

## ARCHÉOLOGIE ET INFORMATIQUE AUJOURD'HUI: QUELQUES IDÉES POUR UN DÉBAT \*

Il n'est pas question ici de faire l'historique de notre discipline au cours des trente dernières années; mais on peut au moins rappeler que toutes nos recherches sont parties d'une démarche analytique: ainsi les travaux de J. Deshayes et de J.-C. Gardin sur l'outillage en métal de l'âge du bronze, dans les années 50, impliquaient en priorité la mise au point d'un code descriptif. La mécanisation proposée pour la manipulation de ces données reposait, dans un premier temps, sur l'emploi de ce qu'on appelait la mécanographie, puis sur des techniques optiques. Mais c'est l'informatique qui a ensuite ouvert des voies nouvelles, et on a vu alors se développer les recherches et les expérimentations sur les deux étapes essentielles de notre travail d'archéologues: la recherche documentaire d'une part, d'autre part l'aide au raisonnement, avec d'abord l'application de méthodes d'analyse mathématique et statistique, puis par des recherches sur les processus cognitifs et plus particulièrement les systèmes experts.

Au point de vue de la technique, on sait combien les outils ont changé: les matériels et les logiciels dont nous disposons maintenant rendent l'utilisation de l'informatique beaucoup plus facile, plus directe, et, du moins en principe, plus efficace. Les machines sont devenues plus puissantes, et d'accès plus aisé, aussi bien physiquement qu'intellectuellement, les logiciels sont plus efficaces et plus conviviaux, et de nouvelles fonctionnalités nous sont offertes par l'utilisation du mode graphique et par l'enregistrement et la manipulation des images, tandis que les nouveaux supports de stockage nous apportent des possibilités inimaginables il y a encore peu de temps. Tous ces phénomènes sont bien connus, et, pour suivre cette évolution et en tirer le meilleur parti, l'essentiel des recherches a été d'ordre technique: on peut en mesurer les résultats, avec des progrès qui continuent et s'accélèrent. Ce sera donc un des objectifs de ce colloque que de nous tenir au courant.

Mais il faut signaler en outre que, pendant ces trois décennies, la réflexion a porté aussi sur les systèmes descriptifs: beaucoup ont été mis au point, portant sur plusieurs champs de l'archéologie, et ils constituent une avancée pour la discipline, car ils sont largement indépendants des conditions techniques de leur utilisation, et servent même dans la pratique la plus traditionnelle.

Aussi, nous disposons à l'heure actuelle d'outils pour nous aider à peu

\*A la suite d'un problème de transmission entre Paris et Bilbao, le texte de la conférence prononcée par René Ginouvès et Anne-Marie Guimier-Sorbets lors de l'ouverture du 1er Colloque international d'Archéologie et Informatique, tenu à Saint-Germain-en-Laye en Novembre 1991, n'a pu être publié dans les Actes (qui viennent de paraître à la fin de 1995). Nous le reproduisons ici sous sa forme originelle, à la mémoire de René Ginouvès et comme témoignage de l'état, à cette date, de la réflexion sur les bases de données documentaires et les publications en Archéologie.

près à toutes les étapes de notre travail d'archéologues, de l'interprétation des données de prospection jusqu'à l'aide à la publication, en passant par toutes les phases intermédiaires. C'est pourquoi l'objectif de notre colloque, tout comme des précédents, et, nous l'espérons, des suivants, doit être d'abord de nous informer sur les pratiques des uns et des autres, qu'elles en soient au stade opérationnel ou seulement encore à celui de l'expérimentation; mais il nous faudra aussi réfléchir sur l'état actuel de la recherche, et tenter d'en dégager les perspectives. Les exposés, les démonstrations et la table-ronde devraient nous permettre d'atteindre ces deux objectifs. Sans intervenir dès maintenant dans la totalité des perspectives qui seront ouvertes au cours de ces journées, nous nous contenterons ici de proposer quelques idées pour lancer le débat sur les travaux à visée documentaire.

En ce qui concerne ces banques de données documentaires, qu'il s'agisse de l'archivage de résultats de fouilles, d'inventaires de musées ou du patrimoine national, de corpus thématiques, de bibliographies, de photothèques, etc., il faut reconnaître, d'une part, qu'on sait les réaliser et, d'autre part, qu'il en existe qui sont utiles et utilisées effectivement pour la recherche, enfin qu'on continue à mettre à jour celles qui existent et à en entreprendre de nouvelles. Ne voyez aucun triomphalisme dans cette triple constatation; mais comme nous allons devoir souligner les inconvénients que présentent ces banques, pour tenter d'y porter remède, il vaut mieux partir d'un constat positif : si d'ailleurs on ne le faisait pas, il serait inutile d'aller plus loin.

Pourtant, nous savons tous que la réalité n'est pas aussi satisfaisante, que les grandes banques de données conçues dans l'enthousiasme n'ont pas donné tous les résultats escomptés, que certaines ont été abandonnées, ou alors qu'elles sont très critiquées, et qu'on assiste en revanche à la multiplication anarchique des fichiers informatisés de chercheurs, qui ne répondent pas du tout aux mêmes besoins. On a déjà souligné à plusieurs reprises les raisons techniques, institutionnelles et psychologiques de cet état de fait, et point n'est besoin de les rappeler, mais il faut au moins revenir sur les caractéristiques communes de ces entreprises.

Et d'abord celle-ci : à l'heure actuelle, ce sont en majorité des banques faites par des chercheurs et destinées à des chercheurs, soit dans la perspective de l'information spécialisée, c'est-à-dire scientifique et technique, soit dans la perspective traditionnelle de nos publications. En second lieu, ces banques contiennent, outre éventuellement des images graphiques ou photographiques, des informations qui ont été préparées spécifiquement pour la banque, à travers un système d'analyse particulier, qui sert également pour leur interrogation; cette interrogation est généralement de type booléen (avec des possibilités combinatoires plus ou moins étendues selon le type de logiciel), et elle fournit à l'utilisateur des listes de documents pertinents, sous-ensemble des documents de la base qui correspondent exactement aux critères exprimés dans sa question.

Cette démarche est, aujourd'hui encore, la plus efficace. Mais elle entraîne des inconvénients dont il faut être conscient. Pour que la banque de données documentaire apporte le résultat escompté, il faut d'abord que l'information recherchée par l'utilisateur soit enregistrée dans la base, alors que, bien évidemment, des choix ont été nécessairement faits lors de cet enregistrement; il faut ensuite que l'information soit analysée selon un système prédéfini; et il faut enfin que l'utilisateur connaisse ce système. Or, plus la base est riche, et donc comporte de types d'informations analysées, plus elle est lourde à constituer et à mettre à jour, et plus elle est complexe à interroger dans tous ses détails. En outre, le système est fermé, car si on peut et doit mettre à jour l'information retenue, il n'est pas toujours facile – ni même souhaitable – d'ajouter de nouvelles catégories d'information, qui n'intéressent éventuellement que quelques chercheurs.

En conséquence, on se défie maintenant à juste titre des larges entreprises documentaires sans visée d'utilisation immédiate, et on multiplie les fichiers individuels qui, au mieux, répondent aux besoins du chercheur qui les constitue, mais sont totalement perdues pour la communauté scientifique. Et on arrive au paradoxe suivant: nous avons pensé que les banques de données permettraient d'éviter le feuilletage toujours recommencé des mêmes publications par différents chercheurs; maintenant, nous en sommes à l'enregistrement toujours recommencé des mêmes documents par différents chercheurs, pour des perspectives à chaque fois à peine différentes. Et, si on se place dans l'optique d'une institution qui cette fois a une obligation d'archivage, la situation est encore pire, car l'archivage n'est pas vraiment assuré ou, s'il l'est, il présente toute une série de défauts maintes fois soulignés.

\* \* \*

Dans ces conditions, que faut-il faire? On peut, à ce qu'il semble, distinguer deux cas.

Dans le premier, les données ont une "valeur générale", c'est-à-dire sont directement utilisées ou utilisables par un grand nombre de chercheurs: c'est le cas, par exemple, pour les textes anciens avec leur index (comme le *Thesaurus Linguae Graecae*), ou encore de la bibliographie de l'Antiquité. A condition de trouver une institution, publique ou privée, pour constituer la banque, nous avons maintenant des moyens de diffusion qui en permettent l'exploitation directe sur des sites très différents, voire même sa réutilisation dans des entreprises plus spécialisées. La solution actuelle pour cette diffusion est évidemment le CD-Rom, dont le coût de fabrication est amorti seulement si le produit est vendu en un nombre suffisant d'exemplaires: c'est pourquoi cette solution ne peut s'appliquer que pour des données d'intérêt général.

Dans le second cas, où les données sont plus spécifiques, que peut-on faire? Il faudrait, à l'évidence, trouver un type d'enregistrement minimum, qui comporterait des informations "de base", analysées sous une forme "com-

muné” et donc facilement transposable d’une application à une autre. Chacun pourrait ainsi exploiter ces données de base dans des environnements différents, et les enrichir selon sa propre visée d’étude. On sait qu’aujourd’hui les problèmes informatiques de compatibilité sont moins graves qu’ils ne l’étaient, mais ne sont pas résolus pour autant. Il importerait donc, dans la perspective proposée, de se mettre d’accord à l’avance sur des modèles communs d’analyse des données, par domaine scientifique ou par type d’entreprise, et plus généralement si possible. La proposition n’a rien de très révolutionnaire ni de très exaltant, et en outre elle est plus facile à énoncer qu’à réaliser. Pourtant notre Centre, par exemple, a mis au point, pour l’École française d’Athènes, des formats d’échange pour que les données puissent être communiquées entre banques de données factuelles et banques référentielles, qu’il s’agisse de banques appartenant à l’École ou de fichiers personnels construits par ses Membres, dans des environnements aussi bien MacIntosh que Dos ou UNIX; nous avons aussi mis au point, avec un logiciel particulier, la possibilité pour l’utilisateur d’ajouter ses propres informations à celles qu’il tire de la banque générale.

Ce ne sont certes là que des débuts ponctuels, mais l’espoir tient au fait que plusieurs d’entre nous en arrivent à des solutions de ce type. Certains vont ici proposer d’autres pistes de recherche et des commencements de réalisation, en partant soit d’un cadre disciplinaire assez large, par exemple pour les travaux présentés par la Maison de l’Orient à Lyon, ou dans un cadre institutionnel particulier comme pour la Banque des Biens culturels suisses: la discussion sur ces formules devrait être enrichissante.

\* \* \*

Mais peut-être convient-il d’évoquer aussi, même rapidement, un autre aspect de ces entreprises documentaires au sens large, à savoir leurs rapports d’une part avec la publication de type traditionnel, d’autre part avec la diffusion des connaissances auprès des chercheurs, mais aussi d’un public plus large.

Et d’abord, la publication. On sait que l’archéologue doit enregistrer de plus en plus d’informations relatives à ses activités de chercheur, données de fouilles en particulier, description détaillée du matériel, etc.; or il n’est plus possible de publier la totalité de cette information sous la forme de l’imprimé. La tentation est évidemment grande de la mettre dans une machine conçue comme une grande armoire de stockage. Dans quelle mesure doit-on le faire, et surtout sous quelle forme devrait-on alors stocker ces données? Quels outils faut-il mettre au point pour pouvoir ensuite accéder à des données et les exploiter? Nous aurons certainement à débattre des recherches qui sont menées, des solutions qui commencent à être proposées pour répondre à cette fonction de stockage de l’information.

En ce qui concerne l’autre fonction, la consultation de l’information,

les nouvelles technologies, au sens large du terme, nous permettent de manipuler des données de types très divers, images graphiques et photographiques, textes en langage naturel etc., et ce à travers des procédures très diverses aussi, non plus seulement les bases de données de type booléen comme celles dont il a été question d'abord, mais par exemple à l'aide de processus d'hypertexte-hypermédia. On voit aussi se développer des recherches sur des types de consultation tout à fait différents de la consultation classique des bases de données, mais aussi de la consultation traditionnelle de l'imprimé. Il faudra se demander comment nous pouvons les mettre en oeuvre, et cela dans diverses perspectives.

Car, si nous avons les moyens techniques, et si nous reconnaissons l'intérêt scientifique, d'enregistrer des données de types hétérogènes, on peut se demander si nous pouvons en ouvrir la consultation à des publics différents de la clientèle habituelle de nos banques de données scientifiques, qu'il s'agisse d'étudiants en cours de formation ou d'un public plus large intéressé par nos disciplines, et on sait que ce public est de plus en plus nombreux.

Comment construire des systèmes d'information rassemblant des données très hétérogènes quant à leur nature physique et quant à leur contenu, tout en donnant l'accès à des publics aussi divers que le spécialiste du domaine, l'étudiant ou le public intéressé ? Des produits répondant à ces exigences commencent à apparaître, qu'il s'agisse par exemple du *Perseus Project* ou des réalisations du Musée du Louvre. Pourra-t-on aller plus loin ? La recherche devrait alors porter sur les techniques et les outils de mise en oeuvre, mais aussi sur l'identification des publics visés et de leurs besoins. Notre Centre a commencé une recherche sur ce sujet, et nous vous en présenterons un aspect, utilisant les fonctionnalités de l'ingénierie linguistique et de la perspective documentaire non booléenne puisqu'il s'agit de permettre la consultation, en langage naturel, d'un catalogue archéologique de type traditionnel, évidemment rédigé en langage naturel et décrivant des sculptures hellénistiques, domaine difficile à formaliser s'il en est.

Et déjà il est possible de formuler quelques questions: en quoi le type de consultation électronique diffère-t-il de la consultation linéaire de type traditionnel ? Comment peut-on s'en servir pour accéder à des publications traditionnelles ? Ou bien faut-il préparer des produits spécifiques pour ce genre de consultation ? Quel est l'apport relatif des différents types de données (produits formalisés d'analyse, textes en langage naturel, images, etc.) ? D'autres exposés viendront certainement enrichir cette première réflexion, et déjà notre table-ronde devrait nous apporter certaines réponses.

RENÉ GINOUVÈS, ANNE-MARIE GUIMIER-SORBETS  
Centre de Recherche "Archéologie et Systèmes d'Information"  
Université de Paris X - CNRS