

## LES BASES DE DONNÉES : OUTILS DE STOCKAGE ET D'EXPLOITATION DES ÉCHANTILLONNAGES PISCICOLES.

C. SANLAVILLE-BOISSON (1), J.M. OLIVIER (2), M. SIMIER (3), T. CHANGEUX (4),  
G. POIZAT (4), F. TORRE (4), J. BELLIARD (5).

- 
- (1) Université Claude Bernard Lyon 1, Laboratoire de Biologie Micromoléculaire et Phytochimie, 43 boulevard du 11 Novembre 1918, 69622 Villeurbanne Cedex, France.
  - (2) U.R.A. C.N.R.S. 1974, Ecologie des Eaux Douces et des Grands Fleuves, Université Claude Bernard Lyon 1, 43 boulevard du 11 Novembre 1918, 69622 Villeurbanne Cedex, France.
  - (3) O.R.S.T.O.M., Centre de Montpellier, 911 avenue Agropolis, B.P. 5045, 34032 Montpellier Cedex 1, France.
  - (4) U.R.A. C.N.R.S. 1974, Ecologie des Eaux Douces et des Grands Fleuves, Université Claude Bernard Lyon 1, Dynamique Ecologique et Sociale en Milieux Deltaïques, 1 rue Parmentier, 13200 Arles, France.
  - (5) C.E.M.A.G.R.E.F., Division Qualité des Eaux, 14 avenue de Saint-Mandé, 75012 Paris, France.

### RÉSUMÉ

L'étude des relations espèces-environnement à différentes échelles d'espace et de temps impose la manipulation de grands volumes de données et la confrontation de différents types d'informations (biologiques et environnementales). Afin d'optimiser la gestion informatique des données écologiques, plusieurs bases de données (correspondant à des données de structures différentes) ont été élaborées sur micro-ordinateur en utilisant des systèmes de gestion de bases de données différents autour desquels ont été développées des applications spécifiques destinées à faciliter et compléter la gestion de l'information, depuis la saisie des données jusqu'à l'obtention des tableaux à analyser.

**Mots-clés** : systèmes fluviaux, peuplements piscicoles, échelles spatiales et temporelles, gestion des données, bases de données relationnelles.

### USING DATA BASES FOR SAMPLE MANAGEMENT IN FISH ECOLOGY.

### ABSTRACT

The study of species/environment relationships at different spatial and temporal scales requires the processing of large data sets including both biological and environmental information. In order to optimize the management of ecological data in computers, several databases (according to different data structures) were constructed using different database management systems on microcomputers. Together with these systems, specific applications were developed in order to improve information management, from data input to obtaining the tables to be analyzed.

**Key-words** : fluvial systems, fish communities, spatial and temporal scales, data management, relational databases.

### 1. INTRODUCTION

L'évolution récente de la problématique scientifique concernant l'écologie des poissons en milieu lotique nécessite une redéfinition complète des échelles spatio-temporelles et des paramètres physico-chimiques et biologiques à prendre en compte.

En raison des caractéristiques biologiques particulières des espèces composant les peuplements piscicoles des grands cours d'eau (durée de vie élevée, maturité sexuelle tardive, changements ontogéniques d'habitat et mobilité) (SCHLOSSER, 1990 ; WERNER et GUILLIAM, 1984), le rétablissement de la structure des peuplements (richesse spécifique, densité de population, structure en âge et en taille) après une perturbation provisoire nécessite généralement de 10 à 50 ans, voire plus (YOUNT et NIEMI, 1990). Cette période est particulièrement longue lorsque la perturbation induit une modification profonde et durable de l'habitat (endiguements, barrages et par conséquent hydrologie). Si le cadre spatio-temporel des études ichtyologiques en milieu fluvial doit, dans la mesure du possible, intégrer une échelle spatiale suffisante (bassin versant ou hydrosystème), il doit donc surtout s'inscrire sur le long terme (FRANKLIN, 1989 ; SCHLOSSER 1990 ; YOUNT et NIEMI, 1990 ; PERSAT, 1991 ; DETENBECK *et al.*, 1992). Ce type d'études doit permettre de définir l'amplitude de variation naturelle dans le temps de la structure des peuplements (SCHLOSSER, 1990) et de distinguer la part relative de l'aménagement des cours d'eau dans les variations observées (YOUNT et NIEMI, 1990). Peu de données sont actuellement disponibles dans la littérature concernant les études à long terme sur les grands cours d'eau (DETENBECK *et al.*, 1992). Ainsi, à l'initiative du Programme Interdisciplinaire de Recherche sur l'Environnement (P.I.R.) intitulé «Signification des échelles spatio-temporelles dans la variabilité des peuplements piscicoles des grands cours d'eau», plusieurs bases de données (BD) regroupant des informations en provenance de plusieurs fleuves français (essentiellement Rhône et Seine) et tropicaux (Afrique de l'Ouest et Amazonie) ont été élaborées. Ces travaux viennent s'ajouter à la mise en place d'une banque nationale «Hydrobiologie-Poissons» par le Conseil Supérieur de la Pêche. Si l'élaboration par ROUSSEAU d'une base de données ichtyologiques (ROUSSEAU *et al.*, 1985 ; PERSAT, 1988) constitue un travail pionnier, l'organisation en BD dans le cadre du P.I.R., rendue nécessaire par la quantité d'informations à traiter (données biologiques et environnementales), n'a pu être réalisée que grâce au développement récent de logiciels commerciaux ou systèmes de gestion de bases de données (SGBD).

La création d'une BD a pour objectif d'organiser l'ensemble de l'information recueillie sur le terrain et d'en faciliter l'exploitation, et comprend deux étapes :

- l'élaboration du modèle conceptuel des données : elle consiste en une description de la sémantique des données, pour en représenter de façon fiable et non redondante les structures, les contraintes d'intégrité (règles qui doivent être respectées par les valeurs des données pour être considérées comme vraisemblables (FLORY, 1987)) et les règles de gestion globale, indépendamment de toute implantation et de toute utilisation (ANSI, 1975),
- l'implantation physique de la base sur l'ordinateur : elle prend en compte les modalités d'utilisation (nature du matériel et du logiciel, types d'interrogations les plus fréquentes) afin d'optimiser la gestion des données et leur exploitation.

## **2. DE LA NATURE DES PROBLÉMATIQUES AU MODÈLE CONCEPTUEL DES DONNÉES**

La nature des données récoltées en milieu fluvial, dans l'optique de l'analyse des structures spatio-temporelles des communautés piscicoles, renvoie le plus souvent à une structure générale de BD (CHESSEL et AUDA, 1988).

Les structures faunistiques (tableaux relevés-espèces) et les structures du milieu (définissant l'habitat) doivent pouvoir être analysées conjointement. Elles peuvent être approfondies par l'étude des structures en classes d'âge ou en classes de taille au niveau des populations. L'habitat peut être appréhendé à différentes échelles (bassin versant, fleuve, tronçon, secteur, station, point de pêche) et les processus étudiés selon un pas de temps variable (échelle de la décennie, de l'année, de la saison, de la journée). La précision de ces différents paramètres dépend des échelles privilégiées et de la nature des données disponibles.

Ainsi, compte tenu du grand nombre de fleuves pris en considération (français et tropicaux) et de la multiplicité des situations rencontrées, les modèles conceptuels des BD ont été élaborés au niveau de chaque bassin ou de chaque secteur d'intervention.

En France, le Rhône (CHANGEUX, 1994 ; CHANGEUX et ZYLBERBLAT, 1993a et b ; POIZAT, 1993) et la Seine (BELLIARD, 1994) ont été particulièrement bien étudiés. Suivant les problématiques et les données disponibles, plusieurs échelles spatiales imbriquées ont été considérées : le bassin versant, l'axe fluvial, le secteur, la station et le point de pêche.

A titre d'exemple, nous détaillerons la démarche adoptée à propos des données recueillies sur le Rhône par la méthode d'échantillonnage ponctuel d'abondance (E.P.A.) (NELVA *et al.*, 1979 ; PERSAT, 1988). Le point d'échantillonnage est un micro-échantillon défini dans l'espace par la zone d'attraction d'une électrode immergée et constitue l'unité spatio-temporelle de base ; l'unité de rang géographique immédiatement supérieur est la station (regroupant un ensemble de points réalisés dans un bief offrant une physionomie relativement homogène). La pêche est l'entité définie par l'ensemble des points d'échantillonnage effectués dans une station donnée à une date donnée. Les stations sont regroupées en secteurs, le plus souvent apparentés à l'espace affecté par un aménagement hydro-électrique, ou à un lieu géographique particulier (par exemple le secteur d'Arles, situé à l'aval du dernier barrage implanté sur le Rhône). Au niveau de chaque point, pour chaque individu capturé sont enregistrées l'espèce à laquelle il appartient, la longueur totale et dans certains cas la masse. Parallèlement aux données biologiques, les caractéristiques de l'habitat sont enregistrées sous forme de variables qualitatives. Certaines décrivent des paramètres relatifs aux conditions de pêche, d'autres l'habitat au niveau du point d'échantillonnage. Le modèle conceptuel des données correspondant est présenté (Figure 1) sous la forme du modèle entité-association (CHEN, 1976).

Les cours d'eau tropicaux pris en compte dans le P.I.R. regroupent six fleuves (ou groupes de rivières) d'Afrique de l'Ouest et d'Amazonie : le bassin du Bandama (Côte-d'Ivoire), le Niandan (Guinée), des rivières forestières de Guinée, le Tocantins (affluent de l'Amazone), le Careiro (plaine inondée d'Amazonie centrale) et le Sinnamary (Guyane). Parmi eux, trois ont été intégrés à la BD générale mise au point à l'ORSTOM : les rivières forestières de Guinée, le Careiro et le Sinnamary. Cette BD, actuellement opérationnelle mais incomplète, permettra à terme une étude comparative des données recueillies sur différents continents. Le Tocantins a fait l'objet d'une BD particulière.

Le tableau I présente l'ensemble des caractéristiques générales de chacune des BD en définissant les objectifs scientifiques poursuivis lors de leur élaboration.

### 3. DE L'IMPLANTATION PHYSIQUE DE LA BD À SON UTILISATION

Indépendamment du modèle conceptuel, diverses solutions techniques matérielles et logicielles peuvent être utilisées pour l'implantation physique des BD :

— matériel : de simples micro-ordinateurs individuels suffisent :

- MacIntosh pour les BD de cours d'eau français et la BD Tocantins de l'ORSTOM,
- PC en général pour les BD ORSTOM.

— logiciel : la gestion des BD est assurée par divers SGBD relationnels commerciaux :

- ORACLE® (ORACLE Corporation) est utilisé pour les BD universitaires et du Tocantins,
- 4<sup>ème</sup> Dimension (ACI) pour la base Seine,
- FoxPro (Microsoft) pour les bases ORSTOM (sauf Tocantins).

Ces logiciels fournissent les outils permettant de définir les données, de les stocker, de les mettre à jour, de les interroger, de les sauvegarder, et ils assurent l'intégrité et la confidentialité de ces données.

A partir du modèle conceptuel des données est élaboré le modèle logique des données conduisant à une implantation des données adaptée aux types d'interrogations les plus fréquents auxquels seront soumises ces données (FLORY, 1987). Cette démarche est illustrée Figure 2.

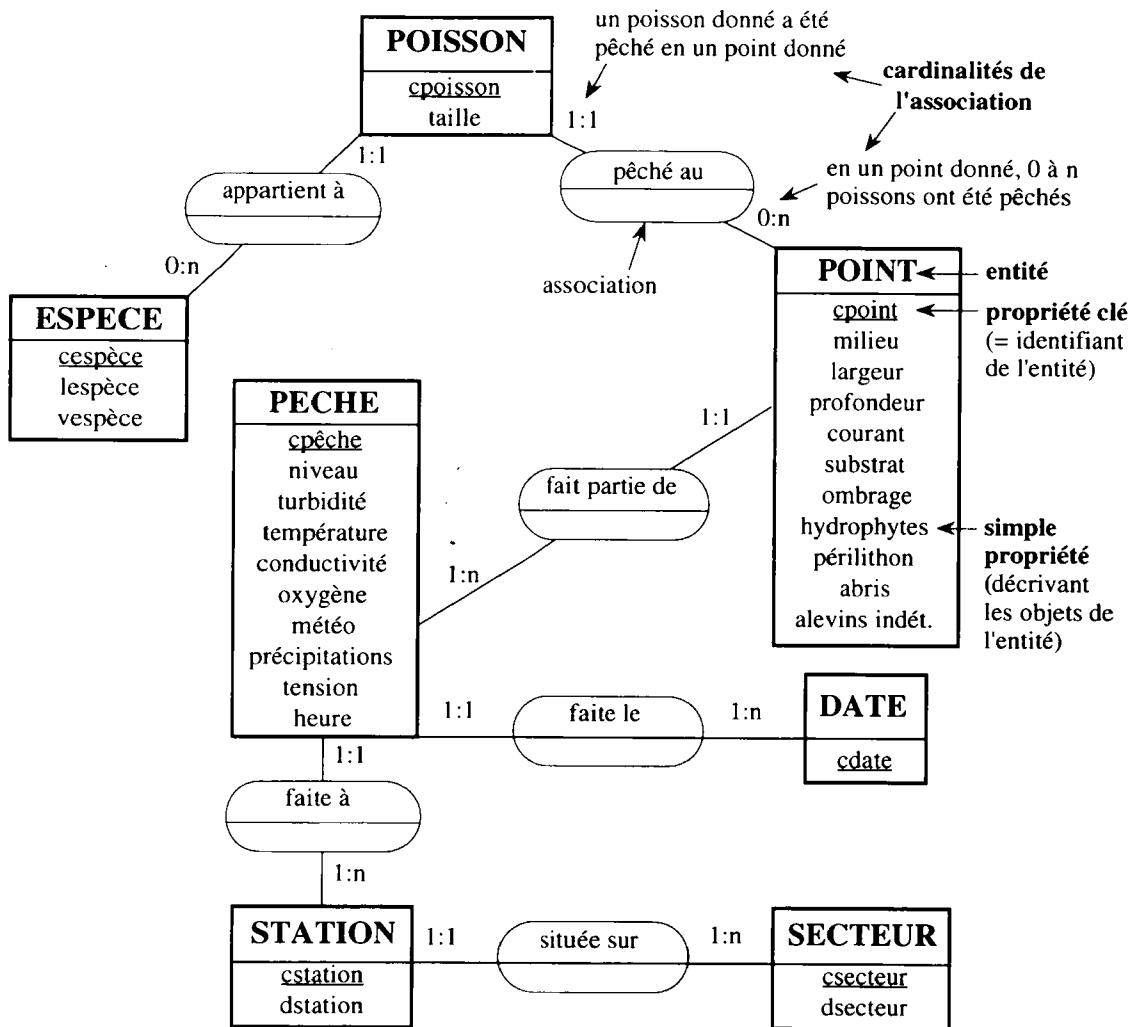


Figure 1 : Modèle conceptuel (entité-association) des données recueillies sur le Haut-Rhône français, modèle construit sur la base des paramètres biologiques et environnementaux déterminés par NELVA et al. (1979).

Figure 1 : Entity/relationship model of the data from the French Upper Rhone, elaborated with the biological and environmental parameters defined by NELVA et al. (1979).

**Tableau I : Présentation des différentes bases de données élaborées au sein du P.I.R.. Légendes des codes utilisés :** ⌘ : indications piscicoles sous forme de pêches électriques (71,9 % des tronçons), de résultats d'enquêtes auprès des pêcheurs (6,8 %) ou de pêches aux filets (0,7 %), \* : données recueillies par pêche électrique, ‡ : pêches réalisées aux filets maillants et/ou au poison, ◊ : résultats des pêches aux engins (pour plus de détails voir CHANGEUX et ZYLBERBLAT (1993a et b)), (1) le tronçon représente l'unité de résolution spatiale adoptée dans les Schémas Départementaux de Vocation Piscicole (SOUCHON et TROCHERIE, 1990), (2) l'opération de pêche est définie comme étant une pêche réalisée en une station à une date donnée et utilisant une méthode donnée (filets maillants, poison...).

**Table I : The different data bases developed in the P.I.R. Interdisciplinary Research Program. Key for the symbols used:** ⌘: information from electric fishing (71,9 % of the selected river stretches), from angler interviews (6,8 %) or from gillnet fishing (0,7 %), \* : information from electric fishing, ‡ : gillnet fishing and/or poisoning, ◊ : information from fishing gear fishery statistics (for more details see CHANGEUX and ZYLBERBLAT (1993a, and b)), (1) this is the geographical unit used in the «Schémas Départementaux de Vocation Piscicole» (SOUCHON et TROCHERIE, 1990), (2) a fishing operation is defined as a fishing excursion performed in a given station at a given date and using a given sampling method (gillnets, poison...).

Problématique	Unité géographique BD	Durée	Nb pêches	Nb unités d'échant.	Nb Poissons	Données environnementales
Structuration des peuplements à l'échelle du bassin (approche multi-échelles) Comparaisons inter-continents	Bassin du Rhône⌘	25 ans	1284	1787	-	Tronçon (1) Station
	Bassin de la Seine*		718	718	6491	
	Rivières de Guinée‡	3 ans	-	-	38277	Opération de pêche (2) Opération de pêche
	Careiro‡	48	1792	38276		
	Synamary‡		17518	356	17518	Opération de pêche
Analyse des statistiques de pêche aux engins : estimation des prélèvements et des stocks de certaines espèces	Bassin du Rhône◊		1677 carnets soit 32466 jours de pêche 430 t de poissons			-
Evolution des structures et dynamique des peuplements et des populations à l'échelle du fleuve	Fleuve Rhône (France) 5 secteurs 16 sous-secteurs*	10 ans	855	855	91523	Sous-secteur
Structuration et dynamique des peuplements à l'échelle du secteur, de la station, étude des profils écologiques, étude multi-échelles	Rhône					
	Haut-Rhône*	16 ans	404	9375	63235	Station et point d'échantillonnage
	Péage-de-Roussillon*	12 ans	222	4463	34850	
	Arles*	3 ans	82	1225	5943	

Autour du noyau de base du SGBD peut être développée une application personnalisée permettant, entre autres :

**\* Le contrôle sémantique des données avant leur chargement dans la BD**

En effet, la saisie informatique est une étape ingrate et souvent source d'erreurs qu'il faut minimiser :

— soit au cours de la saisie, grâce à l'utilisation de masques de saisie, créés par exemple sous 4ème dimension, FoxPro ou la version PC d'ORACLE®. Ces écrans, mimétiques des fiches de terrain, permettent des contrôles (type et format des données, appartenance de la donnée tapée à une liste de valeurs possibles), et fournissent une aide «écran» précieuse (correspondance entre un code et son libellé).

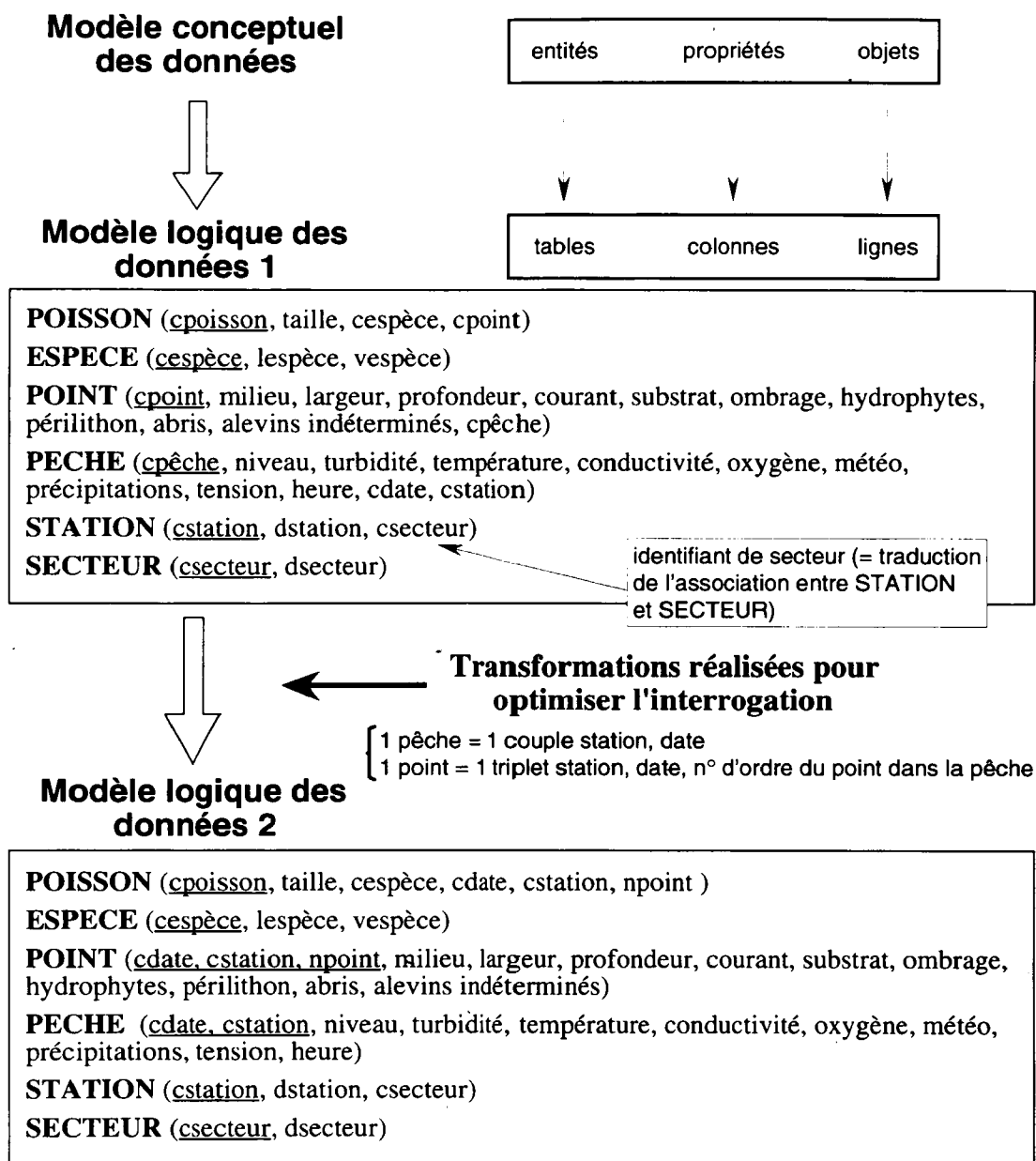


Figure 2 : Modèles logiques des données correspondant au modèle conceptuel des données de la Figure 1.

Figure 2 : Logical data base design (MLD) of the data corresponding to the entity/relationship model presented in Figure 1.

— soit entre la saisie des données — sous forme de fichiers textes — et leur chargement, grâce à un programme «filtre» de relecture de ces fichiers (par exemple, dans l'interface ADE-ORACLE® sur MacIntosh). Les contrôles sont du même type que précédemment, mais l'ergonomie est moins bonne.

#### \* L'interrogation et le traitement des données

Chaque SGBD offre des moyens conviviaux d'interrogation pour les requêtes simples portant sur une ou plusieurs tables : choix de certaines colonnes, utilisation de fonctions de calculs (ex. : sommes, produits...) ou de fonction de groupes (ex. : totaux ou moyennes par pêche sur tous les points de chaque pêche...), sélection des données à éditer selon un ou plusieurs critères, édition selon un ordre donné. L'utilisateur n'a pas besoin d'écrire la requête (génération automatique par le SGBD du texte de la requête) et devient très vite opérationnel.

Mais les progrès récents de la statistique écologique et l'évolution conjointe des problématiques de l'écologie moderne nécessitent la mise en oeuvre d'interrogations plus sophistiquées (ex. : extraction des effectifs par classe de taille et par espèce avec sélection des types de milieu et choix des modalités de variables de point), pour laquelle deux stratégies sont possibles :

— utilisation de requêtes préprogrammées par un informaticien

C'est le cas de la BD ORSTOM sous FoxPro, où l'utilisateur a à sa disposition une dizaine de programmes spécifiques permettant par exemple d'obtenir le tableau des taux de remplissage stomacal par tranche horaire, pour chaque espèce sélectionnée par l'utilisateur, ou le calcul du coefficient de condition K (à partir du b de la relation taille - poids ( $K=10^5 P/L^b$ )), pour une sélection d'individus donnés, etc... Ces programmes sont paramétrés pour certains (l'utilisateur indique au moment de l'exécution la valeur prise par le paramètre (ex. : l'espèce choisie)).

C'est évidemment la solution la plus simple pour l'utilisateur, mais aussi la plus limitante, car le développement de tels programmes nécessite de bonnes connaissances en informatique (langages d'interrogation des données et de programmation). Or, au cours de l'analyse des données, de nouvelles questions voient le jour, donc de nouvelles requêtes ; la collaboration avec un informaticien demeure donc indispensable.

— écriture, par l'utilisateur, de requêtes dans un environnement où l'informaticien a automatisé et rendu transparent un certain nombre d'opérations (gestion de la localisation du dossier de travail, redirection...). Il suffit ainsi à l'utilisateur d'apprendre le langage d'interrogation pour devenir entièrement autonome.

Cette dernière solution a été retenue pour l'interface entre les logiciels ADE (CHESSEL et DOLÉDEC, 1994) et ORACLE®. Cette interface, développée sous HyperCard™, permet à l'utilisateur de réaliser tout son traitement dans le même environnement, depuis la gestion de la BD jusqu'à l'interrogation et l'analyse des données (SANLAVILLE-BOISSON, 1993). Il peut stocker dans cette interface :

- le texte des requêtes (en SQL = Structured Query Language, langage standard de 4<sup>ème</sup> génération),
- de la documentation sur la BD : liste des objets (tables, vues, index) avec leur description, ou résultats synthétiques (ex. : pour chaque espèce répertoriée, l'effectif total, le nombre de points où cette espèce a été trouvée, ses longueurs mini et maxi...),
- des programmes spécifiques nécessaires à la mise en forme des données extraites (ex. : en vue de l'obtention d'une table de contingence).

#### 4. CONCLUSION

L'amélioration récente et rapide de l'efficacité et de l'accessibilité de l'outil informatique (matériel et logiciel, notamment SGBD) mis à la disposition de l'écologue associée à une

interdisciplinarité (informatique et biologie) a entraîné un changement méthodologique radical. En simplifiant et en accélérant la gestion de l'information écologique, cette évolution technique a ouvert de nouveaux champs d'investigations en biologie.

La souplesse des SGBD confère aux BD une capacité d'évolution permanente. Il est possible à tout moment de modifier le modèle conceptuel des données en ajoutant, par exemple des entités ou des propriétés supplémentaires et les échanges de données peuvent être réalisés, même entre SGBD différents.

Ainsi, quelles que soient les caractéristiques du jeu de données, il apparaît toujours préférable de structurer ces informations en BD, plutôt que d'utiliser un tableur, même pour des volumes de données réduits. Il est toujours possible de structurer des données en BD, même si l'échantillonnage n'a pas été organisé dans cette optique ; les BD portant sur les bassins du Rhône et de la Seine en sont un exemple.

En raison des performances des SGBD, l'utilisation des BD devrait connaître un essor important en écologie.

## 5. BIBLIOGRAPHIE

- ANSI (American National Standard for Information Systems), 1975. Interim report ANSI/X3/SPARC study group on data base management systems. *ACM SIGMOD Newsletter*, 7 (2).
- BELLIARD J., 1994. Le peuplement ichtyologique du bassin de la Seine : rôle et signification des échelles temporelles et spatiales. Thèse de doctorat, Université Paris 6, 197 p.
- CHANGEUX T., 1994. Structure des peuplements de poissons à l'échelle du bassin rhodanien. Approche régionale et organisation longitudinale, exploitation des captures par pêche aux engins. Thèse de doctorat, Université Lyon 1, 241 p.
- CHANGEUX T., ZYLBERBLAT M., 1993a. Analyse des statistiques de pêche aux engins dans le bassin du Rhône. Première partie : étude de l'effort de pêche. *Bull. Fr. Pêche Piscic.*, 330, 245-269.
- CHANGEUX T., ZYLBERBLAT M., 1993b. Analyse des statistiques de pêche aux engins dans le bassin du Rhône. Deuxième partie : étude des captures. *Bull. Fr. Pêche Piscic.*, 330, 271-294.
- CHEN P.P., 1976. The Entity-Relationship Model : towards a unified view of data. *ACM Trans. on Database Systems*, 1 (1), 9-36.
- CHEssel D., AUDA Y., 1988. Approche des besoins en analyse des données hydrobiologiques. In DEMONGEOT J., MALGRANGE P., Biologie et économie. Les apports de la modélisation, 155-169, Collection de l'IME, n° 34 Dijon.
- CHEssel D., DOLÉDEC S., 1994. Progammathèque ADE. Analyses multivariées et expression graphique des données environnementales (version 3.7). 9 volumes, Université Lyon 1, 871 p.
- DETENBECK N. E., DE VORE P. W., NIEMI G., LIMA A., 1992. Recovery of temperate-stream fish communities from disturbance : a review of case studies and synthesis of theory. *Environ. Manage.*, 16 (1), 33-53.
- FLORY A., 1987. Bases de données. Conception et réalisation, Economica, Paris, 164 p.
- FRANKLIN J.F., 1989. Importance and justification of long-term studies in ecology. In LIKENS G.E., Long-Term Studies in Ecology. Approaches and alternatives, 3-19, Springer-Verlag, New York.
- NELVA A., PERSAT H., CHEssel D., 1979. Une nouvelle méthode d'étude des peuplements ichtyologiques dans les grands cours d'eau par échantillonnage ponctuel d'abondance. *C. R. Acad. Scie. Paris Série D*, 289, 1295-1298.



- PERSAT H., 1988. De la biologie des populations de l'ombre commun *Thymallus thymallus* (L. 1758) à la dynamique des communautés dans un hydrosystème fluvial aménagé, le Haut-Rhône français. Eléments pour un changement d'échelles. Thèse de doctorat d'Etat, Université Lyon 1, 223 p.
- PERSAT H., 1991. Espace et temps, pierres d'achoppement des études d'impact en milieu fluvial. *Bull. Ecol.*, 22 (1), 203-212.
- POIZAT G., 1993. Echelles d'observation et variabilité des abondances de juvéniles de poissons dans un secteur aval du Rhône. Thèse de doctorat, Université Lyon 1, 217 p.
- ROUSSEAU B., NELVA A., PERSAT H., CHESSEL D., 1985. Constitution d'une base de données ichtyologiques par l'échantillonnage ponctuel d'abondance : application aux peuplements du Haut-Rhône français. *Cybium*, 9 (2), 157-173.
- SANLAVILLE-BOISSON C., 1993. Interface ADE-ORACLE®. Version 1.0. Université Lyon 1, 38 p.
- SCHLOSSER I. J., 1990. Environmental variation, life history attributes, and community structure in stream fishes : implications for environmental management and assessment. *Environ. Manage.*, 14 (5), 621-628.
- SOUCHON Y, TROCHERIE F., 1990. Technical aspects of French legislation dealing with the freshwater fishery (june 1984) : 'fisheries orientation schemes' and 'Fishery resources management plans : In VAN DENSEN W.L.T., STEINMETZ B. and HUGHES R.H., Management of freshwater fisheries (Proceedings of the EIFAC symposium, Göteborg, Sweden), 190-214, Wageningen.
- WERNER E.E., GUILLIAM J.F., 1984. The ontogenic niche and species interactions in size-structured populations. *Annu. Rev. Ecol. Syst.*, 15, 393-425.
- YOUNT J.D., NIEMI G., 1990. Recovery of lotic communities and ecosystems from disturbance - A narrative review of case studies. *Environ. Manage.*, 14 (5), 547-569.