

2m11.3001.9

Université de Montréal

Étude des méthodes de saisie informatique relatives au forage de données appliquées à la prise de décision en urbanisme

par

Sébastien Boyer

Faculté de l'aménagement

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de maître ès sciences appliquées
en aménagement
(option conception, modélisation et fabrication assistées par ordinateur)

novembre 2002

© Sébastien Boyer, 2002



NA
9000
154
2003
15.007

Direction des bibliothèques

AVIS

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :

Étude des méthodes de saisie informatique relatives au forage de données appliquées à la prise de décision en urbanisme

présenté par :

Sébastien Boyer

a été évalué par un jury composé des personnes suivantes :

.....
président-rapporteur

.....
directrice de recherche

.....
codirecteur

.....
membre du jury

Sommaire

Le présent projet de recherche a pour but d'introduire la possibilité d'utiliser de nouvelles approches pour les fins de gestion de l'information urbaine en milieu municipal. Les besoins en la matière sont grandissants en termes de manipulation, de traitement, d'échange, de stockage et de représentation des données. L'origine de ces besoins réside dans la grande variété des informations qui circulent dans les administrations urbaines; de leurs multiples sources de création et de la quantité importante de données qui résulte des interactions entre les humains dans l'espace urbain. Après avoir brossé un tableau des méthodes de gestion actuelles issues des technologies de l'information, mentionnons entre autres les techniques relatives au forage de données, nous décrivons la présente situation de gestion urbaine qui tend à fonctionner davantage à l'aide de logiciels SIURS commerciaux. À un niveau plus expérimental de la recherche, il sera question d'établir les grandes lignes d'un magasin de données informatisé et optimisé supportant les données relatives à l'utilisation du sol urbain. Issu de la technologie ROLAP à cause de sa compatibilité avec les systèmes de bases de données relationnelles en place, il simulera une analyse multidimensionnelle des données par des tables de faits (ou tableaux « Individus-Variables ») prédéfinies. Le magasin en question devra répondre à des requêtes spatiales sur des objets urbains, à l'aide de l'association des données d'utilisation du sol existantes à des attributs de localisation pertinents. Cette information découpée et triée de manière étudiée devra livrer les connaissances attendues suite aux requêtes mathématiques du langage informatique répandu, SQL. La validation de l'hypothèse s'effectuera à l'aide du programme de système de gestion de bases de données *Microsoft Access*. Nous puiserons les données d'analyse dans un échantillon cadastral extrait de la base de données urbaine de la Ville de Saint-Luc. La modélisation des données devra refléter la granularité de l'information qui permettra au magasin de données créé d'évoluer au sein d'un entrepôt de données municipal.

Mots clés : analyse spatiale, base de données spatiale, cadastre informatisé, forage de données, gestion urbaine informatisée, référence spatiale, SIG, SIURS, SQL, systèmes d'information urbaine.

Summary

The present research wants to introduce the possibility of using new approaches in managing urban information through municipalities. The needs for that field are still expanding in terms of manipulation, processing, exchange, stocking and presentation of data. The origins of those needs lie in the various types of information flowing into urban administrations; in the multiple sources the information can come from and mainly in the gigantic loads of data that keep pouring out of human interaction in urban space. After drawing a portrait of the present urban managing situation that uses commercial GIS computer programs, we will try to gain performance with work methods relying on information technologies. Among others, let's mention techniques relative to data mining. At a more experimental level of the research, we will talk about building the guidelines for an optimized computer based data mart that could support land use data. Because of its compatibility with existing relational database systems, the data mart will be build out of the ROLAP technology that simulates multidimensional data analysis through predefined fact tables (double entry matrix). With the help of spatial attributes associations with the existing land use data, that data mart will have to answer information queries on urban objects. That carefully sorted and organized information should deliver the expected knowledge following the mathematical computer assisted queries generated with the well known SQL language. The hypothesis proof will be made with the database managing system computer program *Microsoft Access*. The analysis data will be a sample taken from the City of Saint-Luc computer based land register (graphical and alphanumerical) database. In the end, the data modeling should reflect the optimal information grain that will allow the data mart to evolve into a municipal data warehouse.

Key words : spatial database, computer based land register, Data Mining, computer based urban management, spatial reference, GIS, SRIS, SQL, urban information systems, spatial analysis.

Table des matières :

Sommaire	iii
Summary	iv
Table des matières	v
Liste des tableaux	ix
Liste des figures	x
Liste des annexes	xi
Liste des sigles et abréviations	xii
Introduction	1
Les technologies de l'information au niveau du fonctionnement des villes.....	1
<i>Signification de données, informations et connaissances pour le domaine</i>	1
Problématique de recherche	3
<i>Les besoins en systèmes d'information intégrés adaptés à la gestion municipale</i>	3
La gestion de l'information.....	3
Le travail de l'analyste.....	5
Précision du problème administratif urbain.....	5
Hypothèse.....	7
Objectif de recherche.....	7
Plan de travail.....	7
Partie 1 : Traitement et gestion de l'information	
<u>Chapitre 1 : Le forage de données (data mining)</u>	10
1.1 L'entreposage de données (data warehousing)	10
1.2 Applications spécifiques du forage de données	11
1.3 Approches relatives au forage spatial de données	12
1.4 Les données propices au forage	13
1.4.1 Les tableaux de données.....	14
1.4.1.1 <i>Les tableaux de données recodés</i>	15
1.4.2 Les variables traitées.....	18
1.5 L'utilisation du sol urbain dans les tableaux de données dimensionnels	19
1.5.1 Application de la vision multidimensionnelle à l'information urbaine.....	20
1.5.1.1 <i>Structure de l'information sur les fonctions urbaines</i>	20
1.5.1.2 <i>La vision multidimensionnelle du concept</i>	21
<u>Chapitre 2 : L'assemblage d'un entrepôt de données (data warehouse)</u>	23
2.1 Les bases de données de l'entrepôt de données	24
2.1.1 Types de données constituant les bases de données.....	25

2.2 Le magasin de données (data mart)	25
2.3 Le dictionnaire de métadonnées	26
2.4 L'alimentation de l'entrepôt de données	28
2.4.1 L'extraction et la transformation des données.....	28
2.5 Le concept du multidimensionnel et la technologie OLAP (On-Line Analytical Processing) au sein des bases de données	30
2.5.1 Des technologies issues de l'OLAP.....	31
2.6 Le magasin spécialisé ou l'entrepôt centralisé	33

Chapitre 3 : Gestion du territoire urbain par des outils informatiques de traitement de l'information.....**35**

3.1 Principes de base d'un SIURS	35
3.2 Intégration des données urbaines par un SIURS	38
3.2.1 L'organisation des données contenues.....	38
3.2.2 L'utilisation d'une base géographique urbaine à grande échelle.....	38
3.2.3 Caractéristiques d'opération d'un SIURS.....	40
3.3 Composition du SIURS : l'information supportée concernant un site	41
3.3.1 Les données urbaines.....	41
3.4 L'informatisation de l'information dans le contexte administratif urbain	44
3.4.1 Les types de données traitées et leur gestion.....	44
3.4.2 Relations à établir entre les données urbaines par l'encodage.....	44
3.4.2.1 Selon un identificateur nominal.....	44
3.4.2.2 Selon un identificateur à référence spatiale.....	45
3.4.2.3 Selon un identificateur géographique.....	45
3.5 La codification de l'information géographique urbaine au sein du SIURS	46
3.5.1 Types de codifications municipales les plus courantes.....	46
3.5.2 Édition et stockage dans les bases de données urbaines.....	47
3.5.2.1 Extraction des données relatives à localisation d'un lot.....	48
3.5.3 Traitements de données exécutés par un SIURS.....	50
3.5.3.1 Critères de qualité de l'information traitée par le SIURS.....	51

Chapitre 4: Outils de gestion et de manipulation des connaissances propres à la prise de décision.....**53**

4.1 La localisation de l'information au service du forage spatial de données	54
4.1.1 État des besoins en matière de forage spatial des données.....	54
4.2 Les systèmes de gestion de bases de données spatiaux et relationnels	55
4.3 L'adoption d'un système de requête géographique à deux vitesses	55
4.4 La structuration informatique des données urbaines	56
4.4.1 Les structures de données courantes.....	57
4.4.2 Les possibilités analytiques des données urbaines structurées.....	59
4.4.3 Les données idéales pour le domaine.....	60

Chapitre 5 : Contraintes d'insuffisance sémantique dans les bases de données.....62

5.1 L'information géographique	62
5.1.1 Les métadonnées géographiques.....	63
5.2 L'insuffisance sémantique des structures de données	64
5.3 Les contraintes d'intégrité relationnelles	64
5.4 L'encodage des relations spatiales entre les objets	65
5.5 Le problème à l'échelle de la gestion urbaine	66
5.5.1 Le découpage des informations par objet	66
5.5.1.1 <i>L'organisation hiérarchique des informations</i>	67

PARTIE 2: Expérimentation

Chapitre 6 : Approche méthodologique du problème posé et présentation du cas étudié.....72

6.1 L'approche analytique adoptée pour aborder la problématique	72
6.1.1 Application de l'approche analytique au problème.....	73
6.2 Explication du choix de la solution en fonction de l'approche adoptée	73
6.2.1 Situation du cas dans le processus de fonctionnement des SIURS.....	74
6.2.2 Justification de l'orientation prise pour le traitement des données.....	76
6.2.3 Échelle d'intervention pour le relevé et l'expérimentation	76
6.2.3.1 <i>Collecte des données cadastrales et réglementaires</i>	77

Chapitre 7: Décomposition analytique de la problématique cadastrale par construction du modèle théorique.....80

7.1 Choix du type de magasin de données et de sa source d'alimentation	80
7.2 Composition des tables de faits du magasin de données	81
7.2.1 Définition de la granularité des tables (dimensionnelles) de faits.....	81
7.2.2 Identification des dimensions et des faits de base possibles.....	82
7.2.2.1 <i>Faits pour la localisation et l'identification d'un lot</i>	82
7.2.2.2 <i>Faits pour la documentation des usages prescrits</i>	83
7.2.3 Identification de faits dérivés possibles	84
7.3 Visualisation des modèles dimensionnels construits	86
7.3.1 Plan de composition de l'entrepôt de données municipal possible.....	86
7.3.2 Détail relationnel entre les tables de faits traitées.....	86
7.3.3 Arbre de hiérarchie des attributs dimensionnels et des requêtes.....	87
7.4 Le dictionnaire de données	87

Chapitre 8 : Restitution analytique des éléments du cadastre par validation du modèle dimensionnel.....89

8.1 Construction de la table de données informatisée à partir du modèle	90
8.1.1 Guide de saisie des champs de données pour le cadastre.....	90
8.1.2 Guide de saisie des champs de données pour la grille d'usages.....	93

8.2 Conception des requêtes SQL (plan de forage des données)	96
8.2.1 Cours SQL (Structured Query Language) accéléré.....	97
8.2.2 Formulation des requêtes SQL et obtention des faits dérivés.....	98
8.2.2.1 <i>Journal des requêtes sur la localisation d'un lot</i>	98
8.2.2.2 <i>Journal des requêtes sur la réglementation de zonage</i>	102
8.2.2.3 <i>Agrégation des dimensions les plus combinées</i>	104
<u>Chapitre 9 : Discussions et conclusions sur l'expérimentation</u>	106
9.1 Discussion sur la gestion des éléments géographiques urbains	106
9.2 Discussion sur l'encodage de la grille de réglementation de zonage	108
9.3 Discussion sur la méthode de saisie	109
9.4 Conclusion sur la rencontre des objectifs visés au départ	111
9.5 Un mot sur le trop plein de métadonnées	112
9.6 Conclusion générale sur les SIURS et le but des magasins de données	113
Définitions contextuelles des termes importants employés	116
Bibliographie	118

Liste des tableaux :

I. Méthodes relatives au forage de données suivant un bloc décisionnel (prédictif).....	12
II. Variables quantitatives et qualitatives dérivées.....	18
III. Les cinq facteurs et dimensions impliqués par l'utilisation du sol.....	19
IV. Liste unidimensionnelle des établissements urbains suivant la superficie couverte.....	21
V. Matrice de la superficie couverte des fonctions urbaines selon activité.....	22
VI. Types de données constituant l'entrepôt de données.....	25
VII. Questions résolues à l'aide des métadonnées.....	27
VIII. Opérations de transformation des données.....	29
IX. Différences fonctionnelles entre les technologies MOLAP et ROLAP.....	32
X. Résumé des différences techniques entre l'entrepôt et le magasin de donnée.....	33
XI. Caractéristiques du système d'information spatiale à grande échelle.....	40
XII. Les trois grandes catégories de données urbaines.....	42
XIII. L'information reliée à un site urbain.....	43
XIV. Types de codifications municipales.....	46
XV. Objectifs de la gestion de bases de données relatives à l'utilisation du sol.....	48
XVI. Description des 2 modes de traitement de données dans un système d'information.....	50
XVII. Critères de qualité de l'information urbaine traitée.....	51
XVIII. Problèmes actuels à éliminer par la structuration des données informatiques.....	57
XIX. Possibilités d'analyses effectuées à l'aide des données urbaines.....	60
XX. Faits de base issus de la matrice graphique.....	83
XXI. Faits de base issus de la grille des usages prescrits.....	84
XXII. Faits dérivés issus de requêtes sur la table de localisation d'un lot.....	85
XXIII. Faits dérivés issus de requêtes sur la table de réglementation des usages.....	85
XXIV. Métadonnées sur les dimensions de base et dérivés.....	88
XXV. Les trois niveaux d'application de la granularité de l'information traitée.....	89
XXVI. Tableau de données incomplet montrant des occurrences multiples dans les cellules.....	108

Liste des figures :

1. Graphe de voisinage et matrice logique correspondante pour quatre objets.....	13
2. Tableaux à double entrée « Individus-Variables » génériques.....	14
3. Exemple de tableau logique.....	16
4. Tableau « Individus-Variables » sous forme disjonctive logique.....	16
5. Exemple de constitution d'un tableau de contingence simple.....	17
6. Exemple de tableau de contingence chronologique.....	17
7. Architecture d'un entrepôt de données.....	24
8. La nature du magasin de données au sein de l'entrepôt de données.....	26
9. Liens entre dictionnaire de métadonnées et processus de traitement de l'information.....	27
10. L'acheminement des données vers l'entrepôt de données.....	29
11. Modèle de données hiérarchique relationnel versus multidimensionnel.....	30
12. Schématisation des modes d'analyse relationnels versus multidimensionnels.....	31
13. Le cheminement de l'information dans un SIURS.....	36
14. Modèle théorique des relations thématiques dans un SIURS.....	37
15. Éléments de base constituant une BGU.....	39
16. Démonstration des couches d'information documentées dans un SIG informatisé.....	41
17. Les quatre niveaux de liaison informatiques des dossiers de propriété par géocode.....	47
18. Découpage du globe en fuseaux et application d'une grille de repérage.....	49
19. Construction du matricule de localisation unique d'une propriété.....	50
20. Exemple de requêtes cartographiques interactives.....	56
21. Répartition de l'information sous forme d'arborescence.....	57
22. Répartition de l'information sous forme de réseau.....	58
23. Répartition des information sous forme de tableau relationnel « Individus-Variables ».....	59
25. Index de jointure sous forme de table et de matrice pondérée.....	65
26. Hiérarchie relationnelle des objets constituant une description urbaine.....	68
27. Graphe des relations possibles entre les objets définissant un quadrilatère urbain.....	69
28. Situation de l'intervention dans le processus de fonctionnement d'un SIURS.....	75
29. Extrait commenté d'un plan de cadastre municipal source de la Ville de Saint-Luc.....	77
30. Extraits de la grille des usages et du plan de zonage complémentaires au cadastre.....	78
31. Construction numérique de la table des faits sur l'identification et la localisation d'un lot.....	92
32. Illustration logicielle de la table des faits de localisation d'un lot en deux parties.....	93
33. Construction numérique de la table des faits sur la réglementation de zonage.....	95
34. Illustration logicielle de la table des faits de réglementation de zonage en deux parties.....	95
35. Gestion des pointeurs entre les champs par <i>Microsoft Access</i>	97
36. Les gestionnaires de tables et de requêtes au sein de la base de données.....	97
37. Structure du code informatique d'un énoncé de requête SQL.....	98
38. Résultats informatiques de la requête "matricule du lot".....	99
39. Résultats informatiques de la requête "numéro de lot".....	99
40. Résultats informatiques de la requête "adresse civique".....	100
41. Résultats informatiques de la requête "superficie du lot".....	101
42. Résultats (données de travail) de la requête sur les attributs de "relation d'adjacence".....	101
43. Résultats informatiques de la requête d'interprétation des attributs d'adjacence.....	102
44. Résultats informatiques de la requête sur la dimension des marges latérales totales.....	102
45. Résultats informatiques des requêtes sur les superficies de plancher et de terrain minimales.....	103
46. Résultats informatiques de la requête de CES pour la zone d'expérimentation H1-264.....	103
47. Résultats informatiques de la requête de COS pour la zone d'expérimentation H1-264.....	104
48. Graphe des relations entre les trois éléments géographiques de base en urbanisme.....	107
49. Relations entre les informations de référence et de calcul dans un plan de forage.....	110

Liste des annexes :

Annexe 1 : Planches de travail : banques de données et plan de forage	xii
Annexe 2 : Documents sources	xvii
Annexe 3 : Règles OLAP et code SQL commenté des requêtes	xxi
Annexe 4 : Entrevue avec l'urbaniste de la ville de Saint-Luc (2001/10/03)	xxxii

Liste des sigles et abréviations:

BD : Base de Données
BDM : Base de Données Multidimensionnelle
BDR : Base de Données Relationnelle
BGU : Base Géographique Urbaine
CES : Coefficient d'Emprise au Sol
COS : Coefficient d'Occupation du Sol
FD : Forage de Données
FSD : Forage Spatial de Données
GIS : Geographic Information System
MTM: projection cartographique Mercator Transverse Modifiée
MOLAP : Multidimensional On-Line Analytical Processing
NAD 83 : North American Datum 1983
OLTP: On-Line Transaction Processing
OCDE : Organisation de Coopération et de Développement Économique
ROLAP : Relational On-Line Analytical Processing
SIURS : Système d'Information Urbaine à Référence Spatiale
SGBD : Système de Gestion de Base de Données
SGBDR : Système de Gestion de Base de Données Relationnelle
SGBDM : Système de Gestion de Base de Données Multidimensionnelle
SIG : Système d'Information Géographique
SQL : Structured Query Language

Introduction

Les technologies de l'information au niveau du fonctionnement des villes :

Tel que publié par l'OCDE (Organisation de Coopération et de Développement Économique) et cité par Durlak (1989) il y a une dizaine d'années, les technologies de l'information ont une influence marquée sur plusieurs aspects socio-économiques d'une organisation. On parle de faciliter les restructurations, d'accélérer les transactions diverses et d'améliorer les conditions de travail; tout en innovant au niveau des interactions sociales. Si on applique ces technologies au fonctionnement des villes, on pourrait être témoin de changements organisationnels et techniques au sein de leur administration. Ce sont donc les différents services de la gestion urbaine qui peuvent en bénéficier. On parle de la sécurité publique, de la régulation des circulations automobiles, de la répartition des ressources, du développement économique autant que de l'économie d'énergie, de l'élaboration des plans d'urgence et des services sociaux. Bref, économies, efficacité et qualité des prestations de services ainsi qu'une plus grande flexibilité des capacités d'adaptation ne sont que certains avantages du développement des technologies de l'information.[Durlak, 1989]

Signification de données, informations et connaissances pour le domaine :

Selon Durlak, les données en elles-mêmes n'ont aucune signification particulière (ex : 200 mètres ou 2 voies). Ce n'est qu'à partir du moment où on les combine les unes avec les autres qu'elles deviennent des informations. L'utilisateur qui travaille avec ces informations accroît alors son bassin de connaissances. Peu importe la forme de transmission des informations, (chiffres, lettres, images, son) elles doivent nous aider à construire des connaissances qui nous permettent de mieux comprendre un phénomène donné. Les ordinateurs, programmés par des « êtres pensants » (pour employer l'expression de Durlak), sont en mesure de manipuler les données pour en faire de l'information. C'est le même dessein au niveau des télécommunications qui permettent de transmettre les informations. Un relais d'informations est donc formé entre les êtres

pensants par l'intermédiaire des machines « intelligentes » (ordinateurs) qui, elles-mêmes, échangent entre elles par les télécommunications. Tous ces agents ont besoin de parler le même langage afin de se comprendre à travers les flux de données. [Durlak, 1989]

Justement, au niveau des flux de données, trois pôles d'activités articulent le fonctionnement des systèmes d'information. Il s'agit du **stockage des données**, de leur **traitement** et enfin de leur **transmission** ou **communication**. C'est en l'union de ces trois pôles que les réelles répercussions s'exercent sur les activités urbaines. Une migration s'est faite à partir du traitement des données électroniques en passant par les systèmes d'informations pour arriver finalement aux systèmes de connaissances. Ces transformations, au niveau de la terminologie, entraînent une évolution de la manipulation d'importantes quantités de données à leur structuration. On permet donc à l'utilisateur d'augmenter ses capacités déductives et analytique à l'aide d'une information devenue de plus en plus utile et pertinente. [Durlak, 1989]

L'objet principal des technologies de l'information consiste en l'accroissement de la capacité de mémoire et de raisonnement de l'être humain ainsi que de faciliter ses échanges avec les autres. Plus précisément, dans le cadre administratif municipal, un bon système d'information doit remplir les rôles suivants :

- Optimiser l'utilité des données dans le processus de prise de décision par leur collecte, leur emmagasinage et leur sélection;
- Aider à la prévision des conséquences des programmes publics prévus par la simulation et la visualisation basées sur des modèles mathématiques;
- Permettre l'accès à plusieurs systèmes d'information intégrés contenant des données numériques, textuelles et graphiques;
- Assister l'élaboration de nouveaux matériels et de nouvelles techniques de gestion permettant d'accomplir des tâches courantes auxquelles les administrations municipales ne peuvent échapper;
- Rendre possible la personnalisation de ces systèmes selon les caractéristiques de ses utilisateurs (style d'action posé, de décision et de capacités d'apprentissage);

- Permettre le recours à des systèmes de partage de l'information intelligents offrant des interfaces de résolution de problèmes adaptées.

[Durlak, 1989]

Problématique de recherche

Les besoins en systèmes d'information intégrés adaptés à la gestion municipale :

En matière de gestion du territoire, la liste des données que devraient contenir ces systèmes peut s'étendre à l'infini si on désire répondre aux demandes de chaque instance administrative (manipulation de données autant démographiques, sociales, économiques que géographiques). Malgré l'existence de certaines bases de données et de systèmes de traitement automatisés accessibles à l'époque, on pourrait encore croire aujourd'hui à un système beaucoup trop lourd à gérer technologiquement parlant. [Durlak, 1989]

La gestion de l'information

On débute la démarche de gestion de l'information par une ou plusieurs procédures d'entrée de données, de vérification, de tri et de classification. L'information doit être entreposée de façon sécuritaire de sorte à en minimiser les risques de perte (accident, piratage, usure du temps). La forme de l'information choisie devrait en permettre l'analyse et le traitement pour la production éventuelle de tout autre type d'information nécessaire ou désiré. Il doit être possible d'extraire, de reproduire, de transférer ou de montrer l'information de plusieurs manières. Ce lourd carnet de commande peut s'appliquer en fait à n'importe quel domaine et ne fait pas exception en aménagement du territoire ou en urbanisme. En général, le format informatique (numérique) pour la conservation de l'information est celui qui est actuellement le plus prisé pour son stockage compact et la rapidité de récupération de grands volumes de données.

Ces nouvelles façons de procéder efficaces ont également entraîné leur lot de problèmes. Tel que cité par Dale et McLaughlin [p.158, *Land Information Management*, 1989], ces problèmes en question ont été identifiés par Glenn Morgan de la Banque Mondiale. Voyons-les brièvement :

1. La difficulté à manipuler de grands volumes d'informations hétérogènes : avec l'arrivée de ces nouvelles technologies, les utilisateurs sont conscients des possibilités et demandent l'accès à un plus grand éventail de types d'informations, de formats, d'échelles de cartes, de couverture de relevé, de contenu cartographique, de symbolisation ainsi que de codage d'attributs. Le stockage de données n'étant plus réellement un problème, les capacités de manipulation des utilisateurs sont dépassées.

2. La dépendance aux systèmes informatiques et aux logiciels : la prolifération de plusieurs types d'équipements informatiques, de logiciels et de structures de données signifie que l'échange d'information peut devenir encombrant et très coûteux. Même si des standards internationaux pour l'échange d'information existent, il n'y a toujours pas d'entente universelle sur leur utilisation.

3. La rapidité des changements technologiques : la planification à long terme est obstruée par le changement rapide des technologies. Toutes les ressources disponibles peuvent être investies dans la construction d'un système en particulier alors qu'un nouveau système, meilleur, émergera peu après. De plus, les technologies informatiques utilisées peuvent devenir obsolètes avant d'être complètement fonctionnelles.

4. Le maintien d'obligations administratives et financières : les incertitudes qui surgissent des changements technologiques entraînent le besoin de démontrer qu'un investissement à cet effet comportera une efficacité au niveau du coût et des bénéfices qu'il engendre. Le comportement imprévisible de tels changements ouvre donc ces projets à une critique justifiable. Les bénéfices ne se montrent malheureusement pas immédiatement. On doit investir durant une certaine période où il n'y a pas de retour financier important.

5. La réquisition d'une réforme institutionnelle : les nouvelles technologies requièrent souvent des changements majeurs dans la structure de support organisationnel d'une institution.

6. La pénurie de main d'oeuvre spécialisées : il y a effectivement une pénurie majeure de personnel spécialisé et expérimenté à tous les niveaux et plus particulièrement parmi les gestionnaires de production. Des décisions sur les systèmes et leur développement sont souvent prises par ceux qui ne possèdent pas les qualifications requises.

Le travail de l'analyste :

Enfin, Le grand défi de l'analyste, au niveau du traitement des données urbaines, est de simplifier efficacement la description de l'environnement de sorte à en maintenir l'intégrité. Pour ce faire, la construction d'un modèle statistique doit précéder le modèle visuel. Les deux facteurs principaux à considérer sont le type de données employées et la stratégie d'échantillonnage élaborée. Les données peuvent être collectées sur le site ou issues de dossiers ou d'archives. l'échantillonnage des données sur le site peut être faite au hasard ou de façon étudiée selon les résultats anticipés. À chaque analyste revient le rôle de voir à ce que sa technique soit efficace et réponde à ses attentes. [Smardon, Richard C. 1986]

Précision du problème administratif urbain :

L'information gérée en milieu administratif urbain provient de multiples sources et concerne différents faits urbains. Dans un monde où l'échange d'information est à la base de tout procédé administratif, une exploitation déficiente de l'information devient un handicap certain. Il s'agit ici de pallier au manque en disposant l'information spatiale de façon à rendre plus évidentes les analogies possibles liées à la gestion de l'utilisation du sol. On parle des relations entre les données induites par la coexistence des entités géographiques documentées constituant le monde urbain. Le développement rapide des

technologies informatiques impose maintenant une loi de conformité au niveau de la compatibilité des différentes informations traitées. Les besoins sont variés, allant des capacités d'échange de cette information, selon des formats de données propres à la pratique, à une visualisation relationnelle des plus descriptives et explicites de cette dernière. Les professionnels du domaine cherchent à avoir le contrôle sur l'information qu'ils manipulent par l'obtention des réponses adaptées à leurs différentes requêtes.

Parlons du problème existant à la base d'un point de vue plus pratique. Maintenant, la technologie informatique a donné naissance à plusieurs outils logiciels qui offrent la possibilité de prendre en charge cette multitude de données pour les rendre accessibles visuellement à son utilisateur. Ces outils en question sont des systèmes d'informations urbaines à références spatiales. La difficulté majeure avec ces derniers réside dans le fait que l'accumulation et le stockage des informations s'effectue en "double". C'est-à-dire qu'ils font coexister deux ou plusieurs bases de données distinctes qui décrivent souvent les mêmes objets urbains. On décèle ici une lacune d'intégration des informations urbaines au sein de ces systèmes.

Les municipalités les plus avancées informatiquement sont toujours à la recherche d'un format d'information plus léger et d'une façon de détecter plus rapidement les possibilités d'optimisation de leur banque de connaissances. Par optimisation, on signifie l'allègement des structures de données existantes au sein de la base informatique par la création de liens entre des informations complémentaires au départ. La résolution de ce problème d'interaction, avec les banques de données urbaines, permettrait aux aménagistes du territoire et aux gestionnaires municipaux d'arriver plus efficacement et plus rapidement à des conclusions lorsqu'en situation de prise de décision.

Hypothèse :

Il est possible, par un système d'information suivant une structure de requêtes logiques appliquées à une base de données relationnelle, de décrire et de documenter les relations spatiales entre des éléments géographiques urbains.

Objectif de recherche :

Proposer une méthode d'organisation de l'information, liée à l'identification et à la localisation des éléments géographiques urbains, permettant une description informatique de la régie municipale de ces derniers.

Plan de travail :

Après une documentation complète du domaine et suivant une approche méthodologique analytique du problème, les prochaines étapes ont été franchies pour réaliser le présent projet de recherche:

Mise en contexte

- Définition du problème au sein du processus de cheminement de l'information sur un fait urbain étudié et documenté dans un système d'information à référence spatiale (traitement, saisie, conservation, accès).
- Choix de l'échelle cadastrale municipale comme plan de travail au niveau de la collecte de données urbaines.

Cueillette de données

- Les données traitées sont tirées d'une banque de connaissances déjà composée de tableaux et de champs de données nourrissant un cadastre municipal composé de multiples formes d'informations spatiales.

- L'information cadastrale pour documenter huit lots contigus choisis au hasard a été extraite afin d'être découpée puis réinjectée (saisie) dans des tables formant le nouveau magasin de données.

Conception du modèle théorique

- Sélection d'un type de magasin de données à construire qui sera "multisource" (i.e. comportant plusieurs sources de documentation). Ces sources seront le cadastre municipal et la réglementation d'urbanisme.
- Définition de la granularité des tables des faits ou des champs de données. En fait, on énonce à quelle entité les informations de la table de données se rapporteront. Pour notre part, il sera question du lot foncier et des secteurs urbains régis par les règlements de zonage.
- Des plans des tables constituant le magasin de données sont construits ainsi qu'une projection de l'entrepôt de données dans lequel ce magasin viendrait s'inscrire.
- Génération d'un dictionnaire de métadonnées ou d'un journal de saisie relié aux données recueillies.
- Élaboration de requêtes relationnelles d'extraction applicables aux données.

Validation du modèle théorique

- Construction des tables de données urbaines (cadastrales, réglementation) à l'aide du logiciel de traitement de bases de données *Microsoft Access*.
- Traduction des requêtes relationnelles par le langage informatique *SQL*.
- Mises en forme d'un échantillon de magasin de données *ROLAP* par la formation des liens informatiques appropriés entre les données et l'application des requêtes *SQL* aux tables de données construites (intégration).

Partie 1 :

Traitement et gestion de l'information

Chapitre 1 : Le forage de données (data mining)

Objectifs du chapitre :

L'ouverture à ces techniques d'organisation numérique de l'information est répandue dans plusieurs domaines d'application où l'on retrouve des données informatisées. Étant donné le fait qu'aujourd'hui à peu près tout procédé administratif est informatisé, la gestion municipale n'y fait pas exception. Seulement, certaines règles de forage semblent omises par les concepteurs de banques de données. Il en résulte une mauvaise adaptation du potentiel de certaines techniques aux besoins institutionnels. Le présent chapitre introduira le domaine en clarifiant les différents aspects composant le forage de données ainsi que ses méthodes d'application d'une manière plus détaillée.

Le but de cette discipline est de découvrir les modèles (patrons) d'organisation fonctionnels ($f(x)=y$) ou logiques (règles d'association ou arbre de décision) difficiles à mettre en évidence à cause de l'important volume de données traitées, d'une trop grande quantité de variables à prendre en considération ou des modèles imprévisibles jamais envisagés par l'analyste. Ce processus d'extraction des connaissances comprend des étapes préliminaires telle la construction d'un entrepôt de données (data warehouse). Cet entrepôt est en fait une base de données construite dans un but décisionnel selon des bases de production provenant de plusieurs sources. Le « data warehouse » est en mesure d'archiver des données historisées et actualisées sur une base cyclique soit par « data pull » (interrogation des bases de données) ou par « data push » (envoi automatique des modifications par les serveurs informatiques de bases de données). La base d'un entrepôt de données est de dimension très importante à cause de l'archivage routinier des nouvelles données. [Jambu, 1999]

1.1 L'entreposage de données (data warehousing)

Un système d'information désigne les données collectées par un organisme de façon systématique ainsi que les traitements effectués sur ces dernières pour assurer son fonctionnement. Ces systèmes mettent en jeu des bases de données qui sont en fait des

structures informatiques permettant d'archiver les données et de les mettre à jour. Des réseaux informatiques sont aussi impliqués et permettent la circulation des données. **[Jambu, 1999]**

L'entrepôt de données d'une entreprise contient à la fois des données opérationnelles enregistrées au fil du temps, des données agrégées, des données de pilotage et toutes données externes ayant une relation avec les activités de l'entreprise. Ces données, étant consignées dans une ou plusieurs base(s) relationnelle(s), sont accessibles pour toutes les applications et par des systèmes d'aide à la décision (systèmes de rapports, d'analyses statistiques, de forage de données). L'entrepôt de données se différencie alors du système informationnel ordinaire par l'orientation de l'usage des données vers la décision plutôt que vers le simple archivage. C'est en fait l'idée d'exploiter des données historiques archivées pour en établir des tendances de développement qui fait toute la différence. Du point de vue « utilisateur-décideur », la caractéristique propre d'un entrepôt de données est le fait que les données soient prêtes à l'emploi et qu'elles soient accessibles immédiatement par des requêtes définies de façon claire. Maintenant, soutient l'auteur, les décideurs font de plus en plus appel à des données détaillées plutôt qu'à des données transformées, agrégées ou mises sous forme d'indicateurs. **[Jambu, 1999]**

Enfin, les points clés d'un entrepôt de données, comparativement aux applications usuelles d'un système d'information, consistent en sa capacité d'extraction, de modélisation (qu'elle soit relationnelle ou pas) ainsi que la capacité des logiciels d'analyse de travailler directement sur celui-ci. Les coûts impliqués en développement, en intégration et en maintenance sont plus faibles.

1.2 Applications spécifiques du forage de données

Les méthodes relatives au forage de données sont classées suivant deux catégories, soit un bloc exploratoire plus statistique et un autre décisionnel plus logique. Dans le cadre de la recherche, nous nous appuyerons davantage sur le bloc décisionnel tel que décrit dans le prochain tableau.

Tableau I : Méthodes relatives au forage de données suivant un bloc décisionnel (prédictif)

La recherche de règles de classement d'objets	Une méthode d'apprentissage supervisé qui induit une description permettant de classer les prochaines données à partir d'une base d'exemples et d'un attribut à prédire.
La régression	Permet de prédire la valeur d'une variable en recherchant une fonction mathématique avec une variable à prédire et un résidu possible. C'est un modèle fonctionnel.

[Jambu, 1999]

Ces méthodes démontrent que le forage de données se base en grande partie sur le domaine de l'intelligence artificielle pour les approches plus logiques. On travaille donc essentiellement à améliorer les algorithmes existants afin qu'ils puissent traiter de plus grands volumes de données en appliquant des techniques d'indexation. On tente d'adapter de mieux en mieux au forage de données, les langages de requête tel SQL.

[Jambu 1999]

1.3 Approches relatives au forage spatial de données

Le forage spatial de données intègre complètement l'information concernant la localisation spatiale et les liens de voisinage. Les méthodes de forage de données classique ne considèrent pas ces relations spatiales entre des variables et ne sont donc pas adaptées. L'intégration des techniques SIURS et de forage de données est imminente pour le développement de nouvelles méthodes de FSD. Dans l'image suivante, les relations spatiales en question sont formalisées par ce qu'on appelle un graphe de voisinage entre quatre objets, pouvant également prendre la forme d'une matrice croisée.

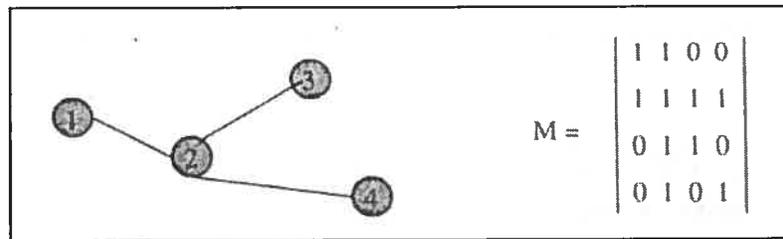


Figure 1 : Graphe de voisinage et matrice logique correspondante pour quatre objets

Cette notion de voisinage peut autant représenter la contiguïté entre des zones ou une proximité sur des points. Les méthodes relatives au forage spatial de données étant des extensions du forage de données classique, engendrent des problèmes de performance importants. Ces derniers découlent des volumes de données imposants générés par le codage des localisations géométriques et de la complexité des calculs des relations spatiales entre les objets encodés dans la base de données. Les recherches amorcées visent donc à optimiser les méthodes d'analyse qui tiennent compte des relations spatiales. [Zeitouni, Yeh, 1999]

Enfin, on considère le forage de données comme étant un processus cheminant des données élémentaires contenues dans un entrepôt de données jusqu'à la décision. Le tout, en intégrant à chaque étape, une valeur informationnelle supplémentaire. Jambu, en 1999, nous présente le forage de données comme un processus qui inclut une manière automatisée d'utiliser les méthodes d'analyse de données. Selon lui, l'entreposage de données devient la logistique de l'information indispensable au forage de données. [Jambu, 1999]

1.4 Les données propices au forage

Tout ensemble de données qui serait non structuré n'est pas analysable par les méthodes de forage de données. Les méthodes mathématiques impliquées ne s'appliquent bien qu'à des situations de tableaux de données. Il faudra donc être en mesure d'extraire d'une situation complexe de données, une situation analysable exprimable en termes de tableaux. [Jambu, 1999]

Tel qu'illustré dans les figures suivantes, un tableau de données est en fait un tableau à double entrée consignant des nombres qui mettent en jeu deux ensembles d'objets. Les rangées du tableau correspondent aux individus, alors que les colonnes correspondent aux variables. Quand des données temporelles sont en jeu, les périodes peuvent être associées aux rangées ou aux colonnes en fonction du cas à l'étude.

		Variables			
		X_1	X_2	X_j	X_p
Individus	I				
	i				
	\vdots				
	n				

		Variables					
		X_1	X_2	\dots	X_j	\dots	X_p
Individus	I	x_{I1}	x_{I2}	\dots	x_{Ij}	\dots	x_{Ip}
	\vdots						
	i	x_{i1}	x_{i2}	\dots	x_{ij}	\dots	x_{ip}
	\vdots						
	n	x_{n1}	x_{n2}	\dots	x_{nj}	\dots	x_{np}

illustrations : [Jambu, 1999]

Figure 2 : Tableaux à double entrée « Individus-Variables » génériques

La case indice correspondant à l'individu et à la variable qui le qualifie, peut contenir une valeur numérique associée à une variable quantitative, comme un nombre représentant le code d'une catégorie. Il existe différents types de tableaux de données « Individus-Variables » soit : quantitatifs (ex : débits), temporels (ex : gains annuels), qualitatifs (ex : rang social), combinaison qualitatif-quantitatif et finalement la combinaison qualitatif-quantitatif-temporel.

1.4.1 Les types de tableaux de données

Les tableaux de données à une variable : ils sont la forme la plus simple que ces derniers peuvent prendre. Ils consistent en l'association d'un seul qualificatif présent ou absent à une série d'individus cible (une seule colonne de variables est associée à celle des individus) .

Les tableaux de données à plusieurs variables : ils permettent de fixer les notions relatives à la manipulation de tableaux de données. Ils étendent le concept du tableau « Individus-Variables » en considérant plusieurs variables notées sur une même population.

1.4.1.1 Les tableaux de données recodés :

Les tableaux de données recodées sont en fait de nouveaux tableaux de données établis à partir d'un tableau source et à l'aide de transformations mathématiques appelées « recodage ». Une bonne définition du recodage relationnel a été écrite par Benzécri en 1977. La voici donc [telle que citée dans **Jambu, 1999, p.31**]:

"L'essence du codage des données est de traduire fidèlement les relations observées entre les choses par des relations entre êtres mathématiques, de telle sorte qu'en réduisant par le calcul la structure mathématique choisie pour image du réel, on ait de celui-ci un dessin simplifié accessible à l'intuition et à la réflexion avec la garantie d'une critique mathématique."-Benzécri, 1977

Voyons maintenant les types de tableaux recodés pouvant être intéressants pour les fins de notre étude.

Les tableaux logiques : ils définissent la présence ou l'absence d'un attribut par 1 ou 0. Parmi ceux-ci, on compte également les tableaux logiques issus de variables à réponses multiples.

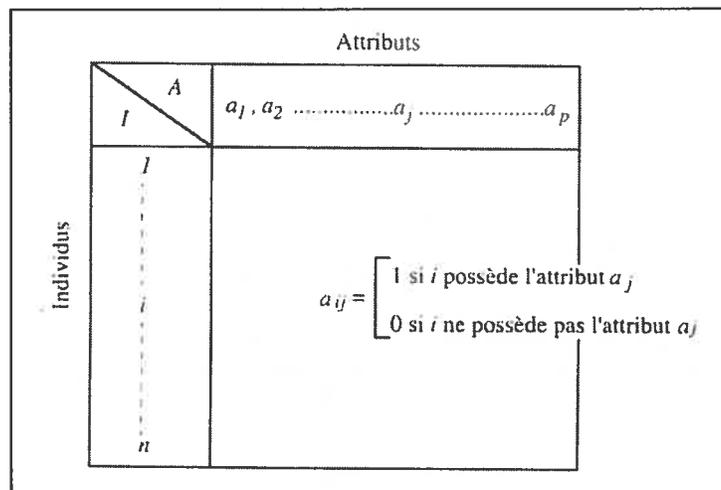


illustration : [Jambu, 1999]

Figure 3 : Exemple de tableau logique

Les tableaux « Individus-Variables » qualitatifs mis sous forme disjonctive : ils consistent en la conversion d'un tableau « Individus-Variables » en un tableau logique référentiel pour chaque variable. Dans l'image suivante, des individus on répondu à des questions comportant chacune trois modalités de réponses. On voit le tableau « Individus-Variables » de départ à gauche mis sous forme disjonctive logique à droite.

		Variables					Variables						
		X	Q_1	Q_2	Q_3	Q_4	Q_5	X	Q_1	Q_2	Q_3	Q_4	Q_5
		i						i	M_1, M_2, M_3	M_1, M_2, M_3, M_4	M_1, M_2, M_3, M_4	M_1, M_2, M_3	M_1, M_2, M_3
Individus	1	1	1	4	2	1	1	1					
	2	2	1	4	2	2	2	1					
	3	2	1	4	2	3	3	1					
	4	2	2	3	1	3	3	1					
	5	1	2	2	3	3	3	1					
	6	3	2	1	1	3	3	1					
	7	3	3	2	2	3	3	1					
	8	3	3	3	2	2	2	1					
	9	1	4	3	3	1	1	1					

illustration : [Jambu, 1999]

Figure 4 : Tableau « Individus-Variables » sous forme disjonctive logique

Les tableaux de contingence simples : ils sont issus du croisement de deux variables qualitatives exclusives (ne mettent pas en jeu de variables à réponses multiples).

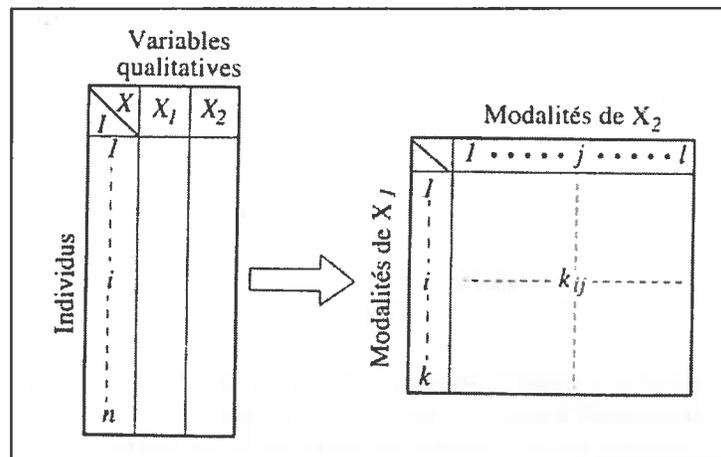


illustration : [Jambu, 1999]

Figure 5 : Exemple de constitution d'un tableau de contingence simple

Enfin, à partir d'un ensemble de variables qualitatives, on peut calculer d'autres types de tableaux pour mettre en valeur diverses singularités. Le temps peut également être étudié à travers des tableaux de contingence chronologiques (formé d'une suite, en rangées et en colonnes, de tableaux de contingences indicés par le temps) et des tableaux « Individus-Variables » qualitatives chronologiques (un cube « individus, variables, périodes » dont les tranches correspondant à chacune des périodes sont mises les unes à côté des autres). En voici une illustration :

		Période 1			Période 2			Période 3		
		X			X			X		
		$X_1^{(1)}$	$X_j^{(1)}$	$X_p^{(1)}$	$X_1^{(2)}$	$X_j^{(2)}$	$X_p^{(2)}$	$X_1^{(3)}$	$X_j^{(3)}$	$X_p^{(3)}$
Individus	1									
	i									
	k									
	n									

illustration : [Jambu, 1999]

Figure 6 : Exemple de tableau de contingence chronologique

Finalement, il existe également des tableaux de données textuelles. Par contre, on exploite rarement les données non numériques. Cependant, ces dernières peuvent être importantes dans le cas de l'aide à la prise de décision. Parmi ces données, on retrouve: les enquêtes qualitatives, les enquêtes à questions ouvertes, les lettres de réclamations ou

de suggestions, la veille technologique et scientifique, la veille concurrentielle et stratégique. [Jambu, 1999]

1.4.2 Les variables traitées

On distingue deux types principaux de variables traitées dans les différents tableaux. Il s'agit des variables quantitatives et des variables qualitatives appelées aussi « variables de catégories ». Les variables quantitatives prennent des valeurs traitables par des opérations arithmétiques comme des différences ou des moyennes. Les variables qualitatives prennent des valeurs qui identifient en fait des catégories. On parle également de modalités de la variable qualitative au lieu de catégories (comme on parle de modalités de réponses à une question).

Tableau II : Variables quantitatives et qualitatives dérivées

Type de variable	Description
chronologiques ou temporelles	possèdent autant d'états, qualitatifs ou quantitatifs, qu'il y a de périodes
logiques ou booléennes	variables qualitatives prenant pour seules valeurs 0 et 1
qualitative à réponses multiples	met en jeu plusieurs modalités possibles
qualitative de rang	met en jeu un classement sur les individus
de préférence	ordonne les réponses d'une variable à réponses multiples
de classe	classifie les valeurs déterminant un intervalle auquel est associée une catégorie

[Jambu, 1999]

Les individus ou les entités (personnes ou objets) quant à eux, appartiennent à une population de référence qui est elle-même définie par des variables de contrôle. On peut de cette façon, cerner un échantillon représentatif. [Jambu, 1999]

Les définitions précédentes, que ce soit en termes de tableaux de données ou des variables elles-mêmes, sont en fait la description des éléments qui sont manipulés lors de la mise en oeuvre d'un entrepôt de données.

1.5 L'utilisation du sol urbain dans les tableaux de données dimensionnels

La contribution majeure à la conceptualisation de l'utilisation du sol a été réalisée par Guttenberg (1959) tel que cité dans Whipple (1976). Il a affirmé que l'utilisation du sol implique généralement cinq facteurs qui entraînent à leur tour les dimensions suivantes:

Tableau III : Les cinq facteurs et dimensions impliqués par l'utilisation du sol

5 facteurs	5 dimensions
un site	le développement général du site
un bâtiment érigé sur le site	l'adaptation du site représentée par des bâtiments ou autres commodités
une activité	l'utilisation actuelle ou le type d'activité pratiquée sur les lieux
une entreprise quelconque qui y est établie	la fonction économique de l'entreprise active
un effet de cet activité	les caractéristiques de l'activité (rythme, domaine, impact matériel)

[Whipple, 1976]

On base ici le développement urbain autour du rayonnement d'une activité économique implantée. À savoir, la structure et la composition des quartiers résidentiels, commerciaux et industriels sur le territoire. Tel que mentionné par Whipple, les définitions « d'activité » et de « fonction » restent à être cernées. Au départ, elles n'étaient pas bien définies par la méthode de Guttenberg. Cependant, elles se rapprochent beaucoup des catégories définies par le recensement fédéral. Le fait de reconnaître l'existence de multiples dimensions et sous-dimensions au sein de l'utilisation du sol est considéré par l'auteur comme un acte de classification. Les relations à faire entre les données doivent être évidentes et donc le système doit être transparent à leur égard. À la base, la structure d'information doit être également propice au stockage et à la manipulation informatique. [Whipple, 1976]

1.5.1 Application de la vision multidimensionnelle à l'information urbaine

Une première approche veut que l'on considère la classification des différentes utilisations du sol urbain suivant deux avenues de traitement. L'une qui se rattache aux fonctions et l'autre aux activités qui en découlent. Prenons la fonction de la vente au détail qui consiste à redistribuer des produits finis aux consommateurs. Elle entraîne des activités incluant le commerce des magasins à rayons pour la vente de toutes sortes de biens. Par exemple, un magasin à rayon peut comporter les fonctions de plancher de vente, d'entrepôt et de siège administratif. Une première tâche consiste en l'établissement de l'échelle d'utilisation du sol (ou l'unité de base) la plus appropriée pour gérer les activités du territoire. On se tourne vers un système de gestion multidimensionnel suivant des principes de classification logiques dont la propriété foncière est l'unité d'enregistrement principale. L'utilisation des établissements, comme unité de base, donne la flexibilité nécessaire à la formation d'ensembles à de plus grandes échelles en fonction de la nature des activités pratiquées et des caractéristiques de l'environnement. Par exemple, des groupements de bâtiments qui rendent possible la détection de relations entre les patrons de localisation des activités urbaines en question peuvent être constitués. [Whipple, 1976]

1.5.1.1 Structure de l'information sur les fonctions urbaines

Regrouper les établissements par types suivant une méthode unidimensionnelle de classification cause un problème. Par exemple, on peut compter plusieurs établissements sur un même pâté de maison. Une même superficie de terrain peut donc contenir plusieurs activités au sein d'une même fonction (ex: la fonction commerciale attribuée à un îlot peut comporter de la vente au détail, de la restauration ou des services). L'arrêt d'une de ces activités, reflété par des chiffres, ne signifie pas nécessairement la réduction de l'espace toujours alloué à cette fonction dans la ville. Pour cette raison, une deuxième approche tente de surmonter ce problème en dupliquant la série de codes liée aux activités. Cette technique se résume en une traditionnelle liste qui ne laisse pas transparaître le détail des informations plus précises reliées à l'activité analysée. Les

sommes des superficies avantagent alors certaines fonctions par rapport aux autres car elles comptent une plus grande diversité d'activités. Si on considère l'occupation du sol comme un exercice de location des lieux, l'information relative à la valeur d'une superficie supportant une occupation particulière est ainsi masquée. [Whipple, 1976]

Voici une liste des établissements urbains et de la superficie couverte par leurs activités suivant une vision unidimensionnelle:

Tableau IV : Liste unidimensionnelle des établissements urbains suivant la superficie couverte

Établissement No.	Description	Superficie couverte	Superficie totale pi.car.
1	alimentation		
	ventes	1200	1200
2	ameublement		
	ventes	18000	
	stationnement	2000	20000
3	bur. Architecte		
	bureau	800	800
4	magasins rayons		
	ventes	150000	
	entreposage	10000	
	bureau	5000	165000
5	vente de souliers		
	ventes	2000	
	entreposage	250	2250
6	tabagie		
	ventes	975	
	manufacture	2000	
	bureau	350	
	entreposage	1250	
	stationnement	600	5175
Superficie de plancher totale			194425

illustration: [Whipple,1976]

1.5.1.2 La vision multidimensionnelle du concept

L'utilisation du sol étant une discipline intégrant plusieurs variables, la vision unidimensionnelle n'est évidemment pas appropriée. Par contre, une matrice à double entrée permet de résoudre ce problème en fournissant des totaux distincts, concernant la superficie pour chaque fonction selon l'activité qui lui est reliée. On peut alors réutiliser les résultats matriciels distincts à d'autres fins d'analyse car on a toujours accès au détail de l'information contenue au sein de chaque fonction. [Whipple, 1976]

Voici les mêmes informations de superficie que celles traitées dans la liste précédente mais cette fois-ci montrée sous le jour d'une matrice à double entrée (remarquons que nous avons accès au détail de l'information en obtenant les totaux de superficie par activité et par fonction):

Tableau V : Matrice de la superficie couverte des fonctions urbaines selon l'activité

Fonction	Activité						Tct. fonct. pi.e.car.
	alimentation	magasin à rayons	vente de souliers	ameublement	tabagie	bur. d'architecte	
Vente au détail	1200	150000	2000	18000	975		172175
Administration		5000			350		5350
Service professionnel						800	800
Entreposage		10000	250		1250		11500
Manufacture					2000		200
Stationnement					2000	600	2600
Total activité pi.car.	1200	165000	2250	20000	5175	800	194425

illustration: [Whipple, 1976]

On doit retenir les morales relatives aux aspects précédents soit: de tenir l'information liée à l'utilisation du sol à l'aide d'un support informatique et aussi qu'il est judicieux de conserver cette dernière sous la forme la plus fragmentée possible pour lui conférer une plus grande flexibilité (lorsqu'on doit réutiliser la même information à plusieurs reprises et en prévision de nouveaux besoins). [Rhind et Hudson, 1980]

Conclusions sur le chapitre :

Après une brève introduction aux applications propices à l'entreposage et au forage de données, les multiples composantes sous forme de données et de tableaux relationnels ont rapidement pris un sens. Ces éléments de manipulation seront le langage d'échange entre « l'utilisateur-concepteur » et le système d'information urbain informatisé. Il faudra, par cet assemblage de tableaux et de variables, traduire et documenter des situations réelles données en fonction du domaine d'application (ici la gestion urbaine), afin d'en automatiser l'ensemble des analyses par l'informatique. Enfin tel sera l'objectif pour l'assistance du système à la prise de décision. Afin de compléter notre exploration, nous verrons dans le prochain chapitre les mécaniques d'assemblage d'un entrepôt de données proprement dit.

Chapitre 2 : L'assemblage d'un entrepôt de données (data warehouse)

Objectifs du chapitre :

L'assemblage d'un entrepôt de données adapté à notre domaine d'application n'est pas une tâche simple. Plusieurs éléments sont à prendre en ligne de compte afin d'être en mesure de se doter d'un outil efficace par la suite (parlons par exemples des données traitées, des bases de données existantes ou des dictionnaires de métadonnées). Le présent chapitre a pour but de définir davantage le rôle d'un entrepôt au sein du forage de données et de détailler plus en profondeur l'assemblage de ses composantes. L'objectif principal étant toujours de concevoir un entrepôt de données sachant répondre aux besoins d'un gestionnaire municipal en matière d'urbanisme, il est important de comprendre la logique derrière l'entrepôt. On veut donc en connaître les concepts technologiques ainsi que les processus informationnels qu'il implique.

Par définition, un entrepôt de données consiste en un stockage intermédiaire des données qui sont elles-mêmes issues d'applications de production. Dans ces dernières, les utilisateurs finaux vont puiser avec des outils de restitution et d'analyse de l'information. En matière d'entrepôt, Bill Inmon nous offre la définition suivante [telle que citée dans **Goglin, 1998, p.27**] : « *Un Data Warehouse est une collection de données thématiques, intégrées, non volatiles et historisées, organisées pour la prise de décisions* ».

Voici une illustration de l'architecture d'un entrepôt de données typique dont nous verrons la signification des composantes :

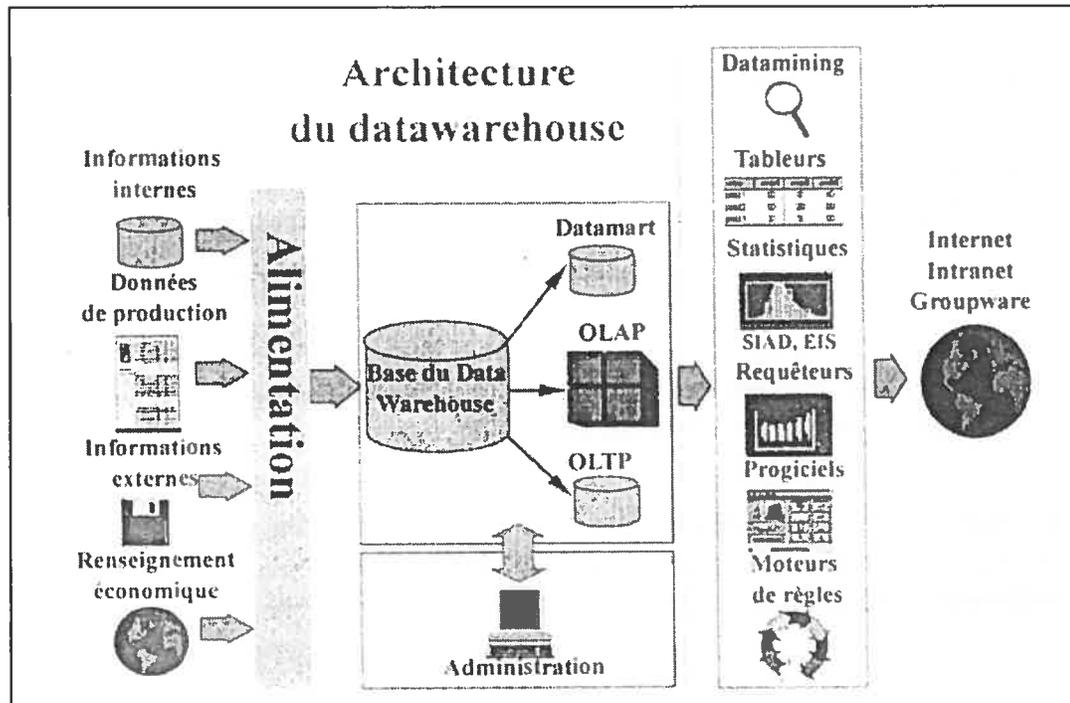


illustration : [Goglin, 1998]

Figure 7: Architecture d'un entrepôt de données

2.1 Les bases de données de l'entrepôt de données

Composantes principales de l'entrepôt, elles renferment les informations extraites des bases de production. En fait, l'entrepôt de données est supporté par une base de données relationnelle, multidimensionnelle ou objet (même si celles-ci sont rarement adaptées à des contextes particuliers). La BD relationnelle présente une organisation de l'information qui existe entre des tables (ex : par la relation entre table des clients et table des factures, on peut avoir la liste des clients facturés). Quant à elle, une BD multidimensionnelle stocke ses données de manière à optimiser l'accès aux informations pour répondre à des requêtes non prévues lors de sa création. [Goglin, 1998]

2.1.1 Types de données constituant les bases de données

Voici un tableau qui résume la description des données composant l'entrepôt de données:

Tableau VI: Types de données constituant l'entrepôt de données

données thématiques	données intégrées	données non volatiles	données historisées
L'objectif principal de l'entrepôt de données est la prise de décision autour des activités majeures de l'organisation ou de l'entreprise. Les informations impliquées seront assemblées par thèmes et sujets.	Cette intégration impliquera une normalisation, une gestion accrue des référentiels communs et de la cohérence, une bonne maîtrise de la sémantique et des règles de gestion qui s'appliquent aux données manipulées.	Afin qu'il soit possible à tout moment de retracer des informations et des décisions prises, les informations qui sont stockées dans l'entrepôt de données sous formes de données ne peuvent être altérées, modifiées ou supprimées.	L'historisation est rendue nécessaire en vue de suivre dans le temps l'évolution des valeurs des indicateurs à analyser (sans tenir compte des problèmes de volumétrie et des capacités de stockage). Un référentiel temporel est donc nécessaire.

[Goglin, 1998]

2.2 Le magasin de données (data mart)

Un magasin de données est en fait une solution départementale d'entrepôt de données qui supporte une partie des données et des fonctions de l'entreprise. On parle ici d'un sous ensemble (fig. 8, p.26) de l'entrepôt qui ne contient que les données relatives à un métier de l'entreprise contrairement à l'entrepôt de données qui contient toutes les données décisionnelles de l'entreprise pour tous les métiers. Donc, dans une administration municipale, le magasin de données « génie civil » contiendra toute les informations nécessaires au service de génie civil. Les informations relatives à ce magasin peuvent bien sûr avoir été générées à l'aide d'autres bases de productions que celles de génie civil uniquement (il y a échange d'information possible entre les magasins de données de l'entrepôt). [Goglin, 1998]

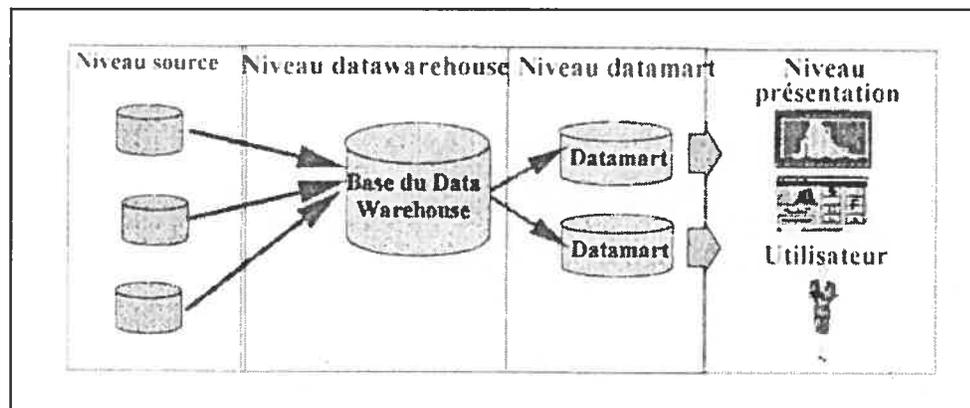


illustration : [Goglin, 1998]

Figure 8 : La nature du magasin de données au sein de l'entrepôt de données

2.3 Le dictionnaire de métadonnées :

Le processus de gestion et d'organisation d'un entrepôt de données mérite un suivi administratif précis afin de finement contrôler le contenu de la base d'informations décisionnelles. Le tout, autant au niveau de son alimentation et de sa mise à jour que de son utilisation. La cohérence logique d'un tel système réparti ne peut être assurée que si ce dernier s'appuie sur un dictionnaire unique qui sait gérer l'ensemble de ses fonctions. Les métadonnées constituent l'ensemble des données qui décrivent des règles ou des procédures attachées aux données. C'est la finalité du système d'information. [Goglin, 1998]

Ce que l'on constitue en fait, par le dictionnaire, est le référentiel pour chaque donnée en considérant son état d'origine, son niveau de granularité, de structure, d'agrégation et finalement sa forme de présentation. Ce référentiel a pour but de décrire toutes les informations nécessaires à la gestion et à l'administration de l'entrepôt de données. Il recèle la description de chacune des étapes de la construction de l'information (en sachant à quelles données de base elles se rattachent). En somme, les métadonnées servent à répondre aux questions suivantes :

Tableau VII:

Questions résolues à l'aide des métadonnées

- | |
|--|
| <ul style="list-style-type: none"> • Comment trouver l'information dont j'ai besoin? • D'où proviennent ces données? • Comment ont elles été agrégées? • Quelles sont les requêtes d'accès disponibles, ont-elles été modifiées et comment? • Comment les définitions des métiers ont-elles évoluées? • De quel historique dispose-t-on? |
|--|

[Goglin, 1998]

Il est alors question de normaliser, d'industrialiser et de piloter les fonctions suivantes : la définition des données, la fabrication des données, le stockage des données, l'accès aux données et la présentation des données. [Goglin, 1998]

Dans l'image suivante, voyons de façon plus détaillée le lien entre le dictionnaire de métadonnées et les étapes de traitement de l'information dans un entrepôt de données :

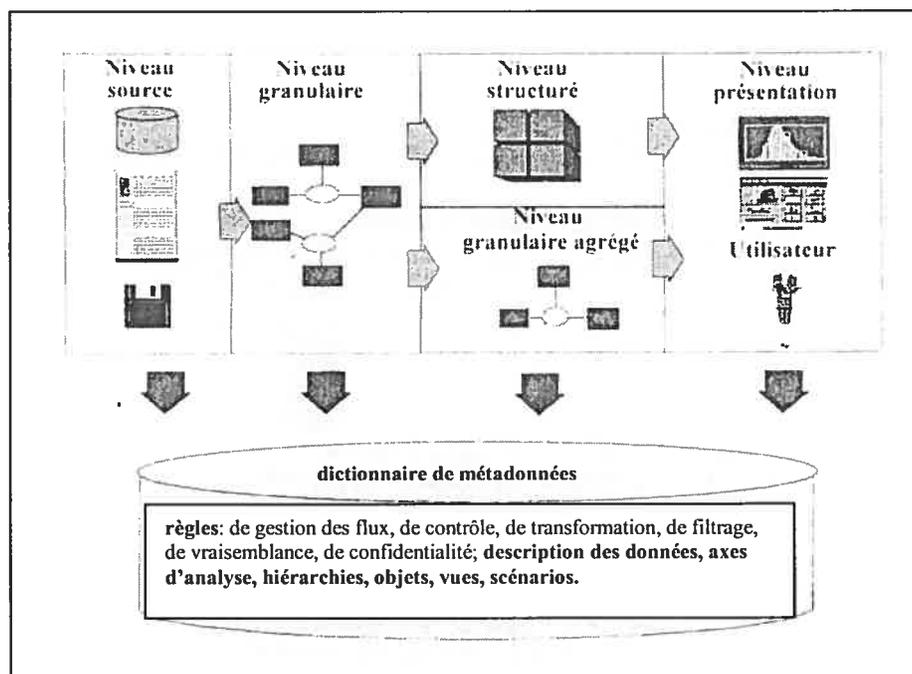


illustration : [Goglin, 1998]

Figure 9 : Liens entre dictionnaire de métadonnées et processus de traitement de l'information

2.4 L'alimentation de l'entrepôt de données

Afin d'alimenter les bases décisionnelles, on doit extraire la matière vive des bases de données opérationnelles. Pour ce faire, les données intéressantes doivent être préalablement identifiées. Des agrégats sont alors définis pour mettre ces dernières en relation. On essaie donc de limiter les redondances en reliant les données complémentaires pour les enrichir et bénéficier ainsi de multiples points d'entrée. L'extraction vers l'entrepôt de données est alors possible. [Goglin, 1998]

2.4.1 L'extraction et la transformation des données :

L'opération d'extraction implique des données hétérogènes (souvent issues de plusieurs SGDB différents), diffuses (logées dans plusieurs environnements matériels différents) et complexes (plusieurs modèles logiques et physiques orientés vers des traitements d'échange). Dans cette étape, il faut bien considérer le facteur d'évolution du système d'information. Les formats de données déjà hétérogènes sont appelés à varier dans le temps par le changement des systèmes de production. Ces derniers auront tendance à intégrer des données supplémentaires autant pour prendre en compte des nouveaux indicateurs que pour obtenir une granularité¹ plus fine de l'information. [Goglin, 1998]

La transformation des données extraites consiste en la deuxième étape du processus d'alimentation. C'est l'obtention d'un ensemble homogène de données, les rendant ainsi comparables et mathématiquement traitables. Par cette affirmation, on voit que l'entrepôt de données est un système intégré qui ne fait pas que juxtaposer des données hétéroclites.

¹ L'échelle de partitionnement absolue de l'information pour l'ensemble des thèmes abordés par l'Entrepôt de Données. [Kimball, 2000]

Tableau VIII : Opérations de transformation des données

L'épuration des données extraites	Cette démarche, qui consiste en le filtrage des données, éliminera les valeurs manquantes ou les valeurs n'ayant pas de signification particulières.
Le dédoubleage des données agrégées	Entraîné par l'agrégation de bases de données distinctes, il interviendra sur chacun des niveaux d'information impliqués pour garantir la cohérence et ainsi éliminer les redondances au sein du data warehouse.
Le restructuration (format) des données	C'est une opération qui vise à normaliser les informations et à préparer leur intégration dans le système cible en les convertissant au format cible (ex : dissection du champ adresse en nom de rue, numéro civique, code postal, ville).
Synchronisation des données	Elle garantit la cohérence entre les agrégats du data warehouse et, du même coup, la pertinence des valeurs calculées.

La difficulté principale réside donc dans la bonne compréhension de la donnée et des règles qui lui sont associées (implicites ou explicites). Certaines données devront alors être isolées pour être traitées à part (définition de modalités, correction ou réaffectation de leur sémantique véritable). Ces opérations peuvent être appliquées à la source et tout au long de l'existence de la BD. [Goglin, 1998]

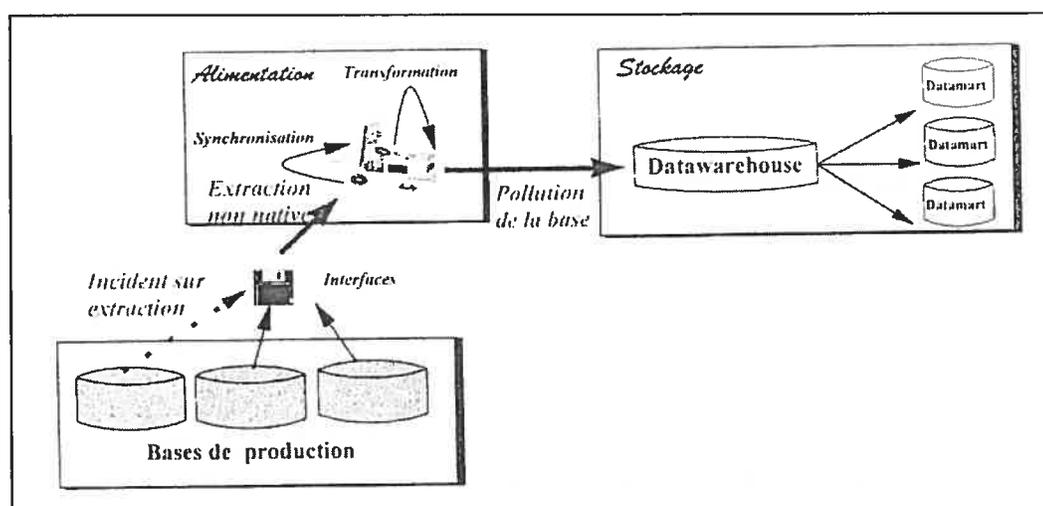


illustration : [Goglin, 1998]

Figure 10 : L'acheminement des données vers l'entrepôt de données

2.5 Le concept du multidimensionnel et la technologie OLAP (On-Line Analytical Processing) au sein des bases de données

Les modèles multidimensionnels organisent les données en les classant suivant les dimensions pertinentes à l'entreprise contrairement aux modèles relationnels qui rangent l'information dans des tableaux matriciels. Au lieu d'ajouter des tables supplémentaires, OLAP ajoute des dimensions (champs de données, variables) supplémentaires. Beaucoup de produits OLAP utilisent des bases de données multidimensionnelles. D'autres font appel à des bases de données relationnelles associées à un schéma en étoile où on stocke les données dans une « table de faits ». On lie à cette dernière des tables d'attributs (tableaux Individus-Variables) par clés de jointures pour chacune des dimensions traitées. [Goglin, 1998]

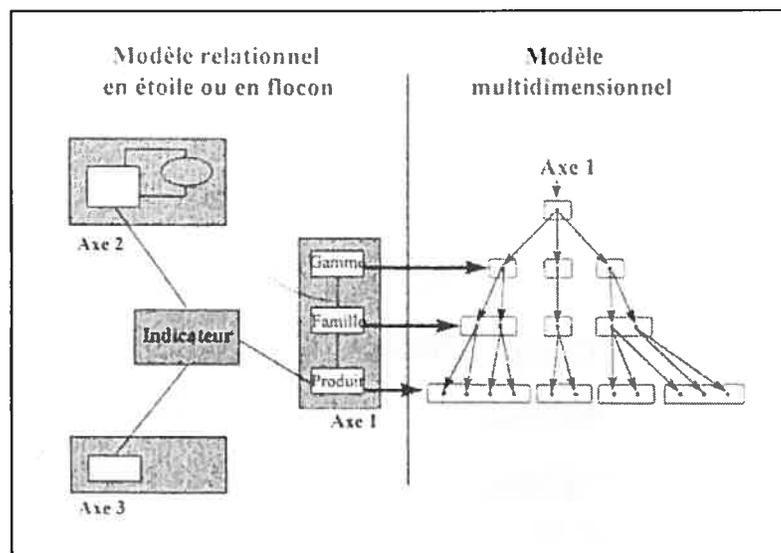


illustration : [Goglin, 1998]

Figure 11 : Modèle de données hiérarchique relationnel versus multidimensionnel

Tel que cité dans Goglin en 1998, le créateur des bases de données E.F.Codd², déclare en 1993 que les systèmes de gestion de bases de données relationnelles n'ont jamais été conçus pour supporter des fonctions d'analyse, de synthèse et de consolidation qui définissent cette analyse multidimensionnelle (mieux représentée sous forme de cube). La base de données traditionnelle permet aux utilisateurs des visions en deux dimensions

² Créateur des bases de données relationnelles dans les années 1970. [Goglin, 1998]

seulement (ex : répartition des produits par région). Le multidimensionnel permet une analyse intégrant plusieurs dimensions (ex : ventes de produits par régions, par couleur, par taille, dans le temps). Cette façon de faire permet de dégager plus rapidement des tendances futures. [Goglin, 1998]

Voici une image de cette équivalence analytique sous forme de cube :

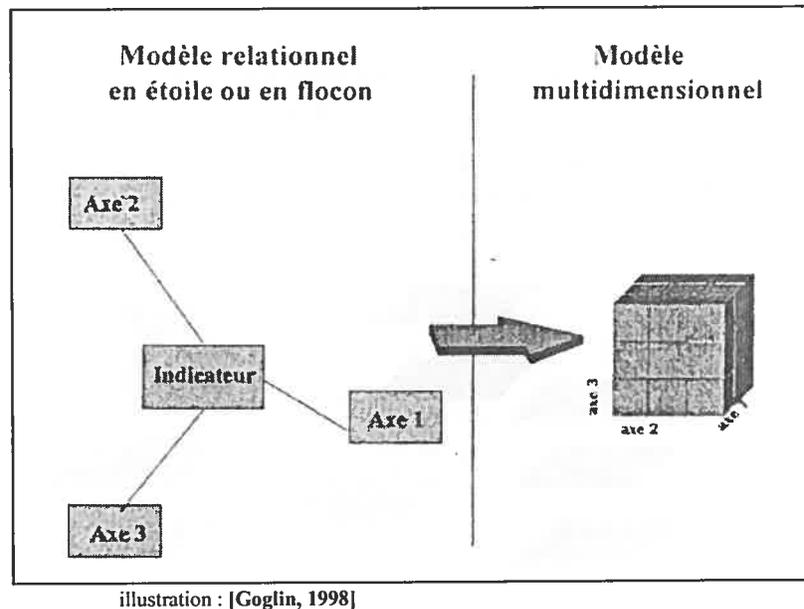


Figure 12 : Schématisation des modes d'analyse relationnels versus multidimensionnels

2.5.1 Des technologies issues de l'OLAP

Progressivement, en fonction des besoins du marché, deux approches ont pris forme autour de la technique OLAP. La première, « MOLAP » s'appuie sur une structure de stockage monolithique sous forme de cube. « Multidimensional OLAP » a été conçue uniquement pour l'analyse multidimensionnelle, avec un mode de stockage optimisé par rapport à des chemins d'accès qui sont prédéfinis. Elle répond en fait aux douze règles OLAP³.

³ Voir l'annexe 3.

La deuxième approche est constituée des bases « ROLAP » (Relational OLAP) qui sont en fait des bases de données OLAP supportant une architecture relationnelle. En fait, un tel outil est capable de simuler le comportement d'un SGBD multidimensionnel en exploitant un SGBD Relationnel classique. Là se trouve tout l'intérêt du système. Il n'est pas nécessaire d'acquérir un nouvel appareillage pour effectuer des analyses multidimensionnelles de l'information. De plus, il est facile de lancer des requêtes complexes directement sur l'entrepôt de données.

Les principales différences entre les deux technologies consistent en les méthodes d'accès et la manière d'agrèger les données utilisées. En voici un portrait :

Tableau IX : Différences fonctionnelles entre les technologies MOLAP et ROLAP

MOLAP	ROLAP
Méthode d'accès aux données: séquentielle indexée (toute donnée est accessible tant que l'index est cohérent et synchronisé).	Méthode d'accès aux données: comme les SGBDR (jointures, index, tri).
Traitement des données : est plus rapide tant que le volume du cube à manipuler est raisonnable.	Traitement des données : est plus lent mais peut supporter de très gros volumes de données en exploitant une technique de partitionnement de données ou d'accès en parallèle.
Administration : plus rapide car il agrège tout par défaut (par contre, rend la validité des agrégations incertaine).	Administration : plus lourd à administrer car il n'agrège rien et qu'il tire parti des agrégats que s'ils existent (ceci implique la création explicite des agrégats possibles).

[Goglin, 1998]

Les principes d'agrégation implicites MOLAP pénalisent lors de la phase de chargement de la base et ce autant au niveau des performances que du volume de données à traiter (gonflement à cause des agrégations par défaut). C'est pour cette raison qu'il est impossible d'appliquer cette technique sur de gros volumes de données brutes. En s'appuyant sur les références du marché, soit les bases relationnelles, ROLAP tire parti de leur évolution (adaptation aux architectures complexes, aux nouvelles extensions des langages de requête). MOLAP présente tout de même des fonctionnalités plus avancées mieux adaptées à la mise à jour, à la simulation et à la prévision. Cependant, MOLAP est

incompatible avec d'autres modes d'accès aux données. Deux bases de données doivent donc y cohabiter pour assurer l'application des règles OLAP. [Goglin, 1998]

2.6 Le magasin de données spécialisé ou l'entrepôt de données centralisé

Rappelons en premier lieu que l'entrepôt de données contient des informations qui concernent tous les services d'une entreprise (exemple pour une municipalité : police, incendie, administration, techniques/inspection, voirie/travaux publics). Le magasin de données, quant à lui, gère les informations propres à un des métiers de l'entreprise. Il traite donc un problème spécifique et rejoint un nombre restreint d'utilisateurs finaux. Soulignons que le magasin de données se situe à un niveau départemental et ne peut être qu'un sous-ensemble de l'entrepôt de données. Une problématique d'intégration d'ensemble émerge alors lorsqu'on se demande si on construit un petit entrepôt (suivant des critères de prix, d'architecture ou pour des questions de logique) auquel peut être éventuellement assimilé un magasin. On a le problème inverse si on démarre avec un magasin de données sans perdre de vue la mise en oeuvre du contenant que sera l'entrepôt de données entier. Premièrement, l'architecture d'un magasin de données est peu évolutive et nuira au développement futur du système suivant une augmentation des utilisateurs. Ensuite, viendront l'incohérence et la redondance générées par la multiplication des magasins de données sémantiquement différents.

Tableau X : Résumé des différences techniques entre l'entrepôt et le magasin de données

Caractéristiques	Datawarehouse	Datamart
Cible utilisateur	toute l'entreprise	un département
Implication du service informatique	élevée	faible ou moyen
Base de données d'entreprise	SQL type serveur	SQL milieu de gamme Bases multidimensionnelles OLAP
Modèle de données	à l'échelle de l'entreprise	à l'échelle du département
Champ applicatif	multi-sujet	quelques sujets
Sources de données	multiples	quelques unes
Stockage	une base de données	plusieurs bases distribuées
Taille courante	centaines de Go et plus	une ou deux dizaines de Go
Temps de mise en place	9 à 18 mois	6 à 12 mois

illustration : [Goglin, 1998]

Il est important, dès le départ de prévoir l'architecture et le référentiel global du futur entrepôt dès la mise en place du premier magasin de données. Comme nous l'avons vu, Nous pouvons être confrontés à deux situations lors de la mise en place d'un projet d'entrepôt de données. On peut vouloir démarrer avec un entrepôt à portée réduite. On peut aussi être tenté de commencer par un magasin de données qu'on enrichira par la suite. Notons qu'il devient plus difficile de consolider les informations provenant de divers magasins en leur donnant une mécanique d'opération transversale. L'auteur propose une solution qui est la suivante : le remède du magasin de données est l'entrepôt de données. Il faut donc commencer par mettre en place un magasin qui présente une architecture à deux niveaux soit : un niveau de structuration des données départementales et un niveau intermédiaire avec un schéma de conception qui permet la transversalité recherchée avec de futurs magasins de données. [Goglin, 1998]

Conclusions sur le chapitre :

Comme nous avons pu le voir, les multiples composantes et méthodes de conception, d'alimentation, d'entretien et de documentation d'un entrepôt ou magasin de données donnent lieu à la mise en place d'une réelle stratégie d'assemblage. Sans quoi, nous pourrions nous retrouver avec un amas de données certes relatives à un domaine d'application précis mais totalement incohérentes. Il en résulterait un système d'information dysfonctionnel complètement inutile lorsqu'en situation de prise de décision. Les professionnels du domaine seraient alors démunis. Dans le but d'éviter cette situation, les prochains chapitres illustrerons les méthodes actuelles de structuration de l'information au sein des systèmes d'information à référence spatiale et des bases de données urbaines. Cet outil de travail privilégié par les municipalités présente certains avantages et inconvénients qui seront présentés.

Chapitre 3: Gestion du territoire urbain par des outils informatiques de traitement de l'information

Objectifs du chapitre :

Dans ce chapitre, nous verrons à mettre en contexte l'environnement procédural des outils de gestion informatique du territoire urbain dans lequel viendra s'inscrire la conception d'un magasin ou d'un entrepôt de données. Cette mise en contexte vise à illustrer les méthodes courantes employées par les administrations municipales pour supporter l'information spatiale relative à l'utilisation du sol au sein des SIURS.

3.1 Principes de base des SIURS

Ces systèmes sont valables pour la gestion de l'ensemble de l'information spatiale liée à un territoire donné. Ils rendent compte de l'évolution d'un territoire ainsi que des interventions qui sont effectuées sur ce dernier. On leur porte un intérêt lorsqu'il s'agit de contrôler de l'information rattachée à un environnement urbain. Leur fonctionnement réutilise toujours les mêmes principes mathématiques, trigonométriques et géométriques. Seulement, leur efficacité est grandement amplifiée par les moyens technologiques et informatiques que l'on connaît aujourd'hui.

Voici deux définitions générales des systèmes d'information urbaine à référence spatiale :

“Geographic information systems- A system of hardware, software, data and people, organizations and institutional arrangements for collecting, storing, analyzing and disseminating information about areas of the earth. [(Ducker and Kejerne 1989, pp. 7-8) Chrisman, 1997]

“Essentially, all these disciplines are attempting the same sort of operation-namely to develop a powerful set of tools for collecting, sorting, retrieving at will, transforming and displaying spatial data from the real world for a particular set of purposes.” [(Burrough, p.6)HuxHold, 1991]*

Les systèmes d'information urbaine à référence spatiale sont donc des outils informatiques de choix pour la gestion des informations relatives à l'utilisation du sol. La raison d'être de ces systèmes consiste essentiellement en la coordination optimale de l'acquisition et de l'utilisation des données urbaines en fonction des besoins locaux. Peu importe les objectifs d'une municipalité, les composantes fonctionnelles de tels systèmes restent les mêmes: un processus d'acquisition et de mise à jour des données, une base de données, une base géographique à grande échelle, un processus de traitement des données, un processus de communication de l'information et un processus de contrôle.

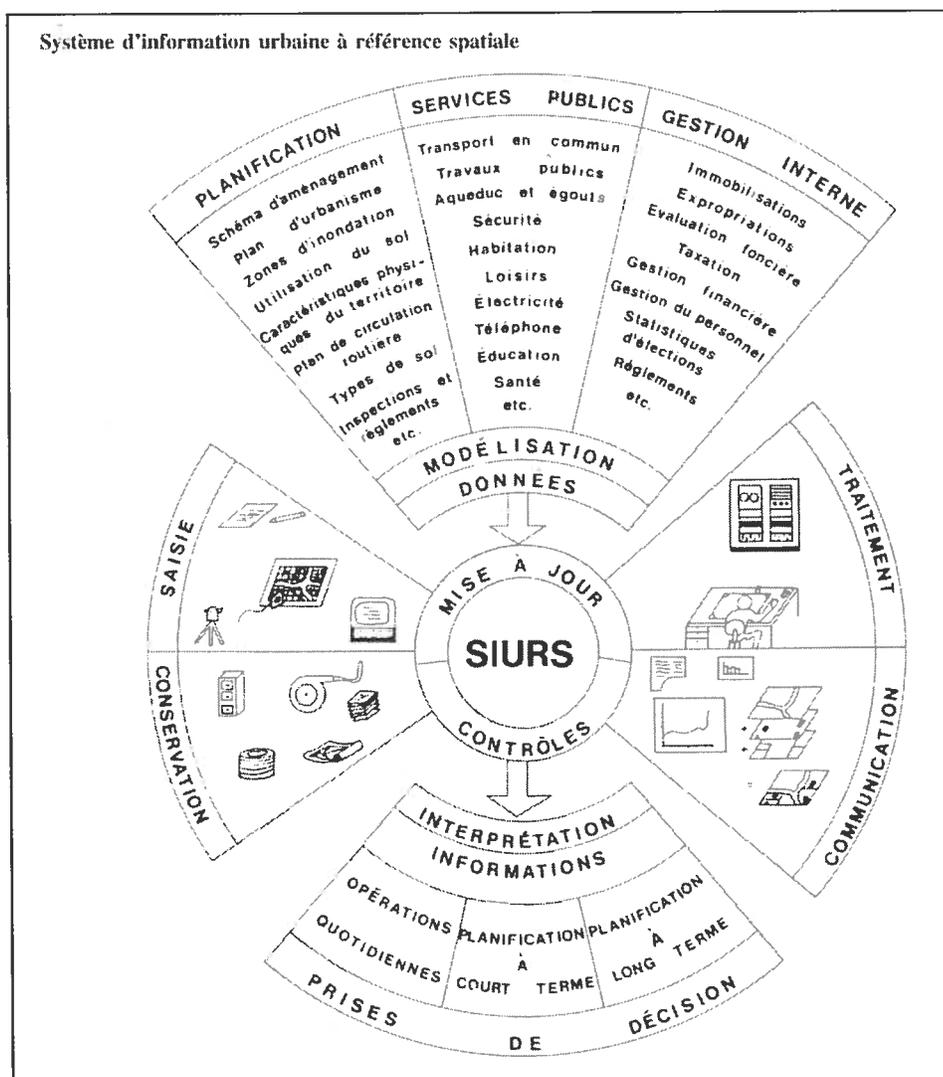


illustration : [MAM, 1984]

Figure 13 : Le cheminement de l'information dans un SIURS

Enfin, les SIURS consistent en une organisation d'actions par lesquelles les humains mesurent l'évolution de phénomènes géographiques⁴ urbains pour, par la suite, en créer un modèle sous forme de bases de données informatiques. L'analyse de ces données laissera place à des relations entre différents problèmes spatiaux soulevés. Ainsi, les représentations cartographiques et les modèles thématiques réalisés à partir des bases de données établies permettront de produire de nouvelles informations sur différents phénomènes (ex : la réalisation d'une étude de desserte pour l'implantation d'une nouvelle bibliothèque). Ce processus permet donc de créer des liens entre les différents phénomènes géographiques étudiés pour entraîner la découverte de solutions aux problèmes rencontrés. [Chrisman, 1997]

Voici l'exemple d'un modèle théorique, propre aux SIURS, des relations entre des phénomènes géographiques d'un territoire par une organisation thématique hiérarchique :

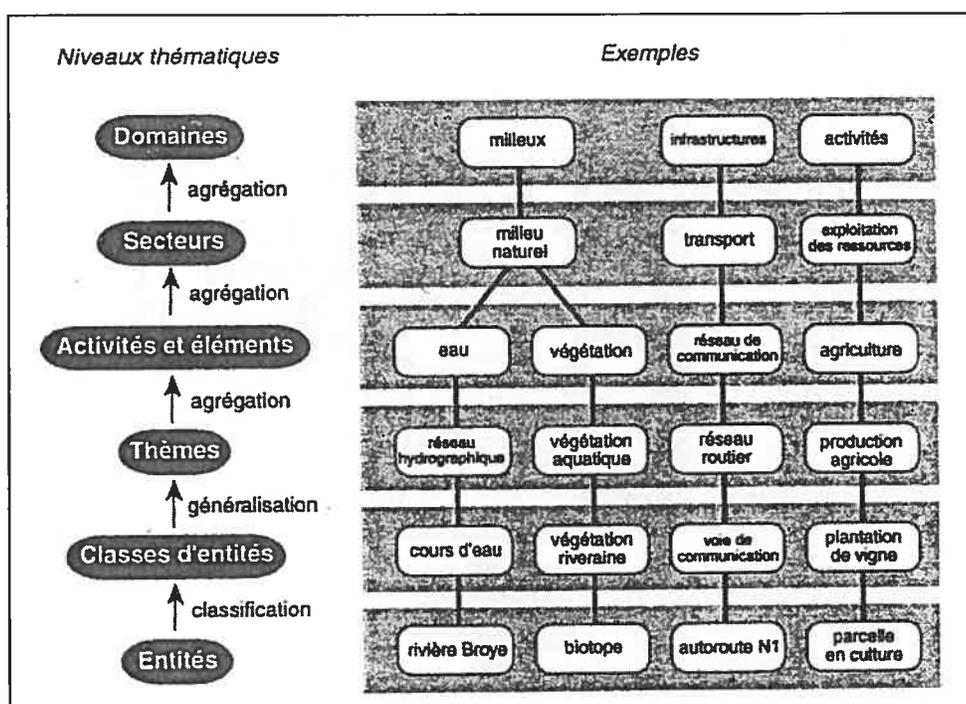


illustration : [Prélaç-Droux, 1995] dans [de Sède, Thériault, 1996]

Figure 14 : Modèle théorique des relations thématiques dans un SIURS

⁴ Activité enregistrable dont on est en mesure de surveiller l'évolution sur un territoire donné. [Rind Hudson, 1980]

3.2 L'intégration des données urbaines par un SIURS

3.2.1 L'organisation des données contenues

Rappelons la première fonction d'un SIURS qui consiste en fournir l'information demandée à celui qui la requiert. Pour y parvenir, il faut satisfaire certaines exigences concernant la duplication des données, leur polyvalence entre utilisateurs, et leur compatibilité entre elles. Ceci amène les auteurs à parler de l'intégration des données à l'intérieur de l'administration municipale. On vise principalement l'intégration des données par l'établissement de liens entre elles, par exemple, en utilisant un géocodage ou un index de référence par adresses civiques. Le Ministère des Affaires Municipales affirme que c'est d'une intégration logique et non physique dont il est question. La distinction entre les deux est simple. La première mise sur la correspondance optimale des identificateurs associés aux informations pour les relier alors que la deuxième mise sur un support matériel informatique toujours plus performant. L'intégration des données à l'intérieur des SIURS se fait présentement à l'aide d'une base géographique urbaine (BGU) commune à tous les services municipaux d'une administration. Ceci ne signifie pas nécessairement que l'on doit entreposer des données dans une base de données centralisée d'envergure. [MAM, 1984]

3.2.2 L'utilisation d'une base géographique urbaine à grande échelle

Beaucoup d'informations convergent vers les propriétés foncières. Les services de génie et des travaux publics demandent un outil de précision pour mener à bien leurs opérations. Il est donc nécessaire d'appuyer le SIURS sur une base géographique à grande échelle. Les échelles de travail considérées varient entre 1:500 et 1:5000 tout dépendant du degré de précision et de la quantité d'information manipulée. Par définition, une base géographique serait un ensemble de cartes comprenant des éléments communs nécessaires à la localisation spatiale des données d'un SIURS. Selon les besoins locaux, la forme de cette base peut varier suivant son contenu, son échelle et sa précision. Elle peut être informatisée ou non et provenir de cartes existantes, ou d'une nouvelle couverture cartographique. Dans le milieu municipal, l'utilisation de la base

géographique varie également selon la nécessité d'illustrer sur des cartes superposables, l'activité immobilière, le réseau d'infrastructures ou autres et ce, en fonction de la spécialisation des tâches dans l'administration municipale. Donc, pendant que certaines municipalités se servent d'une cartographie topographique⁵ (fig.15) très précise associée au réseau géodésique (fig.15) provincial, d'autres utilisent une simple carte de morcellement foncier (fig.15). [MAM, 1984]

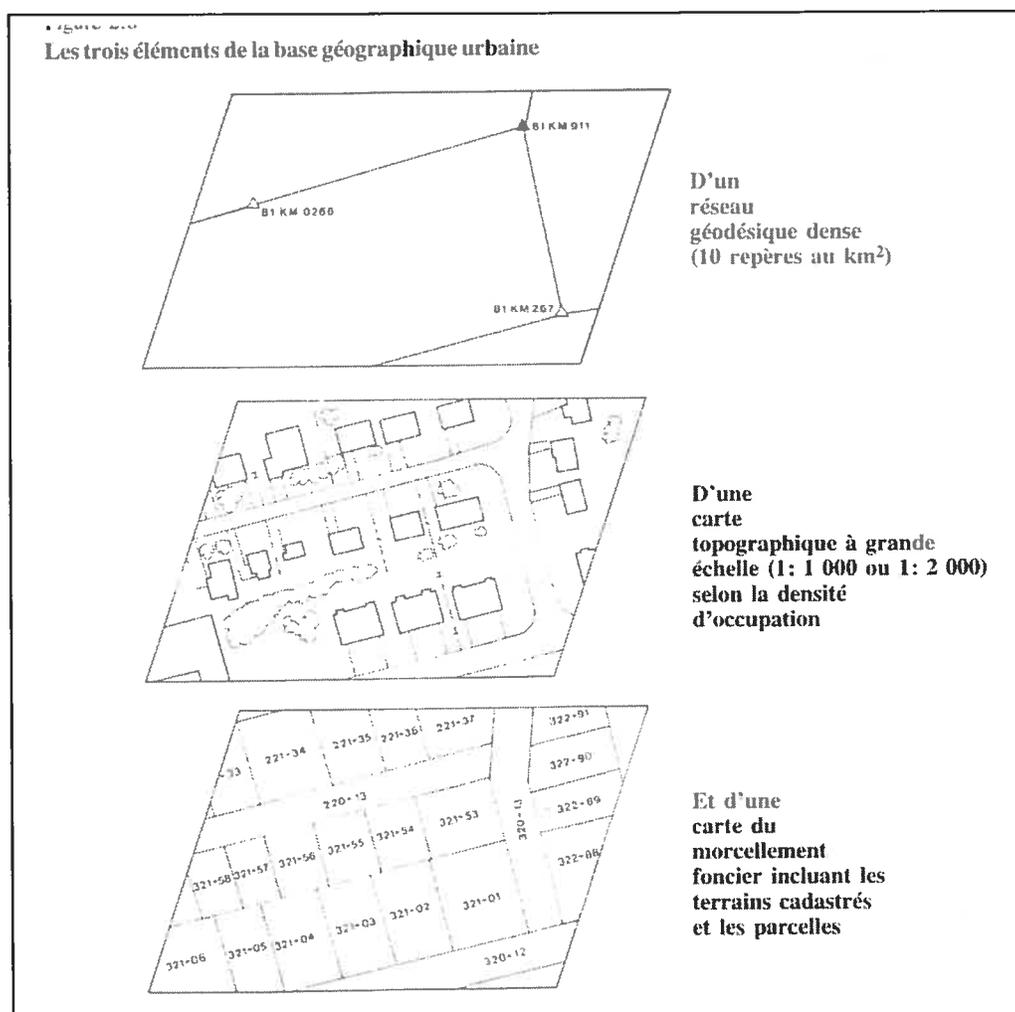


illustration : [MAM,1984]

Figure 15 : Éléments de base constituant une BGU

⁵ Qui tient compte du relief d'un terrain (par cotes de nivellement). [Le Petit Robert]

3.2.3 Caractéristiques d'opération d'un SIURS

Un système d'information spatiale à grande échelle est conçu pour servir autant les institutions que les individus. Ses caractéristiques distinctives sont les suivantes:

Tableau XI :

Caractéristiques du système d'information spatiale à grande échelle
1) Son unité spatiale de base servant à l'organisation est la propriété foncière;
2) Il met en relation une série d'enregistrements propres à un espace découpé en parcelles (mode de tenure, la valeur et l'utilisation);
3) Il est toujours le plus complet possible en termes de superficie de territoire couvert et de la validité des informations;
4) Il constitue un moyen efficace d'accéder à l'information.

[Dale, McLaughlin, 1989]

Le mécanisme de relations entre toutes ces tranches d'information se traduit par un empilage de couches. Ces couches regroupent des données polygonales d'une part et textuelles d'autre part. On y applique ensuite les limites du cadastre (fig.16, p.41), les zones reliées au type d'utilisation du sol, les infrastructures, les informations socio-économiques, les informations qui ont trait à l'environnement. [Dale, McLaughlin, 1989]

Dans le cas d'un système d'information géographique informatisé, on peut distinguer et contrôler chacune des couches d'information suivant le logiciel utilisé. La figure suivante nous montre à gauche une interface affichant l'information cartographique accompagnée d'une table de données. Cette table en question est issue de la base de données alphanumériques qui documente les objets graphiques montrés par des pointeurs référencés. La figure de droite constitue une schématisation de la superposition des couches d'information emmagasinées dans le SIG. La quantité d'information décrite peut alors entraîner le croisement de plusieurs bases de données qu'il faudra coordonner de sorte à les rendre complémentaires

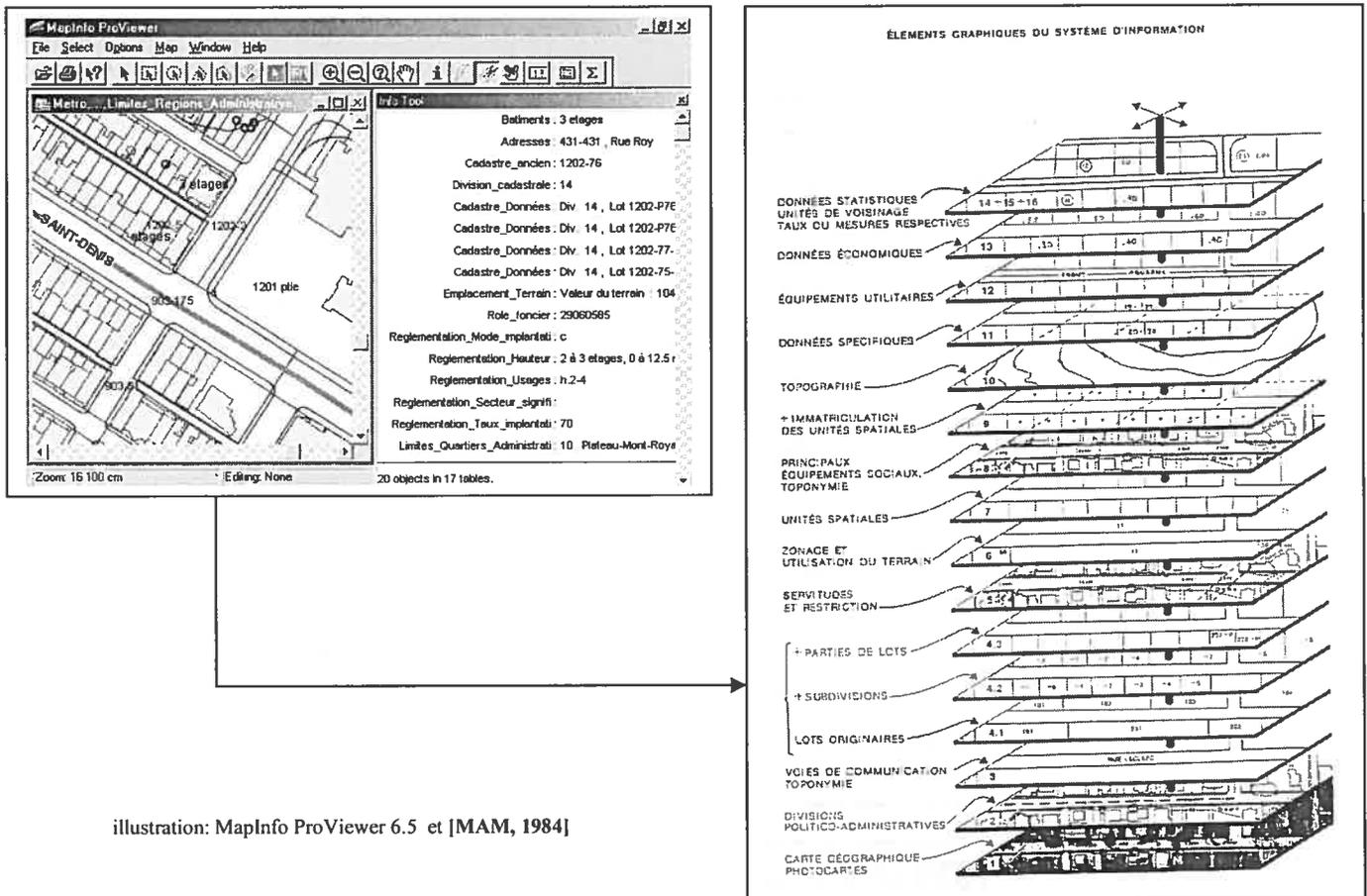


illustration: MapInfo ProViewer 6.5 et [MAM, 1984]

Figure 16 : Démonstration des couches d'information documentées dans un SIG informatisé

3.3 Composition du SIURS : L'information supportée concernant un site

3.3.1 Les données urbaines

L'élargissement des champs d'action municipaux ont donné place à un besoin grandissant en information chez les gouvernements locaux. Il serait ambitieux de vouloir dresser une liste exhaustive de toutes les données qui doivent être incluses dans un système d'information de la sorte et aussi de calibrer leur importance. Les besoins en la matière

peuvent varier d'un endroit à l'autre et ne peuvent être identifiés que par les gestionnaires municipaux eux-mêmes.

Tableau XII : Les trois grandes catégories de données urbaines

La planification	La gestion interne	Les services publics
comprenant des thèmes comme: le développement industriel, le contrôle de la pollution, la gestion du plan d'urbanisme, le schéma d'aménagement régional, le recensement municipal et les inventaires (circulation routière, le couvert végétal, les espaces agricoles, les types de sols, les zones inondables, l'occupation du sol et l'utilisation du sol).	comprenant des thèmes comme: l'évaluation foncière, la matrice graphique, le cadastre, les immobilisations, les expropriations, les servitudes, les élections, les règlements, les permis, les lois, les inspections, la taxation.	<p>-municipaux: comprenant des thèmes comme : la disposition des ordures, les services d'égouts et d'aqueduc, le transport en commun, l'éclairage des rues, la circulation, les loisirs, la culture, la sécurité publique, l'habitation, les travaux publiques, les parcs, etc.</p> <p>-extérieurs: comprenant des thèmes comme la distribution de l'électricité, la distribution du gaz, le téléphone, le câble télé, le transport scolaire l'éducation, la santé.</p>

[Rhind et Hudson, 1980]

Les informations traitées par un système de cadastre informatisé doivent savoir combler les besoins des utilisateurs. Pour notre part, l'information relative à un site urbain nous intéresserait plus particulièrement.

Référons –nous à une grille qui décrit les thèmes ou les champs d'information pouvant être abordés au sein d'un système d'information urbaine :

Tableau XIII : L'information reliée à un site urbain

Thème	Sous-thèmes possibles
La localisation/ le lieu	la topographie, les caractéristiques du sol, la surface du sol utilisable, les marges de recul à respecter, le paysagement, les lieux d'approvisionnement à proximité, les percées visuelles, les accès aux rues et ruelles, les accès aux cours d'eau et aux voies ferrées, les utilités disponibles, la distance avec les marchés, les nuisances voisines, le zonage.
Les dimensions des bâtiments	la surface de plancher, la superficie totale de plancher, la surface louable, le volume du bâtiment, la hauteur du bâtiment, la hauteur du plafond, le dégagement au plafond, le nombre d'étage, le nombre d'appartements.
Le design des bâtiments	l'utilisation à en faire, le style architectural, la forme du bâtiment, le type de toit, la hauteur des étages.
La forme des bâtiments	la surface de plancher (ratio), le nombre de coins.
La qualité de la construction	qualité des matériaux, l'architecture, l'entrepreneur.
Les matériaux de construction	fondations, la structure, les planchers, les murs (intérieurs et extérieurs), les plafonds, le toit.
Autres points relatifs aux bâtiments	le nombre de pièces par type, le chauffage la ventilation, l'air conditionné, la plomberie, les détecteurs de fumée, les rénovations, les galeries et patios, piscines, abris pour l'auto, ascenseur, l'électricité.
L'âge et l'étendue de la dépréciation	l'âge chronologique, l'âge apparent, la valeur sur le marché, la proportion des éléments et parties valables, la condition, les possibilités de modification et de rénovation des bâtiments.

[Dale, McLaughlin, 1989]

On peut également rattacher des informations sur les droits de propriété, les restrictions sur l'usage et le développement du sol, la valeur du sol et l'évaluation foncière, l'utilisation du sol rurale et urbaine, l'information du recensement sur la population, l'information administrative, l'histoire et le patrimoine, la géologie et la géophysique du

terrain, la végétation, la faune, l'hydrologie, le climat, la pollution vs santé vs sécurité, l'industrie et l'emploi, le logement, le transport, les égouts et l'aqueduc, les infrastructures énergétiques et les services d'urgence. Ce n'est pas tout. Nous n'avons qu'à penser aux plans d'ingénierie, aux compilations cadastrales et aux émissions de permis pour réaliser l'ampleur de la quantité de données possibles à traiter. [Dale, McLaughlin, 1989]

3.4 L'informatisation de l'information dans le contexte administratif urbain

3.4.1 Les types de données traitées et leur gestion

La vérification, le classement et le tri de l'information en constituent sa gestion. À l'aide de systèmes informatiques, l'entrée de données est alors une étape critique du processus. Le format des données devrait permettre leur analyse et leur manipulation afin de produire tout type d'information désiré. [Rhind et Hudson, 1980]

Aujourd'hui, avec les systèmes d'informations à référence spatiale (SIURS), trois catégories principales de formats de données coexistent:

- **alphanumériques**: comme des listes de propriétaires ou des caractéristiques des terrains;
- **analogiques graphiques**: cartes, photo aériennes ou images satellites;
- **spatiales numériques**: format vectoriel⁶ ou raster⁷.

3.4.2 Relations à établir entre les données urbaines par l'encodage

3.4.2.1 Selon un identificateur nominal

On peut localiser, sur la base géographique, des données urbaines à l'aide de points, de polygones ou de pixels⁸. Cependant, le repérage à l'aide d'un identificateur nominal est

⁶ Par la définition de vecteurs. ⁷ Par la reconnaissance de pixels.

⁸ La plus petite surface constitutive d'une image numérisée. [Le Petit Robert]

beaucoup plus répandu pour retracer et extraire des données numérisées. Parmi ceux-ci on note l'adresse civique, le nom d'un bâtiment, le numéro cadastral et les numéros de quartier. Cette méthode d'identification de lien est certainement plus facile à appliquer mais elle se prête moins à la manipulation graphique de données relatives au territoire. Un lien pointeur doit encore être établi entre ces localisateurs nominaux et leur coordonnées géographiques. [MAM, 1984]

3.4.2.2 Selon un identificateur à référence spatiale

Une autre méthode consiste en l'utilisation du même identificateur à référence spatiale pour tous les dossiers qui le permettent. Le Ministère des Affaires Municipales avait proposer un formulaire de relevé socio-économique. La fiche comprenait le localisateur traditionnel qu'est l'adresse civique mais elle contient, en plus, le matricule spatial de l'unité d'évaluation (obtenu à l'aide des coordonnées centroïdes⁹ de chaque propriété). Les données issues d'un tel formulaire peuvent être facilement liées au rôle municipal d'évaluation qui utilise le même identificateur pour chaque propriété. On remarque donc une ouverture des données à la cartographie automatisée tout en réduisant les redondances existantes auparavant. [MAM, 1984]

3.4.2.3 Selon un identificateur géographique

Une troisième technique de liaison des données utilise les coordonnées géographiques. La méthode consiste à utiliser les coordonnées des sommets du polygone qui décrit un secteur. En appliquant des formules mathématiques aux coordonnées des centroïdes de lot, on ne retient que les unités comprises dans ledit secteur. Cette façon de procéder minimise la quantité de données à inscrire dans chaque dossier et facilite grandement la mise à jour des relations de voisinage entre chaque propriété constituant le secteur choisi. La base de données sur laquelle on applique ces méthodes est référée spatialement de façon très rapide et à peu de frais. Par contre, le résultat n'est pas parfait car on retrouve des duplications de données et une certaine lenteur d'exécution. [MAM, 1984]

⁹ Point central d'une figure géométrique plane ou d'un volume.

3.5 La codification de l'information géographique urbaine au sein du SIURS

Opération complémentaire à la saisie de données, la codification sert à préparer les données à être intégrées au système. La fonction principale d'un code est de faciliter la manipulation des données. Un code sert à identifier une donnée de façon unique tout en lui conférant une signification particulière. Cette façon de faire nécessite le respect et l'utilisation de normes à l'intérieur du système. Une codification doit satisfaire aux besoins de l'utilisateur, être assez souple pour s'adapter aux nouveaux besoins et comprendre des catégories simples à identifier. Par contre, des codes trop longs sont susceptibles de causer des erreurs. [MAM, 1984]

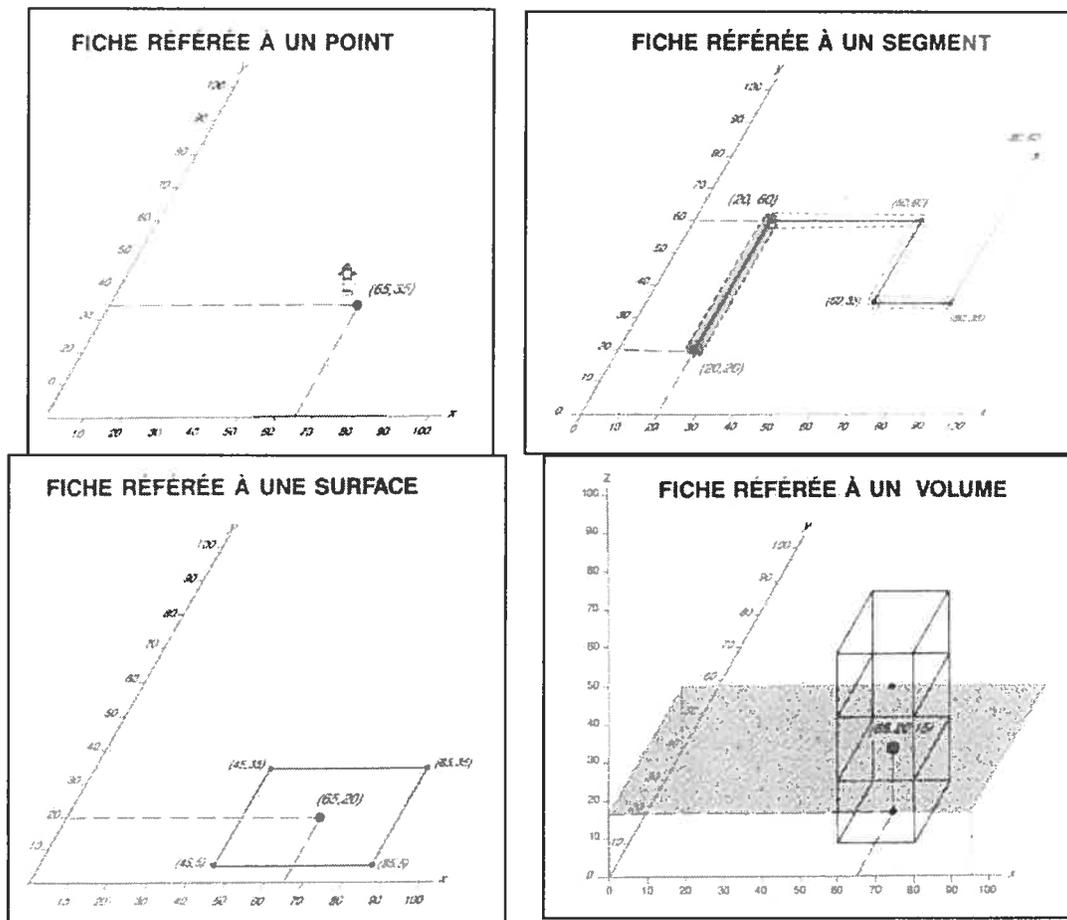
3.5.1 Types de codifications municipales les plus courantes

Tableau XIV : Types de codifications municipales

séquentielle	par blocs	mnémonique	par couleurs
Attribution de lettres ou de numéros consécutifs (ex: pour une classe de bâtiments constituée de 9 types –1 à 9- dans le rôle d'évaluation foncière).	Blocs de caractères ayant différentes significations (ex: le numéro de téléphone comportant le bloc régional, local puis individuel).	Son but est d'aider la mémoire humaine (ex: les codes d'utilisation du sol sur un plan d'urbanisme- Re2 pour résidentiel à faible densité).	Pour les opérations manuelles comportant peu de catégories (ex: classification de dossiers, de cartes, par usages du sol – institutionnel en bleu).

[MAM, 1984]

Cependant, dans le cas particulier des SIURS, il y a un code dont on doit tenir compte. Il s'agit du *Géocode* (un code géographique) qui, possédant une référence spatiale, permet de manipuler plus facilement les données liées au territoire. C'est le code qui constitue souvent la clé d'accès aux dossiers. Le géocode se caractérise par sa référence à une base géographique. [MAM, 1984]



illustrations: [MAM, 1988]

Figure 17 : Les quatre niveaux de liaison informatiques des dossiers de propriété par géocode

3.5.2 Édition et stockage dans les bases de données urbaines

Afin de rendre les données traitées intelligibles pour un ordinateur, on doit les stocker exclusivement sous forme de chiffres. Ce qui est requis comme données à la base, sont la géographie et le caractère de l'utilisation du sol attribués à une entité spatiale¹⁰. Observations faites, nous voyons que l'utilisation du sol est généralement emmagasinée dans l'ordinateur sous forme de points cartésiens représentés par des coordonnées.

Les données sont stockées et organisées de sorte à ce qu'il y ait peu ou pas de redondances et de répétitions au sein de l'information générée. Donc, par la mention du nom d'un propriétaire, ou par l'adresse, on peut obtenir tout le détail de la propriété au

¹⁰ Objet ou individu représentant une forme dans l'espace géographique. [Rhind et Hudson, 1980]

niveau du rôle d'évaluation municipal. L'avantage principal est certainement celui de concentrer l'information de même type en un même endroit de sorte à la mettre à jour qu'une seule fois lorsque des changements surviennent.

Tableau XV :

Objectifs de la gestion de bases de données relatives à l'utilisation du sol
✓ La possibilité d'avoir plusieurs méthodes d'accès à l'information;
✓ Entreposer des données indépendantes des applications exécutées;
✓ Contrôler et sécuriser informatiquement l'accès aux données;
✓ Faciliter la modification et la mise à jour des informations;
✓ Minimiser la redondance des informations.

[Dale, McLaughlin, 1989]

Rappelons qu'il est possible de rassembler l'information sous différentes formes une fois son intégration au système complétée, que ce soit pour des raisons de manipulation ou de présentation. L'édition laisse place à la correction de l'information résultant de l'encodage. De plus en plus, on tente de corriger automatiquement, par l'informatique, certaines erreurs pour ne pas laisser place à celles commises par l'intervention humaine. Dans certains cas, la correction d'erreur nécessite l'organisation de l'information pour que de telles erreurs deviennent évidentes même si, aujourd'hui, la machine joue un plus grand rôle à ce chapitre. [Rhind et Hudson, 1980]

3.5.2.1 Extraction des données relatives à localisation d'un lot

En premier lieu, mentionnons la clé d'accès permettant de relier entre elles les différentes tables d'information et, à la base, de classer tous les dossiers de propriété à l'aide d'un identificateur unique. Il s'agit du matricule de localisation (géocode) qui déjà renferme plusieurs informations sur l'emplacement d'un lot, selon le découpage par feuillets cartographiques défini par le Ministère des Ressources Naturelles. On parle entre autres du numéro de division (1km par 1km) illustrant une tranche des coordonnées (X,Y) sur la grille de découpage appliquée au fuseau terrestre de 3 degrés. Cette gradation correspond ici à la norme de projection MTM (un carreau de la grille fait un kilomètre de côté et cette dernière trouve son origine en X au centre du fuseau et en Y à l'équateur). Ce résultat est associé à une première série de quatre chiffres qui puisent dans les unités de

dix mille des coordonnées exactes du point. Voyons l'exemple suivant pour mieux comprendre l'explication:

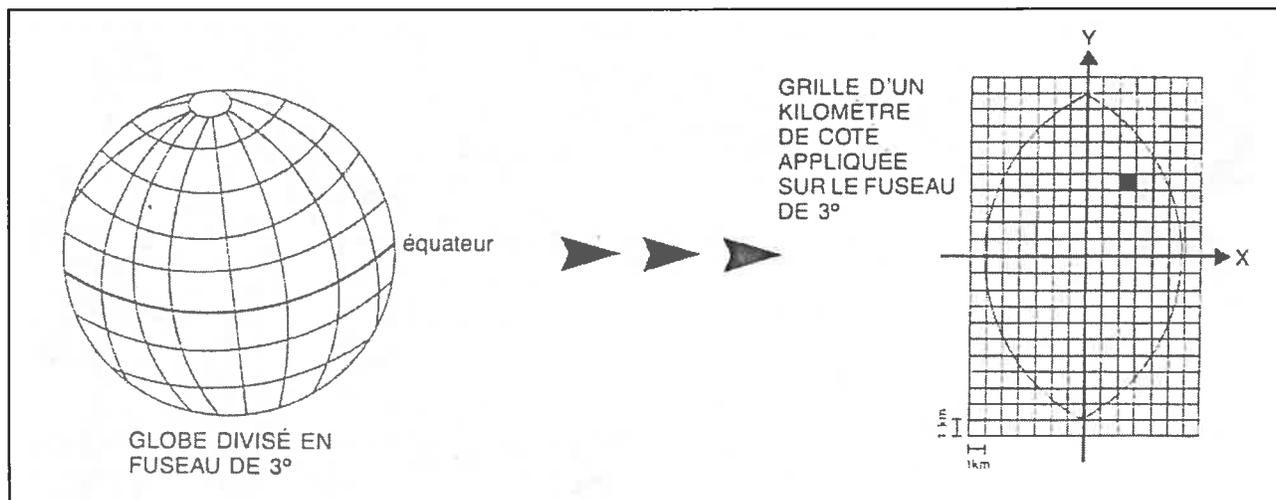


illustration: [MAM, 1984]

Figure 18 : Découpage du globe en fuseaux et application d'une grille de repérage

Le numéro de section (possibilité de dix divisions de 100 m pour l'axe des X et la même chose pour l'axe des Y) consiste donc en deux chiffres correspondant aux unités de cent des coordonnées centroïdes. Finalement, le numéro d'emplacement illustré par le centroïde visuel de la section correspond aux unités des dizaines de ses coordonnées en X et en Y. C'est une dernière série de 4 chiffres pour définir le lot en question (ce qui peut nous donner un numéro qui ressemble à celui-ci: 2546-78-2060). Voyons la situation en détail avec l'illustration de ces explications concernant la construction du matricule et l'énumération des éléments visibles sur le cadastre:

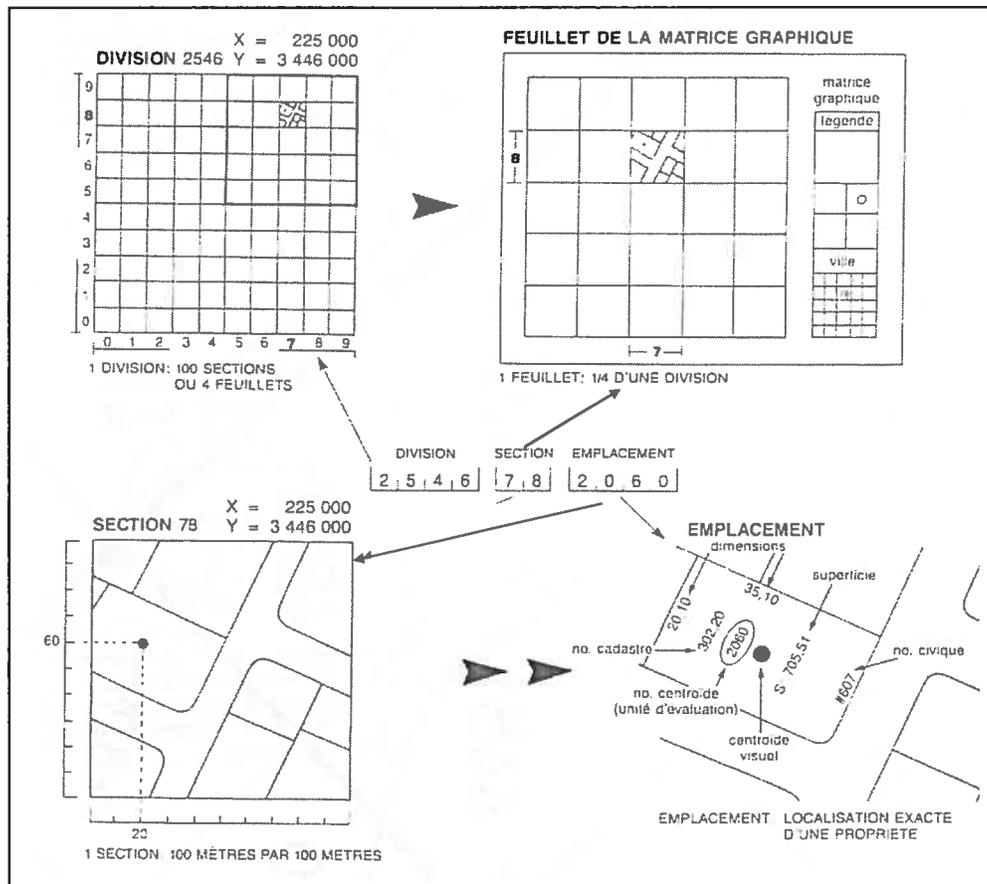


illustration: [MAM, 1984]

Figure 19 : Construction du matricule de localisation unique d'une propriété

3.5.3 Traitements de données exécutés par un SIURS

Le traitement des données au sein des systèmes d'informations est le processus qui varie le plus d'un endroit à l'autre. L'équipement utilisé, la structure des bases de données, l'organisation du système et la nature des informations locales sont des facteurs qui influencent ce processus.

Tableau XVI : Description des deux modes de traitement de données dans un système d'information

l'assemblage	la transformation
les regroupements de données individuelles qui se traduisent en information.(ex: l'association d'un nom de municipalité, d'un numéro de lot et d'une valeur foncière)	l'utilisation d'une donnée à laquelle on fait subir certaines opérations pour obtenir une information dérivée qui ne contient pas la donnée brute (ex: l'obtention de la valeur d'une superficie commerciale à l'aide de l'indice au pied carré).

[MAM, 1984]

La base de données devant être perçue comme un ensemble logique, sa conception prévoit le lien entre les informations des divers services municipaux. L'utilisateur d'une telle base doit être en mesure de saisir la fonction de chaque donnée, sa nature, sa provenance, son mode d'acquisition, son processus de tenue à jour, son mode de représentation dans la base de donnée et les contrôles qui lui sont appliqués. Pour illustrer ce concept, l'auteur évoque l'appellation du « dictionnaire des données ». [MAM, 1984]

3.5.3.1 Critères de qualité de l'information traitée par le SIURS

Autre que la relation avec le sol et la façon dont on l'occupe, l'information traitée doit répondre à quelques critères de qualité de sorte à servir les intérêts de son utilisateur:

Tableau XVII : Critères de qualité de l'information urbaine traitée

Critère	Contexte général	Contexte de la recherche
Précision	L'information doit fournir une mesure adéquate de l'état actuel du terrain.	Les sources d'information à la disposition de l'utilisateur doivent présenter une granularité de l'information à l'échelle du terrain à décrire.
Exactitude	La marge d'erreur provenant de l'information extraite des données doit être très minime ou inexistante. Là où l'erreur existe, la possibilité de la corriger doit être offerte. Si quelqu'un est lésé par ces erreurs, des compensations doivent être disponibles.	La machine qui vient seconder le décideur doit être exempte d'erreur dans l'exécution des requêtes demandées. Sinon, à quoi bon l'aide d'un tel outil dans la prise de décision basée sur des éléments d'information factuelle.
Clarté	L'information doit être libre d'ambiguïté.	L'information utilisée doit être issue d'un consensus entre les personnes impliquées dans la tâche de gestion du territoire. Les champs de données et les codes générés doivent être pertinents aux activités de gestion municipale.
Accessibilité	L'information doit pouvoir être extraite rapidement et facilement.	L'utilisation de tableaux "Individus-Variabes" indexés assure un accès rapide à toute information liée à une propriété.
non biaisée	les données originales ne devraient pas être interprétées ou modifiées de sorte à influencer leur récepteur.	Les données originales, une fois stockées, sont interprétées sur demande par son utilisateur à l'aide de requêtes analytiques.

[Dale, McLaughlin, 1989]

C'est durant l'étape de modélisation que les groupes de données doivent être ainsi évalués. On tente alors de répondre aux priorités des utilisateurs tout en négociant avec les contraintes de réalisation existantes. [Dale, McLaughlin, 1989]

Conclusions sur le chapitre :

Dans ce chapitre, nous convenons que les systèmes d'information urbaine à référence spatiale consistent en des procédures complexes de manipulation et de présentation de l'information cadastrale. Plusieurs contraintes d'ordre physique ou logique s'appliquent lorsque vient le temps de faire coexister les différents types d'information traitée. Ce document informatique intégré et dynamique nécessite un schème d'organisation des données particulier au domaine d'application (ici l'urbanisme municipal) et à la pratique informatique selon les ressources disponibles. Ces outils de travail capables de traiter de grandes quantités d'informations de différentes natures doivent se soumettre à quelques règles d'éthique au niveau de leur tenue et du rendu. Les prochains chapitres discuteront des méthodes en vigueur actuellement pour ce qui en est de l'organisation des données relatives à l'espace géographique urbain et à sa documentation.

Chapitre 4: Outils de gestion et de manipulation des connaissances propres à la prise de décision

Objectifs du chapitre :

Par ce chapitre, on veut être en mesure de cerner les besoins et les potentiels découlant d'une informatisation de l'information. Le format électronique d'emmagasinage des connaissances a entraîné le développement d'outils de manipulation et de traitement spécifiques au domaine. De nouveaux moyens sont nés en matière de gestion de l'information géographique dans les bases de données géomatiques. Ces façons de faire seront revues dans un contexte d'assistance à la prise de décision en urbanisme.

Les besoins engendrés par la tenue de registres informatisés sur l'évolution d'un territoire urbanisé ont entraînés, à leur tour, le développement de plusieurs concepts et techniques de manipulation de ces volumes de connaissances accumulées. Le renouvellement rapide des technologies de traitement informatique a donné naissance à toute une variété de systèmes présentant chacun leurs particularités au niveau du stockage des données, de leur récupération, de leur mise à jour, de leur format d'enregistrement et de classification. On a dû se donner les moyens de rendre toute cette information stockée propre au traitement et à la prise de décision.

Des méthodes nous permettant d'entreposer puis de naviguer dans cette jungle informationnelle ont été développées. Parlons donc en priorité du « forage de données » (data mining) conjointement à « l'entreposage de données » (data warehousing). Dans notre cas particulier des systèmes d'information urbaine, on pourrait ici ajouter le qualificatif « spatial » à ces termes. Si on adopte une logique de requête d'information qui se veut spatiale, il lui faut au préalable une logique d'entreposage qui sera elle même spatiale.

4.1 La localisation de l'information au service du forage spatial de données

Depuis le développement du géocodage¹¹ qui permet la localisation par adresse, les données dites spatiales sont de plus en plus utilisées à l'intérieur d'applications décisionnelles. Dans cet esprit est né le forage de données spatial qui répond au besoin d'exploiter ces données produites. Ces dernières devraient donc être susceptibles de livrer des informations ou des connaissances à l'aide d'outils exploratoires. Ce domaine assez récent réunit des techniques provenant des bases de données spatiales, des statistiques spatiales, du forage de données¹² et des systèmes d'informations. En analyse spatiale, les deux approches préconisées pour ce traitement de l'information sont une approche statistique de même qu'une approche issue de travaux relatifs aux bases de données spatiales. [Zeitouni, Yeh, 1999]

4.1.1 État des besoins en matière de forage spatial de données

L'analyse de données crée en fait trois types de besoins. En premier lieu, il s'agit d'**explorer** les bases de données spatiales afin d'y découvrir de nouvelles connaissances. Plusieurs sources de données sont existantes. Au départ conçues à des fins de stockage, elles renferment une source d'informations utiles que l'analyse exploratoire permet de récupérer. Le deuxième besoin est de **gérer** d'importants volumes de données tout en garantissant des temps de réponse et de traitement acceptables. Des méthodes d'optimisation sont donc nécessaires. Finalement, le besoin qui reflète la spécificité du domaine, est la **prises en comptes des interactions dans l'espace** entre différents objets. Suivant un continuum spatiotemporel, les propriétés d'un endroit sont souvent liées à celles de son entourage. C'est ce qui caractérise les données géographiques. L'analyse de ces données géographiques s'appuierait maintenant sur l'analyse multidimensionnelle de données sans tenir compte des données spatiales. Ce type d'analyse peut être effectué par des méthodes statistiques élémentaires (moyennes, variances, histogrammes) comme par des méthodes (sur une base plus exploratoire) basées sur des analyses factorielles et bi-

¹¹ Autorise la classification d'objets par leur association à un code référencé géographiquement. [MAM,1988]

¹² Habileté à extraire des données précises à partir de grandes masses d'informations. [Jambu, 1999]

variées (corrélations et régressions). Par contre, ces méthodes de traitement s'appliquent toutes à des données quantitatives ou qualitatives mais non à des données spatiales. [Zeitouni, Yeh, 1999]

4.2 Les systèmes de gestion de bases de données spatiaux et relationnels

Rappelons ici que l'outil de travail courant en matière de gestion urbaine d'objets géographiques est le système d'information urbaine à référence spatiale. Ce système est issu des technologies de l'infographie et des systèmes de gestion de bases de données (SGBD). Ils sont caractérisés par cinq fonctionnalités distinctes soit : **l'acquisition de données, l'assemblage de données, l'archivage, l'analyse et l'affichage des résultats.** Dans le cadre du présent projet de recherche, nous nous attarderons à l'assemblage des données de documentation géographiques. Cette étape consiste en l'importation de données de sources différentes et à les fusionner de sorte à constituer une base de données répondant aux besoins d'un SIURS municipal. En intégrant à cette opération la détection de redondance, et la vérification de cohérence¹³ des données, on retrouve des problèmes d'intégration connus au sein des bases de données relationnelles. [Zeitouni, Yeh, 1999]

Malgré le fait que l'on se serve des cinq fonctionnalités mentionnées ci-haut comme référence, elles ne répondent pas pour autant aux spécificités des applications urbaines. L'une des approches actuelles tente de pallier à la situation en interfaçant ce cadre générique à des applications plus spécifiques. Cela est effectué à l'aide d'outils d'analyse statistique ou par des langages de programmation (ex: langage de requêtes SQL).

4.3 L'adoption d'un système de requête géographique à deux vitesses

Les SGBD classiques ont la fâcheuse contrainte ne pas être en mesure de traiter des requêtes géométriques et spatiales qui font appel à des critères plus topologiques, par exemple, le langage SQL ne permet de traiter que des requêtes attributaires (par attributs

¹³ Voir l'annexe 2.

alphanumériques gérés dans des tables). Une requête, spatiale, par commande et interaction graphique (directement sur une carte informatisée à l'écran), pourrait identifier tous les comtés dans un rayon de plusieurs kilomètres autour d'une ville ou identifier des entités géographiques au sein d'une zone d'étude choisie de façon arbitraire tel que le montre l'image suivante :

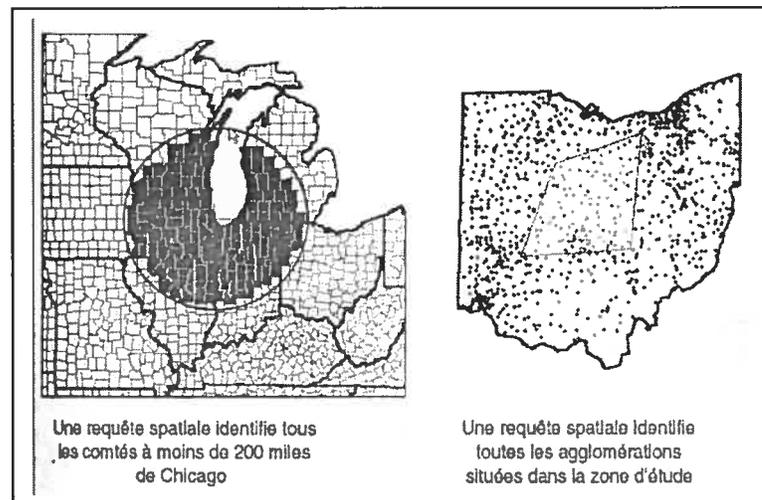


illustration : [GISPLUS, Laurini, Raffort, 1993]

Figure 20 : Exemple de requêtes cartographiques interactives

En fait, il existe différents types de requêtes spatiales ne pouvant pas être résolues par l'algèbre rationnelle. Pour cette raison, on fait appel à la géométrie algorithmique. Deux grandes catégories de requêtes semblent émerger : des requêtes élémentaires (qu'y a-t-il en un endroit précis, suivant un point, une ligne ou une zone?) et des requêtes plus complexes d'analyse spatiale (quel serait l'endroit optimal répondant à certains critères de cheminement ou de découpage de zones?). Selon Laurini et Raffort (1993), des extensions au langage SQL sont inévitables pour le traitement de telles demandes. [Laurini, Raffort, 1993]

4.4 La structuration informatique des données urbaines

La structuration des banques de données relatives à l'utilisation du sol par des systèmes d'enregistrement est nécessaire. Elle pallie à des problèmes de logistique informationnelle au sein des services municipaux qui les utilisent.

Tableau XVIII:

Problèmes actuels à éliminer par la structuration des données informatiques

- Des systèmes en place qui basent leur indexation sur le nom des parties propriétaires d'une parcelle suivant chaque transaction plutôt que sur la parcelle elle-même;
- L'absence d'approche coordonnée pour lier les informations relevées et maintenues par différents services;
- L'absence d'une cartographie de base à échelle moyenne pour usages multiples;
- La cartographie de base urbaine n'était réalisée que par les municipalités en moyens de le faire alors que toutes en ont besoin;
- L'absence d'un système de contrôle et de révision de l'information relevée.

[Larsson, 1991]

4.4.1 Les structures de données courantes

À travers les différents systèmes de gestion de données existants, trois modes de structuration des données existent. Le choix du mode de stockage peut varier en fonction du traitement et de l'utilisation qu'on fait des données. Voici donc les différentes méthodes de structuration de l'information utilisées actuellement:

La structure sous forme d'arborescence : Dans chaque groupe d'information, un champ est considéré comme étant le champ maître ou "racine". Les enregistrements étant répertoriés suivant un ordre sériel, les données sont récupérées en navigant dans les différents niveaux.

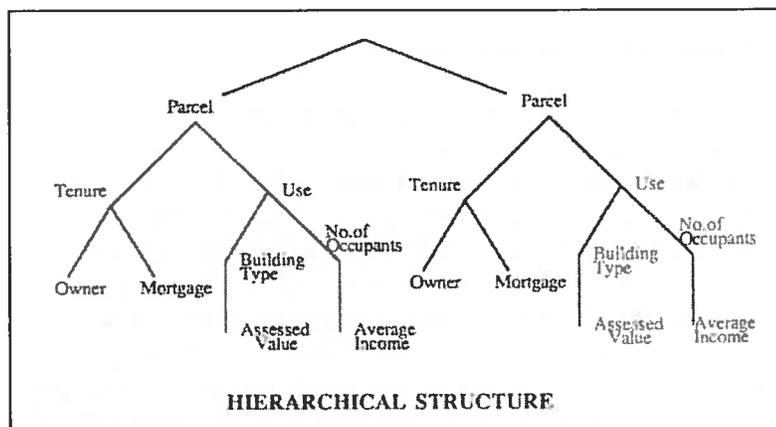


illustration : [Dale, McLaughlin, 1989]

Figure 21 : Répartition de l'information sous forme d'arborescence

La structure sous forme de réseau : Elle consiste en des connections multiples entre les différents fichiers. Ceci permet à l'utilisateur d'accéder à des fichiers précis sans devoir parcourir l'ensemble de la base de données.

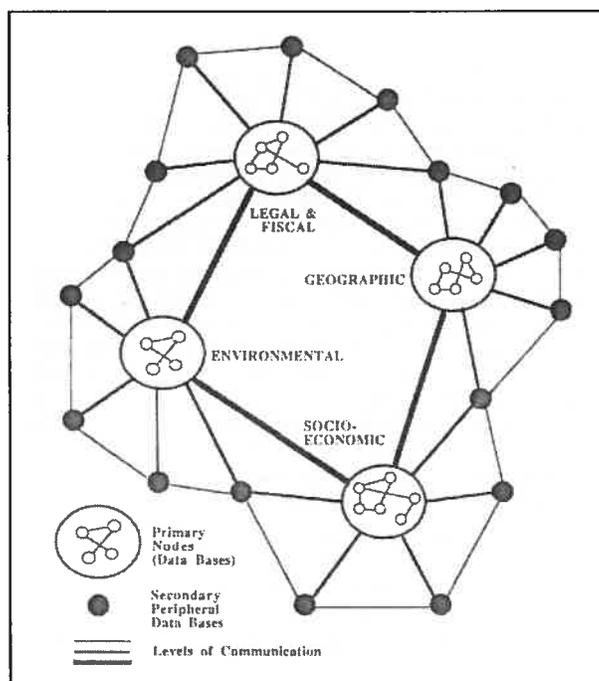


illustration: [Dale, McLaughlin, 1989]

Figure 22 : Répartition de l'information sous forme de réseau

Les structures relationnelles : Ces dernières trouvent leur flexibilité dans l'abolition des hiérarchies de champs de données. Tout champ peut être utilisé comme clé primaire¹⁴ pour extraire de l'information. Un enregistrement n'est plus relié à un champ maître. Au lieu, chaque registre (individu) est considéré comme une rangée faisant partie d'une table à deux dimensions et chaque variable devient une colonne. Cette méthode est aujourd'hui très répandue dans les systèmes informatisés. [Dale, McLaughlin, 1989]

¹⁴ Information référentielle commune à plusieurs tables ou champs de données. [Kimbal, 2000]

Rapport entre la superficie du terrain et l'aire au sol par division, par section, par utilisation générale															
MUNI	DIV	SEC	UTIL	EMPL	SUBDI- LOT VISION	NB BATIM	NB LOGE	NB ÉTAGES	TYPE	CONSTR ANNÉE	SUPERFICIE TERRAIN	AIRE SOL BÂTIMENT	DENSITÉ RSOL/SUP	VALEUR BÂTIMENT	
6331	7181	38	0	3272	27A	0	0	0	0		1 219,313	0,00	0,00	17 500 \$	
			1100	3272	27A	1	1	1	2	1955	0,000	122,83	0,00	3 500 \$	
				3272	27A	1	1	1	2	1960	1 219,313	257,56	0,21	14 000 \$	
				5036	27A	1	1	1	2	1965	1 579,300	224,03	0,14	20 900 \$	
		49	1000	4055	28A	1	1	1	2	1972	666 679,123	311,51	0,00	37 400 \$	
	7182	04	0	3535	27B	0	0	0	0		153 052,109	0,00	0,00	78 700 \$	
			1 000	3535	27B	1	1	1	2	1960	153 052,109	388,62	0,00	25 800 \$	
			1810	3535	27B	1	1	1	2	1960	0,00	390,14	0,00	20 700 \$	
				3535	27B	2	3	1	2	1965	0,000	487,68	0,00	32 200 \$	
		05	1100	7487	27B	1	1	1	2	1970	455,767	221,89	0,49	27 700 \$	
				8664	27B	2	1	1	2	1968	510,671	198,12	0,39	14 700 \$	
				9846	27B	3	1	1	2	1968	639,245	205,74	0,32	14 900 \$	
		06	1100	4665	27	B	1	1	2	1969	1 449,240	405,38	0,28	27 500 \$	
		11	1000	4798	27A		1	1	2	1960	1 321,503	270,66	0,21	31 800 \$	
			1100	3567	27A		1	1	2	1969	2 394,498	233,17	0,10	24 000 \$	
				4274	27A		1	1	2	1940	7 292,650	227,69	0,03	22 800 \$	
		15	1100	0724	27B	4	1	1	2	1962	575,692	205,74	0,27	22 200 \$	
				1501	27B	5	1	1	2	1974	806,093	205,74	0,26	25 700 \$	
	20	10													
		00	3671	27A	2	1	1	1	2	1959	1 827,34 43	281,33	0,15	19 500 \$	
		31	100	1000	8680		1	1	2	2	1950	74 138,409	436,47	0,01	47 300 \$
	7183	02	6911	8510	26B		1	0	1	2	1935	1 858 000	236,22	0,13	5 000 \$
			13	1100	7705	26B		1	1	2	1963	696,750	160,93	0,23	15 900 \$
	9045	40	1900	0000	23A		1	0	0	0	000	743,200	0,00	400 \$	
	9151	14	1900	0010	27B		1	0	0	0	0000	6 038,500	0,00	1 200 \$	
						4356	3933	4126			61 972 763,775	971 774,11		95 795 300 \$	

illustration: [MAM, 1984]

Figure 23 : Répartition des information sous forme de tableau relationnel « Individus-Variables »

4.4.2 Les possibilités analytiques des données urbaines structurées

Tel que cité dans Rhind et Hudson (1980), Jeffers en 1970, décrit deux utilisations principales qui sont faites de l'information relative à l'utilisation du sol. Il s'agit de la production d'estimés des proportions d'un sol utilisé à différentes fins ainsi que de l'évolution de ce dernier à partir du moment où une certaine fonction lui est attribuée. L'autre, se résume en la production de cartes relatives à la zone relevée démontrant la distribution spatiale d'une variété de classes d'utilisation du sol. [Rhind et Hudson, 1980]

Tableau XIX:

Possibilités d'analyses effectuées à l'aide des données urbaines	
✓	fournir des descriptions des volumes d'utilisation du sol à l'intérieur de régions définies, sur une base statistique ou pour associer l'utilisation à un point précis dans l'espace;
✓	superposer des informations relatives à un secteur à différents moments dans le temps afin d'en mesurer l'évolution;
✓	superposer différents types d'informations (ex.: géologies, transports) pour un même secteur afin de faire ressortir des coïncidences et les relations entre les éléments montrés (pistes sur ce qui aurait pu entraîner un type d'occupation sur une parcelle donnée);
✓	supporter l'analyse statistique pour explorer les relations à même les données relevées à l'intérieur et aux limites de la zone choisie (ex: pourquoi deux même types d'usage se côtoient toujours?);
✓	produire facilement des représentations graphiques des données, autant sous formes de cartes que sous forme de graphiques (les SIG permettent déjà depuis longtemps de faire un bon bout de chemin en ce sens).

[Rhind et Hudson, 1980]

4.4.3 Les données idéales pour le domaine

Les données idéales devraient donc décrire la différenciation de l'espace autant sur un plan formel que fonctionnel. On aimerait donc des descriptions du mode d'utilisation du sol et de ce qui le recouvre. Ces données auraient à décrire quelles zones sont homogènes selon leurs formes et leurs fonctions et lesquelles devraient être groupées avec des entités de même nature déjà existantes. Ces données devraient être groupées d'une façon qui ne permettrait pas seulement le réarrangement spatial mais également de façon à pouvoir comparer ces données avec des résultats de relevés ultérieurs. Les auteurs précisent que pour obtenir de telles informations, il faudrait être en mesure de relever l'ensemble du territoire au même moment. Pour l'instant, il existe plusieurs types de relevés et, on peut sonder un même quartier plus d'une fois. [Rhind et Hudson, 1980]

Pour conclure le chapitre, voici une citation qui résume bien la manipulation des informations relatives à l'utilisation du sol:

*“The danger is clear: national land use statistics are, potentially, a meaningless amalgam of figures based on different classifications applied to dissimilar areal units with varying degrees of precision”-[(Dickinson and Shaw, 1978) tel que cité dans **Rhind et Hudson, 1980]***

Conclusions sur le chapitre :

Nous avons vu dans ce chapitre, différents besoins et différents outils liés à la gestion des données et des connaissances urbaines. Certains problèmes émergent de ces besoins en fonction des outils utilisés pour les combler. Pensons à la structuration des données informatiques ainsi qu'à leur consultation. Plusieurs municipalités ne sont pas prêtes à supporter les systèmes d'information urbaine au sein de leur gestion municipale. Un grand travail reste à faire au niveau de la sensibilisation des gestionnaires urbains afin de permettre aux municipalités d'entrer dans la culture des données numérisées. Du moins, pour que cette ressource d'information soit renouvelable et utile à la gestion d'une ville.

Chapitre 5 : Contraintes d'insuffisance sémantique dans les bases de données géomatiques

Objectifs du chapitre :

La construction de bases de données municipales montre certaines lacunes au niveau de la prise en charge de l'espace et de l'utilisation du sol. Il est surtout question de la faible définition sémantique des relations entre les objets urbains répertoriés par l'information géographique. Nous verrons à quel point la cohérence et la documentation (par les métadonnées) de la banque de données prend son importance. Différentes méthodes de gestion des relations spatiales (voisinage des objets) au sein de bases de données relationnelles existent. Nous en explorerons deux. L'une utilise un index de jointure tabulaire, l'autre considère davantage un découpage plus performant (par objet) de l'information de base.

La cohérence des bases de données spatiales doit être vérifiée de façon prioritaire si l'on désire en obtenir des réponses fiables et en extraire des raisonnements valides. Elle consiste en la vérification de l'exactitude des structures géométriques et topologiques des objets spatiaux. Laurini et Raffort (1993) nous brossent un tableau des contraintes d'intégrités spatiales liées aux bases de données géomatiques. Ces dernières sont reliées aux aspects sémantiques qui découlent des principes géométriques et typologiques de ce type de bases de données. Il est question principalement, pour ce qui est de la gestion cadastrale urbaine, d'objets spatiaux modélisés suivant la géométrie euclidienne¹⁵.
[Laurini, Raffort, 1993]

5.1 L'information géographique

l'information géographique se caractérise selon une composante de localisation spatiale qui vient s'inscrire sur la surface terrestre. En général, on parle des dimensions d'un objet représentées par des points, des lignes et des surfaces sous forme de coordonnées planimétriques. Les systèmes actuels traduisent cette information par une projection

¹⁵ Généralisée au traitement des espaces de dimension infinie. [Le Petit Robert]

d'entités au sol. Ces objets géographiques sont définis à l'intérieur du système comme étant des structures qui contiennent des données sémantiques et des données spatiales. Ainsi, on est en mesure de regrouper les objets géographiques qui partagent les mêmes propriétés et les mêmes structures sous forme de couches thématiques (collection homogène). Ce sont en fait ces couches (fig.16, p.41) qui constituent la base de données géographique. Dans ces bases de données, sont décrites également les relations spatiales entre les objets. Ces dernières sont les informations qui traduisent des propriétés essentielles provenant du monde géographique (notions de voisinage). Ces descriptions de l'influence du voisinage sur un endroit peuvent être explicites ou implicites dans la base de données géographique. [Zeitouni, Yeh, 1999]

5.1.1 Les métadonnées géographiques

Selon KAS, 1995 [tel que cité dans Spéry, Libourel, 1998, p.60] ce type de donnée se caractérise comme suit : « *the function of metadata description is to be able to abstract out and capture the essential informations in the underlying data, independant of representational details* ».

Suivant le standard du Federal Geographic Data Committee concernant les métadonnées, la complexité des métadonnées et de l'information qu'elles véhiculent augmentent en fonction de la compétence des utilisateurs. Dans la tendance d'une utilisation plus large des données géographiques, les métadonnées doivent permettre à tout utilisateur de saisir le sens des données qu'il collige. Ce sont des données sur les données. D'autre part, le point axé sur l'échange et la transformation de données est souvent mis de côté. Les métadonnées interviennent donc aux niveaux de la définition du contexte, de la documentation des définitions sémantiques et de la description du processus de production des données géographiques (en vérifiant si des entités jointes sont confondues dans le même objet réel). [Spéry, Libourel, 1998]

Les SIURS appliquent les métadonnées à un lot homogène d'objets graphiques. Elles concernent généralement l'échelle, l'emprise, le référentiel géographique ou le type de

projection, la qualité, la datation. Ces informations sont toutes rassemblées dans un dictionnaire de données (fig. 9, p.27) faisant partie du SIURS. Au-delà du rôle descriptif qui leur est confié, ces métadonnées, selon les auteurs, pourraient intervenir de façon pertinente dans les processus d'intégration et de mise à jour des données, rendant possible la description de l'évolution des entités géographiques. [Spéry, Libourel, 1998]

5.2 L'insuffisance sémantique des structures de données

Partons du fait qu'un objet géographique peut avoir plusieurs représentations géométriques. La réalité inverse est aussi vraie dans le cas où une même modélisation peut représenter des objets spatiaux distincts. Par exemple, prenons un ensemble de points, décrit sous forme relationnelle, modélisant un objet quelconque. Cette modélisation en question pourrait représenter un groupe de points sans lien entre eux, une polygone ouverte ou fermée, ou même un polygone. Donc, des objets spatiaux de type géométrique différent peuvent être modélisés en relationnel de manière identique. Pour différencier ces objets, il s'agit de définir de façon plus précise cette particularité sémantique qui leur est propre. Les auteurs donnent l'exemple de la génération d'un objet entier à partir d'un ensemble de points privilégiés et inversement, la validation de l'appartenance d'un point à un objet. [Laurini, Raffort, 1993]

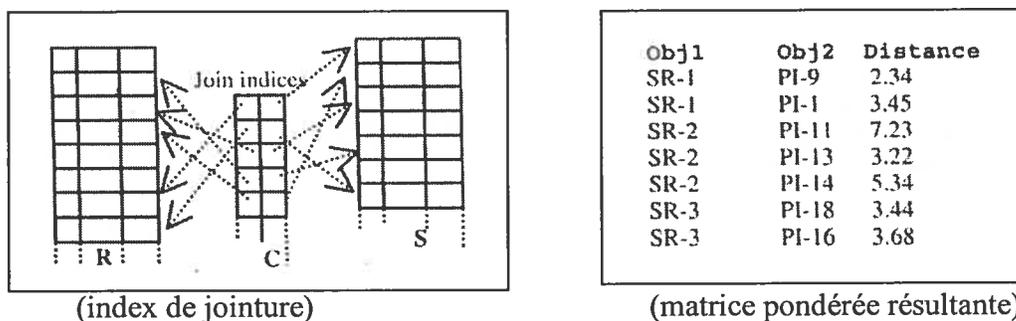
5.3 Les contraintes d'intégrité relationnelles

Les contraintes d'intégrité peuvent être définies au sein des bases de données relationnelles suivant la clause « ASSERT » (permet de certifier les valeurs d'une liste suivant le domaine). Par exemple, est-ce qu'un segment peut posséder le même point pour extrémités? Cette clause est déjà présente dans le langage SQL mais est peu implémentée. On est donc en mesure de vérifier l'intégrité des valeurs (structure adéquate pour une adresse civique), l'intégrité existentielle (un identifiant unique pour chaque entité) et l'intégrité référentielle (clé assurant le lien avec les autres objets de la table). Notons toutefois, selon les auteurs, que l'écriture de certaines contraintes d'intégrité ne

peut se résumer qu'à des formulations de type déclaratif (suivant une liste d'énoncés). Par conséquent, le langage SQL qui n'offre la possibilité de définir des contraintes d'intégrité que par des attributs sur les objets s'avère insuffisant pour les données géographiques. [Laurini, Raffort, 1993]

5.4 L'encodage des relations spatiales entre les objets

Les SGBD spatiaux utilisent des langages d'extraction qui mettent en évidence des relations spatiales par le biais d'opérateurs appropriés. Ces relations sont décrites dans un langage de requête par des prédicats (fonction renvoyant une valeur binaire). Par exemple, on peut définir une relation de voisinage entre deux régions par une adjacence, un chevauchement ou une inclusion. Des extensions au langage SQL permettent ces traitements par une clause « WHERE » dans la requête. Au niveau des traitements, on peut voir les coûts de calcul augmenter de façon importante si les relations spatiales entre les objets géographiques ne sont pas claires et explicites. La méthode des indices de jointure proposée par Valduriez en 1987 ¹⁶ consiste en le stockage de couple d'indices issus de produits cartésiens entre deux collections d'objets spatiaux. Cette technique sert à accélérer les jointures en relationnel et utilise une structure rendue sous forme de tableau. En fait, chaque couple d'indice formé signifie qu'il y a une relations entre les deux objets spatiaux qu'ils représentent. Dans une troisième colonne attachée à celles des couples formés, on note la distance entre les deux objets.



illustrations : [Zeitouni, Yeh, 1999]

Figure 25 : Index de jointure sous forme de table et de matrice pondérée

C'est une pratique adaptée aux bases de données spatiales qui permet de connaître toutes les relations de voisinage existantes entre les objets géographiques enregistrés. De plus, le stockage de ces structures peut simplement emprunter les tables d'un système de bases de données relationnelles. [Zeitouni, Yeh, 1999]

5.5 Le problème à l'échelle de la gestion urbaine

L'article sur l'uniformisation des données décrivant des objets physiques, écrit en coopération par Manon Guité et Claude Parisel à l'Université de Montréal (GRCAO, 2000), cerne bien le problème qui sévit en aménagement du territoire. Voyons-en rapidement les grandes lignes.

Des limites claires existent au niveau de l'établissement de relations entre les informations alphanumériques issues des bases de données urbaines et des données géométriques contenues par les plans cartographiques et les maquettes. Bien entendu, les logiciels de SIURS qui font lentement leur apparition dans la plupart des municipalités de la province, offrent des liens par pointeurs entre les deux types d'information mais seulement pour des descriptions géométriques simples et limitées. Ceci retranche un niveau de réalisme important dans les représentations. Dans le but d'assurer une cohérence des données géométriques nécessaires à l'intégration des informations descriptives, une norme de description des objets assurerait une exploitation plus large des modèles construits. [Guité, Parisel, 2000]

5.5.1 Le découpage des informations par objet

La majorité des logiciels disponibles pour le traitement graphique de l'information offre des possibilités de répartir l'information en catégories suivant le principe des couches (ou calques). Selon les auteurs, cette façon de faire permet de filtrer l'information des catégories auxquelles l'information est reliée mais non l'identification d'objets

¹⁶ Tel que cité dans [Zeitouni, Yeh, 1999]

spécifiques ou les relations entre ces derniers. On peut comparer cette approche à celle retrouvée dans les bases de données relationnelles. [Guité, Parisel, 2000]

5.5.1.1 L'organisation hiérarchique des informations

Faisons tout de suite référence à l'organisation de ces informations. Dans le cadre de l'information à gérer en milieu municipal, nous sommes confrontés à différents types d'objets. Certains d'entre eux jouent un rôle plutôt logique alors que d'autres représentent des formes particulières existantes dans un environnement physique. Un arbre hiérarchique peut être alors constitué par la référence d'un objet à son niveau supérieur. Dans le cas d'une ville, on pourra alors faire des sélections groupées pour les routes entourant un quartier (le quartier jouerait le rôle de l'élément logique rassembleur ou de dénominateur commun pour tous les objets composant ces routes). [Guité, Parisel, 2000]

Voyons une image représentant l'idée des auteurs sur cette question de relation et de hiérarchie entre les objets impliqués dans une description urbaine (les objets en gris sont les objets logiques thématiques alors que les autres sont physiques et représentent des formes particulières existantes):

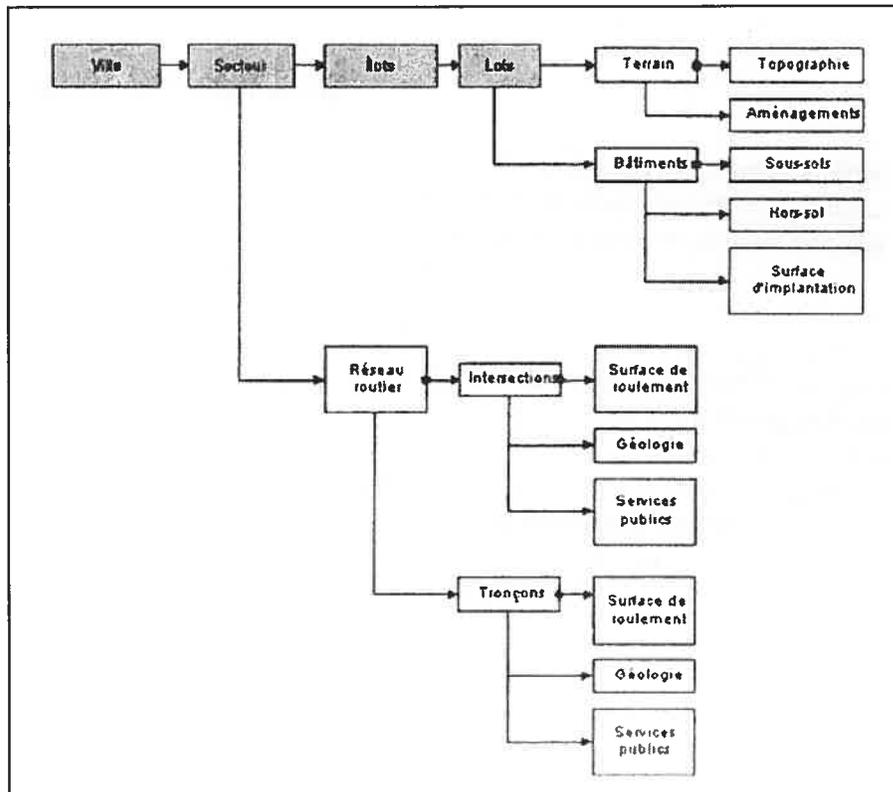


illustration : [Guité, Parisel, 2000]

Figure 26 : Hiérarchie relationnelle des objets constituant une description urbaine

Certains objets entretiennent entre eux des liens de proximité dans l'espace physique réel que les auteurs appellent « relations d'adjacence ». Selon eux, elles sont nécessaires pour traiter l'information de manière intelligente (ex. : une intersection précise avec un tronçon de rue). Chaque objet impliqué dans la modélisation doit être décrit de façon systématique afin de permettre le traitement de l'information qui leur est rattachée. Le format de description peut être propre à chaque objet mais les jonctions doivent être communes à deux objets liés. [Guité, Parisel, 2000]

Voici un exemple de graphe des relations à prendre en compte entre les objets (intersections, tronçons de rues, lots cadastraux) dans le cadre d'une description de quadrilatère urbain quelconque :

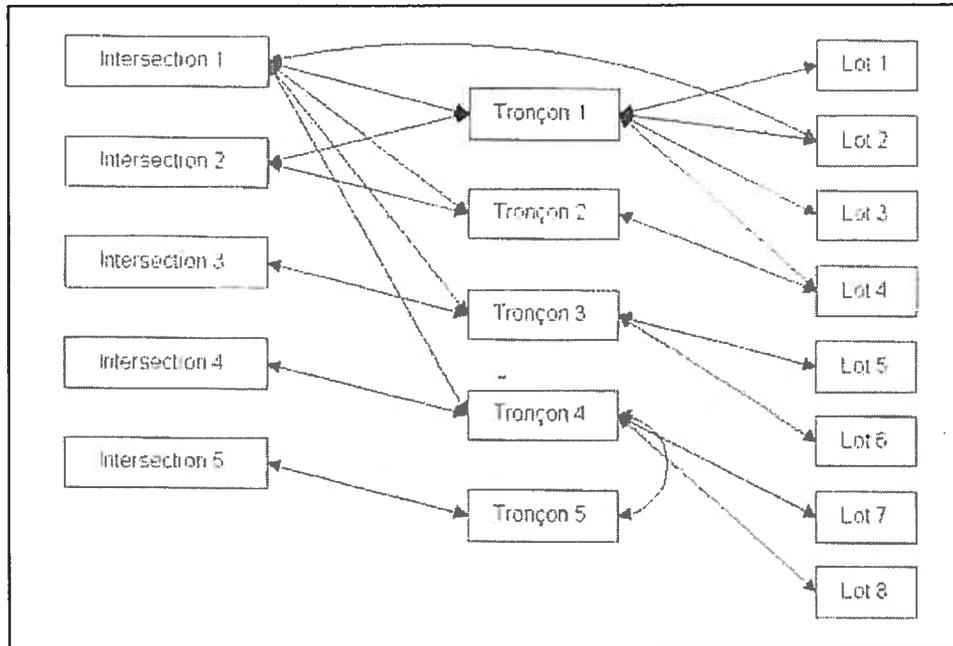


illustration : [Guité, Parisel, 2000]

Figure 27 : Graphe des relations possibles entre les objets définissant un quadrilatère urbain

Au niveau des formats d'enregistrement, il est vital que la banque d'information créée soit précise et organisée de façon à rendre aisée la tâche de l'utilisateur qui opère et met à jour l'information. On doit aussi prévoir les échanges d'informations avec d'autres banques ou utilisateurs. Dès le départ, l'utilisation du modèle d'information à construire doit être connue afin d'identifier les données qui seront nécessaires au traitement. Pour les questions de mise à jour, il n'y a pas d'interférence avec l'interface de saisie qui elle, reste ponctuelle. La saisie traite donc l'enregistrement pour mettre à jour les fichiers utiles à un traitement prochain. C'est en fait la raison pour laquelle des fichiers décrits en formats alphanumériques doivent faire partie d'une structure organisée de manière à répondre à des requêtes. De telles descriptions peuvent être enfin exportées vers des formats propres aux logiciels de CAO couramment utilisés. [Guité, Parisel, 2000]

Conclusion sur le chapitre :

Nous avons appris l'importance de la définition sémantique de l'information et des objets qu'elle décrit. Au sein des municipalités, des insuffisances au niveau de l'utilisation de l'information et des métadonnées géographiques sont au rendez-vous. Les auteurs émettent des possibilité de traitement des données tabulaires à l'aide de clauses se rapportant au langage de requête SQL. La validité des informations traitées est ainsi vérifiée à l'aide des relations de voisinage et d'adjacence. Les deux techniques concernant l'indexation tabulaire et l'organisation des données sémantiques et géographiques de base, pourraient être combinées pour une meilleur performance des bases de données municipales et de la gestion urbaine.

La deuxième partie du mémoire sera dédiée à l'expérimentation de ces méthodes, opérant sous forme de requêtes SQL, au sein d'un entrepôt de données municipal fictif. Toujours dans un contexte d'assistance à la prise de décision, le cadastre et l'utilisation du sol de la Ville de Saint-Luc, sur la rive sud de Montréal, seront à l'étude.

Partie 2 :

Expérimentation

Chapitre 6 : Approche méthodologique du problème posé et présentation du cas étudié

Objectifs du chapitre et de l'expérimentation :

À travers ce chapitre, nous verrons à définir l'approche méthodologique employée pour aborder la problématique de départ concernant la saisie et l'organisation de l'information au sein des bases de connaissances urbaines. Nous situerons exactement le niveau d'intervention au sein des systèmes d'information urbaine à référence spatiale. La solution expérimentée sera exposée. Par cette expérimentation, on aimerait être en mesure de permettre un développement de la gestion efficace de l'information géographique utile à l'urbanisme; de réduire la nécessité des nouvelles saisies par l'exploitation maximale des sources d'information existantes; d'uniformiser l'échange d'information tout en l'allégeant; de faciliter l'accès à l'information et aux données. Les sources de données utilisées pour documenter le cas à l'étude seront présentées.

6.1 L'approche analytique adoptée pour aborder la problématique

Elle consiste en la décomposition du problème simplifié en une multitude d'éléments que l'on traitera en cas par cas pour ensuite les réunir à nouveau. On doit traiter les différents éléments d'un phénomène d'une façon disjonctive. Un peu comme si chaque élément formait un tout à part entière ou comme s'il était autonome. Cette façon d'aborder un phénomène donné accorde une plus grande importance à la compréhension des liens qui rattachent chaque élément de sa composition.

La nouvelle optique, découlant de l'étude des composantes du système d'information géographique urbain en question, nous permettra une meilleure compréhension de ce dernier. L'application de cette approche, en gestion de l'information municipale, nous aidera à développer de nouvelles structures d'information numérique au sein des bases de données actuelles. Nous serons donc en mesure de découvrir le référentiel commun existant en termes de granularité de l'information. Le tout, selon une expérimentation à partir des informations relatives à l'utilisation du sol urbain.

6.1.1 Application de l'approche analytique au problème

Compte tenu du fait qu'il ne peut y avoir qu'une seule information pour régir l'utilisation du sol, la problématique de variété des informations à traiter induit une conjonction évidente des différents types de données colligées. La logique disjonctive de l'approche analytique est impliquée au niveau où chacune des informations qui coexiste, concernant une propriété foncière, est relevée, traitée, stockée et échangée indépendamment de sa nature sous forme de données. Ces dernières viennent toutes s'arrimer à celles en place, en fonction des dimensions de traitement constituant la banque des connaissances de l'administration municipale, afin de former un tout fonctionnel.

On doit trouver des outils capables de virtuellement gérer, appliquer et clarifier ce système de gestion de l'espace créé par l'humain. L'instrument de travail doit le faire de telle sorte qu'il soit compréhensible (clair) pour n'importe quel professionnel du domaine. Ce qui nous intéresse ici, sont les différentes façons de véhiculer une information suivant les nombreuses formes qu'elle peut prendre à travers la base de données SIURS d'une administration municipale. Le ou les outils choisis devront nous permettre de construire un modèle représentatif de la structure des informations et des données supportées par le système d'information actif dans l'institution.

6.2 Explication du choix de la solution en fonction de l'approche adoptée

D'un point de vue plus pratique, parlons du problème existant à la base. En premier lieu, on pourrait en situer la source au niveau de la grande quantité d'informations à traiter dans le domaine de l'urbanisme et la gestion de l'utilisation du sol urbain. Ces différentes informations sont variées dans leurs natures et leurs formes. En fait, il s'agit de traiter des données provenant de multiples sources. Maintenant, la technologie informatique a donné naissance à plusieurs outils logiciels qui offrent la possibilité de prendre en charge cette multitude de données pour les rendre accessibles visuellement à son utilisateur. Ces outils

en question sont des systèmes d'informations urbaines à références spatiales (SIURS). La difficulté majeure avec ces derniers réside dans le fait que l'accumulation et le stockage des informations s'effectue en "double" (ou plus) en raison de la multitude de données qu'on peut y gérer. Ils font coexister deux ou plusieurs bases de données distinctes (pour les attributs alphanumériques et pour les objets graphiques) reliées entre elles par des pointeurs virtuels. Ceux-ci consomment de l'espace disque supplémentaire et alourdissent les temps de traitement pour la résolution de requêtes simples.

La présente proposition suggère une structure d'association et un mode de description alphanumérique de l'information géographique et réglementaire urbaine (zonage) . Et ce, à l'aide des bases de données tabulaires relationnelles conventionnelles (tableaux « Individus-Variables »). Branchés entre eux à la manière traditionnelle (clés de jointure), l'ensemble des tableaux de données peuvent être administrés au sein d'une base de données unique; ce qui en faciliterait la mise à jour, la maintenance et l'échange. On serait en mesure d'effectuer des requêtes spatiales en fonction de règles mathématiques prédéfinies. C'est pour cette raison que l'information devra être structurée à cet effet. C'est également dans cette raison que la granularité des dimensions d'information traitées trouve toute son importance. Il s'agit en fait d'une vision rationnelle qui veut extraire, déduire et documenter le maximum d'information à partir des données de base disponibles pour en minimiser l'entreposage. Ce que l'on remarque au départ dans le milieu administratif urbain, c'est une exploitation déficiente de l'information existante. Un meilleur découpage de cette information ainsi qu'un meilleur agencement en améliorerait le rendement. Par l'établissement de nouvelles relations et règles de traitement entre ces données, un nouveau potentiel s'offre à l'aménagiste qui n'utilisait l'information que pour ce qu'elle était au départ (ex: une longueur, un nom de rue).

6.2.1 Situation du cas dans le processus de fonctionnement des SIURS

À ce point-ci, il est intéressant de savoir à quel niveau, dans le mécanisme de fonctionnement d'un système d'information à référence spatiale, l'intervention en question se situe exactement. En fait, l'intervention en cours aura une influence sur les

étapes de saisie, de conservation, d'accès et de traitement des données au sein des bases de données urbaines. Tel que montré dans le schéma ci-dessous, toutes ces opérations sont liées aux procédures de contrôle, de mise à jour des données ainsi qu'à la possibilité de communication des résultats. On intervient en fait au coeur du système, soit sur l'encodage de l'information à la saisie et sur le stockage optimal (au niveau de la structure), afin de permettre un reconstitution aisée de l'information urbaine. Cette structure des données facilitera alors l'échange d'information et la production de rapports.

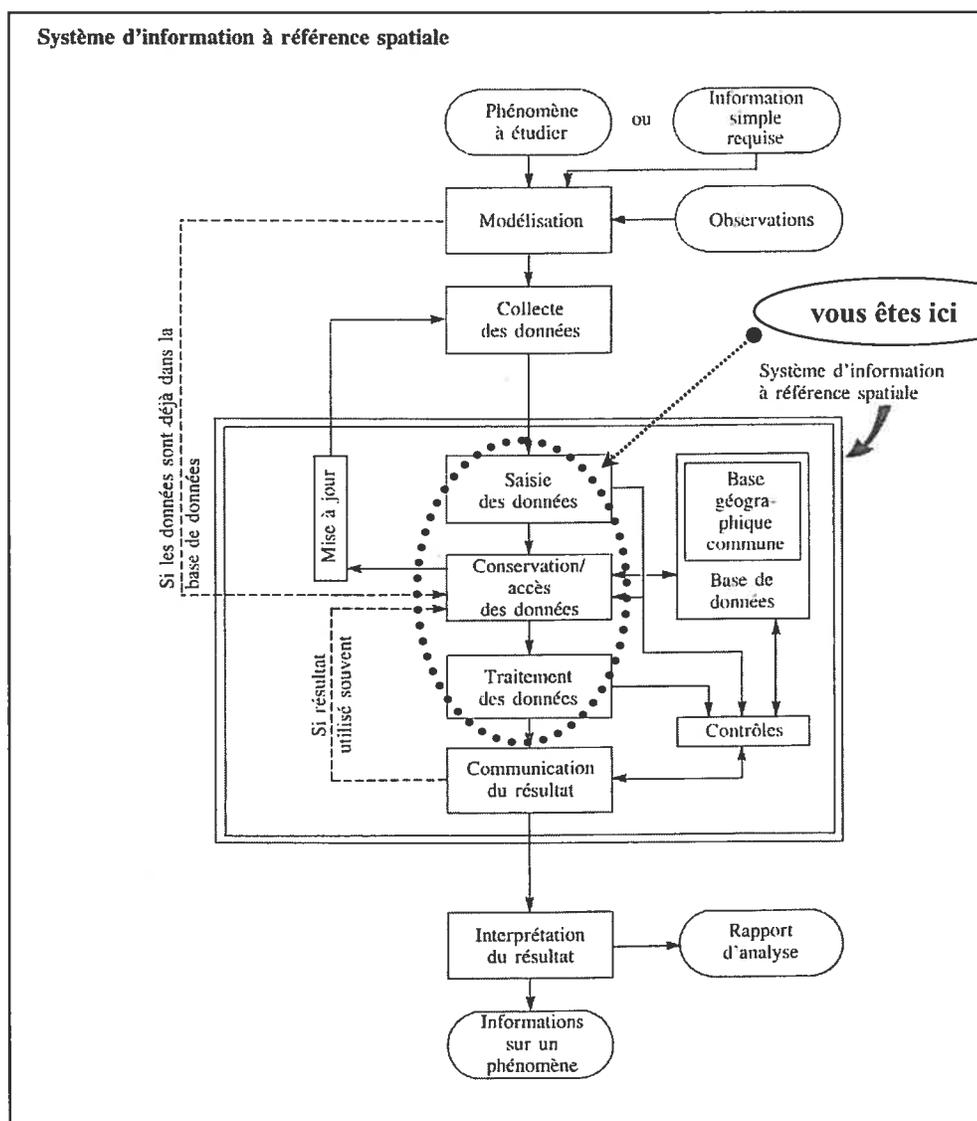


illustration: [MAM, 1984]

Figure 28 : Situation de l'intervention dans le processus de fonctionnement d'un SIURS

6.2.2 Justification de l'orientation prise pour le traitement des données

L'expérimentation adoptera une approche de modélisation de type ROLAP (relational online analytical processing) pour la réalisation du prototype en question. Ce choix découle de différentes raisons. Entre autres, pensons au fait que la gestion relationnelle des tables d'information est utilisée depuis longtemps dans les systèmes de fonctionnement des bases de données. Encore largement répandue, cette méthode de stockage et de manipulation des données informatiques est déjà bien implantée dans les systèmes d'information municipaux. Plus précisément, au niveau de la tenue des informations relatives au rôle d'évaluation municipale. La méthode ROLAP, pouvant simuler le comportement d'une base de données multidimensionnelle, permet la gestion d'un très gros volume de données par partitionnement ou accès en parallèle. Le point principal du projet touche le stockage ainsi que les possibilités d'interprétation de ces données.

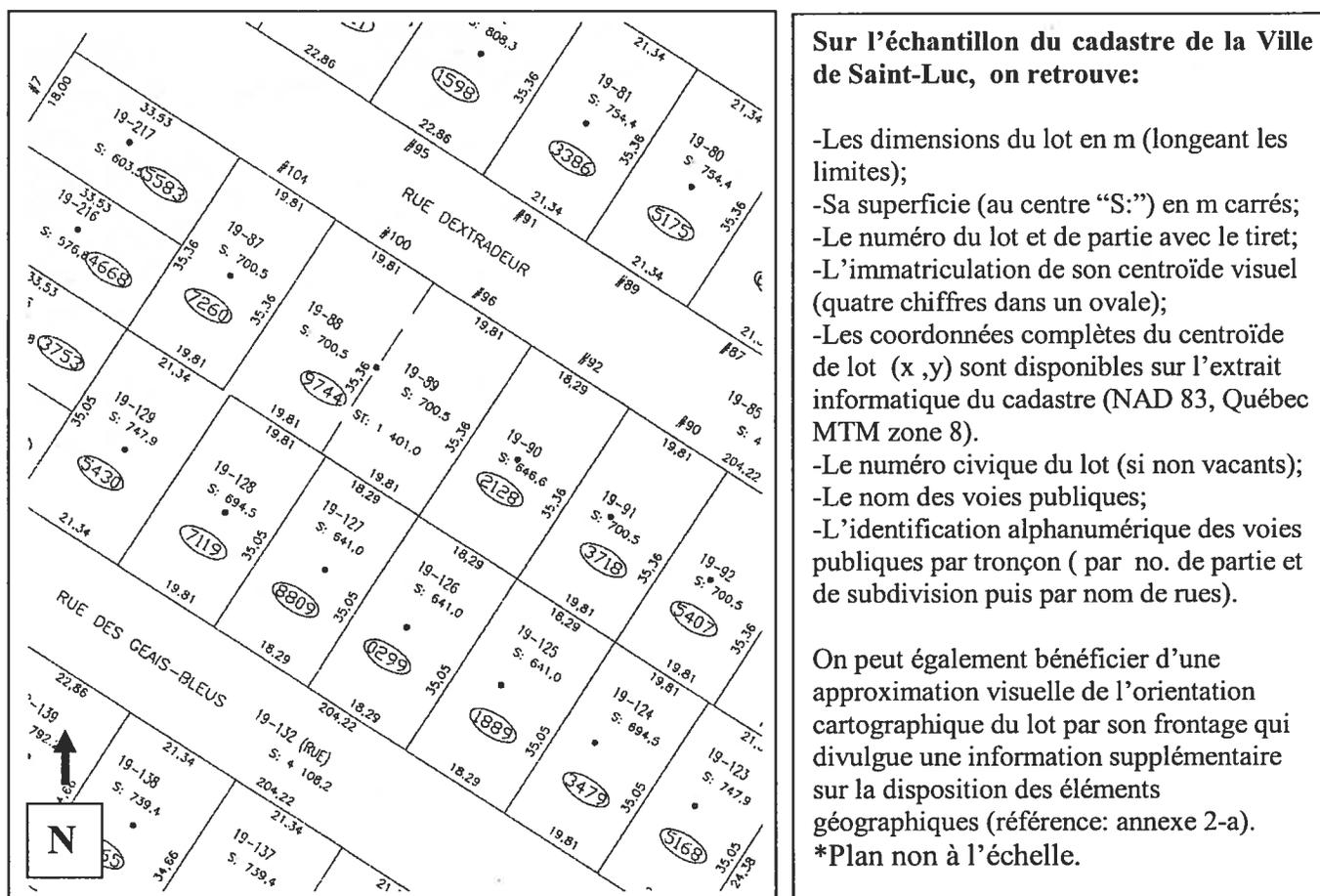
6.2.3 Échelle d'intervention pour la cueillette de données et l'expérimentation

L'échelle d'échantillonnage se situera de façon évidente au niveau municipal urbain. Par contre, l'expérimentation consistera en la gestion de quelques lots de cadastre extraits d'un îlot urbain quelconque. Le temps et les ressources ne nous permettent pas de construire une base de données urbaine complète. Donc, à l'échelle informatique, ce thème de gestion spatiale viendra s'inscrire à l'intérieur d'un magasin de données spécialisé, pour les besoins de la profession d'urbaniste, qui lui-même fera partie de l'entrepôt de données municipal¹⁷. En fonction des résultats obtenus, nous serons en mesure d'imaginer quel serait l'impact d'un tel mode de structure de l'information pour la gestion spatiale d'une ville en entier. En fait, c'est la validité du modèle d'expérimentation qui sera ici mise à l'épreuve.

¹⁷ Voir l'annexe 1-a.

6.2.3.1 Collecte des données cadastrales et réglementaires

Les données utiles à l'expérimentation sont l'ensemble des attributs spatiaux que l'on retrouve sur un plan cadastral et sur la grille des usages prescrits. Parmi ces données, uniquement celles qui correspondent à l'identification géométrique, géographique et réglementaire d'un lot nous intéressent. En fait, toutes les données techniques visibles sur le plan sont candidates. Voyons plus précisément l'information disponible encodée sur un cadastre (matrice graphique).



Les numéros de division et de section complétant le matricule sont disponibles dans le cartouche de la matrice graphique (non visible ici).

Sur un cadastre, le périmètre du lot est évidemment dessiné. Donc, dans une matrice graphique informatisée, les bornes d'un lot ainsi que ses segments sont mémorisés.

Figure 29 : Extrait commenté du plan de cadastre municipal source de la Ville de Saint-Luc

Les informations relatives à l'utilisation du sol seraient considérées davantage comme des attributs ou des dimensions venant se greffer aux entités spatiales en détaillant leur état dans l'évolution du monde urbain. Parmi ceux-ci on peut retrouver le code d'une fonction (ou activité) urbaine particulière, des modalités de droits acquis ou de la réglementation se rattachant à une propriété ou à un secteur en particulier. Ces codes peuvent être construits au gré de la municipalité puisque qu'il n'y a pas de norme qui régit ce type de codification de l'information. L'information relative au zonage est tirée elle aussi d'un plan superposé à la matrice graphique qui délimite chaque zone d'application des règlements sur le territoire. Chaque zone est liée à une grille des usages indiquant les normes spécifiques et les usages permis. Voyons deux extraits des documents

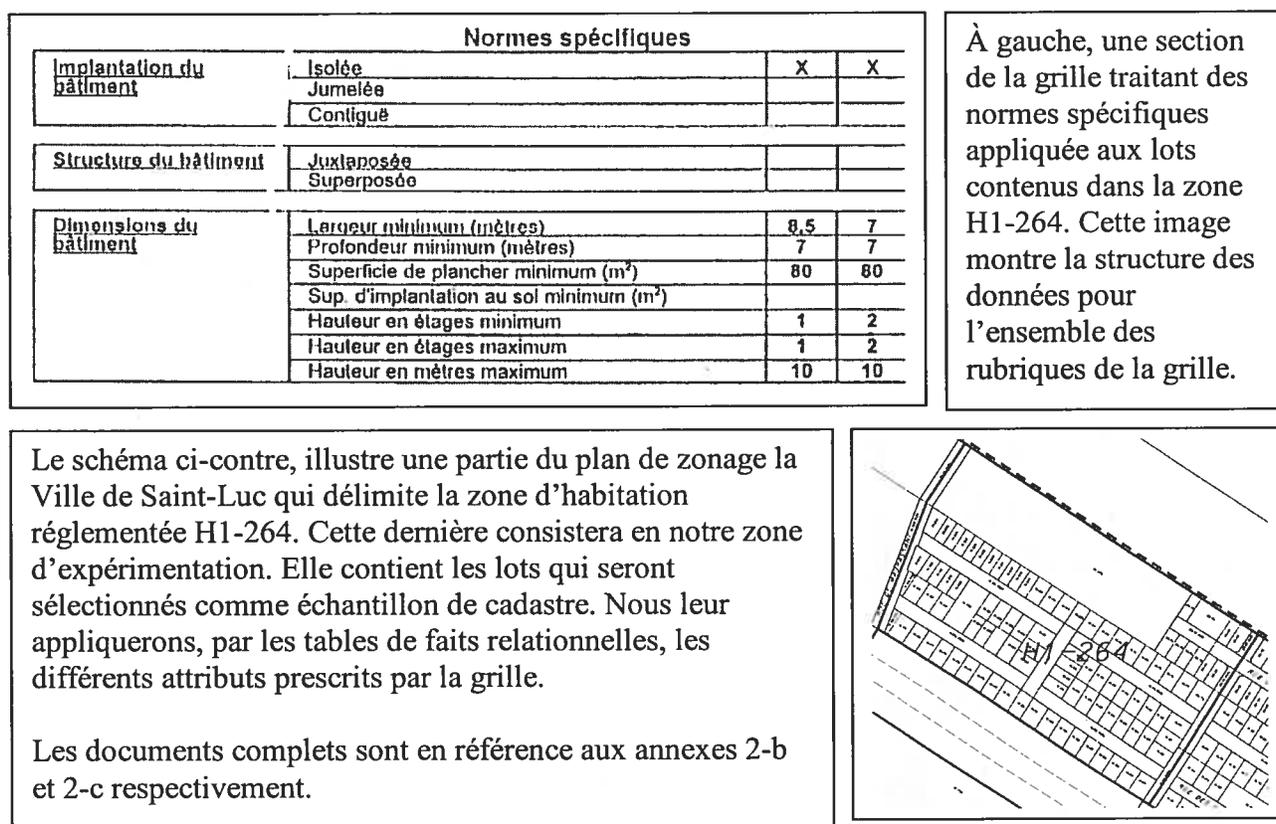


Figure 30 : Extraits de la grille des usages et du plan de zonage complémentaires au cadastre

Les données extraites du cadastre et de la grille des usages seront toutes numérisées à l'intérieur de tables spécifiques au sein du magasin de données. Ces images nous donnent une idée à quel type de données brutes nous aurons à faire face. Il sera question d'en

traduire et d'en documenter le plus efficacement possible, en termes de tableaux, la relation spatiale entre les lots et la zone de réglementation.

Conclusion sur le chapitre:

Nous avons vu dans ce chapitre une présentation de l'approche méthodologique analytique utile pour aborder la problématique de recherche. Cette dernière sera mise en application dans les prochains chapitres. Les données de travail ont également été présentées. Deux formats d'information de base, cartographique et réglementaire, devront être traduits sous formes de tables relationnelles afin d'en permettre une gestion informatique performante. La Ville de Saint-Luc possède déjà son SGBD et son cadastre informatisé. Cependant, l'information qu'il contient n'est pas structurée en fonction d'une opération performante de manipulation des connaissances urbaines. L'idée ici est de reprendre en quelque sorte l'information de base contenue sur la matrice graphique en y joignant des données relatives à la réglementation. Le prochain chapitre est voué à la mise sur pied de cette nouvelle structure de données suivant l'approche analytique.

Chapitre 7: Décomposition analytique de la problématique cadastrale par construction du modèle théorique

Objectif du chapitre:

Rappelons qu'une base de données relationnelle est constituée de plusieurs tables de faits dimensionnelles. Quant à elle, la structure dimensionnelle d'une table de faits est conçue d'ordinaire en quatre étapes soit : par le choix du type de magasin de données dont elle fait partie; de la granularité de cette dernière; des dimensions puis finalement, des faits. Ce chapitre fait la démonstration du modèle théorique de la base de donnée urbaine fictive qui sera appliqué au problème de structuration de l'information de la Ville de Saint-Luc. Chaque source de données sera revue et décomposée afin de pouvoir s'inscrire dans le nouveau processus de forage des données.

7.1 Choix du type de magasin de données et de sa source d'alimentation

Cette tâche consiste à choisir la source de données opérationnelle qui nourrira le magasin de données. Il est plus simple de travailler avec des magasins de données à source unique au départ et de s'assurer que leurs dimensions sont conformes¹⁸. De cette façon, lorsque viendra le temps de les implémenter entre eux, leur connectivité au bus décisionnel¹⁹ de l'entrepôt de données sera assurée.

Le magasin de données à construire sera dans notre cas "multisource" et sera constitué des informations de localisation spatiale et de réglementation d'urbanisme en terme de zonage municipal. Nous l'appellerons donc "le magasin de données du service d'urbanisme" (annexe 1-a). Toutes les informations utiles proviendront de la matrice graphique de la Ville de Saint-Luc ainsi que de la grille des usages prescrits (annexes 2-a et 2-c).

¹⁸ Compatibles avec la granularité de l'information ou le référentiel commun à toutes les tables de l'entrepôt. [Kimbal, 2000]

7.2 Composition des tables de faits du magasin de données

Plusieurs tables de faits sont nécessaires pour le support complet des activités comportant de multiples processus. Chacun des processus engendre à lui seul une ou plusieurs tables. Kimball (2000) affirme que même le magasin de données le plus simple consiste en une série coordonnée de tables de faits dotées d'une structure simple. Plus précisément, on peut dire que le processus de planification urbaine comporte plusieurs étapes. Normalement, la meilleure solution consiste en la construction d'une table des faits pour chacune des étapes du processus (ainsi est constituée la chaîne des tables de faits). Dans le cas à l'étude de la gestion spatiale d'un lot urbain réglementé, il sera question d'une table pour la gestion cadastrale et d'une table pour la gestion des usages du sol (annexe 1-b).

7.2.1 Définition de la granularité des tables (dimensionnelles) de faits

Dans une étape plus technique mais très logique, il est primordial de définir clairement la précision d'un enregistrement au sein de la table des faits dans la dimension concernée. Tel que souligné par Whipple en 1976 et suite au système de gestion de l'information urbaine existant dans les années 1960, l'unité d'enregistrement retenue à la base est la propriété foncière. Voilà donc la granularité de la tables des faits qui comportera autant d'enregistrements qu'il y a de propriétés dans la ville gérée. Il en sera autant pour chacune des dimensions qui verront leurs attributs rattachés à chacune des propriétés enregistrées. Pour être conforme à leur tour, toute table de fait ou dimension supplémentaire qui viendra s'ajouter de façon individuelle ou sous forme de magasin de données devra respecter cette granularité qui définit en fait le bus décisionnel du système. Dans le cas à l'étude une des tables de données concernera la documentation des zones de réglementation des usages et non des lots. Par contre, un lot se rapporte à une zone et, comme nous le verrons, c'est la raison pour laquelle le modèle sera toujours valide.

¹⁹ Conforme à la norme d'échange entre les magasins définies par le grain de l'information stockée.
[Kimball,2000]

Cependant, il existerait une deuxième classe de granularité qui n'est pas à négliger et qui fait toute la différence lors d'opérations de traitement et de manipulation des données. Surtout, quand il est question de forage de données et de requêtes complexes. Il s'agit de la granularité des données elles-mêmes composant les attributs associés à chaque dimension. Plus on montre un niveau de détail fin, plus le système est robuste. L'expérience de Kimball (2000) montre que lorsque les données sont très détaillées, elles réagissent beaucoup mieux à des requêtes inhabituelles et à l'introduction de nouvelles données. Nous verrons plus précisément en quoi tout cela consiste pendant l'établissement des dimensions de la base de données.

7.2.2 Identification des dimensions de base possibles

Le concepteur de la base de données doit avant tout examiner l'ensemble des informations mises à sa disposition et choisir comme dimensions, des descripteurs qui sont de préférence à valeur unique (occurrence simple). La majorité du temps, l'ajout de dimensions est aisée dans la mesure où elle ne modifie pas la granularité de la table des faits. La table des faits d'un modèle dimensionnel consiste en une série de mesures simultanées à un niveau de détail donné. Les mesures les plus intéressantes sont la plupart du temps numériques. Souvent, le concepteur est mis devant le fait accompli d'une série de mesures et doit y déterminer les dimensions qui sont valides (notre cas). La tâche du concepteur consiste donc à assortir cette série de mesures aux moyens de toutes les dimensions qu'il pourra trouver en fonction des besoins des utilisateurs. Dans le présent projet, c'est en fait ce sur quoi il faudra se pencher pour les urbanistes.

7.2.2.1 Faits pour la localisation et l'identification d'un lot

Voici la liste des dimensions sélectionnées pour la construction d'une table des faits de base²⁰ propre à l'identification et à la localisation d'un lot dans l'espace en fonction des informations provenant de la matrice graphique de la Ville de Saint-Luc:

²⁰ Attribut alphanumérique définissant l'entête d'un champs dans une base de données.

Tableau XX:

Faits de base issus de la matrice graphique	
<ul style="list-style-type: none"> • numéro civique • type de voie publique (non montré ci-contre) • nom de la voie publique (non montré ci-contre); • numéro de partie; • numéro de subdivision; • profondeur du lot en m; • frontage du lot en m; • coordonnée en x du centroïde de lot (non visible ci-contre); • coordonnée en y du centroïde de lot (non visible ci-contre); • orientation géographique schématique du lot par son frontage; • <i>application du matricule de la zone de réglementation (référence pour jointure).</i> 	<p>Le diagramme illustre un lot rectangulaire avec les informations suivantes : un numéro civique #92 en haut à droite, une largeur de 18,29 m en haut, une profondeur de 19-20 m à l'intérieur, un matricule circulaire 2129 au centre, et une dimension latérale de 35,26 m à droite.</p>

L'ensemble des dimensions énumérées ci-haut reflète l'information de base affichée sur le cadastre, qu'elle soit écrite ou tracée. Par contre, certaines informations sont en trop lorsqu'on parle d'une intégration à une base de données. Ces dernières pourront être obtenues par l'exécution d'opérations mathématiques sur les faits de base. Les coordonnées (x,y) complètes du centroïde de lot, sont disponibles dans la table d'informations numériques accompagnant la géométrie du cadastre géré par le logiciel AutoCAD Map 2000. La dimension du matricule de la zone de réglementation a été ajoutée pour permettre un lien plus facile entre le lot et les usages qui s'appliquent. Tel qu'expliqué au chapitre cinq (Guité, Parisel 2000), ce champ supplémentaire joue le rôle de lien hiérarchique à la zone de réglementation qui est l'entité supérieure de rattachement du lot cadastral. Notons la différence de granularité de l'information d'une zone de réglementation qui peut contenir plusieurs lots. Rappelons-nous qu'en ce sens, l'ajout de telles informations n'affectera pas la performance du magasin de données.

7.2.2.2 Faits pour la documentation des usages prescrits

Voyons la liste des dimensions sélectionnées pour la construction d'une table des faits de base propre à la réglementation municipale régissant un lot dans l'espace en fonction des informations provenant de la grille des usages prescrits (Annexe "A" du règlement zonage No 870) de la Ville de Saint-Luc qu'on retrouve en référence en annexe 2-c du présent document:

Tableau XXI:

Faits de base issus de la grille des usages prescrits
<ul style="list-style-type: none"> • Code de la zone (matricule identificateur); • Code d'utilisation du sol (une lettre et un chiffre); • structure d'implantation des bâtiments permise en fonction des usages; • largeur minimale du bâtiment autorisée en mètres; • profondeur minimale du bâtiment autorisée en mètres; • hauteur du bâtiment minimale autorisée en nombre d'étages; • hauteur du bâtiment maximale autorisée en nombre d'étages; • hauteur du bâtiment maximale autorisée en mètres; • marge de recul avant minimale permise en mètres selon la structure d'implantation; • profondeur du terrain minimale autorisée en mètres; • marge de recul latérale minimale permise en mètres selon la structure d'implantation; • marge de recul arrière minimale permise en mètres selon la structure d'implantation; • largeur du terrain (façade) minimale autorisée en mètres; • profondeur du terrain minimale autorisée en mètres; • dispositions spéciales si applicable (référence aux articles de réglementation).

Comme nous le verrons, ici encore, nous n'avons pas inscrit l'ensemble des champs de données apparaissant sur le document de base qui supporte l'information. Certaines données pouvant être obtenues à partir des autres, nous les retrouverons dans la liste des faits dérivés propres à des requêtes mathématiques. De plus, pour la numérisation efficace de la grille des usages prescrits nous devons faire face à la problématique des occurrences multiples (plusieurs entrées pour un même champ de donnée). La solution sera décrite lors de la numérisation de l'information.

7.2.3 Identification des faits dérivés

Pour les deux tables de base décrites dans la section précédente, nous traiterons deux types de faits dérivés. Le premier réunit les faits qui sont additifs et proviennent de calculs basés sur les autres faits de base du même enregistrement de table des faits. Le deuxième type de fait dérivé repose sur des calculs non additifs (ratios, association, faits cumulatifs). Ces faits ne peuvent ici être visionnés de façon individuelle sur une table stockée. Ils doivent être calculés au moment d'une requête par l'outil de requête sous forme d'extension virtuelle de la table des faits de base.

Voici les *faits dérivés*²¹ qui seront obtenus à partir des faits de base (issus de requêtes SQL) dans le cadre de la table d'identification et de la localisation spatiale d'un lot:

Tableau XXII :

Faits dérivés issus de requêtes sur la table de localisation d'un lot
<ul style="list-style-type: none"> • Matricule d'emplacement du lot (non additif) (par association); • Adresse civique complète du lot (non additif) (par association); • Numéro du lot complet (non additif) (par association); • Superficie du lot (additifs) (par altération mathématique); • Attribut d'adjacence d'un lot voisin (additifs) (par altération mathématique).

Comme nous pouvons le constater, plusieurs informations qui étaient déjà présentes sur les cadastre se retrouvent sous forme de faits dérivés. Il est question d'optimisation du stockage numérique de l'information. Les méthodes d'obtention et leurs raisons seront décrites lors de la formulation des requêtes.

Maintenant, le choix des faits dérivés obtenus à partir des faits de base (issus de requêtes SQL) du magasin de données propre au service d'urbanisme dans le cadre de la table sur la réglementation de zonage:

Tableau XXIII :

Faits dérivés issus de requêtes sur la table de réglementation des usages
<ul style="list-style-type: none"> • coefficient d'emprise au sol minimal (additif) (par altération mathématique); • coefficient d'occupation du sol minimal (additif) (par altération mathématique); • marges latérales totales (additif) (par altération mathématique); • Superficie de terrain minimale (additif) (par altération mathématique); • Superficie de plancher minimale (additif) (par altération mathématique).

Encore une fois ici, pour des raisons d'optimisation, nous voyons des informations qui apparaissent sur le document de base. En plus, l'ajout d'une information supplémentaire montre qu'il est possible de formuler de nouveaux champs de données au gré des besoins du gestionnaire urbain. Bien sûr, il est certainement possible d'en obtenir davantage. Il ne s'agit dans le cadre de l'expérience, que du traitement d'un échantillon montrant une

²¹ Champs de données obtenus par la combinaison ou par l'altération mathématiques des faits de base.

exploitation plus performante de l'information disponible dans les systèmes d'information urbaine.

7.3 Visualisation des modèles dimensionnels construits

Pour gérer le projet de modélisation dimensionnelle de façon efficace, l'information de conception doit circuler facilement. Nous devons être en mesure d'avoir une image visuelle du système construit. La cartographie de la base de données nous permet d'être en contrôle du système d'information simplement en nous montrant quelles informations sont en relation, où sont elles, quelle est leur signification. Sans quoi, à force d'accumuler et d'introduire de nouveaux tableaux d'informations, on pourrait rapidement se perdre dans une jungle de données numériques.

7.3.1 Plan de composition d'un entrepôt de données municipal fictif

Dans le plan d'entrepôt qui est montré en annexe 1-a, seulement le magasin de données lié à l'urbanisme est projeté avec des tables de faits possibles. Nous travaillerons plus particulièrement sur deux de ces tables soient celles de l'identification et de la localisation d'un lot et celle concernant la réglementation de zonage.

7.3.2 Détail relationnel entre les tables de faits traitées

Le détail relationnel des deux table des faits en annexe 1-b, nous montre la liste complète des dimensions des tables. Nous y retrouvons les faits de base physiquement présents dans la table et les faits dérivés qui seront présentés suivant l'interface du SGBD (faits issus de requêtes). On y présente également les relations entre les dimensions de jointure d'une table à l'autre. Ce diagramme illustre bien de quelle façon les tables pourront dialoguer entre elles.

7.3.3 Arbre de hiérarchie des attributs dimensionnels et des requêtes exécutées

Cet arbre hiérarchique nous fournit une vue d'ensemble des différentes relations entre les attributs de base et dérivés. On parle ici de la définition des chemins de forage par défaut. Les informations véhiculées par le schéma en question comprennent le nom de l'attribut, le chemin des informations desquelles il découle et le chemin vers les attributs auxquels il donne naissance. On trouvera les grilles respectives illustrant le plan de forage de données pour la table sur l'identification et la localisation d'un lot en annexe 1-c et pour la table sur la réglementation de zonage en annexe 1-d.

7.4 Le dictionnaire de données

Ce dictionnaire devrait être en fait, le lieu unique de stockage des informations pilotant les processus de l'entrepôt. Toutes les procédures internes de l'entrepôt de données (du modèle initial jusqu'à la navigation et l'accès aux données) devraient faire appel au dictionnaire des métadonnées. Ma conclusion sur la question rejoint celle de Kimball (2000) qui soutient qu'à l'heure actuelle, une mise en oeuvre aussi parfaite est impossible. On nous suggère tout de même de se procurer un outil qui nous permettra de suivre en partie ces métadonnées et qui nous aidera au moins à les gérer.

Une autre étape de documentation des données manipulées vient donc s'ajouter à la précédente. C'est une étape qui sert à la génération de métadonnées ayant pour but de localiser et de décrire des données dans le cadre de la conception d'un entrepôt de données entier. Vu que nos informations n'ont pas à être collectées et qu'elles sont déjà disponibles, nous n'auront pas à retracer leurs origines. Du moins, tel n'est pas l'objectif de la recherche. Il y a pourtant matière à documenter davantage les dimensions de base au niveau d'un catalogue de métadonnées. Dans le cas à l'étude, les informations suivantes suffiront à la documentation de la base de données:

Tableau XXIV :

Métadonnées sur les dimensions de base et dérivées

Métadonnées relatives aux faits de base :

- **nom de l'attribut** : il doit s'agir du nom officiel dans le cadre de l'activité concernée.
- **type de donnée** : le format de l'attribut sera-t-il textuel ou numérique ou monétaire?
- **définition de l'attribut** : description brève mais significative du champ de données.

Métadonnées plus propices aux faits dérivés :

- **Le nom**: nom sous lequel le fait sera connu de tous et des outils de traitement.
- **Les règles d'agrégation**: règles d'agrégation associées au fait par défaut. Dans certains cas on peut additionner des résultats de calculs, en d'autres cas, on doit faire l'agrégation séparément et recalculer le fait dérivé final.
- **La description**: une brève description du fait.
- **Transformations**: les calculs spécifiques à effectuer lors de la création du fait.

[Kimball, 2000]

L'outil utilisé dans notre expérimentation étant *Microsoft Access*, on est déjà en mesure de supporter automatiquement la gestion de plusieurs des métadonnées citées ci-haut. nous verrons brièvement, à travers l'interface de création des tables de faits de base et des guides de saisie (Chap. 8), les métadonnées prises en considération. Pour ce qui est des faits dérivés nous élaborerons en annexe 3, une description du code des requêtes qui jouera le même rôle qu'un dictionnaire de données. Les auteurs du livre *Concevoir et déployer un Data Warehouse* soutiennent que nous pourrions passer des années à traquer le moindre problème de conflit entre les données. Il s'agit plutôt d'identifier les problèmes majeurs et d'en dresser une priorité de résolution. L'expérience leur dicte que la résolution de certains problèmes peut découler de la disparition de certains autres.

Conclusion sur le chapitre :

Dans ce chapitre, nous avons vu , à travers le processus de forage de données, comment modéliser un système de gestion de base de données au niveau de la structure qu'il aura sous sa forme informatique. De l'entrepôt entier en passant par le magasin, les tables et les métadonnées, nous avons traduit ces méthodes selon nos besoins par une décomposition analytique de notre étude de cas sur la Ville de Saint-Luc. Le prochain chapitre illustrera ces méthodes par la réalisation d'un prototype informatique fonctionnel basé sur ces modèles à l'aide du logiciel SGBD *Microsoft Access*.

Chapitre 8: Restitution analytique des éléments du cadastre par validation informatique du modèle dimensionnel

Objectifs du chapitre :

Nous appliquerons les notions précédentes à la construction d'une partie du magasin de données relatif au service d'urbanisme qui viendrait s'inscrire dans un entrepôt de données municipal fictif. La numérisation des informations sera exécutée à partir de documents de base réels (matrice graphique et grille des usages prescrits) qui seront adaptés au canevas informatique du prototype de SGBD urbain.

L'outil informatique de validation, *Microsoft Access*, prend en charge plusieurs détails (tels que la plupart des métadonnées mentionnées précédemment) issus de détails associées à la construction des bases de données. Ceci nous permettra alors de nous concentrer sur les aspects importants de la problématique d'optimisation et d'intégration des données relatives à l'identification et à la localisation d'un lot cadastral réglementé. Nommons la segmentation des données ainsi que la formulation des requêtes d'information. Plus précisément, le logiciel ne fera que refléter l'efficacité des requêtes SQL appliquées aux faits de base pour l'obtention des faits dérivés, tel que prescrit par notre modèle théorique.

Dans le cadre de cette expérimentation, rappelons les trois niveaux d'application de la granularité des informations avant la construction des tables:

Tableau XXV :

Les trois niveaux d'application de la granularité de l'information traitée
1: les granularités du magasin de données (« multisource » en fonction des départements).
2: la granularité de la table des faits (se rapportant au lot ou à la propriété).
3: la granularité de l'information elle-même (en considérant un groupe de caractères cibles dans la cellule d'un champ de données).

8.1 Construction des tables de données informatisées à partir des modèles

Dans un premier temps, construisons les tables de faits de base appelées “**baseloclot**” (pour table de localisation des lots) et «**usage_zone**» (pour la table de réglementation de zonage) au sein d’une nouvelle base de données appelée “**dataurb**” (pour le magasin de données du service d’urbanisme), à l’aide de l’utilitaire *Microsoft Access*. Selon le nom des dimensions qui ont été choisies précédemment dans le modèle théorique, il est important de refléter, dans la table de données informatique, des en-têtes de champs qui soient toujours significatives. Un code mnémonique sera ici utilisé pour aider l’utilisateur à s’y retrouver. En fait, ces en-têtes doivent également être courtes pour une question de légèreté de gestion informatique et surtout lorsqu’on doit réutiliser ces noms de champs dans le code de requêtes SQL.

8.1.1 Guide de saisie des données pour les faits retenus sur le cadastre

- **numéro civique = no_civ** : on saisit la donnée en entrant le nombre représentant le numéro en question. Considéré comme du texte, donc une valeur non additive (car des adresses civiques ne s’additionnent pas), on prend soins de mettre une barre de soulignement après notre séquence. Sa séparation d’avec les autres composantes de l’adresse nous assure une granularité d’information plus fine et un tri des données plus puissant en vue de nouvelles requêtes. C’est une procédure qui, comme nous le verrons à la définition des requêtes, nous permettra d’obtenir des résultats plus cohérents. Il ne faudra pas oublier, lors d’une requête de sélection par tri, d’intégrer cette même barre de soulignement pour la définition des paramètres (ex: sélection d’une donnée quelconque où no_civ= 640_). Cette barre sert simplement de tampon entre les associations de champs.

- **type de voie publique = typ_voi** : on inscrit le type de voie publique (rue, avenue...) suivi d’une barre de soulignement pour les même raisons que le champ précédent.

- **nom de la voie publique = nom_voi** : on inscrit le nom donné à la voie publique suivi de la barre de soulignement pour les mêmes raisons.
- **numéro de partie = no_part**: ce deuxième chiffre correspond à la partie de la division cadastrale sur laquelle on trouve le lot. Non additif, n'oublions pas la barre de soulignement à la suite. Sa séparation d'avec les autres composantes du numéro de lot nous assure également un tri des données en fonction de ce champ si désiré.
- **numéro de subdivision = no_sub**: correspond à la subdivision de la partie du cadastre. Ce chiffre aussi est suivi d'une barre de soulignement. Sa séparation d'avec les autres composantes du numéro de lot nous assure ici un tri des données en fonction de ce champ (non additif).
- **profondeur du lot en mètres = prof_lot**: on entre ici la valeur numérique proprement dite en mètres (et deux décimales) qui peut être impliquée dans plusieurs calculs ultérieurs. Elle est additive et correspond normalement à la longueur du lot.
- **façade ou frontage du lot en mètres = fron_lot**: on entre ici la valeur numérique proprement dite en mètre (et deux décimales) qui peut être impliquée dans plusieurs calculs ultérieurs et est donc additive (correspond normalement à la largeur d'un lot).
- **coordonnée en x du centroïde de lot = centro_x**: valeur numérique cartésienne en x correspondant à la situation du lot en mètres en son centre. C'est également une valeur de repère additive. On saisit la donnée avec trois décimales de précision.
- **coordonnée en y du centroïde de lot = centro_y**: valeur numérique cartésienne en y correspondant à la situation du lot en mètres en son centre. C'est également une valeur de repère additive. On saisit la donnée avec trois décimales de précision.
- **orientation schématique du lot par sa façade = ori_fron_lot**: indice sur la disposition des éléments cartographiques. Cette valeur textuelle indique la direction

schématique, sur la rose des vents, vers laquelle la façade du lot est orienté sur la carte. On ne saisit ici que les lettres en abrégiation au deuxième degré (ex: N-O).

Pourquoi des barres de soulignement? Pour éviter toute confusion avec le signe de la différence ou le tiret (-).

Relativement à la gestion cartographique des lots, les mesures et les références spatiales s'effectueront en 2D dans un plan cartésien (donc en coordonnées planes) et ce, suivant la projection Mercator transverse modifiée. Le tout, selon la norme de découpage en fuseaux du North American Datum 1983 (Québec MTM zone 8).

Rappelons, comme nous l'avons vu plus tôt, que toutes les informations qui viennent d'être mentionnées sont disponibles sur le plan de cadastre. La fenêtre suivante nous montre la table des faits de base à son état de confection avec les métadonnées éditables que chaque champ de données (chaque fait) implique et qui sont déjà gérées par le logiciel *Microsoft Access* (il es ici question du nom du champ, du type de données qu'il contient ainsi que d'une brève description textuelle de ce dernier):

Field Name	Data Type	Description
no_civ	Text	numéro de porte correspondant à l'adresse civique de la propriété
typ_voi	Text	le préfixe de la voie correspondant à son rang dans la hiérarchie routière
nom_voi	Text	appellation courante de la voie
no_part	Text	identificateur alphanumérique de la partie du cadastre
no_sub	Text	identificateur alphanumérique de la subdivision
prof_lot	Currency	dimension du lot en longueur en mètres
fron_lot	Currency	dimension du lot en largeur en mètres
centro_x	Currency	coordonnées exacte en m sur l'axe des x dans la projection MTM NAD 83
centro_y	Currency	coordonnées exacte en m sur l'axe des y dans la projection MTM NAD 83
ori_fron_lo_schem	Text	orientation géographique schématique du lot sur le cadastre par son frontage
code_zone	Text	zone dans laquelle se situe le lot sur lequel sont appliqués les règlements de zonage

illustration: *Microsoft Access*

Figure 31: Construction numérique de la table des faits sur l'identification et la localisation d'un lot

Voici, en deux parties, le tableau « Individus-Variables » miniature concernant huit lots voisins (des rue Dextrateur et Des Geais-Bleus) et à partir du quel nous exécuterons nos requêtes. Les lots ont été sélectionnés sur la matrice graphique informatisée de la Ville de Saint-Luc (annexe 2-a).

baseloclot : Table						
	no_civ	typ_voi	nom_voi	no_part	no_sub	prof_lot
▶	85_	Rue_	Dextrateur	19_	78	35,36
	87_	Rue_	Dextrateur	19_	79	35,36
	88_	Rue_	Dextrateur	19_	92	35,36
	89_	Rue_	Dextrateur	19_	80	35,36
	90_	Rue_	Dextrateur	19_	91	35,36
	92_	Rue_	Dextrateur	19_	90	35,36
		Rue_	Des Geais-Bleus	19_	124	35,05
		Rue_	Des Geais-Bleus	19_	125	35,05

Record: 1 of 8

baseloclot : Table					
	fron_lot	centro_x	centro_y	ori_fron_lo_sc	code_zone
▶	22,86	322420,651	5024273	S-O	H1-264
	21,34	322401,651	5024285	S-O	H1-264
	19,81	322386,651	5024229	N-O	H1-264
	21,34	322383,651	5024297	S-O	H1-264
	19,81	322369,651	5024240	N-O	H1-264
	18,29	322353,651	5024250	N-O	H1-264
	19,81	322366,651	5024201	S-O	H1-264
	18,25	322350,651	5024211	S-O	H1-264
*	0	0	0		

Record: 1 of 8

illustrations : *Microsoft Access*

Figure 32: Illustration logicielle de la table des faits de localisation d'un lot en deux parties

Soulignons qu'un champ « ID » est créé par le SGBD (*Microsoft Access*) pour jouer le rôle de clé primaire de sorte à ce qu'il puisse reconnaître les enregistrements lorsqu'il effectue des croisements d'une table à l'autre. C'est un nombre incrémenté automatiquement. Nous pourrions définir notre propre clé primaire comme par exemple le matricule du lot dans ce cas-ci, qui serait la clé de jointure unique entre toutes les tables de l'entrepôt. Pourquoi unique? Parce que deux propriétés ne peuvent être au même endroit.

8.1.2 Guide de saisie des champs de données pour la grille des usages

- **code de la zone = code_zone** : on entre l'identificateur alphanumérique de la zone de réglementation telle qu'elle est inscrite sur le plan de zonage. C'est un identificateur unique qui permettra de retracer la zone voulue dans la base.

Considérée comme du texte, on ne peut y appliquer des opérations mathématiques. La répétition de l'enregistrement de ce code une deuxième fois est la réponse fonctionnelle aux occurrences multiples. Notons sur la grille en annexe 2-c, les deux modalités (colonnes) d'utilisation du sol pour la même zone.

- **usage du sol permis = code_usage** : On entre la première lettre de la famille de l'usage permis suivi du sous code numérique d'utilisation du sol. Dans ce cas-ci, « H1 » pour « Habitation Unifamiliale » et « P1 » pour « Public Parcs » tel qu'on peut le voir sur la grille (annexe 2-c). Les deux enregistrements dans le même champ d'occurrence séparés par une virgule. Permettons nous l'occurrence multiple dans ce champs d'information de référence seulement.
- **structure d'implantation du bâtiment = struc_implan_bat** : en remplacement du « x » inscrit sur la grille, on inscrit directement le choix d'implantation sous forme textuelle. Dans le cas présent, « isolée ».
- Pour les champs suivants, on entre la valeur unique qui est inscrite sur la grille pour chaque rubrique désignée sous un format numérique :
 - hauteur minimale du bâtiment en nombre d'étages = haut_bat_min_etag**
 - hauteur maximale du bâtiment en nombre d'étages = haut_bat_max_etag**
 - hauteur maximale du bâtiment en mètres = haut_bat_max_m**
 - marge de recul avant minimale en mètres = marg_av_min_m**
 - marge de recul avant maximale en mètres = marg_av_max_m**
 - marge de recul latérale minimale en mètres = marg_lat_min_m**
 - marge de recul arrière minimale en mètres = marg_ar_min_m**
 - largeur du terrain minimale en mètres = larg_ter_min_m**
 - profondeur du terrain minimale en mètres = prof_ter_moy_m**
- **dispositions spéciales = dispo_spec** : ce champ englobe les rubriques sous *divers* et *amendement* dans la grille. On entre en texte de référence toute note ou numéro

d'article de réglementation municipale pouvant influencer l'application du zonage prescrit dans le secteur.

Les informations ci-haut mentionnées sont toutes disponibles sur la grille des usages permis de la Ville de Saint-Luc (annexe 2-c). Plusieurs d'entre elles ont été escamotés car elles n'étaient pas utiles à la saisie (pas d'information disponible) ou parce qu'elles se retrouveront sous forme de requêtes plutôt qu'en stockage tabulaire. Voyons alors la fenêtre de construction de la table sous le format *Microsoft Access*.

Field Name	Data Type	Description
code_zone	Text	identifiant de la zone
code_usage	Text	définit le type d'usage du sol par un code alphanumérique
struc_implan_bat	Text	structure d'implantation des bâtiments permise en fonction des usages
larg_bat_min_m	Currency	largeur minimale du bâtiment autorisée en mètres
prof_min_bat_m	Currency	profondeur minimale du bâtiment autorisée en mètres
haut_bat_min_etag	Currency	hauteur du bâtiment minimale autorisée en nombre d'étages
haut_bat_max_etag	Currency	hauteur du bâtiment maximale autorisée en nombre d'étages
haut_bat_max_m	Currency	hauteur du bâtiment maximale autorisée en mètres
marg_av_min_m	Currency	marge de recul avant minimale permise en mètres selon la structure d'implantation
marg_av_max_m	Currency	marge de recul avant maximale permise en mètres selon la structure d'implantation
marg_lat_min_m	Currency	marge de recul latérale minimale permise en mètres selon la structure d'implantation
marg_ar_min_m	Currency	marge de recul arrière minimale permise en mètres selon la structure d'implantation
larg_ter_min_m	Currency	largeur du terrain (façade) minimale autorisée en mètres
prof_ter_moy_m	Currency	profondeur du terrain minimale autorisée en mètres
dispos_spec	Text	dispositions spéciales selon l'usage et la structure d'implantation - référence à la réglementation municipale

illustration : *Microsoft Access*

Figure 33: Construction numérique de la table des faits sur la réglementation de zonage

Voyons maintenant la table de données en deux parties qui résulte de la construction précédente.

code_zone	code_usage	struc_implan_bat	larg_bat_min_m	prof_min_bat_m	haut_bat_min_etag	haut_bat_max_etag	haut_bat_max_m
H1-264	H1,P1	isolée	8,5	7	1	1	10
H1-264	H1,P1	isolée	7	7	2	2	10

marg_av_min_m	marg_av_max_m	marg_lat_min_m	marg_ar_min_m	larg_ter_min_m	prof_ter_moy_m	dispos_spec
6	8	1,5	7,5	15	30	
6	8	1,5	7,5	15	30	

illustration : *Microsoft Access*

Figure 34: Illustration logicielle de la table des faits de base créée en deux parties

Encore ici, un champ « ID » (non montré) est créé par *Microsoft Access* et joue le rôle de clé primaire. Il assure ainsi les jointures avec d'autres tables. Nous aurions pu nous-même définir notre clé primaire en créant un identificateur unique pour la zone réglementée. Par exemple, un nom de zone plus complet qui raffine l'occurrence. Maintenant que nous avons construit nos deux tables d'échantillonnage du magasin de données du service d'urbanisme, nous sommes en mesure de procéder avec le plan de forage des données.

8.2 Conception des requêtes SQL (plan de forage des données)

Dériver des faits ou obtenir de nouvelles informations à l'aide de requêtes peut être considéré comme paramétrer l'information. Lorsque les données contenues dans les faits de base seront mises à jour, les faits dérivés seront automatiquement à jour à leur prochain calcul. On pourrait également penser aux calculs inverses et reconstituer des faits de base à partir des résultats d'une requête comme par exemple reconstituer les coordonnées (x, y) d'un centroïde de lot par la décomposition de son matricule.

Dans le cadre de cette base de données miniature, il sera toujours question de réaliser les requêtes avec un affichage de données supplémentaire qui permettra leur association avec l'enregistrement de base respectif dans la table. De cette façon, on obtiendrait toujours toutes les informations sur un lot ou une zone à chaque lancement de requête. Les faits dérivés sélectionnés plus tôt n'existent pas réellement dans la table de données. Ce sont en fait des instructions d'accès mémorisées (sous forme de requêtes) qui sont associées à cette table. Les informations nécessaires, livrées par la requête, sont alors disponibles sur demande de l'utilisateur. Le côté pratique est qu'il est bien moins exigeant pour la mémoire disque de stocker quelques lignes d'instructions qu'un nouveau champ de données pour vingt-cinq mille propriétés dans une ville. L'objectif est, dans la mesure du possible, de réduire au minimum les faits de base nécessaires à l'obtention d'information.

Afin d'assurer les liens entre les tables, nous utiliserons en mode « création » la possibilité d'utiliser des pointeurs entre les tables par l'identificateur d'enregistrement

« ID » créé automatiquement par le logiciel. Dans l'image suivante, on voit l'interface de gestion des pointeurs entre la table « **baseloclot** » et les requêtes de « **numéro de lot** » et de « **matricule du lot** » par ce champ « ID ». Comme nous le verrons, ces dernières viendront compléter l'affichage de la requête « **adresse civique complète** ».

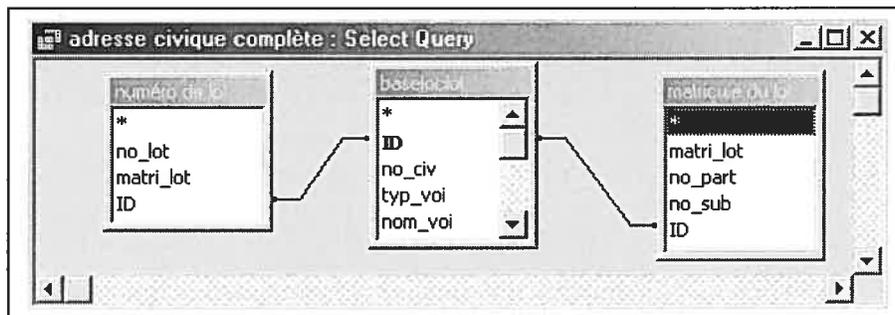
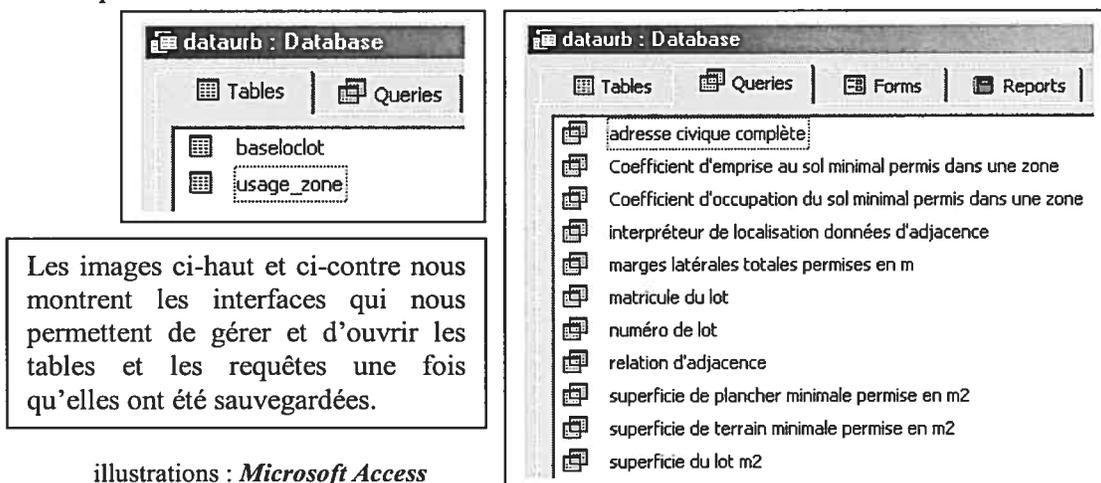


illustration : *Microsoft Access*

Figure 35 : Gestion des pointeurs entre les champs par *Microsoft Access*

Notons que les requêtes construites seront ensuite stockées par nom dans une table à cet effet, permettant à l'utilisateur de les appeler au besoin. Voyons le gestionnaire de tables et de requêtes.



illustrations : *Microsoft Access*

Figure 36 : Les gestionnaires de tables et de requêtes au sein de la base de données

8.2.1 Cours SQL (Structured Query Language) accéléré

Voici un cours accéléré de SQL pour mieux comprendre les requêtes formulées ci-après. À la base, chaque requête est divisée en trois sections (on sélectionne une liste d'attributs

à partir d'une table cible où une condition de sélection doit être rencontrée). On peut en augmenter la complexité par l'utilisation d'opérateurs et de fonctions mathématiques :

```
SELECT <liste d'attributs cibles>
FROM   <liste des tables de données sources>
WHERE  <condition de sélection>.
```

Figure 37 : Structure du code informatique d'un énoncé de requête SQL

Ce langage de programmation se lit de gauche à droite comme du texte. Chaque énoncé est séparé par une virgule (,) et un point virgule (;) termine le code d'une requête.

8.2.2 Formulation des requêtes pour l'obtention des faits dérivés

Dans le cadre de notre base de données miniature qui consiste en fait en une table de données (qui elle fait partie d'un magasin, qui lui, fait partie de l'entrepôt), on veut livrer l'information cadastrale la plus complète possible. On veut de plus, fournir un maximum d'information géographique sur un lot et la réglementation de zonage qui lui est appliquée sans nécessairement avoir recours à une carte pour le localiser. On parle alors de la filtration des informations vitales.

Voyons les requêtes retenues pour les deux tables créées ainsi qu'un exemple de résultat informatique de leur exécution. Le code SQL de chacune des requêtes est décrit et commenté en annexe 3.

8.2.2.1 Journal des requêtes sur la localisation d'un lot

1. Matricule d'emplacement du lot "matri_lot": la clé de jointure (unique) universelle entre toute les tables de l'entrepôt de données municipale. Identificateur géographique auquel devrait se rapporter toute information se rattachant à un lot. Il vaut mieux avoir en mémoire ces quelques lignes d'instructions plutôt que des milliers de matricules stockés en permanence dans un champ de données au sein du système de gestion (tel est le cas

pour la base de données SIURS des valeurs locatives de la Ville de Saint-Luc). Voici en résultats les huit matricules générés pour nos huit lots (on pourrait même faire l'exercice à la main avec les instructions du chapitre trois). Remarquons l'affichage des autres champs de la tables pour assurer un lien futur entre les données :

	matri_lot	no_part	no_sub	ID
▶	2224422073	19_	78	1
	2224420185	19_	79	2
	2224328629	19_	92	3
	2224328397	19_	80	7
	2224326940	19_	91	8
	2224325350	19_	90	9
	2224326601	19_	124	10
	2224325011	19_	125	11

Record: 1 of 8

illustration: *Microsoft Access*

Figure 38 : Résultats informatiques de la requête "matricule du lot"

2. Numéro du lot "no_lot" : nous livre le numéro du lot tel qu'il serait écrit sur un plan de cadastre en réunissant ses deux composantes de base (le numéro de partie et le numéro de subdivision). Encore ici nous affichons le "matricule du lot" et l'identificateur d'enregistrement automatique du logiciel pour des références futures par d'autres requêtes.

	no_lot	matri_lot	ID
	19_78	2224422073	1
	19_79	2224420185	2
	19_92	2224328629	3
	19_80	2224328397	7
	19_91	2224326940	8
	19_90	2224325350	9
	19_124	2224326601	10
▶	19_125	2224325011	11

Record: 8 of 8

illustration: *Microsoft Access*

Figure 39 : Résultats informatiques de la requête "numéro de lot"

3. Adresse civique complète du lot "adres_civ_comp" : cette requête d'union simple nous livre l'adresse du lot comme on l'écrirait sur une lettre. Le tout, en réunissant les trois composantes de base (le numéro civique, le type de voie publique et le nom de la voie publique). On y ajoute encore cette fois-ci l'affichage des champs produits par d'autres requêtes afin de faciliter la compréhension des résultats en fonction du cadastre (ici: le matricule du lot et le numéro de lot). La méthode des pointeurs décrite plus haut est utilisée. Remarquez que les lots vacants n'affichent pas de numéro civique dans le résultat.

	adres_civ_comp	matri_lot	no_lot
	85_Rue_Dextradeur	2224422073	19_78
	87_Rue_Dextradeur	2224420185	19_79
	88_Rue_Dextradeur	2224328629	19_92
	89_Rue_Dextradeur	2224328397	19_80
	90_Rue_Dextradeur	2224326940	19_91
	92_Rue_Dextradeur	2224325350	19_90
	Rue_Des Geais-Bleus	2224326601	19_124
▶	Rue_Des Geais-Bleus	2224325011	19_125

illustration: *Microsoft Access*

Figure 40 : Résultats informatiques de la requête "adresse civique complète du lot"

4. Superficie du lot "superf_lot": calcule, à partir de la profondeur et de la façade de la propriété, sa superficie. Les résultats sont obtenus en mètres carrés pour les lots correspondants. Cette requête évite encore l'ajout d'un fait de base à la table de données qui concernerait la superficie d'un lot à chaque enregistrement. Par liens pointeurs, on ajoute les requêtes du « numéro de lot » et de « matricule du lot » pour faciliter la référence au cadastre.

no_lot	superf_lot	matri_lot
19_78	808,33	2224422073
19_79	754,58	2224420185
19_92	700,48	2224328629
19_80	754,58	2224328397
19_91	700,48	2224326940
19_90	646,73	2224325350
19_124	694,34	2224326601
19_125	639,66	2224325011

illustration: *Microsoft Access*

Figure 41 : Résultats informatiques de la requête "superficie du lot"

5. Relation d'adjacence entre deux lots: en fonction des « delta x » et « delta » y entre les centroïdes des lots, cette requête paramétrée (matricules entrés dans deux boîtes de dialogue par l'utilisateur au lancement la requête) nous donne l'emplacement d'un lot par rapport à un autre, à savoir si le premier lot est au nord, sud , est, ou ouest de l'autre lot choisi et s'il y est adjacent en x et en y. La première partie de la requête consiste en la génération de données de travail qui seront interprétées par la suite pour l'utilisateur par la sous-requête suivante qu'il peut activer au départ sans tenir compte de celle-ci.

Dx	Dy	Dx_comp	Dy_comp	matricule 1	matricule 2
-16	10	19,05	35,36	2224325350	2224326940

illustration: *Microsoft Access*

Figure 8.2.8 : Résultats (données de travail) de la requête sur les attributs de "relation d'adjacence"

6. Interpréteur de données d'adjacence : les résultats des deux lots comparés dans l'exemple sont concluants. Le lot 19-90 sur l'image (en fonction du matricule 1) est bel et bien adjacent (en x et en y) à son voisin et se situe au nord ouest de ce dernier, le lot 19-91. Ce qui est intéressant avec cette requête c'est que même sans carte sous la main on peut rapidement schématiser l'emplacement des lots d'un voisinage. En combinant cette information avec l'orientation de la façade, on est en mesure de schématiser la disposition

des voies publiques associées à ces lots sélectionnés. Cette requête évite le stockage en permanence d'une matrice logique de voisinage pour chaque lot enregistré. Le calcul d'adjacence entre les lots se fait sur demande de l'utilisateur et est rendue sous forme de tableau de données recodé logique (voir le code en annexe 3).

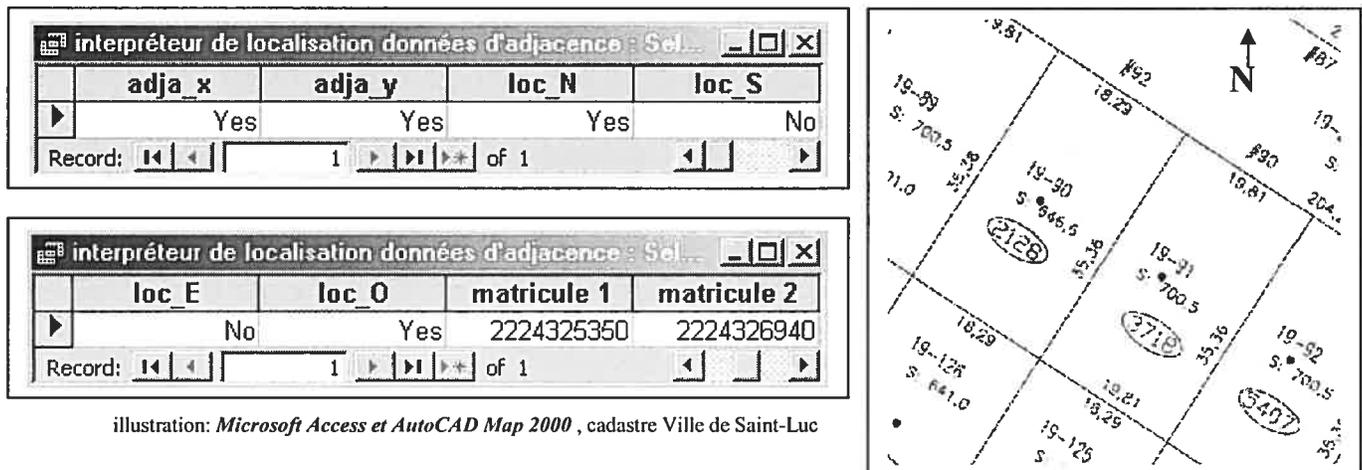


Figure 43 : Résultats informatiques de la requête d'interprétation des attributs d'adjacence

8.2.2.2 Journal des requêtes sur la réglementation de zonage

7. Marges latérales totales permises en mètres « marg_lat_ot_min_m » : cette requête est une simple opération de calcul qui permet d'épargner l'espace mémoire du disque en fabricant l'information au besoin. Sur la grille des usage (annexe 2-c), la donnée sur les marges latérales totales est disponible. Nous l'escamotons pour la transformer en requête. On peut l'obtenir à partir de la donnée de base qui est disponible. On parle de la marge latérale minimale permise que l'on aura qu'à multiplier par deux pour les deux côtés du terrain. C'est aussi une requête paramétrée qui donne la possibilité à l'utilisateur d'inscrire un choix de zone dans une boîte de dialogue (annexe 3) montrée au lancement.

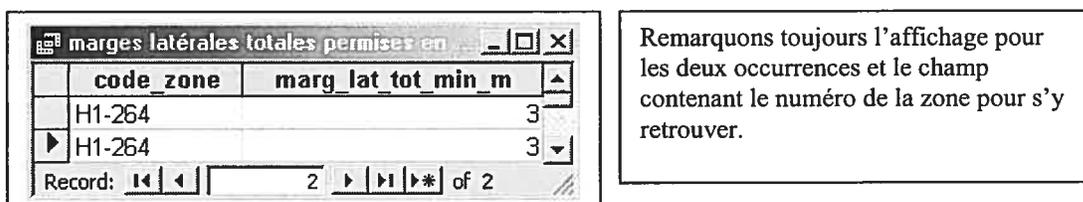


illustration : *Microsoft Access*

Figure 44 : Résultats informatiques de la requête sur la dimension des marges latérales totales

8. **Superficies, de plancher et du terrain, minimales permise en mètres carrés (« sup_plan_min_m2 »,« sup_ter_min_m2 »)** : ces requêtes permettent encore ici de calculer deux informations à partir de données de base simples. Pour cette seule raison, elles n'ont pas leur place dans la grille des usages (annexe 2-c). L'une multiplie les données relatives aux dimensions minimales permises pour un bâtiment, l'autre pour un lot. Ces requêtes attachées à la tables des usages sont paramétrées comme les autres du même type (voir la boîte de dialogue en annexe 3-p.xxix). Ces informations fabriquées seront à leur tour d'une grande utilité pour construire les coefficients d'emprise et d'occupation du sol . On affiche encore ici le code de la zone visée en référence.

code_zone	sup_plan_min_m2
H1-264	59,5
H1-264	49

code_zone	sup_ter_min_m2
H1-264	450
H1-264	450

illustrations : *Microsoft Access*

Figure 45 : Résultats informatiques des requêtes sur les superficies de plancher et de terrain minimales

9. **Coefficient d'emprise au sol minimal permis dans une zone « CES_min »** : Cette requête, appliquée à la table « usage_zone » consiste en le calcul du pourcentage de la surface du bâtiment ancrée au sol par rapport au terrain. Dans notre cas, c'est une requête paramétrée qui demande le code de la zone pour laquelle l'utilisateur aimerait que le calcul soit fait. Sur la grille, cette case n'était pas remplie (annexe 2-c). Nous l'avons stocké sous forme de requête. On calcule la superficie minimale couverte par le bâtiment que l'on divise par le calcul de la superficie minimale de terrain permise en mètre carrés (champ CES_min).

code_zone	CES_min
H1-264	0,132222222
H1-264	0,108888889

illustration : *Microsoft Access*

Nous observons ici les résultats possibles pour les deux situations d'occurrence distincte (annexe 2-c) de la zone d'usage au sein de la table de la réglementation de zonage "usage_zone".

Figure 46 : Résultats informatiques de la requête de CES pour la zone d'expérimentation H1-264

10. Coefficient d'occupation du sol minimal dans une zone « COS_min » : Cette requête, également appliquée à la table « usage_zone » consiste en le calcul du pourcentage de densité d'occupation du sol générée par le bâtiment construit sur le terrain. C'est une requête paramétrée qui demande le code de la zone pour la quelle l'utilisateur aimerait que le calcul soit fait. Sur la grille, cette information était inexistante (annexe 2-c). On calcule la superficie minimale couverte par le bâtiment que l'on divise par le calcul de la superficie minimale de terrain permise en mètres carrés. On multiplie le résultat par le nombre d'étages minimal du bâtiment pour obtenir le coefficient (COS_min).

code_zone		COS_min
	H1-264	0,1322222222
▶	H1-264	0,2177777778

Record: 2

Encore une fois, nous observons ici les résultats possibles pour les deux situations d'occurrence distincte (Annexe 2-c) de la zone d'usage au sein de la table de la réglementation de zonage "usage_zone".

illustration : Microsoft Access

Figure 47 : Résultats informatiques de la requête de COS pour la zone d'expérimentation H1-264

Notons que nous aurions pu aussi paramétrer les quatre premières requêtes pour aller chercher l'information voulue en fonction d'un lot précis plutôt que pour l'ensemble.

8.2.2.3 Exercice d'optimisation par agrégation des dimensions

Les agrégats sont des véhicules permettant d'améliorer les performances du système. Sujets à de nombreux changements ils seront régulièrement supprimés et ajoutés. La technique recommandée par les experts (Kimball, 2000) est de construire quelques agrégats et de tester leur performance avant de définir les agrégats de production. Pour ce faire, on doit se baser sur la liste des requêtes les plus fréquentes et des dimensions les plus associées. Dans le cadre de notre base de données urbaine miniature, nous ne pourrions expérimenter beaucoup cet aspect. En exemple, il pourrait se produire que la municipalité utilise beaucoup la requête de « l'adresse civique complète ». Ceci impliquerait une fusion de ses trois composantes en un nouveau champ de données ou fait de base.



Conclusions sur le chapitre :

Ce chapitre consistait en fait en un dictionnaire de données ou en un lot de métadonnées décrivant et expliquant la gestion d'un plan de forage possible pour le prototype de « magasin de données du service d'urbanisme ». Nous y avons vu en détail l'application du modèle théorique à l'aide de l'outil de validation informatique *Microsoft Access*. Chaque champ de données y a été documenté et chaque relation entre les données a été traduite informatiquement. L'exploitation du langage de requête SQL a été requis pour manipuler les connaissances de ce système de requêtes spatiales miniature. À l'aide des tables relationnelles nous avons été en mesure de décrire informatiquement les relations géographiques entre des entités urbaines. On a également répondu à un besoin municipal réel en définissant une méthode de saisie de l'information adaptée à l'utilisation d'un résumé de la réglementation de zonage prescrite sous forme de grille (non conforme au traitement relationnel). Ces différents points seront discutés davantage dans le prochain chapitre des discussions et des conclusions.



Chapitre 9 : Discussions et conclusions sur l'expérimentation

9.1 Discussion sur la gestion des éléments géographiques urbains

Le lot cadastral est associé à une représentation hiérarchique des entités auxquelles il est rattaché. On parle évidemment des autres éléments urbains qui sont décrits dans un système d'information géographique géré par une municipalité. Il est question des voies publiques et, dans certains cas, des bâtiments. Ces liens sont constitués de multiples couches. C'est-à-dire que plusieurs types de descriptions, à des échelles différentes, peuvent se prêter à la même entité géographique. Par exemple, au sein d'une administration municipale, l'urbaniste peut avoir besoin d'informations sur les voies publiques, le lot et les bâtiments, qui ne seront peut être pas nécessairement utiles au génie civil et à l'architecte.

Ceci nous amène à traiter ces informations de façon différente pour chacun de ces objets autant dans la façon de les illustrer que dans le détail de leur description. Deux scénarios s'offrent à nous. Soit on construit des tables de description spécialisées pour chacune des disciplines (tables qui pourront être questionnées individuellement ou en parallèle à l'aide de requêtes simples) ou une base de données centralisée comportant les trois niveaux de description intégrés; ceci nécessitant des requêtes plus complexes pour satisfaire les différents domaines professionnels. L'approche expérimentée suggère le premier scénario dans un premier temps, ce qui n'empêchera pas la métamorphose de cette dernière. Autrement dit, il pourrait y avoir des tables de faits concernant la description des bâtiments, de la voirie et des lots propres au travail quotidien de l'urbaniste, du génie civil et de l'architecte et ce dans trois magasins de données différents. Ne perdons pas de vue notre granularité universelle, au sein des tables de données, qui est le lot ou la propriété foncière.

Voici un diagramme qui décrit les liens entre les entités discutées et qui satisfait à la représentation de l'environnement urbain évoquée dans ce travail et en urbanisme :

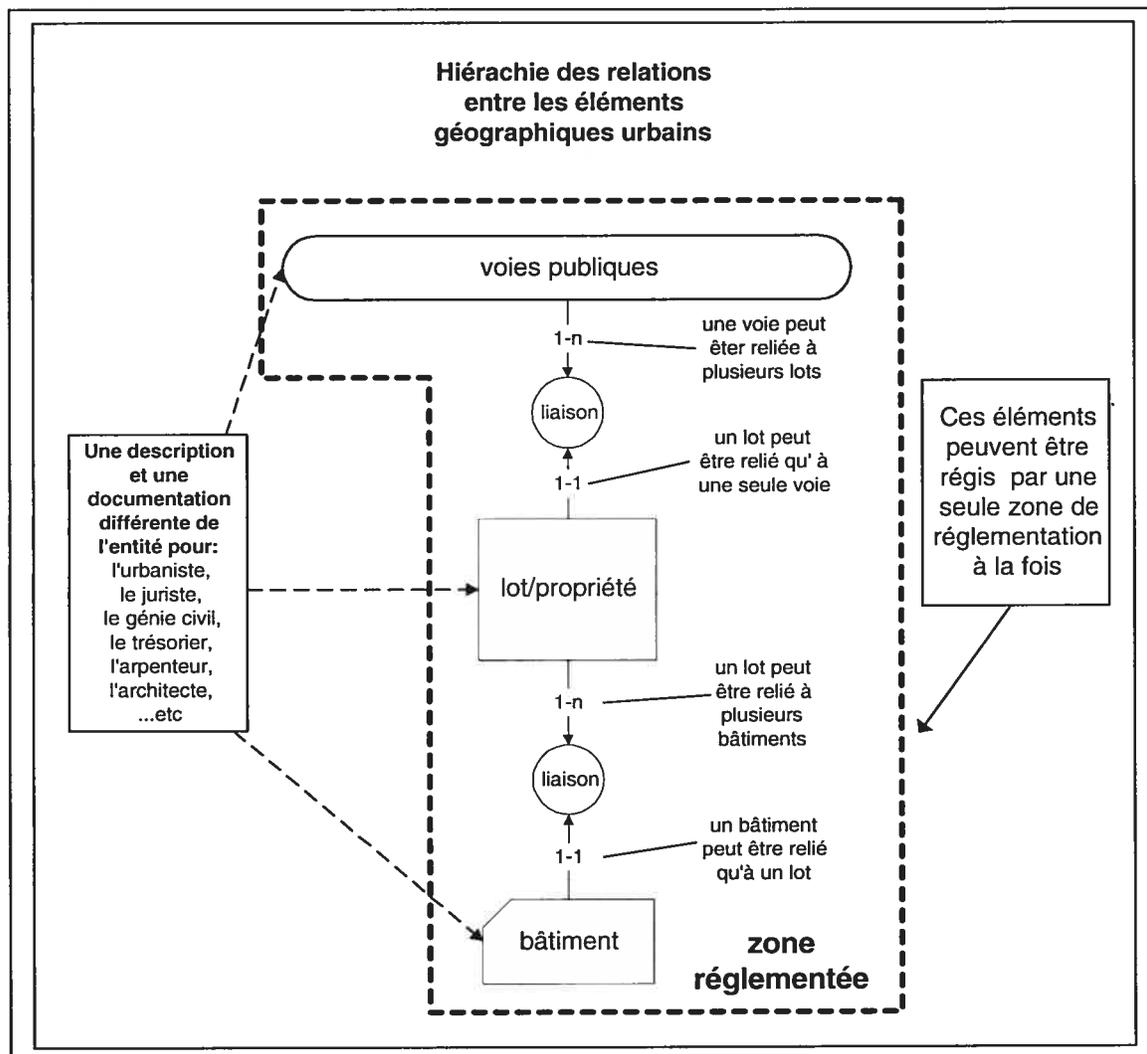


Figure 48 : Graphe des relations entre les trois éléments géographiques de base en urbanisme

Ce diagramme vient appuyer notre choix du lot (propriété foncière) comme granularité du magasin de données lié à l'urbanisme. Entre autres parce qu'il est au coeur des deux entités spatiales importantes que sont les bâtiments et les voies publiques. Il est donc en mesure d'assurer un lien entre ces dernières. Les règles simples de liaison des entités qui y sont décrites peuvent facilement être traduites mathématiquement et informatiquement pour vérifier l'intégrité de tout tableau d'association contenant de l'information urbaine

géoréférencée sur l'un des trois objets représentés. Chacun d'eux peut se voir soumis à la réglementation municipale prescrite par une zone.

9.2 Discussion sur l'encodage de la grille de réglementation de zonage

L'intégration de réglementation municipale au sein du système de gestion semblait une issue intéressante pour la tenue informatique des questions de zonage relatives aux lots cadastraux. La transformation de cette information en tableaux « Individus-Variables » semblait simple au départ vu que l'information concernant une zone est déjà résumée dans une grille construite à cet effet. La difficulté principale réside en le fait que les usages prescrits dans chaque zone sont enregistrés suivant des occurrences multiples. C'est à dire que chaque rubrique d'utilisation du sol peut comporter plusieurs modalités d'utilisation au sein d'une même zone. Dans notre, cas la zone à l'étude, "H1-264", comporte deux variantes de réglementation possible au niveau des normes spécifiques applicables à un lot cadastral. Malgré le fait que les autres sections de la grille comportent les mêmes informations dans les deux cas, on peut difficilement entrer, en une seule cellule, des informations distinctes utiles au calcul d'autres données. Notons au départ que l'entreposage de ces informations s'effectue dans un volumineux cartable sur des grilles qui sont remplies à environ 25% de leur capacité (annexe 2-c).

Tableau XXVI : Tableau de données incomplet montrant des occurrences multiples dans les cellules

	usage	façade m	marges latérales m
zone 1	h1; i4; c3;	19; 200, 10;	4;30;;
zone 2	h2;	7;	
zone 3	p1;		

illustration : *Microsoft Excel*

Tels étaient les deux problèmes soient, une difficulté au niveau de la construction d'un tableau optimal et la dissociation des informations utiles au calculs mathématiques des requêtes. La réussite de l'encodage a eu un prix qui a été celui de répéter une entrée d'enregistrement (le code de la zone) et de laisser des cellules vacantes à l'occasion à cause du champ des "dispositions spéciales" qui n'est pas utilisé dans toutes les zones. Le

code de la zone venait couronner chacune des occurrences et le champ des dispositions spéciales devait apparaître malgré le fait qu'il ne soit pas utilisé ici. Évidemment, la table "**usage_zone**" comporterait l'information pour toutes les autres zones de la municipalité qui pourrait, en fonction de la structure des requêtes, présenter des nombres d'occurrences différents pour chaque zone. Notons qu'un seul champ de la table « bénéficie » d'entrées à occurrences multiples. Par contre, cette information n'est pas utile au calcul mais plutôt à la consultation (à titre de référence). Le fait que plusieurs entrées coexistent en une cellule n'est pas dramatique. Au mieux, cette cellule peut subir des requêtes de reconnaissance de chaînes de caractère pour retrouver l'information.

9.3 Discussion sur la méthode de saisie

La méthode de saisie employée pour l'encodage de cette grille nous amène à réfléchir sur deux catégories d'informations bien distinctes qu'il faut prendre en compte. Il s'agit de *l'information de calcul* et de *l'information de référence*. La distinction entre les deux est simple. Disons que *l'information de calcul* est utilisée pour construire de nouvelles informations et de nouvelles données qui pourront à leur tour être réutilisées. Cette dernière peut être intégrée à des calculs mathématiques et apparaît en majeure partie sous forme numérique. Les cellules qui la contiennent au sein du tableau de données doivent être exemptes d'occurrences multiples. C'est-à-dire que selon les capacités du SGBD (dans notre cas *Microsoft Access*) il est de mise, lorsqu'on gère l'information de calcul, de ne stocker qu'une seule valeur numérique dans les champs de sorte à simplifier l'exécution des requêtes ultérieures et la compréhension des données par l'utilisateur.

L'information de référence quant à elle, ne se plie pas aux mêmes restrictions. Comme son nom le dit, c'est une information utilisée à titre complémentaire. Sous forme normalement textuelle, elle ne représente pas de valeur et n'a pour rôle que de renseigner son utilisateur sur un fait quelconque. Les opérations qui peuvent généralement être appliquées à ce type d'information sont des opérations d'association ou de repérage comme nous l'avons fait avec l'adresse civique ou le matricule d'un lot. C'est également

pour cette raison, comme on peut le voir dans les diagrammes de forage (annexe 1), qu'il y a toujours un peu moins de requêtes de ce type reliées à ces champs de données.

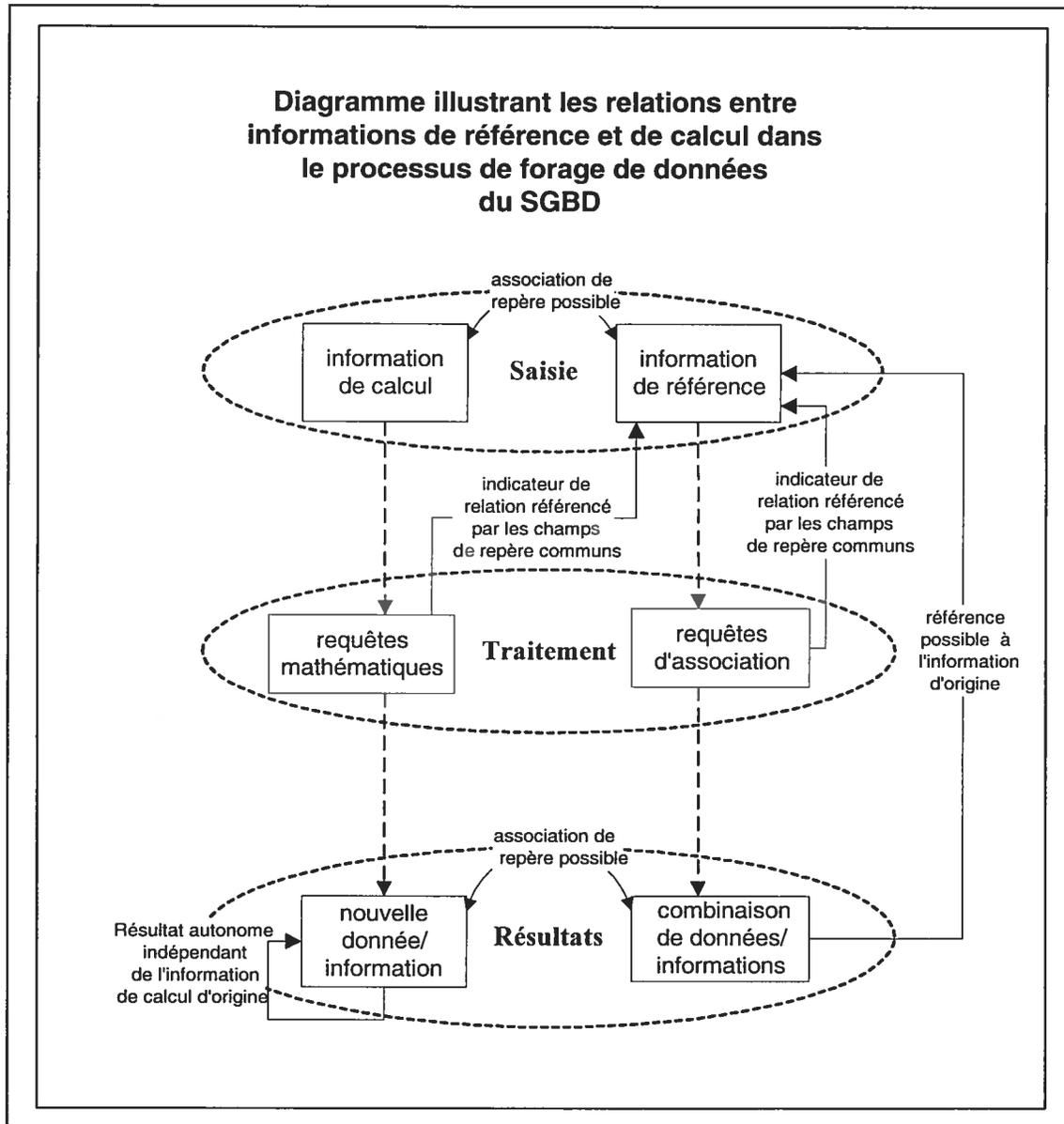


Figure 49 : Relations entre les informations de référence et de calcul dans un plan de forage

Notons que ces deux types d'information sont au coeur du système d'information urbaine et selon les résultats de l'expérimentation peut-être au coeur de tous les SGBD. Nous avons vu que le contexte de saisie de l'information est bien différent lorsqu'on construit un tableau de données selon nos besoins (documentation du lot cadastral) que lorsqu'on « traduit » de l'information provenant d'un système d'information existant et fonctionnel

(encodage de la grille de zonage). Pour la question de l'informatisation de procédures existantes, on doit être en mesure de retrouver au moins les performances qu'offrait l'ancien support ou mieux de préférence. Nous avons respecté la saisie propre à la construction d'un entrepôt de donnée et d'un plan de forage précis et c'est très efficace. Maintenant, il serait intéressant, par des expérimentations ultérieures, d'observer le rendement d'une implantation de ces méthodes dans une municipalité.

9.4 Conclusion sur la rencontre des objectifs visés au départ

Rappelons que l'objectif de recherche principal visait la proposition d'une méthode d'organisation de l'information permettant une description informatique des relations spatiales entre les éléments géographiques urbains. Ces derniers étant soumis à une réglementation municipale, une meilleure organisation de l'information concernant leur identification, leur localisation et documentation semblait essentielle à leur gestion. Dans cette optique, par cet objectif principal que nous avons certainement atteint, des sous objectifs fortement liés à ce dernier ont également été atteints :

- ✓ **L'uniformisation des format d'échange** : par cette expérimentation on assure l'uniformité du format d'échange de l'information en ne manipulant que des tables de données descriptives dissociées de tout langage graphique spécifique propre à un format informatique donné.
- ✓ **La documentation géographique** : on fourni de façon optimale, grâce aux requêtes créées à partir des données de base, un maximum d'information de documentation sur la disposition des lots dans le plan ainsi que sur les zones de réglementation auxquelles ils sont rattachées.
- ✓ **La portabilité des informations** : on assure une portabilité de l'information par une segmentation de la base de données par domaines de traitement et puis par tables de faits spécifiques à un cas de traitement particulier de ce domaine. Ceci nous permet d'exporter des tables précises plus légères.

- ✓ **La rationalisation des données :** cette expérience nous a également permis de rationaliser la quantité totale de données (réduction de l'espace disque requis) nécessaire à la compréhension de l'identification et de la localisation d'éléments géographiques urbains.
- ✓ **La solution d'aspects techniques :** à l'aide des requêtes SQL, les relations spatiales entre les lots urbains ont pu être encodées informatiquement en évitant le stockage de matrices logiques de correspondance. La méthode de saisie de la grille des usages a su adapter l'information pour qu'elle puisse être supportée par les tables relationnelles en minimisant les occurrences multiples et les cellules vides.

9.5 Un mot sur le trop plein de métadonnées et l'intégrité sémantique

On peut dire qu'à la base, c'est l'échange d'information qui cause des difficultés. C'est normal, tous les utilisateurs doivent être en mesure de comprendre l'information reçue. On devrait peut-être alors échanger moins. Ne distribuer l'information seulement à ceux pour qui elle sera pertinente et donc à des gens du domaine qui n'auront pas ou peu besoin des métadonnées pour la comprendre. Sinon, il devrait être possible de générer des métadonnées sur demande seulement lorsque nécessaire pour ainsi ne pas avoir à les gérer (un peu comme des requêtes). Dans l'expérience réalisée, il y avait autant sinon plus de métadonnées à gérer que de données réelles. De plus, pour les besoins de l'expérimentation, elles n'ont pas été toutes générées. Ce sont en fait les métadonnées à véhiculer qui alourdissent beaucoup les procédés de gestion.

La granularité maximale de l'information (en tout temps) sous forme tabulaire éliminerait probablement le problème à la base en nous montrant une information à l'état brute. Cette information serait donc libérée de toute contrainte de liaison (notée sur un support externe à la base de données). Évidemment, un lot minimum de métadonnées est nécessaire pour expliquer ce que l'on traite et son origine. Cette situation souhaitée s'avère donc impossible. Par contre, la description sémantique des données peut être enrichie par l'ajout de champs de données sous forme d'attributs d'appartenance. Par

exemple, comme il en a été le cas dans la table « **baseloclot** », on a ajouté un champ qui vient préciser l'appartenance d'une propriété à une zone réglementée. Nous pourrions ajouter des champs pour définir les quartiers d'appartenance des lots et même des tronçons de rue précis qui leur sont associés. Malheureusement, cette façon de faire entraînerait d'importantes répétitions de données au sein des tables. Du point de vue de la validation sémantique, l'expérimentation démontre que l'intégrité des informations entrées par les utilisateurs, au niveau des champs et des requêtes, est automatiquement vérifiée par un nombre prédéfini de caractères à entrer et l'échec d'une requête paramétrée mal renseignée. Le tout, sans l'ajout de métadonnées ou de clause de vérification « ASSERT » supplémentaire (sous forme de routine SQL).

9.6 Conclusion générale sur les SIURS et le but des magasins de données

On a fait des bases de données SIURS une question technique concernant l'espace disque et le temps de calcul alors que le tout correspondrait à une question d'ordre et de pertinence de l'information traitée. Il faut donc penser à rationaliser les données et métadonnées impliquées. On pourrait ainsi définir quel est le coût (en espace disque, en temps de relevé, en temps de saisie et en toute autre ressource externe) de l'ajout d'une information pour répondre à une requête logique à laquelle il nous manquerait des informations spécifiques décrites au sein de la base de données actuelle.

Le but de ce magasin de données était d'amener un certain détail spatial au sein des bases de données alphanumériques. Étant conscient du fait que des mots et des chiffres pourraient difficilement remplacer une carte graphiquement illustrée, la puissance que représentait l'absence des objets graphiques dans l'échange d'information semblait une victoire sur le problème. Du moins, il est bien possible de se renseigner sur un nombre de faits spatiaux relatifs à un lot cadastral ou à une zone de réglementation urbaine sans pour autant consulter une carte à chaque instant. En ce sens, il y aurait encore matière à développer le couplage des attributs alphanumériques à un moteur de génération graphique. Sans mémoriser les objets graphiques sur des disques durs, on n'archiverait alors que des tables de données interprétées ponctuellement.

Dans le cas étudié, ce n'est pas tant le langage de requête qui est puissant ou non mais bien le découpage de notre information. On peut donc briser les informations tant que les fragments ont encore une signification pour le domaine d'application. Rappelons que les différents niveaux de gestion de l'information au sein du SGBD sont présents pour segmenter et distribuer l'information de la façon la plus efficace possible suivant l'utilisation projetée. Le but ultime étant de toujours viser la base de données la plus compacte et la plus rationnelle capable de satisfaire aux besoins de la gestion urbaine.

Un projet de magasin de données en urbanisme serait à sources multiples étant donnée la variété et la nature des informations traitées. En plus de la géographie des propriétés foncières, il serait également question de gérer la géométrie des bâtiments et des voies publiques. Aujourd'hui, peu importe le domaine d'application des technologies de l'information, il faudra savoir traiter les données manipulées de façon optimale afin de restreindre les informations de base nécessaires au fonctionnement du SGBD. Sans quoi, l'accumulation de données et d'enregistrements de toutes sortes sur les champs d'activités qui nous intéressent deviendra un jour trop lourde à gérer même à l'aide des ordinateurs de pointe (comme c'est déjà le cas à certains endroits). Les outils de traitement informatique de qualité qui sont disponibles à l'heure actuelle nous amènent maintenant à nous demander non pas lequel d'entre eux serait le meilleur pour répondre à nos besoins mais plutôt comment pourrions-nous en être de meilleurs utilisateurs?

Enfin, cette recherche m'a amené à comprendre l'importance de la gestion de l'information et ce peu importe le domaine d'application. j'ai pu découvrir la puissance de l'extrapolation et de la déduction d'information à partir de faits existants pour ainsi construire de nouveaux faits. Ce phénomène est une boucle qui tourne tant qu'il y a des possibilités d'agencements des données et des informations. Ce qui fait la complexité du domaine est l'impossibilité de le maîtriser immédiatement ou d'en avoir une vue d'ensemble pour une application concrète. En d'autres termes, c'est une utilisation fréquente des outils qui nous fait découvrir peu à peu leur potentiel.

Afin de poursuivre cette recherche, nous pourrions penser à d'autres moyens efficaces d'organiser l'information au sein des systèmes administratifs urbains. D'autres modèles de SGBD que celui utilisé dans cette expérimentation existent. Des méthodes de stockage des données permettant d'intégrer efficacement (suivant une logique d'information paramétrée) les données relatives aux autres domaines professionnels (architectes, génies civils, trésoriers) de la gestion municipale peuvent être proposées. Cette recherche pourrait également être utilisée comme point de départ pour l'élaboration d'une base de données centralisée optimale documentant la génération de maquettes numériques d'un espace urbain. On parle bien sûr de la représentations graphiques des objets utiles à la gestion urbaine, soit : les voies publiques, les bâtiments et autres infrastructures importantes. Les possibilités informatiques actuelles nous permettraient un couplage entre les tables d'informations construites et des moteurs logiciels de génération graphique qui viendraient y puiser. Considérons alors cette recherche comme étant une première étape de compréhension du stockage et du traitement de l'information numérique municipale ouvrant la porte à des modélisations en 2D et en 3D de l'environnement urbain.

Définitions contextuelles des termes importants employés

- **base/banque de données/de connaissances** : toute base de données contemporaine est informatisée et composée de tables relationnelles montrant des champs d'attributs associés à des individus ou à des objets (tableaux :« Individus/Variables »). [Jambu,1999]
- **cadastre** : pouvant aussi porter le nom de matrice graphique, son rôle est de présenter sous la forme de cartes indexées en feuillets, les informations relatives à la localisation d'un lot dans l'espace (références géodésiques, matricules, dimensions, subdivisions, bornes, lots voisins). Chaque municipalité en possède un qui couvre ainsi l'ensemble de son territoire.
- **description de l'information** : consiste en un plan d'organisation clair de l'information traitée pour la rendre propice à un stockage optimal ainsi qu'à des manipulations ultérieures plus efficaces (dans la présente recherche, nous effectuons cette description à l'aide de tableaux).
- **donnée** : unité de saisie représentée sous la forme d'un code ou d'un symbole rattachée à un élément d'information. Le tout, selon un formalisme prédéfini. [Fondin, 1998]
- **élément géographique urbain** : toute entité constituant le monde urbain, que ce soit au niveau des infrastructures, des bâtiments ou des propriétés foncières (lots).
- **exploitation des données**: tirer profit du découpage de l'information par son intégration à un cycle d'utilisation, assurant ainsi le rendement d'une activité quelconque documentée et gérée.
- **fait urbain** : événement ou construction marquant l'histoire d'un lieu dans la ville.
- **frontage** : anglicisme utilisé en urbanisme qui désigne la façade sur rue d'une propriété foncière.
- **géocodage** : action d'associer des données ou des informations avec un code à référence spatiale (ex : coordonnées géographiques).
- **géomatique** : combinaison des mots géographie et informatique, désignant tous les éléments d'information informatisés sous forme de données à références spatiales.
- **information** : composées de multiples données agrégées par le biais de relations et de requêtes, elle comporte une signification pour son utilisateur principal. [Fondin, 1998]

- **métadonnées** : consiste en des informations sur les données stockées (entre autres : leur source, leur dimension, leur nature, leur forme, leur utilité, leur destination).
- **occurrence multiple** : véhicule plusieurs informations relatives à un même enregistrement par l'entrée de toutes les données pertinentes au sein d'une cellule d'attribut tabulaire unique.
- **organisation de l'information/des données** : arrangement des éléments d'information de manière à répondre aux attentes d'un domaine en rendant possibles des analogies par les liens entre les données.
- **présentation de l'information** : résultats d'analyses informatiques des données montrés clairement à l'utilisateur suivant un médium de sortie (à l'écran ou sur un rapport imprimé) prévu à cet effet.
- **relations entre les données/informations** : rapprochements effectués entre deux éléments d'information significatifs par le biais d'associations (attributaires dans le cadre du projet).
- **saisie de données** : transformation des informations, partant de leur état d'origine au relevé, sous une forme numérique pouvant être emmagasinée dans un ordinateur.
- **système d'information** : tout processus permettant d'échanger, divulguer, archiver et entreposer des données ou des informations. Un tel système repose sur la gestion et la manipulation de l'information, peu importe sa nature (géographique, juridique ou autres). [MAM,1984]
- **traitement de l'information/des données** : procédures et manipulations informatiques variées appliquées à différents groupes d'informations ou de données une fois qu'elles ont été saisies.
- **utilisation/occupation du sol** : consiste en l'activité humaine en sol urbain. On y fait référence lorsqu'on veut gérer qui fait quoi, où et comment dans la ville. De son côté, l'occupation du sol décrit la façon dont on plante et investit (divise, construit) l'espace.

**AutoCAD Map 2000 d'Autodesk, MapInfo Professional et MapInfo ProViewer ainsi que Microsoft Access et Microsoft Excel sont des marques déposées.*

Bibliographie :

- ❑ **Angelo**, Mark, *The use of computer graphics in the visual analysis of the proposed sunshine ski area* dans **Elsner**, Gary H. dir. *Proceedings of our national landscape*. États-Unis, Incline Village NA, United States Department of Agriculture. 1979. Pp 439-446.
- ❑ **Agostini**, Frédiane et al. (Direction générale de l'urbanisme et de l'aménagement du territoire), *Paysage, rue, architecture et affichage*, Canada, Québec, [min. des affaires municipales] Publications du Québec, 1991. 77 pages.
- ❑ **Arancibia**, Gabriel, **Cavayas**, François, *Système prototype à référence spatiale pour la gestion du réseau routier* (études et recherches en transports), Canada, Québec, Direction des communications du ministère des transports du Québec, 1991. 59 pages.
- ❑ **Anderson**, Paul F., *Regional landscape analysis*. USA, Reston VA, Environmental Design Press. 1980. 248 pages.
- ❑ **Bacon**, Warren R., *The visual management system of the forest service, USDA* dans **Elsner**, Gary H. dir. *Proceedings of our national landscape*. États-Unis, Incline Village NA, United States Department of Agriculture. 1979. Pp 660-665.
- ❑ **Barcelo**, Alain-Michel, *Planifications territoriales* (notes de cours), Canada, Montréal, Presses de l'Université de Montréal, 1998. 384 pages.
- ❑ **Beaulieu**, Denis et al. (O.A.G.Q.), *Guide de la géomatique : La géomatique au service de la municipalité et de la M.R.C.*, Canada, Sainte-Foy (Qué), Ordre des arpenteurs-géomètres du Québec, 1990, 80 pages.
- ❑ **Berry**, Joseph K., *Beyond Mapping, Concepts, Algorithms and Issues in GIS*, U.S.A, CO., G.I.S. World Books Inc. , 1993. 246 pages.
- ❑ **Charvolin**, Florian, **Veitl**, Philippe, *L'administration du paysage et le document photographique : les deux exemples de la société pour la protection des paysages de France et de l'observatoire photographique du paysage* dans **Corbin**, Alain dir. : *Évolution et représentation du paysage de 1750 à nos jours-dans le cadre du festival d'histoire de MontBrison du 28 septembre au 6 octobre 1996-*. France. Édité par le festival d'histoire de MontBrison-Imprimerie de Laplaine. 1996. Pp 407-413.
- ❑ **Corvisier de Villèle**, Marie-Anne, **Bousquet-Bressolier**, Caroline, *À la naissance de la cartographie moderne : la commission topographique de 1802* dans **Corbin**, Alain dir. : *Évolution et représentation du paysage de 1750 à nos jours-dans le cadre du festival d'histoire de MontBrison du 28 septembre au 6 octobre 1996-*. France. Édité par le festival d'histoire de MontBrison-Imprimerie de Laplaine. 1996. Pp 393-399.

- **Choko**, Marc H., **Gaudreau**, Marcel, *L’Affichage Commercial à Montréal* (Affichage commercial sur le domaine public), Canada, Montréal, INRS-Urbanisation, 1989. 65 pages.
- **Chrisman**, Nicolas, *Exploring Geographic Information Systems*, U.S.A., New-York, John Wiley & Sons Inc., 1997. 298 pages.
- **Clawson**, Marion, **Stewart**, Charles L., *Land Use Information, a critical survey of U.S. statistics including possibility of greater uniformity*, USA, Baltimore, John Hopkins Press, 1965, 402 pages.
- **Côté**, René, **Jolivet**, Christian et al., *La géomatique et ses enjeux juridiques*, Canada, Québec, Les Publications du Québec, 1993. 252 pages.
- **Dale**, Peter F., **Mclaughlin**, John D., *Land Information Management, an introduction with special reference to cadastral problems in Third World countries*, USA, New York, Oxford, Clarendon press, 1988, 266 pages.
- **Dieng**, Rose, *Méthodes et outils pour la gestion des connaissances*, Paris, Dunod, c2000, 302 pages.
- **Durlak**, Jérôme T., *Application des technologies de l’information et des communications au fonctionnement des villes, perspectives canadiennes*, Can, Ottawa, SCHL, 1989, 53 pages.
- **Duffey-Armstrong**, Marilyn, *The generation of criteria for selecting analytical tools for landscape management* dans **Elsner**, Gary H. dir., *Proceedings of our national landscape*. États-Unis, Incline Village NA, United States Departement of Agriculture, 1979. Pp 305-311.
- **Dureau**, Françoise, **Weber**, Christiane dir., *Téledétection et systèmes d’informations urbains*, France, Paris, Éditions Anthropos, 1995. 375 pages.
- **Fondin**, Hubert, *Le traitement numérique des documents*, France, Paris, Éditions Hermès, 1998, 382 pages.
- **Glasson**, John et al., *Introduction to Environmental Impact Assessment*, England, London, UCL Press Limited (University College of London), 1995. 342 pages.
- **Goglin**, Jean-François, *La construction du datawarehouse du datamart au dataweb*, Paris, Éditions Hermès, 1998, 244 pages.
- **Guité**, Manon, **Parisel**, Claude, *Réflexion sur une uniformisation des données pour la description d’objets physiques*, dans **De Paoli** et **Tidafi** dir., *Modélisation architecturale et outils informatiques entre culture et représentation du savoir-faire*, Saint-Laurent, Éditions Fides, 2000, Pp.79-99.

- **Hackos, JoAnn T. et al.,** *User and task analysis for interface design*, É-U (N.-Y.), Can (Toronto), Wiley Computer Publishing, 1998, 488 pages.
- **Hansen, Yvonne M.,** *Visualization for thinking, planning and problem solving*, dans **Jacobson, ed.,** *Information Design*, U.S.A., MIT Press, 1999, Pp. 163-189
- **Heikkila, Eric J. et al.,** *Using expert systems to check compliance with municipal building codes*, Canada, Société Canadienne d'Habitation et de Logement, 1991. 20 pages.
- **Horn, Robert E.,** *Information design, emergence of a new profession*, dans **Jacobson, ed.,** *Information Design*, U.S.A., MIT Press, 1999, Pp. 2-27
- **Huxhold, William E.,** *An Introduction to Urban Geographic Information Systems*, U.S.A., New-York, Oxford University Press Inc., 1991, 337 pages.
- **Jackle, John A.,** *The visual elements of landscape*, États-Unis, MA, The University of Massachuset Press, 1987. 193 pages.
- **Jambu, Michel,** *Introduction au Data Mining, Analyse intelligente des données*, France, Paris, Éditions Eyrolles, 1999, 114 pages.
- **Kim, Uk,** *Formalization of building codes : a knowledge-based system for building code compliance checking as a desing tool*, U.S.A., MI, UMI Dissertation Information Service, 1987. 157 pages.
- **Kimball, Ralph et al.:** *Concevoir et déployer un data warehouse (guide de conduite de projet)*, Paris, Éditions Eyrolles, 2000, 576 pages.
- **Landscape Architecture Research Office,** *The interaction between urbanization and land: quality and quantity is environmental planning and design*, USA, Cambridge, Massachussets, Graduate School of Design, 1974, 263 pages.
- **Langran, Gail,** *Time in Geographic information systems*, U.S.A., Bristol, Taylor & Francis Ltd., 1992. 189 pages.
- **Laurini, Robert, Milleret-Raffort, Françoise,** *La base de données en géomatique*, Paris, Éditions Hermès, 1993, 340 pages.
- **Larsson, Gherard,** *Land registration and cadastral systems : tools for land information and management*, Harlow, Essex, England, New York, Longman Scientific and Technical, Wiley, 1991. 175 pages.

- Lesca, Humbert et Lesca, Elisabeth, *Gestion de l'information: qualité de l'information et performances de l'entreprise*, Paris, France, Éditions Litec, 1995. 209 pages.
- Léveillé, Laurent, Dr Urb., *Systèmes d'information à référence spatio-temporelle*, Québec, Ministère des Affaires Municipales, Gouvernement du Québec, 1988. 44 pages.
- **Ministère des affaires municipales**, *Introduction au système d'information urbaine à référence spatiale (SIURS)*, Québec. Direction générale des publications gouvernementales, Gouvernement du Québec, 1984. 75 pages.
- Montgomery, Glenn E., Schuch, Harold C., *GIS Data Conversion Handbook*, U.S.A., Fort Collins, GIS World Books, 1993. 292 pages.
- Nagel, E., Newman, James R. (Kurt Gödel) traduit par Girard, Jean-Yves, *Le théorème de Gödel*, France, Paris, Éditions du Seuil, 1989, 178 pages. .
- Newman, William M. et al., *Interactive System Design*, Rank Xerox Research Center, Cambridge, Addison-Wesley Publishing Company, 1995, 468 pages.
- Olsen, Dan R., *Developing user interfaces*, USA, California, San Francisco, Morgan-Kauffman Publishers Inc., 1998. 414 pages.
- Parisel, Claude, dir. GRI&A, *Essai sur la codification graphique des immeubles pour des fins de gestion*, Montréal, Presses de L'Université de Montréal, 1990. 65 pages.
- Paulsson, Bengt, *Urban Applications of Satellite Remote Sensing and GIS Applications*, U.S.A., Washington, The World Bank, 1992. 60 pages.
- Pornon, Henri, *Systèmes d'information géographiques, pouvoirs et organisations*, France, Paris, Montréal, Éditions l'Harmattan, 1998. 255 pages.
- Prax, Jean-Yves, *Le guide du knowledge management: concepts et pratiques du management de la connaissance*, Paris, Éditions Dunod, c2000. 266 pages.
- Preece, Jenny, *A guide to usability: human factors in computing*, England, The Open University, Addison-Wesley Publishing company, 1993. 144 pages.
- Rhind, David et Hudson, Ray, *Land Use*, USA, New-York, Methuen & Co., 1980. 272 pages.
- Smardon, Richard C. et al., *Foundations for visual project analysis*. New York, Toronto, Wiley Editions. 1986. 346 pages.

- Spéry, Laurent, Libourel, Thérèse, *Les nouveaux usages de l'information géographique: vers une structuration des métadonnées*, Revue internationale de géomatique. Volume 8-no.1-2, 1998. pages 59 à 74.
 - Stockinger, Peter: *Les nouveaux produits d'information : conception et sémiotique du document*, Paris, Hermès Science Publications, c1999. 341 pages.
 - Sutte, Donald T., *The Appraisal of Outdoor Advertising Signs*, U.S.A., Chicago, The Appraisal Institute, 1994. 93 pages.
 - Turner, Tom, *Landscape Planning and Environmental Impact Design 2nd Edition*, England, London, UCL Press Limited, 1998. 425 pages.
 - Ward, Peter, *Picture composition for film and television*, É-U(Boston), Focal Press, Reed Educational and Professional Publishing Ltd, 1996, 161 pages.
 - Whipple, R.T.M (Reginald Thomas Milner), *Classifying urban land use: practice and principles*, USA, New Jersey, Princeton University, Urban Planning Research Center, 1976. 28 pages.
 - Worboys, Michael F. dir., *Innovations in GIS*, England, London, Taylor & Francis Ltd., 1994. 267 pages.
 - Zeitouni, Karine, Yeh, Laurent, *Le Data Mining Spatial et les bases de données spatiales*, Revue internationale de géomatique. Volume 9-no.4,1999. pages 389 à 423.
 - Zube, Ervin H. (University of Arizona), *Environmental evaluation : perception and public policy*, États-Unis, CA. Books/Cole Publishing Company. 1980. 148 pages.
-

ANNEXES

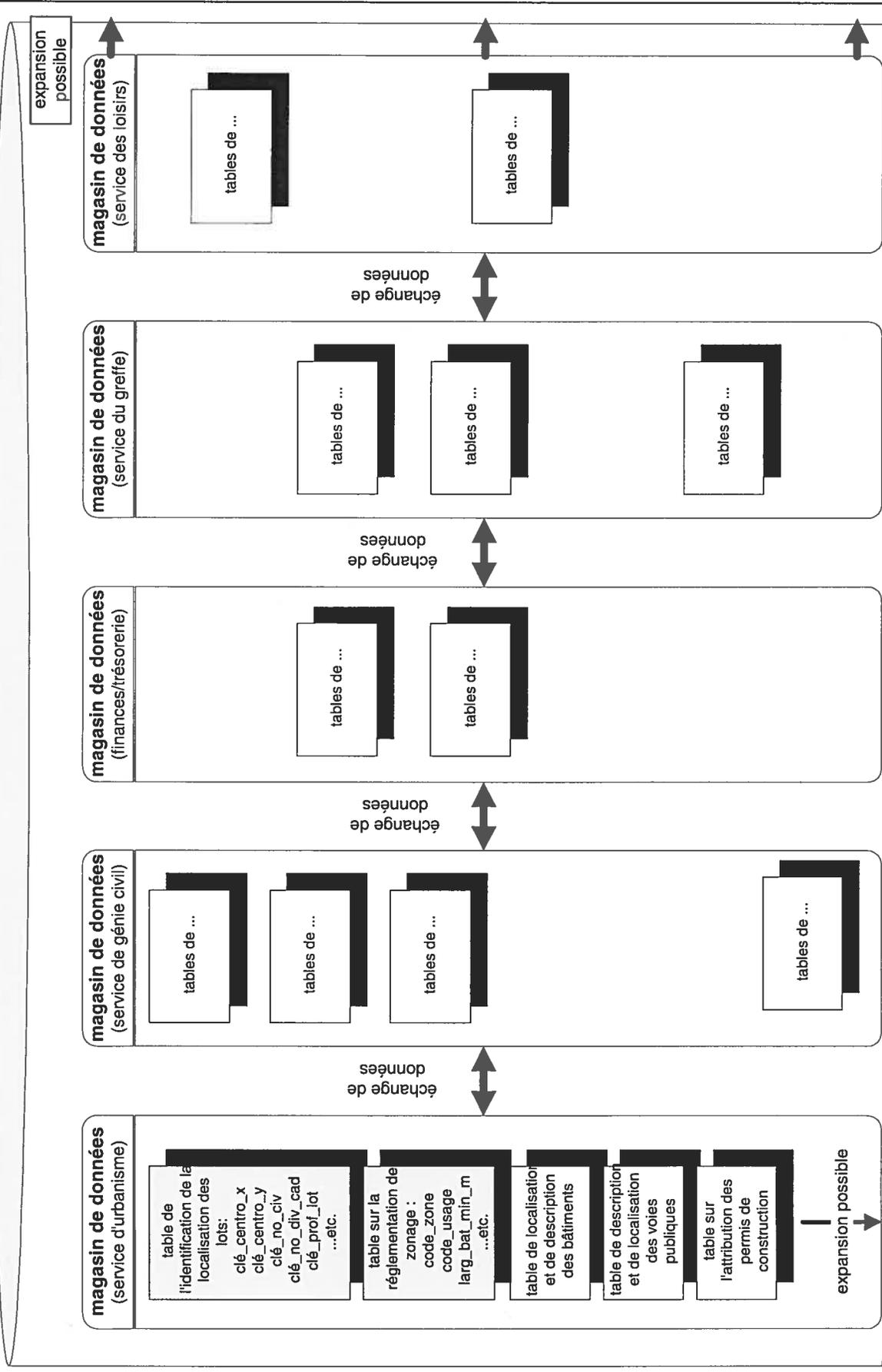
ANNEXE 1 :

-Planches de travail-

Plan d'entrepôt, détails relationnels des tables de faits, détail hiérarchique de la table sur la localisation et l'identification d'un lot, détail hiérarchique de la table sur la réglementation de zonage.

Entrepôt de données
(de l'administration municipale)

Les tables en grisé du magasin de données du service d'urbanisme montrent les tables qui seront considérées pour l'expérimentation



- magasin de données (service d'urbanisme)**
- table de l'identification de la localisation des lots:
 - clé_centro_x
 - clé_centro_y
 - clé_no_civ
 - clé_no_div_cad
 - clé_prof_lot
 - ...etc.
 - table sur la réglementation de zonage :
 - code_zone
 - code_usage
 - larg_bat_min_m
 - ...etc.
 - table de localisation et de description des bâtiments
 - table de localisation et de description des voies publiques
 - table sur l'attribution des permis de construction
- expansion possible →

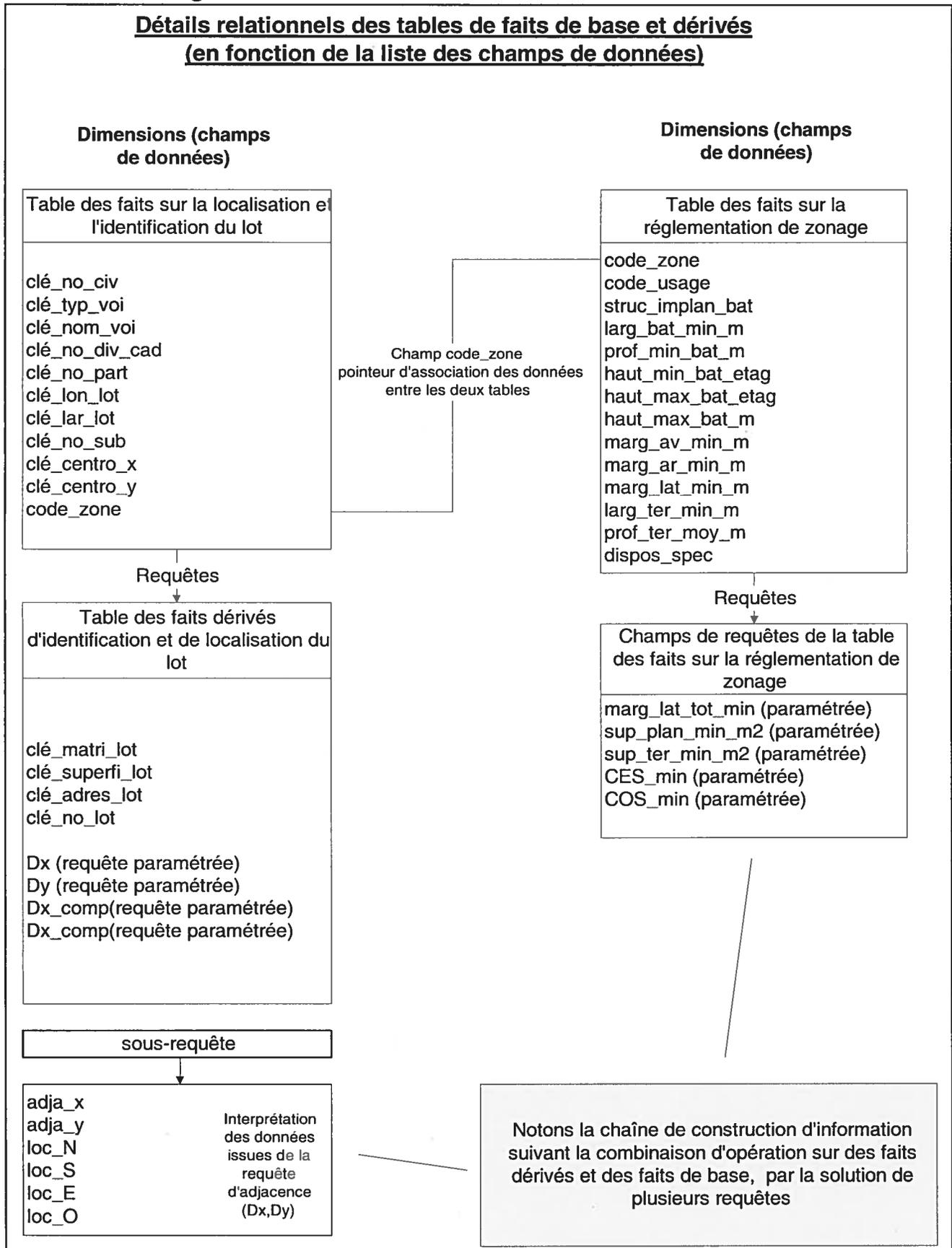
- magasin de données (service de génie civil)**
- tables de ...
 - tables de ...
 - tables de ...
 - tables de ...

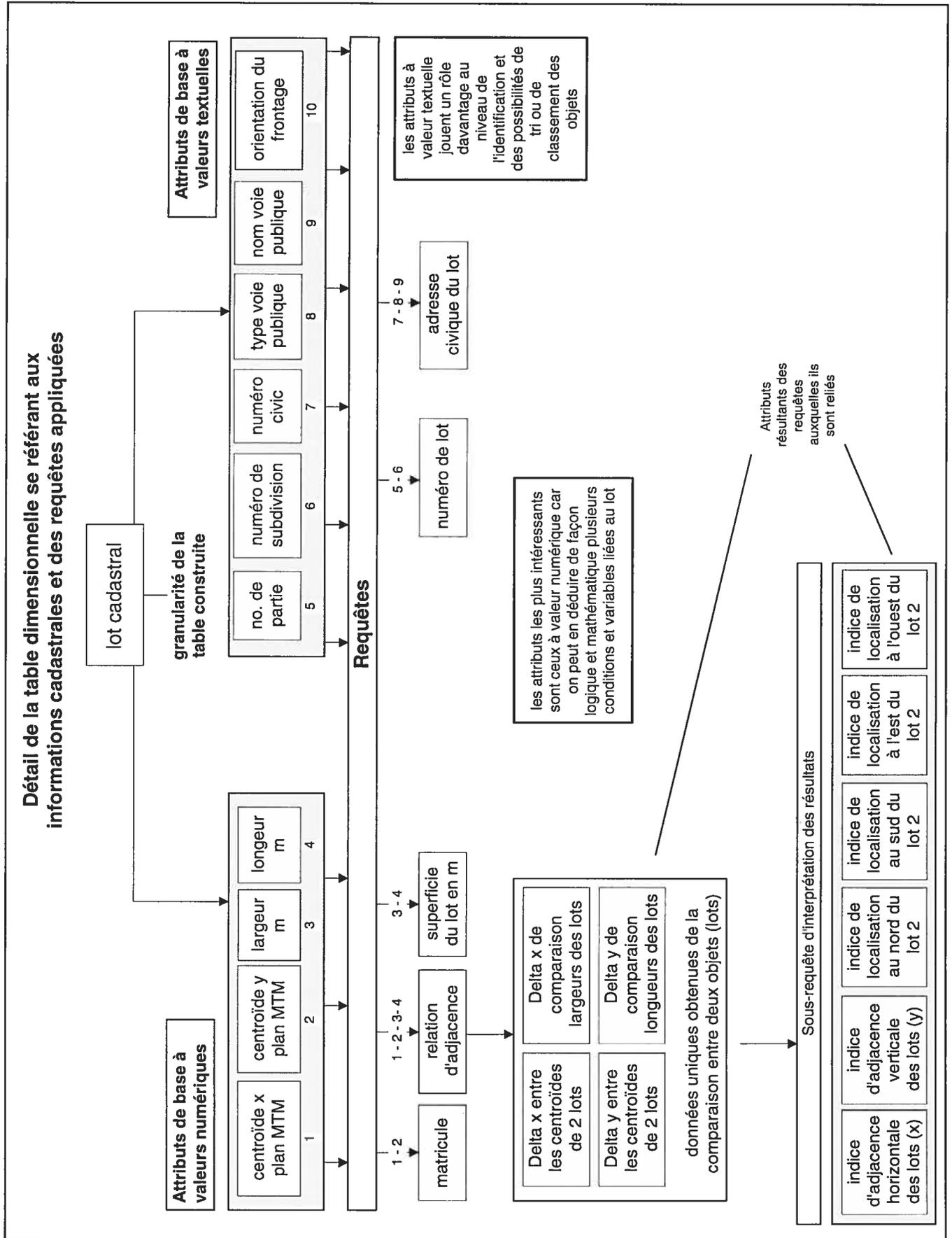
- magasin de données (finances/trésorerie)**
- tables de ...
 - tables de ...

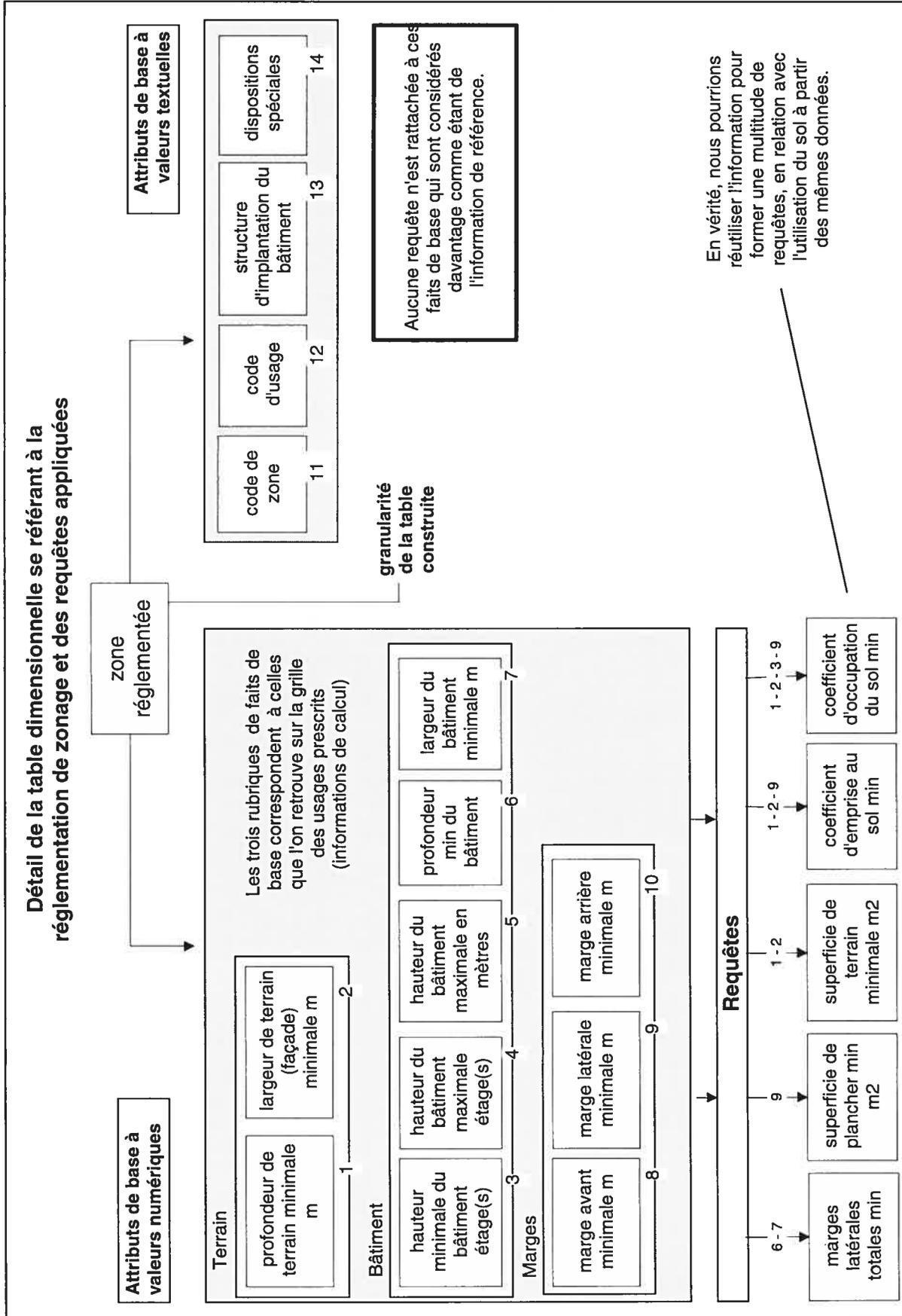
- magasin de données (service du greffe)**
- tables de ...
 - tables de ...
 - tables de ...

- magasin de données (service des loisirs)**
- tables de ...
 - tables de ...
- expansion possible ↑

Annexe 1-b: Diagramme des relations entre les tables au niveau des faits de base et dérivés





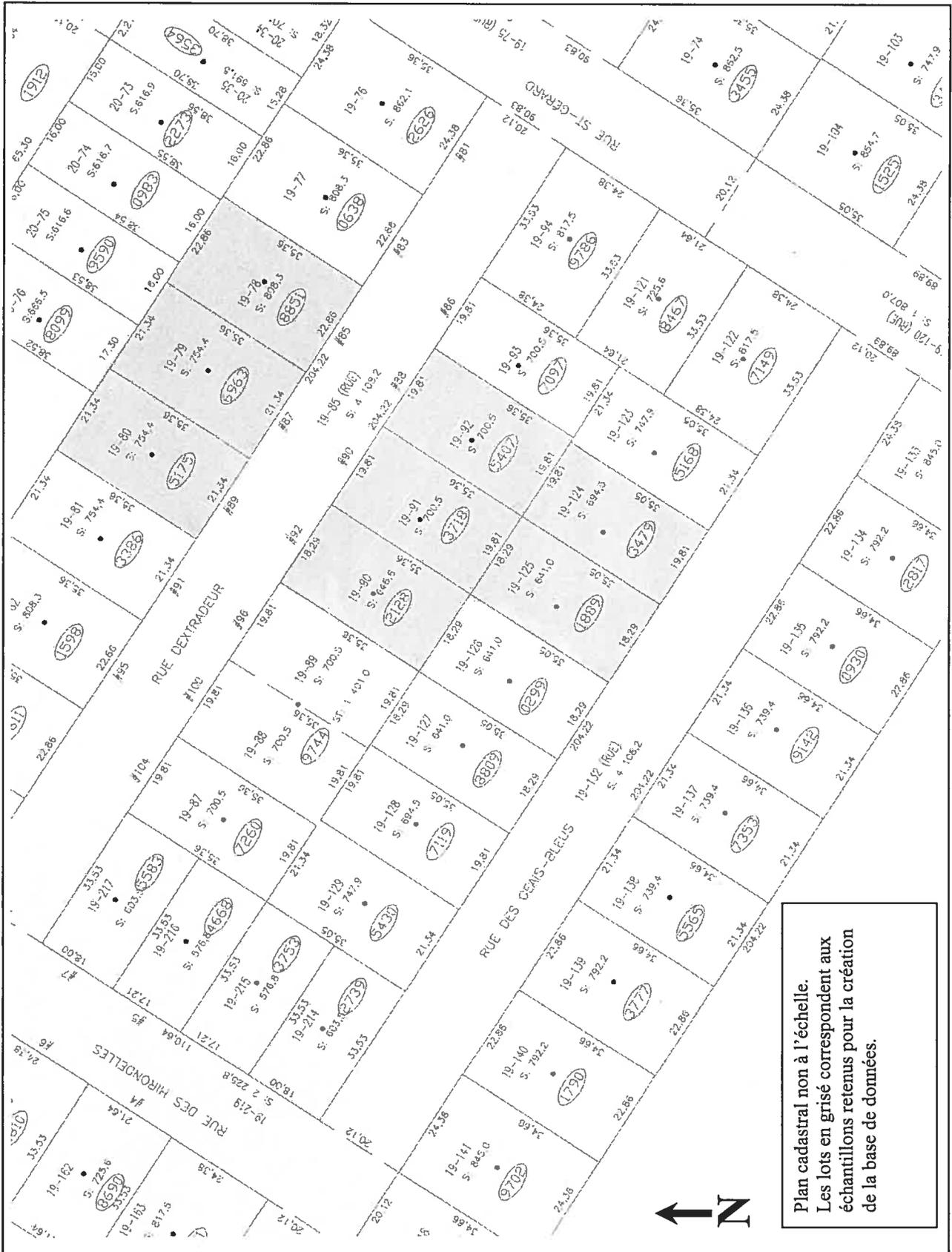


ANNEXE 2 :

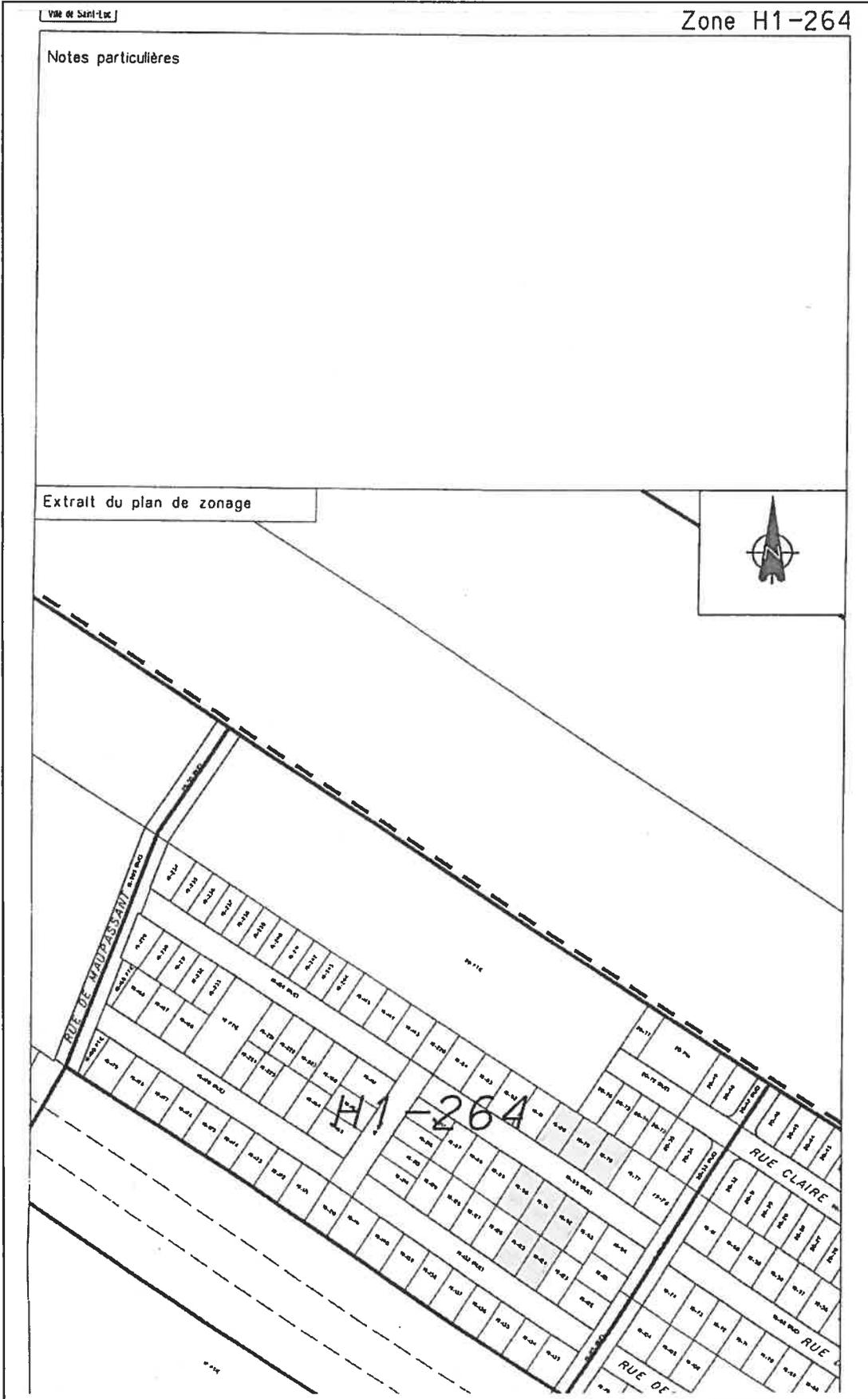
-Documents sources-

Plan cadastral, Plan de zonage et grille des usages.

Annexe 2-a: Extrait du plan cadastral de la Ville de Saint-Luc



Annexe 2-b: Extrait du plan de zonage de la Ville de Saint-Luc



Le plan n'est pas à l'échelle.

Les zones en grisé correspondent à l'échantillon étudié contenu dans la zone des usages prescrits, H1-264 délimitée par le trait noir.

Annexe 2-c: Grille des usages complémentaire à l'extrait du plan de zonage

Vie de Saint-Luc

modalité d'utilisation du sol
double pour la même zone

Zone H1-264

Usages permis

Habitation	Usages permis					
	1	2	X	X		
1: Unifamiliale						
2: Bifamiliale						
3: Trifamiliale						
4: Multifamiliale (4 à 6 logements)						
5: Multifamiliale (7 logements et +)						
6: Maison mobile						

Ce phénomène entraîne l'occurrence multiple au sein d'une base de données tabulaire.

Commerce	Usages permis					
	1	2				
1: Commerce de voisinage						
2: Commerce de quartier						
3: Service professionnel et spécialisé						
4: Commerce local						
5: Commerce régional						
6: Commerce de grande surface						
7: Commerce de divertissement						
8: Commerce d'amusement						
9: Commerce récréo touristique						
10: Service relié à l'automobile, cat. A						
11: Service relié à l'automobile, cat. B						
12: Commerce de faible nuisance						
13: Commerce de forte nuisance						

Industrie	Usages permis					
	1	2				
1: Haute technologie						
2: Faible nuisance						
3: Forte nuisance						
4: Extractive						
5: Industrie déchets/matières recyclables						

Public	Usages permis					
	1	2	X	X		
1: Parc/terrain de jeux/espace naturel						
2: Service public						
3: Infrastructure et équipement						

Rural	Usages permis					
	1	2				
1: Culture						
2: Consolidation résidentielle						

Agricole	Usages permis					
	1	2				
1: Culture						
2: Élevage						
3: Élevage en réclusion						
4: Service et transformation						

Notons la disparition de plusieurs des titres de rubriques à gauche, au profit des en-têtes de champs de données dans la table de données "usage_zone"

Usages spécifiques	Usages permis					
	Permis	Exclus				
Permis						
Exclus						

Normes spécifiques

Implantation du bâtiment	Usages permis					
	1	2	X	X		
	Isolée					
Jumelée						
Contiguë						

Structure du bâtiment	Usages permis					
	1	2				
Juxtaposée						
Superposée						

Dimensions du bâtiment	Usages permis					
	1	2				
Largeur minimum (mètres)			0,5	7		
Profondeur minimum (mètres)			7	7		
Superficie de plancher minimum (m ²)			80	80		
Sup. d'implantation au sol minimum (m ²)						
Hauteur en étages minimum			1	2		
Hauteur en étages maximum			1	2		
Hauteur en mètres maximum			10	10		

Marge	Usages permis					
	1	2				
Avant minimum (mètres)			6	6		
Avant maximum (mètres)			8	8		
Latérale minimum (mètres)			1,5	1,5		
Latérales totales minimum (mètres)			3	3		
Arrière minimum (mètres)			7,5	7,5		

Lotissement

Terrain	Usages permis					
	1	2				
	Largeur minimum (mètres)			15	15	
Profondeur moyenne (mètres)			30	30		
Superficie minimum (m ²)			450	450		

Divers

Notes particulières	Usages permis					
	1	2				
Notes particulières						

ANNEXE 3 :

-Règles OLAP -

- Code SQL commenté des requêtes-

Métadonnées sur les requêtes SQL concernant la localisation et l'identification d'un lot et la réglementation de zonage.

Actuellement, les règles d'OLAP sont celles qui devraient être suivies pour la gestion de bases de données décisionnelles.

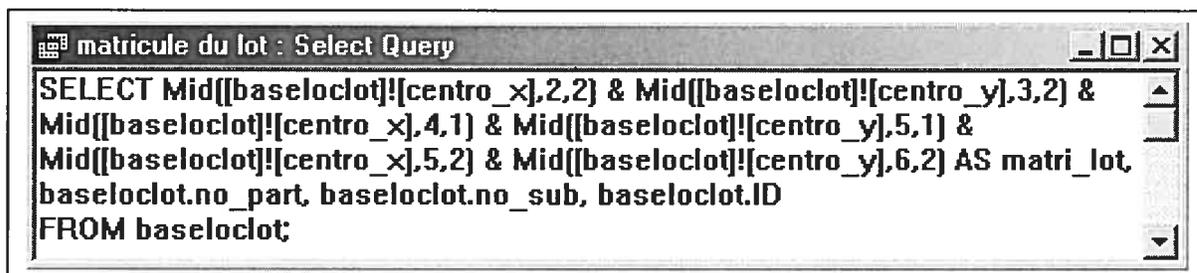
Règles d'OLAP (E. F. Codd)

- ✓ *avoir au départ une optique conceptuelle multidimensionnelle;*
- ✓ *une volonté de transparence du système (un système transparent à travers d'outils tels que des tableurs);*
- ✓ *une optique d'accessibilité aisée(alimentation par de multiples sources de données possibles);*
- ✓ *une intention d'amélioration continue des performances de rapport(le fonctionnement de la base reste le même alors que les structures peuvent changer);*
- ✓ *une architecture " client-serveur" (serveur puissant pour un client presque passif);*
- ✓ *d'une dimensionnalité générique (toute donnée doit être accessible peu importe la dimension de travail);*
- ✓ *la prise en charge dynamique des matrices de dispersion (les espaces vides de la matrice cubique générée par le produit des axes cartésiens ne doit pas gêner le temps d'accès);*
- ✓ *un support "multiutilisateur" simultané (autant en lecture qu'en écriture);*
- ✓ *des opérations dimensionnelles croisées non restreintes (toute donnée ou tranche de données, doit être accessible et visualisable dans le cube toujours conforme aux règles de gestion);*
- ✓ *une manipulation intuitive des données (possibilité d'agir directement sur la cellule contenant la donnée sans intermédiaire);*
- ✓ *des options de rapport flexibles (présentation des données);*
- ✓ *des niveaux d'agrégation dimensionnels illimités (pas de limite au nombre de dimensions et niveaux d'agrégation requis).*

[Goglin, 1998]

Code SQL commenté pour les requêtes de la table sur l'identification et la localisation d'un lot « baseloclot »

1. Code de la requête de sélection "matricule du lot"



```

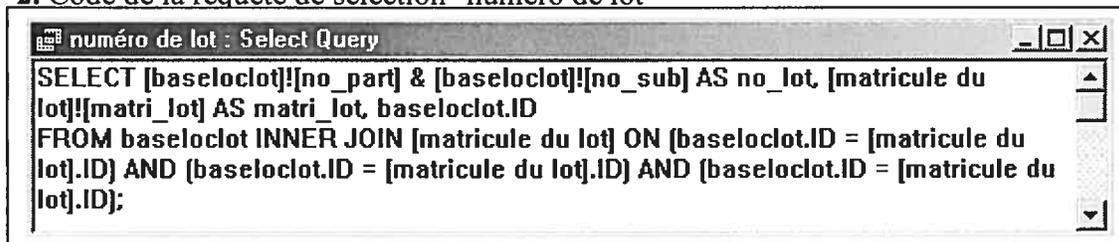
SELECT Mid([baseloclot].[centro_x],2,2) & Mid([baseloclot].[centro_y],3,2) &
Mid([baseloclot].[centro_x],4,1) & Mid([baseloclot].[centro_y],5,1) &
Mid([baseloclot].[centro_x],5,2) & Mid([baseloclot].[centro_y],6,2) AS matri_lot,
baseloclot.no_part, baseloclot.no_sub, baseloclot.ID
FROM baseloclot;

```

illustration: *Microsoft Access*

Rappelons-nous maintenant le cours accéléré de SQL ainsi que la façon dont un matricule de lot est obtenu à partir des coordonnées d'un centroïde de lot. Sous la première partie de la requête "SELECT", l'opérateur "Mid" indique au système, dans la première expression (deuxième ligne du code), qu'il doit aller chercher dans l'attribut **centro_x** (coordonnée en x du centroïde de lot), les deux chiffres consécutif qu'il y trouve à partir du deuxième caractère contenu dans la cellule. Ainsi de suite en alternance avec "centro_y" de la table "baseloclot". Finalement, l'expression "AS" indique au système de nous servir le résultat sous un champ de données nommé "matri_lot". On remarque après cette dernière virgule "baseloclot.no_part", "baseloclot.no_sub" et "baseloclot.ID" nous indiquant également que le numéro de partie, le numéro de subdivision et le champ de la clé primaire seront montrés avec le résultat (pour des jointures futures). La section "FROM" nous assure que tous les attributs questionnés proviennent de notre table "baseloclot".

2. Code de la requête de sélection "numéro de lot"



```

SELECT [baseloclot].[no_part] & [baseloclot].[no_sub] AS no_lot, [matricule du
lot].[matri_lot] AS matri_lot, baseloclot.ID
FROM baseloclot INNER JOIN [matricule du lot] ON (baseloclot.ID = [matricule du
lot].ID) AND (baseloclot.ID = [matricule du lot].ID) AND (baseloclot.ID = [matricule du
lot].ID);

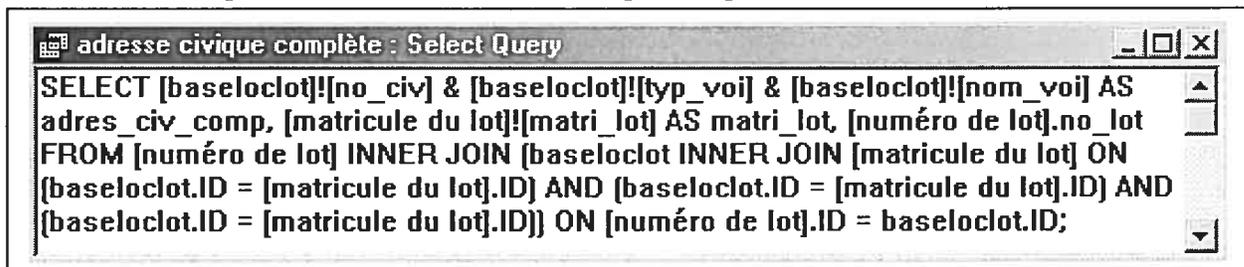
```

illustration: *Microsoft Access*

Exactement sur le même principe que la requête précédente, on indique au système, dans l'énoncé "SELECT", d'unir les trois champs de base de la table "baseloclot" correspondant au numéro de partie (**no_part**) et au numéro de subdivision (**no_sub**). Le tout sous ("AS") le nouveau champ "no_lot" pour "numéro de lot". On ajoute au résultat l'identificateur automatique (**ID**) et le matricule du lot correspondant (**matri_lot**).

La section "FROM" s'interprète exactement de la même façon que dans la requête précédente. Les raisons d'une telle requête sont également les mêmes soit : une puissance de tri et de classement accrue et un accès rapide à l'information complète.

3. Code de la requête de sélection "adresse civique complète du lot"



```

SELECT [baseloclot].[no_civ] & [baseloclot].[typ_voi] & [baseloclot].[nom_voi] AS
adres_civ_comp, [matricule du lot].[matri_lot] AS matri_lot, [numéro de lot].no_lot
FROM [numéro de lot] INNER JOIN (baseloclot INNER JOIN [matricule du lot] ON
[baseloclot.ID = [matricule du lot].ID] AND [baseloclot.ID = [matricule du lot].ID] AND
[baseloclot.ID = [matricule du lot].ID]) ON [numéro de lot].ID = baseloclot.ID;

```

illustration: *Microsoft Access*

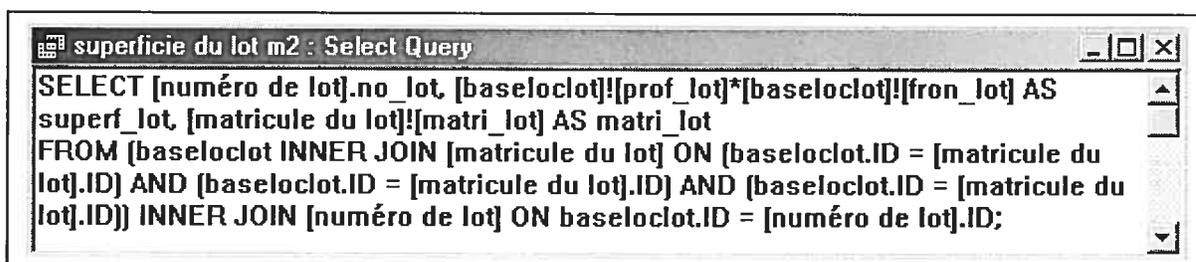
Dans cette requête, on indique simplement au système, dans la partie "SELECT" du code, d'aller chercher les champs de la table "baseloclot" qui correspondent aux numéro civique (**no_civ**), au type de voie publique (**typ_voi**) et au nom de la voie publique (**nom_voi**) pour les unir dans un même champ de données ("AS") appelé "adres_civ" (pour adresse civique). On ajoute comme prévu (dans le dernier énoncé de la section "SELECT") les résultats de la requête de matricule du lot toujours dans le champ "matri_lot" et les résultats de la requête sur le numéro de lot "no_lot" pour assurer des jointures ultérieures.

La section "FROM" nous indique que les jointures des résultats avec les requêtes "matricule du lot" et "numéro du lot" s'effectuent à partir d'un joint interne "INNER JOIN" au niveau ("ON") de l'équivalence de l'identificateur automatique du système "baseloclot.ID" au sein des deux tables générées par les requêtes "numéro de lot.ID" et

“**matricule du lot.ID**”. Ces dernières répètent les mêmes enregistrements que dans la table “**baseloclot**”.

Remarquons les valeurs que l’on retrouvait dans la table d’origine qui sont séparées dans le champ cible par les barres de soulignement (_). Cette requête nous permet d’obtenir l’adresse complète d’un lot sur demande et nous permet de bénéficier de champs de base plus puissants pour le triage des données.

4. Code de la requête de sélection "superficie du lot"



```

superficie du lot m2 : Select Query
SELECT [numéro de lot].no_lot, [baseloclot].[prof_lot]*[baseloclot].[fron_lot] AS
superf_lot, [matricule du lot].[matri_lot] AS matri_lot
FROM [baseloclot INNER JOIN [matricule du lot] ON (baseloclot.ID = [matricule du
lot].ID) AND (baseloclot.ID = [matricule du lot].ID)] INNER JOIN [numéro de lot] ON baseloclot.ID = [numéro de lot].ID;

```

illustration: *Microsoft Access*

L’énoncé « **SELECT** » s’interprète facilement. Toujours dans la table « **baseloclot** », on cherche à unir les deux champs de données « **prof_lot** » et « **fron_lot** » mais cette fois-ci par l’entremise d’une opération mathématique de multiplication (« * » à la première ligne). Cette dernière, dans le cas d’un lot octogonal symétrique nous permettra d’avoir la superficie du lot sous (« **AS** ») le champ créé « **superf_lot** ». Le matricule du lot et l’identificateur automatique sont toujours montrés ici comme on peut le voir sur la deuxième ligne du code (**matri_lot** et **baseloclot.ID**).

La section « **FROM** » du code s’interprète toujours de la même façon. Dans la table « **baseloclot** », on associe les champs résultants par jointure interne (« **INNER JOIN** ») suivant (« **ON** ») l’identificateur automatique de la table « **baseloclot** » (« **ID** ») qui concorde avec tous les champs joints.

5. Code de la requête d'attributs d'adjacence entre deux lots choisis par l'utilisateur

```

SELECT [baseloclot].[centro_x]-[baseloclot_1].[centro_x] AS Dx,
[baseloclot].[centro_y]-[baseloclot_1].[centro_y] AS Dy,
([baseloclot].[fron_lot]/2+[baseloclot_1].[fron_lot]/2) AS Dx_comp,
([baseloclot].[prof_lot]/2+[baseloclot_1].[prof_lot]/2) AS Dy_comp,
[matricule du lot].matri_lot AS [matricule 1], [matricule du lot_1].matri_lot
AS [matricule 2]

FROM baseloclot INNER JOIN [matricule du lot] ON (baseloclot.ID =
[matricule du lot].ID) AND (baseloclot.ID = [matricule du lot].ID), baseloclot
AS baseloclot_1 INNER JOIN [matricule du lot] AS [matricule du lot_1] ON
baseloclot_1.ID = [matricule du lot_1].ID

WHERE ((([matricule du lot].matri_lot)=[matricule du premier lot]) AND
((([matricule du lot_1].matri_lot)=[matricule du deuxième lot]));

```

illustration: *Microsoft Access*

La requête la plus intéressante nous présente une relation des plus significatives (géographiquement parlant) entre les données. L'énoncé « **SELECT** » de la requête nous propose de calculer les « **delta X** » et « **delta Y** » entre deux lots, au choix de l'utilisateur, provenant de la table de données « **baseloclot** ». Pour ce faire, le système crée une table « **baseloclot_1** » identique à « **baseloclot** » entre lesquelles il pourra effectuer ses opérations de calcul et de comparaison suivant les deux lots choisis. Cette requête est alors paramétrée. Prenons la première opération (1ère ligne du code). Pour calculer le « **delta X** » entre les deux lots, le système va puiser dans la table « **baseloclot** » pour le « **centroïde X** » du premier lot entré par l'utilisateur (l'entrée de donnée se fait à l'aide du numéro de matricule) et dans la table « **baseloclot_1** » pour le « **centroïde X** » du deuxième lot qu'on retranche ensuite du premier ($[baseloclot].[centro_x]-[baseloclot_1].[centro_x]$). On obtient le « **delta X** » entre les deux lots sous le champ (« **AS** ») « **Dx** ». La requête procède de la même façon pour le « **delta Y** ».

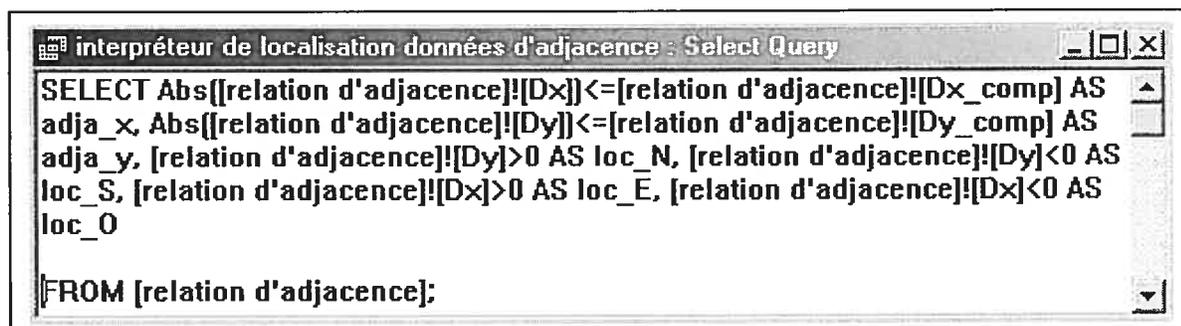
Maintenant, pour les « **deltas** » de comparaison, on veut savoir quel serait le delta maximal que pourrait parcourir un centroïde pour se rendre à celui du lot qui lui serait adjacent. Pour ce faire, prenons l'exemple du « **delta X** » encore une fois. À la troisième ligne du code, on prend les mesures des façades associées à leurs lots respectifs (dans

chacune des deux tables), que l'on divise individuellement en deux puis que l'on additionne pour obtenir le « **delta X** » de comparaison « **Dx_comp** » ($(([\text{baseloclot}]![\text{fron_lot}]/2+[\text{baseloclot_1}]![\text{fron_lot}]/2) \text{ AS } \text{Dx_comp})$). La même procédure est employée pour le « **delta Y** » de comparaison (ligne 4) excepté qu'on utilise les profondeurs de lot. À la suite de notre requête (lignes 5 et 6) on affiche les matricules entrés pour garder une trace de notre comparaison ($([\text{matricule du lot}].\text{matri_lot AS } [\text{matricule 1}], [\text{matricule du lot_1}].\text{matri_lot AS } [\text{matricule 2}])$)

Dans la section « **FROM** », le joint interne « **INNER JOIN** » s'interprète de la même façon que dans les autres requêtes. Seulement, cette fois-ci il est double à cause de la « table-miroir » « **baseloclot_1** » générée par le système. Les deux tables sont couplées avec la requête de génération de matricule et suivant (« **ON** ») l'identificateur automatique du système « **ID** » car c'est de cette façon seulement que l'utilisateur peut questionner la base de données aux lignes huit et neuf du code (**baseloclot INNER JOIN [matricule du lot] ON (baseloclot.ID = [matricule du lot].ID) AND (baseloclot.ID = [matricule du lot])**)).

Finalement, la condition « **WHERE** » est très simple car on n'effectue le calcul seulement pour les deux lots choisis par l'utilisateur suivant un premier et un deuxième matricule (la jointure avec la requête de matricule est très importante). Cette requête est couplée à une sous requête d'interprétation de ses résultats pour faciliter la tâche de l'utilisateur dont nous verrons le code dans la figure suivante.

6. Code de la requête d'interprétation des attributs d'adjacence résultants



```

SELECT Abs([relation d'adjacence]![Dx])<=[relation d'adjacence]![Dx_comp] AS
adja_x, Abs([relation d'adjacence]![Dy])<=[relation d'adjacence]![Dy_comp] AS
adja_y, [relation d'adjacence]![Dy]>0 AS loc_N, [relation d'adjacence]![Dy]<0 AS
loc_S, [relation d'adjacence]![Dx]>0 AS loc_E, [relation d'adjacence]![Dx]<0 AS
loc_O

[FROM [relation d'adjacence];

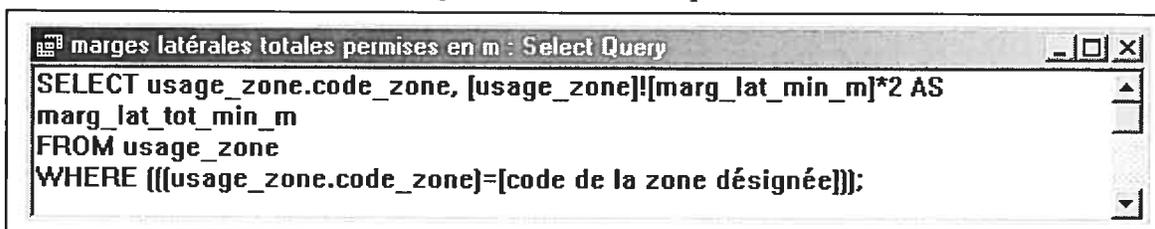
```

Illustration: *Microsoft Access*

La section « **SELECT** » du code reprend directement les valeurs générées par la requête précédente ([**relation d'adjacence**!]). Sur la première ligne, dans le cas où la valeur absolue (« **Abs** ») du « **delta X** » (« **Dx** ») est plus petite ou égale au « **delta X** » de comparaison (« **Dx_comp** »), le système nous livrera une réponse positive sous (« **AS** ») le champ d'adjacence en x (« **adja_x** »). C'est le même scénario pour l'adjacence sur l'axe y. Pour la localisation du premier lot entré par rapport au deuxième, par exemple, pour savoir si un lot est situé au nord de l'autre, la valeur non absolue du « **delta y** » (« **Dy** ») doit être plus grande que « 0 » (à la ligne 3 du code). Si tel est le cas, une réponse positive (« **Yes** ») sera affichée sous (« **AS** ») le champ « **loc_N** » (pour localisation au nord). Un « **delta Y** » plus petit que « 0 » nous donnera une réponse positive pour une localisation au sud. Le même scénario se répète dans les deux autres directions (E et O) pour le « **delta X** ». La section « **FROM** » nous indique logiquement que toutes les informations sources proviennent de la relation d'adjacence ([**relation d'adjacence**!]).

Codes pour les requêtes de la table sur la réglementation de zonage « usage zone »

7. Code de la requête sur les marges latérales totales permise



```

SELECT usage_zone.code_zone, [usage_zone]![marg_lat_min_m]*2 AS
marg_lat_tot_min_m
FROM usage_zone
WHERE (((usage_zone.code_zone)=[code de la zone désignée]));

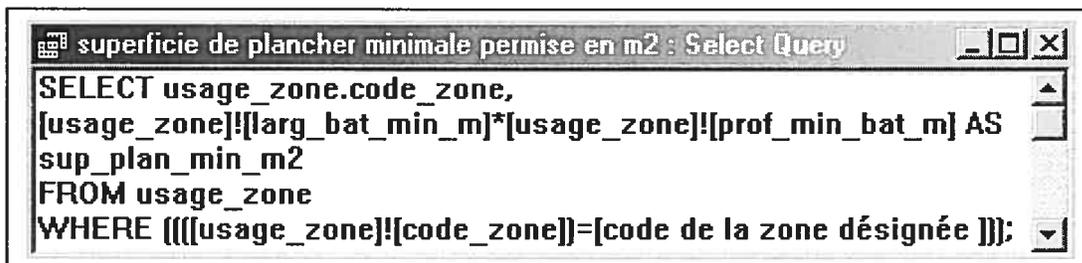
```

illustration: *Microsoft Access*

C'est une information disponible sur la grille des usage que nous avons ici tournée en requête pour des question d'efficacité. Très simple, la section « **SELECT** » indique qu'en fonction du champ (« **code_zone**») de la table [**usage_zone**], on multiplie (*) par "2" la valeur du champ de la marge latérale minimale en mètre permise [**marg_lat_min_m**] et on présente le résultat sous (« **AS**») "**marg_lat_tot_min_m**". La section « **FROM** » nous indique la table d'origine des données, « **usage_zone** ». Cette requête est paramétrée et

comporte une condition « **WHERE** » qui donne la possibilité à l'utilisateur de choisir la zone réglementée qu'il désire consulter. Dans notre cas, c'est la seule zone (H1-264).

8. Code des requêtes sur les superficies de plancher et de terrain minimales permises:



```

superficie de plancher minimale permise en m2 : Select Query
SELECT usage_zone.code_zone,
[usage_zone].[larg_bat_min_m]*[usage_zone].[prof_min_bat_m] AS
sup_plan_min_m2
FROM usage_zone
WHERE ((([usage_zone].[code_zone]=[code de la zone désignée]));

```



```

superficie de terrain minimale permise en m2 : Select Query
SELECT usage_zone.code_zone,
[usage_zone].[larg_ter_min_m]*[usage_zone].[prof_ter_moy_m] AS sup_ter_min_m2
FROM usage_zone
WHERE ((([usage_zone].[code_zone]=[code de la zone désignée]));

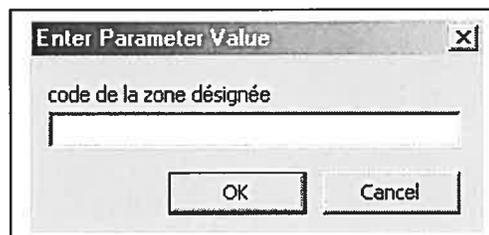
```

illustrations: *Microsoft Access*

Ces deux requêtes sont identiques seulement, elles ne traitent pas les mêmes champs de données au sein de la table “usage_zone”. En mode « **SELECT** », la première multiplie (“*”) les largeurs [**larg_bat_min_m**] et longueurs [**long_bat_min_m**] minimales permises pour un bâtiment. La deuxième fait la même opération avec les largeurs [**larg_ter_min_m**] et profondeurs [**prof_ter_moy_m**] minimales permises pour un terrain. Les résultats que l’on obtient respectivement pour chacune des requêtes sont les superficies de plancher minimales en mètres carrés [**sup_plan_min_m2**] et superficies minimales de terrain en mètres carrés [**sup_ter_min_m2**] permises. Pour les deux requêtes, la section « **FROM** » signifie que les données proviennent de la table [**usage_zone**] et la section « **WHERE** » laisse encore le paramètre au choix de l'utilisateur pour la zone à questionner (ici H1-264 dans la boîte de dialogue).

Ci-contre, la boîte de dialogue “type” permettant d’entrer un paramètre lors du lancement d’une requête. Dans le cas présent, on demande le paramètre au niveau de la zone de réglementation ciblée pour l’exécution de la requête.

Illustration: *Microsoft Access*



Enter Parameter Value

code de la zone désignée

OK Cancel

9. Code de la requête sur le coefficient d'emprise au sol minimal permis dans une zone :

```

SELECT usage_zone.code_zone,
(((usage_zone![larg_bat_min_m]*usage_zone![prof_min_bat_m])/((usage_zone![larg_ter_min_m]*usage_zone![prof_ter_moy_m])) AS CES_min
FROM usage_zone
WHERE (((usage_zone![code_zone])=[code de la zone désignée]));

```

illustration : *Microsoft Access*

On prend la requête à sa forme la plus simple. La section « **SELECT** » montre qu'on va puiser les informations en fonction de la colonne (« **code_zone** ») de la table (« **usage_zone** »). On calcule d'abord la superficie couverte par le bâtiment qui consiste en la largeur minimale du bâtiment (« **larg_bat_min_m** ») multipliée par la profondeur minimale du bâtiment (« **prof_min_bat_m** »). On divise le tout par la superficie minimale de terrain permise obtenue par la multiplication entre la largeur du terrain minimale (« **larg_ter_min_m** ») et la profondeur de terrain moyenne en mètres (« **prof_ter_moy_m** »). Toutes ces données provenant de la même table, on obtient « **FROM** » (« **usage_zone** »). Le pourcentage sous forme de coefficient est ensuite obtenu dans le champ (AS) (« **CES_min** »). La section « **WHERE** » offre la possibilité d'entrer un paramètre. La valeur du champ (« **code_zone** ») sera déterminée par l'utilisateur qui entrera le numéro de la zone qu'il veut questionner au sujet du « **CES minimal** » lors du lancement de la requête.

10. Code de la requête sur le coefficient d'occupation du sol minimal permis

```

SELECT usage_zone.code_zone,
(((usage_zone![larg_bat_min_m]*usage_zone![prof_min_bat_m])/((usage_zone![larg_ter_min_m]*usage_zone![prof_ter_moy_m]))*usage_zone![haut_bat_min_etag]
AS COS_min
FROM usage_zone
WHERE (((usage_zone![code_zone])=[code de la zone désignée]));

```

illustration : *Microsoft Access*

Cette requête montre la possibilité de créer des informations qui n'étaient pas prévues par l'administration. La section « **SELECT** » est la même que pour la requête précédente. En fait, toutes les sections sont identiques, seulement, on multiplie le résultat qui donne le « **CES minimal** » par le nombre d'étages minimal permis (« **haut_bat_min_etag** ») pour obtenir le pourcentage d'occupation du sol sous forme de coefficient (AS) (« **COS_min** »).

Annexe 4:

Entrevue avec l'urbaniste de la Ville de Saint-Luc (2001/10/03)

Sujet: la situation municipale face aux systèmes informatisés de gestion de l'information géoréférencée

1. Avec quel type d'information travaillez-vous quotidiennement?

- **description précise des données traitées (quantité, forme)**
- réglementation : texte écrit (texte de loi numéroté avec alinéas)
- zonage: cartographie, réglementation (texte), grille d'occupation du sol (matrice à double entrée)
- permis de construction: formulaires (texte écrit de façon formelle administrative)
- permis de lotissement: formulaire (texte écrit de façon formelle administrative)
- Cartes thématiques: zones inondables; glissements de terrain; zones agricoles; aménagements commerciaux; réseaux d'infrastructures; ortho-images et photos satellites.

2. Comment sont actuellement organisées ces informations pour votre travail ?

- **stockage** : archives en filières, archives informatiques
- **support** : papier, mémoire informatique sur serveur, CD-ROM (ou disquette).
- **accès** : tous peuvent y avoir accès dans les divers services de la municipalité par l'interface des logiciels Access 97 ou AutoCAD Map 2000 utilisés pour la conception et le fonctionnement du système(suivant une optique de diffusion de la même information à tout le monde).

3. Y a-t-il des informations auxquelles vous n'avez pas accès ou qui ne vous sont pas disponibles qui vous aideraient dans votre travail?

- **quelles sont elles?** Au niveau de l'utilisation du sol.
- **Sous quelle forme en auriez-vous besoin?** Selon le code d'utilisation du sol par secteur, on aimerait savoir où sont répartis sur le territoire, les bâtiments, en fonction de leur nombre d'étages. On voudrait intégrer informatiquement en format Access la grille de l'occupation du sol pour la lier au plan de zonage en AutoCAD (mise à jour ainsi simplifiée).

Actuellement, ce type de requête ne questionnerait pas une banque d'objets graphiques numérisés mais bien la base de donnée du rôle d'évaluation à laquelle il manquerait un champs de données comportant le nombre d'étage de chaque construction sur les lots construits.

4. Pour quelles raisons avez-vous informatisé votre système de gestion de l'information?

- **A-t-on imposé la nouvelle technologie d'un pallier administratif supérieur?** Oui, mais la normalisation imposée par le MRN (ministère des ressources naturelles) ne convenait pas aux besoins de la municipalité.
- **Est-ce la volonté du milieu à l'interne?** Oui, ce sont les gens ayant des connaissances et un intérêt marqué pour l'informatique qui ont décidé de passer à l'action à temps perdu.
- **Avez-vous eu de l'aide (financière et conception) extérieure pour la migration?** Non, la municipalité a déboursé l'ensemble des frais pour les honoraires professionnels des consultants externes impliqués lorsque nécessaire.

5. Quelles procédures employez-vous afin d'effectuer les mises à jour de ces informations maintenant numérisées?

- **les étapes de la procédure :** la mise à jour des données s'effectue souvent à partir de documents papiers ou de relevés de terrain récents, par l'interface des logiciels de gestion de données (graphique et numérique) utilisés. On s'efforce de tenir à jour surtout les bases de données vouées au services des finances et de la trésorerie qui ont une valeur légale (surveillées par la loi). On parle donc en majeure partie du rôle d'évaluation et de la matrice graphique qui sont écrasés mensuellement par une version toujours plus à jour (l'ancienne est archivée sur CD-ROM).

- **ressources impliquées par ces étapes** : chaque service est responsable de mettre à jour l'information qui lui est rattachée à même la base de donnée centrale (ex: le service des finance et de la trésorerie, le service d'urbanisme et de génie). Parfois, l'évaluateur municipal, tant qu'à manipuler des données relatives au rôle d'évaluation, mettra lui-même à jour ce qu'il traite.

6. Depuis l'informatisation, remarquez-vous une différence notable dans la performance ou le rythme du traitement de l'information (à l'interne)?

Oui, beaucoup d'améliorations ont été ressenties à plusieurs niveaux dans les activités quotidiennes notamment:

- Au niveau de la capacité de sortir facilement des extraits de plans pour n'importe quel besoin d'illustration cartographique;
- Pour monter les dossiers du CCU, on est en mesure rapidement de superposer le cadastre sur des photographies aériennes afin d'illustrer de façon accrue l'alignement des bâtiments par rapport à la rue par exemple;
- Le système accélère grandement la mise à jour des plans de zonage (autrefois papier);
- Le coût de production cartographique est réduit au 1/5 de sa valeur d'origine (chaque besoin en carte est donc comblé presque sans hésitation).
- Le traitement et l'intégration d'images satellites à tout document cartographique est précis et simplifié.

7. Y a-t-il des actions ou des fonctions que vous ne réalisez pas parce que le système de gestion ne vous le permet pas (par un manque d'outils par exemple)?

Les limites du système se traduisent par des seuils de performance fixés par les logiciels autant que par les machines:

- *Microsoft Access 97* peut gérer une base de donnée dont la grosseur est limitée pour les besoins de la municipalité et bientôt de la grande ville vu les fusions prochaines (gel du système).
- *Microsoft Access 97* limite également le nombre d'utilisateurs pouvant puiser dans la base de données, limitant donc sa diffusion par le fait même (gel du système).
- Les ordinateurs quant à eux, ne suivent pas et sont rapidement essouffés par la lourdeur des bases de données.

Quelles solutions coûteuses envisager? Migrer vers un autre logiciel de base de données plus performant mais compatible avec celui-ci, pour récupérer la structure des bases en place, ou, renouveler le système Access pour une version récente et remplacer la flotte d'ordinateurs? Cette situation pourrait nous forcer à travailler avec une base de données scindée en plusieurs morceaux.

8. Qu'est-ce qu'on aime et qu'est-ce qu'on aime pas du système implanté ?

Permet un échange rapide des données avec les parties impliquées dans divers projets d'implantation d'infrastructures (par exemple arpenteurs, entrepreneurs) ainsi qu'avec d'autres municipalités informatisées. Desquels on peut maintenant exiger des documents sous le format numérique "*.dxf" répandu dans l'industrie et les bureaux d'ingénieurs. On réduit l'attente pour les transactions d'informations et on réduit le fardeau sur la mémoire humaine par des procédures informatiques de suivi des dossiers en cours. Les mauvais côtés du système sont relatés par la dépendance aux produits informatiques tel que décrit à la question #7.

9. Est-ce que les fusions et la rationalisation des services influent sur vos procédures de gestion de l'information ?

Oui, les objectifs de chaque municipalité impliquée dans la fusion ne sont pas les mêmes, leurs attentes ne sont pas les mêmes et leurs procédures de gestion de l'information ne sont pas les mêmes non plus. Voilà plusieurs points non compatibles qui ralentiront les ardeurs d'évolution du système d'information municipal en place pour un bon bout de temps afin de faire les ajustements nécessaires (question de mise à niveau et d'intégration de tous les partis impliqués). Par exemple, l'une de ces municipalités fait réaliser plusieurs de ses documents par des firmes externes. Voilà beaucoup d'argent investi pour qu'il ne reste en bout de ligne que les documents nécessaires plutôt qu'une méthode de travail acquise et accrue. La rationalisation des ressources humaines entraînée par les fusions ne pourra que remettre en question la quantité d'information mise à jour suivant un cycle régulier, quitte à s'en tenir qu'au minimum nécessaire.

10. En amont de tout ça, comment traduit-on les problèmes rencontrés au niveau de l'organisation informatique de l'information en langage de gestion municipale?

- **En quoi sentez-vous que la gestion informatisée de l'information peut influencer votre travail?**

La dépendance au système, une fois qu'il est actif, devient très grande. Une panne informatique sérieuse paralyserait les services à tous les niveaux.. La vitesse de traitement des dossiers gagnée serait reperdue quitte à revenir aux méthodes traditionnelles (à la mitaine). On devra donc retourner sur le terrain pour relever, constater, affirmer, infirmer une information proposée ou manquante alors que nous l'avons maintenant disponible dans plusieurs banques de données.

Les coûts liés à l'implantation de ces systèmes étant élevés, nous sommes toujours face aux mêmes dilemmes. Par exemple, oui, nous auront les budgets pour monter notre cartographie informatisée mais en revanche, nous ne pourrons la tenir à jour. C'est ce qui

est arriver à l'une des municipalités fusionnées. Il faut donc constamment doser la précision de notre information car une information plus précise coûte plus cher à mettre à jour. C'est encore pour cette raison qu'on doit toujours optimiser les relations entre les informations quitte à l'enrichir le plus possible par elle même et, du même coup minimiser les manipulations humaines répétées de cette dernière (réf: l'évaluateur qui génère et numérise lui-même ses certificats).

Commentaires sur le système et ajouts :

Les objectifs de la Ville de Saint-Luc sont clairs. Il ne veulent pas dépendre de consultants externes en informatique qui prétendront répondre à leur besoins en systèmes d'informations. La ville préconise à cet égard une approche plutôt autonome ou autosuffisante. Ce qui n'est pas le cas de toutes les villes impliquées dans la fusion. Elle veut centraliser l'information pour l'ensemble de ses services de sorte à la diffuser le plus rapidement possible à l'interne. Cette optique favorisera la communication entre les départements en informant chacun des avois de l'autre en information. On évitera donc des dépenses inutiles en acquisition ou en production de données que l'on a déjà. La municipalité se donne comme objectif l'année 2002, pour se doter d'une fenêtre informatique permettant à chaque service de la boîte de consulter des données ou des couches graphiques, pertinentes à leurs fonctions, à même la base de données centrale.

Le logiciel Access 97 est le choix pour la gestion des bases de données. Il est le plus utilisé et le mieux utilisé (au meilleur de ses capacités). Le rôle d'évaluation sur un ancien format a été transféré en "*.mdb". L'idée au départ était de ne pas être obligé, pour chaque demande de permis, de remplir la même information qui se répète sur trois ou quatre formulaires différents. Bien sûr, l'utilisation du système ne s'arrête pas là aujourd'hui car on l'a poussé plus loin que prévu au fil des avancements du projet. Plusieurs requêtes peuvent être effectuées et rendues sous forme d'état, suivant une interface adaptée et personnalisée par des gens, à l'interne, répondant à leur propres besoins dans l'environnement informatique Access.

Le logiciel AutoCAD Map 2000 supporte tout ce que l'on peut relier à la matrice graphique en termes de cartographie et d'imagerie satellite. Le système a été mis en place par un consultant externe qui a complété la polygonisation et la référence détaillée avec points de repère (centroïde et numéro de matricule lié au lot). Les fonctions d'AutoCAD les plus utilisées au niveau graphique sont à mon grand désarroi le calcul de distances (addition de tronçon de réseau) ainsi que le calcul de surfaces groupées. Pour le reste, on ajoute dans un répertoire, un à la suite des autres, les dessins techniques exigés aux firmes lors de la réalisation de projet (avec un nom de fichier correspondant à son emplacement réel). En fait, au niveau graphique, l'automatisation est très faible. Ici, tel qu'anticipé, l'informatique ne fait que remplacer le papier. Oui, on gère plusieurs couches que l'on ajoute à la carte de fond mais ce ne sont en fait que des calques (zonage, infrastructures). Au niveau de ce logiciel, ce n'est pas la censure par manque de fonction qui limite son utilisation mais plutôt un manque de connaissance des capacités de ce dernier, ainsi que des ressources limitées en frais de possibilités d'exploration du domaine. D'un point de vue plus urbanistique, la tâche du service consiste en l'identification de polygones par numéros de lots cadastrés et par numéros de permis. L'imagerie satellite leur permet seulement de comparer plusieurs états dans le temps.

Nous avons discuté la pertinence de numériser l'ensemble des règlements municipaux sous forme de base de données. L'urbaniste n'en voyait pas d'utilité pour l'instant, ne serait-ce que pour l'amendement mensuel de règlements qui nécessiteraient une mise à jour supplémentaire. La bonne vieille méthode du cartable de règlement fait apparemment toujours l'affaire. Selon lui, il n'y aurait pas beaucoup d'analyses possibles que l'on pourrait faire plus rapidement à l'aide de requêtes informatiques sur les règlements. La question de lier dans une base de données ces règles à leurs numéros de zones respectifs, sur le plan de zonage numérique, ne lui semble pas pertinente. Selon lui, on a rarement eu besoin jusqu'à maintenant de sortir toutes les règles de zonage applicables à une zone précise au même moment. De toute façon, le projet de numérisation de la grille des usages (donc de la matrice à double entrée) remplirait un peu le même rôle et serait beaucoup plus légère à traiter et à archiver.