

“ODYSSÉE”: REPRÉSENTER L'INFORMATION AU SEIN D'UN SYSTÈME D'INFORMATION ARCHÉOLOGIQUE POUR LA RECHERCHE SUR L'HABITAT RURAL DU MOYEN ÂGE

1. INTRODUCTION

Cet article répond à une double préoccupation au croisement de deux des activités de l'auteur: celle de chercheur, comme doctorant au sein de l'équipe Archéologies Environnementales de l'UMR 7041, ArScAn, et celle d'enseignant-chercheur en archéologie médiévale à l'université de Paris 1 – Panthéon-Sorbonne.

À travers le partage de l'expérience du développement d'Odyssee¹, un système d'information pour la recherche sur l'habitat rural du Moyen Âge, il s'agit de commencer une réflexion sur l'établissement de standards dans la représentation de l'information au sein des systèmes d'information archéologiques.

2. AVANT-PROPOS

“Normes et standards”, le thème adopté pour les Journées Informatique et Archéologie de Paris en juin 2010 et dont est issu cet article, est révélateur. Il existe aujourd'hui profusion d'outils de conception informatique facilitant la mise en œuvre par tout un chacun d'une base de données accessible *via* Internet. Le paradigme du logiciel libre ajoute à cette facilité technique une gratuité bienvenue pour notre discipline, surtout dans le contexte actuel.

La gestion de l'information est une pratique intrinsèque à l'archéologie. Pour l'illustrer, il n'est besoin que d'évoquer les premiers travaux de Christian Jürgensen Thomsen au Musée National danois dans la première moitié du XIX^e siècle (LUNDBECK-CULOT 2005). Son premier objectif était de mettre de l'ordre dans les collections, ce qui lui permit au passage de proposer un système de classement s'appuyant sur les «trois âges»². Une réflexion méthodologique sur les systèmes d'information archéologiques a été proposée récemment dans un travail de synthèse d'Anne Chaillou (CHAILLOU 2003).

Facilité, gratuité et nécessité s'associent pour rendre compte d'une tendance de fond dans la pratique actuelle de l'archéologie: le développement de base de données à un niveau individuel ou groupusculaire, par opposition au niveau institutionnel. À un tel point qu'aujourd'hui, à l'université, du moins

¹ O.D.Y.S.S.É.E: Outil pour la Description, en vue de sYnthèSes, des Sites archÉologiques.

² Les âges de la Pierre, du Bronze et du Fer.

à l'université de Paris 1, la conception d'une base de données en archéologie fait partie du *cursus* de Master.

Or, concevoir et faire évoluer un système d'information ne s'improvisent pas. Il s'agit de surcroît d'un métier à part entière (on parle alors d'"urbanistes"). La disponibilité d'outils de conception de bases de données peut alors s'avérer être un leurre pour le chercheur. En effet, il risque de perdre temps et énergie à surmonter des difficultés d'ordre conceptuel ou technique, auxquelles il n'est pas ou peu préparé.

Or, tant pour l'étudiant que pour le chercheur confirmé, l'objectif final est le traitement de l'information archéologique. Partant du constat de la diffusion des outils de base de données, il nous semble donc à propos d'inviter à une réflexion collective sur un ensemble de normes en archéologie. Il ne s'agit cependant pas de réfléchir sur le "quoi", mais sur le "comment". Ainsi nous ne nous interrogerons pas sur quelles informations une base doit contenir (enjeu des normes décrites par Anne Chaillou dans sa thèse: CHAILLOU 2003, 52s), mais sur comment représenter ces informations.

En bref, il s'agit de réfléchir sur les modèles conceptuels de données formant le squelette d'une base de données. Toutefois, nous ne proposerons pas un schéma unique ou universel ni même n'y aspirerons³. Loin de nous cette idée, car le projet peut sembler illusoire. Il peut exister en effet un schéma différent par problématique de recherche et donc par système d'information, ceci à la convenance du chercheur.

Notre idée est donc plutôt de raisonner par modules d'information. C'est-à-dire de proposer une "boîte à outils" dans laquelle irait piocher l'archéologue. Les points d'entrée sont des questions certes simples, mais récurrentes lors des phases de définition d'un système d'information en archéologie: comment représenter une opération, un site, des informations de datation, etc.? Libre après au chercheur de combiner ces briques de base comme il le souhaite pour répondre à sa problématique.

Pour amorcer cette réflexion, nous nous proposons de soumettre à la critique en première instance les modèles de données utilisées au sein d'Odysée. Notre discours sera divisé en trois mouvements: la genèse d'Odysée, sa conception et les modèles en eux-mêmes.

3. LA GENÈSE D'ODYSSÉE

Le projet de système d'information Odysée est né à l'automne 2007 par la conjonction de trois circonstances favorables: la relance du Programme Collectif de Recherche (PCR) sur l'habitat rural du haut Moyen Âge en

³ Pour une tentative, voir la "base minimale" proposée par CHAILLOU 2003, 206-238.

Île-de-France par les chercheurs de l’INRAP (Institut National de Recherches Archéologiques Préventives), en particulier Gaëlle Bruley-Chabot et François Gentili; le début du PCR sur l’habitat rural du Moyen Âge en région Centre sous la direction d’Anne Nissen-Jaubert (Université François Rabelais de Tours) et de Sébastien Jesset (INRAP); le début de la propre thèse de l’auteur sur l’habitat rural du haut Moyen Âge en Île-de-France (BRULEY-CHABOT 2003, GENTILI *et al.* 2003, NISSEN-JAUBERT 2006, CATTEDU *et al.* 2007).

La caractérisation des habitats ruraux du Moyen Âge est apparue comme un dénominateur commun aux trois approches. Une même préoccupation s’est aussi imposée: comment traiter la masse de documentation archéologique accumulée depuis le développement des fouilles préventives? Soit, autour du millier de sites pour l’Île-de-France pour le haut Moyen Âge et près de cent cinquante du Moyen Âge pour la région Centre. Le bien-fondé d’une approche commune pour concevoir un système d’information était évident. Le projet Odyssée était lancé.

4. LA CONCEPTION

L’objectif principal d’Odyssée est d’aider à la caractérisation des habitats ruraux du Moyen Âge. Il s’agit donc d’un système d’information focalisé sur la recherche par opposition aux systèmes plus orientés vers la gestion administrative comme Patriarche par exemple. L’objectif secondaire est d’assurer à terme un recensement complet des sites du Moyen Âge pour les régions concernées. Ainsi il sera possible d’évaluer l’importance d’un site pour un axe d’étude donné (artisanat, bâtiments, structures agraires...) et de faciliter les recherches futures.

L’auteur, ayant une certaine expérience dans le développement de logiciel, s’est proposé pour assurer la conception et la réalisation du système. Toutefois, la réflexion a été menée de manière collégiale, et en particulier avec Gaëlle Bruley-Chabot, responsable d’opération à l’INRAP. Afin que le système convienne au plus grand nombre, en pratique un cycle de développement logiciel en V a été suivi (DESMOULINS 2009). Ce mode implique qu’à chaque grande étape une validation des utilisateurs ait lieu. Certes, un temps plus important est nécessaire du fait des adaptations intermédiaires, mais nous nous assurons ainsi de la satisfaction finale des utilisateurs. De plus, les réunions régulières d’information ont permis de garder intacte l’implication des personnes au projet.

Avant d’aborder les modèles de données en eux-mêmes, nous allons évoquer les choix techniques. La nécessité d’autoriser la saisie des données par un nombre élevé de contributeurs amène à privilégier une solution en ligne, c’est à dire accessible par Internet (moyennant une identification). En centralisant l’information, cette solution permet en effet d’éviter les problèmes

associés à la réplication de la base de données sur les postes contributeurs: en particulier le risque de doublons et le fastidieux travail de fusion des bases de données de chacun. De plus les interfaces web, *via* l'utilisation de navigateurs (Firefox, Safari, IE...) constituent aujourd'hui le moyen le plus familier pour renseigner ou consulter l'information sous forme électronique. L'utilisation des standards de saisie et de présentation en ligne assure ainsi la facilité d'utilisation du système d'information. Enfin, l'adoption d'une technologie web permet l'utilisation de logiciels libres dont l'efficacité est à la mesure de leur large diffusion. Ces logiciels présentent différents avantages⁴. En particulier, ils respectent des standards ouverts assurant la pérennité des informations qui leur sont confiées. En effet, les formats utilisés par les logiciels propriétaires pour stocker de manière numérique l'information sont pour la plupart fermés: seule la société éditrice du logiciel sait décrypter ces informations. L'utilisateur est alors tributaire de cette dernière pour accéder à ses données (nécessité d'achat de licences par exemple). Par leur caractère ouvert et leur succès, les logiciels libres permettent de disposer d'une large documentation accessible en ligne: en particulier de nombreux exemples d'implémentation d'architectures diverses permettant de gagner du temps dans la mise en œuvre de la solution. Enfin, ces logiciels sont fournis gratuitement. Leur utilisation n'entraîne aucun coût d'achat de licences.

Pour des raisons de coûts, de pérennité de l'information et de facilité de développement, des logiciels libres ont donc été utilisés: Linux pour le système d'exploitation du serveur; le serveur Web Apache; le générateur de pages web PHP; le système de gestion de base de données MySQL; le progiciel de gestion intégré Dolibarr comme pont entre PHP et MySQL.

Enfin Firefox est utilisé pour accéder au système (consultation, modification).

5. LES MODÈLES DE DONNÉES

Nous présentons ici les différents modules d'information qui constituent Odysée. Comme annoncé en introduction, il s'agit d'une première ébauche soumise ici à la critique et qui, nous l'espérons, pourra déclencher une réflexion sur un ensemble de référence en archéologie.

Si nous avons pu évoquer dans le paragraphe précédent des points techniques relevant de l'implémentation de la solution, il convient de souligner que le cœur de notre discours concerne la conception. En effet, les différentes modélisations proposées ci-dessous ne préjugent aucunement des choix te-

⁴ Pour une présentation détaillée: AFUL, *Qu'est-ce qu'un logiciel libre?*, Association Francophone des Utilisateurs de Linux et des Logiciels Libres, <http://www.aful.org/ressources/logiciel-libre/>, page consultée le 28 avril 2008.

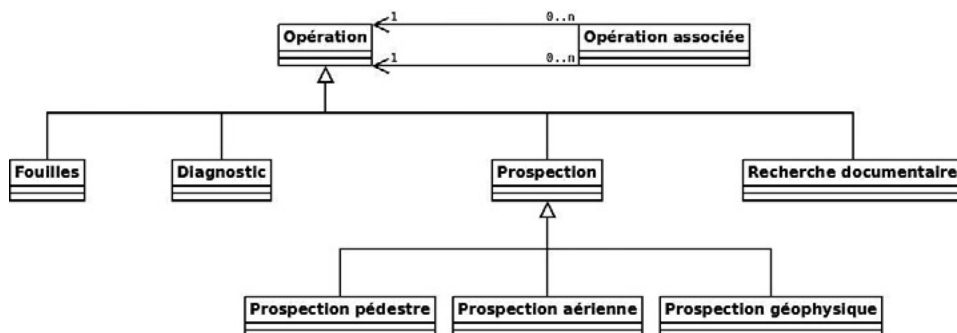


Fig. 1 – Module “opération”.

chniques retenus pour les implémenter. La séparation de la conception et de l’implémentation est une règle de base de l’ingénierie informatique. Elle vise à s’affranchir de normes techniques qui évoluent très rapidement.

Les 9 modules fonctionnels constituant Odyssee sont présentés ci-dessous. La syntaxe graphique utilisée pour formaliser les différentes composantes de la base de données utilise le standard de conception logicielle UML “Unified Modeling Language” (MULLER 2000).

5.1 L’opération (Fig. 1)

En archéologie, l’opération correspond très souvent à la pierre d’angle du système d’information. Selon le type, elle pourra être déclinée en fouilles, diagnostic, prospections pédestres, aériennes, géophysiques...

La classe “opération associée” met la classe “opération” en relation avec elle-même. Ceci permet de traiter le cas où des opérations successives concernent le même espace; par exemple un diagnostic suivi d’une fouille. On dispose ainsi d’un moyen de détecter des informations dupliquées, comme par exemple une structure qui apparaît au diagnostic et est reprise en fouille.

5.2 La localisation (Fig. 2)

Les informations de localisation sont associées à l’opération et relèvent principalement de deux catégories: l’espace administratif (dans le cas présent, communes/département/région/pays) et les coordonnées géographiques (dans le cas présent le type de projection Lambert, les coordonnées elles-mêmes étant mentionnées au niveau de la classe “opération”). D’autres catégories pourraient néanmoins être ajoutées.

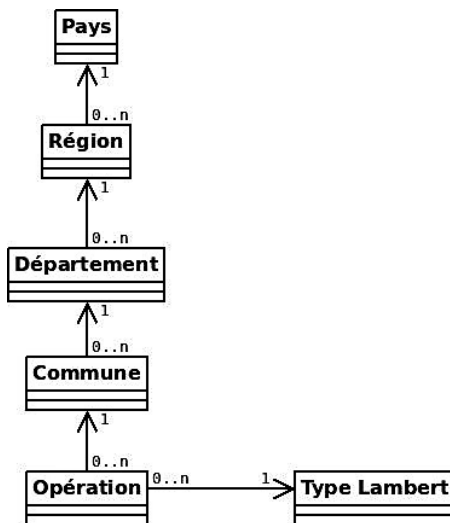


Fig. 2 – Module “localisation”.

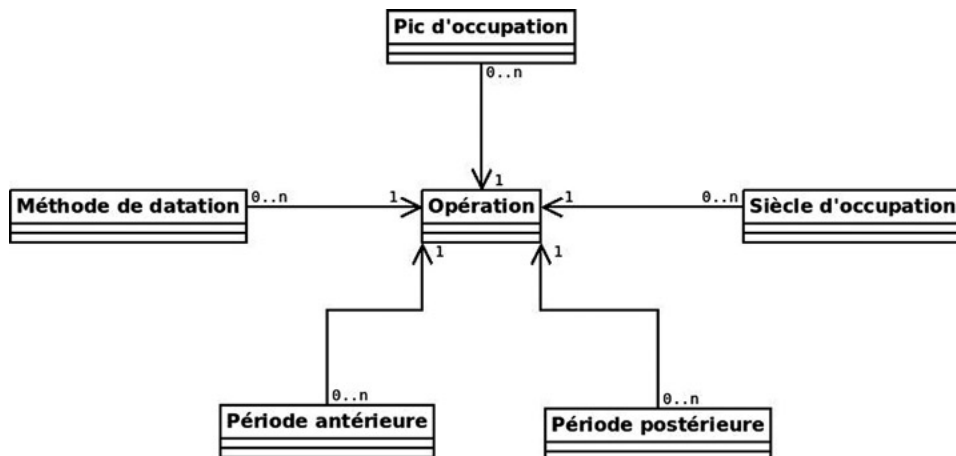


Fig. 3 – Module “datation”.

5.3 La datation (Fig. 3)

Ce module est lui aussi associé à l’opération. Il permet de renseigner (1) les méthodes de datation (2) les datations elles-mêmes, soit les périodes pour lesquelles une occupation est attestée.

Dans ce dernier cas, la période centrale à la recherche est distinguée.

Il est possible de mentionner si des périodes antérieures et postérieures sont attestées et les siècles (ou un intervalle inférieur ou supérieur) peuvent être précisés pour la période centrale. Pour pouvoir distinguer un simple “bruit de fond” d’une occupation dense, il est aussi possible de mentionner les siècles au sein d’une classe “pics d’occupation”.

5.4 *Le matériel archéologique* (Fig. 4)

Le matériel archéologique mis au jour lors d’une opération est mentionné dans ce module. Trois catégories ont été distinguées :

- les matières;
- le mobilier;
- les activités.

La dernière catégorie est plus interprétative puisqu’elle suppose qu’une fonction a été déduite du mobilier trouvé.

5.5 *Les structures* (Fig. 5)

Associé à l’opération, un décompte des structures par type est réalisé au sein de ce module. Deux décomptes peuvent être distingués :

- un décompte général par type;
- un décompte par type et par phase (siècle de début, siècle de fin ou tout autre intervalle pertinent).

5.6 *Le bâti* (Fig. 6)

Ce module permet de mentionner les différents bâtiments distingués par le responsable d’opération. Plusieurs points sont à noter :

- les bâtiments sont associés à une opération;
- à un bâtiment peut être associé un certain nombre de techniques de construction et de matériaux;
- enfin une plage d’occupation en siècles peut être définie pour chacun des bâtiments.

5.7 *Les études spécialisées* (Fig. 7)

Si des études spécialisées ont été réalisées en sus de l’opération elle-même, elles sont renseignées ici : anthracologie, archéozoologie, céramologie...

5.8 *Le système agraire et l’environnement* (Fig. 8)

Ce module rassemble les informations concernant le système agraire et l’environnement que l’opération a permis de déduire. Ces informations sont

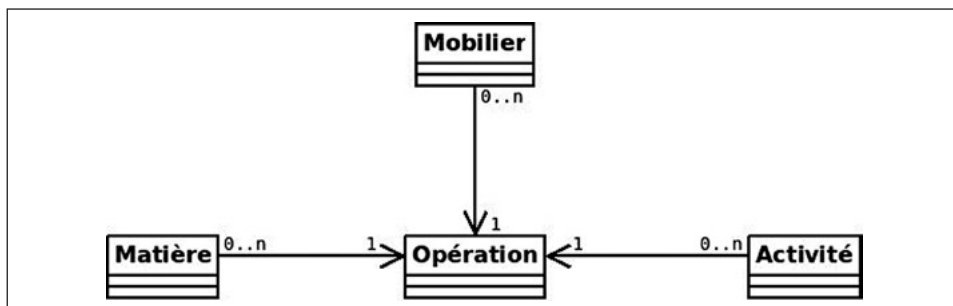


Fig. 4 – Module “matériel archéologique”.

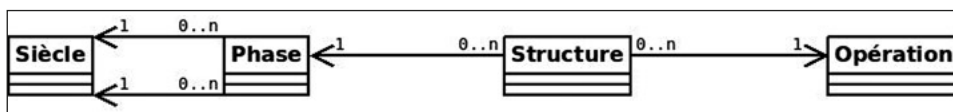


Fig. 5 – Module “structures”.

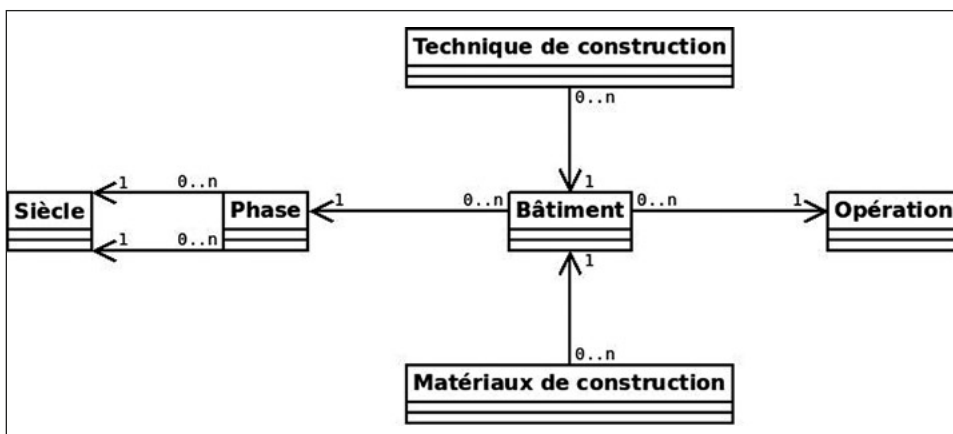


Fig. 6 – Module “bâti”.

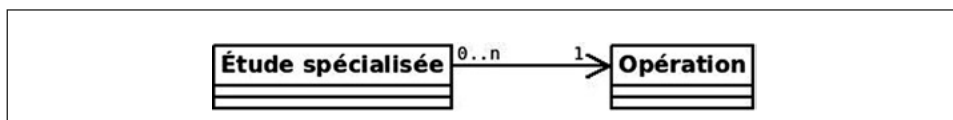


Fig. 7 – Module “études spécialisées”.

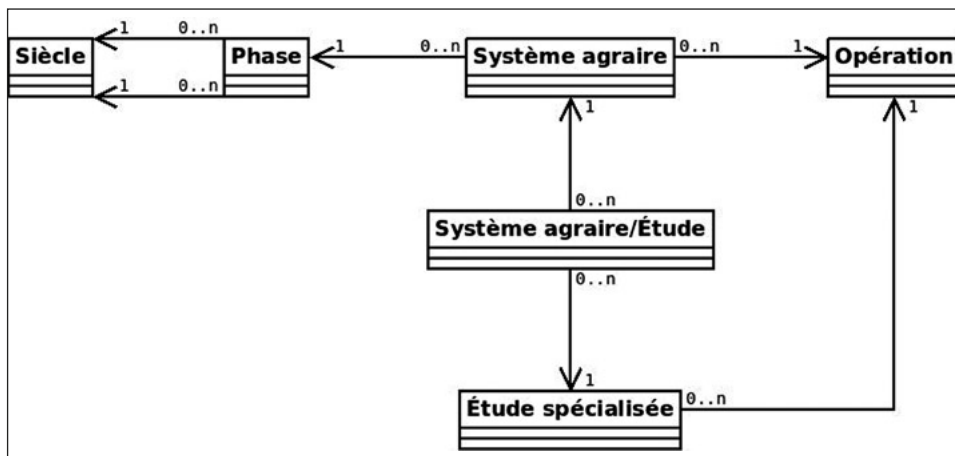


Fig. 8 – Module “système agraire et environnement”.

bien souvent associées à des études spécialisées (carpologie, palynologie...) qu’il est donc possible de mentionner (lien avec la classe “étude spécialisée” *via* la classe “système agraire/étude”). Enfin une phase chronologique peut être définie (lien avec la classe “phases”).

5.9 Le site et les ensembles spatiaux remarquables (Fig. 9)

Il est important de bien distinguer la notion de site de celle d’opération. Un site correspond à l’appariement spatial de structures au sein de mêmes phases temporelles. Il s’agit d’un niveau interprétatif plus élevé que celui guidant le renseignement des opérations. Un site est modélisé par différents moyens.

Un site peut être référencé sous plusieurs opérations. De même, une opération peut référencer plusieurs sites (lien avec la classe “opération” *via* la classe “site/opération”). Et ceci pour tout type d’opération (diagnostics, fouilles, prospections). Un site peut comprendre plusieurs phases d’occupation (lien avec la classe “phases” *via* la classe “site/phase”). Enfin, si un site est constitué de structures et de bâtiments, ceux-ci peuvent faire l’objet au préalable de regroupements au sein d’“ensembles spatiaux remarquables”.

Il ne s’agit pas ici d’un artefact de modélisation, mais de la traduction de réflexions collectives menées au sein des deux PCR sur l’habitat rural du Moyen Âge (cfr. *supra*) par les chercheurs, en particulier ceux de l’INRAP⁵: des aires spécifiques sont délimitées au sein des sites (du moins pour les plus

⁵ Je remercie ici François Gentili (INRAP) pour les fructueuses réflexions dont il m’a fait part.

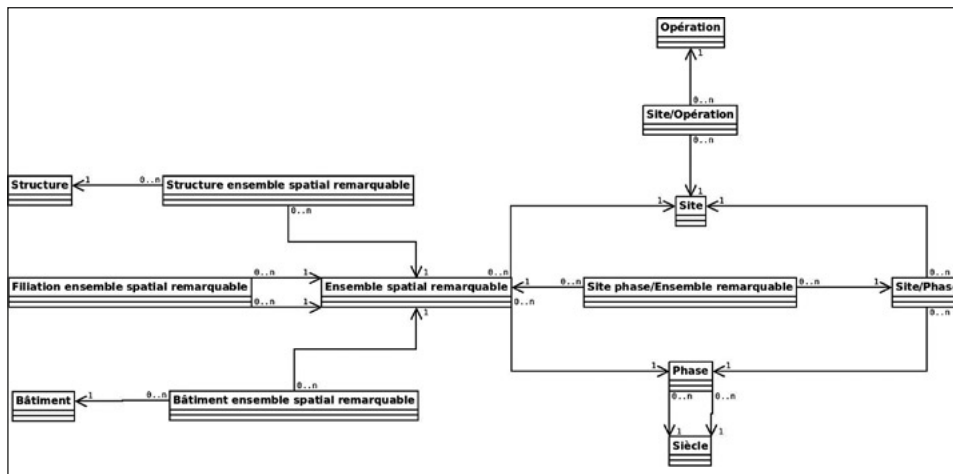


Fig. 9 – Module “site et ensembles spatiaux remarquables”.

importants). Il s’agit ici de faciliter l’analyse spatiale inter-sites en permettant de comparer rapidement et aisément des groupes de structures et bâtiments fonctionnant ensemble.

Au cours des différentes phases d’occupation d’un site, les ensembles spatiaux remarquables sont susceptibles d’évolutions. Pour traiter cet aspect, il est possible d’établir un lien de “filiation” entre ensembles remarquables. Les modalités de cette évolution peuvent être nuancées spatialement: aire identique, extension, diminution, décalage.

Une remarque importante doit toutefois être faite ici. En terme spatial, l’agencement des structures et des bâtiments, les uns par rapport aux autres, constitue une information primordiale. C’est pourquoi il est important que le système d’information permette de sauvegarder un plan d’agencement des structures et des bâtiments réalisé selon une légende normalisée.

Dans le cadre des PCR concernés, le standard de représentation est toujours en voie d’élaboration. L’objectif à terme serait d’établir une typologie de ces regroupements, aidant ainsi à la caractérisation des habitats ruraux du haut Moyen Âge en Île-de-France et du Moyen Âge en région Centre.

6. CONCLUSION

Aujourd’hui en archéologie, l’étudiant, l’enseignant-chercheur ou le chercheur conçoivent leur propre système d’information. Cette tendance de fond a été rendue possible par la démocratisation de l’informatique à tous les niveaux: sur le plan matériel évidemment, mais surtout sur le plan logi-

ciel, avec la mise à disposition de système d’information gratuit de qualité professionnelle.

Réaliser et mettre en service son système d’information deviendra de plus en plus aisé. Toutefois, il existe une étape primordiale dont la difficulté ne pourra jamais être éludée: la conception même du système et la réflexion préalable qu’elle impose. Soit, le modèle de données présidant à la représentation de l’information au sein de la base de données. Or, chaque système d’information peut *a priori* être unique, car adapté à une problématique précise. Cependant, il est possible de dégager des schémas récurrents au sein des modèles de données, car répondant aux mêmes questions élémentaires: comment représenter une opération, des informations de localisation, de datation, un site...?

À travers cet article, nous avons donc proposé une série de descriptions de modules d’information pour les bases de données en archéologie. Issus certes du système Odyssée pour l’étude de l’habitat rural du haut Moyen Âge, ces modèles de données se veulent avant tout génériques pour pouvoir être réutilisés. À terme, l’objectif est de constituer un référentiel pour aider l’archéologue, étudiant ou chercheur plus confirmé, à concevoir son système d’information. Pour ce faire, nous espérons qu’une réflexion collective puisse s’initier et déboucher sur un ensemble de normes traitant des modèles de données en archéologie.

ANTOINE BOURROUILH

Université de Paris 1 Panthéon-Sorbonne

UMR 7041 Archéologies et Sciences de l’Antiquité

BIBLIOGRAPHIE

- BRULEY-CHABOT G. 2003, *Les fours culinaires en Île-de-France*, in *L’habitat rural du haut Moyen Âge en Île-de-France, PCR bilan 2002/2003*, 1^{er} suppl. au Bulletin Archéologique du Vexin français, Guiry-en-Vexin, 25-31.
- CATTEDDU I., GENTILI F., JESSET S. 2007, *Le renouvellement des connaissances sur l’habitat et l’espace rural au haut Moyen Âge*, in J.-P. DEMOULE (dir.), *L’archéologie préventive dans le monde. Apports de l’archéologie préventive à la connaissance du passé*, Paris, La Découverte – INRAP, 82-92.
- CHAILLOU A. 2003, *Nature, statut et traitements informatisés des données en archéologie: les enjeux des systèmes d’informations archéologiques*, thèse de Doctorat de sciences humaines et sociales, Université Lumière Lyon 2, 2 t. (<http://tel.archives-ouvertes.fr/tel-00137986/fr/>) [mis en ligne en Mars 2007].
- DESMOULINS N. 2009, *Maîtriser le levier informatique: accroître la valeur ajoutée des systèmes d’information*, Paris, Pearson, 95-96.
- GENTILI F., LEFÈVRE A., MAHÉ N. 2003, *Introduction*, in *L’habitat rural du haut Moyen Âge en Île-de-France, PCR bilan 2002/2003*, 1^{er} suppl. au Bulletin Archéologique du Vexin français, Guiry-en-Vexin, 5-9.
- LUNDBECK-CULOT K. 2005, *Recherche sur la fondation de l’archéologie danoise. Une réflexion sur les influences réciproques entre le Danemark et la France*, thèse «nouveau régime» soutenue à l’Université de Paris 1 – Panthéon-Sorbonne sous la direction d’Alain Schnapp, Atelier National de Reproduction des Thèses, Lille, 274 ss.

MULLER P.-A. 2000, *Modélisation objet avec UML*, Paris, Eyrolles.

NISSEN-JAUBERT A. 2006, *Le haut Moyen Âge*, in A. FERDIÈRE, V. MATTERNE, P. MÉNIEL, A. JAUBERT, B. PRADAT, *Histoire de l'agriculture en Gaule*, Paris, Errance, 141-197.

ABSTRACT

The large amount of documentation made available by “preventive” archaeology led to the setting up of several collective research programs. Aimed at rural settlements during the Middle Ages on a regional level, both “région Île-de-France” and “région Centre” PCR have to deal with handling this documentation. In line with the current PhD work of the author about settlements and agrarian system during the early Middle Ages, a shared reflection on a common information system was launched. Beyond approaches aimed at evaluating archaeological potential or solely focused on inventory, here the goal is to acquire a tool dedicated to research. Moreover, if databases are used on a daily basis by the archaeologist, their use in a collective way leads to the standardization of the manner information is encoded. The aim of this paper is to present the conclusion of this shared reflection, focusing especially on settlement patterns, environmental data and the estimate of reliability of information. The technical choices which helped to minimize costs are also presented.