

Réflexions

Christian Dutheil

Directeur des Services techniques du CNIC

L'INFORMATION EN CHIMIE

ESSAI DE SYNTHÈSE

POUR les chimistes, pour les utilisateurs de produits chimiques, pour les acteurs de l'entreprise, les besoins d'information en chimie sont nombreux et diversifiés. Il convient d'y adjoindre l'information sur la chimie à destination des scientifiques des autres disciplines et, plus généralement, du public. Il faut alors prendre un langage dépouillé des jargons professionnels afin d'être compréhensible par tous.

Le recensement de ces besoins d'information, l'étude des solutions documentaires qui tentent d'y répondre avec un degré de satisfaction plus ou moins élevé selon le domaine ont très tôt mobilisé l'intérêt des spécialistes de l'information et des documentalistes de la chimie. La création du Centre national de l'information chimique (CNIC) concrétisait, dès 1972, les initiatives des professionnels (UIC¹, grandes entreprises) et des pouvoirs publics dans ce domaine. Celles-ci s'intensifiaient en 1978 par la restructuration du CNIC, afin de lui donner une représentativité nationale et de lui permettre des activités internationales fructueuses. Ainsi la chimie est la seule discipline, en y associant l'industrie pétrolière avec l'IFP², à disposer d'un centre national de l'information, dont l'un des buts est de permettre à la communauté française l'accès à toutes les sources d'information disponibles.

L'information en chimie

Les problèmes rencontrés en information dans le domaine de la chimie sont suffisamment nombreux, complexes et importants pour justifier l'organisation de colloques spécialisés (1, 2).

Compte tenu de l'évolution rapide des services offerts, il est difficile de dresser un inventaire exhaustif des banques de données existantes (tout inventaire dans ses détails n'a qu'une validité temporaire). En dehors des grands systèmes d'information, il est bien difficile aux petites banques de données d'exister. Elles ont ainsi parfois des « vies publiques » tumultueuses, (accès aux utilisateurs), malgré leur grand intérêt.

Sur plus de 3 500 systèmes d'information disponibles en ligne, environ 350 concernent la chimie, à divers égards. Les principaux domaines couverts sont l'information scientifique et technique générale, la thermodynamique et les propriétés physico-chimiques, les spectrométries, les propriétés nucléaires, la sécurité et la toxicité des produits chimiques, la réglementation, la propriété industrielle, l'ingénierie et le génie chimique, la technico-économie, les utilisations spécifiques.

Ainsi le chimiste semble être privilégié par rapport aux autres scientifiques. Cependant, de nombreuses lacunes subsistent : les couvertures ne sont que partielles et certains systèmes proposés ne sont pas exempts de reproches. J'ai utilisé le terme système d'information de préférence à banque ou base de données, car il est bien difficile de décrire la réalité à l'aide des classifications qui doivent être de plus en plus précises alors que les systèmes documentaires se diversifient en prenant des natures hybrides. Par exemple, les systèmes EURECAS et CAS ONLINE sont constitués de banques de composés chimiques (EURECAS, POLYCAS, REGISTRY FILE), de bases de données bibliographiques (CAS, CA FILE) de banques de données factuelles (CA OLD). Pour intéressantes qu'elles soient, les tentatives de classification sont trop rigides (3,4,5). Il est plus simple d'utiliser les définitions générales officielles (6) : banques de données bibliographiques signalant l'information primaire sous forme de références (titre, auteur, indexation, résumé...), banques de données factuelles donnant un accès direct à l'information primaire, parmi lesquelles on distingue banques numériques (données chiffrées), banques textuelles (données alphanumériques), banques en texte intégral (texte original d'articles, de dépêches, d'encyclopédies).

De par leur nature, la chimie et la biologie sont deux domaines fondamentaux qui se trouvent au cœur de toute activité industrielle ou de recherche, de production ou de protection de l'homme et de son environnement. En tant que sciences, elles font appel à toutes les autres disciplines; en tant que techniques, elles sont impliquées dans tous les secteurs d'activité. De plus, il existe entre ces deux sciences de base de nombreuses inter relations (biochimie, bio-

1. Union des industries chimiques
2. Institut français du pétrole

technologie, médicaments...) qui ont souvent des implications conjointes sur les autres domaines (interactions médicamenteuses, pharmacologie des molécules...).

La pérennité des informations relatives aux millions de substances connues (plus de 8 millions répertoriées dans les CAS) constitue un problème majeur et spécifique de l'information chimique. Il convient plutôt de parler d'information en chimie car il est nécessaire de répondre aux besoins du chimiste qui ne se limitent pas aux informations scientifiques et techniques, mais concernent aussi tous les besoins d'informations de son entreprise en matière de propriété et d'environnement industriels.

Pour le non-chimiste il faut être en mesure de fournir une information traduite en langage clair et des données générales sur la chimie et les produits chimiques.

Il est bien évident que la plupart du temps les besoins du chimiste et de l'entreprise se confondent. Il est inconcevable que la direction scientifique d'une entreprise se lance dans une recherche ou une production, sans avoir au préalable examiné tous les aspects de la propriété industrielle, de la sécurité ou de la réglementation. Les investissements de développement et de production sont tels qu'au sein des entreprises on confère une importance et une responsabilité particulière aux spécialistes de l'information.

Les besoins des non-chimistes peuvent être variés; ils se décomposent en trois grandes familles :

— ceux des professionnels d'une industrie utilisant des produits chimiques au cours d'un procédé ou d'une technologie particulière, dont les besoins en information sont les propriétés physico-chimiques, les réactivités, les dangers éventuels et la réglementation;

— ceux des consommateurs qui cherchent à s'informer sur les dangers éventuels, l'impact sur l'environnement et la santé, en relation avec l'information sur la sécurité et la réglementation;

— enfin, ceux des économistes et des services douaniers qui s'intéressent aux produits chimiques en tant que biens de consommation sujets à des échanges commerciaux.

Compte tenu de ces besoins très variés, il est difficile de définir en quoi consiste l'information en chimie et préférable d'évoquer

divers domaines d'information en chimie.

Les domaines couverts

A l'heure actuelle il existe plus de 350 banques de données d'accès public en ligne, mais cette abondance masque des problèmes réels: certains des domaines ne sont couverts que partiellement (quelques aspects, nombre limité de composés ou d'événements, inadéquation aux besoins réels).

On peut répartir les banques de données en: information générale, thermodynamique et propriétés physico-chimiques, spectrométries, propriétés nucléaires, réactions chimiques, ingénierie et génie chimique, utilisations spécifiques (médicaments, arômes et additifs alimentaires, produits agro-alimentaires, pesticides, peintures, caoutchouc et plastiques, biotechnologies..., propriété industrielle, sécurité-toxicité et réglementation, technico-économie.

Au niveau des utilisations spécifiques, il est difficile de préciser les frontières de la chimie (en fait,

peut-on les délimiter?), car les producteurs de banques de données et les services d'information ont une acception très large du terme chimie, à l'instar du Chemical abstracts service.

Les banques de données bibliographiques sont les plus satisfaisantes. Elles sont dotées d'outils documentaires permettant des recherches performantes (codes, classifications, hiérarchies, indexations...). Une longue pratique permet d'en maîtriser l'organisation et les outils logiciels. Il n'en va pas de même avec les banques en texte intégral, plus récemment introduites, pour lesquelles de nouveaux outils logiciels ont été développés. Leur utilisation actuelle est empirique et coûteuse. La meilleure solution pour accéder à ces banques serait mixte: une indexation classique permettant une pré-sélection, puis une recherche sur le texte intégral, ce que proposent certains systèmes sur disques compacts.

Les banques factuelles, numériques et textuelles sont très nombreuses. Le développement des télécommunications et de la tech-

Les besoins du chimiste et de l'entreprise

Besoins du chimiste

- scientifiques:
 - spécialisés
 - multidisciplinaires
 - autres sciences en relation
- techniques:
 - applications de techniques à la chimie
 - utilisation des produits chimiques
 - analytiques
- technologiques:
 - ingénierie et génie chimique
 - procédés industriels et leur contrôle
 - informatisation (modélisation, contrôle)
 - automatisation
 - catalogues industriels (produits chimiques, matériels, équipements, compétences de laboratoires...)

Besoins de l'entreprise

- information sur la propriété industrielle: brevets, marques, législation

— information technico-économique:

- données factuelles et numériques sur les marchés, les coûts, les productions, les capacités d'usines, les nouvelles installations, le prix de vente...
- opportunités de savoir-faire, de licences, de savoir-vendre...

— information économique:

- mercuriales
- macro et micro-économie

— information sur la sécurité et la toxicité des produits:

manipulation au laboratoire, production, hygiène industrielle, stockage, transport, diffusion, sécurité du consommateur...

— information sur la réglementation:

produits, installations, stockages, utilisations, étiquetage...

— données sur les entreprises: juridiques et financières

— données et information générales

R

nologie des micro-ordinateurs permet désormais à tout laboratoire de produire ses propres banques de données. Le « Numeric data group » de l'ICSTI³ (7), pour lequel j'ai récemment effectué un inventaire, fait évaluer à plus de deux mille le nombre de banques factuelles en chimie. De cet ensemble, seuls quelques systèmes sont commercialisés⁴. Toutefois se pose le problème de la diffusion, et donc de l'accès à l'information, qu'on peut résumer par le schéma ci-dessous.

Les problèmes rencontrés

Ils se situent à différents niveaux :

La classification

L'ICSTI vient d'achever un travail de révision de la classification de la physique et tente de l'étendre à la chimie, pour homogénéiser les vocabulaires employés par les différents producteurs. Cette classification est toutefois difficile à

mettre en œuvre en chimie à cause de la dualité substance-produit/propriétés-utilisation. Selon les systèmes documentaires, on trouve des classements par produit ou des classements par propriété ou utilisation. Les classifications qui se veulent mixtes introduisent des redondances néfastes, contraires au but recherché. A l'heure actuelle il faut se contenter des classements spécifiques (sections des Chemical abstracts; plan de classement Pascal).

La prise en compte des substances chimiques

C'est la nécessité de décrire et d'indexer les substances qui distingue l'information en chimie de celle des autres domaines. L'indexation peut se faire par le nom chimique (trivial, commercial, nomenclature), par un numéro d'enregistrement (RN de CAS, Uniterm de CLAIMS), par la structure du composé sous forme définie, génétique, ou de formules de Markush.

Il est le plus souvent nécessaire d'interroger les noms chimiques par fragments sémantiques (qualitatifs ou quantitatifs) et les struc-

tures des composés par sous-structures.

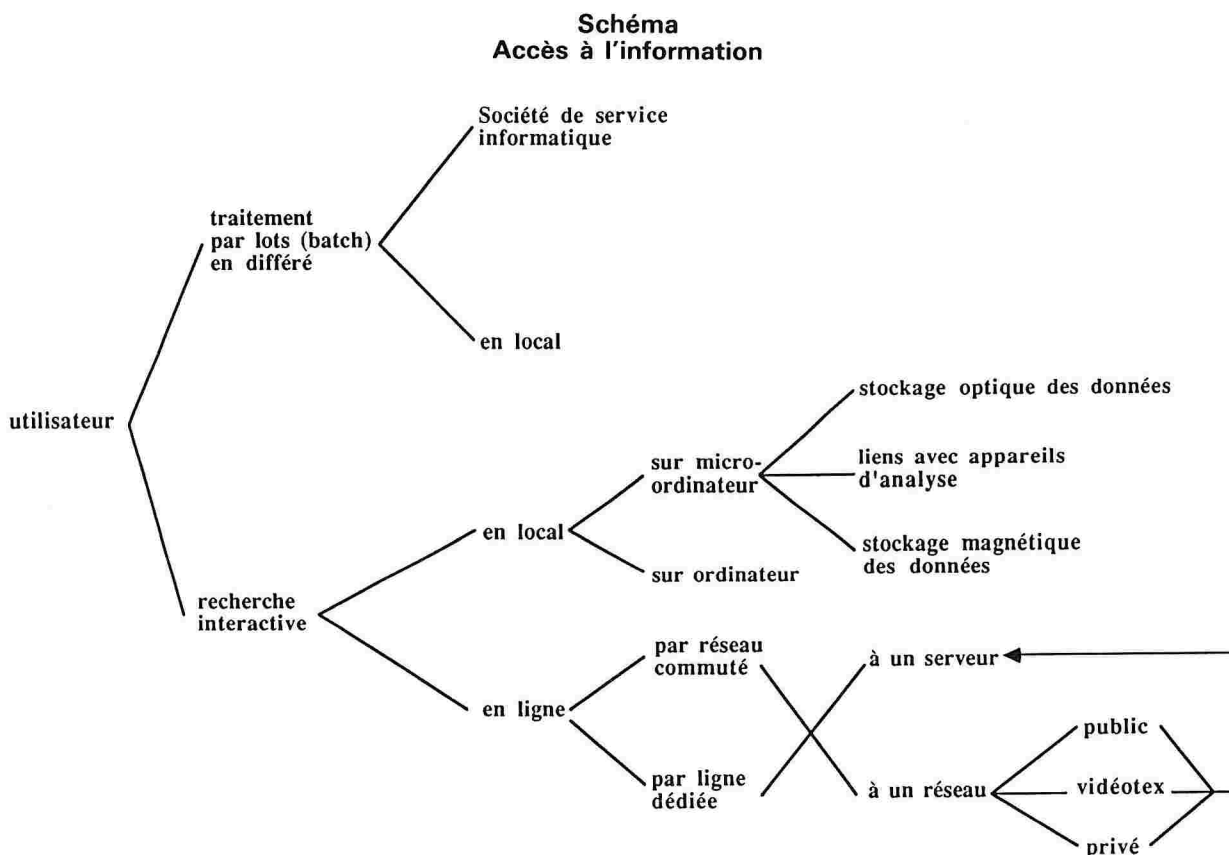
Le nom chimique attribué par un individu à une substance dépend en général de sa formation scientifique et de l'époque à laquelle il a fait ses études. Ainsi la même substance pourra être nommée formol, formal, formaldéhyde ou aldéhyde formique. Les nomenclatures, si elles sont correctement attribuées, permettent d'identifier sans problème la substance.

Au niveau des recherches par familles de composés, elles sont peu efficaces, car les règles d'attribution des nomenclatures sont basées sur un jeu de hiérarchies entre systèmes cycliques et groupes fonctionnels. Quelquefois une modification mineure du point de vue chimique entraîne un changement radical du nom. De plus, même au sein de la nomenclature IUPAC⁵, il existe des nomenclatures substitutives et alternatives

3. International committee for scientific and technical information

4. Ils sont rapportés dans les tableaux en annexe

5. Union internationale de chimie pure et appliquée



pour le choix des noms. A l'exception de celle des Chemical abstracts, les nomenclatures se prêtent mal aux manipulations informatiques.

Ces difficultés ont conduit, en 1965, Chemical abstracts service à attribuer un numéro de registre unique RN⁶ à chaque nouvelle substance. Plus de 8 millions de RN ont ainsi été attribués aux composés décrits depuis 1962. Malheureusement ce numéro séquentiel ne contient aucune information sur la structure du composé. Il est devenu le numéro d'identité de la substance. Son emploi est demandé par de nombreux organismes nationaux et internationaux, ainsi que dans les banques de données. Certains organismes ont développé leur propre système d'enregistrement (ONU, CEE...).

Pour l'enregistrement et l'interrogation des structures chimiques, différents outils ont été mis au point (8) :

— *les codes fragmentaires* (CPI de Derwent, Ring code du pharmadokumentation Ring, GREMAS de l'IDC, Uniterm-fragment de Claims...). Exploités, à l'origine, de manière mécanographique, ils sont d'une manipulation souvent malaisée en informatique, mais ils sont propres à prendre en compte aussi bien les structures totalement définies que celles qui ne le sont que partiellement ou d'une manière générique (famille de composés, formule de Markush). Le codage est unique, mais les combinaisons de fragments ne conduisent pas à des restitutions univoques.

— *les codes linéaires* (WLN Wiswesser line notation) présentent les mêmes inconvénients et complexités que les nomenclatures. Ils ont l'avantage d'être plus compacts qu'elles et plus facilement manipulables par l'informatique. Ils peuvent également être transposés (transcodage) en codes fragmentaires, (avec appauvrissement du contenu) et, dans certaines conditions, en codes topologiques.

— *les codes topologiques* (matrice CAS, code DARC...) permettent un enregistrement et une restitution bi-univoques des struc-

tures définies. Les développements des logiciels d'exploitation de ces codes ont conduit à des langages d'interrogation très conviviaux, très proches du langage du chimiste.

Il est également possible de formuler des questions génériques, c'est-à-dire comprenant certaines alternatives dans la définition des atomes, des liaisons, des substituants. La prise en compte codage-restitution est bi-univoque. Pour des systèmes de gestion de banques de données structurales en interne, de nombreux codes ont été développés (MACCS de Molecular design), mais, au niveau des serveurs d'information en ligne, seuls les systèmes de DARC (développé par l'équipe du professeur J.-E. Dubois à l'Université Paris VII et par Télésystèmes-Questel) et CAS ONLINE (développé par Chemical abstracts service) sont disponibles. Certains systèmes, indépendants des serveurs, ont été développés sur micro-ordinateur pour générer automatiquement les codes à partir du dessin de la structure.

Avec les formulations génériques, ce sont les représentations de Markush, si caractéristiques des brevets en chimie, qui sont approchées. Une formule de Markush est une représentation compacte d'un ensemble de composés ayant des éléments structuraux ou fonctionnels en commun. Certains éléments chimiques (atomes) de la formule développée peuvent être remplacés par une variable dont les valeurs possibles sont spécifiées, soit de manière précise (chlore, méthyl), soit de manière générique (alkyl, aryl). Dans une telle formule, la position d'un substituant peut être localisée de façon plus ou moins précise. Par le jeu de la combinatoire, des milliers, voire des millions, de composés définis peuvent être générés à partir d'une seule formule de Markush.

La stéréochimie et l'isotopie sont parfaitement gérées au niveau des systèmes locaux (DARC *in house*), mais ne sont pas utilisables sur les services en ligne, car ces informations, à l'heure actuelle, ne sont pas prises en compte totalement dans le codage initial des structures (matrices de connectivité).

Cette approche structurale est très performante pour les recherches documentaires concernant les substances chimiques. Il est toutefois quelques domaines pour

lesquels elle est encore insuffisante : chimie minérale, polymères (9), peptides et acides nucléiques, biomolécules de grosse taille. Les codes spéciaux (Uniterm de Claims, code Plasdoc de Derwent, système Diapason de Rhône-Poulenc) et les codes structuraux spécifiques (codes stéroïdes et peptides du Ring) ne constituent que des outils d'approche.

La transcription des problèmes chimiques

La transcription des problèmes d'isomérisation, de délocalisation, de tautomérie et de mésomérisation est en général résolue par des conventions de représentation, (on choisit systématiquement une forme), ce qui ne facilite pas l'utilisation des systèmes d'information par l'utilisateur final ou le néophyte. Des systèmes d'équivalences automatiques, développés au niveau du logiciel d'interrogation, devraient voir le jour rapidement pour résoudre ces problèmes.

La description des réactions

Bâtir un système d'information sur les réactions chimiques, c'est tout d'abord faire un choix sur le type de données signalées et sur leur organisation (10) :

— banque bibliographique décrivant, par des outils documentaires classiques, les principaux éléments d'une réaction;

— banque factuelle signalant les valeurs, numériques ou textuelles, des différents paramètres de la réaction;

— système de gestion de réactions permettant de prendre en compte l'ensemble des paramètres et des étapes d'une réaction et de la décrire sous ses différents aspects.

Pour décrire une réaction, de nombreux paramètres doivent être pris en compte : produits de départ et réactifs, produits et sous-produits synthétisés, milieu réactionnel (solvant, additifs...), inducteurs de réaction, catalyseurs ou modérateurs, conditions opératoires, stœchiométrie, contrôle, schéma réactionnel avec ses différentes voies, étapes réactionnelles avec leurs enchaînements, rendement, documentation et références bibliographiques.

La description des structures doit se faire en termes de structure définie, de formule de Markush ou de site réactionnel, et doit tenir

6. Il est unique à un instant donné

compte de la stéréochimie et de l'encombrement stérique. Les différents aspects de la compréhension de la réaction doivent être abordés : mécanisme réactionnel, aspects énergétique et thermodynamique, aspect cinétique.

A ces données factuelles, qui doivent être validées (car leur qualité essentielle est la fiabilité) et respecter les conventions habituelles du chimiste, vient s'ajouter une indexation (typologie et dénomination des réactions). Dans le signalement des informations, il convient de différencier les réactions de laboratoire des réactions industrielles (ces dernières devant être accompagnées d'informations sur le contrôle et le pilotage du procédé). Il n'existe pas, à l'heure actuelle, de système qui rassemble ces éléments.

Les informations graphiques

Des informations graphiques, autres que les structures, sont nécessaires au chimiste. Elles concernent la cristallographie, les spectres, les schémas et dessins du génie chimique. Ce type d'information, quand il est pris en compte, ne l'est qu'au niveau des systèmes d'information spécialisés.

Les nouvelles représentations des molécules

Au niveau de la recherche, tant fondamentale qu'appliquée, on constate une orientation de la représentation de la molécule par une autre image que la structure développée. Des modèles quantiques, volumiques (rayon atomique, Van der Waals), isoélectriques (courbes de niveau), géométriques interactifs ont été développés. Ils ne sont utilisés que dans des systèmes de conception assistée par ordinateur (CAO et *Drug design*) ou à titre de gadget dans certains systèmes en local.

Inventaire

Banques bibliographiques spécialisées

Le domaine de la chimie est largement couvert par Chemical abstracts qui, depuis 1907, signale la littérature mondiale (revues, livres, congrès, brevets...). Tous les aspects de la chimie y sont pris en compte : biochimie, chimie minérale, chimie macromoléculaire, chimie physique, chimie analytique, ingénierie, radiochimie..., ainsi que les sciences en relation avec la chimie (physique,

mathématique...) et les applications industrielles de la chimie (pétrochimie, agro-alimentaire, métallurgie et traitements de surface, pharmacie et cosmétique, polymères et plastiques, énergie, environnement).

Depuis 1967, le fonds est automatisé (CASEARCH) et accessible sur 7 serveurs dans des configurations différentes :

- base bibliographique unique (BRS, ESA-IRS);

- dictionnaires de composés interrogeables par fragments de nomenclature, associés aux fichiers bibliographiques (DIALOG, SDC, DATA-STAR, STN INT);

- fichiers structuraux (EURECAS et CAS ONLINE) interrogeables par structures et sous-structures (à l'aide de DARC sur QUESTEL, de MESSENGER sur STN INT) associés aux fichiers bibliographiques (CAS sur QUESTEL, CA File sur STN INT). Ce fonds représente plus de 8 millions de composés et 9 millions de références, en un seul ou plusieurs fichiers selon le cas. Sur le serveur STN INT, un fichier de correspondance entre numéro de référence et numéros de registre des composés cités par cette référence permet de compléter l'antériorité sur la période 1962-1966 (CA OLD).

Il faut regretter la disparition du fichier Index chemicus de l'Institut of scientific information (ISI) qui complétait la panoplie de l'organicien (3 millions de composés; 3 millions de références) sur QUESTEL. Ce fait traduit la difficulté d'exister des systèmes d'information. Au-delà de l'intérêt scientifique se situent les impératifs économiques (interrogation insuffisante) et politiques (diffusion sous forme exploitable par micro-information) des producteurs.

Les fonds Beilstein et Gmelin, dont la mise en ligne est souhaitée de longue date, seront prochainement disponibles. Parmi les fonds plus ciblés, citons ZLC (Zinc, Lead, Cadmium abstracts) produit par Zinc development association (GB) et accessible sur INFOLINE, qui concerne la chimie et la métallurgie des trois éléments. Un fichier des composés organiques du silicium, de l'étain, du germanium et du plomb est réalisé par l'UA 35 du CNRS/Université de Bordeaux 1.

Banques en texte intégral

Trois ouvrages essentiels pour le chimiste sont disponibles en texte intégral :

- les revues publiées par l'American chemical society depuis 1982, soit plus de 50 000 articles, sont interrogeables sur STN INT et BRS. Avec une mise à jour bi-hebdomadaire, le chimiste peut accéder à des textes qu'il n'a pas encore reçus sous forme imprimée;

- l'*Encyclopedia of chemical technology*, Kirk-Othmer (12 000 articles, 6 000 tableaux, 5 000 figures), éditée par Interscience, est disponible sur BRS, DATA-STAR et DIALOG;

- le fichier Heilbron, qui rassemble le *Dictionary of organic compounds* (5^e éd.) et le *Dictionary of organometallic compounds*, édités par Chapman and Hall Ltd, est disponible sur DIALOG.

Banques bibliographiques multidisciplinaires

De grands fonds multidisciplinaires contiennent une part importante d'information chimique :

- PASCAL, produit par le Centre de documentation scientifique et technique du CNRS accessible sur QUESTEL, ESA-IRS et DIALOG (antériorité 1973). Depuis 1980, dans certaines sections, des suffixes permettent de préciser les attributions d'un composé chimique dans son contexte (indicateur de rôle);

- SCISEARCH, le *Science citation index* de ISI, accessible sur DIALOG et DIMDI (antériorité 1974);

- *Comprehensive disssertation abstracts*, produit par University microfilm international (États-Unis), accessible sur BRS et DIALOG (antériorité 1861);

- NTIS, produit par le National technical information service de l'US department of commerce, accessible sur DIALOG, ESA-IRS (antériorité 1964), SDC et DATA-STAR (antériorité 1970), CEDOCAR et STN INT;

- SSIE, produit par NTIS, accessible sur DIALOG, BRS, SDC (antériorité 1974);

- NASA, produit par l'agence aérospatiale américaine, accessible sur ESA-IRS et INKA (antériorité 1962);

- FIESTA, produit et servi par le CEDOCAR (FR) (antériorité 1972).

Il convient également de citer RINGDOC (de Derwent), les fichiers de l'American petroleum institute (APILIT et APIPAT) et, en atteignant les frontières de la chimie METADEX (de l'American society for metals), World aluminum abstracts (de l'ASM), Non ferrous metals abstracts (du British non ferrous metals technology center) pour les métaux, RAPRA (de Rubber and plastics Assoc. of GB), DKI (du Deutsches Kunststoff Inst.) pour les polymères, PAPER-CHEM (Institute for paper chemistry, États-Unis) pour le papier, World textiles (de Shirley institute de Manchester) et TITUS (de l'Institut textile de France) pour l'industrie textile. Ces exemples ne sont pas, bien entendu, limitatifs.

Banques de composés chimiques

C'est encore Chemical abstracts service qui fournit le plus important fichier de composés : le RNSS comprend plus de 9 millions de RN et plus de 10 millions de noms chimiques interrogeables. Dans le tableau 1 sont écartés volontairement les fichiers d'activité et les fichiers commerciaux (nommés fichiers de produits), ainsi que les banques de séquences de nucléotides. A l'exception de ceux qui constituent une véritable banque de composés, ils seront cités plus loin.

Ingénierie et génie chimique

Dans le domaine du génie chimique, chemical engineering abstracts de la Royal society of chemistry (GB), 75 000 références depuis 1970 (servi par INFOLINE, ESA-IRS et DATA-STAR) et DE-CHEMA de la Deutsches Gesellschaft für Chemisches Apparatewesen (RFA), 85 000 références depuis 1976 (servi par STN INT et FIZ TECHNIK), complètent avantageusement Chemical abstracts. Le fichier spécialisé le plus important, COMPENDEX, ne concerne que peu la chimie.

Banques de propriétés physico-chimiques et thermodynamiques

La plupart des propriétés physico-chimiques peuvent être calculées à partir des valeurs thermodynamiques, ce qui justifie le regroupement de ces deux types de banques. Les données obtenues sont, selon le cas, mesurées, calculées, ou évaluées.

Certaines banques enregistrent les valeurs dans un état de référence et permettent de les calculer

dans d'autres conditions (température, pression), THERMODOATA par exemple; PPDS offre même le choix de la méthode de calcul. D'autres se limitent aux paramètres de calcul (fichiers du National bureau of standards). Malheureusement la nécessaire homogénéité des informations oblige les producteurs à définir un domaine de couverture très limité. Le nombre de composés de chaque banque est donc réduit. L'accès par des serveurs grand public est rare. De nombreuses banques ne sont disponibles que sur disquettes à exploiter par micro-informatique (7) (cf. tableau 2).

Banques analytiques, spectroscopiques et cristallographiques

Parmi les nombreuses techniques analytiques, seules quelques-unes, particulièrement les spectroscopies, ont donné lieu à la réalisation de banques de données. Le travail effectué dans le cadre du NDG⁷ de l'ICSTI (7) a permis de dénombrer dans ce domaine plusieurs dizaines de systèmes dont quelques-uns sont accessibles en ligne. Beaucoup ne sont que des collections de données sans procédure de validation et ne peuvent, quelle que soit leur valeur intrinsèque, constituer de systèmes d'information crédibles (cf. les tableaux 3 à 6 pour les banques spectroscopiques, le tableau 7 pour les banques cristallographiques, le tableau 8 pour les banques cinétiques, le tableau 9 pour les banques de chromatographie et le tableau 10 pour les banques diverses).

Banques de propriétés nucléaires

Ces propriétés physiques particulières méritent un classement à part, d'autant qu'elles sont utilisées pour des applications spécifiques : les besoins en information et la présentation sont différents de ceux des autres banques (cf. le tableau 11).

Banques de réactions chimiques

Dans le fichier PASCAL, pour les sections de chimie (ex. 170, 171 et 172), la notion de réaction peut être abordée grâce aux indicateurs de rôle. Avec les systèmes

documentaires bâtis sur *Chemical abstracts*, des recherches par composés sur les fichiers dictionnaires ou structuraux peuvent être combinées sur les fichiers bibliographiques pour retrouver les signalements bibliographiques des réactions. Pour intéressante qu'elle soit, cette approche n'est pas satisfaisante, car elle ne prend pas en compte les paramètres de la réaction. Aussi des fichiers spécifiques ont-ils vu le jour :

— *Chemical reaction documentation service* (CRDS), de Derwent, qui correspond aux volumes 1 à 30 de *Synthetic methods of organic chemistry*, éd. par W. Theilheimer, depuis 1942, puis au *Journal of synthetic methods* (littérature et brevets) à partir de 1975. Ce fichier, d'abord réservé aux souscripteurs sur SDC, est ouvert depuis peu à tous;

— KETO-REACT, produit et servi par l'ARDIC, contient 3 000 documents sur les synthèses de cétones.

Avec les systèmes de gestion de réactions en local, des banques de données sont souvent proposées :

— FIZER and FIZER par Télésystèmes - Questel avec DARC *in house*;

— THEILHEIMER et sous-fichiers de ISI par Molecular design avec MACCS.

Banques de congrès et fournisseurs de documents originaux

De nombreux serveurs ont mis en place, avec la collaboration des producteurs de banques et les grandes bibliothèques publiques, des services de commande en ligne des documents originaux. Il est, de plus, souvent utile de consulter en ligne :

— CASSI (Chemical abstracts service source index) sur ORBIT

— WORLD TRANSINDEX, produit par International translation centre de Delft et le CDST du CNRS, sur ESA-IRS

— CCN (catalogue collectif national des publications en série) et TELETHESES, produits par la DBMIST sur SUNIST

— il faut encore citer : Books information; Books in print, LC Line, LC Marc, Remarc, UKMarc, Ulrich's int. period. directory, ISTP & B de ISI, et les catalogues des éditeurs.

7. Numeric data group

L'INSERM propose une revue des sommaires scientifiques sur minitel.

Pour les congrès, il convient d'ajouter: Conference paper index (Cambridge scientific abstracts), Conference proceeding index (British library), EI (Engineering information) et Meeting agenda (CEN-Saclay).

Banques sur la sécurité et la toxicité des substances

Un inventaire complet a été réalisé par la CNIC (11) qui recense aussi bien les fonds spécialisés que les fonds multidisciplinaires ou généraux. Deux serveurs, tous deux américains, paraissent spécialisés dans ce type d'information: Chemical information system, CIS, et la National library of medicine, NLM, représentée en France par l'INSERM, dont le réseau TOXNET propose une version révisée et enrichie du fichier Toxicology data bank (TDB données validées) et rassemble des données plus nombreuses mais non nécessairement validées (HSDB, Hazardous substances data bank).

Au Canada, deux systèmes d'information orientés vers le poste de travail sont accessibles: INFOCHIM (1 000 substances) et NM, constitué à partir des fiches de sécurité (18 000 produits) par le CCHST (Centre canadien d'hygiène et de sécurité du travail), et INFOTOX (3 200 substances; 2 300 préparations) par le CSST (Commission de la santé et de la sécurité du travail).

Le fichier européen ECDIN, séduisant dans son cahier des charges, pêche par manque massif de données. Quant aux informations retrouvées, elles se présentent le plus souvent sous forme de fiches signalétiques, avec les propriétés physico-chimiques en relation avec la sécurité et la toxicité, l'identité de la substance et les données toxicologiques (intoxication aiguë, expérimentale et chronique, écotoxicité).

Dans la plupart des banques, le RN est un des critères d'interrogation.

Normes et réglementation

L'aspect normes et réglementation est primordial dans un contexte industriel. Ces deux types d'informations sont rassemblés sur les banques, car la tendance du législateur est d'appuyer la réglementation par les normes. Le

CNIC a développé, en collaboration avec l'AFNOR, un complément d'indexation de NORIANE et réalisé un dictionnaire des substances soumises à réglementation. Les fichiers nationaux de normes sont progressivement chargés sur les serveurs. (cf. tableau 12).

Banques technico-économiques

Les systèmes documentaires dans le domaine technico-économique et dans celui des affaires, accessibles en ligne, se multiplient rapidement. Ce sont actuellement les plus nombreux. Il suffit pour s'en persuader de consulter les annuaires de banques de données. Ils représentent économiquement la part la plus rentable des banques de données. Pour répondre aux besoins de l'industrie, il faut ajouter les informations financières, les mercuriales et les informations sur les sociétés. Les problèmes de ce domaine spécifique sont décrits par Marie-Françoise Mazières (12). Pour la technico-économie, on considère en général qu'un tiers seulement de l'information est publiée et que l'on n'en retrouve que le tiers dans les banques de données. Ces systèmes offrent des banques spécifiquement consacrées à la chimie:

— CIN, Chemical industry notes, produit par Chemical abstracts service, servi par DIALOG, ORBIT et DATA-STAR;

— CBNB, Chemical business news base, produit par la Royal society of chemistry, servi par INFOLINE, DIALOG et DATA-STAR, qui contient des références bibliographiques et des données numériques et factuelles;

— CHEM-INTELL, produit par Chemical intelligence service (GB), servi par DATA-STAR et INFOLINE, qui propose des rapports statistiques, sur 10 ans, d'une centaine de pays pour une centaine de composés organiques;

— Chemical age project file, produit et servi par INFOLINE, qui propose 15 000 données textuelles et numériques sur les usines chimiques et para-chimiques depuis 1980 (couverture internationale);

— Chemical economics handbook online, produit par SRI international et servi par ORBIT, fournit des données textuelles et numériques sur 1300 produits chimiques commerciaux.

Parmi les nombreux fichiers de PREDICASTS, le PROMT (Predictcasts overview of market and technology) contient une part importante d'information sur les composés chimiques manufacturés. Pour répondre avec le maximum d'exhaustivité aux questions, il est aussi nécessaire de consulter les fichiers du Centre français du commerce extérieur sur le serveur CISI, ceux de l'OCDE, entre autres, sur le serveur DRI, ceux de la CEE sur les serveurs ECHO et EURIS, ainsi que ceux du serveur français spécialisé GSI-ECO.

Les domaines stratégiques de la chimie, comme la pétrochimie (PE/NEWS de l'API) et la chimie pharmaceutique (PNI), ont leurs propres banques spécialisées.

Banques d'opportunités, de compétences et catalogues industriels

Les systèmes d'information sur l'innovation, les transferts de technologies, les appels d'offres et les compétences de laboratoire sont intéressants bien que limités. Les catalogues industriels devraient se multiplier dans un futur proche, d'autant que ce type d'information convient parfaitement à l'interrogation par le réseau vidéo-text (cf. tableau 13).

Banques par applications des produits

Des répertoires encyclopédiques ont été constitués par domaine d'utilisation des produits. Ils apparaissent progressivement sur les serveurs.

Banques en biotechnologies

De nombreuses banques de données spécialisées dans le domaine des biotechnologies apparaissent. Ce concept récent recouvre pourtant des technologies anciennes, voire ancestrales, que les autres fichiers ont toujours signalées (cf. Chemical abstracts) L'information en biotechnologies peut revêtir quatre aspects:

— l'information scientifique et technique (13), cf. tableau 15;

— l'information technico-économique (14);

— les banques de séquences (15), cf. tableau 16;

— les collections de cultures (16), cf. tableau 17.

Information sur la propriété industrielle

Compte tenu du volume et de la pérennité des informations en chimie, les brevets y prennent une importance particulière. Il existe deux types de fichiers (17) : les fichiers administratifs (type fichiers de l'INPI), qui aident remarquablement à suivre la « vie » d'un brevet, et les fichiers documentaires (type WPI de Derwent), qui ajoutent une indexation permettant une recherche plus efficace que la classification internationale des brevets.

La plupart des grands pays ont mis, ou sont sur le point de mettre, leurs fichiers brevets en accès télématique. LEXPAT permet l'interrogation en texte intégral des brevets américains. INPADOC et EDOC décrivent les familles de brevets relevant d'une même invention. Pour les fichiers plus spécifiques de la chimie, cf. le tableau 18.

Les fichiers de marques déposées (nationales et internationales) apparaissent sur les serveurs (INPI-Marques, TMINT), de même que les fichiers relatifs aux aspects juridiques liés aux brevets (JURINPI, PATLAW). Des accords entre producteurs et serveurs se développent pour traiter spécifiquement les brevets de la chimie (INPI-DERWENT-QUESTEL). Enfin, de nombreuses banques de données prennent en compte les brevets, particulièrement Chemical abstracts.

Informations sur les pôles géopolitiques

Il existe des banques de données recensant l'information sur — et en provenance de — certains pays ou pôles géopolitiques. Ces systèmes sont particulièrement utiles (bien que fragmentaires) pour les pays peu représentés dans les banques internationales, ou dont la langue est inexploitable pour la plupart des utilisateurs (URSS, Chine, Japon). La banque de données JOIST offre des extraits (en version anglaise) des principales banques japonaises. Un bureau du Japan information center for science and technology (JICST) est installé à Paris, au CDST du CNRS. Une cellule « Japon » fonctionne au CNRS. Des sociétés privées (EURALIA, par exemple) fournissent des bulletins de surveillance sur l'IST japonaise. Au Japon, les banques spécialisées sont décrites dans des rapports de mission (18, 19, 20).

Pour l'URSS, on peut consulter les banques Soviet science and technology (produite par IFI Plenum et servie par DIALOG), et East European monitor — the Chemical industry (produite par Business international et servie par DATA-STAR).

Pour l'Amérique latine, BIBLAT (produite par l'Université de Mexico) est disponible sur QUESTEL.

Banques de données et intelligence artificielle

Les systèmes d'intelligence artificielle (IA), particulièrement les systèmes experts, font appel à des banques de connaissance qui regroupent des « faits » ainsi que les « règles » des « inférences » permettant de les relier pour simuler le raisonnement d'un homme de l'art dans un domaine spécifique. De nombreux travaux, employant des techniques d'IA, sont en cours sur l'extraction d'informations contenues dans les banques de données, pour « instruire » des bases de connaissance. Les réalisations opérationnelles sont pour un futur proche.

Systèmes de corrélations et de modélisations

Des programmes de modélisation, de corrélation et d'interprétation sont accessibles au sein de systèmes documentaires sur certains serveurs (ou sur micro-ordinateur). Ces programmes permettent d'élargir le champ d'action du spécialiste de l'information et surtout de l'utilisateur final. Ceci est important car l'accès à une bibliothèque de programmes autorise l'utilisation ponctuelle de logiciels de calculs scientifiques sans avoir à les acquérir (avec toute l'infrastructure idoine). Au niveau du calcul des propriétés physico-chimiques citons : PROSIM de l'Institut de génie chimique de Toulouse et GC DATA de l'Ecole supérieure de chimie de Marseille (5, 7).

Observations générales

Il est difficile de dégager des généralités sur les systèmes documentaires, tant est spécifique l'appréhension des informations relatives à chaque secteur technique. Les besoins (nature et forme de présentation de l'information) et les marchés sont très variés. Des observations généra-

les peuvent néanmoins être dégagées :

— par essence, une banque spécialisée est relative à un domaine très limité. Le nombre d'événements est faible. Si l'élargissement de la couverture est tenté, deux risques apparaissent : l'hétérogénéité des données et les « trous » d'information;

— la crédibilité des informations contenues dans une banque dite factuelle est assujettie à des règles essentielles : validation des données, homogénéité et cohérence des « valeurs », mises à jour en fonction des progrès de la technique et de la précision des mesures, indication de la valeur nominale avec sa précision pour les valeurs évaluées, indication de la méthode, référence de la source;

— la dispersion des banques sur les différents serveurs entraîne des difficultés particulières qui s'ajoutent aux problèmes de la connaissance du fonds.

Dans un avenir qui ne peut être précisé, des systèmes experts servant d'interface entre l'utilisateur final et le serveur seront disponibles. Ils relayeront le spécialiste de l'information pour permettre le dialogue entre l'utilisateur final et le serveur sur des questions simples.

Il faut savoir ce qu'on peut attendre des systèmes d'information en ligne et ne pas en aborder l'utilisation avec trop de naïveté. Le choix des banques de données à interroger pour répondre à une question nécessite la connaissance du fonds documentaire (couverture, politique d'indexation, outils documentaires, organisation) et la connaissance de l'organisation en banque de données sur un ou plusieurs serveurs (informations interrogeables, langage du serveur). Il est difficile d'être compétent pour toutes les banques. Et pour des informations particulières, il faut bien faire appel à des banques de données que l'on utilise peu. Les annuaires et répertoires (21, 22, 23), imprimés ou en ligne (CUADRA, REBK), sont alors très utiles bien que limités dans leur description.

Le devenir d'une banque de données en ligne est totalement lié à son succès commercial, à des problèmes financiers et juridiques entre serveur et producteur (politique de distribution, exclusivité...). C'est un monde en perpétuel mouvement, et l'accroisse-

ment du nombre de banques disponibles ne doit pas masquer le taux important de renouvellement. En fait, un grand nombre de banques disparaît chaque année.

Besoins non satisfaits et potentialités

Dans les domaines passés en revue, de nombreuses lacunes ont été constatées :

— en physico-chimie : *handbook* de propriétés, diagrammes pH-potentiels, diagrammes de phases, cinétiques, propriétés de surface...;

— en analytique : au sein même des méthodes spectroscopiques, qui sont bien représentées, les lacunes sont nombreuses : UV, RPE, RAMAN, spectro laser, fluorescence, phosphorescence, spectro micro-ondes, spectro radiométrique, ESCA, ESR, NQR... Les autres méthodes mériteraient que des banques largement accessibles leur soient consacrées : chromatographie, HPLC, polarographie, électrophorèse, radiochimie, immunochimie. Or, il n'existe aucune compilation des couplages de méthodes et de leurs interférences;

— en technico-économie, il serait urgent qu'apparaisse, pour la France et l'Europe, un fichier du type *Fine chemical directory*. Sa réalisation technique serait relativement simple, par fusion des catalogues des producteurs et fournisseurs;

— pour les applications : un fichier des principaux composés et substances industriels (5 000 substances) serait du plus haut intérêt, tant pour la médecine que pour rechercher des diversifications sur les marchés des produits. Un fichier des membranes et des résines échangeuses d'ions, avec leurs caractéristiques et leurs applications, aurait de multiples utilisateurs.

Les potentialités sont nombreuses. A n'en pas douter, toutes sortes de fichiers, sur fiches ou sur micro-ordinateur, ont été constitués pour des usages spécifiques et individuels (7). Toutefois, ces banques ne sont qu'exceptionnellement utilisables directement pour un usage public. Les laboratoires de recherche publics (CNRS, INRA, BRGM, IRCHA, INSERM, Universités...) et privés (grandes entreprises, experts, ingénieurs conseils...), ainsi que des associations professionnelles et de nombreuses agences nationales et in-

ternationales détiennent un volume considérable de données (25, 26, 27, 28, 29). Une liste détaillée des potentialités intéressant un large public a d'ailleurs été présentée au 1^{er} colloque CNIC (5).

Un double langage

Un serveur confronté aux problèmes de diffusion des banques de données se trouve dans l'obligation de satisfaire deux catégories d'utilisateurs dont les besoins sont apparemment antinomiques :

— répondre aux exigences des spécialistes de l'information en développant des outils informatiques de plus en plus sophistiqués pour exploiter au mieux toutes les ressources, même potentielles, d'interrogation des banques de données;

— dialoguer avec l'utilisateur final ou le néophyte en lui proposant un accès simplifié (avec un minimum de dégradation des taux de pertinence et de rappel) et assisté (logiciels par menus; assistance en ligne...).

Ces problèmes peuvent être résolus en partie par les techniques d'intelligence artificielle. L'utilisation de la micro-informatique conduit à la production *in situ* d'une information élaborée multiparamétrée (scientifique, technique, économique) et sélectionnée, triée (doublons éliminés), et présentée de manière agréable et synthétique (manipulation de l'information par traitement de texte).

Exploitation de l'information des banques

Il convient d'aller plus loin et de dépasser la seule amélioration des possibilités d'interrogation pour aboutir à une véritable exploitation des données. Celle-ci suppose l'utilisation des possibilités du téléchargement sur micro-ordinateur pour combiner entre eux les éléments variés des banques, les sélectionner, les rendre accessibles dans un format unique. Afin de tirer tout le sens de cette information, il faudra ensuite la soumettre à des traitements particuliers, notamment des analyses de données (24). L'application des méthodes de la statistique et de l'analyse de données aux informations extraites des banques de données donne naissance à de nouvelles disciplines : bibliométrie, technométrie,

scientométrie. Ces méthodes sont indispensables à l'exploitation stratégique de l'information.

Des outils de dénombrement sont apparus sur certains serveurs : GET sur INFOLINE, MEMTRI sur QUESTEL, ZOOM sur ESA-IRS. Certains d'entre eux proposent des outils plus élaborés : CISI, CITI2. Des logiciels sont disponibles sur micro-ordinateur, soit pour des usages non spécifiques, soit pour exploiter les données de systèmes documentaires particuliers : PATSTAT de Derwent pour les brevets. Des méthodes plus élaborées peuvent être utilisées, car leur mise en œuvre ne nécessite plus de moyens informatiques importants. Il est possible de les classer par rapport à leur degré de performance :

- l'analyse en composantes principales,
- les classifications automatiques (hiérarchiques ou non),
- l'analyse factorielle des correspondances,
- l'analyse discriminante.

Il faut toutefois interpréter les résultats obtenus avec beaucoup de prudence. Ce ne sont jamais des « oracles » mais des « signes » qui doivent être soumis à critique (analytique, documentaire, scientifique), afin de discriminer les corrélations des artefacts. Cela nécessite la collaboration de trois compétences rarement réunies par un seul individu : la compétence documentaire, pour établir un corpus de données fiables, exhaustives et pertinentes, la compétence mathématique du spécialiste de la méthode et de sa mise en œuvre informatique, enfin la compétence scientifique de l'expert spécialiste du domaine analysé.

Ces outils intelligents d'exploitation de l'information ne relèvent pas du domaine du rêve. Certaines entreprises et le CNIC s'efforcent de les mettre à la disposition de leurs utilisateurs (30, 31), et deux associations se sont constituées pour traiter de ces problèmes et de problèmes connexes : la Société française de bibliométrie appliquée, (SFBA), et la Société pour le développement de la scientométrie et de la technométrie (ADEST).

L'information est vitale pour le développement et l'innovation, tant dans les grandes entreprises que dans les PME, et pour la recherche, publique ou privée, fondamentale ou appliquée. Certes, les éléments tirés des ban-

ques de données doivent être complétés par la documentation manuelle; mais par le balayage systématique des données qu'elles permettent, elles constituent le moteur qui déclenche une analyse plus approfondie. Le bilan des possibilités existantes est résolument positif, même si nous avons pu en mesurer les limites. Les autres fonds pouvant être mis rapidement en ligne représentent un volume très impressionnant d'informations. A partir de ces éléments et des technologies existantes, il serait possible de bâtir des solutions réalistes, pour un serveur de banques de données spécialisé en chimie. Toutefois, au-delà des aspects techniques, il faudrait considérer les aspects économiques, culturels et politiques qui échappent au cadre de cet article.

RÉFÉRENCES

1. 1^{er} colloque national sur l'information en chimie, Paris, 13-14 novembre 1984, CNIC, 1985.
2. 2^e colloque sur l'information en chimie, Lyon, 13-14 novembre 1986, CNIC, 1987.
3. Marx, Bernard, « Banques de données », *Les Techniques de l'ingénieur*, T7101.
4. Chambaud, Serge et d'Arcy, Marie-Christine, « Banques de données », *Les Techniques de l'ingénieur*, H 3900.
5. Dutheuil, Christian, « Le Point sur les banques de données spécialisées en chimie », *1^{er} colloque*, op. cit., p. 61-62; *Parfums, arômes et cosmétiques*, déc. 1984, n° 60, p. 37-52; *Informations chimie*, janv. 1985, n° 256, p. 159-170; *Travail et maîtrise*, avril 1986, n° 423, mai 1986, n° 424, juin-juil. 1986, n° 425, mises à jour.
6. Arrêté du 22 décembre 1981, *Journal officiel de la République française*, 17 janvier 1982.
7. Dutheuil, Christian, « Numeric databanks in chemistry », *Conf. annuelle ICSTI, NDG report*, New-York, 1987.
8. Barnard, S.M., « Computer handling of generic chemical structure », *Conf. Chem. Struct. Assoc.*, Univ. Sheffield (GB), 26-29 May 1984, Ed. Gower, 1985.
9. Kaback, S.M., « Polymer patent information systems could be even better? », *J. Chem. Inf. Comput. Sci.*, 1985, 25 (4).
10. Willett, P., « Modern approach of chemical structure », *Conf. Chem. Struct. Assoc.*, Univ. York (GB) 8-11, 1985, Ed. Gower, 1986.
11. Dutheuil, Christian et Caron, N., *Les Banques de données en toxicologie*, CNIC, 1986.
12. Mazières, Marie-Françoise, « Les Bases et banques de données technico-économiques et leurs limites », *2^e colloque*, op. cit. *Informations chimie*, mai 1987, n° 283, p. 123-132.
13. Crafts-Lightly, A., *Information sources in biotechnology*, Ed. VCH, 1986.
14. Dutheuil, Christian, « Les Biotechnologies dans les banques de données en information scientifique et technique », *stage CNIC-BIOFUTUR*, CRC, 1987.
15. Id., « Banques de données factuelles en biotechnologies », *stage CNIC-BIOFUTUR*, CRC, 1987.
16. Id., « Les Collections de culture », *stage CNIC-BIOFUTUR*, CRC, 1987.
17. Id., « L'Information en matière de brevet et son exploitation », *2^e colloque sur l'information en chimie*, Lyon, 13-14 novembre 1986: notes de synthèse et résumés des communications.
18. Lucas, N., « Bases et banques de données japonaises », *Documentaliste*, mars-avril 1984, vol. 21 (2), p. 60-67.
19. Marx, Bernard, « Banques de données scientifiques et techniques au Japon », *Le progrès scientifique*, 1982, n° 27, p. 19-24.
20. Rosselin, J., *Rapport de mission*, MIDIST, 1983.
21. ANRT, *Répertoire des banques de données en conversationnel*, 10^e éd., diff. Lavoisier, 1987.
22. GFFIL, *Annuaire des banques de données françaises*, A jour, 1986.
23. CUADRA, *Directory of online databases*, New-York, Elsevier, 1987.
24. Creysseil, P., « Le Troisième âge de l'information: perspectives de l'information scientifique et technique », *Recherche technologique*, 1986, n° 2, p. 36-46.
25. Sandino, D., « Bases et banques de données en chimie analytique », *Les techniques de l'ingénieur*, p. 14.
26. Déroulède, A. et Dutheuil, C., *Information chimique: inventaire des ressources en France*, CNIC, 1983.
27. « Application des méthodes de la micro-informatique à la chimie », *Journées scientifiques nationales du CNRS, INPT-ENS de chimie de Toulouse*, 26-27 mai 1982.
28. « Computer in chemistry », *EUCHEM conf.*, Univ. technol. de Compiègne, 11-13 octobre 1983.
29. 1^{er} colloque, op. cit., p. 78-81.
30. Dutheuil, C., « Vers une exploitation stratégique de l'information »: *22^e rencontres internat. de chimie thérapeutique*, Clermont-Ferrand, 3-5 septembre 1986, à paraître.
31. Id., « Applications de l'analyse factorielle à l'expertise scientifique à partir des banques de données », *Journées Intelligence artificielle, systèmes experts, documentation*, ADBS, Paris, 3 mars 1987.

Tableau 1
Fichiers de substances

nom du fichier	producteur	serveur	volume	critères d'interrogation
EURECAS (non-polymères) POLYCAS (polymères)	CAS	QUESTEL	8 · 10 ⁶	(3), (4)
REG file/ de CAS ONLINE		STN Int.		(1), (2), (3), (4), (6)
CHEMDEX (2 fichiers)		ORBIT		(1), (2), (3), (6)
CHEMSIS CHEMNAME CHEMSEARCH CHEMZERO		DIALOG		(1), (2), (3), (6)
CNAM		DATA-STAR		(1), (2), (3), (6)
SANSS (65 sources)	NIH-EPA	CIS	350 000	(1), (2), (3), (4), (6)
CHEMLINE	NLM/CAS	NLM DIMDI BLAISE	715 000	(1), (2), (3), (6)
ISI-IC (index chemicus)	ISI	QUESTEL*	3 · 10 ⁶	(4)
CLAIMS Compound registry	IFI/Plenum	DIALOG ORBIT	14 000	(1), (5)
SPECTRA (spectro masse)	NBS	QUESTEL	38 000	(3), (4)
ECOIN	CEE	DATA CENTRALEN	68 000	(1), (3)
RTECS	NLM	CIS NLM DIMDI	63 000	(1), (2), (3)
TSCA	EPA	DIALOG ORBIT	58 000	(1), (3)
HSDB (125 sources)	NLM Toxnet	NLM	4 200	(1), (3)
Dictionnaire substances réglementées	CNIC	--	6 000	(1),(2),(3),(6)
PHARMSEARCH (brevets pharmaceutiques)	INPI	--		(1),(2),(3),(4)

* fichier déchargé fin 86

critères d'interrogation :

(1) noms triviaux, commerciaux, nomenclatures

(2) fragments de nomenclature

(3) numéro de registre (RN de CAS)

(4) structure des composés

(5) code fragmentaire

(6) éléments complémentaires (type de composé, formule moléculaire, comptage d'élément, analyse cyclique...)

Tableau 2

Propriétés physico-chimiques et thermodynamiques

THERMODATA	<i>produit et servi par Thermodata (F), propr. thermodyn. de 3 000 comp. de la chimie organique</i>
THERMDOC	<i>produit et servi par Thermodata, réf. bibliogr. de la thermodyn. inorganique</i>
IFP-TH	<i>produit par l'Institut français du pétrole, servi par QUESTEL, 45 000 données depuis 1970, sur 250 produits, 91 propriétés, 16 références d'enthalpie</i>
DETERM	<i>produit par DECHEMA (RFA), servi par INKA, comprend plusieurs modules dont SDC, 10 000 références de propr. phys. de 550 produits, associées à un programme de calcul; SDR, 10 000 réf. bibliogr. relatives aux propr. phys.</i>
GAPHYOR	<i>produit par le Laboratoire de physique des gaz et plasmas de l'Université de Paris-Sud, Orsay, servi par GAPHYOR, 165 000 données depuis 1970 sur les propr. des petites molécules gazeuses et des collisions</i>
PPDS	<i>produit par Institution of chemical engineers (GB), servi par CISI, GEIS, Aquitaine service, et CDC, propr. phys. et thermodyn. de 850 composés, 18 constantes, 20 variables, entre -50°C et 1000°C, mélanges jusqu'à 20 composants, composé de différents modules</i>
DIPPR	<i>Design institute for physical property data (Etats-Unis), co-produit par différents partenaires américains sous la responsabilité de l'American institute of chemical engineers (Aiche), servi par STN int, 765 réf. fournissant des données sur plus de 1 000 substances commercialisées</i>
Technical data	<i>Différents fichiers (coef. de partage, log P, pression de vapeur, solubilité) produits par texas A à M University et servis par Technical data service int.</i>
CHEMIX	<i>produit et servi par CISRO (Port-Melbourne, Australie), données sur les équilibres</i>
FACT (ou FAIT)	<i>produit par R. Mill Coll Canada et Thermfact Ltd, servi par CISTI et Mac Gill University, propr. de 3 800 comp. inorganiques et systèmes binaires aqueux, programme d'évaluation des propriétés</i>
CESARS	<i>données et programme d'évaluation de propriétés produit par Dept. Nat. Res. office materials control, State of Michigan (Etats-Unis), et servi par CIS</i>
CHRIS	<i>propr. phys. et thermodyn. programme d'évaluation des paramètres de sécurité, produit par US Coast guard de Washington et servi par CIS</i>
HYDROGENE DATA	<i>produit par ENSPC de Paris, servi par Thermodata, données sur l'hydrogène et la fragilisation des métaux, 7 300 réf. depuis 1980</i>
EPIC	<i>produit par l'Institut de génie chimique industriel de l'Université de Liège, servi par CISI et l'Université de Liège, 723 composés, propr. et évaluation</i>
MANLABS NPL	<i>produit par Manlab Inc, Cambridge (Etats-Unis), 2 000 composés inorganiques, métaux, alliages, gaz en solution...</i>
DORTMUND VLE/LLE	<i>produit par Friedrich Ubdde GmbH, Université de Dortmund, servi par Chemshare corp et l'université de Dortmund, 13 000 données sur 1 200 composés pour les équilibres vapeur-liquide et liquide-liquide</i>
CHEMTRAN	<i>produit et servi par Chemshare corp, propr. phys. des corps purs et mélanges</i>
NBS collections	<i>nombreuses tables thermodynamiques et compilations de propriétés, programme d'évaluation, exploitables sur micro-ordinateur</i>
MELODIC	<i>produit par le laboratoire de cinétique appliquée, CNRS-Nancy, 6 000 réf. et 3 000 données thermodyn., cinétiques, génie des réactions, critères de sécurité</i>

Tableau 3
Banques de spectres RMN

C13 NMR	<i>produite par Fiz Energie Physik Math et BASF (RFA), servie par INKA, 55 000 spectres carbone 13</i>
CNMR	<i>produite par Netherlands information combine (GB), servie par CIS, 15 000 spectres sélectionnés dans la littérature</i>
NMR-LIT	<i>fichier bibliographique produit par NIH, EPA, Preston publications inc, servi par CIS, 43 000 réf. depuis 1984</i>
RMN DATA	<i>produite et servie par ITODYS, Université Paris VII, 18 000 spectres avec la structure du composé interrogeable par DARC</i>
MEDAGRA, RMN	<i>produite par le laboratoire de biophysique RMN de la Faculté de médecine de Rennes, applications biomédicales et agro-alimentaires de la RMN</i>
FNMR	<i>produite et distribuée par Preston scientific Ltd, Fraser Williams (GB), spectres RMN du Fluor 19 de 3 000 composés, exploitable par micro-informatique</i>

Tableau 4
Banques de spectres de masse

MASS SPECTRA	<i>collections compilées du NBS (Washington), EPA (Washington) et NIH (Bethesda) plus de 42 000 spectres, servi par CIS (fichier MSSS + logiciel d'exploitation micro-ordinateur MASCOT), IC Inc., QUESTEL (fichier SPECTRA)</i>
MASS SPECTROMETRY BULLETIN	<i>fichier bibliographique produit par la Royal society of chemistry (GB), 160 000 réf. depuis 1966, servi par INFOLINE, ESA-IRS,</i>
SPECMA	<i>produite par la Faculté des Sciences de Marseille-St Jérôme, comprend plusieurs fichiers :</i> <ul style="list-style-type: none"> — MS, 1 200 spectres et indices de Kovats pour les composés hétérocycliques volatils et à caractère aromatique — IF, 5 000 spectres et indices de Kovats pour des composés volatils à flaveur et fragrance — VPC-MS, données sur le couplage de la chromatographie gazeuse et de la spectrométrie de masse pour les composés volatils.
WMSSS	<i>Willey mass spectral search system, produit par J. Wiley and sons Inc, 72 000 spectres de 50 000 composés</i>
MASS DATA	<i>produite et servie par ITODYS, Université Paris VII, 38 000 spectres</i>

Tableau 5
Banques de spectres IR et UV

IRSS	<i>produite par EPA et Institut B Kidric (Yougoslavie), 4 000 spectres IR, servie par CIS</i>
Nicolet collection	<i>produite par Nicolet instrum. corp., Madison (Etats-Unis), banque de spectres IR de référence, système d'acquisition de données et programmes d'élucidation</i>
SPIR	<i>produite par ASTM, Philadelphie (Etats-Unis), 145 000 spectres IR à partir de 10 collections de données et de la littérature, servie par CISTI et IRGO</i>
GEISA	<i>produite par le laboratoire de météorologie dynamique de l'Ecole polytechnique de Palaiseau, 23 000 données spectrales sur des molécules de l'atmosphère terrestre et des planètes</i>
IRIS	<i>produite par Sadtler research lab. (Etats-Unis), servie par Univ computing Co., 200 000 spectres IR</i>
ESTE-DBS	<i>produite par Univ. technologie de Wraclaw (PL), spectrométrie électronique d'émission (UV, visible)</i>
IR data collection	<i>produite par Elsevier Ltd (NL)</i>
GAMS	<i>Le centre d'information spectroscopique du GAMS donne accès à une collection de 25 000 spectres IR et une collection de 35 000 spectres UV (+ masse et RMN)</i>

Tableau 6
Banques des spectres autres

TTAS	<i>produite par l'Université Notre-Dame (Etats-Unis), triplet-triplet absorption, spectres de 1 200 composés</i>
RATES	<i>produite par l'Université Notre-Dame (Etats-Unis), données sur les radicaux libres produits par les photons dans 4 000 réactions de procédés radicalaires en solution aqueuse</i>
SDBS	<i>produite par les laboratoires nationaux de chimie industrielle de Higoshi et Ibaraki (J), 1 300 spectres ESR (+ spectro RAMAN)</i>
RADIATION CHEMISTRY AND PHOTOCHEMISTRY	<i>produite par l'Université Notre-Dame (Etats-Unis), données numériques et factuelles, réf. bibliogr. sur les processus chimiques initialisés par les radiations ionisantes en solution, 78 000 données</i>
METASTABLE CHEM SPECIES IN SOL	<i>produite par l'Université Notre-Dame (Etats-Unis), données photochimiques, physicochimiques et radiochimiques pour les intermédiaires métastables en solution</i>
PROTON AFFINITY	<i>produite par le NBS (Etats-Unis), 6 000 composés</i>

Tableau 7
Banques cristallographiques

CRYSTAL DATA	<i>produite par le Crystallographic data Center de Cambridge (GB) et complétée par ITODYS (serveur), 30 000 spectres X</i>
PDSM	<i>produite par Joint committee on powder diffraction standards (JCPDS), Sandia nat. lab. Albuquerque (Etats-Unis), diagrammes de poudres pour 32 000 composés inorganiques (diffraction électronique, X)</i>
CRYSTALLOGRAPHIC DATA	<i>produite par le Crystallographic data Center de Cambridge (GB), 38 000 spectres et 42 000 réf. bibliogr. de structures cristallines et moléculaires, servie par CISTI (CRYSTOR), CIS (CRYST)</i>
CRYSTDAT	<i>produite par le NBS, Washington (Etats-Unis), servie par CISTI, 95 000 données cristallographiques extraites depuis 1912 de 1 300 journaux (X, diffraction électronique, diffraction de neutrons)</i>
ICSD	<i>produite par l'Institut de chimie inorganique de l'Université de Bonn (RFA), servie par INKA (fichier ICSD) et CISTI (fichier CRYSTIN), 23 000 données cristallographiques sur les solides inorganiques</i>
CRYSMET	<i>produite par Nat. res. council d'Ottawa, servie par CISTI, 6 000 données métallographiques obtenues par des méthodes de diffraction depuis 1913</i>
SCIF	<i>Single crystal identification file, produite par NBS (Etats-Unis), servie par CIS (X-TAL) CISTI (CRMSTAD) et IClnc (XTAL), paramètres de la cellule cristalline réduite pour 95 000 composés</i>

Tableau 8
Banques en cinétique

—	<i>paramètres cinétiques et mécanisme des réactions de décomposition thermique, table des $\ln(x)$, produite par Université de Calcuta (Inde)</i>
—	<i>cinétique chimique des réactions de combustion (300 à 2 500°K), produite par NBS, Gaithersburg (Etats-Unis)</i>
Chem Kinetics code package	<i>produite par Sandia lab. Livermore (Etats-Unis), banque associée au logiciel INTERPRETER de Chemkin</i>
—	<i>banque de données cinétiques pour les réactions de cracking (ind. pétrochimique), produite par l'ECAM, Châtenay-Malabry</i>

R

Tableau 9
Banques de chromatographie

VFC-MS	<i>produite par la faculté des Sciences de Marseille Saint-Jérôme, composante du système SPECMA (cf. tableau 4)</i>
—	<i>banque sur micro-ordinateur pour HPLC, produite par Burrouhgs Wellcomes (GB)</i>
—	<i>banque des indices de rétention pour HPLC, produite par l'Université technol. de Loughborough (GB)</i>
—	<i>banque des indices de rétention et temps de rétention pour la chromatog. gazeuse, produite par Mtd Police d'Ottawa (Canada)</i>

Tableau 10
Banques diverses en analytique

ANALYSES CHIMIQUES	<i>produite par l'Institut de géologie de l'Université de Rennes, 1 000 données sur les teneurs moyennes et les traces des éléments</i>
ARTEMISE	<i>produite par le Centre de recherches pétrographiques et géochimiques, CNRS Nancy, 25 000 données sur l'analyse chimique des roches</i>
DEPHRA	<i>produite par le Centre de recherche en pétrographie et géochimie, CNRS Nancy, 8 000 données sur les ophiolites et roches associées</i>
ROCHES GEOLDOC	<i>produite par l'Institut de géodynamique, Université Bordeaux III, données sur la physique, la chimie et la géologie des carbonates naturels</i>
ANALYTICAL ABSTRACTS	<i>produite par la Royal society of chemistry (GB), servie par INFOLINE, DATA-STAR, ORBIT, 75 000 réf. bibliogr. depuis 1980</i>

Tableau 11
Banques de propriétés nucléaire

INIS	<i>International nuclear information system, produite par l'Agence internationale de l'énergie atomique de Vienne (Autriche), servie par ESA-IRS, STN Int., BELINDIS, de 700 000 à 1 million de réf. bibliogr. (selon serveur) depuis 1972, informations d'origine Etats-Unis, URSS, NL, RFA, GB, F, J</i>
ENDSF	<i>Evaluation nuclear structure data file, produite par National nuclear center de Brookhaven (Etats-Unis) et Fiz Karlsruhe (RFA), 14 000 données depuis 1930 sur les propriétés, la structure nucléaire et la période des éléments, servie par INKA</i>
NSR	<i>Nuclear structure reference, produite par le National nuclear center de Brookhaven et FIZ Karlsruhe, servie par INKA, 86 000 réf. depuis 1910</i>
MED LIST	<i>produite par le Nuclear data project, Oak Ridge (Etats-Unis) et FIZ Karlsruhe, servie par INKA, 3 000 données sur les caractéristiques des radiations nucléaires et atomiques produites par tous les radio-isotopes connus</i>
DRF-DRE	<i>produite par le CEA, servi par CISI, paramètres d'isotopes radio-actifs</i>
NUCLEAR	<i>produite par FIZ Karlsruhe, servi par INKA, 500 000 réf. sur les recherches nucléaires et radiochimiques</i>

Tableau 12
Banques de normes

NORIANE	<i>produite par AFNOR, servie par QUESTEL, normes et réglementation technique françaises, CEE et internationales depuis 1976, 45 000 réf., sera complétée par une indexation des composés chimiques et associée à un dictionnaire chimique (cf. tableau 1), réalisés par le CNIC</i>
NORMATERM	<i>produite par AFNOR, servie par GSI ECO, banque terminologique bilingue (F-GB)</i>
NORM	<i>produite par l'Institut belge de normalisation, servie par BELINDIS, 3 200 normes et réglementations techniques s'appliquant à la Belgique</i>
STANDARDS AND SPECIFICATIONS	<i>produite par National standard assoc. (Etats-Unis), servie par DIALOG et ESA-IRS, 115 000 normes et réglementations américaines depuis 1950</i>
INDUSTRY AND INTERNATIONAL STANDARDS	<i>produite par Information handling serv. (Etats-Unis), servie par BRS, normes et réglementations techniques américaines et internationales (NBS, ANSI, IHS)</i>
STANDARD	<i>produite par Norwegian standard assoc, servie par NSI, 2 500 normes et 650 réglementations norvégiennes</i>
BSI STANDARDLINE	<i>produite par le British standard institution, servie par INFOLINE</i>
DITR	<i>produite par Deutsches information Zentrum für teschnisches Regeln (RFA), servie par FIZ TECHNIK, 40 000 normes DIN + réglementations techniques</i>
NORMADRID	<i>produite par Empresa provincial de informatica de Madrid (EPIMSA), servie par EPIMSA, 7 000 normes espagnoles (en espagnol)</i>
STANDARD SEARCH	<i>produite par Society of automotive engineers (Etats-Unis), servie par ORBIT, 15 000 documents ASTM et SAE</i>

Tableau 13
Catalogues industriels

TRANSINOVE	<i>produit par l'INPI, banque d'opportunités des brevets et licences cessibles</i>
LABINFO	<i>banque des connaissances et des techniques du CNRS et de l'ANVAR, répertoire des services et compétences scientifiques et techniques des laboratoires français de recherche publics et professionnels, servie par QUESTEL</i>
TELELAB	<i>produit par la DBMIST, répertoire des laboratoires de recherche universitaires et publics français, servi par SUNIST et accès vidéotex</i>
FINE CHEMICAL DIRECTORY	<i>produit et distribué sur disquettes et CD-ROM, par FRASER-WILLIAMS (GB), catalogue des produits chimiques et de leurs producteurs distributeurs en Grande-Bretagne</i>
QUEMQUEST	<i>produit par Pergamon infoline Inc, servi par INFOLINE, catalogues de 50 producteurs de produits chimiques européens et américains, 174 000 réf., constitue une extension du Fine chemical directory</i>
JANSSEN	<i>catalogue des 11 500 produits de chimie fine du producteur Janssen, servi par QUESTEL</i>
CSCHEM CSCORP	<i>produit par Directories publishing company, Inc. (Etats-Unis), catalogue de 110 000 produits fabriqués ou distribués aux USA, servi par STN Int.</i>
TECHNOTEC	<i>produit par Control data corp (Etats-Unis), banque d'opportunités pour les transferts de technologie, servie par CDC</i>
DEQUIP DETEQ	<i>répertoire d'équipements et d'appareillages en ingénierie et génie chimique, en ingénierie de l'environnement, produits par DECHEMA (RFA), servis par STN Int., correspondent aux catalogues AHEMA</i>
SOUTRAITEL	<i>banque d'appels d'offre des grandes entreprises et du secteur public</i>
MALLET	<i>catalogue de produits distribués par MALLET SA, accessible sur vidéotex</i>
ENSCHI	<i>banque de données sur l'enseignement de la chimie, produite par l'Ecole normale supérieure de Saint-Cloud</i>

R**Tableau 14**
Banques d'application**Produits pharmaceutiques et médicaments**

DIOL	<i>Drug information on line, produite par Excerpta medica, 30 000 composés</i>
DIF	<i>Drug information full text, produite par l'American society of pharmacists, servie par DIALOG et BRS, 50 000 médicaments</i>
MARTINDALE INDEX	<i>produite par the Pharmaceutical society of Great-Britain, servie par DATA-STAR, dictionnaire en texte intégral de 5 130 médicaments</i>
BIAM	<i>produite par le BIAM, servie par CITI2, banque de données sur les médicaments (10 000 données), 2 600 principes actifs, interactions médicamenteuses</i>
INTERMEDE	<i>produite par le Centre de médecine préventive de Nancy, 5 000 réf. bibliogr. et 2 000 sur les interactions médicamenteuses</i>
MAJIM	<i>produite par le CHU de Rennes, 5 000 données sur les interactions médicamenteuses</i>
RINGDOC	<i>produite par Derwent publication Ltd (GB), servie par ORBIT et DIALOG, 900 000 réf. bibliogr. depuis 1964</i>
IPA	<i>International pharmaceutical abstracts, produite par American society of hospital pharmacists, servie par BRS, DIALOG et ESA-IRS, 120 000 réf. bibliogr. depuis 1970</i>
PNI	<i>Pharmaceutical news index, produite par Data Courier Inc. (Etats-Unis), servie par DIALOG, 175 000 réf. bibliogr. depuis 1976, données économiques, juridiques et législatives</i>
MERCK INDEX	<i>produite par MERCK (Etats-Unis), servie par CIS et QUESTEL, 10 000 composés utilisés dans l'industrie pharmaceutique</i>
De HAEN DRUG DATA	<i>produite par P. DE HAEN Int. Inc. (Etats-Unis), servie par DIALOG, 80 000 données sur les médicaments, produits utilisés, produits en expérimentation, produits à l'étude, réactions et interactions</i>
THERIAQUE	<i>produite par le Centre national d'information sur les médicaments hospitaliers (CNIMH) accessible par vidéotex, répertoire des médicaments utilisés en milieu hospitalier</i>

Agrochimie

PESTDOC I, II	<i>produite par Derwent publication Ltd (GB), servie par ORBIT et DIALOG, 127 000 réf. bibliogr. depuis 1968 + 15 000 réf. de congrès et rapports depuis 1975</i>
AGROCHEMICALS	<i>produite par la Royal society of chemistry (GB), servie par DATA-STAR, 20 000 données de l'Agrochemicals handbook et de l'European directory of agrochemical products</i>
PESTICIDE DATABANK	<i>produite par British crop. protect. council (GB), servie par INFOLINE, 3 000 données sur plus de 600 composés</i>
PHYTOTOX	<i>produite par l'Université de l'Oklahoma (Etats-Unis), servie par CIS, 70 000 données sur les effets biologiques des produits chimiques</i>
STANDARD PESTICIDE FILE	<i>produite par Derwent publication Ltd (GB), servie par ORBIT, 3 000 données textuelles et numériques sur les pesticides</i>

Divers

GLASSFILE	<i>produite par Stazione speumentale del Vetro de Murano (It), servie par ESA-IRS, 50 000 réf. bibliogr. et données textuelles ou numériques sur les propriétés des composés à l'état vitreux</i>
CORROSION	<i>produite par Orbit information technol., Santa-Monica (Etats-Unis), servie par ORBIT, 2 500 données textuelles et numériques sur la corrosion des métaux, du carbone, des verres et matériaux vitreux, des alliages, des caoutchoucs plastiques et polymères à l'égard de 600 substances chimiques</i>
AROM	<i>banques de 1400 molécules à caractère odorant, prop. phy., prop. chimiques, toxicité et caractéristiques odorantes, produite par CNRS-Muséum d'histoire naturelle, SPIIA (F), distribuée par le CNIC</i>

Tableau 15
Banques en biotechnologies

nom de la banque	producteur	serveur	
BIOTECHNOLOGY PASCAL BIOTECHNOLOGIES	Derwent CDST/CNRS	ORBIT QUESTEL ESA-IRS	125 000 réf. bibliog. 10 000 réf. depuis 1982
CURRENT BIOTECHNOL. ABSTRACTS	RSC	INFOLINE DATA-STAR ESA-IRS	17 000 réf. depuis 1983
TELEGEN SUPERTECH	EIC intelligence	DIALOG ESA-IRS DIMDI	117 000 réf. depuis 1973
BASE BIBLIOG. EN BIOTECHNOLOGIES	INSERM	CITI2	100 000 réf. extraites de 65 revues, indexées par des chercheurs
BIOTECH UPDATE	Worldwide videotex	NEWSNET	lettre d'information sur les biotechnologies en texte intégral
CURRENT AWARENESS IN BIOL. SCIENCES	Pergamon press Ltd	INFOLINE	620 réf. bibliogr. depuis 1983
CELL	Biocommerce data Ltd	DATA-STAR	30 000 réf. bibliogr. depuis 1981
BIOMASS	IEA Biomass conversion info. center	STN INT.	23 000 réf. bibliogr. 1980
CT-BIO	CIB Marseille	VIDEOTEX SUNIST	Centre transfert des bio-industries, compé- tences des laboratoires et industries de la région PACA
LIFE SCIENCE COLLECTION	Cambridge scientific abstracts	DIALOG BRS	constituée de sous-fichiers parmi lesquels — Biotechnology research abs., 7 000 réf. depuis 1984 — Microbiology abs., 105 000 réf. depuis 1978 — Virology abs., 63 000 réf. depuis 1978
BIOREG	ORGANIBIO CNIC	CNIC	informations sur la sécurité et les réglemen- tations des biotechnologies dans les pays de l'OCDE, exploitable sur micro-ordinateur

R

Tableau 16
Banques de séquences

EMBL	<i>Nucleotide sequence data library, produite par European molecular biology laboratory (RFA), 750 réf. de données sur 1 500 séquences représentant 1,7 million de bases nucléiques, servie par BIONET, INTELLIGENETICS, CITI2</i>
NBRF-NIH	<i>banque de protéines, Protein sequence database, produite par National biomedical research foundation du Georgetown university medical center de Washington (Etats-Unis) 3 000 données, servie par BIONET, INTELLIGENETICS et CITI2</i>
Nucleic acid sequence database	<i>produite par le National biomedical research foundation servie par BIONET et INTELLIGENETICS</i>
Restriction Enzyme database	<i>produite par Dr. R. J. Roberts, servie par BIONET et INTELLIGENETICS</i>
GENBANK	<i>Genetic sequence database, produite par Bolt, Berant, Newman Inc. + EMBL de Heidelberg (RFA), servie par INTELLIGENETICS, CITI2 et PROPHET, 6 000 séquences analysées représentant plus de 6 millions de bases nucléiques</i>
BISANCE	<i>servi par CITI2, le système est constitué de EMBL + NBRF + Genbank, de programmes de séquençage</i>
ACNUC	<i>système de restructuration des données (accès hiérarchique) pour l'analyse des séquences d'animo-acides, développé par l'Institut d'évolution moléculaire de l'Université de Lyon et mis en œuvre sur CITI2 pour le système BISANCE</i>

Tableau 17
Collections de cultures

MIRDAB	<i>Microbiological resources databank, Excerpta Medica (EFCVC, WFCC), banque sur les cultures de cellules animales (dont hybridomes), végétales, bactéries, levures et champignons, virus, gènes individualisés</i>
TROPICAL DATABASE	<i>produite par l'Institut de Campino (Brésil), cultures existantes et centre d'expertise</i>
MICIS	<i>produite par le Laboratory of government chemist LGC, regroupe une dizaine de collections anglaises et européennes, données sur 50 000 organismes accès vidéotex</i>
GRIN	<i>produite par Germplasm resources information network, 600 000 données depuis 1984 sur la génétique des plantes.</i>

Tableau 18
Brevets en chimie

systemes doc.	fichiers	description	producteur	serveur
API	APIPAT	160 000 brevets depuis 1964 du domaine du pétrole et de l'énergétique (+ forage depuis 1981)	American petroleum institute	SDC
CLAIMS	USPAT	1 600 000 brevets américains de la chimie (depuis 1950) et tous domaines (depuis 1963)	IFI/Plenum DATA Company	DIALOG SDC STN Int.
	UNITERM	600 000 brevets américains de la chimie depuis 1950		
	Reassignment Reexamination	1 420 000 données sur les brevets américains réexaminés ou cédés (depuis 1950 pour la chimie, 1963 pour l'ensemble)		DIALOG
	COMPOUND	répertoire de 150 000 unitermes composés cités au moins 5 fois depuis 1950		DIALOG
	CITATIONS	2 000 000 de références brevets internationaux entrés dans les brevets américains depuis 1947		DIALOG
	COMPREHENSIVE	brevets américains de la chimie depuis 1950 indexés en détail, réservés aux souscripteurs de COMPREHENSIVE DATA BASE		DIALOG
IDC patent data file	PDB	9 400 000 brevets internationaux de la chimie depuis 1960, composés interrogeables par le code GREMAS Réservé aux adhérents de IDC	International Dokumentationsgesellschaft für chemie (RFA)	IDC
World patent index	WPI WPIL	3 300 000 brevets des 28 principaux pays (+ brevets européens et PCT) depuis : 1963 pour la pharmacie 1965 pour l'agriculture 1966 pour les polymères 1970 pour la chimie 1974 pour l'ensemble des brevets indexé par les « Punch codes » et « codes structuraux fragmentaires ».	Derwent publication Ltd (GB)	DIALOG SDC QUESTEL
	CPI	1 750 000 brevets depuis 1970, le Central patent index constitue un sous-fichier de WPI		SDC
PHARMSEARCH		brevets français européens américains de la chimie pharmaceutique et du médicament, structures interrogeables par DARC.	INPI	projet