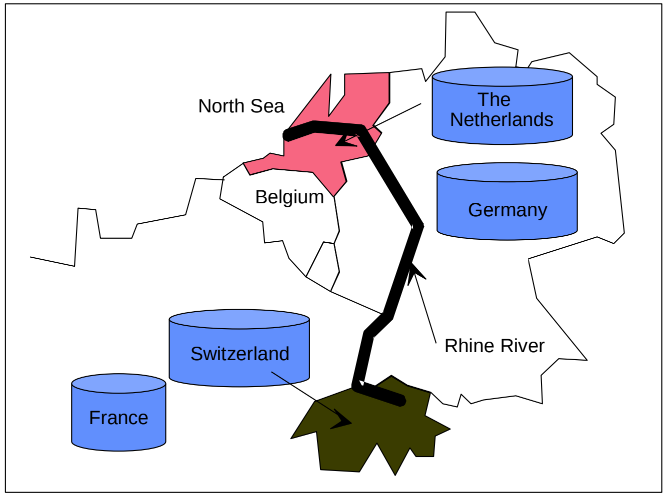
FEDERATING GEOGRAPHIC DATABASES: A STEP TOWARDS INTEROPERABILITY

Robert LAURINI
INSA of Lyon



When is a GIS Federation Worthwhile ?

- Natural or technological risks
- Street repairs
- Environmental monitoring and studies
- International transportation
- Huge public works
- Marine cartography (navigation)
- etc.

I - Introduction

II - Spatial Schema Integration

III - Multidatabase Spatial Querying and Indexing

IV - GIS Interoperability

V- Conclusion

"within a few years, isolated GIS will be seen as dinosaurs"

GIS Interoperability

A Dream for Users
A Nightmare for System Developers

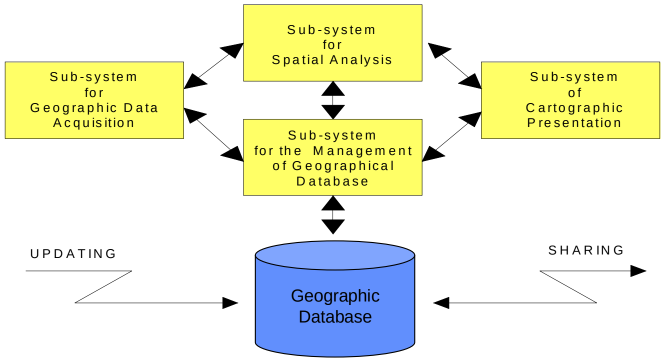
Open GIS Consortium

- OGIS "*Open Geodata Interoperability Specifications*"
- Cross platform compatibility
- Necessity of standardization
- http://www.opengis.org

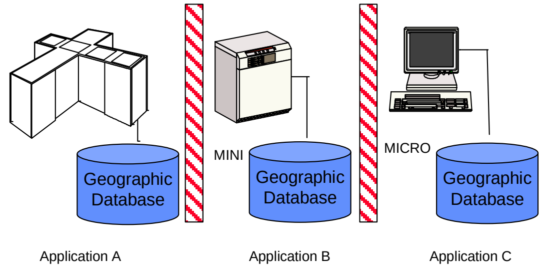
I - Introduction

- 1.1. Why Distributed Geographic Databases ?
- 1.2. Definitions / Classifications
- 1.3. Specificities of Geographic Databases
- 1.4. Metadata
- 1.5. Fragmentation

Structure of a GIS



Separation of Geographic Databases



1.1. Why Distributed Geographic Databases ? Advantages, Drawbacks

- cost of data re-acquisition, updating
- remote sites
- coupling different GIS belonging to several services/institutions
  - same GIS product (ex Arc-Info)
  - different GIS products (ex Arc-Info, SmallWorld, SICAD, etc.)

- splitting of information
  - security
  - responsibility
    - different layers (each user having his own layer)
    - same layers, but different spatial coverages
  - easiness of pereniality (each user is responsible of updating his own information)

1.2. Distributed Databases and alike, Definitions

- Remote Databases
- Distributed Databases
- Federated Databases
- Distributed DBMS
- Homogeneous Distributed Databases
- Heterogeneous Distributed Databases
- Data Dictionary

- Schema
- Local Query
- Distributed Query
- Horizontal Fragmentation
- Vertical Fragmentation
- Mixed Fragmentation
- Interoperability
- Client-server

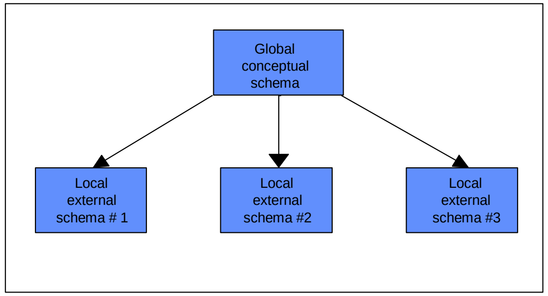
Schemata

- local schema
- global schema
- export schema
- import schema
- external schema

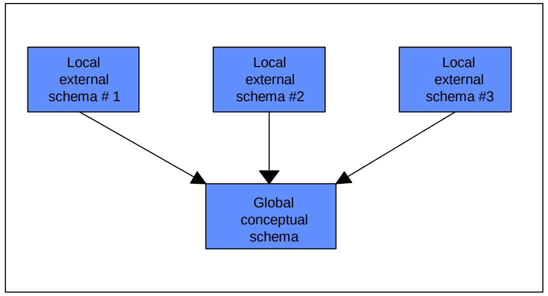
Views

- transparent to the user
- located on one site
- located on several sites

Decomposition of Schemata according to Several Local Sites: Distribution



Schema Integration of several already Existing Databases: Federation



1.3. Specificities of Geographic Databases

1.3.1 Homogeneous GIS

- Same GIS on all sites and same spatial representations
  - vertical fragmentation
  - horizontal fragmentation

||
||

===> **> zonal fragmentation**
> layer fragmentation

1.3.2. Heterogeneous GIS

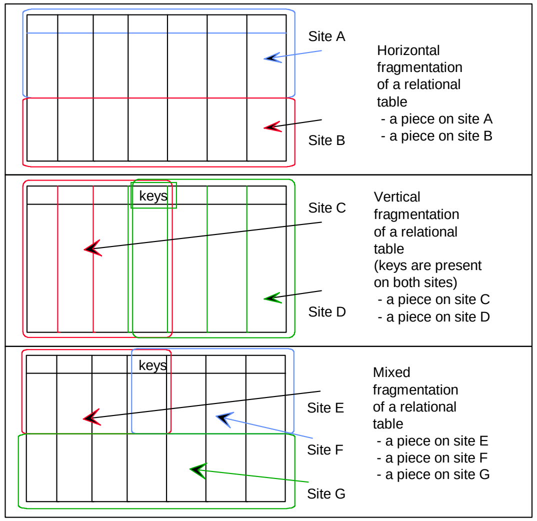
- in addition to the previous ones:
  - Scales and accuracy
  - Multiplicity of spatial representations
    + standardization; conversion
  - > Spatial queries
    + distributed spatial analysis

1.4. Metadata

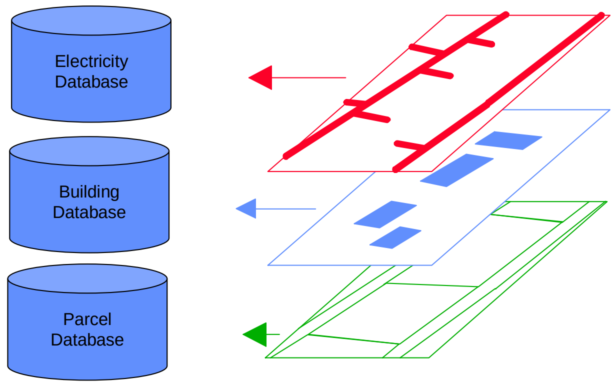
- Data about data : global / local
- Spatial metadata contents
  - data name/identifier - definition
  - creator of the data
  - spatial representation - spatial coverage
  - year validity
  - sites - access rights
  - updating procedure
  - duplicates (location)
  - authorized users for modification

1.5. Fragmentations

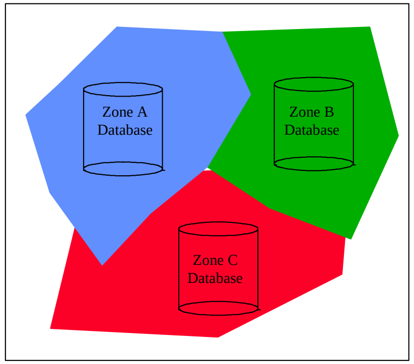
- Relational domains :
  - Horizontal, Vertical, Mixt
- Spatial domains :
  - Zonal, Thematic, Heterogeneous



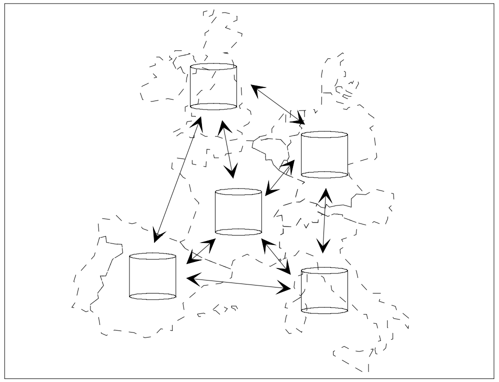
Layer Fragmentation Thematic Partitioning



Zonal Fragmentation
Geographic Partitioning



Example of Zonal Fragmentation in a Client-server Architecture



II - Spatial Schema Integration

2.1. Generalities about Schema Integration

2.2. Semantic and Topological Discrepancies

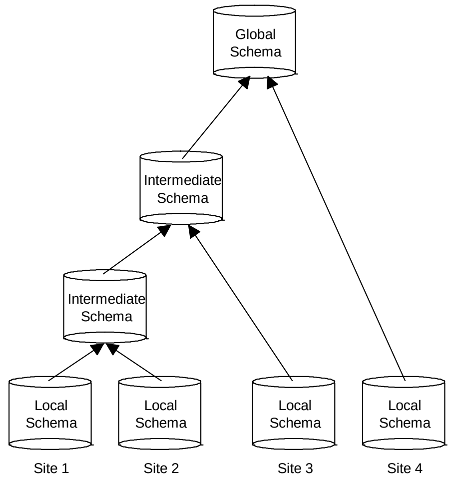
2.3. Implications for zonal and layer fragmentations

2.1. Integration of schemata

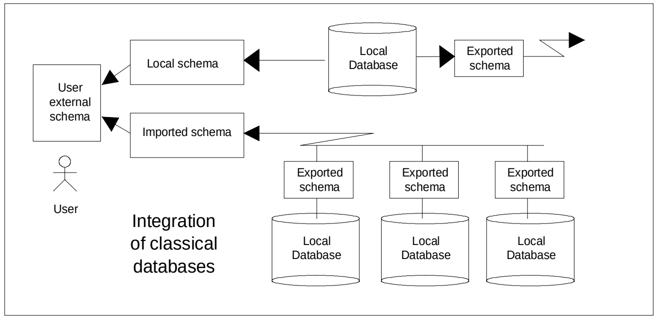
Scope : starting from different local schemata, obtain the global schema

Idea : intermediate schemata

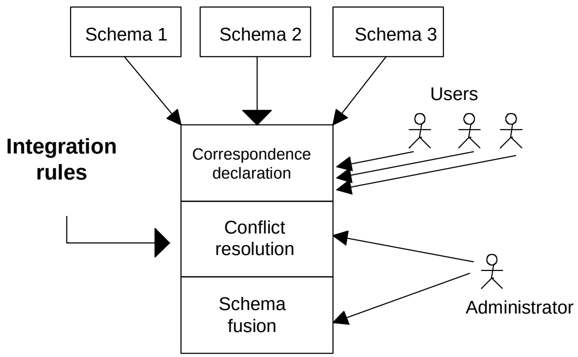
Approach : Pure binary integration
Binary ladder integration
N-ary integration



Conventional Schema Integration



Schema Integration Procedure (Parent and Spaccapietra)



Transparent Access to Distant Relations

```
SELECT * FROM
Peter.L_BLOCK@gis1.paris.fr
UNION
SELECT * FROM
Jim.L_BLOCK@gis2.athens.gr
```

On the London Site

```
CREATE VIEW BLOCK (...) AS
SELECT * FROM Peter.L_BLOCK@gis1.paris.fr
UNION
SELECT * FROM
Jim.L_BLOCK@gis2.athens.gr
```

2.2. Semantic and Topological Discrepancies

- Semantic discrepancies in distributed databases
  - type discrepancies
  - definition discrepancies
  - format discrepancies
  - unit discrepancies
  - data acquisition period discrepancies
  - encoding discrepancies
  - interpretation of "nulls" meaning varying over sites
  - existence of synonyms and polysems
  - data capture errors
  - replicate updating discrepancies

- Semantic discrepancies in distributed geographic databases
  - diversity of geometric representation
  - diversity of coordinate values (scales)
  - presence generalization/detailization
  - diversity of spatio-temporal samplings
  - variability of definition over time and space

Gas Company Database (G-site)

G-STREET	*(#street, street_name, (#axis_segment, width)*)*
G-SEGMENT	*(#segment, #point1, #point2)*
G-POINT	*(#point, x, y)*
G-PIPE	*(#edge, #node1, #node2)*
G-NODE	*(#node, x, y, z, type)*

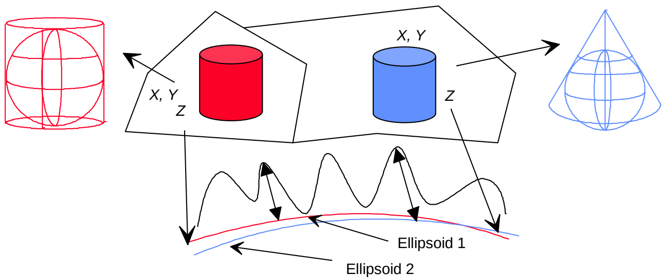
Water Company Database (W-site)

W-STREET	*(#street, (#right_segment, order)*, (#left_segment, order)*)*
W-SEGMENT	*(#segment, #from_point, #to_point)*
W-POINT	*(#point, x, y)*
W-PIPE	*(#edge, #from_node, #to_node)*
W-NODE	*(#node, x, y, z, (#edge)*, category)*

Street Repair Company Database (SR-site)

SR-STREET	*(#street, street_name, (#parcel_segment)*,(kerb_segment)*)*
SR-SEGMENT	*(#segment, #point1, #point2, begin_address, end_address)*
SR--POINT	*(#point, x, y)*
SR-G-PIPE	*(#edge, #node1, #node2)*
SR-G-NODE	*(#node, x, y, depth, type)*
SR-W-PIPE	*(#edge, #node1, #node2)*
SR-W-NODE	*(#node, x, y, depth, type)*

Integration of Coordinates



2.3. Implications for Zonal and Layer Fragmentations

- Problems
  - measurement errors
  - boundaries do not coincide
  - coordinates not aligned
- Constraints
  - no modifications on any database
  - solving when querying (updating problems)

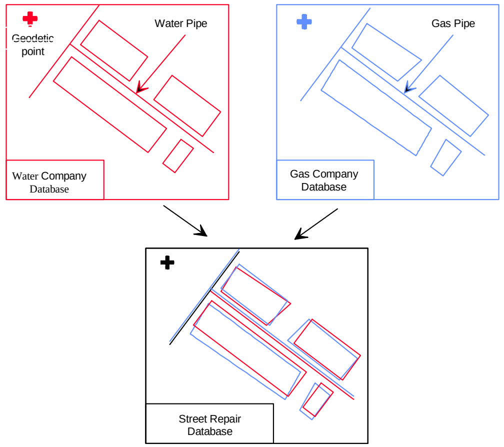
Elastic Transformations (Rubber-sheeting)

- based on homologous points
- force-fitting of all points in the plan
- possibly with constraints

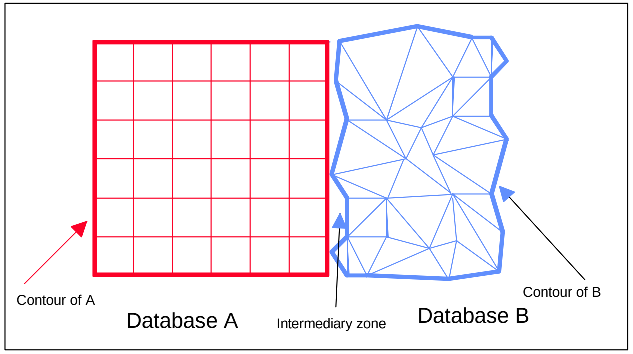
Example of a simple rubber-sheeting function

- 4 control points to force-fit; no constraints
- *x*, *y*: old coordinates; *X*, *Y*: new coordinates
  - $X=axy+bx+cy+d$
  - $Y=a'xy+b'x+c'y+d'$
- 8 unknowns
- 8 equations in total (4 for X's, 4 for Y's)
- 8×8 matrix to invert to get the parameters

Example of Geometric Discrepancies in Layer Fragmentation



Example of Terrain Integration



Example for Terrain Integration

Database A :	***Database B :***
- Grid file representation - UTM co-ordinates - Type A ellipsoid - Convention for sea-level (z=0) in Jackson Harbour - Relations: *A_Terrain (#terrain, # mesh)* *A_Mesh (#mesh, #nw_pt, #ne_pt, #sw_pt, #se_pt)* *A_Point (#point, x, y, z)*	- TIN's - Gauss co-ordinates - Type B ellipsoid - Convention for sea-level (z=0) in Johnson Harbour - Relations: *B_Terrain (#terrain,#triangle)* *B_Triangle (#triangle, #pt1, #pt2, #point3)* *B_Point (#point, x, y,z)*

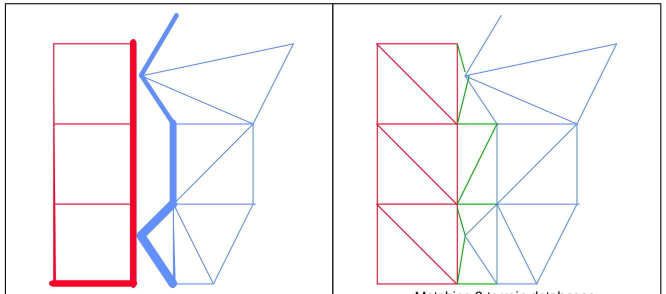
Federation AB:

- Triangulated irregular networks
- Gauss co-ordinates
- Type B ellipsoid
- Convention for sea-level (z=0) in Johnson Harbour

- Relations

AB_Terrain (#terrain, #triangle)
AB_Triangle (#triangle, #point1, #point2, #point3)
AB_Point (#point, x, y,z)

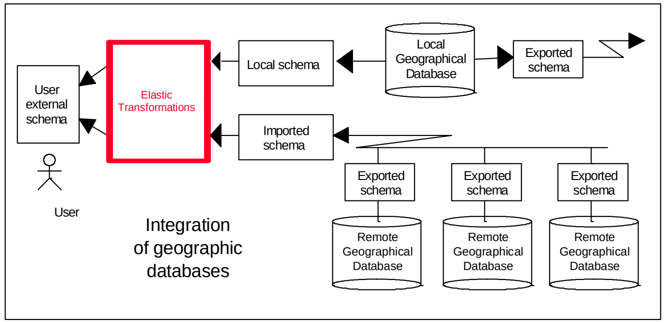
Database Terrain Matching
Terrain Continuity



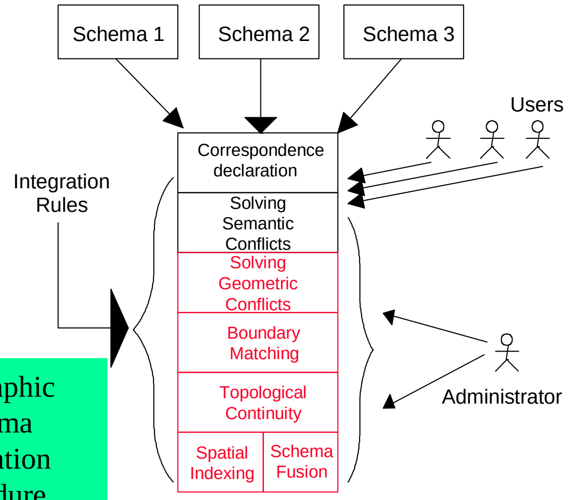
Excerp of 2 terrain databases which are to be federated and matched

Matching 2 terrain databases by transforming squares into triangles and adding some intermediary triangles

Geographic Schema Integration



Geographic Schema Integration Procedure



III - Multidatabase Spatial Querying and Indexing

3.1. Boundary Matching

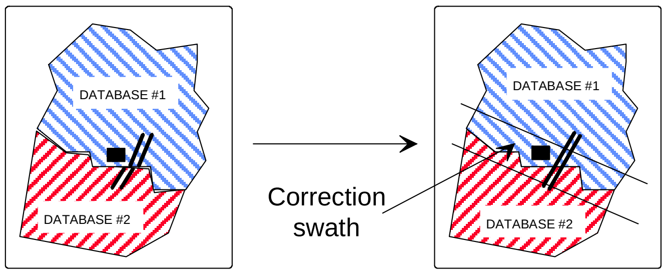
3.2. Interdatabase querying

3.3. Multidatabase Spatial Indexing

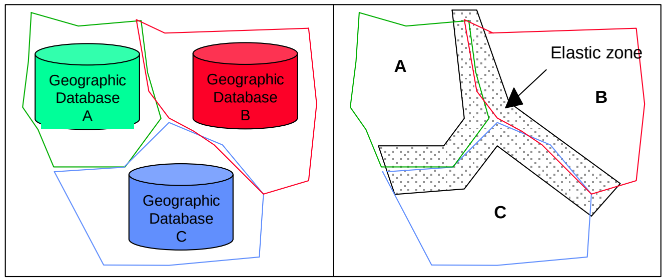
3.4. Using Indexing for Some Spatial Queries

3.5 Conclusions about Multidatabase querying and indexing

3.1 Boundary Discrepancy and Matching



Matching the Boundaries Elastic Swaths



Correction Swaths

- Swaths at the borderline of each database
- Outer swaths are better
- Necessity of homologous points
- Necessity of control points

Practical Consequences

- (i) - the points located in zone A have no modification
- (ii) - the points located in zone B, but outside the elastic swath have no modification
- (ii) - only points located in zone B and within the elastic swath will be modified (outer swath).

- **Advantages**
  - maps look good
  - no content modification
  - no replication
  - updates are possible (even for homologous pairs)
  - distribution transparency
  - delimitation of the outer swath via a visual interface
- **Limitations**
  - automatic finding of homologous points
    - (visual interfaces)
  - impossibility of handling some constraints

Procedures to be Performed

- When entering the federation
- When running a query

- Scope : Maintaining of updates near the boundary

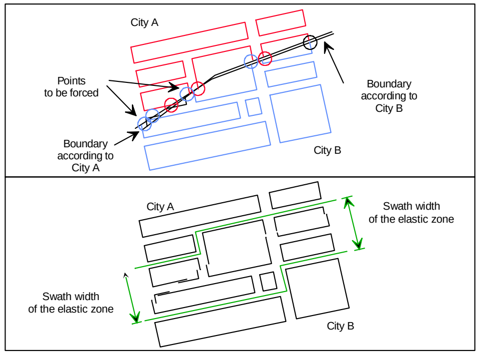
Initialization of the Federation

- Transforming coordinate systems (when necessary)
- Exported and imported schemata
- Schemata mapping (spatial representations)
- Determination of the outer swath
- Homologous points
- Computation of the elastic function

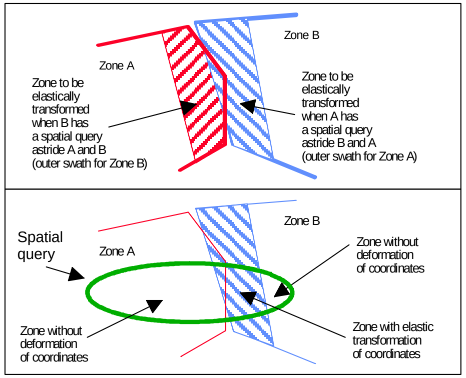
Running of the Federated System (when Querying)

- Running formulae for coordinate transformation
- Running formulae for elastic transformation

Fit-forcing Points at the Boundary



3.2. Multidatabase Querying

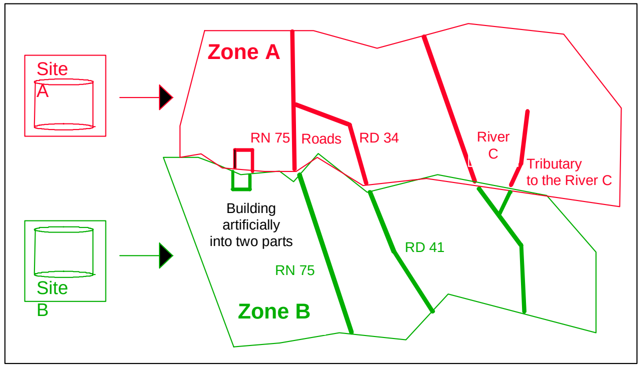


- If same query from A and B

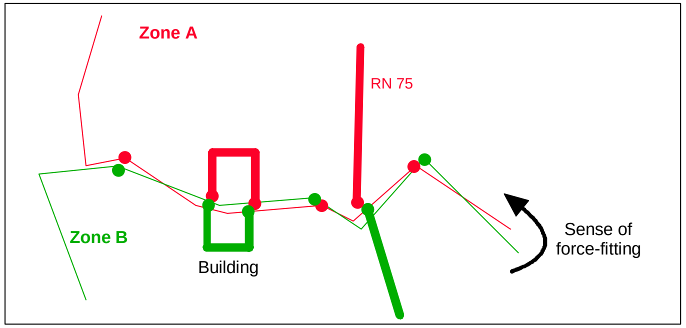
Results not superimposable

- Geometric continuity
- Maps look good

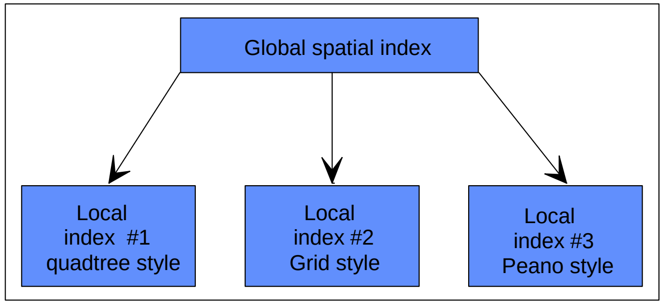
Semantic and Topological Continuity



Fit-forcing



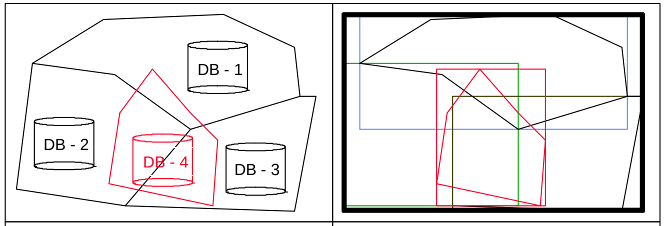
3.3. Multidatabase Spatial Indexing



Multidatabase Spatial Indexing

- Rapidly locate pertinent sites
- Can be based on
  - Peano-keys
  - R-trees
  - etc.

Minimum Bounding Rectangles of Databases R - trees Structure of Indices



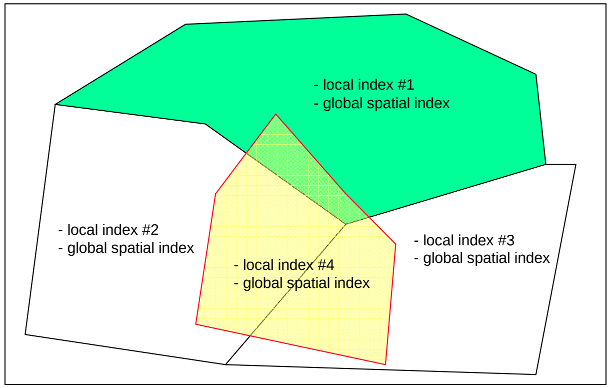
(a) Various databases to federate

(b) Bounding rectangles for each database

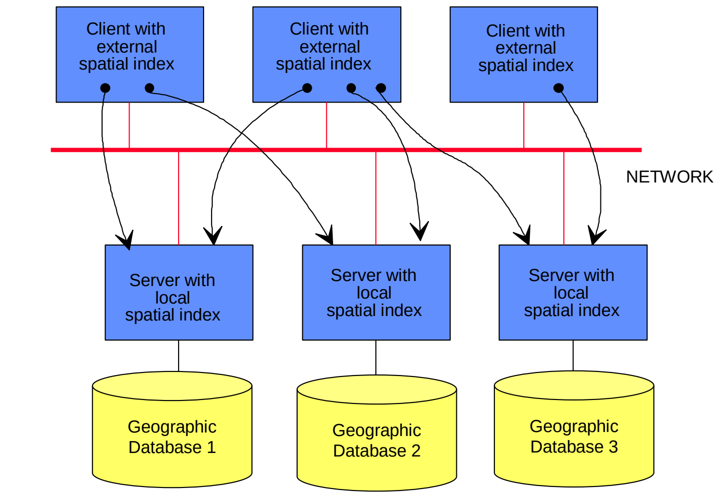
Where to put the Global Index?

- Parallel with the location of data dictionary
- Two possibilities:
- Either a single copy on one privileged site (but violation of Date's rule)
- Or one copy per site

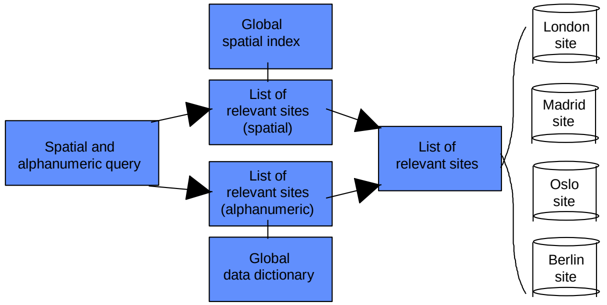
Local Indices and Global Index Location



Indexing in a client-server architecture



Using dictionaries and spatial indices



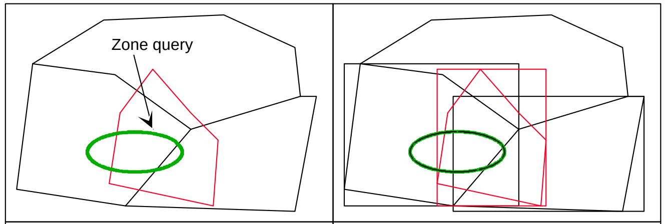
3.4. Using Multidatabase Indexing

- Point and Region query
- Buffer zone query
- Path query

Point Query

- Pertinent sites directly located by the global spatial index
- One single site when zonal fragmentation
- Several sites when layer or heterogeneous fragmentations
- Problems when within correction swaths or holes

Region Query



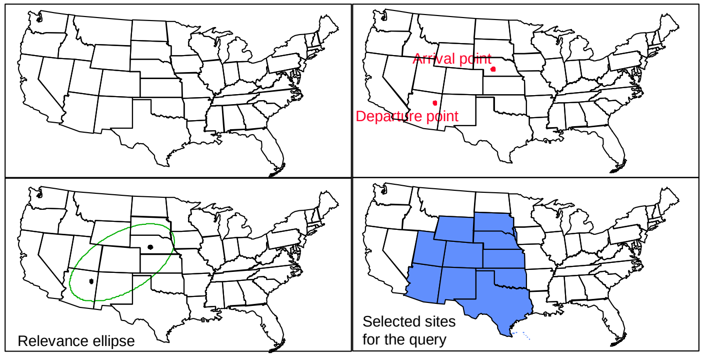
(a) Example of a zone query against several databases

(b) Only three rectangles and so three databases are concerned by the zone query

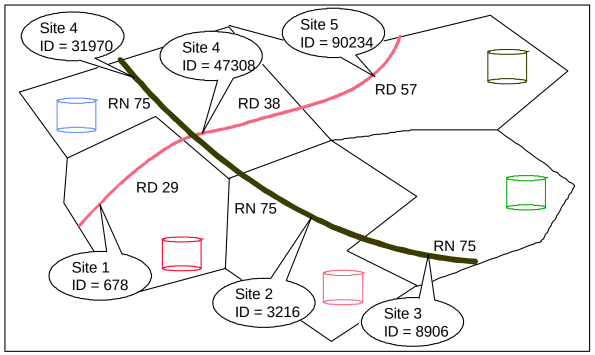
Buffer-zone Query

- Definition of buffer-zone
- Some parts of the buffer-zones are located on another databases
- Site vicinity matrix for each site
- Adjacent sites
- Using spatial index afterwards

Solving a Path Query



Topological Continuity Between Databases



Interdatabase Continuity

- Semantic continuity
- Geometric continuity
- Topological continuity

Ensuring Continuity Between Sites

- Continuity table between sites
- Located in each sites, for its own objects

Continuity-Table (#object, (#site, #site_object))*

- Local and global identifiers

3.4 Conclusions on Multidatabase Spatial Indexing

- Location of global spatial indices
- Necessity of simulation
- Taking boundary errors into account
- Taking continuity into account
- Links with spatial data dictionaries
- Field-Orientation and federation

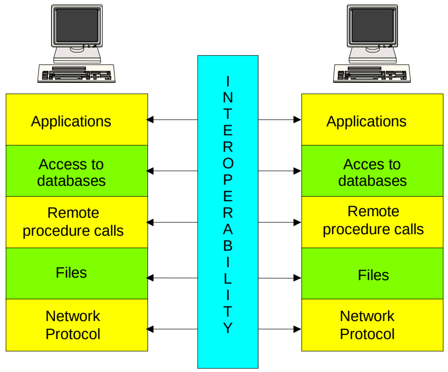
IV - GIS Interoperability

- *Legacy systems*
- variety of softwares and applications
- difficulty of re-writing and re-use of existing programs (re-engineering)
- easy connection between several sites

Interoperability
Cooperation between Softwares

- 4.1 - Introduction
- 4.2 - Semantic cooperation
- 4.3 - Interoperability based on ontologies
- 4.4 - Conclusion about interoperability

Levels of interoperability



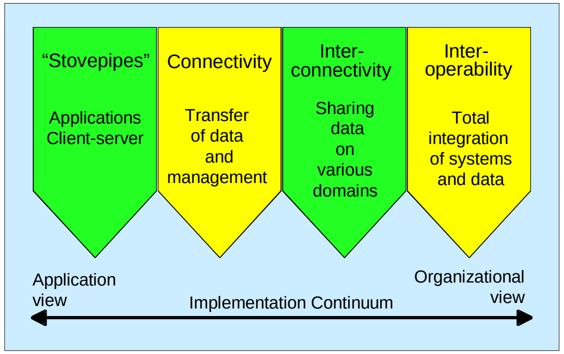
4.1 - Introduction

- Vocabulary
- Fundamentals of interoperability
- Introduction to semantic interoperability
- Metadata

Definition of interoperability

Technical capacity of software applications
for cooperating without conflicts
neither of systems,
nor of contents,
between several organizations.

Continuum of interoperability



Various types of interoperability

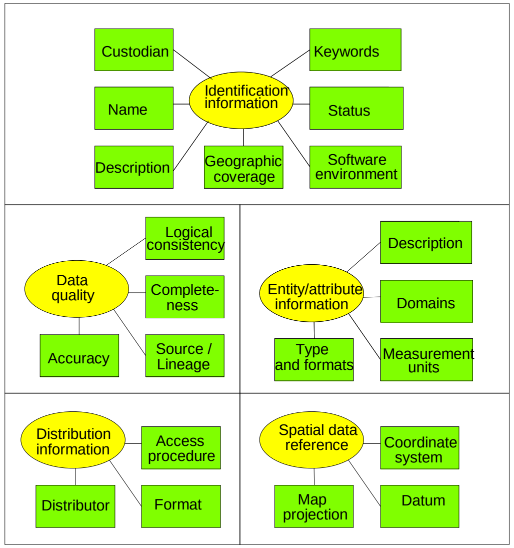
- Interoperability protocols for facilitating the connection between islands of objects differently implemented:
  - syntactic interoperability (high level): CORBA
  - semantic interoperability (more difficult), for which it is necessary to know the domains, and the behaviors of services and objects

4.2 - Semantic cooperation

- Metadata
- Mediators
- Ontologies
- Multi-agent

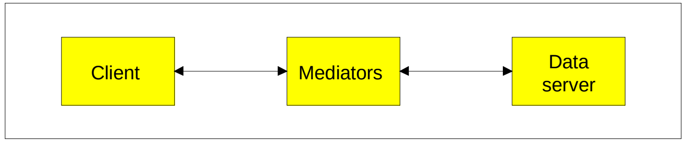
Metadata

- *Data about data*
- = Data dictionary
- *Metadata are information which allow the description of any kind of data: nature, definition, origin, organization, availability, updating, usage, consistency, etc. .*



Overview of the FGDC Metadata Content Standard

Mediators

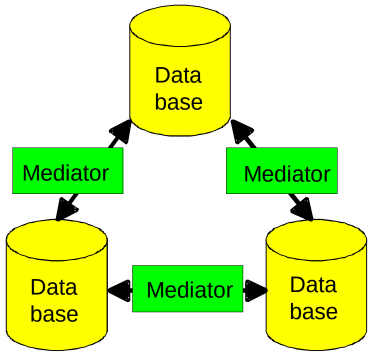


Mediator = a data transformer located on the network

Examples of mediators

- support conversion of supports
- structure conversion
- unit conversion
- attribute encoding
- name translation
- object classification
- semantic clustering (layers)
- etc.

Mediator-based interoperability



Interoperability based on mediators

- **Mediator** = software module which solves the schematic and semantic conflicts
- **Wrapper** = software module which
  - provides the services for data access thanks to a language common to databases and mediators:
  - translates queries,
  - formats the results and
  - transmits them to mediators

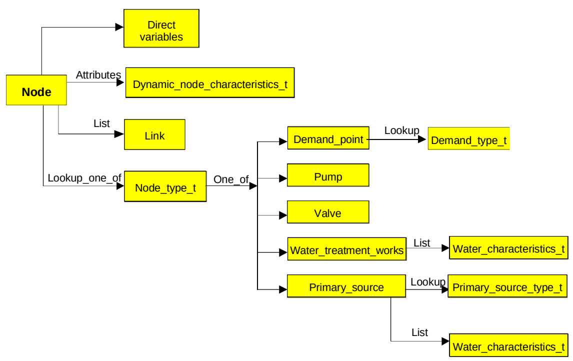
Integration methodology based on mediators

- Principle : small modules distributed along the network
- find data couples which are similar in each database
- write conversion function (generally, one mediator per attribute)
- install those mediators at appropriate locations

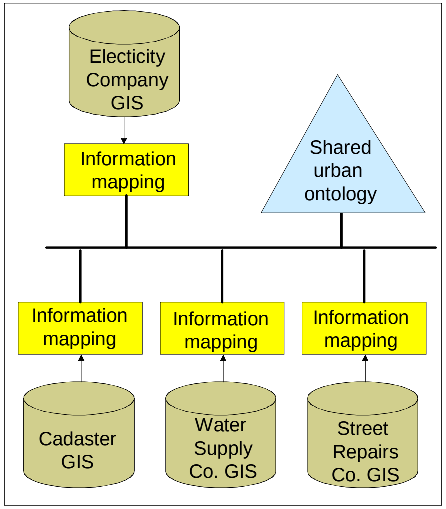
4.3 Ontologies

- Formal vocabulary for describing data and situations
- *Ontological commitment*
- Languages : Ontolingua, KIF, some extensions of XML

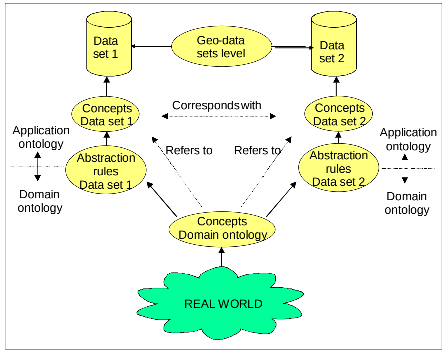
Example of ontology



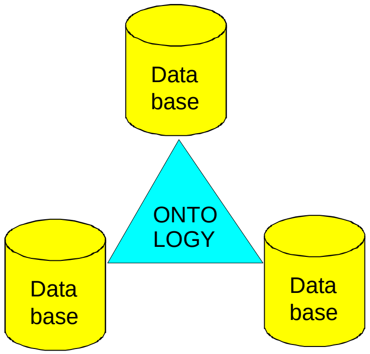
Ontology sharing



Mappings



Ontology-based interoperability



Integration methodology based on ontology

- Principle: the ontology must pre-exist.
- Find mappings between the data base contents and the ontology vocabulary
- Define the transformations
- Dynamic resolution of conflicts

4.4 - Conclusions about GIS interoperability

- Very important for all applications
- Increasing importance of the ontology approach
- Sometimes difficulties in writing the mappings between attributes
- Necessity of defining a complete ontology of geographic features

V - General Conclusions

- "*within a few years, isolated GIS will be seen as dinosaurs*"
- Cost of data acquisition
- Real-time data sharing
- Standards, SQL etc..
- Integration of yet-existing heterogeneous databases
- Handling measurement errors
- Inter-institutional agreements
- Multidatabase spatial indexing
- GIS interoperability

"The GIS's of the future will be federated and interoperable"

Thanks for your attention!

"***Information Systems for Urban Planning: A Hypermedia Co-operative Approach***"

http://lisi.insa-lyon.fr/~laurini