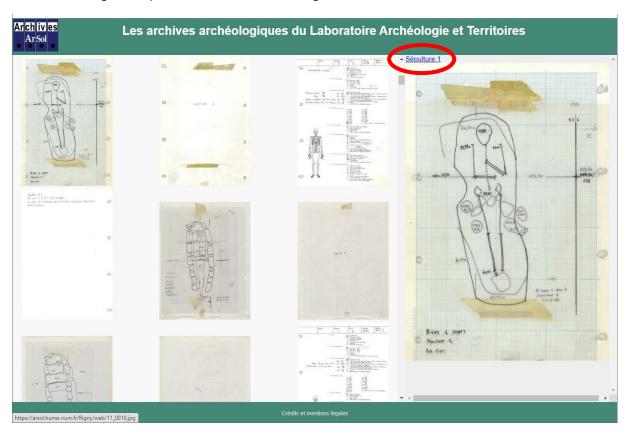
La sélection d'un site archéologique permet de filtrer les unités d'archives pour ne voir que celles qui concernent ce site.



L'ouverture d'une unité d'archive permet d'accéder à l'ensemble des vues sous forme de vignettes. Un clic sur une vignette permet d'en avoir une vue agrandie. Un clic sur la vue agrandie permet de la passer en plein écran. Il s'agit de vues réduites pour le web par rapport aux scans bruts. Un lien sera prochainement fait pour accéder aux scans bruts stockés sur une Huma-Num Box (serveur de stockage de fichiers volumineux proposé par la TGIR Huma-Num.

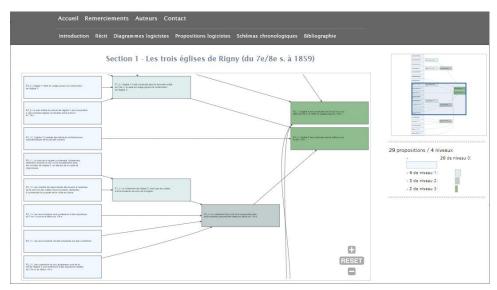


En outre, quand l'indexation indique une référence à un élément dans la base de données de terrain ArSol (une sépulture, un mur, une unité stratigraphique, un objet, etc.), un lien est automatiquement créé audessus de la vue agrandie pour accéder à la fiche en ligne de cet élément.



Exploitation scientifique

La mise en ligne de ces fonds d'archives permet aussi une valorisation scientifique. La <u>publication des fouilles</u> <u>de Rigny</u> est paru au printemps 2020 et cette publication a pour particularité d'être au format logiciste (précepte de Jean-Claude Gardin) permettant de formaliser tout le raisonnement de l'archéologue depuis l'observation de terrain jusqu'aux conclusions d'interprétation du site. De ce fait, le lecteur souhaitant consulter le détail de l'enregistrement de terrain sur lequel s'appuie une proposition de l'auteur aura accès à la donnée primaire. L'accès public à ce fonds documentaire participe donc directement au processus de publication et de valorisation des recherches archéologiques en assurant l'administration de la preuve depuis les synthèses.



Ce qu'il reste à faire...

(En date du 10 juin 2021)

La migration de la base ArSol actuellement sous un système libre en PHP-mySQL représente un chantier conséquent programmé sur 2 ans. Pendant cette migration, le site actuel des archives archéologiques du LAT sera directement intégré aux données de terrains d'ArSol, au lieu d'avoir comme actuellement un lien hypertexte permettant de relier les deux bases.

Quelques unités d'archives sont encore à numériser et à indexer pour compléter ce fonds, notamment des documents que les chercheurs ont chez eux pour travailler.

De même, les métadonnées concernant les sites archéologiques ne sont pas complètes (dates d'intervention, responsabilité de l'opération) et il nous faut explorer les archives en détail pour retrouver certaines de ces informations, ce qui se révèle fastidieux même si la numérisation a justement facilité ce travail.

En outre, il est également prévu de déposer l'ensemble du fonds au CINES et nous étudions également la possibilité d'ouvrir une instance sous NAKALA afin de bénéficier d'identifiants ARK automatiques. Nous étudions également la possibilité d'obtenir des identifiants ARK pour chaque ressource (fonds, collections, unités d'archives et vues) directement avec l'Université de Tours, ou la MSH Val de Loire dont dépend le Laboratoire Archéologie et Territoires.







Olivier Marlet olivier.marlet@univ-tours.fr
Le 10 juin 2021

Soutiens

(en plus de CollEx-Persée)







