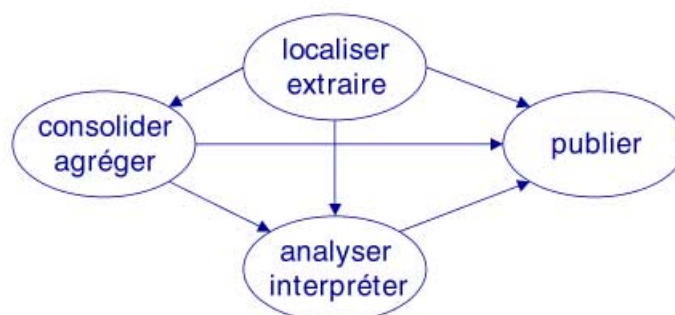


# Recherche d'information dans un réseau de sources de données scientifiques hétérogènes et autonomes

Éric Simon

## Contexte et besoins

- Réseau de sources d'information hétérogènes et autonomes : fournisseurs, consommateurs, intermédiaires



## Cas du projet européen Thetis

---

- Réseau de sources d'information :
  - données hétérogènes (biologie marine, océanographie, etc.)
  - programmes de traitement d'images, de modélisation, etc.
- Exemple :

Pb : obtenir une simulation du transport d'un certain polluant sur une zone et une date précises



## Solution 1 : intégration réelle d'information

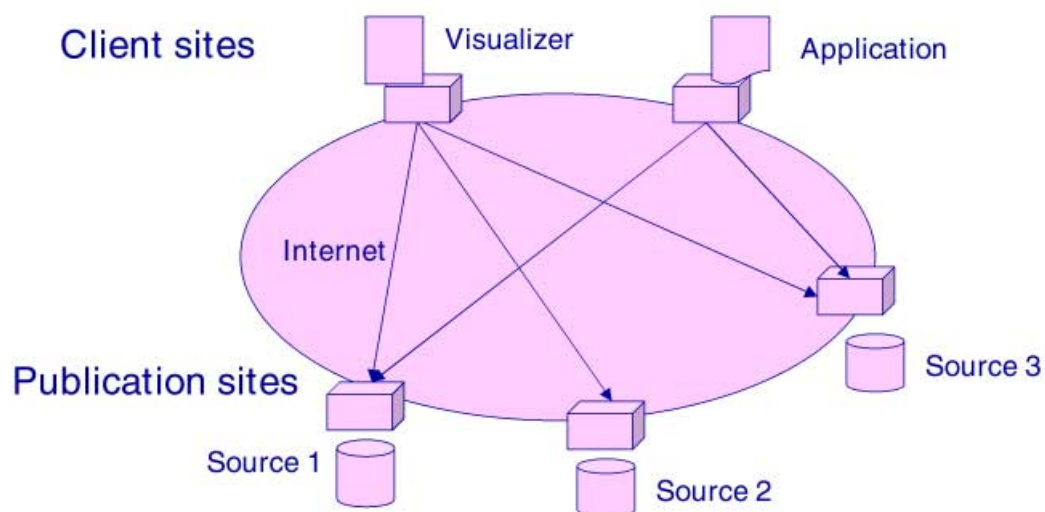
---

- Création d'un entrepôt d'information :
  - dupliquer, consolider et agréger données sur un site
  - installer programmes nécessaires sur ce site
- Avantages :
  - source d'information intégrée uniforme et cohérente
  - service d'accès indépendant
- Difficultés :
  - résoudre les incohérences entre données
  - maintenir à jour l'entrepôt

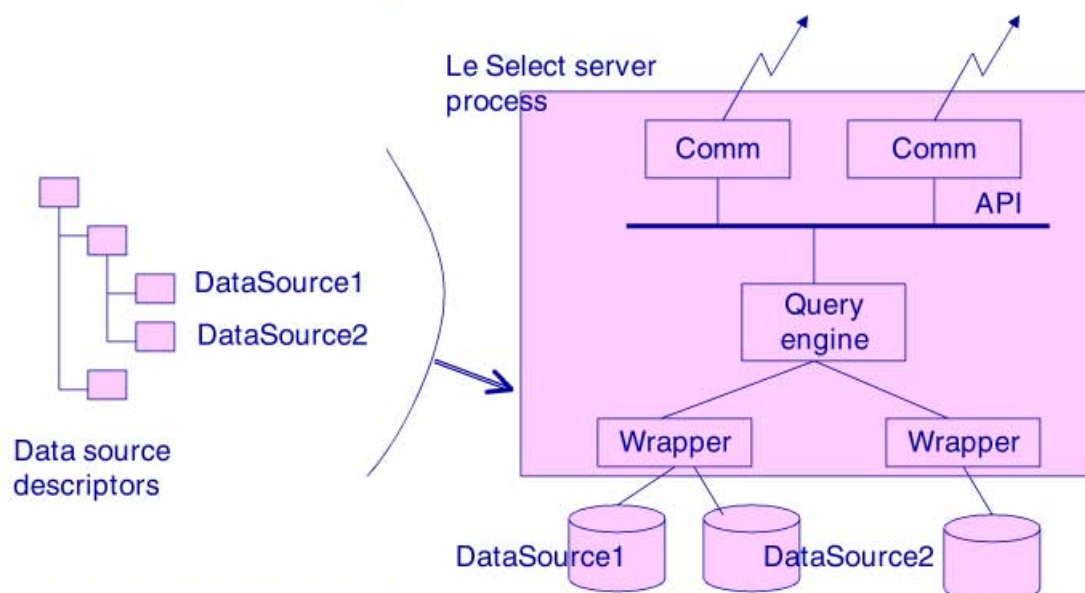
## Solution 2 : intégration virtuelle

- **Création d'une « vue » uniforme de l'information**
  - traduction des données et programmes dans un format commun
  - traitement transparent des accès à la vue
- **Avantages :**
  - respect de l'autonomie des sources
  - pas de maintien à jour
  - facilité de mise en œuvre
- **Inconvénients :**
  - risques d'incohérence potentielle

## Le Select : système d'intégration virtuelle



## Site de publication



## Traduction des sources

- Données : modèle relationnel (tables)
- Programmes : fonctions ou tables
- Traduction s'effectue via des adaptateurs (« wrappers »):
  - accède données dans leur forme d'origine
  - traduit données en tables
  - offre interface standardisée aux données
- Les traducteurs sont écrits par ceux qui publient

## Exemple de traductions

### ■ Source 1 :

émissions	(ident,	date,	zone,	description)
em100	10-04-1998	(x1,y1)(x2,y2)	em100.gif	
em101	10-04-1998	(x1,y2)(x2,y3)	em101.gif	

select on date, zone

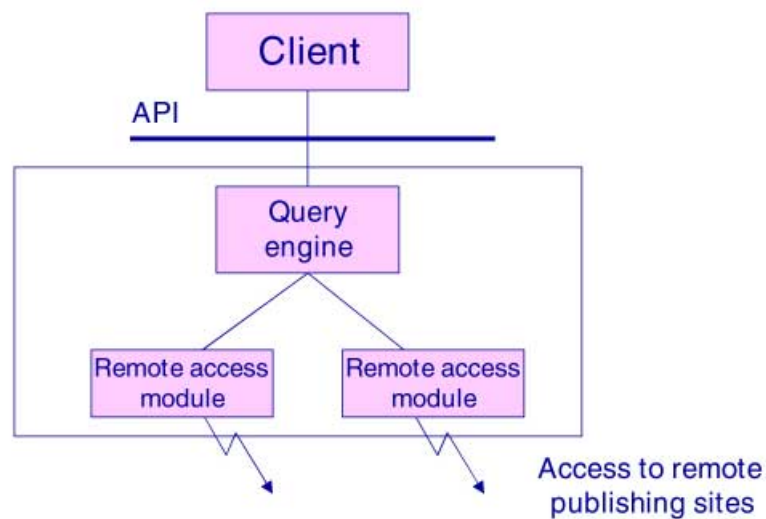
### ■ Source 2 :

courants (ident, date, zone, description)  
any select

### ■ Source 3 :

waste-transp (courant, polluant, sequence)  
input: courant, polluant; output: sequence

## Architecture d'un site client



## Accès aux données

---

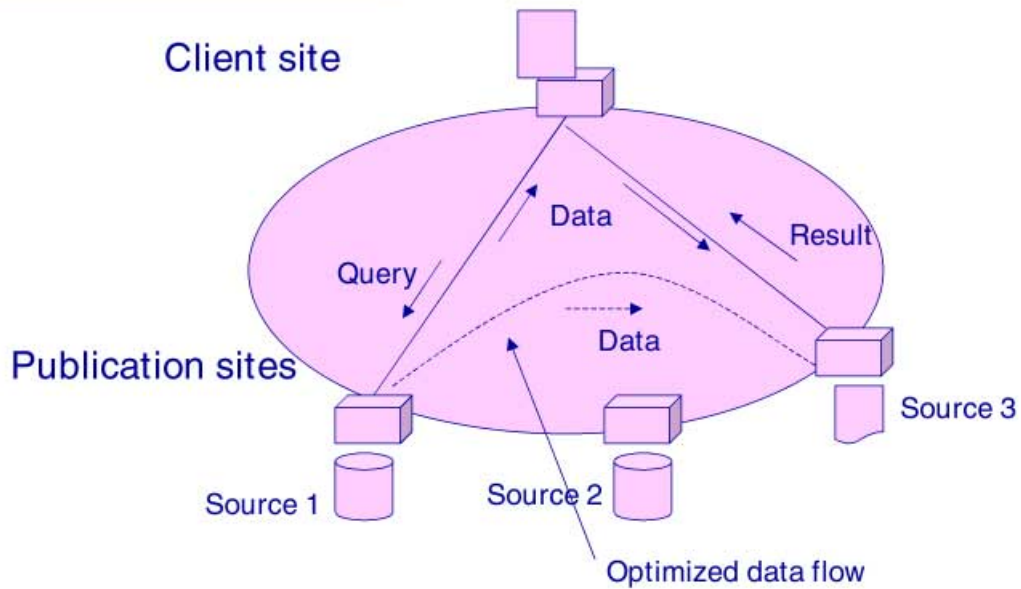
- Requetes de type SQL :
  - requête est décomposée en sous-requetes locales (query engine)
  - sous-requetes retournent données (via wrapper)
  - requête est exécutée (query engine)
- Accès au query engine
  - JDBC,
  - C/C++ via Corba, RMI, ftp

## Exemple de requête

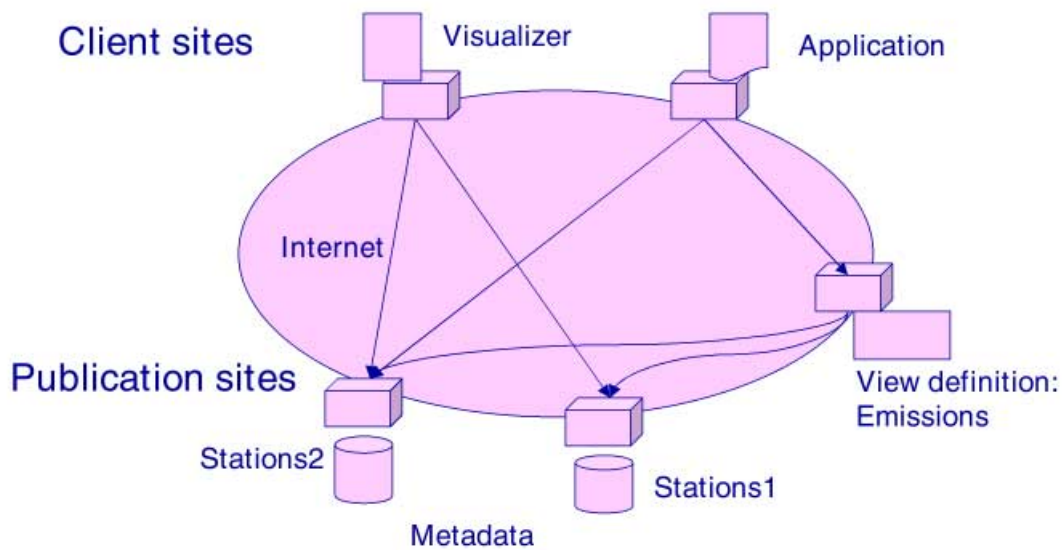
---

```
Select s3.sequence
from  leselect://www.source1.fr/wrapper/emissions s1
      leselect://www.source2.gr/wrapper/courants s2
      leselect://www.source3.uk/wrapper/waste-transp s3
where s1.date = s2.date = ' 10-04-1998 ' and
      same (s1.zone, s2.zone) and
      s3.polluant = s1.description and
      s3.courant = s2.description
```

## Exécution des programmes



## Données dérivées



## Méta-données

---

- “Documents” attachés aux données
  - méta-information sur les données
  - format simple (attribut/valeur) ou élaboré (XML)
- Définition des documents
  - fichiers de description des wrappers
  - tables particulières
- Accès aux sources via documents
  - “search engines”
  - requêtes

## Projet Caravel à l’Inria

---

- Conçoit et expérimente des systèmes de médiation d’information
  - médiateurs d’accès (par ex. Le Select)
  - médiateurs d’intégration
  - systèmes de gestion de catalogues sur le Web
- Domaine d’application privilégié sur l’environnement



# 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 Sismar, 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