

# La fourniture de documents électroniques

## L'expérience de l'INIST

par Christian Lupovici

*Directeur, assistant du directeur général de l'INIST*

### Introduction

La création de l'Institut de l'information scientifique et technique (INIST) en 1988, partait de la volonté du ministère de la Recherche de développer, en France, un centre multidisciplinaire de fabrication de bases de données associé à un centre de fourniture de documents pour la diffusion de l'information scientifique et technique, d'abord pour les chercheurs français, mais également au service des chercheurs européens et dans le monde.

Pour cela, il fut décidé de mettre en œuvre les moyens technologiques les plus modernes. A chaque stade de la production d'information, une technologie spécifique a été adaptée et a été intégrée dans le système général d'information, depuis l'acquisition des documents jusqu'à la fabrication des bases de données PASCAL pour la science, la technologie et la médecine et FRANCIS pour les sciences humaines et sociales, et à la gestion de la fourniture de documents primaires.

La chaîne de production de l'INIST est divisée en plusieurs systèmes qui jalonnent les étapes de transformation de l'information :

- le traitement des documents et leur gestion, effectué par un système GEAC 9 000 : la saisie de l'information bibliographique est effectuée en format UNIMARC pour lequel les zones locales de gestion ont été énormément développées pour gérer les transformations successives de l'information jusqu'aux produits documentaires livrables aux utilisateurs. Ainsi tous les types de traitement liés au titre de la publication en série sont préprogrammés ;
- le catalogage analytique des articles de périodiques et la saisie des résumés effectués sur un réseau de stations de travail est sous-traité à une société privée : l'information bibliographique générale (de type bibliothèque) est livrée en UNIMARC. Elle contient de plus des indications de traitement pour l'en-

semble des articles du titre. Le système reformate l'information bibliographique en format parenthésé type SGML (Standard Generalized Mark-up Language) pour la traiter et l'augmenter du dépouillement des articles et des résumés. Si la saisie du catalogage article est encore effectuée manuellement, la saisie des résumés d'auteur est effectuée par reconnaissance optique de caractères (OCR) ;

- l'indexation des articles par les documentalistes sur des postes de travail permettant des gestions et des contrôles de vocabulaire, voire une indexation assistée par ordinateur ;
- un système de contrôle pour la fabrication des bases de données ;
- un système de tri pour la production des produits bibliographiques ;
- la numérisation ;
- le stockage des documents sur support optique ;
- la gestion des commandes de la fourniture de documents.

### Les collections numérisées

#### *L'historique*

L'INIST a été membre de deux projets européens : TRANSDOC et ADONIS. De ces deux expériences, il est résulté un projet opérationnel : le Système d'archivage numérique (SAN). Du projet TRANSDOC (1984-1986), l'INIST a retenu que le transfert par satellite des textes numérisés était difficile à mettre au point et coûteux, mais que les techniques de numérisation image et de stockage sur disque optique numérique étaient maîtrisables à un coût raisonnable.

Du projet ADONIS, l'INIST a retenu que la technique de numérisation *a priori* et intégrale des textes devait se gérer facilement en association avec la base de données.

#### *Le projet SAN*

L'INIST a choisi 1 800 périodiques (sur son fonds de 27 000 titres) pour être numérisés systématiquement. Ces titres ont été choisis parmi les titres les plus demandés en fourniture de documents et représentent donc les périodiques cœurs de la littérature internationale.

Si l'on considère que la demande de documents est, pour une moitié, extrêmement dispersée (sur les titres et les années) et pour l'autre moitié concentrée sur un peu plus de 5 000 titres, il était logique de prendre parmi ces derniers, les titres dont on sait d'avance que chaque fascicule est sollicité pour la fourniture de documents.

En dehors du défi technique que cela représentait en 1990, la numérisation permettait de résoudre simultanément plusieurs problèmes :

- l'assurance d'une disponibilité permanente des documents, à la fois pour la fourniture de documents et pour l'analyse à la base de données ;
- une garantie de qualité constante de la reproduction et une rapidité de transfert du document ;
- un processus automatisé ;
- l'espoir de réduire l'espace de stockage à terme.

#### *La technique employée*

La numérisation est effectuée en mode image selon les normes recommandées par le groupe international sur l'échange de documents électroniques (GEDI), dont l'INIST est membre : une résolution de 300 BPI, l'organisation des fichiers en format TIFF et la compression des images en groupe IV CCITT.

La société Jouve qui effectue la numérisation des 90 fascicules par jour, trie les documents en fonction des problèmes de qualité :

- 30 % des documents qui contiennent beaucoup de photographies sont traités systématiquement en « mode mixte » (technique de compression proche de la norme JPEG) permettant de simuler

une trame et convertis en groupe IV (avis CCITT) ;

- 70 % des documents peuvent être numérisés de façon classique pour économiser de la place. Seules les pages d'images posant problème sont reprises en « mode mixte ».

Les images sont copiées sur des disques optiques numériques ATG double face qui contiennent 70 000 pages. Les disques sont chargés sur un juke-box (CYGNET) qui contient 131 disques, muni de 2 lecteurs de disques (ATG 60001). Le système est géré par un ordinateur SUN 4/370 (sous ORACLE/UNIX) qui communique avec le système de fourniture de documents sur un IBM 9121 par l'intermédiaire d'un réseau local ETHERNET.

Trois micro-ordinateurs permettent la transmission des documents sur les réseaux de télécommunication en fax groupé IV ou groupe III. De la réception de la demande à l'INIST à la réception du document chez l'utilisateur, quelques minutes suffisent, et sans intervention humaine.

## La gestion des commandes et la fourniture de documents

L'INIST traite environ 600 000 demandes de fourniture de documents par an, activité annuelle en progression d'environ 8 %. Pour assurer ce service avec des délais courts (2 heures, 24 heures ou 48 heures selon le niveau de service demandé), l'INIST a informatisé la gestion des commandes.

### Les différents moyens de commande

#### Les commandes sous forme électronique

Les clients de l'INIST peuvent utiliser différents moyens de commande :

- la messagerie des serveurs (QUESTEL, ESA, STN...);
- les systèmes de prêt entre bibliothèques tels que le PEB en France ou OCLC ;
- la saisie directe sur le système INIST par terminal vidéotex (minitel en France) ou VT 100. Dans ce cas, l'utilisateur a accès en ligne au catalogue des documents (titres fascicules et articles de périodiques) de l'INIST et le système

remplit automatiquement l'écran de commande lorsque le document a été identifié.

L'INIST a également développé des procédures de transfert de fichiers à partir du système du client ou à partir d'une grille de saisie formatée sur un poste de travail, ou enfin à partir d'une sélection de références bibliographiques capturées sur les CD-ROM Pascal. Les protocoles de communication adoptés permettent un accès par les réseaux de la recherche en messagerie électronique ou avec FTP sur TCP/IP (type INTERNET) ou par le RNIS avec STUTEL.

#### Les commandes sous forme traditionnelle

L'INIST reçoit encore beaucoup de commandes sous forme non électronique : formulaires papier, fax, coups de téléphone.

Toutes les informations sont, dans ces cas, saisies dans le système pour que toutes les commandes soient gérées électroniquement.

### La fourniture de documents

Le système trie les commandes selon la cote en magasin des documents et adresse les commandes aux bons étages de magasin où elles sont imprimées pour traitement. Le Système d'archivage numérique (SAN) est le magasin prioritaire. Toutes les demandes portant sur des documents numérisés lui sont adressées. Ce n'est qu'en cas de panne qu'un reroutage dans les magasins est prévu.

Pour les documents numérisés, le SAN traite les commandes par lots, toutes les deux heures. C'est l'équilibre actuel entre la nécessité de grouper les demandes par adresse physique sur les disques et la nécessité d'une réponse rapide au client.

Pour tout problème de référence incomplète ou ambiguë, le système ou les magasiniers adressent l'enregistrement au fichier d'un service spécialisé dans l'identification et la localisation des documents.

L'INIST travaille en coopération avec 100 bibliothèques de recours pour obtenir les documents qu'il ne possède pas. La localisation de certains de ces documents est faite dans le fichier INIST qui est conçu comme un catalogue collectif interne pour permettre une orientation externe automatique.

La saisie des éléments de fabrication (nombre de pages) sert d'indication de fin de traitement.

## L'évolution de la fourniture de documents

La fourniture de documents se fait de plus en plus sous forme électronique. Mais cette forme électronique est principalement du mode image. Ce mode permet de transmettre des documents dont l'origine est sous forme papier, mais ses gros inconvénients sont :

- de prendre beaucoup de place en mémoire et du temps en transfert, cela serait encore pire si les documents étaient numérisés en couleurs ;
- de ne permettre aucune recherche sur le texte.

C'est pourtant sur cette base que travaillent l'INIST et ses partenaires européens dans le projet EDIL (Electronic Document Interchange between Libraries) auquel des bibliothèques universitaires vont participer. C'est aussi selon ces mêmes techniques que sont présentés les documents du projet TULIP avec l'éditeur Elsevier ou Right Pages de Springer Verlag, pour ne citer que ceux-là.

Déjà de nouveaux projets de distribution des documents en mode caractères, en particulier en normes SGML commencent à voir le jour. L'INIST et le ministère de l'Enseignement supérieur et de la Recherche y sont attentifs car c'est la voie de l'avenir. Des expériences appliquées à la littérature grise permettront de rendre opérationnelle une chaîne de diffusion électronique.

L'accès au texte intégral et l'augmentation considérable de la puissance des systèmes de recherche documentaire vont bouleverser l'approche que nous avons jusqu'alors à la fois de la recherche documentaire dans les bases de données bibliographiques et de la fourniture de documents.

Les bases de données bibliographiques devront évoluer très vite pour ne pas devenir obsolètes par la concurrence de l'accès direct au texte intégral. L'INIST, à la fois producteur de bases de données et centre national de fourniture de documents, est bien placé pour apprécier tous les éléments du problème et saura les intégrer pour que de la convergence du signalement bibliographique et du document primaire naissent de nouveaux produits plus efficaces et adaptés à l'attente du public.