

# *Projet VERSO*

*Bases de Données*

*Rocquencourt*

THÈME 3A



*R*apport  
*d'Activité*

1999



## Table des matières

<b>1</b>	<b>Composition de l'équipe</b>	<b>3</b>
<b>2</b>	<b>Présentation et objectifs généraux</b>	<b>4</b>
<b>3</b>	<b>Fondements scientifiques</b>	<b>5</b>
3.1	Bases de données et le Web . . . . .	5
3.2	Données multidimensionnelles et complexes . . . . .	6
3.3	Vues Actives pour le commerce électronique . . . . .	7
<b>4</b>	<b>Domaines d'applications</b>	<b>7</b>
4.1	Xyleme : exploitation des données du Web . . . . .	8
4.2	Intranets et mémoire d'entreprise . . . . .	8
4.3	Applications culturelles . . . . .	9
4.4	Applications multidimensionnelles . . . . .	9
<b>5</b>	<b>Résultats nouveaux</b>	<b>10</b>
5.1	Bases de Données et le Web . . . . .	10
5.1.1	Données semi-structurées . . . . .	10
5.1.2	Documents virtuels . . . . .	11
5.1.3	Médiation sémantique de données . . . . .	11
5.1.4	Xyleme : stockage et interrogation de documents XML . . . . .	12
5.2	Données multidimensionnelles et statistiques . . . . .	12
5.2.1	Bases de données avec contraintes . . . . .	13
5.2.2	DEDALE . . . . .	13
5.2.3	Représentation de la topologie et requêtes topologiques . . . . .	14
5.2.4	Données agrégées . . . . .	14
5.3	Vues actives pour le commerce électronique . . . . .	15
<b>6</b>	<b>Contrats industriels (nationaux, européens et internationaux)</b>	<b>15</b>
<b>7</b>	<b>Actions régionales, nationales et internationales</b>	<b>16</b>
7.1	Actions nationales . . . . .	16
7.1.1	Projet RNRT GAEL . . . . .	16
7.1.2	Autres Collaborations Nationales . . . . .	16
7.2	Actions financées par la commission européenne . . . . .	16
7.2.1	Projet C-WEB . . . . .	16
7.2.2	Projet TMR Chorochronos . . . . .	17
7.2.3	GDR-PSIG Cassini . . . . .	17
7.3	Réseaux et groupes de travail internationaux . . . . .	17
7.4	Relations bilatérales internationales . . . . .	18
7.4.1	Europe . . . . .	18
7.4.2	Moyen-Orient . . . . .	18
7.4.3	Amérique du Nord . . . . .	18

7.4.4	Asie et océan Pacifique . . . . .	18
7.5	Accueil de chercheurs étrangers . . . . .	19
<b>8</b>	<b>Diffusion de résultats</b>	<b>19</b>
8.1	Actions d'enseignement . . . . .	19
8.2	Participation à des colloques . . . . .	20
8.2.1	Conférences invitées, tutoriels, cours, etc. . . . .	21
8.2.2	Animations scientifiques . . . . .	22
<b>9</b>	<b>Bibliographie</b>	<b>23</b>

---

## 1 Composition de l'équipe

### Responsable scientifique

Serge Abiteboul [DR, Janv-Mai]

Sophie Cluet [DR, Juin-Déc]

### Assistante de projet

Danièle Moreau [AJA, en commun avec le projet MEVAL]

### Personnel Inria

Stéphane Grumbach [DR]

Anne-Marie Vercoustre [DR]

Luc Segoufin [CR, Sept-Dec]

### Conseillers scientifiques

Claude Delobel [Professeur, Univ. Paris 11]

Michel Scholl [Professeur, CNAM]

### Collaborateurs extérieurs

Bernd Amann [Maître de Conférence, CNAM]

Brendan Hills [CSIRO, Australie, Janv-Mai]

### Chercheurs invités

Leonid Libkin [Chercheur, Bells Labs-USA]

Victor Vianu [Professeur, UC San Diego, Juin-Dec]

### Ingénieurs experts

Guy Ferran [IE Inria, Sep-Déc]

### Chercheurs doctorants

Vincent Aguilera [Ingénieur Ministère Equipement, Ecole des Ponts]  
Irina Fundulaki [Boursière Inria, CNAM]  
Amélie Marian [Boursière MENRT, Paris-Dauphine, Oct-Déc]  
Laurent Mignet [Boursier MENRT, CNAM]  
Pierangelo Veltri [Boursier INRIA, U. Paris 11]  
Fanny Wattez [Boursière Cifre Ardent-Software (Jan-Nov), INRIA (Déc), Paris 1]

### Stagiaires

Sébastien Ailleret [Polytechnique, Avril-Sept]  
Frédéric Hubert [U. Caen, Avril-Sept]  
Sandrine Jacqmin [DEA Paris-Dauphine, Avril-Sept]  
Amélie Marian [DEA Paris-Dauphine, Avril-Sept]  
Bruno Tessier [Polytechnique, Avril-Sept]

## 2 Présentation et objectifs généraux

Les données sont de plus en plus complexes, distribuées, hétérogènes, répliquées, multi-formes, changeantes. L'objectif du projet est l'étude des problèmes fondamentaux posés aux systèmes de gestion de bases de données existants et le développement de solutions novatrices appropriées. Notre but est d'obtenir des systèmes plus ouverts à des données plus riches (documents, cartes), ouverts vers le réseau. Suivant une tradition bien établie, nous étudions les problèmes sur le plan théorique et sur le plan pratique, dans un effort permanent d'adapter nos résultats fondamentaux aux réalités du monde industriel avec lequel nous avons des partenariats.

L'émergence de XML comme format d'échange standard pour Internet ouvre des perspectives particulièrement intéressantes pour l'utilisation des données du WWW. En effet, à l'opposé de HTML, XML apporte de l'information sur la structure des documents et, par là même, sur leur sémantique. Il nous paraît primordial de mettre tout en œuvre pour exploiter cette nouvelle propriété des données du Web. Nous voulons notamment permettre une recherche plus intelligente sur le Web et une meilleure utilisabilité des informations trouvées. Les techniques que nous avons développées ces dernières années autour des données hétérogènes et semistructurées s'appliquent parfaitement à cette nouvelle préoccupation de l'équipe. Nous poursuivons également nos travaux sur les fondements théoriques des bases de données et nos activités sur les bases de données spatiales et le commerce électronique.

Notre action de recherche s'articule cette année autour de trois thèmes : (i) bases de données et le Web (ii) gestion de données multidimensionnelles, et (iii) systèmes de règles actives pour le commerce électronique.

## 3 Fondements scientifiques

### 3.1 Bases de données et le Web

**Mots clés** : XML, DTD, stockage, indexation, architecture, ontologie, versions, requête temporelle, notification, moteur de recherche, intégration, World Wide Web, schéma, document, hétérogénéité.

Malgré le formidable développement du Web et son utilité pour de nombreuses communautés, la recherche et l'exploitation d'informations s'y révèlent encore très difficiles.

Les causes des limitations actuelles sont multiples. Le langage HTML utilisé pour écrire les pages Web est très pauvre : il n'apporte aucune information structurelle ou sémantique. De ce fait, les moteurs de recherche fournissent des réponses trop larges, desquelles il est quasiment impossible d'extraire les données pertinentes. De plus, le protocole du Web n'est pas adapté à l'écriture simple d'applications qui requièrent l'accès à plusieurs sites.

Verso démarre une action de recherche, nommée Xyleme, dont le but est de tenter de résoudre ces problèmes. Cette action s'inscrit dans la continuité des nombreux travaux de l'équipe sur ce thème et s'appuie sur l'hypothèse que le langage XML deviendra le langage de définition et de description des données du Web. En levant les limitations de HTML, ce langage ouvre la voie à de nouveaux services sur le Web. Verso, par l'intermédiaire du projet Xyleme, explore et fait avancer ces nouveaux domaines d'application. Notamment, il devient possible d'envisager des langages de requêtes très riches qui dépassent de loin les interfaces de recherches par mots-clés que proposent les moteurs de recherche.

Cependant, d'autres problèmes demeurent : comment "intégrer" des données, certes XML mais hétérogènes, et comment évaluer efficacement des requêtes multi-sites? Xyleme y répond en proposant d'intégrer les données du Web en un seul endroit. Cet entrepôt de données devient alors la plate-forme idéale pour exécuter des applications très sophistiquées qu'il serait impossible d'envisager directement sur le Web, soit parce qu'elles seraient trop difficiles à écrire, soit parce qu'elles seraient trop lentes.

Ce projet ambitieux aborde plusieurs aspects. Afin d'en démontrer la faisabilité, il veut intégrer les résultats de recherche dans un prototype réaliste.

Le projet se divise en sous-projets.

#### – **Stockage**

Il s'agit d'étudier les modèles de stockage des données XML et des méta-données. L'indexation des données est un facteur déterminant pour l'efficacité des couches logicielles qui accèdent aux données.

#### – **Langage de requêtes**

Partant de notre expérience sur les langages OQL, POQL et Lorel et des travaux issus du W3C, il s'agit de définir un langage de requêtes complet dont l'algèbre inclut de manière efficace toutes les opérations nécessaires à la sélection d'arbres XML, leur filtrage et la construction de résultats dans le même modèle.

#### – **Evolutions temporelles**

Il s'agira de démontrer les plus-values apportées par Xyleme par rapport au Web actuel. De nouveaux services seront étudiés : gestion de versions de documents, requêtes temporelles, abonnement sur requête et notification.

– **Intégration sémantique**

La nature hétérogène des données implique un traitement sémantique de celles-ci afin de les intégrer ou de les classer. On explorera les possibilités qu'offrent les ontologies dans ce domaine.

– **Acquisition de données**

La technologie des moteurs de recherche sera adaptée pour rechercher des données XML sur le Web. Mais plus généralement, on se posera la question de la manière dont un serveur peut "exporter" des données XML.

– **Architecture**

Cette tâche transversale étudiera les problèmes spécifiques liés à la taille de l'entrepôt (de l'ordre du tera octet) : regroupement, partitionnement, multi-processeurs, réseau, parallélisme.

### 3.2 Données multidimensionnelles et complexes

**Mots clés** : multidimensionnel, contrainte, topologie, cartographie, spatio-temporel, agrégation, données statistiques.

Les données multidimensionnelles, comme les données spatiales et temporelles, posent des problèmes fondamentaux aux systèmes de gestion de bases de données, non seulement à cause de la taille considérable et en croissance constante de ces données, mais aussi à cause de la complexité des relations, en particulier, géométriques et topologiques qu'entretiennent ces données entre elles.

Nous nous sommes intéressés à la définition de modèles de données et de langages de requêtes de haut niveau d'abstraction pour le temps, l'espace et plus généralement les données liées à l'espace et au temps, comme les données géographiques (*e.g.* précipitation, nappe phréatique, pollution, etc.) qui sont en général définies par interpolation sur une base finie.

Le cas du temps seul diffère fortement de celui de l'espace, puisqu'il bénéficie d'une longue tradition de formalisation en logique. Pour les données spatiales, nous étudions un modèle de données qui permet de représenter explicitement des objets spatiaux, tels que les polygones, comme l'ensemble infini des points qu'ils contiennent. Ce modèle est basé sur l'approche par contraintes, dans laquelle un objet géométrique est représenté par une formule de la logique du premier ordre. Dans les systèmes d'information spatiale, il n'y a pas de couche logique qui permette à l'utilisateur d'avoir une vue des données indépendante de la représentation physique. Le modèle contraintes permet précisément de définir cette couche logique. Il permet aussi d'unifier le mode de représentation des données spatiales et des autres données alphanumériques. Dans les systèmes d'information spatiale, ces deux types de données sont manipulés par des systèmes différents. Enfin, il permet une modélisation de l'interpolation linéaire des données transparentes pour l'utilisateur.

L'effort principal du projet dans ce domaine est consacré depuis quelques années à la réalisation, en collaboration avec le CNAM, d'un prototype, DEDALE, dont la première version est implémentée sur le SGBD O<sub>2</sub>. Le modèle de données est celui des bases de données avec contraintes linéaires sur les rationnels, ce qui suffit à la modélisation des données. Une deuxième version du prototype est en cours de développement sur un gestionnaire d'objets développé au CNAM.

DEDALE permet aujourd'hui d'écrire des requêtes spatiales complexes qui ne sont pas toujours exprimables dans les systèmes existants qui ne gèrent pas les objets de dimension supérieure à 2. Le défi actuel est de réaliser l'optimisation de ces requêtes qui sont exprimées indépendamment de leur stratégie de calcul. On propose, pour cela, des méthodes basées sur des critères tels que la dimension, la géométrie et la topologie des données.

Par ailleurs, nos efforts ont porté sur un autre type de données, les données agrégées, qui peuvent être modélisées dans un formalisme logique avec arithmétique réelle et des opérateurs d'agrégation, comme le comptage et la somme ou le produit itéré.

### 3.3 Vues Actives pour le commerce électronique

**Mots clés :** règle de production, base de données active, parallélisme, déduction, datalog, workflow, commerce électronique, calcul relationnel, non-déterminisme.

Il s'agit de mieux comprendre des applications mettant en jeu du partage de données entre des clients qui interagissent via un système de "vues actives". Des vues actives permettent à un client de voir les données qui le concernent et d'interagir avec le serveur et d'autres clients par un mécanisme de notification. Les aspects novateurs de nos travaux s'articulent autour de (i) l'utilisation de données semiestructurées, (ii) l'étude de descriptions logiques des connaissances (celles du catalogue et des profils d'utilisateurs), et (iii) une spécification déclarative des notifications. (Voir Projet RNRT GAEL.)

Ces travaux se situent dans le cadre de la spécification déclarative d'applications distribuées.

## 4 Domaines d'applications

Les bases de données n'ont pas de champ d'application privilégié. En effet, toute application mettant en jeu une quantité importante de données ou d'information se doit d'utiliser des bases de données. Verso choisit de cibler des applications qui présentent des challenges particuliers. Cela est le cas pour les données spatiales et temporelles que l'on retrouve notamment dans des applications géographiques sur lesquelles nous travaillons (voir 7.2.2), ou pour le répertoire des documents XML du Web (voir 3.1), au-dessus duquel nous envisageons de fournir des services.

Nous nous contenterons de mentionner plus en détail quatre applications à titre d'illustration. La première porte sur l'exploitation intelligente des données du Web. La seconde concerne également le Web, mais dans une vision Intranet. Nous nous intéressons notamment à des sous-Web pour des communautés spécifiques (C-Web), en particulier pour gérer la mémoire d'entreprise. La troisième considère la gestion de données culturelles pour lesquelles nous voulons créer des médiateurs *intelligents* s'appuyant sur des descriptions sémantiques des sources

et du domaine (ontologies, thésauri). La quatrième, enfin, illustre les applications des bases de données multidimensionnelles.

Pour conclure cette excursion dans les domaines d'applications, nous noterons que Verso étudie aussi une application très à la mode : le commerce électronique. Certains aspects de cette application mêlent la gestion de données hétérogènes distribuées ou des notions de transactions, familières dans une problématique bases de données. D'autres aspects plus originaux mettent en jeu des spécifications déclaratives des flots d'information entre les acteurs de telles applications. (Voir 5.3 et Projet RNRT GAEL)

#### 4.1 Xyleme : exploitation des données du Web

**Mots clés** : internet, World Wide Web.

Depuis son adoption par le W3C comme format d'échange standard pour les données de l'Internet, XML a séduit tous les grands acteurs du marché informatique.

Contrairement aux données HTML, il est possible d'extraire la structure d'un document XML. Cette structure peut être utilisée de multiples façons. Notamment, elle permet d'interroger les documents de façon plus intelligente. Par exemple, il est concevable en XML d'évaluer une requête de type : "donner les documents contenant des produits dont le prix est supérieur à 10000 Euros". Sur un ensemble de documents HTML, la requête la plus proche de celle-ci serait : "donner les documents contenant les mots produits et prix". Il est également possible d'envisager l'intégration automatique d'un ensemble de documents portant sur le même sujet mais dont les structures sont légèrement différentes. Ceci facilite l'interrogation mais également la manipulation des données par des applications (par exemple, une application sur des données génomiques collectées en différents endroits de la planète).

Verso parie sur XML. Nous pensons que ce format va gagner encore en popularité et que dans un futur relativement proche, la majorité des données publiées sur le Web seront XML. Nous comptons être alors présent avec des outils permettant une bonne exploitation de ces données. Nous nous proposons plus particulièrement de construire une base de données géante intégrant toutes les données XML du Web. Cette base servira de support à un ensemble de services tels que : interrogation conviviale et pertinente, abonnement de requêtes (prévenez-moi si un nouveau projet apparaît sur le site de l'INRIA) ou requêtes temporelles (quelle est l'évolution du nombre de projets à l'INRIA depuis 1980).

#### 4.2 Intranets et mémoire d'entreprise

**Mots clés** : intranet, World Wide Web, communauté.

De plus en plus d'entreprises utilisent les technologies de l'Internet pour stocker, gérer et échanger leurs informations internes ; c'est ce qu'on appelle *l'intranet*. On peut considérer que, pour la plupart des sociétés, l'intranet contient déjà plus d'informations que leurs bases de données (c'est certainement le cas à l'Inria). Les intranets sont donc un instrument vital pour

les sociétés et il est important de savoir les gérer de façon sûre et efficace pour en faire une vraie *mémoire d'entreprise* réutilisable.

Le développement de technologies basées sur XML est une opportunité qui peut changer radicalement le Web lui-même et son utilisation dans les entreprises. Nous pensons que XML permettra l'échange, non plus seulement de documents, mais de données et de connaissances, en explicitant et en encodant l'information sous une forme flexible pour l'homme et utilisable par la machine.

Le projet C-Web vise à développer une plate-forme générique, basé sur des standards ouverts (en particulier XML et RDF), pour créer des échanges sur le Web adaptés et spécialisés pour une communauté ou une entreprise particulière (voir 7.2.1). Le système devra répondre à des requêtes portant sur des concepts du domaine, des structures argumentatives ou plus généralement des liens typés, et retourner des réponses précises assemblées dans un nouveau document.

### 4.3 Applications culturelles

**Mots clés :** culturel, musée, patrimoine, bibliothèque électronique.

Les applications culturelles concernent principalement la gestion de documents électroniques dans les domaines de l'art (e.g., musées), du patrimoine national (e.g., monuments), des bibliothèques, etc. Le but est d'ouvrir ces collections de documents aux chercheurs et au public le plus large. On retrouve les pôles d'intérêt de Verso : grande quantité d'informations multi-média (image, graphique, texte), distribuées géographiquement sur des supports (systèmes et formats) divers.

Verso s'appuie dans ce domaine sur sa maîtrise de la technologie objet. Dans le cadre du contrat Inventaire97 avec le ministère de la Culture, Verso travaille sur la navigation à l'aide de cartes géographiques (IGN, cadastre) pour accéder efficacement une base de documents du patrimoine national. Nous avons également commencé à travailler sur un médiateur intelligent qui s'appuie sur une description sémantique de ces collections et du domaine considéré (voir 7.2.1).

### 4.4 Applications multidimensionnelles

**Mots clés :** temps, espace, agrégation..

Les applications des bases de données impliquent de plus en plus souvent des données multidimensionnelles, comme les cartes (par exemple géographiques), les plans d'occupation des sols, les descriptions 3D du sous-sol pour la géologie et l'exploitation minière, les objets mobiles (comme le déplacement des voitures), ou encore pour les objets de forme variable (comme les fleuves). La modélisation de ces données et le développement de langages permettant d'extraire et de manipuler efficacement l'information stockée constituent un véritable défi au coeur des pôles d'intérêt de Verso : modélisation, langages de requête, optimisation.

Deux applications tournent aujourd'hui sur le prototype DEDALE [20]. La première, cartographique, porte sur une carte de la région d'Orange fournie par l'IGN et permet de répondre

à un ensemble de requêtes typiques des systèmes d'information géographique. La seconde application est multidimensionnelle et consiste en des objets mobiles dans une station de ski, sur des données fournies par le LAMA de Grenoble (voir 7.2.3). L'interrogation se fait via une interface graphique dont la nouvelle version a été entièrement écrite en Java afin de permettre son utilisation à distance. D'autres applications sont à l'étude avec des contraintes temporelles plus fortes.

Le projet TMR Chorochronos vise explicitement les applications spatio-temporelles. L'objectif de ce projet est le développement d'une architecture idéale pour la modélisation et l'interrogation des données multidimensionnelles (voir 7.2.2).

## 5 Résultats nouveaux

### 5.1 Bases de Données et le Web

**Participants :** Serge Abiteboul, Vincent Aguilera, Sébastien Ailleret, Bernd Amann, Sophie Cluet, Claude Delobel, Irini Fundulaki, Stéphane Grumbach, Sandrine Jacqmin, Amélie Marian, Laurent Mignet, Michel Scholl, Anne-Marie Vercoustre, Victor Vianu, Fanny Wattez.

**Mots clés :** hétérogénéité, intégration, World Wide Web, schéma, document, semi-structuré, SGML, XML, optimisation de requêtes.

**Glossaire :**

**Document virtuel** Document dont au moins certaines parties sont générées dynamiquement

**Résumé :** *L'importance croissante du Web comme source d'information rend crucial le traitement de données hétérogènes et leur intégration avec des bases de données. Depuis déjà plusieurs années, Verso travaille sur différents thèmes liés à cette problématique comme les données dites semi-structurées ou la réutilisation de données hétérogènes sous forme de documents virtuels. Cette année, nous abordons le problème sous deux nouveaux angles. D'une part, nous avons commencé plusieurs travaux sur la médiation intelligente pour des données culturelles et pour la mémoire d'entreprise. D'autre part, nous projetons d'intégrer toutes les données XML du Web dans une base unique: Xyleme.*

#### 5.1.1 Données semi-structurées

Les données accessibles sur le réseau vont du très structuré dans des bases relationnelles au totalement non structuré dans des fichiers de texte. De plus, l'intégration de données est aussi souvent source d'irrégularité, des données similaires étant souvent représentées avec des structures différentes dans des sources indépendantes. On utilise le terme *semi-structuré* pour ces données qui ne sont pas vraiment structurées mais qui présentent, pourtant, une certaine structure même si celle-ci est souvent peu régulière et implicite.

Nous avons publié un livre sur les données du Web en considérant des données depuis le très structuré jusqu'au semi-structuré et à XML [8]. Dans [29], nous considérons le problème général de la spécification de vues pour des données XML et l'évaluation de requêtes sur ces vues dans [42]. Dans [32], nous considérons l'utilisation de caches sémantiques pour optimiser les requêtes de type *attribut = valeur*, requêtes fréquentes dans le contexte semi-structuré. On notera aussi des travaux sur des expressions de chemin régulières sur ce type de données [13]. Dans [38], nous proposons le langage Ozone, une extension de OQL [43] et de Lorel permettant de manipuler à la fois des données structurées et semi-structurées. Enfin, l'évolution dans le temps de telles structures et une extension temporelle du langage de requêtes temporel Lorel sont étudiés dans [16].

Par ailleurs, de nombreux sites contiennent des données assez structurées mais dont la structure n'est pas directement accessible. C'est le cas, par exemple, des bases de données accessibles via le Web. Dans [33], on étudie sous quelles conditions la structure des données peut être reconstruite à partir de données sous forme HTML, par exemple.

### 5.1.2 Documents virtuels

Nous voyons les documents virtuels comme un moyen d'intégrer et de réutiliser des données venant des bases de données ou d'autres sources d'information semi-structurées, comme le Web, des documents SGML ou Latex. Ceci permet par exemple la génération de rapports ou de pages Web à partir de bases de données ou d'autres pages XML ou HTML avec une approche centrée "document".

Nous utilisons un modèle générique d'arbres étiquetés pour assurer l'intégration des données hétérogènes. Ce modèle est bien adapté à l'interrogation et l'intégration de documents XML.

Nous avons utilisé cette approche pour la génération d'un CD-ROM à partir de descriptions de photos exprimées en XML (métadonnées) [41]. Nous avons utilisé le modèle de méta données du Dublin Core (<http://purl.oclc.org/dc/>), étendues par des descriptions dans un schéma défini pour l'application considérée (habitat rural). Le CD-ROM peut être généré automatiquement en mode compilation (toutes les pages sont instanciées), ou en mode dynamique (les pages sont générées quand elles sont accédées).

Ce travail est réalisé en étroite collaboration avec le CSIRO-CMIS, à Melbourne (Australie), où nous avons également étudié l'application des documents virtuels à la synthèse de vidéo interactive [40].

### 5.1.3 Médiation sémantique de données

L'objectif de ce projet est d'offrir à l'utilisateur, spécialiste ou non, une interface lui permettant de chercher l'information en utilisant le vocabulaire de son domaine d'intérêt (par exemple des données culturelles). Un but du projet est la réalisation d'un médiateur entre des requêtes utilisateur exprimées en terme de concepts d'un domaine et plusieurs sources d'information indépendantes. Plus précisément, ce système doit être capable d'identifier les sources de données importantes pour la requête posée, traduire les sous-requêtes dans le langage de chaque base et intégrer les résultats (partiels) obtenus.

Dans un premier temps, nous nous intéressons à des sources de documents XML. Nous avons

étudié les possibilités d'utilisation de structures sémantiques (ontologies, thesauri) existantes pour la spécification de médiateurs spécialisés dans un domaine d'application (par exemple, un médiateur pour "la poterie Française du XVIII<sup>e</sup> siècle"). Nous avons défini un modèle d'intégration entre ontologies et thesauri [30] qui permet de générer un schéma RDF. Nous avons commencé un prototype basé sur ce travail qui intégrera une ontologie et des thesauri du service de l'Inventaire au Ministère de la Culture, pour la description de monuments et d'œuvres d'art, dans le cadre d'une collaboration avec le CNAM.

Nous sommes en train d'étudier différentes applications de ce modèle pour la spécification de médiateurs à partir de la description sémantique et structurelle des sources, dans le contexte des projets C-Web (voir 7.2.1) et Xyleme (voir 3.1).

#### 5.1.4 Xyleme : stockage et interrogation de documents XML

Nous avons commencé à étudier différents modèles de stockage pour XML :

- Le modèle dit structuré reflète la structure des documents, par un schéma correspondant à la DTD (Document Type Description). Cette approche permet, entre autres, une évaluation plus efficace des requêtes et la validation des documents.
- Le modèle semi-structuré est utile lorsqu'il n'y a pas de DTD ou si l'on veut une approche flexible permettant d'intégrer des documents variés. Il représente toutes les données de façon identique par un arbre DOM.
- Le modèle mixte permet de concilier le meilleur des deux mondes et est une approche encore peu étudiée.

Dans un premier temps, nous avons développé un chargeur flexible de documents XML permettant de stocker un document XML dans la base, en mode structuré, en mode semi-structuré, en mode hybride, ou en "blob" (sous-arbre conservé comme un texte XML). Le mode de chargement est spécifié par des annotations ajoutées à la DTD. Nous avons commencé à évaluer les différents modes de stockage sur un jeu de requêtes test. Le coût des requêtes, sur des documents relativement réguliers comme les articles d'un catalogue, est deux fois plus élevé en mode générique qu'un mode spécifique. Par ailleurs, le volume de stockage en générique est 3 voire 4 fois plus important qu'en spécifique.

En partenariat avec la société Ardent Software, nous avons également étudié comment améliorer les performances des requêtes sur des documents (plus généralement, sur des arbres) stockés en mode spécifique[45]. Nous avons étudié de façon extensive différents algorithmes et établi les règles permettant de choisir l'un plutôt que l'autre en fonction du contexte et de la requête.

## 5.2 Données multidimensionnelles et statistiques

**Mots clés** : multidimensionnel, contrainte, topologie, cartographie, agrégation, données statistiques.

**Participants** : Stéphane Grumbach, Leonid Libkin, Michel Scholl, Luc Ségoufin, Pierangelo Veltri, Victor Vianu.

**Résumé :** *Les bases de données contenant des informations scientifiques, du type géométrique, temporel, géographique, statistique, etc. posent des problèmes particuliers que nous abordons dans le projet sous différents angles. Nous étudions le modèle de bases de données avec contraintes, qui connaît un intérêt grandissant dans la communauté, et que nous utilisons pour l'élaboration d'un prototype en cours de développement pour les bases de données multidimensionnelles.*

### 5.2.1 Bases de données avec contraintes

Les données multidimensionnelles et, en particulier, les données spatiales conduisent à des représentations infinies (e.g. sous-espace du plan réel) mais admettant une représentation effective finie. De telles données peuvent être représentées à l'aide de contraintes sur des domaines numériques, par exemple, les contraintes polynomiales sur les réels ou les contraintes linéaires sur les rationnels. Les bases de données avec contraintes généralisent les bases de données relationnelles et un certain nombre d'outils du modèle relationnel peuvent être utilisés dans ce contexte étendu. Un livre faisant un état de l'art sur les bases de données avec contrainte est en cours de finition [10]. Le projet Verso a participé à la rédaction de nombreux chapitres de ce livre : [24, 14, 15, 17, 18, 20, 23, 22, 25].

Les bases de données avec contraintes permettent une modélisation robuste des données spatiales. Dans ce cadre, il est souvent nécessaire d'avoir des requêtes calculant la surface ou le volume des objets manipulés. Le rajout de tels prédicats aux langages de requêtes existants pose problème car cela permet de définir des ensembles non représentables par les techniques classiques. Deux approches sont alors possibles : restreindre le langage de requêtes de départ ou faire un calcul seulement approché du volume. Ces deux approches ont été étudiées dans [31].

### 5.2.2 DEDALE

Un effort important a été consacré à la réalisation d'un prototype de base de données spatiales, DEDALE, implémenté sur le SGBD O<sub>2</sub>, en collaboration avec le CNAM (M. Scholl, P. Rigaux) et basé sur le modèle avec contraintes [20]. Les contraintes sont linéaires, ce qui suffit à la modélisation des données spatio-temporelles [21]. Deux applications tournent déjà. L'interrogation se fait via une interface graphique Java afin de permettre son utilisation à distance. Le défi actuel est de réaliser l'optimisation de ces requêtes, qui sont exprimées indépendamment de leur stratégie de calcul. On propose, pour cela, des méthodes basées sur des critères tels que la dimension, la géométrie et la topologie des données [18]. Dans [36], on montre comment évaluer des requêtes sur des objets de dimension supérieure à 2 à l'aide de techniques d'approximation. Ces techniques, assez complexes, permettent de faire une évaluation des requêtes avec une complexité linéaire en la dimension des objets (celle-ci est exponentielle en général), moyennant certaines restrictions sur les données utilisées. Dans [35], on montre comment il est possible de modéliser et d'interroger des données interpolées de manière complètement transparente pour l'utilisateur. On montre aussi que l'évaluation des requêtes sur de telles données peut, en fait, se ramener à la manipulation d'objets de dimension 2, réduisant encore une fois la complexité exponentielle en la dimension en une complexité linéaire. Dans [44], on propose

une évaluation en deux étapes, la première rapide mais approximative servant de filtre à la seconde, plus coûteuse en temps.

### 5.2.3 Représentation de la topologie et requêtes topologiques

Dans certaines applications, la forme des régions est importante alors que, dans d'autres, on ne s'intéresse qu'à leurs propriétés *topologiques*. Dans [26, 27, 23], on étudie les requêtes topologiques sur des bases de données spatiales en dimension 2, où les régions sont spécifiées par des inégalités polynomiales avec des coefficients entiers. Un résumé des résultats connus sur ce sujet est présenté dans [23]. Plus précisément, dans [26] on étudie différents langages de requêtes permettant d'exprimer des propriétés topologiques et on montre que toute l'information topologique peut être regroupée dans une base de données relationnelle finie appelée invariant topologique et vue comme une annotation topologique des données spatiales brutes.

Il est possible de répondre à n'importe quelle requête topologique en utilisant cet invariant topologique. Comme il a généralement une taille bien moins importante que la base de donnée spatiale elle-même, cette méthode induit une stratégie d'évaluation des requêtes topologiques potentiellement plus efficace. La traduction d'une requête sur des données spatiales en une requête équivalente sur l'invariant topologique est examinée dans [27]. Il est montré que le langage "fixpoint+counting" exprime exactement les requêtes topologiques PTIME sur l'invariant topologique. Cela suggère que les invariants topologiques sont des structures particulièrement bien adaptées par rapport à la logique descriptive.

### 5.2.4 Données agrégées

Nous étudions une logique avec compteurs étendue à l'ensemble de tous les opérateurs d'agrégation. Dans [37], nous montrons que ce formalisme satisfait des théorèmes de localité analogues à ceux de Hanf et Gaifman. Nous considérons ensuite un dialecte de SQL avec les fonctions standards d'agrégation, ainsi que les clauses GROUPBY et HAVING, exprimable dans la logique, ce qui permet de montrer des bornes d'expressivité, qui ne dépendent pas de la classe de fonctions arithmétiques, et qui généralisent les résultats précédents. Dans [34], nous étudions la complexité des requêtes sur des données agrégées, et montrons en particulier la faisabilité des requêtes dans une logique du premier-ordre avec l'arithmétique réelle, et les fonctions d'agrégation, compteur et somme et produit itérés.

Nous étudions aussi l'évaluation incrémentale des requêtes SQL. Il existe des requêtes non-exprimable dans SQL, mais qui peuvent être évaluées incrémentalement dans SQL. L'exemple le plus complexe est une requête PTIME-complete. Dans [39], nous montrons que, grâce à l'agrégation, SQL peut évaluer incrémentalement toute requête dont la complexité dans les données est dans la hiérarchie polynomiale. De plus, certaines requêtes PSPACE-complete peuvent aussi être évaluées.

Dans [34], nous étudions aussi un problème fondamental pour l'interrogation des données agrégées et l'optimisation des requêtes, le problème de l'équivalence de requêtes. Ce problème est indécidable, en général. Nous montrons que, pour certaines requêtes simples, mais importantes en pratique, l'équivalence peut être décidée en temps non-déterministe polynomial. Ces résultats sont ensuite utilisés pour répondre à des requêtes sur des données statistiques.

### 5.3 Vues actives pour le commerce électronique

**Participants :** Serge Abiteboul, Sébastien Ailleret, Bernd Amann, Sophie Cluet, Amélie Marian, Laurent Mignet, Bruno Tessier, Anne-Marie Vercoustre.

**Mots clés :** règle de production, base de données active, parallélisme, déduction, datalog, workflow, commerce électronique, calcul relationnel, non-déterminisme.

**Résumé :** *Nous nous intéressons ici aux applications Internet nécessitant la manipulation de gros volumes de données et une forte interaction entre différents utilisateurs (par exemple, le commerce électronique).*

Nous avons poursuivi cette année nos travaux sur les "vues actives". Il s'agit de pouvoir offrir des vues d'une base de données définies déclarativement et qui incluent des aspects actifs comme d'être notifié de certains événements ou d'être averti quand certaines valeurs de la vue ont (peuvent avoir) été modifiées dans la base de données. Le but à long terme est de pouvoir offrir un tel outil sur le Web. Nous travaillons sur des modules comme : (i) le maintien de trace de l'activité du système interrogeable avec le langage de requête standard, (ii) un système de règles actives permettant au système de réagir à des notifications.

Nous considérons, en particulier, deux grandes applications. D'abord, le développement de catalogues de vente électronique sur le Web. Ceci fait l'objet du projet GAEL labellisé RNRT en commun avec le LRI (Marie-Christine Rousset) et EZEcom, une start-up en cours de création. Il s'agit de faciliter à des non-informaticiens le développement de boutiques sur le Web en s'appuyant sur des langages déclaratifs. La seconde concerne les systèmes de gestion d'information pour la fabrication. Ceci est l'objet d'un projet commun AFIRST avec l'Université de Tel Aviv (Tova Milo).

Une première ébauche du système AViews a été présentée à la conférence VLDB [28] et un prototype du système a été démontré dans cette même conférence (<http://www-rocq.inria.fr/verso/ACTIVEVIEWS/>). A. Marian a travaillé sur la surveillance de documents et de requêtes, S. Ailleret sur le module de règles actives, B. Hills et B. Tessier sur l'interface.

## 6 Contrats industriels (nationaux, européens et internationaux)

**Ardent Software** Depuis sa création, nous collaborons avec la société Ardent Software (ex O<sub>2</sub>Technology) qui développe et commercialise le système de gestion de bases de données objet O<sub>2</sub>. Cette année, cette collaboration a porté sur l'optimisation de requêtes sur des arbres via une bourse CIFRE.

Verso n'a pu que constater la fermeture du centre de Versailles d'Ardent Software. Pour être résolument positif, ceci nous permet d'avoir des contacts avec de nouvelles sociétés (iMediation, oXyMeL) et surtout de bénéficier de la présence dans l'équipe de Guy Ferran (ex-directeur technique d'O<sub>2</sub>).

## 7 Actions régionales, nationales et internationales

### 7.1 Actions nationales

**Résumé :** *En 1999 nous avons lancé le projet RNRT GAEL*

#### 7.1.1 Projet RNRT GAEL

Ce projet réunit une start-up Ezeecom, l'équipe de M.C. Rousset au LRI (U. Paris Sud) et le projet Verso.

L'objectif du projet est de concevoir un générateur de catalogues électroniques et de services associés, permettant à des non informaticiens de développer leurs propres catalogues (boutiques, galeries marchandes) électroniques sans gros effort de programmation et en les personnalisant à leurs besoins. Pour ce faire, on combinera des techniques de représentation des connaissances et d'agents intelligents. L'effort de recherche portera principalement sur une spécification déclarative des connaissances relatives au commerce électronique (les différents types de produits, de profils clients, d'actes commerciaux), ainsi que des informations présentées aux clients en XML. Nous utiliserons des agents intelligents pour doter les catalogues électroniques de comportements dynamiques permettant de réagir de façon adéquate aux achats en cours des clients en fonction du contenu du catalogue, de la stratégie commerciale spécifiée de façon déclarative par le vendeur, ainsi que du profil des clients.

#### 7.1.2 Autres Collaborations Nationales

Des liens étroits existent avec le Labri (N. Bidoit, B. Courcelle), le LRI (C. Delobel, E. Waller, M.C. Rousset) et le Cedric au CNAM (M. Scholl, B. Amann, P. Rigaux).

### 7.2 Actions financées par la commission européenne

#### 7.2.1 Projet C-WEB

Le projet C-Web est un projet européen qui comprend comme partenaires l'INRIA (équipes Verso et Acacia) et EDW, une PME italienne spécialisée dans les systèmes de gestion d'informations d'entreprise, basés sur SGML/XML.

Dans la première phase du projet (un an), il s'agit de définir une plate-forme générique et ouverte pour la gestion de Webs communautaires, ou de mémoire d'entreprise. Le projet final doit inclure d'autres partenaires industriels.

Le serveur de C-Web peut être consulté à <http://cweb.inria.fr>.  
Le responsable Verso pour C-Web est A.M. Vercoustre.

### 7.2.2 Projet TMR Chorochronos

Chorochronos est un programme d'échanges européen TMR (Training and Mobility of Researchers) sur les bases de données spatiales et temporelles. La collaboration entre les 10 noeuds du réseau porte sur les thèmes suivants (i) Structure et représentation de l'espace et du temps, (ii) Modèles et langages pour les SGBD spatio-temporels (SGBDST), (iii) Interfaces pour SGBDST, (iv) Exécution et optimisation de requêtes, (v) Structures de données et indexation spatio-temporelle, et (vi) Architecture d'un SGBDST. Dans ce projet, Verso est le noeud français. Ses contributions essentielles, en collaboration avec le groupe bases de données Vertigo du Cedric au CNAM, sont l'étude de modèles et langages pour les données multidimensionnelles, l'optimisation de l'évaluation des langages et le développement du prototype DEDALE pour manipuler des données multidimensionnelles. Un certain nombre d'applications spatio-temporelles sont étudiées pour valider ces études. Le projet TMR Chorochronos a été prolongé d'une année jusqu'en juin 2000.

Les responsables Verso pour Chorochronos sont S. Grumbach et M. Scholl.

### 7.2.3 GDR-PSIG Cassini

Cassini est un programme de recherche national pluridisciplinaire (informaticiens, géographes) financé par le CNRS et l'Institut Géographique National (IGN) sur l'information géographique. Dans ce programme, le projet Verso, en collaboration avec le CNAM d'une part et avec le LSR de l'IMAG et le LAMA à Grenoble, d'autre part, étudie les problèmes de modélisation d'information spatiale évoluant au cours du temps. L'étude d'une application fournie par le LAMA (station de ski) a permis de valider l'approche bases de données contraintes et une avancée significative sur le prototype DEDALE.

Le responsable Verso pour Cassini est M. Scholl.

## 7.3 Réseaux et groupes de travail internationaux

Verso participe à Pastel, un groupe de travail financé par la communauté européenne dont l'objectif est la réalisation de systèmes pour des applications persistentes. Ce groupe de travail fait suite au projet Esprit B Fide2 auquel Verso participait.

Verso est membre du réseau d'excellence Compulog (logic programming) et DELOS (European Digital Libraries) et participe au groupe "Bases de données" de l'Ercim. Verso participe également à un working group d'Ercim sur la programmation avec contraintes en cours de création. A.M Vercoustre participe aussi au groupe de travail pour l'internalisation du Dublin Core. A.M.Vercoustre a représenté l'INRIA dans les groupes ERCIM pour les Bibliothèques Electroniques et a participé au projet DELOS pour la création de ETRDL (ERCIM Technical

Reference Digital Library). Une démonstration a été présentée à Amsterdam, pour les 10 ans d'ERCIM, en novembre 99.

## 7.4 Relations bilatérales internationales

### 7.4.1 Europe

Verso et le centre de recherche en informatique FORTH (en Crète) ont soumis une proposition de collaboration dans le cadre du programme de recherche PLATON'2000 du ministère des Affaires étrangères (en liaison avec le ministère de l'éducation nationale, de la Recherche et de la Technologie). Les objectifs scientifiques de cette demande concernent la médiation de données culturelles 4.3. La réponse est attendue pour novembre 1999.

### 7.4.2 Moyen-Orient

Financé par l'Association Franco-Israélienne pour la Recherche Scientifique et Technique (AFIRST), le projet "Les Usines du Futur" nous permet de collaborer étroitement avec l'Université de Tel Aviv. Le contrat porte sur l'utilisation des vues actives pour permettre le contrôle de fabrication (voir 5.3 et [11]) et nous a déjà permis de développer un prototype au-dessus du système AXIELLE développé par la société Ardent Software. Nous continuons à travailler sur ce prototype, notamment sur les aspects d'optimisation de requêtes.

### 7.4.3 Amérique du Nord

En Amérique du Nord, des travaux en commun sont en cours avec l'Université de Stanford (J. Widom), Pennsylvanie (P. Buneman), UC Santa Barba (J. Su) [22], et UC San Diego (V. Vianu), ATT (Sihem Amer-Yahia, Divesh Srivastava, Dan Suciu), Lucent-Bell (Jérôme Siméon), U.C. Berkeley (Christos Papadimitriou).

### 7.4.4 Asie et océan Pacifique

En 1999, A.-M. Vercoustre a passé deux semaines au CSIRO, à Melbourne, en Australie, dans le groupe "Technologies for Electronic Documents". La collaboration se poursuit sur le thème des documents virtuels, et nous avons accueilli Brendan Hills pour 5 mois en 1999. Il a travaillé avec nous sur des vues actives pour XML. Plus généralement, A.-M. Vercoustre est chargée de coordonner les relations de l'Inria avec l'Australie.

## 7.5 Accueil de chercheurs étrangers

Cette année, nous avons accueilli :

- Brendan Hills, CSIRO, Australie, pour travailler sur ActiveViews pour XML (5 mois).
- Leonid Libkin, Chercheur à Bells Labs, USA (1 an).
- Victor Vianu, Professeur, UC San Diego (6 mois).
- Timos Sellis, NTU-Athens- Grèce (1 mois).

Nous avons également accueilli pour de courtes visites d'autres chercheurs et professeurs étrangers avec lesquels nous avons des collaborations suivies : Vassilis Christophides (chercheur, FORTH-Crète, 3 semaines), Tova Milo (professeur, U. Tel Aviv, 15 jours), Guido Moerkotte (professeur, U. Manheim, 15 jours), Dan Suciu (professeur, ATT Laboratory-USA, 15 jours), Agnès Voisard (professeur, U. Berlin, 20 jours).

## 8 Diffusion de résultats

### 8.1 Actions d'enseignement

- S. Abiteboul est maître de conférence à l'Ecole Polytechnique.
- B. Amann est maître de conférence au CNAM-Paris.
- C. Delobel est professeur à l'Université de Paris 11.
- M. Scholl est professeur au CNAM.
- V. Vianu est professeur à UCSD.

Les cours suivants ont été assurés par plusieurs membres de l'équipe.

#### *SGBD relationnels,*

CNAM-Paris, Cycles A et B, B. Amann  
MIAGE et nouvelle formation d'ingénieurs, Paris 11, C. Delobel.

#### *SGBD avancés,*

Polytechnique, S. Abiteboul;  
CNAM-Paris, M. Scholl et B. Amann;  
DEA Systèmes Informatiques, cohabilité Paris 6-Telecom-CNAM, B. Amann, M. Scholl.

*Bases de données avec contraintes,* DEA programmation en convention avec l'Inria, et Université de Calabre, S. Grumbach.

*Modèle et Langages de Requêtes à objets,* DEA Systèmes Informatiques de Paris VI, S. Cluet.

*Systèmes d'intégration de données,* DEA Informatique de Paris-Dauphine, S. Cluet.

*Algorithmique et Complexité*, Polytechnique, S. Abiteboul.

*Théorie des modèles finis*, DEA Paris VII, S. Abiteboul et S. Grumbach

## 8.2 Participation à des colloques

L'équipe a eu de nombreuses publications dans des conférences internationales et des colloques (voir la bibliographie). Enfin, certains membres de l'équipe ont participé à des comités de programmes. La liste en est donnée ci-dessous.

S. Abiteboul

- 1999 : EDBT Summer School'99 (Président du comité de programme), PODS'99 (ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems), ECDL'99 (European Conference on Digital Libraries - Président du comité de programme), VLDB'99 (International Conference on Very Large Databases), l'école Dagstuhl sur "Foundations for Information Integration."

B. Amann

- 1999 : INFORSID'99, DEXA'99 (Ninth International Conference on Database and Expert Systems Applications), BDA'99 ( Journées Bases de Données Avancées).

S. Cluet

- 1999 : WebDB (ACM-Sigmod Workshop on the Web and Databases - co-présidente du comité de programme), PODS'99 (ACM SIGACT-SIGMOD-SIGART Symposium on Principle of Database Systems), CIKM (Conference on Information Knowledge and Management), ICDE (International Conference on Data Engineering), SIGMOD'99 (ACM Conference on the Management Of Data).

S. Grumbach

- 1999 : SSD'99 (6th International Symposium on Large Spatial Databases), JFPLC'99 (8e Journées Francophones de Programmation Logique et Programmation par Contraintes), BDA'99 (Journées Bases de Données Avancées), DBPL'99 (Workshop on Database Programming Languages), STDBM'99, (Workshop on Spatio-Temporal Database Management);
- 2000 : WAIM'2000 (International Conference on Web-Age Information Management).

L. Libkin

- 1999 : DaWaK (International Conference on Data Warehousing and Knowledge Discovery ), LPAR (Logic for programming and automated reasoning), DBPL'99 (Database programming languages);

- 2000 : FOIKS (Foundations Of Information and Knowledge Systems), PODS'99 (ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems).

M. Scholl

- 1999 : HT'99 (10th ACM conference on Hypertext), CAISE'99 (11th Conference on Advanced Information Systems), SSD'99 (6th International Symposium on Spatial Databases), COSIT'99 (Conference on Spatial Information Theory), DEXA'99 (Ninth International Conference on Database and Expert Systems Applications), ADBIS'99 (Third East-European Symposium on Advances in Databases and Information Systems), STDBM'99 (Co-PC chair - Workshop on Spatio-Temporal Database Management), ER'99 (International Conference on Entity-Relationship Model).

A.-M. Vercoustre

- 1999 : WWW8 (Co-Présidente - 8th International World Wide Web Conference), ECDL'99 (Third European Conference on Digital Library - co-organisatrice), H2PTM'99 (Hypertexts and Hypermedia: Products, Tools, Methods), IHM'99 (Onzièmes journées sur l'Interaction Homme-Machine), ACDS'99 (3ème Australian Computing Document Symposium),
- 2000 : WWW9 (9th International World Wide Web Conference), HT2000 (ACM Conference on Hypertext and Hypermedia), ECDL2000 (European Conference on Digital Libraries).

V. Vianu

- 1999 : Workshop on Query Processing for Semistructured Data and Non-Standard Data.

### 8.2.1 Conférences invitées, tutoriels, cours, etc.

S. Abiteboul a présenté une conférence invitée à ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, 1999. Il a présenté un cours sur le commerce électronique à l'école d'été sur "Extending Database Technology" à La Baule, 1999.

S. Grumbach a fait un exposé invité sur les bases de données contraintes à JFPLC'99, Huitièmes Journées Francophones de Programmation Logique et Programmation par Contraintes

L. Libkin a présenté une conférence invitée à KGC'99, Kurt Goedel Colloquium, Espagne, 1999, sur les bases de données avec contraintes et les modèles finis encastrés. Il a présenté une autre conférence invitée aux journées de bases de données avancées (BDA99), Bordeaux, 1999, sur le pouvoir expressif des langages de requêtes. Il a aussi présenté un cours sur les modèles finis encastrés au DIMACS tutorial and workshop on Logic and Cognitive Science, USA, 1999. Il est éditeur associé des archives de SIGMOD.

V. Vianu a présenté plusieurs conférences invitées sur les requêtes topologiques dans les bases de données topologiques géographiques : au Annual meeting of the Association of Symbolic Logic, San Diego, mars 1999 ; à la Federated Logic Conference, Trento, juillet 1999 ; et à

la Conference on Computer Science Logic, Madrid, septembre 1999; ainsi qu'aux universités de Freiburg, Allemagne, novembre 1999 et de Caen, novembre 1999. Il a aussi donné une conférence invitée sur la logique en tant que langage de requêtes au Meeting of the American Association for the Advancement of Science, Anaheim, California, juin 1999.

A.-M. Vercoustre a fait une présentation sur "Web et Bases de Données" au séminaire de La Villette, janvier 1999. Elle a été co-éditeur avec S. Abiteboul des actes de la conférence ECDL99, et est éditeur invitée pour un numéro de "Journal of Digital Libraries".

M. Scholl a présenté un tutoriel sur les bases de données spatio-temporelles à l'école d'été européenne EDBT en mai à La Baule. Il est membre du comité de lecture du journal *Geoinformatica*.

S. Abiteboul et S. Cluet ont participé à la série *Reminiscences on Influential Papers* du Journal SIGMOD Record.

V. Vianu est éditeur d'une colonne sur la théorie des bases de données dans SIGACT News. Il est membre du comité de lecture du "Journal of Discrete Mathematics and Theoretical Computer Science", du "SIGMOD Digital Reviews", ainsi que du "ACM Transactions on Computational Logic".

S. Abiteboul est éditeur de "ACM Transactions on Database Systems", de "Information and Computation", de "Journal of Digital Libraries" et de "Theory and Practice of Object Systems" (TAPOS).

### 8.2.2 Animations scientifiques

S. Cluet a organisé avec T. Milo le second workshop international sur le Web et les bases de données qui s'est tenu à Philadelphie en mai. Ce workshop était parrainé par l'INRIA et l'ACM.

S. Abiteboul était président du comité de programme de la conférence européenne sur les bibliothèques numériques organisé par l'INRIA et la Bibliothèque Nationale de France à Paris en 1999 [9]. Ce colloque a été organisé par les Relations Extérieures de l'INRIA. S. Abiteboul, A.-M. Vercoustre et B. Amann ont participé à cette organisation.

S. Abiteboul était responsable de l'organisation scientifique de l'école d'été européenne "Extending Database Technology" à La Baule en 1999. Il a participé à son organisation en collaboration avec l'Université de Nantes (N. Mouaddib). S. Abiteboul a aussi participé à l'organisation scientifique de l'école Dagstuhl sur "Foundations for Information Integration".

S. Abiteboul est membre du comité de coordination des STIC (Ministère de la Recherche).

A.M. Vercoustre a été élue trésorière de SIGWEB (ACM SIG on Hypertext, Hypermedia and the Web). Elle est co-organisatrice d'un Workshop sur les "Documents virtuels personnalisés", à la conférence IHM99, Montpellier, Nov. 99.

L. Libkin est membre du comité d'organisation de la conférence Logic in Computer Science (LICS) et coorganisateur de la conférence Federated Logic Conference (FLoC99), Italie, 1999.

V. Vianu est General Chair du ACM SIGMOD-SIGACT-SIGART Symp. on Principles of Database Systems, et membre du Comité Exécutif de SIGMOD. Il est co-organisateur du séminaire Dagstuhl "Finite Model Theory, Databases, and Computer-Aided Verification", octobre 1999.

## 9 Bibliographie

### Ouvrages et articles de référence de l'équipe

- [1] S. ABITEBOUL, P. BUNEMAN, D. SUCIU, *Data on the Web: From Relations to Semistructured Data and XML*, Morgan-Kaufman, New York, 1999.
- [2] S. ABITEBOUL, R. HULL, V. VIANU, *Foundations of Databases*, Addison-Wesley, 1995.
- [3] S. ABITEBOUL, V. VIANU, « Computing With First-Order Logic », *Journal of Computer and System Sciences (JCSS)* 50(2), 1995, p. 309–335.
- [4] S. GRUMBACH, P. RIGAUX, L. SEGOUFIN, « The DEDALE System for Complex Spatial Queries », *in : sigmod*, 1998.
- [5] S. GRUMBACH, J. SU, « Finitely representable databases. », *Journal of Computer and System Sciences (JCSS)* 55, 2, 1999, p. 273–298.
- [6] S. GRUMBACH, C. TOLLU, « On the Expressive Power of Counting », *Journal of Theoretical Computer Science (TCS)*, 1995.
- [7] M. SCHOLL, A. VOISARD, J.-P. PELOUX, L. RAYNAL, P. RIGAUX, *SGBD Géographiques*, International Thomson Publishing, 1996.

### Livres et monographies

- [8] S. ABITEBOUL, P. BUNEMAN, D. SUCIU, *Data on the Web: From Relations to Semistructured Data and XML*, Morgan-Kaufman, New York, 1999.
- [9] S. ABITEBOUL, A. VERCOUSTRE (éditeurs), *Research in Advanced Technology for Digital Libraries, Proceedings of the Third European Conference, ECDL'99, 1696*, Paris, France, Springer-Verlag, 1999.
- [10] G. KUPER, L. LIBKIN, J. PAREDAENS, *Constraint Databases*, Springer-Verlag, 1999, à paraître.

**Articles et chapitres de livre**

- [11] S. ABITEBOUL, S. CLUET, A. EYAL, P. MOGILEVSKY, J. SIMÉON, S. ZOHAR, T. MILO, «Tools for Data Translation and Integration», *IEEE Data Engineering Bulletin* 22(1), 1999, p. 3–8.
- [12] S. ABITEBOUL, L. HERR, J. V. DEN BUSSCHE, «Temporal Connectives Versus Explicit Timestamps to Query Temporal Databases», *Journal of Computer and System Sciences* 58, 1, 1999, p. 54–68.
- [13] S. ABITEBOUL, V. VIANU, «Regular Path Queries with Constraints», *Journal of Computer and System Sciences* 58, 3, 1999, p. 428–452.
- [14] M. BENEDIKT, L. LIBKIN, «Expressive Power: the Finite Case», in: *Constraint Databases*, G. Kuper, L. Libkin, et J. Paredaens (éditeurs), Springer-Verlag, 1999, à paraître.
- [15] M. BENEDIKT, L. LIBKIN, «Query Safety and Constraints», in: *Constraint Databases*, G. Kuper, L. Libkin, et J. Paredaens (éditeurs), Springer-Verlag, 1999, à paraître.
- [16] S. CHAWATHE, S. ABITEBOUL, J. WIDOM, «Managing historical semistructured data», *Theory and Practice of Object Systems* 5, 3, 1999, p. 143–162.
- [17] J. CHOMICKI, L. LIBKIN, «Aggregate Languages for Constraint Databases», in: *Constraint Databases*, G. Kuper, L. Libkin, et J. Paredaens (éditeurs), Springer-Verlag, 1999, à paraître.
- [18] S. GRUMBACH, Z. LACROIX, P. RIGAUX, L. SEGOUFIN, «Query Evaluation and Optimization», in: *Constraints Databases*, G. Kuper, L. Libkin, et J. Paredaens (éditeurs), Springer-Verlag, 1999, à paraître.
- [19] S. GRUMBACH, T. MILO, «An Algebra for Pomsets», *Information and Computation* 150, 1999, p. 268–306.
- [20] S. GRUMBACH, P. RIGAUX, M. SCHOLL, L. SEGOUFIN, «The Design and Implementation of a Constraint-Based System: DEDALE», in: *Constraints Databases*, G. Kuper, L. Libkin, et J. Paredaens (éditeurs), Springer-Verlag, 1999, à paraître.
- [21] S. GRUMBACH, P. RIGAUX, L. SEGOUFIN, «Spatio-Temporal Data Handling with Constraints», *GeoInformatica*, 1999, à paraître.
- [22] S. GRUMBACH, J. SU, «Expressive power in the infinite», in: *Constraints Databases*, G. Kuper, L. Libkin, et J. Paredaens (éditeurs), Springer-Verlag, 1999.
- [23] B. KUIJPERS, V. VIANU, «Topological Queries», in: *Constraints Databases*, G. Kuper, L. Libkin, et J. Paredaens (éditeurs), Springer-Verlag, 1999, à paraître.
- [24] G. KUPER, L. LIBKIN, J. PAREDAENS, «Introduction», in: *Constraint Databases*, G. Kuper, L. Libkin, et J. Paredaens (éditeurs), Springer-Verlag, 1999, à paraître.
- [25] G. KUPER, M. SCHOLL, «Geographic Information Systems», in: *Constraints Databases*, G. Kuper, L. Libkin, et J. Paredaens (éditeurs), Springer-Verlag, 1999, à paraître.
- [26] C. PAPADIMITRIOU, D. SUCIU, V. VIANU, «Topological Queries in Spatial Databases», *Journal of Computer and System Sciences*, 1999.

- [27] L. SEGOUFIN, V. VIANU, «Querying Spatial Databases via Topological Invariants», *Journal of Computer and System Sciences*, 1999, invité pour un numéro spécial.

### Communications à des congrès, colloques, etc.

- [28] S. ABITEBOUL, B. AMANN, S. CLUET, A. EYAL, L. MIGNET, T. MILO, «Active Views for Electronic Commerce», in: *Int. Conf. on Very Large DataBases (VLDB)*, Edinburgh, Scotland, septembre 1999.
- [29] S. ABITEBOUL, «On views and XML», in: *Proc. ACM Symp. on Principles of Database Systems*, 1999.
- [30] B. AMANN, I. FUNDULAKI, «Integrating Ontologies and Thesauri to Build RDF Schemas», in: *Proceedings of the Third European Conf. on Digital Libraries (ECDL'99)*, Paris, France, septembre 1999.
- [31] M. BENEDIKT, L. LIBKIN, «Exact and Approximate Aggregation in Constraint Query Languages», in: *Proceedings of ACM Principle of Database Systems*, p. 102–113, 1999.
- [32] S. CLUET, O. KAPITSKAIA, D. SRIVASTAVA, «Using LDAP Directory Caches», in: *PODS*, 1999.
- [33] S. GRUMBACH, G. MECCA, «In search of the lost schema», in: *Proc. of Intl. Conf. on Database Theory*, 1999.
- [34] S. GRUMBACH, M. RAFANELLI, L. TINININI, «Querying Aggregate Data», in: *PODS*, 1999.
- [35] S. GRUMBACH, P. RIGAUX, L. SEGOUFIN, «Modeling and Querying Interpolated Spatial Data», in: *Bases de Données Avancées*, 1999.
- [36] S. GRUMBACH, P. RIGAUX, L. SEGOUFIN, «On the Orthographic Dimension of Constraint Databases», in: *ICDT*, 1999.
- [37] L. HELLA, L. LIBKIN, J. NURMONEN, L. WONG, «Logics with Aggregate Operators», in: *Proceedings of the 14th IEEE Symposium on Logic in COmputer Science*, p. 35–44, 1999.
- [38] T. LAHIRI, S. ABITEBOUL, J. WIDOM, «Ozone: Integrating Structured and Semistructured Data», in: *Proc. International Workshop on Database Programming Languages*, 1999.
- [39] L. LIBKIN, L. WONG, «On the power of incremental evaluation in SQL-like languages», in: *Preliminary Proceedings of the 7th Workshop on Database Programming Languages (DBPL'99)*, R. Connor, A. Mendelzon (éditeurs), Univ. of Glasgow, 1999. Full post-conference proceedings will appear in Springer LNCS, in 2000.
- [40] C. LINDLEY, A. VERCOUSTRE, «Generic Viewer Interaction Semantics for Dynamic Virtual Video Synthesis», in: *third International Conference on Visual Information Systems (Visual99)*, Amsterdam, juin 1999.
- [41] A. VERCOUSTRE, F. PARADIS, «Metadata for Photographs: from Digital Library to Multimedia Application», in: *Third European Conference on Digital Libraries (ECDL'99)*, Paris, France, septembre 1999.

**Divers**

- [42] V. CHRISTOPHIDES, S. CLUET, J. SIMÉON, «On Wrapping Query Languages and Efficient XML Integration», Rapport Verso numéro 170, Octobre 1999.
- [43] S. CLUET, S. GAMERMAN, «OQL: le solide challenger 100% objet», Programmez!, Janvier 1999.
- [44] S. GRUMBACH, P. RIGAUX, P. VELTRI, «Hierarchical Optimization of Linear Constraint Processing», 1999.
- [45] F. WATTEZ, S. CLUET, V. BENZAKEN, G. FERRAN, C. FIEGEL, «Benchmarking Tree Queries: Learning the Hard Truth the Hard Way», Rapport Verso numéro 169, Octobre 1999.