les banques de données, une réponse au défi des informations

LE SYSTEME FRANÇAIS D'ACCUMULATION DE DONNEES LOCALISEES

L'informatique aura-t-elle une autre vertu dans le monde de demain ? Celle de permettre à l'homme de maîtriser la connaissance qu'il a accumulée.
Le pire des « encombrements » de l'avenir ne sera pas l'encombrement physique ou même démographique, mais l'encombrement des informations.
La mémoire de l'homme n'est plus suffisante pour les enregistrer.
La diversité des données est telle qu'elles ne pourraient plus être mobilisées pour les décisions de plus en plus globales que la société se doit de prendre.
Banques de données et centres de documentation automatique répondront-ils à ce défi ? Quelles différences y a-t-il entre eux ?
Jean-Paul Trystram, qui a créé en France l'un des premiers systèmes d'accumulation de données, analyse ces instruments nouveaux.

Jean-Paul Trystram, directeur de Recherches à l'Institut de recherches d'informatique et d'automation (Iria - France), et un terminal du Sad.

Banque de données et centre de documentation : les différences

Dans un sens large, tout centre de documentation pourrait être considéré comme une banque de données, mais il semble préférable de restreindre la signification du terme ; on peut alors dire que, parmi les centres de documentation, on appellera « banques de données » ceux qui utilisent un ordinateur, aussi bien pour stocker l'information que pour la retrouver. Cette définition exclut alors les centres où seule la recherche du document nécessite l'utilisation d'un ordinateur, celui-ci restant stocké sous une forme traditionnelle. C'est le cas des centres de documentation automatique qui permettent de retrouver des références à des documents, ces références étant parfois seulement un numéro d'ordre qui renvoie à une liste que l'on consultera séparément, ou, dans d'autres cas, une fiche bibliographique plus complète, pouvant comporter un résumé du document. Mais — et c'est là ce qui fait toute la différence — le document lui-même n'est pas entièrement enregistré dans la mémoire de l'ordinateur.

La banque de données se distingue donc du centre de documentation automatique par le fait que la totalité de l'information y est enregistrée.

Par contre, ce n'est pas la nature de l'information qui permet de définir les banques de données ; elles peuvent tout aussi bien concerner des composants électroniques que des corps chimiques, des maladies que des stocks industriels. D'une façon très générale, une banque de données a pour base un ou plusieurs fichiers, comportant, pour chaque unité — représentée par une fiche —, un nombre plus ou moins important de caractères qui permettront de sélectionner la ou les fiches correspondant à une demande d'information.

On peut encore classer les banques de données dans deux grandes catégories : celle où le produit final est la fourniture de l'information telle qu'elle est enregistrée, et celle où des opérations peuvent être faites sur cette information de base et où l'utilisateur obtient donc, s'il le désire, aussi bien des informations élémentaires que des informations traitées.

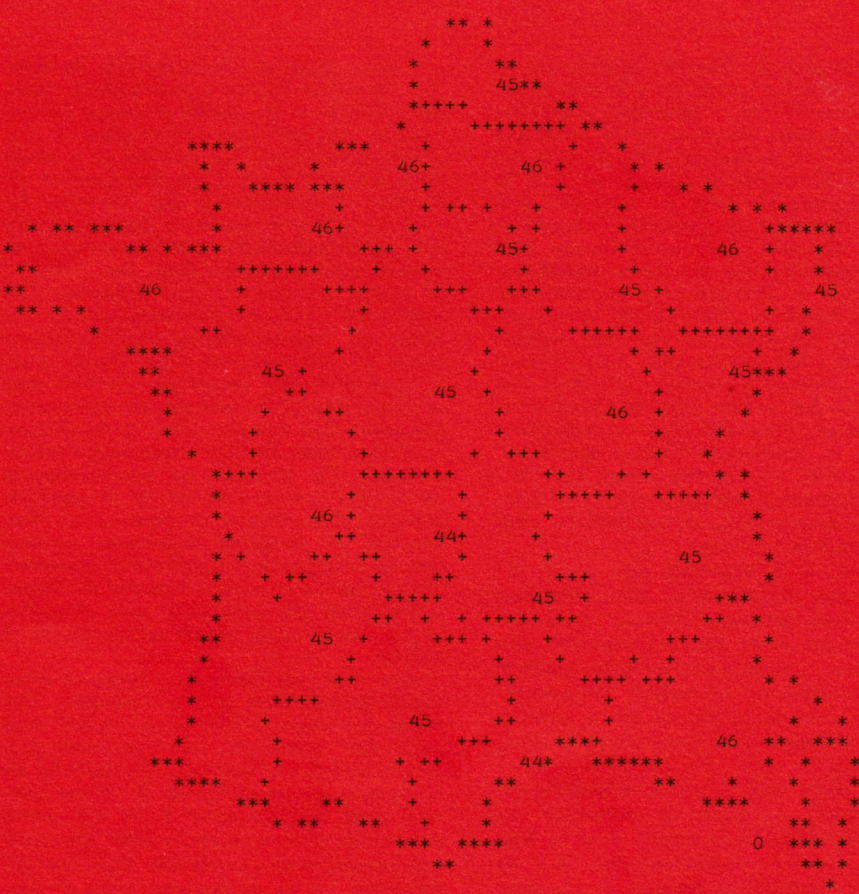
Les limites de la banque de données

On entend parfois des critiques systématiques des banques de données ; elles sont le plus souvent mal placées. Certains détracteurs feignent de croire que les instigateurs des banques de données voient dans l'outil qu'ils forgent l'unique source de progrès des sciences humaines, alors qu'ils sont les premiers à ramener les enthousiasmes à un niveau plus modeste, connaissant aussi bien les possibilités que les faiblesses de leur outil. Pas plus que l'ordinateur ne peut remplacer l'intelligence humaine, la banque de données ne peut remplacer une juste conceptualisation des problèmes. Il y a d'ailleurs, à ce point de vue un changement d'échelle qui peut rendre possible des travaux que l'on ne pouvait envisager auparavant. Obtenir des informations très variées dans une fraction de minute, permet de tester des idées — à condition d'avoir assez d'imagination — en évitant les pertes de temps considérables qui sont encore, à l'heure actuelle, inévitables pour rassembler des informations manuellement.

La critique dépasse d'ailleurs souvent les banques de données, elle s'adresse à toute utilisation des statistiques, et même, jusqu'à un certain point, à l'intérêt de la recherche scientifique. Les remises en cause de la société contemporaine et une importance plus grande donnée à l'action politique orientent actuellement les esprits dans d'autres directions que celle de la patiente analyse des situations à travers des outils aussi objectifs que possible. D'autre part, les banques de données souffrent dans le grand public de la crainte superstitieuse que certains ont encore devant l'ordinateur, et du souci (qui n'est pas toujours inutile) de protéger la liberté individuelle contre la mise en fiche systématique de toutes les activités humaines.

Les banques de données ne sont que des outils ; elles n'ont de valeur que par l'habileté de ceux qui les utilisent. C'est actuellement surtout en amont et en aval que de graves problèmes se posent ; en amont, c'est tout l'appareil statistique qui est mis en cause : toute banque de données dépend étroitement de la qualité des bases de données ; en aval, on se heurte encore, en plus des obstacles psychologiques rappelés ci-dessus, à la faiblesse des méthodes qui permettraient d'utiliser au mieux des masses importantes de données. C'est surtout des progrès réalisés dans ces deux directions que l'on peut espérer une amélioration de l'utilisation des banques de données.

J.-P. T.



Carte établie par un terminal imprimant (type clavier de télex) du Sad. Les renseignements peuvent être demandés par un utilisateur situé à une grande distance de la banque de données (un maire, un économiste peuvent interroger un centre régional).

Un exemple français de banque de données : Le système d'accumulation des données localisées

Les banques de données statistiques se prêtent particulièrement bien à des traitements, étant donné le caractère numérique de l'information de base et la puissance de calcul des ordinateurs. C'est dans cet esprit que la banque de données de la Délégation à l'aménagement du territoire et à l'action régionale (le Système d'accumulation des données : Sad), a été conçue. Elle vise à connaître par des séries statistiques territoriales la diversité géographique française, et l'évolution des régions, des départements et des villes les unes par rapport aux autres ; elle permet aussi bien de fournir des données statistiques de base que d'effectuer des calculs simples (corrélation, pourcentage, taux d'évolution, projections linéaires et exponentielles, etc.). Les résultats peuvent être imprimés sous trois formes différentes : tableaux, cartes géographiques, diagrammes, ou mémorisés pour une utilisation ultérieure.

Lorsque le besoin s'en fait sentir, les informations statistiques de base, aussi bien que les informations traitées, peuvent être réintroduites dans des programmes plus complexes. Ces programmes sont indépendants de ceux de la banque de données auxquels ils sont reliés par un module de passage permettant de transmettre les données utiles.

On a choisi cette structure modulaire des programmes, de façon à minimiser les temps d'accès à l'information dans la banque de données proprement dite. Il a en effet paru essentiel de faciliter, le plus possible, l'utilisation de la banque de données par toute personne intéressée. De ce principe découlent plusieurs conséquences, dont l'une d'elles est la rapidité d'accès à l'information. Une option essentielle dans ce sens a été l'utilisation d'un système de « temps partagé » commercial, comme support du Système d'accumulation des données. Le « temps partagé » permet à des utilisateurs occasionnels de disposer de la puissance de grands ordinateurs pendant des fractions de temps très courtes, sans dépendre d'un ordinateur particulier n'ayant qu'un nombre limité de moyens d'accès, habituellement situés à proximité même de l'ordinateur. Le Sad peut être utilisé aujourd'hui, sous sa forme la plus évoluée par toute personne disposant d'un terminal, d'une prise de courant électrique, et d'une ligne téléphonique normale. Le terminal est essentiellement un télétype formé d'un clavier de machine à écrire, d'un rouleau de papier pour l'impression et d'organes techniques permettant la relation avec l'ordinateur, complété parfois par un lecteur-perforateur de bande. Certains de ces terminaux sont portables et sont logés dans deux valises faciles à manier.

Avec quelques limitations dues aux vitesses de transmission et à la dimension plus petite des chariots d'écriture, le Sad peut aussi être utilisé par toute personne possédant un télex classique. L'accessibilité de la banque de données de la Datar permet donc théoriquement à un très grand nombre de personnes de l'utiliser sans qu'il soit nécessaire de s'encombrer d'un matériel important, et ce en quelque point du territoire que l'on se trouve.

Une banque de données à la portée de tous

De même qu'on a cherché à faciliter l'accès sur le plan matériel, on a fait un effort tout particulier pour que l'apprentissage soit aisé. L'un des grands problèmes des banques de données est de faire en sorte que les « décideurs » s'en servent et posent les questions convenables. Le Sad peut être utilisé par des personnes qui ignorent tout de l'informatique.

Il est certes nécessaire d'apprendre à utiliser le terminal et de connaître les mots de reconnaissance qui permettent d'obtenir la connexion avec l'ordinateur ainsi que la façon de lui poser des questions. Mais il s'agit de codes faciles à mémoriser : tous les ordres sont composés d'une partie alphabétique de trois lettres qui sont presque toujours les trois premières lettres d'un mot français : une addition sera demandée par l'ordre ADD, une corrélation par COR, l'impression d'un tableau par IMP, l'établissement d'une carte par CAR, etc. Le Système d'accumulation des données est utilisé en « mode conversationnel » c'est-à-dire que l'utilisateur entreprend une véritable conversation avec l'ordinateur. Les réponses à une question l'incitent à en poser une autre qu'il n'avait pas prévue au début d'une séance de travail.

Il faut aussi savoir interpréter ensuite ces constatations ou corrélations ; la banque de données ne fait rien d'autre que faciliter les calculs, l'interprétation des résultats qui n'est pas toujours aisée, est ensuite du ressort de l'utilisateur.