

Développement d'un modèle et d'un langage pour identifier et spécifier des contraintes d'intégrité dans les cubes de données spatiales

Yvan Bédard
Extraits adaptés de la soutenance de thèse de
Mehrdad Salehi

Département des Sciences Géomatiques
Université Laval

5 mars 2010

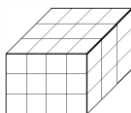


Contexte

- Les entrepôts des données spatiales font partie des systèmes d'aide à la décision
- Ces entrepôts des données sont souvent structurés en cube de données



+



=

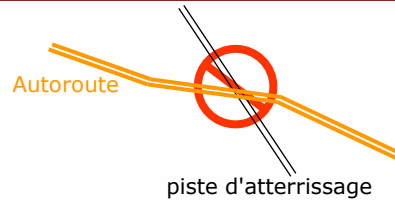
Cube de données spatiales

- La qualité des données dans ces cubes est importante
- CI jouent un rôle essentiel dans l'amélioration d'un des éléments de la qualité des données: la cohérence logique (ISO/TC211 - 19113)

Exemples

CI sont des règles qu'on impose aux données afin d'éviter les incohérences dans les bases de données.

Une autoroute ne doit pas traverser une piste d'atterrissage.



Un barrage hydroélectrique doit être sur un cours d'eau.



3

Problématique

Il n'existe aucun:

- modèle de cubes de données spatiales qui inclut explicitement les CI
- langage formel pour spécifier des CI des cubes de données spatiales

Problématique : La difficulté d'identifier et de spécifier des CI de cubes de données spatiales

4

Hypothèse

Il est possible de faciliter l'utilisation des CI dans les cubes de données spatiales:

- Identification plus facile des CI des cubes de données spatiales
- Spécification plus formelle des CI des cubes de données spatiales

5

Objectif principal et objectifs spécifiques

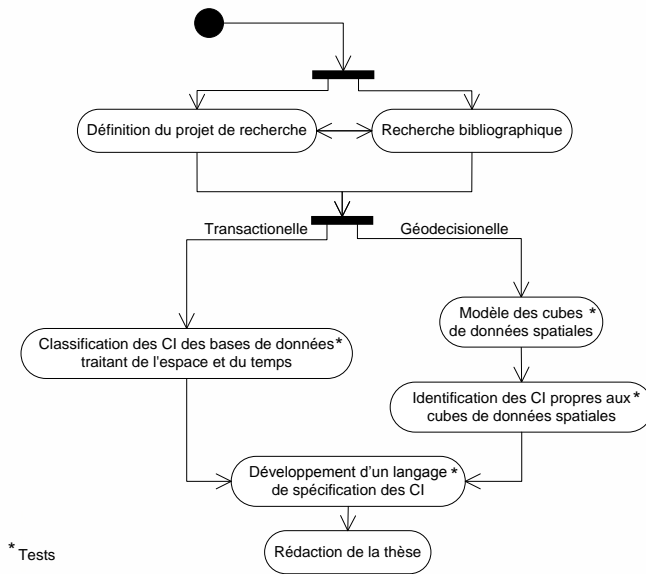
L'**objectif principal** de cette recherche est d'identifier plus facilement et de spécifier formellement des CI des cubes de données spatiales.

Objectifs spécifiques:

- Classifier les CI des bases de données traitant de l'espace et du temps
- Proposer un modèle formel des cubes de données spatiales
- Identifier les types de CI des cubes de données spatiales
- Développer un langage formel de spécification des CI des cubes de données spatiales

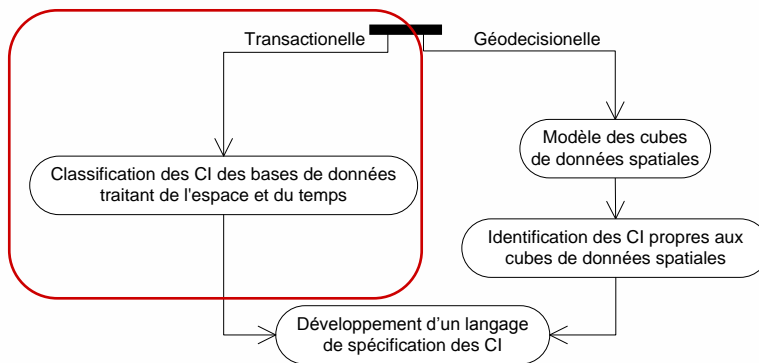
6

Méthode de recherche



7

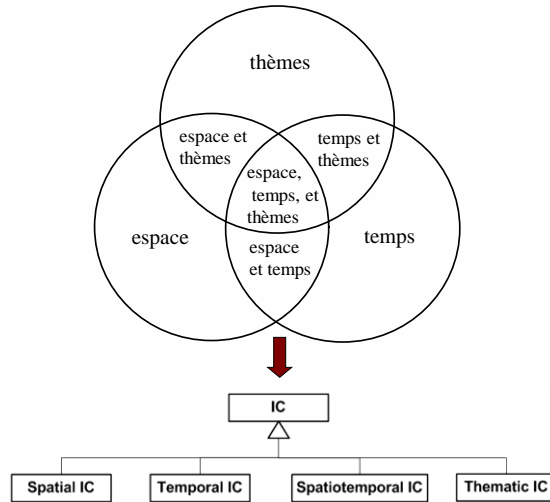
Plan



8

Classification formelle des CI des bases de données traitant de l'espace et du temps

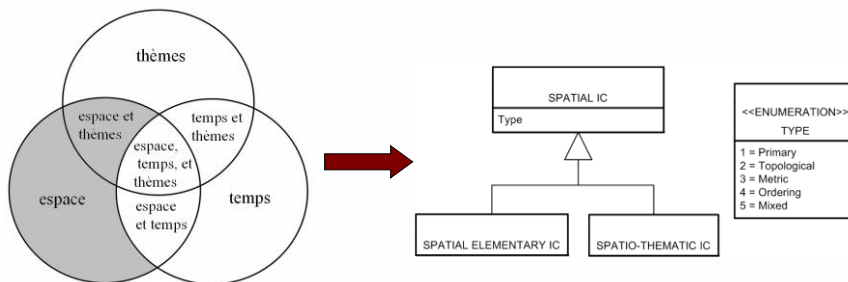
Les cubes de données spatiales incluent généralement les données spatiales, temporelles, et thématiques.



9

Classification formelle des CI des bases de données traitant de l'espace et du temps

CI Spatiale : Il s'agit d'une CI dont la définition ne contient que des concepts spatiaux et peut-être des concepts thématiques.



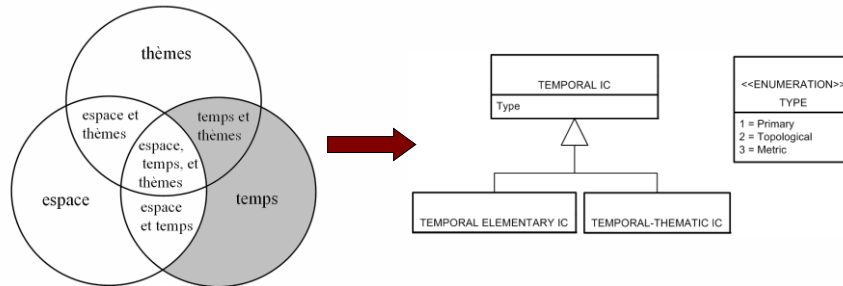
Exemple:

La distance entre une école et une station d'essence doit être supérieure à 300 mètres (CI spatio-thematic metric).

10

Classification formelle des CI des bases de données traitant de l'espace et du temps

CI Temporelle: Il s'agit d'une CI dont la définition ne contient que des concepts temporels et peut-être des concepts thématiques.



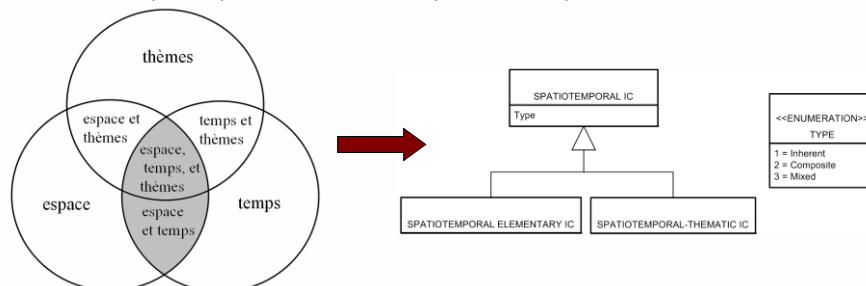
Exemple:

La date de construction d'un bâtiment doit être avant la date de rénovation (CI temporel-thematic topological).

11

Classification formelle des CI des bases de données traitant de l'espace et du temps

CI Spatiotemporelle: Il s'agit d'une CI dont la définition contient des concepts spatiotemporels ou des concepts spatiaux et des concepts temporels en même temps, et peut-être des concepts thématiques.



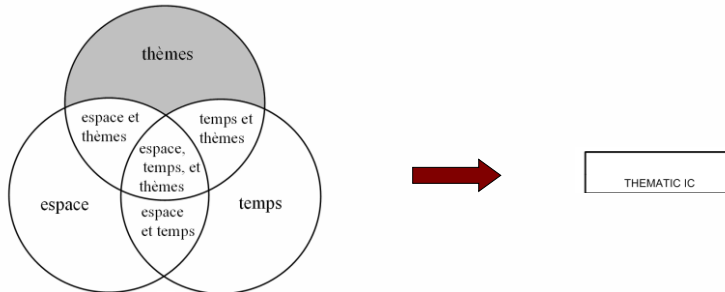
Exemples:

- La distance entre une école et une station d'essence doit être de plus de 300m après 1970(CI spatiotemporal-thematic composite).
- Si l'accélération d'un *moving-point* est supérieure à zéro, sa vitesse doit augmenter (CI spatiotemporal elementary inherent).

12

Classification formelle des CI des bases de données traitant de l'espace et du temps

CI Thématique : Il s'agit d'une CI dont la définition ne contient que des concepts thématiques.

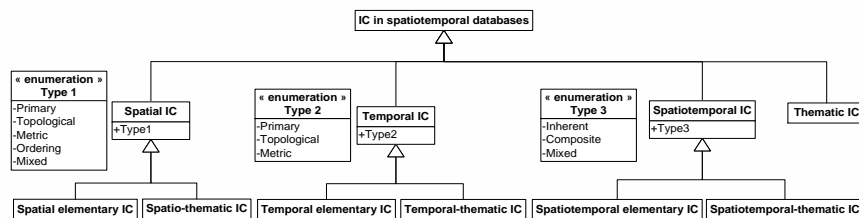


Exemple:

L'âge d'un parent doit être supérieur à celui de son enfant.

13

Classification formelle des CI des bases de données traitant de l'espace et du temps



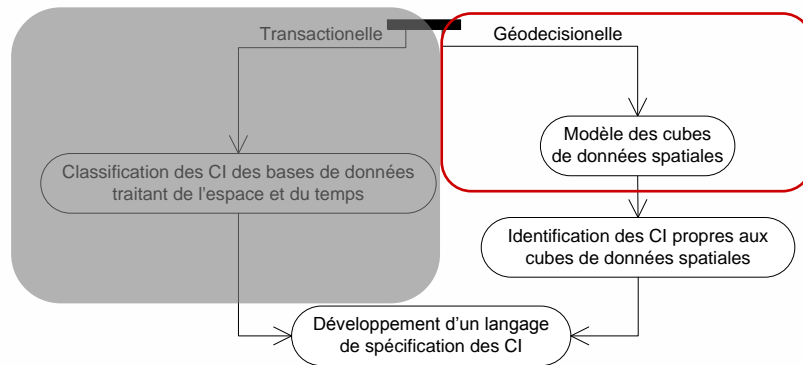
Le résultat de cette classification à été utilisé pour analyser:

- **Qui** spécifie la CI (ex. utilisateur, développeur des logiciels)
- **Où** la CI est spécifiée (ex. dans un schéma ou un dictionnaire)
- **Comment** ces CI sont spécifiées (ex. avec les concepts spatiaux, temporels, spatiotemporels)

Test: Un modèle conceptuel qui s'appelle "gestion des catastrophes" a été défini et plusieurs CI de cet exemple ont été classifiées

14

Plan



15

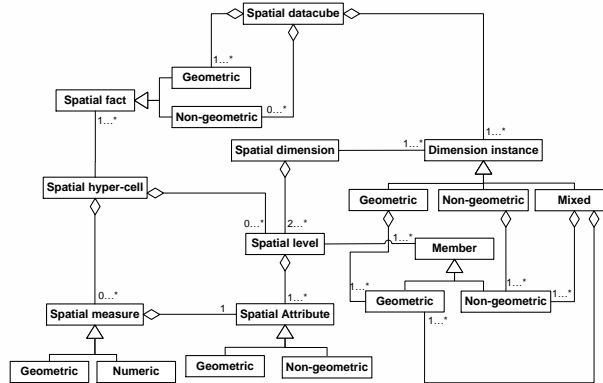
Un modèle formel pour les cubes de données spatiales

Composante	Définition	Formalisation
Attribut	décrit un niveau	a_i
Niveau	granularité de l'analyse sur une dimension	$l = \{a_1, \dots, a_n\}$
Dimension	un axe d'analyse	$d = (\{l_1, \dots, l_n\}, <)$
Instance d'une dimension	instancie une dimension	$di = (\{m_1, \dots, m_n\}, \leq)$
Membre	instance d'un niveau	$m = (\{a_1, \dots, a_k\}, \{v_1, \dots, v_k\}, :)$
Mesure	attribut analysé en tenant compte des niveaux des dimensions	ms_i
Hyper-cell	modélise plusieurs faits	$hc = (L, MS)$
Schéma de cube de données	structure d'un cube	(D, MS, HC)
Fait	décrit un événement pour la prise de décision	$f = (M, V)$
Cube de données	contenu d'un cube	(DI, F)

16

Un modèle formel pour les cubes de données spatiales

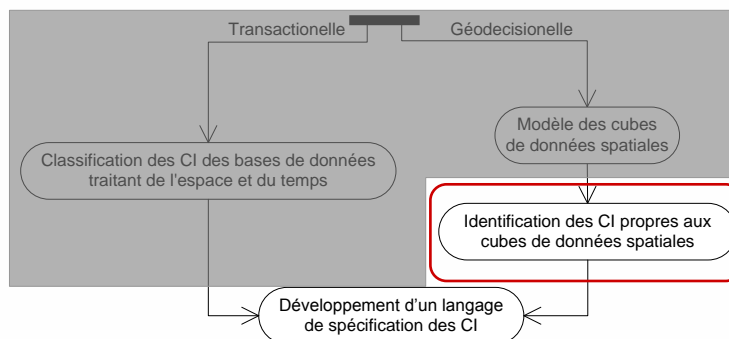
Différentes composantes "spatiales" d'un cube de données spatiales ont été identifiées et formellement définies



Test: Le modèle proposé a été utilisé pour définir un cube de données qui s'appelle "feu de forêt"

17

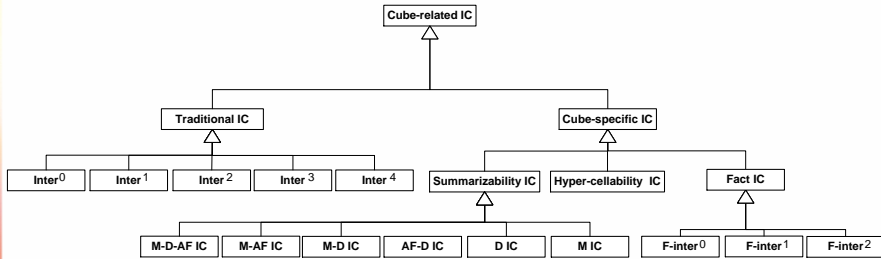
Plan



18

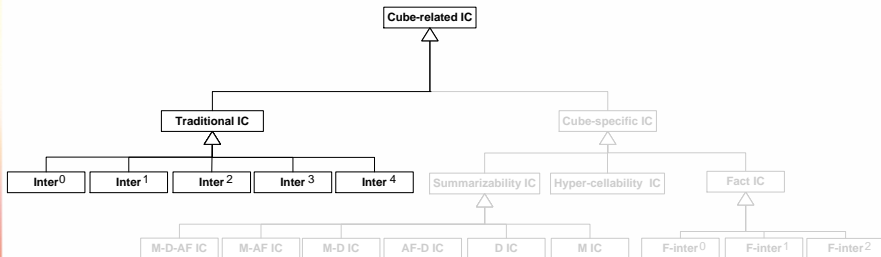
Identification des CI des cubes de données spatiales

Différentes catégories des CI:



19

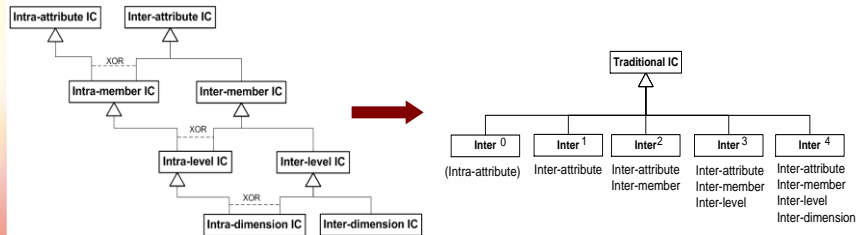
Identification des CI des cubes de données spatiales



20

Identification des CI des cubes de données spatiales

CI Traditionnelles: Permettent d'éviter les incohérences qui pourraient être causées lors de l'intégration des membres

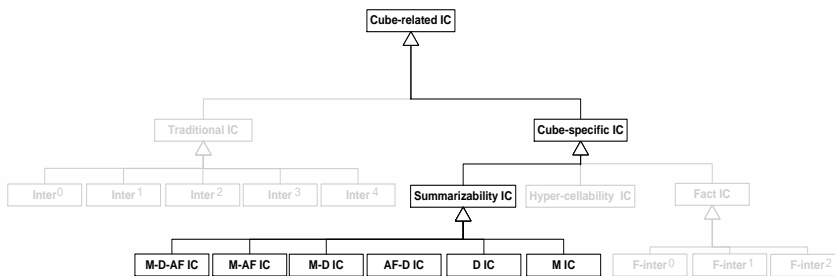


Exemple:

dimension "transport" avec le niveau "station = {position, existence, type}"

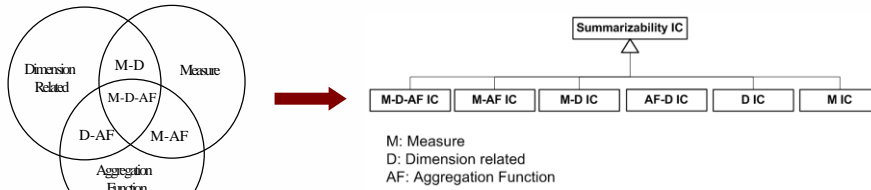
CI inter²: si les deux membres de niveau "station" ont le type «aéroport» et qu'ils existent sur la même période de temps, la distance entre les positions de ces deux membres doit être supérieure à 40 km. 21

Identification des CI des cubes de données spatiales



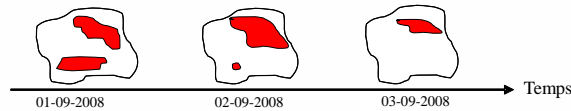
Identification des CI des cubes de données spatiales

CI Summarizability: Permettent d'éviter l'agrégation sans signification



Exemple:

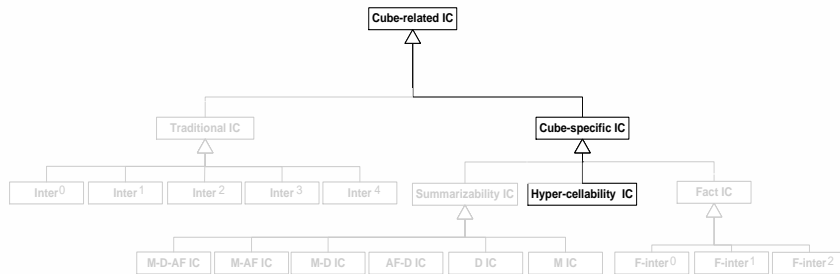
dimension: temps = {jour, mois}, mesure: superficie du feu, fonction agrégation: SUM



CI M-D-AF: L'agrégation de la mesure «superficie du feu» avec la fonction d'agrégation «SUM» sur la dimension «temps» n'a pas de sens.

23

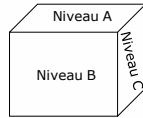
Identification des CI des cubes de données spatiales



24

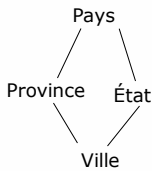
Identification des CI des cubes de données spatiales

CI Hyper-cellability : Permettent d'éviter la création des hyper-cells qui n'ont pas de sens

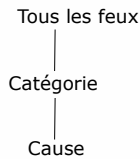


Est-ce que la combinaison des niveaux A,B, et C a du sens?

Exemple:



Dimension Région

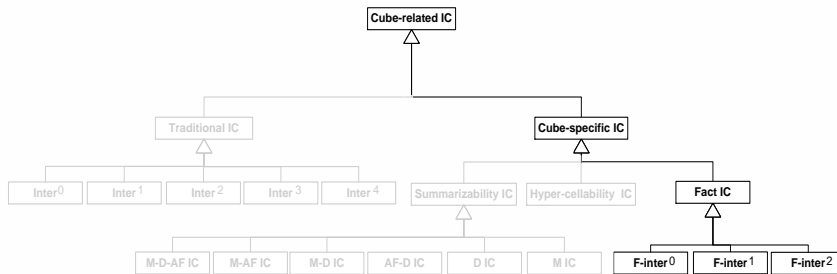


Dimension Feu

CI hyper-cellability: il n'y a pas de hyper-cell qui comprend à la fois les niveaux "Province" et "Cause".

25

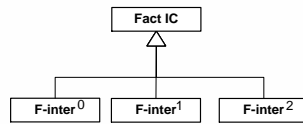
Identification des CI des cubes de données spatiales



26

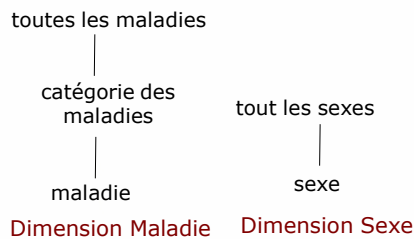
Identification des CI des cubes de données spatiales

CI Faits: Permettent d'éviter les faits qui ne sont pas souhaitables



	Fact		Hyper-cell	
	Inter	Intra	Inter	Intra
F-inter ⁰ IC				
F-inter ¹ IC				
F-inter ² IC				

Exemple:



CI F-inter⁰: il n'y a pas de fait qui comprend la maladie "cancer de Prostate" et le sexe "femme" .

27

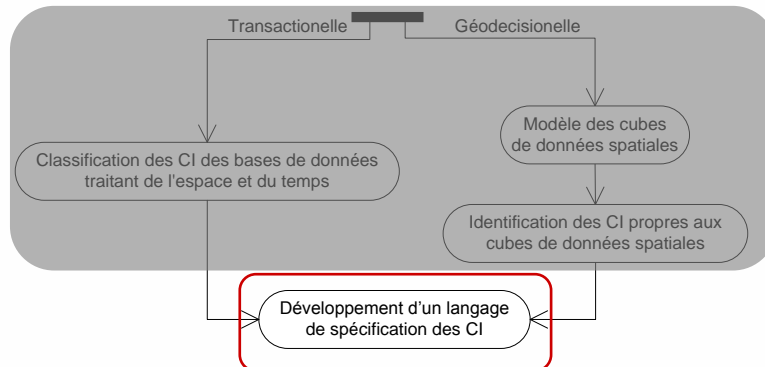
Identification des CI des cubes de données spatiales

Test:

- En utilisant plusieurs exemples de CI, les problèmes qui se produisent en l'absence de chaque catégorie de CI ont été illustrés, ainsi chaque catégorie a été testée
- Inversement, nous n'avons pas réussi à trouver d'exemples qui ne pouvaient pas entrer dans une des catégories définies

28

Plan



29




Développement d'un langage de spécification des CI

- Les langages de spécification des CI existants ont été catégorisés:
 - Langage naturel (libre, contrôlé)
 - Langage visuel
 - Langage logique
 - Langage visuel hybride
 - Langage naturel hybride (avec symboles, avec pictogrammes)
- Ces catégories de langages ont été comparées en se basant sur leur:
 - Pragmatisme
 - Richesse
 - Qualité sémantique
- Résultat: les langages suivants sont plus appropriés pour la spécification des CI au niveau conceptuel:
 - Langage naturel contrôlé
 - Langage naturel hybride avec les pictogrammes

30

Développement d'un langage de spécification des CI

Un langage pour exprimer les CI des cubes de données spatiales (traditionnelles, summarizability, hyper-cellability, et faits) a été développé

- Ce langage inclut:
 - Les concepts de cube
 - Dimension, niveau, membre, attribut, hyper-cell, mesure, fait
 - Les concepts spatiaux
 - topologiques, métriques, ordre
 - Les concepts temporels
 - topologiques, métriques
 - Les pictogrammes (, , )

31

Développement d'un langage de spécification des CI

Exemples:

- CI traditionnelle
 - (transport.station.station₁.existence, station₁.type = "aéroport"), → *Identifie membres*
(during), → *CI temporelle*
(transport.station.station₂.existence, station₂.type = "aéroport"), → *Identifie membres*
(station₁.position \square distance station₂.position \square > "40"km) → *CI spatiale*
- CI summarizability
 - (measure: "superficie du feu", aggregation function: "SUM", dimension: "temps")
is not summarizable.
- CI hyper-cellability
 - hyper-cell = (I_1, I_2, M) does not exist where → *Identifie hyper-cell*
 I_1 is the level "Région.Province" AND I_2 is the level "Feu.Cause". → *Identifie niveaux*

32

Développement d'un langage de spécification des CI

Test: Plusieurs CI ont été spécifiées avec le langage proposé

33

Limites et perspectives de recherche

- Réaliser plusieurs validations « grandeur réelle » de l'approche (ex. cube(s) réel(s), dans différents domaines) au fil des ans
- Développer un outil facilitant la spécification des CI des cubes de données spatiales
- Développer des stratégies pour implémenter les CI des cubes de données spatiales
- Développer un modèle et un langage pour les CI des cubes raster
- Introduire "CI Soft" vs "CI Hard"

34

Conclusion

- Les CI jouent un rôle important pour l'amélioration de la qualité des données
- L'absence d'un cadre pour identifier les CI des cubes de données spatiales et l'absence d'un langage pour les spécifier faisaient que les CI étaient traitées de façon non-systématisées et ad-hoc
- Notre solution: Le développement d'un modèle et d'un langage formel pour les CI de cubes de données spatiales
- Cette solution permet:
 - D'identifier ces CI a priori et de façon systématique
 - D'exprimer ces CI de façon formelle et utilisable
- Nous avons donc enrichi le corpus théorique du domaine

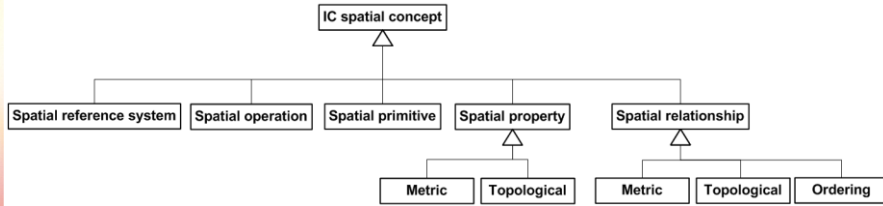
35

Remerciement

- Chaire de recherche industrielle de base de données géospatial décisionnelle (CRSNG et 9 partenaires)
- Faculté de foresterie, de géographie et de géomatique
- Fondation Avenor
- Institut Canadien des sciences géomatiques
- Département des sciences géomatiques
- Équipe de recherche du Dr. Yvan Bédard

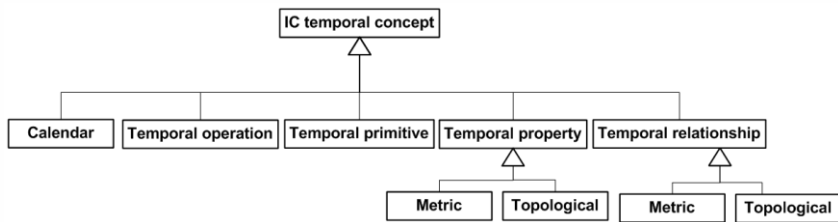
36

Concepts spatiaux des CI



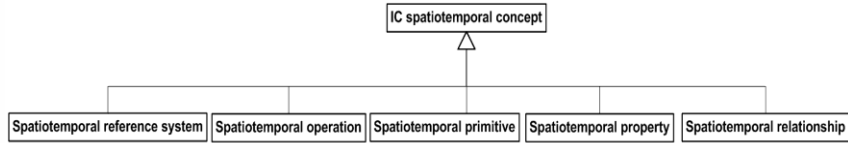
37

Concepts temporels des CI



38

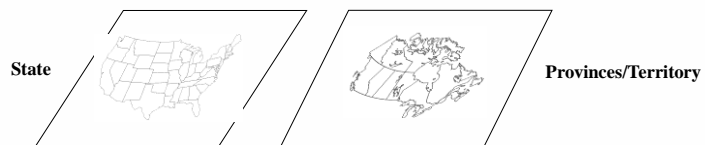
Concepts spatiotemporels des CI



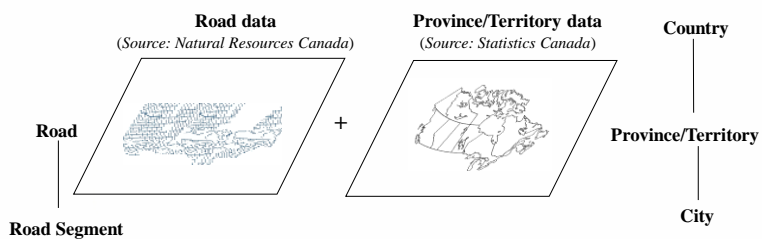
39

Intégration des membres

• Intégration horizontale



• Intégration verticale



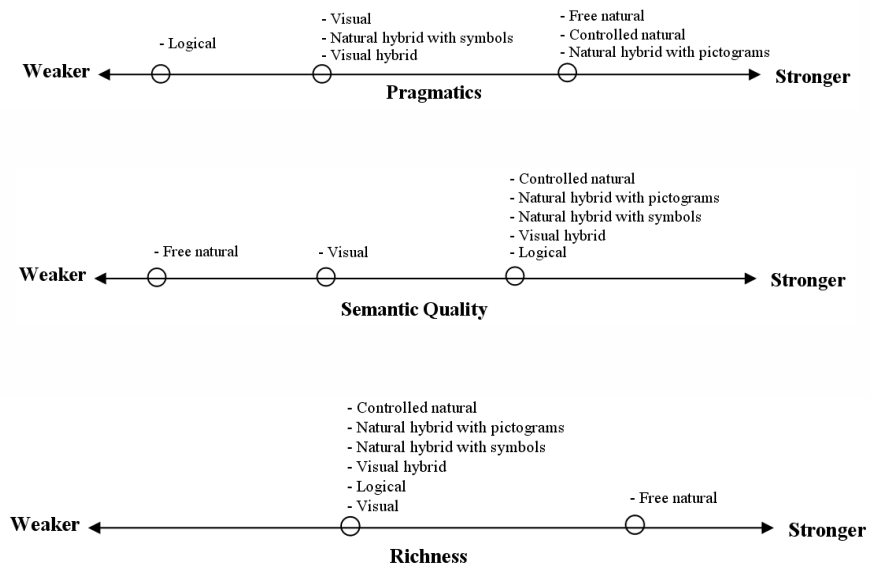
40

CI traditionnelles

	Attribute		Member		Level		Dimension	
	Inter	Intra	Inter	Intra	Inter	Intra	Inter	Intra
Inter ⁰								
Inter ¹								
Inter ²								
Inter ³								
Inter ⁴								

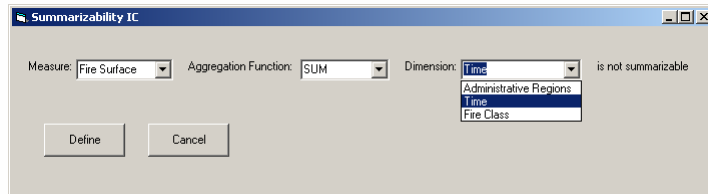
41

Comparaison des catégories des langages de spécification des CI



42

Un outil facilitant la spécification des CI des cubes de données spatiales



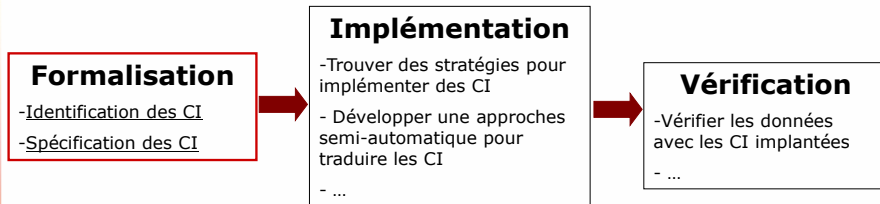
43

Caractéristiques du modèle formel pour les cubes de données spatiales

- Proposer des définitions explicites et précises pour les composantes spatiales
- Séparation entre la structure et le contenu
- Supporter plusieurs attributs pour un même niveau
- Supporter un ensemble de mesures pour un même fait
- Introduire la composante "hyper-cell"

44

Les processus de formalisation, implémentation, et vérification des CI



CI des cubes de données spatiales

