

Bases de données pour les géosciences

Un effort de connaissance et de prospective

Philippe Waldteufel

Initiative CNES + INSU

La lettre de mandat (fin 98) :

- La gestion des données est un **enjeu important** en géosciences,
 - pour les institutions (mission, coût),
 - pour les chercheurs
 - **Tout évolue** : enjeux (ex. global change), sources de données, projets, techniques ...
- ➔ Proposer des axes pour un **schéma directeur** à l'échelle de 5 à 10 ans

Commentaires

Le coût consolidé de la gestion des données scientifiques en France est mal connu ; l'ordre de grandeur est de plusieurs centaines de MF par an.

S'agissant des données spatiales, les projets sont foisonnants et hétérogènes, et aboutissent à des solutions au coup par coup, s'agissant tant des choix que de la façon dont ils sont construits. Semble manquer une doctrine.

Groupe de travail :

- **Quatre chercheurs** : Jean-Claude ANDRÉ (Cerfacs), Annick BRICAUD (Villefranche), Pierre BRIOL (IPGP), Philippe WALDTEUFEL (IPSL).
- **Avec** : Cathy BOONNE (IPSL), Alain GABORIAU (Cnes), Claude HUC (Cnes).
- **Puis** : Jean-Pierre ANTIKIDIS (Cnes), Paul KOPP (Cnes), Jean-Paul REBERT (IRD)
- Echéance : en principe avant l'été 99

Commentaires

Le groupe était supposé couvrir l'ensemble des géosciences : atmosphère, océan, cryosphère, terre solide, sciences de l'environnement.

Dans les faits, il n'a pas été possible de traiter à fond les sciences de la Terre solide.

Le calendrier était optimiste : le document final a été rendu en novembre 1999.

Détail du mandat :

- Inventaire national de l'existant
- Inventaire international sommaire
- Inventorier la demande non satisfaite des utilisateurs (horizon 10 ans)
- Évaluer les axes politiques des organismes nationaux concernés
- Discuter les questions de politique de distribution
- Proposer des schémas de collaboration interorganismes
- Discuter la problématique de centralisation

Remarques

- Beaucoup de travail !
- L'enjeu central : **comment optimiser l'utilisation des ressources nationales consacrées à la gestion des données ?**
- ... en tenant compte de la **diversité** (acteurs, nature, niveaux des données)

Commentaires

Le simple énoncé des termes du mandat fait apparaître qu'il n'est pas seulement demandé au groupe de réfléchir ! La réflexion doit se baser sur des éléments d'information qui sont *a priori* peu ou mal connus, et qu'il faut aller chercher et rassembler.

Méthodologie de départ : **collecte d'informations**

1. Les (15) institutions nationales : **entretiens** (selon un fil directeur) : BRGM, CEA, CLS, Cnes, CNRS/Insu, Ifen, IFP, Ifremer, IGN, Inra, IRD, Médias-France, Météofrance, Shom, Spot.
2. Les bases : **inventaire sommaire** (sources d'info : WEB, contacts, doc. ; cas part. ESA)
 - (qu'est-ce qu'une base ? L'institution ? Le serveur ?)
 - Lesquelles choisir ? —> cf (3).
 - Au total une trentaine (selon un fil directeur)
3. Les utilisateurs : comment connaître leurs besoins sans leur demander ?
 - ➔ **Enquête** sur WEB
 - **Qui** sont-ils ?
 - Quelles **bases** utilisent-ils ? Comment les **apprécient-ils** ?
 - **Besoins**, problèmes présents et futurs

Commentaires

Institutions : ont des activités multiples en rapport avec les données, mais aussi parfois des vocations dominantes : collecte, conservation, utilisation, ou encore parfois intervention dans la constitution et la gestion des bases.

Utilisateurs : il est intéressant de les positionner (formation de recherche, thématique, situation, investissement personnel éventuel dans des bases) .

Leur opinion est déterminante pour piloter les choix des efforts à accomplir.

La constitution de l'inventaire impose une classification sommaire des types de données (incluant les données produites par simulation numérique), ainsi qu'une classification de types de bases dans laquelle interviennent principalement l'extension du cercle d'utilisateurs, l'articulation entre production et distribution, la durée de conservation et de mise à disposition des données.

Statut au 04/05

Entretiens avec institutions : 14/15

- Très **différentes** quant à l'équilibre des missions : fournisseur, gestionnaire, utilisateur, intervenant intermédiaire
- Ouverture fréquente à des initiatives coordonnées
- La filière de mise en place d'une base et ses métiers..

Inventaire : en cours ‡ 15/06

Enquête : close 15 Mars ; 138 réponses

(42 formations de recherche)

- Délais **d'identification** des bases
- Tropisme imprévu pour **réponses collectives** (198 déclarés pour 138)
- Questions pas optimisées ; formulaire trop long...

Commentaires

Le panorama des institutions met en évidence le questionnement des institutions à dominante utilisatrice (Insu du Cnrs, IRD, Inra) qui collectent néanmoins aussi des données (campagnes de terrain, séries d'observatoires).

Est également identifié le problème spécifique des données spatiales pour lesquelles, schématiquement, l'agence nationale (le Cnes) hésite entre une interprétation de sa mission restreinte au programme spatial national et un investissement dans les autres données (issues de missions européennes ou étrangères) qui intéressent la communauté française. En outre, dans tous les cas, l'implication des chercheurs dans la constitution des bases est indispensable.

L'enquête auprès des utilisateurs fait ressortir quelques tendances dominantes. Cependant les appréciations sur les bases ne fournissent guère de discrimination, et la faiblesse numérique de l'échantillon fragilise sa représentativité.

Ensuite...

Quelques enjeux mis en évidence

- **s'adapter à la diversité** des données, donc des problèmes
 - Équilibrer le poids du spatial (et de ses outils)
 - Les campagnes
 - Les données de simulation (en plus)
- Que faire **en France** (sur 343 bases citées : 50 % françaises) ?
- Selon **quels schémas** (institution, pôle, labo...) ?
- Quels **moyens collectifs** (compétences, outils, information, intervenants, visibilité) ? internes ? externes ?
- Que faire à l'échelle de la **formation de recherche** ?

Conclusions

- « Ils » feront mieux la prochaine fois !
- Mais on va faire ; rendez-vous à l'automne sur le WEB

Commentaires

Le document final mettra l'accent sur une vision des données à usage scientifique comme un patrimoine collectif, avec les conséquences déontologiques qui en résultent.

Le tour d'horizon fait ressortir quelques propositions simples sur des questions spécifiques (par ex. localisation des bases) ; il conduit à prôner une concertation interorganismes, ainsi qu'une coordination technique entre gestionnaires de bases.



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