

# Du XML au multidimensionnel : Conception de magasins de données

Yasser Hachaichi, Jamel Feki, Hanene Ben-Abdallah

Laboratoire MIRACL, Faculté des Sciences Economiques et de Gestion de Sfax  
Route de l'Aérodrome km 4, B.P. 1088, 3018 Sfax, Tunisie  
{Yasser.Hachaichi, Jamel.Feki, Hanene.Benabdallah}@fsegs.rnu.tn

**Résumé.** Avec l'ouverture des entreprises sur l'Internet, les sources de données englobent davantage des données échangées avec les partenaires et/ou issues du Web. Dans ce cadre organisationnel ouvert, les documents XML constituent des sources de données plus utilisées aussi bien pour le stockage que pour les échanges transactionnels. Ce nouveau format de données a motivé des propositions de modèles pour les entrepôts de données XML et des démarches de conception d'entrepôts à partir de sources XML. Cependant, les démarches proposées requièrent une intense intervention qui nécessite une double expertise dans le domaine des sources XML et celui d'entreposage de données. Cet article propose une démarche automatique de conception de schémas multidimensionnels ayant un potentiel analytique à partir de sources XML. Il détaille les étapes de la démarche et l'illustre sur un exemple.

## 1 Introduction

Dans un contexte organisationnel sans cesse complexe et mouvant, les entreprises sont confrontées à une concurrence sévère et à des clients de plus en plus exigeants. Pour faire face, l'entreprise doit prévoir et planifier en s'appuyant sur l'analyse d'informations soigneusement collectées et préparées. Dans ce contexte, l'informatique décisionnelle offre des bases de données rigoureusement conçues pour ces analyses, appelées entrepôts de données (ED) et magasins de données (MD). Traditionnellement, les ED/MD sont alimentés par des informations pertinentes issues de sources de données opérationnelles, internes à l'entreprise.

Récemment, et avec l'ouverture des entreprises sur l'Internet, les sources de données englobent davantage des données échangées avec les partenaires et/ou issues du Web. Dans ce nouveau cadre organisationnel ouvert, les documents XML constituent des sources de données de plus en plus fréquentes, utilisées aussi bien pour le stockage que pour les échanges transactionnels. En effet, grâce à sa flexibilité quant à la modélisation des données, XML est devenu le langage élite pour décrire des informations et les échanger avec les partenaires. Par conséquent, il convient de disposer de mécanismes adéquats pour extraire et structurer les informations pertinentes à la prise de décision à partir de sources XML.

Face à cette problématique, plusieurs travaux ont proposé soit des modèles pour les entrepôts de données XML (*cf.*, (Hümmer et al., 2003) et (Boussaid et al., 2006)), soit des démarches semi-automatiques de conception de ED/MD à partir de sources XML (*cf.*, (Golfarelli et al., 2001), (Vrdoljak et al., 2003) et (Jensen et al., 2007)). Etant semi-automatiques, les démarches proposées requièrent une intense intervention pour