

## Introduction

Ce séminaire est le deuxième séminaire de prospective organisé par la Commission scientifique des sciences de l'Ingénieur et de la Communication.

Il a été organisé par Jean-Michel Kornprobst, et par François Le Verge qu'il convient de saluer particulièrement.

Quels sont les objectifs de ces deux journées ?

Tout travail scientifique (à l'exception peut-être des mathématiques) repose sur des données mesurées. Elles peuvent être de toute nature. Elles sont des mesures physiques ou des observations sur l'Homme, sa santé, son comportement. Donc tout institut scientifique est un organisme dont le travail repose fondamentalement sur des données, de toute nature, acquises par ses membres, transmises, transformées. C'est d'ailleurs la plus-value sur cette transformation qui est le but et la raison d'être de ces instituts.

D'autre part, les possibilités d'acquisition, stockage, transfert et analyse des données vivent une période révolutionnaire. En effet, depuis quelques années, la possibilité de numériser tout type de données ouvre la porte à des possibilités totalement inédites et impensables il y a vingt ans seulement. La numérisation ouvre la porte à des possibilités inédites de stockage, avec le grand avantage d'une normalisation universelle. Elle ouvre la porte au transfert immédiat en tout point du globe. Enfin le traitement de quantités immenses de données qu'il s'agisse de sélection, filtrage, ou d'analyse, est aujourd'hui réalisable.

L'ensemble des instituts de recherche, parmi bien d'autres organismes, vit donc une révolution à laquelle leur raison d'être tient de très près. L'extraordinaire gain de temps que les techniques de numérisation apportent donne des potentialités de production scientifique considérablement accrues par rapport à ce qu'elles pouvaient être il y a vingt ans. Cela n'est pas sans mettre en jeu, de façon plus crue, le rôle économique des données. Car, si les instituts sont des producteurs et consommateurs de données, le coût de ceux-ci permet d'évaluer globalement le coût des données acquises et produites. Cela pose corrélativement des questions juridiques associées au transfert de données.

Enfin, le rôle de la Commission scientifique est directement concerné, s'il doit y avoir une action à la hauteur des changements techniques, cela n'ira pas sans restructurations et examen de carrières adaptées. Il convient également d'envisager ces aspects.

Telles sont les questions que nous nous proposons de débattre durant ces deux journées, avec des représentants d'organismes scientifiques concernés par la même problématique. Nous les remercions d'être venus parmi nous.

Bon travail !

**Marcel Raffy**

Président de la CS7

Cela me fait très plaisir de vous voir aussi nombreux ce matin et je vous remercie tous bien sincèrement d'être venus et pour certains d'entre vous, de très loin.

L'organisation de ce séminaire sur les données scientifiques a été une belle histoire, mais comme toutes les belles histoires, elle n'a pas été le fait d'un seul homme, évidemment, mais est le résultat d'un travail fait en commun, dans un même but, mené par un groupe de personnes motivées et déterminées et que je tiens à remercier tout particulièrement.

D'abord ce séminaire est dû à une initiative du président de notre Commission scientifique 7, Marcel Raffy, professeur à l'université de Strasbourg et c'est donc lui que je tiens à remercier en premier. Ensuite, s'il m'est incombé l'organisation générale de ce séminaire, celle-ci a, en réalité, été menée à bien grâce à un comité de cinq personnes, Marcel Raffy et moi-même, François Le Verge, qui est responsable de la documentation à l'Ifremer, membre nommé de la Commission scientifique 7 et qui animera la séance de demain matin. À ces trois membres extérieurs à l'IRD, le comité d'organisation s'est adjoint deux membres de l'IRD, Monsieur Francis LALOË, bio-statisticien au centre IRD de Montpellier et Monsieur Patrick SÉCHET, directeur du service des Assemblées et des Indicateurs scientifiques au siège de l'IRD, rue La Fayette. Par ailleurs, il n'y a pas de comité efficace sans secrétariat efficace et celui-ci a été mené avec une grande compétence par trois personnes de l'IRD que je citerai par ordre alphabétique : Brigitte ALEXANDRE, Catherine FONTAINE et Marie-Hélène PERROT.

Les paroles s'envolent, les écrits restent, c'est bien connu ! Pour toute manifestation, qu'elle soit scientifique, culturelle ou d'une autre nature encore, il est essentiel d'en assurer la pérennité par un document écrit de type classique, c'est à dire imprimé, ou par un document de nature contemporaine, c'est à dire électronique. Vous le savez, c'est la forme d'édition que nous avons choisi – un CD-ROM et je tiens à remercier ici Marie-Christine PASCAL, secrétaire du laboratoire Halieutique et Ecosystèmes aquatiques au centre IRD de Montpellier, qui assuré tout le travail de prépresse, ainsi que Renée Lecouffe et Marie-Odile Schnepf du service des éditions de l'IRD à Bondy pour la mise en forme finale du CD-ROM.

Enfin, et j'y tiens beaucoup, il me reste à remercier le directeur du centre IRD de l'Ile de France, Monsieur Jean-François Turenne, qui nous accueille ici chez lui et, *last but not the least*, Monsieur Jean-Pierre Muller, directeur général de l'IRD qui honore ce séminaire de sa présence. Sachant combien votre temps est compté, Monsieur le directeur général, je vous suis très reconnaissant d'avoir accepté d'ouvrir ce séminaire sur les données scientifiques.

**Jean-Michel Kornprobst**

Professeur à l'université de Nantes

Je suis heureux de participer à l'ouverture de ce séminaire sur les bases de données. Je félicite et remercie la commission Sciences de l'Ingénieur et de la Communication de l'IRD pour cette heureuse initiative.

Oui, l'Institut de recherche pour le développement est concerné au premier chef par votre réflexion. En effet, l'IRD, en temps qu'institution de recherche, utilise et alimente des bases de données, des systèmes d'information géographiques et cartographiques, des réseaux de communication. Il se sert de procédures de modélisation et de simulation et contribue à les faire évoluer. En bref, il est fortement engagé dans tous les processus modernes d'acquisition, de gestion et de transfert d'informations de tous ordres, analytiques et synthétiques.

L'ensemble de ces dispositifs va, à coup sûr, continuer à se modifier de façon très rapide au cours des années à venir, et l'IRD entend bien participer de façon active à ces mutations. Sa situation particulière d'institut aux implantations réparties sur toute la surface de la planète, au demeurant, l'y contraint, sauf à perdre rapidement pied dans l'exercice même de ses responsabilités fonctionnelles.

Plus spécifiquement, l'IRD envisage, au cours des quatre années à venir, de développer sa présence et son action sur deux axes privilégiés :

- 1) faciliter l'accès des pays du Sud et des collectivités territoriales françaises d'outre-mer aux richesses de ses fonds documentaires ;
- 2) utiliser délibérément les nouvelles technologies de l'information pour favoriser l'accès en ligne des grandes banques de données et de ressources documentaires à ses équipes de recherche et, plus généralement à tous ses partenaires du Sud, cela afin d'améliorer l'efficacité de la recherche et de l'expertise.

Ce faisant, l'IRD - comme la plupart des autres organismes - est et sera inévitablement confronté à plusieurs types de problème. Je ne ferai ici qu'ouvrir une liste - non exhaustive - des questions qui me viennent à l'esprit :

- comment sauver, archiver, valoriser les données acquises depuis des dizaines d'années et engrangées sous des formes extrêmement diverses, tels que des tableaux de mesures hydrologiques tapés à la machine à écrire ou stockées sur des fiches mécanométriques ou de nombreuses cartes pédologiques, souvent coloriées à la main dans les années 50 ? Il s'agit là d'autant de pièces - passez-moi l'expression - d'un « trésor inestimable sur lequel l'IRD est assis ».
- Comment standardiser les techniques d'acquisition et d'archivage des données acquises aujourd'hui et demain ?

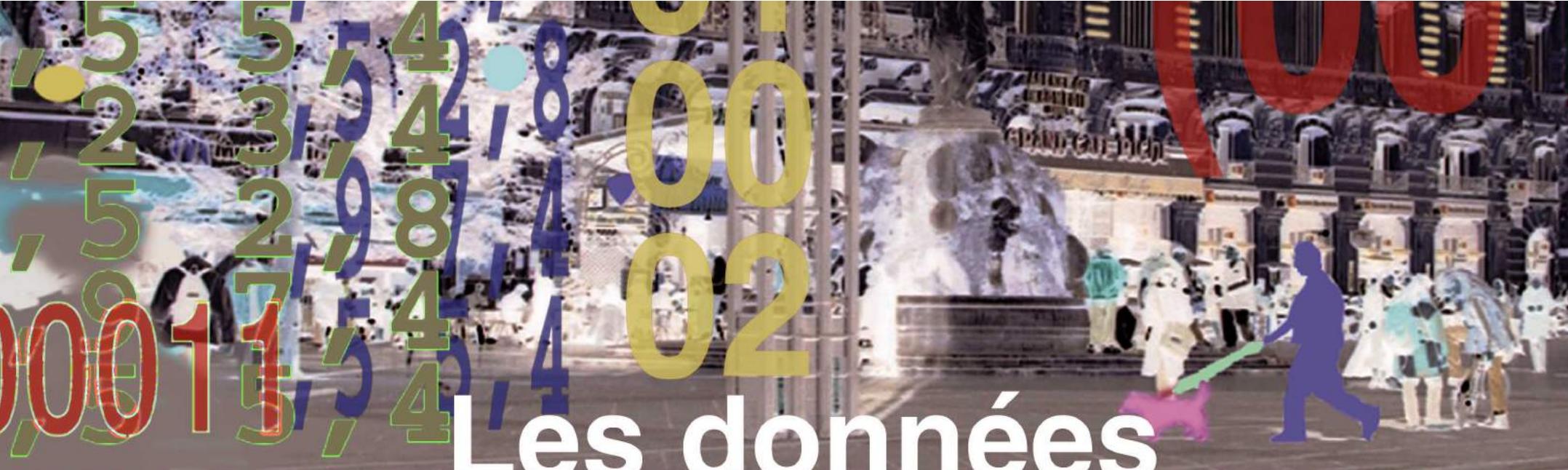
- Comment assurer la communication, l'échange, la protection, la vente de toutes ces données ? S'agissant *in fine* de rendre toutes ces données accessibles en réseau, l'IRD n'a aucun intérêt à s'attaquer seul à ces problèmes. D'ailleurs nous soutenons d'ores et déjà activement la mise en place de banques de données inter-institutionnelles telles Médias France pour les données océanographiques et climatiques et, j'espère bientôt, hydrologiques.

Je ne peux donc que saluer l'initiative prise par la commission « Sciences de l'Ingénieur et de la Communication » de l'IRD d'organiser un séminaire consacré à l'acquisition, la gestion, l'exploitation et la diffusion des données scientifiques, et qui devrait débattre aussi des questions juridiques liées à la propriété et à l'utilisation par des tiers de données obtenues en propre ou en partenariat.

Je sais gré aux organisateurs de ce séminaire - qui suit opportunément un séminaire organisé en novembre dernier par la même commission et qui, était consacré aux nouvelles technologies de l'information et de la communication - d'avoir su attirer des spécialistes de presque tous les EPST et EPIC pour réfléchir à ces questions. Je me réjouis enfin que des mots clés tels que « qualité » et « partage » auxquels la direction de l'IRD est particulièrement sensible, figurent dans l'intitulé de ce séminaire. La direction de l'IRD sera donc attentive aux recommandations que vous pourriez lui adresser à l'issue de ce séminaire de deux jours.

Tout en remerciant les organisateurs, MM. M. Raffy, J.-Michel Kornprobst, J.-François Turenne et leurs collaborateurs, et en souhaitant la bienvenue à nos hôtes des autres institutions, je déclare « ouvert » le séminaire sur les données scientifiques : base de progrès des connaissances - quelle qualité ? quel partage ? quel avenir ? en formulant pour ce séminaire les vœux de réussite ... qu'il mérite.

**Jean-Pierre Muller**  
Directeur général de l'IRD



# 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<sup>re</sup> 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<sup>e</sup> 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<sup>e</sup> 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