

Directives sur les bonnes pratiques en matière de sauvetage des données climatologiques

Édition 2016

TEMPS CLIMAT EAU



ORGANISATION
MÉTÉOROLOGIQUE
MONDIALE

OMM-N° 1182

Directives sur les bonnes pratiques en matière de sauvetage des données climatologiques

Édition 2016



ORGANISATION
MÉTÉOROLOGIQUE
MONDIALE

OMM-N° 1182

NOTE DE L'ÉDITEUR

La base de données terminologique de l'OMM, METEOTERM, peut être consultée à l'adresse <http://public.wmo.int/fr/ressources/meteoterm>.

Il convient d'informer le lecteur que lorsqu'il copie un hyperlien en le sélectionnant dans le texte, des espaces vont apparaître après <http://>, <https://>, <ftp://>, <mailto:>, et après les barres obliques (/), les tirets (-) et les séquences de caractères (lettres et chiffres). Il faut supprimer ces espaces de l'URL ainsi recopiée. L'URL correcte apparaît lorsque l'on place le curseur sur le lien. On peut aussi cliquer sur le lien et copier l'adresse qui s'affiche dans le ruban du navigateur.

On trouvera des informations complémentaires sur les bonnes pratiques en matière de sauvetage des données, notamment des illustrations et des photos, des précisions sur les méthodes employées et l'équipement nécessaire, des cas d'école, des hyperliens, des adresses, etc. sur le portail international I-DARE sur le sauvetage des données (www.idare-portal.org). Toute demande de renseignement supplémentaire doit être adressée via I-DARE.

REMERCIEMENTS

Nous remercions Madame Nancy Westcott, qui a travaillé au Centre climatologique régional du Midwest, *Illinois State Water Survey*, Université de l'Illinois (États-Unis d'Amérique), pour sa contribution exceptionnelle à l'élaboration de la présente publication.

OMM-N° 1182

© **Organisation météorologique mondiale, 2016**

L'OMM se réserve le droit de publication en version imprimée ou électronique ou sous toute autre forme et dans n'importe quelle langue. De courts extraits des publications de l'OMM peuvent être reproduits sans autorisation, pour autant que la source complète soit clairement indiquée. La correspondance relative au contenu rédactionnel et les demandes de publication, reproduction ou traduction partielle ou totale de la présente publication doivent être adressées au:

Président du Comité des publications
Organisation météorologique mondiale (OMM)
7 bis, avenue de la Paix
Case postale 2300
CH-1211 Genève 2, Suisse

Tél.: +41 (0) 22 730 84 03
Fax: +41 (0) 22 730 81 17
Courriel: publications@wmo.int

ISBN 978-92-63-21182-8

NOTE

Les appellations employées dans les publications de l'OMM et la présentation des données qui y figurent n'impliquent, de la part de l'Organisation météorologique mondiale, aucune prise de position quant au statut juridique des pays, territoires, villes ou zones, ou de leurs autorités, ni quant au tracé de leurs frontières ou limites.

La mention de certaines sociétés ou de certains produits ne signifie pas que l'OMM les cautionne ou les recommande de préférence à d'autres sociétés ou produits de nature similaire dont il n'est pas fait mention ou qui ne font l'objet d'aucune publicité.

TABLE DES MATIÈRES

	<i>Page</i>
CHAMP D'APPLICATION	v
1. INTRODUCTION – IMPORTANCE SCIENTIFIQUE DU SAUVETAGE DES DONNÉES CLIMATOLOGIQUES	1
2. PRÉSENTATION GÉNÉRALE DU SAUVETAGE DES DONNÉES CLIMATOLOGIQUES: ARCHIVAGE, NUMÉRISATION EN MODE IMAGE ET NUMÉRISATION EN MODE TEXTE.....	2
3. ARCHIVAGE DES SUPPORTS PAPIER	2
3.1 Recherche et localisation.	2
3.2 Conservation et stockage	4
3.3 Création d'inventaires électroniques des données conservées sur papier/ microfilm et des images	5
4. NUMÉRISATION EN MODE IMAGE DES SUPPORTS ORIGINAUX	7
4.1 Grandes étapes du processus de numérisation en mode image	7
4.2 Validation et stockage des fichiers images	8
5. NUMÉRISATION DES DONNÉES.....	9
5.1 Limiter les erreurs de saisie	9
5.2 Comparaison avec les bases de données internationales	10
6. ARCHIVAGE DES SUPPORTS NUMÉRIQUES(IMAGES ET DONNÉES).....	11
6.1 Recoupement des supports papier, des images et des jeux de données numériques	11
6.2 Rafraîchissement des supports.....	11
6.3 Recommandations relatives à l'archivage des supports numériques	11
APPENDICE 1. UNE ACTIVITÉ PLANÉTAIRE	13
APPENDICE 2. ÉVALUATION DU SAUVETAGE DES DONNÉES	15
APPENDICE 3. INFRASTRUCTURE, ÉQUIPEMENT, FOURNITURES ET PERSONNEL	16
APPENDICE 4. CONSEILS RELATIFS À LA CONSERVATION DES ARCHIVES PAPIER.....	18
APPENDICE 5. PARAMÈTRES À PRENDRE EN CONSIDÉRATION POUR DÉFINIR LA HIÉRARCHIE DES DONNÉES À ENREGISTRER SOUS FORME D'IMAGES ÉLECTRONIQUES ET À NUMÉRISER.....	19
APPENDICE 6. TECHNIQUES DE L'IMAGERIE ÉLECTRONIQUE.....	21
APPENDICE 7. AUTRES MÉTHODES DE NUMÉRISATION.....	24
APPENDICE 8. ERREURS DE NUMÉRISATION ET ASSURANCE QUALITÉ.....	26
APPENDICE 9. ENSEIGNEMENTS	28
APPENDICE 10. APPLICATIONS PERMETTANT DE DÉMONTRER LA VALEUR DES DONNÉES	30
APPENDICE 11. LES DIFFÉRENTES ÉTAPES DU SAUVETAGE DES DONNÉES.....	32
APPENDICE 12. GLOSSAIRE.....	34
APPENDICE 13. BIBLIOGRAPHIE.....	36

CHAMP D'APPLICATION

Le présent document technique résulte de la mise à jour de la publication intitulée *Guidelines on Climate Data Rescue* (2004) (WMO/TD-1210, WCDMP-55). Il prend en compte l'évolution de la technologie depuis douze ans et tire les enseignements des opérations de sauvetage des données les plus récentes menées dans le monde. On y trouvera une présentation générale du concept de sauvetage des données, suivie de chapitres consacrés à l'importance de cette activité, à l'archivage des supports d'origine, à leur numérisation en mode image, à la numérisation des données et à l'archivage des images et des données numériques. Des informations complémentaires sont fournies dans douze appendices.

L'objectif des directives sur le sauvetage des données climatologiques est de recommander un ensemble de bonnes pratiques. Compte tenu de la diversité des Services météorologiques et hydrologiques nationaux (SMHN) en termes de taille et de niveau de développement technologique, ainsi que de la multiplicité des conditions météorologiques et climatiques des Membres de l'OMM, ces bonnes pratiques ne seront pas pertinentes pour tous. Le présent document propose toutefois une large gamme de conseils qui devraient contribuer à l'organisation et à la mise en œuvre des opérations de sauvetage des données, et présente des solutions technologiques générales applicables par tous les Membres. On trouvera des informations techniques plus spécifiques, ainsi que des illustrations et photographies instructives sur le [portail international I-DARE sur le sauvetage des données](#) géré par l'OMM avec le soutien de l'Institut météorologique royal des Pays-Bas et de l'Équipe d'experts pour le sauvetage des données de la Commission de climatologie de l'OMM.

Si les bonnes pratiques présentées ici concernent le sauvetage des données météorologiques et climatologiques, elles peuvent également s'appliquer à d'autres domaines scientifiques, qu'ils relèvent ou non du mandat de l'OMM. Il est à noter en particulier que le sauvetage des données hydrologiques, maritimes et environnementales connexes obéit globalement à des principes et à des pratiques semblables et entre donc dans le champ d'application des présentes directives. On veillera toutefois à déterminer et à prendre en compte la spécificité de ces données, en coopération étroite avec les milieux concernés, parmi lesquels la Commission d'hydrologie de l'OMM et la Commission technique mixte OMM/Commission océanographique intergouvernementale (UNESCO) d'océanographie et de météorologie maritime.

DIRECTIVES SUR LES BONNES PRATIQUES EN MATIÈRE DE SAUVETAGE DES DONNÉES CLIMATOLOGIQUES

1. INTRODUCTION – IMPORTANCE SCIENTIFIQUE DU SAUVETAGE DES DONNÉES CLIMATOLOGIQUES

Le sauvetage des données climatologiques consiste à structurer et à sauvegarder les relevés instrumentaux et autres données qui risquent d'être perdus. Cette démarche est cruciale si l'on veut que les générations futures de scientifiques et autres utilisateurs de données puissent avoir accès à toutes les informations nécessaires à l'évaluation de la variabilité du climat et du changement climatique et offrir un ensemble de services climatologiques. Les données en question permettent de faire la jonction entre les données paléoclimatiques et les observations actuelles. Si les jeux de données climatologiques disponibles dans le monde renferment le plus souvent des données numériques remontant jusqu'aux années 1940 et 1950, bien des observations datant du XIX^e et du début du XX^e siècle n'existent que sur support papier et doivent être sauvegardées et numérisées. De plus, bien qu'on dispose, pour certaines régions de la planète, de relevés remontant jusque dans les années 1800, c'est loin d'être le cas pour toutes. La validité des modèles climatiques et des observations sur les données paléoclimatiques (anneaux de croissance des arbres, carottes de glace ou pollen par exemple) pourrait être grandement renforcée par l'accès à de longues séries chronologiques de mesures instrumentales issues de toutes les régions du monde.

Au XXI^e siècle, il reste encore des régions où les données sont rares et où il est donc fondamental d'obtenir et de vérifier les relevés les plus récents. La fiabilité des observations météorologiques peut être grandement influencée par l'emplacement et l'exposition des stations et dépend aussi du contrôle de la qualité des observations. La comparaison des données d'une station avec celles de sites voisins et la possibilité de consulter les documents contenant les observations d'origine (et les informations relatives aux méthodes d'observation de l'époque) permettront aux climatologues de valider avec davantage de certitude les régimes météorologiques et climatiques passés et présents.

S'agissant des archives climatologiques, la possibilité de remonter plus loin dans le temps et de combler les lacunes relatives à certaines époques ou à certaines régions, est une avancée positive à plusieurs égards:

- a) Elle permet de renforcer la crédibilité des modèles numériques dans les domaines de l'agrométéorologie, de l'hydrologie, de la climatologie et de la propagation des maladies, et d'améliorer la qualité des projections climatologiques;
- b) L'association des données sauvegardées et des données disponibles permet de mieux situer les conditions météorologiques et climatiques actuelles dans une perspective historique;
- c) On dispose de données de référence permettant d'évaluer la sensibilité, par le passé, des systèmes naturels et anthropiques à la variabilité du climat et au changement climatique, et il est donc possible de procéder à une estimation plus précise de l'incidence de la variabilité et de l'évolution du climat à l'avenir.

Ces évaluations peuvent éclairer les décideurs lors de la mise en place de mesures visant à atténuer les pertes dues aux catastrophes d'origine météorologique et constituent une source d'information supplémentaire à l'appui du développement économique.

Pour que les prestataires de services climatologiques et les chercheurs puissent les consulter, les données doivent être disponibles sous forme numérique (images et valeurs). En outre, toutes les données actuelles doivent également être sauvegardées et mises à disposition en format

numérique, puisqu'elles constituent les données historiques de demain. Les documents d'origine sont à cet égard d'une importance fondamentale car ils permettent de confirmer les valeurs numériques.

2. **PRÉSENTATION GÉNÉRALE DU SAUVETAGE DES DONNÉES CLIMATOLOGIQUES: ARCHIVAGE, NUMÉRISATION EN MODE IMAGE ET NUMÉRISATION EN MODE TEXTE**

Le sauvetage des données climatologiques consiste à structurer et à sauvegarder les données climatologiques que l'on risque de perdre en raison de la détérioration et de la destruction des relevés, de la négligence, de l'obsolescence technique ou de la simple dispersion des données au fil du temps. Les données non numérisées sont menacées par la vulnérabilité des relevés papier d'origine. Le sauvetage des données recouvre les opérations suivantes: classement et numérisation en mode image des relevés sur papier, microfilm et microfiche; saisie des données numériques et textuelles et numérisation des données issues des diagrammes à déroulement continu dans un format utilisable; et archivage des données, des métadonnées et des résultats du contrôle de la qualité et des procédures correspondantes. On trouvera dans le tableau 1 un récapitulatif des différentes étapes d'une opération de sauvetage des données climatologiques.

La plupart des pays se sont dotés de solides programmes d'observation météorologique, les mesures étant enregistrées, soit manuellement soit automatiquement, puis transcrites sur des supports papier ou copiées sur microfilm, avant d'être saisies sur un support informatisé (dans le meilleur des cas un système de gestion des données climatologiques ou SGDC) pour qu'elles puissent être ensuite facilement consultées et analysées. Malheureusement, les données n'ont pas toujours été retranscrites à partir des relevés papier, par manque de moyens financiers et/ou humains, ou parce qu'elles étaient conservées ailleurs, parfois à l'étranger. Bon nombre de ces formulaires papier et copies sur microfilm ou microfiche risquent d'être perdus en raison de la détérioration rapide des supports ou parce qu'une partie des archives a été déplacée, sans qu'on garde la trace de ce déménagement.

On trouvera dans les présentes directives des conseils sur la marche suivre pour le classement des relevés climatologique, leur numérisation en mode image, leur numérisation en mode texte et leur conservation.

3. **ARCHIVAGE DES SUPPORTS PAPIER**

3.1 **Recherche et localisation**

La première étape d'une opération de sauvetage des données consiste à les localiser. Avant d'entamer les procédures de conservation, il est fondamental de retrouver les jeux de données anciens présentant un intérêt sur le plan climatologique. Ces données peuvent se trouver dans les archives nationales et les ministères, dans des stations d'observation ou dans des collections privées, etc., sans que les professionnels de la météorologie en aient toujours connaissance. Il se peut que certaines données soient conservées dans des archives situées à l'étranger, notamment sur le territoire de pays au passé colonial. En faisant l'économie de ce type de recherche, on risque de se priver de l'accès à des jeux de données pouvant présenter un grand intérêt, mais dont l'existence restera ignorée de tous. Outre les données déjà classées, numérisées et conservées comme il convient dans le SGDC du SMHN ou du service compétent, on pourra retrouver auprès des organismes et des milieux énumérés ci-après de précieuses données sur le climat qui auraient pu tomber dans l'oubli:

- a) SMHN et autres services ou organismes publics, qui disposent souvent de données;
- b) Universités, établissements d'enseignement supérieur, lycées et particuliers (employés de longue date d'organismes météorologiques ou affiliés);

Tableau 1. Les différentes étapes d'une opération de sauvetage des données climatologiques

<i>Étapes d'une opération de sauvetage des données</i>	<i>Activités</i>	<i>Mots clés</i>
Archivage des supports papier et microfilm/microfiche (section 3)	Recherche et localisation	SMHN, sites d'observation, universités, organismes chargés de l'aviation et des affaires maritimes, organisations agricoles, bibliothèques et bases de données internationales, archives nationales
	Conservation et stockage	Nettoyage des supports; dépôt dans des boîtes à archives en carton non acide étiquetées, à l'abri de la poussière, de l'humidité et des parasites
	Création d'un inventaire électronique des fonds d'archives papier/microfilm	Inventaire de l'ensemble des supports papier; estimation de la charge de travail représentée par la numérisation en mode image et la numérisation en mode texte
Supports image (section 4)	Création d'un inventaire de référence des images	
	Création et validation des images électroniques Actualisation de l'inventaire de référence des images Création d'inventaires des fichiers images dans les répertoires de chaque CD/DVD ou ordinateur	Mise à jour de l'inventaire de référence des images (incluant les métadonnées) après vérification de leur lisibilité Recoupement des inventaires des fichiers des CD/DVD avec l'inventaire de référence des images
Numérisation des données (section 5)	Création d'un inventaire des données numériques	
	Saisie, tracé des diagrammes	Saisie des données dans le SGDC
	Contrôle de la qualité	Mise à jour de l'inventaire des données numériques à mesure que les données sont numérisées et soumises à de divers tests de contrôle de la qualité
Archivages des supports numériques (section 6)	Recoupement entre les supports papier, les images et les données numériques	Comparaison des inventaires des images et des données numériques avec l'inventaire électronique des données sur papier/microfilm d'origine
	Sauvegarde des supports électroniques	Opération quotidienne
	Distribution de multiples copies des archives des images et des données numériques	Vers des sites différents
	Renouvellement des supports et migration technologique	Tous les 5 à 10 ans

- c) Centres internationaux de données;
- d) Bibliothèques ou historiens locaux pouvant disposer de données sur support papier, sur microfilm ou microfiche, ou même de copies numériques de ces données;

- e) Acteurs du monde agricole, tels que les grands groupes du secteur de l'agroalimentaire qui ont possédé des exploitations dans le monde entier et disposent de relevés météorologiques remontant à plusieurs centaines d'années;
- f) Les organismes militaires sont également susceptibles de posséder des archives accessibles, tout comme les ministères des transports qui détiennent les relevés des données météorologiques destinées à l'aviation et au transport maritime;
- g) Organisations religieuses, notamment celles qui sont actives dans les domaines de l'observation ou de la science, tels que les ordres enseignants et les membres individuels du clergé;
- h) Sociétés scientifiques, en particulier les organisations du passé (Sociétés d'histoire naturelle);
- i) Musées;
- j) Les livres de bord des navires, les archives de presse et les journaux intimes constituent également d'excellentes sources d'informations météorologiques et climatologiques.

Au long de cette recherche, il convient également de procéder à un recoupement avec les données numériques disponibles, afin de déterminer si les données retrouvées ont déjà été numérisées, de recenser les lacunes importantes et d'obtenir des indications sur le moment où ont démarré les programmes d'observation dans certaines régions spécifiques. On veillera notamment à consulter les bases de données internationales, telles que:

- a) Le Réseau mondial de données climatologiques anciennes;
- b) *L'International Surface Pressure Databank*;
- c) *L'Integrated Surface Database*.

En outre, les bibliothèques nationales renferment des jeux de données dignes d'intérêt (on trouvera les liens vers les sites de ces organismes sur le [portail I-DARE](#)).

3.2 Conservation et stockage

Les données sur support papier et microfilm/microfiche doivent être classées de manière logique puis conservées dans des boîtes à archives en carton non acide que l'on entreposera sur des étagères solides ou dans des armoires de classement:

- a) En fonction de l'état des diagrammes ou des formulaires et du lieu où ces documents étaient conservés précédemment, il pourra se révéler nécessaire d'épousseter ou de nettoyer à l'aide d'un aspirateur les supports papier avant leur stockage;
- b) On veillera au respect des règles d'hygiène et de sécurité lors de la manipulation de relevés qui peuvent être contaminés;
- c) Il est nécessaire de prévoir un équipement spécialisé et une formation adaptée pour les personnes chargées d'épousseter les documents fragiles;
- d) On veillera tout particulièrement à la sécurité des opérateurs si les relevés sont moisissus ou ont été traités au moyen d'inhibiteurs de moisissures ou de pesticide;
- e) La manipulation de documents fragiles impose le port de gants;
- f) Le port de vêtements de protection et de masques peut se révéler nécessaire en fonction du volume de poussière dégagée.

On prévoit généralement des locaux spéciaux pour le stockage des archives sur papier et microfilm:

- a) Dans l'idéal, il faut pouvoir réguler la température et l'humidité de la pièce;
- b) Les archives doivent être protégées des insectes, des rongeurs, de la moisissure, de l'incendie, des inondations, de la poussière, du vol et de tous les autres aléas;
- c) On évitera la présence de bois dans la pièce pour réduire au maximum les risques d'incendie et d'infestation par les insectes;
- d) Il peut être utile de faire appel à un archiviste professionnel pour encadrer les opérations et il importe de désigner un responsable chargé de l'inspection régulière des locaux.

Les relevés sont généralement conservés dans des boîtes à archives en carton non acide ou dans des armoires de classement, en les regroupant par catégories (type de formulaire ou de diagramme par exemple) et par station, année et mois:

- a) La manière la plus logique de classer les données sur support papier (par année ou par station) dépendra du nombre d'années de données et de stations dont on dispose;
- b) Il convient d'étiqueter les boîtes à archives et/ou les armoires de classement et d'indiquer, dans l'inventaire électronique des fonds papier/microfilm, le lieu de stockage (numéro de la boîte ou du classeur à tiroirs), ainsi que le nom de la station, l'année, le mois, le type de support et de formulaire, afin de pouvoir facilement retrouver les données pour éventuellement procéder à leur enregistrement sous forme d'images électroniques et à leur numérisation.

Pour un stockage à long terme, on veillera à tenir compte des éléments suivants:

- a) Une fois que les données importantes ont été enregistrées sous forme d'images numériques, il ne faut pas jeter les documents originaux. Si l'on ne dispose pas des moyens nécessaires à la poursuite de leur stockage, on s'adressera aux archives nationales, aux bibliothèques des ministères ou aux universités. Si ces organismes ne sont pas en mesure de conserver les archives, et sous réserve de l'aval du service des archives nationales, on prendra contact avec des bibliothèques ou des institutions internationales (voir l'appendice 4 pour plus de détails);
- b) Les microfilms ne constituent pas un support adapté à un archivage permanent et ils se détériorent s'ils ne sont pas conservés dans des locaux adaptés, climatisés et équipés de déshumidificateurs;
- c) Les images numériques ne constituent pas non plus des archives permanentes (voir la section 6.2).

On trouvera à l'appendice 3 des éléments d'information supplémentaires sur l'infrastructure, les fournitures et le personnel requis et à l'appendice 4, des conseils relatifs à la conservation des archives papier.

3.3 **Création d'inventaires électroniques des données conservées sur papier/microfilm et des images**

Le classement et le stockage des supports papier doit aller de pair avec la réalisation d'une série d'inventaires, afin de faciliter le transfert des copies papier en mode image, puis en mode texte numérique. Il faut en effet pouvoir accéder aux relevés papier, après leur localisation et leur classement, afin de procéder à leur enregistrement sous forme d'image. Une fois cette opération réalisée, on doit être en mesure de retrouver les images pour procéder à la numérisation des données qu'elles contiennent. Par la suite il faut pouvoir facilement retrouver aussi bien les données originales sur support papier que les images électroniques, aux fins de contrôle

Tableau 2. Étapes de la création et de la maintenance des inventaires électroniques

Rassembler la totalité des relevés et des métadonnées aux fins de stockage	
Classer les relevés conformément à un plan d'archivage logique	Par station/type de formulaire/année Par type de formulaire/station/année Par année/station Pour les données figurant sur des diagrammes, par élément météorologique (vent, précipitations, etc.) également
Création d'un inventaire électronique des données sur papier/microfilm	Mentionner la station, l'année et l'emplacement de la boîte/du classeur à tiroirs, le type de support, le type de formulaire Inclure les types de variables Déterminer les périodes manquantes Estimer la quantité de types de support, de types de formulaires et le volume d'images afin d'établir un plan de travail
Création d'un inventaire de référence des images à partir de l'inventaire électronique des données	Indiquer la station, l'année, le mois, le type de support, le type ou le numéro de formulaire Définir les périodes des relevés Estimer le nombre de pages à numériser en mode image
Création, sur chaque CD/DVD ou répertoire d'ordinateur, d'inventaires des fichiers images au fur et à mesure de leur création, afin de pouvoir facilement les localiser	
Inclusion de pages contenant uniquement les métadonnées (informations sur les réseaux, les pratiques d'observation, etc.)	
Actualisation de l'inventaire de référence des images au fur et à mesure de leur création	Saisir le nombre de pages reproduites en mode image après validation de ces dernières Saisir l'emplacement des CD/DVD ou des ordinateurs
Recoupement entre l'inventaire de référence des images et les inventaires des fichiers images pour vérifier le nombre d'images enregistrées sur les CD ou les autres supports de stockage	

de la qualité des valeurs numérisées. Dans l'idéal, le protocole régissant la désignation des produits créés à chaque étape doit permettre de conserver autant d'informations que possible. Le tableau 2 résume les étapes nécessaires à la création d'une archive des données sur leur support d'origine (papier et microfilm/microfiche) et à leur enregistrement sous forme d'image numérique, au classement des images et à la localisation des fichiers images.

En résumé, dès que l'on retrouve des données climatologiques, il faut les incorporer dans un inventaire électronique des données sur support papier/microfilm, ce qui permet de savoir de quelles données on dispose, quelles mesures de sauvetage ont déjà été prises et le travail qui reste encore à accomplir. Pour chaque document, l'inventaire précise la station, l'année, le mois, le type de support, le numéro ou le type de formulaire, le type d'élément observé, le nombre de pages (y compris les pages de métadonnées) et en particulier l'emplacement de la boîte/du classeur à tiroirs et l'identifiant unique qui lui a été attribué. Il convient d'inclure également le type de variable mesurée.

Une fois créé cet inventaire électronique des relevés climatologiques anciens, il convient de le tenir à jour à mesure que l'on reçoit de nouvelles données. L'inventaire de référence des images doit également être complété au fur et à mesure de l'avancement de la numérisation des documents en mode image. L'inventaire de référence des images peut servir à estimer le progrès de ce travail de numérisation. On trouvera davantage de détails sur cette opération à la section 4.

4. NUMÉRISATION EN MODE IMAGE DES SUPPORTS ORIGINAUX

Une fois que l'on a recueilli et catalogué les données sur support papier et microfilm, on peut lancer le travail de planification pour déterminer dans quel ordre elles vont être enregistrées sous forme d'images puis numérisées. Ce sont les besoins spécifiques de chaque SMHN qui détermineront par où il faut commencer et la manière dont les opérations vont se dérouler. Il est probable que la hiérarchisation des priorités sera différente pour la création d'images et pour la numérisation des données. On trouvera à l'appendice 5 une analyse des éléments à prendre en compte pour déterminer ces priorités.

Les pages des relevés sur support papier et/ou microfilm sont photographiées ou scannées afin de créer des images ou des photographies numériques qui peuvent ensuite être préservées de toute nouvelle détérioration et rendues largement accessibles. On doit veiller à reproduire l'intégralité des pages des documents météorologiques originaux au moyen d'un appareil photo numérique ou d'un scanner optique. Il s'agit de garantir le sauvetage de toutes les métadonnées disponibles (pratiques d'observation, sites et description des données, par exemple). On trouvera à l'appendice 6 une comparaison des avantages respectifs du scanner et de l'appareil photo numérique, ainsi que les fonctionnalités souhaitables de ce dernier.

Dans tous les cas, les métadonnées (description des données et des pratiques d'observation, emplacement et caractéristiques du site) doivent être sauvegardées et numérisées en même temps que les images et les valeurs numériques.

4.1 Grandes étapes du processus de numérisation en mode image

Si les pratiques et les techniques employées pour enregistrer les relevés sous forme d'images numériques peuvent varier, il importe dans tous les cas de veiller au suivi des documents et des fichiers. Il est vivement recommandé d'adopter les procédures de base suivantes:

- a) Comme indiqué dans le tableau 2, créer un inventaire de référence des images à partir de l'inventaire électronique des données sur support papier ou microfilm. Inclure le nombre de pages à numériser en mode image. Actualiser l'inventaire de référence au fur et à mesure de la numérisation et de la validation des images, en indiquant le nombre de pages traitées et validées, ainsi que le répertoire de l'ordinateur ou l'emplacement du CD/DVD où elles se trouvent (fournir un inventaire type);
- b) Valider les fichiers images au fur et à mesure pour vérifier que l'image est lisible et qu'elle correspond au nom du fichier;
- c) Le nombre de fichiers images créés au moyen de l'appareil photo numérique ou du scanner sera très élevé. Dès que l'image est enregistrée sur l'ordinateur, le logiciel de l'appareil photo ou du scanner donne à ce fichier un nom qui n'est pas forcément en rapport avec son contenu. Il se peut en outre que le même nom soit donné à des images différentes si un très grand nombre d'images est enregistré sur un ordinateur le même jour. Il est donc recommandé d'avoir recours à un logiciel spécialisé pour générer des noms de fichier en rapport avec l'image qu'ils contiennent (voir par exemple ci-dessous, d)). Certains appareils photo ou scanners sont dotés de logiciels de ce type, mais on peut également les acheter séparément. Voir le [portail I-DARE](#) et l'appendice 6;
- d) Le nom du fichier image doit comporter le numéro d'identification de la station (SID), un acronyme désignant le type de formulaire (ACR), la date (YYYYmmDD) et le numéro de page (PPP). Il s'agit généralement de fichiers de type .png, .jpg ou .tif comme par exemple: SID_ACR_YYYYmmDD_PPP.png;
- e) Conserver les images dans un répertoire organisé selon les paramètres suivants: station/année/mois. Tenir une liste de référence des noms de fichiers inclus dans le répertoire principal, ainsi que des listes du même type dans les répertoires individuels ou sur les CD/DVD, afin de pouvoir localiser et récupérer les images facilement;

- f) Actualiser l'inventaire de référence des images en indiquant le nombre de pages créées et comparer le nombre de pages comptabilisées au départ avec celui qui a été effectivement produit et enregistré sur CD/DVD ou sur disque dur. Le nombre de pages décomptées et saisies dans les inventaires doit être le même que le nombre d'images produites. Saisir l'emplacement des images (CD/DVD ou nom de répertoire);
- g) Il est fondamental de garder une trace documentaire de la procédure suivie pour créer les images, et notamment des règles appliquées pour nommer les fichiers, une rotation du personnel chargé de ces tâches étant toujours possible.

4.2 Validation et stockage des fichiers images

Visualiser et vérifier les fichiers images avant de les enregistrer sur CD/DVD, disque dur ou dans un SGDC :

- a) Il se peut que les images soient floues à cause du positionnement de l'appareil photo;
- b) Il se peut que les images soient trop claires et que les corrections inscrites au stylo ou au crayon ne soient pas visibles;
- c) Il se peut que les images soient trop sombres à cause du réglage de l'appareil photo ou du scanner;
- d) Il se peut qu'une partie des données soit rendue illisible par une lumière trop vive. En effet, quand on utilise un éclairage direct, que ce soit le flash ou un éclairage placé sur le support de l'appareil, la lumière peut se refléter dans l'objectif et effacer certaines des données à l'image, surtout si le support est un papier brillant.

Une bonne méthode de contrôle de la qualité consiste à vérifier périodiquement (toutes les 300 pages par exemple) la lisibilité de plusieurs clichés. Il est également conseillé de comparer quelques images avec les archives papier correspondantes, afin de s'assurer qu'il ne manque aucune page. Il est important de confirmer la concordance du nom du fichier image avec son contenu au moins une fois par mois/année/station.

Si l'on utilise des CD ou des DVD, il faut veiller à les étiqueter. Par ailleurs:

- a) Quand on utilise les noms des fichiers images, il faut les saisir dans un inventaire des fichiers accompagnés des métadonnées permettant de les identifier (nom et numéro de station, année, mois, type de formulaire, nombre de pages) et de l'étiquette du CD ou DVD;
- b) Une copie des noms de fichiers associés à chaque CD ou DVD doit être incluse dans le CD ou DVD en question, ainsi que dans le répertoire de référence;
- c) Le libellé de l'étiquette du CD doit être saisi dans l'inventaire de référence des images.

Il est absolument indispensable de créer des copies des archives d'images sur CD/DVD, sur disque dur externe et/ou dans un SGDC. Quand on copie des CD/DVD sur disque dur ou dans un SGDC, il faut prévoir un répertoire pour chaque CD/DVD et copier également tous les inventaires des fichiers. En raison de leur plus grande capacité de stockage, il est recommandé d'utiliser des DVD plutôt que des CD.

Pour une bonne gestion des données, il est impératif d'entreposer les copies de sauvegarde des données et des métadonnées hors site, pour se prémunir contre toute perte en cas de dégâts occasionnés aux bâtiments et aux ordinateurs par des catastrophes naturelles ou d'origine anthropique. Ainsi:

- a) Il est important de distribuer des copies de la totalité des images électroniques archivées et de l'inventaire de référence des images;

- b) Si les données sont traitées dans un centre régional ou multinational, il faut en distribuer des copies au pays d'origine, tout en en conservant également plusieurs au centre. En amont du traitement des données dans un centre régional, un accord doit être conclu entre les pays, indiquant les mesures à prendre si les bâtiments ou les ordinateurs sont endommagés par une catastrophe naturelle ou d'origine anthropique. Le Secrétariat de l'OMM peut assister les pays dans la conclusion de tels accords.

5. NUMÉRISATION DES DONNÉES

On entend par numérisation le tracé des diagrammes à déroulement continu et la transcription des données textuelles dans un format numérique pouvant être stockés sous forme de nombres lisibles par ordinateurs. La première opération est traitée à l'appendice 7. On traitera ici de la numérisation des relevés alphanumériques par saisie. La numérisation des données est souvent réalisée au SMHN, mais il arrive aussi qu'elle soit effectuée par des entreprises privées dans des centres régionaux ou dans le cadre de projets de type «sciences citoyennes». La démarche est généralement pilotée par des spécialistes tels que des climatologues. Si les données imprimées sont retrouvées dans une archive nationale, c'est sans doute là que les opérations de saisie devront s'effectuer. Lors de la phase de planification du traitement en dehors des murs du SMHN, voire même à l'étranger, un accord devra être conclu au préalable entre les parties concernées. Le Secrétariat de l'OMM peut là aussi apporter son assistance aux pays.

La numérisation des données par saisie exige en général beaucoup plus de temps que leur enregistrement sous forme d'image électronique: s'il faut une heure pour enregistrer des données en mode image, il en faut une vingtaine pour les numériser par saisie. Faute de moyens, il se peut donc que l'on ne puisse pas numériser toutes les données. La détermination des données prioritaires doit se faire sur la base de paramètres scientifiques, techniques et socio-économiques. Lorsqu'il n'est pas possible de saisir numériquement la totalité des informations figurant sur un formulaire, il est particulièrement important de conserver les originaux et leurs copies. On trouvera à l'appendice 5 des éléments supplémentaires sur la hiérarchisation des priorités en matière de numérisation.

Si l'on ne dispose pas des ressources nécessaires à la saisie de grandes quantités de données, on peut, dans certains cas, opter pour la solution de la production participative: des bénévoles saisissent les données dans des formulaires électroniques ou réalisent des tracés à partir de séries de relevés chronologiques sur des sites Internet conçus spécialement à cet effet. Une autre solution consiste à recourir à un logiciel de reconnaissance optique de caractères (OCR) pour numériser les données imprimées. Cette méthode est en phase d'expérimentation et pourrait gagner en efficacité et se généraliser à l'avenir. Actuellement, l'OCR ne présente qu'un intérêt limité car elle nécessite l'emploi de formulaires spécialisés pour garantir la lisibilité. Pour plus d'informations sur la production participative, les diagrammes à déroulement continu et l'OCR, on se reportera à l'appendice 7 et au [portail I-DARE](#).

La première étape de la numérisation consiste à créer un inventaire des données numériques semblable à celui que l'on a établi lors de la numérisation des documents en mode image, afin de pouvoir suivre la progression du travail. Il faut classer les données par station, année, mois et catégorie.

5.1 Limiter les erreurs de saisie

Il existe plusieurs techniques permettant de réduire au minimum les erreurs de numérisation. La première est de procéder à la double ou à la triple saisie des données (c'est-à-dire faire saisir les mêmes données par au moins deux personnes). Ceci peut consister à demander à plusieurs opérateurs de saisir exactement les mêmes données puis à comparer les valeurs saisies. On peut également faire saisir les données quotidiennes par un ou plusieurs opérateurs, tandis qu'un ou plusieurs autres travaillent sur les données mensuelles, puis évaluer la cohérence du total

mensuel et des valeurs moyennes. La double saisie peut être réalisée en dehors d'un SGDC, les données étant importées une fois la comparaison effectuée. De cette manière, les données peuvent être saisies directement dans le SGDC.

Une seconde technique utilisée pour limiter les erreurs est de créer, pour la saisie, un formulaire qui reprend le modèle de l'original. De cette manière l'opérateur sait exactement où trouver les données sur la page, ce qui réduit le nombre d'erreurs. Les données ainsi saisies peuvent être directement importées dans une base de données. Même lorsque l'on ne saisit qu'une partie des données d'un document, il est recommandé de créer un nouveau formulaire type à chaque fois.

En troisième lieu, il est fondamental de transcrire les données exactement comme elles sont imprimées (de façon absolument littérale). On parle de «*key as you see*» en Australie ou de «saisir ce que l'on voit» en France. On évitera toute forme de codage ou de conversion d'unité «improvisée». Les ordinateurs sont très efficaces pour recoder les données, changer les unités, calculer les moyennes et même recoder à nouveau. En saisissant exactement ce qui est imprimé sur les formulaires, on peut être sûr de conserver une trace fidèle des formules utilisées pour recoder les données. Si l'on doit utiliser un codage, on veillera à le signaler clairement et à faire en sorte que le processus soit réversible (c'est-à-dire que l'on puisse retourner aux données d'origine). Les données doivent être saisies exactement comme elles apparaissent sur les formulaires d'observation originaux, même s'il est manifeste que l'observateur a commis une erreur. Ces documents sont parfois utilisés dans le cadre de procédures juridiques d'où l'importance de conserver les données originales. Il convient bien entendu de signaler les valeurs erronées et de les corriger.

La double ou la triple saisie, la création d'un formulaire type et la transcription littérale sont des méthodes permettant de réduire grandement le nombre d'erreurs. Une fois les données saisies ou importées dans le SGDC en tant que données provisoires, elles doivent être soumises à un contrôle qualité (voir l'appendice 8). On ne saurait trop insister sur l'importance de cette étape.

Enfin, il est recommandé de saisir l'intégralité des données figurant sur une page traitée: il est en effet peu probable que l'on ait une autre occasion d'y revenir et de saisir les données omises (latitude, longitude, nom de l'observateur, heure de l'observation, autres variables météorologiques et toutes les données numériques et textuelles par exemple). La numérisation de la totalité des variables permet aussi un meilleur contrôle de la qualité des données.

5.2 Comparaison avec les bases de données internationales

L'acquisition de données climatologiques numériques conservées dans des bases de données internationales alimentées par les messages SYNOP ou METAR peut être une option intéressante pour les SMHN. Il est souvent arrivé que ces données numériques ne soient pas conservées en interne au moment de leur production et il peut s'avérer difficile d'y accéder en raison de la bande passante requise pour télécharger des fichiers de données très volumineux. Ces données sont intéressantes car elles sont déjà numérisées et peuvent couvrir davantage de périodes et de variables que les bases de données climatologiques internes. Dans bien des cas toutefois, elles sont contaminées par des erreurs de saisie, de codage et de transmission. Il est cependant probable qu'elles aient été soumises à des procédures d'assurance qualité. Toutefois, si une valeur n'apparaît pas clairement sur les formulaires papier originaux, elle peut être récupérée à partir du message SYNOP. La numérisation des supports d'origine et la comparaison avec les données issues des bases de données internationales permet de disposer d'une version plus complète et précise des données climatologiques. La provenance des données est encore renforcée si l'on dispose de données issues des bases de données internationales, des données nouvellement numérisées et des images, dans la mesure du possible.

6. ARCHIVAGE DES SUPPORTS NUMÉRIQUES (IMAGES ET DONNÉES)

6.1 Recoupement des supports papier, des images et des jeux de données numériques

Il est utile de suivre systématiquement la progression du projet, en recoupant le nombre de mois d'images et de données numériques créées avec le total attendu sur la base du nombre de pages enregistrées sous forme d'images numériques ou devant l'être, ainsi qu'avec la période pendant laquelle les relevés ont été effectués pour chaque station. Au début du processus de sauvetage des données, on a créé un inventaire électronique des supports papier/microfilm à partir des relevés d'origine. En fin de projet, on compare la quantité de données numériques créées (calculée à partir des inventaires des fichiers images et des données numérisées) avec l'inventaire électronique initial, pour déterminer s'il manque des données.

6.2 Rafraîchissement des supports

Le sauvetage des données est tout aussi important pour les données numériques que pour les données papier, puisque les supports sur lesquels elles sont conservées ne sont pas permanents. En effet, les bandes magnétiques perdent peu à peu leur magnétisme, surtout dans un environnement chaud et humide. Il faut plusieurs siècles avant que le papier ne commence à s'effriter, mais les bandes magnétiques et les autres supports lisibles par ordinateur ne sont plus utilisables au bout de quelques dizaines d'années seulement. Les bandes magnétiques, les disques optiques et les cassettes 8 mm sont généralement considérés comme obsolètes. Si les données numériques sont encore stockées sur ces supports, il faut contacter les archives nationales ou les universités pour procéder à leur transfert vers des outils plus modernes, tels les DVD ou les disques durs, faute de quoi il risque de ne plus rester, en guise de données, qu'une suite incompréhensible de chiffres et de lettres.

Avec l'évolution de la technologie informatique, les ordinateurs, les systèmes d'exploitation, les langages informatiques et les logiciels employés pour lire les anciens supports deviennent également obsolètes. Or, ce problème ne va pas disparaître. À l'avenir, il est vivement recommandé de procéder à la migration des données climatologiques qui ont été enregistrées sous forme d'images, puis numérisées, vers la nouvelle génération de supports dont la fiabilité a déjà été établie (comme les CD et les disques durs). Les CD et DVD sont largement utilisés aujourd'hui, mais ce ne sera peut-être plus le cas dans un avenir proche. Il est également recommandé de conserver de nombreuses copies de sauvegarde, étant donné que les ordinateurs (ainsi que les CD et les disques durs) ne sont pas à l'abri de graves défaillances.

Il faut le souligner, il est essentiel de veiller à la migration, vers des ordinateurs et des systèmes d'exploitation modernes, du logiciel permettant d'enregistrer, de stocker, de lire et d'interpréter les données climatologiques. Le logiciel lui-même devra être mis à jour régulièrement. En procédant de la sorte, on sauve les données avant même qu'un problème ne se pose et qu'on ne les perde. La question de la migration des supports doit sans doute être posée tous les deux à cinq ans et celle de la modernisation des ordinateurs et des logiciels tous les cinq à sept ans.

Enfin, il est recommandé de faire le choix du format ASCII pour les fichiers de données texte, afin de faciliter la lecture des données numériques et le transfert entre ordinateurs. Si l'on utilise une base de données, il faut pouvoir exporter facilement ces dernières en format ASCII (fichier .csv par exemple). Dans la perspective d'une opération de migration, il est également important que le code numérique dans lequel sont stockés les fichiers soit simple (c'est-à-dire qu'aucune opération de décodage ne soit nécessaire) et consigné par écrit.

6.3 Recommandations relatives à l'archivage des supports numériques

- a) Conserver aussi bien les fichiers des images que des données numériques. Il se peut que les fichiers images soient considérés comme les archives officielles et il importe donc de les protéger;

- b) Classer les données pour en faciliter la consultation à l'avenir (sur CD ou après recherche dans l'architecture du répertoire: station, année, mois). Conserver un inventaire des fichiers sur chaque CD et dans le répertoire de référence;
 - c) Veiller à une sauvegarde quotidienne du SGDC;
 - d) Stocker les images et les données numériques sur des supports appropriés pour garantir leur disponibilité sur le très long terme, en procédant aux migrations nécessaires des jeux de données en fonction des besoins, afin de contrecarrer les risques d'obsolescence et de détérioration;
 - e) Stocker les supports numériques hors site pour préserver les images électroniques et les données numériques. Pour protéger les données en cas de catastrophe naturelle ou d'origine anthropique, entreposer des copies de l'intégralité desdites données en de multiples endroits, notamment dans un centre climatologique régional situé à l'extérieur de la région d'origine.
-

APPENDICE 1. UNE ACTIVITÉ PLANÉTAIRE

1. QUI DOIT PROCÉDER AU SAUVETAGE DES DONNÉES?

Toute personne ou groupe possédant des données climatologiques (sur support papier, microfilm ou support numérique) doit s'efforcer de contribuer à leur sauvetage. Les responsables de la gestion des archives météorologiques d'un pays ont un rôle particulier à jouer dans ce domaine, puisqu'ils sont mieux placés pour apprécier et juger de la valeur des données à sauvegarder et savoir celles qui sont les plus importantes. Ces professionnels travaillent souvent dans les services de climatologie au sein des SMHN et le sauvetage des données fait partie de leurs attributions en matière de conservation. On peut toutefois trouver des artisans du sauvetage des données dans bien d'autres institutions, tant publiques (services chargés de l'agriculture) que privées (universités, exploitations agricoles, entreprises du secteur agroalimentaire). De plus, des organisations de bénévoles, telles que l'Organisation internationale de sauvetage des données (IEDRO) et le projet de reconstitution de jeux de données sur la circulation atmosphérique (ACRE) ont été mises sur pied pour encourager le sauvetage des données. Parmi les autres parties intéressées, il faut citer les climatologues en activité ou à la retraite, les bibliothécaires, les historiens et les étudiants. Ces personnes peuvent se révéler très utiles lors de la mise en œuvre d'opérations de production participative.

2. OÙ DOIVENT SE DÉROULER LES OPÉRATIONS DE SAUVETAGE DES DONNÉES?

Le choix du site des opérations de sauvetage des données peut être envisagé sous plusieurs angles. On peut décider de choisir un lieu situé dans le pays où les données à sauver sont entreposées. Il peut s'agir du siège du SMHN et/ou de ses bureaux en province. La meilleure solution peut être de réunir les données en un lieu unique, ou, si les compétences requises y sont disponibles ou peuvent y être mobilisées, d'organiser les opérations de saisie en mode image et de conservation dans les bureaux de province. Il est également possible de collaborer avec des organismes régionaux ou multinationaux, tels que les centres climatologiques régionaux de l'OMM. Cette solution a pour avantage de mettre à la disposition des pays des compétences et des équipements qui ne sont peut-être pas à leur portée. On peut en outre envisager de faire appel, pour réaliser les opérations de numérisation en mode image ou de saisie des données, à un tiers situé en dehors du pays ou de la région, tous les documents étant ensuite restitués au pays d'origine. La meilleure solution dépend du contexte individuel. Avant tout transfert des données à un tiers, le Secrétariat de l'OMM se tient à la disposition des Membres pour les accompagner dans la planification du site des activités de sauvetage des données.

3. LE PORTAIL INTERNATIONAL POUR LE SAUVETAGE DES DONNÉES (I-DARE)

Le [portail I-DARE](#) est une ressource accessible en ligne destinée aux personnes intéressées par la conservation, le sauvetage et la numérisation des données. Le portail permet de centraliser, à l'échelle du globe, les informations concernant le statut des projets de sauvetage passés ou en cours, les données à sauvegarder et les méthodes et technologies en vigueur. Il s'agit d'une véritable plateforme de mutualisation de l'information sur tous les aspects du sauvetage des données, et notamment les techniques établies ou émergentes en la matière. Ces ressources peuvent également trouver une application dans d'autres disciplines scientifiques.

Ses objectifs étant de renforcer la visibilité des activités de sauvetage des données en cours, d'encourager de nouvelles et de mieux coordonner les opérations menées à l'échelon international, le [portail I-DARE](#) constitue un précieux outil de communication. Chaque projet qui y est recensé permet de faire connaître l'excellent travail accompli par chaque pays à l'ensemble des parties intéressées. L'accent mis sur les activités passées et présentes contribuera également à mettre

en évidence les lacunes et les nouvelles possibilités, permettra de donner la priorité aux régions où les besoins en matière de sauvetage des données se font le plus sentir et aidera les parties prenantes à mobiliser des fonds pour les projets.

APPENDICE 2. ÉVALUATION DU SAUVETAGE DES DONNÉES

Le tableau 3 donne un aperçu des mesures à prendre pour évaluer la situation en matière de sauvetage des données en vue de mettre en place un nouveau programme de sauvegarde.

Tableau 3. Liste des mesures à prendre

1. Prendre connaissance, sur le portail I-DARE , des ressources techniques et des bonnes pratiques en matière de sauvetage des données
2. Télécharger tous les inventaires connus de données à partir des bases de données internationales, des bibliothèques, des publications scientifiques, afin d'évaluer l'étendue des données météorologiques connues du pays
3. Déterminer l'état actuel du sauvetage des données dans le pays
4. Visiter et évaluer les salles d'archives existantes
5. S'entretenir avec le directeur et le personnel du SMHN afin d'établir pourquoi le sauvetage des données leur paraît important et quelles sont les priorités en la matière
6. Prodiguer, si nécessaire, des conseils relatifs à l'élaboration d'un inventaire des sources, des données et des métadonnées d'origine
7. Prodiguer des conseils relatifs au stockage du papier, aux besoins en locaux et en fournitures, et à l'hygiène et à la sécurité sur le lieu de travail
8. Prodiguer des conseils relatifs aux méthodes de création d'images numériques et à l'équipement nécessaire
9. Évaluer les compétences du SGDC et l'infrastructure entrant dans le cadre de la gestion des données (âge, système de sauvegarde, contenu)
10. Passer en revue le processus actuel de numérisation et de contrôle de la qualité afin de déterminer comment les données nouvellement numérisées peuvent s'y intégrer et s'il convient de modifier le processus
11. Prodiguer des conseils relatifs aux méthodes de numérisation et à l'équipement nécessaire
12. Prodiguer des conseils relatifs à l'inventaire des fichiers images et données numérisées
13. Prodiguer des conseils relatifs au recueil des métadonnées, à l'archivage et à la création d'images numériques (WMO/TD-1186)
14. Prodiguer des conseils relatifs à la saisie des données
15. Prodiguer des conseils relatifs à l'utilisation possible des données sauvegardées afin d'en montrer l'intérêt aux bailleurs de fonds/SMHN
16. Prodiguer des conseils relatifs au personnel requis pour les opérations de numérisation en mode image et en mode texte
17. Estimer le calendrier associé à chaque étape du sauvetage des données, en fonction de l'ampleur de la tâche et du nombre de personnes impliquées
18. Élaborer un plan de sauvetage des données sur la base des priorités du SMHN
19. Afficher des informations sur la mission de sauvetage des données sur le portail I-DARE

APPENDICE 3. INFRASTRUCTURE, ÉQUIPEMENT, FOURNITURES ET PERSONNEL

Dans tous les cas, le stockage, la numérisation en mode image et la numérisation des données doivent se dérouler dans un environnement sûr, bien éclairé, à l'abri de la poussière et dont l'atmosphère est régulée (ce qui correspond au minimum à la présence d'un climatiseur et d'un déshumidificateur). Les achats d'équipement et de fournitures dépendront des besoins de chaque pays et des méthodes choisies. Le tableau 4 peut servir de point de départ à la planification d'un projet de sauvetage des données climatologiques.

Pour garantir la réussite d'un projet de sauvetage des données climatologiques, il est essentiel de s'assurer les services d'une équipe compétente. Un climatologue semble la personne idéale pour procéder à la recherche des données climatologiques et prendre les décisions relatives à l'organisation des archives. Quant aux procédures de numérisation en mode image, il vaut mieux les confier à une personne intéressée par la conservation, éventuellement à un amateur de photographie. Le soin et la minutie sont les qualités que l'on recherchera chez les opérateurs chargés de l'enregistrement des relevés sous forme d'images et de la numérisation des données, la rapidité et l'exactitude comptant beaucoup dans ce domaine. On peut organiser un roulement entre les opérateurs chargés de la numérisation en mode image et de la saisie des données, afin d'éviter l'ennui occasionné par un travail répétitif et de ne pas être tributaire d'une seule personne. Les opérations d'enregistrement en mode image et de numérisation des données avancent plus rapidement lorsque les opérateurs possèdent une bonne connaissance de l'informatique et quand ils travaillent en tandem, l'un étant chargé de l'acquisition de l'image et l'autre de vérifier sa qualité, tandis que deux autres personnes saisissent ou tracent les données issues de la même page (double saisie). Il est vivement recommandé de prévoir des séances de travail d'une demi-journée ou de quatre heures, étant donné qu'elles requièrent une concentration totale.

Le nombre d'appareils photo utilisés et le choix du banc de reproduction dépendent de la manière dont les relevés se présentent, en volumes reliés ou sous forme de feuilles volantes. Quand on photographie une page à la fois, il est conseillé de prévoir deux appareils photo, l'un servant uniquement d'appareil de réserve. Le banc de reproduction permet de positionner aussi bien l'appareil photo que l'éclairage. En revanche, pour photographier les données figurant dans un volume relié, on a besoin de deux appareils sur le banc de reproduction, ce système permettant de traiter deux pages à la fois sans abimer la reliure. On trouvera à l'appendice 6 et sur le [portail I-DARE](#) des informations complémentaires sur les bancs de reproduction.

Tableau 4. Éléments à prendre en compte dans la planification des missions de sauvetage des données

	<i>Organisation</i>	<i>Numérisation en mode image</i>	<i>Numérisation en mode texte</i>
Infrastructure	Pièce bien éclairée, climatisée et sécurisée	Pièce équipée d'un nombre suffisant de prises électriques et d'un éclairage adéquat, ainsi que d'un nombre suffisant de tables pour trier les formulaires papier	Pièce équipée de suffisamment de prises électriques et d'un éclairage adéquat
Équipement	Boîtes à archives en carton non acide Étagères double largeur (2 m de large et 2,5 m de haut) Escabeau	Au moins deux appareils photo (voir fonctionnalités et accessoires à l'appendice 6) Banc de reproduction PC Lutrin Bureau Manuels techniques	SGDC

	<i>Organisation</i>	<i>Numérisation en mode image</i>	<i>Numérisation en mode texte</i>
Fournitures	Gants, blouses, masques, chiffons à poussière	CD, DVD et/ou deux des plus gros disques durs disponibles Câble USB ou WIFI	
Personnel chargé du sauvetage des données	Deux personnes chargées de nettoyer les documents, d'organiser et de concevoir les archives et d'établir des inventaires	Deux personnes chargées de la photographie et du contrôle de la qualité, du stockage et de la mise à jour des inventaires	Deux personnes chargées de numériser les données et de mettre à jour les inventaires
	Formation à la gestion de l'hygiène et de la sécurité au travail en cas de contamination des relevés		
Personnel chargé du suivi	Personne qualifiée chargée de la mise à jour des inventaires et du suivi des conditions ambiantes	Climatologue chargé de superviser l'introduction de données et métadonnées nouvelles	Climatologue chargé du contrôle de la qualité et de l'élaboration de produits

APPENDICE 4. CONSEILS RELATIFS À LA CONSERVATION DES ARCHIVES PAPIER

Les archives papier doivent être conservées dans une salle spéciale, qui ne se situera de préférence ni en sous-sol ni sous les toits. La pièce doit être propre et il convient d'y veiller à l'absence de poussière, de moisissures et de parasites. Il faut empêcher que les documents soient en contact avec le sol en les rangeant dans des boîtes à archives en carton non acide sur des étagères solides. La pièce doit être bien éclairée et les boîtes étiquetées et faciles d'accès. Une personne doit être expressément désignée pour vérifier régulièrement l'état de la salle.

On conservera les documents papier au moins jusqu'à ce que a) ils aient été enregistrés sous forme d'images; b) on ait procédé à la vérification de la qualité et de la complétude des images; et c) les images aient été copiées. Il est important de réaliser plusieurs copies électroniques des inventaires et de l'intégralité des archives d'images et de les conserver dans plusieurs endroits hors site.

Il est préférable de toujours conserver les copies papier:

- a) Les copies papier peuvent comporter des indications d'erreurs de mesure ou d'instrument (ratures au crayon ou au stylo) qui ne sont pas lisibles sur les copies électroniques;
- b) Il est important de pouvoir consulter les originaux agrémentés de corrections et de notes pour la confirmation des valeurs saisies;
- c) Une panne informatique peut entraîner la perte des copies électroniques;
- d) Jusqu'à présent, les copies papier ont une durée de vie supérieure à celle des supports numériques et électroniques.

Il est prudent de conserver les archives papier en tant que copie de sauvegarde.

Quand l'espace nécessaire existe, le mieux est de conserver les archives papier dans les locaux du SMHN. Si toutefois les ressources disponibles ne le permettent pas, il faut contacter les archives nationales ou universitaires et les bibliothèques. Il se peut que les services nationaux des archives définissent des règles contraignantes sur la durée de conservation des données (sur support papier et électronique). En cas de manque d'intérêt de ces organismes, on contactera des archives internationales.

APPENDICE 5. PARAMÈTRES À PRENDRE EN CONSIDÉRATION POUR DÉFINIR LA HIÉRARCHIE DES DONNÉES À ENREGISTRER SOUS FORME D'IMAGES ÉLECTRONIQUES ET À NUMÉRISER

Si l'idéal serait sans doute de procéder à l'enregistrement sous forme d'images numériques et à la saisie de toutes les données en même temps, il faut reconnaître que l'archivage est un processus de longue haleine et qu'il faut bien commencer quelque part. On pourra par exemple choisir de remédier d'abord aux lacunes constatées dans les séries chronologiques existantes. Les participants à la Réunion internationale sur le sauvetage des données (Genève, 11-13 septembre 2001) ont conclu qu'il convenait de sauver en priorité les données climatologiques anciennes et actuelles qui:

- a) Étaient de grande qualité;
- b) Revêtaient une importance nationale, régionale et mondiale et risquaient d'être perdues;
- c) Permettraient de combler des lacunes dans les jeux de données établis.

1. NUMÉRISATION EN MODE IMAGE

Quand on numérise les documents papier sous forme d'images, il semble logique de commencer par les documents les plus anciens, les plus rares ou les plus fragiles et de procéder station par station. Toutefois, la qualité du papier et de l'encre ayant baissé ces dernières années, il peut s'avérer nécessaire de procéder d'abord au traitement de documents plus récents. Quand on établit l'ordre de priorité des diagrammes à déroulement continu, on peut commencer par ceux qui revêtent le plus d'importance pour le SMHN (précipitations, pression, vent ou ensoleillement, éventuellement). Il faut enregistrer sous formes d'images la totalité des pages d'un ouvrage (registre mensuel par exemple) afin de veiller à sauver la totalité des métadonnées disponibles (dont la description du site et des données, les photographies, les informations sur les instruments, les étalonnages, les méthodes d'observation et les procédures de contrôle qualité).

2. NUMÉRISATION EN MODE TEXTE

La numérisation des données prend du temps, aussi est-il important de définir des priorités en la matière. On tiendra compte des facteurs énumérés ci-après pour la numérisation des observations en surface:

- a) Pour démontrer l'intérêt du sauvetage des données à des bailleurs de fonds potentiels, il peut être utile de commencer par une station affichant les plus longs relevés en continu ou de se concentrer sur une année ayant été marquée par des conditions climatiques particulières (sécheresse ou inondation par exemple);
- b) Il peut être prudent de commencer par des stations dont certaines données ont déjà été numérisées et qui sont toujours en service, ou qui affichent la plus longue période de relevés;
- c) À l'échelle de la planète on dispose de moins de données pour la période antérieure aux années 1950. Toutefois, dans certains pays, les données recueillies avant 1990 peuvent revêtir une importance cruciale;
- d) La localisation géographique est également un facteur à prendre en compte: il peut être utile de choisir des stations représentatives des différentes régions climatiques d'un pays ou situées sur des sites présentant un intérêt socio-économique particulier;

- e) La disponibilité des données peut être un autre facteur à prendre en considération. On peut choisir de commencer par une station affichant les données les plus fiables et le moins de données manquantes;
- f) Il est important de disposer des données de stations voisines, même pour une période de courte durée, afin de corroborer les mesures;
- g) La résolution temporelle (mensuelle, quotidienne, infraquotidienne) des données actuelles ou anciennes, ou la nécessité de disposer de paramètres météorologiques spécifiques peuvent contribuer à déterminer les données à numériser en priorité;
- h) Si l'on s'intéresse à l'étude des variations diurnes de la température, des précipitations ou du vent, il est crucial de pouvoir disposer de données issues de diagrammes à déroulement continu ou d'observations horaires (infraquotidiennes);
- i) Pour étudier les variables climatologiques essentielles, il convient de choisir des variables météorologiques dont on puisse calculer la valeur mensuelle, telles que la température et les précipitations (Bojinski *et al.*, 2014);
- j) Dans les zones présentant une discontinuité spatiale des données, on pourra choisir certaines stations spécifiques;
- k) La climatologie des orages, de la grêle, des tornades, des tempêtes de poussière, du brouillard et des tempêtes de verglas requiert des observations météorologiques visuelles réalisées sans instruments.

Ce sont généralement les observations en surface que l'on envisage d'enregistrer sous forme d'image électronique puis de numériser, mais les sondages en altitude, les cartes en surface et en altitude et les observations maritimes revêtent également une grande importance:

- a) Les sondages individuels fournissent des informations précieuses sur l'évolution de la température, de la pression et de l'humidité en fonction de l'altitude;
- b) Les données des sondages individuels n'étant plus toujours disponibles, les cartes en altitude peuvent être le seul moyen de les obtenir par déduction;
- c) Les cartes en surface peuvent également revêtir une grande importance car il arrive que les formulaires d'origine se soient égarés après le tracé des données d'observation.

Les directives et les paramètres proposés ici pour établir la hiérarchie des priorités en matière de sauvetage des données ont un caractère général. Chaque pays ou centre de données doit définir lui-même ses priorités en tenant compte de ses besoins fondamentaux et de l'état des données climatologiques dont il dispose. Une fois qu'on a fait le choix d'une station et d'une année, il est judicieux de tester les procédures de numérisation en mode image et en mode texte. On pourra ainsi définir les besoins en personnel et l'organisation rationnelle du processus.

On terminera par un rappel au sujet de la numérisation des formulaires: il est vivement conseillé de saisir la totalité des données figurant sur une page donnée, y compris les métadonnées (car l'occasion ne se présentera sans doute pas d'y revenir pour saisir les données éventuellement omises). Le mieux pour ce faire est de créer des modèles pour la numérisation des différents types de formulaires. Dans tous les cas, il convient de sauver et de numériser les métadonnées, la description des données et les pratiques d'observation (nom de l'observateur, heure de l'observation), l'emplacement (latitude, longitude) et les caractéristiques du site, en même temps que les valeurs numériques ou le fichier image.

APPENDICE 6. TECHNIQUES DE L'IMAGERIE ÉLECTRONIQUE

La création d'images électroniques est vivement conseillée car elle réduit la nécessité de manipuler les supports papier ou les microfilms. Cela permet également de redimensionner et d'accentuer l'image et d'améliorer la lisibilité des mentions écrites au moment de la saisie des données. Le format JPEG est généralement adapté. Certains spécialistes estiment qu'une taille de 1 600 x 1 200 pixels est suffisante.

1. SCANNERS

Le scanner est souvent une solution envisagée pour numériser en mode image un document papier. Il existe deux grands types de scanners: les scanners à plat (couramment employés pour les livres) et les scanners à chargeur de documents (souvent utilisés pour les pages individuelles comme les cartes). Les scanners constituent une solution de substitution intéressante aux appareils photo, car le volume d'équipement requis est souvent moins important. Toutefois, ils présentent un certain nombre d'inconvénients:

- a) L'utilisation d'un scanner à plat peut endommager la reliure d'un livre et il arrive que les données situées à proximité de la reliure soient floues ou ne soient pas reproduites. Ce problème ne se pose pas avec un appareil photo;
- b) Il est nécessaire d'utiliser un scanner spécial pour les microfilms et les microfiches;
- c) Les scanners comportent de nombreuses pièces mobiles et nécessitent généralement des opérations de maintenance et une licence;
- d) Il est souvent nécessaire de suivre une formation spécialisée pour utiliser les scanners plus sophistiqués. Ainsi certains projets ont-ils été interrompus en raison de la rotation du personnel. Dans d'autres cas, une formation insuffisante et des problèmes de licence ont empêché le recours aux scanners pour la création d'images électroniques;
- e) À l'exception des modèles très coûteux, les scanners sont généralement lents et d'un emploi peu commode.

2. APPAREILS PHOTOGRAPHIQUES

Les appareils photo constituent la solution la plus couramment choisie pour la création d'images numériques. Ils donnent de bons résultats, aussi bien pour les documents reliés que pour les feuilles volantes, ainsi que pour les documents de grande dimension comme les cartes. Les appareils photo numériques sont très répandus et même les appareils automatiques peuvent produire des images fiables, surtout si l'éclairage est constant et de bonne qualité. En utilisant un pied fixe et un système d'éclairage, une télécommande pour prendre les photos et un adaptateur de courant alternatif, il suffit de tourner les pages, le processus est rapide et efficace. De plus il existe désormais des logiciels bon marché permettant de renommer les images numériques et de les transférer vers un ordinateur. L'archivage des images est également très rapide.

3. FONCTIONNALITÉS DE L'APPAREIL PHOTOGRAPHIQUE

On trouvera dans le tableau 5 une liste des équipements requis pour mettre en place un dispositif de prise de vues. On aura d'abord besoin d'une table, sur laquelle on posera l'ordinateur et l'équipement photographique (y compris un trépied ou un support). Il est essentiel de disposer d'un nombre suffisant de prises électriques reliées à la terre. L'éclairage naturel ou la lumière du jour constituent l'option idéale pour les prises de vue, du moment que la lumière est diffuse

Tableau 5. Équipement photographique nécessaire

1	Table
2.	Prises électriques
3.	Support appareil photo/trépied
4.	Éclairage
5.	Appareil photo et accessoires
6.	Ordinateur

et que l'on évite les ombres; cependant, il n'est pas toujours possible de disposer d'une pièce bénéficiant d'un éclairage naturel. Le dispositif de prise de vues inclut donc souvent deux ou quatre lampes fixées au trépied ou au support de l'appareil qui le maintient en position et permet de réaliser un grand nombre de clichés très rapidement. Il est important de vérifier fréquemment la qualité de l'éclairage, étant donné que certains supports papier sont brillants et reflètent la lumière dans l'objectif, ce qui rend illisibles certaines des données. Le recours à une télécommande évite au photographe d'avoir à se placer à côté de l'appareil et de risquer de le faire bouger. Il est indispensable de disposer d'un ordinateur pour stocker les images numériques.

Les batteries des appareils photo se déchargent relativement vite. Avec un adaptateur secteur, on peut effectuer un nombre illimité de prises de vue et réaliser de considérables économies de temps et d'argent puisque l'on n'a pas à acheter et à remplacer les batteries. S'il n'est pas possible d'utiliser un adaptateur secteur avec l'appareil photo dont on dispose, il est recommandé de se procurer un jeu de batteries rechargeables et un chargeur. Si l'on ne peut travailler avec un adaptateur, il est important de choisir un appareil dont on puisse changer les batteries sans le déplacer de son support. La plupart des appareils numériques se fixent sur les supports au moyen d'un pas de vis en plastique qui tend à s'user au bout de quelques dizaines de manipulations pour remplacer les batteries.

Il faut disposer d'un logiciel et d'un câble USB ou avoir accès à un réseau de communication sans fil (Wifi) pour transmettre automatiquement l'image de l'appareil à l'ordinateur. Si le logiciel permet d'actionner l'obturateur de l'appareil au moyen de la souris, un câble de télécommande ne sera pas forcément nécessaire. En l'absence de port USB ou d'accès au Wifi, il faut se procurer une carte mémoire de très haute capacité. Les appareils photo numériques sont souvent équipés de logiciels qui génèrent un nom de fichier pour chaque image et transmettent directement les photographies à l'ordinateur, ce qui permet au photographe et à l'archiviste de réaliser d'énormes gains de temps. Il peut également être utile de disposer d'un zoom et de pouvoir modifier les réglages de l'appareil. On trouvera dans le tableau 6 une liste des accessoires photographiques recommandés pour garantir le bon déroulement des opérations.

Tableau 6. Accessoires photographiques propices au bon déroulement des opérations

1.	Banc de reproduction pour un appareil photo et deux à quatre lampes
2.	Adaptateur secteur pour alimenter l'appareil photo
3.	Télécommande pour prendre les photos ou logiciel décrit au point 4
4.	Logiciel intégré à l'appareil permettant de prendre des photographies et de les transférer directement vers un ordinateur au moyen d'un câble USB ou d'une connexion Wifi
5.	À défaut, la plus grande carte mémoire disponible
6.	Capacité de renommer les fichiers images et logiciel correspondant
7.	Zoom, si disponible, et objectif macro pour les gros plans
8.	Possibilité de modifier les réglages de l'appareil (sur l'appareil ou sur l'ordinateur)
9.	Vitre antireflet pour aplatir délicatement les archives reliées

Tableau 7. Fonctionnalités courantes sur les appareils photo permettant de garantir la qualité de l'image

1.	Possibilité de choisir la dimension de l'image
2.	Possibilité de désactiver le flash
3.	Mode macro pour les gros plans
4.	Possibilité de régler la vitesse d'obturation
5.	Possibilité de régler la balance des blancs

La taille de l'image est un paramètre à prendre en compte, surtout si l'on utilise une carte mémoire. En fonction de la qualité de l'image, il ne sera probablement pas nécessaire d'opter pour la taille maximale et une image de taille moyenne suffira. On constate qu'un appareil photo d'une définition de 4 mégapixels seulement suffit pour produire des fichiers JPEG dont la résolution permet la lecture des valeurs alphanumériques. Or, de nos jours, même les appareils photo des téléphones portables ont une définition d'environ 12 mégapixels. On se souviendra toutefois qu'une résolution très fine, si elle n'est pas obligatoire pour la récupération des données alphanumériques, peut être nécessaire si l'on utilise un logiciel de numérisation automatique. Généralement les photographies sont prises en couleur et non en noir et blanc. La couleur permet une lecture plus aisée des corrections et des annotations portées sur les carnets d'origine.

D'autres fonctionnalités à envisager sont l'option de désactivation du flash, un objectif macro pour les gros plans, et la possibilité de régler la vitesse d'obturation, ainsi que la balance des blancs. Le tableau 7 reprend les fonctionnalités courantes permettant d'obtenir une image de qualité.

APPENDICE 7. AUTRES MÉTHODES DE NUMÉRISATION

On trouvera ici des considérations générales sur les méthodes de numérisation autres que la saisie par un opérateur. Elles nécessitent toutes un contrôle de la qualité du résultat final, puisque toute transcription est susceptible de s'accompagner d'erreurs.

1. **RECONNAISSANCE OPTIQUE DE CARACTÈRES, RECONNAISSANCE INTELLIGENTE DE CARACTÈRES ET RECONNAISSANCE INTELLIGENTE DE MOTS**

La reconnaissance optique de caractères (OCR) est une technologie logicielle courante permettant de numériser un texte imprimé, mais qui ne peut prendre en charge l'écriture manuscrite. La reconnaissance intelligente de caractères (ICR) cible les caractères manuscrits individuels et prend en charge différents types de polices et de styles d'écriture. Cette technologie fonctionne au mieux pour la lecture de formulaires structurés et peut atteindre une précision supérieure à 97 %. La reconnaissance intelligente de mots permet de transcrire des mots manuscrits entiers, qu'ils soient en caractère d'imprimerie ou en cursives. Les formulaires anciens ne sont généralement pas compatibles avec les technologies ICR/OCR, qui ne permettent pas de les numériser actuellement. Ces deux méthodes utilisent la technologie de l'apprentissage machine en intégrant automatiquement dans leurs bases de données les nouveaux types d'écriture manuscrite rencontrés.

À l'heure actuelle, ces technologies sont hors de prix et, comme toutes les méthodes de numérisation, elles nécessitent un contrôle de la qualité pour garantir l'exactitude de la transcription. Elles continuent à évoluer et leur fiabilité va sans doute s'améliorer au fil du temps. Actuellement la numérisation par saisie reste la méthode la plus couramment utilisée puisque l'on estime que le cerveau humain est l'outil le mieux adapté à la transcription de textes manuscrits.

2. **PRODUCTION PARTICIPATIVE**

Dans les projets de production participative, l'information est traitée en ligne par des bénévoles, souvent qualifiés de «citoyens pour la science». Un certain nombre de programmes bien établis relatifs au climat, comme OldWeather et le réseau communautaire collaboratif «pluie, grêle et neige» (*Community Collaborative Rain, Hail & Snow* ou CoCoRaHS), ont recours à cette méthode. Les bénévoles du projet OldWeather saisissent le texte de livres de bord de navires (datant pour la plupart du début du XX^e siècle) sur un formulaire en ligne. Chaque livre de bord est saisi par trois personnes dont le travail est comparé avant publication du résultat final (voir sur le [portail I-DARE](#) le lien vers le projet OldWeather). Les bénévoles du projet CoCoRaHS mesurent les précipitations dans des pluviomètres de 10 cm, situés généralement à proximité de leur domicile, et saisissent quotidiennement ces données sur un formulaire en ligne. Elles sont ensuite introduites dans une base de données puis cartographiées.

Bien que la saisie soit effectuée par des bénévoles, la production participative a tout de même un coût. Il faut du temps, des compétences et des fonds pour concevoir et gérer l'infrastructure de saisie des données en ligne, créer la base de données, contrôler la qualité, élaborer les instructions destinées aux opérateurs de saisie et aux observateurs, et mener à bien des actions de communication destinées à attirer de nouveaux bénévoles. Comme toutes les autres technologies, la production participative ne peut faire l'économie du contrôle qualité.

3. NUMÉRISATION DES DIAGRAMMES

La numérisation d'un diagramme consiste généralement pour les observateurs à transcrire les valeurs qui y figurent dans un tableau, ce qui est un travail fastidieux, surtout lorsque le diagramme comporte des fluctuations rapides, comme les anémographes, ou plusieurs courbes. S'il est possible de lire sur un diagrammes les valeurs correspondant à des résolutions temporelles fines, il arrive souvent que seules les valeurs maximales ou quotidiennes soient relevées. Avec un logiciel de numérisation spécialisé, il suffit de tracer une courbe reprenant les points du diagramme pour que ces derniers soient convertis en valeurs numériques. La résolution temporelle des données va de la minute à l'heure. Dans les années 1990, plusieurs programmes de numérisation des données pluviométriques (pluviographes) ont été mis au point par des groupes individuels, mais ils ne fonctionnent que sur des ordinateurs et des systèmes d'exploitation spécifiques. Le code source de ces logiciels est accessible, mais il n'est pas mis à jour (pour tenir compte des langages de programmation et technologies informatiques plus récents) et devrait être adapté par un spécialiste. De plus, il faut une formation pour gérer et utiliser ces logiciels.

Plusieurs nouveaux logiciels de numérisation ont été développés récemment et sont opérationnels. Ces programmes permettent notamment de calculer le volume des précipitations à différentes échelles temporelles (heures à minutes), de lire différents types de diagrammes et de traiter plusieurs courbes sur un même diagramme. On trouvera la documentation de ces logiciels sur le [portail I-DARE](#). Certains de ces programmes sont disponibles en utilisation libre à condition que les données soient saisies dans une base de données ouverte et accessible à tous.

APPENDICE 8. ERREURS DE NUMÉRISATION ET ASSURANCE QUALITÉ

Le présent appendice passe brièvement en revue la question du contrôle de la qualité dans le contexte des erreurs de numérisation concernant essentiellement la saisie. Il se peut que certaines, voire la totalité des fonctions de contrôle qualité décrites ici soient prévues dans le SGDC d'un pays Membre. Dans ce cas, les erreurs sont signalées au moment de la saisie des données dans le SGDC, puis corrigées. Si le SGDC n'est pas doté de ces fonctionnalités, il est fortement conseillé de remédier à cette lacune. Le contrôle initial de la qualité des données numériques nouvellement saisies est axé sur les erreurs associées à la saisie, à l'instrument, à l'observateur et aux métadonnées, et pas sur les analyses plus fines réalisées au cours de l'homogénéisation des données, ce que l'on qualifie parfois de contrôle qualité «en mode différé». Il convient de considérer (et de signaler) les données saisies comme provisoires jusqu'à leur vérification par des tests de contrôle de la qualité.

Si les stratégies de la double saisie, de l'emploi de modèles et de la saisie littérale limitent les erreurs de saisie des valeurs individuelles, elles ne les éliminent cependant pas complètement. Les problèmes dus à l'imperfection des métadonnées (nom et numéro d'identification des stations pour l'essentiel), à la qualité des images scannées, au dysfonctionnement des instruments, et à des pratiques d'observation non normalisées continuent à se poser. Pour garantir la représentation numérique la plus fidèle possible des données issues des formulaires météorologiques, il convient de procéder à des tests de contrôle qualité pour signaler les valeurs douteuses, aussi bien des métadonnées (nom ou identifiant de la station, année ou mois, type d'élément par exemple) que des variables météorologiques. Les valeurs douteuses sont signalées avant de faire l'objet d'une vérification manuelle, le type d'erreur et le résultat de la vérification étant ensuite spécifiés.

Les erreurs de saisie courantes sont les suivantes: chiffres intervertis, virgules de décimale manquantes ou mal placées, chiffres manquants (un zéro par exemple), périodes d'accumulation incorrectes et valeurs manquantes. D'autres inexactitudes se rencontrent fréquemment: a) erreurs de l'observateur (unités non conventionnelles par exemple); b) dysfonctionnement de l'instrument; c) problèmes de métadonnées (incohérence dans l'identifiant ou la désignation de la station, année ou mois incorrect); d) interversion ou duplication d'un élément particulier pour un mois et une station donnée; et e) manque de clarté des formulaires. Ces erreurs peuvent être le signe de problèmes touchant au formatage ou, plus systématiquement, à la qualité des données.

Des indicateurs sont utilisés pour déterminer l'origine de chaque erreur. Après examen de l'erreur signalée, la personne chargée de l'évaluation indique le résultat de cette vérification, qui peuvent se décliner comme suit: a) valeur correcte, correspondant au formulaire, raisonnable, inchangée; b) une erreur, signalée mais n'entraînant pas de modification des données; c) une erreur, mais valeur de remplacement spécifiée; d) valeur ne correspondant pas au formulaire d'origine, corrigée en conséquence; e) une erreur, valeur manquante; ou f) valeur à saisir, puis saisie. Généralement, quand une valeur signalée est ambiguë, mais pourrait être correcte, on ne la modifie pas.

Les tests et les procédures de contrôle de la qualité donnent aux utilisateurs la garantie que les données numérisées sont une représentation exacte des observations enregistrées dans les documents d'origine. Si les tests ne sont pas effectués automatiquement par le SGCD, il convient de créer un inventaire du contrôle qualité sur tableur, afin de suivre les tests effectués pour chaque station. Les valeurs aberrantes de chaque test doivent faire l'objet d'une vérification manuelle et les éventuelles corrections apportées ensuite. Après chaque étape du contrôle de la qualité des données d'une station, il importe de mettre à jour le tableur.

Les tests de contrôle de la qualité incluent des contrôles par fourchette (des valeurs quotidiennes et des totaux et moyennes mensuels), des vérifications de cohérence interne (en comparant les mesures entre les types de données) et le contrôle des valeurs extrêmes (réalisé sur des valeurs

individuelles). Pour plus d'informations, on se reportera au *Guide des pratiques climatologiques de l'OMM* (OMM-N° 100 (2011), section 3.4) et à Westcott *et al.* (2011) ainsi qu'à d'autres sources disponibles sur le [portail I-DARE](#).

APPENDICE 9. ENSEIGNEMENTS

1. **PROMOUVOIR UNE COOPÉRATION PLUS ÉTROITE ENTRE LES DIVISIONS DE LA PRÉVISION ET DE LA CLIMATOLOGIE AU SEIN DES SERVICES MÉTÉOROLOGIQUES ET HYDROLOGIQUES NATIONAUX**

La Division de climatologie d'un SMHN est souvent chargée de la supervision des opérateurs de terrain sur les sites météorologiques, du contrôle de la qualité des données climatologiques nationales (tableaux horaires, quotidiens), ainsi que de la saisie d'une partie des données dans son propre SGDC. En règle générale, les données en code SYNOP sont transmises par les observateurs météorologiques via le Système mondial de télécommunications de l'OMM, mais ces données numériques ne sont en général accessibles que par la Division des prévisions du SMHN et souvent uniquement en temps réel. Il est courant que seuls les tableaux manuscrits horaires et quotidiens des données soient envoyés à la Division de climatologie, mais pas les données numériques codées. De plus, si la Division des prévisions archive les données météorologiques décodées des stations SYNOP ou des stations automatiques, il arrive souvent que la Division de climatologie n'y ait pas accès. En conséquence, cette dernière doit consulter les bases de données internationales pour obtenir les données nationales. L'idéal serait d'archiver les données numériques transmises dans une base de données centrale consultable aussi bien par les prévisionnistes que par les climatologues et à partir de laquelle seraient réalisés tous les produits. Une meilleure coordination entre ces groupes permettrait à la Division de climatologie de disposer d'un jeu de données plus actuel et surtout plus complet pour l'élaboration d'analyses climatologiques. On pourrait également réaliser des comparaisons avec les rapports papier afin de valider les valeurs extrêmes.

2. **EXAMINER LES PRATIQUES ACTUELLES EN MATIÈRE DE NUMÉRISATION, DE COLLECTE DES MÉTADONNÉES ET DE CONTRÔLE DE LA QUALITÉ; BESOINS EN MATIÈRE DE FORMATION**

La Division de climatologie d'un SMHN se dote souvent de sa propre méthode de saisie des données et de contrôle de la qualité, qui évolue en fonction des effectifs disponibles, de ses ressources dans les domaines de la communication et de l'informatique, et du nombre et du type de stations météorologiques existant dans le pays. Il est conseillé de faire régulièrement le point sur les pratiques en vigueur et d'intégrer les directives pertinentes de l'OMM dans l'organisation de l'inventaire, de la saisie et du contrôle de la qualité des données actuelles. On se reportera à la section du [portail I-DARE](#) sur les bonnes pratiques.

Il est recommandé de prendre en compte toutes les informations glanées lors des opérations de sauvetage des données qui sont susceptibles de renforcer cette démarche ou le processus d'intégration des données actuelles, ou qui pourraient permettre de localiser d'autres sources de données à sauvegarder. La qualité des données climatologiques est indissociable de la rigueur des observateurs et des analystes.

Les personnes chargées des opérations de sauvetage des données seront appelées à travailler avec des équipements et des logiciels nouveaux qui ne leur seront pas familiers. Il faudra qu'elles bénéficient d'une formation et acquièrent de l'expérience dans l'utilisation d'appareils photo numériques, de scanners et de logiciels de numérisation, la consultation des bases de données et la réalisation du contrôle de la qualité. En raison de la rotation du personnel, il convient de former plusieurs personnes sur le tas au moyen de guides pratiques, pour éviter tout point de défaillance. Ces guides doivent être mis à jour en fonction de l'évolution des programmes et des utilisateurs.

3. DÉMONTRER L'INTÉRÊT DU SAUVETAGE DES DONNÉES

La réorganisation des relevés, la création d'images électroniques, la numérisation des données et les mesures de contrôle de la qualité inhérentes au sauvetage des données peuvent être coûteuses et prendre beaucoup de temps. Ceci explique sans doute en partie pourquoi peu de SMHN ont jusqu'à présent mis en place ce type de démarche. Le lancement et la pérennisation d'un programme de sauvetage des données commence par le SMHN, qui doit clairement en démontrer la nécessité. Ceci peut passer par la mise en avant de questions pratiques auxquelles on sera en mesure de répondre quand on disposera de données climatologiques de meilleure qualité. Il se peut par exemple que les données climatologiques anciennes permettent au SMHN de répondre rapidement et avec précision aux questions d'ordre hydrométéorologique d'un ministre sur la fréquence des épisodes de sécheresse ou de fortes pluies; le SMHN pourrait aussi indiquer la direction dominante du vent au cours des cent dernières années afin d'orienter correctement la nouvelle piste d'un aéroport d'une valeur de dix millions de dollars des États-Unis d'Amérique. Ces informations constituent un produit «à valeur ajoutée» que les directeurs des SMHN pourraient monnayer, dotant ainsi leur institution d'une source supplémentaire de revenus. De plus, l'organisation des données climatologiques peut permettre de récupérer des centaines de mètres carrés dans les locaux du SMHN sans pour autant se débarrasser d'archives historiques précieuses. L'espace ainsi libéré peut être utilisé pour créer des bureaux supplémentaires ou des salles de formation. Enfin, si le SMHN est prêt à partager les données sauvées et numérisées, il gagnera une réputation internationale de précurseur dans ce domaine.

Les activités de sauvetage des données nécessitent souvent un soutien externe sur une durée de plusieurs années. On trouvera à l'appendice 10 d'autres arguments en faveur de ces activités. Les personnes ou les groupes ayant une expérience du sauvetage des données peuvent servir de porte-parole, donner des conseils et contribuer à convaincre d'éventuels bailleurs de fonds.

En entrant en contact avec d'autres parties intéressées par le sauvetage des données, on peut obtenir des informations techniques et des solutions possibles à des problèmes inattendus rencontrés en cours de projet. On trouvera sur le [portail I-DARE](#) des informations techniques ainsi que des contacts possibles. Les SMHN peuvent s'adresser au Secrétariat de l'OMM pour solliciter l'aide d'autres Membres.

APPENDICE 10. APPLICATIONS PERMETTANT DE DÉMONTRER LA VALEUR DES DONNÉES

1. VALIDATION DE VALEURS EXTRÊMES DE PARAMÈTRES MÉTÉOROLOGIQUES ET DE VALEURS INDIRECTES

On estime qu'un climat changeant et plus variable débouchera à l'avenir sur des valeurs plus extrêmes des températures, des précipitations, de la pression atmosphérique et du vent. Ces valeurs futures ne pourront être validées comme constituant des «records» si l'on ne connaît pas les valeurs extrêmes du passé. La préservation des documents d'origine et la numérisation des relevés météorologiques anciens faciliteront le processus de validation des valeurs extrêmes actuelles et futures. En disposant des valeurs vérifiables des paramètres météorologiques, on aura une meilleure connaissance de la distribution globale des valeurs extrêmes et l'on pourra déterminer si elle évolue. Les études climatologiques réalisées par le passé sur des périodes ou dans des régions pour lesquelles on ne dispose que de peu de données météorologiques, voire d'aucune donnée, se sont appuyées sur des données indirectes (anneaux de croissance des arbres, carottes de glace). Les données résultant du sauvetage de relevés anciens sont d'une grande utilité pour vérifier la validité de ces méthodes.

2. INTÉGRATION DES DONNÉES DANS DES ÉTUDES, LIENS AVEC DES PROJETS DE RECHERCHE

La meilleure façon de démontrer l'intérêt que revêt la récupération des données climatologiques anciennes est de les employer dans des travaux de recherche ou d'inciter d'autres à le faire. Les longues séries chronologiques permettent de mettre en évidence et d'étudier la gravité et la fréquence de phénomènes passés, tels que les sécheresses et les inondations, les vagues de chaleur et les périodes de froid, les tornades, les cyclones tropicaux et les tempêtes de grêle et de poussière. Si les météorologues et les climatologues sont surtout intéressés par l'étude des causes des phénomènes météorologiques, d'autres scientifiques le sont par leur impact sur la société et même sur l'histoire d'un pays. Ces retombées, ainsi que les réactions humaines à ces phénomènes passés, peuvent contribuer au travail de préparation à de futurs phénomènes météorologiques extrêmes. Le recours à des données anciennes pour étudier les phénomènes climatiques du passé pourrait fort bien susciter l'enthousiasme des bailleurs de fonds potentiels, surtout si ces mêmes phénomènes se produisent actuellement ou sont attendus à l'avenir. Si l'on parvient à démontrer une climatologie (évolutive) de ces phénomènes, on sera peut-être mieux à même de prédire leur probabilité future. L'accès à des séries chronologiques complètes de données permet également d'ajuster et de vérifier les modèles climatiques.

3. MISES À DISPOSITION DES DONNÉES AUX UTILISATEURS

Il est important de permettre aux utilisateurs de consulter facilement les données climatologiques. La mise à disposition de données anciennes fiables est propice au développement d'échanges entre chercheurs et à la contribution des étudiants. Les exploitants agricoles et les responsables de la gestion de l'eau sont des alliés potentiels dans la défense du sauvetage des données climatologiques; il se peut d'ailleurs qu'ils disposent eux-mêmes de relevés ou de ressources propres leur permettant d'y contribuer. Ils considéreront très probablement les données sauvegardées comme étant essentielles pour évaluer la sensibilité des systèmes naturels et anthropiques aux phénomènes et extrêmes climatiques. De plus, la création

de produits susceptibles d'intéresser le grand public (comme la climatologie des jours fériés) peut susciter un intérêt pour la météorologie et contribuer à la mise en place de groupes de «citoyens pour la science».

APPENDICE 11. LES DIFFÉRENTES ÉTAPES DU SAUVETAGE DES DONNÉES

1. Rechercher et localiser les jeux de données anciennes: relevés papier, microfilms/microfiches et données numériques se trouvant dans les bibliothèques et les centres climatologiques locaux, régionaux, nationaux et internationaux
2. Organiser et nettoyer les relevés sur papier et les microfilms/microfiches et les entreposer dans des boîtes à archives étiquetées rangées sur des étagères ou dans des armoires de classement, dans des salles sécurisées équipées de systèmes de régulation de la température et de l'humidité
3. Créer un inventaire électronique des relevés sur papier ou microfilm, en indiquant la station, l'année, le mois, le type de support, le type de formulaire et l'emplacement de la boîte ou du classeur à tiroirs
4. Créer un inventaire des images électroniques sur un tableur en indiquant ce qui a été reproduit sous forme d'image (ou pas) et validé (ou pas)
5. Créer sur un tableur un inventaire électronique des données numériques en indiquant ce qui a été numérisé (ou pas), et soumis à un contrôle de qualité (ou pas)
6. Déterminer l'ordre de priorité dans la numérisation en mode image et dans la numérisation en mode texte
7. Faire le point sur les mesures en vigueur en matière de numérisation et de contrôle de la qualité, afin de déterminer comment le flux de données nouvellement numérisées peut s'intégrer au travaux du SMHN et s'il faut modifier les mesures en vigueur
8. Définir les besoins en matière d'informatique (SGDC), de personnel, d'espace, de numérisation en mode image et de numérisation en mode texte
9. Élaborer un plan de travail pour la numérisation en mode image et la numérisation en mode texte
10. Obtenir, si nécessaire, les équipements requis et engager du personnel
11. Mettre au point un processus de numérisation en mode image qui inclut l'archivage des données, des tableurs pour les inventaires, ainsi que le contrôle de la qualité des images et des métadonnées
12. Reproduire sous forme d'image toutes les copies des données et des métadonnées et archiver de manière appropriée les images ainsi obtenues
13. Inclure l'inventaire des images dans les répertoires des ordinateurs et sur des CD
14. Réaliser des copies des archives d'images et les stocker hors site
15. Conserver les relevés originaux sur papier ou microfilm
16. Mettre au point un processus de numérisation des données incluant la création d'un formulaire modèle et d'une base de données, un inventaire actualisé et le contrôle de la qualité des données et métadonnées
17. Faire le point sur les processus de numérisation en mode image et de numérisation en mode texte au fur et à mesure que le travail avance pour déterminer s'il est possible de les améliorer

18. Procéder à en recoupement entre, d'une part, les données saisies sous forme d'images et numérisées et, d'autre part, les relevés effectués durant la même période dans les stations, pour garantir leur exhaustivité
 19. Faire quotidiennement des copies de sauvegarde des données numériques de l'ordinateur/SGDC
 20. Mettre à jour régulièrement l'inventaire des supports et les inventaires des données numériques en intégrant les données et métadonnées actuelles
 21. Créer de multiples copies des données numériques et les stocker sur plusieurs sites
 22. Élaborer un plan de migration technologique et veiller à la migration régulière de toutes les données électroniques (images et données numérisées) vers de nouveaux supports de stockage
 23. Créer des produits climatologiques à partir des données dont la qualité a été contrôlée
 24. Mettre les données à la disposition des utilisateurs d'une manière pratique et accessible
-

APPENDICE 12. GLOSSAIRE

ASCII. Code américain normalisé pour l'échange d'informations: c'est le format informatique le plus couramment utilisé pour les fichiers de texte.

Banc de reproduction. Dispositif permettant de maintenir les appareils photo et les lampes en position pour photographier des documents.

DARE. Sauvetage des données. Acronyme utilisé par l'OMM pour désigner ses activités de sauvetage des données climatologiques.

Diagramme à déroulement continu. Diagramme créé par un enregistreur à tambour mû par un mouvement d'horlogerie qui affiche automatiquement des mesures de paramètres météorologiques tels que la température, la direction et la vitesse du vent et la pression atmosphérique (identique à Diagramme d'appareil enregistreur).

Diagramme d'appareil enregistreur. Diagramme créé par un enregistreur à tambour mû par un mouvement d'horlogerie qui affiche automatiquement des mesures de paramètres météorologiques tels que la température, la direction et la vitesse du vent ou la pression atmosphérique (identique à Diagramme à déroulement continu).

Données numériques. Données saisies dans un ordinateur et stockées sous forme de nombres, et non pas en format analogue ou sous forme de diagrammes. Ces nombres sont saisis soit dans des formulaires en format texte brut soit dans une base de données relationnelle de type Microsoft ACCESS ou ORACLE, ou dans celle d'un SGDC.

Image. Reproduction d'un document comportant des données climatologiques réalisée au moyen d'un appareil photo numérique ou d'un scanner.

Inventaire électronique. Tableur contenant des informations sur des archives de données.

Métadonnées. Ensemble d'attributs ou d'éléments nécessaires pour décrire une ressource. Dans le domaine du sauvetage des données, ce terme désigne aussi bien les informations relatives aux stations que celles relatives à l'inventaire des données climatologiques.

Microfiche. Type de microfilm ayant la forme d'une feuille rectangulaire.

Microfilm. Méthode de reproduction des images des relevés dans un format très réduit sur une pellicule photographique.

Migration. Démarche visant à contrecarrer l'obsolescence technologique des ordinateurs et des logiciels et consistant à transférer des données d'un support informatique à un autre afin de préserver le contenu intellectuel de l'objet numérique.

Numérisation. Processus consistant à transposer des données analogues sous forme numérique afin de permettre leur traitement informatique. Passe généralement par la saisie des données textuelles ou le tracé de points à partir d'un diagramme à déroulement continu.

Préservation. Protection des données contre les risques de perte ou de détérioration.

Rafraîchissement. Copie d'informations d'un support de stockage vers lui-même ou vers un autre support sans modification du format.

Reformatage. Copie d'informations d'un support de stockage à un autre ou conversion d'un fichier dans un autre format.

Saisie. Transfert de données vers un ordinateur réalisé au moyen d'un clavier.

Scanner. Produire, au moyen d'un scanner, un fichier image à partir d'un relevé analogue ou de la copie papier d'un relevé numérique.

Scanner. Dispositif employé pour scanner des documents sous forme de fichiers numériques.

SGDC. Système de gestion des données climatologiques.

SMHN. Service météorologique et hydrologique national.

Support. Papier, microfilm, microfiche, CD/DVD, bandes, disques durs ou ordinateurs sur lesquels sont enregistrées des données.

Tableur. Type de format fréquemment utilisé pour la production d'inventaires. Ce terme désigne ici les logiciels Microsoft EXCEL ou Lotus 1-2-3.

APPENDICE 13. BIBLIOGRAPHIE

- Bojinski, S., M. Verstraete, T.C. Peterson, C. Richter, A. Simmons et M. Zemp, 2014: The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Amer. Meteor. Soc.*, 95:1431–1443.
- Brunet, M. et P. Jones, 2011: Data rescue initiatives: bringing historical climate data into the 21st century, *Climate Research*, 47:20–40.
- Westcott, N.E., K. Andsager, L. Stoecker, M.L. Spinar, R. Smith, R. Obrecht, M.C. Kruk, R.T. Truesdell et D. O’Connell, 2011: Quality control of 19th century weather data. Midwestern Regional Climate Center Contract Report 11–04, Illinois State Water Survey, 68 p.
- Organisation météorologique mondiale (OMM), 2003: *Guidelines on Climate Metadata and Homogenization*. WMO/TD-1186, OMM, Genève.
- , 2004: *Guidelines on Climate Data Rescue*. WMO/TD-1210, OMM, Genève.
- , 2009: *Guidelines on Analysis of Extremes in a Changing Climate in Support of Informed Decisions for Adaptation*. WMO-TD 1500, WCDMP-No. 72, OMM, Genève.
- , 2010: *Guide des instruments et des méthodes d’observation météorologiques*. (Guide de la CIMO), OMM–N° 8 (mise à jour de l’édition 2008).
- , 2011: *Guide des pratiques climatologiques*. OMM–N° 100 (édition 2011), OMM, Genève.
-

Pour de plus amples informations, veuillez vous adresser à:

Organisation météorologique mondiale

7 bis, avenue de la Paix – Case postale 2300 – CH 1211 Genève 2 – Suisse

Bureau de la communication et des relations publiques

Tél.: +41 (0) 22 730 83 14/15 – Fax: +41 (0) 22 730 80 27

Courriel: cpa@wmo.int

public.wmo.int