

# Ioana MANOLESCU-GOUJOT

## 1. Etat Civil

Chargée de recherche 1re classe, EPI Gemo, INRIA Saclay-Île-de-France  
Bâtiment G, 4, rue Jacques Monod, 91893 Orsay Cedex  
Courriel: ioana.manolescu@inria.fr, web: <http://www-rocq.inria.fr/~manolesc>  
33 ans, mariée, 2 enfants

## 2. Titres Universitaires

**Thèse de doctorat:** “Techniques d’optimisation pour l’interrogation de données hétérogènes et distribuées”, sous la direction de Patrick Valduriez, décembre 2001, Université. de Versailles Saint-Quentin. Mention: très honorable avec les félicitations du jury.

**Magistère de Mathématiques Fondamentales et Appliquées et d’Informatique** en juillet 1999, Ecole Normale Supérieure de Paris et Université de Paris VI. Mention: bien.

**Diplôme d’Etudes Approfondies en Systèmes Informatiques** en septembre 1998, Université de Paris XI. Mention: très bien.

**Maîtrise en Informatique** en septembre 1997, Ecole Normale Supérieure de Paris et Université de Paris XI. Mention: bien.

**Licence en Informatique** en février 1997, Ecole Normale Supérieure de Paris et Université de Paris XI. Mention: assez bien.

## 3. Parcours

<b>12/2005-présent</b>	Chargée de recherche 1re classe, projet Gemo, INRIA Saclay-Île-de-France.
<b>09/2002-11/2005</b>	Chargée de recherche 2e classe, projet Gemo, INRIA Futurs.
<b>03/2002-08/2002</b>	Post-doc, Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italie.
<b>01/2002-02/2002</b>	Ingénieur expert, Projet Gemo, INRIA Futurs.
<b>12/1998-12/2001</b>	Allocataire de recherche, Université de Versailles (thèse effectuée dans le projet Rodin/Caravel, INRIA Rocquencourt).
<b>09/1996-09/1998</b>	Elève étrangère à l’Ecole Normale Supérieure à Paris, boursière de la fondation Soros et du Ministère des Affaires Etrangères.

#### 4. Activités d'enseignement

Tous les volumes horaires sont indiqués en équivalent TD.

**Traitement de requêtes XML** (20h), cours dans le Master Recherche en Informatique (2e année) de l'Université de Paris XI, 2002-2006 et 2008. J'ai élaboré les supports de cours sur le stockage de données XML, optimisation algébrique et traitement de requêtes XQuery.

**Services Web** (18h), intervention dans le Master Recherche en Informatique (2e année) de l'Université de Paris IX, 2006 (<http://www-rocq.inria.fr/~manolesc/WS>). J'ai élaboré les supports de cours sur les protocoles de communication à l'aide des services Web (SOAP, WSDL), orchestration (BPEL4WS) et conversations (WSCL).

**Bases de données avancées** (20h), cours et encadrement de projets pour des élèves en 4e année à l'Institut de Formation d'Ingénieurs de Paris Sud, 2005-2006.

**Introduction aux bases de données** (30h), cours de Maîtrise à l'Université de Versailles-Saint Quentin, 2002.

**Bases de données et Internet** (36h), cours de 3e année d'école d'ingénieurs à l'Institut des Sciences et Techniques des Yvelines, de l'Université de Versailles-Saint Quentin, 2001.

**Bases de données avancées** (40h), cours de 3e année à l'École Nationale Supérieure de Telecommunications (ENST) de Paris, 2001.

**Introduction aux bases de données** (150h), cours de 3e année de Maîtrise de Méthodes Informatiques Appliquées à la Gestion des Entreprises (MIAGE) à l'Université de Paris XI, 1998-2000.

#### 5. Activités liées à l'administration

Membre élue du Comité du Centre INRIA Saclay-Île-de-France depuis mai 2008.

Membre de la commission de spécialistes externe du laboratoire PRISM de l'Université de Versailles Saint-Quentin depuis 2006.

Membre du Conseil d'Orientation Scientifique et Technologique (COST) de l'INRIA depuis 2007.

#### 6. Activités liées à la recherche

##### 6.1 Prix déjà reçus pour un article ou la thèse

N/A

##### 6.2 Participation à des comités, jurys, editorial boards, organisation de colloques, séminaires etc.

**Présidente** de l'initiative *SIGMOD Repeatability* 2008, visant à vérifier (reproduire) les expériences menées sur des logiciels dans le cadre des soumissions des articles scientifiques à la conférence *ACM SIGMOD*, réputée parmi les meilleures au niveau mondial en bases de données. J'ai organisé et supervisé la reproduction des expériences décrites dans 80 articles. Cet effort a été mené pour la première fois en 2008.

**Editeur** du numéro spécial 33(2) (2008) de la revue *Elsevier Information Systems* comprenant des articles invités du workshop ExpDB 2006.

**Présidente** de la section démonstrations de la conférence *Extending Database Technologies (EDBT)* 2008

**Co-présidente** de sections de conférences internationales:

- Section *Demonstrations* de la conférence *International Conference on Data Engineering (ICDE)* 2010
- Section *Web Engineering* de la conférence *World Wide Web Conference (WWW)* 2009
- Section *Databases* de la conférence *ACM Conference on Information and Knowledge Management (CIKM)* 2008

**Co-présidente des workshops internationaux**

- *Programming Languages for XML (PLAN-X)* 2008, en coopération avec ACM SIGPLAN
- *Web and Databases (WebDB)* 2008, en coopération avec ACM SIGMOD
- *Experimental Evaluation in Databases (ExpDB)* 2006, en coopération avec ACM SIGMOD
- *XQuery Implementation, Experience and Perspectives (XIME-P)* 2004, en coopération avec ACM SIGMOD.

Avec les co-présidents, nous sommes à l'origine de la création des workshops XIME-P (2004) et ExpDB (2006), en coopération avec ACM SIGMOD.

**Membre des comités de programmes des conférences internationales**

- *ACM SIGMOD International Conference on Management of Data (SIGMOD)* 2009 et 2004
- *Very Large Databases (VLDB)* 2008, 2007 et 2005
- *International Data Engineering Conference (ICDE)* 2008, 2006, 2005 et 2004
- *Extending Database Technologies (EDBT)* 2006
- *Conference on Innovative Database Research (CIDR)* 2009
- *World Wide Web Conference (WWW)* 2006 et 2005

**Membre des comités de programmes des workshops internationaux** *ACM Workshop on Web Information and Data Management (WIDM)* 2007, 2005 et 2004, *Programming Languages for XML (PLAN-X)* 2009, XIME-P 2006 et 2005, WebDB 2006 et 2005, *XML Symposium* en coopération avec VLDB 2005 et 2003 etc.

**Membre des comités de programmes de la conférence nationale** *Journées de Bases de Données Avancées* (BDA) 2008, 2007, 2005 et 2004.

**Lecteur** de soumissions aux revues: *ACM Transaction on Database Systems*, *ACM Transactions on Internet Technologies*, *VLDB Journal*, *Data and Knowledge Engineering Journal*, *International Web Engineering Journal*, *Elsevier Journal on Information Systems*, *Ingénierie des Systèmes d'Information*.

### 6.3 Programmes d'échanges, collaborations, réseaux internationaux, projets nationaux et européens

#### Projets nationaux

**Codex (2008-2011):** projet ANR DEFIS "Domaines émergents". *Je suis la responsable du projet.*

Le projet réunit 7 partenaires (INRIA Saclay, INRIA Lille, U. Paris VII, U. Paris XI, U. Blois, Laboratoire d'Informatique de Grenoble, et la start-up Innovimax) autour de la thématique des traitements efficace et hautement distribués sur des données XML. Le projet comporte trois volets: (i) traitements efficace sur des données XML, dans un contexte centralisé ou distribué; (ii) prise en compte et exploitation de la dynamique des données XML, en particulier présentes sur le Web; et (iii) interopérabilité et composition des programmes XML dans un contexte distribué.

**EDIFlow (2008-2011):** projet Digiteo Émergence "Interactions, visualisation et réalité virtuelle". *Je suis la responsable du projet.*

Ce projet est une nouvelle collaboration entre deux projets INRIA (Gemo et Aviz, spécialisé dans la visualisation) et l'équipe BD du LRI, autour de la thématique des workflows interactifs à usage intensif de données. Le but du projet est de proposer des modèles et un langage de workflow capables d'intégrer l'utilisateur et ses interventions au coeur de la spécification des processus. Il s'agira par la suite de déployer une plate-forme capable de supporter des processus spécifiés selon ce modèle, et efficace sur de grands volumes de données. Le domaine d'application visé en priorité est celui des processus de la gestion de données scientifiques.

**DataRing (2008-2011):** projet ANR "Réseaux du Futur et Services".

Le projet est une collaboration entre l'EPI Atlas de l'INRIA Rennes, Gemo, le LIG et le LIRMM autour de la gestion de données en réseaux pair-à-pair de très grande échelle. Les sujets abordés sont: données incertaines, intégration sémantique, mécanismes de redondance (réplication, cache).

**WebContent (2005-2008):** projet RNTL "Réseaux d'Information et de Connaissances". *Je suis responsable du lot "Gestion et interrogation de gros volumes de contenus en pair à pair".*

WebContent vise à fournir une plateforme configurable de gestion des données du Web sémantique. Le projet réunit des acteurs tels que le CEA, EADS, LIMSI, et des équipes INRIA, afin de concevoir un environnement complet pour l'acquisition, l'enrichissement, et l'exploitation des données sémantiques. Une architecture pair à pair permet le passage à l'échelle en termes de capacités de stockage et de traitement [CI1, CI2, D1].

**WebStand (2005-2008):** "ACI Jeunes Chercheuses & Jeunes Chercheurs", en collaboration avec B. Nguyen (U. Versailles), D. Colazzo (U. Paris XI) et A. Vion (sociologue, U. Aix-Marseille II). Le projet vise à proposer un modèle, une architecture et un système pour la construction

et gestion d'un entrepôt de données du Web en XML par des sociologues. Une plate-forme a été construite pour une application cible et validée avec les utilisateurs [c5].

**Tralala (2004-2007):** “Langages pour la manipulations de documents XML: fondements et pratique”, Actions Concertée Incitative “Masses de Données”. J’ai contribué à élaborer un modèle d’optimisation de requêtes *générique*, autant par le support de plusieurs langages de requêtes, que par la prise en compte de modèle de stockage très hétérogènes. Ce travail a été mené en collaboration avec V. Benzaken (U. Paris XI) et a constitué le sujet de thèse d’A. Arion.

**NTIC (2004-2007):** “Normalisation dans les technologies de l’information et de la communication”, Action Concertée Incitative en Sciences Sociales. Le projet a étudié l’établissement de normes et spécifications techniques en informatique. Il a été piloté par des chercheurs de l’Institut d’Etudes Politiques de Paris. En collaboration avec B. Nguyen (U. Versailles) et P. Senellart (Gemo), nous avons proposé un modèle et un premier outil permettant la création et l’exploitation d’un entrepôt de données XML à partir des listes de diffusion des groupes de travail du *World Wide Web Consortium (W3C)* [c7].

**MDP2P (2003-2006):** “Masses de données en pair-à-pair”, Action Concertée Incitative “Masses de Données”. *J’ai été responsable scientifique pour le projet Gemo.*  
Le but de MDP2P est de définir et mettre en œuvre des modèles et des plate-formes permettant la gestion facile de grands volumes de données en mode pair-à-pair. Dans MDP2P, Gemo a développé des algorithmes et une plate-forme pour la gestion de documents XML à très grande échelle dans des réseaux DHT [CN3, D6]. Ce travail s’est inscrit dans la thèse de N. Preda.

### Projets européens

**WebDam (2008-2011):** ERC Advanced Grant (S. Abiteboul en est *principal investigator*).

Le projet porte sur le développement de méthodes formelles pour la gestion des données du Web, en s’intéressant à l’échange, partage, intégration, interrogation et mise à jour des données. Je compte y travailler sur des nouveaux modèles de calculs distribués sur des données XML.

**eDos (2004-2007):** projet européen centré sur la gestion des applications distribuées telles que, par exemple, la gestion du développement de logiciels composant la distribution Linux de Mandriva. Gemo a étudié l’utilisation du modèle ActiveXML pour spécifier, mettre en œuvre et optimiser de tels systèmes [CI6].

**WebSI (2002-2003):** “Data Centric Web Services Integrator”, projet européen FP5 (<http://www.ibm.com/websi/html/home.htm>). J’ai collaboré avec le groupe de chercheurs de Politecnico di Milano, dirigée par Prof. Stefano Ceri. Nous avons développé une nouvelle méthodologie de modélisation et de déploiement automatique d’applications Web utilisant des services Web et permettant la mise en place des processus applicatifs [RI5, RI6, RI7, T4]. Ces résultats ont été intégrés dans le produit commercialisé par la compagnie WebRatio (<http://www.webratio.com>).

### Projets en collaboration avec un centre de recherche étranger:

**Van Gogh (2005-2007):** “Programme d’Action Integre (PAI)” en collaboration avec l’Université d’Amsterdam. Le projet a porté sur des aspects théoriques et de performance liés aux traitements sur des données XML [W3].

**XClean (2004-2007):** “Programme d’Action Intégrée (PAI)”, en collaboration avec V. Benzaken de l’U. Paris XI) et l’Institut d’Ingénierie des Systèmes et Ordinateurs de Portugal ([www.inesc-id.pt](http://www.inesc-id.pt)). *J’ai été responsable scientifique du côté français.*

Dans des applications telles que la gestion de données scientifiques, expérimentales, bibliographiques etc., un même objet peut être représenté dans des formats XML différents, avec des erreurs et/ou omissions. Il est alors nécessaire de pouvoir *nettoyer les données XML*, ce qui consiste à en extraire les informations correctes et non-redondantes. XClean propose une algèbre et un prototype pour le nettoyage de données XML en s’appuyant sur des requêtes XQuery [CI4, D3].

#### 6.4 Actions de valorisation, brevets, logiciels/matériels diffusés, autres réalisations.

J’ai contribué en 1998-2000 à la réalisation de la plate-forme d’intégration de données *LeSelect*, en développant un optimiseur de requêtes distribuées utilisant des données et des programmes de traitement des données (fonctions) relationnelles. Ma participation était estimée à 18% en 2001 lorsque j’ai quitté le projet. Par la suite, *LeSelect* a été transféré à la start-up *Mediance*, acquise par *BusinessObjects*.

#### 6.5 Administration liée à la recherche (coordinateur de projet, chef d’équipe/labo, etc.)

Je suis la responsable des projets ANR *Codex* et *Digiteo EDIFlow*, ainsi que la responsable du Lot 5 (pair-à-pair) du projet *RNTL WebContent*. J’ai été la responsable pour *Gemo* des projets PAI *XClean* et *ACI MDP2P*.

## 7. Encadrement

Ma participation à l’encadrement est indiquée en % entre parenthèses.

*Stages au niveau M2 et thèse:*

1. Ravi Vijay (U. Californie à San Diego, 2007): *Materialized views in peer-to-peer systems*. Ravi a effectué son stage d’été après son Master (100%).
2. Melanie Weis (U. Technique de Berlin, 2004-2005): *Nettoyage déclaratif de données XML*. Melanie a effectué son stage dans le cadre de sa thèse, qui imposait une visite de 6 mois dans un laboratoire étranger [CI4, D3] (100%).
3. Nicolaas Ruberg (MS, U. Rio de Janeiro, Brésil, 2003-2004): *Optimisation de la matérialisation de documents ActiveXML*. Ce travail a conduit à la publication de [CI9, RT1] (80%).
4. Gabriela Ruberg (U. Rio de Janeiro, 2004): *Modèle de coût pour ActiveXML*. Gabriela a effectué sa thèse entre l’Université Fédérale de Rio de Janeiro et l’INRIA. Je lui ai proposé un travail qui a pris 2-3 mois de son année chez *Gemo*, et a conduit à la publication de [CI9] (100%).
5. Andrei Arion (DEA I3, U. Paris XI, 2004): *Optimisation de requêtes XML à base d’index*, co-encadré avec V. Benzaken (U. Paris XI) (50%).

6. Nicoleta Preda (DEA I3, U. Paris XI, 2004): *Entrepôts de données en pair-à-pair*, co-encadré avec S. Abiteboul (mars-août 2004). Le stage a donné lieu à la publication de [W6] (50%).

#### *Stages de fin d'études d'ingénieur:*

1. Thai-Tho Nguyen (IFIPS Hanoi, Vietnam, 2006): *Choix automatiques de vues matérialisées dans ULoad* (80%).
2. Asmae Fahoum (Ecole Mohammadia d'Ingenieurs, Rabat, Maroc, 2006): *Librairie d'opérations élémentaires de mises à jour de vues matérialisées XML* (80%).
3. Andrei Arion (U. Polytechnique de Bucarest, Roumanie, 2003): *Traitement de requêtes XML dans le système XQueC*. Le stage a donné lieu aux publications [CI7, CN4, D8]. Andrei a obtenu le prix SIGMOD Undergraduate Award 2004 pour les réalisations de son stage (100%).
4. Cosmin Cremarencu (U. Polytechnique de Bucarest, Roumanie, 2003): *Pair ActiveXML sur un PDA*. Ce travail a contribué à la démonstration [D9] (50%).
5. Florin Drăgan (U. Polytechnique de Bucarest, Roumanie, 2003): *Replication et mises à jour dans ActiveXML*. Ce travail a contribué à la démonstration [D9] (50%).
6. Nicoleta Preda (U. Polytechnique de Bucarest, Roumanie, 2003): *Traitement de requêtes distribuées dans ActiveXML*. Ce travail a contribué aux publications [CI8, D9] (50%).
7. Lucian Precup (U. Polytechnique de Bucarest, Roumanie, 2001): *Intégration d'un module de re-écriture de requêtes dans Le Select* (50%).
8. Dan Olteanu (U. Polytechnique de Bucarest, Roumanie, 2000): *Re-écriture de requêtes dans le système Agora*. Ce travail a contribué à la démonstration [D11] (50%).

#### *Thèses:*

1. Konstantinos Karanasos (depuis 2009): *Interrogation techniques for XML and Semantic Web data in peer-to-peer networks*.
2. Wael Khemiri (depuis 2008): *Workflows interactifs et efficace pour des données scientifiques*, co-encadré avec V. Benzaken (U. Paris XI) et J.-D. Fekete (INRIA Aviz) (40%).
3. Spyros Zoupanos (depuis 2006): *Gestion de données distribuées en pair-à-pair*, co-encadré avec S. Abiteboul (INRIA Gemo) [CI2, D1, D2, W1] (80%).
4. Andrei Arion (2004-2007): *Modules d'accès XML: vers l'indépendance physique dans les bases de données XML*, co-encadré avec V. Benzaken (U. Paris XI)[RI3, RI4, CI3, CI5, CI7] (50%).
5. Nicoleta Preda (2004-2008): *Construction et gestion d'entrepôt de ressources du Web*, co-encadrée avec S. Abiteboul (INRIA Gemo) [CI1, D4, D6, W5, CN3] (50%).

#### *Post-doc:*

1. Daniele Braga (Politecnico di Milano, 2006): *Algorithmes de mise à jour de vues matérialisées XML* (100%).

## 8. Synthèse de travaux et description de principaux résultats

L'essentiel de mes travaux de recherche s'organise autour de l'*optimisation de requêtes pour des données XML*, en poursuivant notamment les trois axes suivants: optimisation du stockage de données XML; gestion efficace de données XML distribuées; intégration de la compression de données dans un entrepôt XML. Ces trois axes sont développés par la suite, mettant en évidence les résultats les plus importants obtenus.

### 8.1 Optimisation du stockage de données XML

Ce travail a été au cœur de la thèse d'Andrei Arion.

Les architectures des systèmes d'interrogation de données XML peuvent être classifiées en deux catégories, selon la manière dont les données sont gérées. Une première classe de systèmes prend en charge le stockage persistant des données, ce sont les systèmes couramment appelés "bases de données XML" par analogie avec les bases de données relationnelles classiques. Une deuxième classe de systèmes ne prend pas en charge le stockage persistant des données, se contentant de charger (une représentation) des documents XML en mémoire chaque fois qu'il faut en interroger le contenu. Les bases de données XML présentent les avantages de la persistance et de la robustesse. En échange, ils sont comparativement "lourds", et posent la question du modèle et du format de stockage utilisé à l'intérieur de la base, puisque le modèle XML lui-même, irrégulier et très verbeux, n'est pas approprié pour un modèle de stockage.

Des systèmes différents ont proposé des modèles de stockage guidés soit par le schéma du corpus XML lorsqu'il existe, soit par les caractéristiques statistiques de ce corpus. En complément des modèles de stockage, des schémas d'indexation ont été proposés, basés sur les chemins rencontrés dans les arbres XML, sur les valeurs présentes dans les feuilles etc. De manière générale, ces modèles de stockage et d'indexation souffrent d'une trop grande *rigidité*, qui prescrit telles ou telles structures à matérialiser sur disque sans donner suffisamment de liberté de définir les structures les plus adaptées à des jeux de données et requêtes potentiellement très variés.

Dans ce contexte, nous avons proposé un modèle générique de structures de stockage, appelé XML Access Modules ou XAM, permettant de décrire une classe très large de structures de stockage, indexation, et caches (ou vues matérialisées) XML précédemment proposées, ainsi que de nombreuses autres structures. Le modèle XML est un modèle de motifs d'arbres, enrichi avec la possibilité d'utiliser: des arêtes optionnelles; des attributs spécifiant ce qui est stocké pour chaque noeud; et le regroupement (imbrication) des données correspondant à certains noeuds, dans les données correspondant à leurs ancêtres. Fait essentiel pour son usage, le modèle XAM a une sémantique bien définie dans une algèbre relationnelle imbriquée [W4].

En nous appuyant sur ce modèle, nous avons décrit un algorithme de re-écriture de requêtes XQuery à l'aide de vues matérialisées, qui utilise des contraintes structurelles sur le document XML interrogées [CI3, CI5]. Ces contraintes sont exprimées par un résumé structurelle (ou DataGuide), constituant une alternative intéressante aux schémas XML, et pouvant être construit à bas coût lors du chargement de données dans la base. Cet algorithme comporte deux étapes essentielles:

- Décomposition algébrique d'une requête exprimée dans un sous-ensemble conjonctif de XQuery, vers (i) un ensemble de XAMs maximales, i.e., on obtient des motifs d'arbre aussi peu nombreux et aussi larges que possible et (ii) une expression algébrique décrivant les opérations algébriques à effectuer sur ces XAMs afin d'obtenir le résultat de la requête. Cette étape ne



prend pas en compte les connaissances sur la structure des données [CI5].

- *Re-écriture de chaque XAM de la requête, considérée comme une requête motif d'arbre, à l'aide des XAMs décrivant les structures persistantes disponibles:* stockage proprement dit, index, ou vues matérialisées. Cet algorithme s'appuie sur les connaissances structurelles fournies par le DataGuide, qui permettent de proposer certaines re-écritures qui ne seraient pas valides dans l'absence de ces connaissances. De par l'usage du modèle XAM pour décrire le stockage, les index, et les vues matérialisées, leur traitement est complètement unifié. Ainsi, l'algorithme de re-écriture peut être vu comme effectuant la sélection du meilleur chemin d'accès dans des bases de données XML [CI3].

Le modèle XAM et l'algorithme de re-écriture ont été mis en œuvre dans la plate-forme ULoad [D5]. Afin d'en démontrer la généralité, ULoad a été intégré à deux types de stockage: un modèle relationnel simple dans Postgres, et un stockage "natif" développé à l'aide de la librairie BerkeleyDB. Récemment, nous avons rajouté à ULoad la possibilité d'exploiter les capacités de serveurs de bases de données récents, tels que Oracle 12, d'interroger à la fois des structures tabulaires et des structures XML à l'aide de l'extension SQL/XML. Dans tous les cas, les structures de stockage sont décrites à l'optimiseur par des XAM.

## 8.2 Gestion efficace de données XML distribuées

La distribution des données pose des défis supplémentaires pour leur exploitation. Le défi principal lié à la performance est dû aux transferts de données qui peuvent s'avérer nécessaires pour répondre à une requête portant sur des données de plusieurs sites. A un niveau plus conceptuel, il est généralement souhaitable de gérer la distribution de manière transparente, à savoir, ne pas imposer à l'utilisateur de connaître l'emplacement des données dans le système, ni ceux de la distribution du traitement des requêtes sur plusieurs sites.

Le développement des technologies XML et en particulier des services Web, ainsi que des plateformes d'échange de données en pair-à-pair, offrent des ingrédients utiles pour la mise en place de nouvelles plateformes distribuées pour la gestion de données. Mon travail de recherche dans cette direction a étudié deux thématiques, détaillées par la suite.

*KadoP: gestion efficace des documents XML à l'aide de DHTs*

Ce travail a été au cœur de la thèse de N. Preda.

Une classe importante de systèmes d'échange d'information en pair-à-pair est celle s'appuyant sur une *distributed hash table* ou DHT, qui fournit une interface abstraite basée sur des opérations *put(clé, valeur)* et *get(clé)*, tout en distribuant équitablement le contenu de cette table de hash virtuelle entre les pairs du réseau. La conception d'un système de gestion de données basé sur une DHT commence par le choix d'un modèle d'indexation, à savoir les clés qui décriront les données du système, et les valeurs à associer à ces clés.

Nous avons fait le choix de la transparence à la localisation des données, nous étudions donc des requêtes ne spécifiant pas le(s) sites où les données pourraient se trouver. Dans ce contexte, une première fonction de l'index distribué sur la DHT est de permettre la localisation des sources de données intéressantes pour la requête. Nous souhaitons que l'index soit *complet*, c'est-à-dire qu'il fournisse au moins l'ensemble complet des sources de données utiles pour une certaine requête. L'index est *précis* pour un langage d'interrogation cible, s'il retourne exactement les documents utiles pour une requête formulée dans ce langage.

Nous avons conçu et développé le système *KadoP*, le premier système de gestion de données XML basé sur une DHT, utilisant un index complet et précis pour des requêtes de motifs d'arbres conjonctifs<sup>1</sup> [D6, CN3]. *KadoP* se distingue des autres plate-formes XML basées sur des DHT aussi par le fait que c'est un système complètement développé et testé sur des tailles importantes de données (dizaines de milliers de documents). Comme nous en avons fait l'expérience, le développement d'un système réel pose des problèmes de performance très sérieux, qu'une plate-forme partiellement simulée ne permet pas de détecter. Ces problèmes concernent principalement: les communications bloquantes de la DHT, et la gestion peu efficace des composantes locales de l'index distribué. Nous avons proposé des solutions à ces problèmes et validé le passage à l'échelle du système [CI1].

### *Optimisation pour ActiveXML*

Ce travail a été au cœur des deux premières années de la thèse de Spyros Zoupanos.

Le langage ActiveXML a été initialement proposé en 2001 dans l'équipe Gemo, juste avant mon arrivée dans le groupe. Il a depuis fait l'objet de nombreux travaux de recherche, et plusieurs plate-formes le supportant ont été réalisées. *J'ai été impliquée dans la conception et le développement des plate-formes pour PC v1.0 [D10] et v2.0 [D2], et de celle pour PDA [D9].*

Le langage ActiveXML sert à spécifier des documents *intensionnels*, dont le contenu est défini à l'aide d'appels de services Web. Lorsqu'un service est invoqué, sa réponse vient s'ajouter au document, à côté de l'appel. Un document ActiveXML peut ainsi être vu comme une spécification d'un calcul à effectuer sur les données. Un document ActiveXML est stocké sur un *pair ActiveXML*, capable de comprendre et d'effectuer des appels de services. La première plate-forme d'évaluation de documents ActiveXML proposait une *stratégie d'évaluation naïve*, où tous les appels de service étaient effectués depuis le pair où le document était stocké. Les appels contenus dans un même document sont évalués tels quels, dans un ordre spécifié par la syntaxe ActiveXML employée.

Dans ce contexte, ma recherche a contribué à identifier *des techniques d'évaluation optimisée* de documents ActiveXML. Ainsi, un modèle de *cache et réplication* pour des données intensionnelles et une *stratégie optimisée d'exécution sur des données distribuées* sont décrits dans [CI10]. Une première *technique d'évaluation paresseuse* de requêtes sur des documents ActiveXML, restreignant les appels de services à activer à ceux qui produisent des résultats utiles pour la requête a été proposée dans [CI8]. D'autres techniques d'évaluation optimisée ont fait l'objet d'autres travaux dans l'équipe et ailleurs.

Par la suite, nous avons établi un premier *modèle algébrique de gestion de données distribuées basé sur ActiveXML* [CI6]. Ce modèle part d'un sous-langage dit "positive ActiveXML" ayant de bonnes propriétés (telles qu'existence et unicité d'un point fixe etc.) et lui rajoute des services élémentaires (prédéfinis), ou *primitives d'évaluation*. En s'appuyant sur ces services, nous avons formulé des *règles d'équivalence algébriques* qui peuvent être utilisées pour re-écrire un document ActiveXML dans un autre, qui aura le même contenu à la fin de l'évaluation que le premier, mais dont l'évaluation est potentiellement plus performante. Le modèle a été raffiné, des *stratégies d'optimisation* (décrivant comment les règles d'optimisation peuvent s'appliquer) ont été proposées, évaluées et mises en œuvre dans le cadre de la thèse en cours de S. Zoupanos [CI2, D2].

Les résultats de ma recherche sur l'optimisation dans des systèmes XML distribués peuvent être vus comme se situant à deux niveaux: celui de l'infrastructure de stockage et calcul (*KadoP*), et celui de l'optimisation algébrique à plus haut niveau (optimisation ActiveXML). *Les deux niveaux sont complémentaires, et nous avons récemment démontré les bénéfices de leur intégration dans*

*la plate-forme distribuée en pair-à-pair WebContent [D1].* Ici, l'optimiseur ActiveXML permet de choisir comment mieux utiliser différents index XML en pair-à-pair (nous avons utilisé KadoP et PathFinder, développé à l'U. Versailles) afin de traiter des requêtes sur un entrepôt distribué de données et connaissances.

### 8.3 Compression de données dans les bases de données XML

Ce travail s'est déroulé dans le cadre du stage de Master et de la thèse d'Andrei Arion.

Le format XML a suscité de l'intérêt dans les bases de données pour sa flexibilité et sa grande popularité. Une bonne partie des problèmes de performance posée par la gestion des données XML dans un environnement de type "serveur" est due à la redondance du format XML (un même fragment XML peut apparaître plusieurs fois) et à son caractère verbeux (la présence des noms d'attributs et d'éléments répétés). La compression des données ayant été prouvée utile pour améliorer les performances de serveurs de données relationnelles, plusieurs travaux ont exploré son applicabilité au domaine des documents XML.

Les techniques issues de ces recherches peuvent être différenciées selon leur degré d'intégration avec le processeur de requêtes XML. A un extrême, un document peut être compressé de manière très compacte, mais totalement opaque pour le processeur de requêtes. Le traitement de requêtes s'en trouve allourdi, car avant le traitement habituel il faut décompresser le document. Des techniques plus légères réduisent la place des noms d'éléments etc. mais laissent intacte la structure du document. Ceci permet de manipuler un document de moindre taille, mais ne permet pas d'exploiter les répétitions dans la structure du document, ni pour réduire davantage sa taille, ni pour simplifier le traitement de requêtes.

Nous avons proposé *un modèle de compression pour des documents XML, une architecture, et développé un prototype*, appelé XQueC, prenant le parti de rajouter la compression dans une architecture "serveur classique" de bases de données XML. Techniquement, cela consiste à compresser, d'un côté, la structure du document à l'aide d'un résumé structurel (ou DataGuide), et de l'autre, à *explorer un espace d'alternatives pour choisir le meilleur algorithme de compression des valeurs (textes ou mots) rencontrées sur différents chemins dans le document.* Le meilleur choix relève d'un compromis: affecter un algorithme donné à un ensemble de valeurs très uniforme rend la compression très efficace, en même temps la recherche de tels sous-ensemble optimaux a un coût prohibitif, ce qui motive l'usage d'heuristiques. Une autre considération à prendre en compte concerne les opérations à appliquer sur les valeurs situées sur un certain chemin: certains algorithmes de compression permettent l'évaluation de certaines opérations (p.ex. comparaison d'égalité, d'inégalité etc.) directement sur les valeurs compressées, d'autres non. Notre modèle a été décrit pour la première fois dans [CI7] et le système démontré dans [D8]. Nous avons mieux exploré les problèmes techniques posés par la compression des données dans [RI4], et nous avons décrit des bonnes propriétés des résumés structurels dans les architectures modernes de serveurs XML dans [RI3].

## 9. Publications

### 9.1 Tableau récapitulatif de toutes les publications

Chapitres de livres	1
Articles dans des revues internationales avec comité de lecture	9
Articles dans des revues nationales avec comité de lecture	2
Articles dans des conférences internationales avec comité de lecture	15
Articles dans des workshops internationaux avec comité de lecture	7
Démonstrations dans des conférences internationales avec comité de lecture	11
Tutoriels dans des conférences et écoles d'été internationales	6
Articles dans des conférences nationales avec comité de lecture	9
Posters, articles courts, communications orales dans des conférences	13
Rapports de recherche, articles soumis	6

### 9.2 Chapitres de livres

**CL1** S. Abiteboul, O. Benjelloun, I. Manolescu, T. Milo et R. Weber: “Active XML: a data-centric perspective on Web services”, chapitre du livre “Web Dynamics: Adapting to Change in Content, Size, Topology and Use”, ISBN 3-540-40676, Springer-Verlag, octobre 2003, pp. 275-300.

### 9.3 Articles dans des revues internationales avec comité de lecture

- RI1** I. Manolescu, L. Afanasiev, A. Arion, J. Dittrich, S. Manegold, N. Polyzotis, K. Schnaitter, P. Senellart, S. Zoupanos et D. Shasha: “The repeatability experiment of SIGMOD 2008”, dans *SIGMOD Record* vol. 37, no. 1, pp. 39-45, mars 2008.
- RI2** P. Michiels, I. Manolescu et C. Miachon: “Toward microbenchmarking XQuery”, dans *Information Systems Journal*, vol. 33, no. 2, pp. 182-202, avril 2008.
- RI3** A. Arion, A. Bonifati, I. Manolescu et A. Pugliese: “Path summaries and path partitioning in modern XML databases”, dans *World Wide Web Journal*, vol. 11, no. 1, pp. 117-151, mars 2008.
- RI4** A. Arion, A. Bonifati, I. Manolescu et A. Pugliese: “XQueC: a query-conscious compressed XML database”, dans *ACM Transactions on Internet Technologies (TOIT)*, vol. 7, no. 4, novembre 2007.
- RI5** M. Brambilla, S. Ceri, P. Fraternali et I. Manolescu: “Process modelling in Web applications”, dans *ACM Transactions on Software Engineering and Methodology*, vol. 14, no. 4, octobre 2006.
- RI6** I. Manolescu, M. Brambilla, S. Ceri, S. Comai et P. Fraternali: “Model-driven design and deployment of service-enabled Web applications”, dans *TOIT*, vol. 5, no. 3, août 2005.

- RI7** M. Brambilla, S. Ceri, S. Comai, P. Fraternali et I. Manolescu: “Specification and design of workflow-driven hypertexts”, dans *Journal of Web Engineering*, vol. 1, no. 2, pp. 163-182, avril 2003.
- RI8** M. Brambilla, S. Ceri, S. Comai, P. Fraternali et I. Manolescu: “Model-driven specification of Web service composition and integration with data-driven Web applications”, dans *Data Engineering Bulletin*, vol. 25, no. 4, pp. 53-60, décembre 2002.
- RI9** A. Schmidt, F. Waas, M. Kersten, D. Florescu, M. Carey, I. Manolescu et R. Busse: “Why and how to benchmark XML databases”, dans *SIGMOD Record*, vol. 3, no. 3, pp. 27-32, septembre 2001.

#### 9.4 Articles dans des revues nationales avec comité de lecture

- RN1** A. Arion, A. Bonifati, I. Manolescu et A. Pugliese: “Un modèle de stockage de données XML basé sur les séquences”, dans *Ingénierie des Systèmes d’Information (ISI)*, vol. 10, no. 2, pp. 9-37, 2005.
- RN2** I. Manolescu, L. Bouganim, F. Fabret et E. Simon: “Interrogation efficace de ressources distribuées dans des systèmes de médiation”, dans *Technique et science informatiques (TSI)*, vol. 22, no. 10, pp. 1271-1296, 2003.

#### 9.5 Articles dans des conférences internationales avec comité de lecture

- CI1** S. Abiteboul, I. Manolescu, N. Polyzotis, N. Preda et C. Sun: “XML processing in DHT networks”, dans la conférence *International Conference on Data Engineering (ICDE)* 2008, pp. 606-615.
- CI2** S. Abiteboul, I. Manolescu et S. Zoupanos: “OptimAX: Optimizing Distributed ActiveXML Applications”, dans la conférence *International Conference on Web Engineering (ICWE)* 2008, pp. 299-310.
- CI3** A. Arion, V. Benzaken, I. Manolescu et Y. Papakonstantinou: “Structured materialized views for XML queries”, dans la conférence *Very Large Databases (VLDB)* 2007, pp. 87-98.
- CI4** M. Weis et I. Manolescu: “Declarative XML data cleaning with XClean”, dans la conférence *Advanced Information Systems Engineering (CAISE)* 2007, pp. 96-110.
- CI5** A. Arion, V. Benzaken, I. Manolescu, Y. Papakonstantinou et R. Vijay: “Algebra-Based tree pattern extraction in XQuery”, dans la conférence *Flexible Query Answering Systems (FQAS)*, LNAI no. 4027, 2006, pp. 13-25.
- CI6** S. Abiteboul, I. Manolescu et E. Taropa: “A framework for distributed XML data management”, dans la conférence *Extending Database Technologies (EDBT)*, LNCS no. 3896, 2006, pp. 1049-1058.
- CI7** A. Arion, A. Bonifati, G. Costa, S. D’Aguanno, I. Manolescu et A. Pugliese: “Efficient query evaluation over compressed XML data”, dans la conférence *Extending Database Technologies (EDBT)* 2004, pp. 200-218.

- CI8** S. Abiteboul, O. Benjelloun, B. Cautis, I. Manolescu, T. Milo et N. Preda: “Lazy query evaluation for Active XML”, dans la conférence *ACM Special Interest Group on Management of Data (SIGMOD)* 2004, pp. 227-238.
- CI9** N. Ruberg, G. Ruberg et I. Manolescu: “Towards cost-based optimization for data-intensive Web service computations”, dans la conférence *Braslian Database Symposium (SBBD)* 2004.
- CI10** S. Abiteboul, A. Bonifati, G. Cobéna, I. Manolescu et T. Milo: “Dynamic XML documents with distribution and replication”, dans la conférence *SIGMOD* 2003, pp. 527-538.
- CI11** A. Schmidt, F. Waas, M. Kersten, M. Carey, I. Manolescu et R. Busse: “XMark: a benchmark for XML data management”, dans la conférence *VLDB* 2002, pp. 974-985.
- CI12** I. Manolescu, L. Bouganim, F. Fabret et E. Simon: “Efficient querying of distributed resources in mediator systems”, dans la conférence *Cooperative Information Systems (CoopIS)* 2002, pp. 468-485.
- CI13** I. Manolescu, D. Florescu et D. Kossmann, “Answering XML queries over heterogeneous data sources”, dans la conférence *VLDB* 2001, pp. 241-250.
- CI14** D. Florescu, I. Manolescu et D. Kossmann: “Integrating keyword search into XML query processing”, dans conférence *World Wide Web Conference (WWW)* 2000, paru dans *Computer Networks* no. 33, vol. 1-6, pp. 119-135.
- CI15** D. Florescu, A. Levy, I. Manolescu et D. Suciu: “Query optimization in the presence of limited access patterns”, dans la conférence *SIGMOD* 1999, pp. 311-322.

### 9.6 Articles dans des workshops internationaux avec comité de lecture

- W1** I. Manolescu et S. Zoupanos: “XML materialized views in P2P”, dans le workshop *Database Technologies for Handling XML Information on the Web*, 2009.
- W2** I. Manolescu, C. Miachon et P. Michiels: “Towards micro-benchmarking XQuery”, dans le workshop *Experimental Evaluation of Data Management Systems (EXPDB)*, 2006.
- W3** L. Afanasiev, I. Manolescu et P. Michiels: “MemBeR: A micro-benchmark repository for XQuery”, dans le workshop *XML Symposium* 2005.
- W4** A. Arion, V. Benzaken et I. Manolescu: “XML Access Modules: towards physical data independence in XML databases”, dans le workshop *XQuery Implementation, Experience and Perspectives (XIME-P)* 2005.
- W5** S. Abiteboul, I. Manolescu, B. Nguyen et N. Preda: “A test platform for the INEX heterogeneous track”, dans le workshop *Information Retrieval for XML (INEX)*, 2004.
- W6** S. Abiteboul, I. Manolescu et N. Preda: “Constructing and querying peer-to-peer warehouses of XML resources”, dans le workshop *Semantic Web and Databases*, en coopération avec la conférence *VLDB*, 2004.

**W7** S. Comai et I. Manolescu: “Conceptual modeling issues in Web applications enhanced with Web services”, dans le workshop *E-Services and the Semantic Web*, en coopération avec la conférence WWW, 2003.

### *9.7 Demonstrations de logiciels dans des conférences internationales avec comité de lecture*

- D1** S. Abiteboul, T. Allard, P. Chatalic, G. Gardarin, A. Ghitescu, F. Goasdou, I. Manolescu, B. Nguyen, M. Ouazara, A. Somani, N. Travers, G. Vasile et S. Zoupanos. “WebContent: Efficient P2P Warehousing of Web Data”, dans la conférence *VLDB* 2008.
- D2** S. Abiteboul, I. Manolescu et S. Zoupanos. “OptimAX: efficient support for data-intensive mash-ups”, dans la conférence *International Conference on Data Engineering (ICDE)* 2008, pp. 1564-1567. Cette démonstration a également été présentée à *BDA* 2007.
- D3** M. Weis et I. Manolescu: “XClean in Action”, dans la conférence *Innovative Database Research (CIDR)* 2007, pp. 259-262.
- D4** B. Butnaru, F. Dragan, G. Gardarin, I. Manolescu, B. Nguyen, R. Pop et N. Preda: “P2PTester: Testing P2P Platform Performance”, dans la conférence *International Conference on Data Engineering (ICDE)* 2007, pp. 1501-1502.
- D5** A. Arion, V. Benzaken, I. Manolescu et R. Vijay: “ULoad: Choosing the Right Storage for your XML Application”, dans la conférence *VLDB* 2005, pp. 1330-1333.
- D6** S. Abiteboul, I. Manolescu et N. Preda: “Peer-to-peer Warehousing of XML Resources”, dans la conférence *International Conference on Data Engineering (ICDE)* 2005, pp. 1122-1123. Cette démonstration a également été présentée à *BDA* 2004.
- D7** M. Brambilla, S. Ceri, S. Comai, M. Dario, P. Fraternali et I. Manolescu: “Declarative Specification of Web Applications exploiting Web Services and Workflows”, dans la conférence *SIGMOD* 2004, pp. 909-910.
- D8** A. Arion, A. Bonifati, G. Costa, S. D’Aguanno, I. Manolescu et A. Pugliese: “XQueC: Pushing XML Queries to Compressed XML Data”, dans la conférence *VLDB* 2003, pp. 1065-1068. Cette démonstration a été également présentée à *BDA* 2003.
- D9** S. Abiteboul, J. Baumgarten, A. Bonifati, G. Cobéna, C. Cremarenco, F. Drăgan, I. Manolescu, T. Milo et N. Preda: “Managing Distributed Workspaces with Active XML”, dans la conférence *VLDB* 2003, pp. 1061-1064. Cette démonstration a été également présentée à *BDA* 2003.
- D10** S. Abiteboul, O. Benjelloun, I. Manolescu, T. Milo et R. Weber: “Active XML Demonstration: Peer-to-peer Data and Web Services Integration”, dans la conférence *VLDB* 2002, pp. 1087-1090.
- D11** I. Manolescu, D. Florescu, D. Kossmann, D. Olteanu et F. Xhumari: “Agora: Living with XML and Relational”, dans la conférence *VLDB* 2000, pp. 623-626.

### *9.8 Tutoriels dans des conférences et écoles d’été internationales avec comité de lecture*

- T1** I. Manolescu et S. Manegold: “Performance evaluation in database research: principles and experience”, tutoriel à la conférence *ICDE* 2008.
- T2** I. Manolescu et Y. Papakonstantinou: “XQuery midflight: emerging database-oriented paradigms and a classification of research advances”, tutoriel à la conférence *ICDE* 2005.
- T3** I. Manolescu: “XML query processing: storage and query model interplay”, tutoriel à l’école d’été *EDBT* 2004.
- T4** A. Lerner et I. Manolescu: “The nuts and bolts of DBMS construction: building your own prototype”, tutoriel à la conférence *SBBD* 2003.
- T5** S. Ceri et I. Manolescu: “Constructing and integrating data-centric Web Applications: methods, tools, and techniques”, tutoriel à la conférence *VLDB* 2003.
- T6** I. Manolescu: “Adaptive and self-tuning query processing”, tutoriel à l’école d’été *EDBT* 2002.

### *9.9 Articles dans des conférences nationales avec comité de lecture*

- CN1** S. Abiteboul, I. Manolescu et S. Zoupanos: “OptimAX: Optimizing Distributed ActiveXML Applications”, dans *Journées de Bases de Données Avancées (BDA)* 2008 (version étendue de [CI2]).
- CN2** S. Abiteboul, I. Manolescu, N. Polyzotis, N. Preda, C. Sun: “XML processing in DHT networks”, dans *BDA* 2007.
- CN3** S. Abiteboul, I. Manolescu et N. Preda: “Sharing Content in Structured P2P Networks”, dans *BDA* 2005.
- CN4** I. Manolescu, A. Arion, A. Bonifati et A. Pugliese: “Path Sequence-Based XML Query Processing”, dans *BDA* 2004.
- CN5** I. Manolescu, L. Bouganim, F. Fabret et E. Simon: “Efficient Querying of Distributed Resources in Mediator Systems”, dans *BDA* 2002 (version préliminaire de [CI13]).
- CN6** S. Abiteboul, O. Benjelloun, I. Manolescu, T. Milo et R. Weber: “Active XML: A Data-Centric Perspective on Web Services”, dans *BDA* 2002.
- CN7** I. Manolescu, D. Florescu et D. Kossmann, “Answering XML Queries over Heterogeneous Data Sources”, dans *BDA* 2001 (version étendue de [CI13]).
- CN8** D. Florescu, I. Manolescu et D. Kossmann: “Integrating Keyword Search into XML Query Processing”, dans *BDA* 2000. (version étendue de [CI14]).
- CN9** D. Florescu, A. Levy, I. Manolescu et D. Suci: “Query Optimization in the Presence of Limited Access Patterns”, dans *BDA* 1999 (version étendue de [CI15]).

### *9.10 Posters, articles courts, communications orales dans des conférences internationales et nationales*



- c1** I. Manolescu, Y. Papakonstantinou, V. Vassalos: “XML tuple algebra”, article dans *Encyclopedia of Database Systems*, à paraître chez Springer, 2008 (éditeur: S. Amer-Yahia)
- c2** D. Barbosa, I. Manolescu et J. Y. Xu: “XML benchmarking”, article dans *Encyclopedia of Database Systems*, à paraître chez Springer, 2008 (éditeur: S. Amer-Yahia)
- c3** D. Barbosa, P. Bohannon, J. Freire, C.-C. Kanne, I. Manolescu, V. Vassalos et M. Yoshikawa: “XML Storage”, article dans *Encyclopedia of Database Systems*, à paraître chez Springer, 2008 (éditeur: S. Amer-Yahia)
- c4** I. Manolescu et S. Manegold: “Performance Evaluation and Experimental Assessment - Conscience or Curse of Database Research?”, description d’une table ronde dans la conférence *VLDB 2007*, pp. 1441-1442.
- c5** D. Colazzo, F.-X. Dudouet, I. Manolescu, B. Nguyen et A. Vion: “Traiter des corpus dinformation sur le Web. Vers de nouveaux usages informatiques de l’enquête”, dans le *9e Congrès de l’Association Francaise de Sciences Politiques (AFSP)*, 2007.
- c6** A. Arion, A. Bonifati, I. Manolescu et A. Pugliese: “Path Sequence-Based XML Query Processing”, dans la conférence *WWW 2006*, pp. 1077-1078.
- c7** F.-X. Dudouet, I. Manolescu, B. Nguyen et P. Senellart: “XML Warehouse Meets Sociology”, dans *IADIS International Conference on WWW/Internet*, 2005.
- c8** I. Manolescu et Y. Papakonstantinou: “Report on the First “XQuery Implementation, Experience and Perspectives” (XIME-P) Workshop”, dans *SIGMOD Record*, vol. 33, no. 4, 2004.
- c9** Z. Ives, B. Ludäscher et I. Manolescu: “Reminiscences of Influential Papers”, dans *SIGMOD Record*, vol. 33, no. 3, septembre 2004.
- c10** M. Garofalakis, I. Manolescu, M. Mesiti, G. Mihaila, R. Schenkel, B. Thuraisingham et V. Vassalos: “What’s Next in XML and Databases?”, *EDBT Workshops 2004*, pp. 318-324.
- c11** I. Manolescu, S. Ceri, M. Brambilla, P. Fraternali et S. Comai: “Exploiting the Combined Potential of Web Applications and Web Services”, poster dans la conférence *WWW 2003*.
- c12** S. Ceri, M. Brambilla, S. Comai, P. Fraternali et I. Manolescu: “Specification and Design of Workflow-Driven Hypertexts”, poster dans la conférence *WWW 2003*.
- c13** M. Brambilla, S. Ceri, S. Comai, P. Fraternali et I. Manolescu: “Conceptual Modeling of Web Services-Enabled Web Applications”, poster dans la conférence *Conceptual Modelling / Entity-Relationship Approach (E-R) 2003*.

### 9.11 Brevets

J’ai contribué à un brevet logiciel obtenu en 2001 sur la version 1.0 du système de médiation de données distribuées LeSelect, développé dans l’équipe Caravel de l’INRIA Rocquencourt. Ma contribution était alors estimée à 18%.

### 9.12 Rapports de recherche, papiers soumis etc.

- RR1** A. Arion, V. Benzaken, I. Manolescu, R. Vijay et Y. Papakonstantinou: “Algebra-based Tree Pattern Extraction in XQuery”, rapport INRIA HAL no. 1147, 2006 (version étendue de [CI5]).
- RR2** A. Arion, Angela Bonifati, I. Manolescu et Andrea Pugliese. “Path Sequences and Path Partitioning in Modern XML Databases”, rapport technique Gemo no. 473, 2005.
- RR3** F.-X. Dudouet, I. Manolescu, B. Nguyen et P. Senellart: “Sociological Analysis of the W3C Standardization Process: XML Warehouse Meets Sociology”, rapport technique Gemo no. 369, 2004.
- RR4** N. Ruberg, G. Ruberg et I. Manolescu: “Towards cost-based optimization for data-intensive Web service computations”, rapport de recherche INRIA no. 5222, 2004 (version étendue de [CI9]).
- RR5** I. Manolescu, L. Bouganim, F. Fabret et E. Simon: “Efficient Data and Program Integration Using Binding Patterns”, rapport de recherche INRIA no. 4239, 2004 (version étendue de [CI12]).
- RR6** I. Manolescu, D. Florescu et D. Kossmann: “Pushing XML Queries inside Relational Databases”, rapport de recherche INRIA no. 4112, 2001.
- RR7** D. Florescu, A. Levy, I. Manolescu et D. Suci: “Query Optimization with Limited Access Patterns”, rapport de recherche INRIA no. 3969, 1999 (version étendue de [CI15]).