

Madame, Messieurs,

Au cours de mon DEA portant sur la refonte de bdB, base de données Beuvray, je me suis aperçue qu'une réflexion plus générale serait à menée sur le traitement de l'information archéologique. J'ai donc envisagé de préparer une thèse sur ce thème. Après consultation, Bruno Helly m'a proposé un sujet sur lequel je travaille depuis 5 ans et dont le titre est : "Nature, statut et traitements informatisés des données en archéologie : les enjeux des systèmes d'informations archéologiques".

L'objectif que je souhaitais donner à ma réflexion est devenu évident lorsque j'ai relu l'intervention de Jean-Claude Gardin au colloque du CNRS de Marseille de 1972. J'ai eu l'impression de lire ce que j'étais en train de vivre. J'étudiais à cette époque des bases de données. Je butais sur des problèmes d'homogénéités de toutes sortes pour lesquels je ne trouvais aucune réponse dans la littérature à ma disposition. J'en ai conclu qu'en trente ans, aucune réflexion de fond n'avait été menée, ou alors que cette réflexion était restée intimiste. Il était clair que je devais avoir pour objectif de répondre concrètement aux enjeux des bases de données. Ces enjeux sont d'assurer la pérennité des données et d'effectuer un croisement des données pour une exploitation plus large et plus riche de celles-ci.

Cette volonté de donner une réponse matérielle à mon travail vient de ma formation d'architecte. En effet en tant qu'architecte, je ne peux pas concevoir de mener un travail de réflexion sans proposer une solution concrète. À savoir ici une structure de bases de données répondant aux enjeux que j'ai mis en avant. Je propose donc un outil d'archivage des données archéologiques, outil fonctionnel et utilisable par tous. Cette formation m'a aussi apporté une méthode de travail qui je crois était nécessaire pour traiter ce type de sujet. L'architecte est amené à analyser les besoins de ses clients pour créer un programme collant au mieux à leurs desiderata. Sans oublier bien-sûr de prendre en compte la faisabilité. Et avant surtout toute création ou mise sur le papier. C'est cette capacité de programmation "purement intellectuelle" si j'ose dire, qui m'a permis de finaliser ce travail. Un archéologue se serait heurté à ses habitudes, ses pratiques de terrains, ses tics. Mais aussi il serait tombé dans le piège de se lancer dans la réalisation tout de suite et surtout sans avoir une idée assez claire et fiable du résultat final.

Pour répondre à cet objectif, j'ai mené en parallèle deux types de recherches. L'un portant sur la mise en place d'un vocabulaire, d'analyses nécessaires pour mener à bien une réflexion de fond. L'autre sur un groupe de bases de données et de normes utilisées.

Il me semblait indispensable de poser les définitions des termes spécifiques que j'utilisais. Non pas pour les imposer. Simplement pour être sûre que mon discours serait compris sans ambiguïté. Je devais définir quelle était la nature des données de l'archéologue, quels statuts elles pouvaient avoir et bien sûr comment elles étaient exploitées et archivées. Pour cela j'ai mené mon petit chemin épistémologique comme le montre ma bibliographie. Cette réflexion personnelle apparaît de manière brève dans mon mémoire. Je m'en explique. Je ne me sentais pas la capacité d'entreprendre pour un public plus large une réflexion que même les archéologues "confirmés" hésitent à entreprendre. Il est très rare et même pratiquement impossible de trouver dans les ouvrages de méthodologie de l'archéologie un discours réellement construit sur ce point. Je savais que si je rentrais dans cette voie de l'épistémologie, les archéologues se focaliseraient dessus et oublieraient le but premier que je fixais à mon travail. Du reste, aller au fond de ce travail épistémologique n'aurait, je crois, rien apporté de plus à mes interrogations. Je me suis donc limitée aux éléments qui m'étaient indispensables pour formuler mes idées.

Les données archéologiques pouvant être traitées au sein d'une base de données peuvent être rassemblées en deux grands types : les données primaires et les données traitées.

- **Les données primaires** sont toutes les données qui proviennent de l'observation directe du site, de la fouille et de la documentation que l'on peut en tirer. Ces données seront donc les analyses du mobilier, des objets et des couches, les liens directs entre ces différents éléments ainsi que les enregistrements des photographies, des minutes de terrain, des journaux de fouille, tout ce qui concerne la topographie du lieu.
- **Les données traitées** sont le résultat d'études menées sur les données primaires, d'assemblages de plan, de dessins d'objet ainsi que de tout ce qui touche à l'interprétation de ces données.

Une base de données, quant à elle, se définit comme un ensemble d'informations hétérogènes, organisées et évolutives susceptible de répondre rapidement et sélectivement à une demande spécifique.

Le deuxième exercice mené en parallèle, est l'étude de bases existantes destinées à l'archivage des données de fouilles. 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. Les bases de données étrangères ainsi que celles des institutions françaises à l'étranger ont été écartées. En effet, celles que j'ai pu observer ne sont pas si différentes de celles du corpus et n'apportent pas de points nouveaux significatifs.

Ces bases sont les suivantes :

- Base\_DFS du service régional de l'archéologie de la région Aquitaine, qui pourrait être considérée comme un archivage simple concernant surtout les données primaires ;
- bdB utilisée à Bibracte, et SysDA, à la maison de l'archéologie de Chartres qui ont pour but un archivage permettant de mettre à disposition du plus grand nombre le plus d'information possible dans un temps court ;
- ArchéoDATA par son utilisation par l'UTAH qui s'oriente vers un système universel ;
- Et SYSLAT qui est à la fois une méthode de fouille et une aide à la publication.

Ces études ont été menées sur des bases fournies par les concepteurs ou les utilisateurs. Suite à une analyse base par base, j'ai pris le parti d'associer entre eux les éléments de chaque base ayant les mêmes caractéristiques. J'ai pu extraire de ces associations des synthèses thématiques. Ces synthèses ont permis de mettre en avant les similitudes et les différences entre ces bases. J'ai grâce à cela pu faire apparaître deux composantes permettant de répondre aux enjeux définis ci-dessus.

La première composante consiste à 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. Vient ensuite la description homogène d'un même type de donnée. En effet pour pouvoir exploiter une base, il faut que toutes les données d'un même type 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. Il faut associer à ces champs le quatrième élément, le langage. Il est capital, pour s'y retrouver, de décrire des données identiques avec les mêmes mots ou termes.

Par l'examen de différentes normes disponibles, je peux dire qu'aujourd'hui il n'existe aucune homogénéité. PATRIARCHE comme les normes proposées par le CIDOC ne descendent pas au niveau de la donnée de fouille.

La deuxième composante est un tronc commun. Ce tronc commun regroupe les données et systèmes d'archivage minimaux qui sont nécessaires à la gestion des données issues d'une fouille archéologique. Il serait le prolongement de PATRIARCHE, l'échelon suivant qui intègre toutes les données issues de la fouille. Il ne s'agit pas, bien sûr, de faire une base nationale unique comme PATRIARCHE mais d'avoir des éléments communs de travail. Les bases seraient donc semblables, puisqu'elles utiliseraient les mêmes éléments de fondation. Ces bases seraient aussi indépendantes puisqu'elles pourraient évoluer à partir de ce tronc commun en fonction des besoins de chaque fouille.

J'ai décidé de traiter ce tronc commun comme la base de données de départ indispensable lorsque l'on commence l'archivage des données sur une fouille. Cette base minimale répond aux objectifs suivant :

- Recueillir les références de toutes les données et de la documentation issues de la fouille suivant des codes d'inventaires uniques pour chaque donnée.
- Organiser ces données et cette documentation, les décrire par l'intermédiaire de champs et d'un langage adaptés, ...
- Traiter 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.

Ce tronc commun contient les fichiers, rubriques et liens minimaux à la bonne gestion des données. Les fichiers sont au nombre de huit :

- entité archéologique PATRIARCHE
- chantier
- photographie
- documentation graphique
- UF
- mobilier
- objet et prélèvement
- sources documentaires

Au fur et à mesure que des besoins spécifiques apparaîtront, ceux-ci seront traités suivant deux protocoles possibles :

- soit par l'intermédiaire de nouvelles rubriques au sein d'un fichier existant,
- soit par l'ajout d'un ou de plusieurs nouveaux fichiers à la structure de la base de départ.

L'archivage de ces nouvelles données devra se référer au système de codes d'inventaires mis en place.

L'association de ces deux composantes permet d'atteindre les enjeux essentiels précités : la mise en cohérence des bases de données archéologiques. Les éléments homogènes permettent de faire des recherches croisées. Ces éléments et le tronc commun assurent la pérennité des données. En effet les données d'une base pourront toujours être transférées vers une autre base, quel que soit le support de celle-ci.

Beaucoup s'étonneront que cette solution se limite à une structure papier et à un inventaire complet des rubriques de chaque fichier. L'explication est très simple. J'ai pu constater combien le choix d'un logiciel pour l'exploitation de l'information et d'une charte graphique pour la présentation de cette information, pouvait jouer contre un outil qui disposait d'atouts non négligeables. Recommander tel ou tel de ces deux instruments n'est pas l'essentiel, pour nécessaires et évidents qu'ils soient l'un ou l'autre. À chacun donc de choisir selon ses préférences sachant, je le rappelle, que la pérennité des bases n'est pas assurée par un support informatique particulier.

Je terminerai en disant que cette base minimale a été conçue en suivant étapes par étapes un protocole de création de bases de données que j'ai mis en place. Ce protocole permet de gérer toutes questions soulevées lors de la création d'une base. Il est traité en onze points hiérarchisés et chronologiques qu'il est indispensable de suivre pour créer une base de données logique et structurée. Il se base sur le long parcours suivi par les concepteurs des bases du corpus qui se sont plus ou moins tous heurtés aux mêmes impasses et aux mêmes problèmes. Ce protocole, s'il est suivi à la lettre, devrait permettre d'éviter ces impasses, de limiter les écueils, les retards et les dysfonctionnements d'une base. Il a bien sûr été testé lors de la mise en place d'une base de données portant sur l'étude de l'iconographie hindoue des 11<sup>e</sup>-13<sup>e</sup> siècles. Cette base de données, nommée LILA, est effective depuis un an et répond parfaitement aux besoins exprimés.

Je tiens à profiter de la présence dans cette salle de nombreuses personnes qui m'ont aidé dans la réalisation de ce travail, pour les remercier encore pour leurs appuis, conseils et critiques. Sans eux, et là je pense surtout aux concepteurs qui m'ont fourni sans réserves les bases de données que j'ai étudiées, ce travail n'aurait pas abouti aux résultats concrets que je viens de vous présenter.

Je vous remercie de votre attention.