Table des matières

Résumé ii

Avant-Propos iii

Table des matières v

Liste des tableaux vii

Liste des figures viii

Chapitre 1 : Introduction 1

1.1- Mise en contexte 1
1.1-1. La géomatique et la découverte de connaissance 1
1.1-2. L'opportunité archéologique : l'analyse spatio-temporelle de données 3D 2
1.2- Problématique 4
1.3- Objectifs 6
1.4- Méthodologie 7
1.5- Description des chapitres du mémoire 9

Chapitre 2 : Revue des concepts 10

2.1- Introduction 10
2.2- La donnée : de sa création à la génération de modèles 3D 10
2.2-1. De la réalité aux différents types de données 11
2.2-2. L'objet tridimensionnel 15
2.2-3. La modélisation géométrique 3D 17
2.3- L'analyse de données – revues de définitions 19
2.3-1. Les différentes catégories d'analyses 20
2.3-2. Les opérateurs d'analyse: 23
2.4- Les outils d'exploitation des données spatiales 29
2.4-1. Les systèmes transactionnels 30
2.4-2. Les systèmes d'analyse 32
2.5- Conclusion du chapitre 37

Chapitre 3 : L'optimisation des outils d'analyse en ligne 38

3.1- Introduction 38
3.2- La révision des données d'interprétation : nouveaux concepts 41
3.2-1. Principe 41
3.2-2. Mise à jour vs révision 42
3.2-3. Les processus de révision et de mise à jour, et les systèmes d'analyse en ligne 45
3.2-4. Révision de structure vs révision de données 48
3.3- La révision des données d'interprétation : sélection d'une structure multidimensionnelle 51
3.3-1. Le choix de l'architecture ROLAP 51
3.3-2. Le choix du schéma en étoile 52
3.3-3. Le choix de la structure Parent-enfant d'une dimension 54
3.3-4. Construction et reconstruction des agrégations 55
3.4- La révision des données d'interprétation : les règles d'application. 58

3.4-1. La révision des dimensions 58
3.4-2. La révision de la table des faits 63
3.5- Le SOLAP 3D 67
3.5-1. Le SOLAP 3D : de la théorie à la pratique 67
3.5-2. La nécessité d'un SOLAP 3D 71
3.6- Conclusion 74

Chapitre 4 : Application des concepts du SOLAP 3D à un contexte de fouille archéologique 75

4.1- La constitution des données archéologiques 75
4.1-1. Les objets archéologiques 76
4.1-2. L'interprétation des objets archéologiques 77
4.1-3. Les données d'observation 79
4.1-4. Les données d'interprétation 80
4.1-5. Les modèles 3D archéologiques 82
4.2- L'analyse des besoins des archéologues 83
4.2-1. Les besoins généraux sur l'analyse des données 83
4.2-2. Un besoin spécifique : la compatibilité cognitive 86
4.3- La révision de la structure multidimensionnelle : de la théorie à la pratique 89
4.3-1. Présentation du prototype 90
4.3-2. Analyse des résultats 95
4.4- Conclusion 96

Chapitre 5 : Discussion et conclusion 98

5.1- Présentation des résultats et discussion 98
5.1-1. L'exploration 3D avec les solutions OLAP et/ou SOLAP 98
5.1-2. L'élargissement des concepts de mises à jour de données d'une structure multidimensionnelle 99
5.1-3. L'application archéologique 101
5.2- Développement et recherches futures 102

Bibliographie 103

Annexe A Dimensions d'analyses utilisées en archéologie I

Annexe B Dimensions d'analyses non -utilisée en archéologie IV

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

Liste des tableaux

Tableau 2-1 : La dimension géométrique des objets adaptée de Brisebois 2003 15
Tableau 3-1 : Exploitation des données descriptives et des données spatiales des technologies actuelles 70
Tableau 4-1 : Exemple d'analyses spatiales métriques 3D appliquées à l'archéologie.....83
Tableau 4-2 : Exemple d'analyses spatiales topologiques 3D appliquées à une UF 84
Tableau 4-3 : Exemple d'analyses temporelles topologiques 3D appliquées à l'archéologie 85
Tableau 4-4 : Les mesures utilisées et leur fonction d'agrégation 91
Tableau 4-5 : Exemple de liste de modifications.....93

Liste des figures

Figure 1-1 : Extrait du « Processus de construction de connaissances adapté de (Gahegan et al., 2001) » par (Brisebois, 2003) 1

Figure 1-2 : Méthodologie adoptée dans la présente recherche (formalisme UML : diagrammes d'activités) 7

Figure 2-1 : De la donnée à la connaissance inspiré de (Bellinger *et al.*, 1997; Gahegan et al., 2001) 11

Figure 2-2 : La caractérisation d'une réalité inspirée de (Bédard, 2003a) 12

Figure 2-3 : La modélisation de la réalité en classe d'objets inspiré de (Bédard, 2003a) 13

Figure 2-4 : des classes d'objets aux données inspiré de (Bédard, 2003a) 14

Figure 2-5 : Comparaison de l'objet 3D défini par la géométrie et défini par l'univers adapté de (Brisebois, 2003) 16

Figure 2-6 : Objet 3D défini par son rectangle englobant adapté de (Brisebois, 2003) 17

Figure 2-7 : L'analyse, de la réalité à la compréhension, adaptée de (Champoux, 1991) 20

Figure 2-8 : Les opérateurs temporels topologiques 26

Figure 2-9 : Navigation : zoom et déplacement de l'écran - exemple d'Autocad 2004 28

Figure 2-10 : Exemple d'utilisations de graphiques et d'histogrammes 29

Figure 2-11 : Exemple de cubes, dimensions et mesures inspiré de (Caron, 1998; Bourgon *et al.*, 1999) 33

Figure 3-1 : Le processus de compréhension de la réalité au travers des données d'observation et d'interprétation 44

Figure 3-2 : Comparaison des efforts de plusieurs groupes d'utilisateurs par rapport à des opérations particulières d'une approche multidimensionnelle inspirée des travaux de Bédard 2005 et Lambert 2005 46

Figure 3-3 : La donnée d'interprétation dans le processus d'un système d'analyse en ligne inspirée de Lambert (2005) 47

Figure 3-4 : Niveau conceptuel d'une dimension 48

Figure 3-5 : Exemples de processus de révision de la structure multidimensionnelle inspiré de (Vaisman *et al.*, 2002) 49

Figure 3-6 : La relation père-fils : une relation clé primaire – clé étrangère 49

Figure 3-7 : Exemples de processus de révision des données d'interprétation 50

Figure 3-8 : Architecture ROLAP inspirée de (Rivest, 2000) 52

Figure 3-9 : Exemple de modélisation en étoile (ici avec 2 dimensions seulement) 54

Figure 3-10 : Arbre balancée vs Arbre non-balancée 54

Figure 3-11 : Exemple de la relation « parent-enfant » 55
Figure 3-12 : Exemple de construction d'agrégations d'une configuration ROLAP 56
Figure 3-13 : Exemple de changement de père avec aucune répercussion.....59
Figure 3-14 : Exemple de changement de père avec des répercussions 60
Figure 3-15 : L'ajout d'un niveau : l'avantage de la relation « parent-enfant ».....63
Figure 3-16 : Exemple de modifications de clés étrangères 64
Figure 3-17 : Exemple de modifications de mesures.....65
Figure 3-18 : Exemple de couplage intermédiaire.....73
Figure 4-1 : Exemples d'Unités de Fouille–Construction (gauche) et – Débris (droite).....76
Figure 4-2 : Les quatre catégories de matériel archéologique retiré des unités de fouille d'un chantier archéologique.....77
Figure 4-3 : Exemple simplifié de reconstitution de l'aire d'activité 78
Figure 4-4 : Du chantier de fouille aux données d'observation.....79
Figure 4-5 : Les données d'interprétations générées au cours de l'opération catalogage d'éléments de la culture matérielle.....80
Figure 4-6 : Les données d'interprétations archéologiques issues des unités de fouille 81
Figure 4-7 : Les données d'interprétations issues des dimensions (exemple de la couleur) 82
Figure 4-8 : Exemples de modèles 3D archéologiques (Tell 'Acharneh) (Losier, 2005).....82
Figure 4-9 : Exemple d'échelle temporelle archéologique.....85
Figure 4-10 : Exemple d'analyses visuelles 3D appliquées à des « Unités de Fouille » (Image tirée de la modélisation de Losier 2005) 86
Figure 4-11 : Les quatre étapes de la compréhension d'un site archéologique 87
Figure 4-12 : Interface « Évolution des clés étrangères et mesures ».....92
Figure 4-13 : Les mesures et les dimensions 92
Figure 4-14 : Les différents types de modifications 93
Figure 4-15 : Interface « Modification des dimensions ».....94
Figure 4-16 : Liste des modifications des dimensions.....95

Chapitre 1 : Introduction

1.1-Mise en contexte

1.1-1. La géomatique et la découverte de connaissance

L'analyse de données s'inscrit dans un processus de découverte de connaissance. La découverte de connaissance (Knowledge Discovery en anglais) est une extraction complexe à partir de la donnée d'une information implicite, au départ inconnue et potentiellement utile (Frawley *et al.*, 1992). Le processus de découverte de connaissance, explicité en Figure 1-1 est de nature itérative et permet de progresser depuis l'exploration des données jusqu'à la présentation de modèles. Ce processus en quatre phases aide l'usager à construire une connaissance dotée d'une composante spatiale (Gahegan *et al.*, 2001; Brisebois, 2003). La phase d'exploration (phase I) a pour objectif la recherche des patrons, de structures et de relations implicites dans un jeu de données. La phase suivante (II) consiste à proposer des hypothèses. Ces dernières sont par la suite validées ou invalidées au moyen d'analyses statistiques (phase III). Lorsque les résultats sont concluants (phase IV), ils sont communiqués en prenant soin de délimiter leur portée.

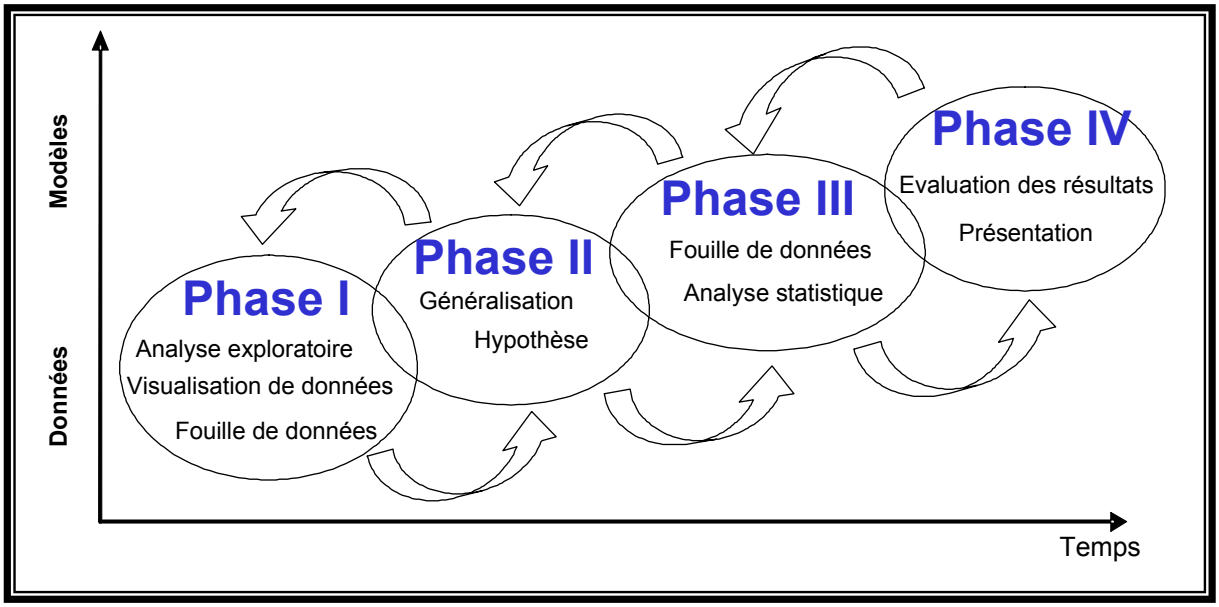


Figure 1-1 : Extrait du « Processus de construction de connaissances adapté de (Gahegan et al., 2001) » par (Brisebois, 2003)

La géomatique est une « discipline ayant pour objet la gestion des données géographiques et qui fait appel aux sciences et aux technologies reliées à leur acquisition, leur stockage, leur traitement et leur diffusion » (OQLF, 2006). Pour être compatible avec le processus de découverte de connaissance, la géomatique a dû inclure d'autres aspects : l'aide à la décision caractérisée, entre autres, par la technologie OLAP (On-Line Analytical Processing). Ces outils de découverte de connaissances ou d'analyse en ligne visent à assister l'usager dans son analyse en lui facilitant l'exploration de ses données et en lui donnant la possibilité de le faire rapidement (Bédard, 2005). Les travaux de ces dix dernières années sur l'analyse en ligne ont fait ressortir une perspective d'évolution intéressante. Ainsi, Caron, en 1998, a démontré que l'OLAP possède un potentiel réel pour supporter l'analyse spatio-temporelle (Caron, 1998). Cependant, sans volet cartographique, il est impossible de visualiser la composante géométrique des données spatiales. L'ajout de ce volet cartographique a conduit à l'élaboration de l'outil SOLAP (OLAP spatial) : « une plate-forme visuelle supportant l'exploration et l'analyse spatiotemporelle rapides des données selon une approche multidimensionnelle à plusieurs niveaux d'agrégation » (Rivest, 2000). Par la suite, la potentialité de la tridimensionnalité de la technologie SOLAP a été explorée (Brisebois, 2003). Suite aux constats de Brisebois et aux difficultés à manipuler la tridimensionnalité des données (Zlatanova *et al.*, 2002; Lachance, 2005), l'archéologie est apparue comme un excellent domaine d'application. Elle offre des situations permettant de se pencher sur la manipulation des données tridimensionnelles dans un contexte de découverte de connaissance.

1.1-2. L'opportunité archéologique : l'analyse spatio-temporelle de données 3D

Lors de la fouille d'un site archéologique, l'archéologue/fouilleur prend en compte la localisation sur le terrain (x, y, z) des données qui y sont découvertes. Suivant une pratique courante, elles sont enregistrées en regard d' « unités de fouille » (**UF**) qui se trouvent être des volumes (tridimensionnels) de terre archéologique qui sont retirés de différents endroits du site lors des opérations de fouille et qui contiennent des artéfacts, des écofacts, des restes humains ou des vestiges architecturaux.

Rapport-Gratuit.com

Lors du processus d'analyse, les « unités de fouille », et par conséquent le matériel archéologique qu'elles contiennent, sont mises en relation avec les différentes couches de terre stratigraphiques formant le site. Ces corrélations tiennent une place importante dans le processus de découverte de connaissances archéologiques. En effet, la compréhension - ou l'interprétation - d'un site archéologique passe, dans un premier temps, par la reconstitution de l'ordre d'accumulation au sol de ces couches stratigraphiques. Puis, dans un second temps, grâce à des jalons chronologiques fournis par le matériel archéologique, la séquence de déposition des couches est datée d'une manière relative – les unes par rapport aux autres - mais aussi absolue : points fixes dans le temps. A noter que les vestiges archéologiques enregistrés sur un chantier de fouilles sont des traces laissées au sol par des humains dans le passé. Le but ultime du fouilleur est de reconstituer la séquence événementielle qui a mené à la formation du site archéologique qu'il a fouillé : « Un site archéologique est un lieu où, dans le passé, des actions humaines et sociales ont été réalisées » (Barceló *et al.*, 2003).

Les méthodes les plus répandues à l'heure actuelle parmi les archéologues de terrain pour procéder à la construction (ou reconstruction) des connaissances spatio-temporelles sur un chantier de fouilles sont la combinaison des dessins en plan et en coupe, ainsi que la « Matrice de Harris ». Cependant, ces moyens « traditionnels » agissent dans un environnement 2D et non 3D comme le souhaiteraient les archéologues dans leur optique de reconstruire une réalité 3D (Green *et al.*, 2001; Malinverni *et al.*, 2002; Barceló et al., 2003; Cattani *et al.*, 2004; Day *et al.*, 2004; Losier, 2005).

Des catégories de logiciels gèrent déjà des données géographiques 3D : ce sont les outils de types Conception Assistée par Ordinateur (plus communément appelés CAO ou CAD en anglais) et les Systèmes d'Information Géographique (plus communément appelés SIG ou GIS en anglais). Plus spécifiquement, de nombreux travaux ont été réalisés dans le domaine des SIG 3D en urbanisme (De La Losa, 2000; Ramos, 2003) ou en géologie (Apel, 2004). Cependant, les technologies actuelles en logiciel SIG ne permettent pas de répondre à la problématique particulière de l'archéologie citée précédemment (Wheatley *et al.*, 2002; Zlatanova *et al.*, 2004; Barceló, 2005). En effet, jusqu'à présent, l'évolution de la recherche s'est plus orientée vers la modélisation et la navigation 3D au détriment d'une analyse de

données 3D (Zlatanova et al., 2004). Ainsi, force est de constater que, pour l'instant, tous les outils permettant la gestion des données tridimensionnelles ont plus mis l'emphase sur la visualisation et la représentation 3D que sur une analyse de données (descriptive, temporelle, spatiale ET visuelle) rapide et simple.

Cette situation lance des défis à la géomatique notamment la nécessité d'analyser simultanément l'espace et le temps, dans un contexte où l'un peut être influencé par l'autre, et de le faire rapidement et facilement dans un environnement tridimensionnel afin de mieux supporter l'interprétation des données, la création d'hypothèses et la découverte de nouvelle connaissance archéologique.

1.2-Problématique

Initialement, notre intérêt s'était porté vers une amélioration des fonctionnalités tridimensionnelles des outils SOLAP dans un domaine – l'archéologie de terrain – où l'on connaît l'existence d'une problématique d'analyse tridimensionnelle spatio-temporelle. Même si la pertinence d'un outil SOLAP 3D en la matière a déjà été démontrée (Brisebois, 2003), l'efficacité réelle d'un tel outil pour les utilisateurs n'a pas été testée en profondeur puisqu'il s'agissait d'une première exploration technologique du potentiel d'adapter le SOLAP au 3D ainsi que d'une première tentative d'amener un système de type SOLAP en archéologie (Rageul, 2004). L'amélioration des fonctionnalités tridimensionnelles devait passer par une étude de la gestion des données spatio-temporelles tridimensionnelles dans un contexte SOLAP. La mise en évidence des variables graphiques utiles au système d'analyse de données 3D, l'intégration d'un modèle géométrique volumique aux systèmes SOLAP étaient des exemples d'amélioration. Nous devions nous servir de l'opportunité archéologique afin d'étudier la faisabilité et l'efficacité de tels outils. Plus particulièrement, l'étude de leur efficacité devait permettre aux utilisateurs potentiels d'évaluer l'outil en fonction de sa facilité à extraire l'information des données descriptives, temporelles et spatiales issues du processus d'acquisition des données et de permettre une utilisation et une rapidité d'exécution visuelle compatible avec le processus analytique des archéologues.

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

Cependant, le cours de recherche a permis de mieux comprendre les attentes des archéologues quant à l'analyse de leurs données. « Un problème dans la réalisation de projets SOLAP est la difficulté des utilisateurs à exprimer leurs besoins et leurs attentes en début d'analyse de système ». (Guimond, 2005). Ainsi l'analyse des attentes des archéologues a permis la découverte d'un autre besoin, plus immédiat et intéressant tout autant pour eux que pour la géomatique. La découverte de ce besoin a permis de réorienter la problématique initiale vers une nouvelle problématique tout en offrant certains éléments de réponse à la première. Ce nouveau besoin n'a jamais été rencontré auparavant et il présente un défi primordial pour l'archéologue : **la possibilité de modifier les données incluses dans un SOLAP « durant » l'analyse**. Cette modification n'entre ni dans la mise à jour des données terrain, ni dans l'optique d'une correction d'erreur mais plutôt dans une nouvelle catégorie de problème. Dans la Figure 1-1 relative aux phases de découverte de nouvelles connaissances, nous pouvons situer ce besoin principalement dans la boucle de rétroaction entre les phases 1 (analyse exploratoire) et 2 (génération d'hypothèse).

Il existe en effet des situations exceptionnelles (comme en archéologie) où la donnée incluse dans l'outil SOLAP doit être « modifiée », « **révisée** », i.e. réinterprétée afin d'être améliorée. Il s'agit d'une amélioration que l'utilisateur veut apporter à la donnée car en fait, et c'est le cas en archéologie, l'analyste ne peut valider efficacement le peuplement de l'outil SOLAP qu'en effectuant son analyse avec l'outil en question ! En d'autres termes, c'est en explorant ses données qu'il fait émerger des idées, de nouvelles hypothèses, et c'est grâce à celles-ci qu'il peut améliorer son interprétation du site et qu'il « construit » la réalité d'autrefois. Dans l'outil SOLAP, la qualité des données s'améliore donc au fur et à mesure de son analyse et ce sont justement ces données qui fournissent la clé aux archéologues.

Or, actuellement, les systèmes OLAP servent principalement à exploiter une base de données ne subissant qu'occasionnellement des ajouts (ce qui est différent des modifications) et dont la structure est aussi rarement modifiée. En raison d'une dénormalisation volontaire, pour satisfaire une réponse rapide à des requêtes plus ou moins complexes, les outils OLAP traditionnels sont plutôt inadaptés pour exploiter des données

subissant des ajouts fréquents (Bédard, 2005). Les systèmes OLAP « temps réels » ont justement été développés pour répondre spécifiquement à ce besoin. Cependant ces systèmes ne sont pas encore optimisés à la mise à jour complète (ajout, suppression et modification) des données ce qui se ferait normalement suite à un changement sur le terrain dans la réalité. En effet, le processus de mise à jour consiste à changer la valeur des données pour les faire correspondre le plus possible à l'état actuel de la réalité (d'où le terme également utilisé d'actualisation des données) (Pouliot *et al.*, 2004). Les systèmes OLAP « temps réels » ne peuvent également pas forcément permettre la révision interactive de la structure du système et des données avec des modifications des caractéristiques de celles-ci. La possibilité de pouvoir modifier le contenu du SOLAP pendant une phase d'analyse interactive, voire même la structure complète de l'outil, ne semble pas avoir été explorée à ce jour dans le monde des données spatiales. De plus, les concepts de mise à jour de données ont été enrichis, mais dans le sens de rafraîchissement (ajout) de données suite à la reconstruction de système pour un entrepôt de données (Lambert, 2005). A la différence du projet de Lambert, notre problématique de recherche consiste à **étudier l'évolution, qui plus est interactive, des données qui en plus sont tridimensionnelles, dans une structure SOLAP**. Cette évolution ne concerne pas exclusivement le rafraîchissement et la reconstruction des données, mais aussi une révision possible des caractéristiques d'une donnée ou de la structure du système SOLAP, tout en laissant à l'utilisateur, l'opportunité de le faire pendant sa démarche de découverte de connaissance.

1.3-Objectifs

L'objectif principal de notre recherche est de **proposer des mécanismes qui permettraient l'évolution des données 3D spatio-temporelles (ajout, suppression et modification) dans une structure multidimensionnelle** directement durant la phase d'analyse de l'utilisateur.

De cet objectif principal découlent des objectifs secondaires :

- Faire état des réflexions actuelles sur les possibilités de combiner l'exploration 3D avec les solutions OLAP et/ou SOLAP afin de faire des recommandations pour le développement futur de ce genre de système.

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

- Elargir les concepts d'évolution de données et de structure dans une structure multidimensionnelle
- Identifier les besoins des archéologues quant à l'analyse de leurs données, mettre en évidence leur processus analytique et faire des propositions pour mettre au point un système de découverte de connaissance mieux adapté à leur contexte.

1.4-Méthodologie

La méthodologie suivie au cours de la présente recherche est exposée en Figure 1-2.

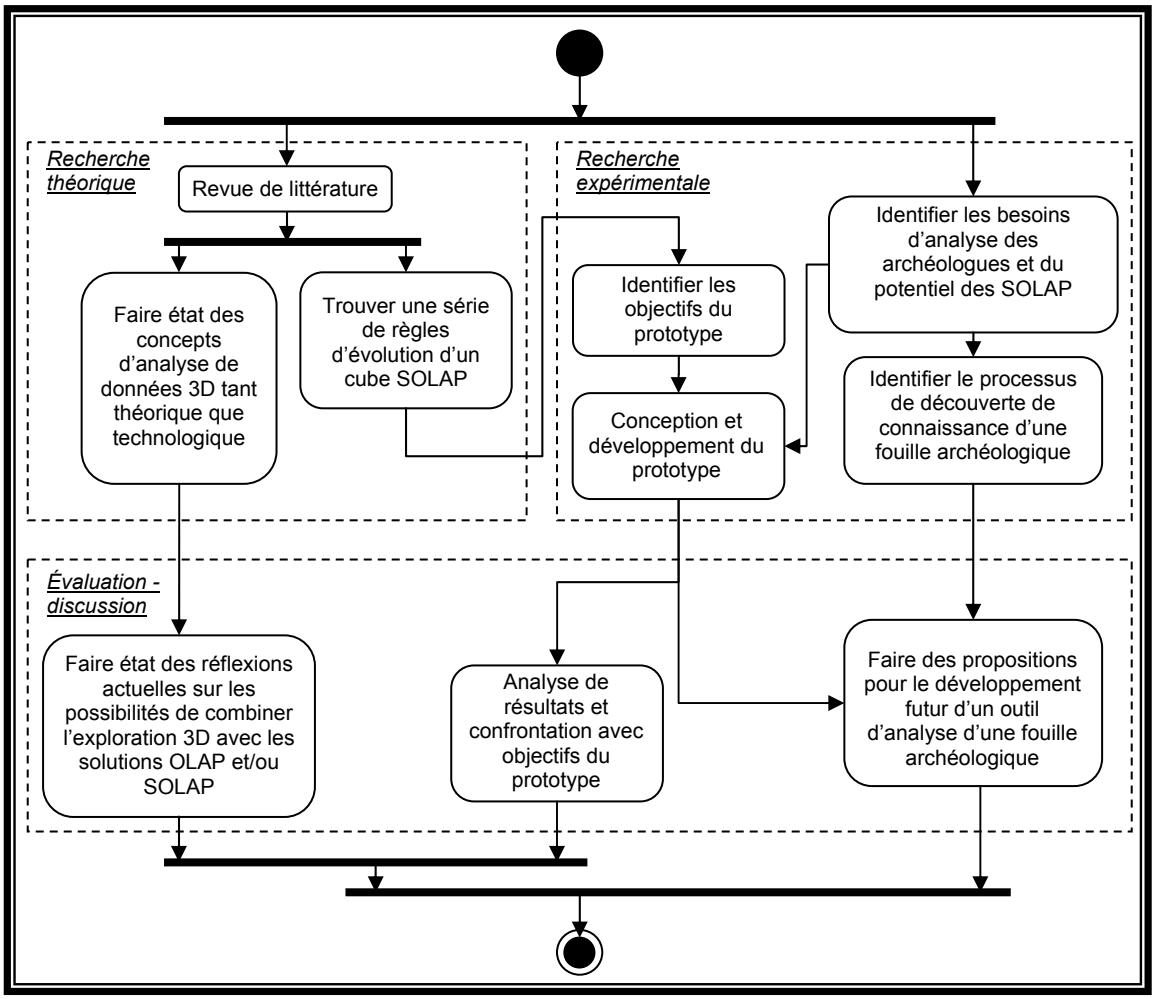


Figure 1-2 : Méthodologie adoptée dans la présente recherche (formalisme UML : diagrammes d'activités)

La *recherche théorique* s'est limitée à faire état des concepts d'analyse de données 3D tant théorique que technologique. La revue de littérature s'est portée sur l'analyse de données

graphiques et descriptives, sur les différents types de technologies à potentiel 3D : SIG 3D, CAO 3D entre autres, mais aussi sur les diffusions de données graphiques telles que la réalité virtuelle comme outil de visualisation 3D. Cette recherche concernait aussi la mise en place d'une série de règles d'évolution d'un cube SOLAP.

La *recherche expérimentale* s'est portée tout d'abord sur l'identification des besoins des archéologues. Diverses rencontres avec des archéologues ont permis d'identifier une partie du processus de découverte de connaissance d'une fouille archéologique. Ces rencontres ont également permis d'étudier le potentiel du SOLAP appliqué à l'archéologie à l'aide d'une maquette SOLAP résultante du projet de géomatisation du processus de fouille archéologique (Rageul, 2004) et présentée à des archéologues. La recherche expérimentale a aussi conduit à la mise en place d'un prototype d'évolution de données d'un cube SOLAP. Le projet étant appliqué à l'archéologie, l'étude préalable des besoins d'analyse des archéologues a été nécessaire.

L'*évaluation et la discussion concluent* la recherche sur les trois thèmes abordés au cours de cette maîtrise :

- la mise en place d'architectures justifiées combinant l'exploration 3D avec les solutions OLAP et/ou SOLAP,
- l'évolution des données et de la structure pendant la phase d'analyse dans une structure multidimensionnelle
- le contexte archéologique (identification des besoins d'analyse, exploration du processus de découverte de connaissance d'une fouille archéologique et recommandations pour le développement futur d'un outil d'analyse adapté à leurs besoins.

1.5- Description des chapitres du mémoire

Ce mémoire est divisé en 5 chapitres. Après la présente introduction (chapitre 1), le chapitre 2 expose une synthèse des notions théoriques pertinentes à notre projet de recherche. Cette synthèse caractérise la donnée spatiale tridimensionnelle, en examinant quelles sont les catégories d'analyses effectuées et les technologies disponibles pour gérer, manipuler, analyser et visualiser cette donnée. Nous y avons introduit les concepts du SOLAP et de l'approche multidimensionnelle et plus particulièrement les dernières optimisations (temps réel et 3D) relatives à notre projet. Le chapitre 3 décrit une approche théorique d'optimisations des outils SOLAP. La première optimisation concerne la prise en compte de l'évolution de données pendant la phase d'analyse. La seconde optimisation porte sur les architectures potentielles d'un SOLAP 3D et comprend une discussion sur les difficultés à les mettre en œuvre. Le chapitre 4 explicite, quant à lui, l'approche expérimentale de la première optimisation dans le contexte archéologique. Nous nous attacherons aussi dans ce chapitre à comprendre le processus de découverte de connaissance archéologique. Enfin, le chapitre 5 conclut ce mémoire en discutant les résultats de nos approches tant théorique qu'expérimentale, et en offrant des pistes de recherche qui pourraient être explorées dans le cadre d'autres projets de recherche qui seraient susceptibles de voir le jour au regard de nos conclusions.

Chapitre 2 : Revue des concepts

2.1- Introduction

Afin d'introduire certains concepts théoriques jugés nécessaires à la compréhension de notre travail de recherche, nous allons brièvement discuter sur les données et leur rôle prépondérant dans la représentation numérique 3D. Ce chapitre introduit également les concepts d'analyse, processus permettant de déduire l'information des données pour mieux comprendre la réalité. Ces notions sont importantes car elles permettent de poser les bases tant dans notre approche théorique d'optimisation (chapitre 3) que notre approche expérimentale (chapitre 4).

Il nous a également semblé pertinent de faire une présentation des différents systèmes informatiques capables d'exploiter les données, et tout particulièrement celles qui sont tridimensionnelles. Ces systèmes informatiques tiennent une place importante dans la démarche de l'analyste lorsque celui-ci est dans une phase de compréhension de la réalité. Nous mettrons l'accent sur le système SOLAP ainsi que deux optimisations : la première concerne la notion de temps réel dans un SOLAP et la deuxième concerne la gestion de la donnée tridimensionnelle. Nous verrons par la suite dans le chapitre suivant comment ces deux optimisations influence directement notre recherche d'optimisation.

2.2- La donnée : de sa création à la génération de modèles 3D

Comme le souligne Aamodt et Nygard, la donnée, l'information et la connaissance sont trois notions extrêmement dépendantes (Aamodt *et al.*, 1995). Les Figures 1-1 et 2-1 représentent le processus de construction de connaissances, la donnée servant de base à la génération de modèle et à l'élaboration de nouvelle connaissance. La **donnée**, en géomatique, est la représentation d'une information codée dans un format permettant son traitement par ordinateur (OQLF, 2006). La donnée peut être un nombre, un texte ou un symbole [(Longley et al., 2005) p 11] non interprétés (Davenport, 1997; Van der Spek *et al.*, 1997). Elle est physique et peut donc être stockée. Lorsque nous donnons un sens, une signification, une interprétation spécifique à une donnée, nous la transformons en **Information** (Davenport, 1997; Van der Spek et Spijkervet, 1997; Choo *et al.*, 2000;

Pouliot, 2005). Dès qu'on décode les données dans leur contexte d'utilisation et selon la personne qui les lit, une même donnée peut recevoir diverses interprétations et peut donc conduire à différentes informations (ou bien à aucune information si on n'en comprend ni le caractère, ni le symbole). L'information est un élément de connaissance concernant un phénomène et qui, pris dans un contexte déterminé, a une signification particulière. Le cadre de référence qui détermine cette interprétation est constitué de la somme des **connaissances** (ensemble de faits, d'événements, de règles d'inférence et d'heuristiques) et des expériences de la personne qui effectue l'interprétation (OQLF, 2006).

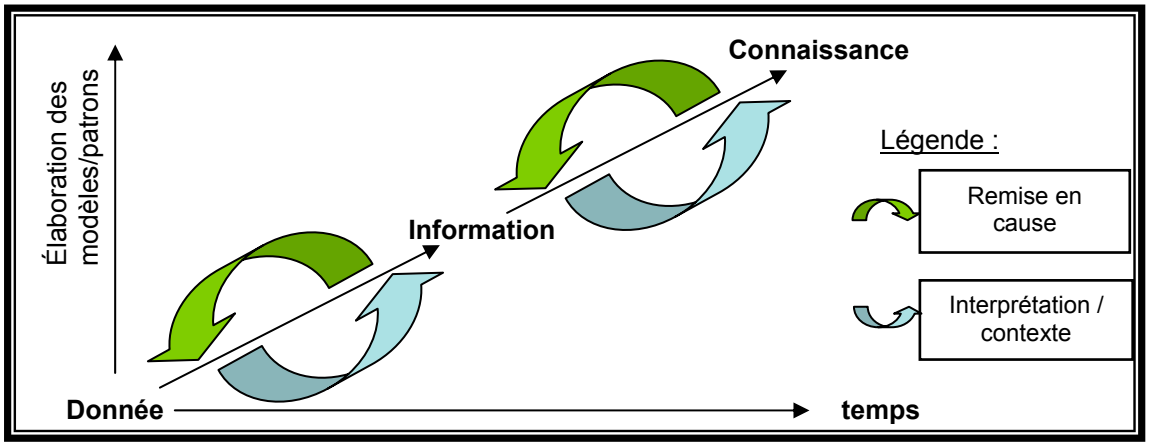


Figure 2-1 : De la donnée à la connaissance inspiré de (Bellinger *et al.*, 1997; Gahegan et al., 2001)

Afin de comprendre comment la donnée pourra nous informer dans les divers systèmes de découverte de connaissance, nous allons montrer la terminologie utilisée pour passer de la réalité à sa représentation numérique tridimensionnelle.

2.2-1. De la réalité aux différents types de données

Le « Robert » définit la réalité comme étant constituée de « choses réelles et de faits réels », c'est à dire qui « existent en fait ». En géomatique, la réalité est constituée de phénomènes et de relations entre ces phénomènes. L'OQLF (2006) définit le phénomène comme la réalité qui se manifeste à la conscience, que ce soit par l'intermédiaire des sens ou non. Le phénomène constitue donc la réalité première. Comme le montre graphiquement la Figure 2-2, le phénomène peut prendre l'apparence d'un objet tangible comme une maison, d'un événement comme un accident, d'un concept comme la notion d'unité de fouille en archéologie ou finalement l'apparence d'un être vivant comme une personne. L'autre

élément constitutif de la réalité est la relation entre les phénomènes, comme par exemple : « Monsieur Jalon possède une maison ».

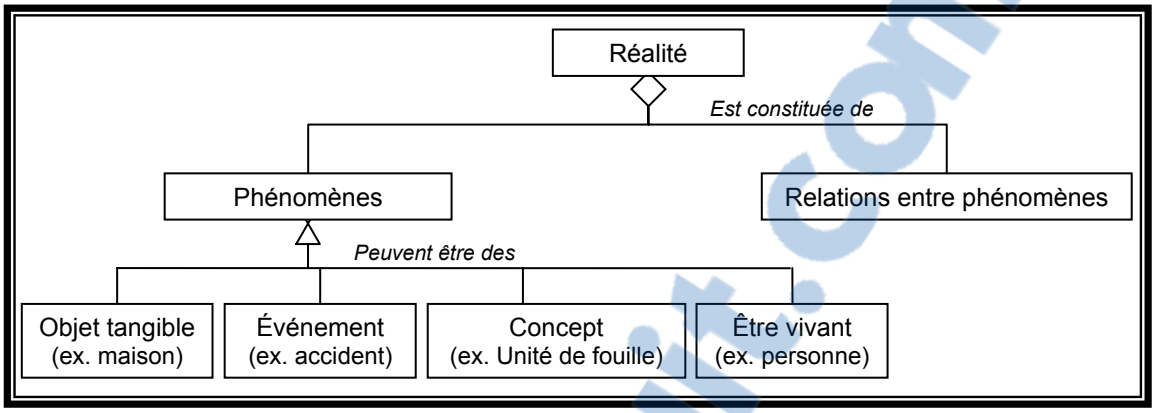


Figure 2-2 : La caractérisation d'une réalité inspirée de (Bédard, 2003a)

Afin de comprendre et de représenter numériquement cette réalité, nous allons chercher à la modéliser. La modélisation est une description dans un langage compréhensible à la fois par l'humain et par l'ordinateur de la forme, du mouvement et des caractéristiques d'un objet ou d'un ensemble d'objets qui crée un modèle, c'est-à-dire une représentation simplifiée de l'objet (OQLF, 2006). Cependant, un modèle reste une abstraction spécifique de la réalité et il doit faire ressortir de manière explicite les caractéristiques de ce qu'il est sensé modéliser. C'est en effet sur le modèle que l'analyse se fait d'où un contrôle permanent de la validité du modèle comme étant une représentation conforme à la réalité (Wilsey, 1999).

Étant donnée que les systèmes d'analyses visés par notre recherche sont principalement constitués de base de données, il n'est pas inutile de définir la démarche de modélisation ; démarche nécessaire pour passer d'une réalité à sa représentation : « Une méthode de modélisation est un ensemble de procédures et de règles à suivre permettant de capter la

Rapport-Gratuit.com

réalité du client et de concevoir une base de données correspondant aux besoins de cette réalité » (Bédard, 2003a). Au niveau conceptuel, c'est à dire indépendamment de la technologie utilisée, les modèles[1] sont constitués de classes d'objets - ou d'entités suivant la méthode utilisée - décrites par des attributs, des opérations, des associations entre les classes d'objets, des agrégations de classes d'objets simples en classes d'objets complexes et des généralisations ou spécialisations d'une super-classe en ses sous-classes. Les phénomènes sont ainsi modélisés en classes d'objets. Cet objet est l'élément de base qui sert à construire des logiciels. Il est la matérialisation de la classe dont il reproduit les caractéristiques (OQLF, 2006).[2] Les relations entre phénomènes, quant à elles, sont modélisées par des associations qui peuvent être de surcroît spatiales et/ou temporelles et décrites dans les classes d'objets (cf. Figure 2-3).

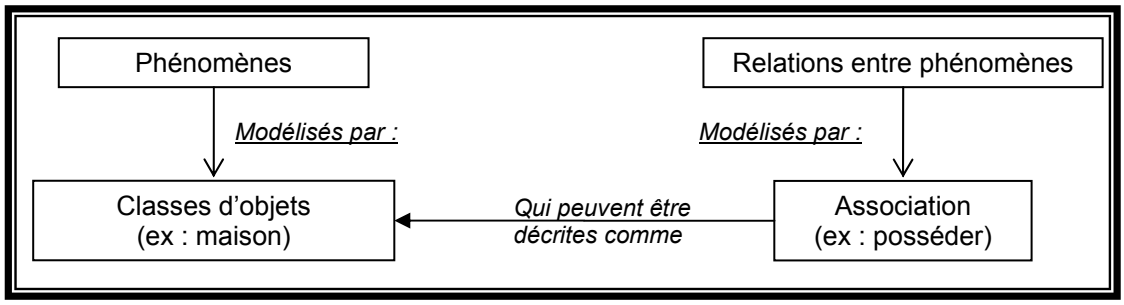


Figure 2-3 : La modélisation de la réalité en classe d'objets inspiré de (Bédard, 2003a)

Ainsi, les classes d'objets (ou entités), éléments essentiels d'un modèle conceptuel de données (MCD), sont décrites par des attributs et des opérations (cf. Figure 2-4). L'attribut est une « composante d'un modèle conceptuel de données, qui représente une caractéristique propre à un phénomène ou une caractéristique propre à une relation entre phénomènes » (OQLF, 2006). On y retrouve les attributs descriptifs comme le nombre d'étages d'un bâtiment, les attributs temporels comme la date de construction du même bâtiment, les attributs géométriques comme la position en (x,y,z) et les attributs graphiques comme la couleur. La donnée se retrouve ainsi stockée dans les valeurs d'attribut.

[1] Modèle conceptuel de données pour la méthode « Entité/Relationnel » ou modèle d'analyse pour la méthode « Orienté Objet »

[2] Cette notion d'objet se différencie de l' « objet tangible » présent en Figure 2-2. A partir de la section 2.2-2. la notion d'objet sera considéré comme un objet tangible, ie que l'on peut connaître en touchant.

Les **données descriptives** (cf. Figure 2-4) sont des données relatives à un des attributs d'une entité ou d'une relation, à l'exclusion de sa position et de sa forme (OQLF, 2006). Cela comprend les **attributs descriptifs** comme le nombre d'étages d'un bâtiment par exemple, mais aussi les **attributs temporels** comme la naissance ou la mort de l'objet. On peut aussi rencontrer dans la littérature la notion de **données temporelles** comme étant une donnée qui peut avoir un lien avec toute notion de temps (durée ou instant) (Korte, 2001).

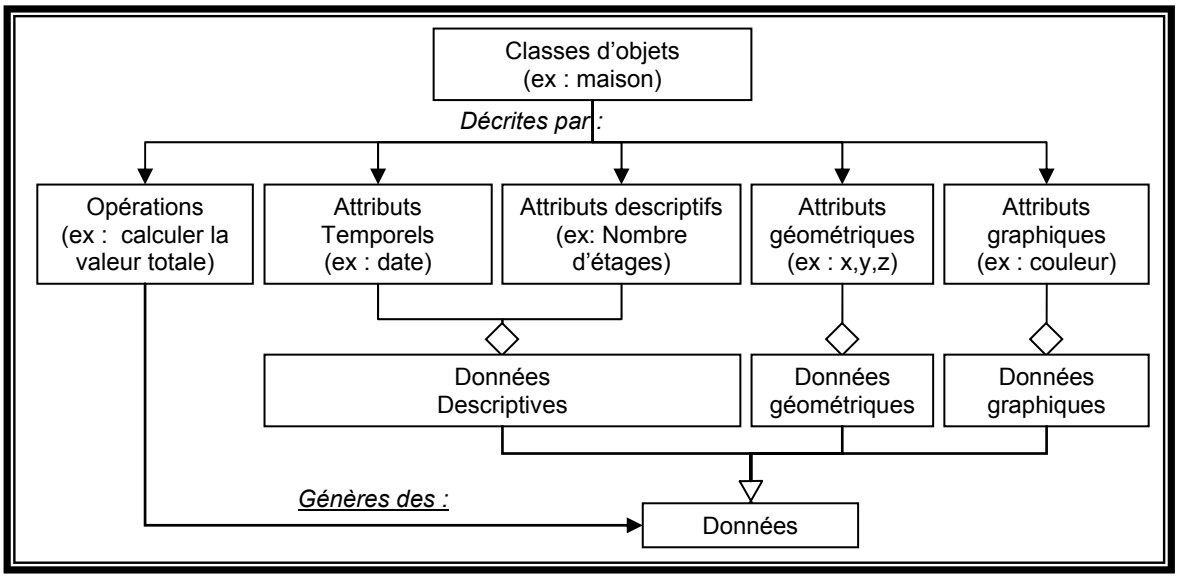


Figure 2-4 : des classes d'objets aux données inspiré de (Bédard, 2003a)

Les **données géométriques** (cf. Figure 2-4) renseignent sur la position ou la forme d'une entité géométrique. Les primitives géométriques (point, ligne, polygone) et les primitives topologiques (nœud, vecteur, surface) en sont des exemples (OQLF, 2006).

Les **données graphiques** (cf. Figure 2-4) portent sur la représentation visuelle d'une entité géométrique On peut ainsi mentionner, comme exemple, la couleur et la largeur du trait. On parle de données graphiques pour les représentations des entités spatiales sur des photos, des cartes ou des plans : la représentation cartographique des limites d'une propriété par exemple (OQLF, 2006).

Les **données spatiales** portent sur les entités spatiales et leurs relations, dans une application géomatique. L'objet peut alors être localisé par rapport aux autres [(Haining, 2003) : p4]. Les données spatiales comprennent l'ensemble des données géométriques, des

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

données descriptives et des métadonnées utilisées dans l'application géomatique (OQLF, 2006). Les données spatiales archéologiques étant des données spatiales tridimensionnelles, nous avons tenté de trouver une définition d'une **donnée spatiale 3D**. Cependant, celle-ci n'existe pas de manière **explicite** et **uniforme** dans la littérature principalement à cause de divergences liées à la définition même de « 3 » et « D » (Larrivée *et al.*, 2005). La section qui suit a pour but de présenter le rôle de la donnée pour la représentation numérique 3D.

2.2-2. L'objet tridimensionnel

L'objet est un « phénomène quelconque du monde extérieur ou intérieur qu'un homme observe (ou peut observer) à un instant déterminé » (OQLF, 2006). Il peut être décrit selon diverses dimensions : géométriques (en lien avec les données géométriques) ou descriptives (en lien avec les données descriptives). La notion de dimension est utilisée dans différents domaines (dessin technique, économie, électricité, imprimerie, mathématique, philosophie, informatique, physique, statistique…). Cependant, elle n'est pas clairement définie (Brisebois, 2003; Lachance, 2005; Larrivée *et al.*, 2006; Pouliot *et al.*, 2006) Face à cette multitude de définitions, nous utiliserons deux notions de dimensions liées à l'objet multidimensionnel (espace) pour définir l'objet 3D tel que nous l'utiliserons pour ce mémoire. Ce sont la dimension géométrique d'objet en lui-même et la dimension de l'univers dans lequel l'objet est positionné.

La **dimension géométrique de l'objet** est exprimée en fonction du nombre de directions ou de lignes suivant lesquelles le corps s'étend (Brisebois, 2003). D'après (Bédard *et al.*, 2002), l'objet 3D (encadré pointillé dans le Tableau 2-1) sera associé à une dimension géométrique 3D, c'est à dire un *objet volumique.*

Tableau 2-1 : La dimension géométrique des objets adaptée de Brisebois 2003

	Géométrie 0D	Géométrie 1D	Géométrie 2D	Géométrie 3D
Exemples d'objets				

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

La **dimension de l'univers** dans lequel évolue l'objet est exprimée en fonction du nombre d'axes, spatiale et/ou temporelle, nécessaire et suffisant pour positionner l'objet. D'après (Kennedy, 2004), l'objet 3D (encadré pointillé dans le Figure 2-5) sera associé à une dimension de l'univers 3D, c'est à dire à trois axes spatiaux et où les objets sont positionnés dans le volume infini et imaginaire, formé par les 3 axes.

Dépendant de la littérature étudiée, un objet peut donc être considéré comme 3D (Brisebois, 2003; Lachance, 2005; Pouliot, 2005) si la géométrie de l'objet est une géométrie 3D (Bédard *et al.*, 2002) ou si l'univers dans lequel évolue l'objet est un univers 3D (Kennedy, 2004) (Figure 2-5) ou encore si l'enveloppe des objets, c'est-à-dire le rectangle englobant de l'objet, est en 3D (Bentley, 2002) (cf. Figure 2-6).

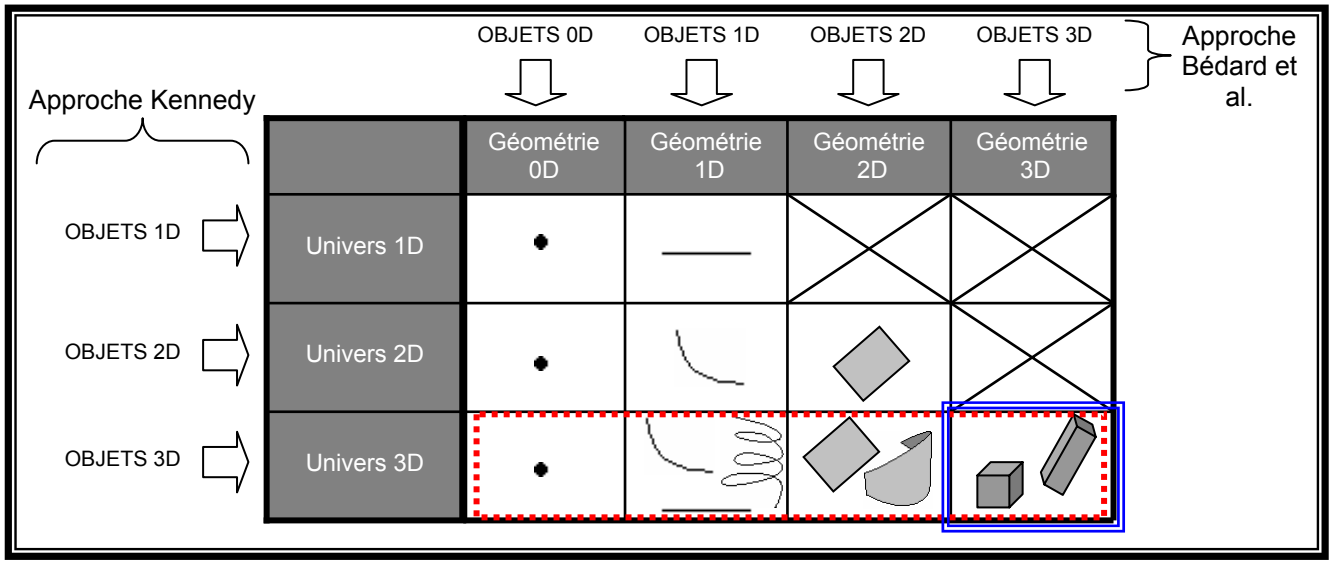


Figure 2-5 : Comparaison de l'objet 3D défini par la géométrie et défini par l'univers adapté de (Brisebois, 2003)

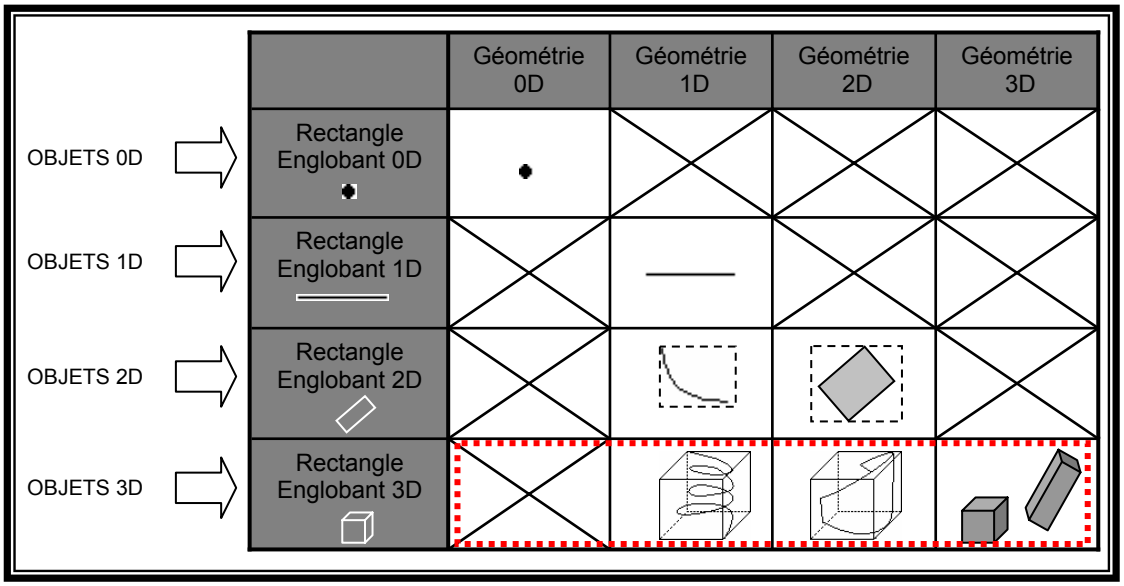


Figure 2-6 : Objet 3D défini par son rectangle englobant adapté de (Brisebois, 2003)

Pour le présent mémoire, nous associerons le terme d'objet 3D à **tout objet à trois dimensions géométriques (longueur, largeur et hauteur) et positionné dans un univers 3D (x, y, z) et référencés temporellement** (encadré double sur la Figure 2-5). En effet, en archéologie, lors des opérations de fouille, les seules données spatiales sont enregistrées en regard d' « unités de fouille » (*UF*) qui se trouvent être des volumes (tridimensionnels) de terre archéologique qui sont retirés de différents endroits du site (univers 3D).

2.2-3. La modélisation géométrique 3D

Les Figure 2-2, Figure 2-3 et Figure 2-4 ont montré le cheminement qu'il faudrait suivre pour reconstituer la réalité. À partir des données, la forme, le mouvement et les caractéristiques des objets sont reconstitués sous forme de modèle. Étant donné que nos travaux chercheront à examiner tout particulièrement l'évolution des informations dans un contexte d'analyse spatiale 3D et parce que l'archéologie propose déjà des modèles géométriques diversifiés (Green et al., 2001; Barceló et al., 2003; Nigro *et al.*, 2003; Cattani et al., 2004; Day et al., 2004; Mngumi *et al.*, 2004; Losier *et al.*, 2007) il nous apparaît d'expliquer en quoi consiste la modélisation géométrique tridimensionnelle.

La modélisation tridimensionnelle rend compte « des propriétés géométriques de l'objet » (OQLF, 2006) et s'intéresse à la représentation de la géométrie des systèmes étudiés et s'interroge sur la représentation spatiale proprement dite et sur la manière de les représenter (Pouliot, 2005). Cependant, il n'existe pas une seule façon de construire un modèle géométrique [(Longley et al., 2005) : p178]. Cette différence marque l'importance de bien spécifier quel type de modélisation géométrique est utilisé, puisque la modélisation influence de près l'analyse des données et l'affichage de celles-ci (Pouliot, 2005). Beaucoup d'auteurs (De La Losa, 2000; Ramos, 2003; Apel, 2004; Lachance, 2005) ayant déjà fait un état de l'art sur les différentes modélisations, l'intérêt de notre synthèse se porte plutôt sur une classification, même si elle ne se prétend pas exhaustive, qui permettrait de

mieux cerner les avantages et les inconvénients de chacune. À ce sujet, nous avons retenu la même classification que celle utilisée par Pouliot qui propose une classification basée sur les approches orientées espace et orientées objets (Pouliot et al., 2006).

L'**approche orientée espace** est utilisée lorsque le partitionnement de l'espace est arbitraire (par exemple des pixels ou des triangles) (Pouliot et al., 2006). Cette approche est souvent associée à la structure matricielle ou raster. D'après Longley, le modèle de données raster utilise des cellules (par exemple des pixels en 2D ou voxel en 3D) pour représenter les objets du monde réel [(Longley et al., 2005) : p181]. Il existe plusieurs manières de découper l'espace. La première, le découpage régulier, se fait sous la forme de cubes élémentaires fixes ou variables : le voxel. Ce découpage permet une représentation non ambiguë, unique et simple de la réalité (Pouliot, 2005; Thalmann, 2006). La précision, la granularité du modèle dépend directement de la résolution du pixel/voxel ou du nombre de niveaux hiérarchiques de l'octree. Cependant, cette précision se fait au détriment d'une taille de plus en plus conséquente du fichier de stockage. La deuxième manière, le découpage irrégulier, permet d'avoir une meilleure résolution avec un meilleur stockage. Le meilleur rapport résolution/stockage est principalement dû au fait que le découpage peut être lié à l'orientation de certains objets (modèles Binary Space Partitioning) ou en fonction de la densité des points échantillonnés (modèles en décomposition en cellules).

L'**approche orientée objet** est utilisée lorsque le partitionnement de l'espace est fonctionnel (par exemple la frontière d'une route, d'un bâtiment) (Pouliot et al., 2006). Une prémisse importante de cette approche consiste à connaître les frontières de l'objet. Il faut donc avoir ces frontières et être capable de les mesurer et de les estimer. Cette approche est souvent associée à la structure vectorielle. Les objets sont construits à partir des coordonnées des points et des arrêtes décrivant leurs position et forme. Il existe plusieurs approches pour représenter de façon discrète l'enveloppe des objets 3D ou leurs parties. L'*approche basée sur la frontière* est caractérisée par les modèles « fils de fer » (ou wireframe en anglais) et les modèles B-rep (Boundary Representation) La modélisation *fil de fer* est la plus simple des modélisations 3D [(Bertoline *et al.*, 2002) : p 304] car elle ne contient que les informations sur les points (vertex) et les lignes /courbes (edge) [(Foley,

1995) : p560]. Cette simplicité dans la modélisation permettant de stocker les informations géométriques la rend performante mais en fait aussi un modèle ambigu, confus et non-unique [(Saxena *et al.*, 2005) : p258 et 259]. La modélisation *B-rep* (boundary representation), modélise les faces des objets (orientées ou non) pour représenter les solides. [(Bertoline et Wiebe, 2002) : p314] sans pour autant avoir une idée sur ses propriétés volumiques (De La Losa, 2000). L*'approche basée sur des formes paramétrables* est caractérisée par les modèles Constructive Solid Geometry (plus communément appelés modèles CSG), les modèles par primitive Instancing et les modèles par balayage. La construction des objets avec la méthode *CSG* consiste à l'élaboration d'un modèle géométrique complexe à partir d'un jeu de primitives simples (Skibniewski *et al.*, 1997) comme des cubes, des sphères, des cylindres, des cônes ou des tores [(Gasparini, 2005) : p63] qui sont emboités après translation, rotation et facteurs d'échelle puis auxquels on effectue des unions, des intersections et/ou des différences (De La Losa, 2000). La primitive instancing est une approche indépendante de la représentation des objets solides. Cette modélisation est basée sur la notion de famille d'objets ou chaque membre de la famille est distingué par des paramètres [(Foley, 1995) : p539](Skibniewski et Kunigahalli, 1997) et ne se chevauche pas entre eux (Lattuada, 2005). Finalement, la modélisation par balayage consiste à représenter un objet en « balayant », une aire définie ou un volume, le long d'une trajectoire définie (Lattuada, 2005).

2.3-L'analyse de données – revues de définitions

Afin de comprendre la réalité et d'améliorer sa perception du monde, les analystes essaient de reconstruire cette réalité sous forme de modèle. L'élaboration de modèles se fait à l'aide des différentes données recueillies. Ces données sont caractérisées en fonction de leurs positions spatiale et temporelle, de leur description et de leur représentation graphique. L'aspect spatial est d'autant plus important puisqu'il sert de base à la modélisation géométrique. L'archéologie est une discipline exploitant au maximum cette reconstruction de la réalité au moyen de modèles géométriques (Losier, 2005). Une fois les données et le processus de modélisation décrits, nous allons maintenant expliquer les principaux types d'analyse à disposition de l'analyste (cf. chapitre3). Cela nous permettra de mieux mettre

en évidence les tenants et aboutissants d'un tel processus dans le contexte spécifique de l'analyse des données issues d'un chantier de fouille archéologique (cf. chapitre 4).

L'analyse de données correspond à l'ensemble des méthodes statistiques permettant de visualiser, de classer et d'expliquer les données (OQLF, 2006). Il s'agit d'un processus cognitif itératif qui utilise différents opérateurs et qui vise à mieux faire connaître un phénomène et à le situer dans un plus grand ensemble, afin d'améliorer notre modèle cognitif de la réalité (cf. Figure 2-7) (Champoux, 1991). Ce processus cognitif permet à l'analyste de déduire l'information qu'il désire obtenir à partir des données présentées par son outil d'analyse (Caron, 1998).

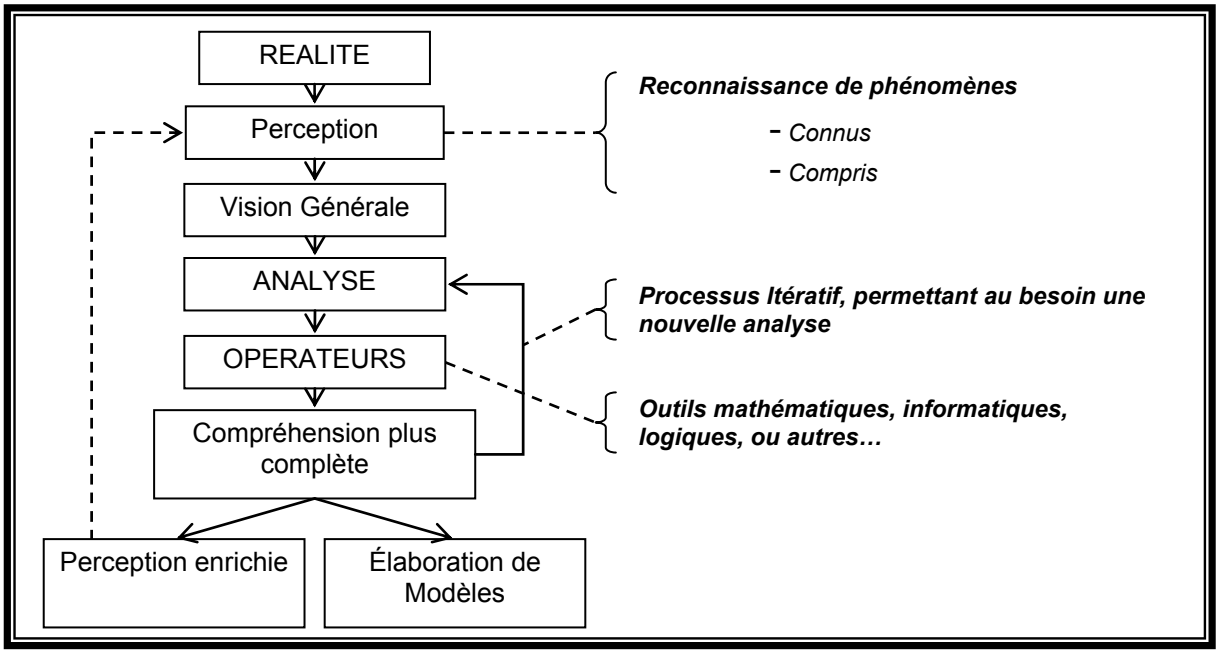


Figure 2-7 : L'analyse, de la réalité à la compréhension, adaptée de (Champoux, 1991)

2.3-1. Les différentes catégories d'analyses

L'**analyse spatiale** est un terme couramment utilisé dans les SIG. En 1987, Goodchild a défini l'analyse spatiale comme un ensemble de méthodes analytiques qui requiert l'accès à la localisation ainsi qu'aux attributs des objets analysés (Goodchild, 1987). En 1989, Aronoff complète cette définition en ajoutant que les fonctions d'analyse spatiale utilisent les attributs spatiaux et non spatiaux pour répondre à des questions sur le monde réel (Aronoff, 1989). En 1991, Champoux, définit l'analyse spatiale comme un « processus cognitif et itératif qui utilise différents opérateurs, dont au moins un spatial pour déduire

des caractéristiques descriptives ou spatiales d'un phénomène isolé ou regroupé, réel ou simulé dans l'espace. » (Champoux, 1991). Finalement, en 1995, Proulx précise l'analyse spatiale comme étant un « processus cognitif et itératif qui utilise différents opérateurs, dont *au moins un spatial, mais aucun temporel*, pour déduire les caractéristiques descriptives, spatiales ou temporelles d'un phénomène isolé ou regroupé, réel ou simulé dans l'espace » (Proulx, 1995). Cette dernière définition étant celle qui convient le mieux à l'archéologie, considérant leur nécessité à différencier l'analyse spatiale de l'analyse temporelle, nous retiendrons donc la définition de Proulx pour ce mémoire.

Parallèlement à l'analyse spatiale d'un objet qui porte sur la partie géométrique, l'**analyse temporelle** porte sur la datation, sur l'existence ou non, sur l'évolution ou non, de l'objet. L'analyse temporelle est une analyse de données en relation avec le temps. Cette analyse est très souvent utilisée par exemple dans le domaine de la criminalité [(Mena, 2002) : p47], mais aussi sur des modèles de simulation, sur des suivis de trafics routiers. Nous utiliserons la définition de Proulx pour définir l'analyse temporelle. L'analyse temporelle est un processus cognitif et itératif qui utilise différents opérateurs, *dont au moins un temporel, mais aucun spatial*, pour déduire les caractéristiques descriptives, spatiales ou temporelles d'un phénomène isolé ou regroupé, réel ou simulé dans le temps (Proulx, 1995).

Nous utiliserons aussi la définition de Proulx pour expliciter l'**analyse spatio-temporelle**. L'analyse spatio-temporelle est un processus cognitif et itératif qui utilise différents opérateurs, *dont au moins un temporel et au moins un spatial*, pour déduire les caractéristiques descriptives, spatiales ou temporelles d'un phénomène isolé ou regroupé, réel ou simulé dans l'espace et dans le temps (Proulx, 1995). Cette analyse spatio-temporelle est particulièrement importante en archéologie. Elle permet d'analyser simultanément l'espace et le temps, dans un contexte où l'un peut être influencé par l'autre.

L'**analyse descriptive** est d'abord un moyen de faire ressortir des unités statistiques. En effet, cette analyse a pour objectif de résumer quantitativement ou qualitativement l'information récoltée sur l'échantillon de données étudié. L'explication n'est alors pas

recherchée ; l'objectif est seulement de décrire, en dégageant les faits essentiels (Juillard, 2005; Sanaa, 2005). Traditionnellement, la donnée descriptive tient compte non seulement du contenu descriptif d'une donnée mais aussi du contenu temporel (Proulx, 1995; Korte, 2001; ESRI, 2006). L'archéologie nécessitant de traiter tant l'aspect spatial que temporel de la donnée, il est impérieux de différencier la temporalité de l'aspect descriptif des données. La notion de données descriptives existant dans la littérature, elle nous amène à une définition de l'analyse descriptive évidente : *analyse qui porterait sur les données descriptives* et donc qui ne porterait ni sur l'aspect spatial, ni sur l'aspect temporel de la donnée. Cependant, malgré une révision exhaustive de la littérature, nous n'avons trouvé aucune définition explicite sur l'analyse descriptive pour les bases de données autre que celle-ci. « Une analyse descriptive est un processus cognitif et itératif qui utilise les opérateurs descriptifs pour déduire les caractéristiques descriptives d'un phénomène » (Proulx, 1995). Par exemple, l'entité spatiale municipalité possède certains attributs descriptifs: nom, population, activité économique, etc. Il est donc possible d'effectuer une analyse descriptive du genre "Quelles sont les villes de plus de 100 000 habitants?" (Boursier, P., Mainguenaud, M., 1992).

La dernière catégorie d'analyse que nous allons présenter est **l'analyse visuelle ou graphique**. L'homme est meilleur pour interpréter visuellement des données que pour interpréter des chiffres [(Longley et al., 2005) : p273]. De plus, l'image tient une place importante dans la réflexion, l'interprétation et la compréhension car quelque soit la représentation (diagrammes, icones, géométries,…), ce sont des symboles qui ressemblent plus à ce qu'ils sont censés représentés que le texte. (Thagard *et al.*, 1997; Kovalerchuk *et al.*, 2005). L'expérience de générations de scientifiques comme Bohr, Einstein, Faraday et Watt ont montré que la représentation et le raisonnement visuels peuvent grandement améliorer la recherche de nouvelles hypothèses [(Kovalerchuk et Schwing, 2005) : p74]. L'analyse visuelle s'appuie sur ce que Bertin appelle la Graphique [(Bertin, 1992) : p20]. Elle poursuit deux objectifs : traiter les données pour comprendre et en tirer l'information et communiquer s'il y a lieu cette information ou inventaire de données élémentaires. Ces deux objectifs sont ce que Bertin appelle la Graphique de traitement et la Graphique de communication. La *Graphique de traitement* comporte deux impératifs auxquels la

Rapport-Gratuit.com

graphique de communication n'est pas soumise: elle doit transcrire toutes les données du tableau, c'est-à-dire les données exhaustives et elle doit répondre à toutes les questions pertinentes et permettre de simplifier les deux composantes du tableau de données. La *Graphique de communication* est un moyen de fixer et de dire aux autres ce que l'on a découvert d'une manière simple et rapide. Elle utilise les propriétés de l'image visuelle pour faire apparaître les relations de ressemblance et d'ordre entre les données. Elle permet alors une perception rapide et éventuellement la mémorisation de l'information d'ensemble.

2.3-2. Les opérateurs d'analyse:

Les opérateurs sont des outils mathématiques, informatiques, logiques ou autres qui permettent à l'utilisateur d'effectuer des requêtes (Champoux, 1991). Un requête est un ensemble de commandes dont l'exécution permet d'obtenir un résultat (OQLF, 2006). Plus spécifiquement, un opérateur est l'élément qui exécute un calcul (Ginguay *et al.*, 1993) lors d'analyses descriptive, spatiale, temporelle ou spatio-temporelle (Proulx, 1995). A cette liste d'opérateurs, on peut rajouter les opérateurs d'analyse graphique ou visuelle ce qui permet alors de distinguer 4 grandes familles d'opérateurs : les opérateurs spatiaux, les opérateurs temporels, les opérateurs descriptifs et les opérateurs visuels.

Les opérateurs spatiaux

D'après l'OQLF (2006), un opérateur spatial est un opérateur permettant de traiter la position et la forme d'une entité spatiale, ou les relations spatiales qui existent entre les entités spatiales. Que l'on travaille avec des données 2D ou des données 3D, il existe deux grandes familles d'opérateurs spatiaux : (Champoux, 1991; Bédard, 2003b) : les opérateurs spatiaux métriques et les opérateurs spatiaux topologiques.

Les **opérateurs spatiaux métriques 2D ou 3D** sont des opérateurs qui permettent de traiter tout ce qui se mesure (Golay *et al.*, 2001). D'une part, on y retrouve les simples mesures comme une distance, une direction, une moyenne, un calcul de périmètre, de surface, de volume,… D'autre part, on y retrouve aussi tout ce qui peut toucher à l'analyse de position complexe notamment les rayons d'inclusion, les corridors, les enveloppes,…

Les **opérateurs spatiaux topologiques 2D et 3D** permettent de traiter les relations spatiales établies entre des entités spatiales (OQLF, 2006). Ceux-ci sont utilisés pour établir la présence ou l'absence de certaines relations spatiales entre les entités géométriques. (Champoux, 1991) La topologie est une branche des mathématiques traitant des relations de voisinage entre des figures géométriques, relations qui ne sont pas altérées par la déformation des figures (OQLF, 2006). Cela signifie que même après des transformations comme une translation, une rotation ou un changement d'échelle, la propriété topologique des figures restent invariante (Lachance, 2005). Kemp complète en spécifiant que dans le contexte d'une structure vectorielle, la topologie est la description explicite de relations spatiales entre des entités stockées (Kemp, 2006). Dans les SIG, la topologie peut être utilisée pour valider la géométrie vectorielle (c'est à dire, vérifier qu'un des polygones se ferme et que toutes les lignes du réseau se ferment ensemble) et pour certains types d'opérations (analyse réseau et test d'adjacence) [(Longley et al., 2005) : p186].

Il existe différentes manières de définir le nombre de relations topologiques entre des objets (Pullar *et al.*, 1988; Egenhoffer *et al.*, 1991; Clementini *et al.*, 1993). D'après Normand et al., le modèle reconnu par la norme ISO TC 211 – norme ISO 19107 sur le schéma spatial (www.iso.org) et basée sur le concept d'Egenhofer (Egenhoffer et Herring, 1991; Egenhoffer *et al.*, 1992; Egenhoffer *et al.*, 1994) – peut être regroupé en 5 catégories (Normand et al., 2003). La **disjonction** est utilisée lorsqu'il n'y aucun contact entre l'intérieur et la limite des objets. L'**adjacence** est utilisée lorsqu'il y a contact avec que les limites des objets. L'**intersection** est utilisée lorsqu'il y a contact entre l'intérieur d'un objet et la limite avec (intersection intérieure) ou sans (intersection limite) l'intérieur des autres objets. L'**inclusion** est utilisée lorsqu'un objet contient complètement un autre objet, avec une limite commune (inclusion limite) ou non (inclusion totale). L'**égalité** est utilisée lorsque les deux objets se recouvrent exactement.

Le logiciel *ArcMap* d'ESRI (www.esri.com) avec les opérateurs topologiques « Intersect, are completely within, completely contain, touch the boundary of, …» ou MGE d'intergraph (www.intergraph.com) avec les opérateurs « overlaps, entirely contains, passes trough, meets, Contained by, on boundary of,… » ou le logiciel Map Info

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

(www.mapinfo.com) avec les opérateurs « contains, within, contains entire, contains part, intersect,… » sont autant de preuves de la présence des logiciels actuels à gérer les opérateurs spatiaux topologique 2D. En 3D, les opérateurs topologiques, en théorie, sont relativement similaires avec le modèle en 9 intersections d'Egenhofer. Cependant, ils sont bien plus difficiles à mettre en place puisqu'ils exigent une structure topologique 3D de l'objet, plus complexe qu'une structure topologique 2D (De La Losa, 2000; Zlatanova, 2000; Zlatanova et al., 2002; Billen *et al.*, 2003; Ramos, 2003; Apel, 2004; Zlatanova et al., 2004; Lachance, 2005; Pouliot, 2005; Pouliot et al., 2006).

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

Les opérateurs temporels

Pour que des relations temporelles puissent être considérées entre deux objets, il faut supposer le temps T comme quelque chose de continu. L'objet temporel est alors considéré dans ce temps définit. Il est caractérisé par un intervalle temporel I définit par deux instants t1 et t2 (t1<t2) (Claramunt *et al.*, 2000). Les opérateurs temporels sont alors basés sur le même principe que les opérateurs spatiaux. [(Swiaczny *et al.*, 2001) : p139]. On peut donc étendre la classe opérateurs métriques et topologiques aux requêtes temporelles (Gagnon, 1993; Bédard, 2003b).

Gagnon (1993) les **opérateurs métriques temporels** comme la position d'un objet dans le temps ou la durée de celui-ci. A l'instar des opérateurs spatiaux topologiques, les opérateurs temporels topologiques concernent également la relation immuable, au niveau temporel, entre les objets. Les concepts d'Allen (Allen, 1984) sont souvent utilisés pour définir les relations topologiques temporels (equal, before after, meets, overlaps, during, starts, finishes) et leurs inverses respectifs à l'exception d' « equal » qui est un opérateur symétrique (after, met, overlapped, contain, started, finished) (cf. Figure 2-8) (Claramunt et Jiang, 2000; Jian-feng *et al.*, 2005).

	Début(i_1)<Début(i_2) ET Fin(i_1)<Fin(i_2)	i_1 *égale* i_2	Légende : Ligne du temps
	Fin(i_1)<Début(i_2)	i_1 *avant* i_2 i_2 *après* i_1	: Intervalle i_1
	Fin(i_1)<Début(i_2)	i_1 *rejoint* i_2 i_2 *est rejoint par* i_1	: Intervalle i_2
	Début(i_1)<Début(i_2)<Fin(i_1) ET Fin(i_1)<Fin(i_2)	i_1 *recouvre* i_2 i_2 *est recouvert par* i_1	
	Début(i_1)>Début(i_2) Fin(i_1)<Fin(i_2)	i_1 *pendant* i_2 i_2 *contient* i_1	
	Début(i_1)=Début(i_2) ET Fin(i_1)<Fin(i_2)	i_1 *débute* i_2 i_2 *a débuté* i_1	
	Début(i_1)>Début(i_2) ET Fin(i_1)=Fin(i_2)	i_1 *termine* i_2 i_2 *a terminé* i_1	

Figure 2-8 : Les opérateurs temporels topologiques

Les opérateurs descriptifs

Proulx (1995) a défini les opérateurs descriptifs comme des opérateurs utilisant des attributs descriptifs de la base de données comme critères de sélection.

On distingue deux catégories d'opérateurs descriptifs (Proulx, 1995; Bédard, 2003b) :

- les opérateurs mathématiques qui sont appliqués exclusivement aux valeurs quantitatives. On dit d'une valeur qu'elle est quantitative si cette valeur peut-être mesurée ou comptée (OQLF, 2006). Ces valeurs, numériques seront par exemple de type *float*, *double* ou *integer* (Diansheng, 2003). Les opérateurs « +, -, *, / » sont des exemples d'opérateurs mathématiques
- les opérateurs logiques qui sont appliqués aux valeurs qualitatives mais aussi quantitatives. Les valeurs qualitatives peuvent être soit de nature nominale (les valeurs sont des noms dont l'ordre n'a pas d'incidence comme *commercial, résidentiel, industriel*) ou de nature ordinale (les valeurs sont des noms qui suivent une progression comme *pauvre, moyen riche*) (Bédard, 2003a; Pouliot, 2003). Ces valeurs, alphabétiques, seront de types *char* ou *varchar* (Diansheng, 2003). Les opérateurs « <, >, =, <=, >=, <> » sont des exemples d'opérateurs logiques. Les valeurs qualitatives de nature nominale utiliseront exclusivement les opérateurs logiques « = » et « <> ». De plus, d'après Proulx (1995), les opérateurs logiques comprennent les opérateurs ensemblistes (ex: "and", "or").

Les fonctions de visualisation de l'information spatiale

L'interprétation visuelle des données peut se faire à l'aide d'opérateurs ou des fonctions de visualisation de l'information spatiale (Golay et Caloz, 2001). La liste d'opérateurs visuels ne peut être ni exhaustive ni hiérarchisée (du plus important au moins important) en raison de l'aspect subjectif du sens de la vue ; chaque personne ayant sa propre perception et son propre sens de l'analyse visuelle. Cependant, on peut retenir les principales fonctions de visualisation suivantes :

- l'élaboration de cartes thématiques – avec l'importance de la légende et de l'orientation
- la visualisation tridimensionnelle – il est estimé que 50 % des neurones humains sont réservé à la vision et que l'affichage 3D stimule plus de neurones, utilisant

ainsi une plus grande portion du cerveau permettant la résolution plus rapide et facile des problèmes (Swanson, 1996).

- la navigation – zoom et déplacement de l'écran (cf. Figure 2-9)
- l'utilisation de graphiques et d'histogrammes (cf. Figure 2-10)
- l'usage d'une sémiologie graphique appropriée ; celle-ci remplace l'inventaire classique des formules graphiques par une analyse des moyens et des buts et par un ensemble de règles impératives qui commandent la rédaction graphique, c'est-à-dire le choix des correspondances entre les sensibilités visuelles disponibles et les éléments de l'information. (Bertin, 2005).

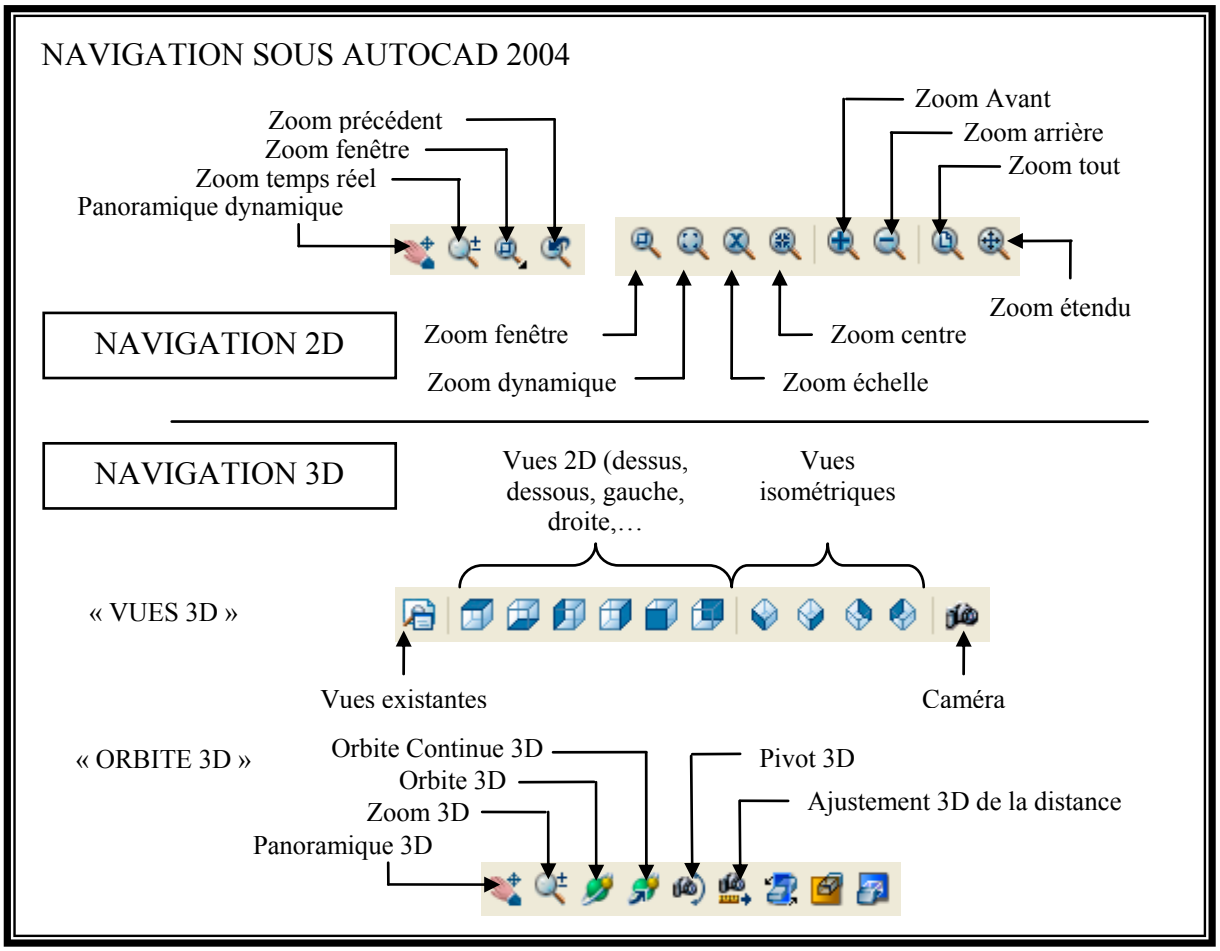


Figure 2-9 : Navigation : zoom et déplacement de l'écran - exemple d'Autocad 2004

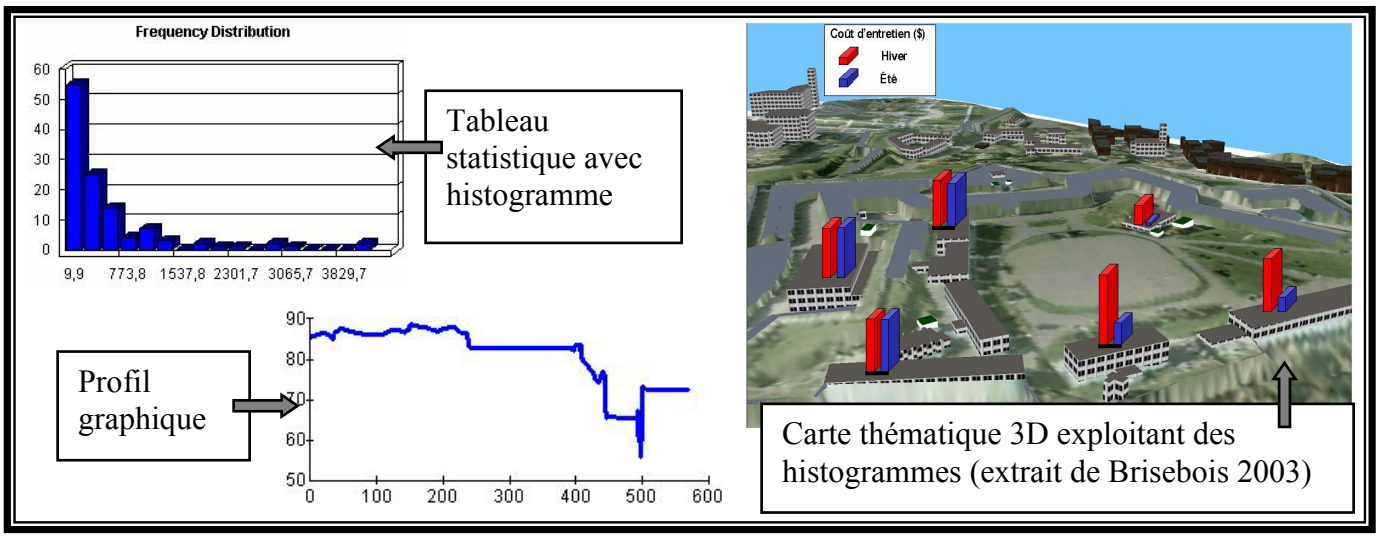


Figure 2-10 : Exemple d'utilisations de graphiques et d'histogrammes

Cette deuxième partie a pu montrer les différentes notions de l'analyse de données et des opérateurs utilisables pour effectuer ces analyses. Notre recherche bibliographique a permis de mettre en avant les principales notions afin de pouvoir optimiser et rechercher de nouveaux concepts d'analyse de données 3D. De surcroît, nous nous sommes intéressés principalement aux analyses spatiales et temporelles (plus spécifiquement aux analyses topologiques) puisque ces analyses intéressent particulièrement les archéologues avec l'analyse stratigraphique.

2.4-Les outils d'exploitation des données spatiales

On entend par le terme « exploitation de données », une exécution des instructions d'un programme par l'unité centrale, qui se traduit en une série d'opérations logiques ou d'opérations de calcul effectuées sur des données ou des informations, entre le moment où celles-ci sont entrées dans un système informatique et celui où elles en sortent (OQLF, 2006). Le terme « traitement de données » (ou traitement des données) est également employé pour désigner le déroulement logique de ces opérations (collecte, enregistrement, fusion, tri, recherche, affichage, modification, impression, etc.), selon un processus prédéterminé (OQLF, 2006). L'exploitation des données passe donc par la gestion (ajout, suppression, modification), l'affichage et l'analyse (descriptive, spatiale, temporelle et

visuelle). Les développements informatiques montrent qu'il existe deux grandes catégories de systèmes : les systèmes transactionnels et les systèmes analytiques.

2.4-1. Les systèmes transactionnels

Les systèmes transactionnels sont des systèmes informatiques dont le fonctionnement repose essentiellement sur le traitement transactionnel, c'est-à-dire sur le traitement immédiat des données (ajout modification suppression), suite à une demande ou à une intervention de l'utilisateur, chaque transaction étant vue comme un ensemble de sous-opérations qui forme un tout fini, avec un début et une fin. La différence entre les concepts de « traitement transactionnel », « traitement transactionnel en ligne » (« OLTP ») et « traitement en temps réel » est si mince, que, la plupart du temps, on trouve les termes « système transactionnel » et « système OLTP » (« On-line Transaction Processing System » (OQLF, 2006).

L'un des systèmes transactionnels les plus connus dans le monde de la géomatique est le **Système d'Information Géographique (ou SIG)**. En effet, le SIG est couramment utilisé pour analyser les données, et les récents développements dans ce domaine sont trop volumineux pour les énumérer (Longley et al., 2005; Nyerges, 2006), notamment puisque les domaines d'applications, dont l'archéologie fait partie (Wheatley et Gilling, 2002), sont riches et vastes. De plus, l'indication du nombre de logiciels[3] et des ouvrages sur le sujet[4] en est une aussi grande preuve. Les SIG sont un outil performant pour : acquérir, gérer, traiter, analyser et diffuser des données géographiques (Bédard, 2003a; Nyerges, 2006). En effet, afin de permettre au système transactionnel d'effectuer l'intégralité des transactions (ajout, modification, suppression) de façon facile, rapide et sûre, la plupart de ces systèmes sont implantés selon une structure relationnelle normalisée de telle sorte que la duplication des données soit à son minimum.

La littérature a montré que les systèmes transactionnels de type SIG performent mieux pour la manipulation de données que pour l'interrogation spatiale (Caron, 1998; Bédard, 2005;

---

[3] Plus de 70 répertoriés par Geoplan (http://www.geoplan.ufl.edu) en 2001

[4] Sur books.google.com la recherche « GIS » offre plus d'une dizaine de pages d'ouvrages ayant le mot « GIS » dans son titre.

Nyerges, 2006). En effet, les SIG ont aussi une faiblesse lorsqu'il s'agit de traiter un gros volume de données ce qui d'après Caron arrive lorsque les données méritent d'être analysées (Caron, 1998). De plus, pour éviter la duplication des données, normaliser la structure relationnelle revient souvent à augmenter le nombre de tables. Ainsi un nombre important de tables (et donc de jointures entre les tables) et un gros volume de données sont souvent les raisons de la faiblesse des systèmes transactionnels pour analyser les données. En effet, principalement pour des requêtes de types agrégatives et pour celles nécessitant un nombre élevé de jointures entre les tables, une lenteur dans les réponses est toujours constatée lorsque vient le temps d'interroger les données (qui se chiffre en jours plutôt qu'en minutes) (Bédard, 2005; Nyerges, 2006). Ces mêmes requêtes complexifient aussi l'élaboration des requêtes (Bédard, 2005). En effet, même si les dernières années ont vu plusieurs travaux de recherche visant à optimiser et à accélérer l'interrogation, à partir d'un SIG, des bases de données, en utilisant par exemple les requêtes graphiques (Aufore-Portier, 1995; Favetta *et al.*, 2000), une majorité d'outils SIG nécessite encore que l'utilisateur maîtrise tant le langage d'interrogation comme SQL (Egenhoffer *et al.*, 1993) que la structure de la base de données (Langran, 1992).

Bien que les SIG aient d'abord été construits pour gérer des données 2D, ils restent cependant performants quant à la représentation tridimensionnelle des données (Brisebois, 2003). Des effets de perspective et d'ombrage à l'utilisation possible de lunette stéréoscopique en passant par une possibilité d'animation 3D, rendent les SIG comme un outil de visualisation 3D opérationnel. Mais, bien que les SIG offrent une analyse visuelle performante (Nyerges, 2006) et une analyse spatiale bien appropriée au 2D, les SIG commerciaux actuels, aptes à manipuler les données 3D (par exemple : ArcGis - www.Esri.com, MapInfo – www.mapinfo.com, Geomedia – www.intergraph.com), sont plutôt limités à l'analyse des surfaces et à la visualisation 3D. De plus, toutes les analyses se font en 2.5D et non dans un véritable environnement 3D (Tet Kuan *et al.*, 2005) car les SIG, bien qu'ils permettent d'accomplir plusieurs tâches importantes offrent très peu de possibilités de représentation 3D principalement en raison d'une difficulté à structurer les données 3D tant spatiales que descriptives (Lachance, 2005) et à leur incapacité à gérer la plupart des primitives géométriques non planaires (MNT non compris) (Brisebois, 2003).

2.4-2. Les systèmes d'analyse

Les outils d'exploitation de données spatiales devraient servir de support au processus de décision des organisations et permettre de gérer l'historique des données, fournir des réponses rapides et des interfaces à l'usager faciles à utiliser (Bédard, 2005). Les outils qui exploitent les entrepôts de données, le marché de données, les requêteurs et rapporteurs, le forage automatique de données spatiales et les outils basés sur l'approche multidimensionnelle sont des exemples de systèmes d'analyse. L'outil d'analyse, à la manière d'un processus cognitif, doit présenter les données dans une forme proche de la façon dont l'analyste conçoit mentalement le phénomène à analyser. Le travail de déduction qu'il doit effectuer est moins laborieux et son analyse est facilitée (Caron, 1998). De plus, d'après Caron (1998), la **rapidité** et la **facilité** sont les deux conditions que doit rencontrer un système d'analyse pour ne pas nuire au processus cognitif d'analyse de l'usager tel que définit par Champoux (1991). La rapidité permet à l'usager de garder son attention dans sa recherche et son interrogation des données et son processus cognitif (cf. Figure 2-7) n'est pas interrompu. La facilité permet à l'usager de se concentrer sur l'analyse en cours et non sur la manière de formuler son interrogation.

La structure multidimensionnelle : l'analyse rapide et facile des données

Parce qu'elle offre une perspective d'évolution intéressante et qu'une première tentative d'application à l'archéologie a été réalisée (Rageul, 2004), **l'approche multidimensionnelle** sera examinée. À l'intérieur d'une base de données destinée à l'analyse, la dénormalisation des données est utilisée afin de limiter le nombre de tables et le nombre de jointures nécessaires entre les différentes tables dans le but d'améliorer la rapidité d'exécution des requêtes. A la différence des structures transactionnelles qui structurent les données sous la forme de listes unidimensionnelles, la structure multidimensionnelle gère les données sous forme de matrices multidimensionnelles où chaque axe représente un phénomène différent (Caron, 1998) (cf. Figure 2-11).

Rapport-Gratuit.com

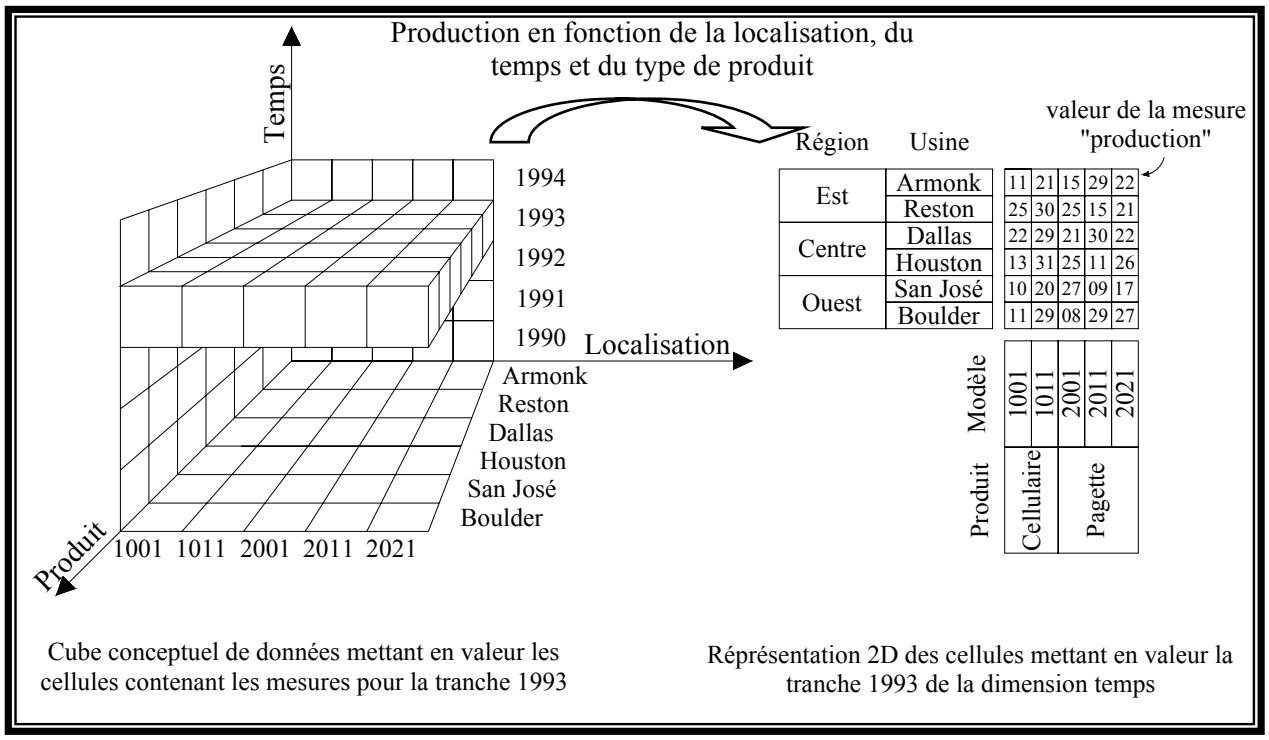


Figure 2-11 : Exemple de cubes, dimensions et mesures inspiré de (Caron, 1998; Bourgon *et al.*, 1999)

Les thèmes d'analyses, ou **dimensions,** sont définis comme les « axes d'analyse selon lesquels on veut étudier des données observables qui, soumises à une analyse multidimensionnelle, donnent aux utilisateurs des renseignements nécessaires à la prise de décision » (OQLF, 2006). D'après la Figure 2-11, la « localisation », le « temps » ou le « produit » sont des exemples de dimensions. Ces dimensions peuvent être hiérarchisées à plusieurs niveaux d'agrégations. (Niveau 1 : Région [Est; Centre; Ouest] - Niveau 2 : Usine [Armonk, Dallas,,…] pour la dimension *Localisation* ou Niveau 1 : Produit [Cellulaire, Pagette] – Niveau 2 : Modèle [1001, 1011,…] pour la dimension *Produit*). À la différence des dimensions où les données peuvent être agrégées, les **mesures** sont les variables non-agrégeables qui font l'objet de l'analyse (dans l'exemple choisi, « production » est une mesure). Elles permettent d'assurer l'unicité de chaque enregistrement (ligne). Finalement, le **Fait** représente la valeur d'une mesure. Il correspond à l'association entre les mesures et les combinaisons de un ou plusieurs membres d'une ou plusieurs dimensions. Dans notre exemple, un fait serait : « L'usine de Dallas de la région du centre a produit 22 cellulaires du modèle 1001 durant l'année 1993 ».

Les **outils OLAP** ou On-Line-Analytical Processing sont définis comme une « catégorie de logiciels axés sur l'exploration et l'analyse rapide des données selon une approche multidimensionnelle à plusieurs niveaux d'agrégation » (Caron, 1998). L'avantage de tels outils renvoi directement aux deux conditions que doit rencontrer un système d'analyse pour ne pas nuire au processus cognitif d'analyse de l'usager ; la facilité et la rapidité. Comme l'analyse se fait à l'aide des dimensions, l'usager n'a plus à maîtriser les langages d'interrogation. Il peut alors interroger les données de manière plus intuitive contrairement aux interfaces souvent complexes des autres systèmes. Quant à la rapidité, les outils OLAP exploite une dénormalisation maximale des données, sous la forme de pré-agrégation stockée. L'usager peut alors obtenir ses réponses dans la seconde même lorsqu'il s'agit d'interrogation portant sur les agrégations et pas seulement sur les données détaillées. Concernant l'analyse de données, il n'y a pas de différence entre un débutant, un expérimentaliste ou un analyste car le principe très intuitif d'outils d'analyse en ligne permet à un n'importe quel utilisateur de devenir opérationnel rapidement.

Les outils OLAP utilisent des opérateurs visuels particuliers afin de « naviguer » dans les cubes multidimensionnels (Rivest, 2000) :

- *Pivoter (pivot, swap)* : Permet d'interchanger deux dimensions. Cette opération est nécessaire comme toutes les dimensions ne peuvent être visualisées en même temps
- *Forer (drill-down)* : Permet de descendre dans la hiérarchie de la dimension, comme par exemple passer d'une analyse des usines du « Centre » à l'analyse des usines de « Dallas »
- *Remonter (drill-up, roll-up)* : Permet de remonter dans la hiérarchie de la dimension et passer d'un niveau détaillé à un niveau plus agrégé, comme par exemple passer d'une analyse des modèles « 1011 » à l'analyse du produit « Cellulaire »
- *Forer latéralement (drill-across)* : pour soit passer d'une mesure à l'autre comme par exemple, visualiser le nombre de personnes travaillant le produit par site au lieu du nombre de produits par sites ou soit passer d'un membre de dimension à un autre comme par exemple, visualiser les données de « Dallas » au lieu de celles de « San José »

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

Une base de données multidimensionnelle à référence spatiale intègre une ou des classes d'objets spatiaux sous la forme de dimensions spatiales (les membres hiérarchiques de la dimension peuvent avoir ou non une géométrie associée) ou de mesures spatiales dont l'objectif est de pointer vers les objets spatiaux. (Brisebois, 2003). L'avantage d'utiliser une approche multidimensionnelle spatiale avec plusieurs dimensions spatiales, permet à l'utilisateur d'user d'opérateurs visuels, non seulement sur des tableaux ou des graphiques comme le proposerait les outils OLAP, mais aussi sur des occurrences cartographiques, principalement si la dimension spatiale est une dimension spatiale géométrique (i.e. tous les membres de chaque niveau possède une primitive géométrique). Tels sont les concepts des **systèmes SOLAP** ou Spatial OLAP qui permettent alors d'exploiter la composante spatiale des données. Bédard définit ainsi l'outil SOLAP comme « une plate-forme visuelle supportant l'exploration et l'analyse spatio-temporelle faciles et rapides des données selon une approche multidimensionnelle à plusieurs niveaux d'agrégation via un affichage cartographique, tabulaire ou en diagramme statistique » (Bédard, 2005).

La structure multidimensionnelle : la faiblesse pour la gestion des données en temps réel et l'exploitation de la 3ème dimension

Un **système d'analyse temps réel** doit non seulement offrir une compatibilité cognitive avec les utilisateurs potentiels (performance du système : accès rapide et facile à l'interrogation de données), mais il doit aussi leur permettre de pouvoir profiter d'un accès à l'information courante en tout temps (White, 2002). A ce titre, certaines organismes nécessitent un outil d'analyse avec une actualisation plus importante que celle proposée par ce dernier qui se résume généralement à une actualisation périodique (mensuelle, hebdomadaire et rarement quotidienne) et non sur demande. Traditionnellement, il existe aucune connexion directe entre le système d'acquisition et le système d'analyse (Chountas *et al.*, 2004). Cependant, les travaux de Lambert (Lambert, 2005) et White (White, 2002), ont permis de justifier la pertinence d'un outil d'analyse en ligne en temps réel. Cela signifie qu'avec une structure multidimensionnelle en temps réel, le cube ramasserait les données issues de la transaction de façon instantanée (synchronisation entre le système d'acquisition et le système d'analyse concernant l'ajout des données), au lieu d'user d'un processus explicite [(Lachey, 2005) : p29 et p 43]. Cependant, en pratique, l'actualisation des données dans une structure multidimensionnelle correspond juste à l'ajout de données

avant une phase d'analyse et non une gestion entière des données (modification ou suppression).

Concernant **l'exploitation de la troisième dimension**, notre revue exhaustive de littérature indique qu'il ne semble y avoir eu que les travaux de Brisebois relatifs au SOLAP 3D (Brisebois 2003). Tout d'abord, les caractéristiques de l'outil SOLAP 3D sont composées d'une extension des caractéristiques d'outil SOLAP 2D :

1. un traitement de tous les types de primitives géométriques,
2. une analyse spatiale possible.

Les caractéristiques de l'outil SOLAP 3D a aussi des caractéristiques qui lui sont propres :

3. une présence de vues 3D,
4. une navigation tridimensionnelle au sein des vues 3D,
5. une navigation assistée pour explorer le résultat d'une requête, comme le positionnement automatique de l'usager sur le résultat de la requête,
6. une aide à l'orientation,
7. une glissière altimétrique,
8. une source d'éclairage,
9. une fenêtre de manipulation d'objet de petite taille,
10. une génération de vues 2D à partir d'une vue 3D (« slicer »),
11. une possibilité d'impression d'immersion

Cependant, le prototype proposé par Brisebois, présente deux volets : un volet descriptif et un volet cartographique. Ce dernier exploite la composante spatiale géométrique des dimensions et des mesures grâce à des outils de visualisation cartographique de SIG. Or, comme nous l'avons vu à la section 2.4-1. , les SIG sont encore assez limités quant à la gestion explicite de la 3e dimension même s'ils permettent la visualisation dans un univers 3D.

2.5-Conclusion du chapitre

Ce chapitre a pu couvrir l'ensemble des notions que nous avons jugées importantes et pertinentes pour comprendre, optimiser et expérimenter l'analyse de données qui plus est 3D. Le contexte archéologique pose plusieurs défis à la géomatique.

Le premier défi concerne l'aspect tridimensionnel des données. Face à une diversité dans la notion de 3D, nous avons associé, pour le présent mémoire, le terme d'objet 3D à tout objet à trois dimensions géométriques (longueur, largeur et hauteur) et positionné dans un univers 3D (x, y, z) et référencés temporellement. La première partie de ce chapitre a aussi mis en avant le rôle prépondérant de la donnée dans la représentation numérique de cet objet 3D.

Le deuxième défi que le contexte archéologique lance à la géomatique concerne la nécessité d'analyser simultanément l'espace et le temps, dans un contexte où l'un peut être influencé par l'autre, et de le faire rapidement et facilement dans un environnement tridimensionnel afin de mieux supporter l'interprétation des données. Ainsi, la deuxième partie de ce chapitre a apporté les principales notions théoriques et pratiques concernant l'analyse des données. De surcroît, nous nous sommes intéressés principalement aux analyses spatiales et temporelles (plus spécifiquement aux analyses topologiques) puisque ces analyses intéressent particulièrement les archéologues avec l'analyse stratigraphique. La troisième partie de ce chapitre a aussi permis de mettre en évidence deux approches globales pour exploiter les données spatiales. Ces deux approches se distinguent essentiellement lorsqu'il s'agit de traiter un gros volume de données tridimensionnelles. **Les systèmes transactionnels** sont des systèmes informatiques dont le fonctionnement repose essentiellement sur le traitement immédiat des données (ajout, modification et suppression) au détriment d'une analyse rapide et facile. **Les systèmes analytiques**, quant à eux, à la manière d'un processus cognitif, doivent présenter les données dans une forme proche de la façon dont l'analyste conçoit mentalement le phénomène à analyser.

Chapitre 3 : L'optimisation des outils d'analyse en ligne

3.1-Introduction

Suite aux éléments mis en évidence dans le chapitre 2 quant à l'identification des forces et des faiblesses des technologies actuelles pour l'analyse de données spatiales, nous avons décidé de procéder à la sélection d'une approche d'analyse qui sera par la suite optimisée pour un contexte particulier, soit l'archéologie de terrain.

L'archéologie est une discipline des sciences humaines/sociales dont l'objet d'étude est l'ensemble des vestiges matériels laissés par les anciennes sociétés humaines (Barceló, 2002). Dans bien des cas, pour étudier ces vestiges, les archéologues doivent dans un premier temps les retirer de terre. Ils utilisent alors une technique précise : la fouille. Lors de ce processus d'acquisition de leurs données, que ce soient des structures ou des objets, ils doivent tenir compte de leur localisation exacte au sein de la superposition de différentes couches de terre (= stratigraphie) formant le site archéologique qu'ils fouillent. Aussi, pour comprendre les relations stratigraphiques entre les différents vestiges, l'archéologue recherche un outil d'analyse performant qui répondrait entièrement à ses besoins.

Suite à diverses rencontres avec des archéologues, l'analyse de leurs besoins a montré que, globalement, les archéologues désirent interroger fréquemment la base de données et visualiser, par une représentation cartographique tridimensionnelle, les « unités de fouille » pour faciliter leur interprétation. Des requêtes complexes ont été pointées comme d'intérêt pour eux. Des requêtes de type :

- Comparer les unités de fouille entre elles en fonction de leur couleur, de leur granulométrie, de leur consistance et de leur position géographique et stratigraphique avec en plus la possibilité de comparer ces unités avec une céramique particulière (type et temporalité) contenue dans celles-ci »
- Où se trouvent les unités de fouille dont la matrice (la terre) est gris claire, loameuse et molle et qui contiennent des fragments de col de jarre (céramique) du bronze ancien au sein d'un contexte majoritairement de l'âge bronze ?

Ces requêtes sont complexes car :

- elles font intervenir plusieurs catégories de données (ex : les unités de fouille, la céramique) et des attributs différents dans une même requête ;
- elles font aussi intervenir tant des données détaillées (ex : l'âge de bronze ancien,) que des données agrégées (ex : l'âge de bronze est l'agrégation de l'âge de bronze ancien, moyen et récent - la consistance molle est l'agrégation de la consistance meuble, friable, friable variée)

Le contexte de fouille archéologique propose donc des défis intéressants et particuliers pour les outils d'analyse de données spatiales, notamment la nécessité:

- d'analyser simultanément l'espace et le temps, dans un contexte où l'un peut influencer l'autre ;
- de ne pas être ralenti par le logiciel d'analyse dans leur processus analytique, processus jugé très intuitif, peu structuré et peu typé ;
- de faire évoluer les données dans le système (les interprétations de l'archéologue suivant des heuristiques qui ne peuvent pas toujours être formalisées de façon absolue) ;
- de posséder une interface d'analyse conviviale, simple et accessible par des non spécialistes en informatique ;
- puis d'exploiter au mieux les environnements tridimensionnels afin de mieux supporter l'interprétation des données, la création d'hypothèses et la découverte de nouvelle connaissance archéologique.

Cette analyse des besoins nous a donc amené à identifier trois critères retenus pour faire la sélection de l'approche d'analyse: la rapidité et la facilité d'utilisation, la possibilité de faire évoluer les données dans le système et la visualisation tridimensionnelle.

Les éléments du chapitre 2 ont mis en évidence que les systèmes basés sur une approche transactionnelle sont optimisés pour certaines opérations telles l'ajout, la suppression et la modification de données. Ainsi, la structure transactionnelle n'est pas optimisée pour la

découverte de connaissances en raison d'une complexité récurrente dans les requêtes et une lenteur dans les réponses.

Nous avons aussi vu que de leur côté, les systèmes basés sur une approche analytique multidimensionnelle sont optimisés pour une analyse dite multidimensionnelle où, grâce à une base de données dénormalisée et aux axes d'analyse (les dimensions), les requêtes même celles de types agrégatives sont simples et leurs réponses sont rapides (dans la seconde). Cependant, les systèmes analytiques sont conçus de telle manière à ne pas permettre la modification de l'organisation conceptuelle de la base de données. Ainsi, afin de conserver l'intégrité de la base de données, les systèmes analytiques permettent aux utilisateurs d'interroger leur base de données pour faire ressortir l'information rapidement mais ne peuvent pas leur permettre de réécrire ou modifier leurs données. Notre revue de littérature spécifique à l'examen de travaux incorporant les concepts d'évolution de données, nous indique que des travaux ont eu lieu afin d'optimiser la mise à jour des données avec la notion de SOLAP temps-réel (Lambert, 2005). Cependant, cette mise à jour représentait exclusivement l'ajout de données avant une phase d'analyse. De plus, les manipulations restaient complexes pour, par exemple, modifier la hiérarchie d'une dimension ou bien tout simplement prendre en compte l'évolution des données surtout si celles-ci sont sujettes à des modifications récurrentes comme le sont certaines données archéologiques.

Du point de vue de la gestion de 3e dimension, nous avons vu que les systèmes transactionnels ont été construits pour gérer des données 2D. Même s'ils permettent la visualisation dans un univers 3D, ils sont encore assez limités quant à la gestion explicite de la 3e dimension, sauf dans le cas spécifique des MNT où ces derniers représentent un assemblage de surfaces (i.e. des triangles). Les systèmes analytiques ont aussi été construits pour gérer des données 2D. Notre revue de littérature spécifique à l'examen de travaux incorporant la 3e dimension dans les systèmes analytiques nous a révélé qu'il n'y avait que les travaux de Brisebois, relatifs au SOLAP 3D (Brisebois 2003). Cependant, d'après le prototype conçut par Brisebois (2003), le volet cartographique étant supporté par un SIG, la gestion du 3D n'était alors pas optimale.

A partir de ces éléments, nous avons donc décidé d'examiner de plus près les systèmes analytiques car ils ont été jugés plus adéquats dans le contexte d'analyse de données issues d'une fouille archéologique (Rageul, 2004). Reste maintenant à l'optimiser. Pour cela, nous avons identifié deux grands axes jugés possédant des perspectives de développement intéressant soit : le processus de révision des données et l'introduction de la 3e dimension. Ces deux axes marquent en effet les principales faiblesses de l'approche analytique.

3.2- La révision des données d'interprétation : nouveaux concepts

3.2-1. Principe

Dans le contexte des bases de données spatiales, la **mise à jour** de données consiste à changer la valeur des donnée spatiales pour les faire correspondre le plus possible à l'état actuel de la réalité (Pouliot et al., 2004). Ce changement peut se concrétiser par un ajout, une modification, une suppression de données. La **correction des données,** quant à elle, consisterait à remplacer une donnée que l'on pense erronée par une nouvelle donnée que l'on sait meilleure en se basant sur des croyances assurées.

La **« modification volontaire »** de données se distinguerait de la mise à jour par le fait que celle-ci n'implique pas un changement de la réalité ou une erreur de manipulation comme c'est le cas pour la correction de données. Cette modification se ferait simplement pour « *voir ce que cela donnerait* » et revenir éventuellement à l'ancienne donnée. Ce processus de modification volontaire est très important dans une démarche de recherche et de découverte de nouvelle connaissance, comme c'est entre autres, le cas en archéologie. En effet, la fouille archéologique est une action « destructrice ». Il n'y a donc pas de mise à jour possible par rapport à la réalité qui ne peut plus changer. Seule l'interprétation de l'archéologue sur ces données va le pousser à **réviser** certaines des données. La distinction que nous faisons ici entre mise à jour et révision est très importante à comprendre. Ce processus de révision consiste alors à effectuer des **processus itératifs** de modifications volontaires dans le but d'aboutir à la **meilleure** interprétation possible du site étudié. Cela influencera directement la manière de gérer cet évènement dans le système analytique.

Actuellement, il est difficile de comprendre comment se produit la révision dans un outil d'analyse en ligne pendant son utilisation. En effet, typiquement, un utilisateur ne va jamais écrire dans un cube de données (du moins, selon les principes théoriques). Par contre, il existe des situations exceptionnelles (comme en archéologie) où la donnée mise dans le cube doit être "modifiée" à la volée, i.e. réinterprétée afin d'être améliorée. Il ne s'agit pas d'une erreur à corriger dans le système transactionnel, ni d'une mise à jour puisque la situation terrain n'a pas changé. Il s'agit d'une amélioration que l'utilisateur veut apporter à la donnée car en fait, et c'est le cas en archéologie, l'analyste ne peut valider efficacement le peuplement du cube de données qu'en effectuant son analyse ... avec le cube! Comme typiquement les données agrégées résultent de l'interprétation des données fines ET d'autres données agrégées, il est impensable ou même impossible de remonter aux données sources pour effectuer les changements désirés.

De plus, comme les interprétations de l'archéologue suivent des heuristiques qui ne peuvent pas toujours être formalisées de façon absolue, il faut laisser à l'analyste archéologue la possibilité de revoir les données agrégées produites par algorithmes lors du peuplement initial du cube ET de conserver les résultats de cette nouvelle interprétation (qui ne suit pas les règles de l'algorithme d'agrégation utilisé initialement pour peupler le cube). Nous allons étudier dans la section suivante les différents types de données : celles sur lesquelles il sera possible d'effectuer des mises à jour et celles sur lesquelles il sera possible d'effectuer des révisions. Ainsi, cette différenciation, nous a permis de trouver deux grandes catégories de données : les **données d'observation** et les **données d'interprétation** qui seront deux données distinctes dans un système d'analyse en ligne.

3.2-2. Mise à jour vs révision

La **mise à jour de données** va se porter sur des données que l'on va qualifier de **données d'observation.** Ce sont des données sources qui sont issues de systèmes d'acquisition des données ou de systèmes transactionnels. Un récepteur de données GPS, une image satellite, un outil relevant les ventes et les achats en temps réel d'une entreprise, une prise de sang pour une analyse médicale sont des exemples de ces systèmes d'acquisition de données d'observation. Ces données sont habituellement traitées avant d'être utilisables dans les

Rapport-Gratuit.com

systèmes transactionnels. Elles peuvent ensuite être ajoutées au système d'analyse et ne peuvent pas être modifiées au cours de l'analyse car elles représentent « la réalité observée ». Une correction de la donnée pourra cependant être envisagée si une mauvaise « saisie » de la réalité avait été effectuée.

La **révision de données** va se porter sur des données issues d'interprétation et qualifiées de **données d'interprétation.** L'expression "données d'interprétation" correspond d'une certaine manière à l'expression "données dérivées" dans le contexte traditionnel de base de données, mais que dans le contexte actuel d'archéologie, ce terme "dérivées" ne serait pas approprié car les données archéologiques ne sont pas le résultat de traitements algorithmiques, mais bel et bien le résultat d'interprétations par un expert, d'une expertise non programmable. C'est pour mieux refléter cette nuance que nous avons préféré l'expression "données d'interprétation" à l'expression "données dérivées". De plus, ces données d'interprétation étant saisies explicitement dans le système plutôt qu'obtenues par traitements, l'expression "données dérivées" aurait été fausse. Un budget prévisionnel, l'interprétation d'une analyse médicale, la recherche de nouveaux concepts,… sont autant d'exemples de processus permettant la création de données d'interprétation.

Ces données ne proviennent pas du système d'acquisition dans le sens traditionnel du terme, mais sont directement issues d'un besoin de l'utilisateur à revoir son analyse afin d'enrichir, de comprendre et/ou d'expliquer sa perception de la réalité, ceci sans mettre à jour les données d'observation. De plus, le degré de certitude des données d'interprétation varie en fonction de la façon dont celles-ci sont interprétées. Par exemple, si elles sont directement issues d'une combinaison de données observables, elles auront un degré de certitude élevé de même que si ces données sont issues de phénomènes connus et/ou compris. En prenant l'exemple de l'analyse médicale, une donnée d'interprétation sur l'état de santé du patient va être générée en fonction de plusieurs données d'observation (tension, température corporelle, présence ou non d'infection,…) et d'autres données issues d'interprétation (stress, hygiène alimentaire, activité physique,…).

La Figure 3-1 présente les données d'observation, d'interprétation ainsi que leur évolution respective (mise à jour pour la donnée d'observation et révision pour la donnée d'interprétation) dans le processus complet de compréhension de la réalité.

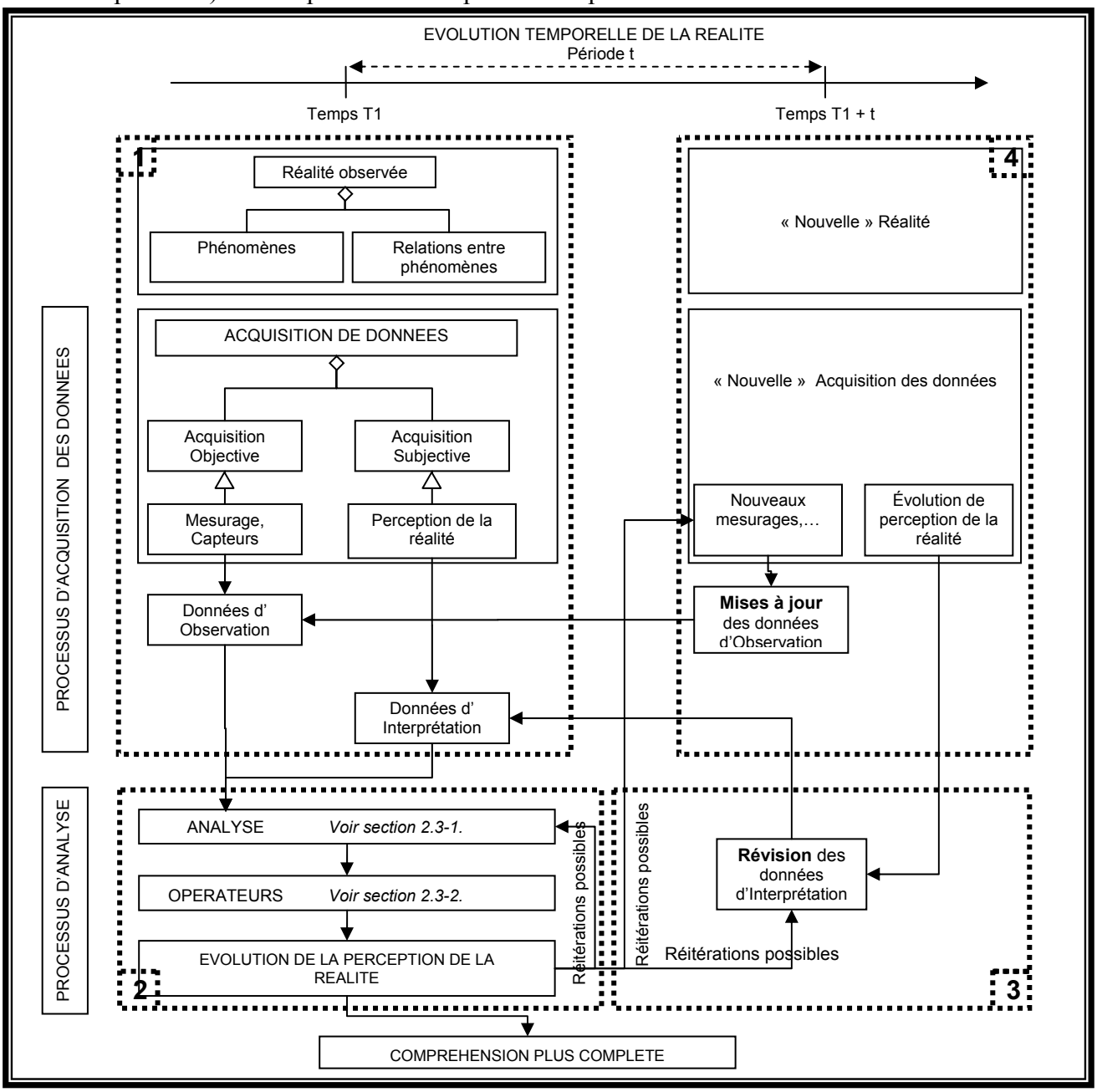


Figure 3-1 : Le processus de compréhension de la réalité au travers des données d'observation et d'interprétation

A une date donnée (encadre 1 de la Figure 3-1), la réalité observée, par l'intermédiaire de capteurs, va générer des données d'observation via ce que nous appellerons le **processus objectif d'acquisition**. L'utilisateur va avoir aussi une propre perception de la réalité et en

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

fonction de ces besoins, il va générer des données d'interprétation via ce que nous appellerons le **processus subjectif d'acquisition**. Le processus d'analyse tel que défini par Champoux (1991) peut alors se faire sur ces données (encadre 2 de la Figure 3-1). Suivant l'évolution de la perception de la réalité en fonction des résultats de l'analyse, l'utilisateur peut réviser ces données d'interprétation (encadré 3 de la Figure 3-1). De surcroit, suivant l'évolution de la réalité, ou si l'utilisateur souhaite réitérer l'ensemble de son processus objectif et/ou subjectif à un temps ultérieur, les données d'observation sont mises à jour et les données d'interprétation peuvent être révisées (encadré 4 de la Figure 3-1).

Quoique la distinction entre une mise à jour de données d'observation et une révision de données d'interprétation n'ait pas été observée dans la revue de littérature, elle s'est avérée utile pour la présente recherche. Nous en avons d'ailleurs fait un concept théorique central au processus d'analyse de l'archéologue et nous en sommes servi pour le développement de notre solution : supporter la révision des données d'interprétation ne se traite pas de la même façon que supporter la mise à jour des données d'observation. De plus, cette distinction permet de mieux situer les présents travaux avec ceux de Lambert (2005). Lambert en 2005 avait explicité le processus de mise à jour de données dans une structure multidimensionnelle avec l'outil SOLAP temps réel. Nous allons voir plus en détails comment le processus de révision se positionne dans un système d'analyse en ligne.

3.2-3. Les processus de révision et de mise à jour, et les systèmes d'analyse en ligne

La Figure 3-2 présente les différentes opérations que l'on peut effectuer avec une approche multidimensionnelle : la mise à jour des données d'observation, la révision des données d'interprétation, l'analyse des données détaillées et des données agrégées d'après les concepts de (Bédard, 2005; Lambert, 2005). Cette figure montre les efforts que l'utilisateur et la machine devront produire pour ces opérations. Concernant l'analyse de données, il n'y a pas de différence entre un débutant, un expérimentaliste ou un analyste car le principe très intuitif d'un outil d'analyse en ligne permet à n'importe quel utilisateur de devenir opérationnel rapidement. Cette figure nous montre aussi que l'outil d'analyse en ligne reste stable (rapide et facile) pour des analyses sur des données détaillées et/ou agrégées (Bédard, 2005). Concernant la mise à jour des données d'observation, les travaux de Lambert ont permis de mettre en évidence la notion d'actualisation d'une structure multidimensionnelle

associée à la notion de temps réel (Lambert, 2005). Cependant, aucune référence ne permet de quantifier l'effort que devrait mettre en œuvre tant l'utilisateur que la machine sur le processus de révision des données d'interprétation ni même la révision de la structure multidimensionnelle.

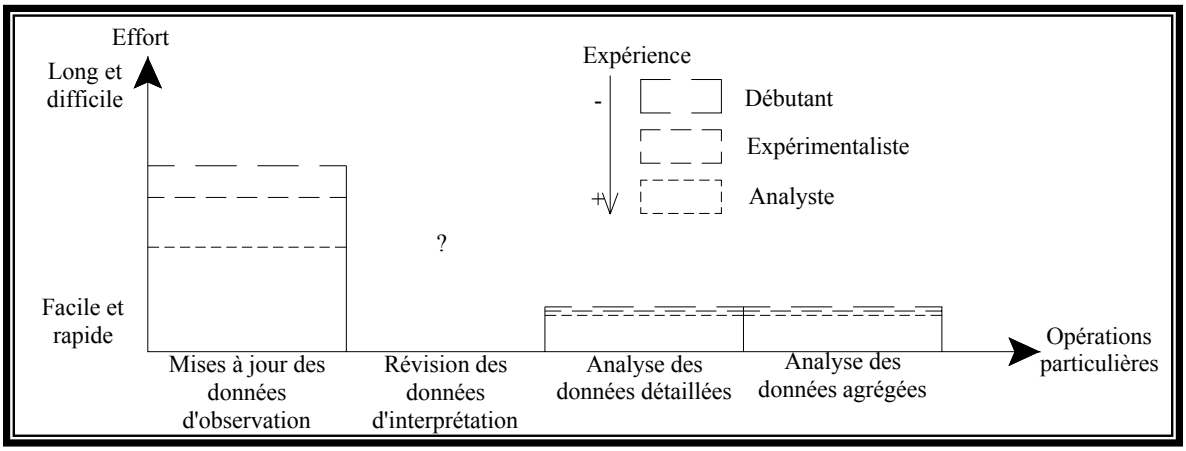


Figure 3-2 : Comparaison des efforts de plusieurs groupes d'utilisateurs par rapport à des opérations particulières d'une approche multidimensionnelle inspirée des travaux de Bédard 2005 et Lambert 2005

En partant des concepts de Lambert (2005) sur le SOLAP temps réel (cf. section 2.4-2.), la Figure 3-3 présente comment on peut conceptualiser, dans un système d'analyse en ligne, le lien qui unit les données d'interprétation avec les données d'observation. Ainsi, même si les données d'observation et les données d'interprétation ne sont pas issues de la même source « d'acquisition des données », il n'y a aucune différenciation dans l'entrepôt de données.

On remarque aussi que pour un outil d'analyse en ligne, il y a deux phases bien distinctes pour l'utilisateur :

- **la phase de pré-analyse :** il s'agit de l'ajout et suppression de données, traitement des données, rafraîchissement et reconstruction des agrégations. D'après Lambert (2005), cette phase peut être automatique (temps réel et quasi-réel) ou bien peut être fait sur demande (juste à temps).
- **la phase d'analyse :** l'utilisateur va interroger le cube reconstruit. En fonction des résultats et de ces besoins qui peuvent évoluer en cours d'analyse, il peut alors choisir de réviser ces données d'interprétation. L'outil permettant la révision de ces

données doit être, tout comme le système d'analyse, compatible avec le processus cognitif de l'utilisateur. Cet outil doit présenter les données, et les modifications possibles dans une forme proche de la façon dont l'analyste conçoit mentalement le phénomène à modifier dans le but d'une meilleure révision et par conséquent d'une meilleure interprétation.

Finalement, on remarque que grâce à l'approche de Lambert, les processus de révision et de mise à jour peuvent se porter à deux endroits de l'outil d'analyse en ligne (double flèche sur la Figure 3-3) : en amont du traitement (cela nécessite une reconstruction des agrégations) ou en aval du traitement (directement sur le cube reconstruit).

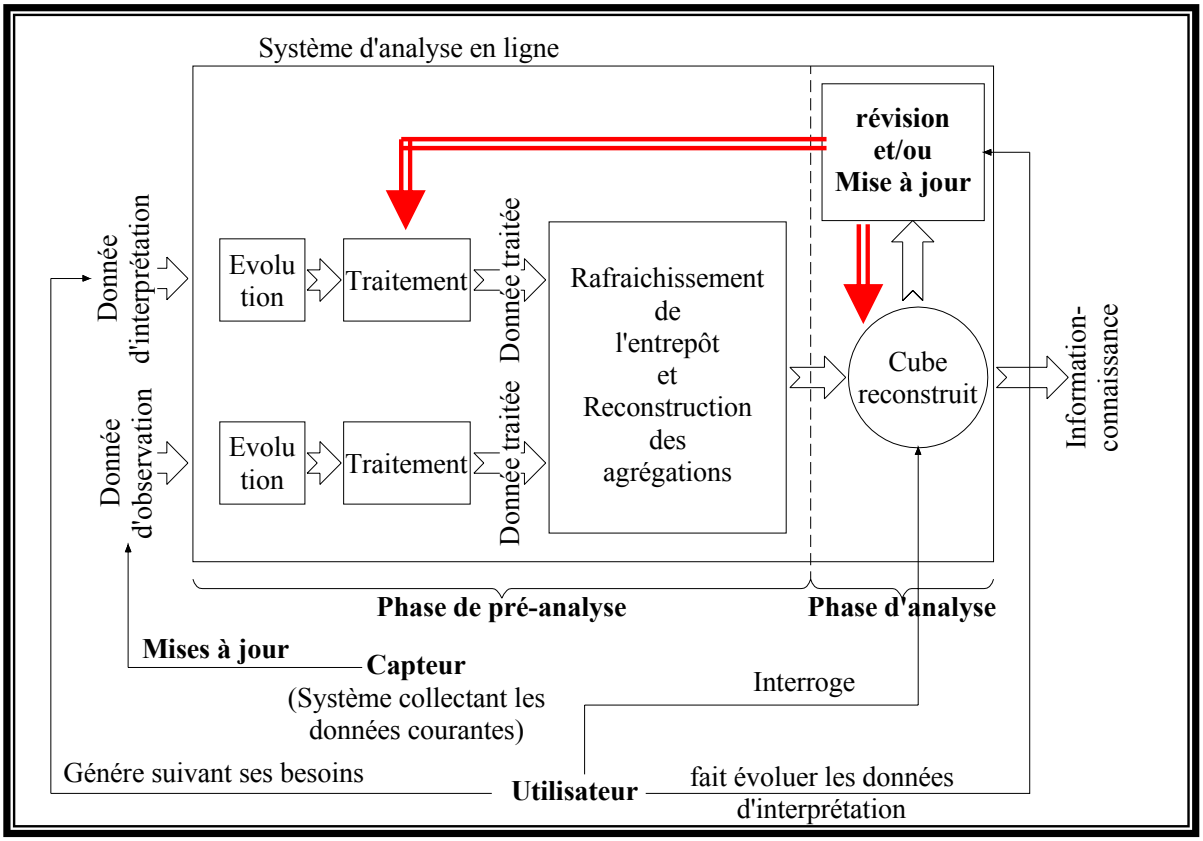


Figure 3-3 : La donnée d'interprétation dans le processus d'un système d'analyse en ligne inspirée de Lambert (2005)

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

3.2-4. Révision de structure vs révision de données

Une structure multidimensionnelle est constituée de membres organisés en dimensions (cf. Figure 3-4), de mesures et de faits (cf. section 2.4-2.). Un fait est principalement constitué de l'ensemble des mesures, des identifiants des membres des dimensions et d'un identifiant qui lui est propre. Ce fait est stocké dans une table que l'on appelle la « table des faits ».

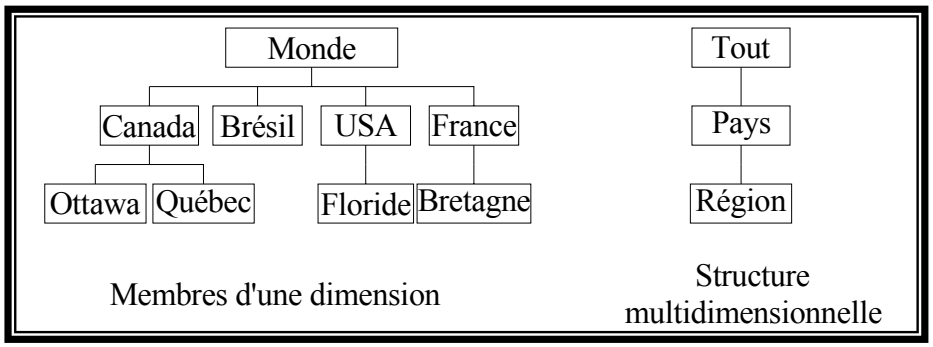


Figure 3-4 : Niveau conceptuel d'une dimension

D'un point de vue conceptuel, le processus de révision peut se produire à deux niveaux : la structure multidimensionnelle et les données elles-mêmes. La Figure 3-4 présente un exemple de structure multidimensionnelle à trois niveaux fixes: le grand-père « Tout », le père « Pays » et le fils « Région » et un exemple de membres associés à cette structure. La révision de la structure multidimensionnelle concerne l'ajout d'une dimension, la création d'une dimension alternative, l'ajout de niveaux à une dimension existante, l'ajout d'une mesure dans la table des faits, etc.

La Figure 3-5 présente deux exemples de révisions de structure multidimensionnelle :

1. La dimension est constituée de quatre niveaux hiérarchiques : « Ville » a pour père « Province » qui a pour père « Pays » qui a pour père « Tout ».
2. <u>Première révision de la structure multidimensionnelle :</u> Création d'un niveau alternatif dans la dimension spatiale de la structure du cube « Région » a pour père « Pays » et a pour fils « Ville ».
3. <u>Deuxième révision de la structure multidimensionnelle :</u> Ajout d'un niveau hiérarchique « Continent » a pour père « Tout » et a pour fils « Pays ».

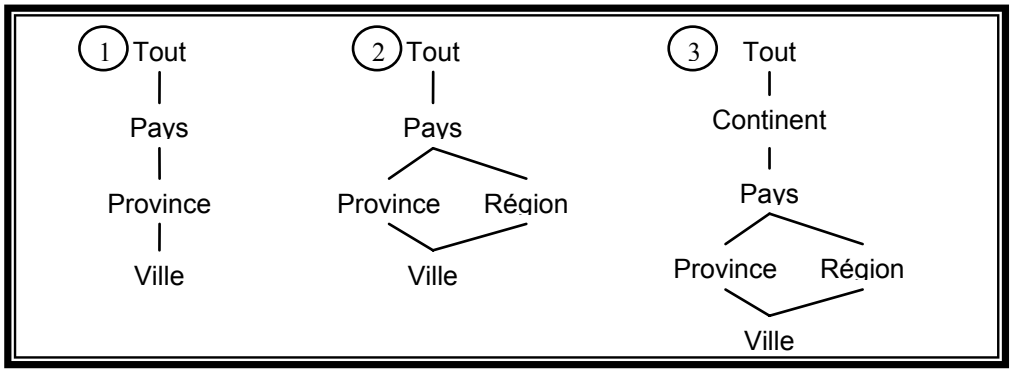
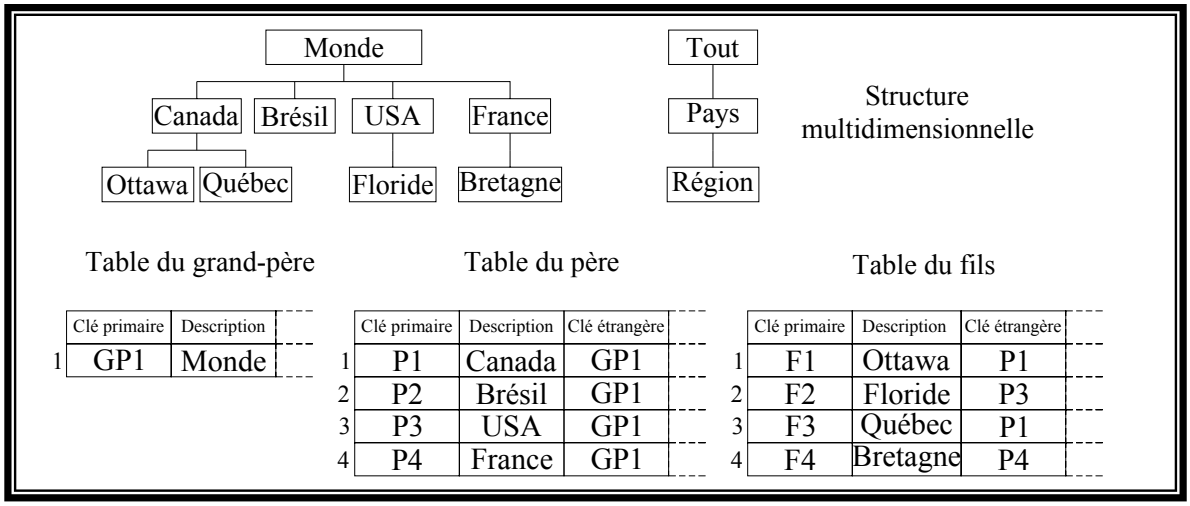


Figure 3-5 : Exemples de processus de révision de la structure multidimensionnelle inspiré de (Vaisman *et al.*, 2002)

Le processus de révision de données d'interprétation revient à modifier les membres d'une dimension lorsque la structure de la dimension existe. La Figure 3-6 montre l'exemple de trois niveaux <u>fixes</u> caractérisés par le grand-père, le père et le fils. Chaque niveau est caractérisé par une table dont chaque occurrence est identifiée par une clé primaire et fait référence, hormis le plus haut niveau hiérarchique, à la clé primaire de son ascendant via une clé étrangère.



	Clé primaire	Description
1	GP1	Monde

	Clé primaire	Description	Clé étrangère
1	P1	Canada	GP1
2	P2	Brésil	GP1
3	P3	USA	GP1
4	P4	France	GP1

	Clé primaire	Description	Clé étrangère
1	F1	Ottawa	P1
2	F2	Floride	P3
3	F3	Québec	P1
4	F4	Bretagne	P4

Figure 3-6 : La relation père-fils : une relation clé primaire – clé étrangère

La Figure 3-7 présente des exemples de révisions de données d'interprétation :

1. Structure multidimensionnelle fixe de la dimension ;
2. Exemple de membres de la dimension « Canada » a pour père « Tout », « Bretagne » a pour père « France » qui a pour père « Tout ».
3. <u>Première révision de données d'interprétation :</u> Création d'une donnée d'interprétation « Autres » dans le niveau « Pays ». « Autres » a pour père « Tout »

4. <u>Deuxième révision de données d'interprétation :</u> « Bretagne » qui a pour père « France » change de père pour « Autres »

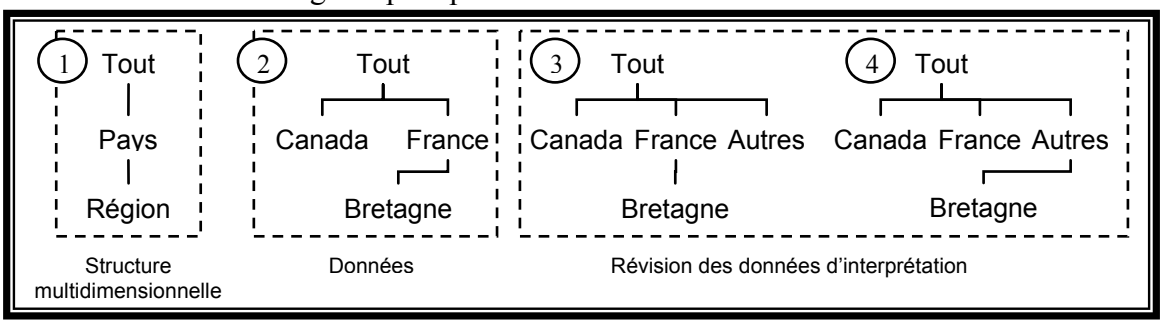


Figure 3-7 : Exemples de processus de révision des données d'interprétation

En résumé, pour notre projet, nous avons différencié deux types de données : les données d'observation qui sont extraites de la réalité et directement mesuré via le processus objectif d'acquisition et les données d'interprétation qui sont issues d'interprétation de données d'observation et/ou d'autres données d'interprétation et estimé via le processus subjectif d'acquisition. Ces données possèdent la même caractérisation telle que définie au chapitre 2. Cependant leur évolution est différente. La donnée d'observation va être actualisée à partir de mise à jour. Dans une structure multidimensionnelle, la mise à jour des données d'observation a été étudiée par Lambert (Lambert, 2005). La donnée d'interprétation, quant à elle, évolue en fonction du modèle que construit l'utilisateur du monde réel et des résultats de l'analyse. Son évolution est plus complexe que l'évolution des données d'observation car il s'inscrit dans des processus de révisions. **Le processus de révision consiste à effectuer des processus itératifs de modifications volontaires** dans le but d'aboutir à une meilleure interprétation possible de la réalité, ou à une meilleure gestion d'évènements (ou phénomènes) à venir. Il existe deux niveaux de révisions dans une structure multidimensionnelle. Le premier concerne la révision de la structure et le deuxième concerne la révision des données proprement dites. Nous avons décidé de n'étudier que la révision de données d'interprétation et de considérer la structure multidimensionnelle comme invariable. La Figure 3-3 présente comment devrait s'« exécuter » une révision dans une structure multidimensionnelle et plus particulièrement la révision des données d'interprétation. Cependant, pour mieux comprendre ce processus de révision et comment il s'exécute, nous verrons dans la section suivante, la justification du choix de structure multidimensionnelle utilisée.

3.3- La révision des données d'interprétation : sélection d'une structure multidimensionnelle

Dans cette section, nous allons comparer les structures multidimensionnelles possibles permettant de favoriser et comprendre le processus de révisions des données d'interprétation.

3.3-1. Le choix de l'architecture ROLAP

Lors de la construction d'une application OLAP, plusieurs architectures sont possibles (MOLAP, ROLAP ou HOLAP). Notre choix s'est porté sur l'architecture ROLAP (Relational OLAP). La principale raison est une capacité qu'a l'architecture ROLAP à répondre aux attentes de flexibilité et d'accessibilité des données relationnelles stockées dans la structure multidimensionnelle. En effet, l'approche ROLAP ne nécessite pas un serveur intermédiaire puisqu'il peut travailler directement avec la base de données relationnelle. Il peut donc offrir des modifications et analyses à la volée performantes et surtout permettre de comprendre comment le processus de révision se comporterait dans une structure multidimensionnelle sous une architecture ROLAP.

De son coté le MOLAP (Multidimensional OLAP) utilise des serveurs souvent plus rapides et performants à effectuer une reconstruction totale des agrégations. Par exemple, le logiciel Analysis Services de SQL Server 2005 avec la fonction « proactive catching » permettrait de balancer les données qui sont stockées en ROLAP ou MOLAP lors des mises à jour des cubes afin de permettre l'interrogation des données lors de la reconstruction des cubes. Toutefois, nous avons constaté qu'il n'existait aucun renseignement supplémentaire permettant d'analyser le principe de cette fonction. De surcroit, le format propriétaire de SQL Server 2005, et plus spécialement, la structure propriétaire de l'architecture MOLAP empêcherait de comprendre, sur l'aspect technologique, comment le processus de révision se comporterait dans une structure multidimensionnelle.

Quant au HOLAP (Hybrid OLAP), il offre aussi des performances intéressantes puisqu'il offre tous les avantages du MOLAP en améliorant sa capacité à gérer des gros volumes de données. Il permet aussi de combiner les données relationnelles et les données multidimensionnelles. Cependant, les implantations HOLAP sont généralement supportées

par un seul distributeur ce qui rend le déploiement et la maintenance extrêmement délicate [(Kahn, 2003) : p67]. De plus, ils peuvent être parfois restrictifs en termes de mobilité qui rend son exploitation pour notre projet plus que problématique. Enfin, une telle solution souffre des mêmes restrictions que la solution MOLAP pour les données agrégées stockées dans le cube multidimensionnel.

3.3-2. Le choix du schéma en étoile

L'architecture ROLAP choisie est la suivante (cf. Figure 3-8). Cette architecture se distingue par :

- une base de données qui doit supporter les données agrégées ou résumées au moyen d'une simulation de structure multidimensionnelle.
- un serveur ROLAP qui permet de gérer tant la structure multidimensionnelle que l'accès aux données de la part des usagers.
- un client SOLAP qui permet aux usagers de manipuler et d'explorer les données aux moyens de graphismes, de tableaux et/ou d'occurrences cartographiques

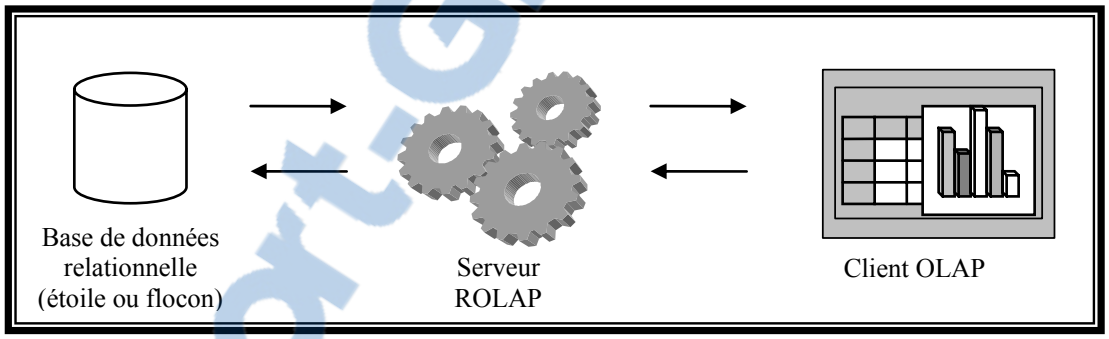


Figure 3-8 : Architecture ROLAP inspirée de (Rivest, 2000)

La simulation de structure multidimensionnelle est nécessaire pour une configuration ROLAP afin de mieux répondre aux besoins multidimensionnels. Il existe plusieurs types de simulation, mais notre choix s'est porté sur la simulation dite **schéma en étoile** pour plusieurs raisons.

La première raison concerne l'intérêt d'user du modèle en étoile. Ce modèle est le plus populaire puisqu'il offre une performance dans la rapidité d'analyse des données en réduisant au maximum le nombre de tables et le nombre de joints entre les tables. Il offre aussi, pour notre projet, un maximum de flexibilité des dimensions puisque chaque

Rapport-Gratuit.com

dimension est représentée par une seule table. Les modifications des dimensions deviennent plus simples. Sa simplicité (une table centrale appelée table des faits autour de laquelle gravitent les dimensions) peut lui permettre de faire des analyses complexes sans grande difficulté.

La deuxième raison concerne la lacune des autres modèles (modèle en flocon et modèle en constellation) à optimiser le processus de révision dans un outil d'analyse en ligne. Le **modèle en flocon** est une variante du schéma en étoile. Une des différences majeures est que les tables des dimensions du modèle en flocon sont normalisées afin d'éviter la redondance. Une dimension peut être alors représentée par une succession de tables. Cette succession de tables peut le rendre à première vue plus facile à mettre à jour et sauve de l'espace de stockage [(Han *et al.*, 2000) : p50]. Cependant, cela ne concernerait que la mise à jour des données d'observation (donc modification de données de façon occasionnelle) et non le processus de révision des données d'interprétation (donc modification de données de façon courante). Si cela n'était pas le cas, cela influencerait grandement la performance et demanderait un effort supplémentaire parce qu'il y a plus de tables à mettre à jour et à réviser. En effet, si on choisit le schéma en flocon, cela signifie que la structure de la dimension est clairement définie, ce qui n'est pas le cas avec la plupart des dimensions archéologiques. Finalement, le **modèle en constellation** est aussi une variante du schéma en étoile. Même si chacune des dimensions est représentée avec une table unique, le modèle contient plusieurs tables des faits partageant les dimensions. Augmenter le nombre de table des faits, augmente aussi les efforts de mises à jour et de révision des tables.

La troisième raison est directement liée au caractère novateur de ce type de projet appliqué à l'archéologie. En effet, le dernier modèle, **modèle mixte**, est un croisement entre le schéma en étoile et le schéma en flocon. Utiliser un tel modèle revient à connaître quelles sont les dimensions faisant appels aux données d'observation et quels sont les dimensions faisant appels aux données d'interprétation. Il peut être envisagé, de se tourner vers ce genre de structure lorsque les besoins des utilisateurs et certaines structures de dimensions sont clairement définis ce qui n'était pas notre cas en début de projet.

3.3-3. Le choix de la structure Parent-enfant d'une dimension

La modélisation en étoile est donc composée d'une table centrale dénormalisée autour de laquelle gravitent les dimensions représentées par des tables uniques (cf. Figure 3-9). La table des faits contient les données quantitatives à analyser. Il s'agit des valeurs des mesures : les faits. Cette table contient aussi les clés étrangères des dimensions.

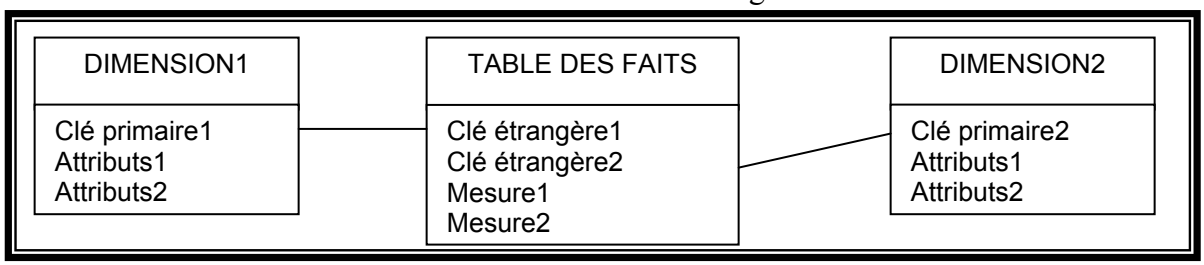


Figure 3-9 : Exemple de modélisation en étoile (ici avec 2 dimensions seulement)

Notre dernier choix va concerner la structure de la dimension. Comme le montre la Figure 3-9, la table de dimension doit impérativement contenir des clés primaires ainsi que les attributs définissant les membres des niveaux hiérarchiques de cette dimension. Nous avons répertorié deux grandes façons de structurer une table de dimension (cf. Figure 3-10). La première façon n'est pas adaptable à notre projet puisqu'il s'agit d'une dimension à hiérarchie balancée, c'est-à-dire avec des niveaux bien déterminés et équilibrés. Cette hiérarchie présuppose une non-évolution de la dimension qui maintient l' « arbre » de la dimension équilibrée. Cette hiérarchie est inadaptée à notre projet puisque la liberté d'évolution de la dimension n'est alors plus garantie.

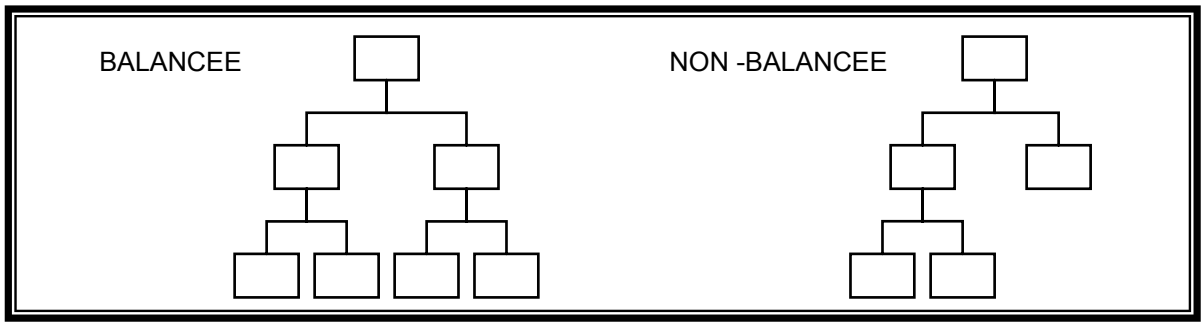
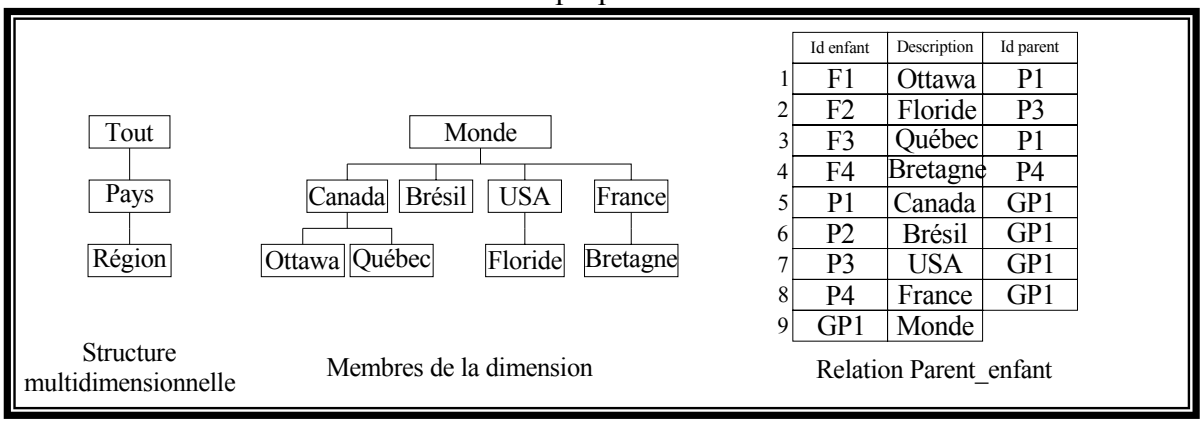


Figure 3-10 : Arbre balancée vs Arbre non-balancée

La deuxième structure, structure non balancée, n'a pas la contrainte de demeurer balancée ce qui donne plus de liberté pour réviser les données d'interprétation. Elle est donc naturellement choisie en utilisant la relation caractéristique de cette structure : la **relation**

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

parent-enfant. Cette structure offre toutes les possibilités d'évolution des données (d'observation et d'interprétation) puisqu'elle ne donne aucune contrainte à la dimension. L'implantation de cette relation dans une table de dimension nécessite au moins trois champs : un champ de clé primaire correspondant à l'identifiant de l'enfant, un champ de description de l'enfant et un champ correspondant à l'identifiant du parent. Le parent se retrouvant lui aussi enfant, sa description n'est alors pas nécessaire car elle se retrouve dans le membre où il est enfant. Cela peut ainsi éviter une redondance non nécessaire. (cf. Figure 3-11). Finalement, la Figure 3-6 explicitait la relation « père –fils » entre les membres d'une dimension. La structure « parent-enfant », est la seule structure où la modification de parenté, et éventuellement l'ajout d'un niveau dans la structure multidimensionnelle, revient à la simple modification d'une donnée (modification du champ présentant l'identifiant du parent, « Id parent » dans la Figure 3-11) et non pas à l'ajout de nouvelles tables ou de nouvelles colonnes comme le proposerait les autres structures.



	Id enfant	Description	Id parent
1	F1	Ottawa	P1
2	F2	Floride	P3
3	F3	Québec	P1
4	F4	Bretagne	P4
5	P1	Canada	GP1
6	P2	Brésil	GP1
7	P3	USA	GP1
8	P4	France	GP1
9	GP1	Monde	

Figure 3-11 : Exemple de la relation « parent-enfant »

3.3-4. Construction et reconstruction des agrégations

Pour permettre de comprendre le processus de révision, nous estimons que l'architecture OLAP doit être une architecture ROLAP, que la simulation multidimensionnelle pour les bases de données relationnelles doit être le schéma en étoile et que la structure d'une dimension doit être basée sur la relation « parent-enfant » caractéristique des dimensions à hiérarchie non-balancée. L'approche ROLAP ne nécessitant pas un serveur intermédiaire, lui permet de travailler directement avec la base de données relationnelle. Grâce à cela, nous pouvons mieux comprendre comment s'effectue la construction et la reconstruction des agrégations d'une structure multidimensionnelle.

La **construction des agrégations** est présentée en Figure 3-12.

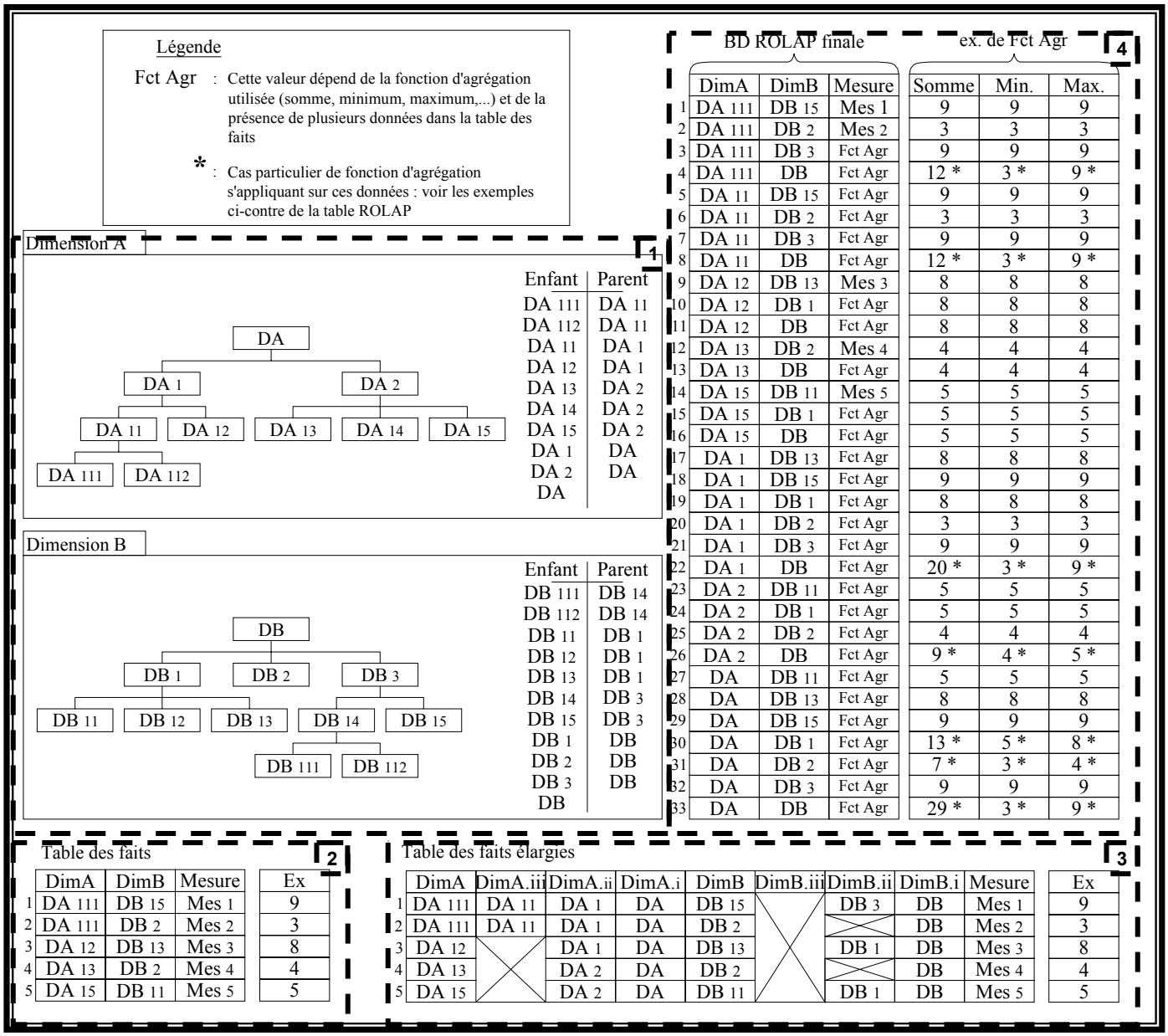


Figure 3-12 : Exemple de construction d'agrégations d'une configuration ROLAP

L'encadré 1 de cette figure présente deux exemples de dimensions structurées avec la relation parent-enfant. L'encadré 2 présente la table des faits composée de cinq enregistrements. Comme expliqué précédemment, chaque enregistrement possède les clés étrangères des dimensions A et B ainsi qu'une mesure. L'encadré 3 présente la première étape d'une construction d'agrégations : la constitution d'une table des faits élargie. Sur chaque enregistrement, on ajoute les clés étrangères de chaque parent pour chaque dimension. Par exemple, pour le premier enregistrement, on ajoute trois champs

supplémentaires pour la dimension A (DimAiii, DimAii, DimAi). La donnée DA_{111} a pour père DA_{11} (DimAiii=DA_{11}) qui a pour père DA_1 (DimAii=DA_1) qui a pour père DA (DimAi=DA). On fait de même avec les autres dimensions du même enregistrement et on remplit toute la table de cette façon. La deuxième et dernière étape de la construction consiste à effectuer l'ensemble des combinaisons entre les données détaillées et les données agrégées grâce à la table élargie précédemment créée (encadré 4 de la Figure 3-12). La nouvelle table duplique considérablement les enregistrements de la table des faits initiale, mais elle permet d'avoir l'ensemble des combinaisons possibles. La table des agrégations, ou table ROLAP, ainsi créées possède la même structure que la table des faits. De surcroît, les mesures sont calculées à l'aide de la fonction d'agrégation. Suivant le choix de la fonction d'agrégation des mesures, la valeur des mesures peut varier. Ainsi, en prenant l'exemple de la ligne 4 de la table ROLAP :

« Mesure DA111 DB = Fonction_agregation(DA111 DB=9 ; DA111 DB=3)
= 12 (si Fonction_agregation=somme)
= 03 (si Fonction_agregation=minimum)
= 09 (si Fonction_agregation=maximum) »

Ainsi, nous constatons ce qui fait effectivement la force et la faiblesse de la structure multidimensionnelle. Plus il y a de dimensions et de niveaux hiérarchiques plus l'analyse se fera rapidement et simplement comparativement aux structures transactionnelles qui seront de plus en plus difficiles et longues. Cependant, plus il y aura de dimensions et de niveaux hiérarchiques, plus la maintenance de la structure (la reconstruction des agrégations et le recalcule des mesures) sera longue et difficile. On remarque aussi que cette maintenance se ferait sur deux niveaux. Le premier niveau serait la dimension. Une évolution de la dimension aurait des répercutions directes sur la reconstruction des agrégations et le calcul des mesures. Le deuxième niveau serait la table des faits. Cette table présente les « faits », c'est à dire le croisement des clés étrangères des dimensions entre elles avec les mesures associées. Une modification de ces clés étrangères engendrait irrémédiablement des agrégations différentes.

La **reconstruction partielle des agrégations** tire son intérêt à ne pas reconstruire toute la table des agrégations mais seulement la partie qui est affectée par le changement. Notre

exemple de la Figure 3-12 ne représente qu'un échantillon de données représentatives de la structure type d'une dimension et de table des faits. A ce niveau, il n'y a aucune différenciation entre une reconstruction totale et une reconstruction partielle. Cependant dans la réalité, le nombre de niveaux d'une dimension ainsi que le nombre total de dimensions seront plus conséquents et vont accroître considérablement et de manière exponentielle la table ROLAP. Au final, celle-ci sera de plusieurs millions d'enregistrements plutôt que d'une petite trentaine. Nous supposons alors que des reconstructions partielles seront plus fréquentes que des reconstructions totales.

Toutes les reconstructions partielles varient en fonction du processus de révision de l'utilisateur pendant sa phase d'analyse. Nous allons voir dans la section suivante, les différentes révisions de données d'interprétations possibles dans un outil d'analyse en ligne et ses répercutions sur la reconstruction des agrégations. L'ensemble des exemples de reconstruction partielle sera basé sur la construction présentée en Figure 3-12.

3.4- La révision des données d'interprétation : les règles d'application.

La table des faits comprend l'ensemble des clés étrangères (faisant référence aux clés primaires présentes dans les dimensions) ainsi que les mesures, variables qui font l'objet de l'analyse (cf.3.3-3.) Cette table matérialise l'association entre les mesures et les combinaisons de un ou plusieurs membres d'une ou plusieurs dimensions. Nous allons voir à présent l'ensemble des règles de révision qui portera tant sur les dimensions que sur la table des faits.

3.4-1. La révision des dimensions

Renommer le membre d'une dimension

Cette évolution est la plus facile à mettre en œuvre. En effet, la table des faits et les dimensions communiquent au moyen de clé primaire et de clés étrangères et non en fonction des attributs des membres de la dimension. Ainsi, tout changement d'attribut n'influencera en rien la correspondance clé primaire – clé étrangère et AUCUNE agrégation même partielle n'est à envisager.

Ajouter un membre de l'arbre

L'ajout d'un membre doit répondre à l'un des deux besoins suivants. Soit l'utilisateur a besoin d'une donnée détaillée supplémentaire même si celle-ci n'a pas d'influence sur la reconstruction des agrégations. Aucune clé étrangère de la table des faits n'y fait référence. Cependant, elle peut être utile pour justifier que même cette donnée n'influence pas l'analyse. Soit l'utilisateur a besoin de créer une donnée supplémentaire pour effectuer des modifications ultérieures relatives à cette donnée (en en modifiant une ou plusieurs clés étrangères vers la clé primaire de la nouvelle donnée (cf. Modifier les clés étrangères) ou en modifiant la structure de la dimension en effectuant des changements de père (cf. Changer de père). Ainsi, aucune reconstruction totale ou partielle n'est à envisager avec l'ajout d'un membre. Cependant, une grande majorité des ajouts nécessite une révision supplémentaire qui nécessite quant à elles des reconstructions totales ou partielles des agrégations

Changer de père :

Nous avons répertorié trois cas. **1er cas : le changement de père n'a aucune répercussion sur les agrégations construites** (cf. Figure 3-13). Les parties qui sont concernées par le changement de père n'affectent aucune des clés étrangères de la table ROLAP. Aucune reconstruction n'est alors envisagée. Cette modification peut cependant être intéressante surtout si elle précède un ajout, ou une modification de la table des faits faisant intervenir la donnée qui vient de subir le changement de père.

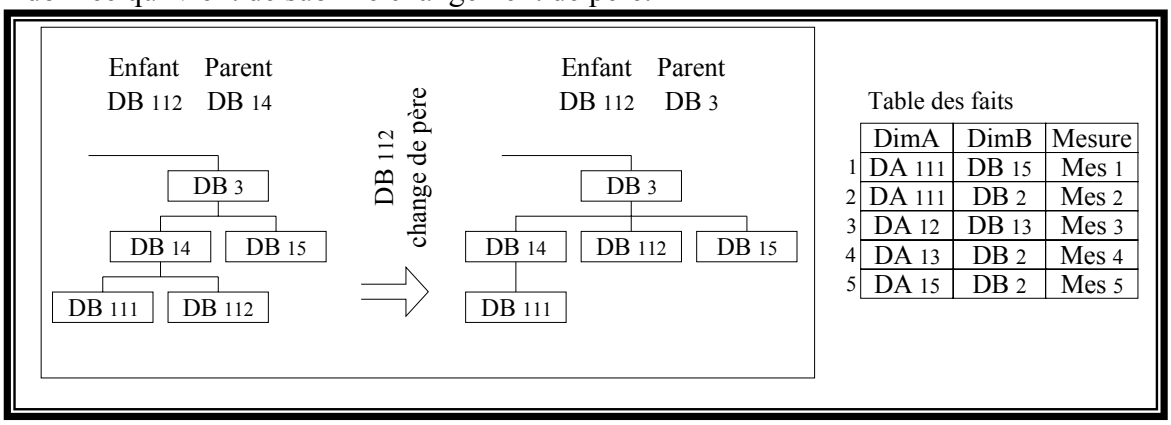


Figure 3-13 : Exemple de changement de père avec aucune répercussion

2ème cas : les parties concernées par le changement de père affectent certaines clés étrangères. La reconstruction partielle devient alors très intéressante puisqu'elle va

concerner une partie des enregistrements et ainsi ne reconstruire qu'une partie des agrégations. La Figure 3-13 présente des exemples de changement de père avec répercussion sur la table des agrégations. Ce changement de père peut se faire sur deux types de données.

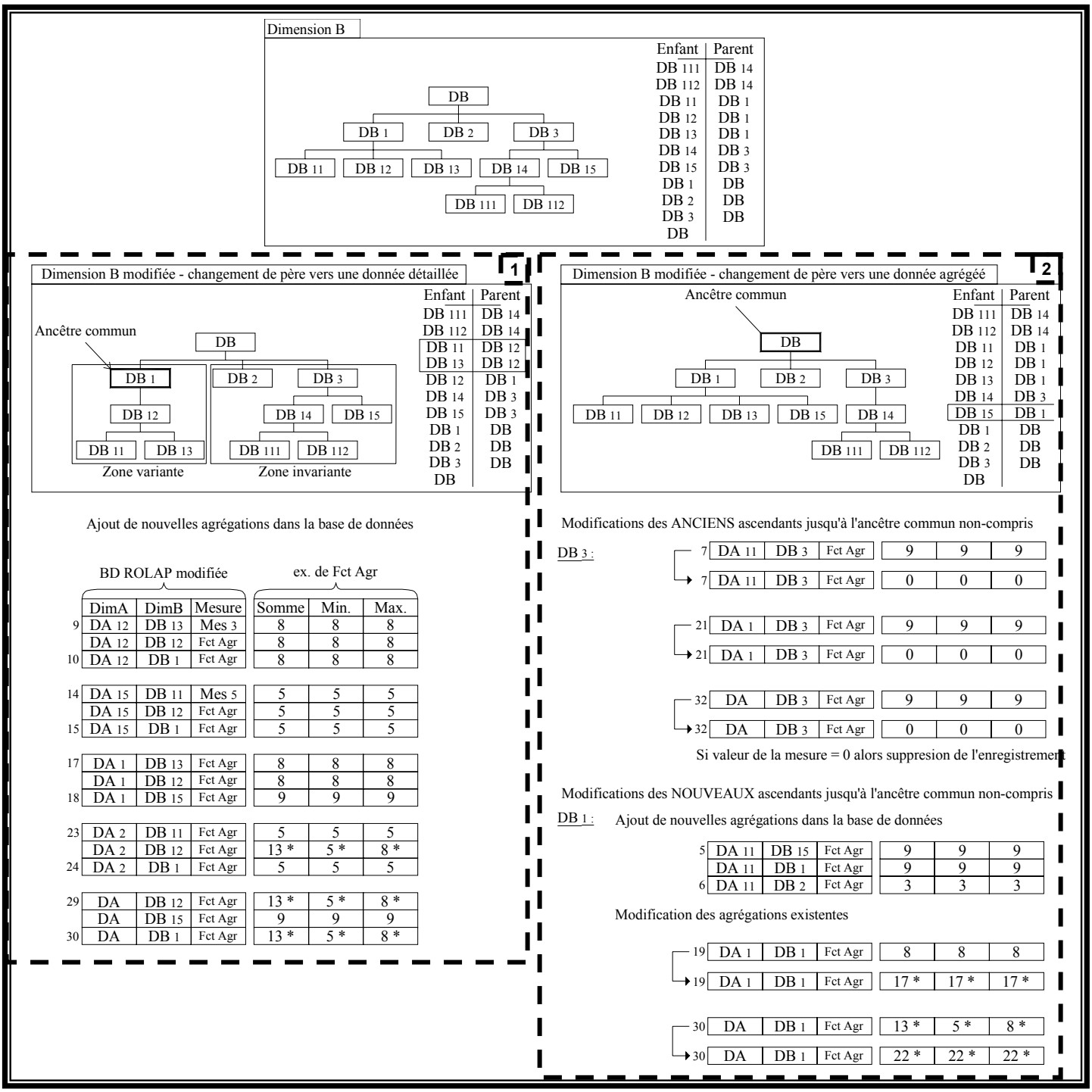


Figure 3-14 : Exemple de changement de père avec des répercussions

Tout d'abord, comme le montre l'encadré 1 de la Figure 3-13, le changement peut se faire sur une **donnée détaillée**. En premier lieu, il faut vérifier si le futur père, comme il s'agit d'une donnée détaillée, a une correspondance en clé étrangère dans la table des faits. Si une correspondance existe, il faut alors, soit interdire l'accès à la parenté à cette donnée, soit permettre à l'utilisateur de contourner ce problème en manipulant un peu plus ces dimensions : 1- en renommant la donnée qui aurait du être le futur père (cf. Renommer le membre d'une dimension), 2- en ajoutant un membre à la dimension qui porte le premier nom de la donnée qui aurait du être le futur père (cf. Ajouter un membre de l'arbre), 3- en affectant le changement de père sur la nouvelle donnée ajoutée. Une fois le contrôle validé, la reconstruction partielle peut s'effectuer. Dans l'exemple de l'encadré 1 de la Figure 3-13, la reconstruction partielle va consister à ajouter des enregistrements dans la table des agrégations.

Finalement, le changement peut se faire sur une **donnée agrégée.** Par définition, aucune clé étrangère n'y fait référence dans la table des faits. La reconstruction partielle ne nécessite aucun « pré-contrôle » et peut-être envisagée. Comme le montre l'encadré 2 de la Figure 3-13, la reconstruction partielle doit d'abord modifier les anciennes agrégations de la donnée (anciens ascendants de la donnée), puis ajouter et/ou modifier les agrégations concernées (nouveaux ascendants de la donnée).

Ces deux exemples nous montrent aussi que le niveau hiérarchique de l'ancêtre commun du nouveau et de l'ancien père tient une place importante dans l'effort de reconstruction des agrégations. En effet, plus l'ancêtre commun des deux pères est éloigné du sommet de l'arbre hiérarchique et moins il y aura d'effort dans la reconstruction partielle d'agrégation. Au contraire, si l'ancêtre commun aux deux pères se rapproche du sommet de l'arbre hiérarchique, plus la reconstruction partielle sera importante et longue.

3ème cas : les parties qui sont concernées affectent toutes les clés étrangères. Ce troisième cas sera fréquent si la dimension n'est pas très hiérarchisée (2 niveaux par exemple), si elle n'a pas beaucoup de membres ET si tous les enregistrements de la table des faits font référence à cette dimension. Une reconstruction partielle n'est plus envisagée

car toutes les agrégations se trouveraient modifiées. Il faudrait alors reconstruire l'ensemble des agrégations.

Supprimer un membre de l'arbre

La suppression d'une donnée doit être extrêmement contrôlée car supprimer un membre d'une dimension revient à supprimer une clé primaire. Il faut alors faire attention à l'intégrité entre les clés primaires et étrangères. Ainsi il y a alors plusieurs possibilités. Si le membre supprimé est un **membre agrégé**, il n'y a alors aucun problème de contrainte d'intégrité des clés puisque le lien entre la table des faits et les dimensions se fait avec les données détaillées. Les enfants héritent directement du grand père et la dimension est révisée. Si les enfants de la donnée agrégée supprimée n'ont aucun lien avec la table des faits, alors aucune reconstruction partielle n'est effectuée. Sinon, si au moins un des enfants a un lien avec la table, alors les agrégations ne sont pas pour autant reconstruites mais certains enregistrements sont supprimés

Si le membre supprimé est un **membre détaillé**. Il se peut qu'il y ait un problème de contrainte d'intégrité. En effet, s'il n'y a AUCUN lien entre la clé primaire à supprimer et AU MOINS une clé étrangère, le membre peut être supprimé en toute sécurité. Cependant, s'il y a UN lien entre la clé primaire et AU MOINS une clé étrangère, deux possibilités s'offrent alors à l'utilisateur. Soit la suppression de la donnée est interdite. Soit on permet à l'utilisateur de contourner ce problème en lui laissant la possibilité de manipuler TOUTES les clés étrangères qui font référence à la clé primaire à supprimer (cf. révision des clés étrangères) afin de rendre la donnée à supprimer libre de tout lien avec la table des faits.

Ajouter un niveau à l'arbre (changement de structure)

Nous avons décidé de garder la structure multidimensionnelle non révisable. Cependant nous avons remarqué que la structure « parent-enfant » était la seule structure où l'ajout d'un niveau revient, tout simplement, à l'ajout d'un enregistrement (et donc de données) dans la table de la dimension déjà existante. Cela ne revient pas à l'ajout de nouvelles tables ou de nouvelles colonnes comme le laisserait supposer les relations « père fils » (cf. Figure 3-15).

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

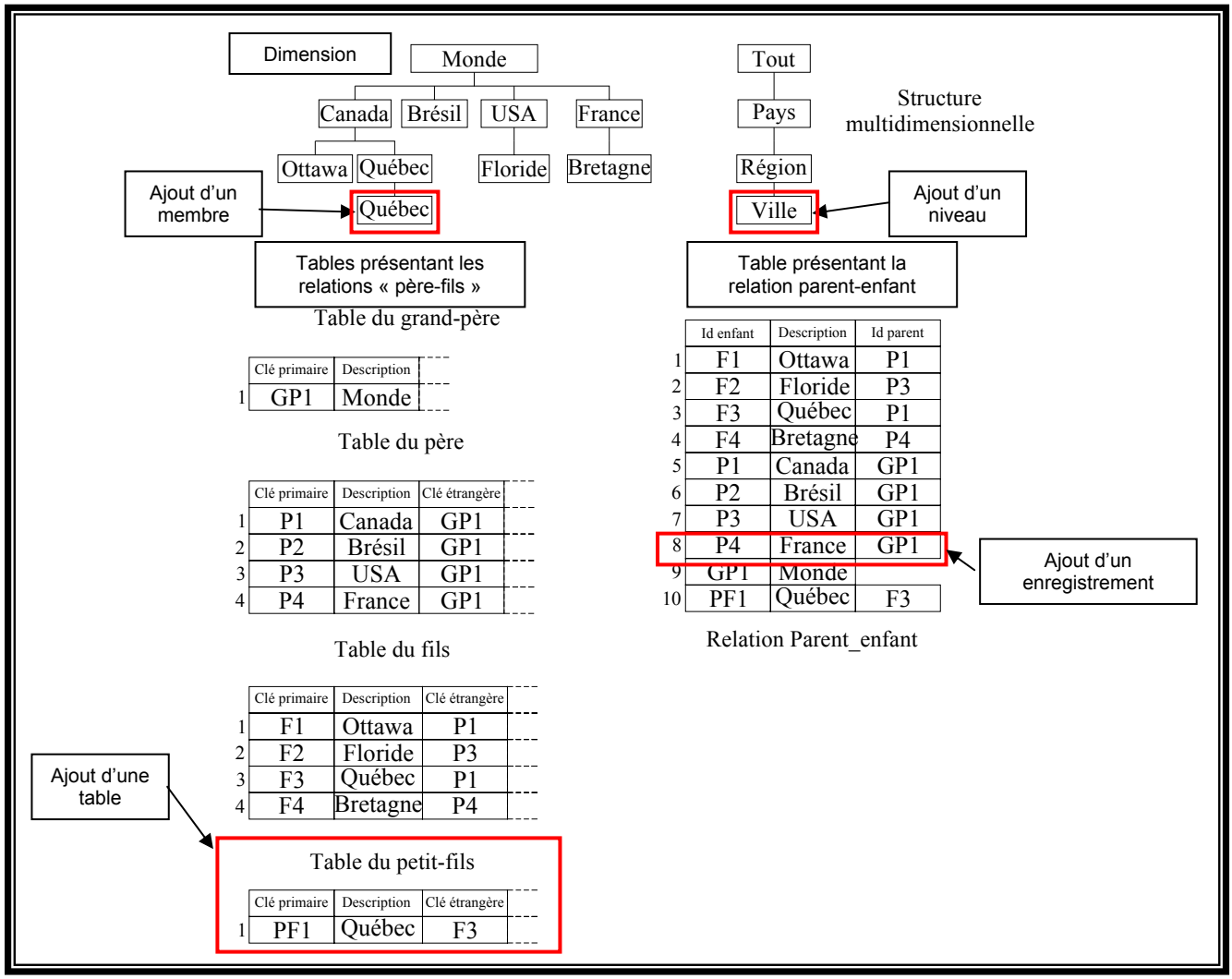


Figure 3-15 : L'ajout d'un niveau : l'avantage de la relation « parent-enfant »

3.4-2. La révision de la table des faits

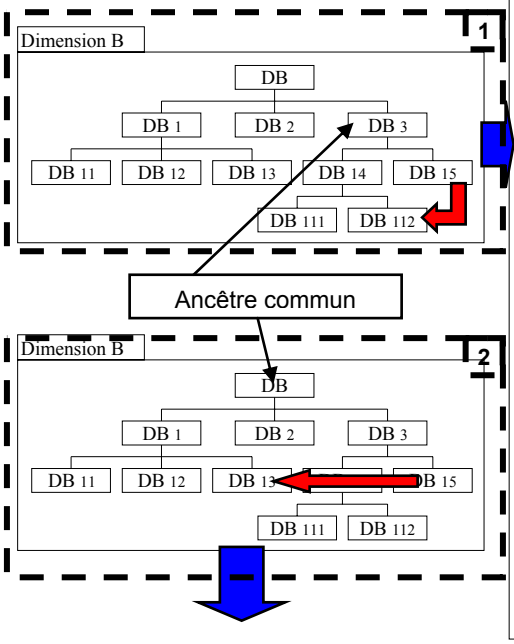
Modifier les clés étrangères

L'utilisateur, choisissant de réviser les clés étrangères, va obliger une reconstruction partielle des agrégations. Une évolution des clés étrangères est obligatoire à partir du moment où, quelques soient les raisons, l'attribution n'est pas celle souhaitée. La révision d'une clé étrangère doit se faire sur une autre clé primaire d'un membre détaillé. Comme pour un changement de structure (cf. Changer de père), l'effort de reconstruction des agrégations sera fonction de la position de l'ancêtre commun des clés primaires qui faisaient référence à l'ancienne et la nouvelle clé étrangère. La Figure 3-16 présente deux exemples de révision de clé étrangère (encadrés 1 et 2 sur la Figure 3-16).

Table des faits

	DimA	DimB	Mesure	Ex
1	DA 111	DB 15	Mes 1	9
2	DA 111	DB 2	Mes 2	3
3	DA 12	DB 13	Mes 3	8
4	DA 13	DB 2	Mes 4	4
5	DA 15	DB 11	Mes 5	5

Révision de la donnée DB15



	BD ROLAP gardé			modifié			ajouté		
1				DA 111	DB 112	Mes 1			
							DA 111	DB 14	Fct Agr
2	DA 111	DB 2	Mes 2						
3	DA 111	DB 3	Fct Agr						
4	DA 111	DB	Fct Agr						
5				DA 11	DB 112	Fct Agr			
							DA 11	DB 14	Fct Agr
6	DA 11	DB 2	Fct Agr						
7	DA 11	DB 3	Fct Agr						
8	DA 11	DB	Fct Agr						
9	DA 12	DB 13	Mes 3						
10	DA 12	DB 1	Fct Agr						
11	DA 12	DB	Fct Agr						
12	DA 13	DB 2	Mes 4						
13	DA 13	DB	Fct Agr						
14	DA 15	DB 11	Mes 5						
15	DA 15	DB 1	Fct Agr						
16	DA 15	DB	Fct Agr						
17	DA 1	DB 13	Fct Agr						
18				DA 1	DB 112	Fct Agr			
							DA 1	DB 14	Fct Agr
19	DA 1	DB 1	Fct Agr						
20	DA 1	DB 2	Fct Agr						
21	DA 1	DB 3	Fct Agr						
22	DA 1	DB	Fct Agr						
23	DA 2	DB 11	Fct Agr						
24	DA 2	DB 1	Fct Agr						
25	DA 2	DB 2	Fct Agr						
26	DA 2	DB	Fct Agr						
27	DA	DB 11	Fct Agr						
28	DA	DB 13	Fct Agr						
29				DA	DB 15	Fct Agr			
							DA	DB 14	Fct Agr
30	DA	DB 1	Fct Agr						
31	DA	DB 2	Fct Agr						
32	DA	DB 3	Fct Agr						
33	DA	DB	Fct Agr						

	BD ROLAP gardé			modifié			ajouté			supprimé			
1				DA 111	DB 13	Mes 1							
							DA 111	DB 1	Fct Agr				
2	DA 111	DB 2	Mes 2										
3										DA 111	DB 3	Fct Agr	n'existe plus
4	DA 111	DB	Fct Agr										
5				DA 11	DB 13	Fct Agr							
							DA 11	DB 1	Fct Agr				
6	DA 11	DB 2	Fct Agr										
7										DA 11	DB 3	Fct Agr	n'existe plus
8	DA 11	DB	Fct Agr										
9	DA 12	DB 13	Mes 3										
10	DA 12	DB 1	Fct Agr										
11	DA 12	DB	Fct Agr										
12	DA 13	DB 2	Mes 4										
13	DA 13	DB	Fct Agr										
14	DA 15	DB 11	Mes 5										
15	DA 15	DB 1	Fct Agr										
16	DA 15	DB	Fct Agr										
17				DA 1	DB 13	Fct Agr							
18										DA 1	DB 13	Fct Agr	redondance
19				DA 1	DB 1	Fct Agr							
20	DA 1	DB 2	Fct Agr										
21										DA 1	DB 3	Fct Agr	n'existe plus
22	DA 1	DB	Fct Agr										
23	DA 2	DB 11	Fct Agr										
24	DA 2	DB 1	Fct Agr										
25	DA 2	DB 2	Fct Agr										
26	DA 2	DB	Fct Agr										
27	DA	DB 11	Fct Agr										

Le reste n'a pas pu être mis par manque de place

Figure 3-16 : Exemple de modifications de clés étrangères

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

Modifier les mesures

La révision d'une mesure ne modifie en rien la construction des agrégations car la mesure n'influence pas toute la construction des agrégations En effet, tous les croisements effectués restent et resteront car les croisements dépendent de l'agrégation des données et donc des dimensions. Par contre, des modifications à l'endroit où la mesure est stockée seront faites. Tous les ancêtres de la donnée à laquelle cette mesure fait référence subiront des modifications qui dépendront de la fonction d'agrégation. La modification sera obligatoire si la fonction est une somme mais pas forcément si la fonction est un maximum ou un minimum (cf. Figure 3-17 : la valeur Mes1 n'est plus 9 mais 2).

	BD ROLAP finale			ex. de Fct Agr			ex. de Fct Agr		
	DimA	DimB	Mesure	Somme	Min.	Max.	Somme	Min.	Max.
1	DA 111	DB 15	Mes 1	9	9	9	2	2	2
2	DA 111	DB 2	Mes 2	3	3	3			
3	DA 111	DB 3	Fct Agr	9	9	9	2	2	2
4	DA 111	DB	Fct Agr	12 *	3 *	9 *	5 *	2 *	3 *
5	DA 11	DB 15	Fct Agr	9	9	9	2	2	2
6	DA 11	DB 2	Fct Agr	3	3	3			
7	DA 11	DB 3	Fct Agr	9	9	9	2	2	2
8	DA 11	DB	Fct Agr	12 *	3 *	9 *	5 *	2 *	3 *
9	DA 12	DB 13	Mes 3	8	8	8			
10	DA 12	DB 1	Fct Agr	8	8	8			
11	DA 12	DB	Fct Agr	8	8	8			
12	DA 13	DB 2	Mes 4	4	4	4			
13	DA 13	DB	Fct Agr	4	4	4			
14	DA 15	DB 11	Mes 5	5	5	5			
15	DA 15	DB 1	Fct Agr	5	5	5			
16	DA 15	DB	Fct Agr	5	5	5			
17	DA 1	DB 13	Fct Agr	8	8	8			
18	DA 1	DB 15	Fct Agr	9	9	9			
19	DA 1	DB 1	Fct Agr	8	8	8			
20	DA 1	DB 2	Fct Agr	3	3	3			
21	DA 1	DB 3	Fct Agr	9	9	9	2	2	2
22	DA 1	DB	Fct Agr	20 *	3 *	9 *	13 *	2 *	8 *
23	DA 2	DB 11	Fct Agr	5	5	5			
24	DA 2	DB 1	Fct Agr	5	5	5			
25	DA 2	DB 2	Fct Agr	4	4	4			
26	DA 2	DB	Fct Agr	9 *	4 *	5 *			
27	DA	DB 11	Fct Agr	5	5	5			
28	DA	DB 13	Fct Agr	8	8	8			
29	DA	DB 15	Fct Agr	9	9	9	2	2	2
30	DA	DB 1	Fct Agr	13 *	5 *	8 *			
31	DA	DB 2	Fct Agr	7 *	3 *	4 *			
32	DA	DB 3	Fct Agr	9	9	9	2	2	2
33	DA	DB	Fct Agr	29 *	3 *	9 *	22 *	2 *	3 *

Figure 3-17 : Exemple de modifications de mesures

Ajouter et supprimer des enregistrements (clés étrangères et mesures) dans la table des faits.

Le dernier point de révision de la table des faits s'intéresse aux ajouts et aux suppressions d'enregistrements dans la table des faits. Ces deux actions ont un impact direct dans la reconstruction des agrégations : la « construction » de nouvelles agrégations pour l'un et la

« destruction » d'agrégations existantes pour d'autres. Dans le cas d'un **ajout d'un nouvel enregistrement** (par exemple : 6 | DA 112 | DB 112 | Mes 5 | 13), le membre ajouté doit :

- Tenir compte de l'évolution des dimensions précédentes (i.e. ne pas associer la clé primaire d'une ancienne donnée détaillée qui est devenu agrégée, ne pas associer la clé primaire d'une donnée détaillée supprimée,…)
- Tenir compte des agrégations déjà existantes à modifier. Dans notre exemple, DA11 DB3, DA1 DB3, DA DB3, DA11 DB, DA1 DB, DA DB existent déjà avec DA111 DB15, le premier enregistrement de la table des faits.
- Ajouter toutes les autres agrégations qui n'existent pas dans la table ROLAP

Dans le cas de la **suppression d'un enregistrement** de la table des faits, les parents des clés étrangères qui n'ont aucun autre enfant que la donnée supprimée (les enregistrements sans les * dans notre exemple de la Figure 3-12) peuvent être supprimés sans risque. Cependant, les parents qui ont plusieurs enfants (i.e., qui sont la combinaison de plusieurs agrégations comme le sont les données DA et DB) ne peuvent être supprimés, mais sont juste modifiés en tenant compte de la donnée supprimée.

Ainsi la phase d'analyse d'un cube ROLAP a pu mettre en évidence sept modifications majeures du cube. Quatre concernent les dimensions et trois concernent la table des faits. On remarque que plusieurs facteurs vont compliquer la maintenance du cube ROLAP : le nombre d'opérations de modification, le nombre de dimensions que l'on souhaite modifier, le niveau dans la hiérarchie d'une dimension en voie d'être modifiée (plus l'ancêtre commun sera proche du sommet de l'arbre hiérarchique et plus la maintenance du cube sera longue et difficile), le nombre de clés étrangères impliquées dans les dimensions modifiées, le nombre de clés étrangères modifiées dans la table des faits.

3.5- Le SOLAP 3D

Le deuxième axe d'optimisation d'un outil d'analyse en ligne concerne l'introduction de la 3e dimension. Partant des travaux de Brisebois (2003) sur l'élaboration d'un outil SOLAP 3D, nous proposerons des réflexions qui permettront de présenter ce que l'introduction de la 3ème dimension exigerait d'un outil d'analyse en ligne de type SOLAP. La décision d'introduire ou non, la 3ème dimension est directement liée à la possibilité de surmonter deux barrières actuelles:

- la barrière technologique, pour appliquer les concepts théorique du SOLAP 3D,
- la barrière conceptuelle pour valider la réelle nécessité d'un SOLAP 3D.

3.5-1. Le SOLAP 3D : de la théorie à la pratique

Les difficultés de mettre en place, technologiquement, les concepts du SOLAP 3D concernent directement les concepts inspirés des SOLAP 2D (cf. section 2.4-2.) :

Un traitement de tous les types de primitives géométriques ?

La modélisation géométrique 3D est ni uniforme, ni conventionnée et ni normalisée bien que cela soit en cours (Bédard, 2006). En 2D, on distingue deux types de structures de données géométriques : la structure matricielle où la primitive est le pixel et la structure vectorielle (simple ou topologique) où les primitives sont les points, lignes, polygone. En 3D, on peut aussi faire cette différence en distinguant les approches orientées espaces et orientées objets (cf. section 2.2-3.). Cependant, on a pu constater que ces deux approches sont particulièrement complexes et apportent l'une et l'autre des avantages et des inconvénients. Actuellement, il n'existe aucun système informatique capable de traiter toutes les primitives géométriques 3D, même si certains logiciels comme GOCAD, offrent des possibilités de traiter certaines primitives volumiques comme les tétraèdres et les voxels et des primitives basées sur l'approche orientée objet comme des points, lignes et courbes (Pouliot, 2005) .

Une analyse spatiale possible ?

L'analyse spatiale 3D, comme nous avons pu le voir à la section 2.3-1. est possible si elle offre suffisamment d'opérateurs spatiaux 3D fonctionnels et si la structure des données le

permet. Comparativement aux opérateurs spatiaux 2D qui s'affinent, les opérateurs spatiaux 3D métriques et topologiques (cf. section 2.3-2.) n'en sont qu'à leur balbutiement car ces opérateurs dépendent directement du type de structures géométriques 3D. Comment effectuer une analyse topologique à une structure de données non topologique ? Il existe des structures topologiques de données 3D comme le résume Lachance (Lachance, 2005), mais force est de constater que l'emphase des outils technologiques actuels est plus portée sur la visualisation et la représentation 3D (donc sur l'analyse visuelle 3D) que sur l'analyse spatiale 3D et plus particulièrement l'analyse topologique 3D. L'avancé de la visualisation des données 3D au détriment de l'analyse spatiale 3D s'explique par le fait que la majorité des manipulations de ces données se fait sur un support 2D et ne laisse à l'utilisateur que la possibilité d' « apprécier et non de mesurer » (Pouliot, 2005). On comprend alors mieux l'intérêt des technologies actuelles d'optimiser l'analyse visuelle 3D et pourquoi les structures topologiques de données 3D ne sont pas encore génériques, même si des avancées dans ce domaine sont de plus en plus présentes.

La manipulation interactive des données géométriques 3D ?

L'interactivité est une « propriété d'un programme informatique qui permet à l'utilisateur d'interagir avec le système en modifiant le déroulement du contenu du programme » (OQLF, 2006). La manipulation interactive des données géométriques consiste à reformuler des requêtes à partir de résultats d'autres requêtes (soit des occurrences cartographiques, soit un tableau, soit un graphique) comme le propose les outils OLAP et plus spécifiquement les outils SOLAP. Ces derniers proposent en effet de pouvoir sélectionner tant dans la partie descriptive, tabulaire ou graphique que directement sur les occurrences cartographiques afin d'améliorer l'interactivité avec l'utilisateur.

En 3D, cette interactivité est plus difficile à mettre en place pour plusieurs raisons. La première concerne le point de vue de l'utilisateur (lorsque des objets se chevauchent…) et sa possibilité à cliquer sur les occurrences cartographiques 3D. La deuxième concerne le problème de la notion d'objets associés aux occurrences cartographiques. Cette notion d'objets est caractéristique des systèmes d'information géographique. Cependant, les CAO permettent une modélisation plus performante, donc une visualisation 3D optimisée, mais

ne sont pas optimisées quant à la notion d'objets qui rend problématique la manipulation interactive. La troisième raison concerne le type de modélisation choisi. Des modélisations comme le fil de fer ou la représentation par balayage rende difficile le concept de « sélection » de l'occurrence cartographique. De plus, comme il est difficile de sélectionner une occurrence cartographique qui est représentée sous forme de pixel en 2D, le concept de sélection d'une occurrence cartographique représenté sous forme de voxel en 3D est tout aussi complexe.

La faisabilité d'un SOLAP 3D : les couplages technologiques possibles

Le Tableau 3-1 présente les technologies actuelles comparées entre elles par rapport à leur manière d'exploiter les données descriptives et d'exploiter les données spatiales. Comme l'intégration de la 3ème dimension dans un outil SOLAP nous intéresse, nous avons distingué dans l'exploitation des données spatiales 3D : l'aspect visualisation, représentation 3D – l'aspect modélisation 3D – et l'aspect manipulation interactive 3D. Les logiciels sont :

- les Systèmes de Gestion de bases de données (**SGBD**). La fonction de ces systèmes transactionnels non spatiaux est d'assurer la gestion automatique d'une base de données et de permettre la création, la modification, l'utilisation et la protection des données (OQLF, 2006)
- les **SIG** (cf. section 2.4-1.),
- les outils **OLAP**, les outils **SOLAP** (cf. section 2.4-2.),
- les outils Conception assistée par Ordinateur (**CAO**). Ces outils tels les logiciels de conception comme Autocad, Microstation, Gocad, sont spécialisés pour la modélisation et la conception de dessin, d'objets géométriques. Les outils CAO offrent une analyse spatiale métrique 3D poussée [(Szalapaj, 2001) : p42] et, parce qu'ils sont capable de modélisation (principalement CSG et par frontière), la représentation 3D est plus efficace qu'un SIG. Malheureusement, les outils CAO ne permettent de gérer ni la notion explicite d'objet ce qui rend difficile d'y associer les attributs ni la topologie [(Longley et al., 2005) : p180] (Kemp, 2006).
- Les **visualiseurs 3D**. Ces visualiseurs sont principalement le VRML (Virtual Reality Markup Language) et ses « concurrents » directs X3D et Java 3D. Le X3D

(Extensible 3D) est un format de description de graphismes 3D pour le Web qui a été normalisé en 2005. Il est considéré comme le successeur direct du VRML avec pour avantages principaux un espace de stockage (quelques centaines de ko comparé aux Mo du VRML) et l'utilisation d'un langage basé sur le XML. Le format X3D (ainsi que VRML 2.0) peut être utilisé à l'aide de visualiseurs tels que Démotride (http://www.demotride.com) ou xj3D (http://www.xj3d.org/). Java 3D, quant à lui, fournit un ensemble de classes d'extension Java destiné à créer des scènes 3D en réalité virtuelle. Jview3DPro est un exemple d'outil permettant l'utilisation de Java 3D.

Tableau 3-1 : Exploitation des données descriptives et des données spatiales des technologies actuelles

Exploitation des données spatiales / Exploitation des données descriptives		NON	OUI			
			2D	3D		
				Vis 3D	Mod 3D	Manip Interact
OUI	Analytique	OLAP	SOLAP			
	Transactionnel	SGBD	SIG			
NON			CAO			
				Vis 3D		

Ainsi, concernant la barrière technologique du SOLAP 3D, il est, tout d'abord, actuellement technologiquement impossible de pouvoir effectuer une manipulation interactive des données géométriques 3D tel que définie plus tôt. De plus, le seul couplage technologique pour obtenir un premier prototype SOLAP 3D a été le couplage [OLAP + SIG]. Ce couplage avait pour objectif d'étudier la potentialité d'un outil SOLAP 3D. Le Tableau 3-1 nous montre qu'il existerait plusieurs combinaisons possibles. Cependant, afin de ne pas dépasser le cadre de la maîtrise, nous n'avons pas pu les mettre en œuvre.

3.5-2. La nécessité d'un SOLAP 3D

Le SOLAP est né de cette nécessité à vouloir et pouvoir interroger de manière interactive les données descriptives à l'aide entre autres, des occurrences cartographiques. Le SOLAP a vécu au Centre de Recherche en Géomatique de l'Université Laval au travers de plusieurs architectures distinctes successives – notamment : SGBD + SIG ([Access + SoftMap]), OLAP+SIG ([SQL Server (Analysis Services) + ProClarity + Geomedia] – dans différents domaines d'application (Sport, Foresterie, Santé,…). Nous allons examiner la nécessite et la pertinence d'un outil SOLAP 3D et nous allons le faire à travers trois axes de discussion:

1. L'influence de l'interrogation interactive des données descriptives sur des outils 3D
2. L'influence de la visualisation 3D pour l'analyse de données 3D
3. L'influence d'une synchronisation des outils d'analyses

Axe 1 : l'influence de l'interrogation interactive des données descriptives sur des outils 3D

Pouvoir juger de l'influence d'un outil d'interrogation interactif des données descriptives sur les outils de visualisation 3D revient à poser la question suivante : *« Est-ce que la navigation interactive des données descriptives amène une plus value à une représentation cartographique 3D ? »*. Cette question a souvent été posée avec les outils cartographiques 2D et les diverses comparaisons SIG, OLAP+SIG et SOLAP ont pu prouver la nécessité d'user d'une navigation interactive de données descriptives.

Des travaux ont été effectués pour coupler des outils performants de modélisation 3D avec des SGBD [SGBD+CAD] (Apel, 2004). Il peut être alors utile de voir la pertinence de coupler une technologie de navigation interactive de données descriptives avec un outil à fort potentiel 3D, ce qui n'a jamais été fait hormis l'approche avec un SIG de Brisebois (2003). Pour cela, une comparaison de plusieurs architectures serait nécessaire :

- OLAP + CAD 3D
- OLAP + Visualiseur 3D
- OLAP + CAD 3D + Visualiseur 3D

Plusieurs critères de comparaisons, non exhaustifs, sont alors à envisager :

- la faisabilité technologique : compatibilité des langages de programmation

- l'efficacité de l'analyse spatiale et la représentation tridimensionnelle non diminuée
- l'intérêt d'effectuer une analyse interactive de données descriptives sur un outil de modélisation 3D.

Axe 2 : l'influence de la visualisation 3D pour l'analyse de données 3D

Pouvoir juger de l'influence d'une visualisation 3D pour l'analyse de données 3D revient à se poser les questions suivantes :
« Est-il nécessaire de voir en 3D pour l'analyse de données 3D ? Est-ce que l'exploration interactive graphique 3D des données amènerait une plus value à l'analyse de données 3D ? Des coupes 2D, représentant différentes vues particulières, que l'on manipulerait interactivement peuvent elle être suffisantes pour l'analyse de données 3D ? »
Afin de répondre à ces questions, l'étude de plusieurs architectures et leur comparaison à l'outil SOLAP est à envisager :

- SOLAP + CAD 3D
- SOLAP + Visualiseur 3D
- SOLAP + CAD 3D + Visualiseur 3D

Actuellement la manipulation interactive 3D, tel que définie plus tôt, n'est pas possible, mais la manipulation interactive 2D existe dans l'outil SOLAP. Plusieurs critères de comparaisons, non exhaustifs, sont alors à envisager :

- la pertinence d'une navigation tridimensionnelle comparée à une manipulation interactive 2D
- la pertinence d'une représentation de plusieurs vues fixes 2D comparé à la navigation tridimensionnelle.

Axe 3 : l'influence d'une synchronisation des outils d'analyse

Quand bien même la nécessité d'user d'une exploration interactive spatiale et descriptive peut être étudiée, les besoins de synchronisation des deux outils d'explorations peuvent eux aussi être soulevés. Trois approches sont alors à envisager :

- le **couplage direct** des navigations (synchronisation complète des deux navigations comme le propose actuellement l'outil SOLAP mais pour le 2D : interrogation de manière interactive les données descriptives et graphique et analyse spatiale

Rapport-Gratuit.com

possible). Ce couplage est le plus difficile à mettre en œuvre puisqu'il exige une compatibilité totale entre les deux systèmes tant au point de vue de l'accessibilité que sur la faisabilité

- le **couplage sens unique** - navigation descriptive vers navigation graphique. Par exemple, on peut exporter un fichier issu de l'analyse descriptive qui soit lisible par l'outil spatial 3D. Ce couplage peut être plus simple à mettre en œuvre au détriment d'une interactivité entre les deux outils. Le couplage sens unique - outil spatiale 3D vers navigation descriptive marquerait la prédominance des analyses graphiques, géométriques et visuelles par rapport à l'analyse descriptive ce qui semblerait être moins pertinent que le premier couplage sens unique.
- le **couplage intermédiaire**. Ce couplage permet de faire un couplage direct entre une ou plusieurs parties d'un ou des deux outils d'exploration. Cette solution intermédiaire aurait pour objectif une meilleure compatibilité et accessibilité des outils. Un exemple de couplage intermédiaire est présenté en
- Figure 3-18.

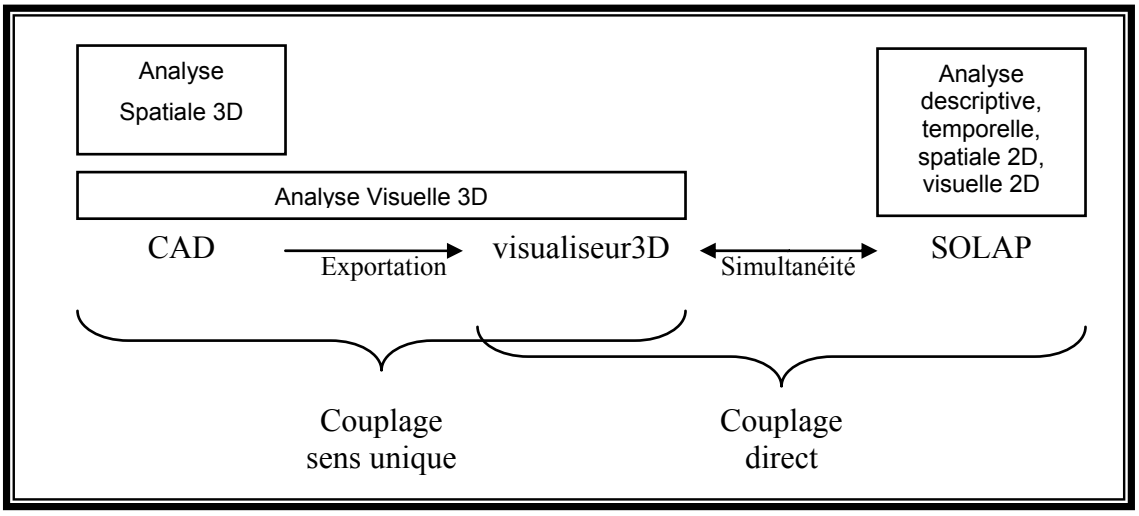


Figure 3-18 : Exemple de couplage intermédiaire

3.6- Conclusion

Pour optimiser un système d'analyse en ligne, nous avons identifié deux grands axes possédant des perspectives de développement intéressant soit : le processus de révision des données et l'introduction de la 3e dimension.

Concernant le processus de révision de données, nous avons présenté de nouveaux concepts sur le processus de révision ainsi que sur les données susceptibles d'être révisées. Ces données, que nous avons qualifiées de données d'interprétation se différencient des données d'observation par leur capacité à évoluer pendant la phase d'analyse. Les données d'observation quant à elles ne seront mises à jour qu'à partir du moment où la réalité évoluera et que des capteurs auront saisis cette modification. D'un point de vue conceptuel, le processus de révision d'un outil d'analyse en ligne se produit sur deux niveaux : la structure multidimensionnelle et les données elles-mêmes. Nous avons choisi de n'étudier que le processus de révision de données. Toutes les structures multidimensionnelles des outils d'analyse en ligne offrent la possibilité de manipuler les données d'interprétation. Cependant, une seule, la structure ROLAP, sous la simulation de schéma en étoile avec la modélisation des dimensions en « parent-enfant », offre les meilleures caractéristiques pour optimiser et comprendre le processus de révision des données d'interprétation. Cependant, nous avons constaté que la construction des agrégations est un travail qui demande beaucoup de temps mais semble t'il nécessaire à chaque révision d'une donnée d'interprétation. Or, la notion de temps est essentielle puisqu'une reconstruction trop longue pourrait stopper le processus de réflexion de l'utilisateur. Ainsi nous avons pu remarquer qu'une reconstruction partielle engendrerait un gain de temps considérable.

Concernant l'introduction de la 3ème dimension, nous avons constaté la difficulté à mettre en œuvre les concepts du SOLAP 3D principalement en raison de la faiblesse des technologiques actuelles en matière de 3D. Face à ces difficultés technologiques, des questions sur la nécessité d'un SOLAP 3D ont été soulevé à savoir : pour analyser des données 3D, l'interrogation interactive des données descriptives est-elle utile aux outils 3D ? La visualisation 3D, à la différence d'une visualisation 2D judicieuse, est elle nécessaire pour une analyse de données 3D ?

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

Chapitre 4 : Application des concepts du SOLAP 3D à un contexte de fouille archéologique

Voyons maintenant comment nous avons mené une expérimentation afin d'optimiser le processus d'analyse en ligne de données 3D dans un domaine en particulier : l'archéologie. Avant de présenter en détails ces expérimentations et le prototype développé, ce chapitre propose un rappel des différents types de données archéologiques recueillies sur un chantier de fouilles archéologiques, des besoins spécifiques des fouilleurs en regard de l'analyse de leurs données 3D et plus particulièrement d'un outil d'analyse compatible avec le processus de compréhension d'un site archéologique. Nous conclurons en examinant les réelles capacités du prototype et des concepts sous-jacents à répondre à notre objectif initial qui consistait à gérer adéquatement l'évolution des données (spatiales ou non) dans une structure multidimensionnelle et ce, simultanément à l'étape d'analyse du spécialiste.

4.1-La constitution des données archéologiques

Pour notre expérimentation, nous avons utilisé les données fournies par la fouille du site de Tell 'Acharneh, en Syrie, dirigé par Michel Fortin, professeur d'archéologie à l'Université Laval. Il s'agit d'un site de grandes dimensions : 70 hectares environ (1,2 km du nord au sud X 500-650 m d'est en ouest) qui se trouve dans la moyenne vallée de l'Oronte, à 35 km environ au nord-ouest de la ville moderne de Hama. L'idée d'inclure les technologies et les approches géomatiques dans un processus de fouille archéologique intéresse depuis plusieurs années le professeur Fortin. Certains travaux de *géomatisation*[5] ont été lancés ces dernières années en vue d'uniformiser le processus d'acquisition des données sur un chantier de fouilles archéologiques (Lachance *et al.*, 2006), d'inclure dans cette étape d'acquisition un balayeur laser (Marchand *et al.*, 2006)de modéliser en 3D les données extraites d'un tel chantier (Losier, 2005; Losier *et al.*, 2007)et même d'utiliser un SOLAP pour les besoins d'analyse (Rageul, 2004). Les données qui sont décrites ici sont donc celles propres au chantier de Tell 'Acharneh, un site dans un contexte proche-oriental.

[5] Le terme *géomatisation* sera utilisé dans ce document pour faire référence à l'introduction de méthodes et d'outils principalement exploités dans le monde de la géomatique (incluant ici plus spécifiquement les technologies de l'information, les disciplines de la cartographie numérique, des GPS, des SIG, des traitements numériques d'images).

Cependant, les mêmes catégories de données se retrouvent sur tous les sites archéologiques, peu importe les aires culturelles, à quelques différences mineures près.

Il va de soi, qu'étant donnée la nature de ce mémoire, la présentation de ces données sera ici sommaire et simplifiée. Parmi ces données, nous verrons celles qui sont d'observation et celles qui sont d'interprétation d'après les concepts tels que précédemment proposés comme classification au chapitre 3. Plus spécifiquement, ces données, et plus particulièrement les données d'observation ne sont pas encore exhaustives car la classification de données en « données d'observation » dépend directement de l'acquisition des données. Cette phase du projet n'étant pas encore complètement opérationnelle, ces données peuvent encore évoluer par rapport à la réalité du terrain et à l'intérêt porté par l'archéologue-fouilleur sur ce qu'il observe et ce qui l'intéresse pour son analyse.

4.1-1. Les objets archéologiques

Pour le chantier de Tell 'Acharneh, les fouilleurs reconnaissent deux classes d'objets archéologiques :

- les unités de fouille ou « UF » (auparavant appelées « lots »)
- les éléments de la culture matérielle inclus dans les UF et formant le catalogue

Les **unités de fouille** sont des entités volumétriques de terre (pour les UF-débris) ou de matériaux de construction (pour les UF-constructions) formant un site archéologique. Les UF sont arbitrairement délimitées par les fouilleurs en fonction de données observables sur le terrain (voir Figure 4-1). En somme, un site archéologique, lors de sa fouille, est fragmenté ou segmenté en un amalgame d'unités de fouille.

Figure 4-1 : Exemples d'Unités de Fouille–Construction (gauche) et – Débris (droite)

La **culture matérielle** trouvée dans ces unités de fouille peut être regroupée en quatre grandes catégories (Figure 4-2). Les **artéfacts** sont des objets qui ont été façonnés par un humain; ceux fabriqués à partir d'un matériau d'origine minérale sont plus nombreux étant donné leur haut degré de préservation (matériau non-périssable). Bien qu'étant en réalité un artéfact, la **céramique** est mise dans une catégorie à part en raison du nombre impressionnant de fragments recueillis lors d'une fouille archéologique. Les **écofacts** sont des restes naturels, non travaillés, d'un organisme végétal ou animal; un fragment d'os ou une tige de bois transformés en objets utilitaires ou esthétiques sont considérés comme des artefacts. Les **restes humains** sont différenciés des ossements d'animaux et forment une dernière catégorie. L'ensemble de la culture matérielle forme le catalogue.

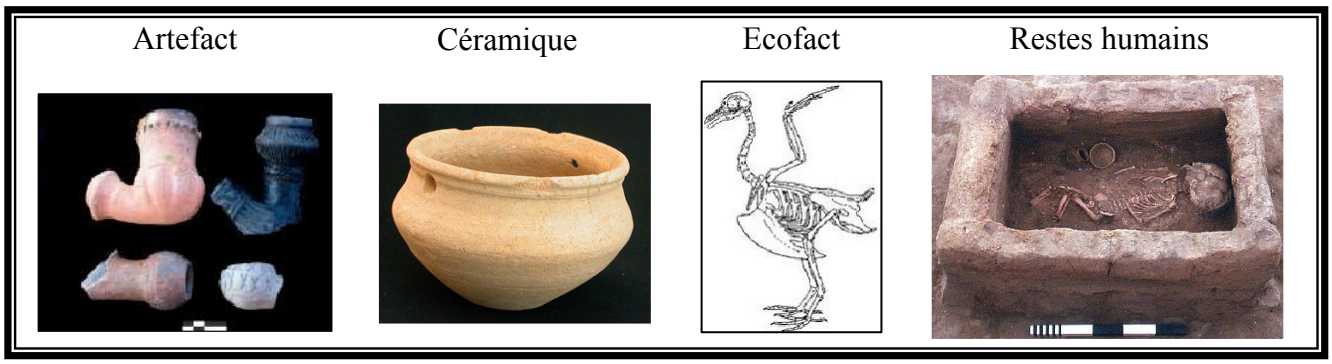


Figure 4-2 : Les quatre catégories de matériel archéologique retiré des unités de fouille d'un chantier archéologique

Autour de ces classes d'objets, vont graviter des données d'observation retenues pour l'analyse et des données d'interprétation.

4.1-2. L'interprétation des objets archéologiques

La compréhension d'un site archéologique passe dans un premier temps par la compréhension de la culture matérielle. Les spécialistes de la culture matérielle (ex. céramologue) inventorient le matériel archéologique retirés des unités de fouille puis le trient afin de retenir (cataloguer) les éléments qui apportent le plus d'informations (on élimine donc certaines données d'observation). On rappelle qu'une donnée d'observation est une donnée issue de systèmes d'acquisition des données ou de systèmes transactionnels (cf. chapitre3). Le matériel archéologique, notamment la céramique, est ensuite analysé afin de constituer les données d'interprétation qui les composent. On rappelle qu'un donnée d'interprétation est une donnée directement issue d'un besoin de l'utilisateur à revoir son analyse afin d'enrichir, de comprendre et/ou d'expliquer sa perception de la réalité, ceci

sans mettre à jour les données d'observation (cf. chapitre3). Le catalogue regroupe donc des données d'observation (jugées pertinentes pour l'analyse) et des données d'interprétation du matériel archéologique.

Dans un deuxième temps, la compréhension d'un site passe par la reconstitution virtuelle de ses constructions et de ses aires d'activité (cf. Figure 4-3). Les « Constructions » (ex. bâtiment) sont formées à partir de l'agrégation d' « Éléments de Constructions » (ex. murs, planchers, etc.) qui eux-mêmes résultent de l'agrégation d'UF-Constructions. Les aires d'activité sont constituées à partir de l'agrégation de « Constructions » et d'UF-débris ainsi que les éléments de culture matérielle que ces dernières contiennent. Toutes ces données ont été générées par une activité sociale dans le passé.

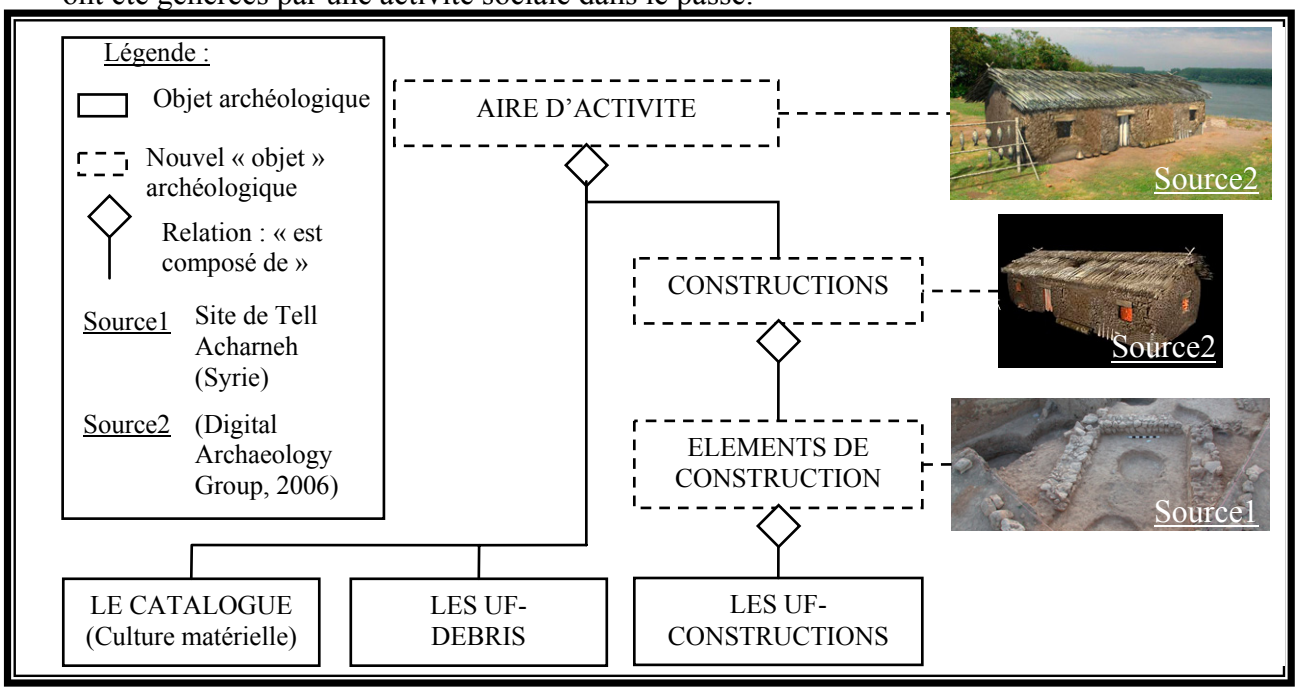


Figure 4-3 : Exemple simplifié de reconstitution de l'aire d'activité

En pratique, cependant, toutes les UF-constructions d'un site, ne pourront pas nécessairement être agrégées en « Éléments de Constructions ». Pareillement, des « Éléments de Constructions » ne pourront peut-être pas être agrégés en « Constructions ». Ainsi, en pratique, une aire d'activité pourra facilement inclure aussi des UF-constructions et des « Éléments de Constructions » non agrégés. Finalement, même si nous n'en tiendrons pas compte dans notre recherche, la dernière étape d'interprétation d'un site archéologique passe par la reconstitution de ses « Niveaux d'Occupation » superposés les uns aux autres.

Chaque « Niveau d'Occupation » regroupe des aires d'activités d'une même période temporelle ainsi que les « Constructions », « Éléments de Construction », « UF Constructions », « UF-débris » et « culture matérielle ». Il est pertinent de signaler ici que tous les éléments constitutifs d'un niveau sont contemporains, d'une part, et, d'autre part, que la datation d'un niveau est fournie par les éléments de culture matérielle.

4.1-3. Les données d'observation

La description des unités de fouille et des différentes catégories de la culture matérielle se fera à partir des données d'observation recueillies pendant la fouille. La Figure 4-4 montre, dans le contexte d'un chantier subdivisé en carrés de fouille, des exemples de données d'observation propres aux UF-Construction et aux UF-Débris. Elles comportent des attributs descriptifs (ex. consistance, granulométrie, couleur de la terre pour les UF-débris, sortes de matériaux pour les UF-constructions et mesures qui contiendront ultérieurement des données telles que l'archéologue les observera sans autre jugement de sa part). Toutes ces données sont des données d'observation car elles sont enregistrées par les archéologues/fouilleurs au moment même de la fouille, soit lors de la cueillette des données, avant qu'ils puissent leur attribuer une fonction ou une interprétation quelconque. Ce sont des données brutes, non-interprétées ou analysées.

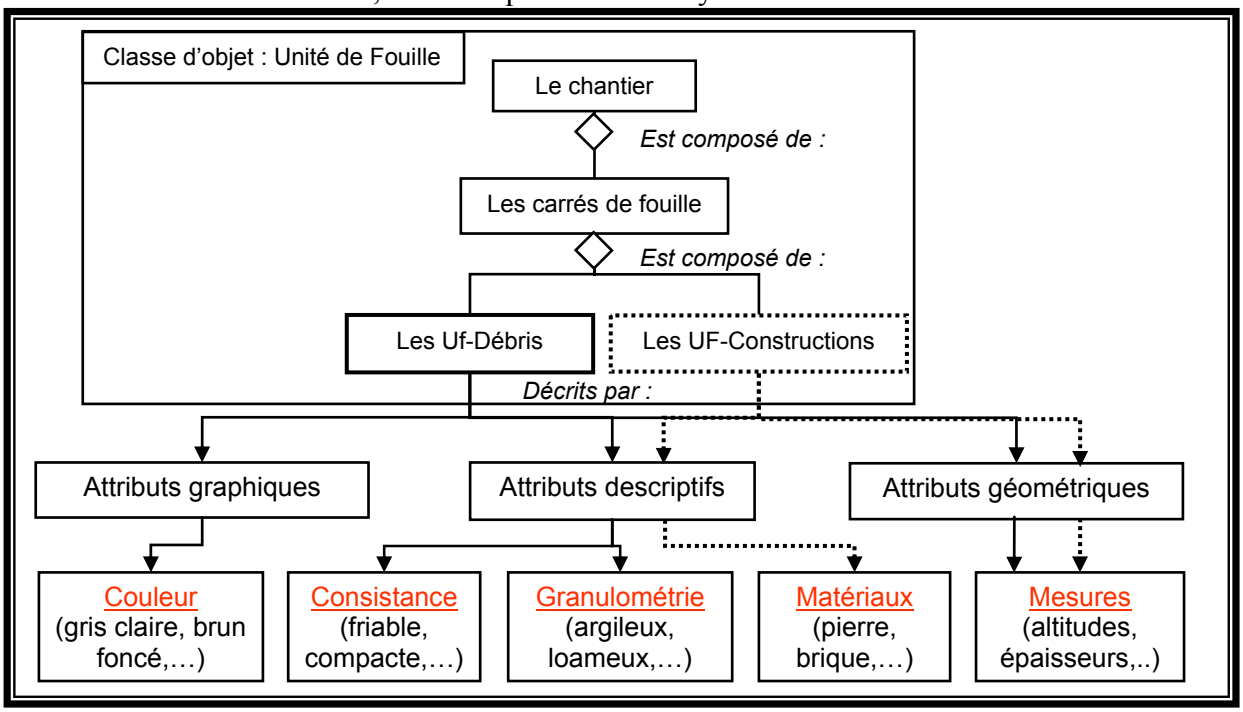


Figure 4-4 : Du chantier de fouille aux données d'observation.

Par conséquent, à cette étape, très peu de données d'observation seront notées en regard du matériel archéologique (culture matérielle); ce seront essentiellement :

- l'UF dans laquelle le matériel archéologique est trouvé. En effet, le matériel archéologique est enregistré comme « composante » de l'UF et non comme un objet spatial à part entière avec ses propres coordonnées;
- le « nombre total » des éléments de chaque catégorie de la matériel archéologique

4.1-4. Les données d'interprétation

Les données d'interprétation, comme son nom l'indique, sont issues de l'interprétation des objets archéologiques. Elles regroupent donc les données produites lors du catalogage des éléments de la culture matérielle (artéfacts) (cf. Figure 4-5) et la reconstitution des aires d'activités.

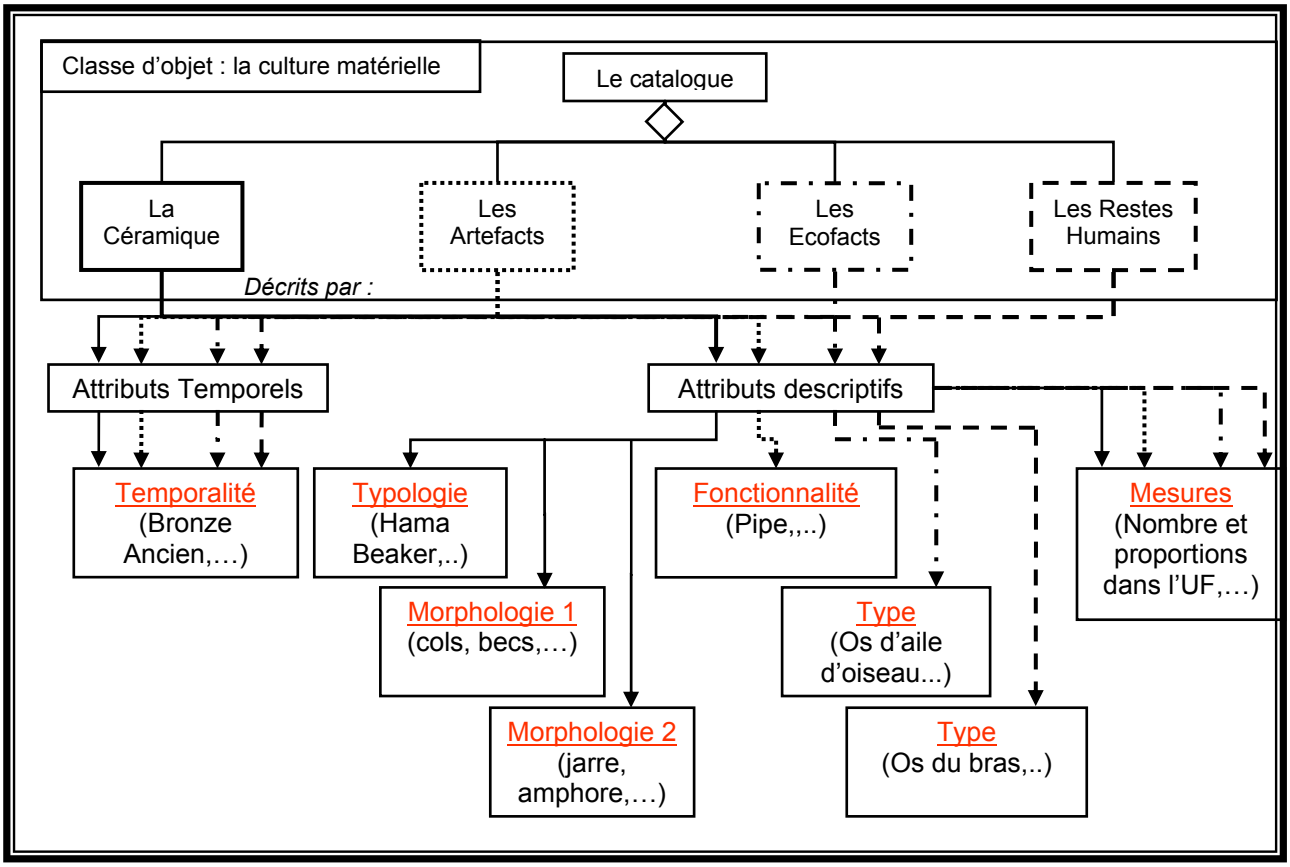


Figure 4-5 : Les données d'interprétations générées au cours de l'opération catalogage d'éléments de la culture matérielle

On remarquera sur la Figure 4-5 que les attributs temporels et les mesures sont omniprésents dans les différents objets constituants la culture matérielle. Plus spécifiquement, la céramique se distingue par des attributs descriptifs relatifs à la typologie et la morphologie (qui passe par une double interprétation). Les « Artefacts », quant à eux, se démarquent par un attribut descriptif portant sur leurs fonctionnalités alors que les « Ecofacts » ou les « Restes Humains » se différencient par leur type.

Concernant la classe d'objet « Unité de Fouille », la Figure 4-6 montre quelles peuvent être les données d'interprétation issues de l'analyse des unités de fouille. Ainsi, les attributs temporels résultent bien évidemment de l'interprétation de données temporelles relatives à la culture matérielle contenue dans cette UF. Les « Éléments de Construction » sont décrits par des attributs descriptifs de fonctionnalité et les « Constructions » sont quant à elles décrites par des attributs descriptifs d'inventaire. Ces « Constructions » et « Éléments de Construction » étant formés à partir d'UF-Constructions de même temporalité, ils héritent chacun de ces attributs temporels. Nous avions déjà vu à la Figure 4-4 une classe d'objet « Unité de Fouille ». Cependant cette classe d'objet était arbitraire puisque qu'elle dépendait du découpage en carrés de fouille du site archéologique. Il devient alors nécessaire de composer une nouvelle classe d'objet « unité de fouille » qui permettra la reconstitution des « Constructions ».

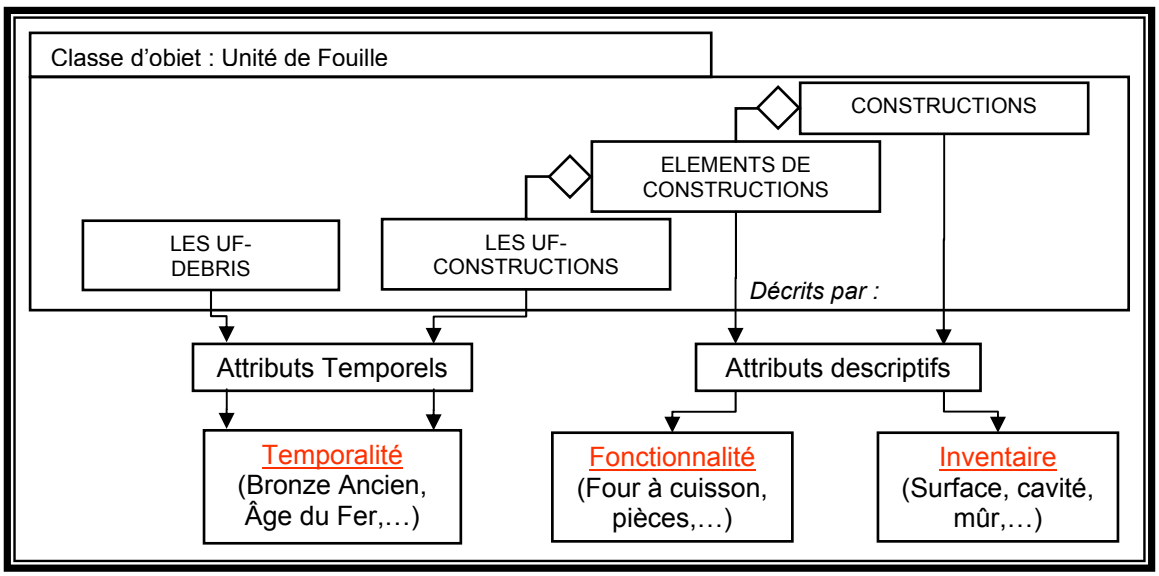


Figure 4-6 : Les données d'interprétations archéologiques issues des unités de fouille

Dans une structure multidimensionnelle, d'autres données d'interprétation seront les membres agrégés des dimensions (cf. Figure 4-7). Les membres les plus détaillés (i.e. le niveau le plus bas) de ces dimensions seront les données d'observation issues de l'observation des objets archéologiques et les données d'interprétation provenant de l'interprétation des mêmes objets.

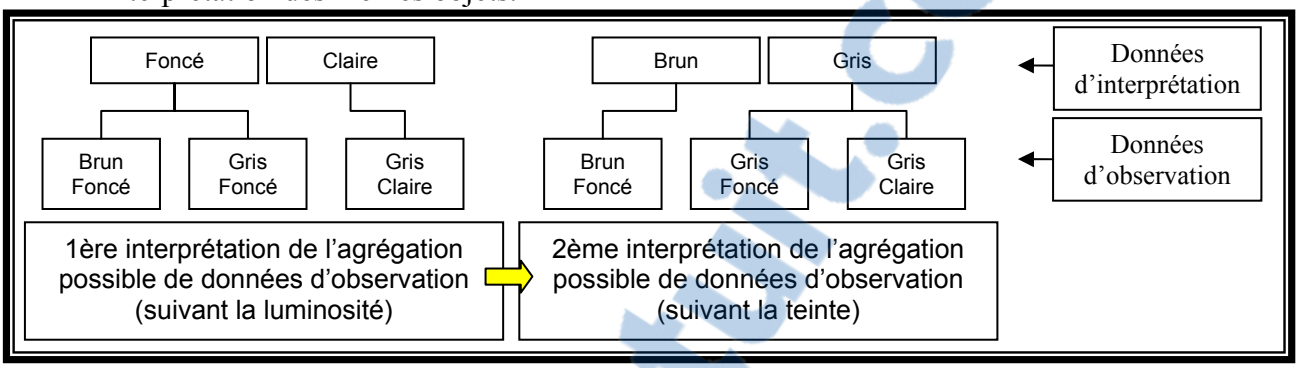


Figure 4-7 : Les données d'interprétations issues des dimensions (exemple de la couleur)

4.1-5. Les modèles 3D archéologiques

Pour l'aider dans son interprétation du site et sa reconstitution des aires d'activité, l'archéologue peut, dans des conditions idéales, documenter son site en constituant des modèles tridimensionnels. Ces modèles sont construits à partir des attributs géométriques et éventuellement graphiques des unités de fouille. En se fondant sur des données issues du chantier de Tell 'Acharneh, Losier (2005) a étudié les différents modèles géométriques qui pouvaient être générés pour un chantier de fouilles archéologiques. Parmi les trois modèles étudiés (surfacique, tétraédrique et voxel, cf. Figure 4-8), le modèle tétraédrique semble le mieux adapté à l'archéologie car il permet la gestion de la transparence, est peu volumineux pour un jeu conséquent de données, léger à manipuler, rapide à l'affichage (Losier, 2005; Losier et al., 2007).

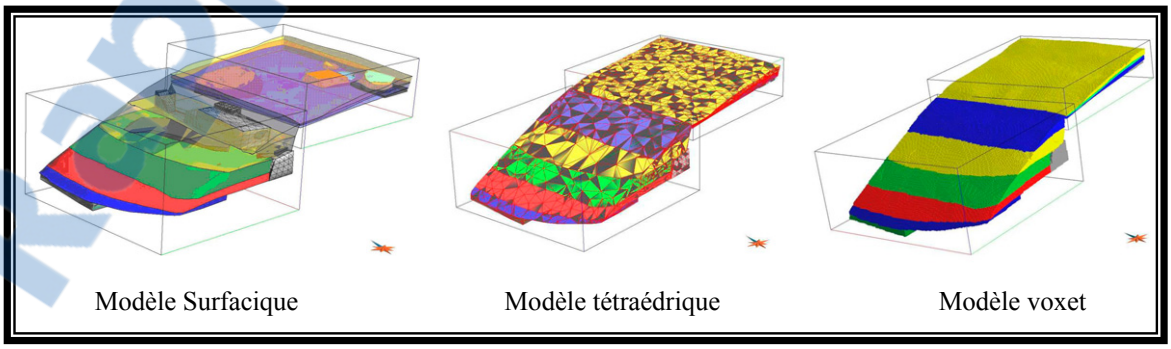


Figure 4-8 : Exemples de modèles 3D archéologiques (Tell 'Acharneh) (Losier, 2005).

Rapport-Gratuit.com

4.2-L'analyse des besoins des archéologues

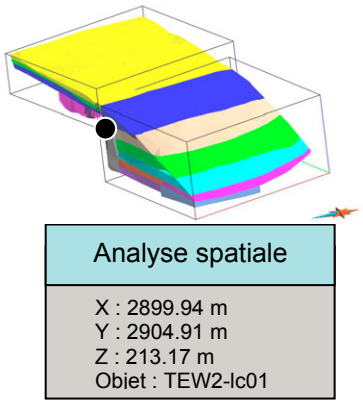
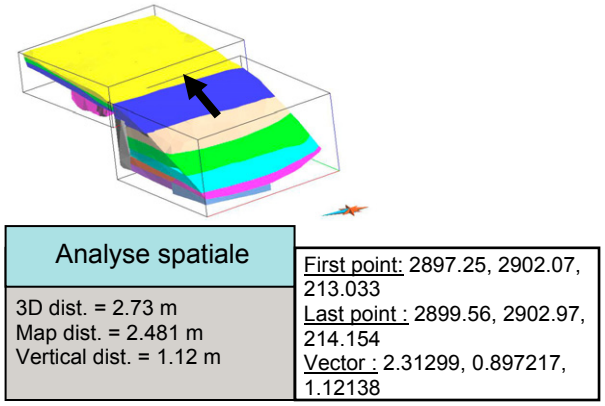
L'objectif des archéologues/fouilleurs est de reconstituer les aires d'activité constituant le site archéologique à partir des données acquises et décrites dans la section précédente. Cette partie présente les besoins recensés des archéologues quant à l'analyse de leurs données.

4.2-1. Les besoins généraux sur l'analyse des données

Les archéologues ont un premier besoin sur le type d'analyse (cf. sections 2.3-1.) à effectuer. L'**analyse descriptive** porte autant sur l'UF (« *Existe-t-il des unités de fouilles ayant une consistance, une granulométrie particulière ?* »,…) que sur la culture matérielle (« *De quelle morphologie est la céramique qui est en plus grand majorité dans une unité de fouille particulière ?* », « *Existe-t-il des unités de fouille ayant une couleur particulière avec un céramique d'un type particulier ?* »,…). Cette analyse est très importante puisqu'elle permet, par l'analyse des caractéristiques de la céramique, de dater celle-ci et par la même occasion d'attribuer une datation à l'UF.

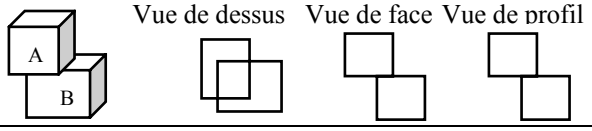
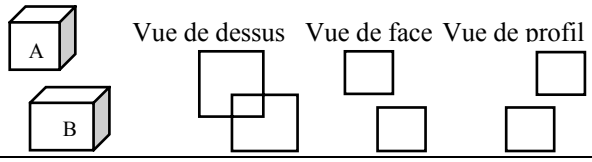
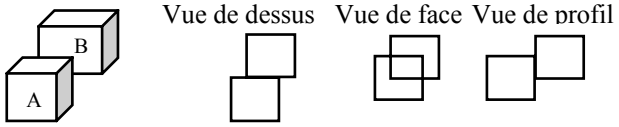
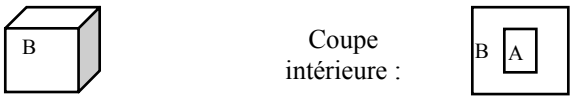
L'**analyse spatiale** s'applique exclusivement aux UF puisque ce sont les seules données qui portent la référence spatiale. Le besoin est tant sur l'aspect métrique (« *De combien de mètres sont distancés les deux unités de fouille* », « *Où sont positionnées les UF ayant une couleur, une consistance et une granulométrie particulière* »,…) que topologique de l'analyse spatiale. Les opérateurs spatiaux métriques utilisés sont des opérateurs métriques 3D : la position, la longueur/distance, l'aire/superficie, le volume et le corridor (cf. Tableau 4-1).

Tableau 4-1 : Exemple d'analyses spatiales métriques 3D appliquées à l'archéologie

Analyse métrique de position	Analyse métrique de distance
	

Des opérateurs spatiaux topologiques 3D peuvent aussi être appliqués à une « Unité de Fouille ». Ils proviennent d'une adaptation des opérateurs topologiques 2D. Comme les opérateurs métriques, les opérateurs spatiaux topologiques concernent exclusivement les unités stratigraphiques (cf. Tableau 4-2).

Tableau 4-2 : Exemple d'analyses spatiales topologiques 3D appliquées à une UF

Les opérateurs topologiques 3D		**Exemple**
Disjonction		Aucun lien entre les deux objets
Adjacence	Adjacence verticale avec contact	
	Adjacence verticale sans contact	
	Adjacence horizontale	
Inclusion		
Égalité		Les deux objets sont confondus

Comme l'analyse spatiale, l'**analyse temporelle** est aussi extrêmement importante pour l'analyse de données issues d'une fouille archéologique et principalement pour l'analyse stratigraphique. Cette analyse se fait suivant une échelle chronologique (cf. Figure 4-9) qui n'est pas figée puisqu'elle dépend de l'interprétation des archéologues en fonction, entre autres, de la localisation du site en question, des phases évolutives déjà connues, des différentes groupes humains ayant vécu dans la région ou se trouve le site fouillé et des découvertes ayant été faites par d'autres archéologues ayant fouillés dans la même région. Les époques, et les périodes inclues dans cette chronologie le seront en regard des *matériaux* utilisés par l'Homme, des *artefacts*, en regard également de ses *modes de vie*, voire même en tenant compte des noms des peuples qui y ont vécu et dont les noms nous sont connus par des écrits anciens (ex. période perse ou romaine). L'analyse métrique temporelle ne repose pas nécessairement sur des dates absolues car pour les périodes anciennes ces dates ne sont pas toujours connues avec exactitude. Donc à la différence

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

d'une analyse métrique de distance dont l'unité de mesure est le nombre d'années par exemple, on pourrait avoir une analyse temporelle métrique de distance dont l'unité de mesure serait les époques.

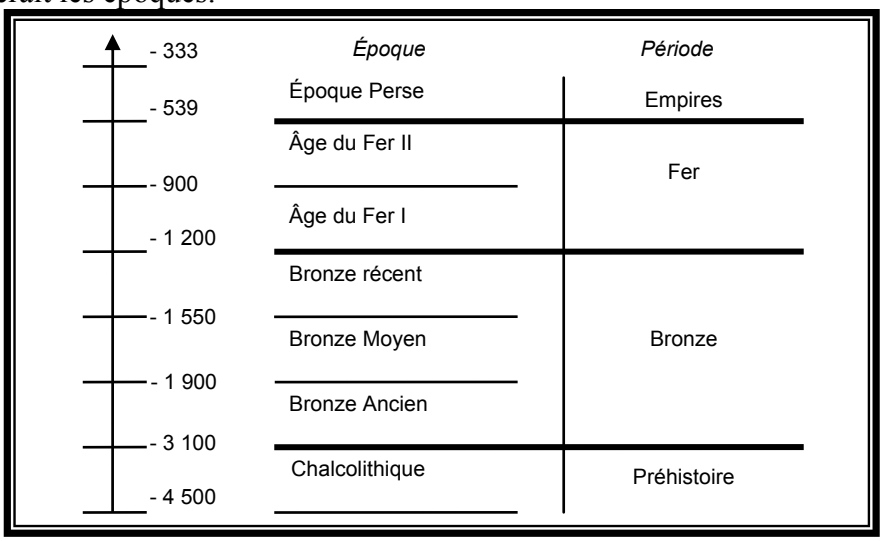


Figure 4-9 : Exemple d'échelle temporelle archéologique

Les opérateurs temporels topologiques peuvent aussi être adaptés à l'archéologie. Les opérateurs sont plus simples puisqu'ils mettent en évidence des durées qui ne se chevauchent pas. Les opérateurs retenus sont : avant sans continuité, avant avec continuité, pendant, après avec continuité, après sans continuité. De plus, les archéologues sont attachés tant sur les époques que sur les périodes. Nous pouvons alors coupler les différents opérateurs temporels ce qui nous donne le Tableau 4-3.

Tableau 4-3 : Exemple d'analyses temporelles topologiques 3D appliquées à l'archéologie

Opérateurs – Période	**Opérateurs – Époque**	**Exemple (cf. Figure 4-9)**
Avant Discontinuité	Avant avec discontinuité	Chalcolithique→Age Fer I
Avant Continuité	Avant avec discontinuité	Chalcolithique→ Bronze Récent
	Avant avec continuité	Chalcolithique→ Bronze Ancien
Pendant	Avant avec Discontinuité	Bronze Ancien → Bronze Récent
	Avant sans Discontinuité	Bronze Ancien → Bronze Moyen
	Pendant	Bronze Moyen → Bronze Moyen
	Après sans discontinuité	Bronze Récent → B.M.
	Après avec discontinuité	Bronze Récent → B.A.
Après continuité	Après avec continuité	Empire Perse→ Age Fer II
	Après sans continuité	Empire Perse → Age Fer I
Après sans continuité	Après sans continuité	Empire Perse → Bronze Récent

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

L'**analyse visuelle** est finalement tout aussi importante pour les archéologues dans l'analyse de leurs données. Elle nécessiterait entre autres : une présence de vues 3D, un forage graphique 3D (manipulation interactive graphique), de générer des vues 2D par slicer (coupes interactives 2D à partir d'un solide 3D), une fenêtre de manipulation des données 3D, une navigation assistée pour explorer le résultat d'une requête (cf. Figure 4-10).

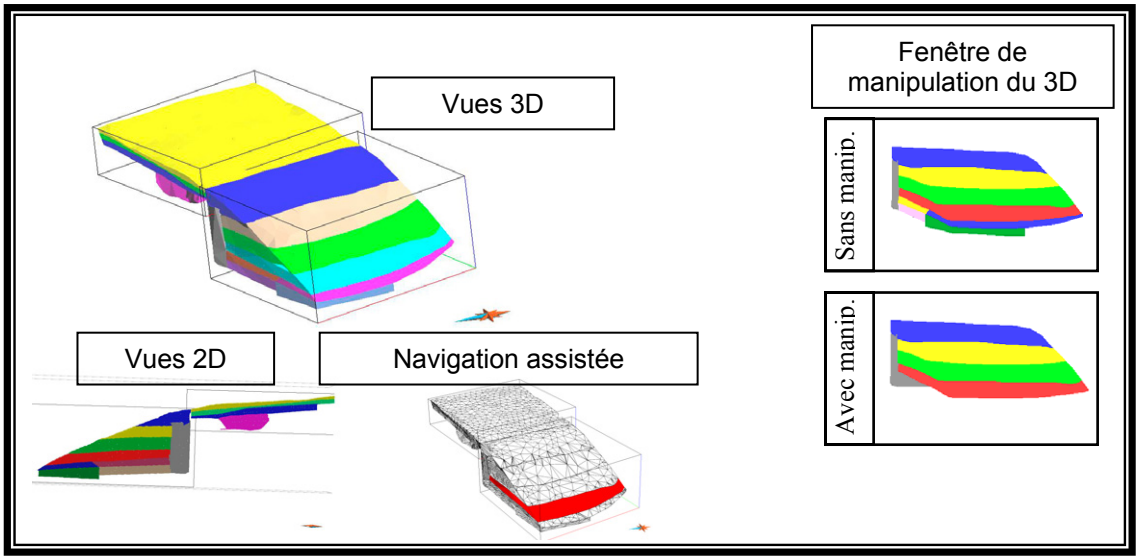


Figure 4-10 : Exemple d'analyses visuelles 3D appliquées à des « Unités de Fouille » (Image tirée de la modélisation de Losier 2005)

4.2-2. Un besoin spécifique : la compatibilité cognitive

Un besoin spécifique des archéologues concerne la manière d'analyser les données. La recherche archéologique relative à la compréhension et l'interprétation des données issues d'un chantier de fouilles archéologiques s'inscrit dans un processus de **découverte de connaissance**. Nous prétendons que ce processus est constitué de quatre grandes étapes itératives (cf. Figure 4-11) :

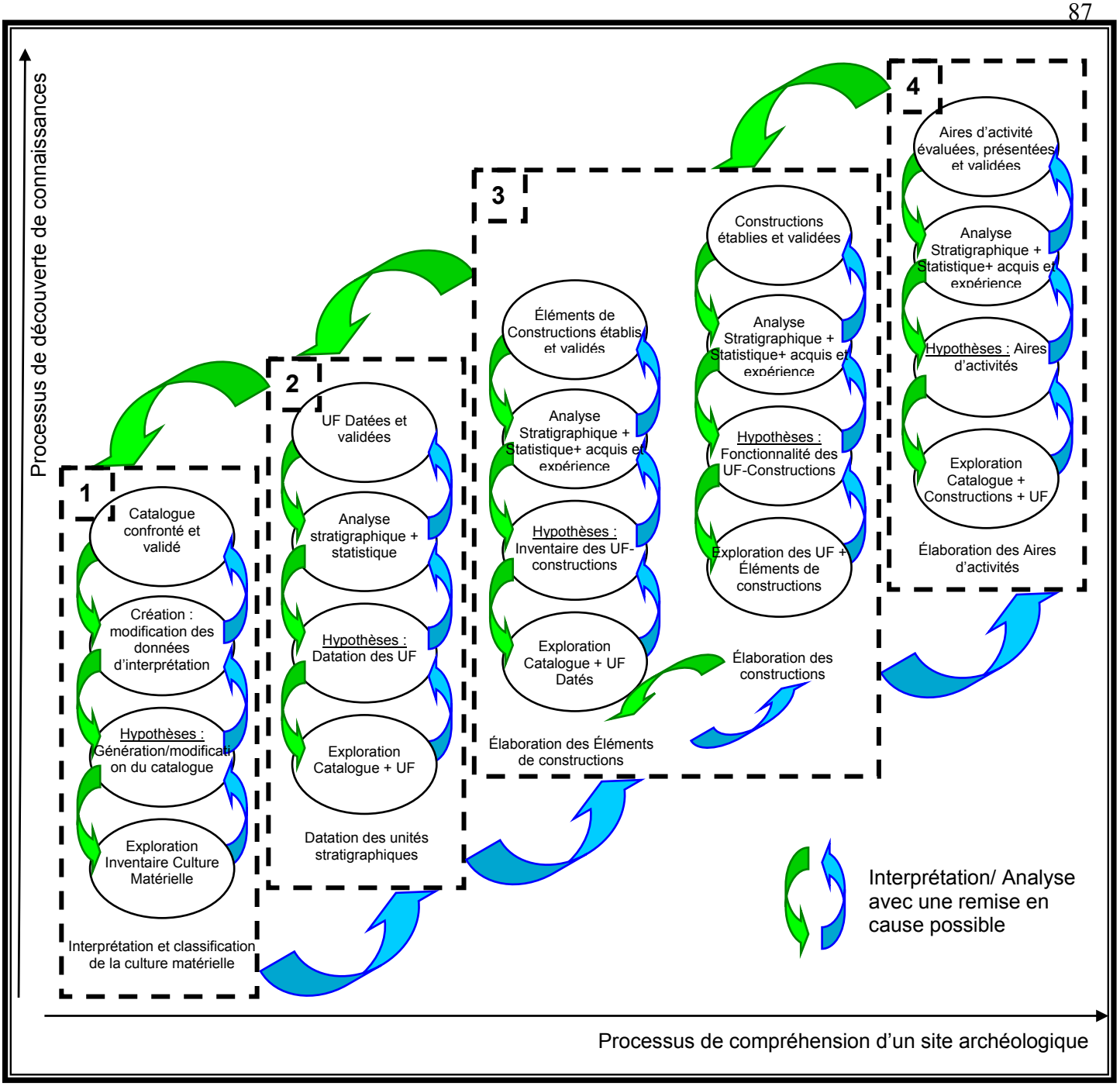


Figure 4-11 : Les quatre étapes de la compréhension d'un site archéologique

Étape 1 : Interprétation et classification de la culture matérielle (cf. encadré 1 Figure 4-11). Cette étape permet de mettre en évidence ce qui est appelé le catalogue, qui regroupe toutes les données (d'observation et d'interprétation) concernant la culture matérielle (ex. « Artefacts ».

Étape 2 : Analyse stratigraphiques pour la datation des « Unités de fouille » (cf. encadré 2 Figure 4-11). L'archéologue/fouilleur va chercher à établir des relations implicites entre la culture matérielle et les unités de fouille (- construction et/ou - débris) qui la contiennent et, d'autre part, la relation entre les UF. Cette dernière analyse est plus simple par une analyse spatiale. Il s'aide pour cela d'une analyse topologique spatiale et temporelle afin de comparer les UF en fonction de leur voisinage et de vérifier la cohérence notamment en ce qui nous concerne de la datation des UF en cause. Ce type d'analyse est souvent appelé une analyse stratigraphique par les fouilleurs car leur préoccupation première est de reconstituer la superposition stratigraphique du site fouillée qui témoigne d'une séquence événementielle passée. De nouvelles analyses statistiques sur la culture matérielle sont effectuées au cours du processus et une révision des données d'interprétation relatives au matériel archéologiques peut remettre en cause l'Étape 1 via une nouvelle interprétation.

Étape 3 : Interprétation des UF-constructions pour former des « Éléments de Construction » et des « Constructions » (cf. encadré 3 Figure 4-11). Au moment des analyses stratigraphiques, statistiques et de voisinage, et fort des acquis et expérience de l'analyste, les UF-Construction s'agrègent en Éléments de Construction. Un inventaire (surface, cavité, mur,…) de ces nouveaux objets archéologiques est dressé. Puis de nouvelles explorations des données et principalement concernant les UF-construction non agrégés et les « Éléments de Construction » nouvellement créés sont faites. De nouveaux objets archéologiques, les « Constructions », apparaissent et des interprétations sur leur fonctionnalité (four à cuisson, mur de défense,…) sont élaborées. Il est très important de noter qu'une mauvaise datation des Unités de Fouille (retour à l'étape 2) et/ou un catalogue mal interprété (retour à l'étape 1) peuvent remettre en cause le processus d'interprétation et de restitution des « Constructions » voire même l'ensemble du processus de compréhension du site fouillé.

Étape 4 : Reconstitution de l'aire d'activité (cf. encadré 4 Figure 4-11). L'archéologue va ensuite poursuivre son processus analytique en agrégeant toutes les UF (incluant les artéfacts compris dans ces UF), les « Éléments de Construction », les « Constructions » qui sont contemporains afin de reconstituer des aires d'activités appartenant à un seul et même niveau d'occupation du site fouillé. Pour cela, il va devoir analyser à nouveau la culture

matérielle, visualiser les données spatiales (UF-débris, UF-constructions, « Éléments de Constructions », « Constructions ») précédemment interprétées, donc datées, et s'aider des données d'interprétation (inventaire et fonctionnalité) qu'il s'est créé. Des hypothèses sont alors posées quant aux agrégations possibles des différents objets archéologiques. Toutes ces analyses sont utiles et nécessaires à l'archéologue pour l'aider dans sa recherche. Une nouvelle analyse stratigraphique sur les Unités de Fouille, de nouvelles statistiques sur la culture matérielle, l'expérience de l'archéologue/fouilleur,… sont autant de facteurs permettant de confirmer ou de réfuter l'hypothèse. De mauvaises interprétations sur les constructions et/ou éléments de constructions (retour à l'étape 3), sur la datation (retour à l'étape2) et sur la culture matérielle (retour à l'étape 1), sont autant de possibilités pour réitérer les étapes.

Cette découverte de connaissance doit donc être favorisée par une **compatibilité cognitive** entre l'outil d'analyse et les réflexions de l'archéologue. Comme nous l'avons vu au chapitre 2, les premiers critères de sélection de ces outils sont la rapidité des résultats et la facilité d'utilisation. Les outils d'analyse en ligne répondent parfaitement à ces deux critères. L'autre critère de sélection est la possibilité de réviser les données d'interprétation dans le but de comprendre le site archéologique. La section suivante présente une expérimentation du processus de révision en archéologie.

4.3- La révision de la structure multidimensionnelle : de la théorie à la pratique

Il apparait, donc, que tout le processus analytique relatif à l'interprétation des données issues d'un chantier de fouille archéologique en vue de la compréhension du site fouillé et la restitution de son aspect originel est un processus d'analyse qui exige des révisions constantes des données fondamentales. Un prototype a été construit et expérimenté afin de vérifier nos observations théoriques sur le sujet.

Les objectifs du prototype développés doivent permettre de vérifier simplement si les efforts pour réviser une donnée d'interprétation peuvent être compatibles avec les efforts requis pour un SOLAP afin d'en conserver la fluidité d'exploration interactive. Les outils de type SOLAP sont caractérisés par leur rapidité dans les réponses aux requêtes et leur

facilité d'utilisation (Caron, 1998). De plus, ces outils donnent accès à une variété de vues de n'importe quelles informations de façon rapide, concise et interactive (Cios *et al.*, 2005). C'est pourquoi, les objectifs du prototype devront se porter sur les mêmes critères qui caractérisent l'outil SOLAP. En effet, l'intérêt d'un tel prototype peut être remis en question s'il bloque le processus cognitif de l'utilisateur. C'est pourquoi, les objectifs vont se porter tant sur la **faisabilité** d'un tel outil (est-il possible de garder la simplicité et la facilité d'utilisation ?) que sur l'**efficacité** de celui-ci (est-il possible de garder la rapidité d'exécution ?).

4.3-1. Présentation du prototype

Un SOLAP 3D archéologique

Comme nous l'avons vu, la structure multidimensionnelle sera de type ROLAP sous la simulation de schéma en étoile avec la modélisation des dimensions en « parent-enfant ».

Nous nous inspirerons d'un premier travail réalisé par Rageul (Rageul, 2004). La structure MOLAP avait alors été utilisée. Nous avons gardé les mêmes dimensions (8) avec quelques légères modifications pour certaines (cf. Annexe A pour une description détaillée de chaque dimension) :

- la couleur de la terre de l'unité de fouille ;
- la granulométrie de la terre de l'unité de fouille ;
- la consistance de la terre de l'unité de fouille ;
- le numéro d'unités de fouille (contient la référence spatiale) ;
- la temporalité de l'unité de fouille ;
- le type de tesson ou type de céramique ;
- la temporalité de la céramique ;
- la proportion de la céramique dans l'unité de fouille en fonction de la temporalité et/ou de ses caractéristiques ;

D'autres dimensions, comme par exemple la relation spatiale topologique, la relation temporelle topologique et l'élaboration des constructions, aires d'activités et niveaux, ont été envisagées mais n'ont pas été implantées dans la structure multidimensionnelle sur

laquelle nous avons élaboré le prototype. Celles-ci sont présentées à titre indicatif en annexe B.

Les mesures utilisées pour notre prototype sont présentées dans le Tableau 4-4

Tableau 4-4 : Les mesures utilisées et leur fonction d'agrégation

Mesures utilisées	Fonction d'agrégation
Nombre de céramiques	Somme
Nombre d'unités de fouille	Somme
Altitude Supérieure de l'unité de fouille	Maximum
Altitude Inférieure de l'unité de fouille	Minimum
Épaisseur de l'unité de fouille	Moyenne

Présentation du prototype de révision des données d'interprétation

Le prototype a été conçu au travers d'un langage de programmation *Visual Basic* et le système de gestion de base de données *Microsoft Access* 2003. Ce prototype permet tout d'abord de reconstruire intégralement le cube ROLAP à partir des données entrantes. Cette reconstruction se compose de trois étapes :

- Uniformiser les tables, insérer les clés étrangères des dimensions dans les tables uniformisées et créer la table des faits (12 secondes)
- Créer la table des faits élargie pour générer un fichier texte lisible par le logiciel SAS (29secondes)
- Exécuter le fichier texte sous le logiciel SAS (environ 40 heures)

Le prototype permet aussi de modifier le cube ROLAP et les tables entrantes. Deux interfaces sont prévues à cet effet. (cf. Figure 4-12 et Figure 4-15). La première (cf. Figure 4-12) qui fait aussi « interface d'accueil », présente l'évolution des clés étrangères (tous les clics de souris offrent un affichage dans la seconde sauf le module D).

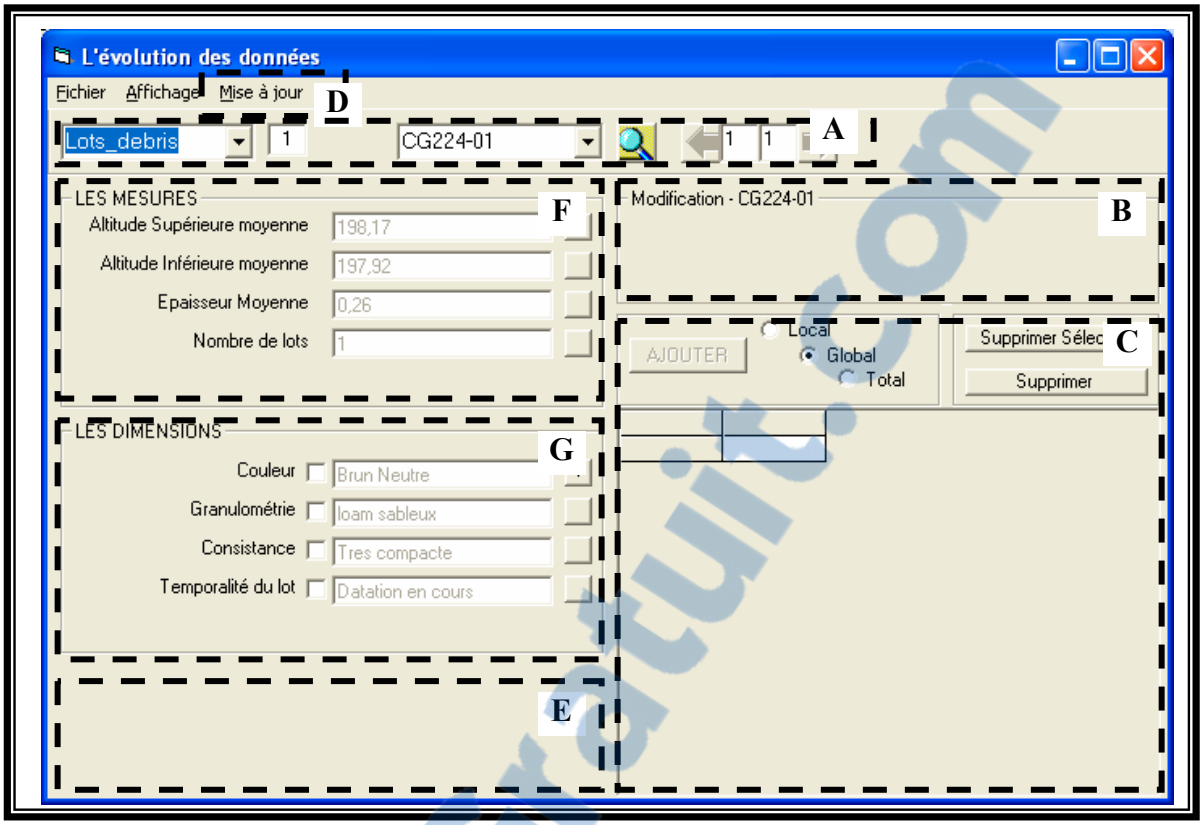


Figure 4-12 : Interface « Évolution des clés étrangères et mesures »

A) Permet à l'utilisateur de rechercher les éléments qu'il souhaite modifier et entraîne automatiquement les mesures (F) et les dimensions (G) associées (cf. Figure 4-13)

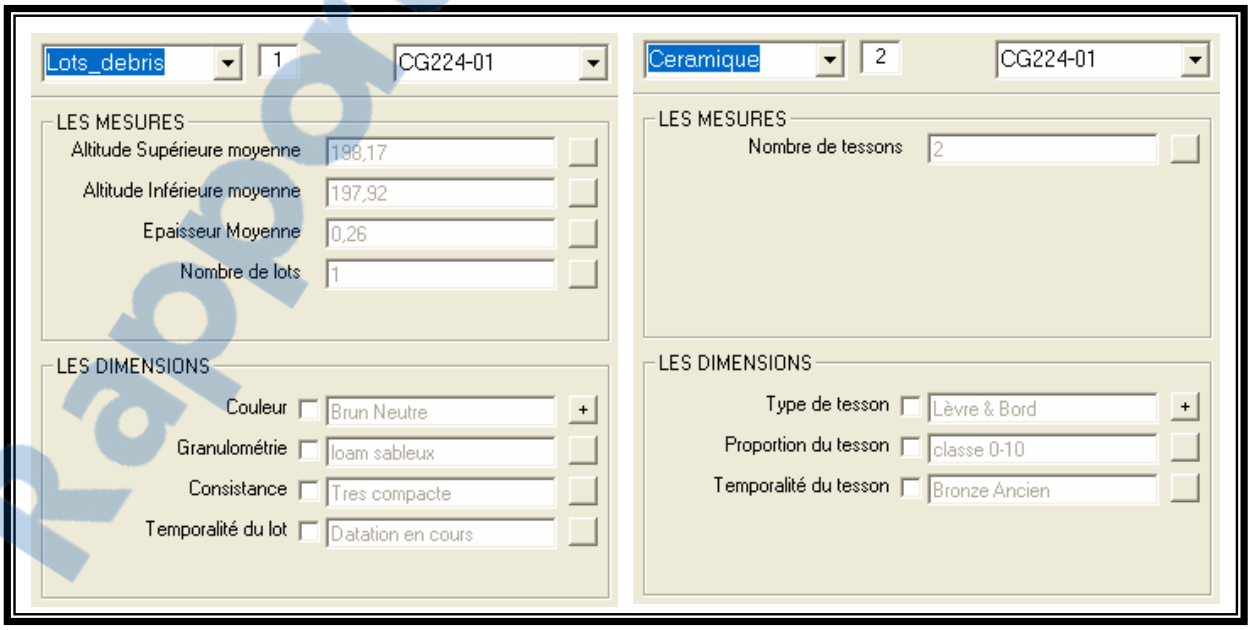


Figure 4-13 : Les mesures et les dimensions

Rapport-Gratuit.com

B) Affiche le type de modification : c'est dans cette partie du programme que l'on pourra effectuer les modifications souhaitées (cf. Figure 4-14) :

B.1) L'évolution des mesures

B.2) L'évolution des dimensions

B.2.a. S'il s'agit d'une clé étrangère

B.2.b. S'il s'agit de la hiérarchie de la dimension

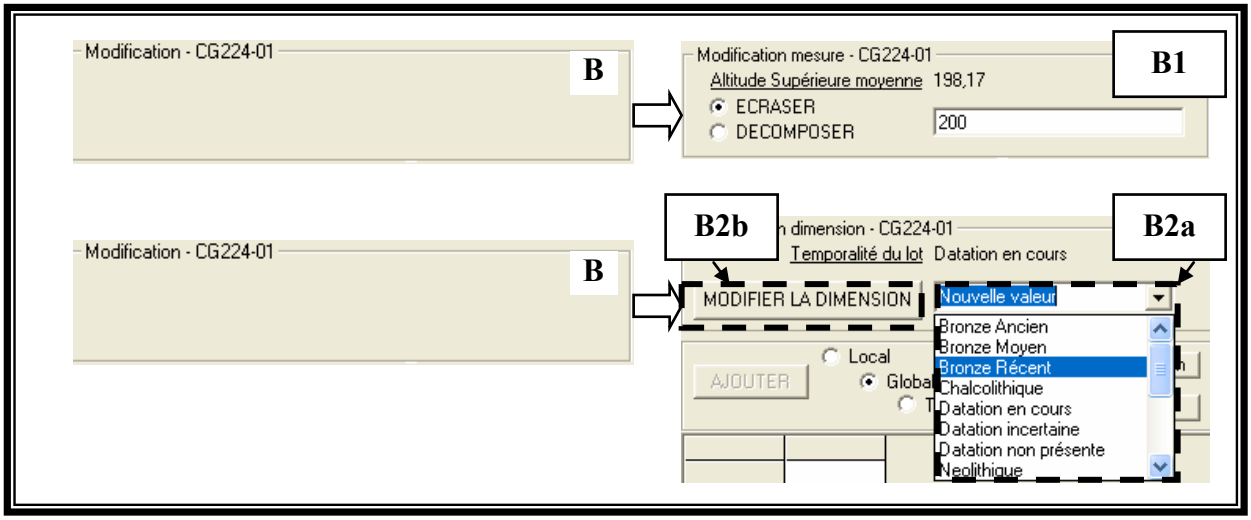


Figure 4-14 : Les différents types de modifications

C) Affichent la liste de toutes les modifications (cf. Tableau 4-5)

Tableau 4-5 : Exemple de liste de modifications

N°	Id_Pri	Typ	Nom	Avant	Apres
1	CG224-01	Mes	Altitude Supér	198,17	200
2	CG224-01	Dim	Temporalité d	Datation en co	Bronze Moyen
3		Rer	Couleur	Brun	Marron
4	Tenon	Chg	Type de tesso	Secondaire	Principal

N°1 : Pour l'UF CG224-01 : écrase la valeur de la mesure « altitude supérieure » actuelle (198,17) par la valeur « 200 » ;

N°2 : Pour l'UF CG224-01 : change la clé étrangère de la dimension temporalité de l'UF « datation en cours » par la clé étrangère « Bronze moyen » ;

N°3 : Renomme dans la dimension Couleur, la donnée interprétée« brun », par la donnée « marron » ;

N°4 : Dans la dimension « type de tesson », la donnée « tenon » change de père (passe de « secondaire » à « principal ») ;

D) Exécute les mises à jour. Les mises à jour sur toutes les tables se font dans la seconde (dimensions, table des faits, etc…). Faute de temps, nous n'avons pas pu tester la connexion avec la table finale ROLAP. Cependant, nous estimons que l'actualisation de la table ROLAP reste largement inférieure aux 40 heures que nécessite sa mise en place. En effet, des requêtes SQL (simples et complexes) ont été effectuées sur cette table de 7,5 millions d'enregistrements. Les requêtes offrent des résultats sous la minute, ce qui peut être plus intéressant que les 40 heures de reconstruction sous le logiciel SAS.

E) Affiche l'historique de la révision des données. En effet, le caractère itératif que nécessite le processus de révision de données peut impliquer aussi de revenir sur des révisions faites par l'analyste. Un rappel de l'historique peut simplifier cette démarche.

La deuxième interface (cf. Figure 4-15) permet de modifier la dimension. Cette interface est accessible via B2b de la Figure 4-14.

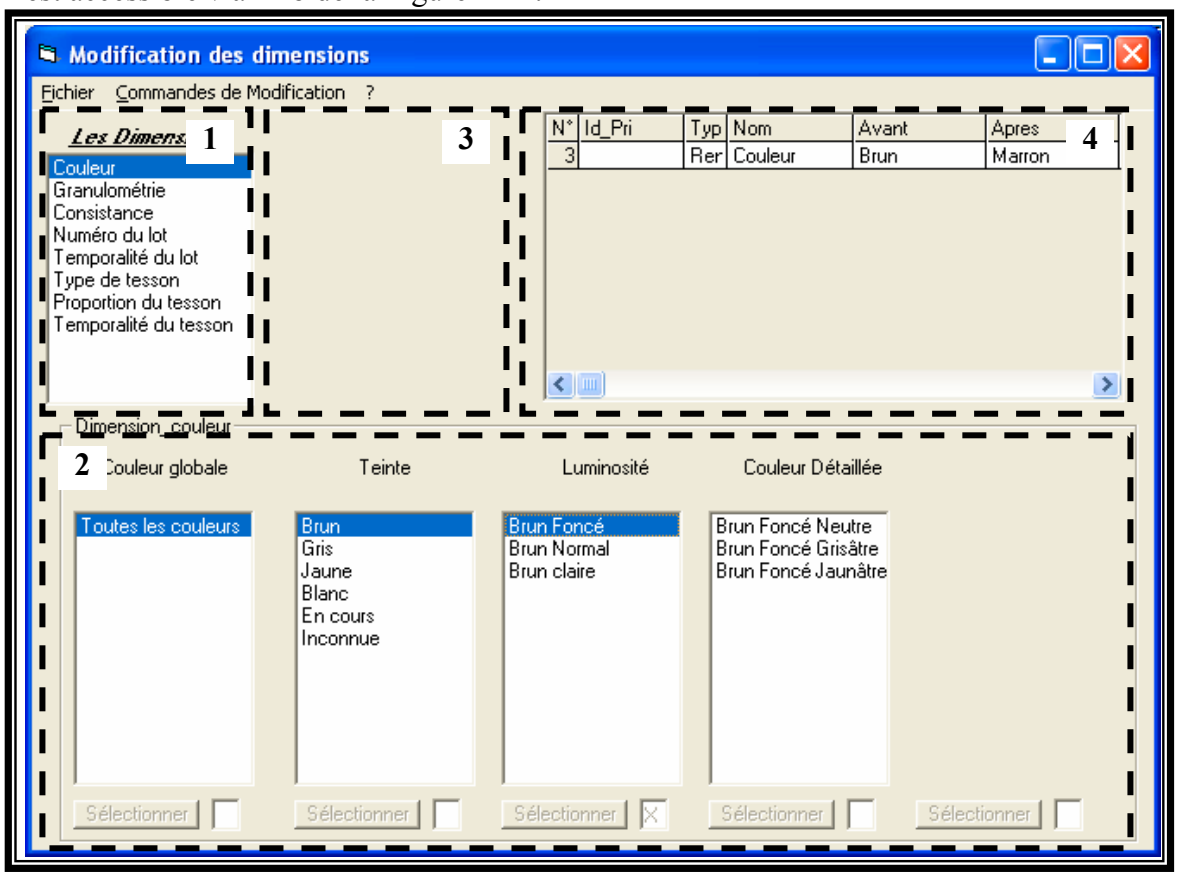


Figure 4-15 : Interface « Modification des dimensions »

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

1. Liste des dimensions
2. Navigation dans les dimensions
3. Type de modification : c'est dans cette partie du programme que l'on pourra effectuer les modifications souhaitées (cf. Figure 4-16)
4. Affiche la liste de toutes les modifications

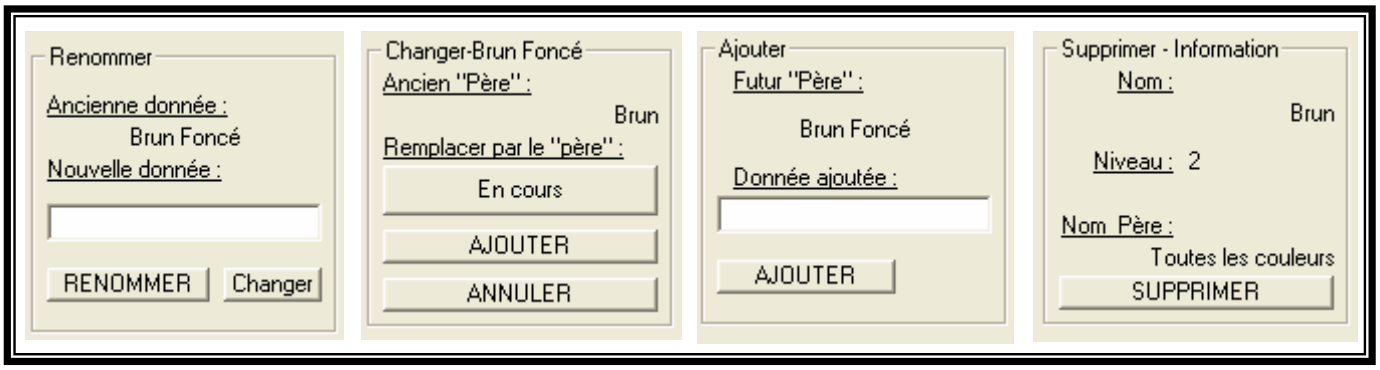


Figure 4-16 : Liste des modifications des dimensions

4.3-2. Analyse des résultats

Nous avons décidé d'élaborer un prototype afin de pouvoir explorer techniquement le processus de révision des données dans une structure multidimensionnelle. Cette maîtrise n'avait pas pour objectif principal d'élaborer un prototype fonctionnel sur tous les points de vue mais plutôt de tester et comprendre ce processus dans une structure multidimensionnelle. Après quatre mois de recherche et de programmation dans notre expérimentation, des conclusions sur l'évolution de données d'interprétation dans une structure multidimensionnelle peuvent en être tirées.

Le premier point concerne la faisabilité. Nous avons pu conserver ce qui fait la force d'une structure multidimensionnelle : l'intuition et la facilité d'utilisation (navigation par clic et non par requête SQL). Il est donc possible pour un usager, qu'il soit analyste ou débutant, de pouvoir effectuer tous les changements nécessaires avec une extrême facilité.

Le deuxième point concerne l'efficacité. Même si la connexion avec le ROLAP n'a pas pu être faite, l'ensemble des fonctionnalités du prototype offre des résultats sous la minute (de l'uniformité des tables à la reconstruction du fichier lisible sous SAS : 41 sec.) et dans la

seconde pour toutes les modifications des tables. De ce point de vue le prototype a pu montrer une performance évidente puisqu'il gagne 35 secondes à chaque modification au lieu des 41 secondes prévues. En effet, seulement 6 secondes sont nécessaires soit 3 secondes pour effectuer le changement - dépendant du type de modification - et 3 secondes pour reconstruire le fichier lisible sous SAS. La reconstruction totale des agrégations quant à elle se fait en 40 heures. Après un test sur des requêtes SQL simples (portant sur un champ) ou même complexe (portant sur une combinaison de champs), le temps de réponse résultant a toujours été sous la minute. Même si nous n'avons pas pu expérimenter complètement cet aspect du prototype, nous avons pu montrer où seraient les possibilités d'optimisation et d'évolution du prototype en tenant compte des différentes règles théoriques émises au chapitre 3.

4.4- Conclusion

Pour bien optimiser l'analyse de données issues d'une fouille archéologique, nous nous sommes attachés à connaître le type de données issues de cette fouille et à comprendre le processus d'élaboration de connaissance archéologique. Cette compréhension est essentielle pour pouvoir aider les archéologues dans l'analyse de leurs données. Bien que nous ayons pu la rationaliser un petit peu, force est de constater que le processus de compréhension d'un site archéologique est extrêmement complexe et itératif. Il demande souvent beaucoup d'efforts de connaissances et d'association entre diverses connaissances pour l'archéologue car, à ce jour, aucun système informatique n'a permis de clairement les aider dans l'analyse de leurs données. C'est d'ailleurs grâce à cette compréhension, que de nouveaux concepts géomatiques d'analyse de données, tels que vus au chapitre 3, ont pu être mis en évidence.

Un objectif, vis à vis de cette science humaine/sociale, était de comprendre les besoins d'analyse des archéologues, de les expliciter et de proposer une optimisation possible. Tout d'abord, nous nous sommes attachés à étudier les différents types d'analyse et d'appliquer certains opérateurs d'analyse à l'archéologie. De plus, l'étude des besoins en analyse des archéologues nous a permis d'étudier un concept nouveau pour une structure multidimensionnelle : la révision des données d'interprétation pendant la phase d'analyse.

Puisque très peu explorée à ce jour, la modification des données d'une structure multidimensionnelle pendant son utilisation est difficilement quantifiable du point de vue de l'effort (facilité et rapidité) que doivent fournir tant l'outil que l'utilisateur. Le prototype conçu et présenté dans ce chapitre, a permis de répondre en partie à cette quantification des efforts en étudiant tant la faisabilité (l'évolution des données doit se baser, pour l'utilisateur, sur des clics et non des requêtes SQL) que l'efficacité de notre outil (efficacité sur la mise en place : temps de construction du cube SOLAP et efficacité sur la révision : le temps de modification de la table devant être strictement inférieur au temps de reconstruction).

Chapitre 5 : Discussion et conclusion

L'objectif principal de cette recherche était de **proposer des mécanismes qui permettraient l'évolution des données 3D spatio-temporelles (ajout, suppression et modification) dans une structure multidimensionnelle** directement pendant la phase d'analyse de l'utilisateur. Pour amener des éléments de réponse à notre étude, trois objectifs spécifiques ont du être abordés :

- Faire état des réflexions actuelles sur les possibilités de combiner l'exploration 3D avec les solutions OLAP et/ou SOLAP afin de faire des recommandations pour le développement futur de ce genre de système.
- Élargir les concepts de mises à jour de données d'une structure multidimensionnelle
- Identifier les besoins des archéologues quant à l'analyse de leurs données, mettre en évidence leur processus analytique et faire des propositions pour mettre au point un système de découverte de connaissances mieux adapté à leur contexte.

5.1-Présentation des résultats et discussion

5.1-1. L'exploration 3D avec les solutions OLAP et/ou SOLAP

Les éléments du chapitre 2 ont pu mettre en évidence que les systèmes transactionnels sont encore assez limités quant à la gestion explicite de la **3e dimension**. Les systèmes analytiques ont été construits pour gérer des données 2D. De plus, notre revue de littérature spécifique à l'examen de travaux incorporant la 3e dimension dans les systèmes analytiques nous a révélé qu'il n'y avait que les travaux de Brisebois, relatifs au SOLAP 3D (Brisebois 2003). Cependant, d'après le prototype conçu par Brisebois (2003), le volet cartographique étant supporté par un SIG, la gestion du 3D n'était alors pas optimale.

Partant des travaux de Brisebois (2003) sur l'élaboration d'un outil SOLAP 3D, nous avons, tout d'abord, pu mettre en évidence (chapitre3) les **difficultés à mettre en pratique un tel outil**. Un des résultats de cette maîtrise a été de porter un diagnostic sur ces difficultés. Ainsi, la décision d'introduire ou non, la 3ème dimension est directement liée à la possibilité de surmonter deux barrières actuelles : la **barrière technologique**, pour appliquer les concepts théorique du SOLAP 3D et la **barrière conceptuelle** pour valider la

réelle nécessité d'un SOLAP 3D (pour analyser des données 3D, l'interrogation interactive des données descriptives est-elle utile aux outils 3D ? La visualisation 3D, à la différence d'une visualisation 2D judicieuse, est elle nécessaire pour une analyse de données 3D ?)

L'ampleur du projet géomatique « élaborer un SOLAP 3D opérationnel » a été un aspect qui a limité la recherche. Notre projet de recherche est né d'une volonté d'aider les archéologues à analyser leurs données. Nous pensions que l'exploration des données spatiales, surtout concernant des données de géométrie 3D, devait d'abord passer par une étude des attentes du futur utilisateur pour être compatible avec le processus analytique de celui-ci. Après avoir appliqué le SOLAP 3D tel que conçu par Brisebois, des conclusions avaient pu être tirées quant à une problématique possible : trouver et justifier de manière rigoureuse la meilleure solution ou combinaison de solutions permettant de développer un prototype intégrant le mieux possible les besoins des archéologues sur l'analyse et l'exploration de données spatiales. En d'autre termes, selon la définition de "qualité" de l'ISO (cf. "fitness for use"), il s'agissait de réaliser à partir de technologies à potentiel 3D (ex. SIG, CAD, SOLAP), le prototype de la meilleure qualité possible pour les besoins des archéologues interprétant les données de fouilles. Cette optimisation est toujours faisable et justifiable, mais nous n'avons pu la mettre en œuvre. En effet, afin de rester dans le cadre d'une maîtrise, nous n'avons effectué qu'un diagnostic sur les difficultés à mettre en place de ce genre de prototype.

5.1-2. L'élargissement des concepts de mises à jour de données d'une structure multidimensionnelle

Les éléments du chapitre 2 ont aussi mis en évidence que les systèmes basés sur une **approche analytique multidimensionnelle** sont optimisés pour une analyse dite multidimensionnelle où les requêtes même celles de types agrégatives sont simples et leurs réponses sont rapides. Cependant, afin de conserver l'intégrité de la base de données, les systèmes analytiques permettent aux utilisateurs d'interroger leur base de données pour faire ressortir l'information rapidement mais ne peuvent pas leur permettre de réécrire ou modifier leurs données.

L'un des résultats importants des travaux de cette maîtrise consiste à avoir proposé un nouveau traitement, celui de la **révision**, nécessaire pour une analyse optimale des données. Le processus de révision consiste à effectuer des **processus itératifs** de modifications volontaires dans le but d'aboutir à la **meilleure** interprétation possible. Dans un premier temps, nous avons étudié les données susceptibles d'évoluer pendant la phase d'analyse. Nous avons qualifié ces données d'interprétation pour les différencier des données d'observation qui sont relevées lors de la phase d'acquisition objective. L'évolution des données d'interprétation s'inscrit donc dans le processus de révision et d'acquisition subjective de la réalité. Afin de bien différencier la notion de révision des données avec la notion de mise à jour, nous avons clarifié la définition de la mise à jour comme correspondant à l'évolution des données d'observation. Elle fait suite à un changement, sur le terrain, de la réalité et non pendant la phase d'analyse.

Pour permettre de comprendre le processus de révision dans une structure multidimensionnelle, nous avons estimé (chapitre 3) que l'architecture OLAP devait être une **architecture ROLAP**, que la simulation multidimensionnelle pour les bases de données relationnelles doit être le **schéma en étoile** et que la structure d'une dimension doit être basée sur la **relation « parent-enfant »** caractéristique des dimensions à hiérarchie non-balancée. L'approche ROLAP ne nécessite pas un serveur intermédiaire. Cela lui permet de travailler directement avec la base de données relationnelle. Grâce à cela, nous pouvons mieux comprendre comment s'effectuent la construction et la reconstruction des agrégations d'une structure multidimensionnelle. Nous avons alors pu constater que la plupart des ajouts, suppressions et/ou modifications qui pouvaient être envisagés ne nécessitaient pas une reconstruction complète de la structure multidimensionnelle mais plutôt une reconstruction partielle demandant, en théorie, moins de temps.

La recherche expérimentale a permis l'élaboration d'un prototype (en langage Visual Basic et avec le logiciel de gestion de base de données Microsoft Access 2003). Cette recherche a d'abord été le fruit d'une envie d'aller étudier la faisabilité et l'efficacité de notre approche théorique concernant la révision des données d'interprétation pendant la phase d'analyse. Les résultats du prototype sont assez significatifs, même si celui-ci n'est pas entièrement

opérationnel. La mise en place d'une structure multidimensionnelle ROLAP nécessite une quarantaine d'heures alors que les révisions ne nécessitent que quelques minutes et peuvent être faites facilement en laissant à l'utilisateur le caractère intuitif qui faisait la force des outils d'analyse en ligne.

Le facteur temps a été un aspect limitant la recherche et le développement de notre prototype. En effet, faute de temps et parce que ce n'est pas le but principal d'une maîtrise, nous n'avons pas pu approfondir le prototype, ni sa connexion avec la base de données ROLAP, ni l'utilisation de système permettant une meilleure gestion de la base de données comme le propose Microsoft SQL Server. Cependant, nous avons estimé que l'élargissement des concepts de mises à jour tel que proposé dans cette maîtrise, avec la création d'un nouveau traitement, la révision, pouvait servir de bases fondamentales au développement d'un prototype plus générique que celui proposé.

5.1-3. L'application archéologique

L'analyse des besoins des archéologues (chapitre 4) nous a permis d'identifier trois critères retenus pour faire la sélection de l'approche d'analyse: la rapidité et la facilité d'utilisation, la possibilité de faire évoluer les données dans le système et la visualisation tridimensionnelle.

Nous avons profité de la recherche expérimentale pour aider les archéologues à optimiser l'analyse de leurs données. Nous avons adapté certains opérateurs d'analyse à l'archéologie en mettant l'emphase sur les analyses spatiales et temporelles. Finalement, nous avons essayé de rationaliser en partie leur démarche de compréhension d'un site archéologique, cependant celui-ci est extrêmement complexe et itératif. En effet, il est apparu que tout le processus analytique relatif à l'interprétation des données issues d'un chantier de fouille archéologique est un processus d'analyse qui exige des révisions constantes des données fondamentales. A ce titre, cette exigence sur l'analyse des données archéologiques nous a permis de légitimer notre démarche d'élargissement des concepts de mises à jour de données d'une structure multidimensionnelle.

Malheureusement, l'ampleur du projet archéologique « la géomatisation d'une fouille archéologique » a été un facteur limitant la recherche sur ce sujet. Ce projet est caractérisé

par plusieurs projets qui sont pour la plupart toujours en cours et dont la phase d'analyse dépend notamment la phase d'acquisition des données.

5.2- Développement et recherches futures

Les pistes de recherche que ce projet peut amener dépendent justement des limites que nous avons eues :

L'archéologie offre des domaines de recherche très intéressants. Même si les technologies actuelles ne permettent pas de pleinement manipuler les données 3D, l'archéologie est un domaine d'application complet et riche nécessaire à la géomatique. De plus, les avancées de la géomatique dans l'exploitation des données tridimensionnelles enrichiront pleinement l'archéologie.

Pour la géomatique, les difficultés à appliquer les concepts théoriques du SOLAP 3D sont autant de pistes de recherche possible : le traitement de toutes les primitives 3D, l'analyse spatiale 3D (métrique et surtout topologique), la manipulation interactive 3D des données tridimensionnelles… De plus, les architectures SOLAP 3D proposées sont pour chacune d'entre elles, une manière de faire avancer la recherche d'architecture idéale, comme l'avait été le SOLAP 2D actuel. Ainsi, des recherches sur l'influence de l'interrogation des données descriptives sur des outils 3D ou bien l'influence de la visualisation 3D pour l'analyse de données 3D peuvent être des pistes envisageables

D'un point de vue conceptuel, le processus de révision se produit sur deux niveaux : la structure multidimensionnelle et les données elles-mêmes. Pour notre recherche, nous avons aussi décidé de considérer la structure multidimensionnelle fixe, il serait intéressant d'étudier ce processus avec une structure révisable. Concernant l'évolution des données pendant la phase d'analyse, il pourrait être intéressant d'envisager le développement d'un tel outil opérationnel basé sur les concepts que nous avons établis dans cette présente recherche. Finalement, il pourrait être utile d'étudier notre approche concernant le processus de révision pendant une phase d'analyse dans d'autres domaines d'application où la recherche de nouvelles connaissances est essentielle (ex. géologie, histoire, géographie, océanographie, physique, gestion des risques environnementaux …).

Rapport-Gratuit.com

Bibliographie

Aamodt, A. et Nygard, M.
1995, «Different Roles and Mutual Dependencies of Data, Information and Knowledge - An AI Perspective on their integration », dans *Data & Knowledge Engineering 16*. Elsevier, North Holland, pp. 191-222.

Allen, J. F.
1984, « Towards a general theory of actions and time », *Artificial Intelligence*, v. 23, pp. 123-154.

Apel, M.,
2004, « A 3D Geoscience Information System Framework », *Thèse de doctorat*, Instityt National Polytechnique de Lorraine, pp. 93.

Aronoff, S.,
1989, « Geographic Information Systems: a management perspective », WDK Publications (eds): Ottawa, Canada

Aufore-Portier, M.-A.
1995, «Human-Computer Interaction Aspects of visual-Interactives Locational Analysis », dans *Cognitive Aspects of Human-Computer Interaction for Geographic Information Systems*, Springer (éds). pp. 163-177.

Barceló, J. A.
2002, « Archaeological Thinking: Between Space and Time », *Archeologia e Calcolatori*, v. 13, pp. 237-256.

Barceló, J. A.
2005, « Multidimensional spatial analysis in Archaeology », papier présenté au *GIS Symposium "Reading the Historical Spatial Information in the World"-studies for Human Cultures and Civilizations based on Geographic Information System*, à Kyoto, Japon, 7 février 2005.

Barceló, J. A., De Castro, O., Traver, D. et Vicente, O.
2003, «A 3D Model of an Archaeological Excaviation », dans *The Digital Heritage of Archaeology. Computer Applications and Quantitative methods in Archaeology*, M. Doerr et A. Sarris (éds).

Bédard, K.,
2006, « La construction de modèles géologiques 3D à l'ère de la normalisation », *mémoire de maîtrise*, Université Laval, pp. 113.

Bédard, Y.

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

2003a, « SIG et Analyse Spatiale, section Notions de base en SIG », *note de cours*, Cours à l'Université Laval, filière Géomatique.

Bédard, Y.
2003b, « SIG et Analyse Spatiale, section opérateurs spatiaux dans les systèmes de gestion de données spatiales », *note de cours*, Cours à l'université Laval, filière Géomatique.

Bédard, Y.
2005, « Base de données, Section entrepôt de données spatiales, OLAP et SOLAP », *note de cours*, Cours à l'Université Laval, filière géomatique.

Bédard, Y., Pouliot, J., Normand, P. et Frenette, P.,
2002, « Optimisation de la modélisation et du traitement graphique de données 3D urbaines », *rapport de recherche*, Université Laval, Québec, pp.

Bellinger, G., Castro, D. et Mills, A.
1997, « Data, Information, Knowledge and Wisdom »,source: *http://www.outsights.com/systems/dikw/dikw.htm*, consulté le:

Bentley
2002, « Documentation de Microstation SE », *Documentation en ligne*.

Bertin, J.,
1992, « La graphique et le traitement graphique de l'information », frammarion (eds): Paris, pp. 272

Bertin, J.,
2005, « Sémiologie graphique: les diagrammes - les réseaux - les cartes », Les réimpressions (eds), pp. 454

Bertoline, G. R. et Wiebe, E. N.,
2002, « Technical Graphics Communication, 3rd Edition », McGraw-Hill Professional (eds), pp. 1312

Billen, R. et Zlatanova, S.
2003, « 3D Spatial Relationships model: a useful concept for 3D cadastre? », *Computers, Environment and Urban Systems*, v. 27, pp. 411-425.

Bourgon, J. F., Brodeur, J., Lemieux, F., Marchand, P. et Rivest, S.,
1999, « Fondements théoriques et pratiques des serveurs universels pour l'exploitation multidimensionnelle des très grandes bases de données », Québec, pp. 69.

Brisebois, A.,
2003, « Analyse du potentiel d'extension du concept SOLAP pour l'exploration des données spatiales tridimensionnelles », *mémoire de maîtrise*, Université Laval, pp. 143.

Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

Caron, P.-Y.,
1998, « Etude du potentiel de OLAP pour supporter l'analyse spatio-temporelle », *mémoire de maîtrise*, Université Laval, pp. 109.

Cattani, M., Fiorini, A. et Rondeli, B.
2004, «Computer applications for a reconstruction of archaeological stratigraphy as a predictive model in urban and territorial contexts, », dans *Enter the Past. the e-way into the four Dimensions of Cultural Heritage.CAA 03 Computer Applications and Quantitative Methods in Archaeology*, Börner W. Ausserer K.F., Goriany M. & Karlhuber-Vöckl L (éds). pp. 299-303.

Champoux, P.,
1991, « Etude sur les fonctions danalyse spatiale à utiliser dnas un SIRS appliqué à l'exploration minière », *mémoire de maîtrise*, Université Laval, pp.

Choo, C. W., Detlor, B. et Thumbull, D.,
2000, « Web zork: Information Seeking and Knowledge Work on the World Wide Web », Kluzer Academic Publishers (eds): Dordrecht

Chountas, P., Petrounias, I., Vasilakis, C., Tseng, A., El-Darzi, E., Atanassov, K. T. et Kodogiannis, V.
2004, «On uncertainty and Data-Warhouse Design », dans *Advances In Information Systems*, Springer (éds). pp. 4-12.

Cios, K. J. et Kurgan, L. A.
2005, «Trends in data mining and knowledge », dans *Advanced Techniques in Knowledge Discovery and Data Mining*, Springer (éds). pp. 1-26.

Claramunt, C. et Jiang, B.
2000, «A representation of relationships in temporal spaces », dans *Gis and Geocomputation*, Taylor & Francis (éds). pp. 41-54.

Clementini, E., Di Felice, P. et OOsterom, P. V.
1993, « A small set of formal topological relationships suitable for end-user interaction », *International Symposium on large spatial databases*,Singapore, pp. 277-295.

Davenport, T. H.,
1997, « Information Ecology: Mastering the Information and Knowledge Environnement », Oxford University Press (eds)

Day, W., Cosmas, J., Ryan, N. et Vereennoghe, T.,
2004, « Integrated 3D and Harris Matrix Data Entry, Query and Visualisation Tools ».

De La Losa, A.,
2000, « Modélisation de la troisième dimension dans les bases de données géographiques », *thèse de doctorat*, Université de Marne-La-Vallée, pp. 174.

Diansheng, G.
2003, « a geographic visual query composer for accessing federal databases », *Proceedings of National conference for digital research*, pp. 397-400.

Digital Archaeology Group
2006, « Digital Archaeology »,source: *http://online-archaeology.com/HouseReconstruction.htm*, consulté le: 15 Décembre 2006

Egenhoffer, M. J. et Al-Taha, K.
1992, «Reasoning about Gradual Changes of Topological Relationships », dans *Theories and Models of Spatio-Temporal Reasoning in Geographic Space. Lecture Notes In Computer*, I. Campari and U. Formentini A. Franck (éds), v. 639. Berlin, pp. 196-219.

Egenhoffer, M. J., Clementini, E. et Di Felice, P.
1994, « Topological relations between regions with holes », *International Journal of Geographical Information Systems*, v. 8(2), pp. 129-144.

Egenhoffer, M. J. et Herring, J. R.
1991, « Categorizing binary topological relationships between regions, lines and points in geographic databases », *Department of Surveying Engineering*, Rapport de recherche,University of Maine, Orono,Me.

Egenhoffer, M. J. et Herring, J. R.
1993, «Querying a geographical information system », dans *Human Factors in geographical Information System*, D. Medyckyi-Scott et H.M. Hearnshaw (éds). Belhaven Press, London (UK), pp. 124-135.

ESRI
2006, « GIS Dictionary »,source: *http://support.esri.com/*, consulté le: 7 mai 2006

Favetta, F. et Aufore-Portier, M.-A.
2000, «About Ambiguities in Visual GIS Query Languages: A Taxonomy and Solutions », dans *Advances in Visual Information Systems: 4th International Conference, VISUAL 2000*, Robert Laurini (éds), v. 1929. Lyon, France, pp. 154-165.

Foley, J. D.,
1995, « Computer graphics », Addison-Wesley Professional (eds), pp. 1200

Frawley, W., Piatesky-Shapiro, G. et Matheus, C.
1992, « Knowledge Discovery in Databases: An overview », *AI Magazine*,Fall, pp. 213-228.

Gagnon, P.,
1993, « Concepts fondamentaux de la gestion du temps dans les SGDL », *mémoire de maîtrise*, Université Laval, pp. 160.

Gahegan, M., Wachowicz, M., Harrower, M. et Rhyne, T.-M.
2001, « The integration of Geographic Visualisation with Knowledge Discovery In Databases and Geocomputation », *Cartography and Geographic Information Science*, v. 28(1), pp. 29-44.

Gasparini, C.,
2005, « CAD Tutor 3D », Gasparini Editore (eds), pp. 202

Ginguay, M. et Lauret, A.,
1993, « Dictionnaire d'informatique 5ème edition », Masson (eds), pp. 371

Golay, F. et Caloz, R.
2001, « Systèmes d'information à référence Spatiale1 - Section Traitements des données spatiales », *note de cours*, Ecole Polytechnique Fédérale de Lausanne.

Goodchild, M. F.
1987, « A spatial Analytical perspective on geographical information systems », *International Journal of Geographical Information Systems*, v. 1, pp. 327-334.

Green, D., Cosmas, J., Itegaki, T., Waelkens, M., Degeest, R. et Grabczewski, E.
2001, « A real time 3D stratigraphic visual simulation system for archaeological analysis and hypothesis testing », *In Proceedings ACM Virtual Reality, Archaeology and Cultural Heritage (VAST 2001)*, (Novembre 2001).

Guimond, L.-E.,
2005, « Conception d'un environnement de Découverte des besoins le développement de solutions SOLAP », *Mémoire de maîtrise*, Université Laval, pp. 124.

Haining, R.,
2003, « Spatial Data Analysis », Cambridge University Press (eds), pp. 452

Han, J. et Kamber, M.,
2000, « Data Mining - concepts and techniques », Morgan Kaufmann (eds), pp. 550

Jian-feng, Y., Zahn-huai, L. et Chen-juan, G.
2005, «Temporal semantic abstraction based on layered multimedia data model », dans *Information Systems Development - Advances in Theory, Practice, and Education*, Springer (éds). pp. 353-364.

Juillard, M.
2005, « Cours de Statistiques descriptives », *note de cours*, Département d'économie, Université Paris 8, (12 Octobre 2005).

Kahn, A.,
2003, « Data Warehousing 101 - Concepts and Implemantation », iUniverse (eds), pp. 136

Kemp, K. K.
2006, «Spatial Databases: Sources and Issues », dans *Environmental Modeling With Gis*, Oxford University Press US (éds). pp. 363-371.

Kennedy, H.,
2004, « Data In Three Dimensions », Thomson Delmar Learning (eds), pp. 209

Korte, G. B.,
2001, « The GIS Book », Thomson Delmar Learning (eds), pp. 400

Kovalerchuk, B. et Schwing, J.,
2005, « Visual and Spatial Analysis - Advances in Data Mining, Reasoning, and Problem Solving », Springer (eds), pp. 576

Lachance, B.,
2005, « Développement d'une structure topologique de données 3D pour l'analyse de modèles géologiques », *Mémoire de maîtrise*, Université Laval, pp. 115.

Lachance, B., Fortin, M. et Pouliot, J.
2006, «La fouille archéologique: un processus à repenser », dans *La mesure du passé: contributions à la recherche en archéométrie (2000-2006)*, Jacques Chabot et Marcel Moussette Allison Bain (éds). British Archaeological Reports. (collectif de 22 articles) *à paraitre*.

Lachey, t.,
2005, « Applied Microsoft Analysis Services 2005 », Prologika Press (eds), pp. 696

Lambert, M.,
2005, « Développement d'une méthode de peuplement en temps quasi-réel d'un outil SOLAP à partir de mesures GPS: application aux activités sportives en plein air », *Mémoire de maîtrise*, Université Laval, pp. 126.

Langran, G.,
1992, « Time in geographic Information Systems », Taylor&Francis (eds), pp. 180

Larrivée, S., Bédard, Y. et Pouliot, J.
2005, « Fondement de la modélisation conceptuelle des bases de données géospatiales 3D », *Revue internationale de géomatique*,Paris.

Larrivée, S., Bédard, Y. et Pouliot, J.
2006, « Modélisation conceptuelle des bases de données géospatiales pour des applications 3D », *Revue internationale de géomatique: Information géographique tridimensionnelle*,Paris, v. 16, (1), pp. 9-27.

Lattuada, R.
2005, «Three-Dimensional Representations and Data Structures in GIS and AEC », dans *Large-Scale 3D Data Integration*, CRC Press (éds). pp. 57-86.

Longley, P. A., Goodchild, M. F., McGuire, D. J. et Rhind, D. W.,
2005, « Geographic Information Systems and Science », John Wiley and Sons (eds), pp. 536

Losier, L.-M.,
2005, « Modélisation géométrique 3D de données extraites d'un chantier de fouilles archéologiques », *Essai de maîtrise*, Université Laval, pp.

Losier, L.-M., Pouliot, J. et Fortin, M.
2007, « geometrical modeling of archaeological excavations at Tell 'Acharneh (Syria) », *Journal of Archaeological Science*,Elsevier, v. 34, (2), pp. 272-288.

Malinverni, E. S., Gagliardini, G. et Fangi, G.
2002, « Virtualisation of an archaeological site », présenté pour *Commission V symposium Close-range Imaging, Long Range Vision*, à Corfu, Grêce, p. 407-412

Marchand, J., Daniel, S., Pouliot, J., Gauvin, R. et Fortin, M.
2006, « Potentiel d'un scanneur laser dans le cadre de fouilles archéologiques », présenté pour *Conference Géomatique 2006 de l'association canadienne des sciences Géomatiques*, à Montréal, 25-26 octobre 2006,

Mena, J.,
2002, « Investigative Data Mining for Security and Criminal Detection », Elsevier (eds), pp. 272

Mngumi, E. A. et Ruther, H.
2004, « Solid modeling for heritage documentation », papier présenté au *Worshop - Archaelogical survey - FIG (Fédération Internationale des Géomètres)- working week 2004*, à Athènes, grèce, 22-27 Mai 2004.

Nigro, J. D., Ungar, P. S., de Ruiter, D. J. et Berger, L. R.
2003, « Developing a Geographic Information System (GIS) for mapping and analysing fossil deposits at swartkrans, Gauteng Province, South Africa », *Journal of Archaeological Science*, v. 30, pp. 317-324.

Normand, P., Bédard, Y., Larrivée, S. et Proulx, M.-J.
2003, « SIG et Analyse Spatial, section: Modélisation des contraintes d'intégrité spatiale », *note de cours*, Cours à l'Université Laval, filière Géomatique.

Nyerges, T. L.
2006, «Understanding the Scope of GIS: Its relationship to Environmental Modeling », dans *Environemental Modeling with GIS*, Oxford Universty Press US (éds). pp. 75-93.

OQLF,
2006, « Le grand dictionnaire terminologique, Office Québecois de la langue française ».

Pouliot, J.
2003, « Intégration des données spatiales - Concepts et pratique », *note de cours*, Cours à l'Université Laval, filière géomatique.

Pouliot, J.
2005, « Travaux de synthèse en modélisation 3D », *note de cours*, Cours à l'Université Laval, filière géomatique.

Pouliot, J., Bédard, Y., Caron, C., Larrivée, S., Nadeau, M., Levasseur, D., Métévier, R. et Monaghon, D.
2004, « M@gic: expérimentation d'une approche incrémentielle de gestion et d'échange de mises à jour de données géospatiales », *Geomatica*, v. 58 n°2, pp. 119-132.

Pouliot, J., Lachance, B. et Kirkwood, D.
2006, « L'importance de la modélisation géométrique pour les SIG-3D: Exemple du développement d'un modèle topologique 3D pour une application géologique », *Revue internationale de géomatique*, v. 16/1.

Proulx, M.-J.,
1995, « Développement d'un nouveau langage d'interrogation de bases de données spatio-temporelles », *mémoire de maîtrise*, Université Laval, pp.

Pullar, D. et Egenhoffer, M. J.
1988, « Towards formal definition of topological relations among spatial objects », *Third International Symposium on Spatial Data Handling*,Australia, pp. 225-242.

Rageul, N.,
2004, « Développement d'une application d'exploration de données géospatiales comme support à la fouille archéologique », *mémoire pour l'obtention du diplome d'ingenieur*, INSA Strasbourg, Strasbourg, pp. 59.

Ramos, F.,
2003, « Modélisation et validation d'un système d'information géographique 3D opérationnel », *Thèse de doctorat*, Université de Marne La Vallée, pp. 174.

Rivest, S.,
2000, « Investigation des modes d'intégration physique entre un serveur de base de données multidimensionnelle et un SIG », *essai de Maîtrise*, Université Laval, pp. 57.

Sanaa, M.
2005, « Biostatistique section - Statistique Descriptive version 2005 »,source: *http://www.vera.fr/biostatistique/STATDES.ht*, consulté le: 20 mai 2006

Saxena, A. et Sahay, B.,
2005, « Computer Aided Engineering Design », Springer (eds), pp. 394

Skibniewski, M. J. et Kunigahalli, R.
1997, «Chapitre 17: Automation in Concrete Construction », dans *Concrete Construction Engineering Handbook*, CRC Press (éds). pp. 18.

Swanson, J.
1996, « The Three Dimensional Visualization & Analysis of géographic data »,source: *http://maps.unomaha.edu/Peterson/gis/Final_Projects/1996/Swanson/GIS_Paper.html*, consulté le: 20 mai 2006

Swiaczny, F. et Ott, T.,
2001, « Time-Integrative Geographic Information Systems », Springer (eds), pp. 234

Szalapaj, P.,
2001, « CAD Principles for Architectural Design », Architectural Press (eds), pp. 242

Tet Kuan, C. et Abdul Rahman, A.
2005, «3D Buffering: A visualization tool for Disaster Management », dans *Geo-Information for Disaster Management*, Springer (éds). pp. 841-867.

Thagard, P. et Shelley, C.
1997, «Abductive reasoning: Logic, visual thinking, and coherence », dans *Logic and Scientific methods*, M.-L. Dalla Chiara et al (éds). Dordrecht: Kluwer, pp. 413-427.

Thalmann, D.
2006, « Computers Graphics section "la modélisation géométrique" », *note de cours*,Ecole Polytechnique Fédérale de Lausanne, pp. 48.

Vaisman, A. A., Mendelzon, A. O., Ruaro, W. et Cymerman, S. G.
2002, «Supporting Dimension Updates in an OLAP Server », dans *Advanced Information Systems Engineering*, Springer (éds). pp. 67-82.

Van der Spek, R. et Spijkervet, A.,
1997, « Knowledge Management: Dealing Intelligently with Knowledge », CIBIT (eds): Utrecht

Wheatley, D. et Gilling, M.,
2002, « Spatial Technology and Archaeology. The Archaeological Applications of GIS », (eds), v. London and New York, Taylor & Francis, pp. 269

White, C.
2002, « Intelligent Business Strategies: Real-Time Data Warehousing Heats Up »,DMReview, Août 2002, source: *http://www.dmreview.com/article_sub.cfm?articleId=5570*, consulté le: 26 juin 2006

Wilsey, P. A.

1999, «Modeling, analysis, and simulation of computer and telecommunication systems », dans *Encyclopedia of Computer Science and Technology*, Marcel Dekker (éds). pp. 147-160.

Zlatanova, S.
2000, « On 3D topological relationships », présenté pour *Proceedings of the 11th International workshop on Database and Expert System applications*, à Greenwich, London, p. 913-919

Zlatanova, S., Rahman, A. et Pilouk, M.
2002, « 3D GIS: current status and perspectives », présenté pour *Proceedings of ISPRS*, à Ottawa, Canada, 8-12 juillet 2002, p. 8

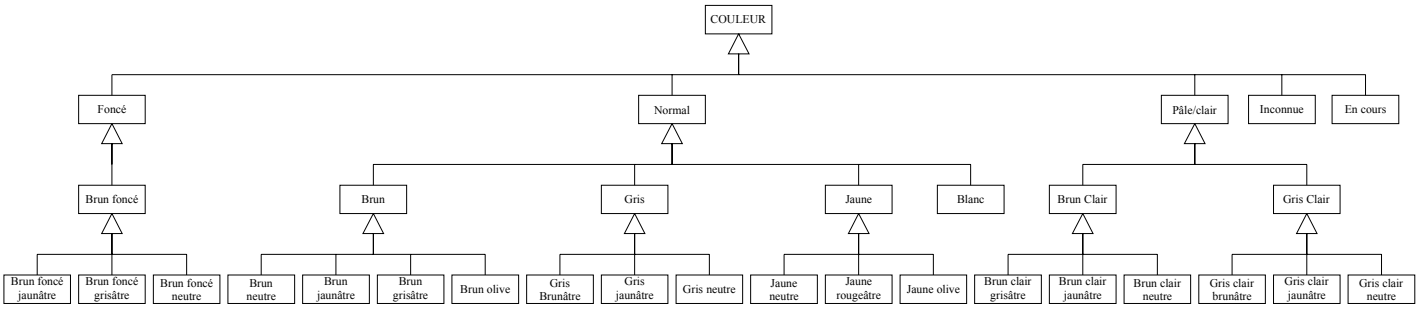
Zlatanova, S., Rahman, A. et Shi, W.
2004, « Topological Models and Frameworks for 3D spatial Objects », *Computer&geosciences*, v. 30, n°4, May.

Rapport-Gratuit.com

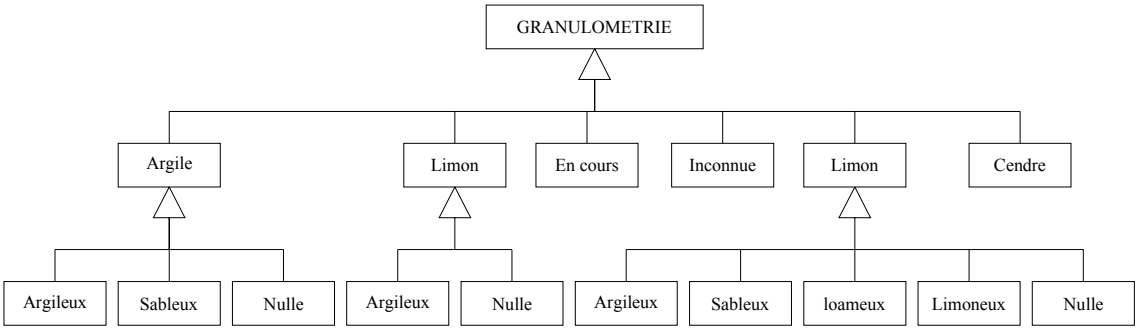
Annexe A Dimensions d'analyses utilisées en archéologie

Cette annexe présente les hiérarchies (membres et niveaux) des dimensions d'analyses qui composent la BDMRS utilisé par le prototype d'évolution de données.

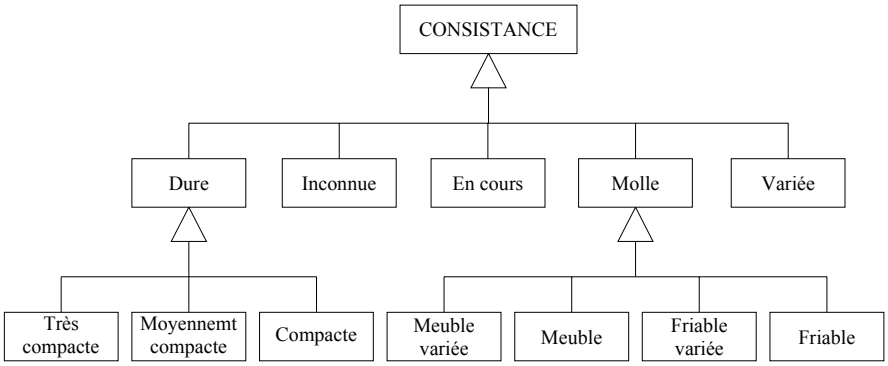
La dimension « COULEUR D'UNE UNITE DE FOUILLE»



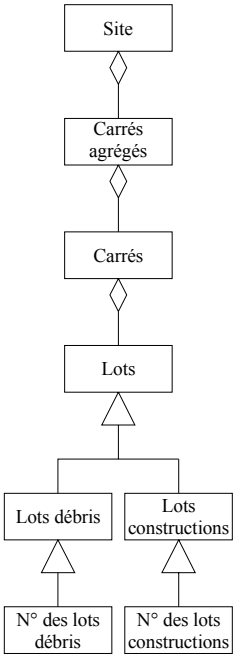
La dimension « GRANULOMETRIE D'UNE UNITE DE FOUILLE »



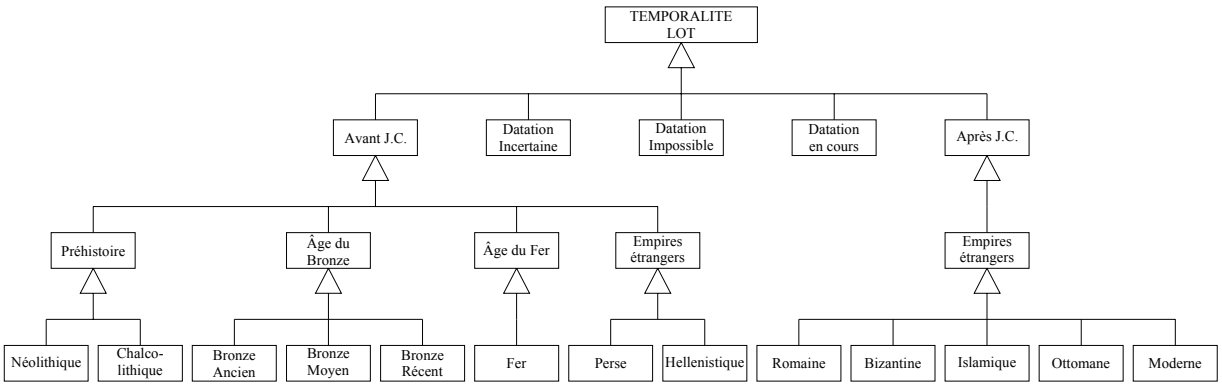
La dimension « CONSISTANCE D'UNE UNITE DE FOUILLE»



La dimension « NUMERO DES UNITES DE FOUILLE»

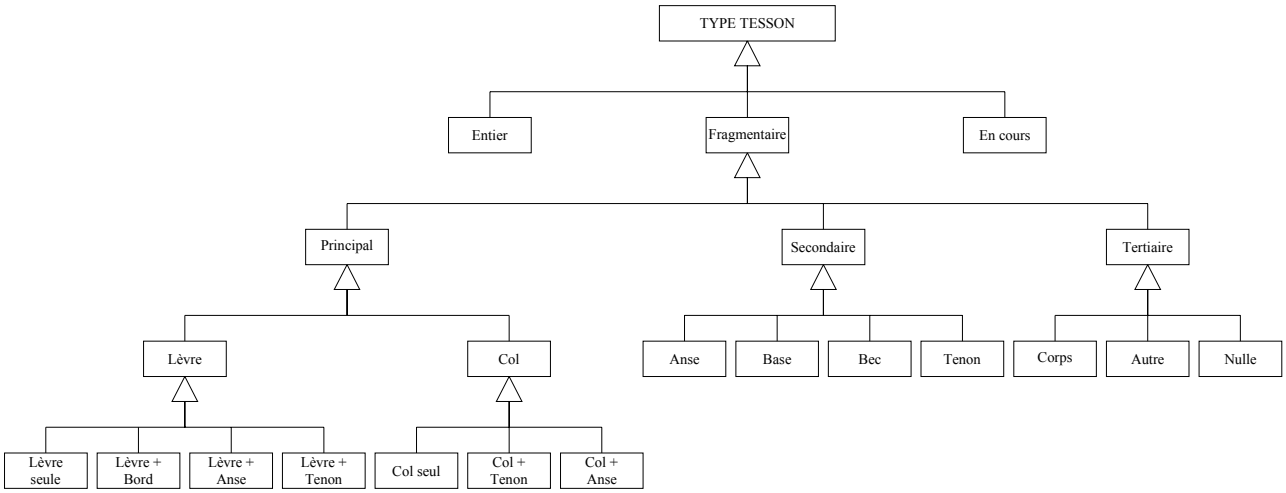


La dimension « TEMPORALITE D'UNE UNITE DE FOUILLE»

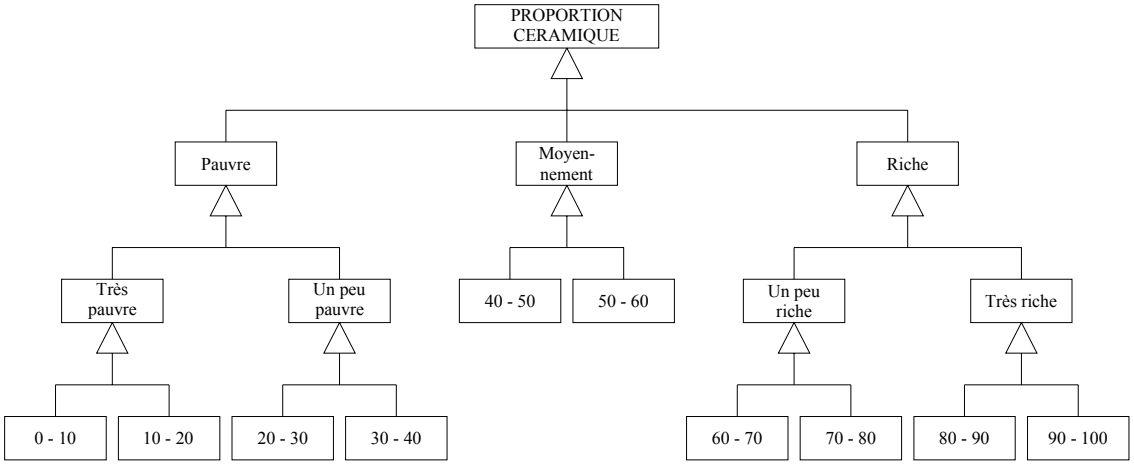


Rapport-gratuit.com
LE NUMERO 1 MONDIAL DU MÉMOIRES

La dimension « TYPE DE TESSONS»



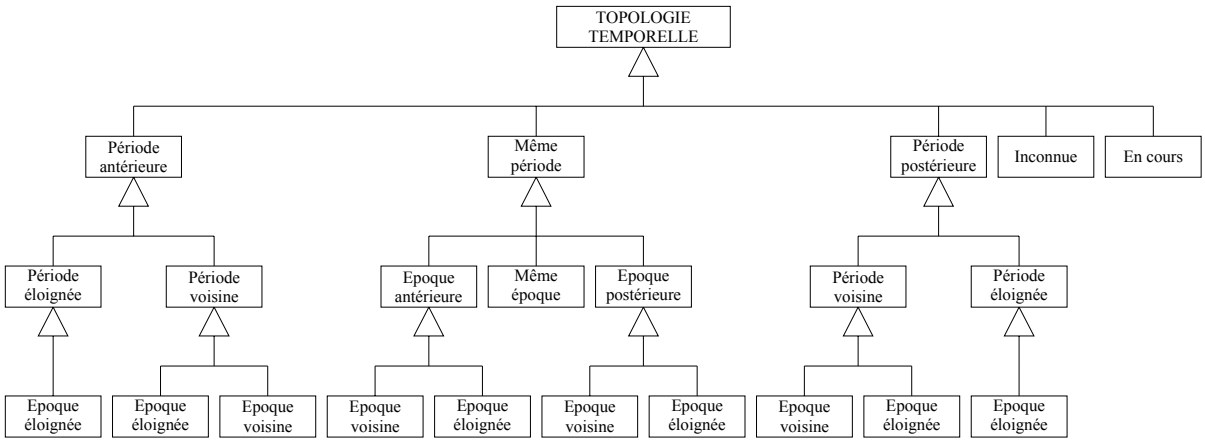
La dimension « PROPORTION CERAMIQUE »



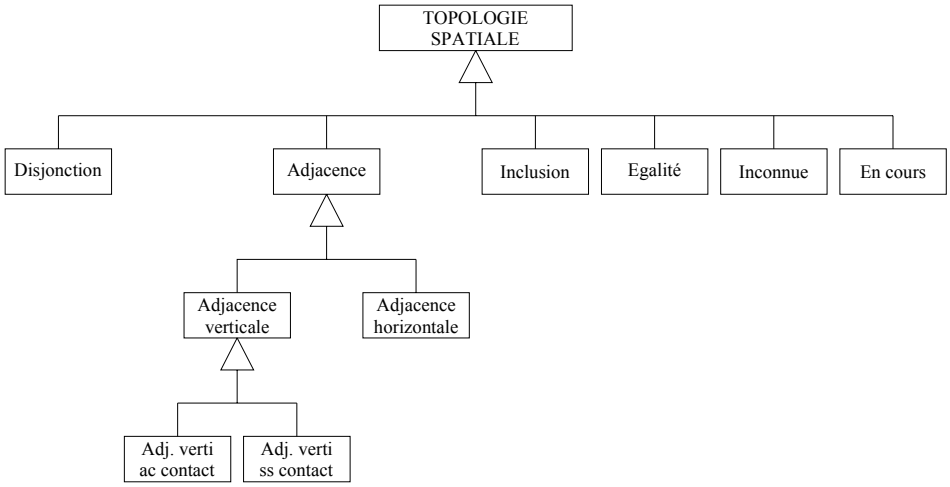
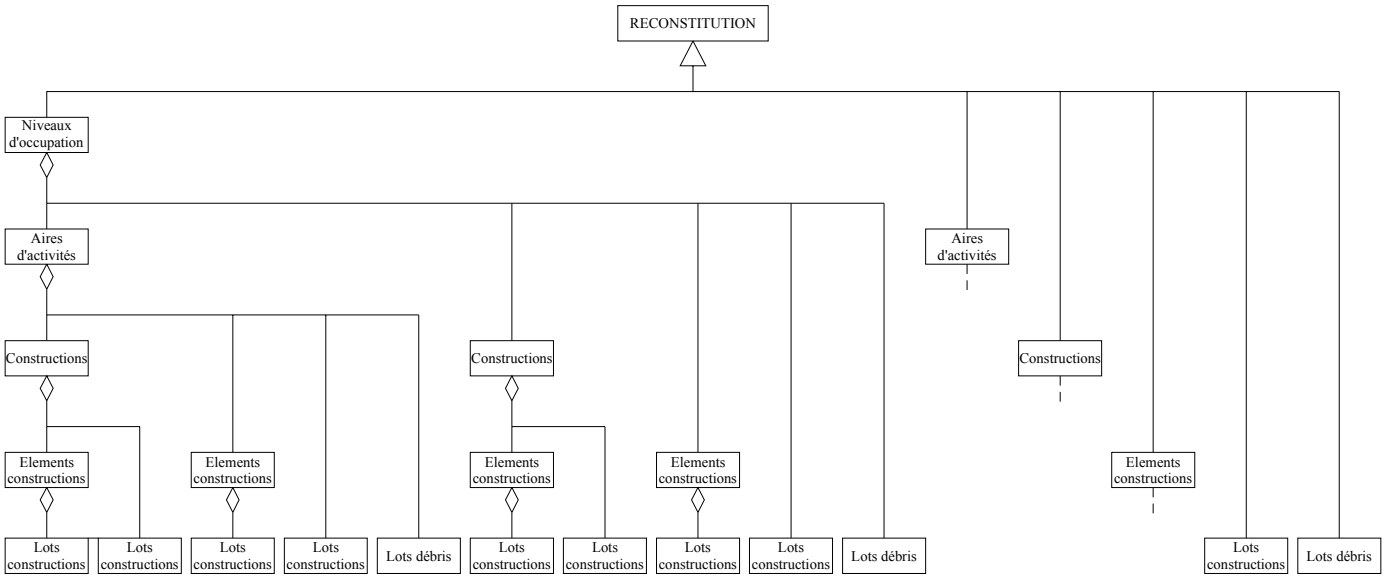
Annexe B Dimensions d'analyses non -utilisée en archéologie

Cette annexe présente les hiérarchies (membres et niveaux) de certaines dimensions d'analyses qui auraient pu composer la BDMRS

La dimension « TOPOLOGIE TEMPORELLE »



La dimension « TOPOLOGIE SPATIALE »



La dimension « RECONSTITUTION DES CONSTRUCTIONS, AIRES D'ACTIVITES et NIVEAUX D'OCCUPATION »