

## MARS : *Multimedia Archaeological Research System*

Patrick SEMAL, Els CORNELISSEN & Nicolas CAUWE

### Résumé

MARS est un projet du Programme pluriannuel de soutien au développement de la société de l'information (2001-2008) initié par la politique scientifique fédérale (projet I2/AE/212). Il réunit dans sa première phase trois établissements scientifiques de l'État (Musées royaux d'Art et d'Histoire, Institut royal des Sciences naturelles de Belgique, le Musée royal de l'Afrique Centrale) et une société de service informatique spécialisée en technologie *Open Source* (Aragne). L'objectif du projet est de développer un outil informatique souple et performant qui peut servir de base à la recherche scientifique tout en gérant de manière rationnelle les objets de collections au niveau de leur conservation, de leur étude ou de leur exploitation muséologique. Les représentations multimédias doivent illustrer mais aussi, dans la mesure du possible, préserver les pièces originales en les substituant par leur représentation virtuelle.

*Mots-clés* : numérisation, collections archéologiques, *Open Source*, *Zope*, *Plone*, *Archetypes*, *Multimedia*, *Spy*.

### 1. Introduction

De par leur histoire, les collections archéologiques et anthropologiques sont physiquement dispersées entre différents établissements scientifiques fédérales (ESF), communautaires, régionales, communales ou encore privées. Dans le cas de fouilles faites à l'étranger, le morcellement est encore plus important puisque les collections sont partagées entre différents pays ou continents. L'aspect multidisciplinaire des produits de fouilles entraîne également leur morcellement physique entre les collections d'archéologie, d'anthropologie, de paléontologie, etc ... Ces collections sont, par essence, uniques et irremplaçables puisqu'elles sont le seul témoignage matériel d'une humanité disparue. De même, la fouille entraînant la destruction du site, la préservation de toute forme d'information est capitale dans une perspective de gestion patrimoniale.

La numérisation est de ce fait la seule méthode « réaliste » pour reconstituer des ensembles unitaires cohérents. Elle ne peut s'arrêter à un simple catalogue électronique des objets. L'ensemble des données associées est également requis afin de ne pas en perdre l'information contextuelle (enregistrements de terrain, contextes géologiques et sédimentaires, analyses chimiques, biologiques et physiques, etc...).

Pour une meilleure compréhension, le lecteur peut se référer à un glossaire des termes techniques principaux situé à la fin de l'article.

### 2. Contraintes techniques et économiques

L'une des contraintes techniques majeures du projet est son fonctionnement collaboratif au travers d'une application *Web*. Le produit développé doit être multi-plateforme tant au niveau du serveur d'application (*Windows*, *Linux*) que du client (*Windows*, *Mac*, *Linux*); on utilisera donc les fonctions des navigateurs *Internet* (*Mozilla/Netscape* et *Internet Explorer*). La configuration matérielle requise doit être variable et adaptable selon les besoins. En effet, l'utilisation de la technologie par les différents publics cibles ne doit pas être limitée par la configuration matérielle requise ou le système d'exploitation.

De manière générale, la technologie développée doit permettre la gestion efficace de bases de données multi-média de taille comprise entre quelques milliers et un million d'enregistrements. Ceci comprend autant les interfaces d'entrée et de consultation que la gestion de la base de données proprement dite et de ses utilisateurs.

L'aspect multi-langues est essentiel, tant au niveau de l'indexation des données que de leur consultation. De ce fait, nous pensons que la gestion dynamique du portail d'accès et du moteur de recherche est un avantage. Il doit toutefois rester compatible avec des pages *HTML* statiques respectant la norme *xhtml* 1.0. De même, la gestion des polices de caractères étendus « *Unicode* » est nécessaire en vue de la diffusion simultanée d'alphabets différents (alphabets latins avec ou sans signes diacritiques, slave, scandinave, cyrillique, grec,...).

L'encodage des données doit être possible sous forme d'objets ou sous forme de tables relationnelles. La base de données peut être gérée en interne ou liée à un système extérieur via des connecteurs *SQL*. Un processus de comparaison et consultation de thésaurus lors de l'entrée des données doit permettre l'uniformisation des données tout en respectant les spécificités régionales ou continentales. Les thésaurus doivent autoriser une recherche croisée multi-langues : la recherche du mot « néolithique », par exemple, doit fournir les entrées « néolithique » mais aussi « *neolithic* », « *neolithikum* » ou « *неолит* ».

La mise à jour du site et des données doit pouvoir être effectuée par un personnel non spécialisé en informatique via des mécanismes sécurisés et ce à l'aide de protocoles standards.

Les formats multimédia doivent permettre l'archivage et l'illustration de l'ensemble des informations (documents, sons, vidéos, images, animations, modèles 3D). L'ouverture du système à différentes catégories d'utilisateurs étant essentielle, nous ne comptons pas utiliser des formats nécessitant l'acquisition d'une licence particulière. Tous les formats envisagés seront accessibles via un navigateur *internet* équipé de *plugins* en accès libre. Dans certains cas, comme la description vocale d'un objet au format mp3, le support multimédia doit représenter une alternative transitoire ou définitive à un encodage complet des données et permettre l'utilisation des seuls mots-clés pour les recherches.

Les représentations 2D/3D des pièces remarquables doivent autoriser une première approche scientifique en permettant une représentation à l'échelle sur n'importe quel écran de consultation. Des mesures et des comparaisons sont également envisagées. De même, la combinaison des données graphiques (photos, cartes, plans) et des données encodées doit permettre la génération de représentations comme des cartes de répartition des objets en fonction de différents critères. La connexion à des sources d'information extérieures (*GIS*, bibliographie) est également nécessaire.

### 3. Technologie adoptée

L'utilisation de technologies *Open Source* préserve l'autonomie intellectuelle et financière par rapport aux grands standards commerciaux. Ce choix autorise un développement et une évolution souple et financièrement maîtrisée. Il garantit l'accès ultérieur à la technologie par tout autre public que celui ciblé lors de ce projet de démonstration, ainsi que la pérennité du système en dehors de toute société de service spécialisée.

L'analyse des différents systèmes possibles montrent que *Zope* semble pouvoir évoluer en fonc-

tion des besoins et des choix de la structure des bases de données (système intégré, système découplé et recherche sur fichiers plats). *Zope*, est une plate-forme de développement d'applications *Web* écrite et extensible en *Python*. Son utilisation est répandue tant dans le monde académique/publique (*NASA*, Ministère de la culture en France, *OTAN*) que privé (*BMW*, *CBS*). La nature « objet » d'une partie de la structure et sa gestion dynamique permet de gérer un seul contenu et de le présenter différemment en fonction du public visé, des choix linguistiques et de ses restrictions de diffusion associées.

Le produit *MARS* est développé sur une base de serveur *Zope* (version 2.7.3 compilé pour *Python* 2.3.4) enrichi de *Plone* (version 2.1). La sécurité des transactions est assurée à l'aide d'un serveur *web* « *Apache* » disposé en façade qui prend en charge le cryptage des données à l'aide du module *SSL*.

L'application *MARS* distingue clairement trois couches : la « Couche Données » (le stockage des données proprement dit), la « Couche Logique » (l'ensemble des méthodes appelées pour effectuer le travail sur les données) et la « Couche Présentation » (la génération de pages *web* permettant de mettre en œuvre l'interface avec l'utilisateur).

Concernant la « Couche Données », le stockage est effectué directement dans la base de données objet intégrée à *Zope* (la *ZODB*). Ce choix facilite l'installation du produit *MARS* au sein de petites entités qui ne bénéficient pas de compétences techniques étendues en matière d'administration de systèmes de base de données. Afin d'alléger le système, les fichiers multimédias sont stockés hors du système proprement dit et adressés par liens. De même, la possibilité de stocker les données dans des tables externes, sous des logiciels *SBDD Open Source* comme *Postgress SQL* ou *MySQL*, est retenue dans une seconde phase pour les collections des *ESF* en raison du très grand nombre d'objets.

Pour la « Couche Logique », *Plone* apporte une solution qui permet de contrôler la publication des données au sein d'un « *workflow* » (différents états de publication – privé, soumis à publication, publié, etc – liés aux différents rôles que les intervenants remplissent au sein du projet) et de bénéficier d'outils comme « *Archetypes* » qui permet de réduire le temps d'écriture du code de l'application tout en respectant la séparation des trois couches (données, intelligence et restitution).

Enfin, la « Couche Présentation » est écrite principalement à l'aide de *ZPT* (*Zope Page Templates*), un langage qui permet à la fois de respecter la norme *xhtml*, selon les recommandations du *W3C*, et de respecter le modèle en couche évoqué plus haut. En outre, les *ZPT* facilitent l'intégration du produit final à *Plone*.

#### 4. Les utilisateurs et les accès

Les accès sont gérés par des protocoles standards. Pour un serveur MARS, le niveau *internet* comprend l'ensemble des utilisateurs au sein d'une institution, tant en encodage qu'en consultation. Le niveau *extranet* correspond aux échanges entre utilisateurs des différentes institutions et avec des utilisateurs extérieurs. Différents niveaux de privilèges sont envisagés et permettent l'encodage et la consultation ou la consultation seule. Enfin, le niveau *internet* couvre les accès de consultation par le public extérieur.

#### 5. Requis des utilisateurs et modèle de données

L'un des objectifs du système MARS est d'améliorer la gestion et l'exploitation scientifique des collections. Le système favorise la collaboration et le partage de données entre chercheurs mais autorise une utilisation strictement privée. Chaque utilisateur enregistré dispose ainsi d'un espace personnel au sein duquel il peut stocker et éditer des documents privés. Il peut ensuite décider de les partager avec d'autres utilisateurs définis ou encore de les soumettre à publication afin de les rendre accessibles à l'ensemble de la communauté scientifique. L'auteur de la donnée possède les droits et responsabilités intellectuels relatifs à celle-ci. Chaque donnée publiée peut être l'objet d'un commentaire par un autre utilisateur; pour cette raison, une donnée ne peut être effacée ou modifiée par son créateur après sa publication. Chaque information est traçable au niveau de son auteur, de sa date de création et de publication et des commentaires associés. En utilisant le système MARS dès le début du processus d'étude d'un objet, il n'est pas nécessaire de ré-encoder les informations ultérieurement afin de les rendre accessibles, il suffit d'en modifier le statut de privé à public. Il s'agit donc d'un système collaboratif sécurisé.

L'unité de base du système est l'objet de collection. MARS permet de définir les informations relatives à des objets anthropologiques (restes isolés, individus, assemblages), archéologiques (objets isolés, individus, assemblages), de faune (restes isolés, individus, assemblages), de flore (restes isolés, individus, assemblages) et géologiques (restes isolés, individus, assemblages). Il peut être lié à d'autres objets de collections ou à d'autres objets comme une structure, une couche stratigraphique, une stratigraphie, une campagne de fouilles ou encore un site. Chacun de ces 15 types d'objets de collection est défini par environ 50 variables. Certaines sont obligatoires afin de rendre possible l'exploitation des données par le système. D'autres sont facultatives, elles permettent de fournir une information additionnelle. Les informations sur chaque

objet comportent des données historiques (origine, découverte, campagne(s) de fouille(s), historique de la collection), des informations administratives (collections, institutions, personnes, gestion et conservation), des données de description et de mesure, les analyses associées, les références et les documents multimédia (fig. 1).

La gestion des références bibliographiques constitue l'un des piliers de toute démarche scientifique. Le projet MARS intègre un module bibliographique performant qui allie la puissance des banques de données

#### OBJECT INFORMATION

<b>COLLECTION OBJECT</b>  Human remains Archeological object Fauna flora Geological object	<i>Identification</i> <i>Discovery coordinates</i> <i>Physical description</i> <i>Administrative parameters</i> <i>Deposit</i> <i>Description</i> <i>Taphonomy</i> <i>Measures</i>
<b>REFERENCE SAMPLE</b>	<i>Identification</i> <i>Description</i> <i>Data</i> <i>Measures</i>
<b>ANALYSE</b>  Dating Other analyse	<i>Identification</i> <i>sample, sampling</i> <i>laboratory, author</i> <i>protocol</i> <i>result(s)</i> <i>comment(s)</i>

#### GEO INFORMATION

<b>SITE</b>	<i>Location, discovery</i>
<b>EXCAVATION</b>	<i>Date(s), excavators, unit(s)</i>
<b>STRATIGRAPHY</b>	<i>Substratum, layers, historical background</i>
<b>STRUCTURE</b>	<i>Identification, components, cultural attribution</i>

#### ADMINISTRATIVE INFORMATION

<b>INSTITUTION</b>	<i>Identification, description</i>
<b>PEOPLE</b>	<i>Identification, contact</i>
<b>COLLECTION</b>	<i>Identification, history</i>
<b>CURATION</b>	<i>Curator, description</i>

<b>ARCHIVES</b>  Bibliography Documents Pictures Multimedia	<i>Identification</i> <i>Author</i> <i>Date</i> <i>Copyright</i> <i>Link to file</i>
--	--

partagées avec la flexibilité de la gestion personnelle d'une liste bibliographique liée à une thématique particulière ou à un article en cours de rédaction. Le module bibliographique de *MARS* permet l'importation et l'exportation vers de nombreux formats tels que *BibTex*, *Medline* ou *Endnote*. Il gère les références personnelles au sein d'une base de données collective et lie la référence au fichier *PDF* associé, privé ou public. Il autorise la génération de listes bibliographiques selon des profils de présentation prédéfinis ou personnalisés.

La recherche au sein du système *MARS* peut être effectuée par le biais d'un interface de recherche de « type *Google* » ou par une recherche avancée plus spécifique. L'ensemble des données implémentées dans le système *MARS* peuvent être recherchées, y compris le contenu des fichiers *PDF*, *Word* et *Excel*. La recherche peut utiliser le thésaurus pour les correspondances linguistiques et autorise de ce fait une recherche multi-langues.

## 6. Suivi et développement(s) du projet *MARS*

Le comité de suivi est composé de représentants des ESF partenaires (IRSNB, MRAH, MRAC), des universités (ULg, KUL, UG, VUB), des administrations régionales (Région wallonne) et des musées (Cedarc, Préhistosite de Ramioul, *Provinciaal Gallo-Romeins Museum*). Il peut suivre et intervenir dans le développement du système *MARS* afin de rencontrer les exigences de l'ensemble des utilisateurs potentiels. À la fin de la première phase (2005), les membres du comité de suivi remettront un avis relatif à l'adéquation de l'application développée par rapport à leurs besoins propres. Le cas échéant, ils pourront être intégrés à la deuxième phase du projet afin de participer au développement de *MARS*. Le système développé pourra être utilisé par l'ensemble des opérateurs travaillant dans le domaine de l'archéologie.

Il est essentiel d'attirer l'attention sur la volatilité du support électronique de l'information, l'inconvénient majeur de ce médium. Durant des siècles, l'information a été conservée sous forme manuscrite ou imprimée. L'apparition du média électronique décuple les capacités de stockage et de recherche, mais l'absence de support matériel persistant rend ce média excessivement fragile. La faillite d'une société privée hébergeant une base de données ou un site *web* entraîne souvent la disparition de la source d'information et de son contenu.

Le rôle du pôle public est ici essentiel, afin de garantir la pérennité de l'information et son accès par les différents publics cibles. Il est difficile de demander à une petite a.s.b.l. d'investir dans un matériel coûteux

afin de sauvegarder ses informations sur support électronique. Durant sa seconde phase, *MARS* testera la synchronisation et la sauvegarde entre les serveurs. Ainsi les collections archéologiques numérisées à l'aide du système *MARS* appartenant à des collections extérieures aux ESF pourront être sauvegardées via le système *MARS* sur un serveur central.

## 7. Le projet pilote

Nous testons le système *MARS* en grandeur nature à l'aide des collections provenant de la grotte de Spy (Jemeppe-sur-Sambre, Prov. de Namur). Celles-ci sont parfaitement représentatives des collections d'un site archéologique et de leur complexité à savoir la dispersion physique, la pluridisciplinarité et l'appartenance chronologique multiple.

La grotte de Spy est le site préhistorique le plus célèbre de Belgique et l'un des gisements paléo-anthropologiques de référence au niveau européen. Le gisement fut fouillé à de nombreuses reprises entre 1879 et 1980. Le produit de ces travaux est réparti entre diverses Institutions scientifiques et musées (Musées royaux d'Art et d'Histoire, Institut royal des Sciences naturelles de Belgique, Université de Liège, Université catholique de Louvain/Leuven ?, Musée archéologique de la ville de Namur et Musée Curtius de la Ville de Liège) sans compter les nombreuses collections privées provenant des prospections réalisées par des chercheurs amateurs. Le matériel préhistorique récolté se rapporte à toutes les périodes comprises entre le Paléolithique moyen et le Néolithique. La collection dépasse plusieurs dizaines de milliers d'objets (anthropologiques, archéologiques et paléontologiques) et comprend plusieurs milliers de documents d'archive (iconographiques et alphanumériques).

Cet éclatement des collections explique partiellement l'absence d'une étude à caractère monographique et récente sur le site de Spy, malgré la réputation mondiale et la très grande richesse de ce gisement. Le produit de certaines campagnes de fouilles est encore largement inédit (Rougier *et al.*, 2004). Seul un programme complet de numérisation peut permettre la création d'une collection « virtuelle » unique.

La numérisation complète des objets (photographies numériques, acquisitions 3D, photographies en relief) ainsi que celle des archives iconographiques et alphanumériques permettra la conservation du patrimoine mais aussi la réalisation d'une nouvelle étude de ce site remarquable. Un programme d'étude pluridisciplinaire de l'IrSnB (2004-2008) doit conduire à une réétude complète de ces collections. Ce programme de recherche est effectué en partenariat avec la Direction de l'Archéologie de la Division générale de l'Aménage-

ment du Territoire, du Logement et du Patrimoine (DGATLP) de la Région wallonne afin de réaliser les investigations et les analyses nécessaires à une meilleure compréhension du site et de ses collections.

Le choix de la collection de Spy s'inscrit également dans le calendrier scientifique international des prochaines années. En effet, le 150<sup>e</sup> anniversaire de la découverte de l'Homme de Néandertal à Feldhof (1856) sera commémoré en 2006 par le biais de plusieurs expositions et réunions scientifiques. Dans cette perspective, *TNT (The Neanderthal Tools)* (Semal et al., 2004), projet du 6<sup>ème</sup> programme cadre de l'Union européenne, a pour objectif de créer la plus grande base de données mondiale relative aux Néandertaliens en développant des outils multimédias et collaboratifs novateurs. L'IrSnB est partenaire du projet *TNT* qui regroupe 8 partenaires européens au sein d'un consortium qui allie expertise technologique, diffusion multimédia et accès à des collections néandertaliennes majeures. L'IrSnB assure la collecte des informations et la numérisation des collections néandertaliennes de Belgique.

### Remerciements

Nous tenons à exprimer nos remerciements à David Convent et Louis Wannijn qui jour après jour construisent la tour de Babel des collections archéologiques belges. Merci également à tous ceux qui au sein des ESF et du comité de suivi, favorisent la gestation intellectuelle du projet par leur attention et leurs suggestions.

### Références

ROUGIER H., CREVECOEUR I., FIERS E., HAUZEUR A., GERMONPRÉ M., MAUREILLE B. & SEMAL P., 2004. Collections de la *Grotte de Spy* : (re)découvertes et inventaire anthropologique. *Notae Praehistoricae*, 24 : 181-190.

SEMAL P., KIRCHNER S., MACCHIARELLI R., MAYER P. & WENIGER G. C., in press. *TNT : The Neanderthal Tools*. In : K. Cain, Y. Chrysanthou, F. Niccolucci and N. Silberman (eds), *The 5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST(2004) - Short Presentations* : 2 p.

#### Zope

main site : <http://zope.org>

zope book : [http://www.plope.com/Books/2\\_7Edition](http://www.plope.com/Books/2_7Edition)

#### Plone

main site : <http://plone.org>

plone book (The Definitive Guide to Plone by Andy McKay - covers Plone 2.0+) :

<http://his.biologie.hu-berlin.de/PloneBook>

code and errata : <http://plone-book.agmweb.ca>

#### Archetypes

Documentation : <http://plone.org/documentation/archetypes>

#### TNT - The Neanderthal Tools

main site : <http://www.the-neanderthal-tools.org>

#### Glossaire des principaux termes techniques utilisés

Apache : Serveur HTTP (protocole de transfert des données sur le web) gratuit et *Open Source*.

Archetypes : Couche logicielle permettant de créer rapidement des nouveaux types de contenu adaptés à l'environnement *Plone*.

Champ : Partie définie par l'utilisateur ou le concepteur d'un programme, qui est réservée à l'enregistrement d'informations spécifiques dans une base de données.

CMF : *Content Management Framework*, surcouche de *Zope* intégrant les fonctions de base d'un CMS, comme la gestion intégrée des états de publication (*workflow*) et la prise en charge des méta-données définies par le *Dublincore*. *Plone* est considéré comme une surcouche du CMF dont il exploite les fonctionnalités.

CMS : *Content Management System*, système de gestion de contenu (au sens large).

BiBTeX : Format de sauvegarde de données bibliographiques. Extension du langage de mise en page *LaTeX*.

Dublincore : Le *Dublin Core* est un set de méta-données.

HTML : *Hyper Text MarkupLanguage*, langage à balises à l'aide duquel sont écrites les pages consultables par un navigateur web.

LaTeX : Langage de mise en page de très haut niveau, particulièrement utilisé pour ses capacités de mise en forme de formules mathématiques ou physiques.

MEDLINE : la base de donnée de la Bibliothèque Nationale de Médecine des USA qui contient plus de 11 millions de références bibliographiques d'articles de journaux dans le domaine des sciences médicales.

Objet : Ensemble unitaire d'informations, créé à l'aide d'une application, qu'on peut insérer dans un autre document et qu'on manipule au moyen de ses propriétés. En programmation orientée objet, module structuré et limité constitué d'un ensemble de données et de procédures (appelées *méthodes*) qui décrivent leur manipulation.

Open Source : Principe par lequel un développeur de logiciel laisse le code source du programme accessible à d'autres programmeurs, et soumet son travail à une licence publique.

Plone : Système de gestion de contenu basé sur Zope et le CMF.

Python : Langage portable, dynamique, extensible, gratuit et qui permet (sans l'imposer) une approche modulaire et orientée objet de la programmation.

SBDD : Système de base de données.

SQL : *Structured Query Language*, langage standard pour adresser des requêtes à une banque de données relationnelle.

SSL : *Secured Sockets Layer*, protocole qui permet la sécurisation des communications en réseau en exploitant un système de cryptage développé en *open source*.

Variable : Entité symbolique à laquelle une valeur peut être attribuée.

W3C : *World Wide Web Consortium*, le consortium qui définit les standards de l'*internet* et les meilleurs pratiques à adopter pour les manipuler.

Web : *Word Wide Web* (WWW), la toile constituée de l'ensemble des pages consultables sur l'*internet*.

Workflow : Technologie qui permet de suivre le statut d'objets/documents dans un système informatique. Au sens de *Plone*, structure adaptée à la gestion des documents (des types de contenu) au sein d'un processus de publication.

XML : Langage permettant de concevoir la structure des données qu'un document doit contenir à l'aide de balises.

ZODB : *Zope Object Data Base*, banque de données de Zope (orientée objet).

ZOPE : Serveur d'application Web. Logiciel *Open source* écrit en *python* et largement extensible à l'aide de modules.

ZPT : *Zope Page Template*, outil de génération de pages web dynamiques (dont le contenu est généré en fonction du contexte d'affichage).

Patrick Semal  
Section Anthropologie et Préhistoire  
Institut royal des Sciences naturelles de Belgique  
29, rue Vautier  
BE - 1000 Bruxelles (Belgique)  
patrick.sem@naturalsciences.be

Els Cornelissen  
Afdeling Prehistorie en Archeologie  
Koninklijk Museum voor Midden-Afrika  
13, Tervuurse steenweg  
BE - 3080 Tervuren (Belgique)

Nicolas Cauwe  
Musées royaux d'Art et d'Histoire  
10, Parc du Cinquantenaire  
BE - 1000 Bruxelles (Belgique)