



## NRC Publications Archive Archives des publications du CNRC

### Bases de données numériques scientifiques

Wood, Gordon H.

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

<https://doi.org/10.4224/40000002>

### NRC Publications Record / Notice d'Archives des publications de CNRC:

<https://nrc-publications.canada.ca/eng/view/object/?id=c340f559-caa3-4aa5-8e4d-5fa3f5564ebc>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=c340f559-caa3-4aa5-8e4d-5fa3f5564ebc>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

**Questions?** Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

**Vous avez des questions?** Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.



Bases de données numériques scientifiques

---

Gordon H. Wood

Conseil national de recherches du Canada

Ottawa, Canada

---

ASTED - XIe Congrès

Ottawa

le 9 novembre 1984

NRCC 23951

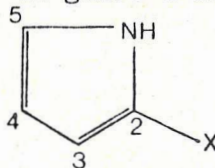
Gordon H. Wood

Institut canadien de l'information scientifique et technique  
Conseil national de recherches du Canada  
Chemin de Montréal  
Ottawa, Canada K1A 0S2

I. Introduction

Afin de mieux évaluer la puissance et l'utilité des bases de données numériques scientifiques (BDNS) comme outils de recherche pour les scientifiques et les ingénieurs, il convient de présenter trois questions typiques qui peuvent se poser dans un milieu de recherche:

1. Il faut obtenir des données sur le plus grand nombre possible de composés dans lesquels une chaîne fermée d'un atome d'azote et de quatre atomes de carbone est présente (connu sous le nom de noyau pyrrol,  $C_4H_5N$ ). N'importe quel atome (X), ou chaîne d'atomes, peut être substitué à la place de l'un des atomes d'hydrogène relié à l'un des atomes de carbone 2 ou 5. En particulier, de quelle façon varie la longueur de liaison carbone-azote et l'angle carbone (2)-azote-carbone (5) avec le genre d'atome (ou chaîne d'atomes) X?



2. Le personnel d'un laboratoire analytique possède un spectre infrarouge (un tableau indiquant l'intensité de la radiation infrarouge qui est absorbée par un composé à diverses longueurs d'ondes) d'une substance qu'il essaie d'identifier. Est-ce qu'il existe une façon de comparer ce spectre à un nombre raisonnablement élevé de spectres de composés connus?
3. Un métallurgiste s'intéresse au gadolinium, une terre rare coûteuse. Il veut savoir quels composés peuvent se former entre le gadolinium, l'oxygène et le chlore, la forme de ces composés et à quelles températures les mesures existantes ont été prises. Effectuer ces mesures lui-même serait à la fois coûteux et long; est-ce qu'il existe des données de ce genre?

Une recherche manuelle dans les collections appropriées de spectres et dans les manuels de données thermophysiques respectivement répondrait aux questions 2 et 3. Par contre, la consultation d'une base de données numériques scientifiques serait une façon beaucoup plus efficace d'y arriver. À toutes fins utiles, la consultation d'une base de données numériques scientifiques est la seule façon de répondre à la question 1. Ces questions seront abordées un peu plus en détail à la section III où il sera démontré de quelle façon les bases de données numériques scientifiques peuvent résoudre ces problèmes et d'autres problèmes connexes de façon efficace et économique.

Les scientifiques et ingénieurs connaissent l'utilisation de l'ordinateur pour effectuer des calculs et automatiser la prise de mesures; certains connaissent aussi l'utilisation des ordinateurs pour l'interrogation de vastes bases de données bibliographiques; très peu cependant pensent à utiliser l'ordinateur comme moyen d'accès à des données scientifiques pondérées tirées de la documentation mondiale. Le spécialiste de l'information, en se familiarisant davantage avec les possibilités et avantages des bases de données numériques scientifiques, peut établir des rapports profitables entre ses clients scientifiques et ingénieurs et les bases de données numériques scientifiques.

La section II définit une partie de la terminologie de base et propose certaines perspectives des bases de données numériques scientifiques. Les possibilités et avantages des bases de données numériques scientifiques sont illustrés à la section III. La section IV traite des questions d'ordre pratique d'accessibilité en direct et la section V conclut avec un rapide survol des bases de données et des services offerts par l'ICIST par l'entremise du réseau CAN/SND.

## II. Définitions

Le système de classification utilisé dans le "Directory of Online Databases"<sup>1</sup> semble obtenir une reconnaissance générale et servira ici à décrire les divers genres de bases de données offertes.

### A. Bases de données références/sources

Les bases de données de références servent à fournir ou signaler à l'utilisateur une autre source, souvent un document, pour donner plus de détails ou le texte au complet. Ce groupe peut être divisé en deux catégories: (1) les bases de données bibliographiques (renfermant essentiellement des références à de l'information publiée comme des articles de revues, des rapports, des brevets, des dissertations, des comptes rendus de conférence et des livres) et (2) les bases de données d'orientation (renfermant essentiellement des références à des sources d'information inédite comme des organismes, des personnes, des documents audiovisuels et des moyens de communication non imprimés).

Les bases de données de sources, qui englobent les données complètes ou le texte au complet de l'information originale de la source, sont divisées de façon pratique comme suit: bases numériques, textuelles-numériques et texte au complet. Les bases de données numériques englobent des représentations de données originales ou statistiquement manipulées, ou les deux; les bases de données textuelles-numériques contiennent un mélange de données numériques et d'information textuelle connexe (commentaires au sujet des données ou des références bibliographiques, ou les deux, aux données originales); les bases de données de texte au complet englobent des notices du texte complet.

B. Les bases de données numériques scientifiques

Classées comme bases de données de sources dans les catégories numérique ou textuelle-numérique, les bases de données numériques scientifiques sont une collection ordonnée de chiffres dont les valeurs:

- 1) correspondent à des propriétés, des paramètres ou des attributs d'éléments, de substances ou de systèmes différents
- 2) sont évaluées et corrigées par des spécialistes avant d'être versées dans la base de données.

Les bonnes bases de données numériques scientifiques ne doivent donc pas être considérées comme de simples compilations de chiffres. Les fonctions importantes et coûteuses de révision, d'évaluation et de correction, qui sont souvent absentes des bases de données bibliographiques, rendent les données plus fiables que celles trouvées dans la documentation en général et plus utiles à cause de la rationalisation de facteurs comme des énoncés d'incertitude et des unités de mesure.

En outre, l'utilisateur d'une base de données numériques scientifiques est plus souvent un scientifique, un ingénieur ou un technicien dans l'exercice de son travail qu'un spécialiste de l'information. Il en est ainsi en raison de la nature spécialisée des données elles-mêmes et de la formation requise pour présenter des interrogations ainsi que pour interpréter et manipuler les résultats. Dans plusieurs cas, le spécialiste de l'information repère la base de données numériques scientifiques pertinente et aide à en obtenir l'accès. Cependant, l'interrogation, le repérage et l'analyse elle-même, comme il vient d'être mentionné, sont faits par les spécialistes de la discipline.

Le tableau suivant donne un ordre de grandeur de l'abondance relative des bases de données numériques scientifiques (en se basant sur les entrées de la référence n° 1):

<u>Genre de bases de données</u>	<u>Science/Génie</u>		<u>Autre</u>		<u>Total</u>	
	<u>1983</u>	<u>1984</u>	<u>1983</u>	<u>1984</u>	<u>1983</u>	<u>1984</u>
Sources	86	86	977	961	1063	1047
Références	198	178	472	576	670	754
					<u>1733</u>	<u>1801</u>

Ainsi, en 1984, les bases de données de sources dans les disciplines des sciences et du génie représentent environ uniquement 5% du total. Ce ne sont pas toutes ces bases de données de sources en science et génie qui seraient strictement classées comme base de données numériques scientifiques, ni celles d'ailleurs qui sont en voie d'élaboration mais qui ne sont pas encore publiquement accessibles.

### C. Les systèmes de bases de données numériques scientifiques

Afin d'éviter la confusion, le terme système de bases de données numériques scientifiques (SBDNS) devrait servir à décrire un ensemble d'une ou plusieurs bases de données numériques scientifiques combinées à un groupe de progiciels permettant au scientifique ou à l'ingénieur d'interroger une base de données, de repérer des données d'intérêt et de les manipuler de diverses façons. L'utilisation d'un système de bases de données numériques scientifiques rapporte donc bien plus que de simplement parcourir électroniquement un manuel afin de trouver une entrée précise comme la prochaine section l'explique.

### III. Exemples d'utilisation - Possibilités et avantages

L'utilité et l'efficacité d'un SBDNS peuvent être illustrées à l'aide des trois questions posées dans l'introduction.

Selon l'auteur, il n'existe aucune façon pratique de répondre à la question 1 en utilisant des bases de données bibliographiques de références ou des ouvrages imprimés de référence à l'exception d'une recherche documentaire excessivement exhaustive et un grand nombre de calculs manuels. Cependant, en consultant un SBDNS comme la base de données cristallographiques de Cambridge<sup>2</sup>, la réponse à un tel problème peut être trouvée en moins d'une ou deux heures de façon simple et systématique. Il suffit pour l'utilisateur de décrire la façon dont les atomes désirés sont interconnectés (la connectivité chimique) et de demander au système de comparer cette connectivité à celle de tous les composés de sa collection. Le résultat de cette interrogation, une liste de tous les composés renfermant un fragment de pyrrol, peut par la suite faire l'objet d'opérations par un programme intégré afin d'exécuter les analyses géométriques requises.

Le client pourrait bien sûr obtenir la réponse à la question 2 en employant la manière forte, en cherchant manuellement dans les compilations et les atlas de spectres infrarouges et en cherchant celui qui ressemble au spectre inconnu. Une façon d'améliorer cette recherche consisterait à utiliser un système manuel comme, par exemple, le "Spec-Finder" commercialisé par Sadtler Research Laboratories<sup>3</sup>. Avec ce système, le spectre est divisé en 27 intervalles et le pic ou la bande le plus fort dans chaque intervalle, s'il y en a un, est codé. Un répertoire classé selon le pic le plus marqué en général indique alors à l'utilisateur les spectres dans les collections de cette entreprise qui correspondent le mieux à l'inconnu. Une autre amélioration serait obtenue en utilisant un SBDNS comme FIRST-1<sup>4</sup>, SPIR<sup>5</sup>, IRGO<sup>6</sup> ou IRIS<sup>3</sup> qui font tous appel à une base de données de spectres compilées par la American Society for Testing and Materials. Dans ces systèmes, le spectre est codé selon les emplacements de ses pics, de ses bandes et des zones sans bande et entré dans l'ordinateur pour former un "masque". Ce

"masque" est alors automatiquement comparé au grand nombre de spectres (environ 140 000) dans la base de données et l'utilisateur reçoit une liste de composés cibles dont les spectres ressemblent le mieux à l'inconnu. Dans la plupart des cas, l'utilisateur doit quand même consulter une copie papier du spectre connu afin d'établir une comparaison détaillée. Il a été néanmoins possible d'épargner beaucoup de temps et de plus, l'utilisateur est certain que la recherche est exhaustive et porte sur les données accessibles actuellement.

Il ne serait pas aussi facile de répondre à la question 3 en consultant simplement les manuels de données thermophysiques. Même en trouvant les données pour les éléments d'intérêt, la probabilité est très grande que les valeurs indiquées ne soient pas à la température voulue ni dans l'ensemble approprié d'unités de mesure. Ainsi, les données devraient être manipulées puis entrées dans un ordinateur local (ajoutant la possibilité d'erreurs de transcription) pour être traitées par certains programmes de calculs thermodynamiques. En faisant appel à un SBDNS comme FACT<sup>7</sup>, par exemple, l'utilisateur a accès immédiatement à un ensemble assez complet des données les plus récentes, dans toutes les unités appropriées, avec des programmes intégrés d'interpolation afin de couvrir les divers éventails de température. Chose toute aussi importante, il est alors possible d'avoir automatiquement accès à ces données par des routines établies de calcul thermodynamique éliminant par la suite l'ennuyeuse réentrée des données avec la possibilité d'erreurs.

#### A. Possibilités

En gardant ces exemples en mémoire, il est maintenant utile d'ébaucher la portée des fonctions qu'un SBDNS peut accomplir.

1. Repérer des données de façon rapide, exhaustive et précise à partir de vastes collections de données; repérer selon des critères pour lesquels il n'existe pas d'index; repérer des genres d'information trop détaillée et trop ennuyeuse que l'esprit humain ne puisse traiter aisément (ex. la recherche de connectivité décrite plus haut).
2. Manipuler et analyser les données de diverses façons, par exemple:
  - a) faire correspondre des courbes pour quantifier des rapports
  - b) interpoler ou extrapoler pour faciliter la comparaison de nouvelles mesures
  - c) générer des graphiques de tendance ou établir des comparaisons statistiques
  - d) produire des tracés de géométrie moléculaire
3. Simuler des expériences avec des modèles mathématiques, explorer théoriquement des procédés comme des réactions chimiques, évitant alors la nécessité d'exécuter les véritables expériences ou de construire de l'équipement prototype.

4. Formuler de nouvelles idées à partir d'observations et d'inférences statistiques sur les données elles-mêmes. La base de données cristallographiques de Cambridge est un vaste ensemble de données fondamentales fiables et les auteurs l'ont utilisée pour obtenir de l'information sur les effets de substituants, sur la réactivité chimique, sur la flexibilité moléculaire et les forces intermoléculaires (voir, par ex. 8, 9 et 10).

#### B. Avantages

Dans un exposé de synthèse en 1981, V. Hampel<sup>11</sup> traite de la plupart des avantages économiques découlant de l'utilisation d'une base de données numériques scientifiques ou d'un SBDNS. Par soucis d'exhaustivité, il est utile de résumer ici quelques-uns de ces avantages.

Les économies directes de temps du scientifique, de l'ingénieur ou du spécialiste en information sont évidentes. Voyons quelques exemples:

- 1) le temps et l'argent investis pour mesurer inutilement certaines propriétés d'une substance qui sont déjà connues
- 2) les efforts consacrés pour trouver les données cherchées
- 3) les efforts consacrés à la représentation de telles données pour utilisation ou manipulation ultérieure, sans mentionner la possibilité d'erreurs inhérentes dans l'entrée des données, le transfert de la copie papier à la forme magnétique.

Il existe d'autres facteurs, moins facilement quantifiables mais néanmoins de grande importance, comme l'à-propos des données (à la fois nouvelles et corrigées) et l'assurance que les données sont généralement plus fiables que celles trouvées dans la documentation ou dans les compilations générales.

Pour avoir une meilleure idée de la valeur d'un SBDNS, il suffit de consulter une étude<sup>12</sup> commanditée par le United States Department of Energy. Comme mesure choisie pour juger la valeur de leur base de données bibliographiques sur l'énergie, les responsables ont évalué les économies globales de temps, de matériaux ou d'équipement que l'on peut obtenir à partir de frais engagés par un scientifique qui lit un article de revue ou un rapport pertinent. Le rapport indique que l'économie moyenne pour chaque article ou rapport lu se chiffrait à 590 \$ (US) et 1280 \$ (US) respectivement. Il n'est pas déraisonnable de suggérer que la valeur des données repérées et analysées à l'aide d'un SBDNS serait à tout le moins comparable.

En bref, un SBDNS peut optimiser la proportion de temps passé par le scientifique ou l'ingénieur à des activités créatives et, en fait, peut servir de puissant outil pour ces activités.

#### IV. Accessibilité

Étant généralement plus petits que les bases de données de références, les SBDNS sont parfois diffusés par des droits de licence pour utilisation privée à l'aide d'un ordinateur local ainsi que par les réseaux publics plus conventionnels d'accès en direct. Il n'est pas nécessaire d'aborder pour les fins de la présente discussion la question de l'accès d'un SBDNS sous forme de licence privée. Il faut toutefois mentionner que cette méthode exige habituellement la conclusion d'un genre d'entente juridique avec le producteur de la base de données et nécessite habituellement une grande compétence en informatique de la part du titulaire de la licence.

En général, les SBDNS montés sur des ordinateurs principaux publics sont accessibles de la même façon et avec le même équipement que les bases de données de références. Ainsi, un bon de commande est rempli afin de fournir à l'utilisateur un code d'identification et le terminal de l'utilisateur est relié à l'ordinateur principal par l'entremise d'un modem commuté et d'un réseau de télécommunication. Par souci d'exhaustivité, la suite de cette section traite des types de terminaux appropriés et des moyens utilisés pour les télécommunications.

##### A. Les terminaux

À l'exception des applications nécessitant l'utilisation en direct de routines de traçage, tout terminal compatible avec le code ASCII (American Standard Code for Information Interchange) peut être utilisé. Cela comprend la plupart des terminaux offerts sur le marché allant du télétype de base, aux appareils vidéotextes, à l'écran cathodique intelligent.

D'autres possibilités comprennent les machines de traitement de texte et les microordinateurs munis d'un tableau de communications RS232 et d'un logiciel d'émulation de terminal. Le raccord à divers ordinateurs principaux est toujours plus facile si le terminal possède des réglages faciles à choisir comme le duplex, l'avancement automatique d'un interligne, le mode, la parité et la vitesse (110, 300 ou 1200 bits par seconde - 10, 30 ou 120 caractères par seconde - sont communes.

Pour profiter pleinement des possibilités de traçage en direct dont sont dotées certains SBDNS, il faut utiliser un terminal graphique. (Aux fins du présent exposé, un "terminal graphique" signifie un terminal bidirectionnel capable d'envoyer des caractères ASCII et de recevoir des caractères ASCII ou des codes graphiques. Les caractères ASCII sont affichés comme à l'habitude; les codes graphiques entraînent le traçage des diagrammes.) La base de données cristallographiques de Cambridge<sup>2</sup>, par exemple, possède un ensemble de traçage (PLUTO78) qui permet à l'utilisateur de générer une grande variété de diagrammes de molécules organiques. Il est important de souligner que les instructions nécessaires au fonctionnement de PLUTO78 sont toutes sous forme de caractères; l'option graphique d'un terminal n'est utilisée que pour afficher les diagrammes.

Donc, un utilisateur dont le terminal n'a pas l'option graphique peut quand même faire appel à PLUTO78 pour présenter les diagrammes désirés et les faire tracer en différé sur un appareil graphique à l'emplacement de l'ordinateur principal. Une telle procédure est certes plus lente que le mode en direct mais quand même réalisable.

Certains SBDNS, dans le but d'éviter le recours à un terminal graphique, utilisent des caractères ASCII judicieusement espacés et situés comme -, \_, x, |, etc. afin de produire des diagrammes bruts. De tels diagrammes sont acceptables lorsque leur utilité est essentiellement schématique, comme pour la connectivité chimique ou les diagrammes de phase, mais ne sont pas acceptables lorsque l'échelle et les détails sont importants.

Jusqu'à présent, il n'y a pas de code graphique équivalant à la norme ASCII. En conséquence, les fabricants de terminaux graphiques ont tenté d'élaborer leurs propres systèmes de codage de sorte, qu'en général, un ordinateur principal donné doit adopter le code qu'il produit pour le terminal graphique avec lequel il communique. Heureusement, du point de vue de l'utilisateur, un certain nombre de fabricants produisent leurs terminaux graphique dotés de la possibilité d'imiter la norme PLOT 10 de Tektronix Corporation avec laquelle la plupart des ordinateurs principaux sont compatibles.

Les terminaux vidéotextes ou les microordinateurs émulant le vidéotexte peuvent également servir à afficher des graphiques. Ces appareils utilisent le système de codage alpha-numérique de la norme NAPLPS (North American Videotext/Teletext Presentation Level Protocol Syntax) - une amélioration du système élaboré dans le cadre de Telidon. Bien qu'ils soient moins coûteux que les terminaux graphiques "classiques", les terminaux vidéotextes ont généralement une résolution moins bonne.

En bref, tout terminal qui peut communiquer avec une base de données de références peut servir à accéder à un SBDNS; un terminal graphique peut être utile dans certains cas mais n'est pas essentiel.

## B. Les télécommunications

Il est pratique de supposer que le terminal utilisé, ou l'appareil servant de terminal, soit équipé d'une interface RS232 et que la liaison avec l'ordinateur principal à distance soit établie avec un réseau public. Étant donné ces conditions, l'utilisateur n'a qu'à se préoccuper du genre de réseau à utiliser et les moyens physiques de connexion du terminal à ce réseau.

Dans les pays où ces services sont offerts, le seul choix raisonnable de liaison est le soit disant réseau de commutation de paquets. Ces réseaux, conçus pour la transmission de données, présentent des frais peu élevés, une vitesse suffisamment élevée et une très grande intégrité des données comparativement aux lignes conventionnelles des

appels interurbains. Les utilisateurs au Canada, par exemple, peuvent accéder au réseau DATAPAC (administré par Telecom Canada) par n'importe lequel des noeuds situés dans environ 80 villes. Si une personne habite dans une ville possédant un noeud et désire communiquer avec un ordinateur dans une autre ville comportant également un noeud, il n'y a aucun frais d'appel interurbain. Les frais du réseau DATAPAC tiennent compte de la quantité de données échangées et, depuis les grands centres, les frais sont très peu fonction du temps et de la distance comparativement aux coûts des appels interurbains conventionnels. Comme illustration, le tableau ci-dessous est tiré des tarifs courants publiés pour un appel de dix minutes et la transmission de 5000 caractères à Ottawa depuis les villes indiquées.

<u>Ville</u>	<u>Coûts (\$)</u>	
	<u>DATAPAC</u>	<u>Voix</u>
Halifax	0,47	9,00
Montréal	0,46	5,71
Toronto	0,47	6,21
Winnipeg	0,48	10,00
Vancouver	0,49	11,00

L'autre caractéristique importante des réseaux de commutation de paquets, la grande intégrité des données, est particulièrement intéressante pour les utilisateurs des SBDNS, car un chiffre erroné dans un tableau, par exemple, peut être plus difficile à déceler qu'un mauvais caractère alphabétique dans une chaîne de caractères d'un texte. L'intégrité des données est également d'une extrême importance dans la transmission des graphiques où une seule erreur peut ruiner tout un diagramme.

Pour brancher un terminal à un réseau, il faut utiliser un appareil électronique appelé modulateur/démodulateur (modem). Sa fonction première est de convertir (moduler) les signaux émis par le terminal en une forme appropriée pour la transmission dans le réseau jusqu'à l'ordinateur principal et de reconvertir (démoduler) les signaux envoyés par l'ordinateur au terminal. Avec certains types de modem, le raccord physique au réseau se fait acoustiquement en plaçant le combiné de l'appareil, qui est utilisé pour rejoindre le noeud du réseau, dans un réceptacle spécialement conçu sur le modem. Ce genre de modem est habituellement limité à des vitesses de transmission de 300 bits par seconde ou moins et a tendance à être plus perturbé par les parasites électriques que les modems commutés câblés où le téléphone est quand même utilisé, mais le raccord physique se fait par des conducteurs électriques. Il est donc recommandé d'utiliser des modems commutés pour communiquer avec les SBDNS en raison de leur plus grande intégrité des données et des plus grandes vitesses de transmission possibles, 1200 bits par seconde. Cette vitesse de transmission est particulièrement intéressante lors de l'envoi de graphiques à cause de l'économie du temps passé au terminal.

## V. Les SBDNS à l'ICIST

Dans le cadre de son mandat, l'ICIST a mis sur pied un programme en 1980 visant à rendre les SBDNS accessibles à la communauté scientifique canadienne. Cette section brosse un rapide survol de ce programme qui s'est fait connaître sous le nom de système CAN/SND (bases de données numériques scientifiques).

### A. Composantes et services

La création, l'accès, l'orientation et la diffusion sont les principales composantes du système CAN/SND.

Le premier élément touche aux activités concernant la production des bases de données numériques scientifiques au Canada. À titre d'exemple, l'accord avec l'université McMaster pour l'édition et la fourniture de données qui, en retour, sont échangées avec un groupe en République fédérale d'Allemagne qui produit une base de données cristallographiques pour les composés inorganiques. Un autre exemple, la collaboration entre la division de chimie du CNRC et l'ICIST qui produisent et la diffusent un système de base de données cristallographiques pour les métaux connu sous le nom de CRISTALMET.

L'accès englobe des activités touchant l'aide apportée aux Canadiens afin qu'ils puissent avoir accès aux divers SBDNS offerts à l'échelle internationale. Dans certains cas, il faut négocier des ententes d'accès; entre autres, il s'agit simplement de promouvoir une liaison entre l'utilisateur canadien et le fournisseur approprié.

L'orientation comporte la tenue de répertoires courants de bases de données qui sont accessibles à l'échelle mondiale de sorte que les Canadiens puissent obtenir de l'aide pour trouver les bases de données pertinentes.

La diffusion se traduit par la mise à la disposition des Canadiens du SBDNS que l'ICIST gère. Même si ce rôle est d'abord assumé par un réseau en direct faisant appel au Centre de calcul du CNRC via DATAPAC, la diffusion se fait aussi par voie de location de bandes et de recherches personnalisées. La location de bandes, offertes dans le cas de certains SBDNS, permet à un organisme-client de monter un SBDNS sur un ordinateur local pour sa propre utilisation. Les recherches personnalisées sont faites moyennant des frais minimes par le personnel du système CAN/SND pour les clients à qui ce mode d'accès convient mieux, ou s'il s'agit de leur seul mode d'accès. Le réseau en direct offre un accès bidirectionnel en français ou en anglais. Les manuels d'utilisation fournissant des exemples ainsi que des fichiers d'explications en direct réduisent au minimum la nécessité de la formation de l'utilisateur. À ce stade-ci cependant, des cours structurés de formation ne sont pas encore offerts.

## B. Les bases de données accessibles

En ce moment, uniquement deux SBDNS peuvent être utilisés en direct: SPIR (Spectres infrarouges) et CRISTALOR (base de données cristallographiques de Cambridge<sup>2</sup>). La base de données SPIR, qui englobe une collection d'environ 140 000 spectres infrarouges de quelque 96 000 composés, est essentiellement un outil analytique. L'utilisateur peut fournir le spectre infrarouge d'un composé inconnu et SPIR comparera ce spectre à sa propre collection et signalera les 20 composés dont les spectres correspondent le mieux à l'inconnu. La base de données CRISTALOR englobe des informations structurales et bibliographiques sur tous les composés organiques et organométalliques dont la structure de cristal a été publiée depuis 1935. Un outil de recherche très polyvalent, CRISTALOR peut être interrogée soit à partir de la connectivité chimique ou des zones bibliographiques comme le nom du composé, la formule, l'élément, l'auteur, etc. Les données numériques repérées peuvent être par la suite traitées par des programmes d'analyse géométrique ou servir à fabriquer des diagrammes moléculaires comme mentionné précédemment.

Trois autres SBDNS, présentement dans les dernières étapes d'élaboration et de vérification, seront bientôt accessibles pour utilisation générale. Il s'agit de la base de données<sup>13</sup> des structures de cristaux inorganiques (CRISTALIN), de la base de données<sup>14</sup> cristallographiques sur les métaux du CNRC (CRISTALMET) et de la base de données<sup>15</sup> sur l'identification des cristaux du National Bureau of Standards (CRISTALDAT). CRISTALIN et CRISTALMET sont de conception semblable à CRISTALOR, mais portent sur des composés différents comme leur nom l'indique. CRISTALDAT sert d'abord à identifier des composés à partir de leurs unités cristallographiques répétitives. La force actuelle du système CAN/SND qui réside dans le domaine de la cristallographie est tout à fait unique. À la connaissance de l'auteur, aucun autre service en direct n'offre un ensemble de SBDNS aussi exhaustif ou si à jour.

D'autres bases de données sont présentement à l'étude et des plans prévoient leur mise en oeuvre dès que les négociations seront terminées et que les ressources le permettront.

## C. Utilisation

Puisque le système CAN/SND est relativement nouveau (CRISTALOR fut d'abord accessible en novembre 1981) et s'adresse à une clientèle plutôt spécialisée, le nombre d'utilisateur est restreint comparativement aux systèmes bien implantés de bases de données de références. Des quelque 50 comptes existants 38 sont réellement actifs. Plus de 80% des utilisateurs, c'est-à-dire les personnes qui s'assoient au terminal, sont des scientifiques ou des techniciens; les autres sont des spécialistes en information.

VI. Conclusion

Étant donné leurs possibilités et leurs avantages, les bases de données numériques scientifiques s'avèrent des outils de recherche efficaces pour les scientifiques, les ingénieurs et les techniciens. Leur accessibilité pratique en direct avec des terminaux ordinaires et les canaux de télécommunication, combinée à leur facilité d'utilisation, rend leur accessibilité de plus en plus facile tant au pays qu'à l'échelle internationale.

## Références

1. Directory of Online Databases, Spring 1984 edition, Cuadra Associates Inc., California, U.S.A., p. 7
2. F.H. Allen et al, The Cambridge Crystallographic Data Centre: Computer-based Search, Retrieval, Analysis and Display of Information, Acta Cryst, Vol. B35, 1979, p. 2331
3. Sadtler Research Laboratories, 3316 Spring Garden St., Philadelphia, PA 19104, U.S.A.
4. DNA Systems Inc., P.O. Box 1424, Saginaw, Michigan 48605, U.S.A.
5. ICIST, Conseil national de recherches du Canada, Ottawa, Canada, K1A 0S2
6. Chemir Laboratories, 761 W. Kirkham, St. Louis, MO 63122, U.S.A.
7. Thermfact, 447 Berwick Ave., Mont-Royal, Québec, Canada, H3R 1Z8
8. R. Taylor et O. Kennard, Crystallographic Evidence for the Existence of C-H...O, C-H...N and C-H...Cl Hydrogen Bonds, J. Am. Chem. Soc., Vol. 104, 1982, 5063
9. E. Bye, W.B. Schweizer et J.D. Dunitz, Chemical Reaction Paths. 8. Stereoisomerization Path for Triphenylphosphine Oxide and Related Molecules: Indirect Observation of the Structure of the Transition State, J. Am. Chem. Soc., Vol. 104, 1982, 5893
10. R.E. Rosenfield, Jr. et P. Murray-Rust, Analysis of Atomic Environment of Quaternary Ammonium Groups in Crystal Structures, Using Computerized Data Retrieval and Interactive Graphics: Modeling Acetylcholine - Receptor Interactions, J. Am. Chem. Soc. Vol. 104, 1982, 5427; P. Murray-Rust et al, Intermolecular Interactions of the C-F Bond: The Crystallographic Environment of Fluorinated Carboxylic Acids and Related Structures, J. Am. Chem. Soc., Vol. 105, 1983, p. 3206
11. V. Hampel, Fact Retrieval in the 1980's, AGARD Conference Proceedings No. 304, Technical Information Panel Specialists' Meeting, Munich, September 1981, p. 6-2 to 6-4
12. D.W. King et al, Value of the Energy Database, U.S. Dept. of Energy, March 1982, DOE/OR/11232-1, (DE 82014250)
13. G. Bergerhoff et al, The Inorganic Crystal Structure Database, J. Chem. Inf. Comp. Sci. Vol. 23, 1983, pp. 66-69
14. L.D. Calvert et J.R. Rodgers, The Metal Data File, Computer Physics Communications, (to be published)
15. J.R. Rodgers et A.D. Mighell, Searching the NBS Crystal Data File, American Crystallographic Association Series II, Vol. II, Part I, 1983, p. 15