

Contrôle de la qualité des données

Application à un observatoire socio-économique spatialisé

Michel Passouant

Dans le cadre d'une coopération avec le Cirad ¹, et sur financement du ministère français de la Coopération, la SAED ² a mis en place un système de suivi-évaluation afin de collecter, centraliser, capitaliser, redistribuer, des informations statistiques de suivi du développement des périmètres irrigués dans le delta du fleuve Sénégal.

Les données s'accumulent dans le système depuis bientôt 10 ans, la question de la qualité de l'information se pose. Nous allons présenter ci-après une méthode utilisée pour réaliser une première évaluation de cette qualité.

Plan de l'exposé

L'exposé va aborder successivement :

- ◆ La problématique va présenter d'une part le système d'information et d'autre part les questions posées en terme de qualité des données.
- ◆ L'approche qualité à proprement parler va être abordée selon deux angles, d'une part la méthode adoptée, fondée sur un essai de structuration de la démarche, et d'autre part une présentation de sa mise en œuvre et de l'instrumentation utilisée.
- ◆ En conclusion, en prenant un peu de recul vis-à-vis de l'objectif de qualité nous ouvrirons vers un besoin de structuration des informations descriptives des données utilisées pour cette démarche qualité et donc vers le concept de méta données.

Problématique

Le système de suivi-évaluation

Pour assurer sa mission de planification du développement régional et de coordination du développement rural intégré, la SAED doit disposer d'informations régulières, fiables et représentatives des activités socio-économiques dans la vallée.

Dans un contexte de fort désengagement, elle a donc réalisé, avec l'appui du CIRAD, la mise en place d'un système d'information dédié au suivi-évaluation du développement de l'agriculture irriguée dans la vallée du fleuve Sénégal.

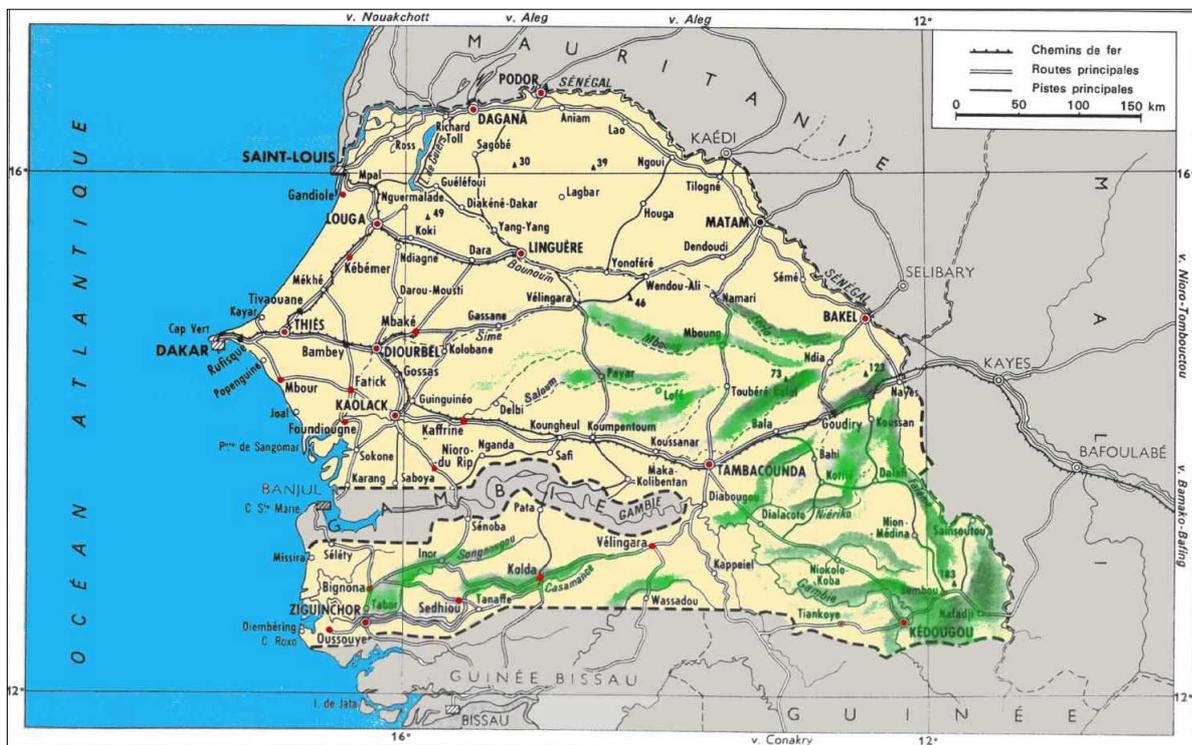
Le souci de maîtriser l'information est à l'origine de cette démarche organisée en deux grandes composantes :

- ◆ Une base de données spatialisées de description et de suivi exhaustifs des aménagements hydro-agricoles selon les cinq objectifs suivants : Collecter, gérer et conserver l'information, analyser et présenter les données.

- ◆ Des enquêtes représentatives auprès des principaux acteurs économiques (suivis d'exploitations, de rizeries, sondages de rendement...).

La base de données spatialisées, qui a fait l'objet de la démarche présentée ici, se décline en deux éléments distincts tant en terme de gestion opérationnelle que d'outils informatiques :

- ◆ Une base de données de description et de suivi annuels de la mise en valeur des périmètres irrigués, mise en place et opérationnelle depuis 1991, sans interruption. La zone couverte concerne l'ensemble de la zone d'intervention de la SAED.
- ◆ Et une représentation spatialisée des entités de la base de données ci-dessus au sein d'un Système d'Information Géographique (SIG) installé en 1996 et reprenant rétroactivement des données annuelles depuis 1993. La zone couverte concernait initialement la seule délégation de Dagana et est en cours d'extension à la délégation de Podor.



La structure du système d'information

Le système d'information s'appuie sur une organisation tripolaire des données descriptives et de suivi : les unités sociales (regroupant les divers acteurs du développement), les unités géographiques spatialisées (essentiellement à caractère hydraulique) et le croisement des deux types précédents avec les parcelles³ attribuées aux acteurs.

Ce dernier type d'entité a fait l'objet d'une définition particulière sous la dénomination d'unités de mise en valeur ou UMV.

Pour ces entités, deux types d'observations sont possibles :

- ◆ des descriptions de caractères permanents, regroupées sous des entités dites signalétiques ;
- ◆ des suivis à des pas de temps soit annuel, soit saisonnier, avec deux calendriers l'un fondé sur l'année civile et l'autre sur l'année agricole.

Parmi ces entités, deux sont spatialisées et sont dotées d'une représentation géographique : les aménagements et leur subdivision en parcelles attribuées ou unités de mise en valeur (UMV). Le modèle de données a fait l'objet au cours des années d'adaptations, mais sans remise en

cause son architecture de base présentée ci-dessous. Ces évolutions ont concerné soit les attributs des entités existantes (ajout ou suppression), soit l'introduction de nouvelles entités soit explicitement au sein du même modèle, soit par le développement de modèles connexes thématiquement spécialisés, comme les suivis de rizeries par exemple. Notre analyse ne porte ici que sur le modèle principal et dans sa version actuelle.

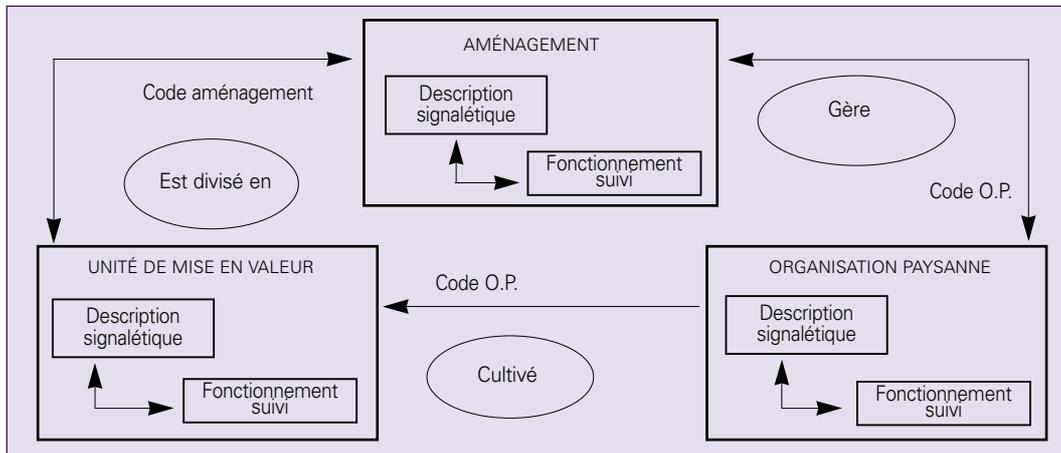


Schéma tripartite des données.

Acquisition annuelle des données

Le système a vocation d'exhaustivité aussi bien en terme de recensement des unités d'observation, que du suivi de leur fonctionnement. Au niveau de la collecte annuelle cet objectif est atteint selon qu'il s'agit :

- ◆ de données de signalétique, par l'enregistrement, chaque année, seulement des modifications d'état ou des apparitions ou disparitions d'unités. On travaille en « différentiel » ;
- ◆ de données de suivi, par la saisie des données décrivant le fonctionnement de l'ensemble des unités actives. On travaille ici sur la totalité de la population.

Sur la base d'un cycle annuel d'acquisition des données, deux méthodes de collecte de données sont utilisées pour alimenter le système :

- ◆ En s'appuyant sur son dispositif d'encadrement et de conseil agricole, la SAED réalise chaque année, depuis 1990, des enquêtes exhaustives pour :
 - le recensement des acteurs : sans descendre au niveau de l'agriculteur individuel, l'ensemble des acteurs organisés en groupes structurés et intervenant dans l'agriculture irriguée dans la vallée est identifié et se voit affecter un code identifiant. À ce jour quelque 3 450 acteurs de l'ensemble de la vallée sont présents dans la base de données ;
 - l'identification des périmètres irrigués : tous les périmètres sont identifiés par un code unique et décrits par un certain nombre de variables. La base comporte actuellement 2 450 aménagements, réunissant là aussi bien périmètres à proprement parler qu'adducteurs ;
 - recensement des attributions des terres et suivi de leur mise en culture : c'est le domaine qui requiert la plus lourde charge d'acquisition de données, tant en ce qui concerne l'attribution de terres que surtout leurs mises en culture annuelles. La table d'attribution de terres comporte 9 160 enregistrements alors que celle de suivi de mise en culture en comprend 19 650.

En complément du dispositif de terrain sur lequel s'appuie le processus de collecte ci-dessus, la SAED a mis en place un suivi par télédétection dans un premier temps du seul delta, puis une extension à la délégation de Podor. Il s'agit d'une procédure d'acquisition systématique d'images SPOT en période d'hivernage (octobre). À partir de ces images traitées, l'équipe Télédétection réalise, par contrôles terrain, : identification, localisation et dessin des

périmètres irrigués et des parcelles attribuées. Les données depuis l'année 1993 sont collectées par ce procédé qui utilise les mêmes codifications que la base de données.

Moyens mis en œuvre

Les principes de base qui ont guidé les choix technologiques pour la conduite de ce projet résultent de la prise en compte des moyens mobilisables par la DPDR (direction de la planification et du développement rural à la SAED) qui a la responsabilité de l'opération. C'est dans l'explicitation de ces contraintes que l'on trouvera une partie des raisons de la démarche de qualité des données engagées ici : le refus de recourir à des moyens à technologie plus sûre, mais trop contraignante a pu engendrer des défauts dans la qualité des données.

Ces contraintes peuvent sembler extrêmement pénalisantes sur le plan de la qualité des données. Cependant c'est leur acceptation qui est à l'origine du succès de l'opération : le choix d'outils à technologie robuste, le respect de contraintes institutionnelles et la maîtrise de l'outil par les utilisateurs constituent des raisons de la pérennité du système. Depuis 1991, les données sont accumulées, sans interruption, avec une production régulière d'états et de cartes statistiques.

Moyens techniques

Tout d'abord, il faut resituer le projet dans le temps, son début se place en 1990. À cette époque les seuls ordinateurs disponibles étaient des PC de type 286, fonctionnant sous MS/DOS. Si au cours des années certains services se sont vus dotés de matériels plus puissants, suivant en cela les évolutions technologiques qu'a connus le monde de la micro-informatique, ces ordinateurs de 1990 sont restés en exploitation jusqu'à très récemment.

Une forte contrainte a pesé sur les moyens matériels informatiques. Corrélativement à ces dotations en matériel, il a été choisi des logiciels supportés par ces machines et donc de classe « bureautique évoluée ». Cela s'est traduit par le fait que l'on est resté très longtemps avec une application construite à partir de la version 1.0 de Foxpro, et que la composante géographique est exploitée sous Mapinfo. Le passage sous Access 97 vient d'être effectué.

L'utilisation de tels systèmes est à l'origine d'une partie des problèmes. Avec un grand nombre de degrés de liberté pour résoudre des situations exceptionnelles ils autorisent le contournement des contrôles de saisie (validité et intégrité) en tolérant des accès directs aux tables de données.

Moyens organisationnels

La deuxième contrainte à prendre en compte après les moyens techniques informatiques concerne l'organisation que la SAED a souhaité adopter. En relation avec l'implantation sur le terrain, il a été institué une totale décentralisation de la gestion de sa base de données auprès de ses structures opérationnelles que constituent les quatre délégations réparties au niveau des quatre départements de la vallée du Sénégal (Dagana, Podor, Matam, Bakel). Cette option de rapprocher au maximum le stockage et la gestion des données de leur lieu d'acquisition et de validité a favorisé une redistribution locale prolifique des informations collectées. Cependant, en revers de la médaille, cela a pour conséquence de diluer sur cinq sites géographiques les compétences et la maîtrise de l'outil.

Par contre, en raison des moyens matériels et logiciels nécessaires à l'activité de télédétection, celle-ci a été confiée à un service dépendant directement du siège. Ceci a pour conséquence le partage des données par deux services, et même si des procédures et des règles rigoureuses définissant droits et devoirs quant aux accès en consultation et modification des dérivés ont pu entraîner des lacunes dans la qualité des données.

Moyens humains

La dernière contrainte concerne les hommes chargés de la collecte, de la gestion et de la diffusion des données. On a déjà parlé de leur isolement au sein des délégations, mais en plus de cela, leur grande mobilité a une conséquence immédiate à la fois sur leur connaissance du terrain et sur leur maîtrise des outils de suivi dont ils disposent. Cet état de fait a nécessité de nombreux efforts d'animation et de formation pour assurer la permanence et l'efficacité du protocole de collecte.

Enfin, la faiblesse des effectifs des équipes n'a pas permis, jusqu'en 1998, la présence de spécialistes informaticiens.

Qualité

Définition Afnor

Avant d'engager une analyse et une évaluation de qualité des données contenues dans la base, il faut être clair sur le sens donné à ce terme. Nous reprendrons donc une définition de l'Afnor (provenant de la norme ISO 8402) qui définit la qualité comme "*l'ensemble des propriétés et caractéristiques d'un produit ou service qui lui confèrent l'aptitude à satisfaire des besoins exprimés ou implicites.*"

Cette définition, bien que très générale, va nous servir de guide dans la construction de la méthode d'évaluation de cette qualité de la base de données.

Démarche ex post

L'analyse de la qualité de l'information est engagée par un gestionnaire de base de données dont les objectifs immédiats concernent :

- ◆ L'évaluation, l'identification des défauts afin d'engager une procédure de correction fondée soit sur une analyse croisée de sources concurrentes, soit par un retour au relevé de terrain.
- ◆ Prévenir les nouvelles erreurs, à partir d'une mise en évidence et d'une analyse des causes des défauts constatés.

Cette démarche est engagée pour satisfaire des utilisateurs du système d'information, qu'il s'agisse des services directs de la SAED, ses tutelles ou tout utilisateur extérieur, qui attendent :

- ◆ Evaluation de la pertinence de l'information qui leur est proposée.
- ◆ Mesure de la validité des données qu'ils utilisent.

Donc une double motivation à la démarche menée sur des données déjà collectées et guidée d'abord par un objectif de fidélité de représentation d'une réalité et ensuite pour répondre à des attentes d'utilisateurs du système, complètement étrangers au processus de collecte, voire extérieurs à la société.

Différentes formes de qualité

Pour approcher cette qualité de l'information, il est nécessaire de pousser plus avant la définition Afnor ci-dessus. À cet effet, et pour mener notre analyse, il a été choisi deux approches complémentaires de la qualité :

- ◆ La première, qui est la plus facile à mettre en œuvre, est qualifiée de « qualité syntaxique » et est relative à l'aspect formel du système. On parlera aussi de respect de l'intégrité de la base. Reprenant la dichotomie classique en SIG entre données attributaires et géographiques, et ce d'autant plus que ces deux types de données sont gérés par deux applications informatiques différentes, on distingue donc :
 - pour les données attributaires, respect du modèle de données, matérialisé par la structure relationnelle de la base mais que le logiciel Foxpro utilisé jusqu'il y a peu ne contrôlait que partiellement ;
 - pour les données géographiques, respect des contraintes spatiales qui se caractérise au sein d'une topologie, mais que le logiciel utilisé, Mapinfo, ne sait pas prendre en compte.
- ◆ La seconde approche concerne la « *qualité sémantique* », c'est-à-dire le sens de l'information et donc son aptitude à restituer une image exacte du monde réel qu'elle est censée décrire. Là encore, on peut distinguer deux types d'entrées :
 - l'exhaustivité, c'est-à-dire la capacité à décrire l'ensemble des éléments du domaine. Dans le cas qui nous préoccupe, on peut envisager des questions du type : tous les périmètres irrigués ont-ils bien été identifiés et codifiés dans la base ;
 - l'exactitude concerne la capacité à décrire exactement ou fidèlement les éléments du domaine : elle concerne donc directement les données enregistrées, qu'elles que soit leur nature (numérique, date, caractères...). Ainsi par exemple la superficie enregistrée pour une parcelle est-elle conforme à la réalité de terrain ?

L'intégrité référentielle

La première approche a donc concerné l'évaluation de l'intégrité des données de la base, en comparant les données réelles aux spécifications prévues dans le modèle conceptuel.

Modèle relationnel

Trois types de contraintes inscrites dans le modèle relationnel, construit sur la base de la méthode entité/association, sont vérifiés et contrôlés :

- ◆ **Domaine de valeurs** : les données enregistrées ont-elles la forme prévue ? Pour des variables qualitatives cela veut dire n'utiliser que les codifications prévues dans les nomenclatures, pour les variables quantitatives et les dates, respecter les plages de valeurs acceptables définies, et pour les variables littérales respecter les contraintes sur les structures de ces chaînes de caractères.
- ◆ **Intégrité existentielle** : les champs identifiants sont déclarés dans le modèle, et sur cette base, dans une table donnée il ne peut y avoir de doublons.
- ◆ **Intégrité référentielle** : l'approche conceptuelle définit des relations entre les entités. Ceci se traduit par la nécessité pour des enregistrements de tables de données de trouver des correspondants dans d'autres tables, et le respect de ces règles de correspondance est une exigence de base du modèle. À ce niveau il faut vérifier aussi les liens entre la base de données attributaires à proprement parler et les couvertures géographiques.

Normalement, ces règles d'intégrité sont vérifiées à la saisie des données. Cependant, pour un ensemble de raisons développées ci-avant, ces contrôles ont pu ne pas avoir été réalisés, et de ce fait une série d'analyses a posteriori a été mise en place pour détecter les erreurs et en mesurer l'ampleur.

Méthode de contrôle

Compte tenu du volume de données accumulé d'une part, et de la complexité du modèle conceptuel d'autre part, il ne peut être envisagé de contrôler les données par un examen individuel, par pointage, de tous les enregistrements de toutes les tables.

Les contraintes de valeurs ont été contrôlées avec des outils de statistiques simples, qui ont permis, variable par variable, de mettre en évidence :

- ◆ liste de valeurs pour les variables qualitatives directement obtenues par construction de tableaux de fréquences ;
- ◆ intervalles de valeurs prises, directement issus de calculs de paramètres statistiques de tendance centrale et de dispersion, avec en complément distribution statistique ;
- ◆ pour les variables caractères, extraction de sous-chaînes, statistiques et analyses des occurrences de ces variables élaborées.

Les recherches de doublons, c'est-à-dire identifiants dupliqués, sont proposées en standard dans les logiciels modernes de gestion de base de données. Il a donc suffi de réaliser des importations de tables et de lancer les recherches. Cela a fait apparaître un taux de doublons moyen sur l'ensemble des tables de la base inférieur à 0.3 %.

Le contrôle des relations a fait l'objet du développement d'un utilitaire spécifique à l'application et qui a permis le contrôle du respect de l'intégrité relationnelle : chaque enregistrement trouve-t-il bien les divers correspondants qui lui sont associés par le modèle de données ?

Les erreurs, limitées en nombre, mises en évidence provenaient essentiellement de recodifications effectuées sur certains aménagements sans avoir été répercutées sur les suivis de parcelles.

Modèle spatial

Cette partie concerne la composante géographique du système d'information. Il s'agit ici de vérifier :

- ◆ Contrôle des propriétés géométriques de l'objet en regard de ses propriétés structurelles : les périmètres irrigués doivent avoir une représentation surfacique, les pompes sont matérialisées par un point... Ces contrôles sont effectués en exploitant des fonctions géométriques du logiciel SIG.
- ◆ Contrôles topologiques : vérifier les contraintes de non-recouvrement de polygones entre eux, de relation d'agrégation entre objets appartenant à deux niveaux emboîtés, entre les périmètres et les parcelles attribuées par exemple. Là encore, des fonctionnalités standards, ou des outils spécialisés, du logiciel SIG sont utilisés.
- ◆ Correspondances entre les occurrences base de données et les occurrences SIG : vérifier que la relation entre données attributaires et données géographiques est bien respectée.

Le manque d'outils de contrôles topologiques n'a pas permis de mener à bien ces analyses.

Sémantique

Contrairement à la démarche précédente qui relevait de la conformité des données à un modèle clairement et indiscutablement défini, l'analyse de la qualité sémantique est beaucoup plus difficile à mettre en œuvre. En effet il est très ardu d'évaluer la qualité d'une représentation d'un phénomène connu par cette seule représentation.

Face à cette difficulté, on ne peut que s'engager vers des méthodes faisant appel à des :

- ◆ Recoupement de sources concurrentes ou complémentaires. Dans notre cas précis enquêtes au sol et imagerie spatiale ont été utilisées conjointement.
- ◆ Règles théoriques issues de modèles sémantiques, par opposition aux modèles syntaxiques ci-dessus, et valables au niveau de chaque enregistrement :
 - déterministes, traduisant ainsi une relation stricte entre plusieurs variables : la superficie récoltée ne peut être supérieure à la superficie semée par exemple ;
 - statistiques, dans lesquelles la relation fait intervenir une partie déterministe, fixe, et une partie aléatoire sous forme d'un écart résiduel à la norme : relation entre la superficie géographique (mesurée par SIG) et la SAU par exemple ;

- logiques, traduisant ainsi une relation logique entre propriétés d'un objet : une UMV doit appartenir à un périmètre irrigué par exemple.

Enfin, il faut être bien conscient que la qualité d'une donnée est fonction de l'utilisation qu'il va en être faite, et en particulier du niveau d'agrégation utilisé pour les restitutions : une surface d'une parcelle identifiée peut être fautive, mais agrégées au niveau d'un périmètre les surfaces ainsi calculées deviennent correctes.

Qualité sémantique : exhaustivité

Le premier objectif à atteindre est la couverture intégrale du domaine :

- ◆ Exhaustivité : tous les éléments présents dans la réalité ont-ils bien été recensés ? La référence à la géographie est utilisée en comparant les valeurs de superficie selon les sources de données enquête et télédétection.
- ◆ Redondance, recherche des doubles comptes. Identifiants différents pour plusieurs occurrences d'un même élément. À cet effet on peut avoir recours à des tris et pointages pour repérer des enregistrements différents concernant le même objet.
- ◆ Comptage des données manquantes, afin de contrôler variable par variable des taux de remplissage, en distinguant néanmoins les situations données manquantes de non-concerné.
- ◆ Dernier point de l'exhaustivité à contrôler, c'est l'actualité des données. Le système dispose pour cela d'un enregistrement de la date de transaction.

Qualité sémantique : exactitude et précision

Dans la progression de la démarche qualité, on arrive ici dans la partie la plus importante, mais aussi la plus complexe à mettre en œuvre puisque c'est ici que prend toute sa place le paradoxe d'évaluer la fiabilité d'une représentation qui est la seule connaissance disponible de la réalité. Dans la démarche engagée ici, on a envisagé deux façons d'aborder la qualité ici, en exploitant ce que l'on a introduit comme des règles issues de modèles :

- ◆ Analyse de la cohérence par des requêtes conditionnées par l'expression d'une ou de plusieurs règles. Tous les enregistrements ne respectant pas la contrainte sont sélectionnés, comptés et analysés.
- ◆ Analyse de la vraisemblance, par des calculs de statistique des dispersions monovariées et multivariées, avec éventuellement calculs de résidus relativement à un modèle.

Conclusion : la nécessité de Méta données

Dans la gestion d'une base de données comme celle de la SAED présentée ici, les deux rôles distincts de « producteur » et d'« utilisateur » sont bien apparus. La démarche qualité entreprise ici a été menée par le producteur de données afin de garantir un service à l'utilisateur. Dans le cas de base complexe, ce dernier ne peut avoir connaissance de tous les détails du processus de fabrication de l'information (collecte, saisie, contrôle...). Que ce soit pour le producteur dans la conduite de sa démarche qualité, ou pour l'utilisateur dans son exploitation des données, il est nécessaire de disposer d'information descriptive ou explicative des données fournies. Ces méta-données peuvent être classées selon leur usage :

- ◆ pour les collecteurs sur le terrain : protocole de relevé, sémantique avec une description précise et claire des unités d'observations et variables mesurées ;
- ◆ pour les préparateurs de données : méthodes de calculs des données élaborées, algorithme et condition de mise en œuvre ;

- ◆ pour les administrateurs : structure de la base de données, avec modèle relationnel, description des domaines de validité...;
- ◆ Pour les producteurs : procédures de chargement des données dans la base, localisation des données originelles...;
- ◆ Pour l'utilisateur : sémantique, localisation, valeurs.

Le problème posé est donc maintenant de l'intégration de ces méta-données dans la base de données afin de faciliter gestion et utilisation des données.

Bibliographie

FRANCO J.-M., PROMÉTHÉUS E.-I., 1997 — *Le Data Warehouse, Le Data mining*, Eyrolles.

LAURINI R., MILLERET-RAFFORT F., 1993 — *Les bases de données en géomatique*. Paris, Hermès.

LIBOKO-NDABAGNA S.-L., 1997 — Analyse de la Qualité des informations de la base de données de la SAED sur les aménagements hydroagricoles et les organisations paysannes. SAED, CIRAD-SAR, ENEA : 23.

MULLON C., PIRON M., 1998 — Sur la méthodologie de mise en place des observatoires socio-économiques. *De l'observation à l'analyse, implication de la biométrie dans les pays en développement*. F. Laloë et X. Perrier, Société française de Biométrie. 15 : 139.

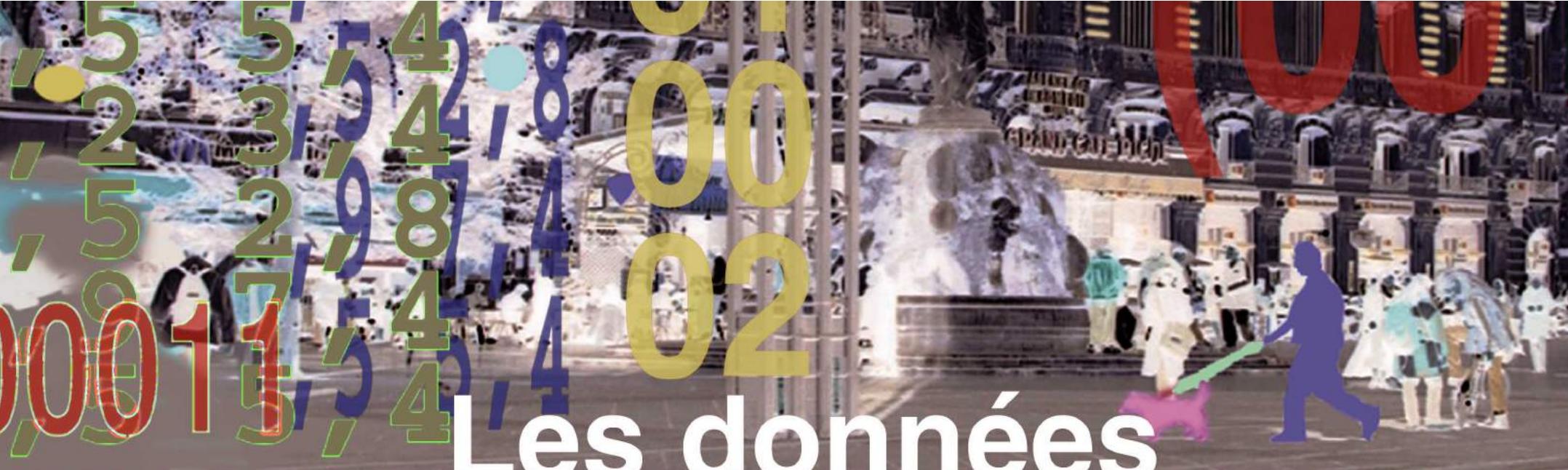
PASSOUANT M., 1998 — Statistiques et systèmes d'information. *De l'observation à l'analyse, implication de la biométrie dans les pays en développement*. F. Laloë et X. Perrier, société française de Biométrie. 15 : 139.

TARADELLAS E.I., 1997 — Meta-information statistique et nouvelles technologies. xxix^e Journées de statistique, Carcassonne.

¹ Cirad : Centre de coopération internationale en recherche agronomique pour le développement - Département TERA - BP 5035 - 34032 Montpellier Cedex.

² SAED : Société nationale d'Aménagement et d'Exploitation des terres du delta du Sénégal et des vallées du fleuve Sénégal et de la Falémé – direction de la Planification et du Développement rural - BP 74 - route de Khor - Saint-Louis - Sénégal.

³ Le parcellaire correspond ici à la superficie attribuée à une organisation paysanne ou à un exploitant agricole, regroupant le plus souvent plusieurs parcelles hydrauliques et/ou foncières non nécessairement contiguës.



Les données scientifiques

Bases de progrès des connaissances
Séminaire tenu à l'IRD Ile-de France, Bondy les 4 et 5 mai 1999

Éditeurs scientifiques Jean-Michel Kornprobst, Marcel Raffy

Sommaire

Introduction

Marcel Raffy, professeur, université de Strasbourg

Jean-Michel Kornprobst, professeur, université de Nantes

Jean-Pierre Muller, directeur général de l'IRD

1^{re} partie : acquisition et stockage des données

Coordinateur : Francis Laloë, IRD, Montpellier

Les données : expérience, observation et traitement
Francis Laloë

Les enjeux de l'information dans le domaine des pêches
Pierre Chavance

Contrôle de qualité des données. Application à un observatoire socio-économique spatialisé
Michel Passouant

Recherche d'informations dans un réseau de sources de données scientifiques hétérogènes et autonomes
Éric Simon

La manipulation de pétaoctets de données en physique des hautes énergies
Joseph Le Foll

Coordinateur, François Le Verge, Ifremer, Brest

Le contrôle qualité dans les centres de données
François Le Verge, Alain Laponche

Les aspects techniques de la pérennité des données scientifiques

Claude Huc, Danièle Boucon

Video and graphic broadcasting information system for research vessels

Présentation de l'application SDIV (Système de diffusion d'information et de vidéo) du navire océanographique Thalassa

Fabrice Lecornu, Armel Rué, Didier Lavoine

Utilisation des techniques avancées : base de données relationnelles, catalogues en ligne www, logiciels expert de contrôle qualité pour l'archivage, la gestion et la diffusion des données océanographiques

Catherine Maillard

Numérisation, transmission, acquisition et traitement de données géophysiques au département Analyse, Surveillance, Environnement du CEA

Pascal Dallot

2^e partie : gestion et valorisation des données

Coordinateur : Jean-Michel Kornprobst

Diffusion des données géographiques : valorisation et aspects juridiques

Pierre Peltre

Le partage et la diffusion des données et résultats scientifiques

Dominique Vuillaume



Les données scientifiques : de l'inconduite scientifique à la démarche qualité

Françoise Souyri



Bases de données pour les géosciences : un effort de connaissance et de prospective

Philippe Waldteufel



Conclusion des débats et synthèse

Marcel Raffy



La gestion informatique des chroniques en hydrologie

Michel Lang



Gestion et valorisation de données sur l'environnement global, avec l'exemple de Médias-France

Michel Hoepffner, Éliane Cubero-Castan, J.-L. Boichard



3^e partie : aspects juridiques et stratégiques

Coordinateur : Patrick Séchet, IRD, Paris

Les chercheurs peuvent-ils continuer à ignorer le droit ?

Patrick Séchet



Aspects juridiques de la diffusion des données scientifiques

Sébastien Lafargue



Diffusion des données de l'INPI

Bernard Marx



La CNIL et les fichiers de recherche médicale :
Les nouvelles procédures de formalités dans le secteur
de la recherche médicale

Jeanne Bossi



Adresse des auteurs

Jean-Luc **Boichard**, informaticien, Météo-France/Médias, BP 2102, 18, avenue E. Belin, 31401 Toulouse cedex 4.

Jeanne **Bossi**, secteur santé, CNIL, 21, rue St-Guillaume, 75007 Paris.
e-mail : jbossi@cnil.fr

Danièle **Boucon**, ingénieur CNES, 18, av. Edouard Belin, 34401 Toulouse cedex 4.

Eliane **Cubero-Castan**, informaticienne, Médias-France, BP 2102, 18, avenue Edouard Belin, 31401 Toulouse cedex 4.

Pierre **Chavance** IRD, BP 1386, Dakar, Sénégal.
e-mail : Pierre.Chavance@ird.sn

Pascal **Dallot**, assistant informatique, CEA/DAM, Analyse, surveillance, environnement, B.P. 12, 91680 Bruyères-le-Châtel.
e-mail : dallot@dase.bruyeres.cea.fr

Michel **Hoepffner**, hydrologue, IRD-Médias, BP 2102, 18, av. E. Belin, 31401 Toulouse cedex 4.
e-mail : Michel.Hoepffner@medias.cnes.fr

Claude **Huc**, ingénieur, département Valorisation et gestion des données spatiales, CNES, 18, av. Edouard Belin, 31401 Toulouse cedex 4.
e-mail : claude.huc@cnes.fr

Jean-Michel **Kornprobst**, professeur université de Nantes, vice-Président de la CS7, ISOMer, Laboratoire de chimie marine, BP 92208, 2, rue de la Houssinière, 44322 Nantes celex 3.
e-mail : jean-michel.kornprobst@wanadoo.fr

Sébastien **Lafargue**, juriste, Ifremer, Technopolis 40, 155, rue J.J. Rousseau, 92138 Issy-les-Moulineaux.
e-mail : Sebastien.lafargue@ifremer.fr

Francis **Laloë**, IRD, Halieutique et Écosystèmes Aquatiques, BP 5045, 34032 Montpellier cedex 1.
e-mail : laloe@mpl.ird.fr

Michel **Lang**, hydrologue, Cemagref, Division hydraulique, 3 bis, quai Chauveau, CP 220, 69009 Lyon cedex.
e-mail : michel.lang@cemagref.fr

Alain **Laponche**, ingénieur Sismar, Ifremer, centre de Brest, BP 70, 29280 Plouzané.

Didier **Lavoine**, ingénieur réseau, 2 bis, rue R. Le Ricollais, 44000 Nantes.

Fabrice **Lecornu**, ingénieur informaticien, Ifremer, centre de Brest, BP 70, 29280 Plouzané.
e-mail : Fabrice.Lecornu@ifremer.fr

Joseph **Le Foll**, informaticien, CEA/DSM/DAPNIA, CE Saclay, 91191 Gif-sur-Yvette cedex.
e-mail : lefoll@hep.saclay.cea.fr

François **Le Verge**, chef du service de la documentation, Ifremer, centre de Brest, BP 70, 29280 Plouzané.
e-mail : fleverge@ifremer.fr

Catherine **Maillard**, ingénieur de recherche Ifremer, centre de Brest, BP 70, 29280 Plouzané.
e-mail : Catherine.Maillard@ifremer.fr

Bernard **Marx**, INPI, service DDI, 26 bis, rue de Saint-Pétersbourg, 75008 Paris.

Jean-Pierre **Muller**, pédologue, directeur général de l'IRD, 209-213, rue La Fayette 75480 Paris cedex 10.

Michel **Passouant**, statisticien Cirad, Campus International de Baillarguet, Bât. F, 34398 Montpellier cedex 4.
e-mail : michel.passouant@cirad.fr

Pierre **Peltre**, géographe, IRD, 32, avenue Henri-Varagnat, 93143 Bondy cedex.
e-mail : peltre@clarke.bondy.ird.fr

Marcel **Raffy**, professeur, université de Strasbourg, président de la CS7, ULP-CNRS, Parc d'innovation, 5, bd S. Brandt, 67400 Illkirch-Graffenstaden.

Armel **Rué**, ingénieur réseau, Ifremer, centre de Brest, BP 70, 29280 Plouzané.

Patrick **Séchet**, informaticien, IRD, 209-213, rue La Fayette, 75480 Paris cedex 10.
e-mail : sechet@paris.ird.fr

Éric **Simon**, directeur de recherche en informatique, Inria, BP 105, 78153 Le Chesnay.
e-mail : eric.simon@inria.fr

Françoise **Souyri**, directeur de recherche, MENRT-CSDR, 5, rue Descartes, Paris cedex 05.
e-mail : francoise.souyri@dr.education.gouv.fr

Dominique **Vuillaume**, économiste de la santé, Service du partenariat pour le Développement, Inserm, 101, rue de Tolbiac 75654 Paris cedex 13.
e-mail : vuillaume@tolbiac.inserm.fr

Philippe **Waldteufel**, climatologue, CNRS-IPSL, 10-12, avenue de l'Europe, 78140 Vélizy.
e-mail : Philippe.Waldteufel@ipsl.uvsq.fr