

L'INFORMATIQUE DOCUMENTAIRE A L'ORSTOM

■ Interview de Jean-Paul Roux-Fouillet, responsable du secteur documentaire

Pour mener à bien ses missions, l'Orstom dispose de nombreux centres de documentation et d'une informatique documentaire qui s'est largement développée en France et à l'étranger. Bases de données internes, serveur ouvert à la clientèle ou CD/Rom, composent une large palette d'outils documentaires informatisés. Nous avons interviewé M. Jean-Paul Roux-Fouillet, responsable du secteur documentation et administrateur de la base bibliographique Horizon pour faire le point sur l'informatique documentaire à l'Orstom.

Archimag. Quel est le rôle de la documentation à l'Orstom ?

Jean-Paul Roux-Fouillet. Il est très important. Il ne faut pas oublier que l'Orstom réalise essentiellement de la recherche fondamentale. 50 % du personnel est composé de chercheurs. De plus, nous comptons un très grand nombre d'ingénieurs et de techniciens. Leur travaux s'exercent sur tous les continents. La recherche, orientée vers le développement, en coopération, touche des disciplines aussi stratégiques que l'agronomie, les maladies tropicales, la géologie, l'ethnographie. Nous avons même des linguistes qui travaillent

sur le dialecte Peul afin d'aider à la créolisation de ce langage, c'est-à-dire une fusion/mélange entre cette langue d'origine et la langue environnante. Pour toutes ces missions, les chercheurs ont besoin d'une importante documentation. C'est notre premier axe, classique, qui vise à rassembler et mettre à leur disposition ouvrages, articles, périodiques, rapports et littérature grise (1). Le second axe, et j'insiste sur ce point, vise à valoriser les publications, monographies et réalisations de l'Orstom, entre autre par la mise en

(1) La littérature dite grise concerne tout ce qui n'est pas commercialisés : rapports et publications internes.

ligne sur notre propre serveur de la base Horizon, accessible via Transpac. Celle-ci recense la production scientifique de l'Orstom, y compris des articles de chercheurs publiés à l'étranger. Elle compte actuellement 25 000 références et montera à 35 000 fin 91. Nous avons également produit en 1989, un CD/Rom, en collaboration avec l'INRA (Institut national de la recherche agronomique), le CIRAD (recherche agronomique tropicale), le BPRA, etc. Le CD/Rom Sésame est un produit thématique sur l'agronomie et l'agriculture. Une deuxième version est prévue pour octobre 1990. Enfin, nous exerçons un rôle d'ingénierie documentaire et de trans-

fert de savoir-faire pour réaliser des centres de documentation et des bases de données, clé en main, formation comprise. Nous aidons aussi à la reconstitution de fonds documentaires locaux, pays par pays, avec transfert de documents, tout cela afin de favoriser l'autonomie locale dans le cadre des politiques nationales. Exemple : à Madagascar et au Cameroun.

De quelle informatique documentaire disposez-vous ?

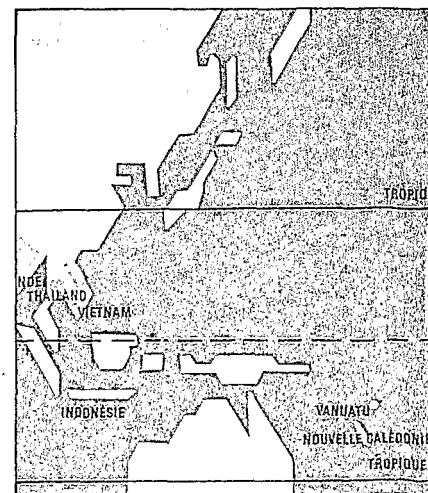
Nous avons démarré par des bases de données sous MS/Dos et sur micro-ordinateur. Elles sont été multipliées en divers points en France et à l'étranger.

On connaît bien le CNRS mais on connaît moins l'Orstom. L'Orstom, Institut Français de Recherche Scientifique pour le Développement en Coopération, est un Etablissement Public Scientifique et Technique (EPST) dépendant des ministères de la Recherche et de la Coopération. Il couvre quarante pays avec cinquante implantations dans l'Afrique Francophone, l'Océan Indien, le Pacifique, l'Amérique Latine et le Sud-Est avec un total d'emplois de 2 500 si on y ajoute tous les

contrats en cours dans les zones concernées.

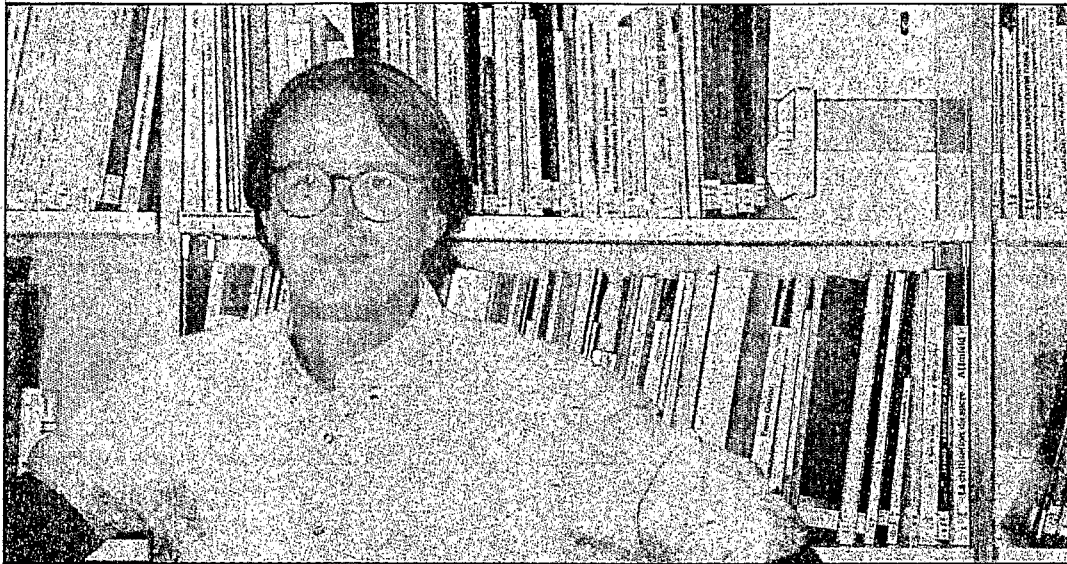
L'Orstom exerce une activité de recherche fondamentale et de recherche appliquée contribuant au développement. Ces recherches couvrent les sciences de la terre, les sciences économiques et sociales et les sciences de la vie, c'est-à-dire aussi bien l'océanographie, le climat, la santé que l'agronomie, la linguistique, l'urbanisation, la consommation, etc.

Deux tiers du personnel travaillent à l'étranger.



ORSTOM Fonds Documentaire

NTAIRE



Les chercheurs ont besoin d'une importante documentation.

Nous avons petit à petit évolué vers Unix et les stations de travail Sun. Pour le fonds général Orstom, trois centres de documentation - été mises en œuvre sur micro à travers le monde : Quito, Pointe à Pitre, Yaoundé... Nous travaillons uniquement avec le logiciel Texto, tant sous MS/Dos que dans

chacun étant spécialisé - sont informatisés à Paris, à Bondy et Montpellier. L'ensemble couvre toutes nos spécialités. Deux importants centres de documentation sont également informatisés à Nouméa et à Abidjan. Enfin, vingt autres bases de données non spécialisées mais globales ont

sa version Unix. Des réseaux locaux Ethernet permettent, le cas échéant, la consultation décentralisée. Via Transpac, chacun interroge à distance nos trois serveurs métropolitains. Ils sont accessibles également à l'étranger, exemple : au Cameroun via Campac, puis Transpac. Tout cela, pour

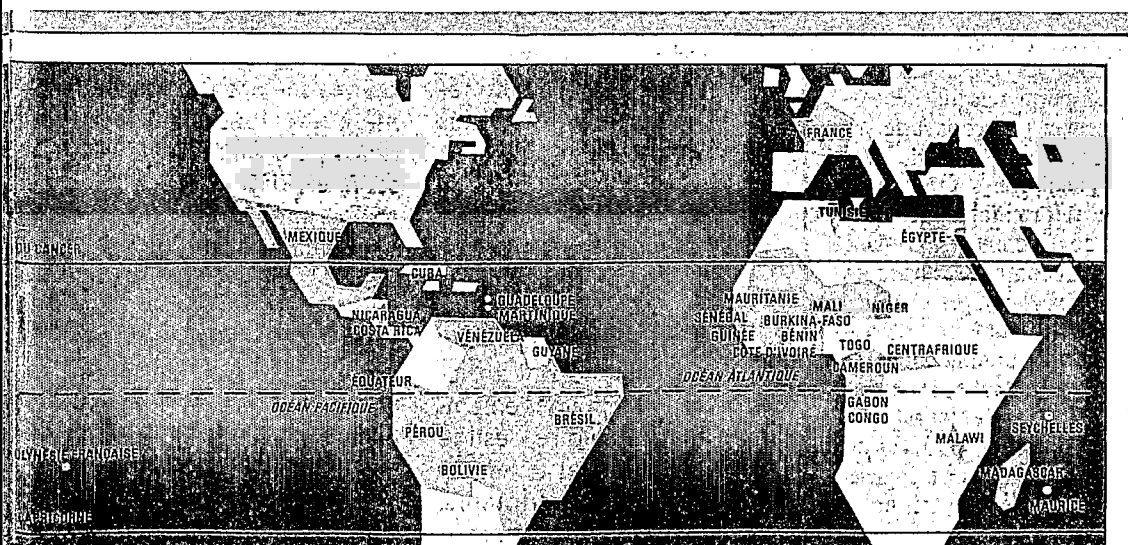
nos utilisateurs externes. Enfin, le serveur externe (et interne) Horizon est à Bondy. Pour tout utilisateur externe comme pour nos chercheurs.

Comment et pourquoi avez-vous effectué ces choix ?

Notre approche a été et reste pragmatique. Nous avons choisi la micro informatique parce que le ticket d'entrée, du point de vue coût, est accessible et que l'appropriation par les utilisateurs est efficace. La micro-informatique peut évoluer simplement au fur et à mesure des besoins vers une architecture répartie en réseau. Lorsqu'une puissance supérieure était ponctuellement nécessaire, nous avons alors intégré des Sun comme serveurs sous Unix. Nous disposons maintenant de cinq stations de travail Sun. Nous utilisons Texto pour plusieurs raisons. Dès 1985, il y avait une version Dos et Unix. Nous pouvions envisager des passages de Dos à Unix et ensuite une large portabilité. C'est un facteur primordial. Texto est aussi un produit éprouvé. D'ailleurs, nous n'avons jamais eu de problèmes de transferts de fichiers. Ensuite, Texto est un logiciel répandu et nombre de nos partenaires l'utilisent. Cela facilite les échanges de fichiers et la réalisation du CD/Rom par exemple. Enfin, ce logiciel couvre efficacement l'ensemble des fonctions documentaires : saisie, recherche, éditions, tris, téléchargement, import-export de données...

Comment utilisez-vous l'informatique documentaire ?

Au niveau de l'organisation ►



► du travail nous ne séparons pas l'indexation et la recherche. Pour des raisons de qualité du travail, de communication et de polyvalence, il y a rotation des tâches. Cela rend également le travail plus attractif et varié. C'est important. Pour la recherche courante d'un poste de consultation à Paris ou à Yaoundé, nous procédons aux interrogations et éditions sur la base locale, sur Horizon, ou sur les domaines documentaires, de Montpellier ou de Bondy. Les postes permettent d'interroger aussi les grands serveurs comme Dialog, Questel et, le cas échéant, de procéder à des téléchargements pour un besoin large. Nous reformatons ces données en format Texto. Nous disposons également d'une messagerie et

sur le même poste pouvons réaliser du traitement de texte.

Quelles évolutions et perspectives envisagez-vous ?

Nous allons d'abord évoluer sur ce principe d'architecture répartie avec passage progressif sur Unix et utilisation ponctuelle de stations Sun. Ensuite, nous avons des projets PAO (Publication assistée par ordinateur) et archivage électronique, DON notamment ; de la PAO aussi bien pour la documentation (retrouver un document) que pour le retravailler, le rééditer. Bien sûr, nous avons donc une conception d'un poste de travail intégré. Enfin, nous allons poursuivre des éditions CD/Rom thématiques, en sciences de la terre, par exemple. □

TEXTO

Texto est un logiciel documentaire ainsi que de gestion de bibliothèques. Il permet la saisie interactive via un fichier éditeur de l'ordinateur ou via un fichier traitement de texte, comme la saisie conversationnelle. Les champs, leurs valeurs et la structure des bases sont paramétrables et modifiables. Les fonctions de création, mise à jour, corrections, suppressions et statistiques sont assurées. L'interrogation est multi-critères et multi-champs avec des opérateurs logiques "et, ou, sauf", les opérateurs de comparaison numérique, la troncature droite, gauche et le masque. Différents tris et formats d'édition et visualisation sont possibles ainsi que des sorties photocomposition et transferts de fichiers. Texto, fonctionne sous Dos et Unix et de

surcroît sur Bull, IBM, HP, Dec, Data General, Control Data en version micro, mini, mainframe. Des progiciels associés enrichissent les possibilités. Logotel permet de personnaliser les menus, de piloter Texto en mode vidéotex, de gérer un thésaurus, des commandes et les prêts. TexLoad permet le téléchargement en provenance de serveurs tels que Questel, Data Star, Dialog, Esa, Orbit, etc... et le reformatage automatique des documents. Dv-One pilote et consulte les banques de données images ou films sur vidéodisques. Enfin AutoX est un programme d'indexation automatique de textes. Texto compte 1 500 références en France et 300 à l'étranger depuis 1977. Le logiciel est disponible en de très nombreuses langues.

THESAURUS : OUI OU NON

Le thésaurus est généralement considéré comme un outil important de la gestion documentaire. Beaucoup de logiciels documentaires intègrent sa gestion : consultation de thésaurus seulement ou utilisation de celui-ci à la recherche. A l'heure actuelle, ils se multiplient. Aussi avons-nous voulu faire le point.

Qu'est-ce qu'un thésaurus ? Thésaurus à l'origine signifie trésor. C'est donc un trésor de vocabulaire, une collection de mots. Dans notre civilisation européenne, les trois premiers furent constitués au Moyen-Age et début de la Renaissance : sur la langue latine, sur la langue grecque et sur la religion. C'est au 19^e siècle, ensuite que l'on retrouve ce terme avec le fameux Rojet's Thésaurus que connaissent bien tous les anglicistes. Ce "dictionnaire" regroupe le vocabulaire par famille sémantique : tous les termes concernant la maison, la famille, la religion, l'agriculture etc. Le Rojet's Thésaurus préfigurait les dictionnaires analogiques.

Dans le domaine documentaire, un thésaurus n'est rien moins qu'une sorte de dictionnaire. Il regroupe l'ensemble du vocabulaire destiné à l'indexation et à l'interrogation : les descripteurs. Le regroupement est organisé suivant des principes

que l'on trouve dans les différents dictionnaires actuels. A chaque principe correspond une relation entre les mots. D'abord, comme dans le Rojet's, la relation hiérarchique suivant le principe de regroupement des mots spécifiques sous un terme générique (TG, TS). Ainsi, le vocabulaire est structuré en plusieurs arborescences, du concept général aux concepts détaillés. Ensuite, à l'instar des dictionnaires, de synonymes, la relation de synonyme (Syn, Cf) permet de mettre en regard d'un terme préférentiel (numériseur), tous les équivalents possibles (scanneur, scanner, digitaliseur). Puis, tout comme le petit Robert, la relation analogique permet des passerelles transversales entre descripteurs, de familles différentes (VA, voir aussi ; TA, terme associé). Enfin, comme dans notre institutionnel Petit Larousse, la définition (ou note d'application, NA), précise l'usage et le sens du descripteur. Signalons aussi les