Fine-tuning the Semantic Interoperability between Geospatial Datacubes

Tarek Sboui
Yvan Bédard

Presented by: Tarek Sboui
SeCoGIS 2010, Vancouver, Canada
### 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»**

***Dimension Road***

<table>
<thead>
<tr>
<th>Level Province</th>
<th>Level City</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quebec</td>
<td>Montreal</td>
<td>LB t</td>
</tr>
<tr>
<td>Victoria</td>
<td>Vancouver</td>
<td>2003</td>
</tr>
<tr>
<td>BC</td>
<td></td>
<td>2002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2001</td>
</tr>
</tbody>
</table>

**Measure value**

***Dimension Lake***

- Lac Beausport
- Lac St-Charles

**Dimension Time**

- 2003
- 2002
- 2001

**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
• 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**

- The interoperability between two geospatial datacubes $C_1$ and $C_2$ is the ability of $C_1$ 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.
Interoperability between geospatial datacubes

National election

Age category

fact 2

Region

Fact 1

Voter-total

Time

Local election

Age category

Fact 2

Region

Voter-total

Electoral quota

Electoral location
Interoperability between geospatial datacubes

Datacube Intersection
- Age category
- Fact A
  - Intersection point
- Region
  - Dimension A

Datacube Accident
- Age category
- Fact B
  - Zone accident
- Region
  - Dimension B

Federated Fact C
- Age category
- Region
- Zone & intersection
Interoperability between geospatial datacubes

**Manpower**

- Age category
- Workforce

**Time**

**Fact A**

**Region**

**Age category**

**Human resources**

**Time**

**Fact B**

Number of employees

**Region**
Interoperability between geospatial datacubes
<table>
<thead>
<tr>
<th>Interoperability of transactional DBs</th>
<th>Interoperability of datacubes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Similarities</strong></td>
<td></td>
</tr>
<tr>
<td>Reusing data</td>
<td></td>
</tr>
<tr>
<td>Facilitates an efficient exchange of information</td>
<td></td>
</tr>
<tr>
<td><strong>Differences</strong></td>
<td></td>
</tr>
<tr>
<td>Deals with the heterogeneities of DB concepts (i.e. tables, attributes, relations, etc.)</td>
<td>Deals with datacubes concepts (facts, measures, dimensions, levels)</td>
</tr>
<tr>
<td>Deals with the semantic heterogeneities of aggregation and summarizing methods and algorithms, including summarizability conditions.</td>
<td></td>
</tr>
</tbody>
</table>
Proposed Approach: A framework for the interoperability between geospatial datacubes

Cube layer

Measure layer

Dimension layer

Level layer

Ontology 1

Ontology 2

Context agent
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.

\[
G_{SP} (K, L) = \left\{ \partial C_K \cap \partial C_L, \partial C_K \cap C_L^\circ, C_K^\circ \cap \partial C_L, C_K^\circ \cap C_L^\circ \right\}
\]
MGsP: Multidimensional Geosemantic Proximity

Gives the possibility to dig into and resolve semantic heterogeneity related to key notions of the multidimensional paradigm.
MGsP: Extending the Geosemantic Proximity

Measure:

\[
\text{measure\_function}_{M_1} \left( \text{hyper\_cell}_{M_1} \cap \text{hyper\_cell}_{M_2} \right) \cap \text{measure\_function}_{M_1} \left( \text{hyper\_cell}_{M_2} \right)
\]
MGsP: Extending the Geosemantic Proximity

Dimension:

\[ \text{hyper}_1 \cap \text{hyper}_2 \cap \text{aggregation} \]

\[ \text{hyper}_1 \cap \text{hyper}_2 \cap \text{aggregation} \]

\[ \text{hyper}_1 \cap \text{aggregation} \]

\[ \text{hyper}_2 \cap \text{aggregation} \]
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 data cubes agents defined according to different layers.

- We proposed an extension to the geosemantic proximity (MGsP).
• A prototype has been developed to experiment the proposed approach.

• Experimentations have been conducted using different geospatial datacubes:
  – 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.
Ontology 1

Generic context
Ontology

Ontology 2

Context agent

Cube layer

Measure layer

Dimension layer

Hierarchy layer

Level layer

C

Ci1 Ci2

Ci

y

er

M

n

Mj1 Mjn

D

in

Mj

Dj1 Djk Djn

H

in

Hj1 Hjk Hjn

Ni

Nik Nin

Nj1 Njk Njn
Introduction: needs to satisfy

Data Warehouses

Data Sources

DBMS

OLAP

Clients (Front-End Tools)

Metadata

Datacube

Extract Transform Load

transactional BD

other sources

Data Warehouse

Datacube

Used by

OLAP server

Used by

Datacube

Used by

Clients (Front-End Tools)

Data Sources

DBMS

OLAP

Clients (Front-End Tools)
Proposed Approach: A framework for the interoperability between geospatial datacubes

Context Agent

Agent A (Datacube A)  
Datacube A ontology
Context ontology A
Datacube A
Context A

Agent B (Datacube B)  
Context ontology B
Datacube B ontology
Datacube B
Context B