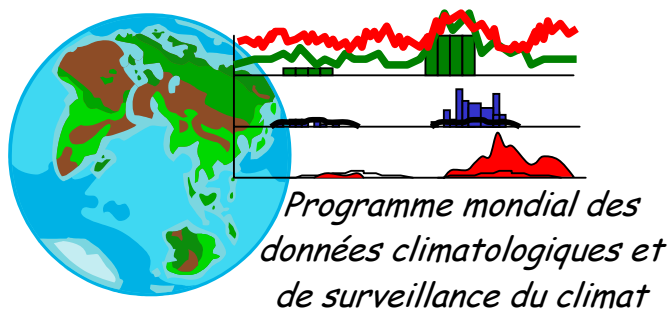


PRINCIPES DIRECTEURS POUR LA GESTION DE DONNEES CLIMATOLOGIQUES

PMDSC n° 60

OMM-DT N° 1376



Organisation météorologique mondiale

(Genève, mars 2007)

(Auteurs : N. Plummer, W. Lipa, S. Palmer, G. Prank, J. Shortridge, D. Stuber)

Révisseurs : Omar Baddour et Hama Kontongomde

©2007 Organisation météorologique mondiale

OMM/DT n° 1376

NOTE

Les appellations employées dans cette publication et la présentation des données qui y figurent n'impliquent, de la part des organismes participants, aucune prise de position quelle qu'elle soit quant au statut juridique des pays, territoires, villes ou zones, ou de leurs autorités, ni quant au tracé de leurs frontières ou limites.

Ce document n'est pas une publication officielle de l'OMM et n'a donc pas été soumis aux procédures éditoriales standard de l'OMM. Les opinions qui y sont exprimées ne nécessitent pas forcément l'approbation de l'OMM.

LA SÉRIE « PRINCIPES DIRECTEURS » DU PMDSC

Reconnaissant combien il est nécessaire pour les Services météorologiques et hydrologiques nationaux (SMHN) d'améliorer leurs services de données climatologiques et de surveillance du climat, la Commission de climatologie (CCI) de l'OMM accorde une priorité de premier plan à la diffusion de principes directeurs aux SMHN.

Le présent document a été préparé par l'Équipe d'experts de la CCI pour la gestion des données climatologiques, faisant partie du Groupe d'action sectoriel ouvert (GASO I) des données climatologiques et de leur gestion, dans le cadre du Programme mondial des données climatologiques et de surveillance du climat (PMDSC). Les principes directeurs faisant l'objet de ce document sont destinés à informer les SMHN sur les meilleures pratiques en vigueur en ce qui concerne la gestion des données climatologiques et à les aider à remplacer d'anciennes bases de données, CLICOM par exemple, par des systèmes présentant une plus grande utilité, sécurité et solidité.

Ce document a été rédigé par un sous-groupe de l'Équipe d'experts de la CCI pour la gestion des données climatologiques et révisé par des intervenants extérieurs. Lors de sa première réunion, qui s'est tenue à Nairobi du 1^{er} au 3 novembre 2006, l'Équipe d'experts, qui avait été créée lors de la quatorzième session de la CCI (Beijing (Chine), 3-10 novembre 2005), a proposé d'effectuer une deuxième révision du document et a fourni quelques mises à jour.

Il convient de rappeler que ce document technique, comme les autres documents techniques publiés dans la série PMDSC de l'OMM, vise à donner des indications sur les meilleures pratiques à adopter par les Membres. Étant donné la diversité des SMHN ainsi que l'étendue et le niveau de développement technologique de chacun d'entre eux, il se peut que ce document ne soit pas très utile à certains Membres en particulier. Il couvre cependant un domaine suffisamment vaste pour apporter tout de même quelques conseils utiles pour tous.

Principes directeurs pour la gestion de données climatologiques

Neil Plummer¹, Wolfgang Lipa², Steve Palmer³, Glenn Prank⁴,
John Shortridge⁵, Denis Stuber⁶

Sommaire

1	INTRODUCTION ET OBJECTIF	1
2	BREF HISTORIQUE DE LA GESTION DES DONNEES CLIMATOLOGIQUES	2
3	GESTION DES DONNEES CLIMATOLOGIQUES : ORGANISATION	3
3.1	Exigences des utilisateurs et besoins à satisfaire en priorité	3
3.2	Systèmes de gestion des données climatologiques : propriétés souhaitables	5
	3.2.1 Modèle de base de données	5
	3.2.2 Principales caractéristiques de saisie	5
	3.2.3 Options d'entrée électronique des données	6
	3.2.4 Éléments liés à la qualité et métadonnées associées	6
	3.2.5 Étendue des contrôles de qualité effectués sur les valeurs observées	6
	3.2.6 Extraction de données	6
3.3	Sécurité	7
3.4	Gestion et surveillance des bases de données	8
3.5	Gestion de la documentation	9
4	ÉLÉMENTS ESSENTIELS DE LA GESTION DES FLUX DE DONNEES CLIMATOLOGIQUES	10
4.1	Enregistrement et gestion des métadonnées	10
4.2	Acquisition, saisie, stockage et archivage des données	11
	4.2.1 Acquisition	11
	4.2.2 Saisie	12
	4.2.3 Stockage et archivage de relevés papier	13
	4.2.4 Stockage et archivage des informations numériques	13
4.3	Gestion des relevés d'origine et sauvetage des données	14
4.4	Assurance et contrôle de la qualité	16
4.5	Échange de données	18
4.6	Accès aux données et développement de produits	19
4.7	Administration et surveillance des données	20
4.8	Gestion des modifications	21
5	TRANSITION VERS UN SYSTEME DE GESTION DE BASE DE DONNEES	22
5.1	Choix d'un système de gestion de bases de données climatologiques : les éléments à prendre en compte	23

¹ Bureau météorologique, Melbourne, Australie

² Institut central de météorologie et de géodynamique (ZAMG), Vienne, Autriche

³ Met Office, Exeter, Royaume-Uni

⁴ Bureau météorologique, Melbourne, Australie

⁵ Bureau météorologique, Melbourne, Australie

⁶ Météo France, Toulouse, France

5.1.1	Prise en compte des besoins	23
5.1.2	Définition du système requis	24
5.1.3	Evolutivité	24
5.1.4	Architecture et technologie	25
5.1.5	Choix du système de gestion de bases de données climatologiques	26
5.2	Considérations relatives à l'architecture de la base de données	26
5.2.1	Considérations relatives à la conception d'une base de données : normalisation	27
5.2.2	Modèles de données utilisés par les systèmes de gestion de bases de données climatologiques	28
5.3	Considérations relatives au logiciel et au matériel informatique	29
5.3.1	Analyse de la situation	30
5.3.2	Définition d'une solution fonctionnelle : évaluation des besoins	30
5.3.3	Définition de la solution technique	33
5.3.4	Aspects liés au changement de système et de services	37
5.4	Transition à partir de CLICOM	38
5.4.1	Niveau de compétences requis	38
5.4.2	Conservation du système CLICOM existant	38
5.4.3	Préparation des métadonnées à importer	38
5.4.4	Préparation des données climatologiques à importer	39
5.4.5	Collecte d'informations sur les réglages du contrôle de la qualité de CLICOM	40
5.4.6	Essais approfondis du nouveau système	41
6	SOUTIEN AUX OPERATIONS DE GESTION DES DONNEES	41
6.1	Ressources nécessaires, notamment en termes de personnel	41
6.2	Formation	42
6.3	Santé et sécurité au travail	42
7	CONCLUSION	42
8	REMERCIEMENTS	43
9	BIBLIOGRAPHIE	44
10	GLOSSAIRE	45
	ANNEXE 1 - DESCRIPTION DES SYSTEMES DE GESTION DE BASES DE DONNEES CLIMATOLOGIQUES PROPOSES	49
	ANNEXE 2 - EXEMPLE DE MISE EN ŒUVRE D'UN SYSTEME DE GESTION DE BASES DE DONNEES CLIMATOLOGIQUES	51
	ANNEXE 3 EXEMPLES DE MATERIEL	54

Introduction et objectif

La série de principes directeurs publiés par le Programme mondial des données climatologiques et de surveillance du climat (PMDSC) et la Commission de climatologie (CCI) fournit des informations et des conseils sur l'organisation et la mise en œuvre de services climatologiques dans des domaines importants pour les Services météorologiques et hydrologiques nationaux (SMHN). Ces principes directeurs sont également destinés à présenter des processus et des solutions technologiques répondant à la situation et aux besoins spécifiques des petits SMHN, qui ne disposent souvent que de ressources limitées. Dans la mesure du possible, ces principes directeurs doivent également :

- s'appliquer à tous les autres programmes de l'OMM ;
- s'adapter à toute une série de besoins ;
- être formulés de manière à pouvoir être modifiés ; et
- être conformes et complémentaires aux principes décrits dans le Guide des pratiques climatologiques.

Ces principes directeurs sont destinés à informer les SMHN sur les meilleures pratiques en vigueur en ce qui concerne la gestion des données climatologiques, informations qui arrivent à point nommé, au moment où de nombreux pays se voient obligés de remplacer leurs anciennes bases de données, CLICOM par exemple, par des systèmes beaucoup plus utiles, plus sûrs et plus solides.

Compte tenu des objectifs généraux de ces principes directeurs, on s'attachera à renforcer les connaissances et les capacités existantes sur les points suivants :

- phases et étapes de la gestion de données numérisées ;
- technologies de bases de données disponibles et choix d'une base de données appropriée ;
- transition vers un système moderne de base de données climatologiques ;
- soutien aux opérations de gestion des données.

On s'attachera délibérément, dans la gestion des données climatologiques, aux aspects qui présentent un intérêt pour les SMHN souhaitant passer à un système plus moderne de gestion des données climatologiques et, ce qui est tout aussi important, aux compétences, systèmes et processus qui doivent être mis en place pour assurer le bon déroulement des opérations. Ces principes directeurs ne s'adressent pas aux SMHN qui envisagent de construire une base de données entièrement nouvelle, mais seront davantage utiles à ceux qui souhaitent étudier et comparer les systèmes existants, candidats pour ce type d'utilisation.

Ces principes directeurs ne font pas mention en détail des pratiques et des normes relatives à la gestion des données car ces dernières figurent déjà dans d'autres documents de l'Organisation météorologique mondiale (OMM), par exemple dans le *Guide des pratiques climatologiques* (OMM 2005) et autres documents (exemple : WMO 2003a, 2003b, 2004). Pour en savoir plus sur les aspects techniques des bases de données, voir les textes spécifiques correspondants (par exemple : Date 1999). Les publications sur le PMSDC (voir <http://web/wcp/wcdmp/html/wcdmpreplist.html>) sont également des sources extrêmement précieuses.

Même s'il n'est pas nécessaire d'avoir des connaissances approfondies de la gestion des données climatologiques, il est probable que ces principes directeurs seront davantage profitables aux personnes qui auront une connaissance pratique du sujet.

Le chapitre 2 présente brièvement l'histoire de la gestion des données climatologiques pour les SMHN. Le chapitre 3 expose les conditions et l'organisation de la gestion des données climatologiques, c'est-à-dire des aspects d'ordre plus général que les gestionnaires de données doivent connaître. Une description plus détaillée du sujet se trouve au chapitre 4, qui porte sur la gestion des flux de données. Le chapitre 5 aborde la transition entre les systèmes de gestion de bases de données, notamment entre le système CLICOM et un système moderne. Le chapitre 6 est consacré aux aspects clés du soutien aux opérations de gestion des données, puis vient la conclusion au chapitre 7.

Bref historique de la gestion des données climatologiques

Depuis des siècles, on note la présence d'informations sur le temps qu'il fait sous forme manuscrite. Les premiers documents faisaient référence à des événements extrêmes, parfois à des catastrophes et à des phénomènes tels que les périodes de gel et de fonte des rivières, des lacs et des mers, données de premier plan de nos jours face aux préoccupations soulevées par les changements climatiques.

Des relevés ou journaux consacrés spécialement à la collecte et à la conservation d'informations climatologiques sont utilisés depuis deux ou trois siècles (OMM 2005). Le développement d'instruments capables de quantifier les phénomènes météorologiques ainsi que tout le soin que mettent les observateurs à conserver des archives méthodiques, fiables et bien documentées, ont ouvert la voie à une gestion organisée des données climatologiques. Depuis les années 1940, on utilise de plus en plus de formes et procédures normalisées, et à partir du moment où les SMHN ont commencé à utiliser des systèmes informatiques, ces formes normalisées ont grandement facilité la saisie de données et, par conséquent, le développement d'archives de données informatisées.

La fin du vingtième siècle a vu s'imposer l'échange régulier de données météorologiques sous forme numérique. Un grand nombre de centres de données météorologiques et de données connexes ont saisi l'occasion d'enregistrer et de stocker directement ces données dans leurs bases. À la fin des années 1950, période au cours de laquelle s'est tenue l'Année géophysique internationale et qui a vu la naissance du programme de Veille météorologique mondiale, on a beaucoup appris sur les méthodes automatiques de collecte et de traitement des données météorologiques. L'élaboration par l'OMM de principes directeurs et de normes internationales sur la gestion et l'échange des données climatologiques a permis aux SMHN d'organiser leurs activités de gestion des données et a indirectement favorisé le développement de bases de données régionales et mondiales. Aujourd'hui, la gestion des relevés climatologiques nécessite une approche systématique, rassemblant les relevés effectués sur papier, sur microfilms et microfiches et les relevés numériques, ces derniers comprenant des fichiers d'images ainsi que des données alphanumériques classiques.

Les ordinateurs ont remplacé les appareils mécaniques qui jouaient un rôle important dans le développement de la gestion des données. Les calculs étaient réalisés à l'aide d'additionneurs à clavier, par exemple, et les résultats étaient enregistrés sur papier. Un grand pas en avant a été franchi avec l'introduction du système Hollerith de cartes perforées, trieuses et tabulatrices. Ces cartes, munies d'une série de trous qui enregistraient les valeurs des variables météorologiques, étaient insérées dans les trieuses et tabulatrices, donnant ainsi des calculs statistiques plus efficaces.

Dans les années 1960 et 1970, plusieurs SMHN ont commencé à utiliser des ordinateurs et à transférer progressivement sur bande magnétique les informations contenues sur des millions de cartes perforées. Ces ordinateurs ont été remplacés par des gros systèmes informatiques de plus en plus puissants, puis les données sont devenues disponibles en ligne, grâce au développement de disques.

Une étape majeure a été franchie avec le projet CLICOM du PMDSC. Ce projet, mis en place en 1985, a entraîné l'installation d'un logiciel de bases de données climatologiques pour PC, accompagné du matériel nécessaire et d'un programme de formation complet, dans plus de 100 SMHN de par le monde. Ce projet a également été à l'origine d'améliorations notoires des services, des applications et des activités de recherche climatologique dans ces pays. À la fin des années 1990, le PMDSC a mis en place un système de gestion des bases de données climatologiques, afin de profiter des toutes dernières technologies visant à répondre aux besoins variés et évolutifs des Membres en matière de gestion de données. Bien que parfois plus onéreux et plus complexes d'un point de vue technologique (nécessitant donc davantage de compétences spécialisées pour les administrer), les nouveaux systèmes de gestion des bases de données climatologiques permettent aux utilisateurs d'accéder plus facilement aux données, qui sont ainsi plus sûres et plus utiles. Les bases de données relationnelles avaient déjà fait leurs preuves dans plusieurs SMHN.

En dehors des progrès réalisés par les technologies de bases de données, la saisie de données est devenue plus efficace entre la moitié et la fin des années 1990, avec une augmentation des stations météorologiques automatiques et des carnets de notes électroniques (ordinateurs portables fixes utilisés pour la saisie, le contrôle de la qualité et la transmission des observations), puis avec l'arrivée de l'Internet et d'autres technologies nouvelles.

Il n'est pas surprenant d'observer un certain nombre de tendances suggérant beaucoup d'autres avantages pour les SMHN en ce qui concerne la gestion des données et les services fournis aux clients. L'Internet améliore déjà considérablement les capacités d'accès aux données et, dans la mesure où les questions de sécurité sont bien maîtrisées, on peut s'attendre à ce que cela crée des opportunités majeures pour les gestionnaires de données dans les cinq à dix prochaines années. En plus, les systèmes de bases de données relationnelles à code source ouvert⁷ pourront également supprimer les obstacles liés aux coûts des bases de données relationnelles pour un grand nombre de SMHN.

Gestion des données climatologiques : organisation

Exigences des utilisateurs et besoins à satisfaire en priorité

Il est indispensable de tenir compte des besoins des utilisateurs actuels et, dans la mesure où ils sont prévisibles, des besoins des futurs utilisateurs, à la fois dans le développement de bases de données climatologiques et dans la mise en œuvre de pratiques de gestion des données. Même si cela semble aller de soi, il peut toujours arriver, par exemple, de développer des structures qui omettent des données importantes pour une application utile, ou bien qu'un centre de données consacre trop peu de ses ressources au contrôle de la qualité des données pour lesquelles les utilisateurs demandent précisément un haut niveau de qualité. Dans la mesure du possible, il est recommandé de stocker toutes les mesures dans le système de gestion des bases de données climatologiques, même si elles ne sont pas utilisées tout de suite.

Pour tout nouveau développement, les gestionnaires de données doivent soit compter au moins un utilisateur de données clé dans leur équipe de projet, soit mettre en place des consultations régulières auprès d'un groupe de partenaires utilisateurs. Les fournisseurs ou utilisateurs de données faisant partie eux-mêmes de l'organisation peuvent également consulter régulièrement les utilisateurs finaux de données (ou d'informations) climatologiques, et il appartient aux

⁷ Un système à code source ouvert est un logiciel dont le code source est librement accessible et peut donc être vu et modifié. Cela signifie que le logiciel est disponible gratuitement (ou parfois seulement au prix de distribution).

gestionnaires de données de se tenir au courant à la fois de l'évolution des besoins et de toute question préoccupant les utilisateurs. En termes simples, la gestion de données nécessite de connaître les besoins des utilisateurs finaux.

À l'heure actuelle, les principaux secteurs demandeurs de données sont les suivants : services liés aux prévisions du climat et aux changements climatiques, agriculture et secteur primaire, santé, gestion des catastrophes / des urgences, énergie, gestion des ressources naturelles (y compris de l'eau), développement durable, urbanisme, finance et assurances. Les gestionnaires de données doivent savoir que l'existence de la fonction de gestion de données dépend du centre qui assure des avantages sociaux, économiques et environnementaux aux communautés d'utilisateurs visées. Il est donc important que le gestionnaire de données encourage et, dans la mesure du possible, participe à des projets qui démontrent la valeur de ses ressources en données. Pour rappeler aux dirigeants des SMHN et convaincre les organismes de financement qu'il est rentable d'investir dans des données, il peut être utile de leur présenter des études montrant, par exemple, les avantages économiques des prévisions du climat ou les avantages sociaux découlant d'une utilisation de données climatologiques dans un système de veille sanitaire. Les données ont encore davantage de valeur lorsqu'elles sont intégrées dans des modèles d'application (exemple : modèles de simulation de culture, modèles économiques). Il convient donc de tenir compte d'aspects liés à l'intégration des données lors de la conception de nouvelles structures de données.

On ne pourra satisfaire pleinement les besoins des utilisateurs que si les structures et les responsabilités nécessaires sont en place. Par exemple, l'organisation devra être ouverte aux avis externes et disposer de moyens de communication efficaces et rentables pour assurer l'échange d'informations en interne.

Les gestionnaires de données doivent veiller également à respecter les principes d'une surveillance du climat à long terme, comme indiqué dans le document GCOS (2003), approuvé par la Commission de climatologie (CCI) et adopté par la Convention-Cadre des Nations Unies sur les changements climatiques. La capacité à assurer la continuité, l'homogénéité et enfin la qualité des données climatologiques dépend en grande partie de la qualité de gestion des réseaux et systèmes d'observation. Pour suivre les dix principes de surveillance du climat (voir WMO (2003a), expliquant dans quelle mesure cela s'applique aux activités d'un SMHN), les gestionnaires de données doivent absolument établir d'étroites relations avec les gestionnaires d'observations et les utilisateurs de données climatologiques. Le gestionnaire de données devra être un intermédiaire efficace entre le gestionnaire d'observations et l'utilisateur de données, en présentant des exigences au premier et en faisant correspondre les besoins du second avec ce qui est économiquement et techniquement faisable. Le document WMO (2003a) expose ces principes plus en détail, au niveau des réseaux et systèmes d'observation du climat.

Pour répondre aux besoins des utilisateurs, il convient de considérer notamment les aspects clés suivants, qui permettent de classer les observations nouvelles ou supplémentaires par ordre de priorité :

- priorités sociales, économiques et environnementales nationales ;
- régions déficitaires en données ;
- paramètres faisant l'objet d'observations insuffisantes ;
- régions particulièrement sensibles aux changements ;
- mesures présentant une résolution temporelle peu satisfaisante.

Systemes de gestion des données climatologiques : propriétés souhaitables

3.2.1 Modèle de base de données

Toute base de données climatologiques reposera sur un modèle de données sous-jacent, très important pour la qualité du système qui sera développé, en particulier pour sa maintenabilité (rappelons que plus de 50% des efforts mis en œuvre pour l'élaboration d'un logiciel sont consacrés à la maintenance et à l'amélioration du système au fil des années). Si le modèle n'est pas adapté, le système sera plus difficile à maintenir. En général, une base de données conçue en priorité pour des travaux météorologiques en temps quasi-réel sera principalement accessible de manière « synoptique », puisque les extractions répondront généralement au besoin de « rassembler toutes les données de quelques endroits ou zones données, ceci pendant une période définie et relativement courte ». En revanche, les applications climatologiques impliquent généralement l'extraction de données d'une ou plusieurs stations, sur une longue période. Ainsi, une approche plus vaste du stockage de données climatologiques implique de stocker ensemble (pour des données quotidiennes) toutes les données liées à une station donnée et à un jour précis. On peut procéder de la même façon pour des données horaires et mensuelles. Une autre approche consiste à stocker séparément des données de type différent, avec un tableau (dans le cas de bases de données relationnelles) contenant toutes les données de température de l'air pour toutes les stations et toutes les dates, etc. Voir paragraphe 5.2 pour plus de détails. Outre la qualité naturelle de la conception, les autres critères à prendre en compte sont notamment la façon dont le modèle de données est documenté et sa facilité d'extension par des programmeurs. L'une des exigences clés d'un modèle de données est qu'il doit être capable de représenter les données transmises par tous les formats de messages standards de l'OMM : TEMP, PILOT, METAR/SPECI, CLIMAT, CLIMAT_TEMP et SYNOP.

Des considérations similaires s'appliquent au « modèle de métadonnées ». Dans un sens, il s'agit d'un domaine plus complexe car les structures de métadonnées peuvent être plus complexes que les structures de données. Il conviendra de définir en temps utile un modèle standard de métadonnées⁸ pour les données climatologiques (élargissant les travaux déjà réalisés sur la norme de l'OMM relative aux métadonnées de base). À ce moment-là, il sera souhaitable que la capacité d'intégration dans ce modèle de données soit un critère de sélection pour les systèmes climatiques.

3.2.2 Principales caractéristiques de saisie

La qualité fondamentale d'un système de saisie de données est de ne pas introduire de défauts évitables dans le processus de saisie. Le personnel chargé de ce travail est généralement très compétent en la matière et mérite qu'on lui assure des conditions de travail productives à tout moment. Cela signifie que le système de saisie de données ne doit pas comporter de défauts agaçants, pouvant ralentir l'opérateur de saisie. De manière idéale, les formes affichées à l'écran pourront être personnalisées de manière à optimiser la saisie de données. Le système doit également, dans la mesure du possible, valider les données à mesure de leur saisie, c'est-à-dire repérer les erreurs et, si possible, proposer des valeurs de remplacement. Il est possible aussi de définir des valeurs par défaut pour certains éléments, ce qui évite des combinaisons de touches inutiles. L'autre caractéristique utile est la double saisie, qui permet de limiter les erreurs (toutes les données sont saisies deux fois, celles saisies par le deuxième opérateur étant automatiquement comparées à celles saisies par le premier).

⁸ Dans ce contexte, les métadonnées font davantage référence aux informations relatives aux données climatologiques (exemple : qualité des données, méthode de dérivation, etc.) qu'à l'historique des informations enregistrées par une station.

3.2.3 Options d'entrée électronique des données

Comme mentionné précédemment, il est souhaitable que le système de gestion des bases de données climatologiques soit en mesure de représenter tout le contenu des messages de format standard de l'OMM (SYNOP, CLIMAT, etc.). Un autre avantage serait qu'il puisse décoder des messages directement dans la base de données climatologiques. Dans certains cas (notamment pour les messages mensuels CLIMAT et CLIMAT TEMP), il est également souhaitable d'avoir des systèmes capables de générer ce type de messages à partir des informations contenues dans la base de données. Parmi les autres options d'entrée électronique utiles, on note la capacité d'entrer des données provenant des stations météorologiques automatiques et la capacité d'incorporer des données exportées du système CLICOM. En général, cette dernière option devrait être réalisée en une seule fois, au moment du changement de système (voir paragraphe 5.4).

3.2.4 Éléments liés à la qualité et métadonnées associées

Une donnée météorologique doit être accompagnée de toute une série d'informations sur sa qualité : qualité attribuée à la donnée en elle-même (commentaires du type « apparemment correcte », « suspecte », « incohérente avec d'autres données », etc.), nature de l'essai ou des essais réalisés pour générer l'indicateur de qualité, etc. Ces informations relatives à la qualité constituent un large éventail de métadonnées, portant, par exemple, sur l'instrument utilisé pour l'observation, sur la piste d'audit des observations faisant état de tous les changements opérés depuis la première saisie des données, sur des informations complètes concernant le site, sur les programmes d'observation en vigueur, etc. À plus long terme, le système sera pleinement en mesure de représenter tout l'historique de la station.

3.2.5 Étendue des contrôles de qualité effectués sur les valeurs observées

Les contrôles appliqués pour déterminer la qualité d'une observation peuvent aller du plus simple au plus complexe. Ces contrôles portent notamment sur les points suivants, présentés ici approximativement par ordre croissant de complexité :

- contrôles syntaxiques (ex : la température de l'air doit être exprimée par une valeur ayant au plus un chiffre après la virgule) ;
- plages numériques (ex : la température doit se situer entre -90 et +70) ;
- contrôles de l'éventail de climats (ex : est-ce que les données sont cohérentes avec la climatologie ?) ;
- cohérence entre les relevés (ex : la température de l'air ne doit pas être inférieure au point de rosée) ;
- cohérence des séries chronologiques (ex : la différence entre deux températures observées successivement sur le même site doit être « plausible ») ;
- cohérence spatiale (ex : les limites d'une différence plausible entre les températures d'une station et de ses environs ne doivent pas être dépassées).

3.2.6 Extraction de données

Un aspect important d'un système de gestion des bases de données climatologiques est à la fois l'étendue et la puissance des fonctions d'extraction et d'analyse des données. De manière idéale, les données peuvent être extraites via une interface graphique utilisateur ou bien via une interface en ligne de commande, selon le cas. Il sera préférable de fournir une interface graphique pour l'immense majorité des utilisateurs, les fonctions liées au langage d'interrogation étant réservées à une poignée d'utilisateurs plus avancés qui ont besoin d'effectuer des extractions non standard. Les utilisateurs qui en sont capables doivent avoir la possibilité de spécifier leurs propres critères d'extraction (généralement via l'interface

graphique), et le système doit être correctement documenté afin de leur donner le maximum d'informations pour ce faire.

Le système doit également proposer un large éventail d'options permettant de déterminer la structure de l'extraction et la présentation des résultats et de spécifier une extraction grâce à une personnalisation des stations, de l'horaire et des détails de présentation. Ces options doivent permettre d'accéder à des listes de données, de résumés tabulaires, d'analyses statistiques (simples et complexes) et de présentations graphiques.

Sécurité

La politique de sécurité et les activités qui s'y rattachent ont pour principal objectif d'éviter la perte ou la détérioration du système de gestion des bases de données climatologiques et de conserver les fonctions de gestion des données dans le meilleur état possible. Pour ce faire, les conditions préalables sont les suivantes :

- le système de gestion des bases de données climatologiques doit se trouver dans un bâtiment abrité et protégé ;
- l'ensemble du personnel doit avoir pleinement connaissance de sa responsabilité professionnelle et de son devoir de veiller au système ;
- les archives et l'environnement des bases de données doivent être sécurisés et protégés contre les incendies, l'humidité, etc. ;
- les manipulations de données (ex : insertion, mise à jour, suppression) ne peuvent être réalisées que par des applications protégées, auxquelles a accès un groupe restreint de personnes ; les personnes qui ont un accès en écriture à la base de données doivent s'engager à ne procéder à aucune transaction en dehors des opérations et pratiques approuvées par le gestionnaire de données ; tout changement apporté aux tables de données doit faire l'objet d'une piste d'audit dont l'accès doit être réglementé ;
- les mots de passe ne doivent être ni communiqués ni écrits où que ce soit ; ils doivent être changés régulièrement et cela s'applique à tous les utilisateurs, de l'administrateur de la base de données à l'utilisateur travaillant sur les applications de manipulation des données ; le mot de passe se compose de lettres et de chiffres qui n'ont apparemment aucun lien entre eux, mais ils peuvent tout à fait être la suite des premières lettres de chacun des mots d'une phrase, ex : je suis assis dans un bureau au 9^{ème} étage -> jsaduba9e ;
- le système de bases de données d'archive doit fonctionner derrière un pare-feu réglé au niveau maximal de sécurité ;
- tous les services annexes des systèmes informatiques de bases de données, tels que la messagerie électronique, doivent être désactivés ; le système informatique des bases de données ne doit utiliser que le noyau du système d'exploitation et le système de gestion des bases de données ;
- l'ordinateur client sur lequel se trouve la base de données doit également être protégé contre les virus et les actes de piratage ;
- des sauvegardes de secours doivent être effectuées à une fréquence suffisante pour ne pas descendre en dessous du niveau maximal autorisé de perte de données en cas de défaillance ; en général, des sauvegardes partielles seront effectués tous les jours et des sauvegardes complètes une fois par semaine ;
- les archives moins fréquentes (généralement mensuelles) du contenu des tables de données doivent être au format ASCII et placées en un lieu sûr et protégé contre les incendies, différent du lieu où se trouve physiquement la base de données climatologiques ;
- un plan de récupération doit être défini, indiquant les sauvegardes et les archives qui seront utilisées en situation d'urgence pour récupérer la base de données ; ce plan envisagera autant de scénarios catastrophes que l'on peut imaginer ;
- le plan de récupération doit être renouvelé et révisé à intervalles réguliers ;

- des sauvegardes de secours du système de gestion des bases de données climatologiques doivent être effectuées avant toute modification du logiciel système, du concept ou des applications contenues dans le système ;
- on procédera, à des intervalles appropriés, à un audit des archives d'origine papier et des archives au format ASCII.

Gestion et surveillance des bases de données

La gestion de base de données vise à garantir l'intégrité de la base de données à tout moment et à assurer que celle-ci contient toutes les données et métadonnées nécessaires pour réaliser les objectifs de l'organisation, au moment présent et à l'avenir. Quant à l'activité de surveillance des bases de données, elle fait appel à des indicateurs de performances pour évaluer comment la base de données et les processus qui l'utilisent se comportent par rapport aux objectifs.

Les données sont utilisées par un certain nombre de processus (voir paragraphe 3.5) : maintenance des métadonnées, alimentation de la base de données, mesures de contrôle de la qualité modifiant la base de données, extraction pour utilisateurs et clients finaux. Chacun de ces processus sera géré par un responsable de processus, chargé du processus en lui-même ainsi que de la surveillance, de l'évaluation et de l'amélioration de ce dernier. Bien évidemment, dans les petits SMHN, toutes ces tâches font partie d'un même processus et, dans les très petits SMHN, la plupart de ces responsabilités sont assumées par une seule et même personne. Il est toutefois utile de bien comprendre chacun de ces rôles et responsabilités et de distinguer les mesures correspondantes. Selon l'un des principes de gestion des risques, lorsqu'un utilisateur réalise une opération quelle qu'elle soit sur la base de données, il doit posséder, dans la mesure du possible, toutes les autorisations nécessaires pour ce faire, et rien de plus.

Les rapports de surveillance comprendront généralement le numéro et le type de la station qui figurent dans la base de données, ainsi que la quantité de données de la base, répartie par station et par type d'élément d'observation. Il sera utile de comparer ces données à celles du programme d'observation, d'identifier les endroits où il en manque – puis on prendra les mesures nécessaires pour en définir les raisons et trouver des solutions. D'autres rapports pourront examiner les mesures de contrôle de la qualité à mettre en oeuvre, vérifier que le contrôle de la qualité (CQ) est en phase avec le flux de nouvelles données et recenser tout regroupement de données faisant l'objet d'un CQ souvent modifié ; ici aussi, des mesures devront suivre pour résoudre les problèmes.

Il est très utile de garder une trace de la quantité et de l'étendue des données extraites suite à des requêtes d'utilisateurs finaux, car on sait ainsi quels sont les jeux de données les plus importants, on peut justifier l'investissement réalisé dans les données climatologiques et on a une idée des domaines à développer à l'avenir. Quand le recouvrement des coûts est appliqué, il est indispensable de connaître les recettes générées par les services utilisateurs des données climatologiques.

La surveillance implique de mesurer des indicateurs objectivement vérifiables (IOV) appropriés, et à chaque IOV est associé un moyen de vérification (MV). Le MV donne une valeur cible à l'IOV, qui mesure les performances par rapport aux besoins. Un MV pourrait être, par exemple, que 90% des rapports mensuels de stations uniquement pluviométriques, présentés sous forme papier, soient disponibles dans la base de données au 15 du mois suivant. Notons qu'il s'agirait là d'un objectif SMART (Spécifique, Mesurable, Atteignable, Réaliste et limité dans le Temps). L'IOV correspondant comparerait alors le nombre de rapports de ce type, présents dans la base de données au 15 du mois, à la valeur attendue, exprimé en pourcentage.

La fréquence ainsi que la période d'établissement des rapports de surveillance dépendra des besoins de l'organisation. S'agissant de l'absorption de données, certains rapports peuvent être

généérés automatiquement chaque jour – pour cela, un rapport journalier de réussites ou échecs est indispensable pour prendre des mesures correctives avant le prochain cycle d'absorption. Beaucoup de données climatologiques suivant un cycle d'un mois calendaire, il conviendrait d'établir des rapports mensuels sur la quantité et la qualité d'absorption des données. Pour procéder à une analyse de recouvrement des coûts et de rendement, il est nécessaire d'avoir des rapports établis au moins une fois par exercice, sur le nombre d'utilisateurs finaux, les recettes générées, la quantité et le type de données impliquées. Les informations nécessaires pour ce faire ne font pas forcément l'objet d'une fonction d'un système de gestion des bases de données climatologiques, mais ce dernier pourrait comprendre des statistiques sur l'utilisation.

Des rapports de surveillance doivent être établis pour assurer un suivi actif du flux d'informations, de métadonnées et de données dans le système de données climatologiques de l'organisation.

Dans le planning de gestion, il convient également de prendre en considération la longue durée de vie des données climatologiques, pour que ces ressources demeurent accessibles aux utilisateurs de demain. Les aspects dont il faudra tenir compte sont notamment la planification de la succession et la formation du personnel, ainsi que le cycle de remplacement du matériel, les coûts de maintenance et de mise à niveau, le support et la formation liés aux logiciels commerciaux.

Gestion de la documentation

Il est indispensable d'établir une documentation sur les processus de gestion et d'utilisation de la base de données, à la fois pour en archiver le concept et pour que cette documentation serve d'instructions d'utilisation et de principes directeurs pour les gestionnaires, utilisateurs et développeurs de la base de données et des applications connexes. Pour les organisations appliquant un système de gestion de la qualité ISO 9000, cette documentation est obligatoire et doit être conforme aux spécifications. Pour les autres organisations, il est fortement recommandé d'établir une documentation suivant les principes de la norme ISO 9000.

Les principes de l'ISO 9000 se résument ainsi :

- dire ce que l'on fait ;
- faire ce que l'on dit ;
- vérifier et surveiller que l'on fait ce que l'on dit ; et
- évaluer et améliorer.

La documentation doit être rédigée dans un style positif afin d'encourager les utilisateurs du processus à faire le maximum.

Pour la gestion et la documentation, il convient de mettre en place une structure à plusieurs processus selon laquelle chaque processus comporte un certain nombre de données d'entrée. Une série d'opérations est alors effectuée sur ces données d'entrée, donnant lieu ensuite à des résultats pour un « client » donné. Ce « client » peut être interne ou externe à l'organisation. Par exemple, le processus de saisie peut comprendre l'enregistrement de formulaires papier, la fourniture de ces formulaires aux opérateurs de saisie, la saisie des données dans le formulaire de la base de données, le chargement du formulaire une fois rempli dans la base de données et la résolution des erreurs, puis l'enregistrement des formulaires papier dans les archives. Les données d'entrée sont les formulaires papier, le personnel et les ressources informatiques. Les résultats quant à eux seraient les données se trouvant dans la base de données et les formulaires papier enregistrés dans les archives. Les clients seraient les utilisateurs de la base de données, y compris d'autres processus comme le contrôle de la qualité (CQ).

Chaque processus doit être attribué à un responsable de processus, chargé de la documentation et de la gestion et de la surveillance globales du processus, qui décide des

modifications à apporter au processus en lui-même et à la documentation et les autorise. Une documentation qui n'est plus à jour ou qui est erronée peut détruire la base de données. De même, si les actions spécifiées dans la documentation ne sont pas effectuées, ceci peut entraîner la perte de la base de données.

Le responsable de processus doit également faire en sorte que le processus soit régulièrement évalué et amélioré afin de s'assurer qu'il réponde toujours aux objectifs de l'organisation.

Éléments essentiels de la gestion des flux de données climatologiques

Enregistrement et gestion des métadonnées

Pour que les données météorologiques soient utiles à de futurs utilisateurs, il est indispensable qu'elles soient accompagnées des métadonnées adéquates. Celles-ci doivent préciser, dans l'idéal, les endroits où ont été réalisées les mesures, par qui et avec quels instruments, le niveau de qualité exigé des données, etc. Le traitement des métadonnées spécifiques à chaque station est décrit en détail dans le rapport intitulé « Guidelines on Climate Metadata and Homogenization » (WMO 2003b) (Principes directeurs concernant les métadonnées climatologiques et l'homogénéisation), qui fait partie des publications du Programme mondial des données climatologiques et de surveillance du climat (PMDSC). Le lecteur peut se reporter à cette publication pour avoir une vision globale de la question, alors que le présent document tente de donner un rapide aperçu des questions relatives aux systèmes.

De manière générale, dans un système idéal, la structure innée des métadonnées sera beaucoup plus complexe que celle des données climatologiques en elles-mêmes. Prenons par exemple le cas d'observations de précipitations. L'essentiel du contenu des données sera simplement du style : « à la station x, les précipitations étaient de p mm pendant une durée e jusqu'à un instant t ». Les métadonnées correspondantes, nécessaires pour interpréter l'intégralité des données, seront par exemple :

- la date de référence utilisée par la base de données (GMT, fuseau horaire, etc.) ;
- le niveau de qualité affecté à l'observation ;
- l'historique des valeurs affectées au paramètre météorologique et tout autre indicateur associé ;
- l'instrument utilisé pour enregistrer l'observation, avec des détails plus précis sur son propre programme de maintenance, ses tolérances, ses paramètres internes, etc. ;
- le nom de l'observateur ;
- des détails complets sur la station et ses archives ;
- le programme d'observation en vigueur et ses archives ;
- l'inventaire des éléments stockés dans la base de données, leurs unités, leurs limites ; et
- des détails sur la topographie et la couverture du sol du site, informations sur la végétation, les constructions environnantes, etc, et sur leur évolution au cours du temps.

L'enregistrement d'informations à ce niveau de détail dans un système informatique nécessite une structure de données complexe et donc un logiciel tout aussi complexe, qui doit être capable de stocker des graphiques (plans de sites, photos, éventuellement images de documents anciens, etc.). Le principe en lui-même n'est pas compliqué, mais il a des répercussions en particulier sur le matériel permettant d'acquérir et de reproduire les images, ainsi que sur le logiciel à utiliser comme système de gestion de base de données.

Il est également nécessaire de fournir des métadonnées au niveau du jeu de données, mais ces détails dépassent le cadre du présent document.

Les données peuvent également faire l'objet d'un travail d'homogénéisation, qui consiste à supprimer d'un jeu de données toutes les influences autres que celles du climat sous-jacent. Les données soumises à ce type d'exercice ont besoin de métadonnées supplémentaires exprimant précisément la nature du travail d'homogénéisation. Les jeux de données homogénéisés devraient être considérés comme des produits du système de gestion des bases de données climatologiques et non pas comme une substitution.

On dispose actuellement d'outils de plus en plus perfectionnés pour accéder aux métadonnées et les gérer (Figure 1).

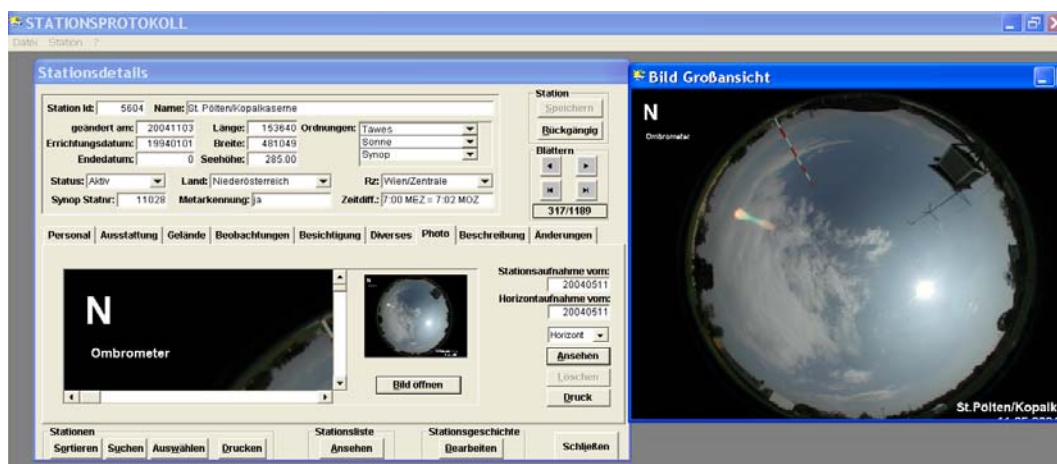


Figure 1. Logiciel de gestion de métadonnées STATPROT de l'Institut central de météorologie et de géodynamique (ZAMG), Autriche

Acquisition, saisie, stockage et archivage des données

4.2.1 Acquisition

L'une des tâches principales d'un SMHN est de faire fonctionner un réseau d'observation du climat comme indiqué dans le document WMO 2003a. Le fonctionnement de ce type de réseau, constituant la force vitale d'un SMHN, a un coût pour l'organisation et n'est généralement pas très rentable dans un premier temps. Mais les SMHN pourraient réduire ces frais généraux en utilisant les réseaux existants d'organisations comme les universités, services publics locaux ou entreprises privées. Les SMHN devraient ainsi recueillir toutes les données climatologiques qui les intéressent, et il serait souhaitable que ces organisations les autorisent à utiliser l'ensemble des données climatologiques qui s'y trouvent, sans aucune restriction, comme s'il s'agissait des leurs. Pour ce faire, des contrats et des protocoles d'accords devront éventuellement être conclus et signés entre les directions générales des organisations concernées.

Il est possible aussi de gérer et de se procurer, pour des besoins commerciaux, des données et informations non conventionnelles, provenant par exemple des nombreuses caméras web appartenant à l'État ou à des personnes privées. Les médias : presse, radio et télévision, sont également des sources d'informations utiles. Si les ressources le permettent, il serait intéressant de réunir tous les rapports et articles portant sur des catastrophes naturelles et autres événements météorologiques dans un système de gestion documentaire, lui-même intégré à une base de données climatologiques élargie.

4.2.2 Saisie

Les données doivent être collectées le plus près possible de leur source. Les stations météorologiques automatiques, y compris celles qui procèdent parfois à des observations manuelles, ainsi que les stations météorologiques partiellement automatiques, doivent collecter leurs données climatologiques ainsi que les messages d'erreur sur site et les transmettre par voie électronique aux systèmes de gestion des bases de données climatologiques, si possible via un autre système de base de données. Les données observées manuellement doivent être collectées et enregistrées sur site, puis transmises dès que possible au système de gestion des bases de données climatologiques. Dans le cas d'une station manuelle, les données doivent être collectées et enregistrées sur site, et il est fortement recommandé de procéder à leur transmission numérique au système de gestion des bases de données climatologiques pendant le jour. Il n'est certes pas toujours possible d'obtenir en une journée toutes les données stockées dans une base de données (par exemple, les retours mensuels des archives papier), mais une collecte journalière de données présente néanmoins les avantages suivants :

- les données seront probablement de meilleure qualité ; par exemple, l'opérateur se souviendra plus facilement du temps qu'il faisait hier que du temps qu'il faisait il y a un mois ;
- plus le coût de transmission des données augmente, plus les efforts humains en termes de contrôle de la qualité diminuent et plus on a de chances d'accéder à davantage de données dans de meilleures conditions ;
- les erreurs techniques seront détectées plus rapidement.

Beaucoup de stations enregistrent encore leurs données uniquement sur papier, ce qui les obligera à les envoyer par la poste une ou plusieurs fois par mois. On conservera les relevés de données sur papier ou sur tableur ou bien – ce serait l'idéal – les données seront générées automatiquement et comparées aux données attendues. Cette dernière méthode permet de détecter facilement les rapports manquants et d'assurer un suivi avec l'observateur. Il conviendra d'évaluer les données avant de les saisir à partir des relevés fournis par des appareils enregistreurs.

Pour le recueil des données, il est recommandé d'utiliser un système qui vérifie les contraintes et le type de données liés à chaque paramètre, avant d'absorber les valeurs dans la base de données.

La saisie des données et le contrôle de la qualité des données sont des opérations étroitement liées entre elles, en effet un contrôle de la qualité est souvent effectué pendant la saisie des données. Toutefois, les données climatologiques numériques doivent subir généralement un autre contrôle de qualité avant d'être classées. Les utilisateurs doivent toujours être informés du niveau de contrôle qualité des données qui leur sont fournies.

Il est important de conserver la valeur des données d'origine (à la fois pour l'évaluation des processus de contrôle de la qualité et pour d'éventuelles demandes officielles ou équivalents – voir 4.3), ainsi que la valeur établie par le dernier contrôle de qualité. Les données d'origine n'auront subi que la vérification du type et de l'étendue du processus d'absorption de la base de données. Les données rejetées par ce processus doivent être tout de même conservées. La valeur du dernier contrôle de qualité peut varier en fonction du stade du contrôle de qualité effectué.

Il doit être envisageable de faire appel à des sources et types de données non conventionnels, selon les pratiques de l'organisation, par exemple d'utiliser des caméras web pour des observations à distance.

Outre les données et métadonnées traditionnelles, le système de gestion des bases de données climatologiques peut contenir des informations diffusées par les médias, photos, etc. Ces informations peuvent être enregistrées sous la forme et selon les conditions suivantes :

- image tirée d'un article de presse, prise par une caméra numérique ou bien scannée ;
- indication de la date, de la région et du type d'événement (inondation, sécheresse, fortes précipitations) et nom de la source médiatique ;
- quelques commentaires sur l'événement ;
- stockage des informations dans un système d'archivage.

Il est possible de recueillir de la même façon des données sur des phénomènes météorologiques extrêmes ou autres événements, diffusées à la radio ou à la télévision.

4.2.3 Stockage et archivage de relevés papier

Tous les relevés papier doivent être stockés dans un environnement contrôlé afin d'éviter toute détérioration et destruction potentielle causées par des conditions extrêmes de température et d'humidité, par des insectes, parasites, incendies, inondations, accidents ou actes de vandalisme. Mais avant l'archivage, ces relevés doivent être enregistrés sous la forme de microfilms ou, de préférence, d'images électroniques, réalisés avec des appareils photos numériques ou des scanners (voir WMO 2004). Il est recommandé d'enregistrer les métadonnées sur le microfilm et/ou sur l'image, lesquels peuvent être intégrés par la suite dans une base de données climatologiques élargie.

Les images électroniques présentent les avantages suivants :

- elles constituent une copie numérique supplémentaire de sécurité ;
- il est facile de les copier plusieurs fois afin de les conserver dans des endroits différents pour des raisons de sécurité ;
- pour consulter l'original, on peut utiliser l'image électronique, ce qui évite d'abîmer le papier ;
- les relevés sont accessibles à un plus grand nombre d'utilisateurs par Internet ou par courrier électronique.

Les relevés fournis par des appareils enregistreurs doivent être traités comme les relevés papier.

Les archives papier doivent être gérées comme suit :

- les données anciennes (datant, par exemple, de plus d'un an) doivent être stockées par ordre chronologique ;
- le reste selon l'ordre des éléments de données synoptiques.

4.2.4 Stockage et archivage des informations numériques

Tel qu'il est employé ici, le terme « stockage de données » désigne la phase transitoire entre la collecte et l'archivage et non pas, comme souvent, un état dans lequel se trouvent les données une fois qu'elles ont subi une sorte de valorisation (ex : contrôle de la qualité, conversion en différentes unités de mesure).

Il est parfois difficile de procéder à une sélection des données pour le stockage et l'archivage. Les stations météorologiques automatiques généreront souvent des données portant sur la qualité des observations, qui ne sont pas des données climatologiques proprement dites (ex : informations sur la tension des batteries d'une station automatique). En règle générale, ces informations devraient être utilisées avant l'archivage des données (c'est-à-dire au cours du processus de contrôle de la qualité) et ne devraient pas figurer dans le système de gestion des

bases de données climatologiques. (Toutes les métadonnées importantes doivent en revanche y être enregistrées, voir 3.2.4, 4.1 et WMO (2003b).) Rappelons que le message d'origine doit être stocké à titre permanent et être accessible aux gestionnaires de données.

L'une des principales tâches du gestionnaire de données est d'estimer les besoins en termes de stockage de données ainsi que leur évolution. Il doit tenir compte des informations complémentaires à inclure dans les relevés de données (ex : indicateurs de contrôle de la qualité, messages d'origine, date/heure de mise à jour des relevés), des besoins en termes de métadonnées et de toute redondance nécessaire pour que les bases de données puissent être correctement récupérées. Il sera difficile d'estimer l'évolution des besoins en termes de stockage. Rappelons que beaucoup de SMHN archivent actuellement 1 et/ou 10 minutes de données de stations météorologiques automatiques et que les données climatologiques non conventionnelles (ex : indices de végétation, d'humidité du sol) sont très importantes pour les climatologues. Les besoins de stockage de données océanographiques et de données de télédétection sont généralement très vastes.

Si l'espace de stockage n'est pas suffisant pour contenir toutes les données brutes d'origine, les données les plus anciennes peuvent être transférées de la base de données vers un système d'archivage à stockage de masse plus lent. Il s'agissait jusqu'à présent d'un système robotisé d'enregistrement sur bandes, mais aujourd'hui de plus en plus sur DVD. Si nécessaire, on devrait pouvoir, par une procédure de réintégration, recharger à partir de ces systèmes de petites parties de données brutes dans l'espace de stockage.

Si, au bout d'un certain temps et après l'application du contrôle de la qualité, on n'a pas constaté de besoin de conserver des archives des données numériques d'origine dans le système de gestion des bases de données climatologiques, ces archives peuvent être supprimées, à condition que des copies des originaux soient conservées ailleurs dans l'organisation.

Gestion des relevés d'origine et sauvetage des données

Le sauvetage des données ne sera traité que brièvement ici car il fait déjà l'objet de plusieurs documents PMDSC (ex : WMO 2002, 2004).

La gestion des relevés d'origine, liée au sauvetage des données, pose toute une série de questions. Pour un certain nombre de raisons, les premières occurrences des observations seront archivées à titre permanent, par exemple la valeur d'un paramètre climatique qui n'aurait pas été modifié, ni ajouté ni supprimé par des processus manuels ou bien automatiques. Premièrement, si les données en temps quasi-réel sont utilisées pour prendre des décisions, il se peut que l'organisation en question doive conserver les valeurs d'origine pour des raisons juridiques. Deuxièmement, le suivi des performances de réseaux et de systèmes d'observation ne sera pas objectif si la preuve du mauvais fonctionnement d'une station est écartée (par exemple si l'on supprime des données erronées).

Dans l'idéal, les valeurs des données d'origine devraient être conservées dans le système de gestion des bases de données climatologiques et être accessibles au personnel interne en tant que valeurs ayant subi un contrôle de la qualité. Pour gérer ces données d'origine, il sera nécessaire d'appliquer avec précaution des indicateurs de contrôle de qualité (voir 4.4). Le gestionnaire de la base de données pourra ainsi décider d'empêcher tout accès externe direct à ces données ou, pour le moins, prendre des mesures qui lui garantiront que les utilisateurs ayant accès à ces données sont parfaitement conscients de leur qualité. S'il n'est pas possible d'enregistrer une copie électronique des données d'origine, il conviendra de conserver les originaux papier d'une manière qui réponde aux recommandations de l'OMM sur le sauvetage des données.

Le document WMO (2002) définit le sauvetage de données comme étant « un processus de sauvegarde de toutes les données qui risquent d'être perdues à cause d'une détérioration de leur support, et la numérisation de données actuelles et passées sous un format informatique compatible pour être facilement accessible ». Cette définition implique que :

- les données soient stockées en tant que fichiers d'image sur des supports régulièrement remplaçables, afin d'éviter toute détérioration du support (cartouches, CD, DVD, etc.) ;
- les données déjà disponibles sur des supports informatiques compatibles migrent en permanence vers des installations de stockage conformes aux nouvelles technologies ;
- les données soient saisies sous une forme utilisable pour des analyses.

Les efforts mis en œuvre dans le cadre du sauvetage des données sont essentiels pour permettre au public d'accéder rapidement et dans une large mesure à des ressources nationales de qualité. D'un autre côté, le sauvetage des données peut être considéré comme une forme de conservation du patrimoine d'un pays. Il est important de sensibiliser les cadres supérieurs des SMHN et d'autres fonctionnaires des gouvernements à cet aspect, ainsi qu'aux nombreux autres avantages que procurent les données climatologiques, pour assurer le financement d'activités de sauvetage des données. Le document WMO (2002) donne quelques exemples, notamment le programme financé par la Belgique qui a permis à plus de 40 pays africains de conserver des millions d'archives sur microfiches.

Le document WMO (2004) expose les principales étapes du sauvetage des données :

- recherche et localisation ;
- création d'inventaires ;
- conservation et stockage ;
- validation des fichiers d'image ;
- saisie et contrôle de la qualité ;
- accessibilité des données.

Ces étapes sont illustrées sur la Figure 2 ci-après.



Tri Analyse Index
 SAUVETAGE DE RELEVES DE DONNEES
 Données
 Images
 Papier Saisie Stockage des données Utilisateur

Figure 2. Etapes du processus de sauvetage de données

Le document WMO (2004) décrit les technologies informatiques, solutions médias et stratégies de migration de données utilisées. En termes d'acquisition d'images électroniques, on utilise de préférence des caméras numériques au lieu de scanners bien que ces derniers soient plus avantageux dans certaines circonstances (par exemple, les scanners multipages sont efficaces pour l'acquisition d'images ayant des formats papier différents). Les projets de sauvetage de données devrait toujours comprendre une composante formation, permettant aux SMHN d'acquérir des compétences sur tous les aspects du sauvetage de données et de s'assurer que les connaissances peuvent être transmises à du personnel déjà familier du sauvetage de données. La santé et la sécurité au travail sont des facteurs particulièrement importants pour les besoins du sauvetage des données. Par exemple, le personnel aura à manipuler des archives papier poussiéreuses et lourdes, dans des locaux de stockage pouvant être habités par des rongeurs et autres animaux.

Assurance et contrôle de la qualité

Le contrôle de la qualité est au cœur du processus de flux de données dans son ensemble. Il doit garantir que les données sont vérifiées et, dans la mesure du possible, qu'elles ne comportent pas d'erreur. Toutes les erreurs dues à l'emplacement de la station, aux instruments/capteurs et aux phases de transmission ou de saisie de données doivent être détectées et éliminées et, si possible, remplacées par des valeurs correctes (tout en conservant les valeurs d'origine). Le contrôle de la qualité des données est un passage obligé dans la gestion de la qualité d'un SMHN.

La première étape consiste à établir un modèle logique de contrôle de la qualité, décrivant les différences entre les essais et vérifications d'une part et la signalisation d'autre part. L'étendue des contrôles de qualité est décrite en 3.2.5.

Le modèle de signalisation devrait indiquer si la valeur est :

- vérifiée ou non ;
- d'origine ou non ;
- suspecte ou non ;
- erronée ou non ;
- correcte ou non ;
- calculée (ou dérivée) ou non.

Le modèle de signalisation peut également indiquée si la valeur est :

- vérifiée automatiquement ou par un opérateur ;
- calculée à partir de valeurs non suspectes ;
- calculée à partir de valeurs non manquantes.

Les données devront passer par l'ensemble du système de contrôle de la qualité et il conviendra d'empêcher tout moyen de contourner ce système. Le processus global devra être illustré par un ou plusieurs diagrammes de flux de données, en commençant par le réseau de la station jusqu'au niveau d'application et en se concentrant particulièrement sur le processus de contrôle de la qualité. Des guides et manuels ainsi que des instructions d'utilisation de l'ensemble du processus (et particulier du processus de qualité des données) devront être établis et suivis au sein de l'organisation. Il est fortement recommandé de décrire les changements apportés au système de contrôle de la qualité des données pour permettre des améliorations par la suite et pour informer les utilisateurs de données intéressés.

En dehors du processus régulier de contrôle de la qualité des données, il convient de réaliser de temps en temps une évaluation de toutes les données historiques de qualité éprouvée. Cette évaluation pourra s'appliquer à chaque paramètre et capteur précis de toutes les stations ou bien à tous les paramètres d'une seule station. Elle fera suite à la demande d'un client prêt à payer un contrôle qualité supplémentaire avant la réception de ses données.

Si, après avoir passé par le processus de contrôle de la qualité, un paramètre spécifique d'une station comporte encore beaucoup de valeurs suspectes, c'est que l'instrument ou le capteur ou le processus d'acquisition des données ne fonctionne pas correctement. Si les valeurs d'une station comportent plus de valeurs suspectes que celles des autres stations, cela est dû à un problème dans le fonctionnement général de la station. Quand le contrôle de la qualité est manuel, il convient de veiller à ce que les procédures soient cohérentes et en cours de validité. Par exemple, un opérateur qui remplacerait toutes les valeurs suspectes par des estimations (ou qui les supprimerait toutes) pourrait endommager davantage la base de données que quelqu'un qui indiquerait toutes les valeurs suspectes comme correctes.

S'il soupçonne un instrument ou un capteur d'être défectueux ou un dysfonctionnement de la station d'être à l'origine de problèmes, l'opérateur chargé du contrôle de la qualité doit agir et, si nécessaire, avertir le responsable de l'instrument mis en cause, par exemple pour :

- remplacer ou régler l'instrument ou le capteur ;
- entreprendre une inspection de la station ;
- formuler des recommandations visant à changer de station ;
- contacter l'observateur afin de confirmer la ou les valeurs extrêmes suspectes.

Pour assurer l'intégrité du processus de contrôle de la qualité, on désignera un responsable de processus qui sera chargé du fonctionnement du processus décrit. Il devra détecter et réviser chaque aberration et chaque variation. Il conviendra de procéder parfois à un audit interne et à un audit externe pour garantir l'intégrité du processus de contrôle de la qualité. Il est essentiel également de prévoir des formations régulières pour tous les opérateurs chargés du contrôle de la qualité.

Les SMHN utilisent actuellement des outils très utiles de contrôle de la qualité (Figure 3).

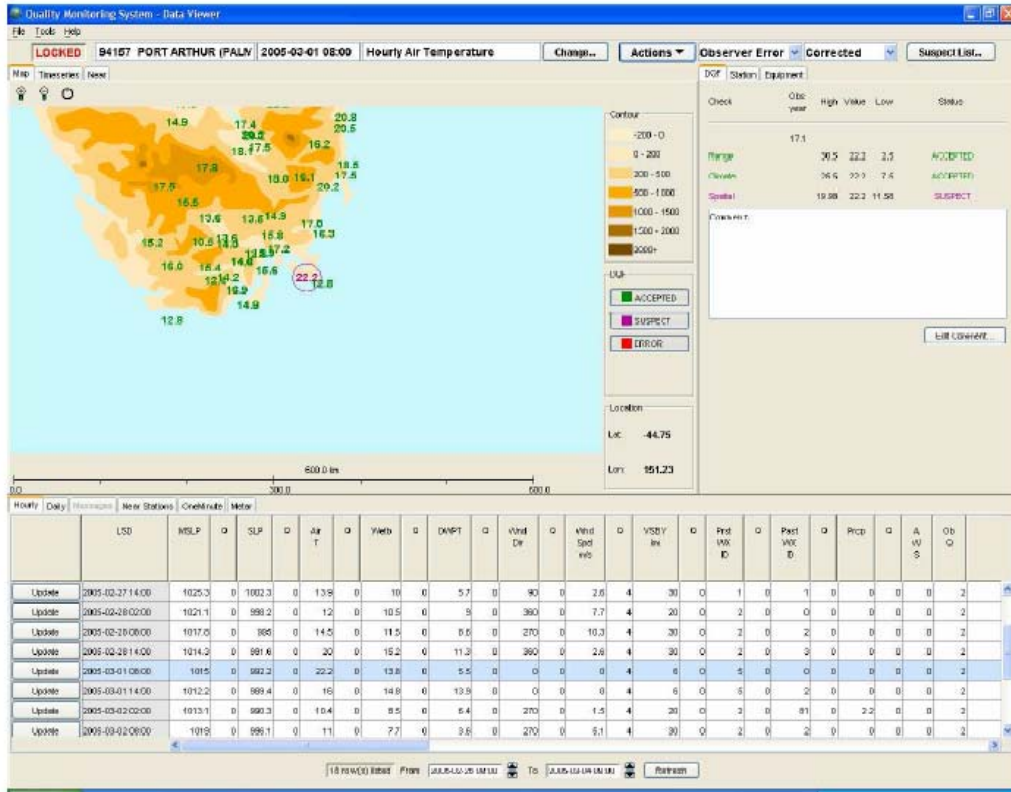


Figure 3. Capture d'écran tirée d'un outil moderne de surveillance / contrôle de la qualité des données : le système de surveillance de la qualité du Bureau météorologique australien.

Échange de données

L'échange de données entre SMHN est essentiel à la climatologie. Il implique à la fois le stockage et l'utilisation de données (et métadonnées) venant d'autres pays dans la base de données d'un SMHN, et la transmission de données à des centres de données mondiaux et régionaux.

Les États Membres de l'OMM ont pour obligation de partager données et métadonnées avec d'autres Membres de l'OMM. Les conditions de transmission de ces données à des tiers sont stipulées dans la Résolution 40 (Cg-XIII) et la Résolution 25 (Cg-XIV) de l'OMM, lesquelles intègrent le concept de données « essentielles » et « supplémentaires » et définissent un jeu de données minimum qui devrait être disponible selon un « accès libre et sans restriction ». Les membres peuvent décider de déclarer « essentielles » un plus grand nombre de données que ce jeu minimum.

Les Membres de l'OMM acceptent qu'une partie de leurs stations appartiennent à des réseaux comme le Réseau de stations d'observation en altitude pour le SMOC (GUAN), le Réseau de stations d'observation en surface pour le SMOC (GSN), le Réseau synoptique de base régional (RSBR) et le Réseau climatologique de base régional (RCBR). Le fait que leurs stations fassent partie de tous ces réseaux obligent les membres à partager leurs données sur le plan international. Sur le GUAN, ce partage passe principalement par des messages CLIMAT TEMP, pour le GSN par des messages CLIMAT. La présentation de messages de ce type est surveillée et les résultats publiés.

Les utilisateurs sont invités à lire attentivement les résolutions précitées avant de mettre en œuvre une politique nationale de données, déterminant les aspects du système de gestion des bases de données climatologiques.

Dans une perspective d'échange de données, on peut distinguer quatre catégories de données climatologiques :

- données en temps réel ou en temps quasi-réel, échangées via le Système mondial de télécommunications (SMT), pour lesquelles la rapidité d'acheminement est essentielle ;
- jeux de données en mode différé qui peuvent être des parties des données précédentes mais qui sont limités en termes de contrôle de la qualité ;
- jeux de données en mode différé qui ont fait l'objet d'un vaste contrôle de qualité et qui ont été éventuellement soumis à une analyse d'homogénéité ;
- autres produits (ex : images satellitaires, métadonnées, produits de surveillance et de prévision du climat, indices climatiques, études climatologiques et analyses nationales / régionales / mondiales, résultats de modèles du climat mondial).

Pour la première de ces catégories, des formats standard OMM ont été définis : les formats CLIMAT et CLIMAT TEMP sont des formats « spécifiques au climat » tandis que d'autres – notamment SYNOP, PILOT, TEMP et Marine – se rapportent au travail effectué sur le climat. Pour plus de détails, voir le document OMM (1995). En ce qui concerne les autres catégories de données, on a défini des formats pour l'échange de données (ex : HDF et NetCDF) mais ces derniers sont d'ordre plus général et ne sont pas « parlants » en soi.

Le projet clé, en cours actuellement, est le projet de SIO (Système d'information de l'OMM), vaste et complexe, qui, pour résumer très brièvement, vise à fournir « une infrastructure unique, coordonnée à l'échelle du globe, pour la collecte et l'échange d'informations nécessaires à tous les programmes de l'OMM et aux programmes internationaux connexes ». Pour ce faire, des protocoles standard sont en cours de développement sur l'échange de données et, ce qui est d'une importance cruciale, sur l'échange de métadonnées. Ceci a une double importance pour les développeurs et les utilisateurs de systèmes de gestion de bases de données climatologiques. D'une part, le fait d'avoir accès à ces installations permettra aux utilisateurs de systèmes suffisamment performants d'obtenir facilement, par des méthodes standard, des données provenant d'autres organisations jouant le rôle de Centres mondiaux du système d'information (CMSI) ou de Centres de production ou de collecte de données (CPCD) au sein du SIO. D'autre part, les développeurs de systèmes de gestion de bases de données climatologiques pourront créer, le cas échéant, des fonctions à l'intérieur de ces systèmes, qui leur permettront d'agir comme un fournisseur de données climatologiques via les protocoles du SIO. Certains travaux de développement suivant cette optique ont déjà été menés à bien.

Accès aux données et développement de produits

Chaque SMHN devrait avoir une politique d'accès aux données qui indique, entre autres, quelles données sont fournies gratuitement aux utilisateurs. Le libre accès aux données d'un réseau national de stations est différent d'un pays à l'autre. Les métadonnées de base sur le réseau de stations, indiquant l'emplacement, les coordonnées géographiques, l'altitude, les capteurs, la disponibilité des données, etc., devraient être fournies gratuitement par les SMHN, de préférence sous la forme de tableaux ou de cartes. Même en Europe, où les SMHN ont pour politique de donner très rarement libre accès aux données et informations des stations, on a tendance à délivrer de plus en plus d'informations sans restriction. Le projet UNIDART (interface unique de recherche de données), mené par le SMHN allemand (Deutscher Wetterdienst) en est un exemple.

Il existe de nombreuses méthodes pour donner accès à une base de données. On peut faire appel notamment au langage SQL (Structured Query Language) avec un système de gestion de base de données relationnelle (SGBDR), ou à d'autres outils permettant à l'utilisateur de

communiquer avec la base de données (tableur, shareware ou logiciels couramment disponibles dans le commerce). La méthode la plus courante consiste à utiliser des applications compatibles avec un navigateur Internet. Ensuite, on peut utiliser les mêmes applications pour l'Internet et pour l'Intranet, avec toutefois des droits d'accès bien définis.

Le personnel responsable de la fourniture des données devra comprendre comment les données ont été traitées – de la source à l'archivage. Auparavant, c'était le personnel chargé de la gestion des données qui était responsable du développement des produits, si bien que les besoins des clients externes étaient rarement pris en compte. Depuis peu, la fourniture de données répond davantage aux exigences des utilisateurs finaux et aux besoins du marché. Des services de fourniture de données se sont développés selon ces deux axes majeurs.

On a d'un côté des produits à valeur ajoutée, qui à la fois apportent aux clients des solutions « taillées sur mesure », répondant précisément à leurs besoins, et constituent des sources de revenus pour les SMHN. Les clients visés ne sont généralement pas intéressés par la quantité des informations, mais par leur pertinence.

D'un autre côté, on a des produits de masse, nécessitant moins d'efforts et une couverture moins sélective, qui ont tendance à être financés par des fonds publics et diffusés via des moyens moins coûteux : les SMHN peuvent les publier sur leur site web, par exemple. En tout cas, tous ces produits doivent être présentés selon un format professionnel car les SMHN seront jugés sur la qualité de leurs produits.

Administration et surveillance des données

Comme indiqué au chapitre 3.1, les gestionnaires de données doivent s'assurer que leurs systèmes et leurs données sont gérés de manière à satisfaire les besoins des utilisateurs (ou des entreprises). Ils doivent par conséquent mettre en place des processus et des moyens permettant de vérifier que les données qu'ils gèrent sont en adéquation avec les attentes des utilisateurs. Pour ce faire, il est utile d'établir une série d'indicateurs de performances bien choisis, reflétant les niveaux attendus.

Les gestionnaires de données doivent tenir compte du fait qu'il arrive parfois que les données recueillies par un instrument ne se retrouvent pas dans la base de données. Il faut alors gérer les processus de flux de données et surveiller de près les données pour savoir lesquelles sont enregistrées dans la base de données. Il convient de recenser les problèmes de flux de données et de les corriger avant qu'ils n'aient des répercussions pour les utilisateurs de données climatologiques. Cette responsabilité incombe directement aux gestionnaires de données même si d'autres parties prenantes peuvent être chargées des mesures correctives (par exemple, l'absence d'observations peut être due à des problèmes liés aux technologies de télécommunications, et on fera alors appel au service informatique pour réparer).

On a défini, au chapitre 3.4, comment mesurer les performances dans le cadre d'une gestion de base de données plus générale. Le principe est le même quand il s'agit plus précisément de données climatologiques : tout indicateur doit être associé à des valeurs cibles permettant de mesurer les performances par rapport aux besoins des utilisateurs. Une fois de plus, ces indicateurs doivent être Spécifiques, Mesurables, Atteignables, Réalistes et limités dans le Temps (SMART).

Que faut-il surveiller ? Du point de vue utilisateur, un gestionnaire de données devrait établir des processus et indices lui permettant de garantir que les données de la base sont :

- appropriées en termes de types de paramètres climatiques ;
- conformes aux directives nationales et internationales relatives à la densité spatiale, la fréquence et la taille des relevés ;
- soumises à un contrôle qualité approprié ;

- de qualité acceptable ;
- disponibles sous forme numérique ;
- stockées de manière optimale en termes de sécurité et d'accessibilité.

Parmi les éléments ci-dessus, certains (ex : paramètres climatiques appropriés) peuvent nécessiter des enquêtes ou bien reposer sur les retours fournis régulièrement par les utilisateurs des données. L'efficacité des procédures de contrôle de la qualité peut être mesurée, par exemple, d'après la proportion de fausses alertes, tandis que l'on déterminera si les niveaux de qualité sont respectés d'après le pourcentage de valeurs de données manquantes et erronées. Les indicateurs destinés à surveiller la proportion de données disponibles sous forme numérique et sous forme papier peuvent être utilisés comme des mesures de l'efficacité des systèmes de traitement de données. Voir exemple plus précis en 3.4. Le suivi des taux de numérisation permet également aux gestionnaires de données d'identifier les problèmes de rentabilité et de prévoir la charge de travail.

Il faut veiller à ce que ces activités de surveillance ne soient pas une charge pour le département responsable de la gestion des données. Elles jouent néanmoins un rôle important, tirant la sonnette d'alarme dès que l'on s'écarte des performances acceptables, et peuvent servir aussi d'outil particulièrement utile de motivation du personnel.

Mais le plus important peut-être, c'est que les systèmes et processus en place doivent donner l'alerte en temps réel, en présence d'activités critiques pour l'économie, par exemple lorsqu'il manque des données pendant la réalisation de produits climatologiques de haut niveau (demandés notamment par des ministres) ou pendant la transmission de données à des clients qui paient un tarif spécial urgence pour avoir un service rapide.

Enfin, il est important de disposer de systèmes de communication efficaces et rentables, capables de diffuser les informations dont ont besoin les personnes chargées des mesures correctives.

Gestion des modifications

Les données climatologiques sont soumises à un grand nombre d'influences non climatiques différentes, d'où la nécessité pour les gestionnaires de données d'adapter leurs bonnes pratiques en matière de gestion des modifications. Les changements climatiques suscitant davantage de préoccupations depuis les dernières décennies, la gestion des modifications des données a pris également plus d'importance.

Pour une bonne gestion des modifications, il est indispensable de disposer de connaissances, de métadonnées et d'une documentation de bonne qualité sur les pratiques existantes de la gestion de données (voir 3.5 et 4.1).

Les types de modifications à gérer sont notamment les suivantes :

- modifications des systèmes et réseaux d'observation ;
- modifications des méthodes d'observation ;
- introduction de nouveaux types de données ;
- modifications des algorithmes de calcul de données dérivées.

Les questions de transition entre systèmes de gestion de bases de données sont abordées au chapitre 5.

Les modifications apportées aux réseaux et systèmes d'observation (ex : déplacements de sites, remplacement de type de capteur) devraient être prises en compte, dans l'idéal, dans les relevés de métadonnées aux stations concernées et, si les modifications s'appliquent à l'ensemble de la station, dans toute autre documentation à large diffusion et déjà prête pour les

futures générations. Lorsqu'un programme d'observation parallèle est en place, par exemple dans le cas d'un changement de site, il est nécessaire d'adapter les flux de données correspondants (par exemple, introduire un nouveau site ou conserver les anciennes données dans une table de base de données séparée). Les modifications apportées aux pratiques d'observation devraient être gérées par le département du SMHN chargé des observations, et la documentation devrait, une fois de plus, être largement diffusée et déjà prête pour les générations futures. Il en va de même pour les modifications impliquant des calculs de données dérivées ; tandis que le département chargé des observations aura pour tâche d'intégrer les modifications dans les algorithmes d'observation liés au système, les gestionnaires de données se chargeront d'enregistrer les modifications en aval, liées à la base de données. Les gestionnaires de données devront toutefois essayer de présenter aux utilisateurs une liste consolidée des modifications.

L'introduction de nouveaux types de données impliquera généralement d'intégrer des processus et des systèmes spécifiques pour le stockage, le contrôle de la qualité, l'établissement de rapports, etc. Si cela représente un volume de travail supplémentaire, on pourra envisager de recruter du personnel ou d'installer de nouveaux matériels ou logiciels, ce qui demande souvent d'importants travaux de planification.

Conformément aux principes de gestion de la qualité, les SMHN devraient envisager de faire appel à des processus officiels pour garantir une gestion en bonne et due forme des modifications. Le document WMO (2003a) décrit les efforts mis en œuvre dans ce sens au sein du Service météorologique du Canada (SMC). Si les structures officielles adoptées par le SMC ne sont pas toujours transposables à d'autres SMHN qui, pour la plupart, sont loin de disposer des mêmes ressources, en revanche les principes généraux méritent d'être retenus, notamment les suivants :

- création d'un groupe au sein de l'organisation, chargé de superviser les modifications apportées aux systèmes, instruments, algorithmes, processus, procédures et documentation connexe, autant de facteurs influençant l'acquisition, le traitement, l'établissement de rapports et l'archivage des observations réalisées depuis les réseaux d'observation ;
- font partie de ce groupe les décideurs des domaines précités et, si nécessaire, le personnel chargé du soutien scientifique et technique ;
- les décisions sont basées sur le consensus dans la mesure du possible ;
- mise en place d'un processus clair et transparent de demande de modification ;
- examen et traitement en temps voulu des demandes de modification ;
- le comité a le pouvoir de créer des groupes de travail chargés d'évaluer les demandes et de formuler des recommandations ;
- facilité d'accès à toutes les demandes, décisions et documents justificatifs.

Transition vers un système de gestion de base de données

Lorsque les outils qu'il utilise pour la gestion et le traitement des données climatologiques ne lui donnent pas entièrement satisfaction, le SMHN peut envisager d'acquérir un nouveau système de gestion de base de données (SGBD), pour diverses raisons : être capable de traiter des volumes de données plus élevés qu'avant, répondre à de nouvelles demandes de la part des utilisateurs (ex : fourniture des données par Internet) ou bien parce que les outils utilisés jusqu'à présent ne sont pas suffisamment évolutifs et ne correspondent aux technologies de pointe du moment.

Le choix du nouveau SGBD ainsi que la phase transitoire qu'il implique sont des étapes cruciales car le SMHN s'engage sur une longue période. Ce chapitre présente une approche « projet » et en décrit les principales phases (analyse de la situation présente, définition d'une solution fonctionnelle et définition d'une solution technique), qui permettront de choisir le système approprié. On trouvera également quelques réflexions sur l'architecture de la base de

données (conception de la base, modèles de données) afin de pouvoir prendre des décisions en connaissance de cause, et enfin une description détaillée des étapes de transition entre le système CLICOM et un nouveau système de gestion de bases de données climatologiques.

Il convient de préciser dès le début le principe à suivre lorsque l'on envisage de changer de SGBD : « acheter si l'on peut, construire seulement si l'on est obligé ». À la lumière de ce principe, le PMDSC préconise le développement de SGBD candidats à examiner par les différents pays (voir Annexe 1 et l'étude de cas présentée à l'Annexe 2). Cet aspect sera traité plus en détail ci-après.

Choix d'un système de gestion de bases de données climatologiques : les éléments à prendre en compte

Ce chapitre utilise une terminologie commerciale même si un grand nombre de SMHN sont entièrement financés par les gouvernements et, par conséquent, ne facturent pas directement leurs services. Ces SMHN doivent cependant prouver leur rentabilité aux organismes qui les financent, en justifiant les prestations fournies aux utilisateurs. Ces organismes sont donc des clients par procuration pour les services fournis aux utilisateurs.

Prise en compte des besoins

Pour choisir un système de gestion de bases de données climatologiques, la première étape consiste à évaluer les besoins de l'organisation ainsi que l'environnement dans lequel le futur système sera utilisé. Ces besoins peuvent être considérés comme des besoins internes ou externes. Les besoins externes couvriront les exigences existantes des clients et des utilisateurs, ainsi que les exigences auxquelles on peut s'attendre à l'avenir. Il faudra procéder à une analyse des produits actuels et des produits souhaités. Par exemple, si l'un des principaux produits est lié au rendement d'une culture en particulier, on pourrait exiger du système de gestion de bases de données climatologiques qu'il enregistre les données à la date de levée du semis, et donc qu'il soit en mesure de spécifier ces types de données.

Dans quelle mesure les besoins peuvent-ils changer pendant la durée de vie du système ? Par exemple, un SMHN peut avoir besoin d'accroître ses revenus commerciaux, ce qui nécessitera de développer de nouveaux services personnalisés. Dans ce cas, il sera important qu'il puisse le faire sans difficulté.

Le domaine de compétence et les limites de l'organisation devront être prises en compte. Son domaine de compétence pourra être limité par le cadre légal prévu pour l'organisation et dépendra aussi d'autres organisations partenaires ou concurrentes.

Par exemple, si l'organisation en question est un SMHN dont les responsabilités relèvent de l'hydrologie, de la sismologie et de l'environnement maritime, elle utilisera un éventail de types de données plus large qu'un SMHN chargé uniquement de la météorologie opérationnelle. Si la météorologie et l'hydrologie relèvent d'organisations différentes, ces dernières devront se mettre d'accord sur celle qui sera responsable des données relatives aux précipitations et sur les processus de collecte, de contrôle de la qualité et d'échange des données.

Un SMHN peut être tenu de stocker des données venant d'autres pays. Dans ce cas, on appliquera des mécanismes spécifiques, permettant de contrôler cet usage aux termes de la Résolution 40 de l'OMM (voir 4.5) ou de tout autre accord.

Définition du système requis

Pour préparer la définition du système requis, il est utile d'établir une liste des problèmes par rapport aux besoins, pour répertorier les problèmes et leur donner un ordre de priorité. Le meilleur moyen est de mettre en place une collaboration entre tous les intervenants du système. Bien entendu, ceci ne pourra pas régler tous les problèmes ni satisfaire tous les besoins.

La principale contrainte sera souvent financière, mais il est important aussi de se pencher sur la disponibilité du personnel et sur ses niveaux de compétences. Il conviendra de prévoir les besoins du personnel en matière de formation. La présence d'infrastructures de formation et d'assistance matérielle et logicielle au niveau local est également très importante. Il n'est pas recommandé de prévoir d'utiliser des matériels et logiciels sans assistance ni support local, sauf si d'autres solutions adaptées peuvent être mises en œuvre durablement pendant tout le cycle de vie du système ; le niveau de redondance demandé sera par ailleurs plus élevé.

Le flux des données dans l'organisation doit être évalué et les modifications correspondantes planifiées. Les quatre étapes principales d'un système de gestion de bases de données climatologiques sont les suivantes : contrôle des métadonnées, saisie des données, contrôle de la qualité et traitement des requêtes. La saisie peut être réalisée de diverses façons, notamment par une assimilation directe à partir de systèmes automatisés et la saisie d'après des formulaires papier (voir 4.2.2). Il conviendra de prendre en compte les moyens utilisés actuellement pour ce faire ainsi que les moyens à mettre en œuvre pour améliorer ces actions dans le nouveau système. Le volume des données a son importance, à la fois pour l'assimilation des données à titre régulier et pour la taille totale de la base de données. Les données climatologiques étant généralement conservées à titre permanent, il faudra envisager une stratégie de sauvegarde et d'archivage pour pouvoir accroître la base de données tout au long de sa vie.

Il est particulièrement intéressant aussi de savoir, pour définir un système de gestion de bases de données climatologiques, si le système fera partie des systèmes synoptiques opérationnels.

Evolutivité

L'évolutivité de la base de données est également un critère à prendre en compte. Peut-elle s'adapter à des changements de l'organisation ? Ces changements impliquent le plus souvent de pouvoir ajouter de nouveaux types de données et de faire face à des volumes de données plus importants : par exemple, pouvoir contenir 10 minutes de données provenant de stations météorologiques automatiques. Généralement, le coût du stockage sur disque dur baisse plus vite que n'augmentent les besoins de données d'observation supplémentaires.

Est-il possible de décentraliser la base de données en fournissant à un bureau régional une version simplifiée qui ne contiendrait que les données applicables à la région en question ?

En général, les données seront de meilleure qualité si leur saisie dans le système est réalisée dans un endroit proche du lieu où elles ont été collectées. Il peut être avantageux de décentraliser la base de données ou d'en donner l'accès à des centres d'information régionaux, dans la mesure où l'on touche ainsi directement les clients et utilisateurs de ces centres, grâce à une plus grande rapidité d'exécution et une meilleure compréhension des besoins des clients et aussi une meilleure utilisation des ressources humaines des centres. L'inconvénient de cette pratique est en quelque sorte une perte de contrôle et de transparence et une hausse des coûts de support. L'établissement d'une architecture commune et d'un système commun de gestion de bases de données climatologiques pourrait également être avantageux pour d'autres organisations telles que les autorités fluviales.

Il conviendrait d'étudier les obstacles et les coûts potentiels de tels changements dans le cadre de la planification à long terme du système de gestion de bases de données climatologiques.

Architecture et technologie

Après avoir défini tous les problèmes et les besoins, on peut mettre en place le système requis pour qu'il réponde au maximum d'entre eux.

La fiabilité et la maintenance sont des critères majeurs pour le matériel. La microinformatique est désormais largement répandue dans tous les pays, si bien qu'ils disposent d'une maintenance matérielle adaptée, au moins pour les technologies standard. La tolérance aux pannes est très importante, et l'alimentation secteur doit comprendre un système d'alimentation électrique à temps de coupure zéro (UPS). Les disques durs du serveur qui héberge la base de données doivent être disposés en réseau RAID (Redundant Array of Inexpensive Disks), comprenant au moins trois disques physiques pour éviter qu'une défaillance de l'un d'entre eux n'affecte les données. Si le système doit être disponible en permanence ou avec des temps de coupure réduits au minimum, on pourra ajouter un serveur redondant qui assurera la continuité des opérations en cas de défaillance du serveur principal.

On peut envisager d'utiliser d'autres architectures matérielles qu'un PC à condition de disposer d'une assistance efficace, capable de rétablir le bon fonctionnement du système dans un délai commercialement raisonnable (pour un système de gestion de bases de données climatologiques, généralement le jour ouvré qui suit la panne).

Tout système de gestion de bases de données climatologiques doit faire l'objet de sauvegardes de secours pour des raisons de sécurité. Ces sauvegardes seront effectuées, dans l'idéal, sur un support lisible par d'autres systèmes de l'organisation, généralement sur des CD et DVD réinscriptibles et éventuellement sur des bandes. Il est recommandé de vérifier régulièrement la lisibilité de ces sauvegardes sur d'autres machines et de conserver une série de sauvegardes sur un site distant.

Le choix du système d'exploitation se fera compte tenu des mêmes considérations de coût, de formation et de support. Microsoft Windows est le système le plus connu, bénéficiant d'une bonne couverture commerciale dans presque tous les pays. En revanche, il est sensible aux virus et autres logiciels malveillants et il est facilement pénétrable pour une utilisation non autorisée. Il est donc nécessaire d'investir dans des logiciels anti-virus et pare-feu pour les connexions extérieures. D'autres systèmes comme Linux demandent généralement un niveau de compétences plus élevé en termes de support, mais s'il se trouve, dans l'organisation, du personnel ayant déjà ces compétences (acquises sur d'autres systèmes), il ne faut pas hésiter à opter pour les systèmes en question car ce sera véritablement rentable.

Le SGBD peut être un logiciel payant, disponible dans le commerce, ou un logiciel à code source ouvert. Voir 5.3.3.4 pour plus de détails sur certains produits particuliers. Les critères essentiels seront la disponibilité et le coût de la formation et du support. On devra déterminer le coût du système de gestion de bases de données climatologiques sur toute sa durée de vie, ce qui peut être complexe. On se demandera notamment si chaque mise à niveau du matériel ou chaque décentralisation du système dans des centres régionaux (voir 5.1.3) donnera lieu à des droits de licence supplémentaires.

Les solutions client-serveur peuvent présenter des avantages significatifs. Des interfaces telles que JDBC (Java DataBase Connectivity) ou ODBC (Open DataBase Connectivity) sont conçues pour la communication entre bases de données et entre les utilisateurs et la base de données. Cela signifie que l'on pourra opter pour une plate-forme principale de serveur pour parvenir à un maximum d'efficacité et de sécurité, tandis que plusieurs clients pourront être définis dans les modules spécifiques aux fonctions commerciales. Des concepts clients

différents peuvent être utilisés d'une part pour les tâches d'assimilation de données et de contrôle de la qualité, et d'autre part pour les requêtes d'utilisateurs finaux. Les systèmes clients utilisés en particulier pour des requêtes d'utilisateurs finaux doivent être extrêmement souples pour concevoir et établir rapidement de nouveaux rapports. Les clients tableurs sont utiles lorsqu'il s'agit de transmettre les données extraites à des utilisateurs finaux, sous la forme d'un fichier de données. De nombreux outils (par exemple, Microsoft Access utilisé en tant que client) permettent de préparer relativement facilement et rapidement des rapports bien présentés.

La qualité de la présentation des rapports ainsi que la facilité avec laquelle de nouveaux produits peuvent être développés sont très importantes pour un SMHN. Une présentation de qualité montre au client ou utilisateur que le SMHN est une organisation scientifique sérieuse et efficace, ce qui valorise son image de marque.

Choix du système de gestion de bases de données climatologiques

La plupart des systèmes de gestion de bases de données climatologiques et des SGBD sur lesquels ils tournent sont évolutifs, à partir d'une plate-forme PC. Comme indiqué plus en détail ci-dessus, le système de gestion de bases de données climatologiques peut être vu comme la combinaison d'un modèle de données et de plusieurs outils réalisant les tâches requises par les besoins commerciaux. De manière idéale, il convient de tester plusieurs systèmes et de sélectionner celui qui convient le mieux. Une autre solution consiste à créer soi-même un système de gestion de bases de données climatologiques, mais, en pratique, c'est beaucoup d'investissements en termes de temps et de travail.

Dans la plupart des cas, ce sera plus efficace de choisir un système existant qui réponde le mieux aux besoins. Il peut être utile de tenir compte de l'expérience d'autres organisations similaires en la matière, mais cela n'aide pas toujours. La décision s'appuiera également sur les coûts liés au système sur l'ensemble de sa durée de vie, au support, à la formation et aux différentes évolutions potentielles. Puis le système devra répondre aux besoins locaux : il faudra généralement ajouter des modèles de saisie pour une utilisation locale, assurer l'intégration automatique de données venant d'autres systèmes et permettre l'élaboration de rapports pour les utilisateurs finaux. Ces fonctions supplémentaires peuvent servir à d'autres utilisateurs du même système, et inversement, on peut reprendre les bonnes idées des autres groupes.

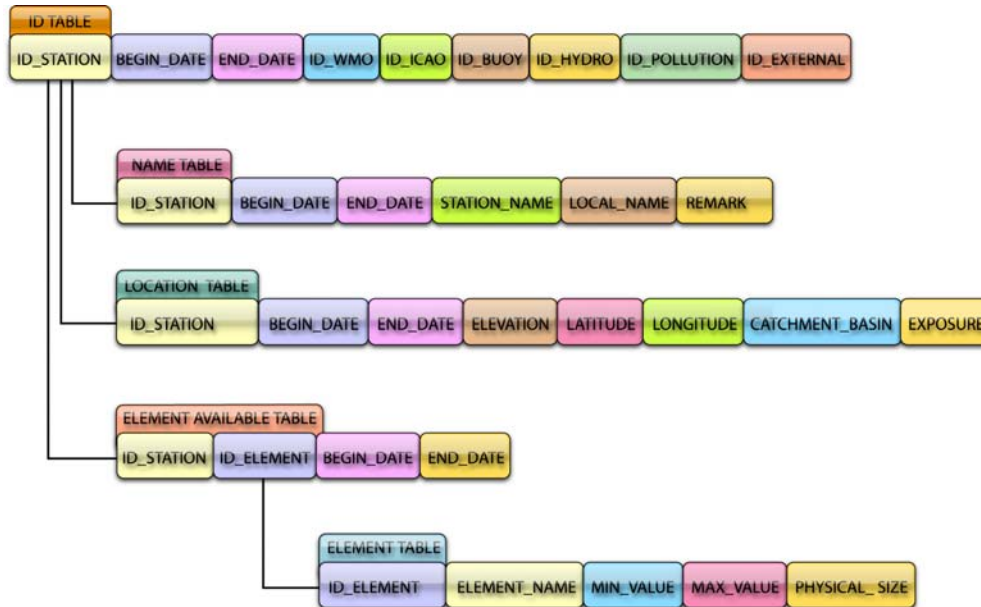
Considérations relatives à l'architecture de la base de données

Ce chapitre se concentre sur le modèle de données relationnelles, laissant de côté les nouvelles technologies, par exemple le « modèle objet » ou le « modèle contextuel », qui, pour l'instant, ne sont pas d'usage courant parmi les SMHN.

Un système de gestion de base de données relationnelle (SGBDR) est un ensemble de données représentées dans des tables composées de rangées et de colonnes. Toutes les relations entre les données sont représentées par des valeurs communes dans les tables correspondantes. Par exemple, on peut créer une table « Id » pour stocker les informations relatives à l'identification de la station, et une solution pourrait être de décomposer l'information Id en plusieurs éléments, par exemple : WMO Id, HYDRO Id et LOCAL Id (voir ci-dessous).



La table Id pourrait alors être reliée à d'autres tables de la base de données, comme indiqué ci-dessus :



La première tâche d'un concepteur de bases de données est de définir un modèle de données et d'optimiser la structure des tables, c'est-à-dire de trouver la structure la mieux adaptée à chaque table et de définir comment ces tables sont reliées entre elles. À partir de là, il peut fonder un système efficace.

Considérations relatives à la conception d'une base de données : normalisation

Pour trouver les meilleures structures de tables, un concepteur de bases de données doit suivre certaines règles, c'est ce que l'on appelle la normalisation.

La normalisation d'une base de données se compose d'une série d'étapes à suivre pour obtenir une base de données permettant un stockage cohérent des données et assurant un accès efficace à ces données. La normalisation est utile pour :

- limiter la redondance des données ;
- limiter les pertes de données ;
- limiter les incohérences entre les données.

La base de données qui en résultera sera un compromis entre performances et flexibilité. En général, il n'existe pas de solution idéale unique, c'est pourquoi les différents systèmes de gestion de bases de données climatologiques existants sont le fruit de solutions différentes.

Modèles de données utilisés par les systèmes de gestion de bases de données climatologiques

La plupart des systèmes de gestion de bases de données climatologiques proposent un modèle de métadonnées pour stocker les informations sur les métadonnées et un modèle de données pour stocker les valeurs d'éléments climatologiques. Le modèle de métadonnées est très variable : entre quelques tables reliées entre elles pour certains systèmes (ex : CLICOM) et une série complexe de tables, pour d'autres. Tout dépend du niveau de détail souhaité.

Le modèle de données, en revanche, est plus homogène. Les systèmes de gestion de bases de données climatologiques utilisent généralement trois modèles de données pour représenter les valeurs d'éléments climatologiques, avec pour chacun ses avantages et ses inconvénients.

Modèle basé sur l'élément

Un modèle basé sur l'élément permet de représenter des données dans des tables avec, dans chaque rangée, **différentes valeurs d'un même élément**, observé sur un même site **à des moments différents**.

Par exemple, les données quotidiennes pourraient être stockées dans une table quotidienne. Chaque rangée correspondrait à une station spécifique, à un mois spécifique et à un élément spécifique. Ces attributs, c'est-à-dire chacune des cellules d'une rangée, permettent de stocker les différentes valeurs de l'élément et de la station pour un mois donné (ex : pour janvier, 31 valeurs).

Tableau 5.1. Table de données quotidiennes pour un modèle basé sur l'élément

ID de la station	Mois/Année	Élément	Valeur Jour 1	Valeur Jour 2	Valeur Jour 3	Valeur Jour 4	Valeur Jour 5	Valeur Jour 31
95123	01/2002	Tmin	23.4	25.2	28.3	26.5	27.8	24.9
66202	01/2003	Tmax						
								

Avantages : on peut facilement ajouter de nouveaux éléments ; le modèle de données reste identique.

Inconvénients : les performances des applications en temps réel risquent d'être médiocres et beaucoup d'opérations de la base de données plus complexes qu'avec d'autres modèles.

Le modèle basé sur l'élément est utilisé, par exemple, dans CLICOM.

Modèle basé sur l'observation

Un modèle basé sur l'observation permet de représenter des données dans des tables avec, dans chaque rangée, les valeurs de **différents éléments**, observé sur un même site **à un moment donné**.

Par exemple, les données quotidiennes pourraient être stockées dans une table quotidienne. Chaque rangée correspondrait à une station spécifique, à un moment donné. Chaque colonne de chaque rangée contiendrait les valeurs des différents éléments observés.

Tableau 5.2. Table de données quotidiennes pour un modèle basé sur l'observation

ID de la station	Jour/Mois/Année	Tmin	Tmax	Pluie	Humidité mini.	Pression mini. au niveau moyen de la mer	Vitesse maxi. du vent	Sens maxi. du vent ..
33220	01/01/2002	24.5	33.4	0	72	1015.6	2.2	160
42500	01/01/2003	15.2	22.3	10.2	80	1013.4	3.3	210

Avantages : performances élevées pour les applications en temps réel ; optimisation du stockage des données.

Inconvénients : nécessité de mettre à jour la structure de la table pour ajouter un nouvel élément qui n'aurait pas été intégré pendant la phase de conception de la base de données.

Le modèle basé sur l'observation est couramment utilisé par les SMHN – voir rapport OMM de Benichou et Lee (1996).

Modèle basé sur la valeur

Un modèle basé sur la valeur permet de représenter les valeurs des données dans des tables avec, dans chaque rangée, **une seule valeur pour un seul élément**, observé sur un même site **à un moment donné**.

Par exemple, les données quotidiennes pourraient être stockées dans une table quotidienne. Chaque rangée correspondrait à une station spécifique, à un moment donné.

Tableau 5.3. Table de données quotidiennes pour un modèle basé sur la valeur

ID de la station	Date	Élément	Valeur
33220	01/01/2002	Tmin	23.4
42500	01/01/2003	Tmax	16.3
22222	01/01/2003		

Avantages : on peut facilement ajouter de nouveaux éléments ; le modèle s'adapte à un large éventail de types de données.

Inconvénients : le stockage des données ne sera pas optimisé correctement, cette méthode n'est donc pas adaptée à des tables contenant d'énormes quantités de données ; mêmes inconvénients que le modèle basé sur l'élément.

En conclusion, les modèles de données ont chacun leurs points forts et leurs points faibles, mais chacun pourrait être utile à un type d'application climatologique. Il devrait être possible, par ailleurs, d'intégrer, dans la même base de données, des tables utilisant différents modèles de données, afin de répondre à des besoins et contraintes spécifiques. Par exemple, un système de gestion de bases de données climatologiques pourrait comprendre un modèle basé sur l'observation pour les données horaires et quotidiennes et un modèle basé sur la valeur pour les données de pollution quotidienne.

Considérations relatives au logiciel et au matériel informatique

Avant de définir en détail les composants d'un nouveau système, il est recommandé de procéder à une analyse de la situation, portant sur les compétences du personnel et le matériel/logiciel disponible, ainsi qu'à une étude des besoins.

Analyse de la situation

Pour décider du matériel ou logiciel adapté à un SMHN, il est nécessaire de connaître et de comprendre parfaitement l'environnement dans lequel le SMHN gère ses données. Informations techniques et compétences en matière de gestion de données devront être réunies pour obtenir une solution fonctionnelle. Les étapes clés sont les suivantes :

Inventaire complet et description détaillée des composants matériels et logiciels disponibles : ordinateurs, réseaux, systèmes d'exploitation, système de gestion de bases de données climatologiques, applications en cours d'utilisation, etc.

Infrastructures de télécommunications dont dispose le pays et/ou la région : lignes de télécommunications nationales et internationales : Internet, SMT, téléphone, radio, etc.

Niveau de compétences de chacun des agents du SMHN en matière de matériel et de logiciel : sur le fonctionnement matériel, la maintenance, les réparations, les systèmes d'exploitation, les systèmes de gestion de bases de données climatologiques, les langages de programmation, etc.

Compétences disponibles dans le pays ou dans des pays proches : fournisseurs et autres organisations connexes, leurs compétences et capacités d'intervention (ex : maintenance, réparation, pièces de rechange).

Définition d'une solution fonctionnelle : évaluation des besoins

Une fois que les capacités techniques et les compétences ont été définies, il convient, dans un deuxième temps, de décrire précisément les besoins du SMHN. Le nouveau système doit généralement remplir au moins les mêmes fonctionnalités utiles que le système en place et être compatible avec les pratiques climatologiques du SMHN. Mais envisager l'acquisition d'un nouveau système est aussi l'occasion de réfléchir à une éventuelle redéfinition des besoins. Les aspects suivants, notamment, mériteraient d'être améliorés : flux des données entre la station et la base de données puis vers les utilisateurs du SMHN, report de l'acquisition des données et de la livraison des produits, sécurité des données, accès aux données, production des données. Ces exigences influenceront directement les choix techniques qui seront faits pour le nouveau système (ex : réseau, matériel/logiciel, système de gestion de bases de données climatologiques).

Les questions suivantes constituent un fil conducteur destiné à aider la personne chargée de définir les besoins du SMHN.

Quelle utilisation et quelles fonctions ?

Le système doit-il stocker une seule et unique base de données climatologiques ? Ou bien doit-il en stocker d'autres (ex : relevés sismiques ou administratifs pour l'organisation) ?

A-t-on besoin de lignes de télécommunications spécifiques ? SMT, Internet, téléphone, SMAS etc.

Quelles sont toutes les fonctions que le système devra assumer, en particulier dans les domaines suivants : acquisition des données, accès aux données, gestion et traitement des données, production et sécurité des données ?

Voir liste complète des critères à l'adresse suivante :

<http://www.wmo.ch/web/wcp/wcdmp/cdmsfuture/html/evaluation.html>.

Quelles contraintes ?

Le nouveau système devra être installé dans un environnement spécifique, c'est-à-dire qu'il devra s'intégrer dans la structure organisationnelle du SMHN et faire face aux contraintes de cette dernière. Par conséquent, les aspects suivants doivent être pris en compte.

Quels utilisateurs ?

Combien d'utilisateurs doivent pouvoir accéder au système de gestion de bases de données climatologiques ? Où se trouvent-ils ? Ont-ils besoin d'accéder aux données en temps quasi-réel ou bien un accès différé est-il suffisant ? Les utilisateurs ont-ils des fonctions particulières ? La réponse à cette question permettrait de déterminer s'il est préférable d'avoir une base de données distribuée (au niveau national et régional) ou une base de données centralisée (unique).

Quel volume de données ?

Quel est le volume de données requis actuellement ? À combien se montera-t-il d'ici cinq ans ? La réponse à cette question est liée à l'évolution future du SMHN : nombre de stations, éléments à stocker, période, fréquence d'acquisition (toutes les minutes, toutes les heures, etc.), politique de saisie et d'acquisition des données, extension des réseaux.

Liens fonctionnels : au niveau interne, national et international ?

Il est particulièrement utile d'établir des liens fonctionnels entre les différentes parties prenantes d'un SMHN, dans le cadre d'une opération de gestion de données, au niveau interne, national et international. Ceci permet de mieux représenter l'ensemble des liens existants et de définir correctement le réseau. Ce type de schéma est particulièrement adapté pour représenter le flux de données entre le lieu d'observation et les produits destinés aux utilisateurs. Il peut également être utilisé dans d'autres domaines comme le fonctionnement des données ou l'accès aux données.

FUNCTIONAL SCHEMA

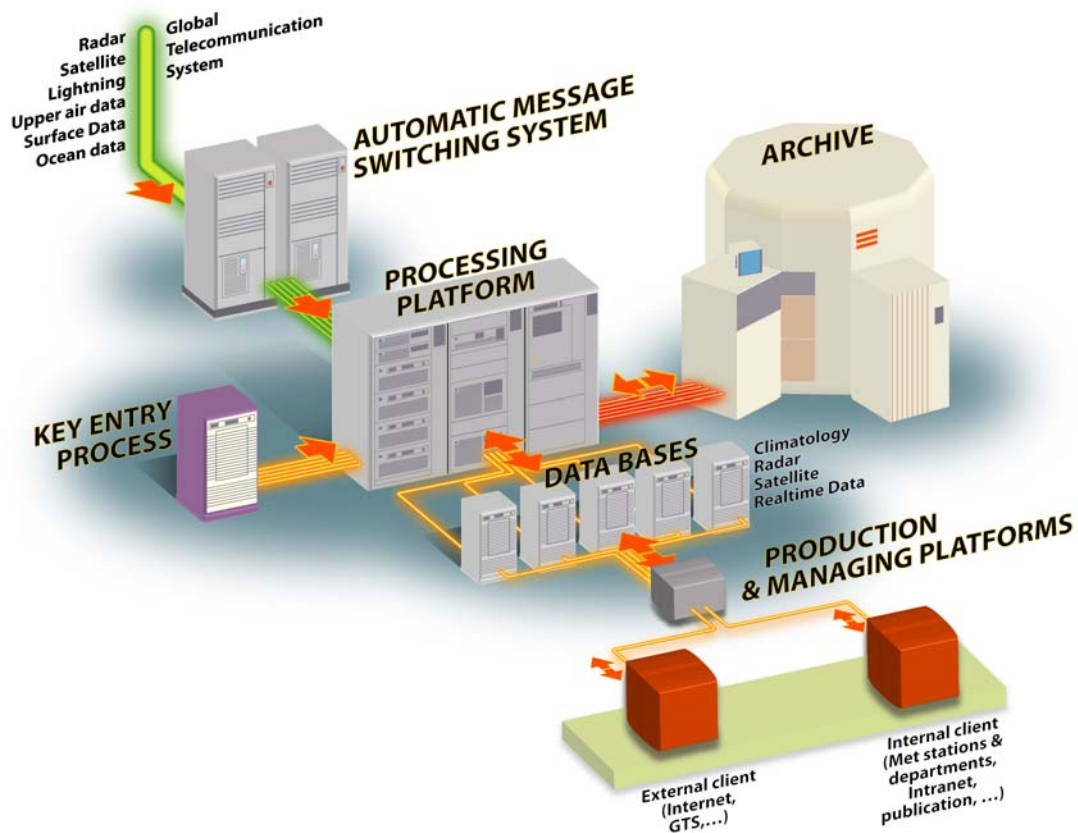


Figure 4. Exemple de schéma fonctionnel indiquant les flux de données et différents liens en jeu dans le cadre d'une opération de gestion de données.

[Traduction des légendes]

SCHÉMA FONCTIONNEL

Radar
 Satellite
 Foudre
 Données aérologiques
 Données de surface
 Données océaniques

Système mondial de télécommunications

SYSTÈME AUTOMATIQUE DE COMMUTATION DE MESSAGES

PROCESSUS DE SAISIE

PLATE-FORME DE TRAITEMENT

ARCHIVE

BASES DE DONNÉES

Climatologie
 Radar
 Satellite
 Données en temps réel

Client externe (Internet, SMT,...)

Client interne (services et stations météo, Intranet, publication, ...)

Disponibilité ?

Des coupures de système sont-elles autorisées ? Le système a-t-il besoin d'un temps minimum pour se mettre correctement hors tension sans entraîner de perte ou de détérioration des données ?

Compatibilité ?

Dans quelle mesure le nouveau système devra-t-il être compatible avec le matériel et le logiciel existants ? Par exemple, certains programmes existants devront-ils fonctionner sur le nouveau système ou certains équipements existants devront-ils communiquer avec le nouveau système ?

Définition de la solution technique

Une fois qu'il dispose des informations ci-dessus et des compétences correspondantes en informatique, le SMHN doit être en mesure de définir une solution technique. Voici quelques facteurs importants à prendre en compte.

Choix budgétaires

Le budget total d'un système n'est pas simplement le coût des différents composants. Il convient d'y inclure plusieurs autres services :

- garantie ;
- coût de formation (par exemple, pour l'utilisation, la maintenance, les réparations) ;
- frais généraux (par exemple les consommables, lignes de communication) ;
- contrat de maintenance ;
- budget pour la maintenance et l'évolution du système : amélioration ou mise à hauteur du matériel et du logiciel, maintenance corrective et adaptative, assistance ponctuelle, développements spécifiques.

Types d'ordinateur

On distingue généralement les ordinateurs par leur taille, du plus modeste au plus performant : micro-ordinateur (ou ordinateur individuel, PC), mini-ordinateur et ordinateur central. On peut parfois aussi les distinguer par leur fonction : simple terminal, station de travail ou serveur. Aujourd'hui, le mini-ordinateur et le simple terminal sont rares sur les marchés traditionnels. Ils sont remplacés de plus en plus par le micro-ordinateur (voir Figure 5), dont les performances ont considérablement augmenté en termes de vitesse, de mémoire et de capacité. On préférera donc se concentrer sur le micro-ordinateur et considérer que l'ordinateur central n'est plus adapté en raison de sa taille et des coûts considérables qu'il générerait, pour répondre aux besoins de la plupart des SMHN.



Figure 5. Le micro-ordinateur (ou ordinateur individuel, PC) est l'outil le plus utilisé par les SMHN pour la gestion des données.

Les micro-ordinateurs peuvent être de plusieurs types :

- **ordinateur de bureau** : utilisé pour des applications bureautiques ;
- **station de travail** : plus puissante qu'un ordinateur de bureau et conçue pour des applications spécifiques ;
- **serveur** : conçu pour des applications critiques dans les entreprises et capable de fournir des informations à plusieurs clients ;
- **grappe de serveurs** : ensemble de plusieurs serveurs interconnectés.

Se reporter au glossaire pour avoir une définition plus précise de chacun de ces termes.

Donc, quel type de micro-ordinateur faut-il choisir ?

Une base de données étant une source d'information importante, c'est-à-dire une partie de l'héritage d'un pays en matière de climat, il est primordial de disposer d'un système solide et sûr.

Un serveur est plus solide, plus évolutif qu'un ordinateur de bureau, mais si le coût du matériel est une contrainte pour le SMHN, on peut éventuellement envisager de convertir certains ordinateurs de bureau en un serveur destiné au système de gestion de bases de données climatologiques.

L'Annexe 3 propose trois solutions, chacune avec un type de micro-ordinateur différent, et compare les avantages et les inconvénients de chacune.

Quel système d'exploitation?

Dans la mesure où l'on opte pour un micro-ordinateur, deux systèmes d'exploitation dominent actuellement sur le marché des systèmes de gestion de bases de données climatologiques : Windows et Linux.

L'environnement Microsoft Windows a presque le monopole mondial de l'ordinateur de bureau. Ce système d'exploitation, disponible dans le commerce, fournit un environnement de serveur aux bases de données et applications. Pour connaître les différentes versions de Windows, consulter le site : <http://www.microsoft.com/windows/>.

L'environnement Linux est celui qui se rapproche le plus d'Unix. Disponible en tant que logiciel à code source ouvert, Linux est couramment utilisé sur les serveurs web et les serveurs de bases de données. Les dernières versions sorties (voir <http://distrowatch.com/> pour les principales distributions) se sont bien améliorées ces dernières années et fournissent désormais des interfaces conviviales aux utilisateurs d'ordinateurs de bureau.

Le choix qui sera effectué entre ces deux systèmes dépendra du SMHN et des réponses aux questions suivantes :

- Quelles sont les compétences du personnel du SMHN ?
- L'environnement technique du système de gestion de bases de données climatologiques ou du SMHN est-il compatible avec Windows et/ou avec Linux ?
- Quel sera le coût réel ?

Il existe une troisième solution, qui a été adoptée par de nombreux SMHN et qui consiste à installer plusieurs ordinateurs avec des composants tournant sous Windows et d'autres sous Linux, selon l'application. Mais cette solution requiert du SMHN un vaste éventail de compétences.

Enfin, pour les SMHN qui utilisent actuellement des systèmes d'exploitation tels qu'UNIX, OS/400, MacOS ou autre système dédié à des ordinateurs centraux ou mini-ordinateurs spécifiques, il reste à analyser précisément s'il convient de continuer à utiliser ces systèmes ou plutôt de passer à Windows ou à Linux.

Quel SGBD ?

La plupart des systèmes de gestion de base de données (SGBD) utilisés actuellement par les SMHN pour gérer les données climatologiques sont des systèmes de gestion de base de données relationnelle (SGBDR). Cependant, vu la rapidité des évolutions technologiques, de nouveaux types de SGBDR apparaissent, comme les SGBD orientés objet, XML ou multidimensionnels.

Le présent document portera uniquement sur les SGBDR. Les plus courants sont répertoriés dans le Tableau 5.4.

Tableau 5.4. SGBRD existants

SGBDR	Entreprise	Système d'exploitation	Code source ouvert	Commentaires
Access	Microsoft	Windows	Non	Facile à utiliser et largement répandu parmi le grand public. Limité en termes de volume. Pas un vrai SGBD client/serveur mais plutôt fichier/serveur.
DB2	IBM	Linux, UNIX, Windows, OS400	Non	Compatible avec OS400, une référence pour beaucoup de grandes entreprises.
FileMaker	FileMaker	Windows, MacOS	Non	Facile à utiliser mais limité en termes d'évolutivité.
Firebird	Firebird	Linux, UNIX Windows	Oui	Basé sur Interbase.
Ingres	Computer Associates International	Linux, UNIX, Windows et OpenVMS	Oui	SGBDR très connu, diffusé avec code source ouvert.
Interbase	Borland	Linux, UNIX Windows	Non	Bon choix pour la gestion de petites et moyennes bases de données.
MaxDB (SAP)	MySQL	Linux, Unix, Windows	Oui/Non	Double licence ouverte et commerciale. Capable de gérer d'énormes volumes de données.
MySQL	MySQL	Linux, Mac OS X, Unix, Windows	Oui	Très répandu sur Internet. N'a pas encore toutes les fonctionnalités de certains autres SGBD.
Oracle	Oracle	Linux, Mac OS X, UNIX, Windows	Non	Comme DB2, c'est une référence pour les entreprises qui ont beaucoup de contraintes et d'énormes volumes de données à traiter.
PostgreSQL	PostgreSQL	Linux, Mac OS X, Unix, Windows	Oui	SGBD solide, l'une des principales références à code source ouvert.
SQL Anywhere Studio, Adaptive Server IQ	Sybase	Windows, Novell NetWare	Non	Version gratuite pour Linux : Adaptive Server Enterprise Express.
SQL Server	Microsoft	Windows	Non	Bon choix pour une organisation de taille moyenne.

Les aspects suivants devront être pris en compte dans le choix d'un SGBDR : fonctions du SGBDR, capacité d'adaptation au milieu du SMHN, coût, compatibilité (matériel, logiciel) et compétences disponibles au SMHN.

Aspects liés au changement de système et de services

La transition vers une nouvelle base de données climatologiques peut faire apparaître de nouveaux outils et de nouvelles tâches susceptibles d'avoir des répercussions sur le travail quotidien du personnel. Cette période transitoire est une bonne occasion de réexaminer la politique de sécurité et le partage du travail au sein de l'organisation.

Sécurité du système et des logiciels

La sécurité est l'une des principales difficultés auxquelles sera confronté un administrateur de base de données. Dans le pire des cas, le SMHN peut perdre l'ensemble de sa base de données si les dispositions de sécurité en place ne sont pas suffisantes.

Voici quelques recommandations de base :

Évitez toute perte de données en effectuant des sauvegardes : sur NAS (Network Attached Storage, système de stockage lié à un réseau), bande (DAT, DLT, LTO), CD-Rom ou DVD.

Évitez toute perte de données grâce à l'archivage : il est possible d'utiliser les mêmes technologies que pour la sauvegarde. Conservez trois copies de la même archive dans trois lieux sûrs différents, si possible dans des villes différentes. Stockez les données climatologiques sous un format facilement lisible (ASCII).

Protégez la salle des serveurs : accès sécurisés, sécurité incendie et contre la foudre. Respectez les consignes de sécurité figurant dans le manuel d'utilisation de tous les dispositifs matériels (humidité, température, poussière, etc.).

Évitez toute défaillance des disques : clonage de disques ou technologie appropriée (RAID).

Alimentation électrique : ajoutez un système d'alimentation à temps de coupure zéro (UPS) au système qui aura ainsi suffisamment de temps pour se mettre correctement hors tension. Envisagez d'utiliser l'énergie solaire si les coupures de courant sont fréquentes. Vérifiez l'électrode de mise à la terre du bâtiment et vérifiez que les lignes électriques comportent le fil à la terre adéquat.

Formez le personnel sur la sécurité et assurez-vous qu'il dispose de la documentation nécessaire.

Ordinateurs : ajoutez une alimentation électrique redondante et un système de refroidissement redondant sur les ordinateurs qui ont une importance capitale. Il conviendrait également d'envisager carrément un système informatique redondant.

Sécurité des logiciels et des réseaux : installez des programmes pare-feu, anti-virus et de protection contre les logiciels malveillants. Assurez-vous que ces programmes sont régulièrement mis à jour. Attribuez à chaque utilisateur un profil avec des droits spécifiques.

Pièces : prévoyez des consommables pour une période suffisante et vérifiez la disponibilité des pièces de rechange.

Fonctionnement au quotidien

Responsabilité des processus : chaque processus devra être décrit et placé sous la responsabilité d'une personne désignée à cet effet. Si nécessaire, le processus pourra être divisé en sous-processus confiés à des agents désignés. Il sera donc relativement courant de trouver, dans un service informatique, un administrateur système responsable de la partie opérationnelle et de la maintenance des systèmes d'exploitation et du matériel, avec à ses côtés un administrateur de base de données qui sera chargé de la conception, de la mise en œuvre et de la gestion des opérations de la base de données. Un administrateur réseau s'assurera que le réseau local (LAN) du SMHN est correctement réglé pour fournir des performances et une sécurité optimales et que ce réseau communique parfaitement avec Internet.

Documentation : chaque processus et enfin toute installation de matériel ou logiciel ainsi que l'historique de ses modifications devront être enregistrés (voir 3.5). Cela concerne en particulier les processus de sauvegarde et d'archivage ou toute modification du système d'exploitation du serveur, indispensables à la survie du système.

Transition à partir de CLICOM

Ce chapitre concerne globalement les SMHN qui utilisent un système CLICOM et qui ont décidé de passer à un nouveau système de gestion de bases de données climatologiques. Cependant, certaines parties de ce chapitre s'appliquent également à des SMHN qui disposent déjà d'une collection de données climatologiques sous un autre format (par exemple, sous la forme de classeurs de tableur) et qui souhaitent convertir ces données dans le nouveau système. On suppose ici que les SMHN ont analysé leur environnement existant (voir 5.3.1), qu'ils ont défini une solution fonctionnelle (voir 5.3.2) et qu'il ont déjà prévu le matériel et le logiciel nécessaires, notamment les SGBD et le système de gestion de bases de données climatologiques (voir 5.1, 5.2 et 5.3).

Maintenant, quelles sont les procédures à suivre pour effectuer la transition ?

Niveau de compétences requis

La transition entre un système CLICOM et un autre système de gestion de bases de données climatologiques fait appel à des compétences aussi bien sur CLICOM que sur le nouveau système choisi pour le remplacer. Premièrement, il est important de bien comprendre les règles climatologiques en place dans les deux systèmes pour que la transition soit cohérente. C'est le cas notamment pour la définition des indicateurs de qualité et pour les méthodes utilisées (contrôle, génération des données) qui doivent refléter, à la fin du processus, les pratiques des SMHN. Deuxièmement, on a besoin aussi de compétences sur les SGBD, notamment d'un savoir-faire parfaitement maîtrisé sur les modèles de données et leurs particularités (intégrité référentielle, clé primaire, contraintes, etc.), afin d'importer correctement toutes les données CLICOM dans les systèmes de gestion de bases de données climatologiques sélectionnés. Certains de ces systèmes peuvent avoir déjà une fonction d'importation à partir de CLICOM, ce qui facilitera cette étape.

D'un point de vue technique, les méthodes de transition sont diverses et peuvent faire appel à différents outils tels que les fonctions d'exportation de CLICOM, le langage DQL (DataEase Query Language) (SGBD CLICOM) et aussi à des fonctions d'importation du SGBD (SQL Loader) ou à des interfaces normalisées entre les bases de données.

Conservation du système CLICOM existant

Il est important de conserver le système CLICOM existant (matériel, logiciel, données climatologiques et métadonnées) en lieu sûr et en condition opérationnelle jusqu'à ce que l'on prouve que le nouveau système fonctionne et délivre les performances souhaitées. Pendant la période transitoire, il peut être nécessaire que les deux systèmes – CLICOM et le nouveau système – tournent parallèlement pendant un certain temps.

Préparation des métadonnées à importer

Les métadonnées sont essentielles pour comprendre parfaitement les données climatologiques et doivent être bien gérées. Généralement, dans un système de gestion de bases de données

climatologiques, les métadonnées doivent être présentes avant la saisie des valeurs climatologiques. Ces informations sont stockées à deux endroits dans CLICOM : l'historique de la station et le dictionnaire des données.

La solution pourrait être d'exporter les métadonnées CLICOM au format ASCII puis de les importer dans le nouveau système de gestion de bases de données climatologiques. La procédure d'importation dépendra du système choisi et du modèle de métadonnées qu'il utilise. Il sera probablement nécessaire de procéder à un traitement spécifique à ce stade, selon le degré de conformité des structures de données CLICOM au nouveau système de gestion de bases de données climatologiques.

Migration des informations relatives à l'historique de la station

Cette partie contient des informations anciennes et actuelles sur l'emplacement de la station, les pratiques d'observation et les instruments, enregistrées sous trois formats DataEase différents : STN GEOGRAPHY, STN OBSERVATION et STN ELEMENT. Les informations présentes sous ces formats peuvent être exportées dans des fichiers ASCII en utilisant de simples rapports DQL.

Les conventions relatives aux noms ou aux codes (par exemple, dans CLICOM, la date de clôture d'une station en service est indiquée sous la forme « 9999-12-31 ») devront être remplacées par celles utilisées dans le nouveau système de gestion de bases de données climatologiques.

Migration des informations du dictionnaire de données

Cette seconde migration porte notamment sur :

- les définitions des éléments, selon le format DataEase « ELEMENT DEFINITION », qui désigne chaque élément climatologique géré par CLICOM ;
- la description des archives CLICOM, selon le format DataEase « DATASET INFORMATION » ;
- les définitions des indicateurs de qualité, utilisées par CLICOM pour le format DataEase « MISC CODE DEFINITION ».

Il convient de se pencher en particulier sur les différences éventuelles entre les indicateurs de qualité CLICOM et ceux du nouveau système de gestion de bases de données climatologiques. Les principales difficultés liées à la transition reviennent à se demander :

- si les résultats de tous les contrôles de qualité réalisés avec CLICOM, c'est-à-dire la valeur des données plus les informations relatives aux indicateurs de qualité correspondants, ont été conservés ;
- si une certaine perte d'informations peut être inévitable et donc si les SMHN doivent se préparer à la gestion des risques.

Préparation des données climatologiques à importer

Conservation des informations relatives aux indicateurs de qualité

Les informations météorologiques liées à chaque valeur de la base de données CLICOM, c'est-à-dire l'indicateur de qualité, doivent être conservées dans la mesure du possible. C'est le cas en particulier des valeurs suspectes (indicateur de qualité D), des valeurs estimées (indicateur de qualité E), des valeurs générées (indicateur de qualité G), des résultats basés sur une

période incomplète (indicateur de qualité I), des valeurs manquantes (indicateur de qualité M) et du volume de précipitations à l'état de traces (indicateur de qualité T). S'il perdait ces informations, le SMHN serait obligé :

- de vérifier une nouvelle fois la qualité de toutes les valeurs dans le nouveau système de gestion de bases de données climatologiques, ce qui impliquerait une hausse importante de sa charge de travail ;
- de se reporter aux données d'origine, qui pourraient être les « archives papier », afin de retrouver des informations telles que des traces de précipitations.

Importation des données CLICOM

Les données CLICOM sont stockées sous les formats DataEase suivants :

- MONTHLY DATA : valeurs d'éléments mensuels ;
- TEN DAY DATA : valeurs d'éléments de 10 jours ;
- DAILY DATA : valeurs d'éléments quotidiens ;
- SYNOPTIC DATA : valeurs de trois éléments horaires ;
- HOURLY DATA : valeurs d'un élément d'une heure ;
- FIFTEEN MINUTE DATA : valeurs d'éléments de 15 minutes ;
- UPPER-AIR DATA : valeurs d'éléments en altitude.

Pendant l'importation des données vers le nouveau système de gestion de bases de données climatologiques, il convient de traiter avec soin les valeurs manquantes dotées de l'indicateur de qualité « M » et de la valeur « 99999 », car ces dernières ont deux significations dans CLICOM :

- valeur absente car aucune observation n'a été réalisée à ce moment-là (par exemple en cas de dysfonctionnement d'un instrument) ;
- valeur absente car la date n'existe pas (par exemple, dans CLICOM, le 31 février est indiquée comme une valeur manquante).

En fonction du modèle de données du nouveau système de gestion de bases de données climatologiques, un SMHN peut exporter ses données climatologiques de CLICOM vers les deux formats ASCII suivants :

- format d'exportation classique CLICOM (via le menu CLICOM # 2.2.3), qui générera un fichier ASCII avec une structure Élément ;
- formation d'exportation d'observation CLICOM (via le menu CLICOM # 2.2.4), qui générera un fichier ASCII avec une structure Observation.

Collecte d'informations sur les réglages du contrôle de la qualité de CLICOM

Les méthodes de contrôle de la qualité qui ont été appliquées sur la base de données CLICOM doivent être enregistrées comme faisant partie des métadonnées du nouveau système de gestion de bases de données climatologiques. Ces méthodes retracent l'historique du contrôle de la qualité mis en place par le SMHN et peuvent s'avérer utiles pour améliorer les algorithmes correspondants ou simplement reprendre les mêmes technique avec le nouveau système.

Les réglages du contrôle de la qualité CLICOM sont stockés dans les fichiers « ELEMCHKS.* » (ELEMCHKS.DLY pour les données quotidiennes, ELEMCHKS.HLY pour les données horaires, etc.), dans le répertoire CLICOM/DATA.

Essais approfondis du nouveau système

Enfin, le nouveau système doit être soumis à des essais approfondis par rapport à l'ancien système. Les vérifications servent à s'assurer que toute perte d'information est acceptable pour l'organisation et que le nouveau système répond aux pratiques officielles du SMHN et aux attentes du client.

Soutien aux opérations de gestion des données

Ressources nécessaires, notamment en termes de personnel

Comme indiqué au chapitre 5 ci-dessus, un SMHN est tout d'abord confronté au choix suivant : « acheter ou construire » un système de gestion de bases de données climatologiques. Dans l'un ou l'autre cas, mais surtout si l'on décide de construire, on devra faire appel à des ressources considérables. Si l'on décide d'acheter, c'est le choix du logiciel qui déterminera, ou du moins influencera le choix du matériel. Le matériel pourra alors être choisi en fonction de ce type de contrainte, sans négliger des aspects tels que le volume des données à gérer, le budget disponible et le degré de solidité requis.

Un des aspects importants à prendre en compte est la gestion des versions. Quel que soit le système adopté, il sera essentiel de gérer les versions de tous les composants logiciels. Malheureusement, de nombreux problèmes peuvent surgir, entraînant des incompatibilités entre des versions de logiciels utilisés en parallèle. C'est pourquoi, dans le cas spécifique de logiciels à code source ouvert, il peut être souhaitable de faire appel aux services d'un fournisseur de solutions au code source ouvert, tels que Red Hat™ ou à l'un de ses concurrents, pour réduire le risque de connaître ce genre de problèmes.

L'établissement d'un système de gestion de données climatologiques demande beaucoup d'investissement en personnel. Le type de personnel requis entre dans plusieurs catégories. Nous avons tout d'abord, si l'on suit le flux des données (et si l'on suppose que les observateurs existent déjà) :

- des opérateurs de préparation des données, chargés de saisir les données qui arrivent sous forme papier ;
- du personnel ayant des connaissances en météorologie, capable d'effectuer le contrôle de la qualité et d'interpréter les besoins météorologiques comme le demande de centre de données (ex : anciens observateurs ou personnes capables de contribuer à l'identification des besoins et à la conception de systèmes tels que de nouveaux systèmes de contrôle de la qualité) ;
- (probablement) du personnel ayant une bonne connaissance de la base de données afin de mener à bien les extractions de données les plus complexes.

On a besoin également de personnel en « arrière-scène » (ne possédant pas nécessairement de connaissances précises en météorologie) :

- administrateurs de base de données ;
- administrateurs système ;
- si le SMHN souhaite réaliser lui-même les extensions du système spécifiques à son pays, il aura besoin de personnel informatique pour maintenir et faire évoluer le logiciel d'application. Bien évidemment, si le SMHN choisit de développer son propre système, il convient de prévoir des investissements plus importants en termes de personnel informatique.

Formation

Les besoins en matière de formation à la gestion de données climatologiques découlent naturellement de l'étendue des responsabilités correspondantes du personnel. Les utilisateurs d'un système ont toujours besoin de formation sur les aspects du système qui les concernent, par exemple :

- le personnel chargé de la saisie des données et du contrôle de la qualité devra comprendre de quelle façon le système réagira à la saisie de données apparemment suspectes ou erronées et savoir quelles mesures correctives devront être prises ;
- s'il est possible qu'un système soit considéré comme une « boîte noire » du point de vue de l'extraction de données ; en revanche, avec des données disponibles uniquement sous un petit nombre de formats prédéfinis et avec des options de sélection prédéfinies, on risque davantage de demander au personnel d'être capable de développer ses propres processus d'extraction ; si le système de gestion de bases de données climatologiques est fondé sur un SGBDR (comme les six systèmes actuels du PMDSC), on pourra extraire les données via le langage SQL, dont le personnel devra alors maîtriser les possibilités et les limites ;
- des compétences seront également requises dans les domaines « infrastructure » de l'administration du système d'exploitation et de l'administration de bases de données ; dans un établissement de petite taille, ces compétences pourront se trouver chez un seul individu (qui aura reçu la formation adéquate) ; cependant, chacune de ces compétences relève d'un domaine à part entière, et dans une organisation plus grande, on exigera un ou plusieurs spécialistes par domaine ;
- le personnel informatique devra se familiariser avec les questions météorologiques, ces connaissances étant fournies par le personnel approprié du SMHN.

Santé et sécurité au travail

Les systèmes de gestion de données climatologiques ne posent pas de problèmes spécifiques liés à la santé et à la sécurité au travail (si ce n'est, à titre exceptionnel, la poussière générée par les archives dans l'exercice de sauvetage de données et les risques liés au portage de lourds documents d'archive). Cependant, il convient de prendre en compte un certain nombre d'aspects valables pour la plupart des systèmes informatiques :

- problèmes d'ergonomie liés aux stations de travail, à la position par rapport au clavier, mobilier, etc. ; un mobilier inadapté et des pratiques de travail non appropriées peuvent entraîner un vaste éventail de problèmes physiques ;
- risques de rayonnement par l'arrière des moniteurs ;
- simples risques physiques dus à la présence de câbles et de conducteurs en grand nombre, comme c'est le cas souvent des installations informatiques ;
- stress, problème commun à de nombreux postes informatiques.

Les pratiques des organisations en matière de santé et de sécurité au travail seront régies par la législation en vigueur dans un grand nombre de pays. Pour avoir plus d'informations, il convient de se renseigner précisément sur ces pratiques auprès des personnes responsables.

Conclusion

Ces principes directeurs sont destinés à fournir des informations aux SMHN sur les meilleures pratiques en matière de gestion des données climatologiques, mettant l'accent en particulier sur l'acquisition de connaissances et de compétences dans les domaines suivants :

- phases et étapes de la gestion de données numérisées ;

- technologies de base de données disponibles et sélection de la base de données qui convient ;
- transition vers un système moderne de bases de données climatologiques ;
- soutien aux opérations de gestion des données.

Dans ce document, on a mis délibérément l'accent sur les aspects de la gestion de données climatologiques présentant un intérêt pour les SMHN qui souhaitent passer à un système de gestion de bases de données climatologiques moderne, et, ce qui est tout aussi important, sur les compétences, systèmes et processus à mettre en place pour s'assurer que les opérations sont correctement soutenues. Ce document vise à aider les SMHN à évaluer leur future orientation stratégique face aux besoins liés à la gestion des données climatologiques.

Il est essentiel que, dès le début, les besoins des utilisateurs existants et futurs, dans la mesure où ils sont prévisibles, soient pris en compte pour le développement de bases de données climatologiques et pour la mise en œuvre de pratiques de gestion de données. Les gestionnaires de données doivent également s'attacher à respecter les principes de surveillance du climat à long terme. La capacité à assurer la continuité, l'homogénéité et enfin la qualité des données climatologiques dépend dans une large mesure de la qualité de gestion des systèmes et des réseaux d'observation.

Il est également indispensable de pouvoir stocker les données climatologiques ayant subi un contrôle de qualité, dans un format facilement accessible, afin de répondre aux besoins des programmes sur le climat au sein des SMHN et d'apporter un soutien à différents programmes de l'OMM. Le relevé climatologique est une ressource pour le présent et pour l'avenir et il est important de sauvegarder ces informations pour les futures générations.

L'une des principales décisions qui devra être prise sur le plan technologique est de savoir s'il faut opter pour un concept existant de système de gestion de bases de données climatologiques, avec le logiciel associé, ou bien s'il faut développer un système sur mesure, à partir de l'un des SGBDR disponibles dans le monde aujourd'hui. Selon toute probabilité, cette décision reposera sur des critères financiers car le développement d'un système de gestion de bases de données climatologiques avec le logiciel associé peut demander un investissement initial conséquent en amont. Cependant, il conviendra aussi de tenir compte d'aspects tels que le flux des données et le contrôle de la qualité, ainsi que de facteurs technologiques. Il est donc fortement recommandé aux SMHN d'envisager très prudemment d'acheter le système plutôt que de le construire car cette dernière option demande beaucoup plus de ressources et reviendra généralement beaucoup plus cher.

Remerciements

Les auteurs tiennent à remercier Amir Delju (OMM) et Suresh Boodhoo (Président de la CIC) pour le soutien qu'ils ont apporté à ce travail. Merci également à Rod Hutchinson et Mike Macaskill (tous les deux du Bureau météorologique australien) pour la révision de ce document.

Bibliographie

Benichou, F. et Lee, D.M. 1996. Use of Relational Management Systems in Climatology in WMO National Meteorological Services, Rapport OMM. Organisation météorologique mondiale, Genève.

Date, C.J. 1999. An Introduction to Database Systems (7th Ed). Addison-Wesley, 938pp.

GCOS. 2003. Second Report on the Adequacy of the Global Observing System for Climate. GCOS Report No 82 WMO-TD 1143. Organisation météorologique mondiale, Genève.

<http://www.wmo.ch/web/gcos/gcoshome.html>

OMM. 1995. Manuel des codes : Codes internationaux, Volume I.1 (Partie A). Organisation météorologique mondiale, Genève.

<http://www.wmo.ch/web/www/DPS/NewCodesTables/WMO306vol-I-1PartA.pdf>

WMO. 2002. Report of the CLICOM-DARE workshop; Report of the international data rescue meeting, WCDMP-No.49/WMO-TD No.1128. Organisation météorologique mondiale, Genève..

http://www.wmo.ch/web/wcp/wcdmp/reports/WCDMP-49/WDMP49_menu.html

WMO. 2003a. Guidelines on Climate Observation Networks and Systems, WCDMP-No.52, WMO-TD No.1185. Organisation météorologique mondiale, Genève.

<http://www.wmo.ch/web/wcp/wcdmp/reports/WCDMP-52.pdf>

WMO. 2003b. Guidelines on Climate Metadata and Homogenization, WCDMP-No.53/WMO-TD No.1186. Organisation météorologique mondiale, Genève.

<http://www.wmo.ch/web/wcp/wcdmp/reports/WCDMP-53.pdf>

WMO. 2004. Guidelines on Climate Data Rescue, WCDMP-No.55/WMO-TD No.1210. Organisation météorologique mondiale, Genève.

<http://www.wmo.ch/web/wcp/wcdmp/reports/WCDMP-55.pdf>

OMM. 2005. Guide des pratiques climatologiques (troisième édition). Organisation météorologique mondiale, Genève. (en préparation).

<http://www.wmo.ch/web/wcp/ccl/guide/guide.3e.shtml>

WMO. 2006. WMO Information System.

<http://www.wmo.ch/web/www/WISweb/home.html>

Glossaire

ACCESS	Systeme de gestion de bases de données relationnelles de Microsoft. Fonctionne généralement sur une station de travail locale, mais peut aussi être installé sur un serveur et donc partagé.
Programme anti-logiciels espions	Programme informatique destiné à détecter les programmes indésirables connus sous le nom de logiciels espions et à s'en protéger.
ASCII	American Standard Code for Information Interchange. Norme de codage de caractères : lettres, chiffres, signes de ponctuation et autres symboles, en informatique.
Octet	La plus petite unité adressable en informatique, composée de 8 bits, qui peut donc avoir 2^8 ou 256 valeurs différentes.
Processeur	Principale unité de traitement d'un ordinateur, désignée généralement par le nombre de cycles de traitement par seconde et la largeur de bande de son bus. Pour une configuration donnée, plus le nombre de cycles est élevé, plus les opérations et les réponses sont rapides.
CLICOM	CL imate COM puting project. – Base de données climatologiques développée sous l'égide de l'OMM et destinée à fonctionner sous un système d'exploitation Microsoft, DOS ou Windows.
Base de données client/serveur	Base de données stockée sur un seul serveur, mais où les applications utilisatrices des données sont exécutées sur les ordinateurs de plusieurs clients.
CLIMAT	Format numérique international d'échange d'informations sur le climat à partir de certains sites d'observation en surface. Les messages sont générés et échangés une fois par mois entre les Membres de l'OMM.
CLIMAT TEMP	Type de message identique à CLIMAT mais avec des données issues d'observations aérologiques.
Données climatologiques	Données utilisées pour décrire ou comprendre le système climatique ; il peut s'agir de données liées aux propriétés physiques, chimiques et biologiques, décrivant des processus atmosphériques, océaniques, hydrologiques, cryosphériques et terrestres. Les attributs de continuité et d'homogénéité ont généralement un niveau de priorité élevé.
Gestion des données climatologiques	Activité consistant à gérer les données climatologiques de manière à fournir des données de qualité fiable aux utilisateurs, stockées généralement dans une base de données informatisée. Cette notion peut englober les outils utilisés pour fournir des produits climatologiques au client.
Grappe de serveurs	Voir Serveur. Ensemble de plusieurs serveurs interconnectés, généralement avec un serveur frontal, qui traite toutes les liaisons avec les utilisateurs. Les grappes de serveurs sont évolutives, ce qui permet d'augmenter relativement facilement leurs capacités.
Type d'ordinateur	Classification générale des ordinateurs en fonction de leur taille et de leurs performances. Le type d'ordinateur va de l'ordinateur central le plus puissant, multi-utilisateurs, au simple PC (ordinateur de bureau), en passant par le milieu de la gamme des mini-ordinateurs.
Modèle de base de données	Type de conception de la base de données. Il existe un certain nombre de modèles possibles, le plus courant étant le modèle relationnel. Autres modèles : modèle objet-relationnel, modèle orienté objet, modèle semi-structuré, modèle associatif, modèle Entité Attribut Valeur (EAV) et modèle contextuel.
Sauvetage de données	Processus de conservation de toutes les données risquant d'être perdues suite à une détérioration du support, et numérisation des données actuelles et passées sous un format informatique

	compatible et facilement accessible.
Base de données	Ensemble d'informations organisées de manière à ce qu'un programme informatique puisse sélectionner rapidement les informations souhaitées. Les informations qui se trouvent dans une base de données sont généralement reliées entre elles.
Système de gestion de base de données (SGBD)	Série de logiciels utilisés pour développer, mettre en œuvre, gérer et maintenir des données stockées dans une base de données.
Bureau	Terme générique désignant l'affichage en arrière-plan d'un ordinateur, où se trouvent les icônes représentant les programmes et les disques et à partir duquel les applications peuvent être lancées.
Ordinateur de bureau	Ordinateur indépendant à usage général. Comme son nom l'indique, il s'installe en fixe sur un bureau, contrairement à un ordinateur portable. Les ordinateurs de bureau sont généralement modulaires, ils n'ont pas qu'une seule et unique configuration possible.
Base de données distribuée	Base de données gérée par un système réparti sur plusieurs nœuds ou processeurs, mais géré en tant qu'entité unique.
Double saisie	Méthode de saisie de données destinée à minimiser les erreurs de frappe. La même donnée est saisie deux fois par deux opérateurs différents, ceci en vertu de la théorie selon laquelle la probabilité que deux opérateurs fassent la même erreur est faible.
Excel	Tableur développé et commercialisé par Microsoft.
Pare-feu	Logiciel conçu pour éviter tout accès non autorisé en provenance ou à destination d'un réseau privé. (Il peut s'agir d'un simple PC.) En présence d'un réseau, ce logiciel est installé sur le serveur passerelle, protégeant le réseau contre des utilisateurs venant d'autres réseaux.
Go	Voir Gigaoctet
SMOC	S ystème M ondial d' O bservation du C limat. Système opérationnel à long terme, orienté utilisateurs, capable de fournir les observations complètes nécessaires pour la surveillance du système climatique, pour détecter et expliquer les changements climatiques, pour évaluer les conséquences de la variabilité du climat, notamment des changements climatiques, et pour permettre aux chercheurs de mieux comprendre, modéliser et prévoir le système climatique.
Gigaoctet (Go)	Unité désignant la capacité de stockage de matériel informatique. Un gigaoctet (Go) est égal à 1024 (2 ¹⁰) mégaoctets (Mo).
GSN	GCOS Surface Network : réseau de stations d'observation en surface pour le SMOC.
SMT	S ystème M ondial de T élécommunications. Réseau mondial de télécommunications utilisé par les Membres de l'OMM pour échanger des données et produits météorologiques et connexes.
GUAN	GCOS Upper Air Network : réseau de stations d'observation en altitude pour le SMOC.
IGU	I nterface G raphique U tilisateur d'un écran d'ordinateur. Système basé sur l'affichage de diverses fenêtres, à partir desquelles on peut lancer directement des actions au moyen de commandes telles que des boutons, cases à cocher, barres de menus et listes.
Disque dur	Dispositif de stockage permanent installé dans un ordinateur et se composant d'une ou plusieurs plaques à surface magnétique. L'espace disponible sur un disque dur est réutilisable. Le système d'exploitation et les autres logiciels sont généralement stockés sur le disque dur.
MacOS	Système d'exploitation utilisé par Apple dans ses ordinateurs Macintosh.

Ordinateur central	Terme désignant généralement un gros ordinateur très puissant, capable de réunir plusieurs utilisateurs (souvent des centaines) et d'exécuter plusieurs programmes simultanément. Les ordinateurs centraux sont généralement installés dans un environnement sécurisé et contrôlé. On note de moins en moins de différence évidente entre un ordinateur central bas de gamme et un mini-ordinateur ou serveur haut de gamme.
Mo	Voir mégaoctet.
Mégaoctet (Mo)	Unité désignant la capacité de stockage de matériel informatique, considérée généralement comme équivalant à 1 048 576 octets (1024^2 or 2^{20}). Selon l'usage, un mégaoctet peut être égal à 1 000 000 ou à 1 024 000 octets.
Métadonnées	Données se rapportant à des données. Il s'agit d'informations relatives aux données. Pour des données météorologiques, il peut s'agir du type d'instrument de mesure, d'informations sur le réglage des instruments, etc. ainsi que des informations sur le contenu et le format des jeux de données.
Paramètre météorologique	Synonyme de donnée météorologique ou élément d'intérêt météorologique tel que la température.
Micro-ordinateur	Synonyme d'ordinateur de bureau et d'ordinateur individuel (voir la définition de ces deux termes).
Mini-ordinateur	Catégorie d'ordinateurs entre les ordinateurs centraux et les micro-ordinateurs. Ce terme est apparu à une époque où l'on distinguait nettement les ordinateurs centraux, les mini-ordinateurs et les micro-ordinateurs. C'est de moins en moins le cas aujourd'hui, et ce terme désigne souvent un serveur ou une station de travail haut de gamme.
MV	M oyen de V érification
SMHN	S ervice M étéorologique et H ydrologique N ational
Normalisation	Processus appliqué à la conception des tables d'une base de données relationnelle, afin d'établir une structure adaptée, minimisant les redondances inutiles. Cinq formes de normalisation ont été définies.
ODBC	O pen D ata B ase C onnectivity : connectivité ouverte de base de données. Norme autorisant l'accès à différents systèmes de base de données, par exemple l'accès à Oracle depuis Microsoft Excel ou Access.
À code source ouvert	Qualité d'un logiciel dont le code source est facilement accessible à d'autres, qui peuvent ainsi le voir et le modifier. Cela signifie que le logiciel est disponible gratuitement (ou parfois seulement à un prix correspondant au coût de distribution).
Système d'exploitation	Logiciel système installé sur un ordinateur et responsable du contrôle et de la gestion du matériel, des opérations et des processus système. Le système d'exploitation contrôle le fonctionnement des logiciels d'application, d'un traitement de texte par exemple. Les systèmes d'exploitation pour PC les plus connus sont Microsoft Windows, Mac OS (Apple) et Linux.
OS/400	Système d'exploitation utilisé par IBM pour sa série d'ordinateurs AS/400.
IOV	I ndicateur O bjectivement V érifiable
Ordinateur individuel, PC	Ordinateur indépendant à usage général, pouvant être utilisé de manière autonome. Ce terme s'applique souvent à des ordinateurs IBM ou compatibles, utilisant les systèmes d'exploitation Microsoft, mais il peut désigner tout ordinateur indépendant à usage général.
Plate-forme	Ensemble composé du matériel, du logiciel et du système d'exploitation utilisés. Souvent, il désigne seulement le système d'exploitation.

AQ	A ssurance de la Q ualité
CQ	C ontrôle de la Q ualité
RAID	R edundant A rray of I ndependent D isks : réseau redondant de disques indépendants. Regroupement de disques de stockage de données en une seule unité assurant une redondance en cas de défaillance de l'un des disques et réduisant ainsi le risque de perdre des données sur ce disque. Généralement, les données présentes sur chacun des disques sont dupliquées sur un ou plusieurs autres disques, agissant ainsi comme des disques miroirs.
RAM	R andom A ccess M emory : mémoire vive. Mémoire informatique permettant un accès plus rapide aux données stockées, en lecture et en écriture, mais ne permettant pas de conserver les données sur le disque une fois l'ordinateur éteint. En général, plus l'ordinateur a de mémoire RAM, plus il sera performant.
RCBR	R éseau C limatologique de B ase R égional
RSBR	R éseau S ynoptique de B ase R égional
SGBDR	S ystème de G estion de B ase de D onnées R elationnelle. Terme souvent utilisé abusivement pour désigner une base de données relationnelle alors qu'il désigne uniquement le système utilisé pour gérer la base de données et non pas la base de données en elle-même.
Base de données relationnelle	Base de données où les données sont stockées dans des tables reliées entre elles et qui permet d'extraire ou de manipuler des données issues de différentes tables car le lien entre ces tables est clairement défini.
Serveur	Ordinateur accessible et utilisable par au moins deux utilisateurs via un réseau. Les serveurs sont généralement évolutifs et peuvent être configurés pour répondre aux besoins d'une organisation et évoluer pour répondre aux besoins à venir. Un serveur peut avoir plusieurs processeurs et périphériques, par exemple plusieurs disques durs.
SMART	Spécifique, Mesurable, Atteignable, Réaliste et limité dans le Temps : propriétés utiles pour mesurer les performances d'une base de données.
Logiciel espion	Terme générique désignant des programmes indésirables généralement installés à l'insu de l'utilisateur, via Internet. Ces programmes pilotent à distance des actions qui sont effectuées sur l'ordinateur et collectent des informations sur l'utilisation qui est faite de l'ordinateur. Ils se manifestent notamment par l'apparition de fenêtres publicitaires (« pop up »), par un ralentissement général de l'ordinateur, par l'apparition de barres de menus intempestives dans les navigateurs et via des logiciels de publicité (« Adware »).
SQL	S tructured Q uery L anguage : langage informatique de requête et d'extraction de données dans et à partir d'une base de données.
UPS	U ninterruptible P ower S upply : système d'alimentation électrique à temps de coupure zéro. Système d'alimentation électrique de secours en courant continu, utilisé en cas de coupure de l'alimentation normale. Garantit ainsi que l'ordinateur continue de fonctionner puis s'arrête correctement si la coupure de courant est plus importante.
OMM	O rganisation M étéorologique M ondiale
Station de travail	Ordinateur haut de gamme à usage général, généralement plus puissant et plus solide qu'un ordinateur de bureau. Une station de travail est souvent utilisée pour des applications d'ingénierie ou des applications graphiques nécessitant beaucoup de mémoire et un processeur rapide. Elle peut comporter plusieurs processeurs et plusieurs disques durs.

Annexe 1 - Description des systèmes de gestion de bases de données climatologiques proposés

Les éléments présentés ici sont tirés, pour la plupart, du document figurant à l'adresse suivante : <http://www.wmo.ch/web/wcp/wcdmp/cdmsfuture/html/evaluation.html>. Les autres systèmes connus font d'objet de courts paragraphes ou bien de liens web permettant d'en savoir davantage.

D'après les nombreuses réunions d'experts qui se sont tenues à ce sujet et les réponses faisant suite au questionnaire envoyé à l'ensemble des Membres de l'OMM, il se trouve que six SMHN ont proposé de remplacer CLICOM par un système de gestion de bases de données climatologiques : la **République tchèque** propose le système *CLIDATA*, la **France** *CliSys*, la **Jordanie** *JCDMS*, la **Fédération de Russie** *CLIWARE*, la **Tunisie** *SDCLIM* et le **Zimbabwe** *ClimSoft* (Tableau A1.1).

Suite aux recommandations formulées lors de la cinquante-troisième session du Conseil exécutif et de la treizième session de la Commission de climatologie (novembre 2001), un atelier a été organisé au Secrétariat de l'OMM du 27 mai au 1^{er} juin 2002, réunissant une équipe d'experts externes, chargés d'évaluer les six systèmes proposés. Pour chaque système, un rapport a été établi, avec un tableau comparatif de critères. Ce rapport est disponible à l'adresse suivante : <http://www.wmo.ch/web/wcp/wcdmp/cdmsfuture/html/evaluation.html>.



Figure A1.1. Équipe d'évaluation du projet de système de gestion de bases de données climatologiques pour l'OMM

Tableau A1.1. Systèmes de gestion de bases de données climatologiques, pris en compte dans le projet de l'OMM

Système de gestion de bases de données climatologiques	SMHN	Lien	Système de gestion de base de données
CLIDATA	République tchèque	http://www.clidata.cz/	Oracle
CLIMSOFT	Zimbabwe	http://www.met-elearning.org/moodle/course/category.php?id=4	Access, SQL/Server
CLISYS	France	http://www.mfi.fr/productsandsolutions.htm	Oracle, PostgreSQL
CLIWARE	Fédération de Russie	http://cliware.meteo.ru/meteo/index_en.html	Oracle, MaxDB, etc.
JCDMS	Jordanie	http://www.jmd.gov.jo/	Oracle
SDCLIM	Tunisie	http://www.meteo.tn	Oracle

D'autres systèmes ont été développés par des SMHN ou par des entreprises privées, en dehors du processus d'évaluation de l'OMM (voir Tableau A1.2).

Tableau A1.2. Systèmes de gestion de bases de données climatologiques développés en dehors du projet de l'OMM

Système de gestion de bases de données climatologiques	SMHN ou entreprise privée	Lien
AMS	Organisation pour l'alimentation et l'agriculture	http://www.foodsecinfoaction.org/News/tools_cw_02.htm
CLDB	Microstep-MIS	http://www.microstep-mis.com/products/meteorology/cldb/english.php
QUALIMET	Ernst Basler + Partner GmbH	http://www.ebp-de.de/en/geschaeftsbereiche/information/projekte/165/

Annexe 2 - Exemple de mise en œuvre d'un système de gestion de bases de données climatologiques

En 1988, le mini-ordinateur utilisé pour la gestion des données climatologiques à Madagascar tombe en panne. Il n'est pas remplacé. La gestion et le traitement centralisés des données climatologiques sont donc stoppés. Heureusement, un projet national est engagé pour récupérer les données du mini-ordinateur, stockées initialement sur bandes, et les copier sur CD-Rom.

Dans les années 1990, avec l'arrivée des micro-ordinateurs, le Département chargé de l'informatique et des bases de données reprend la saisie de données climatologiques mais sur un simple tableur. Cette solution n'est pas suffisante pour recueillir et partager correctement les données au sein des SMHN et entre les différents SMHN, ni de sécuriser les données. Le service météorologique national de Madagascar a donc mis sur pied un projet dans le cadre du Programme de coopération volontaire (PCV) de l'OMM, avec pour objectif de mettre en œuvre une base de données climatologiques nationale et d'installer un système moderne et standard de collecte, stockage, traitement et diffusion des données pour un système de gestion des ressources.

Les avantages attendus étaient les suivants :

- permettre une participation plus efficace au niveau mondial à la mise en œuvre des activités du Programme climatologique mondial, en particulier en ce qui concerne les études sur les changements climatiques régionaux, sous l'influence de facteurs naturels et humains ;
- améliorer la fourniture en temps réel d'informations sur le climat à différents secteurs de l'économie, qui permettront de protéger et de développer les ressources naturelles, humaines et financières.

La France décide alors de financer ce projet et Météo-France se charge de la gestion, en proposant les services de ses agents. Ce projet est mené en deux phases.

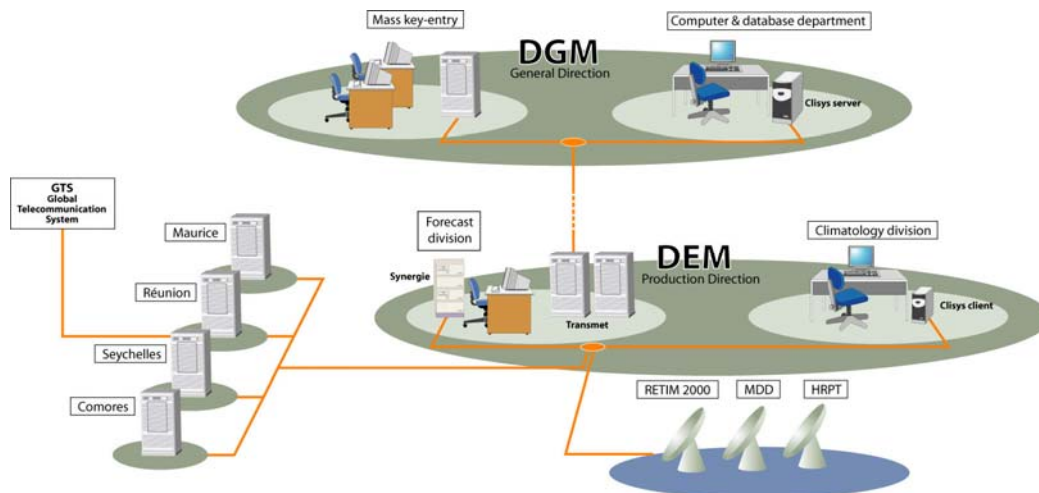
La première phase, qui comprend une étude de l'environnement présent, correspond à la définition du système, lequel comprend un serveur, un micro-ordinateur (répondant à des niveaux de sécurité élevés) assisté de deux stations clients, une imprimante, un scanner, des unités d'archive et de sauvegarde, les consommables correspondants et les manuels de formation. Le système est lié au SGBD Oracle et tourne sous un système d'exploitation de base Linux.

L'intégration du système, son installation et les essais opérationnels ont démarré en septembre 2003 et ont duré trois mois, dans les locaux de Météo-France à Toulouse, sous la surveillance de plusieurs administrateurs système, d'un administrateur Oracle et d'un climatologue. Le système a été accepté par la Direction Générale de la Météorologie (DGM) de Madagascar en mars 2004, ce qui a permis d'engager la mission d'installation et de formation en avril 2004, pendant une période de 15 jours, avec la participation de spécialistes de Météo-France. Cette mission consistait à installer et à démarrer le système dans les locaux du SMHN, à intégrer les données historiques de surface concernant Madagascar et à constituer et former les équipes d'administrateurs système, d'administrateurs de systèmes de gestion de bases de données climatologiques et d'opérateurs de saisie.

La deuxième phase du projet porte sur la consolidation des compétences acquises, sur la poursuite de la formation et le développement de nouvelles procédures.



Figure A2.1. Démonstration du système au personnel du service météorologique de Madagascar



[Traduction des légendes]

Saisie de masse

*DGM
Direction générale*

*Département Informatique et Base
de données*

Serveur Clisys

*SMT
Système mondial de télécommunications*

Division Prévision

Synergie

*DEM
Direction Production*

*Division Climatologie
Client Clisys*

Figure A2.2. Architecture système à Madagascar. Le système de gestion de bases de données climatologiques (CliSys) est relié aux différents services du SMHN. Les données provenant du SMT sont collectées via un SMAS (Transmet) qui fournit les données à l'outil présent sur la station de travail de prévision (Synergie).

Annexe 3 Exemples de matériel

Solutions matérielles basées sur des micro-ordinateurs pour gérer une base de données

Trois solutions matérielles sont indiquées ici pour la gestion d'une base de données climatologiques, allant d'un simple ordinateur de bureau à un solide serveur en passant par une station de travail. Le coût du système, ainsi que sa sécurité et ses performances, augmentent progressivement entre la première et la troisième solution. Cependant, dans ces trois solutions, on a particulièrement mis l'accent sur la sécurité en mélangeant les différentes configurations de sécurité matérielles et en ajoutant, dans chaque cas, des fonctions de sauvegarde et d'archivage.

Sécurité matérielle

Il existe trois façons d'assurer la sécurité matérielle.

La méthode de **clonage** consiste à avoir l'image exacte d'un disque dur, **à un moment donné**, sur un autre disque dur, appelé clone. Lorsque l'administrateur considère que le système fonctionne correctement, un clone du système présent sur le disque dur est créé. Ensuite, si des erreurs se produisent (ex : défaillance du disque dur, perte de fichiers, mauvaise configuration), l'administrateur peut revenir à la situation antérieure, correspondant à la période où le système fonctionnait correctement, grâce au clone. Ce type de sécurité est relativement intéressant pour des données qui ne changent pas trop, par exemple pour le système d'exploitation et les applications logicielles. Mais il n'est pas suffisant pour une base de données qui accumulent des données chaque jour. Dans ce cas, il convient d'ajouter une stratégie d'archivage et de sauvegarde. (Solution 1)

L'**écriture miroir** du disque est une méthode de stockage où les données d'un disque sont dupliquées **simultanément** sur un autre disque, si bien que les deux disques contiennent les mêmes informations. Cette technique, connue aussi sous le nom de RAID de niveau 1, apporte une redondance en cas de défaillance du disque et peut être complétée par un clonage. (Solution 2)

La **méthode de segmentation** protège les données de la corruption en permettant de copier des données sur une série de disques. En cas de défaillance de l'un des disques, toutes ses données sont automatiquement recréées sur les autres disques de la série. Cette méthode est appelée également technique RAID de niveau 5. (Solution 3)

Archivage

L'archivage pourrait être défini comme une méthode de stockage de données **pendant une période aussi longue que possible dans la mesure où la sécurité est assurée**. Cette méthode est spécialement conçue pour des données climatologiques qui, une fois soumises au contrôle de la qualité, sont très peu modifiées. Les supports de stockage doivent résister au temps, ce qui est le cas en particulier des CD, DVD et technologies NAS. Il est recommandé d'archiver les données climatologiques sous un format lisible par la plupart des logiciels, par exemple sous le format ASCII. La durée de vie d'un CD ou d'un DVD n'est pas illimitée. À la fin de chaque session, il est prudent de copier les informations d'un support sur un autre.

Sauvegarde de secours

La sauvegarde de secours peut être définie comme une méthode de stockage des informations **à un moment donné**, permettant de retrouver, en cas de perte, l'ensemble du système ou quelques informations à ce moment donné. Cette méthode est utile lorsque, par exemple, une sauvegarde de secours journalière est en place et que toutes ces sauvegardes journalières

sont conservées à une fréquence hebdomadaire. Dans ces conditions, il est possible d'extraire, en fin de semaine, un fichier qui a été perdu ou endommagé un certain jour de cette semaine.

Enfin, s'il s'avère qu'un seul serveur n'est pas suffisant pour traiter la base de données climatologiques d'un SMHN, il conviendra d'envisager une grappe de plusieurs serveurs et dispositifs de stockage interconnectés, qui formeront, aux yeux des utilisateurs, un système parfaitement accessible.

Tableau A3.2. Trois solutions matérielles pour la gestion des bases de données climatologiques

Fonction	Solution 1 : ordinateur de bureau	Solution 2 : station de travail	Solution 3 : serveur	Remarques
Système d'exploitation	Linux, Windows	Linux, Windows	Linux, Windows	Vérifier la compatibilité entre le système d'exploitation et l'ensemble des dispositifs et logiciels utilisés.
Processeur	1 processeur	1 à 2 processeurs	1 à 128 processeurs	Voir les exigences des différents logiciels utilisés, le nombre de clients et les contraintes du SMN.
Mémoire	Entre 250 Mo et 4 Go	Entre 512 Mo et 8 Go	Entre 512 Mo et 1000 Go	Voir les exigences des différents logiciels utilisés, le nombre de clients et les contraintes du SMN.
Disque dur	Nombre limité de disques durs. Exemple : 2 disques durs, disque 1 : pour l'ensemble du système, disque 2 : clone du disque 1.	Entre 1 et 4 disques. Exemple : 3 disques durs ; disque 1 : pour l'ensemble du système, disque 2 : clone du disque 1, disque 3 : miroir du disque 1 (RAID 1).	Entre 1 et 192 disques. Exemple : 6 disques durs ; disque 1 : pour l'ensemble du système, disque 2 : clone du disque 1, disques 3, 4 et 5 : base de données sur RAID 5, disque 6 : disque de rechange	Analyser les exigences en termes de volume. Choisir le type de disque dur approprié : IDE, SATA, SCSI, et l'architecture appropriée : clone, miroir, segmentation. Prévoir un disque dur de rechange.
Archive	Graveur de CD	Graveur de DVD	NAS	Analyser la technologie la plus appropriée.
Sauvegarde de secours	Graveur de CD	Bande, NAS	Bande, NAS	Analyser la technologie la plus appropriée.
Sécurité électrique	UPS	UPS	UPS	Analyser la technologie la plus appropriée en fonction des contraintes.
Sécurité logicielle				Pare-feu, anti-virus, anti-logiciels espions et toutes les mises à jour. Logiciel de sauvegarde de secours et d'archivage.
Garantie				Demander une garantie sur une période raisonnable.

**REPORTS PUBLISHED IN THE WORLD
CLIMATE DATA PROGRAMME (WCDP)/WORLD
CLIMATE DATA AND MONITORING PROGRAMME (WCDMP) SERIES**

- WCDP-1 WMO REGION III/IV TRAINING SEMINAR ON CLIMATE DATA MANAGEMENT AND USER SERVICES, Barbados, 22-26 September 1986 and Panama, 29 September 3 October 1986 (available in English and Spanish) - (WMO-TD No. 227)
- WCDP-2 REPORT OF THE INTERNATIONAL PLANNING MEETING ON CLIMATE SYSTEM MONITORING, Washington DC, USA, 14-18 December 1987 - (WMO-TD No. 246)
- WCDP-3 GUIDELINES ON THE QUALITY CONTROL OF DATA FROM THE WORLD RADIOMETRIC NETWORK, Leningrad 1987 (prepared by the World Radiation Data Centre, Voeikov Main Geophysical Observatory) - (WMO-TD No. 258)
- WCDP-4 INPUT FORMAT GUIDELINES FOR WORLD RADIOMETRIC NETWORK DATA, Leningrad 1987 (prepared by the World Radiation Data Centre, Voeikov Main Geophysical Observatory) - (WMO-TD No. 253. p. 35)
- WCDP-5 INFOCLIMA CATALOGUE OF CLIMATE SYSTEM DATA SETS, 1989 edition (WMO-TD No. 293)
- WCDP-6 CLICOM PROJECT (Climate Data Management System), April 1989 (updated issue of WCP-I 1 9) - (WMO-TD No. 299)
- WCDP-7 STATISTICS ON REGIONAL NETWORKS OF CLIMATOLOGICAL STATIONS (based on the INFOCLIMA World Inventory). VOLUME II: WMO REGION I - AFRICA (WMO-TD No. 305)
- WCDP-8 INFOCLIMA CATALOGUE OF CLIMATE SYSTEM DATA SETS - HYDROLOGICAL DATA EXTRACT, April 1989 - (WMO-TD No. 343)
- WCDP-9 REPORT OF MEETING OF CLICOM EXPERTS, Paris, 11-15 September 1989 (available in English and French) - (WMO-TD No. 342)
- WCDP-10 CALCULATION OF MONTHLY AND ANNUAL 30-YEAR STANDARD NORMALS, March 1989 (prepared by a meeting of experts, Washington DC, USA) - (WMO-TD No. 341)
- WCDP-11 REPORT OF THE EXPERT GROUP ON GLOBAL BASELINE DATASETS, Asheville, USA, 22-26 January 1990 - (WMO-TD No. 359)
- WCDP-12 REPORT OF THE MEETING ON HISTORICAL ARCHIVAL SURVEY FOR CLIMATE HISTORY, Paris, 21-22 February 1990 - (WMO-TD No. 372)
- WCDP-13 REPORT OF THE MEETING OF EXPERTS ON CLIMATE CHANGE DETECTION PROJECT, Niagara-on-the-Lake, Canada, 26-30 November 1990 - (WMO-TD No. 418)

Note: *Following the change of the name of the World Climate Data Programme (WCDP) to World Climate Data and Monitoring Programme (WCDMP) by*

the Eleventh WMO Congress (May 1991), the subsequent reports in this series will be published as WCDMP reports, the numbering being continued from No. 13 (the last 'WCDP" report).

- WCDMP-14 REPORT OF THE CCI WORKING GROUP ON CLIMATE CHANGE DETECTION, Geneva, 21-25 October 1991
- WCDMP-15 REPORT OF THE CCI EXPERTS MEETING ON CLIMATE CODE ADAPTATION, Geneva, 5-6 November 1991 - (WMO-TD No. 468)
- WCDMP-16 REPORT OF THE CCI EXPERTS MEETING ON TRACKING AND TRANSMISSION OF CLIMATE SYSTEM MONITORING INFORMATION, Geneva, 7-8 November 1991 - (WMO-TD No. 465)
- WCDMP-17 REPORT OF THE FIRST SESSION OF THE ADVISORY COMMITTEE ON CLIMATE APPLICATIONS AND DATA (ACCAD), Geneva, 19-20 November 1991 (also appears as WCASP-18) - (WMO-TD No. 475)
- WCDMP-18 CCI WORKING GROUP ON CLIMATE DATA, Geneva, 11-15 November 1991 (WMO-TD No. 488)
- WCDMP-19 REPORT OF THE SECOND CLICOM EXPERTS MEETING, Washington DC, 18-22 May 1992 - (WMO-TD No. 511)
- WCDMP-20 REPORT ON THE INFORMAL PLANNING MEETING ON STATISTICAL PROCEDURES FOR CLIMATE CHANGE DETECTION, Toronto, 25 June, 1992 (WMO-TD No. 498)
- WCDMP-21 FINAL REPORT OF THE CCI WORKING GROUP ON CLIMATE DATA AND ITS RAPPORTEURS, November 1992 - (WMO-TD No. 523)
- WCDMP-22 REPORT OF THE SECOND SESSION OF THE ADVISORY COMMITTEE ON CLIMATE APPLICATIONS AND DATA (ACCAD), Geneva, 16-17 November 1992 (also appears as WCASP-22) - (WMO-TD No. 529)
- WCDMP-23 REPORT OF THE EXPERTS MEETING ON REFERENCE CLIMATOLOGICAL STATIONS (RCS) AND NATIONAL CLIMATE DATA CATALOGUES (NCC), Offenbach am Main, Germany, 25-27 August 1992 - (WMO-TD No. 535)
- WCDMP-24 REPORT OF THE TENTH SESSION OF THE ADVISORY WORKING GROUP OF THE COMMISSION FOR CLIMATOLOGY, Geneva, 20-22 September 1995 (also appears as WCASP-34) - (WMO-TD No. 711)
- WCDMP-25 REPORT OF THE FIFTH SESSION OF THE ADVISORY COMMITTEE ON CLIMATE APPLICATIONS AND DATA (ACCAD), Geneva, 26 September 1995 (also appears as WCASP-35) - (WMO-TD No. 712)
- WCDMP-26 REPORT ON THE STATUS OF THE ARCHIVAL CLIMATE HISTORY SURVEY (ARCHISS) PROJECT, October 1996 (prepared by Mr M. Baker) - (WMO-TD No. 776)
- WCDMP-27 SUMMARY REPORT OF THE MEETING OF THE THIRD SESSION OF THE CCI WORKING GROUP ON CLIMATE CHANGE DETECTION, Geneva, 26 February - 1 March 1996 - (WMO-TD No. 818)

- WCDMP-28 SUMMARY NOTES AND RECOMMENDATIONS FOR CCI-XII FROM MEETINGS CONVENED TO PREPARE FOR PUBLISHING THE FIFTH AND SIXTH GLOBAL CLIMATE SYSTEM REVIEWS AND FOR A PUBLICATION ON THE CLIMATE OF THE 20TH CENTURY, July 1997 - (WMO-TD No. 830)
- WCDMP-29 CLIMATE CHANGE DETECTION REPORT - REPORTS FOR CCI-XII FROM RAPORTEURS THAT RELATE TO CLIMATE CHANGE DETECTION, July 1997 (WMO-TD No. 831)
- WCDMP-30 SUMMARY NOTES AND RECOMMENDATIONS ASSEMBLED FOR CCI-XII FROM RECENT ACTIVITIES CONCERNING CLIMATE DATA MANAGEMENT, July 1997 (WMO-TD No. 832)
- WCDMP-31 REPORTS FOR CCI-XII FROM RAPORTEURS THAT RELATE TO CLIMATE DATA MANAGEMENT, July 1997 - (WMO-TD No. 833)
- WCDMP-32 PROGRESS REPORTS TO CCI ON STATISTICAL METHODS, July 1997 (prepared by Mr Christian-Dietrich Schönwiese) (WMO-TD No 834)
- WCDMP-33 MEETING OF THE CCI WORKING GROUP ON CLIMATE DATA, Geneva, 30 January - 3 February 1995 - (WMO-TD No. 841)
- WCDMP-34 EXPERT MEETING TO REVIEW AND ASSESS THE ORACLE-BASED PROTOTYPE FOR FUTURE CLIMATE DATABASE MANAGEMENT SYSTEM (CDBMS), Toulouse, France, 12-16 May 1997 - (WMO-TD No. 902)
- WCDMP-35 REPORT OF THE ELEVENTH SESSION OF THE ADVISORY WORKING GROUP OF THE COMMISSION FOR CLIMATOLOGY, Mauritius, 9-14 February 1998 (also appears as WCASP-47) - (WMO-TD No. 895)
- WCDMP-36 REPORT OF THE MEETING OF THE CCI TASK TEAM ON CLIMATE ASPECTS OF RESOLUTION 40, Geneva, Switzerland, 10-11 June 1998 - (WMO-TD No. 925)
- WCDMP-37 REPORT OF THE MEETING OF THE JOINT CCI/CLIVAR TASK GROUP ON CLIMATE INDICES, Bracknell, UK, 2-4 September 1998 - (WMO-TD No. 930)
- WCDMP-38 REPORT OF THE MEETING OF THE WMO COMMISSION FOR CLIMATOLOGY (CCI) TASK GROUP ON A FUTURE WMO CLIMATE DATABASE MANAGEMENT SYSTEM (CDMS), Ostrava, Czech Republic, 10-13 November 1998 and FOLLOW-UP WORKSHOP TO THE WMO CCI TASK GROUP MEETING ON A FUTURE WMO CDMS, Toulouse, France, 30 March-1 April 1999 - (WMO-TD No. 932)
- WCDMP-39 REPORT OF THE MEETING OF THE CCI WORKING GROUP ON CLIMATE DATA, Geneva, Switzerland, 30 November-4 December 1998 - (WMO-TD No. 970)
- WCDMP-40 REPORT OF THE MEETING ON CLIMATE STATISTICS, PRODUCT DEVELOPMENT AND DATA EXCHANGE FOCUSING ON CLICOM 3.1, Geneva, 25-29 January 1999 - (WMO-TD No. 971)
- WCDMP-41 PROCEEDINGS OF THE SECOND SEMINAR FOR HOMOGENIZATION OF SURFACE CLIMATOLOGICAL DATA, Budapest, Hungary, 9-13 November 1998 (WMO-TD No. 962)

- WCDMP-42 REPORT OF THE MEETING OF EXPERTS ON THE CLIMATE OF THE 20TH CENTURY, Geneva, 26-30 April 1999 - (WMO-TD No. 972)
- WCDMP-43 REPORT OF THE TRAINING SEMINAR ON CLIMATE DATA MANAGEMENT FOCUSING ON CLICOM/CLIPS DEVELOPMENT AND EVALUATION, Niamey, Niger, 03 May-10 July 1999, (WMO-TD No. 973)
- WCDMP-44 REPRESENTATIVENESS, DATA GAPS AND UNCERTAINTIES IN CLIMATE OBSERVATIONS, Invited Scientific Lecture given by Chris Folland to the WMO Thirteenth Congress, Geneva, 21 May 1999 - (WMO-TD No. 977)
- WCDMP-45 WORLD CLIMATE PROGRAMME - WATER, DETECTING TREND AND OTHER CHANGES IN HYDROLOGICAL DATA, Zbigniew W. Kundzewicz and Alice Robson (Editors) - (WMO-TD No. 1013)
- WCDMP-46 MEETING OF THE WMO CCI TASK GROUP ON FUTURE WMO CLIMATE DATABASE MANAGEMENT SYSTEMS (CDMSs), Geneva, 3-5 May 2000 (WMO-TD No. 1025)
- WCDMP-47 REPORT ON THE ACTIVITIES OF THE WORKING GROUP ON CLIMATE CHANGE DETECTION AND RELATED RAPORTEURS, 1998-2001 (May 2001, updated from March 2001) (WMO-TD No. 1071)
- WCDMP-48 REPORT OF THE FIRST SESSION OF THE MANAGEMENT GROUP OF THE COMMISSION FOR CLIMATOLOGY (Berlin, Germany, 5-8 March 2002) (also appears as WCASP-55) (WMO-TD No. 1110)
- WCDMP-49 1. REPORT ON THE CLICOM-DARE WORKSHOP (San José, Costa Rica, 17-28 July 2000); 2. REPORT OF THE INTERNATIONAL DATA RESCUE MEETING (Geneva, 11-13 September 2001) (WMO-TD No. 1128)
- WCDMP-50 REPORT OF THE CLIMATE DATABASE MANAGEMENT SYSTEMS EVALUATION WORKSHOP (Geneva, 11-13 September 2001) (WMO-TD No. 1130)
- WCDMP-51 SUMMARY REPORT OF THE EXPERT MEETING FOR THE PREPARATION OF THE SEVENTH GLOBAL CLIMATE SYSTEM REVIEW (7GCSR) (Geneva, 16-19 September 2002) (WMO-TD No. 1131)
- WCDMP-52 GUIDELINES ON CLIMATE OBSERVATION NETWORKS AND SYSTEMS (WMO-TD No. 1185)
- WCDMP-53 GUIDELINES ON CLIMATE METADATA AND HOMOGENIZATION (WMO-TD No. 1186)
- WCDMP-54 REPORT OF THE CCI/CLIVAR EXPERT TEAM ON CLIMATE CHANGE DETECTION, MONITORING AND INDICES (ETCCDMI) (Norwich, UK, 24-26 November 2003) (WMO-TD No. 1205)
- WCDMP-55 GUIDELINES ON CLIMATE DATA RESCUE (WMO-TD No. 1210)
- WCDMP-56 FOURTH SEMINAR FOR HOMOGENIZATION AND QUALITY CONTROL IN CLIMATOLOGICAL DATABASES (Budapest, Hungary, 6-10 October 2003) (WMO-TD No. 1236)

- WCDMP-57 REPORT OF THE RA V DATA MANAGEMENT WORKSHOP (Melbourne, Australia, 28 November-3 December 2004) (WMO-TD No. 1263)
- WCDMP-58 GUIDELINES ON CLIMATE WATCHES (WMO-TD No. 1269)
- WCDMP-59 REPORT OF THE MEETING OF THE RA I WORKING GROUP ON CLIMATE MATTERS (Dakar, Senegal, 22 – 24 February 2006) (WMO-TD No. 1351)
- WCDMP-60 GUIDELINES ON CLIMATE DATA MANAGEMENT (WMO-TD No. 1376)