

Gestion et valorisation de données sur l'environnement global

L'exemple de Médias-France

Michel Hoepffner
Éliane Cubero-Castan
Jean-Luc Boichard

Les recherches dans le domaine de l'environnement global

L'environnement global évolue rapidement, entre autres sous les effets anthropiques, par exemple avec l'augmentation des émissions de gaz à effet de serre. Pour respecter le principe de précaution, et pour mieux comprendre le fonctionnement des systèmes naturels, il faut suivre cette évolution et ses conséquences sur l'environnement humain. C'est ainsi que pourront être proposées des prévisions pour l'avenir, en particulier pour ce qui concerne les modifications de cet environnement à l'échelle régionale (le bassin méditerranéen ou le Sahel africain, par exemple) sur le cycle de l'eau et leurs conséquences dans les domaines de la production agricole, de l'alimentation en eau ou de la préservation de la biodiversité.

Le PMRC (programme mondial de recherches sur le climat de l'organisation météorologique mondiale), le PIGB (Programme international géosphère-biosphère de l'ICSU) et l'IHDP (programme international sur les dimensions humaines des changements globaux de l'environnement de l'International Social Science) et leurs correspondants au niveau national (programme national d'étude de la dynamique du climat, programme atmosphère moyenne et océan superficiel, programme national de télédétection spatiale, programme national de recherche en hydrologie...) ont initié et soutenu de nombreux projets internationaux de recherche dans ce domaine (GEWEX, CLIVAR, BAHC, JGOFS, IGAC, PAGES...), ainsi que des expériences de terrain (HAPEX-Mobilhy et HAPEX-Sahel, FIFE, BOREAS, EXPRESSO, SALSAMEX, FETCH...). Celles-ci ont permis de recueillir des données de toute sorte, acquises au sol, par avion et par satellite, afin de mieux comprendre les mécanismes qui conditionnent l'évolution de notre environnement, de proposer des modèles qui permettent de reproduire leur fonctionnement, et d'essayer de prévoir l'évolution de cet environnement selon les scénarios les plus réalistes, avec les validations que fournissent les données paléoclimatologiques sur les évolutions anciennes du climat.

Les conditions d'acquisition, de stockage et de mise à disposition des données

Les données ainsi recueillies sont nombreuses et variées. Afin de les exploiter dans les meilleures conditions, il a été convenu dans la plupart des cas que ces données devaient être gérées en commun et mises à la disposition des chercheurs selon des règles spécifiques à chaque expérience, et définies de commun accord entre les parties prenantes.

Ces règles de mise à disposition des données doivent aboutir à une utilisation optimale et croisée de ces données, avec la participation sans exclusives de chercheurs d'origine variée (nationalité, organisme, thématique...), en distinguant dans un premier temps ceux qui ont participé effectivement à l'acquisition de ces données, et en ouvrant totalement dans un deuxième temps l'accès de ces données.

Le logiciel à constituer pour stocker toutes les données et permettre aux chercheurs d'acquérir celles qui les intéressent et auxquelles ils ont droit doit tenir compte des meilleures techniques informatiques, mais aussi et surtout être robuste, fiable et évolutif.

Ces données doivent être acquises par le gestionnaire des données au fur et à mesure de leur acquisition. C'est donc préalablement au projet de recherche que doivent être définies les politiques d'acquisition, de stockage et de mise à disposition des données.

L'exemple de Médias-France

Le groupement d'intérêt public Médias-France, qui associe le Cnes, le Cnrs, l'Ird, Météo-France, l'université Paul-Sabatier, Spot-Image et CLS dans le soutien à la recherche sur l'environnement global, notamment dans le bassin méditerranéen et sur le continent africain, a entre autres pour charge de constituer et de gérer les bases de données dans ce domaine.

Il a ainsi aidé (ou aide actuellement) à la constitution ou à la gestion des bases de données :

- ◆ des expériences HAPEX-Sahel, EXPRESSO, FETCH, O3O, ESCOMPTE... ;
- ◆ des réseaux MOZAIC, IDAF, FORMAT, APD, AIRQUAL... ;
- ◆ du projet JGOFS ;
- ◆ météorologiques et océanographiques sur la mer Méditerranée (CD-ROM Méditerranée) ;
- ◆ des sorties du modèle du Centre Européen de Prévision de Reading (EMET-CEP).

Caractéristiques des données « Médias »

Ce sont des données statiques, accessibles uniquement en lecture, durables et précieuses. Elles sont accompagnées de descripteurs (méta-données) indiquant le type de capteur, les modèles utilisés, les conditions d'acquisition et le type de traitement effectué.

La sauvegarde, la disponibilité et la qualité de la documentation associée priment sur tout autre aspect.

L'offre Médias-France

Les prestations possibles sont le support aux activités d'observation et de modélisation, la réalisation de sites Internet, la conception et la réalisation de bases de données, l'édition de livres électroniques, ainsi qu'une possibilité de suivi et d'installation des applicatifs créés.

Le développement comprend la conception du système, le traitement des données avec l'acquisition des données, le formatage et l'archivage, la structuration et le maquettage IHM d'extraction et de visualisation.

Principes directeurs

Principes généraux

La réduction des coûts, l'utilisation maximale de systèmes et de logiciels libres, la réutilisation de composants métiers (les classes Java et widgets communs), l'utilisation de bibliothèques partagées (ListPack, Gd, Zlib...), ainsi qu'une politique d'accès aux données laissée aux donneurs d'ordre sont les principes directeurs que s'est donné Médias-France.

La portabilité

Elle est réalisée avec l'aide de solutions de type « Application Java » ou « Client-Serveur ».

La valorisation des données

Elle se fait grâce au stockage et à l'organisation sur site Internet (avec ou sans mot de passe) et/ou sur CD-ROM, avec des traitements avec interfaces peu coûteuses et faciles à utiliser.

Accès aux données

L'accès aux données est obtenu par une application stricte de la politique de mise à disposition des données définie par le responsable scientifique, ainsi que par une documentation la plus informative possible sur les critères de qualité des données (modes d'acquisition et/ou de traitement...).

La pérennisation est acquise par une souplesse dans le redéploiement des modes de diffusion des données défini par le responsable scientifique, et par une utilisation des IHM (interfaces homme-machine) les plus conviviaux.

Sécurisation

Les systèmes sont sécurisés par des serveurs en redondance, par la pérennité des équipes, par une expérience importante dans le domaine des bases de données permettant de garantir un suivi pour celles qui auront été constituées, enfin par une intégration possible dans des bases « multiproxy ».

Conclusion

La constitution de bases de données dans le cadre d'expériences de collectes de données permet d'en assurer l'accès, et de garantir ainsi que celles-ci seront exploitées au mieux.



Les données scientifiques

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

Éditeurs scientifiques Jean-Michel Kornprobst, Marcel Raffy

Sommaire

Introduction

Marcel Raffy, professeur, université de Strasbourg

Jean-Michel Kornprobst, professeur, université de Nantes

Jean-Pierre Muller, directeur général de l'IRD

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

Coordinateur : Francis Laloë, IRD, Montpellier

Les données : expérience, observation et traitement
Francis Laloë

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

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

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

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

Coordinateur, François Le Verge, Ifremer, Brest

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

Les aspects techniques de la pérennité des données scientifiques

Claude Huc, Danièle Boucon

Video and graphic broadcasting information system for research vessels

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

Fabrice Lecornu, Armel Rué, Didier Lavoine

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

Catherine Maillard

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

Pascal Dallot

2^e partie : gestion et valorisation des données

Coordinateur : Jean-Michel Kornprobst

Diffusion des données géographiques : valorisation et aspects juridiques

Pierre Peltre

Le partage et la diffusion des données et résultats scientifiques

Dominique Vuillaume



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

Françoise Souyri



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

Philippe Waldteufel



Conclusion des débats et synthèse

Marcel Raffy



La gestion informatique des chroniques en hydrologie

Michel Lang



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

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



3^e partie : aspects juridiques et stratégiques

Coordinateur : Patrick Séchet, IRD, Paris

Les chercheurs peuvent-ils continuer à ignorer le droit ?

Patrick Séchet



Aspects juridiques de la diffusion des données scientifiques

Sébastien Lafargue



Diffusion des données de l'INPI

Bernard Marx



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

Jeanne Bossi



Adresse des auteurs

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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