

## SYSTÈMES D'INFORMATION ARCHÉOLOGIQUE: L'EXEMPLE DE ARCH'IS

### 1. INTRODUCTION

Voici plus de quarante ans que l'informatique a fait son apparition dans le petit monde de l'archéologie, principalement, à l'époque, dans le cadre de l'archéologie extraterritoriale (GINOUVÈS 1971). Destiné à l'origine à l'analyse en laboratoire, l'usage de l'ordinateur s'est peu à peu généralisé et surtout diversifié en élargissant le champ d'application à la numérisation des données de fouilles. La conception du logiciel Syslat par Michel Py (PY *et al.* 1991) à la fin des années 1980 illustre cette évolution. Simultanément, les Systèmes d'Information Géographique faisaient leur apparition suscitant un vif intérêt chez les archéologues. La notion de Système d'Information Archéologique s'imposait d'elle-même (colloque Informatique et archéologie de Saint-Germain-en-Laye 1991, DAUNE-LE BRUN 1991). Malgré l'engouement général, les solutions logicielles et matérielles<sup>1</sup> ont tardé à voir le jour. Cette lenteur a conduit à entretenir la disparité qualitative des données de fouilles, notamment en raison de l'absence de normalisation des pratiques et de la sémantique, corolaire inéluctable de l'enregistrement informatisé. Il en découle fréquemment des problèmes de cohérence entre les différentes sources d'informations dont la résolution constitue un travail fastidieux et redondant. Ce problème devient flagrant lorsqu'il s'agit de compiler des données dans le cadre d'une recherche thématique (thèse, PCR, ANR, etc.).

À cela, il convient d'ajouter les difficultés de diffusion de l'information au sein d'une même équipe d'étude. En effet, les solutions utilisées sont la plupart du temps locales et tributaires de la compatibilité entre les différentes plateformes et versions des logiciels en usage. Ainsi, la garantie que chaque chercheur dispose simultanément des mêmes données est, à quelques exceptions près, directement liée au degré d'achèvement du travail d'enregistrement et d'harmonisation. Quant au travail cartographique, il nécessite l'intervention quasi systématique d'un géomaticien, alors que nombre de tâches ne requièrent aucune technicité: elles ne sont que la projection sur un plan des résultats d'une requête effectuée sur une base de données. Il en découle que la capacité d'étude permise par les nouvelles technologies de l'information s'en retrouve largement diminuée par l'absence de simplicité, de travail en temps réel et par les habitudes individuelles.

D'où le projet Arch'IS<sup>2</sup>.

<sup>1</sup> Il faudra attendre le début des années 1990 pour voir lentement se généraliser l'usage de l'ordinateur en archéologie préventive.

<sup>2</sup> Conception et test: Nadine Mahé (INRAP), Sébastien Poignant (INRAP); développement: Sébastien Poignant.

## 2. LE CONCEPT

Au fil des discussions entre archéologues de tous horizons, il est apparu que “l’habitude de” ou encore le “nous on appelle cela comme ça” avaient un caractère décisif dans l’adoption d’un système d’enregistrement et d’exploitation des données. Ainsi, entre adeptes de Macintosh, de Linux, de Windows et au-delà de FileMaker, d’Acess ou autre, l’entente était difficile d’autant que les évolutions rapides des systèmes d’exploitation et des logiciels déstabilisaient les utilisateurs. Les “j’ai une version ancienne, je ne peux pas lire ton fichier...” ou “comment je récupère mes données dans Excel...” sont également des griefs fréquents au sein d’une communauté peu à l’aise avec le monde numérique.

L’objectif du projet Arch’IS était de supprimer, ou pour le moins minimiser ces difficultés. Dans le principe, l’idée est de séparer ce que voit l’utilisateur de ce qui existe réellement, c’est-à-dire distinguer le contenu (les données) du contenant (les interfaces). Dans cette optique, les technologies du Web 2 (Web dynamique) s’avèrent parfaitement adaptées, grâce à l’utilisation de programmes préprocesseurs capables de formater la présentation et l’information, non seulement entre le Serveur Général de Base de Données et l’utilisateur, mais également avec les différentes applications employées parallèlement.

Pour permettre le travail “en temps réel”, l’emploi d’un serveur distant s’impose naturellement. L’application devant être utilisable depuis le terrain jusqu’au formatage du rapport et la production scientifique, il fallait tenir compte de contraintes externes. Les interventions archéologiques ne sont pas toujours menées dans des zones couvertes par un accès inter/intranet, ce qui impose une utilisation “hors connexion”. C’est ainsi que le recours à un système AMP (Apache, MySQL, PHP) est une solution idéale grâce à son indépendance vis-à-vis des configurations matérielles. Ainsi, par le biais d’une interface de mise à jour, il est possible d’utiliser Arch’IS sur un ordinateur portable et de mettre à jour les données sur le serveur central *a posteriori*.

Dans l’optique de l’interopérabilité, les langages employés doivent être conforme aux standards du W3C, notamment ceux employant les syntaxes XML (XML, SVG, KML, VRML, X3D, SLA, IDML,...), mais également ceux permettant le transfert de données graphiques et attributaires avec les SIG (MIF/MID). Grâce à la mise en œuvre combinée des plusieurs langage de programmation et de balisage, Arch’IS peut non seulement importer des données issues d’applications tierces mais surtout exporter des données formatées (vers Word, Excel, Illustrator, Photoshop, ArcGIS, MapInfo, InDesign, Scribus, Acrobat, R, etc.).

## 3. ARCHITECTURE GÉNÉRALE

L’architecture générale est relativement simple, formée de 3 grands composants: le noyau, les interfaces d’entrée et celles de sortie (Figs. 1 et 2).

Le noyau rassemble un ensemble de scripts qui assurent la communication avec le SGBDD, la sécurité, le contrôle et le paramétrage des interfaces. Les interfaces d'entrée regroupent deux types de données que le noyau va utiliser afin de générer les pages html que voit l'utilisateur. Il s'agit des informations dites statiques, c'est-à-dire commune à tous les usagers et celles dites attributaires qui déterminent en fonction du contexte d'utilisation les règles de mise en forme et de contrôle des informations statiques. Enfin, les interfaces de sorties ont une constitution analogue aux précédentes, à la nuance près que le document produit peu aussi bien être une page html qu'un fichier téléchargeable à destination d'un logiciel spécifique.

Cette architecture offre une grande souplesse dans le fonctionnement de l'application, un même objet pouvant revêtir différentes formes selon sa situation. Ceci est rendu possible par l'emploi du langage de programmation PHP qui permet d'écrire des fichiers, temporaires ou définitifs, selon n'importe quelle règle syntaxique. Par exemple, un préprocesseur va écrire et ajuster la présentation d'un formulaire selon des règles de styles (CSS) définie par et pour un utilisateur particulier. L'interface utilisateur permet, selon un procédé identique, d'ajuster la terminologie affectant chaque champ (nom du champ, valeurs de la liste, obligation de réponse, etc.). De même, le lien entre les tables constituant la base de données est directement déterminé par l'utilisateur, ce qui permet de travailler autant à l'échelle de l'US, que de la structure, du fait, de l'UA, du site, d'un secteur, d'une région ou tout autre échelle ou notion. En effet, l'application offre la possibilité d'accueillir un nombre infini de références. Cette orientation utilisateur permet, entre autre, de privilégier la rapidité de réponse de la base de données en construisant l'imbrication des tables en fonction des besoins de l'utilisateur.

#### 4. SÉCURITÉ DES DONNÉES

Comme toute application utilisant un système client-serveur, la sécurité des données et des processus est primordiale. Les contrôles s'opèrent à plusieurs niveaux sur le serveur (Fig. 2). Lorsqu'un utilisateur se connecte, l'intégrité des fichiers est vérifiée grâce à leurs empreintes numériques (MD5: Message Digest, fonction de hachage cryptographique). Une fois l'opération réussie, les droits d'accès sont déterminés par authentification, chaque personne n'ayant pas accès à toutes les données ni fonctionnalités. En cour d'utilisation, des mots de passe complémentaires peuvent être requis pour effectuer certaines opérations sensibles, principalement lors de l'utilisation directe du langage SQL. Chaque information transmise à la base de données est "formatée" de manière à éviter les injections SQL ou l'introduction de séquence de code indésirable. Enfin, un certain nombre d'opération de sauvegarde et de maintenance sur la base de données ont été automatisée afin de garantir l'intégrité et l'accès des données.

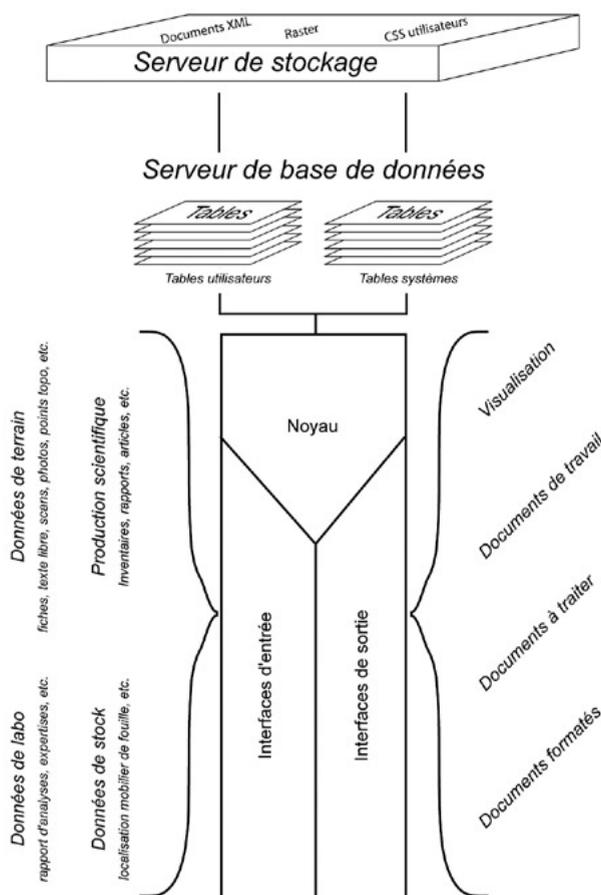


Fig. 1 – Architecture générale du SIA Arch'IS.

## 5. ACQUISITION ET MODIFICATION DES DONNÉES

L'insertion de données dans la base de données s'opère, dans la plupart des cas, par des formulaires. Chacun correspond à un type d'information: la structure, le mobilier, la documentation graphique, etc. Dans Arch'IS, on distingue deux types de formulaires: ceux contenant des champs classiques (texte, liste de valeur, case à cocher, etc.) et ceux constitués d'objets graphiques (Fig. 3). Pour les premiers, il s'agit de pages html classiques où les champs sont agencés selon les règles CSS propres à l'utilisateur et les listes de valeur

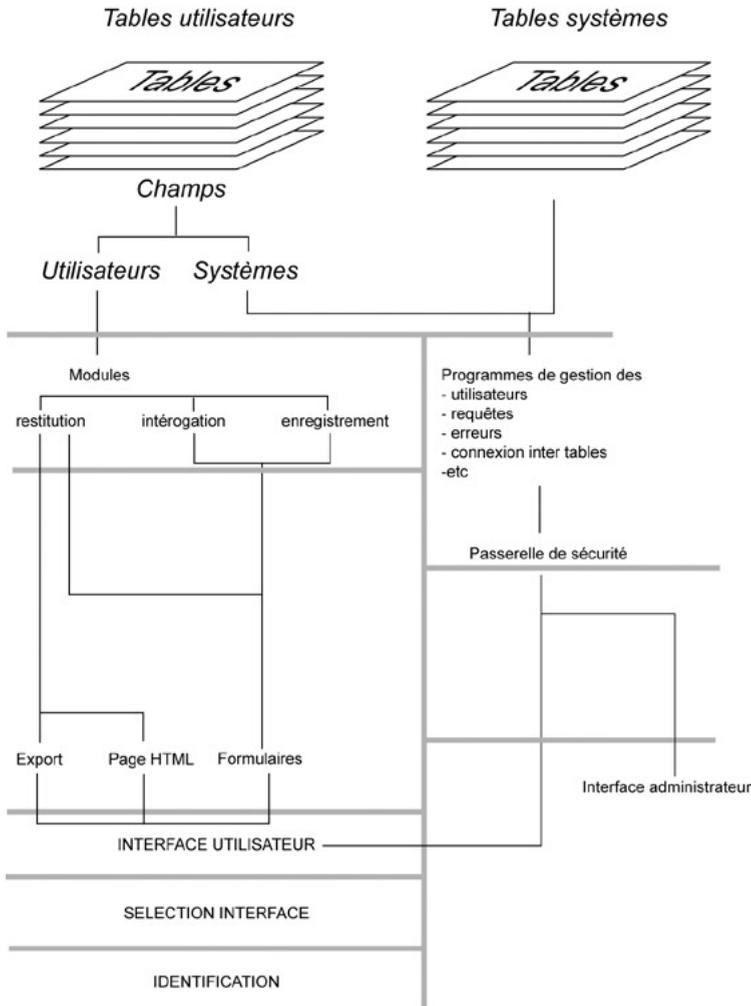


Fig. 2 – Détail de l'architecture générale du SIA Arch'IS.

générées en fonction des occurrences existantes dans la base. Pour les seconds, il est nécessaire de préciser leur nature. Dans certaines circonstances, la prise d'information est réalisée par le biais de schémas commentés. C'est par exemple le cas de certaines fiches anthropologiques sur lesquelles figure un éclaté du squelette humain. Le coloriage des différentes pièces osseuses reflète une information sur la présence et le degré de conservation. Afin d'introduire ces données dans la base, il est nécessaire de les retranscrire dans des champs texte,

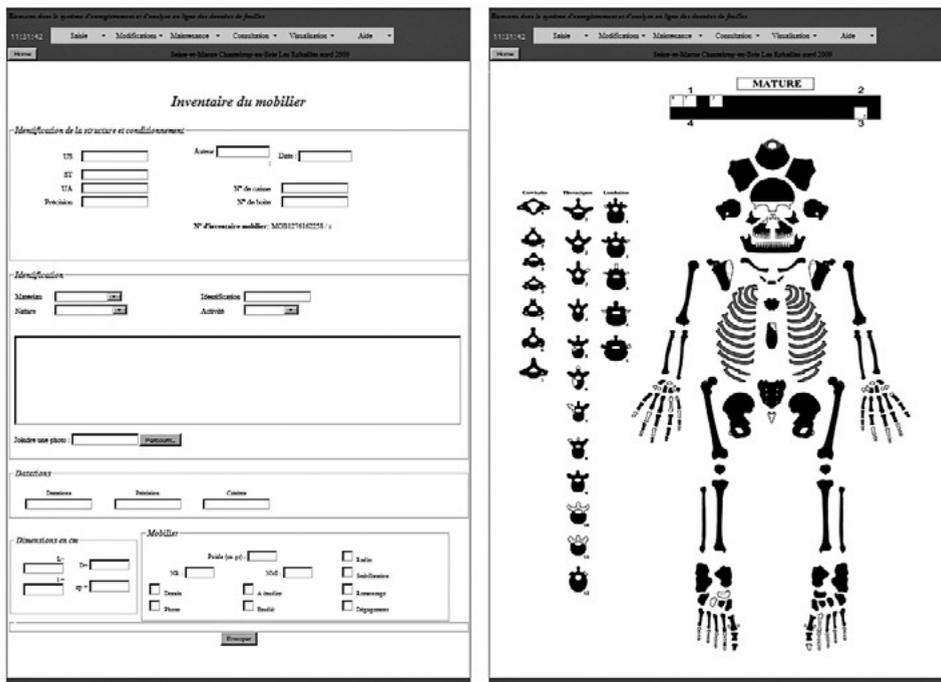


Fig. 3 – Formulaire de saisie ou de modification.

ce qui est parfaitement possible, mais fastidieux. La combinaison au sein d'un DOM (Document Object Model) d'objets graphiques au format SVG (XML décrivant les graphiques vectoriels) et de fonctions en javascript permet de rendre cette étape plus intuitive. En effet, le dessin d'une pièce osseuse peut être modifié par l'utilisateur et devient alors un conteneur d'information, similairement à champ classique. Tous ces formulaires servent également à la modification des enregistrements.

Il arrive parfois que les données à introduire dans la base existent déjà sous une forme numérique, généralement dans une autre base de données ou dans un tableur. Une fonction d'Arch'IS permet de rapatrier ces enregistrements, convertis au préalable au format CSV selon des normes préétablies (ordre des champs, dénomination des en-têtes). Une procédure similaire existe pour récupérer les données attributaires et graphiques provenant des SIG par le truchement du format ouvert de MapInfo, MIF/MID (MIF pour les formes et MID pour les attributs).

La saisie par plusieurs personnes engendre parfois de petites fautes ou divergences d'ordre orthographique ou sémantique. Dans le fonctionnement

d'une base de données, ces erreurs peuvent s'avérer décisives lors d'interrogation, principalement lors de recherches "plein texte". Une interface permet de résoudre ces problèmes grâce à des rapports croisés et des modifications "en chaîne". Dans cette même interface, on accède à des utilitaires de vérification de la cohérence entre les différentes tables (recherche de référence invalides, etc.) et d'harmonisation des informations.

## 6. EXPLOITATION DES DONNÉES

Comme de très nombreux SGBDD, sinon la plupart, la communication avec la base de données MySQL est assurée par un langage informatique spécifique, le SQL. Dans le but de faciliter l'interrogation, Arch'IS dispose de plusieurs interfaces permettant la construction de requête à des degrés de complexité différents. Lorsqu'il s'agit de recherches simples à l'intérieur d'une même table, les formulaires de saisie servent à effectuer la recherche, une dizaine de signes prédéfinis servant d'opérateurs logiques. Pour construire des requêtes plus élaborées, mettant en œuvre plusieurs tables ou impliquant des résultats calculés, un assistant gère la partie syntaxe du langage et permet la jointure de table et l'imbrication de requête sur deux niveaux (Fig. 4). Lors de recherches très complexes, un formulaire permet de soumettre un code SQL écrit indépendamment.

Afin de compléter et faciliter l'accès aux données, des procédures prédéfinies et paramétrables sont disponibles. Elles permettent l'extraction d'un résultat par filtrage et traitement de l'information. Les filtres s'appliquent sur les champs permettant de restreindre les résultats aux seules valeurs désirées. Dans certaines circonstances, le résultat attendu est un calcul comme par exemple une somme de valeur, nombre d'occurrence, etc.

## 7. RENDU DES DONNÉES ET DES RÉSULTATS

Dans Arch'IS, plusieurs mode de rendu existent. Les données peuvent être produites sous la forme de fiches (page html) structurées selon les règles CSS utilisateur. Cette présentation peut également se faire en formulaire ou en tableau (Fig. 5). Dans le cas des résultats de requêtes complexe, seul le mode tableau est disponible, les autres formes n'ayant *a priori* pas de pertinence. Lorsque les données recueillies sont des documents graphiques, la présentation se fait en planches contacts (tableau avec cellules fusionnées).

Parallèlement au rendu texte, Arch'IS dispose d'un processeur graphique qui permet la présentation des résultats sous la forme de cartes ou plans vectoriels géoréférencés (format SVG). La construction de ce document se fait grâce aux données topographiques (X, Y, Z) enregistrées dans la base. Chaque point sert de référence pour la construction automatique du dessin. Ainsi, l'insertion



Fig. 4 – Interface de l’assistant requête.

du levé topographique d’un chantier dans la base produit immédiatement un plan vectoriel (Fig. 6). Chaque élément archéologique est identifié dans le code SVG par son appartenance à une référence de site et son numéro d’entité. Ainsi, chaque objet est lié à ses données présentes dans la base. La réciproque est également possible: à partir du résultat d’une requête, il est possible d’afficher la répartition sur le plan. Les cartes sont divisées en calques ou couches qui rassemblent les éléments de même nature. Chaque couche peut être paramétrée: affichage, couleur, transparence. L’application permettant le travail sur plusieurs sites distincts, un formulaire permet de choisir les chantiers et leur niveau de détail. L’import de données matricielles (raster) est également possible.

## 8. INTEROPÉRABILITÉ

L’un des principaux enjeux de l’informatique actuelle reste l’interopérabilité, c’est-à-dire la capacité des logiciels à communiquer entre eux. Le métalangage XML et son instance XML Schema<sup>3</sup> s’impose aujourd’hui comme la meilleure alternative pour assurer cette fonction. Dans Arch’IS, les résultats d’une requête doivent pouvoir être récupérés dans une autre application plus spécifique. Ces résultats peuvent donc être produits sous la forme de fichiers à télécharger, dans un format compatible avec la destination et la nature des informations (Fig. 7). Le problème est que certains de ces formats nécessitent des bibliothèques absentes en

<sup>3</sup> XSD, remplace les DTD (Document Type Definition), définit la structure du document XML.

Saisie ▾ Modifications ▾ Maintenance ▾ Consultation ▾ Visualisation ▾ Aide ▾

Home Enregistrer le fichier

**6349 (Fosse), Ua :**  
 L= 1440- 1128mm h= 42  
 Dattions: IDCA 6ème - 7ème siècle  
 Fosse sub-rectangulaire à bords irréguliers (tendant plutôt à être évasés) et fond légèrement en cuvette. 6 US. de remplissage : - 6349 US. 01 : couche charbonneuse assez peu profonde (6 cm environ). - 6349 US. 02 : poche très charbonneuse avec cailloux brûlés, céramique et faune reposant sur 6349 US. 01, épaisseur approximative de 6 cm en son centre - 6349 US. 03 : couche limoneuse grise hétérogène reposant sous 6349 US. 01, 12 cm d'épaisseur environ - 6349 US. 04 : sédiment charbonneux gris, 8 cm d'épaisseur environ - 6349 US. 05 : remblais limoneux gris hétérogène, épaisseur allant de 10 à 20 cm - 6349 US. 06 : couche charbonneuse grise placée au fond de la structure, épaisseur de 12 cm environ en son centre.

**6110 (Trou de poteau), Ua 207**  
 L= 0.00- 0.54mm h= 30  
 Dattions:  
 Trou de poteau sub-circulaire à bords quasi-droits et fond irrégulier comblé par un limon gris meuble avec forte concentration de charbon de bois.

**6708 (Remblai), Ua :**  
 L= 3e- 64mm h= 20  
 Dattions:  
 Zone de remblais assez étendus recouvrant une zone de fours. Limon grisâtre avec inclusions de charbons de bois et présence de pierres éparées, épaisseur variable de 5 à 20 cm ; surface totale indéterminée, au minimum 200 m².

Listing

---

17:21:10 Saisie ▾ Modifications ▾ Maintenance ▾ Consultation ▾ Visualisation ▾ Aide ▾

Home Enregistrer le fichier

st	Ni	mat	nat	dat	prec	nr	pds	nmi
0	MOB1259755651 / 1	Alliage cuivreux	Objet			1	5	1
1966	MOB1248349293 / w	Os	humain			5	114	
6000	MOB1259244171 / n	Os	Sépulture					
6000	MOB1259661552 / j	Céramique	Récipient	GR		2	18.5	
6000	MOB1259843136 / n	Scorie				1	6.5	
6001	MOB1259657526 / y	Céramique	Récipient	HMA		4	7.5	
6001	MOB1259677909 / h	Terre cuite	Construction			6	79	
6002	MOB1259658334 / f	Céramique	Récipient	GR		1	8	
6003	MOB1259752894 / v	Scorie				164	4675	
6003	MOB1259757905 / e	Céramique		GR		1	10	
6003	MOB1259656172 / k	Céramique	Récipient			1	6	
6003	MOB1259757561 / r	Scorie				1	5	
6003	MOB1259676533 / x	Terre cuite	Construction			2	58.5	
6006	MOB1259659650 / f	Céramique	Récipient	HMA		1	5.5	
6006	MOB1259228551 / n	Os	Faune			1	1	1
6006	MOB1259843689 / i	Scorie				8	331	
6007	MOB1259662198 / k	Céramique	Récipient	HMA		3	11.5	
6007	MOB1259227871 / l	Os	Faune			4	3.5	1
6007	MOB1259913832 / k	Scorie				7	46.5	
6008	MOB1258463320 / t	Céramique	Récipient	HMA	Mérovingien	1	15	
6010	MOB1259830743 / k	Scorie				6	20.5	
6011	MOB1259828738 / i	Scorie				28	375	
6011	MOB1258469702 / f	Céramique	Récipient	GR		1	26	
6011	MOB1258554797 / n	Os	Faune			5	59	
6012	MOB1259843041 / z	Scorie				17	82	
6013	MOB1259913130 / n	Scorie				20	277	

Fig. 5 – Aperçu des rendus de résultats.



Fig. 6 – Aperçu de l'interface graphique.

Format	Import	Export	type de données	logiciels
.sql	x	x	requêtes SQL	BDD utilisant le SQL
.jpg	x	x	Image raster	
.tiff	x	x	Image raster	
.png	x	x	Image raster	
.csv	x	x	tableau	Open office, Microsoft Office, R, etc
.rtf	x	x	texte mis en forme	Open office, Microsoft Office, R, etc
.svg	x	x	Dessin vectoriel	Illustrator, Inskape
.xml	x	x	données balisées	logiciels gérant le xml strict
.vrml		x	données 3D	Vivacity, etc
.kml		x	données géospatiales	Google Earth, etc
.sla		x	PAO, texte mis en forme	Scribus 1.3.7
.idml		x	PAO, texte mis en forme	In Design v5
.inx		x	PAO, texte mis en forme	InDesign <v5
.mif/.mid	x	x	Données SIG	MapInfo, ArcGis, Grass, etc

Fig. 7 – Formats importés et exportés.

PHP ou indisponibles sur certains serveurs. Cette contrainte limite donc, actuellement, les possibilités aux seuls formats ouverts, généralement dérivé du XML. À titre d'exemple, on citera le SLA (Scribus) ou l'IDML (InDesign), deux formats dédiés aux logiciels de PAO. Des outils développés dans Arch'IS permettent de réaliser des mises en pages automatiques et standardisées, selon des règles de style définies par l'utilisateur. Enfin, grâce à la librairie html2pdf développée par Laurent Minguet, il est possible de convertir en .pdf les documents produits.

## 9. CONCLUSION

L'origine du projet était de construire un système d'information archéologique indépendant des configurations matérielles de manière à limiter à multiplicité des logiciels nécessaires au traitement de l'information. La mise en application d'Arch'IS, dans le cadre de 3 chantiers de fouilles menées entre 2007 et 2009<sup>4</sup>, à montré son efficacité et permis d'apporter des améliorations au système. Cette expérience dévoile plusieurs intérêts d'Arch'IS. Le premier, qui n'est pas une nouveauté, est de rassembler dans un même environnement l'ensemble de la documentation de fouille, dématérialisé par le stockage numérique. La mise en œuvre d'un système client-serveur permet à tous de disposer de l'information dont il a besoin, sans se soucier de l'actualisation des données ou de la transmission de fichiers. Le système de contrôle d'accès, en fixant à chaque site des attributs de communicabilité, offre la possibilité de travailler sur plusieurs gisements pour lesquels le chercheur dispose des droits d'accès. Le second intérêt réside dans la diversité des rendus de résultat qui

<sup>4</sup> Dirigés par Nadine Mahé, INRAP.

peuvent s'exprimer par du texte, des tableaux, des graphiques ou des plans. Le troisième intérêt, de l'ordre de l'ergonomie, est de permettre la personnalisation des interfaces en fonction de la sensibilité de chacun. Enfin, le souci d'assurer l'interopérabilité avec les autres logiciels utilisés, facilite et optimise le travail d'équipe. En définitive, l'expérience d'Arch'IS met en exergue l'intérêt de l'utilisation combinée de plusieurs langages informatiques, chacun spécialisés et complémentaire des autres.

SÉBASTIEN POIGNANT  
Inrap Centre Île-de-France

#### BIBLIOGRAPHIE

- DAUNE-LE BRUN O. 1991, *Colloque Européen Archéologie et Informatique (21-24 November 1991)*, «Paléorient», 17, 2, 178-179.
- GINOUVÈS R. 1971, *Archéographie, archéométrie, archéologie. Pour une informatique de l'archéologie gréco-romaine*, «Revue Archéologique», 93-126.
- PY M. *et al.* 1991, *Système d'enregistrement, de gestion et d'exploitation de la documentation issue des fouilles de Lattes*, Lattara 4, Lattes, Ed. de l'Association pour la recherche archéologique en Languedoc Oriental.

#### ABSTRACT

The project that gave rise to the API ARCH'IS was created in 2008. It was motivated by a discussion on optimizing the registration and use of data searches, particularly in the context of complex, diachronic and a huge variety of structures, sometimes split into several excavation sites. While the use of conventional GIS would have seemed the most appropriate, it raised the twin issues of training and the availability of licenses. In addition, it offered no probative solution for work in "real time". ARCH'IS has been developed by archaeologists to meet their own needs and interests of preservation of information. Its architecture is based on a MySQL SGBDD operated through a modular application implementing various individual languages in the world of the Internet. ARCH'IS integrated an engine map with useful features for most archaeologists, without specific training needs. Experimented on several projects of Marne-la-Vallée, the development plan of this API has now been established for many years.