

ARCHEOLOGIA E CALCOLATORI

19

2008

ESTRATTO

All'Insegna del Giglio

VERS LA MISE EN RÉSEAU DES DONNÉES ET DES CHERCHEURS: LE SYSTÈME D'INFORMATION DE LA PROSPECTION D'ITANOS (CRÈTE ORIENTALE)

1. INTRODUCTION

La prospection d'Itanos fait partie du programme de recherches archéologiques mené à Itanos depuis 1994 sous l'égide scientifique de l'École française d'Athènes et de l'Institut d'Études Méditerranéennes de Réthymno, en collaboration avec la XXIV^e Éphorie des antiquités préhistoriques et classiques (Agios Nikolaos)¹.

Nous ne nous attarderons pas ici sur l'ensemble de la chaîne opératoire ayant présidé à l'inventaire des vestiges archéologiques de la péninsule d'Itanos. Chaque prospection archéologique est différente tant par les conditions naturelles du terrain, que par les objectifs poursuivis. En l'occurrence, l'objectif principal de la prospection était de retracer l'occupation humaine d'Itanos à travers les âges. Il s'agissait tout d'abord d'inscrire l'histoire de la cité grecque du I^{er} millénaire avant J.-C. dans son cadre territorial. Les recherches menées dans la *chôra* d'Itanos ont également mis en évidence une occupation humaine depuis le Néolithique final, avec une phase d'intense activité au cours de l'époque minoenne néopalatiale (XVI^{ème}-XV^{ème} siècles).

Le territoire prospecté permettait donc d'appréhender les modalités d'occupation et d'exploitation des sols à travers les siècles au sein de systèmes politiques et administratifs très différents (minoën, gréco-romain, byzantin). L'aire de prospection couvre environ 20 km², divisée, pour la commodité de l'enquête, en onze zones correspondant aux toponymes modernes. L'abandon complet de ce territoire au cours du VII^e siècle après J.-C., l'absence de perturbation médiévale ou moderne et l'état exceptionnel de conservation des vestiges au sol donnent au paysage prospecté l'apparence d'une carte fossile des établissements antiques. Nous avons donc adopté une méthode de prospection dite extensive, où la définition du site repose essentiellement sur la présence de structures bâties – en bref, des “ruines” – et non sur un seuil de concentration de tessons repérable sur des cartes de densité de distribution céramique.

L'un des objectifs de la publication des résultats de la prospection est d'offrir aux chercheurs un instrument de travail et de diffusion de l'information

¹ Pour une description complète du programme, voir A. DUPLOUY, A. SCHNAPP, A. SCHNAPP-GOURBEILLON, Chr. TSIGONAKI, D. VIVIERS, *Recherches archéologiques à Itanos (Crète orientale)*, «Revue Archéologique», 2008, sous presse.

à la fois commode et complet, par la mise en ligne sur Internet de l'ensemble des données de la prospection d'Itanos: description, photographies, dessins, cartes.

2. CRÉATION D'UN SITE INTERNET

L'ensemble des informations récoltées sur le terrain a d'abord été organisé sous la forme d'une base de données conçue sous FileMaker. Peu évolutive, peu adaptée à la gestion commode de documents graphiques, la solution n'était guère efficace, surtout en terme de partage de l'information entre des collaborateurs dispersés à travers toute l'Europe. L'un des objectifs a donc été de rendre cette base de données rapidement accessible sur Internet. Grâce à la collaboration d'un ingénieur informaticien, un site web a ouvert en novembre 2006 sur le serveur de l'École française d'Athènes: <http://webefa.efa.gr/prospection-itanos/>².

Bien plus qu'une simple présentation de la mission et de ses objectifs, comme il en existe de nombreux exemples, ce site avait pour ambition de fournir à la communauté scientifique, dès l'achèvement des opérations de terrain et sans attendre la publication finale dans la collection des *Études crétoises* de l'École française d'Athènes, un état des lieux complet de la documentation récoltée, mais aussi un véritable instrument de recherche pour les archéologues de l'équipe. Le site comprend en effet deux parties, l'une publique accessible à tout internaute, l'autre privée servant d'interface de rencontre et de travail pour les archéologues de la mission.

2.1 *Partie publique: les sites archéologiques*

Une première base de données concerne les sites archéologiques de la prospection. Sa consultation est ouverte à tout internaute. Chaque fiche de site rassemble les données disponibles, notamment une chronologie détaillée des phases d'occupation du site, une indication de sa (ou ses) fonction(s) probable(s), une brève description archéologique des vestiges visibles au sol, ainsi que les vignettes de toutes les photographies disponibles (aériennes ou au sol). Les coordonnées GPS des sites, exprimées selon le système géodésique grec EGSA87, servent à localiser précisément ceux-ci dans l'espace et sur nos cartes; éléments descriptifs importants, les coordonnées ne sont toutefois pas affichées, afin de se prémunir contre tout pillage intempestif des vestiges.

L'utilisateur a la possibilité de consulter les fiches des sites à travers un formulaire de recherche, combinant éventuellement plusieurs critères dans

² Sa réalisation par Franck Eyraud a bénéficié du soutien du Ministère des Affaires Étrangères (Paris), de l'Institute for Aegean Prehistory (INSTAP), ainsi que de l'École française d'Athènes.

les multiples champs de données ouverts au tri. À partir de l'unique critère proposé dans le formulaire initial, il est possible d'ajouter (ou de retirer) tous les critères nécessaires à la formulation de la recherche complète, y compris dans le même champ de données. Selon le choix de l'utilisateur, les résultats correspondront à tous les critères demandés (opérateur booléen ET, par défaut) ou à au moins un des critères requis (opérateur booléen OU). Le chercheur peut ainsi obtenir la liste de tous les aménagements agricoles de l'époque romaine, de tous les sites de défense du Minoen ancien, de tous les habitats à la fois minoens et hellénistiques, etc.

Des recherches plus complexes, associant plusieurs critères au moyen d'opérateurs booléens différents, nécessitent en revanche d'autres solutions; elles ne correspondent du reste pas aux besoins essentiels du chercheur. Pour simplifier l'utilisation du site et afin d'éviter une syntaxe de recherche compliquée (conforme à la syntaxe booléenne, comme exprimée par exemple sous UNIX), l'utilisateur expérimenté est donc invité à décomposer sa recherche complexe en plusieurs requêtes simples. C'est pour cette raison notamment qu'il est possible d'affiner une recherche, en opérant un nouveau tri sur les résultats obtenus.

Pour le moment, selon la stratégie de prospection adoptée et la notion même de "site" retenue, l'inventaire des sites archéologiques comprend une centaine de fiches. Les champs ouverts au tri sont les suivants: zone, numéro du site, chronologie, fonction et description. Ils se présentent, selon la nature des données, sous la forme de menu déroulant ou d'une zone de texte libre.

Le résultat du tri s'affiche sous la forme d'une liste de sites correspondants aux critères de recherche, chacun des noms donnant accès à la fiche complète du site, en mode consultation uniquement. Sur demande de l'utilisateur, le résultat de la recherche est également proposé sous la forme d'une carte schématique de l'aire prospectée sur laquelle sont localisés (au moyen de leurs coordonnées) les sites correspondant aux critères de recherche. L'interface génère en effet à la volée toutes les cartes archéologiques imaginables par interrogation de la base de données: l'organisation du territoire italien, l'évolution de la distribution des établissements humains ou encore les modalités d'utilisation des sols sont ainsi immédiatement cartographiables. L'échelle relativement petite de la carte obtenue ne permet cependant pas une localisation précise sur le terrain, afin d'éviter les problèmes liés au pillage archéologique.

Toutes les informations proposées en ligne reflètent l'état d'avancement de l'étude. Elles ne peuvent donc actuellement être tenues pour définitives et seront mises à jour régulièrement jusqu'à la publication finale dans la collection des *Études crétoises*. La chronologie des sites, en particulier, repose sur la chronologie de la céramique récoltée sur chaque site, dont l'étude est toujours en cours.

2.2 Partie privée: le matériel archéologique

À côté d'un accès public, le site web comprend également une partie privée, accessible aux seuls utilisateurs référencés (archéologues et dessinateurs de la mission) au moyen d'identifiants personnels (login et password), leur donnant un accès à l'ensemble des fonctions programmées: consultation, édition, création et suppression des fiches de la base de données. Cette partie du site web sert d'interface de rencontre et de travail aux membres de la mission, qui sont dispersés à travers toute l'Europe, ce qui se révèle particulièrement utile et efficace dans la phase actuelle d'étude du matériel. Cette partie privée donne un accès complet aux informations relatives aux sites archéologiques, aux fiches des tessons, ainsi qu'à un ensemble intermédiaire (dénommé US), qui correspond à la fois à une unité de ramassage et de stockage du matériel³.

La répartition du matériel s'étant faite en fonction des spécialités chronologiques de chacun, chaque archéologue est propriétaire de son matériel. Des droits d'accès spécifiques sont donc associés aux identifiants personnels, tandis qu'un administrateur général de la base de données a accès à tous les champs. Utilisateurs spécifiques, les dessinateurs ont accès (en mode édition ou création) au dessin de chaque tesson, mais pas aux données associées. Notons enfin qu'en cas de réévaluation du matériel (ce qui n'est pas nécessairement rare), il n'est pas possible de s'attribuer un tesson sans le consentement de son propriétaire; en revanche, ce dernier a la possibilité de s'en dessaisir au profit d'un autre utilisateur référencé, soit parce que celui-ci le réclame soit parce qu'il est certain qu'il n'appartient pas à sa période. La communication entre archéologues est ainsi encouragée.

Pour chaque tesson se trouvent rassemblées diverses informations utiles (l'année de ramassage et lieu de stockage), sa description détaillée (forme et argile), ses dimensions, son association à une catégorie de matériel, son origine locale ou importée, une chronologie générique, ainsi qu'une date plus précise, reposant sur des parallèles. Pour la commodité du classement, il était en effet nécessaire de distinguer entre une chronologie générique (limitée à des périodes déterminées, plus ou moins larges, cfr. *infra*) et une datation spécifique – et, partant, trop variable pour faire l'objet d'une liste d'autorité. Le dessin de chaque tesson est directement visualisable sur Internet par l'intermédiaire d'une vignette cliquable, permettant de voir rapidement le dessin du tesson (au format.png) dans le navigateur. Les archéologues peuvent du reste télécharger

³ Plusieurs ramassages sélectifs de céramique ont été effectués sur chaque site par des archéologues différents, spécialisés dans l'une ou l'autre catégorie de matériel. Par ailleurs, pour des raisons logistiques et administratives, le matériel a été entreposé, essentiellement selon l'année de ramassage, soit au musée de Sitia soit à l'apothèque de Palaikastro. L'US, à la fois unité de ramassage et de stockage, a donc une valeur logistique, mais également scientifique: les US associées à un même site archéologique peuvent notamment avoir des chronologies radicalement différentes, étant le fruit d'une collecte sélective par divers spécialistes.

le dessin vectoriel (au format Adobe Illustrator), le modifier puis le recharger eux-mêmes sur le site web. Les commandes sont simples et les archéologues gèrent toute la procédure eux-mêmes, en collaboration avec les dessinateurs. Le stockage des données sur le serveur de l'École française d'Athènes est du reste automatique.

Chaque fiche de tesson peut être éditée par son propriétaire et toutes les informations (y compris le dessin) modifiées. Par la même interface, il est également possible de créer de nouvelles fiches, voire d'en supprimer (sans possibilité de retour en arrière, car aucune fonction de sauvegarde n'est prévue dans le programme: un message d'avertissement apparaît alors). Dans le cas d'une création, le propriétaire par défaut de la nouvelle fiche est l'utilisateur connecté; s'il s'agit d'un dessinateur, qui crée une fiche inexistante pour archiver son dessin, celui-ci doit de préférence se dessaisir immédiatement de sa propriété et attribuer le tesson au responsable de son étude. Un tesson, existant ou nouvellement créé, doit par ailleurs être obligatoirement associé à une unité de ramassage/stockage (US) précise et à un site référencé dans la base de données. Il n'est donc pas possible de créer, par mégarde, la fiche de n'importe quel tesson. Seul l'administrateur général de la base de données a la possibilité de rajouter un site archéologique ou des US spécifiques, ouvrant ainsi la voie à l'ajout de certains tessons. La création de doublon est du reste impossible.

Comme pour les sites archéologiques, l'utilisateur référencé a la possibilité d'accéder aux tessons – le matériel actuellement enregistré compte plus de 1000 fiches – à travers un formulaire de recherche, combinant éventuellement plusieurs critères dans les multiples champs de données ouverts au tri. La sélection se fait principalement sur la forme, la description, les dimensions, l'origine, la chronologie, le propriétaire des tessons, ou tout simplement sur leur numéro. Le résultat du tri s'affiche sous la forme d'une liste synthétique de tessons, présentant d'emblée leur chronologie générique et donnant accès à la fiche de chacun d'eux. Un renvoi automatique à l'ensemble des sites archéologiques associés à cette céramique est proposé, de manière à ce que le résultat d'un tri sur l'inventaire des tessons devienne lui aussi immédiatement cartographiable. Réciproquement, une fois identifié, l'utilisateur référencé a accès, depuis la fiche d'un site, à tout le matériel céramique qui lui est associé. Il en va de même pour les photographies du site proposées en pleine définition, alors que pour se conformer à la législation archéologique grecque seules des vignettes sont accessibles dans la partie publique.

Une fois tous les tessons étudiés, leur répartition chronologique générale permettra aux archéologues de préciser définitivement la chronologie de chaque site archéologique. Bien qu'une automatisation ait été envisageable (la chronologie d'un site résultant de l'addition des occurrences chronologiques spécifiques de chaque tesson ramassé sur celui-ci), l'intervention humaine,

supposant une interprétation des données, a été jugée préférable. Ce travail fournira donc, à terme, l'inventaire définitif du matériel archéologique, ainsi qu'une partie de l'interprétation finale, comme les phases chronologiques d'occupation du territoire.

En somme, la base de données en ligne est d'abord un lieu d'archivage commun de toutes les informations récoltées sur le terrain et issues de l'étude du matériel. Chacun des membres de l'équipe peut ainsi bénéficier, en temps réel, des derniers résultats des autres. À terme, tout archéologue intéressé par le projet pourra prendre connaissance de ces données. Il est en effet prévu d'ouvrir à tout internaute (éventuellement après enregistrement) l'essentiel de la base de données en même temps que la publication des résultats dans la collection des *Études crétoises*. L'étude des données sera alors terminée et les résultats tenus, de notre point de vue, pour définitifs. Les avantages d'une telle publication informatique sont nombreux. Les possibilités de tri offertes par une base de données informatique sont presque infinies et les champs ouverts au tri, tant sur les sites que sur les tessons, sont nombreux. À l'inverse, une publication papier traditionnelle impose un classement unique du matériel, généralement par site ou, pour faire écho au travail des céramologues, par typochronologie. Le projet de publication Internet des données de la prospection s'inscrit ainsi dans une démarche de renouvellement des modes traditionnels de publication et de diffusion de l'information scientifique.

3. CHOIX TECHNIQUES ET PROGRAMMATION DU SITE INTERNET

Le schéma structurel des données de la prospection d'Itanos (Fig. 1) est le suivant: l'aire prospectée comprend onze zones, correspondant à des secteurs de prospection; chaque zone a livré une série de sites; chaque site comprend plusieurs unités de ramassage/stockage de la céramique (US); et chaque US comprend plusieurs tessons. Chacune de ces entités (site, US et tesson) est par ailleurs caractérisée au moyen de plusieurs informations.

La structure de ces informations correspond au schéma caractéristique d'une base de données relationnelle (Fig. 2), qui permet de lier différentes classes d'objets entre elles (site, US, tesson), chacune étant décrite de manière complète au moyen de multiples descripteurs ou de liens vers des objets communs (zone, chronologie, fonction, propriétaire). Ainsi, un tesson est défini par une localisation, une chronologie et une US uniques; une US appartient à un seul site, et un site à une seule zone; tandis que plusieurs photographies, fonctions ou chronologies peuvent être attachées à un site.

D'un point de vue technique, les choix principaux ont été l'open source et la programmation générique. La politique du logiciel libre nous a paru offrir de multiples avantages. La gratuité des logiciels – certes tempérée par les coûts d'implémentation spécifique – permet tout d'abord d'éviter

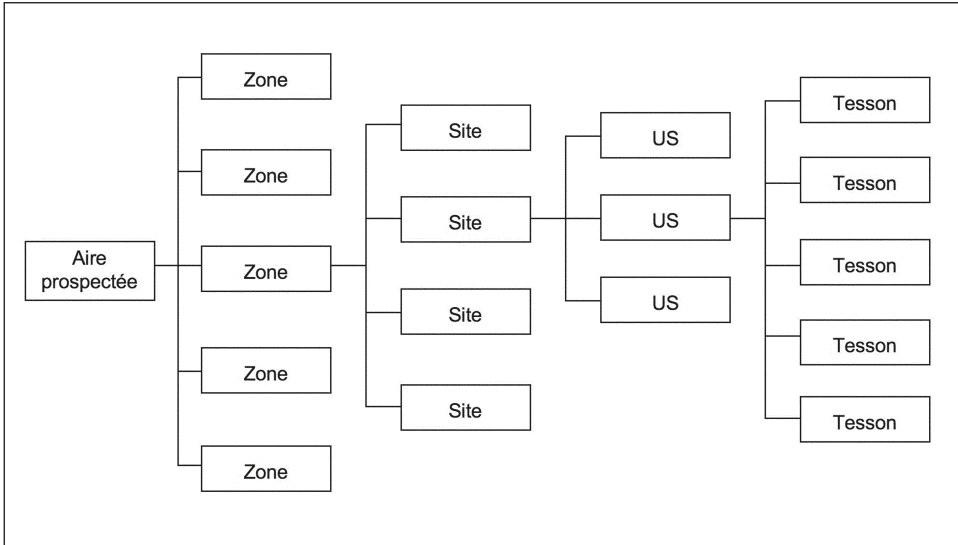


Fig. 1 – Schéma structurel des données.

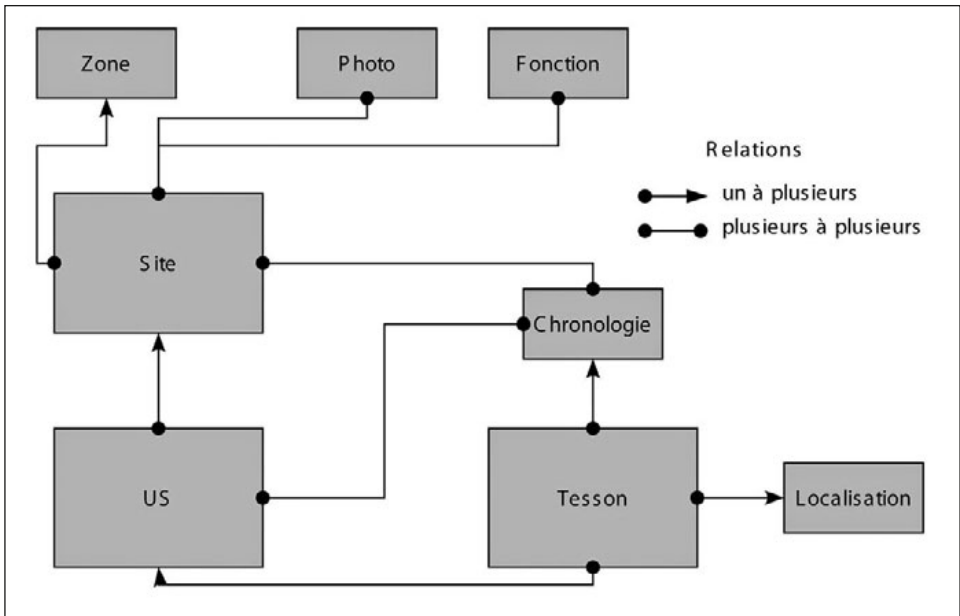


Fig. 2 – Structure de la base "métier".

des dépenses importantes, de surcroît récurrentes en raison des nécessaires mises à jour, liées à l'utilisation de solutions propriétaires. Mais surtout, les intégrations de logiciels propriétaires sont plus difficiles et leur évolution dépend essentiellement du marketing de la société. À l'inverse, les solutions open source offrent une grande interopérabilité. Dans la mesure où les algorithmes utilisés pour la création d'un programme sont directement visibles, la programmation libre permet une grande ouverture des logiciels. La taille de la communauté open source assure du reste une robustesse aux logiciels, qui ont été intensément testés.

Dans le cadre d'une application web, la combinaison de produits la plus largement utilisée est la trilogie serveur web Apache, interpréteur php, base de données MySQL:

- Apache répond aux requêtes reçues des navigateurs qui accèdent à une adresse gérée par le site et décide qu'en faire. Si la requête demande seulement un fichier présent sur le disque, Apache lit ce fichier et le renvoie par le réseau. Si en revanche l'utilisateur souhaite accéder à une page dite "dynamique", c'est-à-dire programmée, alors entre en compte l'interpréteur php;
- php est le langage de programmation dans lequel est écrit le programme. L'interpréteur exécute donc ce programme et génère une page en fonction des paramètres reçus; il fournit le résultat à Apache qui le renvoie au navigateur.
- MySQL gère les données selon le principe des bases de données relationnelles avec SQL comme langage de requête. L'interpréteur php communique avec MySQL par l'intermédiaire d'une librairie d'interface et lui demande les données nécessaires à l'affichage des pages.

Ces trois composantes sont situées sur un même serveur physique, situé à l'École française d'Athènes, ce qui simplifie les opérations de maintenance. Des sauvegardes régulières permettent de se prémunir contre les pertes accidentelles de données; il est nécessaire de copier le contenu de la base de données MySQL où se situe le contenu des fiches, mais également les fichiers et photos associés à ces fiches et qui ont été envoyés par les utilisateurs du programme (ce type de ressource ne pouvant pas être stocké dans un système de base de données).

En eux-mêmes, les programmes Apache, php et MySQL ne fournissent que les outils pour faire une application de n'importe quel type. Pour construire celle-ci, il a été nécessaire d'effectuer une programmation pour la mise en place des fonctionnalités désirées (gestion des utilisateurs, gestion des données, affichage, etc.). Plutôt que d'écrire une application spécifique aux problématiques de la prospection d'Itanos, nous avons opté pour une programmation de type générique. On entend par "programmation générique" le fait de prévoir toujours le cas général au moment de mettre en place une

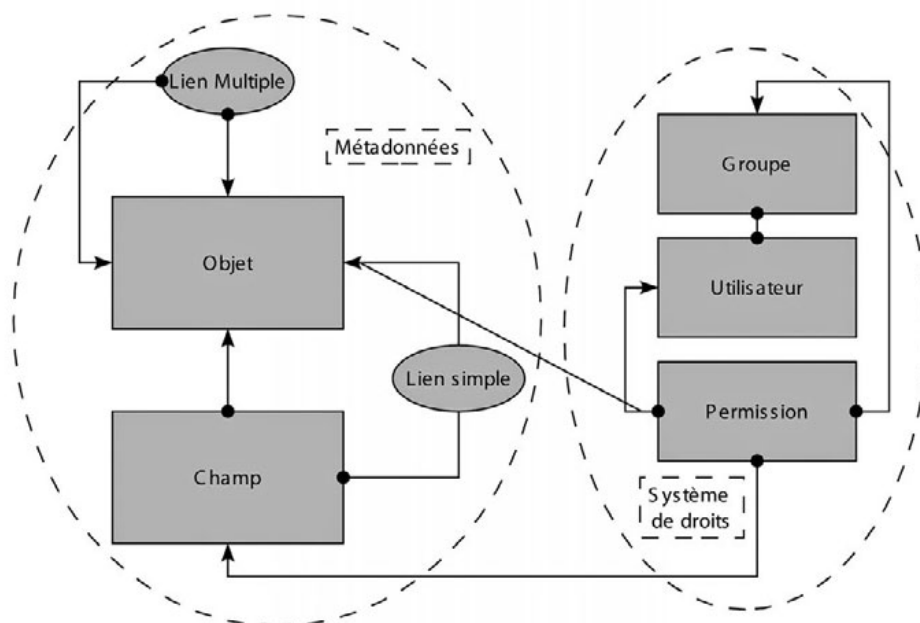


Fig. 3 – Structure de la base "système".

fonctionnalité et pas seulement le cas précis qui nous intéresse. La généralité d'un programme repose sur son indépendance vis-à-vis du type et du nombre des données manipulées. Ce choix de programmation résulte du constat que les applications de gestion de bases de données sont toujours construites sur le même modèle et mettent en place les mêmes fonctionnalités (recherche et consultation de fiches, édition des données, etc.).

De là, l'idée d'écrire un moteur de gestion des données, qui serait paramétrable en fonction des données utilisées – un peu comme l'a fait FileMaker dans le monde du logiciel propriétaire, selon une architecture malheureusement fermée aux développeurs externes. D'une base de données à l'autre, les principes de chargement et d'insertion des données sont en effet communs; seuls les types de données manipulés changent. Ainsi, ce programme écrit pour la présentation des résultats de la prospection d'Itanos pourrait être réutilisé pour les besoins d'une autre mission archéologique ou de tout autre projet.

On distinguera d'une part la base "métier" (Fig. 2) qui contient les données scientifiques directement utiles à l'application et, d'autre part, la base "système". Cette dernière (Fig. 3) contient les données non spécifiques à l'application et se découpe en deux parties:

- les métadonnées (littéralement “information sur les données”) qui permettent une représentation de la structure de la base “métier”;
- les données d’utilisateurs et de droits d’accès qui gèrent le contrôle des permissions et de propriété sur les données.

Les métadonnées contiennent la description des différentes tables utilisées pour représenter les données “métier”. Ainsi, pour le site web de la prospection d’Itanos, la structure des métadonnées de la Fig. 3 est remplie par la description de la Fig. 2 (base de données “métier”). La structure des métadonnées suit le fonctionnement d’une base de données relationnelle classique, soit le mode de représentation de données le plus diffusé et utilisé par MySQL. Chaque objet possède plusieurs champs et est associé à une table de la base de données. Chaque champ est défini par un type de données (entier, texte, mot de passe, fichier, photo, lien) et est associé à une colonne de la table de l’objet qu’il décrit. Par exemple, l’objet “site” contient les champs “numéro du site”, “zone”, “description”, “coordonnées” (X et Y). S’ajoutent aux champs simples une série de liens entre objets; ceux-ci sont soit des “liens simples” (relation un à plusieurs), qui sont dès lors associés à un champ contenant une clé étrangère vers l’objet pointé, soit des “liens multiples” (relation plusieurs à plusieurs), qui sont représentés en base de données relationnelle par une table supplémentaire de deux colonnes contenant chacune une clé étrangère vers les deux objets associés.

Un exemple dans notre cas de lien simple est celui de la zone dans laquelle se situe un site: un site n’est associé qu’à une seule zone, mais une zone peut contenir plusieurs sites. La zone devient donc un paramètre d’un site. Le champ “zone” de l’objet “site” est donc défini comme un lien simple vers l’objet “zone”; la colonne associée à ce champ contient la référence de la zone sous la forme d’une clé étrangère.

Un exemple de lien multiple est celui des fonctions associées à un site: un site peut avoir plusieurs fonctions et, bien sûr, une même fonction peut être associée à plusieurs sites. La base des métadonnées contient donc la définition d’un lien multiple qui associe l’objet “site” à l’objet “fonction” par l’intermédiaire de la table supplémentaire. Le programme est donc capable d’aller chercher dans cette table la liste de toutes les fonctions associées à un site, mais également la liste des sites possédant une fonction donnée.

Grâce aux métadonnées, le programme est capable de générer les pages contenant la liste des différents objets, ainsi que chaque fiche d’objet en détail; il gère en outre automatiquement les modes édition, ajout et suppression. Ces métadonnées sont du reste utilisées pour proposer un formulaire de recherche, avec pour chaque champ des opérateurs propres à son type. En poussant jusqu’au bout la généralité, nous avons également saisi dans la base des métadonnées la description de ces mêmes métadonnées – en d’autres termes inséré la description de la Fig. 3 dans la structure de la Fig. 3 – ce qui au final

a permis la définition des métadonnées propres à la prospection d'Itanos en utilisant l'interface graphique de l'application.

Enfin, le système de gestion des utilisateurs contient les informations relatives à l'identification des utilisateurs (identifiant et mot de passe), aux groupes d'utilisateurs (un groupe contient plusieurs utilisateurs et, à l'inverse, un utilisateur peut appartenir à plusieurs groupes), et aux permissions. Les droits d'accès sont accordés (ou refusés) à un groupe ou à un utilisateur, et concernent l'accès en lecture/écriture/suppression d'une catégorie d'objet, d'un champ ou d'un objet précis: par exemple, droits de lecture de tous les sites, droits de modification uniquement sur certains sites, droits de modification uniquement d'un champ particulier d'un tesson (description, image, dimensions, etc.).

Les principales fonctionnalités programmées dans le cadre du projet Itanos sont les suivantes:

- Gestion des actions sur les objets de la base (consultation, ajout, modification, suppression).
- Gestion des utilisateurs et de l'authentification. Des permissions d'accès et d'action spécifiques sur chaque type d'objets et de champs sont associées à chaque utilisateur enregistré. Par défaut, l'utilisateur non enregistré n'a accès qu'aux fiches des sites en mode consultation par le biais de l'interface de recherche.
- Module de recherche. Les champs ouverts au tri sur les fiches (site, US, tesson) correspondent aux données associées à ces trois objets principaux. Ils se présentent, par référence aux métadonnées, sous la forme de menu déroulant ou d'une zone de texte libre. Pour se conformer aux principes de la programmation générique, la création de la requête SQL correspondant à la recherche de l'utilisateur se fait en utilisant une série d'objets php. Ces objets transforment les critères de l'utilisateur en requête SQL en fonction de la structure de la base de données telle que décrite par les métadonnées. Ainsi, le code SQL n'est pas disséminé un peu partout dans le code source – comme c'est souvent le cas dans les applications de gestion de données – mais est centralisé dans le module qui gère la recherche dans la base.
- Gestion des fichiers et des photos. Pour associer une image ou un fichier à un objet, un mécanisme de chargement de fichier par l'intermédiaire du navigateur a été mis en place, avec création automatique d'une vignette dans le cas des images. Le dessin vectoriel des tessons, enregistré dans un format propriétaire (Adobe Illustrator), est considéré comme un fichier; à charge de l'utilisateur d'y associer une image (format.png) visualisable par le navigateur et traitable par le programme.
- Programmation d'une catégorisation en arborescence pour la chronologie générique. La caractérisation des fiches (site, US, tesson) en terme chronologique a révélé la nécessité d'une représentation de la chronologie par arbre. En effet, certaines chronologies sont précises, alors que d'autres sont plus

larges: un tesson “classique” doit aussi être considéré comme “gréco-romain”. “Gréco-romain” est donc appelé chronologie parente de la chronologie fille “classique”⁴. Lors de l’ajout d’une chronologie fille par l’utilisateur, la chronologie parente est elle aussi automatiquement enregistrée. De même, lors de la recherche d’un objet “gréco-romain”, tous les objets associés aux chronologies filles (géométrique, archaïque, classique, etc.) sont automatiquement inclus dans les résultats du tri. La mise en place de la spécificité de cette problématique a donné lieu à un développement générique, qui pourra être adapté tout autre cas de catégorisation par arborescence.

– Webmapping: La programmation de cet élément a eu recours à la librairie GD, qui permet de manipuler des images par de la programmation. Le principe est de manipuler un objet image sur lequel on effectue diverses opérations (charger une image d’un fichier, y tracer des lignes/cercles/rectangles en choisissant la couleur, ajouter du texte, etc.), avec en général comme opération finale l’exportation de l’objet en un fichier image pouvant être affiché dans un navigateur. Ainsi, pour la génération de la carte des sites correspondant aux critères du tri, le module charge le fond de carte (une image .png de 638×560 pixels), y place les points correspondant aux sites en opérant une corrélation entre les coordonnées GPS et le quadrillage en pixels de l’image, et associe le numéro de chaque site pour la visualisation. Un titre est ajouté en dessous de l’image.

Depuis la création de notre site web à l’automne 2006, plusieurs projets open source ont progressé dans la même direction de la programmation générique. Ils utilisent pour la plupart le mode d’abstraction MVC (Modèle, Vue, Contrôleur), qui permet de séparer la gestion des données (Modèle), le traitement spécifique à l’application (Contrôleur) et la présentation à l’utilisateur (Vue, c’est-à-dire la mise en page). Il s’agit notamment de CakePHP, de Php-On-Trax ou de CodeIgniter. Programmés en langage php, ils sont tous basés sur un projet un peu plus ancien nommé Ruby-On-Rails, référence en la matière, qui utilise en revanche le langage Ruby, moins commun.

Très activement soutenus par une large communauté, ces projets se trouvent actuellement en mesure d’offrir les mêmes fonctionnalités que celles développées en interne pour la prospection d’Itanos et ils ont à présent suffisamment mûri, semble-t-il, pour être directement utilisés comme base d’une application similaire à la nôtre. Ils présentent divers avantages comme certains inconvénients.

⁴ Les chronologies suivantes sont proposées à l’utilisateur: Indéterminée, Pré-/Protohistorique (Néolithique final, Minoen ancien, Minoen moyen, Minoen récent), Gréco-romain (Géométrique, Archaïque, Classique, Hellénistique, Romain, Romain tardif), Médiéval et Moderne. Chacune de ces chronologies parentes et filles correspond à une période spécifique, précisée dans les conventions du site Internet.

Tout d'abord, une grande partie des fonctionnalités génériques que nous avons développées sont présentes, à l'exception du formulaire de recherche automatiquement généré à partir de la structure de la base. Un de leurs points forts est du reste d'utiliser directement comme base de métadonnées la description des tables fournies par le moteur de base de données, là où nous avons choisi pour le projet Itanos de gérer une description redondante à celle de MySQL. Ce choix était délibéré et présentait certaines facilités: rentrer dans la façon spécifique dont MySQL représente ces données aurait demandé plus de temps que de définir notre propre structure.

Tous ces projets conservent néanmoins un inconvénient de taille pour le développement d'une application concrète: ils sont d'abord construits pour simplifier la vie des programmeurs, plus que comme des outils directement accessibles à des personnes ne connaissant pas la programmation. L'objectif de détacher l'archéologue, administrateur scientifique de la base de données, de tout souci d'implémentation n'est pas encore atteint.

4. CRÉATION D'UN SYSTÈME D'INFORMATION GÉOGRAPHIQUE

La mise en place d'une base de données géographique a fait suite à une première série de travaux de cartographie engagés dans le cadre du projet. Chaque site a été, lors sa découverte, localisé géographiquement à l'aide d'un GPS. Les structures les mieux conservées ont bénéficié d'un lever topographique plus complet. Ces informations ont été associées à un fond de carte numérique vectorisé par l'Institute of Mediterranean Studies. L'intégralité de l'information était alors gérée à l'aide d'un logiciel de CAO (Minicad) permettant de visualiser de manière interactive la répartition dans l'espace des différentes informations récoltées durant les phases de terrain.

Le développement d'un véritable système d'information géographique s'inscrit donc comme une suite logique de ces premiers travaux qu'elle reprend et développe. La mise en place de cet outil vise à:

- Développer une base de données géographique permettant de structurer et de gérer les différentes expressions des informations archéologiques collectées durant les années de prospection. Cette base de données doit ouvrir sur un outil d'interrogation et de production documentaire avancé permettant d'aboutir à des productions cartographiques de synthèse destinées aux chercheurs. Ces productions permettront d'aboutir à une modélisation de l'occupation du territoire.
- Permettre une analyse spatiale avancée.
- Aboutir à la mise en ligne de l'ensemble de ces données et cela à la fois au niveau des synthèses que des données archéologiques de base. L'idée est de proposer un outil de consultation et d'interrogation dynamique de l'ensemble du corpus de données permettant de produire sa carte à la carte (cfr. Atlas du patrimoine).

4.1 *La documentation en présence*

Les cartes au 1:5000 fournies par l'Institut d'Études Méditerranéennes (IEM) ont constitué le point de départ de cette opération. Leur vectorisation et la saisie des valeurs d'altitude associées ont permis la réalisation d'un modèle numérique de terrain (MNT) d'une précision importante permettant de travailler à l'échelle de l'inter-site (pixel de 4 m).

Des missions aériennes verticales de l'armée grecque ont été géoréférencées (1945, 1966, 1968, 1992) et nous apportent de précieuses informations sur différents états de l'occupation des sols de ce territoire. De manière complémentaire, un ensemble de prises de vues obliques réalisées pour l'essentiel en 2004 complète ponctuellement sur les sites les plus remarquables les données collectées sur le terrain. Ces informations feront l'objet de travaux d'intégration dans un proche avenir et permettront de confirmer ou de préciser les planimétries déjà existantes.

D'autres données sont venues compléter cet ensemble: les données géologiques, les cartes anciennes, images satellites, cartes climatiques; composant au final un important corpus documentaire complémentaire aux données archéologiques. Au-delà de l'intérêt documentaire, cette information nous a permis d'engager le développement d'un véritable outil d'analyse complémentaire à la base de données décrite plus haut et avec lequel nous espérons aboutir à de nouveaux schémas d'interprétation de la répartition spatiale de nos informations.

Ces données ont été structurées selon un modèle conceptuel de donnée volontairement simplifié nous permettant d'aboutir à un schéma simple mais robuste qui permet d'accueillir une grande diversité d'informations tant thématiques que chronologiques. En outre, ce schéma de donnée a été adopté car suffisamment simple pour être considéré comme générique et donc facilement réadaptable à d'autres situations et à d'autres outils ou plateformes (Fig. 4).

4.2 *L'organisation des données*

Le projet Itanos compte un corpus de données assez varié. Ces données possèdent des natures, des statuts et formalisations différentes. Ces données sont regroupées en différents ensembles:

- les données de référence géographique composent l'ensemble de données de référence pour toutes les opérations de géolocalisation. Ces données proviennent généralement de partenaires ou d'institutions locales (armée grecque) et sont considérées une fois acquises et intégrées comme stables. Elles n'ont pas vocation à être enrichies par le projet mais en constituent le cadre géographique nécessaire;
- les données de référence historique sont des données plus spécialisées. Elles sont le produit des travaux des archéologues et des historiens et composent

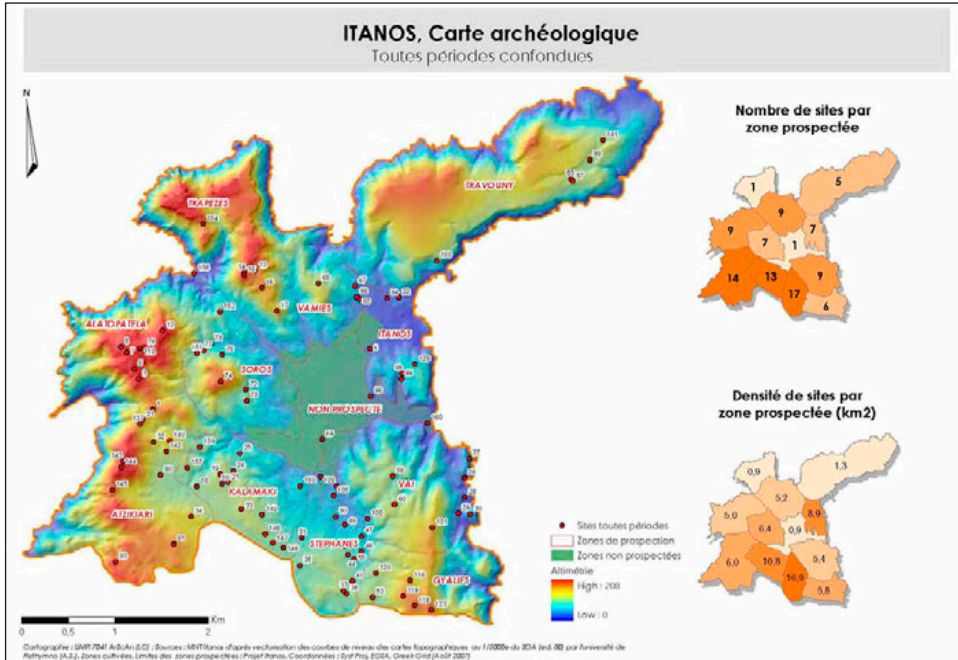


Fig. 4 – Vue générale de la répartition des sites.

le corps véritable du projet. Ce sont ces données qui vont être amenées à être retravaillées et qui vont subir tous les traitements nécessaires permettant de proposer des schémas d'interprétation historique.

4.3 Les données archéologiques et leur expression formelle

La formalisation de la notion de site reste un point sensible en archéologie. Cette notion complexe est classiquement modélisée sous deux formes graphiques:

- d'une part le site, le locus, le gisement quel que soit le nom qu'on voudra bien lui donner réduit à un couple de coordonnées qui trouve son expression en un point celui-ci étant caractérisé par différentes informations: c'est classiquement le modèle retenu de ce que l'on nomme dans le langage courant une "carte archéologique";
- d'autre part un mode d'expression où ce même endroit comportant des vestiges va être modélisé de manière plus complexe sous la forme d'entités (lignes points, polygones) possédant une étendue dans l'espace. Le site devient alors un élément complexe regroupant différentes entités. La structuration de

l'information est alors beaucoup plus complexe et il n'existe à ce jour pas de modèle unique. Parmi les réflexions les plus abouties sur les modèles de données appliqués à l'archéologie, nous avons retenus deux grands modèles d'organisation et de structuration de l'information qui sont le modèle PATRIARCHE⁵ du Ministère de la Culture et le modèle développé par le CNAU pour l'informatisation des DEPAVF.

L'intégration de ces deux visions du site correspondant à deux échelles de modélisation différentes des vestiges a été intégrée au sein d'une seule et même base.

4.4 Le modèle conceptuel de données d'Itanos

Un modèle conceptuel est schéma d'organisation et de relation des différents objets qui vont composer la base de données. Ce schéma d'organisation s'applique en l'occurrence aux données archéologiques et permet de définir la manière dont vont être réparties les différentes données enregistrées sur le terrain.

En nous inspirant des principales notions deux grands modèles de données actuellement appliqués en archéologie nous avons dégagé plusieurs ensembles d'objets correspondant aux objets de la base de données Itanos:

- les sites: la notion de site correspond pour nous à une notion de synthèse. Elle prend la forme graphique d'un point associant une information spatiale (*locus*), chronologique (*datation*) et typologique (*caractérisation fonctionnelle*);
- complémentaire au site on introduit la notion de gisement qui va permettre de traduire le site dans son développement spatial complexe: un gisement correspond à un lieu possédant des limites d'intervention (limites de sondage, limites d'échantillonnage, etc.) composé de vestiges et associé à un site;
- les vestiges, plus petite unité d'enregistrement de l'information, correspondent aux éléments dont on a constaté la présence sur le terrain. Ils peuvent être mobiliers ou immobiliers. Ils sont de différentes natures et leur délimitation est laissée à l'appréciation de l'archéologue. L'association de ces vestiges (mur, pierre, argile, etc.) est réalisée par le biais d'une notion de recrutement qui va permettre de composer des entités de différents niveaux qui vont pouvoir s'emboîter successivement afin de constituer les entités complexes que sont les sites.

Enfin, chaque vestiges est caractérisés par un nom et une datation plancher et une datation plafond. Ce modèle simplissime permet d'une manière relativement simple de construire une base de données structurée autorisant

⁵A. CHAILLOU, J. THOMAS, *L'application Patriarche. Inventaire informatique de la carte archéologique nationale*, «Nouvelles de l'archéologie», 107, 2007, 52-56.

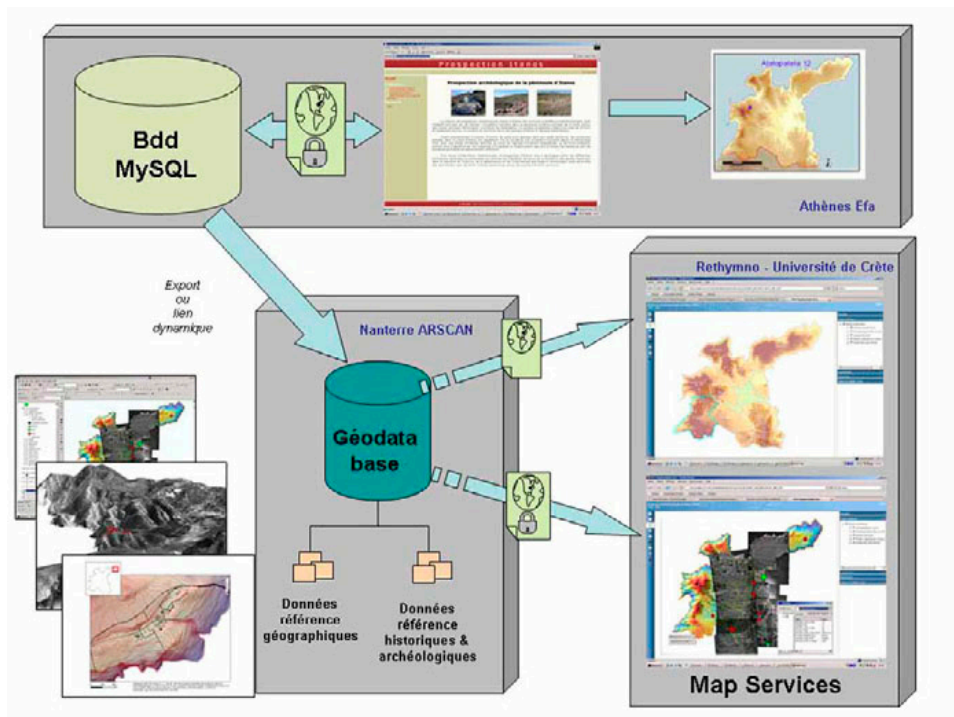


Fig. 5 – Organisation générale des données géographiques.

les synthèses par époque et par type sur l'ensemble des vestiges enregistrés dans le cadre d'une prospection telle que celle d'Itanos (Fig. 5).

5. MISE EN LIGNE DU SIG DE LA PROSPECTION

La volonté de rendre ce SIG accessible au plus grand nombre, de la même manière que la base de données de la prospection archéologique, nous a conduits à chercher à le mettre à disposition des chercheurs sur Internet. Un second site fut donc ouvert au début de l'année 2008 sur le serveur de l'Institut d'Études Méditerranéennes (Rethymno), autre partenaire du projet de recherches archéologiques mené à Itanos, qui a développé un savoir-faire particulier dans l'étude de l'organisation spatiale des données archéologiques et dans la diffusion sur le web des données géographiques (cfr. crête numérique).

Il ne s'agissait certes pas d'offrir une plateforme de recherche identique au SIG développé en interne basé sur une administration en mode monoposte et destiné à l'organisation et à l'analyse de toutes les données géographiques

issues de la prospection. L'ambition était plutôt de fournir un large accès à notre matériel et de permettre une manipulation commode de ces informations par les divers intervenants. L'un des obstacles majeurs posés par l'utilisation des outils SIG demeure en effet leur complexité, tant structurelle qu'ergonomique, pour bon nombre d'archéologues peu habitués à ces "nouvelles" technologies. La popularité du logiciel Google Earth et sa simplicité d'utilisation ont toutefois suscité la curiosité et l'intérêt des chercheurs pour ce genre d'applications. Profitant de cet engouement, il nous a donc paru intéressant d'offrir à tous un outil d'analyse comparable, à la fois simple et performant, consacré à l'étude du territoire d'Itanos.

D'un point de vue technique, l'utilisation d'une solution propriétaire⁶ pour la constitution du noyau central du SIG et sa gestion globale nous a conduits à adopter ces mêmes technologies pour une mise en ligne. L'application ArcGIS Server (version 9.2) a donc été utilisée pour offrir les services d'un SIG sur le web. Un simple navigateur Internet permet ainsi de faire office d'interface client, ce qui rend l'application commodément accessible à tout utilisateur, même inexpérimenté en matière de cartographie informatique.

Très simple, l'interface a été volontairement limitée dans un premier temps à quelques outils basiques (zoom, calcul de distances, pointeur et sélections croisées, création de zones tampons, etc.), offrant à l'utilisateur une navigation aisée et répondant aux principales attentes des archéologues. La sélection des calques permet en outre de visualiser de la manière la plus adéquate l'information recherchée en fonction de l'échelle d'affichage des données. L'interface propose enfin de générer une carte directement imprimable de ses propres sélections selon des paramètres spécifiés (échelle, résolution, dimensions).

Le modèle numérique de terrain et les courbes de niveaux fournissent la structure de base de toute l'information spatiale. D'une résolution de 20 m, celui-ci a été obtenu à partir d'images stéréographiques du satellite SPOT. Les limites de l'aire globale et des zones de prospection, ainsi que l'extension des surfaces actuellement mises en culture complètent le référent géographique.

La version actuelle du site web (http://www.ims.forth.gr/joint_projects/itanos/itanos.html) (Fig. 4), incorporant les derniers éléments relevés sur le terrain en 2007, nous a surtout permis de définir le protocole de transfert depuis le SIG statique et l'architecture générale du projet. Site public, cette version d'évaluation anglophone ne contient que quelques calques de données, n'offrant d'ailleurs guère plus d'informations – d'un point de vue quantitatif du moins – que le webmapping élémentaire proposé par la base de données en ligne. Les sites archéologiques, classés en fonction de leur période chronologique, sont associés à leur numéro d'identification. En marge, la description

⁶ Le logiciel ArcGIS de la société ESRI France est utilisé pour la gestion de la base de données géographique.

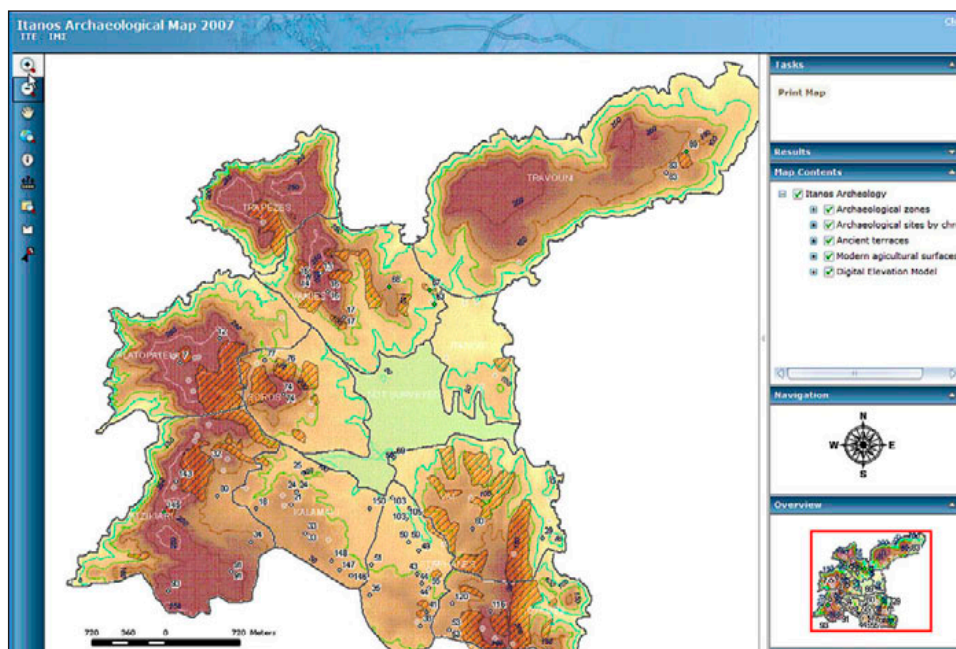


Fig. 6 – Environnement de navigation de l'application webGIS.

des structures reprend l'essentiel des données proposées, de manière certes plus commode, que dans la base de données des sites. Nous y avons enfin ajouté la cartographie détaillée des terrasses de culture antiques, dont les vestiges sont encore très clairement visibles au sol.

Il s'agit là d'une première étape dans la mise en ligne des informations géographiques de la prospection d'Itanos. Une seconde version, actuellement en cours de développement, offrira un accès complet à tous les calques de données intégrés dans le SIG "de bureau" à la fois aux données de bases mais aussi à des données issues d'analyses spatiales plus complexes (telles que les polygones de Thiessens, des modèles de centralités, des couches de données sur les réseaux viaires, une restitution de la topographie antique, etc.). Comme pour la base de données, l'accès à cette carte archéologique dynamique complète sera réservé dans un premier temps aux seuls archéologues de la mission et constituera pour eux un outil de travail essentiel à l'analyse du territoire d'Itanos. Une fois la synthèse historique publiée dans la collection des *Études crétoises*, le site ouvrira ses portes largement à tout utilisateur.

Compte tenu de leur interface relativement différentes, il n'est pas prévu de réunir en une seule application web les deux outils développés sur les serveurs de l'École française d'Athènes et de l'Institut d'Études Méditer-

ranéennes: l'un offre une interface de recherche et de visualisation optimisée pour le texte et les images; l'autre constitue un instrument d'analyse proprement géographique. Tout comme la synthèse historique finale, qui concluera les travaux menés depuis 1994 à Itanos, l'interrogation informatique des données récoltées au fil de ces années s'appuiera sur la complémentarité des disciplines et des approches (Fig. 6).

LAURENT COSTA, FRANCK EYRAUD
CNRS – UMR 7041 ArScAn Archéologies et Sciences de l'Antiquité

ALAIN DUPLOUY
Université de Paris 1 Panthéon-Sorbonne
CNRS – UMR 7041 ArScAn Archéologies et Sciences de l'Antiquité

APOSTOLOS SARRIS
Institute of Mediterranean Studies
Foundation of Research & Technology (FORTH)

ABSTRACT

As part of the archaeological research program conducted in Itanos since 1994, an archaeological survey was carried out on a substantial portion (about 20 sq. km) of the territory of the ancient city. The area surrounding Itanos offers an exceptional site that has not been occupied by any important human settlement since the Middle Ages. It thus appears as a fossilized map of the ancient occupation and presents an exceptional case in the history of the Cretan landscape. During the past few years, archaeologists have collected a huge quantity of information, that must now be presented to the scholarly community. All data have been organized in a database that is fully compatible with a GIS system. Since 2006 this searchable database offering the records of the survey has been made available online on the website of the French School of Archaeology. It is the intent of this article to present the heuristic and technical principles which lie at the basis of the project. Moreover, a GIS platform, offering a digital elevation model and a chronological distribution of all the archaeological sites surveyed, has recently been accessible on the website of the Institute for Mediterranean Studies. This GIS platform will be continually augmented by additional data and will eventually offer a powerful interpretive tool to archaeologists, which will integrate the physical conditions of the landscape which are at the core of historical explanation.