Fine-tuning the Semantic Interoperability between Geospatial Datacubes

Tarek Sboui

Yvan Bédard

UNIVERSITÉ LAVAL

DÉPARTEMENT DES SCIENCES GÉOMATIQUES

Presented by**: Tarek Sboui**

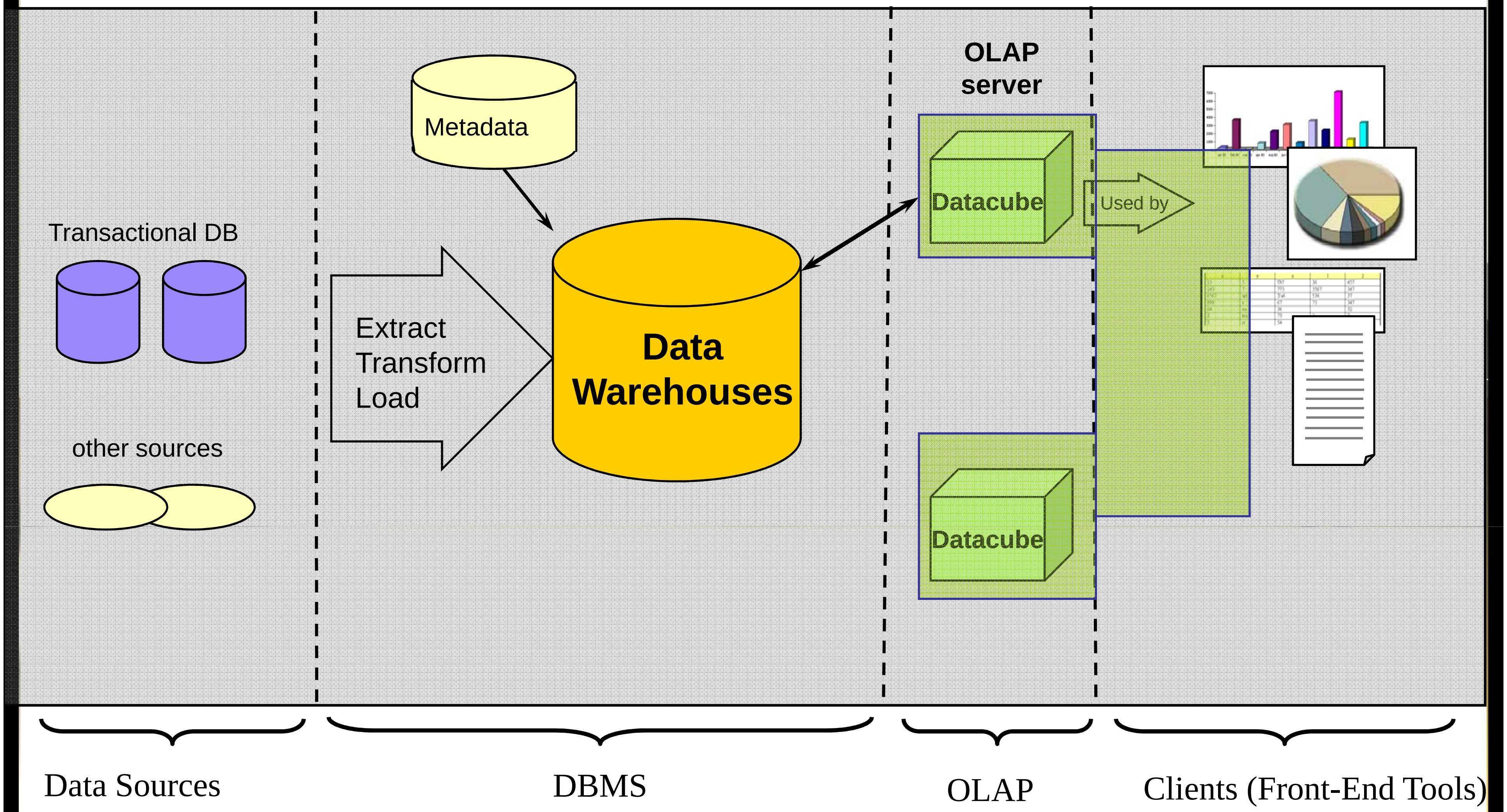
SeCoGIS 2010, Vancouver, Canada

Chaire de recherche industrielle
Bases de données géospatiales décisionnelles

Presentation outline

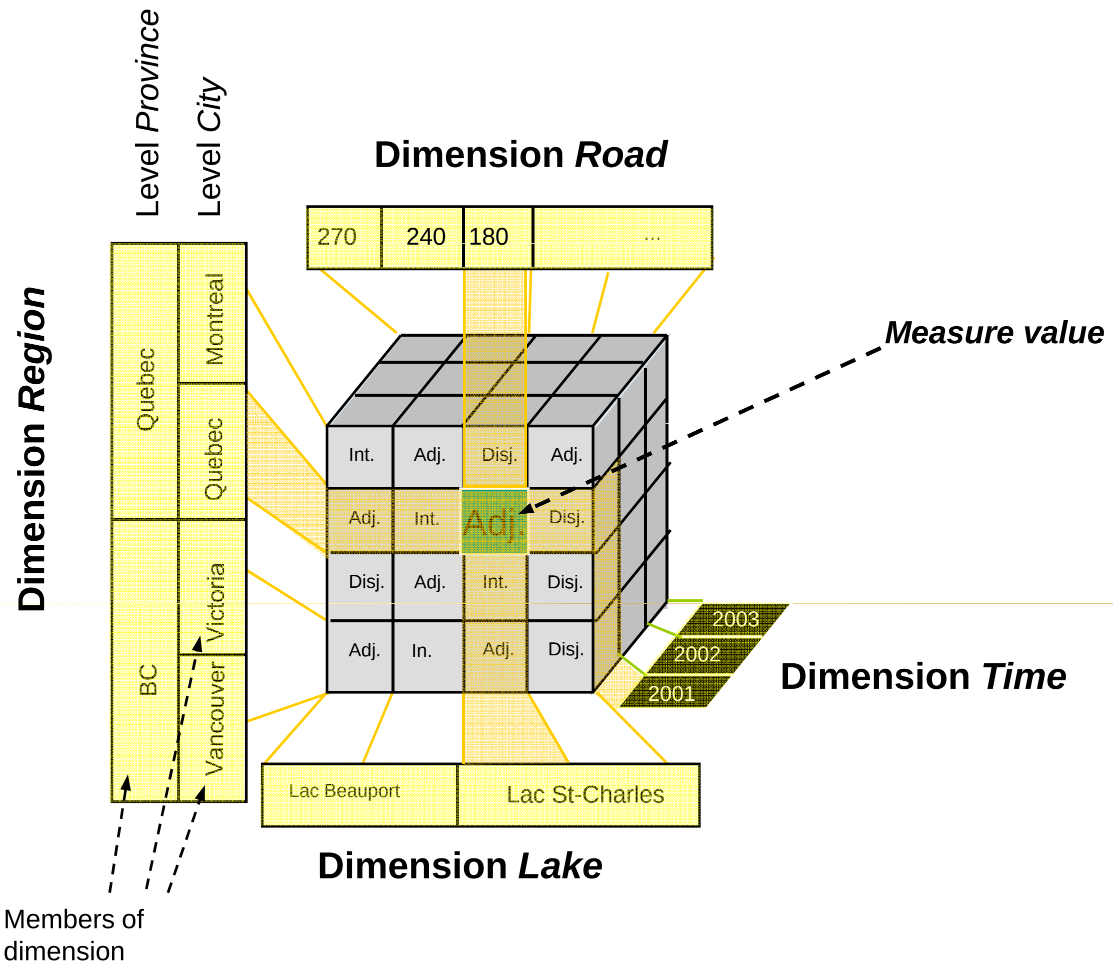
- Introduction
- Challenges
- Defining the interoperability between geospatial datacubes
- Proposed approach
  - Framework to support semantic interoperability between geospatial datacubes
  - MGsP: extending the GsP to support the interoperability between geospatial datacubes
- Conclusion

Introduction: needs to satisfy



Geospatial Datacubes

Datacubes «Roads and Lakes»



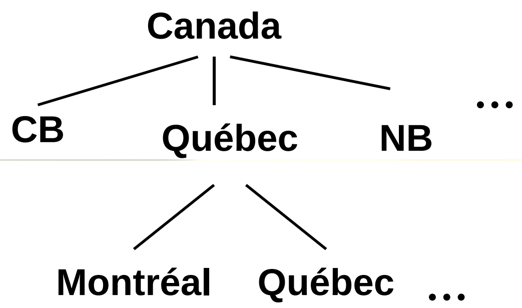
Hyper-cell: A set of combinations of measure values and members.

Geospatial Datacubes

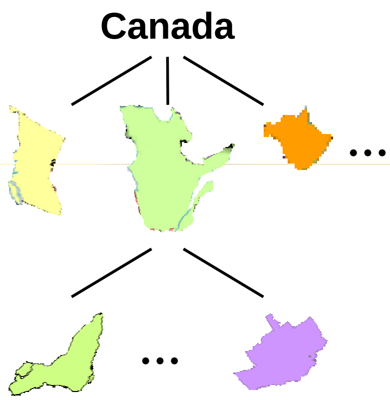
- Both dimensions and measures may contain geospatial components.

Example: Region dimension

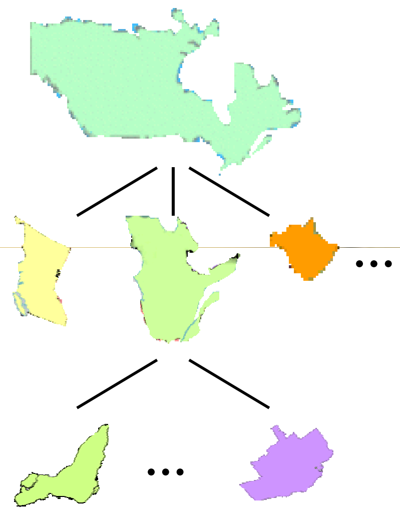
Non-geometric geospatial dimension



Mixed geospatial dimension

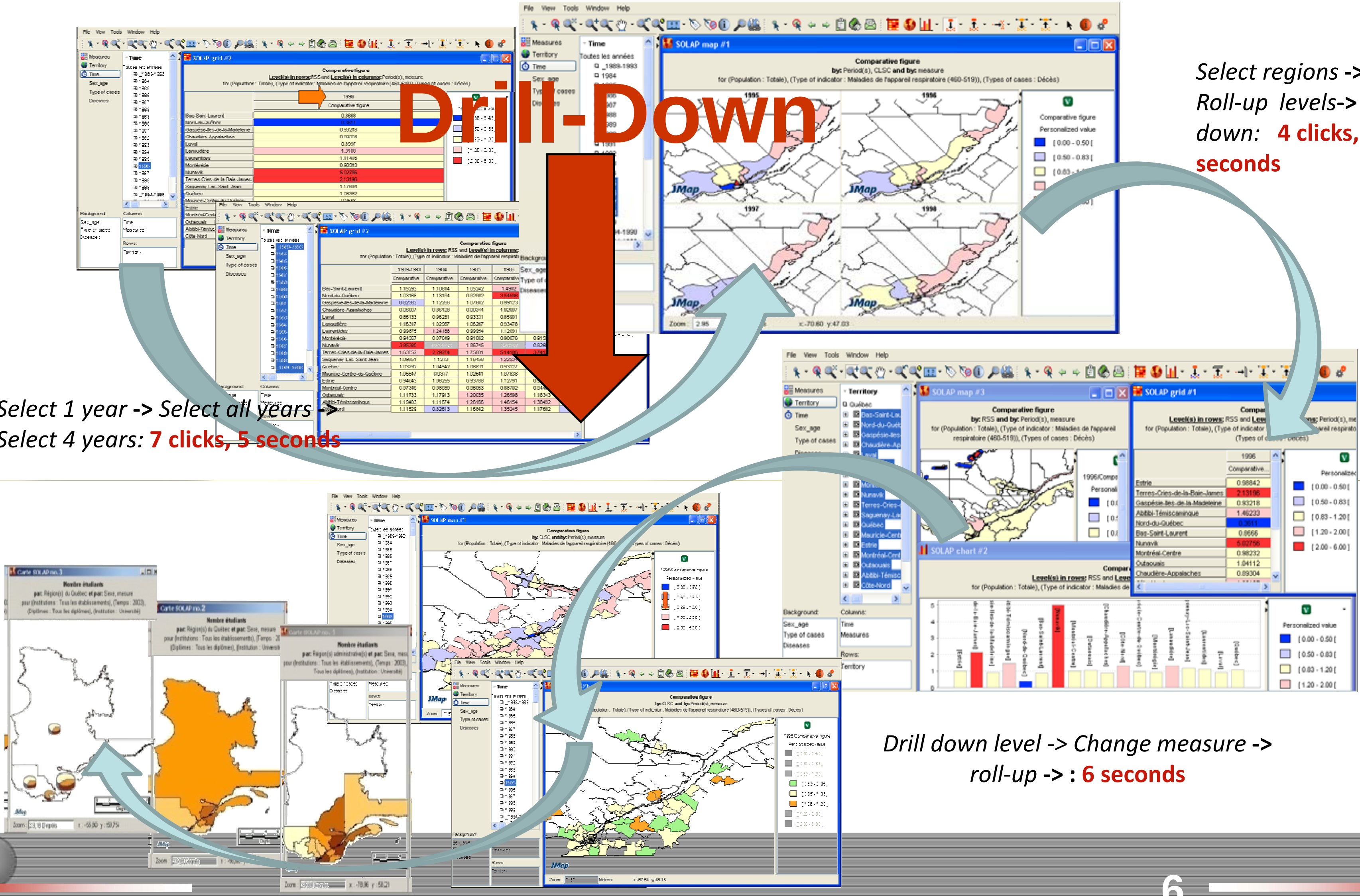


Geometric geospatial dimension



Geospatial Datacubes

- Supports Spatial On-Line Analytical Processing (SOLAP)



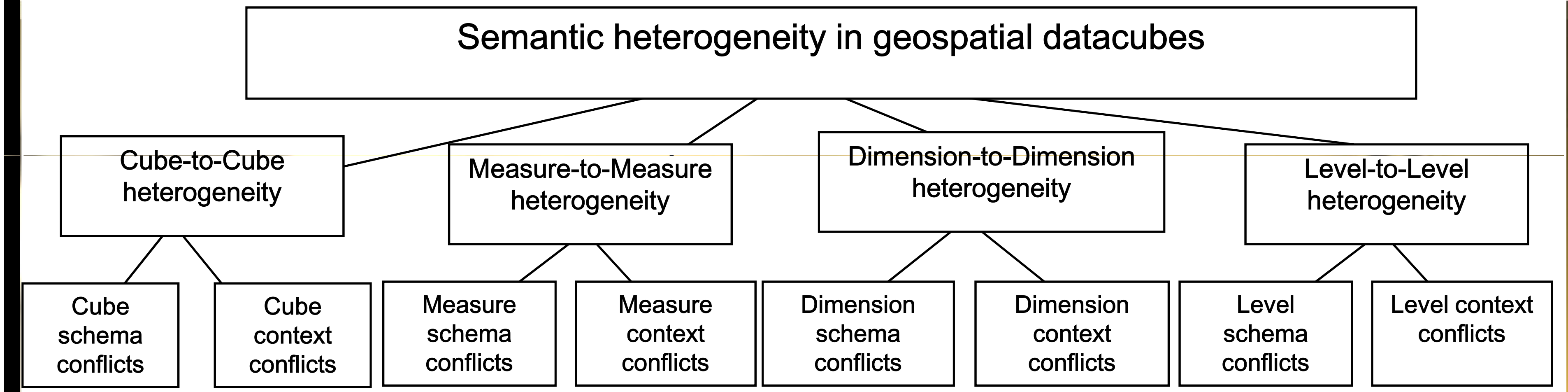
Select regions -> Roll-up levels-> drill-down: **4 clicks, 2 seconds**

Select 1 year -> Select all years -> Select 4 years: **7 clicks, 5 seconds**

Drill down level -> Change measure -> roll-up -> : **6 seconds**

Challenges

- **Geospatial datacubes are usually heterogeneous:**
  - Technical heterogeneities
  - **Semantic heterogeneities**



Challenges

- **Today's interoperability concepts and standards are for transactional systems (they do not support multidimensional concepts).**

Interoperability between geospatial datacubes

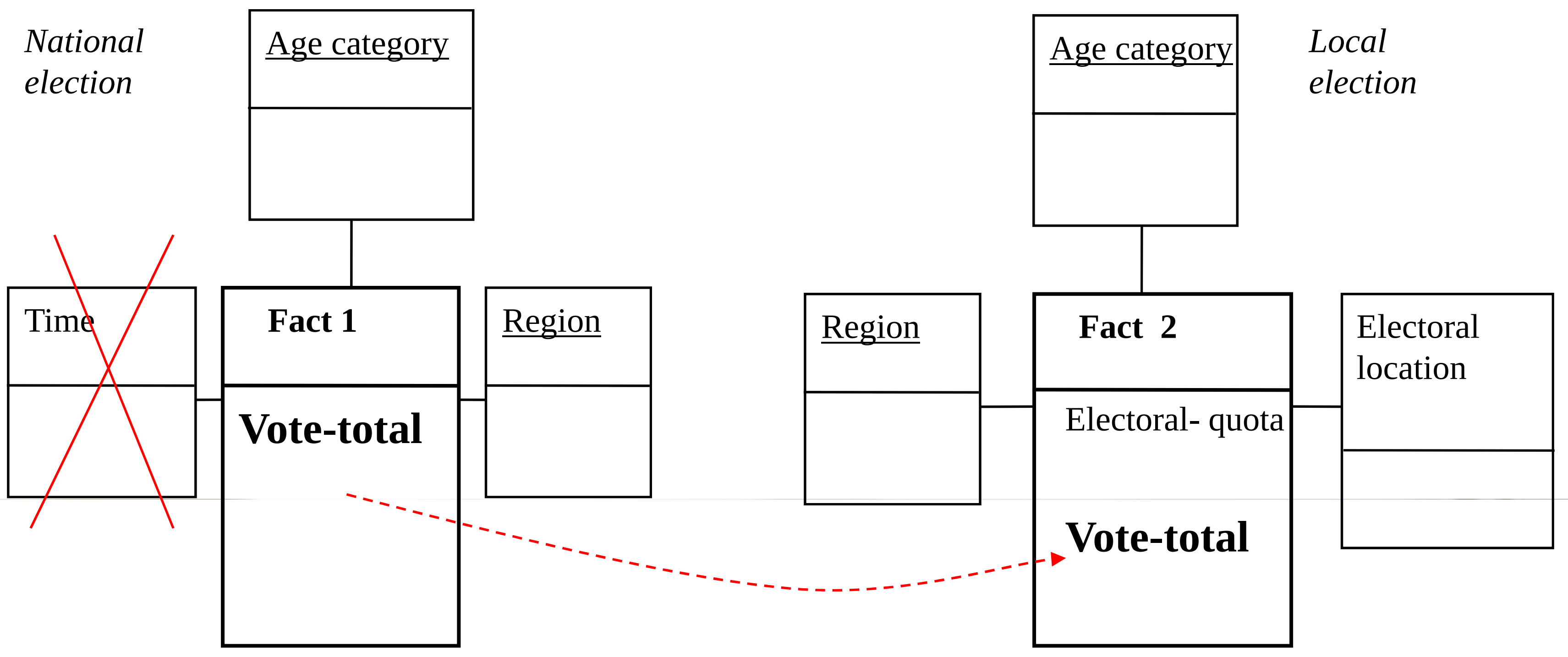
- **T**he interoperability between two geospatial datacubes C1 and C2 is the ability of C1 to request/respond to a service based on a mutual understanding**.**

- **Services could include:**
  - **importing/exporting instances contained in a datacube element (i.e., cube, measure, dimension, or level);**
  - **getting information about a geospatial datacube element (e.g., the type of method used for a geospatial measure);**
  - **verifying the change of a geospatial datacube element (e.g., change of definition, of a geometric representation).**

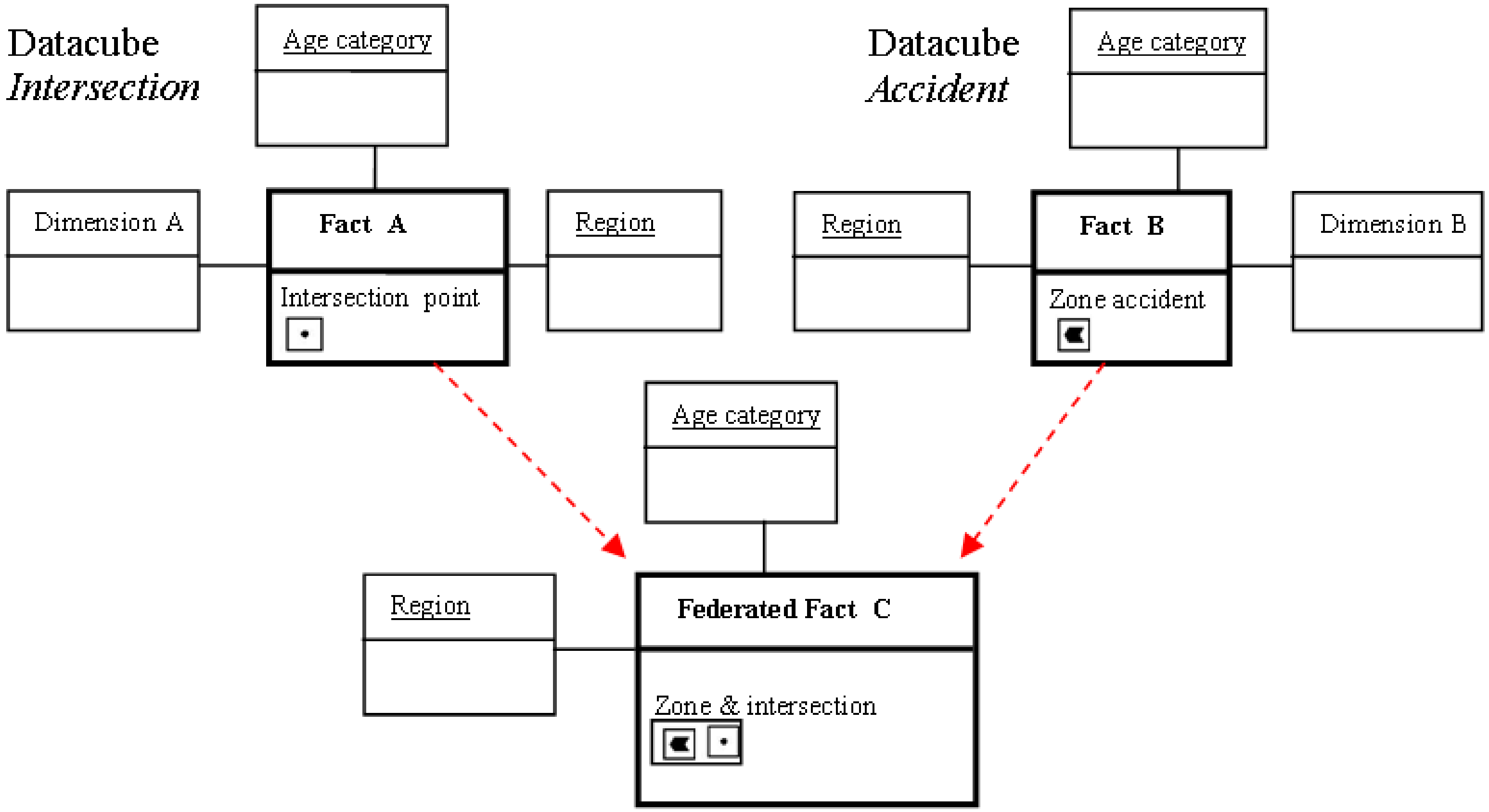
Interoperability between geospatial datacubes

- These services involve one or more of the following categories of actions:
  - Comparing an element of a geospatial datacube against an element of another.
  - Updating an element of a geospatial datacube based on the content of other datacubes involved in the interoperability process.
  - Integrating datacubes involved in the interoperability process. process.

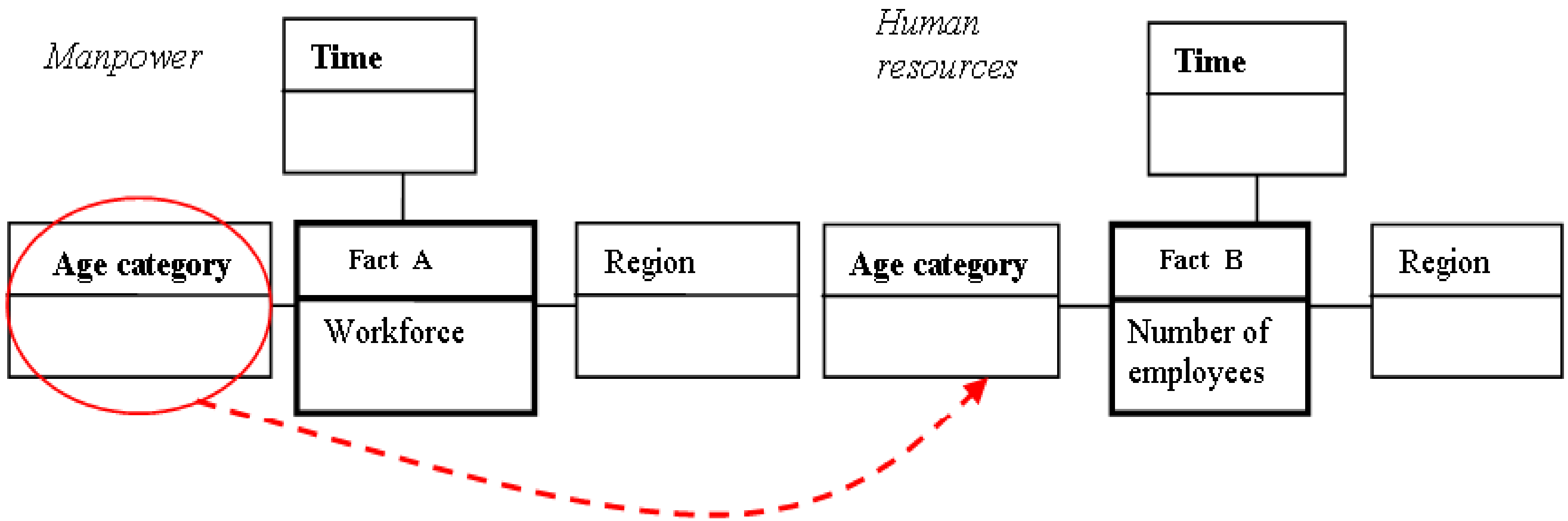
Interoperability between geospatial datacubes



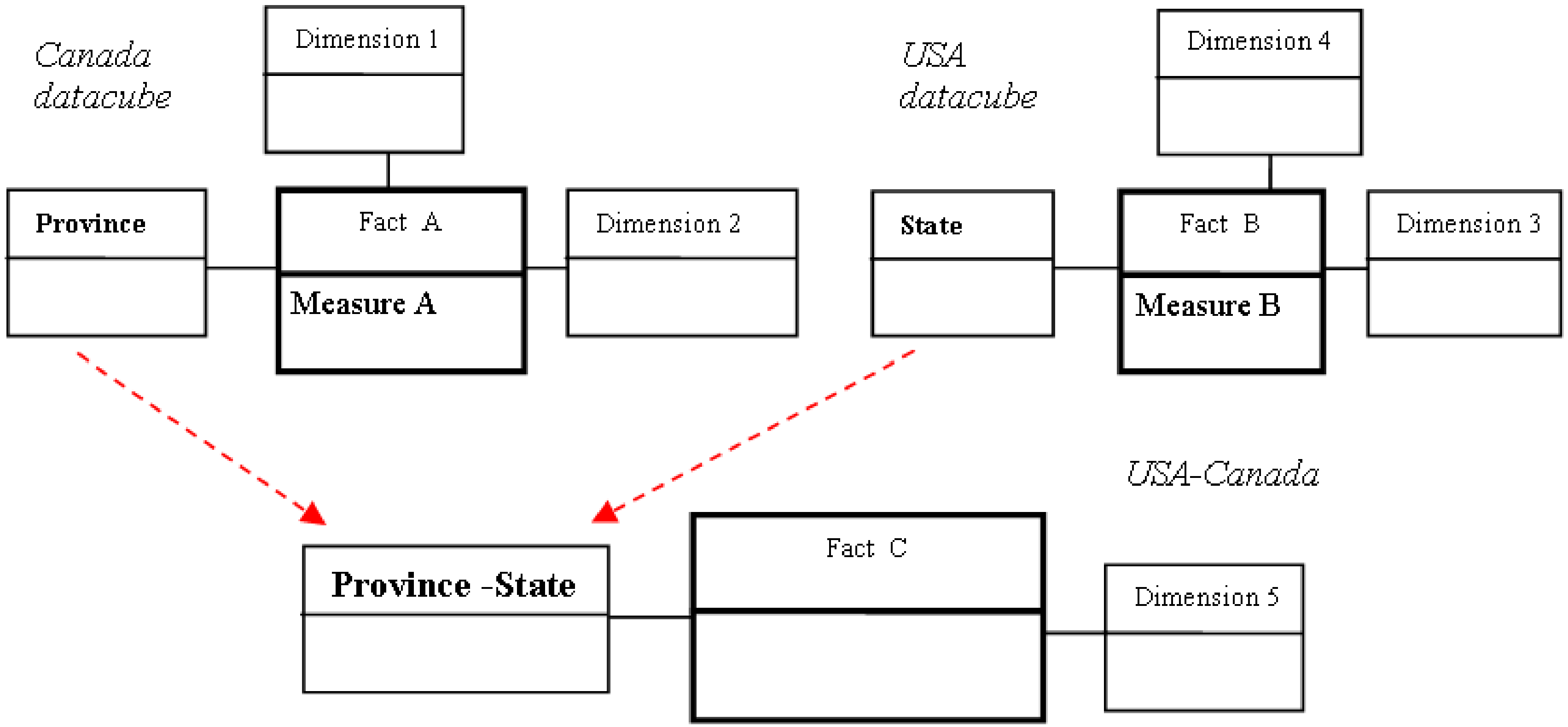
Interoperability between geospatial datacubes



Interoperability between geospatial datacubes



Interoperability between geospatial datacubes



Interoperability of Datacubes VS Interoperability of transactional DBs

Interoperability of transactional DBs	Interoperability of datacubes

Similarities

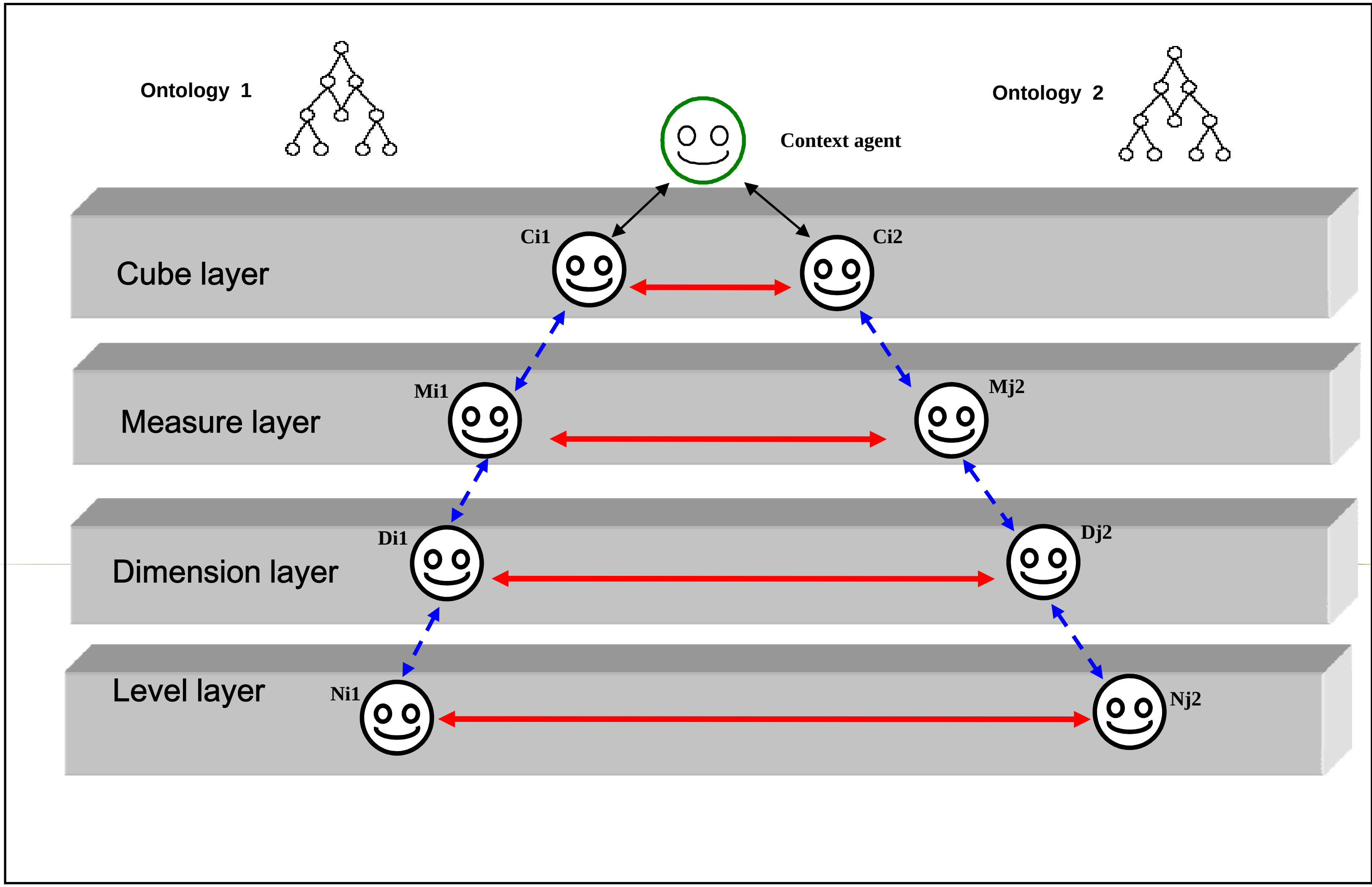
Reusing data

Facilitates an efficient exchange of information

Differences

Interoperability of transactional DBs	Interoperability of datacubes
Deals with the heterogeneities of DB concepts (i.e. tables, attributes, relations, etc.).	Deals with datacubes concepts (facts, measures, dimensions, levels) Deals with the semantic heterogeneities of aggregation and summarizing methods and algorithms, including summarizability conditions.

Proposed Approach: A framework for the interoperability between geospatial datacubes



Data interpretation

Interpreting multidimensional concepts:

- **We use and extend an approach that measures the semantic similarity.**

- **GsP: Geosemantic Proximity (Brodeur 2004):**
  - **Based on human communication**
  - **Deals with geospatial properties of data**

GsP: Geosemantic Proximity (Brodeur 2004)

- Assess the similarity of different geospatial concepts based on their intrinsic and extrinsic properties
- Consists of the intersection between the properties of different geospatial concepts

$$GsP\,(K, L) = \begin{pmatrix} \partial C_K \cap \partial C_L & \partial C_K \cap C_L^{\circ} \\ C_K^{\circ} \cap \partial C_L & C_K^{\circ} \cap C_L^{\circ} \end{pmatrix}$$

MGsP: Extending the Geosemantic Proximity

MGsP: Multidimensional Geosemantic Proximity

Gives the possibility to dig into and resolve semantic heterogeneity related to key notions of the multidimensional paradigm.

<u>GsP</u>

$$\begin{pmatrix} ExP_{B1} \cap GExP_{B2} & ExP_{B1} \cap InP_{B2} \\ InP_{B1} \cap ExP_{B2} & InP_{B1} \cap InP_{B2} \end{pmatrix}$$

$\Rightarrow$

<u>MGsP</u>

$$\begin{array}{c|cc} & MExP_{B2} & MInP_{B2} \\ \hline MExP_{B1} & MExP_{B1} \cap MExP_{B2} & MExP_{B1} \cap MInP_{B2} \\ MInP_{B1} & MInP_{B1} \cap MExP_{B2} & MInP_{B1} \cap MInP_{B2} \end{array}$$

MGsP: Extending the Geosemantic Proximity

Measure:

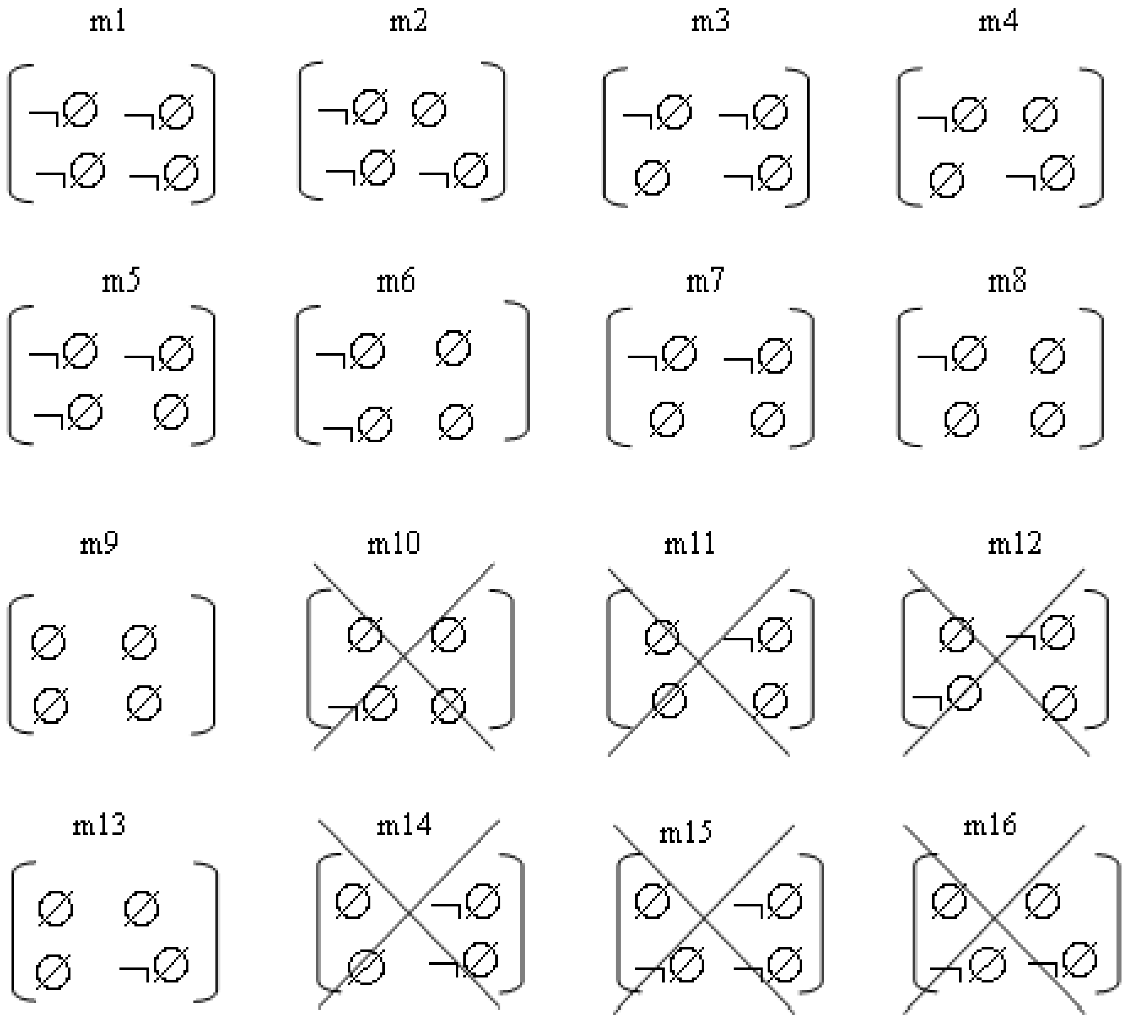
	$hyper\_cell_{M2}$	$measure\_function_{M2}$
$hyper\_cell_{M1}$	$hyper\_cell_{M1} \cap hyper\_cell_{M2}$	$hyper\_cell_{M1} \cap measure\_function_{M2}$
$measure\_function_{M1}$	$measure\_function_{M1} \cap hyper\_cell_{M2}$	$measure\_function_{M1} \cap measure\_function_{M2}$

MGsP: Extending the Geosemantic Proximity

Dimension:

	$hyper\_cell_{D2}$	$aggregation_{D2}$
$hyper\_cell_{D1}$	$hyper\_cell_{D1} \cap hyper\_cell_{D2}$	$hyper\_cell_{D1} \cap aggregation_{D2}$
$aggregation_{D1}$	$aggregation_{D1} \cap hyper\_cell_{D2}$	$aggregation_{D1} \cap aggregation_{D2}$

MGsP: Extending the Geosemantic Proximity



Category 1:

Common hyper-cell (MExP)
& common MInP

Category 2:

Common hyper-cell (MExP)
& no common MInP

Category 3:

No common hyper-cell (MExP)
& no common MInP

Category 4:

No common hyper-cell (MExP)
& common MInP

Conclusion

- We defined a communication framework which is based on datacubes agents defined according to different layers .
- We proposed an extension to the geosemantic proximity (MGsP).

Conclusion

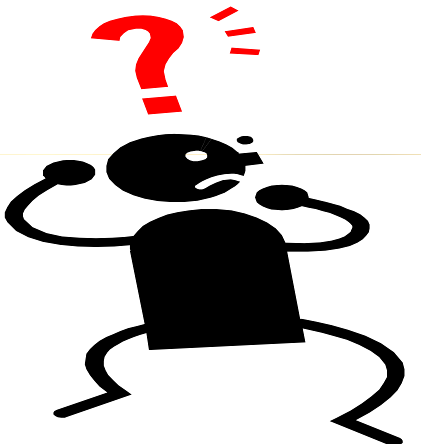
- **A prototype has been developed to experiment the proposed approach.**
- **Experimentations have been conducted using different geospatial datacubess:**
  - **The first datacube is used to determine the distribution of the population in specific areas and periods.**
  - **The second datacube intends to analyze the risk of fire in Canadian forests according to a set of criteria (e.g., time and regions).**
- **They demonstrated the convenience of the MGsP for the interoperability of geospatial datacubes.**

Future Works

- Defining more refined attributes for the MGsP. For example, for the aggregation attribute we can define the aggregation domain and aggregation constraint.

- Use Semantic Web technology to enhance reasoning about the multidimensional concepts.

Thanks !



Need for the interoperability

- **Interoperability between geospatial datacubes may be required in many situations.**
  - *Simultaneous and rapid navigation through different datacubes*: Users from different disciplines may need to access and navigate simultaneously through heterogeneous geospatial datacubes. Navigating separately through each datacube would be an arduous work for users, since they likely need to make extra efforts to manually resolve the problems of heterogeneity between datacubes (e.g., comparing the meaning of concepts and establishing a mapping between them). The principal aim of interoperability is to automatically overcome such differences and, hence, can considerably facilitate the navigation task.

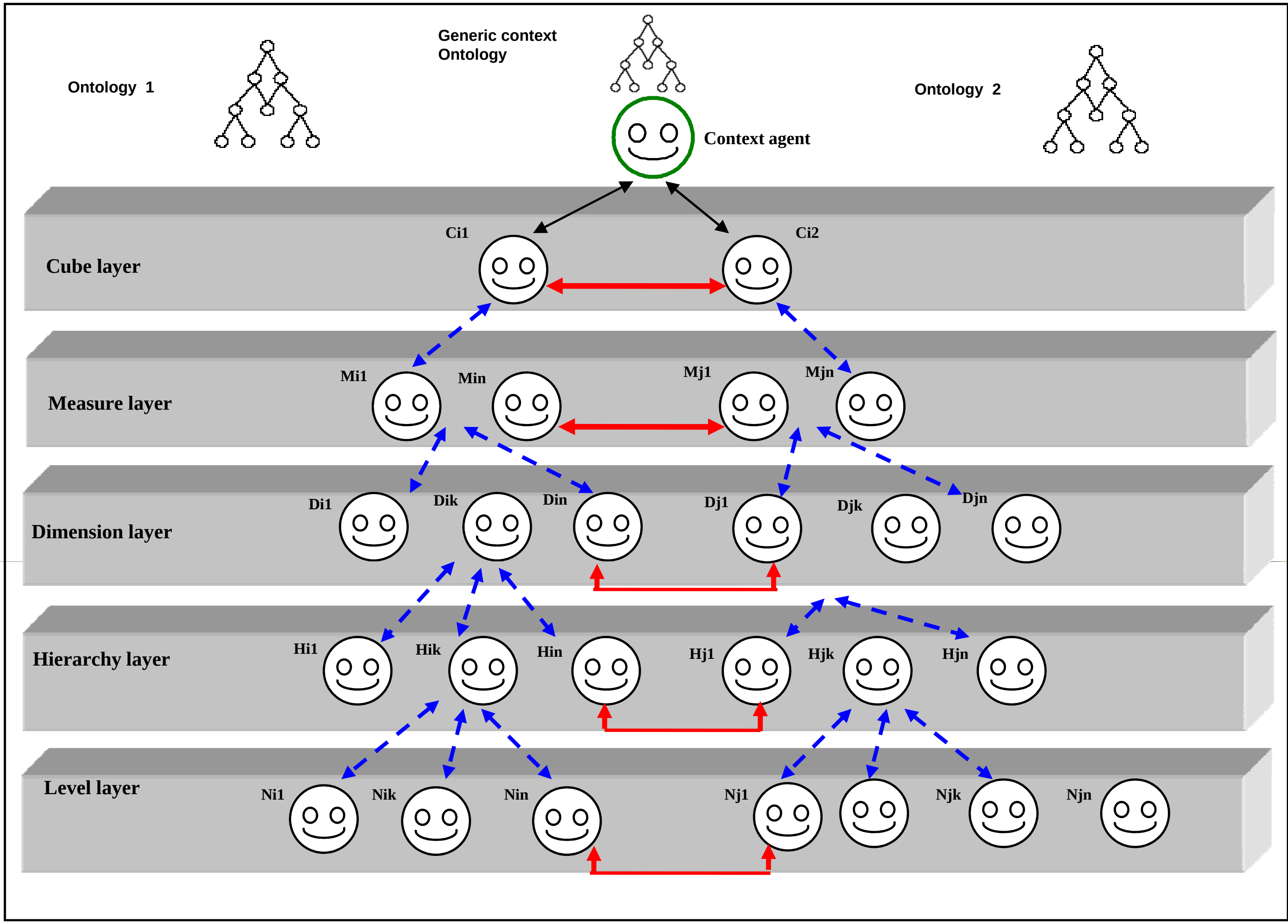
Need for the interoperability

- *Rapid insertion of data in a datacube*: While data in datacubes are usually collected from legacy systems, they can be imported from other heterogeneous datacubes (Bédard and Han 2008). We may need to rapidly insert new data (e.g., measures, members and member properties) in a geospatial datacube from other datacubes.

- *Interactive and rapid comparison of scattered decisional data to analyze phenomena changes*: In order to analyze phenomena change (e.g., forest stand dynamics), we need to compare data describing these phenomena at different epochs. We may need to compare data stored in geospatial datacubes built also at different epochs. Interoperating geospatial datacubes would permit interactively comparing data and analyzing changes.

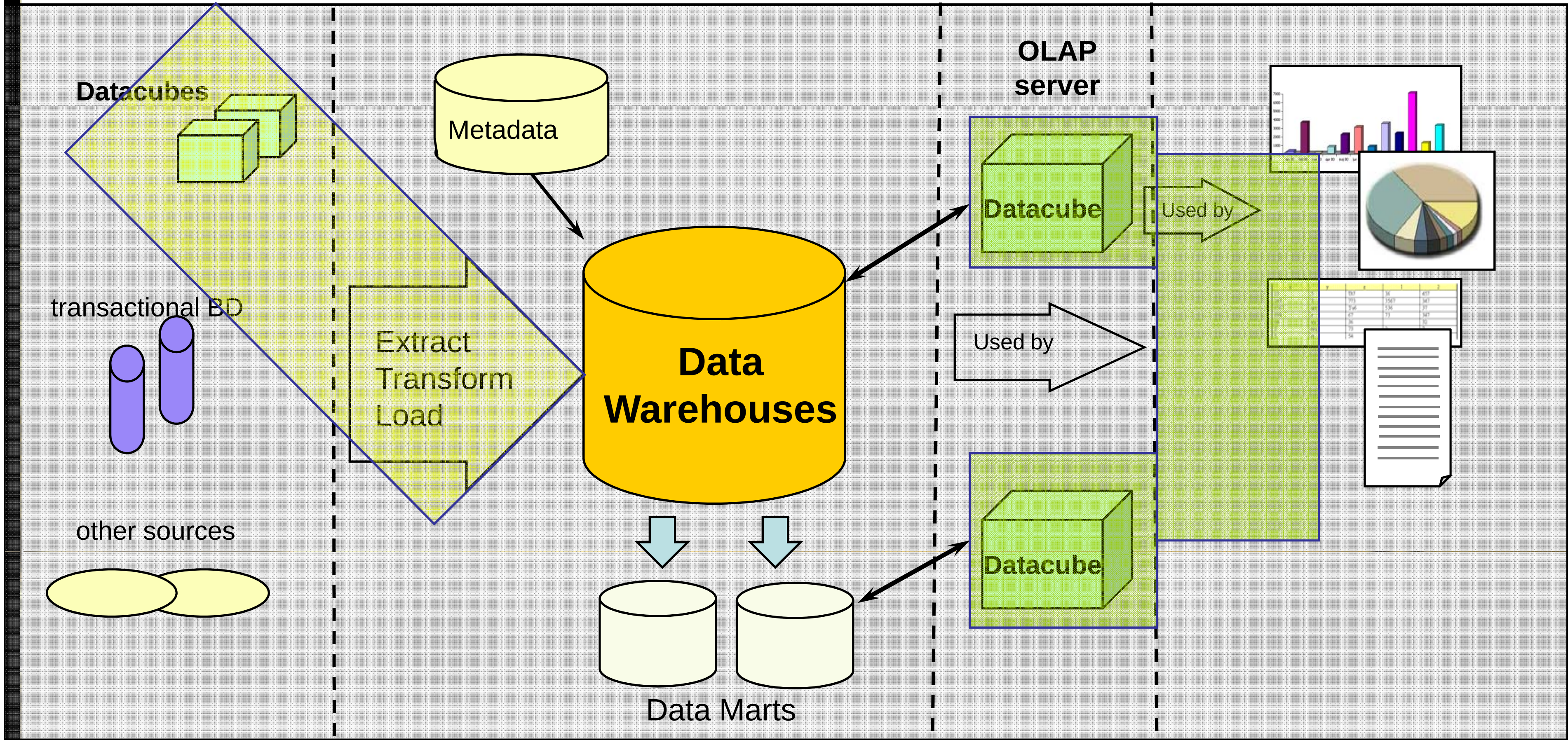
Why not data sources?

- We possibly no longer have access to data source systems from which we created the datacubes due to multiple reasons.
- We need to use data from a long period (i.e., historic data) that usually exist only in datacubes.
- In the context of decision-making, interoperating geospatial datacubes is potentially more efficient than interoperating source systems.



Generic context Ontology
Ontology 1
Ontology 2
Context agent
Cube layer
Ci1
Ci2
Measure layer
Mi1
Min
Mj1
Mjn
Dimension layer
Di1
Dik
Din
Dj1
Djk
Djn
Hierarchy layer
Hi1
Hik
Hin
Hj1
Hjk
Hjn
Level layer
Ni1
Nik
Nin
Nj1
Njk
Njn

Introduction: needs to satisfy



Data Sources | DBMS | OLAP | Clients (Front-End Tools)

Proposed Approach: A framework for the interoperability between geospatial datacubes