

# **“Nature, statut et traitements informatisés des données en archéologie : les enjeux des systèmes d’informations archéologiques”**

**Anne CHAILLOU**

sous la direction de Bruno Helly

L’utilisation des bases de données en archéologie soulève plusieurs problèmes que l’on peut classer en trois catégories. Des problèmes théoriques portent sur la nature des données à intégrer dans la base, leur description et le langage qu’il faut utiliser pour les décrire ; en effet, sans un langage approprié et une gestion cohérente de la description, les informations saisies dans la base ne peuvent être exploitées. La deuxième catégorie de problèmes concerne les questions techniques, en particulier celles qui touchent à la base de données ; il faut construire des bases librement conçues mais fondées sur certains éléments d’homogénéité communs à toutes, permettant des liaisons de l’une à l’autre et une exploitation conjointe. Viennent ensuite les problèmes institutionnels ; ceux-ci doivent être résolus pour permettre d’assurer la pérennité de toutes ces données informatisées. C’est en avançant dans la résolution de ces problèmes que les enjeux des bases de données répondront à l’attente des utilisateurs. Ces enjeux sont doubles : offrir aux chercheurs la possibilité d’une recherche croisée de données dans toutes les bases accessibles en même temps, et assurer la pérennité de ces données par un suivi et un développement des bases dans le temps. L’important n’est pas de faire des bases de données parce que c’est la mode et qu’il faut utiliser l’outil informatique pour donner un semblant d’esprit scientifique à sa recherche, mais parce qu’il est indispensable de permettre un accès et un croisement des données pour une exploitation plus large et plus riche et pour préserver au mieux ces données pour les recherches futures.

Ce travail vise à proposer des solutions à ces problèmes non résolus encore aujourd’hui. Il se développe en trois parties.

La première permet de revenir sur des points de définition, afin de fixer une terminologie aussi claire que possible. La donnée, sa position dans l’histoire de l’archéologie, son statut, les différents types de données archéologiques, leur description, ... y sont ainsi précisés. Une démarche identique traite des bases de données et de leur fondement. Certains éléments doivent être homogènes pour pouvoir référencer, décrire et ensuite exploiter ces données. Ces éléments doivent aussi être homogènes entre toutes les bases pour qu’on puisse assurer une recherche croisée et une pérennité des données par la possibilité de transfert d’une base à une autre. Ces éléments sont l’unité documentaire, les codes d’inventaires, les descriptions et le langage.

La deuxième partie porte sur un état des lieux des bases de données existantes. Le corpus sélectionné comprend cinq bases de données françaises ayant un développement différent, des zones d’interventions plus ou moins grandes, touchant différentes périodes.

Elles sont manipulées par des utilisateurs uniques ou multiples ce qui est un facteur déterminant dans le fonctionnement d'une base de données. Ces bases sont les suivantes : SYSLAT et ArchéoDATA (premières bases dont on peut dire qu'elles se sont données une envergure nationale en étant disponibles sur le marché pour d'autres utilisateurs que le site pour lequel elles ont été conçues), la base de données Beuvray utilisée sur la fouille du Mont Beuvray (Saône-et-Loire/Nièvre) appelée bdB, la base de données SysDA qui est utilisée à la Maison de l'archéologie de Chartres (Eure-et-Loir) et enfin la base de données du service régional de l'archéologie de la région Aquitaine, BASE\_DFS.

La troisième partie s'ouvre sur la proposition d'un protocole de création de bases de données. Ce protocole permet de gérer de A (définition des besoins) à Z (manuel d'utilisation) les questions soulevées lors de la création d'une base. Il a été testé lors de la mise en place d'une base de données sur l'étude de l'iconographie hindoue des 11<sup>e</sup>-13<sup>e</sup> siècles. Dans cette partie, une solution est aussi proposée, parmi sûrement beaucoup d'autres, pour résoudre les problèmes cités précédemment et répondre aux enjeux des bases de données. Elle prend pour base les études de cas et se développe autour de deux actions couplées. Dans la première, il s'agit d'homogénéiser certains éléments clefs de l'archivage des données pour pouvoir mettre en place une norme suivie par tous. Cette norme permettrait d'édifier des liens et des passerelles entre les différentes bases développées. Ces éléments clefs sont au nombre de quatre. Le premier est l'unité documentaire : unité stratigraphique ou unité de fouille. Le second concerne les codes d'inventaires qui permettent d'identifier chaque donnée et de la retrouver. La description homogène d'un type de donnée est le troisième élément clef. En effet pour pouvoir exploiter une base, il faut que toutes les données d'un même type (couches, objet céramique, objet métal, photographies, ...), soient décrites de la même manière et toujours suivant le même protocole, au travers de champs de saisie identiques ou au moins compatibles. C'est grâce à ces champs que l'on pourra faire des recherches concluantes dans plusieurs bases, champs auxquels on aura associé le quatrième élément, le langage. Il est en effet capital, pour s'y retrouver, de décrire des données identiques avec les mêmes mots ou termes. Il faut pour cela mettre en place des listes de termes de vocabulaire par types de données, par disciplines, par catégories d'objets, par époques fouillées, ...

Cette homogénéisation est couplée à la deuxième action : la mise en place d'un tronc commun. Celui-ci contient les fichiers, rubriques et liens minimaux à la bonne gestion des données. Il répond aux objectifs mis en avant par l'étude des bases de données existantes à savoir : recueillir les références de toutes les données et la documentation issues de la fouille suivant des codes d'inventaires uniques pour chaque donnée, organiser ces données et cette documentation, les décrire, permettre le traitement de ces données et cette documentation pour pouvoir réaliser des comparaisons, des recherches, ..., proposer une aide à l'interprétation (datation, regroupement d'US, ...) et une aide à la publication. D'autres données traitées pourront par la suite être associées à ce tronc commun en fonction des besoins divers de chacun.

Cette association *éléments homogènes/tronc commun* permet donc de répondre aux enjeux posés par les bases de données en archéologie. Les éléments clefs homogènes fournissent une identification, un protocole et un langage de description normalisés qui permettent des recherches croisées entre bases, et des études comparées d'objets provenant de fouilles diverses. Le tronc commun, quant à lui, instaure un protocole de sauvegarde des données. Il sera toujours possible, lorsqu'une base ne sera plus exploitée, de transférer ses données vers une base encore en fonctionnement, et ceci grâce à une même structuration des codes d'inventaires, à une même unité documentaire, à des rubriques identiques et à des listes

de termes de vocabulaire communes. Pour obtenir une pérennité complète des données, il faut adjoindre à cette association *éléments homogènes/tronc commun* un traitement des données numérisées (images et textes) dans un langage universel et la mise en place d'un protocole raisonné de recopie des supports de stockage lié à des contrôles de qualité.

Cette structure *éléments homogènes/tronc commun* enfin acquise, des orientations d'évolution peuvent être proposées. La première de ces orientations concerne la consultation des bases de données. Celle-ci est maintenant à envisager via le Web. Il est très simple d'imaginer le bénéfice que cela peut apporter à un chercheur. Néanmoins, cette possibilité de mise en ligne des bases soulève d'autres difficultés qu'il faudra aussi traiter. La deuxième orientation de l'évolution à donner aux bases de données concerne les outils de visualisation des recherches pour aider à l'interprétation des données. Une fois que l'archivage à l'intérieur du tronc commun est réalisé correctement, l'archéologue peut enfin exploiter les données indépendamment de leur origine : recherche, regroupement, association, recoupement, comparaison, réalisation de typologies, ... Ces résultats devraient pouvoir être traités sous différentes formes et non pas seulement sous la forme d'une liste de données comme c'est encore le cas le plus souvent dans les bases de données actuelles. Des graphiques, des courbes, des cartes de répartition, ou des traitements par l'intermédiaire d'un SIG (système d'information géographique), sont autant d'outils de présentation qui devraient être à disposition de l'archéologue. À lui ensuite de choisir le plus pertinent.

Mais pour y parvenir, il faut encore explorer ces outils et donc laisser chacun faire ses propres découvertes et en faire émerger les points les plus importants. Cependant, il faut espérer que l'exemple des bases de données permettra de réfléchir le plus rapidement possible ensemble aux résultats obtenus. C'est à cette condition, la mise en œuvre d'outils de présentation et donc d'aide à l'interprétation, que les bases de données pourront aspirer à la qualification tant convoitée aujourd'hui de Système d'Information Archéologique : SIA.