

Systemes d'information

## **Intégration et interopérabilité des systèmes d'information : un système de médiation de données basé sur l'architecture orientée service (SOA).**

Jérôme Velo<sup>a</sup> – Jacques Labiche<sup>b</sup> – Youssouf Saidali<sup>b</sup>

<sup>a</sup> Faculté de Droit, des sciences Economiques et de Gestion – Barikadimy  
Université de Toamasina – Madagascar.

<sup>b</sup> LITIS, Université de Rouen – UFR Sciences et Techniques, Site de Madrillet  
Avenue de l'Université  
76800 Saint Etienne du Rouvray – France

<sup>a</sup> jvelo5@yahoo.fr, <sup>b</sup> {prenom.nom}@univ-rouen.fr,

**RÉSUMÉ.** Les systèmes de médiation constituent une réponse architecturale pour un accès transparent à des sources de données distribuées. C'est une problématique toujours ouverte mais déjà ancienne qui a donné naissance à de nombreuses méthodologies et de nombreux systèmes et outils. Ces problèmes sont d'autant plus cruciaux lorsque les sources sont nombreuses et hétérogènes. Nous proposons un système de médiation permettant une interopérabilité entre données hétérogènes tant par leur nature que par leur source. Ce système a pour but de permettre des interactions entre utilisateur et système informatique dédié à une plate-forme d'e-Administration à Madagascar.

**ABSTRACT.** The mediation systems constitute an architectural solution for a transparent access to distributed data sources. It is always open but already old problems which gave birth too many methodologies and many systems and tools. These problems are all the more crucial when the sources are many and heterogeneous. We propose a mediation system allowing an interworking between heterogeneous data as well by their nature as by their source. The purpose of this system is to enable interactions between a user and a computerized system dedicated to a platform of E-Administration in Madagascar.

**MOTS-CLÉS :** intégration, interopérabilité, système d'information, médiation de données, Services et Web.

**KEYWORDS:** integration, interworkability, data system, data mediation, Services and Web.

---

## 1. Introduction

Les avancées considérables en matière de réseaux et de bases de données ont conduit à la multiplication et à l'expansion des systèmes d'information à grande échelle. Dans ces systèmes cohabitent des données et les services les manipulant. Le rôle des systèmes d'intégration de données est de répondre aux besoins des utilisateurs à travers des interfaces d'accès uniformes à ces sources de données. D'une part, l'intégration de sources de données n'est pas une tâche facile. En effet, les sources sont souvent hétérogènes car elles ont été définies et/ou structurées indépendamment les unes des autres et avec des sémantiques différentes. D'autre part, les progrès technologiques récents ont donné naissance à de nouvelles applications, qui se caractérisent par un accroissement important du volume des données à acquérir, stocker et traiter et par le caractère distribué et hétérogène de ces données.

Dans ce contexte, nous proposons un système adaptatif d'aide à l'intégration de données hétérogènes dans le but de créer un système agile.

---

## 2. Différentes approches de l'intégration d'applications

Aujourd'hui les différentes organisations nationales doivent faire interopérer les systèmes de bases de données afin de pouvoir réaliser des fonctions critiques. Cela pose des problèmes selon trois aspects : la distribution et l'hétérogénéité des données, l'autonomie des sources, et l'interopérabilité des systèmes qui gèrent les bases. Malgré l'émergence de nouveaux concepts, standards, outils, etc., la complexité de l'intégration d'applications hétérogènes dans une démarche d'urbanisation, de flexibilité et d'évolutivité demeure importante. Le développement des applications a eu tendance à se faire en fonction des besoins immédiats des organisations. Une nouvelle application étant créée à chaque fois qu'un nouveau besoin émergeait dans une organisation. Chaque application traite séparément les données, par conséquent, différentes applications peuvent exécuter plusieurs fois la même fonctionnalité principale, rendant ainsi le système inefficace. L'intégration d'applications est, par nature, la solution à ces problèmes. Grâce à elle, les applications « communiquent » entre elles, réutilisant ainsi les données et les fonctionnalités de chacune d'entre elles. Actuellement, il existe deux types d'approches de l'intégration d'applications : (1) approche matérielle et (2) approche virtuelle.

### 2.1. Approche matérielle

Cette technique consiste à compléter et recopier les données sources pour construire une nouvelle application qui collecte et stocke de manière définitive des informations volatiles provenant d'autres bases de données. Cette nouvelle application s'appelle

*entrepôt de données*. Un entrepôt de données est une structure informatique dans laquelle est centralisé un volume important de données consolidées à partir des diverses bases de données internes et externes d'une organisation ; il est conçu pour offrir un accès rapide à l'information stratégique nécessaire à la prise de décision. Cette technique présente des avantages au point de vue de la performance, la personnalisation des données, la gestion des versions, l'archivage et l'interrogation de ressources passives. Les entrepôts de données, bien qu'ils répondent aux besoins des organisations, ont de nombreux inconvénients car ils sont redondants [12], centralisés [6] et rigides [12], [6]. D'autres solutions ont vu le jour pour améliorer les entrepôts de données, comme les bases de données fédérées. Ce sont des bases de données qui rassemblent en une unique base de données virtuelle – c'est-à-dire qu'elle n'existe pas physiquement – un ensemble de bases de données distribuées, autonomes et hétérogènes [10]. Deux approches sont connues pour intégrer des bases de données à une base de données fédérées. La première approche, appelée *Global-As-View (GAV)* [2], consiste à définir le schéma global en fonction des schémas des sources. Cette approche est utilisée dans les systèmes TSIMMIS<sup>1</sup> [3] et Garlic [9]. La seconde, appelée *Local-As-View (LAV)* [2], consiste à définir les schémas locaux en fonction du schéma global. Cette approche est utilisée notamment dans le projet Information Manifold [7] ou PICSEL. Ces approches apportent une amélioration par rapport aux entrepôts de données néanmoins, elles présentent l'inconvénient d'être difficiles à mettre en œuvre, de repousser le rapatriement des données au dernier moment et elles ne sont pas beaucoup plus flexibles que les entrepôts de données.

En dehors de cet entrepôt de données, il existe le Système d'Information Décisionnel (SID) (Business Intelligence) qui fait l'historique des données essentielles d'une organisation. Il présente l'avantage de permettre de prendre des décisions à l'aide d'un tableau de bord et d'états de synthèse et l'inconvénient de fournir en différé la valeur des données.

## 2.2. Approche virtuelle

Dans cette approche, il existe deux méthodes :

- Système de médiation : qui est défini comme l'intégration de plusieurs sources de données distribuées et hétérogènes. Ce système contient les trois composants essentiels suivants : (i) un médiateur qui identifie les sources à interroger pour une requête globale, réécrit et optimise les requêtes puis intègre les résultats obtenus à partir des sources ; (ii) un adaptateur (wrapper) qui adapte la requête pour les sources et (iii) un mapping qui fait la mise en correspondance [11]. Ce système de médiation fournit ainsi un point d'accès unique à un ensemble de sources de données. Plusieurs projets ont

---

<sup>1</sup> C'est un acronyme de The Stanford-IBM Manager of Multiple Information Sources.

adoptés ce système de médiations de données, citons quelques uns : TSIMMIS [3], STYX et e-XMLMedia. Le système de médiation procure un accès direct aux données mais impose une migration de requêtes vers les sources.

- Architecture pair-à-pair, les systèmes pair-à-pair permettent à plusieurs ordinateurs de communiquer via un réseau, de partager simplement des objets. Dans cette architecture, il n'y a pas de différence entre serveur et client. Ces systèmes permettent de rendre les objets d'autant plus disponibles qu'ils sont fréquemment accédés, et donc répliqués sur un grand nombre de nœuds ce qui facilite l'augmentation du nombre de fichiers dans le réseau. C'est ce qu'on appelle le passage à l'échelle. Cette méthode est utilisée dans plusieurs projets tels que : (i) SenPeer [4], reposant sur une organisation des pairs en domaines sémantiques; et (ii) LeSelect [8] développé dans le cadre du projet Caravel à l'INRIA, permettant aux utilisateurs de publier de manière transparente leurs ressources.

Ces approches virtuelles présentent des avantages tels que la non transformation de données et des désavantages tels qu'une perte de performance, la traduction de requêtes, la dépendance aux capacités des sources.

### **2.3. Architecture orientée service (ou SOA)**

Il existe une nouvelle approche d'urbanisation logicielle, appelée architecture orientée service (Service Oriented Architecture : SOA), permettant de construire une architecture de communication entre applicatifs hétérogènes. SOA est une forme d'architecture de médiation reposant sur l'assemblage de services proposés par différentes applications. Elle permet l'intégration des applications et assure aussi l'interopérabilité de ces applications par l'intermédiaire de ses éléments principaux : le *service*, l'*Enterprise Service Bus* et les *Services Web*. Les différents composants logiciels sont connectés avec un couplage lâche<sup>2</sup>. Selon Xavier Fournier-Morel [5] « *Un service, au sens SOA, met à disposition d'acteurs (humains ou logiciel) intervenants dans des processus métiers, un accès vers une ou plusieurs fonctions métiers* ». Un « *Enterprise Service Bus ou ESB* » est une solution d'intégration fournissant des services ainsi qu'une interopérabilité accrue par l'utilisation systématique des standards comme XML, les Web Services et les normes WS-\*<sup>3</sup>. Un « *Service Web* » est un programme informatique interopérable qui est fondé sur trois concepts : (i) SOAP (Simple Object Access Protocol) : c'est une normalisation des échanges de données utilisant les protocoles de communication HTTP/SMTP/POP ; (ii) WSDL (Web Services Description Language) : c'est une description en XML des Services Web et (iii) UDDI (Universal Description, Discovery and Integration) : c'est un annuaire de services.

<sup>2</sup> Dans un couplage lâche un service n'est pas autorisé à appeler directement un autre service.

<sup>3</sup> WS-\* : Ensemble de normes associées aux Services Web

Cette approche rend très facile l'ajout de nouvelles sources de données et est plus agile [1]. Mais elle a pour inconvénient de demander beaucoup de ressources, qu'elles soient matérielles, humaines ou financières.

### 3. Notre approche

L'intégration et l'interopérabilité entre données hétérogènes sont des problèmes criants pour de nombreuses organisations. En effet, le contexte actuel de restructuration, de fusion, d'organisations publiques et privées engendre des problèmes de gouvernance et requiert la mise en œuvre de processus importants pour assurer l'interopérabilité des systèmes d'informations. Actuellement, Madagascar fait une course contre la montre pour la mise en place de systèmes d'information (SI) dans tous les départements ministériels et institutions. Ceci en l'absence d'un schéma directeur national et/ou d'un organisme public ou privé qui gère ces SI qui souvent contiennent des informations semblables mais à usages différents. De plus, vu le manque d'informaticiens à Madagascar, surtout dans le secteur public, la conception et l'implémentation de ces SI sont confiées à différents prestataires de service qui développent leur propre SI sans prendre en compte les autres.

Pour résoudre ce problème, nous avons adopté la combinaison d'une architecture orientée service (SOA), du système de médiation de données et de l'entrepôt de données pour constituer un système adaptatif d'aide à l'intégration de systèmes d'informations hétérogènes et distribués [11].

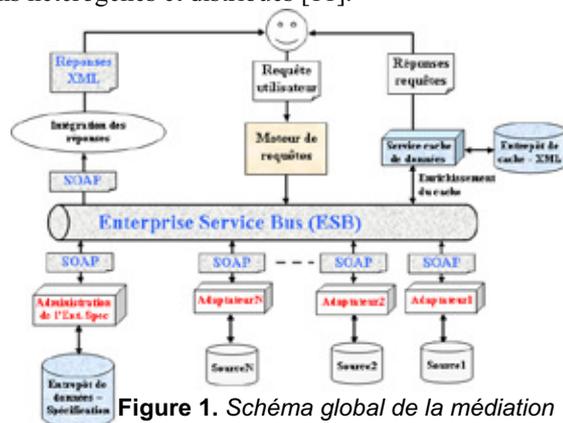


Figure 1. Schéma global de la médiation

Ce système permet l'intégration d'applications existantes et à venir ainsi que l'interopérabilité de ces applications. Son fonctionnement est piloté par l'ESB. Pour argumenter nos choix, nous présentons le rôle des trois approches choisies :

- Entrepôts de données : il s'agit d'un entrepôt de données de spécifications des termes à utiliser dans toutes les sources et d'un entrepôt de cache des réponses. Le premier stocke les informations de spécification conçues par les experts des domaines et les informaticiens, tandis que le deuxième stocke les questions déjà posées par les utilisateurs et leurs réponses.

- Système de médiation : l'introduction du système de médiation dans notre solution a pour but de procurer un accès direct pour interroger les sources d'informations par les biais du médiateur, du mapping et des adaptateurs. Ces derniers assurent la navette à l'aide du protocole SOAP entre le chef d'orchestre (ESB) et les sources. Les rôles du médiateur et du mapping sont assurés respectivement par l'ESB et les services.

- Architecture orientée service (SOA) : cette approche assure l'interopérabilité du système. Le rôle du médiateur est assuré par l'ESB, et toutes les communications de ce système utilisent la technologie Service Web par le biais de son élément SOAP. Ainsi, dès que l'ESB reçoit une requête, il consulte le service cache de données pour savoir si ce dernier a déjà une réponse correspondant à cette même demande. Sinon, l'ESB transmet la requête au service d'administration d'entrepôt de spécification qui envoie les correspondances de la requête aux sources. L'ESB répartit les requêtes aux adaptateurs adéquats. Ces adaptateurs dirigent leur requête vers les sources de destination puis transmettent les réponses à l'ESB qui appelle le service d'intégration des réponses qui les prend en charge.

## 4. Réalisation

Dans cette section, nous présentons sur un exemple les résultats des tests obtenus lors du développement du prototype. Il nous restait à résoudre le problème de la « découverte » des services. Comme les services assurent les fonctionnalités du système et que ces fonctionnalités sont obtenues à partir de l'étude des « cas d'utilisation ». Nous avons étudié le fonctionnement de notre système de médiation de données pour identifier les acteurs et les cas d'utilisation. Soit par exemple la table « EXERCICE » qui appartient à trois sources différentes et est structurée différemment suivant ces trois sources. Cette table est une table clef car elle contient l'information « année d'exercice » qui est nécessaire pour chaque début d'année. Le tableau 1 suivant présente la structure de cette table pour chaque source.

a- Source1			b- Source2		
Nom attribut	NULL ?	Type	Nom	NULL ?	Type
EXERCICE	NOT NULL	VARCHAR(4)	EXERCICE	NOT NULL	VARCHAR(4)
EXER_LIBELLE	NOT NULL	VARCHAR(30)	CODR_CEMT	NOT NULL	VARCHAR(10)
EXER_DT_DEBUT		DATE	EXER_LIBELLE	NOT NULL	VARCHAR(30)
EXER_DT_FIN	NOT NULL	DATE	EXER_DT_DEBUT		DATE
EXER_FLO		NUMBER(3)	EXER_DT_FIN		DATE
EXER_TSI		NUMBER(30)	EXER_FLO		NUMBER(3)
COLONNE1		VARCHAR(4)			
COLONNE2		VARCHAR(4)	c- Source3		
COLONNE3		VARCHAR(4)	Nom	NULL ?	Type
COLONNE4		VARCHAR(4)	EXER_ANNEE	NOT NULL	VARCHAR(4)
COLONNE5		VARCHAR(4)	EXER_LIB	NOT NULL	VARCHAR(30)
			EXER_DT_DEBUT		DATE
			EXER_DT_FIN	NOT NULL	DATE
			EXER_FLO		NUMBER(3)
			EXER_COMB		VARCHAR(255)

Tableau 1. Structure de la table EXERCICE de trois sources différentes

Supposons que, nous voulions connaître l'année et le libellé de tous les contenus de ces tables représentantes. Cette requête se formule comme suit en SQL : SELECT exercice\_annee, exercice\_libelle FROM exercice ; cette requête est décomposée pour les trois sources comme l'indique le tableau 2. ci-dessous.

Source1	Source2	Source3
SELECT exercice, exer_libelle FROM exercice;	SELECT exercice, exer_libelle FROM exercice;	SELECT exer_annee, exer_lib FROM exercice;

**Tableau 2. Réécriture de requêtes aux sources de données**

En exécutant ces requêtes, nous obtenons les résultats suivants :

Source1		Source2		Source3	
exercice	exer_libelle	exercice	exer_libelle	exer_annee	exer_lib
2007	Exer 2006	2005	Exercice 2005	2006	Exercice 2006
2008	Exer 2006	2006	Exercice 2006	2007	Exercice 2007
2009	Exer 2006	2007	Exercice 2007	2008	Exercice 2008
2010	Exer 2006	2008	Exercice 2008		
2011	Exer 2006				

**Tableau 3. Résultat des requêtes reformulées**

Notre prochaine étape consiste à intégrer ces résultats ce qui est l'objet du système de médiation conçu dans la section 3. Le système de médiation devra franchir les étapes que nous avons parcourues dans l'exemple exposé ci-dessus. Pour ce faire, nous devons concevoir le diagramme de classe [11] d'une application permettant de parcourir toutes les sources et toutes ses tables et colonnes et de les traiter si possible. L'intégration par le système de médiation du problème ci-dessus est explicitée par la figure 2 ; il s'agit de l'unification des résultats. Vu le nombre important de colonnes des tables, nous avons choisi de proposer un choix manuel des colonnes contenant les résultats à unifier.

De plus, ce système nous permet d'ajouter des nouvelles sources autant que nous voulons. L'Architecture et les technologies choisies permettent d'ajouter autant de sources que nécessaire ; ce point est primordial pour assurer un déploiement totalement ouvert aux évolutions de l'e-administration à Madagascar.



**Figure 2. L'unification et ses résultats**

---

## 5. Conclusions

L'intégration manuelle des systèmes d'informations est difficilement envisageable en présence d'un grand nombre de sources de données. En réponse à ce problème, nous avons proposé une approche hybride par le biais de la combinaison d'un entrepôt de données, d'un système de médiation et d'une architecture orientée service des systèmes d'information en tenant compte des sources des données et des préférences des utilisateurs. Dans ce contexte, nous avons développé un système simple excluant une transformation des données. Ce système est performant et personnalisable, donc il est plus agile. Il présente aussi l'avantage de la facilité d'ajout de nouvelles sources. Nos travaux actuels traitent les requêtes des utilisateurs en langage SQL. Le système produit [11] est destiné aux utilisateurs qui ont un minimum de connaissances en la matière. Nous allons étendre notre approche pour traiter les requêtes des utilisateurs sans connaissances particulières des domaines administratifs et informatiques, c'est l'objet du grand projet à mettre en place à l'issue de nos travaux actuels.

---

## 6. Bibliographie

- [1] P. Bonnet et al. – *Le système d'information durable : la refonte progressive du SI avec SOA*. – Lavosier, 312p, novembre 2007.
- [2] D. Calvanese et al. – *Data integration in data warehousing*. – Journal of Cooperative Information Systems, Vol. 10, N° 3 :pp. 237-271, 2001.
- [3] S. Chawathe et al. – *The tsimmis project: Integration of heterogenous information sources*. – Journal of Intelligent Information Systems, 8 N°2, 1997.
- [4] D. Faye et al. – *Senpeer : un système pair-à-pair de médiation de données*. Revue Africaine pour la Recherche en Informatique et Mathématiques Appliquées (ARIMA), Vol. 4 :pp. 24-52, 2006.
- [5] X. Fournier-Morel et al. – *SOA : Le guide de l'architecture du SI*. – Dunod, 2ème édition, 350p, septembre 2008.
- [6] D. Geer. – *Federated approach expands database-access technology*. Computer, Vol. 36, N° 5:pp. 18-20, May 2003.
- [7] A. Levy et al. *The information manifold*. – In AAAI Spring Symposium on Information Gathering, 1995
- [8] I. Manolescu et al. – *Le projet LeSelect*, INRIA, avril 2000.
- [9] M. Roth and P. Schwarz. – *Don't scrap it, wrap it! a wrapper architecture for legacy data sources*. In Proc. of the 23th VLDB Conference, Athens, Greece, August 1997.
- [10] A. P. Sheth and J. P. Larson. – *Federated database systems for managing distributed, heterogeneous, and autonomous databases, ACM Computing Surveys (CSUR)*. – ACM Press, New York, NY, USA, Vol. 22 :183-234, 1990.
- [11] J. Velo – *Interactions entre un utilisateur et un système informatique : application à un système d'e-Administration*. – PhD. Thesis, LITIS de l'Université de Rouen, décembre 2009.
- [12] F. Zhu et al. – *Dynamic data integration using Web services*. In Proceedings of IEEE International Conference on Web Services (ICWS'04), IEEE Computer Society, Washington, USA, pp.262-269, 2004