

La banque de données

The data bank

G. Bédriot	Agence financière de bassin « Seine-Normandie », Nanterre
J.F. Nouvelot	Agence financière de bassin « Rhône-Méditerranée-Corse », Pierre Bénite
B. Pister	Arrondissement fonctionnel de l'eau et de l'assainissement du Val de Marne, Créteil
J.C. Scherer	Ministère de l'Agriculture, D.I.A.M.E., Paris
P. Tourasse	Electricité de France, Division technique générale, Grenoble

Une attention particulière est accordée d'abord aux informations descriptives de la fiche signalétique de la station de mesure. Ces renseignements n'ont pas encore fait l'objet d'une informatisation poussée bien qu'ils soient très importants pour qualifier les données des chroniques d'observations. Par ailleurs, la collecte, la saisie, la validation, l'archivage, la mise à jour, la gestion et la mise à disposition des données sont décrits à partir d'exemples pris dans les banques hydrométriques existantes. Avec l'apparition de nouvelles méthodes d'analyse et de conception de systèmes d'information, les prochaines créations de banques de données devraient contribuer à la réunification progressive des informations relatives aux différents domaines du cycle de l'eau.

Particular attention is paid, primarily, to the descriptive data included on the measuring station's specification sheet. Despite the importance of this information to the evaluation of data drawn from chronological observations, it has not yet been the object of full-scale computerizations. Furthermore, the gathering, inputting, vetting, archiving, up-dating and management of data are described, using examples based on existing hydrometric data banks. With the appearance of new analytical methods and information system designs, data banks established in the near future should contribute to the progressive accumulation of data relative to the different stages of the rain cycle.

1. Introduction

Le choix des données à informatiser dépend des traitements que l'on souhaite réaliser. Après plusieurs « itérations » sur données et traitements, on détermine ce qu'il est « raisonnable » de faire pour aller à l'essentiel.

Ainsi la vie de la station, l'évolution de ses équipements, des méthodes de mesures, les changements d'hommes et de services qui la gèrent... ne sont généralement pas informatisés. Dans de nombreux cas cette information

n'existe pas et c'est dommageable pour l'interprétation des mesures. Le premier chapitre consacré à la « fiche signalétique de la station » montre l'importance de ce travail.

Les trois chapitres suivants : « collecte des mesures, leur saisie et leur validation » puis « transmission, archivage, mise à jour et gestion » et enfin « mise à disposition », décrivent des expériences et des projets : ils ne visent pas à l'exhaustivité.

La conclusion dégage quelques traits de la situation actuelle et donne quelques perspectives à moyen et à long termes.

AGENCE DE BASSIN RHONE MEDITERRANEE CORSE

CATALOGUE DES STATIONS HYDROMETRIQUES

CODE OUVRAGE 1366223002 C C A D Y COURS D'EAU VILLEFRANCHE DE CONFLENT STATION
 CODE HYDRO 10 11 CODE EAUX SOUTERRAINES COMMUNE CODES GESTIONNAIRES
 10436400004 620B11 VILLEFRANCHE DE CONFLENT 11 2510431642
 B.V. TOPO km2 B.V. REEL km2 X Y ZONE CARTE 1/25 000 eme 12 13 20 21 24
 60280 60280 3190 3 PRADES 34 2349
 TEXTE :

PERTES IMPORTANTES DANS LES CALCAIRES DEVONIENS ALIMENTANT LA RESURGENCE D'ENGORMER - LES BASSES EAUX OBSERVEES REPRESENTENT MAL L'ECOULEMENT DU BASSIN

CODE	DEBUT	GESTIONNAIRES	STATION	CONTROLE	EQUIPT	COMPLT	ZERO ECH	FIN	RESULTATS	NATURE
12	11/1966	DDA 66	NA	SA	LB		423.72	07/1970	QMJ	NA
12	08/1970	DDA 66	AR	SE	LB		423.72		QMJ	NA
12										
12										
12										
12										

1.

AGENCE DE BASSIN R.M.C (RB273)

DATE 10/09/86 PAGE 1/ 1

STATION HYDROMETRIQUE : FICHE TECHNIQUE

COURS D'EAU : C A D Y

STATION : VILLEFRANCHE DE CONFLENT

CODE OUVRAGE : C 13 66 223 002

CODE HYDRO : Y 043 64 00004

CODES GESTIONNAIRES : 25043642

COMMUNE : VILLEFRANCHE DE CONFLENT

CODE INSEE : 66 223

TYPE DE STATION : MILIEU NATUREL

BASSIN VERSANT TOPOGRAPHIQUE = 60.2 KM2

BASSIN VERSANT REEL =

X = 602.80

Y = 31.90

ZONE COORDONNEES : 3

CARTE : PRADES

NO : 3-4

NO NOUVELLE CARTE :

ECHELLE : 1/25000 EME

TEXTE : PERTES IMPORTANTES DANS LES CALCAIRES DEVONIENS ALIMENTANT LA RESURGENCE

D'ENGORMER - LES BASSES EAUX OBSERVEES REPRESENTENT MAL L'ECOULEMENT DU BASSIN

DEBUT	FIN	GESTIONNAIRES	INSTALLATION	ZERO ECH.	RESULTATS	NATURE
11/1966	07/1970	DDA 66	STATION : NATUREL CONTROLE : SANS CONTROLE	423.72	QMJ	NATUREL
08/1970		DDA 66	STATION : ARTIFICIEL CONTROLE : SEUIL	423.72	QMJ	NATUREL

2.

2. Fiche signalétique de la station

Cette fiche réunit sous forme condensée le maximum d'informations relatives à chaque station de mesures.

2.1. L'identifiant

C'est la propriété qui caractérise chaque station de façon unique. Actuellement, il y a autant de manières d'identifier que de banques existantes. Par exemple, en rivière, l'agence financière de bassin Rhône Méditerranée Corse utilise un code de référence basé sur le code hydrologique complet, fournissant en principe un repérage précis : région (1 lettre), secteur (1 chiffre), sous secteur (1 chiffre), zone (1 chiffre), tronçon (2 chiffres), le point kilométrique et le type d'ouvrage (*figure 1*). Le ministère de l'Environnement, lors de la création de l'inventaire national de la pollution avait défini en 1971 un code à trois chiffres avec une répartition géographique marquée, mais non réellement structurée. Dès 1975, dans le cadre de la préparation de l'inventaire quinquennal de 1976 et en tenant compte de la naissance de réseaux de mesures de la qualité des rivières, la structure du code a été portée à 6 chiffres.

En réseau d'assainissement, les gestionnaires utilisent aussi un identifiant, dont la définition varie d'une collectivité à l'autre.

2.2. Le nom de la station

Il est accompagné de celui du cours d'eau ou de celui du collecteur.

2.3. Le nom du gestionnaire

C'est le producteur de données. Il peut y avoir plusieurs intervenants pour une même station, au cours du temps. Parfois aussi, le gestionnaire sous-traite une partie des opérations à des prestataires de services : par exemple les laboratoires spécialisés d'analyses de la qualité.

2.4. La localisation

2.4.1. La localisation administrative et descriptive

En rivière, le nom de la commune, avec le code INSEE, est accompagné le cas échéant du nom du lieu-dit. Le cas particulier des cours d'eau dont le tracé sert de limite de commune pose un problème le plus souvent mal résolu, notamment pour les stations de mesures de la qualité de l'eau.

En assainissement, on utilise l'adresse postale.

2.4.2. La localisation hydrologique

La codification hydrologique française établie par les circulaires de 1968 permet en principe un repérage précis par rapport au cours d'eau. Son application n'est pas identique dans les six agences de bassin, notamment au niveau des 4^e et 5^e chiffres identifiant les tronçons.

En assainissement, certaines collectivités ont adopté un système similaire pour codifier l'arborescence des principaux collecteurs, mais il n'y a pas de règle générale.

2.4.3. La localisation dans l'espace géographique

La plus répandue se limite au repérage en x, y sur un plan et utilise les coordonnées Lambert qui sont la référence officielle.

Des repérages plus complets comportant l'altitude z sont apparus récemment avec la numérisation de certains réseaux de distribution d'eau ou d'assainissement

En conclusion, l'établissement d'un système unique et général de références spatiales est un problème urgent. Sa solution conditionne les applications à venir en faisant éclater le cadre étroit de chaque fichier voire même de certaines banques et en préparant des passerelles communes entre les différents domaines de l'eau. Dans cet esprit et à terme la localisation serait partie intégrante de l'identifiant dont un des éléments qualifierait la nature de la station (station de prélèvement pour analyse, station automatique de mesure de la qualité, station hydrométrique,...).

2.5. Les caractéristiques du bassin versant à l'amont de la station

En rivière, il s'agit essentiellement de la superficie du bassin versant d'alimentation dont le tracé peut être celui des lignes topographiques de partage des eaux si la station n'est pas influencée par des aménagements hydrauliques ou par des singularités géologiques. Dans le cas contraire, ce sera le bassin actif. Des informations complémentaires concernant les sols, la géologie, l'occupation des sols, la climatologie... peuvent figurer dans un texte libre de quelques centaines de caractères (*figure 2*).

En assainissement : pour les eaux pluviales on indique la surface et le taux moyen d'imperméabilisation du bassin versant, pour les eaux usées le nombre d'habitants théoriquement raccordés.

2.6. L'environnement de la station

En rivière, des précisions sur l'environnement de la station peuvent éventuellement être apportées dans un texte libre.

En assainissement, l'environnement est lié directement à l'ouvrage d'accueil de la station. Cet ouvrage a un code, et une vocation (eaux usées, eaux pluviales, unitaire). Il est caractérisé par une section maximale et une relation entre la hauteur de remplissage et la section mouillée correspondante.

2.7. La description de la station de mesure

En rivière, on décrit l'emplacement des principaux équipements et l'on note l'altitude du zéro de l'échelle limnimétrique avec le système de nivellement correspondant.

En assainissement on situe la section de mesure dans l'ouvrage et l'emplacement (x, y, z) des capteurs. Par ailleurs, le cheminement des câbles vers l'armoire de mesures est relevé avec précision.

2.8. L'équipement de la station

En rivière on note le type de station (naturel ou artificiel), le type de contrôle (seuil, déversoir, rétrécissement,...). Les

caractéristiques des différents appareils de mesures : échelle limnimétrique, limnigraphe, combinaisons d'équipements, ..., de même que les systèmes de mémorisation sur site ou de télétransmission.

En assainissement les mêmes renseignements sont recueillis avec une description plus poussée des capteurs (marque, référence) et de leurs paramètres d'étalonnage avec les dates de mise à jour.

Dans les deux cas des indications concernant la vitesse de déroulement des enregistreurs, les échelles de réduction, le type et le mode d'enregistrement sur mémoire informatique (notamment le pas de scrutation)... sont des éléments indispensables pour apprécier la précision des mesures.

2.9. les méthodes de mesures

Cette rubrique très complexe touche à la fois les appareillages de terrain (moulinets, type de préleveur, ...) la manière de s'en servir, les méthodes d'analyses en laboratoire et les appareillages associés, les méthodes de dépouillement. Sauf cas très particuliers la rigueur dans la description de ce domaine est notoirement insuffisante.

Cependant de nombreux éléments d'information existent et sont quelquefois exploités : ainsi pour une station de jaugeages, on indiquera la forme de la courbe de tarage, le nombre annuel de jaugeages, les débits maximal et minimal mesurés et l'amplitude des extrapolations.

2.10. Le carnet de bord d'exploitation de la station repose sur l'établissement des chroniques relatives aux :

- pannes (périodes de marche, arrêts et cause des arrêts);
- contrôles effectués pour la maintenance technique de la station et pour l'exploitation hydrologique des mesures;
- relevés des observateurs.

2.11. Les influences

Elles peuvent être extraites du carnet de bord de l'exploitation des ouvrages susceptibles d'influencer la station de mesure :

- en rivière, il s'agit de définir les influences des aménagements amont (ou aval);
- en assainissement, l'exploitation datée du réseau inclus dans la zone d'influence concerne des :
 - manœuvres de vannes, de poutrelles,...
 - curages
 - mises en route de stations de pompage de défense contre les crues
 - branchements d'ouvrages neufs

2.12. Les données disponibles

Le type, la période d'observation, la nature et le lieu de l'archivage, les conditions d'obtention... Ce catalogue pourrait être consulté sur Minitel à moyen terme.

2.13. Le dossier de vie de la station consiste à établir la chronologie de l'ensemble des changements intervenus sous chacune des rubriques précédentes avec indication de

leur nature et de leur date. On établit ainsi l'historique des données, sous forme de séquences réputées homogènes.

2.14. Conclusion sur la fiche signalétique

Les renseignements précédents sont indispensables à la fois aux gestionnaires des stations de mesures et pour toute étude hydrologique sérieuse. En principe ils existent au moins sous forme d'archive papier chez les gestionnaires.

A ce jour l'analyse de ces données et leur traitement restent à effectuer à partir des renseignements existants dans les différents organismes concernés.

3. La collecte des mesures, leur saisie et leur validation

3.1. Discussion préalable sur les données élémentaires et les données élaborées contenues dans une banque

3.1.1. Les données élémentaires de la banque

Une donnée élémentaire archivée en banque n'est pas forcément le résultat élémentaire d'un mesurage. Dans un jaugeage au moulinet, un point de mesure du champ des vitesses implique le recueil de l'information suivante : sa position dans la section de mesure (distance à la rive d'origine, profondeur), le nombre de tours de l'hélice pendant une durée déterminée. Pour le jaugeage auquel appartient ce point, on notera la date, les heures de début et de fin d'opération avec les hauteurs correspondantes lues à l'échelle limnimétrique, la largeur de la section, la courbe d'étalonnage du moulinet utilisé. Ces résultats élémentaires du mesurage sont toujours relevés avec soin, quelquefois sur microordinateur qui permet le calcul automatique du débit recherché.

Dans certains cas de reprise des données anciennes, la donnée élémentaire est le débit moyen journalier, voire même le débit moyen mensuel, parce qu'il n'est plus possible de remonter aux mesures élémentaires.

3.1.2. Les données élaborées de la banque

Dans l'exemple d'ARHMA, le débit instantané résultant de l'application de la courbe d'étalonnage à une hauteur instantanée est une donnée élaborée, a fortiori dans le cas du débit moyen journalier [1].

3.2. Les procédures de collecte des données : exemple des hauteurs d'eau et des débits

Plusieurs procédures sont utilisées :

- 1) Les données lues à l'échelle limnimétrique sont transmises par un observateur au gestionnaire qui procède à la saisie.
- 2) Les données sont enregistrées sur un support graphique, le limnigramme, puis traitées par le gestionnaire qui les critique et sélectionne par discrétisation les valeurs à archiver. Un opérateur saisit les coordonnées (hauteur et temps) des points retenus, grâce à une table de lecture

(digitalisation). Les valeurs sont enregistrées sur support informatique (cassettes ou disquettes).

3) Les données sont saisies directement sur le site et transmises au gestionnaire par télétransmission (satellite, radio, téléphone) ou stockées sur place sur mémoire informatique relevée périodiquement.

Les données étant saisies selon une procédure déterminée à l'avance (soit pas de temps de scrutation fixe, soit variation significative de la grandeur mesurée), on procédera à un compactage pour éliminer les informations redondantes après critique par visualisation sur écran graphique ou par procédure informatique.

4) La saisie des débits journaliers déjà calculés est une procédure peu fiable car les informations concernant la courbe de tarage et le calcul des débits ne sont plus prises en compte. Elle n'est donc utilisée que pour les données anciennes.

Dès la fin des années 1960 un système de collecte du débit moyen journalier est mis au point avec 3 chiffres significatifs. Un quatrième chiffre indique la place de la virgule, l'unité choisie étant généralement le mètre cube par seconde. Un code qualité accompagne cet ensemble de 4 chiffres.

3.3. Reprise des données anciennes

La précision des résultats de nombreuses études hydrologiques est étroitement liée à la longueur des chroniques de données disponibles. Il est donc de première importance de pouvoir mettre à la disposition des utilisateurs l'ensemble des séries d'observations collectées, y compris les plus anciennes.

Le rattrapage du fond ancien de données présente toutefois un certain nombre de difficultés :

- le volume de données à traiter serait voisin de 25 000 stations années (un inventaire précis est indispensable) pour l'ensemble du territoire national;

- ces données correspondent pour l'essentiel à des stations, dont les archives sont détenues en majorité par des services dépendant du ministère de l'Environnement, mais qui au cours du temps ont changé plusieurs fois de gestionnaires (grandes forces hydrauliques, circonscriptions électriques, D.D.E., services hydrologiques centralisateurs). Le retour aux observations et aux mesures originales (bordereaux de lectures d'échelles, limnigrammes, feuilles de jaugeages) risque de s'avérer laborieux, voire impossible : dans tous les cas une perte d'information est inéluctable;

- les méthodes d'observations et de mesures ainsi que les méthodes de traitement ont évolué, et ne permettent pas de garantir l'homogénéité des données dans le temps.

Devant l'ampleur de la tâche et les délais nécessaires, la reprise des données anciennes pourrait s'échelonner de la manière suivante :

1^{re} phase : Intégration des fichiers « débits moyens journaliers » existants, ou saisie de ces débits à partir de tableaux manuscrits ou dactylographiés. Ces données permettent de satisfaire plus de 80 % des demandes formulées par les différents utilisateurs. Elles devront faire l'objet de vérifications et critiques en utilisant :

- des tests de contrôle au niveau de la saisie

- l'analyse graphique

- des tests statistiques permettant de vérifier l'homogénéité et la cohérence des données tant au niveau d'une station qu'au niveau régional.

2^e phase : Intégration des fichiers hauteur/temps existants avec utilisation d'un pas de temps fixe, conservant la forme du limnigramme.

3^e phase : Saisie des bordereaux d'observations de hauteurs d'eau ou digitalisation des limnigrammes.

4^e phase : Elaboration des courbes d'étalonnage et des barèmes correspondants, puis si possible, création d'un fichier des jaugeages accessible à certains utilisateurs.

5^e phase : Calcul et validation des débits moyens journaliers : il sera utile de comparer les débits ainsi calculés à ceux archivés lors de la première phase.

3.4. Collecte, saisie, validation en assainissement

3.4.1. Principe de saisie numérique des mesures

Selon son type, une station peut mesurer de 1 à 4 valeurs simultanément (1 mesure de hauteur + 1 à 3 mesures de vitesse selon 1 à 3 transects horizontaux de l'écoulement dans le cas des sondes à ultra-sons).

Les mesures sont saisies à pas de temps variable quand il y a une variation significative de l'un des 4 paramètres mesurés. Le pas de temps, et les différents seuils d'acquisition sont bien entendu paramétrables.

Le choix de ces seuils est délicat. Pour ne pas perdre en précision, ils sont choisis à un niveau assez bas, ce qui provoque une redondance dans les mesures.

3.4.2. Validation des mesures

3.4.2.1. Objectifs

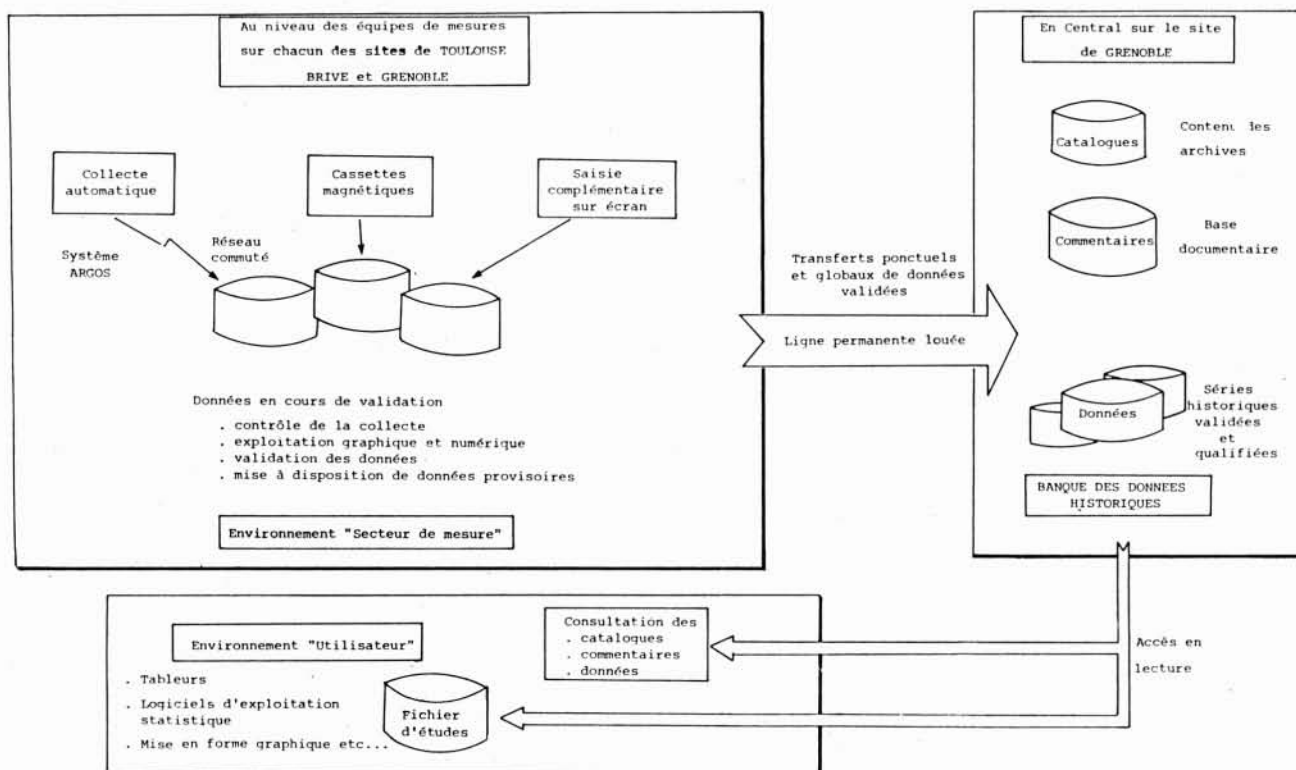
Il est indispensable de s'assurer *in situ* du bon fonctionnement de la station en examinant attentivement les mesures. Que ces dernières soient sur un graphique ou sur un fichier saisi numériquement sur site, seules les mesures ayant reçu un « label qualité » par un opérateur averti sont susceptibles d'alimenter une banque de données.

Il est également souhaitable d'optimiser le nombre d'informations nécessaires pour représenter une séquence de mesures sans que la précision de celle-là n'en soit affectée.

3.4.2.2. Moyens

Pour réaliser correctement la critique et la validation des mesures sur fichiers numériques, il est indispensable de recréer un contact visuel graphique; l'informatique intervient ici pour mettre à la disposition de l'opérateur des moyens conversationnels et interactifs dont la souplesse soit telle que l'opérateur conserve le sentiment de faire un travail d'hydrologue et non pas d'informaticien.

Il doit pouvoir établir aisément une relation entre les mesures de pluies et celles réalisées dans le réseau, distinguer les phénomènes liés à l'exploitation du réseau (dégrillage, pompage, manœuvres de vannes, curage,...), éventuellement lisser une courbe pour diminuer le nombre de mesures.



3. Le circuit des données hydrologiques à la DTG.

3.5. De la saisie sur site à la collecte informatique : l'expérience de l'E.D.F.

Le service ressources en eau de la Division technique générale est chargé de recueillir, mettre en forme et archiver l'ensemble des données hydroclimatologiques qui sont nécessaires pour dimensionner et gérer de manière optimale les aménagements hydroélectriques d'E.D.F.

Il gère pour cela un réseau de près de 1 000 points de mesures de précipitations, d'enneigement, de hauteurs d'eau en rivière, de températures de l'air et de températures de l'eau [2] et [3].

3.5.1. Saisie sur site sur cassettes... bilan d'une expérience

Dans l'espoir de s'affranchir des étapes fastidieuses de dépouillement et de digitalisation des enregistrements, des études ont été menées à bien dès la fin des années soixante pour saisir les données directement sur le site de mesures. Elles conduisirent à la mise au point d'un enregistreur autonome à cassettes magnétiques, à déroulement, qui pouvait résister à des conditions d'environnement sévères et qui était capable de mémoriser sur site à de faibles pas de temps (6 à 12 minutes) un à plusieurs mois de données.

Au fil des années la quasi totalité du réseau a été équipé et aujourd'hui encore près de 450 points de mesures sont dotés de cet enregistreur à cassettes. Les cassettes sont

relevées régulièrement par des observateurs ou directement par les techniciens lorsqu'ils visitent la station. Elles sont acheminées par courrier et dépouillées ensuite par programme sur une baie de lecture qui est connectée au calculateur.

Compte tenu des délais de relevé, d'acheminement, de dépouillement et de contrôle, les données ainsi rassemblées ne sont disponibles qu'après quelques mois. De plus, malgré quelques améliorations apportées au système, la perte d'informations résultant d'incidents de fonctionnement (problèmes d'alimentation, mauvais réglage de l'enregistreur, usure ou détérioration de la cassette,...) reste élevée et nécessite à chaque fois des opérations coûteuses de redigitalisation des enregistrements.

L'expérience a ainsi montré que l'exploitation opérationnelle d'un important réseau de saisie sur cassettes est extrêmement lourde. Les surcoûts de maintenance et de contrôle des valeurs enregistrées sont importants et minimisent alors les avantages certains de la saisie sur site.

3.5.2. La collecte automatique des données... le système SCHTROUMPH

L'automatisation complète du réseau téléphonique public autocommuté, qui est effective sur l'ensemble du territoire français depuis 1980, permet d'atteindre la quasi totalité des stations pluviométriques et hydrométriques par téléphone sans qu'il soit nécessaire, notamment en monta-

gne, de construire des équipements coûteux de relaying.

La banalisation récente d'un certain nombre de dispositifs électroniques permet également de disposer de répondeurs à mémoire interrogeables par téléphone à des coûts relativement faibles lorsqu'ils peuvent être fabriqués en série.

Il est donc possible désormais d'envisager d'automatiser complètement la collecte des données et d'étendre cette automatisation à l'ensemble des points de mesures qui sont accessibles par le téléphone.

3.5.3. Etat d'avancement du projet et avantages du système

Sur un total de 400 points de mesures qui à terme seront équipés de répondeurs à mémoire et pourront donc ainsi être interrogés directement à partir d'un ordinateur, près de 200 points sont déjà installés. Sur cet ensemble une soixantaine de répondeurs sont exploités en temps réel et collectent des données qui alimentent des modèles de prévision de crues de type pluie-débit (prévision des crues sur l'Ardèche, l'Eyrieux, la Ceze, le Fier et la Valserine dans le cadre de la gestion des aménagements hydroélectriques du Rhône) ou débit-débit (prévision des crues sur la Basse-Isère au niveau du poste hydraulique de Vallée de Pizançon. Les autres sont interrogés à la demande pour contrôler le fonctionnement du réseau de mesures ou suivre en temps peu différé l'évolution hydrologique de certains bassins sensibles. Une sélection de postes participe déjà à l'alimentation automatique des fichiers de mesures sous forme d'une scrutation systématique qui est effectuée chaque nuit.

Les multiples avantages attendus de la généralisation de cet équipement sont :

- la disponibilité de tout le réseau de mesures en temps réel pour les prévisions à court terme;
- le contrôle permanent du fonctionnement des capteurs : gain considérable d'efficacité dans la gestion des réseaux de mesures;
- la mise à jour quotidienne des fichiers en ordinateur, supprimant la dualité actuelle et la discordance des valeurs admises en temps réel et des valeurs archivées *a posteriori* (figure 3);
- l'accès multiple à l'information : plusieurs utilisateurs indépendants peuvent à volonté consulter soit directement des mesures individuelles de terrain, soit des synthèses gérées par ordinateur;
- l'importance de la mémoire des postes d'acquisition permet d'éviter la perte des informations en cas d'interruption des liaisons plusieurs jours durant.

3.6. La collecte : une activité intégrée à la production de données

L'évolution récente de l'automatisation de la collecte a rejeté les opérations de validation des données hydrologiques après celles de collecte automatique et de contrôle de transmission. Or, seul le gestionnaire des mesures est compétent pour assurer leur validation hydrologique; de plus en plus, il le fait après avoir réalisé lui-même les contrôles de saisie, autrefois dévolus à l'équipe de gestion des fichiers de données.

**

4. Transmission, archivage, mise à jour et gestion

4.1. L'existant pour les rivières en matière de transmission et d'archivage

4.1.1. Les transmissions

Les procédures sont très variables selon les organismes. E.D.F. utilise le réseau téléphonique commuté et le système ARGOS pour rapatrier les données de ses stations automatiques de mesures, dispose d'une ligne spécialisée permanente entre ses ordinateurs de Toulouse, de Brive et de Grenoble et reçoit des données de la météorologie via le réseau TRANSPAC.

L'agence financière de bassin Rhône Méditerranée Corse dispose d'un réseau local et envisage de communiquer avec les producteurs extérieurs et avec les utilisateurs via TRANSPAC.

La base ARHMA, utilise aussi le réseau TRANSPAC pour la transmission des données. Cette procédure nécessite simplement que le producteur de données dispose d'un microordinateur fonctionnant comme terminal.

4.1.2. L'archivage

L'archivage en banque de données est organisé :

- pour D.T.G.-E.D.F. en fichier à accès direct;
- pour ARHMA en fichier séquentiel indexé;
- pour l'agence financière Rhône Méditerranée Corse en système de gestion de base de données (S.G.B.D.).

4.2. Exemple de projet d'archivage de mesures de débit en réseau d'assainissement; cas du Val de Marne (programme A.ME.DE.)

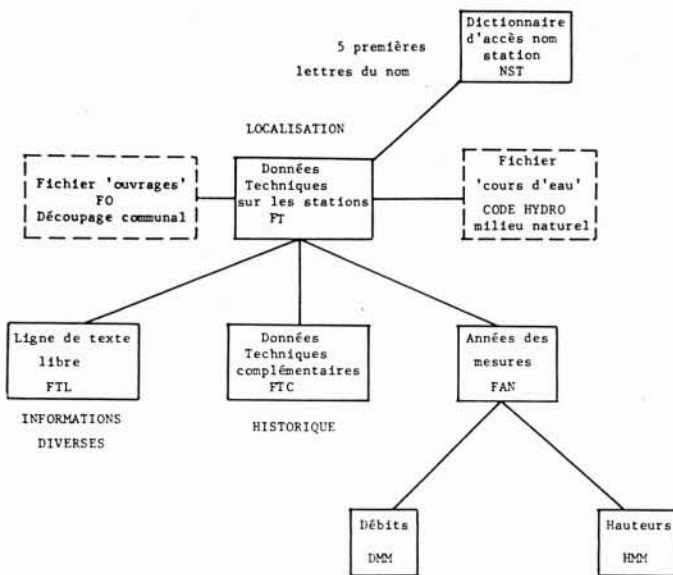
Il s'agit d'un cas où un service opérationnel en assainissement est à la fois producteur et utilisateur de données.

Le programme A.ME.DE. : (Archivage de Mesures de Débit) est au stade de projet qui verra sa réalisation en 1987. Son objet est d'archiver les mesures validées de hauteur, de vitesse et de débit réalisées dans le réseau, en vue des utilisations suivantes :

- études générales d'assainissement, calage de modèles hydrologiques;
- études de fonctionnement;
- études statistiques par temps sec ou non.

La constitution des archives doit être organisée pour permettre des éditions multicritères sur des séquences archivées homogènes.

La nécessaire mise à jour des archives, au fil du temps, met en évidence l'extrême sensibilité du programme permettant la critique et la validation des mesures, point de passage obligé ne devant provoquer ni lassitude ni embouteillage, préjudiciables à la qualité de la banque de données.



4. Banque de données hydrométriques « agence RMC ».

4.3. Une étude de cas concernant la gestion d'une banque de données hydrométriques : l'exemple de l'agence de bassin Rhône Méditerranée Corse

4.3.1. Organisation

La gestion des informations utiles aux différentes missions de l'agence repose sur un système informatique général de base de données (système IDS II : Integrated Data Store, de Bull). Ces informations constituent l'amorce d'une future banque de l'eau du bassin susceptible d'être interrogée tant de l'extérieur que de l'intérieur de l'agence. L'ensemble des données est structuré en plusieurs sous-ensembles :

- les données relatives aux usages de l'eau;
- celles ayant trait aux divers milieux naturels : cours d'eau, lacs et plans d'eau, eaux souterraines;
- les données relatives aux entités administratives compte tenu de leur importance pour la gestion de l'agence : communes, départements,...

La banque de données hydrométriques s'intègre à ce système par l'intermédiaire du fichier commun des ouvrages, dans lequel sont archivés non seulement les ouvrages liés aux différents usages de l'eau (prélèvements, canaux d'irrigation, stations de traitement, assainissement,...) mais également, les stations de mesures et d'observations. Chacun de ces ouvrages est identifié par un préfixe caractérisant son type, par le code INSEE de la commune d'implantation et par un numéro d'ordre dans cette commune.

En plus de son rattachement au découpage administratif, une station hydrométrique est, par l'intermédiaire du code hydrologique (numéro du cours d'eau + point kilométrique), rattachée au fichier commun des milieux naturels (figure 4).

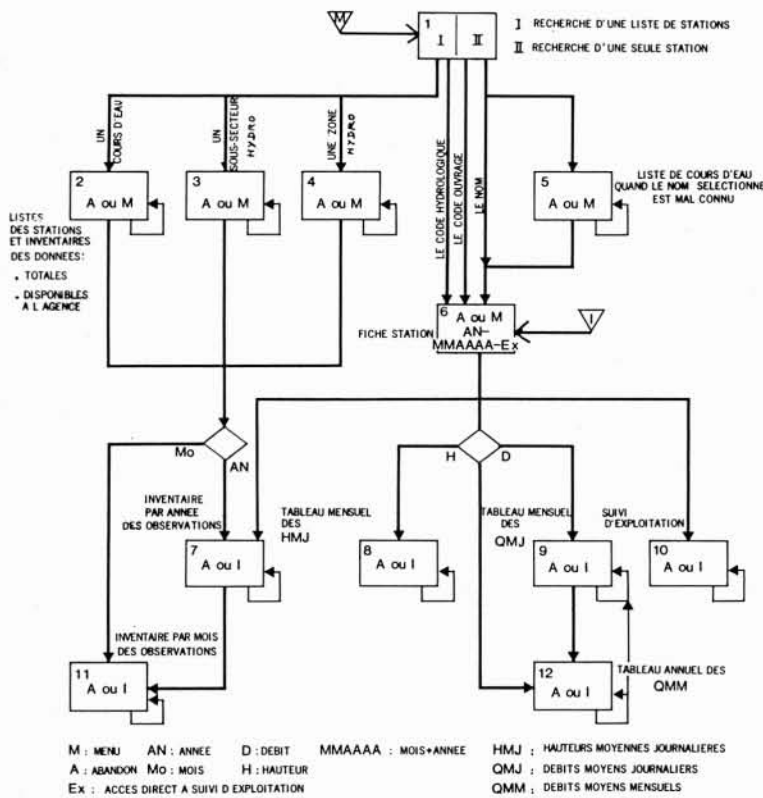
4.3.2. Archivage

Dans une première phase, l'agence a donné priorité à la recherche et à l'archivage des données anciennes. Il a été recensé plus de 1 700 stations en activité ou désaffectées, représentant plus de 20 000 années-stations d'observations et de mesures relatives à des stations implantées directement en milieu naturel (débits observés ou parfois reconstitués) ou appartenant à des aménagements hydrauliques si elles présentent un intérêt pour la connaissance des régimes hydrologiques ou de la ressource en eau (aménagements hydroélectriques par exemple).

Les données archivées sont de deux types :

- 1) Les données d'identification, de localisation et du suivi d'exploitation, regroupés dans un catalogue des stations,
- 2) Les données hydrométriques proprement dites : débits moyens journaliers, débits moyens mensuels, débits maximaux instantanés, débits moyens annuels ou éventuellement hauteurs moyennes journalières en l'absence de débits.

L'agence ne gérant directement que quelques stations hydrométriques, l'alimentation des différents fichiers s'effectue à partir de données fournies par les différents gestionnaires de réseaux du bassin, soit sous la forme de tableaux manuscrits ou dactylographiés, soit sur supports informatiques.



5. Banque de données hydrométriques « agence RMC ». Transactions d'affichage. Enchaînement des écrans.

Le volume d'informations archivées représente :

- environ 1 million de caractères pour le catalogue des stations;
- près de 65 millions de caractères pour les débits et les hauteurs;
- à cela doit s'ajouter 2 millions de caractères en flux annuel.

4.3.3. Mise à jour

Différentes procédures permettent la mise à jour en mode transactionnel, des différents lots d'informations : fiches de stations, débits, hauteurs.

La méthode par « menus » permet d'effectuer des choix, tant sur l'opération à effectuer, création, modification ou suppression que sur les données à traiter.

Lors de « créations » ou de « modifications », de nombreux contrôles sont réalisés avant l'archivage définitif, chaque erreur rencontrée faisant l'objet d'un message affiché à l'écran.

Un code « qualité » peut être introduit manuellement pour chaque donnée journalière. Par contre, les codes « qualité » des moyennes mensuelles et annuelles sont définis et archivés automatiquement à partir des codes des valeurs journalières.

4.3.4. Extraction et transmission des données

4.3.4.1. Modes d'accès aux données

Deux modes d'accès complémentaires permettent d'extraire des données de la banque en répondant aux besoins suivants :

- consultations en conversationnel à l'aide de « menus » pour les demandes peu volumineuses et urgentes (*figure 5*),
- demandes d'états ou de copies d'extraits de fichiers sur bandes magnétiques en système batch, pour des requêtes relativement peu fréquentes mais concernant un volume de données important.

Les extractions peuvent :

- s'effectuer ponctuellement pour une station définie, soit par son code « agence », soit par son nom, soit par son code hydrologique;
- résulter d'une recherche dans le milieu naturel : cours d'eau défini par son nom, entités hydrologiques (zones, sous-secteurs, secteurs) définies par leur code;
- prendre en compte le découpage administratif : communes, départements.

Les informations extraites peuvent concerner :

- les données relatives à l'identification, la localisation et le suivi d'exploitation (historique) des stations;
- les données hydrométriques : débits ou hauteurs;
- des inventaires de toutes les périodes d'observations ainsi que ceux des seules données disponibles à l'agence.

A toutes ces extractions qui ne peuvent s'effectuer qu'à l'aide de logiciels spécifiques, s'ajoutent les nombreuses possibilités offertes par le système d'information relationnel (IQS de Bull) dont dispose l'agence.

4.3.4.2. Types de données fournies et transmission

Les données peuvent être transmises aux différents services de l'agence ou à de nombreux autres usagers exté-

rieurs : administrations, aménageurs, collectivités locales, instituts de recherches, etc. Toutefois, seul le catalogue des stations se veut exhaustif. Ainsi, l'agence peut toujours, dans le cas où elle ne détient pas les données hydrométriques demandées, fournir des informations relatives à la station concernée (identification, localisation, suivi d'exploitation) et plus particulièrement celles permettant d'identifier le producteur de données que l'utilisateur pourra alors consulter.

Comme les délégations régionales de l'agence, implantées à Marseille, Besançon et Montpellier, en ont la possibilité, il est envisagé que des organismes extérieurs puissent effectuer des interrogations à distance en utilisant les systèmes classiques de télétransmission : réseau transpac, réseau auto-commuté, voire des lignes spécialisées.

Dans un laps de temps relativement court l'agence devrait mettre à la disposition des utilisateurs non seulement des données peu ou semi élaborées mais également un certain nombre de logiciels permettant l'exploitation de ces données.

5. Mise à disposition : l'exemple de la base ARHMA

L'utilisation de données est la principale finalité d'une base de données hydrologiques. La mise à disposition de données n'est pas la partie la plus coûteuse, ni en matériel, ni en temps, mais elle est la plus attrayante pour l'utilisateur. De plus, l'expérience montre que l'utilisation des données produites est, pour le gestionnaire, l'aiguillon le plus efficace pour assurer la qualité des données élémentaires.

Les données mises à disposition de l'utilisateur sont parfois classées en données élémentaires et en données élaborées. En réalité, il n'y a pas de solution de continuité entre la donnée brute et le calcul le plus élaboré. Ainsi la base ARHMA fournit [1] :

- la hauteur d'eau de l'Hérault à Laroque le 17 avril 1974 à 16 h 30 : ce n'est que la fourniture pure et simple d'une donnée rentrée en banque sous forme de « point cerise »;
- la hauteur d'eau le même jour à 16 h 50 (qui résulte déjà d'une interpolation linéaire entre 2 points cerises);
- le débit instantané (ou journalier) à cette station (qui nécessite un calcul faisant intervenir le fichier des courbes d'étalonnage);
- la chronique des débits journaliers maximaux, année par année (qui réalise une sélection parmi les valeurs figurant dans le fichier des débits journaliers);
- l'estimation de la crue journalière décennale ou quinquennale avec les bornes des intervalles de confiance à 90 % (qui est l'information synthétique provisoirement la plus élaborée). On peut aller au-delà : la prise en compte d'autres stations et du champ des crues permettrait d'affiner considérablement l'intervalle de confiance.

Le but de la base de données est donc de fournir aussi bien les données élémentaires que le résultat scientifique ultime, l'utilisateur n'ayant plus que la charge du choix financier ou politique (dans notre exemple : la décision de privilégier l'événement quinquennal ou décennal).

Les questions classiques d'organisation d'un serveur — comment, à qui et dans quel but diffuser des informations — dépendent de contraintes matérielles. Dans le cadre technologique actuel l'existence d'une base de données hydrométriques ouverte repose presque obligatoirement sur la construction suivante :

- implantation des données et des logiciels sur des disques connectés à un ordinateur;
- interrogation en temps partagé à partir de terminaux reliés à l'ordinateur central par le réseau TRANSPAC.
- possibilité d'obtenir des sommes de données par bande magnétique.

Ce dispositif technique implique un nécessaire dispositif contractuel. L'utilisateur utilise l'ordinateur du serveur pour les calculs : il faut donc prévoir qu'il lui rembourse les frais occasionnés par son intervention. Pour reprendre l'exemple de la base ARHMA, le coût du calcul de la crue décennale pour une station est de 12 F, pour 4 stations il est de 20 F; ces montants incluent les frais de calcul et de transmission. La fourniture brute des débits journaliers de 4 stations sur 15 ans coûte environ 500 F, les frais de calcul étant moindres mais les frais de transmission nettement plus élevés.

Les utilisateurs possibles de la base de données sont donc tous ceux qui acceptent, d'une part, de se munir des moyens techniques permettant de se connecter à l'ordinateur serveur et, d'autre part, de signer avec l'organisme-serveur un contrat d'usage.

6. Conclusion

6.1. Situation actuelle

En réseau d'assainissement, certains services ont fait d'énormes efforts pour obtenir les plans détaillés de leur réseau et de leurs ouvrages. Peu d'entre eux sont passés à la numérisation de ces documents. Dans le domaine des mesures proprement dites, il y a très peu de banques de données, car les mesures systématiques en réseaux d'assainissement sont très récentes et les problèmes rencontrés y semblent plus complexes qu'en rivière. Les méthodes artisanales de traitement des données ont atteint la saturation. Beaucoup de services auront à prendre des décisions sur le devenir des points de mesures en réseaux : la seule façon de valoriser l'existant et les données à venir consiste à organiser les résultats dans des banques de données, sinon ce sera l'abandon [4].

En rivière, le problème est très différent. Il y a des expériences de banque de données vieilles de 10 ans et plus. Leur conception doit être périodiquement remise en question.

Jusqu'à présent la démarche d'informatisation est avant tout pragmatique. Cela a donné des résultats jugés le plus souvent satisfaisants dans la mesure où les définitions des données élémentaires sont peu nombreuses, où, leur

quantité reste stable ou figée, et où le nombre d'organismes intervenants est limité.

Dans ces conditions, des banques sont réalisées directement à partir de l'expérience acquise sur les fichiers séquentiels. Cette réalisation tient compte des choix préexistants de matériels et de système de gestion, et repose sur une analyse de la demande.

6.2. Perspectives

Cependant, dans des cas de plus en plus nombreux cette démarche ne convient plus parce que le nombre d'objets manipulés s'accroît et que davantage d'organismes sont concernés. Elle doit laisser la place à une autre, faisant abstraction des contraintes matérielles, ces dernières n'étant prises en compte qu'au niveau du cahier des charges de réalisation.

De toutes manières, avec l'évolution actuelle des techniques informatiques, où il n'y a plus de problèmes de capacité d'archivage, où les matériels et les systèmes de gestion évoluent à un rythme rapide, les créateurs de banques de données ont intérêt à s'appuyer sur des cahiers des charges fort complets et dotés d'une analyse fonctionnelle très précise, grâce à de nouvelles méthodes de conception comme la méthode MERISE développée en France vers 1975-1976 à Aix-en-Provence, sous l'égide du ministère de l'Industrie. Encore peu connue des hydrologues, elle vient d'être utilisée par l'IFREMER pour constituer la base de données du réseau national d'observation des eaux littorales. Comme toutes les méthodes de ce type, elle participe à la confection du dictionnaire des données qui est le garant de la stabilité des définitions et des possibilités d'extension de la banque.

Après une première réflexion très sommaire sur les conséquences qu'entraînerait la généralisation de l'emploi de cette méthode aux problèmes de l'eau, on peut prévoir à terme :

- la réalisation d'un dictionnaire des données dont les définitions serviront de référence pour tous les organismes s'occupant de l'eau, de manière à éviter les ambiguïtés nées des synonymes ou des polysèmes;
- la normalisation des identifiants de tous les objets manipulés;
- la généralisation d'un système de références spatiales normalisées, ce qui facilitera l'expression cartographique automatisée;
- la simplification des langages de communication;
- l'effondrement progressif des cloisons qui séparent les différentes spécialités de l'hydrologie. Si l'on considère le domaine de la qualité des eaux, il semble qu'à terme les banques « eaux de surface », « eaux souterraines », « eaux résiduaires », « eaux littorales », « eaux distribuées » pourront être fusionnées, ou du moins dotées de passerelles pour aller de l'une à l'autre sans problème. L'agence Artois-Picardie a déjà réalisé quelque chose de ce genre, concernant la qualité des eaux souterraines et des eaux de surface, depuis de nombreuses années.

Quel immense progrès, si l'on entrevoit ainsi la fin de l'« hydroschizophrénie », particulièrement forte en France où le cycle de l'eau n'a pas place dans la classification des sciences d'Auguste Comte, ni dans celle du C.N.R.S.

Bibliographie

- [1] COLIN DE VERDIÈRE M., DUTILLET J.L., SCHERER J.C., 1984. — L'information dans la gestion des réseaux de mesures hydrométriques; la chaîne ARHMA et la banque de données hydrométriques du ministère de l'Agriculture. Le *bulletin technique d'information du ministère de l'Agriculture*, n° 387-388, Paris, pp. 135 à 152.
- [2] E.D.F.-D.T.G., 1984. — S.C.H.T.R.O.U.M.P.H. système de collecte horaire et traitement de réseaux par ordinateur utilisé pour la modélisation et la prévision de l'hydraulicité — Fiche d'information D.T.G. n° 80, 4 pages.
- [3] DUBAND D., TOURASSE P., 1987. — Le système S.C.H.T.R.O.U.M.P.H. Collecte des données hydrologiques par le réseau commuté et prévision de débits de crues à l'aide de modèles numériques, de transfert « pluie-débit » et « débit-débit ». *Colloque sur la télémessure et la transmission des données hydrologiques*. Tome 1, O.M.M. Toulouse, pp. 37.1 à 37.16.
- [4] BÉDIOT G., 1986. — Collecte et archivage des données. Rapport général de la session A du colloque « Eau et Informatique : l'informatique dans la gestion des réseaux d'eau potable et d'assainissement ». Ecole Nationale des Ponts et Chaussées, Paris, 26 pages.

Discussion

Président : M. P. HERVIO

M. GHIO rappelle que les serveurs de données sont aussi les délégués de bassins ou les services producteurs. Par ailleurs, il souhaite savoir si l'on a une idée de la structure de la demande. M. SCHERER répond que pour l'EDF, toutes les demandes sont prises en compte. Pour ARMA on satisfait toutes les demandes. Par contre il lui est impossible de dire quelles sont les procédures les plus utilisées.

M. TOURASSE poursuit en indiquant que la demande a sensiblement évolué ces dernières années et qu'elle s'oriente de plus en plus vers des données de faibles pas de temps (débits moyens journaliers, chroniques de hauteurs ou de débits horaires, collection d'épisodes pluies-crues discrétisés à des pas horaires ou pluri-horaires pour pouvoir caler des modèles de prévisions...), ce qui nécessite une approche nouvelle au niveau de la collecte et de l'archivage des données. Par ailleurs, la Division technique générale d'EDF (DTG) s'est fixée comme ligne de conduite de ne pas fournir à l'utilisateur des valeurs trop élaborées en lui laissant l'initiative et la responsabilité de ses résultats. Tout résultat ou étude hydrologique nécessite en effet un retour et une critique des données de base et il nous paraît dangereux, à terme, de laisser croire à l'utilisateur qu'il peut s'en dispenser.

Mais on pourrait développer des procédures dites « hydrologie presse bouton » qui permettraient de fournir aux demandeurs des données élaborées. M. THIRRIOT voudrait se projeter dans le futur et savoir si l'on ne craint pas d'arriver à un phénomène de saturation comme on l'a connu dans d'autres domaines.

Peut-on imaginer que ces données soient disponibles à tout le monde à des fins culturelles ? Comment va-t-on réussir à réintégrer dans les bases de données les données historiques dont la valeur est incontestable si l'on veut savoir si oui ou non il y a des dérives séculaires ?

M. BÉDIOT ne sait pas s'il y aura saturation, ou pas. Il ne faudrait pas que l'automatisation conduite à attacher moins de prix au contrôle et à la validation des données. Il ne sert à rien d'engranger des données si l'on n'est pas capable de passer un minimum de temps pour les contrôler. Pour ce qui concerne les données historiques :

— sur Rhin-Meuse, la reprise des données hydrométriques et pluviogéométriques est totale;

— sur Seine-Normandie, la totalité des données hydrométriques est reprise depuis 1876.

Ces problèmes, vont à nouveau, être examinés dans le cadre des futures banques nationales « hydro » et « pluvio ». Pour ce qui concerne la validation des données, M. TOURASSE abonde dans le sens de M. BÉDIOT. Cette validation doit se faire immédiatement après l'acquisition des données, pas dix ans après. Chez EDF, les données enregistrées sur cassette à pas de temps 6 minutes ont été validées pendant très longtemps. Mais cela représentait un travail considérable, et il a fallu limiter le nombre des stations susceptibles de fournir des informations à pas de temps 6 minutes. Il convient de dire que l'informatique permet une validation plus facile des informations, et permet de motiver le personnel par les possibilités de visualisation graphique des résultats.

M. BAUDELET pose le problème de la sauvegarde informatique. Doit-on sauvegarder en parallèle sur support papier ?

M. TOURASSE répond qu'actuellement on est bien content d'avoir les deux sources d'information.

M. NOUVELOT rejoint tout à fait les propos tenus par MM. BÉDIOT et TOURASSE relativement à la validation et au contrôle des données. A l'Agence de Bassin Rhône Méditerranée-Corse on utilise actuellement des logiciels de contrôle au moment de la saisie, on envisage de développer un contrôle visuel par l'intermédiaire de logiciels graphiques puis ensuite de faire des contrôles « amont-aval » ou même des contrôles d'homogénéité des données. Par ailleurs, M. NOUVELOT souhaite connaître l'identifiant utilisé par EDF pour ses stations.

M. TOURASSE indique que c'est un numéro à 4 chiffres correspondant à un codage interne et ancien des stations par bassin. Ce numéro a été conservé pour des raisons de commodité d'usage car les techniciens de la DTG connaissent de mémoire la plupart des numéros des tableaux de mesure qu'ils ont à traiter. Néanmoins, le code hydrologique de la station est stocké en parallèle dans un fichier de caractéristiques pour toutes les stations dites « d'intérêt général ».

