

Une étude sur l'efficacité des méthodes de conception et d'implémentation pour les Entrepôts de Données par une méthodologie « requirement-based »: Cas d'étude de la consommation d'énergie en agriculture

Sandro Bimonte*, Jean-Pierre Chanet*,
Jacques Capdeville**, Aurelie Tailleur***, Marc Luciano****

*Irstea, 9 Avenue Blaise Pascal - CS 20085 - 63178 Aubière
sandro.bimonte@irstea.fr, jean-pierre.chanet@irstea.fr

**Institut de l'Elevage - 31321 Castanet Tolosan
capdeville@idele.fr

***Arvalis - Institut du végétal -44370 La Chapelle St Sauveur
a.tailleur@arvalisinstitutduvégétal.fr

****Arvalis - Institut du végétal -Baziège
m.luciano@arvalisinstitutduvégétal.fr

Résumé. Data Warehouses and OLAP systems allow decision-makers exploring and analyzing huge volumes of data modeled according the multidimensional model, and extracted from heterogeneous data sources. Usually, DW design is a complex, and time and resources consuming task. Then, DW experts are necessary during design and implementation phases. In this paper, we present a new methodology and a tool allowing modelers (DW unskilled users) to design and implement DWs for analyzing simulation results data by themselves, without any intervention of DW experts.

1 Introduction and motivations

Aujourd'hui les Entrepôts de Données (ED) et les systèmes OLAP représentent une solution incontournable pour l'analyse décisionnelle des gros volumes de données (Kimball, 1996). Le succès des projets OLAP repose essentiellement sur la définition de schémas multidimensionnels en accord avec les besoins d'analyse des décideurs. Pour cela, plusieurs méthodologies de conception et d'implémentation ont été proposées en littérature (Romero et Abello, 2009). Ces approches peuvent être classifiées en 3 catégories : « requirements-driven » où le schéma conceptuel est défini à partir des besoins décisionnels des décideurs (i.e. besoins fonctionnels), « sources-driven » où à partir des sources de données des modèles multidimensionnels sont proposés aux utilisateurs finaux, et « mixed » où les deux méthodologies sont conduites en parallèle. Différents auteurs définissent des formalismes et/ou des méthodologies pour l'expression des besoins fonctionnels en utilisant des formalismes graphiques comme UML (Giorgini et al., 2008), et des langages de requêtes comme MDX (Niemi et al.,

2001). Le processus de développement des EDs comprend plusieurs phases différentes étroitement liées : définition des besoins fonctionnels, définition du schéma conceptuel multidimensionnel, son implémentation (dans une architecture ROLAP, MOLAP ou HOLAP) et alimentations de l'ED à partir des sources de données disponibles via des outils de type Extraction-Transforming-Loading (ETL) (Kimball, 1996). Comme pour les systèmes d'information classiques (i.e. OLTP) différentes approches ont été proposées pour la gestion de ces différentes phases (Romero et Abello, 2009) : « cascade » où une validation structurée et systématique lors de chaque étape avant de passer à la suivante est faite ; « spirale » qui est basée sur la génération rapide et incrémentale d'un prototype, etc. L'approche à spirale est souvent implémentée dans une architecture « Model-Driven Architecture » (MDA) où à partir du modèle conceptuel multidimensionnel, les modèles logiques et physiques sont automatiquement implémentés, car elle permet la génération de code exempt d'erreurs (Mazon et Trujillo, 2008). Différents auteurs ont proposé des architectures MDA, la plupart basées sur le formalisme conceptuel UML (Abello et al., 2006). Seulement (Benker et Jürck, 2012) propose une évaluation de cette méthodologie dans le cadre d'un projet OLAP réel, en termes de besoins fonctionnels et non-fonctionnels représentés par le modèle conceptuel multidimensionnel, mais les auteurs ne la comparent pas à d'autres méthodologies. Différentes métriques ont été définies en littérature pour l'évaluation de la pertinence des ED par rapport aux besoins décisionnels (da Chen et al., 2000), ou l'évaluation (testing) des EDs dans les différentes phases du processus de développement (Golfarelli et Rizzi, 2011). Par contre, aucun travail ne présente une évaluation quantitative de l'efficacité de l'approche MDA dans une méthodologie à spirale.

En conclusion, et plus en particulier, à notre connaissance aucun travail ne présente un cadre théorique pour l'évaluation des méthodologies de conception des entrepôts de données requirement driven et leur implémentation. Dans ce travail, qui s'inscrit dans le contexte de deux projets nationaux Energetic et EDEN (Bimonte et al., 2014) (Bimonte et al., 2014) nous nous intéressons à cette évaluation. Nous proposons un cadre formel qui définit les variables qui interviennent dans ce type de projets ainsi que des métriques quantitatives pour l'évaluation de leur efficacité.

L'article est organisé de la façon suivante : La section 2 présente le cadre d'évaluation ; Dans la section 3 différentes méthodologies de conception utilisées dans ces projets sont développées.

2 Cadre d'évaluation

Dans cette section nous présentons un cadre générique pour l'évaluation des méthodologies de conception et implémentation d'entrepôts de données basés sur l'approche « requirement-driven ». Il est important noter que dans cet article, nous ne prenons pas en compte la phase de ETL, mais nous considérons seulement la définition des modèles logiques aux niveaux du SGBD et du serveur OLAP. Comme surligné dans la section précédente dans les méthodologies « requirement-driven » l'utilisateur joue un rôle fondamental car elles le mettent au centre du processus de conception via les besoins d'analyse. Pour cela, les caractéristiques des utilisateurs ainsi que leur cadre de travail deviennent des facteurs primordiaux dans l'évaluation des méthodologies « requirement-driven ». En particulier, il est nécessaire de prendre en compte différents facteurs : (i) les *compétences informatiques et/ou en entrepôts de données* des utilisateurs pour étudier si elles ont un impact sur la qualité du processus de conception. L'échange entre les décideurs et les experts des entrepôts de données s'effectue avec des supports diffé-

rents (formalismes graphiques – e.g. UML, documents textuels, etc.), qui selon nous jouent en rôle fondamental sur l'efficacité du projet ; (ii) le *support d'échange*. Ces échanges sont aussi effectués avec des technologies différentes (e.g. téléphone, réunion, etc.) qui peuvent faciliter (ou non) la compréhension des besoins d'analyse des experts des entrepôts de données ; (iii) les *modalités d'échange*. En plus, il est évident que selon les domaines d'applications le processus de développement peut être plus ou moins complexe ; (iv) le *contexte d'analyse*. Par exemple dans le cadre classiques de l'analyse commerciale des magasins, l'existence de nombreux entrepôts de données peut faciliter l'identification des dimensions et faits d'analyse. En ce qui concerne l'implémentation nous allons identifier trois paramètres : (v) l'*architecture*, (vi) la *méthodologie*, et (vii) le *formalisme*. Les entrepôts de données peuvent être implantés en utilisant différentes architectures ROLAP, MOLAP, HOLAP, etc. toutes avec des avantages particuliers comme par exemple les bonnes performances en terme d'espace pour les entrepôts de données rares dans les systèmes ROLAP. Différentes méthodologies sont présentes en littérature pour l'implémentation des entrepôts de données : en cascade, en spirale, etc. Elles ont un impact important sur l'organisation des procédures de conception et d'implémentation. Enfin, la modélisation conceptuelle est reconnue nécessaire dans les systèmes d'information classiques et aussi dans les entrepôts de données. Il est par contre nécessaire d'analyser l'impact du type de formalisme utilisé (e.g. ad-hoc, UML, logique, etc.) sur le projet. Pour l'évaluation de l'efficacité des projets, nous avons identifiés des métriques associées aux deux grandes phases de développement : (i) *définition de besoins fonctionnels*, et (ii) *implémentation de prototypes*. En particulier pour la globalité du projet nous avons défini trois indicateurs :

1. **Temps** : Durée totale de la définition à la finalisation du prototype final ;
2. **Nombre d'experts ED** : Nombre de personnes affectées à temps complet (1) ou à temps partiel (0.5) à l'implémentation du prototype. Cet indicateur permet d'expliquer des principales différences car évidemment un nombre plus important de personnel se traduit par une meilleure efficacité ;
3. **Complexité du schéma multidimensionnel** : Complexité du schéma en termes de dimensions et du nombre d'indicateurs. Elle est calculée comme la somme du nombre de niveaux et de mesures. Cet indicateur permet d'expliquer des différences principales car évidemment un modèle plus complexe requière plus de temps de conception et développement.

Pour la phase définition de besoins fonctionnels :

1. **Nombre de documents** ;
2. **Nombre de réunions d'échange avec les décideurs**.

Pour la phase implémentation de prototypes

1. **Versions du schéma** : Nombre de versions du schéma conceptuel produites ;
2. **Versions du prototype** : Nombre de versions du prototype implémentées.

Nous ne proposons pas de formules de pondération pour les métriques définies car nous pensons que chaque métrique peut être expliquée indépendamment à partir des paramètres de notre cadre d'évaluation. En utilisant ces paramètres et les métriques identifiées, notre cadre d'évaluation a pour but de répondre aux questions suivantes :

1. *Quels sont les facteurs et comment ils influencent l'échange entre les décideurs et les experts des entrepôts de données ?*

Une étude sur l'efficacité des méthodes de conception et d'implémentation pour les Entrepôts de Données

2. *Quels sont les facteurs et comment ils influencent l'implémentation des entrepôts de données ?*

3 Expérimentation : le cas d'étude des projets Energetic et EDEN

L'expérimentation de notre cadre d'évaluation se fait dans le cadre des projets Energetic et EDEN (Bimonte et al., 2014). Ces projets ont comme but d'élaborer des bilans énergétiques plus fins à l'échelle de la parcelle, de l'atelier de production ou de l'opération dans l'entreprise agricole. Différents experts en ACV (Analyse de Cycle de Vie) ainsi que des agriculteurs ont participé à la conception des EDs en étroite collaboration avec les experts ED. Finalement, trois EDs ont été développés :

- L'ED « Culture » pour l'analyse des consommations énergétiques dans la production des grandes cultures, comme le blé, le colza, etc. ;(Bimonte et al., 2014)
- L'ED « Elevage » pour l'analyse des consommations énergétiques dans le contexte de la production fromagère et laitière ;(Bimonte et al., 2014)
- L'ED « Irrigation » dédié à l'analyse des besoins en énergie pour irrigation des grandes cultures (Figure 5). Ce modèle est décrit avec le profil UML pour les entrepôts de données spatiales décrit par (Bimonte et al., 2013) et comprend deux parties : la première décrit les structures multidimensionnelles SDW (SDW metamodel) et la seconde représente l'agrégation des mesures par rapport à l'analyse des besoins des décideurs (Aggregation metamodel). Le méta-modèle SDW représente de manière conceptuelle des aspects classiques et avancés du modèle spatio-multidimensionnel tels que les hiérarchies multiples et complexes, les relations many-to-many entre les faits et les dimensions, etc.. Le profil définit un stéréotype où une valeur est taguée pour chaque élément spatio-multidimensionnel. Les niveaux de dimension spatiale «SpatialAggLevel» sont représentés avec des attributs géométriques stéréotypés « LevelGeometry », et le stéréotype « AggRelationship » est utilisé pour modéliser les relations d'agrégation qui relient les niveaux de dimension.

3.1 Paramètres

Dans cette section nous décrivons les paramètres de notre cadre d'évaluation pour les trois entrepôts de données. En effet, cette évaluation a été possible car les projets Energetic et EDEN nous ont permis de définir un cadre structurant et de formuler des hypothèses (paramètres) communes aux trois EDs. Dans notre cas d'étude, les décideurs ont une bonne connaissance des données qui peuvent être récoltées et ils présentent tous le même profil : il s'agit de décideurs sans aucune formation sur les EDs et les systèmes OLAP, et des connaissances minimales en systèmes d'information. Cette hypothèse, qui peut sembler inappropriée pour les grandes organisations ou entreprises, représente la réalité des petites et moyennes entreprises. Les décideurs sont les experts métiers, i.e. les agronomes. Tous les échanges entre les décideurs et les experts ED se sont faits via des systèmes de visioconférence et/ou par téléphone vu leurs différentes localisations géographiques. Cette hypothèse représente aussi une réalité pour tout type d'organisation/entreprise vu le développement et l'adoption massive des outils du web 2.0

<i>Paramètre</i>	ED culture	ED élevage	ED irrigation
<i>Support d'échange</i>	Power Point	Word	Excel
Méthodologie	Spirale avec implémentation non automatique	Spirale avec implémentation automatique	Spirale avec implémentation automatique
<i>Formalisme</i>	Extension de ER	UML	UML

TAB. 1 – Paramètres différents

comme les wikis, etc. qui facilitent et permettent efficacement le travail participatif et collaboratif à distance (Bimonte et Kang, 2014). Dans les trois approches nous nous sommes basé sur une méthodologie à spirale comme définie dans la section 1. Il est important de souligner que dans ces projets, les données sont collectées via des réseaux de capteurs après la phase de conception des entrepôts de données (Bimonte et al., 2014), donc nous avons adopté pour les trois EDs la même méthodologie de type « requirement-driven » (Bimonte et al., 2013). En ce qui concerne les données et les indicateurs OLAP, les trois EDs présentent évidemment des différences en termes de dimensions et mesures, mais ils portent tous sur la même thématique : les dépenses énergétiques dans le contexte des productions agricoles. Cela permet aussi d'avoir une vision homogène des problématiques liées au domaine d'application et donc de la complexité du développement des EDs. Enfin, les trois ED ont été implémentés en utilisant la même architecture ROLAP avec PostgreSQL pour le stockage de données multidimensionnelles, Mondrian comme serveur OLAP et JRubik comme client OLAP. Dans la suite nous allons montrer les différences des trois projets selon les autres paramètres d'évaluation (Tableau 1).

L'ED « Culture » a été développé en suivant une approche agile avec les outils suivants : utilisation de fichiers texte décrivant le domaine d'application, et de diapositives PowerPoint pour la description des besoins d'analyse. Le modèle multidimensionnel conceptuel est donc déduit manuellement par l'expert ED et représenté manuellement avec le formalisme graphique Spatio-Multidim (Malinowski et Zimányi, 2007) à l'aide de l'outil de création de graphiques Visio Microsoft. Après plusieurs modèles conceptuels, différents prototypes d'ED ont été implémentés pour permettre la validation des schémas manuellement. En ce qui concerne la traduction des besoins fonctionnels exprimés oralement et à l'aide de PowerPoint, de nombreuses réunions ont été nécessaires pour comprendre comment ces indicateurs pouvaient être représentés dans un modèle spatio-multidimensionnel. Pour le volet élevage, nous avons défini une nouvelle méthodologie et un outil qui l'implémente, ProtOLAP (Bimonte et al., 2013). Les étapes de la méthodologie sont les suivantes :

1. Les décideurs définissent les besoins fonctionnels en utilisant des documents texte ;
2. Les concepteurs créent un schéma conceptuel multidimensionnel avec un profil UML implémenté avec le case tool MagicDraw (Bimonte et al., 2013) ;
3. Le schéma conceptuel est automatiquement traduit en un schéma logique ;
4. Les utilisateurs alimentent manuellement avec des données échantillons réalistes le prototype ;

Une étude sur l'efficacité des méthodes de conception et d'implémentation pour les Entrepôts de Données

Agriculteur	Bloc d'irrigation	Parcelle	Pompe	Type de culture	Position	Année	Conso électrique kWh	Conso d'eau m³	Indicateur kWh/m3
						Année 1	25000	50000	0,5
						Mois (1-12)			
						Mois 1	7500	15000	
						Semaine (1-52)			
					Position la plus favorable	Semaine 1	1700	3400	
						Jour (1-31)			
						Jour 1	200	400	
						Heure (1-24)			
						Heure 1	25	50	
						Année 1			
						Mois (1-12)			
						Mois 1			

FIG. 1 – Besoins fonctionnels avec Excel

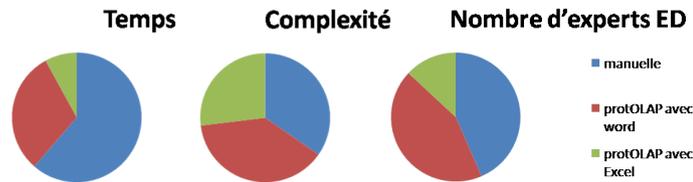


FIG. 2 – Évaluation des paramètres globaux

- Les utilisateurs accèdent et explorent ces données à l'aide de tableaux croisés dynamiques dans le client OLAP JRubik, afin de valider le prototype. Si le prototype n'est pas validé, on retourne à l'étape 1 et on commence une nouvelle itération ;
- Une fois le prototype validé et déclaré conforme aux exigences d'analyse, les données sources sont collectées, l'ETL définitif est implémenté, et l'ED est déployé.

Notons que Word a été utilisé car il permet un travail collaboratif via la possibilité de définir des versions du document ainsi que des commentaires. Un wiki aurait également pu être mis en œuvre. Pour l'entrepôt de données Irrigation nous avons adopté la méthodologie ProtOLAP, mais nous avons demandé aux experts d'exprimer leur besoins sous forme de tableau Excel et de la même façon les requêtes OLAP sont visualisées. Un exemple est montré en Figure 1.

3.2 Métriques

Pour raison de confidentialité nous montrons les valeurs des métriques en pourcentage. L'évaluation des paramètres globaux est présentée en Figure 2. Il est évident que le temps global pour la finalisation de l'ED « Irrigation » réalisée avec la méthodologie ProtOLAP plus le template Excel est le meilleur, même si le nombre d'experts est moins que les autres EDs. Par contre, la complexité des EDs est différente, surtout le ED « Irrigation » semble plus simple que les autres. Le simple comptage des éléments multidimensionnel, comme surligné dans la Sec. 3, pour cette évaluation est négligeable, car beaucoup d'indicateurs de cultures et élevage se ressemblent.

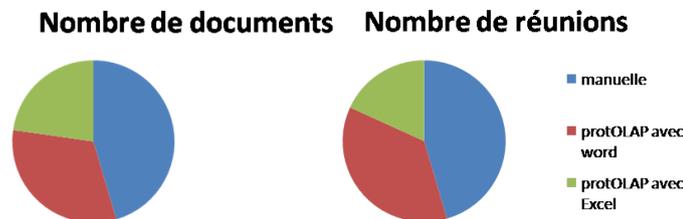


FIG. 3 – Évaluation de paramètres de la phase définition de besoins fonctionnels

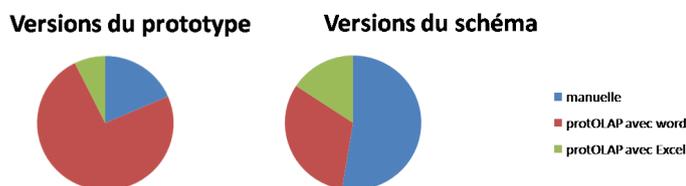


FIG. 4 – Évaluation de paramètres de la phase implémentation de prototypes

Nous considérons maintenant la phase implémentation (Figure 4).

En considérant le nombre de schémas produit, l'ED « Culture » (méthodologie manuelle) semble plus compliqué que les autres EDs qui adoptent la méthodologie ProtOLAP. Ceci s'explique par le fait qu'au début du projet le profil UML n'ayant pas encore été implémenté, des diagrammes manuels étaient produits avec des erreurs dues à l'implémentation manuelle. L'introduction du profil UML a réduit donc le nombre de schémas finaux livrés aux experts mais a augmenté le nombre de schémas intermédiaires utilisés par les experts des ED ce qui signifie que l'utilisation d'un profil UML implémenté dans un CASE tool permet une génération simple et efficace de modèles conceptuels multidimensionnels. En ce qui concerne la production des schémas logiques et physiques, évidemment, un nombre inférieur de schémas conceptuels implique moins de prototypes, ce qui est normal dans les méthodologies avec ProtOLAP. Par contre, il est important de noter que l'utilisation de ProtOLAP permet de définir plus de prototypes que pour l'approche manuelle. En effet, la traduction des besoins d'analyse et la définition du schéma conceptuel et l'implémentation du prototype sont des tâches coûteuses en temps, car l'implémentation de l'architecture ROLAP est très longue car la conception du schéma relationnel et du modèle du serveur OLAP manuellement a été quasiment toujours une source d'erreurs, surtout pour la définition de règles d'agrégation complexes comme pour les mesures semi-additives.

En fin, pour l'ED « Irrigation » (méthodologie ProtOLAP avec Excel) le nombre de schémas produits est bien inférieur aussi car une meilleure définition des besoins fonctionnels facilite l'ensemble du développement du projet, comme expliqué dans la suite. Pour l'analyse de l'efficacité de la phase définition de besoins fonctionnels (Figure 3) il est évident que l'utilisation d'un format simple à utiliser, bien structuré, et familier pour les décideurs comme Excel améliore les échanges avec les experts et facilite ainsi le développement global du projet (en

Une étude sur l'efficacité des méthodes de conception et d'implémentation pour les Entrepôts de Données

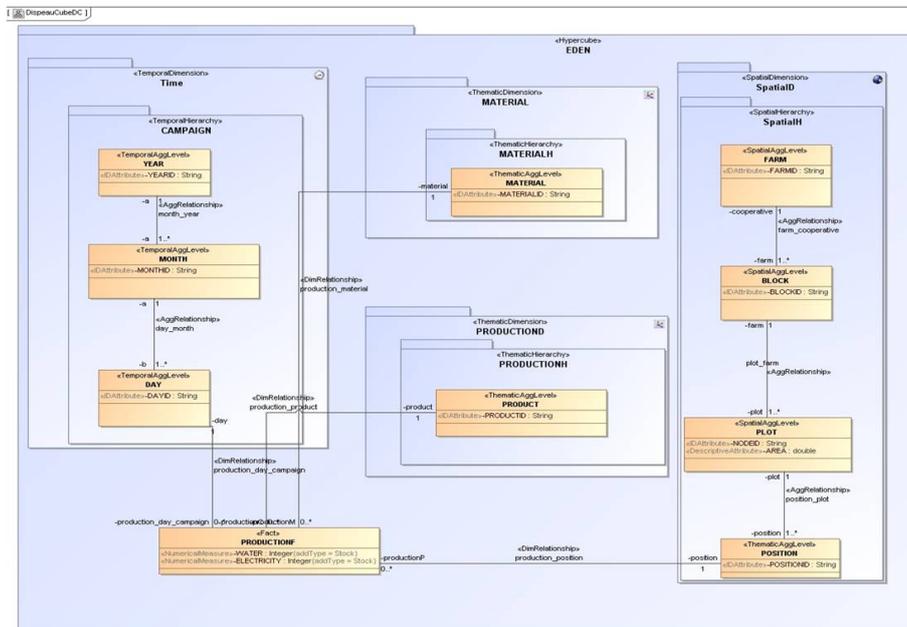


FIG. 5 – ED Irrigation

termes du nombre de documents). En effet, il est très important de remarquer que la description textuelle des besoins est sujette à des ambiguïtés comme par exemple la différence entre « par » et « / » pour définir « l'axe de dimension » et « l'opérateur de division » respectivement. De plus, les décideurs n'ayant pas une vision hiérarchique et multidimensionnelle, définissent systématiquement les indicateurs selon une approche requête et non comme l'ensemble des indicateurs auxquels l'ED doit répondre. Par exemple l'expression « lait par année pour le lot1 » (requête) implicitement fait référence aussi à la quantité de lait par mois et jour (modèle multidimensionnelle). Cela se traduit par un effort très important d'échange avec les experts, et donc de reformulation des indicateurs sous forme hiérarchique. Donc, en utilisant un support plus structuré comme le tableau et qui est familier, les décideurs ont vite intégré les problématiques de hiérarchisation et ils ont pu facilement exprimer leurs formules de calcul pour les agrégations à l'aide des fonctionnalités de calcul d'Excel. Néanmoins, dans cette approche, les décideurs étant libres de construire les tableaux par eux même, quelques réunions ont été nécessaires pour leur demander de bien présenter leurs données/indicateurs sous forme de tableaux croisés. Pour résumer, cette expérimentation nous confirme que l'utilisation d'un langage graphique (e.g. UML) bien implémenté d'un CASE tool ainsi que l'utilisation d'une méthodologie d'implémentation automatique (e.g. MDA) améliorent les temps et les coûts du projet. De plus, dans un contexte particulier où les décideurs ont peu de connaissances en informatique et/ou entrepôts de données l'expression des besoins d'analyse via un paradigme à eux bien familier (e.g. Excel), est très utile.

4 Conclusions et travaux futurs

Dans ce travail et dans le contexte du projet national EDEN, nous décrivons l'efficacité de ces méthodologies. À l'aide des trois EDs différents développés, notre étude montre comment l'utilisation d'outils de conception assistée et d'implémentation automatique améliore les temps globaux de développement du projet. De la même façon, les outils pour la définition des besoins décisionnels apportent une grande amélioration en évitant les ambiguïtés liées au langage naturel ou à la complexité des notations formelles. En particulier, nous avons testé avec succès l'utilisation d'un outil comme Excel qui est familier aux décideurs et permet une définition très structurée des besoins fonctionnels. Actuellement, nous travaillons sur la définition d'un outil de type Tableur pour la définition des besoins utilisateurs qui puisse être intégré directement dans notre outil ProtOLAP. De plus, d'autres expérimentations doivent être conduites pour bien évaluer les possibles impacts sur les projets des autres paramètres qui n'ont pas été pris en compte dans ce travail (e.g. profil utilisateur, architecture, etc.). Nous allons aussi mettre la méthode proposée en parallèle avec les méthodes existantes d'évaluation de la qualité logicielle, comme par exemple Goal Question Metric.

5 Remerciements

Les auteurs tiennent à remercier le ministre français de l'Agriculture pour son soutien de ce projet.

Références

- Abello, A., J. Samos, et F. Saltor (2006). Yam2 : a multidimensional conceptual model extending uml. *Inf. Syst.* 31(6), 541–567.
- Benker, T. et C. Jürck (2012). A case study on model-driven data warehouse development. In *DaWaK*, pp. 54–64.
- Bimonte, S., K. Boulil, F. Pinet, et M. Kang (2013). Design of complex spatio-multidimensional models with the icsolap uml profile : An implementation in magicdraw. In *15th International Conference on Enterprise Information Systems (ICEIS), 04/07/2013-07/07/2013, Angers, FRA*, pp. 310–315.
- Bimonte, S., J. Chanet, J. Capdeville, et Y. Lefrileux (2014). Energetic assessment of dairy activities using olap systems. In *Ageng 2014 International conference on agricultural engineering, 06/07/2014-10/07/2014, Zurich, DEU*, pp. 10.
- Bimonte, S., E. Edoh-Alove, H. Nazih, M.-A. Kang, et S. Rizzi (2013). Protolap : rapid olap prototyping with on-demand data supply. In I.-Y. Song, L. Bellatreche, et A. Cuzzocrea (Eds.), *DOLAP*, pp. 61–66. ACM.
- Bimonte, S. et M. Kang (2014). *WikOLAP : Integration of wiki and OLAP Systems*. John Wang.
- Bimonte, S., M. Pradel, D. Boffety, A. Tailleur, G. Andre, R. Bzikaha, et J. Chanet (2014). A new sensor-based spatial olap architecture centered on an agricultural farm energy-use diagnosis. *International Journal of Decision Support System Technology* 5(4), 1–20.

Une étude sur l'efficacité des méthodes de conception et d'implémentation pour les Entrepôts de Données

- da Chen, L., K. S. Soliman, E. Mao, et M. N. Frolick (2000). Measuring user satisfaction with data warehouses : an exploratory study. *Information & Management* 37(3), 103–110.
- Giorgini, P., S. Rizzi, et M. Garzetti (2008). Grand : A goal-oriented approach to requirement analysis in data warehouses. *Decis. Support Syst.* 45(1), 4–21.
- Golfarelli, M. et S. Rizzi (2011). Data warehouse testing : A prototype-based methodology. *Information & Software Technology* 53(11), 1183–1198.
- Kimball, R. (1996). *The Data Warehouse Toolkit : Practical Techniques for Building Dimensional Data Warehouses*. John Wiley.
- Mazon, J.-N. et J. Trujillo (2008). An mda approach for the development of data warehouses. *Decision Support Systems* 45(1), 41–58.
- Niemi, T., J. Nummenmaa, et P. Thanisch (2001). Constructing olap cubes based on queries. In J. Hammer (Ed.), *DOLAP*, pp. 9–15. ACM.
- Romero, O. et A. Abello (2009). A survey of multidimensional modeling methodologies. *International Journal of Data Warehousing and Mining (IJDWM)* 5(2), 1–23.

Summary

Les entrepôts de données et les systèmes OLAP permettent aux décideurs d'explorer et d'analyser d'énormes volumes de données modélisées suivant un modèle multidimensionnel et extraites de sources de données hétérogènes. En général, la conception d'un ED est compliqué, et demande du temps et de nombreuses ressources. De plus des experts en ED sont nécessaires pendant les phases de conception et d'implémentation. Dans cet article, nous présentons une nouvelle méthodologie et un outil permettant aux modélisateurs (ne connaissant pas les ED) de concevoir et d'implémenter des ED pour analyser eux-mêmes des données résultantes de simulation, sans l'intervention d'expert en ED.