GIS: concepts, methods & tools

Introduction to GIS
Structuring of geographic information
INFORMATION SYSTEM AND ORGANIZATION

Problem → CONTROL SYSTEM → Decision

Requests → INFORMATION SYSTEM → Information

INFORMATION SYSTEM

<table>
<thead>
<tr>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION BASE</td>
<td>INFORMATION PROCESSOR</td>
</tr>
<tr>
<td>Data, rules, and constraints of the external universe</td>
<td>Actions Personnel</td>
</tr>
</tbody>
</table>

EXTERNAL UNIVERSE

Inputs

Outputs
AUTOMATIZATION OF THE INFORMATION SYSTEM

AUTOMATIZED INFORMATION SYSTEM

<table>
<thead>
<tr>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION BASE</td>
<td>INFORMATION PROCESSOR</td>
</tr>
<tr>
<td>Data model</td>
<td>Computer Software: DBMS</td>
</tr>
<tr>
<td>(Database management system)</td>
<td>Personnel</td>
</tr>
</tbody>
</table>

Inputs

FORMALISABLE UNIVERSE

Outputs
GIS: WHAT EXACTLY IS IT?

The term GIS represents, in fact, 3 different concepts:

- An information system about a territory or project
- The databases describing this IS
- The IT solutions used (in particular: software)

One has to return to the basic concept!

Importance of the spatial component of the IS (localization of objects and processes, spatial interactions between elements)
REPRESENTING GEOGRAPHICAL DATA

Duality of Geographical Information:

- **Graphical information:** The geographical objects, their localization, their topological relationships
- **Thematic information:** The descriptors of these objects, of these localizations

‘*semantic*’ data
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

REPRESENTING GRAPHIC DATA

Semantic database

Spatial representation

Entities

Relationships

Tables

Entities

Relationship

Link

Point

Line

Surface

Continuous

Adjacency

Inclusion

Proximity

Path
CODING GEOMETRY: TWO MODES

- **image** vs ‘line drawings’
- **raster** vs **vector**

![Image showing coding geometry: two modes](image)
VECTOR REPRESENTATION: ‘SPAGHETTI’

Point: Id, x, y

Line: Id,
- x₀, y₀
- x₁, y₁
- x₂, y₂
- ... xₙ, yₙ

Surface: Id,
- x₀, y₀
- x₁, y₁
- x₂, y₂
- ... xₙ, yₙ
- (x₀, y₀)

Data redundancy
Undefined relationships
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

VECTOR REPRESENTATION: NETWORK TOPOLOGY

Point: Id, x, y +

Line: Id, (L)
  Na
  x1, y1
  x2, y2
  ......
  Nb

Node: Id, x, y

Surface: Id,
  L1
  L2
  ......
  Ln

No redundancy
Management of connectedness
VECTOR REPRESENTATION: SURFACE TOPOLOGY

Point: Id,x,y

Line: Id, (L), P_g, P_d
   N_a
   x_1, y_1
   x_2, y_2
   ------
   N_b

Node: Id,x,y

Surface: Id,
   L_1
   L_2
   ----
   L_n

Management of connectedness
Management of contiguity
‘SPAGHETTI’ MODEL

• **Entities**
  - point: x, y coordinates
  - line: list of x,y coordinates corresponding to nodes
  - sometimes polygons: set of x,y coordinates, forming a loop: the first coordinate couple = last coordinate couple
    - The spaghetti mode produces a visual **effect** of polygons but, most often, no polygon entity is stored

• **No spatial relationships between objects**
  - unjoined arcs – unclosed polygons
    - polygons cannot form a surface
  - intersections without nodes at the crossing of two arcs
  - Adjacent polygons that overlap or separated by ‘blanks’
TOPOLOGICAL MODEL

• Managing the connectedness and contiguity of objects

• 2 types of topology:
  – Planar topology
  – Network topology

• Topological entities:
  – With permanent storage of the topology
    • e.g.: ArcInfo coverage
  – Taking into account only when editing (creating, updating); this data is called pseudo-topological.
    • e.g.: shp (ArcView) or tab (MapInfo)
  – With defined topological rules, varying depending on context
    • e.g.: feature class within feature dataset in a geodatabase (ArcGIS)
DBMS LINK DESCRIPTIVE DATA

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

1. Weak
2. Strong
3. Hybrid/Integrated
4. LINK
5. STRUCTURING
6. Layer -- Table
RASTER PRESENTATION

- N (orientation)
- dx (Size of the cell)
- dy (geometrical resolution)
- Origin (X,Y)
- extent in X or columns
- extent in Y or rows
- Identifier
- id | attribute
- 1  Wheat
- 2  Grapevine
- 3  Forest
RASTER REPRESENTATION: IMAGE DEPTH

✓ Number of planes

✓ Depth of each plane
  boolean = 1 bit \( (2^1) \)  
  byte = 8 bits \( (2^8) \)  
  integer = 16 bits \( (2^{16}) \)  

...  

2 values: \{0,1\}  
256 values: \{0,255\}  
65536 values: \{-32767,32765\}

Necessity of digital coding!
RASTER REPRESENTATION: LINES AND POINTS

The value chosen for the empty cell is selected by convention.
### RASTER REPRESENTATION: CONTINUOUS DATA

```
11.4  11.5  11.6  11.8  11.9  12.2  12.5  12.4  12.2  12.0
11.5  11.6  11.7  11.9  12.4  12.6  12.6  12.4  12.1  11.9
11.8  12.1  12.4  12.5  12.6  13.0  13.1  12.6  12.1  11.9
12.0  12.2  12.5  12.7  12.8  13.0  13.2  12.7  12.4  12.0
12.5  12.5  12.8  13.1  13.2  13.7  13.9  13.6  13.2  12.6
```

**Direct representation of the variable**

**Associated data: statistical summary**
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

DBMS LINK: DESCRIPTIVE DATA

- **Identifier**
- **Attributes of the point**
  - 1
  - 2
  - 3

- **Identifier**
- **Attributes of the line**
  - 1
  - 2
  - 3

- **Identifier**
- **Attributes of the surface**
  - 1
  - 2
  - 3

- LINK
- Hybrid/integrated
- STRUCTURING
- Layer -- Table
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

RASTER REPRESENTATION: IMAGE DATA

Map (scanned)

Orthophotography

Digital Elevation Model
VECTOR/RASTER CONVERSIONS

FROM VECTOR TO RASTER: RASTERIZATION

FROM RASTER TO VECTOR: VECTORIZATION
VECTOR AND RASTER: COMPLEMENTARITY

• **Raster:**
  – Grid of ‘independent’ cells
  – Matrix calculations
  – Redundancy – volume of information

• **Vector:**
  – Objects +/- structured
  – Information with low or no redundancy (depending on the structuring)
  – Link with external databases

• **Complementarity raster-vector**

• **Improvements in hardware architectures**
  – Processor power / disk space
  – Developments in raster solutions / image processing
SCALE AND ACCURACY - 1

• **Definition:**
  Constant ratio between lengths measured on the map and the corresponding lengths measured on the land.

• **Expression:**
  – Algebraic expression = scale ratio
    map at 1:25,000
  – Graphical expression by a scale bar representing the ratio

  Graphical expression of scale: INDISPENSABLE

• **Semantic confusion**
SCALE AND ACCURACY - 2

• Significance of the scale:
  – Representation ratio
  – Level of analysis of the studied phenomena
  – Geometric accuracy of the information

• Level of approach of geographic space
  – Size of a zone representable at a given scale:
    • A4 sheet: 600 km² at 1:100,000
    • ... : 6 km² at 1:10,000
  – Level of analysis:
    Compartment scale, regional scale, national scale

• Geometric accuracy of the information
SCALE AND ACCURACY - 3

- Scale = representation ratio
  Accuracy = quality of geometric information

- 2 dimensions of accuracy:
  - resolution = size of an ‘elementary’ pixel
  - localization accuracy: localization error

- Scale/accuracy confusion: 2 causes
  - assumption of the reader: each point is significant
  - assumption of the producer: no unnecessary quality

- With ‘undesirable’ consequences
  the zoom syndrome
GEOGRAPHIC DATA QUALITY - 1

- Key-element of any GIS project
- 70 to 80 % of total project costs
GEOREGRAPHIC DATA QUALITY - 2

• Acquiring data from:
  – Institutional: public agencies
  – Private providers
  – External contractors
  – Internal digitalization

  GIS software incorporate tools to create layer:
  – features digitalization,
  – calculation on one or many existing plans,
  – Image processing
GEOGRAPHIC DATA QUALITY - 3

• **Which use ?**
  - To use or extract information from this document
  - As based map (illustration or filling)
  - To take measurements
  - To analyze simultaneously several information plans (spatial analysis)

• **Whatever use, consider :**
  - Precision of imported data
  - Level of detail
  - Reference system used (compatibility with local database)
  - Data property (copyright)
METADATA, TO GUARANTEE DATA QUALITY

- Metadata = "data about data"
  - The procedures followed to acquire the data
  - The precision and methods of measurements
  - The age of the data and update
  - The data coding
  - The geographic referencing
  - The geometry
  - The attributes

- Absence of metadata
  - False interpretation
  - Bad usage
  - Erroneous perception of precision
GIS: concepts, methods & tools

Introduction to GIS
Structuring of geographic information

Pierre BAZILE

Territories, Environment, Remote Sensing & Spatial Information Joint Research Unit Cemagref - CIRAD - ENGREF
INFORMATION SYSTEM AND ORGANIZATION

INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

CONTROL SYSTEM

Problem -> Requests -> Information -> Decision

INFORMATION SYSTEM

STATIC

INPUTS

EXTERNAL UNIVERSE

Outputs

STATIC

INFORMATION BASE

Data, rules, and constraints of the external universe

DYNAMIC

INFORMATION PROCESSOR

Actions Personnel
AUTOMATIZATION OF THE INFORMATION SYSTEM

FORMALISABLE UNIVERSE

Inputs

Outputs

AUTOMATIZED INFORMATION SYSTEM

STATIC

INFORMATION BASE

Data model

DYNAMIC

INFORMATION PROCESSOR

Computer
Software: DBMS
(Database management system)

Personnel
SPATIAL REFERENCE INFORMATION SYSTEM

INFORMATION SYSTEM

<table>
<thead>
<tr>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION BASE</td>
<td>INFORMATION PROCESSOR</td>
</tr>
<tr>
<td>Semantic data model</td>
<td>Personnel</td>
</tr>
<tr>
<td></td>
<td>Software: DBMS</td>
</tr>
<tr>
<td></td>
<td>Computer</td>
</tr>
<tr>
<td></td>
<td>Software: GIS</td>
</tr>
<tr>
<td></td>
<td>(Geographic information system)</td>
</tr>
</tbody>
</table>
The term GIS represents, in fact, 3 different concepts:

- An information system about a territory or project
- The databases describing this IS
- The IT solutions used (in particular: software)

One has to return to the basic concept!

Importance of the spatial component of the IS (localization of objects and processes, spatial interactions between elements)
REPRESENTING GEOGRAPHICAL DATA

Duality of Geographical Information:

- Graphical information: The geographical objects, their localization, their topological relationships
- Thematic information: The descriptors of these objects, of these localizations

‘semantic’ data
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

REPRESENTING GRAPHIC DATA

Semantic database

Spatial representation

Entities

Relationship

Link

Entities

Relationships

Tables

Point
Line
Surface
Continuous

Adjacency
Inclusion
Proximity
Path
CODING GEOMETRY : TWO MODES

- image vs ‘line drawings’
- raster vs vector
VECTOR REPRESENTATION
VECTOR REPRESENTATION: ‘SPAGHETTI’

Point: Id,X,Y

Line: Id,

\[ x_0, y_0 \]
\[ x_1, y_1 \]
\[ x_2, y_2 \]
\[ \ldots \]
\[ x_n, y_n \]

Surface: Id,

\[ x_0, y_0 \]
\[ x_1, y_1 \]
\[ x_2, y_2 \]
\[ \ldots \]
\[ x_n, y_n \]
\[ (x_0, y_0) \]

Data redundancy
Undefined relationships
VECTOR REPRESENTATION: NETWORK TOPOLOGY

Point: ld, x, y

Line: ld, (L)
   Na
   x1, y1
   x2, y2
   Nb

Node: ld, x, y

Surface: ld,
   L1
   L2
   ----
   Ln

No redundancy
Management of connectedness
Vector Representation: Surface Topology

Point: Id, x, y

Line: Id, (L), P_g, P_d
    N_a
    x_1, y_1
    x_2, y_2
    N_b

Node: Id, x, y

Surface: Id,
    L_1
    L_2
    L_n

Management of connectedness
Management of contiguity
‘SPAGHETTI’ MODEL

• **Entities**
  – point: x, y coordinates
  – line: list of x,y coordinates corresponding to nodes
  – sometimes polygons: set of x,y coordinates, forming a loop: the first coordinate couple = last coordinate couple
    ➢ The spaghetti mode produces a visual **effect** of polygons but, most often, no polygon entity is stored

• **No spatial relationships between objects**
  – unjoined arcs – unclosed polygons
    ➢ polygons cannot form a surface
  – intersections without nodes at the crossing of two arcs
  – Adjacent polygons that overlap or separated by ‘blanks’
TOPOLOGICAL MODEL

• Managing the connectedness and contiguity of objects

• 2 types of topology:
  – Planar topology
  – Network topology

• Topological entities:
  – With permanent storage of the topology
    • e.g.: ArcInfo coverage
  – Taking into account only when editing (creating, updating); this data is called pseudo-topological.
    • e.g.: shp (ArcView) or tab (MapInfo)
  – With defined topological rules, varying depending on context
    • e.g.: feature class within feature dataset in a geodatabase (ArcGIS)
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

DBMS LINK DESCRIPTIVE DATA

- Identifier | Attributes of the point
- 1
- 2
- 3

- Identifier | Attributes of the line
- 1
- 2
- 3

- Identifier | Attributes of the surface
- 1
- 2
- 3

- LINK
- Weak
- Strong
- Hybrid/Integrated
- STRUCTURING
- Layer -- Table
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

RASTER PRESENTATION

N (orientation)

Size of the cell (geometrical resolution)

dx

dy

extent
in Y or rows

Origin (X,Y)

extent in X or columns

Identifier

id | attribute
---|---
1 | Wheat
2 | Grapevine
3 | Forest
RASTER REPRESENTATION: IMAGE DEPTH

✓ Number of planes

✓ Depth of each plane
  - boolean = 1 bit \( (2^1) \)  → 2 values: \{0,1\}
  - byte = 8 bits \( (2^8) \)  → 256 values: \{0,255\}
  - integer = 16 bits \( (2^{16}) \)  → 65536 values: \{-32767,32765\}

... 

Necessity of digital coding!
The value chosen for the empty cell is selected by convention.
RASTER REPRESENTATION: CONTINUOUS DATA

<table>
<thead>
<tr>
<th>11.4</th>
<th>11.5</th>
<th>11.6</th>
<th>11.8</th>
<th>11.9</th>
<th>12.2</th>
<th>12.5</th>
<th>12.4</th>
<th>12.2</th>
<th>12.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.4</td>
<td>11.6</td>
<td>11.7</td>
<td>11.9</td>
<td>11.9</td>
<td>12.4</td>
<td>12.6</td>
<td>12.6</td>
<td>12.4</td>
<td>12.0</td>
</tr>
<tr>
<td>11.5</td>
<td>11.7</td>
<td>11.8</td>
<td>11.9</td>
<td>12.4</td>
<td>12.6</td>
<td>12.6</td>
<td>12.4</td>
<td>11.9</td>
<td>11.6</td>
</tr>
<tr>
<td>11.7</td>
<td>11.9</td>
<td>12.0</td>
<td>12.5</td>
<td>12.6</td>
<td>13.0</td>
<td>13.1</td>
<td>12.6</td>
<td>12.3</td>
<td>12.1</td>
</tr>
<tr>
<td>11.8</td>
<td>12.1</td>
<td>12.4</td>
<td>12.5</td>
<td>12.8</td>
<td>13.0</td>
<td>13.2</td>
<td>12.7</td>
<td>12.4</td>
<td>12.0</td>
</tr>
<tr>
<td>12.0</td>
<td>12.2</td>
<td>12.5</td>
<td>12.7</td>
<td>12.9</td>
<td>13.4</td>
<td>13.6</td>
<td>13.1</td>
<td>12.7</td>
<td>12.4</td>
</tr>
<tr>
<td>12.5</td>
<td>12.5</td>
<td>12.8</td>
<td>13.1</td>
<td>13.2</td>
<td>13.7</td>
<td>13.9</td>
<td>13.6</td>
<td>13.2</td>
<td>12.6</td>
</tr>
</tbody>
</table>

Direct representation of the variable

Associated data: statistical summary
### INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

#### DBMS LINK: DESCRIPTIVE DATA

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

**LINK**

Hybrid/integrated

**STRUCTURING**

Layer -- Table
RASTER REPRESENTATION: IMAGE DATA

Map (scanned)

Orthophotography

Digital Elevation Model
VECTOR/RASTER CONVERSIONS

FROM VECTOR TO RASTER: RASTERIZATION

FROM RASTER TO VECTOR: VECTORIZATION
VECTOR AND RASTER: COMPLEMENTARITY

- **Raster:**
  - Grid of ‘independent’ cells
  - Matrix calculations
  - Redundancy – volume of information

- **Vector:**
  - Objects +/- structured
  - Information with low or no redundancy (depending on the structuring)
  - Link with external databases

- **Complementarity raster-vector**

- **Improvements in hardware architectures**
  - Processor power / disk space
  - Developments in raster solutions / image processing
SCALE AND ACCURACY - 1

• Definition:
  Constant ratio between lengths measured on the map and the corresponding lengths measured on the land.

• Expression:
  – Algebraic expression = scale ratio map at 1:25,000
  – Graphical expression by a scale bar representing the ratio

Graphical expression of scale: INDISPENSABLE

• Semantic confusion
SCALE AND ACCURACY - 2

• **Significance of the scale:**
  - Representation ratio
  - Level of analysis of the studied phenomena
  - Geometric accuracy of the information

• **Level of approach of geographic space**
  - Size of a zone representable at a given scale:
    - A4 sheet: 600 km² at 1:100,000
    - ... : 6 km² at 1:10,000
  - Level of analysis:
    - Compartment scale, regional scale, national scale

• **Geometric accuracy of the information**
SCALE AND ACCURACY - 3

• Scale = representation ratio
  Accuracy = quality of geometric information

• 2 dimensions of accuracy:
  – resolution = size of an ‘elementary’ pixel
  – localization accuracy: localization error

• Scale/accuracy confusion: 2 causes
  – assumption of the reader: each point is significant
  – assumption of the producer: no unnecessary quality

• With ‘undesirable’ consequences
  the zoom syndrome
GEOGRAPHIC DATA QUALITY - 1

- Key-element of any GIS project
- 70 to 80% of total project costs

Data
Hardwares
Softwares
Methods
Users
GEOGRAPHIC DATA QUALITY - 2

• Acquiring data from:
  – Institutional: public agencies
  – Private providers
  – External contractors
  – Internal digitalization
    GIS software incorporate tools to create layer:
    – features digitalization,
    – calculation on one or many existing plans,
    – Image processing
GEOGRAPHIC DATA QUALITY - 3

• Which use?
  – To use or extract information from this document
  – As based map (illustration or filling)
  – To take measurements
  – To analyze simultaneously several information plans (spatial analysis)

• Whatever use, consider:
  – Precision of imported data
  – Level of detail
  – Reference system used (compatibility with local database)
  – Data property (copyright)
METADATA, TO GUARANTEE DATA QUALITY

• Metadata = "data about data"
  – The procedures followed to acquire the data
  – The precision and methods of measurements
  – The age of the data and update
  – The data coding
  – The geographic referencing
  – The geometry
  – The attributes

• Absence of metadata
  – False interpretation
  – Bad usage
  – Erroneous perception of precision
GIS: concepts, methods & tools

Introduction to GIS
Structuring of geographic information

Pierre BAZILE
INFORMATION SYSTEM AND ORGANIZATION

Problem → CONTROL SYSTEM → Decision

Requests → INFORMATION SYSTEM → Information

INPUTS

EXTERNAL UNIVERSE

STATIC

INFORMATION BASE

Data, rules, and constraints of the external universe

OUTPUTS

EXTERNAL UNIVERSE

STATIC

INFORMATION BASE

Actions Personnel

DYNAMIC

INFORMATION PROCESSOR

Actions Personnel
AUTOMATIZATION OF THE INFORMATION SYSTEM

AUTOMATIZED INFORMATION SYSTEM

**STATIC**
- INFORMATION BASE
  - Data model

**DYNAMIC**
- INFORMATION PROCESSOR
  - Computer
  - Software: DBMS
  - (Database management system)
  - Personnel

FORMALISABLE UNIVERSE

Inputs

Outputs
**SPATIAL REFERENCE INFORMATION SYSTEM**

<table>
<thead>
<tr>
<th>INFORMATION SYSTEM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STATIC</strong></td>
</tr>
<tr>
<td>INFORMATION BASE</td>
</tr>
<tr>
<td>Semantic data model</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Spatial data model</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

- **Inputs**
- **Outputs**
GIS: WHAT EXACTLY IS IT?

The term GIS represents, in fact, 3 different concepts:

- An information system about a territory or project
- The databases describing this IS
- The IT solutions used (in particular: software)

One has to return to the basic concept!

Importance of the spatial component of the IS (localization of objects and processes, spatial interactions between elements)
REPRESENTING GEOGRAPHICAL DATA

Duality of Geographical Information:

- Graphical information: The geographical objects, their localization, their topological relationships
- Thematic information: The descriptors of these objects, of these localizations

‘semantic’ data
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

REPRESENTING GRAPHIC DATA

Semantic database

Spatial representation

Tables

Entities

Relationship

Entities

Relationships

Link

Point

Line

Surface

Continuous

Adjacency

Inclusion

Proximity

Path
CODING GEOMETRY: TWO MODES

- **image vs ‘line drawings’**
- **raster vs vector**
VECTOR REPRESENTATION

INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION
VECTOR REPRESENTATION: ‘SPAGHETTI’

Point: Id, x, y

Line: Id,
   \[x_0, y_0\]
   \[x_1, y_1\]
   \[x_2, y_2\]
   \[\ldots\]
   \[x_n, y_n\]

Surface: Id,
   \[x_0, y_0\]
   \[x_1, y_1\]
   \[x_2, y_2\]
   \[\ldots\]
   \[x_n, y_n\]
   \[(x_0, y_0)\]

Data redundancy
Undefined relationships
VECTOR REPRESENTATION: NETWORK TOPOLOGY

Point: Id, x, y

Line: Id, (L)
   Na
   x1, y1
   x2, y2
   ...
   Nb

Node: Id, x, y

Surface: Id,
   L1
   L2
   ...
   Ln

No redundancy
Management of connectedness
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

VECTOR REPRESENTATION: SURFACE TOPOLOGY

Point: Id, x, y
Line: Id, (L), P_g, P_d
\[ N_a \]
\[ x_1, y_1 \]
\[ x_2, y_2 \]
\[ N_b \]
Node: Id, x, y
Surface: Id,
\[ L_1 \]
\[ L_2 \]
\[ L_n \]
Management of connectedness
Management of contiguity
‘SPAGHETTI’ MODEL

• **Entities**
  - point: x, y coordinates
  - line: list of x, y coordinates corresponding to nodes
  - sometimes polygons: set of x, y coordinates, forming a loop: the first coordinate couple = last coordinate couple
    - The spaghetti model produces a visual **effect** of polygons but, most often, no polygon entity is stored

• **No spatial relationships between objects**
  - unjoined arcs – unclosed polygons
    - polygons cannot form a surface
  - intersections without nodes at the crossing of two arcs
  - Adjacent polygons that overlap or separated by ‘blanks’
TOPOLOGICAL MODEL

• Managing the connectedness and contiguity of objects

• 2 types of topology:
  – Planar topology
  – Network topology

• Topological entities:
  – With permanent storage of the topology
    • e.g.: ArcInfo coverage
  – Taking into account only when editing (creating, updating); this data is called pseudo-topological.
    • e.g.: shp (ArcView) or tab (MapInfo)
  – With defined topological rules, varying depending on context
    • e.g.: feature class within feature dataset in a geodatabase (ArcGIS)
RASTER PRESENTATION

Size of the cell

Extent in Y or rows

Extent in X or columns

Origin (X,Y)

Identifiers

Attribute table:

<table>
<thead>
<tr>
<th>id</th>
<th>attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wheat</td>
</tr>
<tr>
<td>2</td>
<td>Grapevine</td>
</tr>
<tr>
<td>3</td>
<td>Forest</td>
</tr>
</tbody>
</table>
RASTER REPRESENTATION: IMAGE DEPTH

- **Number of planes**
- **Depth of each plane**
  - boolean = 1 bit \((2^1)\)
  - byte = 8 bits \((2^8)\)
  - integer = 16 bits \((2^{16})\)
  - ...  

2 values: \{0,1\}

256 values: \{0,255\}

65536 values: {-32767,32765}

Necessity of digital coding!
The value chosen for the empty cell is selected by convention.
### RASTER REPRESENTATION: CONTINUOUS DATA

<table>
<thead>
<tr>
<th></th>
<th>11.4</th>
<th>11.5</th>
<th>11.6</th>
<th>11.8</th>
<th>11.9</th>
<th>12.2</th>
<th>12.5</th>
<th>12.4</th>
<th>12.2</th>
<th>12.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Direct representation of the variable

Associated data: statistical summary
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

DBMS LINK: DESCRIPTIVE DATA

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

LINK

Hybrid/integrated

STRUCTURING

Layer -- Table
RASTER REPRESENTATION: IMAGE DATA

Map (scanned)

Orthophotography

Digital Elevation Model
VECTOR/RASTER CONVERSIONS

FROM VECTOR TO RASTER: RASTERIZATION

FROM RASTER TO VECTOR: VECTORIZATION
VECTOR AND RASTER: COMPLEMENTARITY

- **Raster:**
  - Grid of ‘independent’ cells
  - Matrix calculations
  - Redundancy – volume of information

- **Vector:**
  - Objects +/- structured
  - Information with low or no redundancy (depending on the structuring)
  - Link with external databases

- **Complementarity raster-vector**

- **Improvements in hardware architectures**
  - Processor power / disk space
  - Developments in raster solutions / image processing
SCALE AND ACCURACY - 1

• Definition:
  Constant ratio between lengths measured on the map and the corresponding lengths measured on the land.

• Expression:
  – Algebraic expression = scale ratio map at 1:25,000
  – Graphical expression by a scale bar representing the ratio

Graphical expression of scale: INDISPENSABLE

• Semantic confusion
SCALE AND ACCURACY - 2

• **Significance of the scale:**
  - Representation ratio
  - Level of analysis of the studied phenomena
  - Geometric accuracy of the information

• **Level of approach of geographic space**
  - Size of a zone representable at a given scale:
    - A4 sheet: 600 km² at 1:100,000
    - ... : 6 km² at 1:10,000
  - Level of analysis:
    Compartment scale, regional scale, national scale

• **Geometric accuracy of the information**
SCALE AND ACCURACY - 3

- Scale = representation ratio
  Accuracy = quality of geometric information

- 2 dimensions of accuracy:
  - resolution = size of an ‘elementary’ pixel
  - localization accuracy: localization error

- Scale/accuracy confusion: 2 causes
  - assumption of the reader: each point is significant
  - assumption of the producer: no unnecessary quality

- With ‘undesirable’ consequences
  *the zoom syndrome*
GEOGRAPHIC DATA QUALITY - 1

- Key-element of any GIS project
- 70 to 80% of total project costs
GEOGRAPHIC DATA QUALITY - 2

• Acquiring data from:
  – Institutional: public agencies
  – Private providers
  – External contractors
  – Internal digitalization
    GIS software incorporate tools to create layer:
      – features digitalization,
      – calculation on one or many existing plans,
      – Image processing
Which use?
- To use or extract information from this document
- As based map (illustration or filling)
- To take measurements
- To analyze simultaneously several information plans (spatial analysis)

Whatever use, consider:
- Precision of imported data
- Level of detail
- Reference system used (compatibility with local database)
- Data property (copyright)
METADATA, TO GUARANTEE DATA QUALITY

• Metadata = "data about data"
  − The procedures followed to acquire the data
  − The precision and methods of measurements
  − The age of the data and update
  − The data coding
  − The geographic referencing
  − The geometry
  − The attributes

• Absence of metadata
  − False interpretation
  − Bad usage
  − Erroneous perception of precision
GIS: concepts, methods & tools

Introduction to GIS
Structuring of geographic information

Pierre BAZILE
# INFORMATION SYSTEM AND ORGANIZATION

## Problem → CONTROL SYSTEM → Decision

### Requests → Information

## INFORMATION SYSTEM

### STATIC

**INFORMATION BASE**
- Data, rules, and constraints of the external universe

### DYNAMIC

**INFORMATION PROCESSOR**
- Actions
- Personnel

## EXTERNAL UNIVERSE

### Inputs

### Outputs
AUTOMATIZATION OF THE INFORMATION SYSTEM

AUTOMATIZED INFORMATION SYSTEM

<table>
<thead>
<tr>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION BASE</td>
<td>INFORMATION PROCESSOR</td>
</tr>
<tr>
<td>Data model</td>
<td>Computer Software: DBMS</td>
</tr>
<tr>
<td></td>
<td>(Database management system)</td>
</tr>
<tr>
<td></td>
<td>Personnel</td>
</tr>
</tbody>
</table>

FORMALISABLE UNIVERSE

Inputs

Outputs
**Spatial Reference Information System**

<table>
<thead>
<tr>
<th>INFORMATION SYSTEM</th>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INFORMATION BASE</strong></td>
<td><a href="#">Semantic data model</a></td>
<td><a href="#">INFORMATION PROCESSOR</a></td>
</tr>
<tr>
<td><a href="#">Inputs</a></td>
<td><a href="#">Personnel</a></td>
<td></td>
</tr>
<tr>
<td><a href="#">Outputs</a></td>
<td><a href="#">Computer</a></td>
<td></td>
</tr>
</tbody>
</table>

- **Static**
  - Semantic data model

- **Dynamic**
  - Information processor
  - Software: DBMS
  - Software: GIS (Geographic information system)
The term GIS represents, in fact, 3 different concepts:

• An information system about a territory or project
• The databases describing this IS
• The IT solutions used (in particular: software)

One has to return to the basic concept!

Importance of the spatial component of the IS (localization of objects and processes, spatial interactions between elements)
REPRESENTING GEOGRAPHICAL DATA

Duality of Geographical Information:

- Graphical information: The geographical objects, their localization, their topological relationships
- Thematic information: The descriptors of these objects, of these localizations
  ‘semantic’ data
CODING GEOMETRY: TWO MODES

- image vs ‘line drawings’
- raster vs vector
VECTOR REPRESENTATION
VECTOR REPRESENTATION: ‘SPAGHETTI’

- **Point**: `Id,x,y`  
- **Line**: `Id, x_0,y_0, x_1,y_1, x_2,y_2, ... , x_n,y_n`  
- **Surface**: `Id, x_0,y_0, x_1,y_1, x_2,y_2, ... , x_n,y_n, (x_0,y_0)`  

Data redundancy  
 Undefined relationships
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

VECTOR REPRESENTATION: NETWORK TOPOLOGY

Point: Id, x, y

Line: Id, (L)
  N_a
  x_1, y_1
  x_2, y_2

Node: Id, x, y

Surface: Id,
  L_1
  L_2
  ----
  L_n

No redundancy
Management of connectedness
VECTOR REPRESENTATION: SURFACE TOPOLOGY

Point: Id, x, y

Line: Id, (L), P_g, P_d
   N_a
   x_1, y_1
   x_2, y_2
   N_b

Node: Id, x, y

Surface: Id,
   L_1
   L_2
   ...
   L_n

Management of connectedness
Management of contiguity
‘SPAGHETTI’ MODEL

• **Entities**
  - point: x, y coordinates
  - line: list of x,y coordinates corresponding to nodes
  - sometimes polygons: set of x,y coordinates, forming a loop: the first coordinate couple = last coordinate couple
    - The spaghetti mode produces a visual effect of polygons but, most often, no polygon entity is stored

• **No spatial relationships between objects**
  - unjoined arcs – unclosed polygons
    - polygons cannot form a surface
  - intersections without nodes at the crossing of two arcs
  - Adjacent polygons that overlap or separated by ‘blanks’
TOPOLOGICAL MODEL

• Managing the connectedness and contiguity of objects

• 2 types of topology:
  – Planar topology
  – Network topology

• Topological entities:
  – With permanent storage of the topology
    • e.g.: ArcInfo coverage
  – Taking into account only when editing (creating, updating); this data is called pseudo-topological.
    • e.g.: shp (ArcView) or tab (MapInfo)
  – With defined topological rules, varying depending on context
    • e.g.: feature class within feature dataset in a geodatabase (ArcGIS)
DBMS LINK DESCRIPTIVE DATA

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

LINK
Weak
Strong
Hybrid/Integrated
STRUCTURING
Layer -- Table
RASTER PRESENTATION
RASTER REPRESENTATION: IMAGE DEPTH

✓ Number of planes

✓ Depth of each plane
  boolean = 1 bit \((2^1)\)  \(\rightarrow\) 2 values: \{0,1\}
  byte = 8 bits \((2^8)\)  \(\rightarrow\) 256 values: \{0,255\}
  integer = 16 bits \((2^{16})\)  \(\rightarrow\) 65536 values: \{-32767,32765\}
  ...

Necessity of digital coding!
RASTER REPRESENTATION: LINES AND POINTS

The value chosen for the empty cell is selected by convention.
### RASTER REPRESENTATION: CONTINUOUS DATA

<table>
<thead>
<tr>
<th></th>
<th>11.4</th>
<th>11.5</th>
<th>11.6</th>
<th>11.7</th>
<th>11.8</th>
<th>11.9</th>
<th>12.0</th>
<th>12.1</th>
<th>12.2</th>
<th>12.3</th>
<th>12.4</th>
<th>12.5</th>
<th>12.6</th>
<th>12.7</th>
<th>12.8</th>
<th>12.9</th>
<th>13.0</th>
<th>13.1</th>
<th>13.2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11.4</td>
<td>11.5</td>
<td>11.6</td>
<td>11.7</td>
<td>11.8</td>
<td>11.9</td>
<td>12.0</td>
<td>12.1</td>
<td>12.2</td>
<td>12.3</td>
<td>12.4</td>
<td>12.5</td>
<td>12.6</td>
<td>12.7</td>
<td>12.8</td>
<td>12.9</td>
<td>13.0</td>
<td>13.1</td>
<td>13.2</td>
</tr>
<tr>
<td>11.5</td>
<td>11.6</td>
<td>11.7</td>
<td>11.8</td>
<td>11.9</td>
<td>12.0</td>
<td>12.1</td>
<td>12.2</td>
<td>12.3</td>
<td>12.4</td>
<td>12.5</td>
<td>12.6</td>
<td>12.7</td>
<td>12.8</td>
<td>12.9</td>
<td>13.0</td>
<td>13.1</td>
<td>13.2</td>
<td>13.3</td>
<td>13.4</td>
</tr>
<tr>
<td>11.6</td>
<td>11.7</td>
<td>11.8</td>
<td>11.9</td>
<td>12.0</td>
<td>12.1</td>
<td>12.2</td>
<td>12.3</td>
<td>12.4</td>
<td>12.5</td>
<td>12.6</td>
<td>12.7</td>
<td>12.8</td>
<td>12.9</td>
<td>13.0</td>
<td>13.1</td>
<td>13.2</td>
<td>13.3</td>
<td>13.4</td>
<td>13.5</td>
</tr>
<tr>
<td>11.7</td>
<td>11.8</td>
<td>11.9</td>
<td>12.0</td>
<td>12.1</td>
<td>12.2</td>
<td>12.3</td>
<td>12.4</td>
<td>12.5</td>
<td>12.6</td>
<td>12.7</td>
<td>12.8</td>
<td>12.9</td>
<td>13.0</td>
<td>13.1</td>
<td>13.2</td>
<td>13.3</td>
<td>13.4</td>
<td>13.5</td>
<td>13.6</td>
</tr>
<tr>
<td>11.8</td>
<td>11.9</td>
<td>12.0</td>
<td>12.1</td>
<td>12.2</td>
<td>12.3</td>
<td>12.4</td>
<td>12.5</td>
<td>12.6</td>
<td>12.7</td>
<td>12.8</td>
<td>12.9</td>
<td>13.0</td>
<td>13.1</td>
<td>13.2</td>
<td>13.3</td>
<td>13.4</td>
<td>13.5</td>
<td>13.6</td>
<td>13.7</td>
</tr>
<tr>
<td>12.0</td>
<td>12.1</td>
<td>12.2</td>
<td>12.3</td>
<td>12.4</td>
<td>12.5</td>
<td>12.6</td>
<td>12.7</td>
<td>12.8</td>
<td>12.9</td>
<td>13.0</td>
<td>13.1</td>
<td>13.2</td>
<td>13.3</td>
<td>13.4</td>
<td>13.5</td>
<td>13.6</td>
<td>13.7</td>
<td>13.8</td>
<td>13.9</td>
</tr>
<tr>
<td>12.5</td>
<td>12.5</td>
<td>12.6</td>
<td>12.7</td>
<td>12.8</td>
<td>12.9</td>
<td>13.0</td>
<td>13.1</td>
<td>13.2</td>
<td>13.3</td>
<td>13.4</td>
<td>13.5</td>
<td>13.6</td>
<td>13.7</td>
<td>13.8</td>
<td>13.9</td>
<td>14.0</td>
<td>14.1</td>
<td>14.2</td>
<td>14.3</td>
</tr>
</tbody>
</table>

Direct representation of the variable

Associated data: statistical summary
**DBMS LINK: DESCRIPTIVE DATA**

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

**LINK**

Hybrid/integrated

**STRUCTURING**

Layer -- Table
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

RASTER REPRESENTATION: IMAGE DATA

Map (scanned)
Orthophotography
Digital Elevation Model
VECTOR/RASTER CONVERSIONS

FROM VECTOR TO RASTER: RASTERIZATION

FROM RASTER TO VECTOR: VECTORIZATION
VECTOR AND RASTER: COMPLEMENTARITY

• **Raster:**
  – Grid of ‘independent’ cells
  – Matrix calculations
  – Redundancy – volume of information

• **Vector:**
  – Objects +/- structured
  – Information with low or no redundancy (depending on the structuring)
  – Link with external databases

• **Complementarity raster-vector**

• **Improvements in hardware architectures**
  – Processor power / disk space
  – Developments in raster solutions / image processing
SCALE AND ACCURACY - 1

• Definition:
  Constant ratio between lengths measured on the map and the corresponding lengths measured on the land.

• Expression:
  – Algebraic expression = scale ratio map at 1:25,000
  – Graphical expression by a scale bar representing the ratio

Graphical expression of scale: INDISPENSABLE

• Semantic confusion
SCALE AND ACCURACY - 2

- **Significance of the scale:**
  - Representation ratio
  - Level of analysis of the studied phenomena
  - Geometric accuracy of the information

- **Level of approach of geographic space**
  - Size of a zone representable at a given scale:
    - A4 sheet: 600 km² at 1:100,000
    - ... : 6 km² at 1:10,000
  - Level of analysis:
    Compartement scale, regional scale, national scale

- **Geometric accuracy of the information**
SCALE AND ACCURACY - 3

- Scale = representation ratio
  Accuracy = quality of geometric information

- 2 dimensions of accuracy:
  - resolution = size of an ‘elementary’ pixel
  - localization accuracy: localization error

- Scale/accuracy confusion: 2 causes
  - assumption of the reader: each point is significant
  - assumption of the producer: no unnecessary quality

- With ‘undesirable’ consequences
  the zoom syndrome
GEOGRAPHIC DATA QUALITY - 1

- Key-element of any GIS project
- 70 to 80% of total project costs
GEOGRAPHIC DATA QUALITY - 2

• Acquiring data from:
  – Institutional: public agencies
  – Private providers
  – External contractors
  – Internal digitalization

GIS software incorporate tools to create layer:
  – features digitalization,
  – calculation on one or many existing plans,
  – Image processing
GEOGRAPHIC DATA QUALITY - 3

• Which use?
  – To use or extract information from this document
  – As based map (illustration or filling)
  – To take measurements
  – To analyze simultaneously several information plans (spatial analysis)

• Whatever use, consider:
  – Precision of imported data
  – Level of detail
  – Reference system used (compatibility with local database)
  – Data property (copyright)
METADATA, TO GUARANTEE DATA QUALITY

• **Metadata = "data about data"
  - The procedures followed to acquire the data
  - The precision and methods of measurements
  - The age of the data and update
  - The data coding
  - The geographic referencing
  - The geometry
  - The attributes

• **Absence of metadata**
  - False interpretation
  - Bad usage
  - Erroneous perception of precision

Example of metadata's appearance
ArcCatalog (ESRI)
GIS: concepts, methods & tools

Introduction to GIS
Structuring of geographic information

Pierre BAZILE
INFORMATION SYSTEM AND ORGANIZATION

INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

Problem -> CONTROL SYSTEM -> Decision

Requests -> Information

INFORMATION SYSTEM

STATIC

INFORMATION BASE
Data, rules, and constraints of the external universe

EXTERNAL UNIVERSE

Inputs

Outputs

DYNAMIC

INFORMATION PROCESSOR
Actions Personnel

2 / 30
AUTOMATIZATION OF THE INFORMATION SYSTEM

AUTOMATIZED INFORMATION SYSTEM

<table>
<thead>
<tr>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION BASE</td>
<td>INFORMATION PROCESSOR</td>
</tr>
<tr>
<td>Data model</td>
<td>Computer Software: DBMS</td>
</tr>
</tbody>
</table>

INPUTS

FORMALISABLE UNIVERSE

OUTPUTS

Personnel

(Database management system)
## Introduction to GIS: Structuring Geographic Information

### Spatial Reference Information System

<table>
<thead>
<tr>
<th>INFORMATION SYSTEM</th>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION BASE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic data model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial data model</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Personnel</td>
<td>Software: DBMS</td>
</tr>
<tr>
<td></td>
<td>Computer</td>
<td>Software: GIS</td>
</tr>
<tr>
<td>(Geographic information system)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
GIS: WHAT EXACTLY IS IT?

The term GIS represents, in fact, 3 different concepts:

- An information system about a territory or project
- The databases describing this IS
- The IT solutions used (in particular: software)

One has to return to the basic concept!

Importance of the spatial component of the IS (localization of objects and processes, spatial interactions between elements)
REPRESEnTING GEOGRAPHICAL DATA

Duality of Geographical Information:

- Graphical information: The geographical objects, their localization, their topological relationships

- Thematic information: The descriptors of these objects, of these localizations

‘semantic’ data
CODING GEOMETRY: TWO MODES

- image vs ‘line drawings’
- raster vs vector
VECTOR REPRESENTATION
VECTOR REPRESENTATION: ‘SPAGHETTI’

Point: Id, x, y

Line: Id,
    \( x_0, y_0 \)
    \( x_1, y_1 \)
    \( x_2, y_2 \)
    \( \ldots \)
    \( x_n, y_n \)

Surface: Id,
    \( x_0, y_0 \)
    \( x_1, y_1 \)
    \( x_2, y_2 \)
    \( \ldots \)
    \( x_n, y_n \)
    \( (x_0, y_0) \)

Data redundancy
Undefined relationships
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

VECTOR REPRESENTATION: NETWORK TOPOLOGY

Point: Id, x, y

Line: Id, (L)

N_a
x_1, y_1
x_2, y_2

N_b

Node: Id, x, y

Surface: Id,

L_1
L_2
L_n

Network Topology:

No redundancy
Management of connectedness
VECTOR REPRESENTATION:
SURFACE TOPOLOGY

Point: Id, x, y

Line: Id, (L), P_g, P_d
    N_a
    x₁, y₁
    x₂, y₂
    N_b

Node: Id, x, y

Surface: Id,
    L₁
    L₂
    Lₙ

Management of connectedness
Management of contiguity
‘SPAGHETTI’ MODEL

• **Entities**
  - point: x, y coordinates
  - line: list of x,y coordinates corresponding to nodes
  - sometimes polygons: set of x,y coordinates, forming a loop: the first coordinate couple = last coordinate couple
    - The spaghetti mode produces a visual **effect** of polygons but, most often, no polygon entity is stored

• **No spatial relationships between objects**
  - unjoined arcs – unclosed polygons
    - polygons cannot form a surface
  - intersections without nodes at the crossing of two arcs
  - Adjacent polygons that overlap or separated by ‘blanks’
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

TOPOLOGICAL MODEL

• Managing the connectedness and contiguity of objects

• 2 types of topology:
  – Planar topology
  – Network topology

• Topological entities:
  – With permanent storage of the topology
    • e.g.: ArcInfo coverage
  – Taking into account only when editing (creating, updating); this data is called pseudo-topological.
    • e.g.: shp (ArcView) or tab (MapInfo)
  – With defined topological rules, varying depending on context
    • e.g.: feature class within feature dataset in a geodatabase (ArcGIS)
**DBMS LINK DESCRIPTIVE DATA**

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

**Link Types:**
- Link
- Weak
- Strong
- Hybrid/Integrated

**Structuring Layer -- Table**
RASTER PRESENTATION

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3
2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3

Size of the cell
(geometrical resolution)

Identifier

extent in X or columns

extent in Y or rows

N (orientation)

Origin (X,Y)

dx

dy

id | attribute
---|------
1  | Wheat
2  | Grapevine
3  | Forest
RASTER REPRESENTATION: IMAGE DEPTH

✓ Number of planes

✓ Depth of each plane
  boolean = 1 bit \( (2^1) \)  
  byte = 8 bits \( (2^8) \)  
  integer = 16 bits \( (2^{16}) \)  

...  

2 values: \{0,1\}  
256 values: \{0,255\}  
65536 values: \{-32767,32765\}  

Necessity of digital coding!
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

RASTER REPRESENTATION: LINES AND POINTS

The value chosen for the empty cell is selected by convention.
### RASTER REPRESENTATION: CONTINUOUS DATA

<table>
<thead>
<tr>
<th></th>
<th>11.4</th>
<th>11.5</th>
<th>11.6</th>
<th>11.7</th>
<th>11.8</th>
<th>11.9</th>
<th>12.0</th>
<th>12.2</th>
<th>12.4</th>
<th>12.5</th>
<th>12.6</th>
<th>12.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Direct representation of the variable
- Associated data: statistical summary
**DBMS LINK: DESCRIPTIVE DATA**

- **Identifier Attributes of the point**
  - Identifier: 1
  - Attributes: 
  - Identifier: 2
  - Attributes: 
  - Identifier: 3

- **Identifier Attributes of the line**
  - Identifier: 1
  - Attributes: 
  - Identifier: 2
  - Attributes: 
  - Identifier: 3

- **Identifier Attributes of the surface**
  - Identifier: 1
  - Attributes: 
  - Identifier: 2
  - Attributes: 
  - Identifier: 3

**LINK**

Hybrid/integrated

**STRUCTURING**

Layer -- Table
RASTER REPRESENTATION: IMAGE DATA

Map (scanned)

Orthophotography

Digital Elevation Model
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

VECTOR/RASTER CONVERSIONS

FROM VECTOR TO RASTER: RASTERIZATION

FROM RASTER TO VECTOR: VECTORIZATION
VECTOR AND RASTER: COMPLEMENTARITY

• Raster:
  – Grid of ‘independent’ cells
  – Matrix calculations
  – Redundancy – volume of information

• Vector:
  – Objects +/- structured
  – Information with low or no redundancy (depending on the structuring)
  – Link with external databases

• Complementarity raster-vector

• Improvements in hardware architectures
  – Processor power / disk space
  – Developments in raster solutions / image processing
SCALE AND ACCURACY - 1

- **Definition:**
  Constant ratio between lengths measured on the map and the corresponding lengths measured on the land.

- **Expression:**
  - Algebraic expression = scale ratio
    map at 1:25,000
  - Graphical expression by a scale bar representing the ratio

  Graphical expression of scale: INDISPENSABLE

- **Semantic confusion**
SCALE AND ACCURACY - 2

• Significance of the scale:
  – Representation ratio
  – Level of analysis of the studied phenomena
  – Geometric accuracy of the information

• Level of approach of geographic space
  – Size of a zone representable at a given scale:
    • A4 sheet: 600 km² at 1:100,000
    • ... : 6 km² at 1:10,000
  – Level of analysis:
    Compartment scale, regional scale, national scale

• Geometric accuracy of the information
SCALE AND ACCURACY - 3

• Scale = representation ratio
  Accuracy = quality of geometric information

• 2 dimensions of accuracy:
  – resolution = size of an ‘elementary’ pixel
  – localization accuracy: localization error

• Scale/accuracy confusion: 2 causes
  – assumption of the reader: each point is significant
  – assumption of the producer: no unnecessary quality

• With ‘undesirable’ consequences
  \textit{the zoom syndrome}
GEOGRAPHIC DATA QUALITY - 1

Key-element of any GIS project

70 to 80% of total project costs
GEOGRAPHIC DATA QUALITY - 2

- Acquiring data from:
  - Institutional: public agencies
  - Private providers
  - External contractors
  - Internal digitalization

  GIS software incorporate tools to create layer:
  - features digitalization,
  - calculation on one or many existing plans,
  - Image processing
GEOGRAPHIC DATA QUALITY - 3

• Which use?
  – To use or extract information from this document
  – As based map (illustration or filling)
  – To take measurements
  – To analyze simultaneously several information plans (spatial analysis)

• Whatever use, consider:
  – Precision of imported data
  – Level of detail
  – Reference system used (compatibility with local database)
  – Data property (copyright)
METADATA, TO GUARANTEE DATA QUALITY

• **Metadata = "data about data"
  – The procedures followed to acquire the data
  – The precision and methods of measurements
  – The age of the data and update
  – The data coding
  – The geographic referencing
  – The geometry
  – The attributes

• **Absence of metadata**
  – False interpretation
  – Bad usage
  – Erroneous perception of precision
GIS: concepts, methods & tools

Introduction to GIS
Structuring of geographic information

Pierre BAZILE
INFORMATION SYSTEM AND ORGANIZATION

Problem ➔ CONTROL SYSTEM ➔ Decision

Requests ➔ Information

INFORMATION SYSTEM

STATIC

EXTERNAL UNIVERSE

Inputs

Data, rules, and constraints of the external universe

INFORMATION BASE

Outputs

DYNAMIC

INFORMATION PROCESSOR

Actions Personnel
AUTOMATIZATION OF THE INFORMATION SYSTEM

<table>
<thead>
<tr>
<th>AUTOMATIZED INFORMATION SYSTEM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STATIC</strong></td>
</tr>
<tr>
<td><strong>INFORMATION BASE</strong></td>
</tr>
<tr>
<td>Data model</td>
</tr>
<tr>
<td><strong>DYNAMIC</strong></td>
</tr>
<tr>
<td><strong>INFORMATION PROCESSOR</strong></td>
</tr>
<tr>
<td>Computer</td>
</tr>
<tr>
<td>Software: <strong>DBMS</strong></td>
</tr>
<tr>
<td>(Database management system)</td>
</tr>
<tr>
<td>Personnel</td>
</tr>
</tbody>
</table>

**FORMALISABLE UNIVERSE**

- Inputs
- Outputs
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

SPATIAL REFERENCE INFORMATION SYSTEM

<table>
<thead>
<tr>
<th>INFORMATION SYSTEM</th>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION BASE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic data model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial data model</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Inputs

Outputs

Personnel

Software: DBMS

Computer

Software: GIS

(Geographic information system)
GIS: WHAT EXACTLY IS IT?

The term GIS represents, in fact, 3 different concepts:

- An information system about a territory or project
- The databases describing this IS
- The IT solutions used (in particular: software)

One has to return to the basic concept!

Importance of the spatial component of the IS (localization of objects and processes, spatial interactions between elements)
REPRESENTING GEOGRAPHICAL DATA

Duality of Geographical Information:

- Graphical information: The geographical objects, their localization, their topological relationships
- Thematic information: The descriptors of these objects, of these localizations

‘semantic’ data
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

REPRESENTING GRAPHIC DATA

Semantic database

Spatial representation

Tables

Entities

Relationship

Entities

Relationships

Point
Line
Surface
Continuous

Adjacency
Inclusion
Proximity
Path

Link
CODING GEOMETRY : TWO MODES

- image vs ‘line drawings’
- raster vs vector
VECTOR REPRESENTATION
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

VECTOR REPRESENTATION: ‘SPAGHETTI’

Point: Id, x, y

Line: Id,
   \( x_0, y_0 \)
   \( x_1, y_1 \)
   \( x_2, y_2 \)
   \( \ldots \)
   \( x_n, y_n \)

Surface: Id,
   \( x_0, y_0 \)
   \( x_1, y_1 \)
   \( x_2, y_2 \)
   \( \ldots \)
   \( x_n, y_n \)
   \( (x_0, y_0) \)

Data redundancy
Undefined relationships
VECTOR REPRESENTATION: NETWORK TOPOLOGY

Point: Id, x, y

Line: Id, (L)
  \( N_a \)
  \( x_1, y_1 \)
  \( x_2, y_2 \)

  \( N_b \)

Node: Id, x, y

Surface: Id,
  \( L_1 \)
  \( L_2 \)

  \( L_n \)

No redundancy
Management of connectedness
VECTOR REPRESENTATION: SURFACE TOPOLOGY

Point: Id, x, y

Line: Id, (L), P_g, P_d
N_a
x_1, y_1
x_2, y_2
N_b

Node: Id, x, y

Surface: Id,
L_1
L_2
L_n

Management of connectedness
Management of contiguity
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

‘SPAGHETTI’ MODEL

• **Entities**
  – point: x, y coordinates
  – line: list of x,y coordinates corresponding to nodes
  – sometimes polygons: set of x,y coordinates, forming a loop: the first coordinate couple = last coordinate couple
    ➢ The spaghetti mode produces a visual effect of polygons but, most often, no polygon entity is stored

• **No spatial relationships between objects**
  – unjoined arcs – unclosed polygons
    ➢ polygons cannot form a surface
  – intersections without nodes at the crossing of two arcs
  – Adjacent polygons that overlap or separated by ‘blanks’
TOPOLOGICAL MODEL

- Managing the connectedness and contiguity of objects
- 2 types of topology:
  - Planar topology
  - Network topology
- Topological entities:
  - With permanent storage of the topology
    - e.g.: ArcInfo coverage
  - Taking into account only when editing (creating, updating); this data is called pseudo-topological.
    - e.g.: shp (ArcView) or tab (MapInfo)
  - With defined topological rules, varying depending on context
    - e.g.: feature class within feature dataset in a geodatabase (ArcGIS)
### DBMS LINK DESCRIPTIVE DATA

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

**LINK**
- Weak
- Strong
- Hybrid/Integrated

**STRUCTURING**
- Layer -- Table
RASTER PRESENTATION

- **Orientation** (N)
- **Extent in Y or Rows**
- **Extent in X or Columns**
- **Origin (X,Y)**
- **Size of the Cell**
- **Geometrical Resolution** (dx, dy)

**Table:**

<table>
<thead>
<tr>
<th>id</th>
<th>attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wheat</td>
</tr>
<tr>
<td>2</td>
<td>Grapevine</td>
</tr>
<tr>
<td>3</td>
<td>Forest</td>
</tr>
</tbody>
</table>
RASTER REPRESENTATION: IMAGE DEPTH

✓ Number of planes

✓ Depth of each plane
  boolean = 1 bit ($2^1$) → 2 values: {0,1}
  byte = 8 bits ($2^8$) → 256 values: {0,255}
  integer = 16 bits ($2^{16}$) → 65536 values: {-32767,32765}

...
The value chosen for the empty cell is selected by convention.
### RASTER REPRESENTATION: CONTINUOUS DATA

<table>
<thead>
<tr>
<th></th>
<th>11.4</th>
<th>11.5</th>
<th>11.6</th>
<th>11.8</th>
<th>11.9</th>
<th>12.2</th>
<th>12.5</th>
<th>12.4</th>
<th>12.2</th>
<th>12.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.4</td>
<td>11.5</td>
<td>11.6</td>
<td>11.9</td>
<td>12.2</td>
<td>12.5</td>
<td>12.4</td>
<td>12.2</td>
<td>12.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.4</td>
<td>11.6</td>
<td>11.7</td>
<td>11.9</td>
<td>12.4</td>
<td>12.6</td>
<td>12.6</td>
<td>12.4</td>
<td>12.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.5</td>
<td>11.7</td>
<td>11.8</td>
<td>12.4</td>
<td>12.6</td>
<td>12.6</td>
<td>12.4</td>
<td>12.1</td>
<td>11.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.7</td>
<td>11.9</td>
<td>12.0</td>
<td>12.5</td>
<td>13.0</td>
<td>13.1</td>
<td>12.6</td>
<td>12.3</td>
<td>12.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.8</td>
<td>12.1</td>
<td>12.4</td>
<td>12.5</td>
<td>13.0</td>
<td>13.2</td>
<td>12.7</td>
<td>12.4</td>
<td>12.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.0</td>
<td>12.2</td>
<td>12.5</td>
<td>12.9</td>
<td>13.4</td>
<td>13.6</td>
<td>13.1</td>
<td>12.7</td>
<td>12.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.5</td>
<td>12.5</td>
<td>12.8</td>
<td>13.1</td>
<td>13.2</td>
<td>13.6</td>
<td>13.2</td>
<td>12.6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Direct representation of the variable**
- **Associated data: statistical summary**
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

DBMS LINK: DESCRIPTIVE DATA

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

LINK

Hybrid/integrated

STRUCTURING

Layer -- Table
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

RASTER REPRESENTATION: IMAGE DATA

Map (scanned)
Orthophotography
Digital Elevation Model
VECTOR/RASTER CONVERSIONS

FROM VECTOR TO RASTER: RASTERIZATION

FROM RASTER TO VECTOR: VECTORIZATION
VECTOR AND RASTER: COMPLEMENTARITY

• **Raster:**
  – Grid of ‘independent’ cells
  – Matrix calculations
  – Redundancy – volume of information

• **Vector:**
  – Objects +/- structured
  – Information with low or no redundancy (depending on the structuring)
  – Link with external databases

• **Complementarity raster-vector**

• **Improvements in hardware architectures**
  – Processor power / disk space
  – Developments in raster solutions / image processing
SCALE AND ACCURACY - 1

• Definition:
  Constant ratio between lengths measured on the map and the corresponding lengths measured on the land.

• Expression:
  – Algebraic expression = scale ratio
    map at 1:25,000
  – Graphical expression by a scale bar representing the ratio

Graphical expression of scale: INDISPENSABLE

• Semantic confusion
SCALE AND ACCURACY - 2

• **Significance of the scale:**
  - Representation ratio
  - Level of analysis of the studied phenomena
  - Geometric accuracy of the information

• **Level of approach of geographic space**
  - Size of a zone representable at a given scale:
    - A4 sheet: 600 km² at 1:100,000
    - ... : 6 km² at 1:10,000
  - Level of analysis:
    Compartment scale, regional scale, national scale

• **Geometric accuracy of the information**
SCALE AND ACCURACY - 3

• Scale = representation ratio
  Accuracy = quality of geometric information

• 2 dimensions of accuracy:
  – resolution = size of an ‘elementary’ pixel
  – localization accuracy: localization error

• Scale/accuracy confusion: 2 causes
  – assumption of the reader: each point is significant
  – assumption of the producer: no unnecessary quality

• With ‘undesirable’ consequences
  *the zoom syndrome*
GEOGRAPHIC DATA QUALITY - 1

- Key-element of any GIS project
- 70 to 80% of total project costs
GEOGRAPHIC DATA QUALITY - 2

• Acquiring data from:
  – Institutional: public agencies
  – Private providers
  – External contractors
  – Internal digitalization

GIS software incorporate tools to create layer:
  – features digitalization,
  – calculation on one or many existing plans,
  – Image processing
GEOGRAPHIC DATA QUALITY - 3

• Which use?
  – To use or extract information from this document
  – As based map (illustration or filling)
  – To take measurements
  – To analyze simultaneously several information plans (spatial analysis)

• Whatever use, consider:
  – Precision of imported data
  – Level of detail
  – Reference system used (compatibility with local database)
  – Data property (copyright)
METADATA, TO GUARANTEE DATA QUALITY

• Metadata = "data about data"
  – The procedures followed to acquire the data
  – The precision and methods of measurements
  – The age of the data and update
  – The data coding
  – The geographic referencing
  – The geometry
  – The attributes

• Absence of metadata
  – False interpretation
  – Bad usage
  – Erroneous perception of precision

Example of metadata’s appearance
ArcCatalog (ESRI)
GIS: concepts, methods & tools

Introduction to GIS
Structuring of geographic information

Pierre BAZILE

Territories, Environment, Remote Sensing & Spatial Information Joint Research Unit Cemagref - CIRAD - ENGREF
INFORMATION SYSTEM AND ORGANIZATION

Problem → CONTROL SYSTEM → Decision

Requests → Information

INFORMATION SYSTEM

Inputs

EXTERNAL UNIVERSE

Outputs

STATIC

INFORMATION BASE

Data, rules, and constraints of the external universe

DYNAMIC

INFORMATION PROCESSOR

Actions

Personnel
AUTOMATIZATION OF THE INFORMATION SYSTEM

AUTOMATIZED INFORMATION SYSTEM

<table>
<thead>
<tr>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION BASE</td>
<td>INFORMATION PROCESSOR</td>
</tr>
<tr>
<td>Data model</td>
<td>Computer Software: DBMS</td>
</tr>
<tr>
<td></td>
<td>(Database management system)</td>
</tr>
<tr>
<td></td>
<td>Personnel</td>
</tr>
</tbody>
</table>

FORMALISABLE UNIVERSE

Inputs

Outputs
# Spatial Reference Information System

## Information System

<table>
<thead>
<tr>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Information Base</strong></td>
<td><strong>Information Processor</strong></td>
</tr>
<tr>
<td>Semantic data model</td>
<td>Personnel</td>
</tr>
<tr>
<td>Spatial data model</td>
<td>Software: <strong>DBMS</strong></td>
</tr>
<tr>
<td>(Geographic information system)</td>
<td>Computer</td>
</tr>
<tr>
<td></td>
<td>Software: <strong>GIS</strong></td>
</tr>
</tbody>
</table>

**Inputs**

**Outputs**
GIS: WHAT EXACTLY IS IT?

The term GIS represents, in fact, 3 different concepts:

- An information system about a territory or project
- The databases describing this IS
- The IT solutions used (in particular: software)

One has to return to the basic concept!

Importance of the spatial component of the IS (localization of objects and processes, spatial interactions between elements)
REPRESENTING GEOGRAPHICAL DATA

Duality of Geographical Information:

- Graphical information: The geographical objects, their localization, their topological relationships
- Thematic information: The descriptors of these objects, of these localizations

‘semantic’ data
REPRESENTING GRAPHIC DATA

Semantic database

Spatial representation

Tables

Entities

Relationship

Entities

Relationships

Point

Line

Surface

Continuous

Adjacency

Inclusion

Proximity

Path
CODING GEOMETRY: TWO MODES

- **image** vs ‘line drawings’
- **raster** vs **vector**
VECTORS REPRESENTATION
VECTOR REPRESENTATION: ‘SPAGHETTI’

Point: \( \text{Id, } x, y \) +

Line: \( \text{Id, } x_0, y_0 \\
x_1, y_1 \\
x_2, y_2 \\
\cdots \\
x_n, y_n \)

Surface: \( \text{Id, } x_0, y_0 \\
x_1, y_1 \\
x_2, y_2 \\
\cdots \\
x_n, y_n \\
(x_0, y_0) \)

Data redundancy
Undefined relationships
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

VECTOR REPRESENTATION:
NETWORK TOPOLOGY

Point: Id, x, y

Line: Id, (L)
  Na
  x1, y1
  x2, y2
  Nb

Node: Id, x, y

Surface: Id,
  L1
  L2
  ----
  Ln

1

1

2

2

3

3

4

4

5

5

6

6

L1

L1

L2

L2

L3

L3

N1

N1

N1

N1

N2

N2

N2

N2

P1

P1

P2

P2

L1

L2

L3

L1

L2

L3

x1, y1

x3, y3

x5, y5

x2, y2

x4, y4

x6, y6

L1

L2

L3

L1

L2

L3

x1, y1

x3, y3

x5, y5

x2, y2

x4, y4

x6, y6

No redundancy
Management of connectedness
VECTOR REPRESENTATION: SURFACE TOPOLOGY

Point: \( I_d, x, y \)

Line: \( I_d, (L), P_g, P_d \)
\[ \begin{align*}
N_a & \quad x_1, y_1 \\
N_b & \quad x_2, y_2 \\
\end{align*} \]

Node: \( I_d, x, y \)

Surface: \( I_d, L_1, L_2, \ldots, L_n \)

Management of connectedness
Management of contiguity
‘SPAGHETTI’ MODEL

- **Entities**
  - point: x, y coordinates
  - line: list of x,y coordinates corresponding to nodes
  - sometimes polygons: set of x,y coordinates, forming a loop: the first coordinate couple = last coordinate couple
    - The spaghetti model produces a visual *effect* of polygons but, most often, no polygon entity is stored

- **No spatial relationships between objects**
  - unjoined arcs – unclosed polygons
    - polygons cannot form a surface
  - intersections without nodes at the crossing of two arcs
  - Adjacent polygons that overlap or separated by ‘blanks’
TOPOLOGICAL MODEL

• Managing the connectedness and contiguity of objects

• 2 types of topology:
  – Planar topology
  – Network topology

• Topological entities:
  – With permanent storage of the topology
    • e.g.: ArcInfo coverage
  – Taking into account only when editing (creating, updating); this data is called pseudo-topological.
    • e.g.: shp (ArcView) or tab (MapInfo)
  – With defined topological rules, varying depending on context
    • e.g.: feature class within feature dataset in a geodatabase (ArcGIS)
DBMS LINK DESCRIPTIVE DATA

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

- **LINK**
  - Weak
  - Strong
  - Hybrid/Integrated

**STRUCTURING**

Layer -- Table
RASTER PRESENTATION

N (orientation)

Size of the cell
(geometrical resolution)

dx

dy

extent in Y or rows

extent

Origin (X,Y)

extent in X or columns

Identifier

id attribute

<table>
<thead>
<tr>
<th>id</th>
<th>attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wheat</td>
</tr>
<tr>
<td>2</td>
<td>Grapevine</td>
</tr>
<tr>
<td>3</td>
<td>Forest</td>
</tr>
</tbody>
</table>
RASTER REPRESENTATION: IMAGE DEPTH

✓ Number of planes

✓ Depth of each plane
  - boolean = 1 bit \((2^1)\) → 2 values: \{0,1\}
  - byte = 8 bits \((2^8)\) → 256 values: \{0,255\}
  - integer = 16 bits \((2^{16})\) → 65536 values: \{-32767,32765\}

... 

Necessity of digital coding!
The value chosen for the empty cell is selected by convention.
## RASTER REPRESENTATION: CONTINUOUS DATA

<table>
<thead>
<tr>
<th></th>
<th>11.4</th>
<th>11.5</th>
<th>11.6</th>
<th>11.8</th>
<th>11.9</th>
<th>12.2</th>
<th>12.5</th>
<th>12.4</th>
<th>12.2</th>
<th>12.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.4</td>
<td>11.6</td>
<td>11.7</td>
<td>11.9</td>
<td>11.9</td>
<td>12.4</td>
<td>12.6</td>
<td>12.6</td>
<td>12.4</td>
<td>12.0</td>
<td></td>
</tr>
<tr>
<td>11.5</td>
<td>11.7</td>
<td>11.8</td>
<td>11.9</td>
<td>12.4</td>
<td>12.6</td>
<td>12.6</td>
<td>12.4</td>
<td>12.1</td>
<td>11.9</td>
<td></td>
</tr>
<tr>
<td>11.7</td>
<td>11.9</td>
<td>12.0</td>
<td>12.5</td>
<td>12.6</td>
<td>13.0</td>
<td>13.1</td>
<td>12.6</td>
<td>12.3</td>
<td>12.1</td>
<td></td>
</tr>
<tr>
<td>11.8</td>
<td>12.1</td>
<td>12.4</td>
<td>12.5</td>
<td>12.8</td>
<td>13.0</td>
<td>13.2</td>
<td>12.7</td>
<td>12.4</td>
<td>12.0</td>
<td></td>
</tr>
<tr>
<td>12.0</td>
<td>12.2</td>
<td>12.5</td>
<td>12.7</td>
<td>12.9</td>
<td>13.4</td>
<td>13.6</td>
<td>13.1</td>
<td>12.7</td>
<td>12.4</td>
<td></td>
</tr>
<tr>
<td>12.5</td>
<td>12.5</td>
<td>12.8</td>
<td>13.1</td>
<td>13.2</td>
<td>13.7</td>
<td>13.9</td>
<td>13.6</td>
<td>13.2</td>
<td>12.6</td>
<td></td>
</tr>
</tbody>
</table>

- **Direct representation of the variable**
- **Associated data: statistical summary**
DBMS LINK: DESCRIPTIVE DATA

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
RASTER REPRESENTATION: IMAGE DATA

Map (scanned)

Orthophotography

Digital Elevation Model
VECTOR/RASTER CONVERSIONS

FROM VECTOR TO RASTER: RASTERIZATION

FROM RASTER TO VECTOR: VECTORIZATION
VECTOR AND RASTER: COMPLEMENTARITY

• **Raster:**
  – Grid of ‘independent’ cells
  – Matrix calculations
  – Redundancy – volume of information

• **Vector:**
  – Objects +/- structured
  – Information with low or no redundancy (depending on the structuring)
  – Link with external databases

• **Complementarity raster-vector**

• **Improvements in hardware architectures**
  – Processor power / disk space
  – Developments in raster solutions / image processing
SCALE AND ACCURACY - 1

• Definition:
  Constant ratio between lengths measured on the map and the corresponding lengths measured on the land.

• Expression:
  – Algebraic expression = scale ratio map at 1:25,000
  – Graphical expression by a scale bar representing the ratio

Graphical expression of scale: INDISPENSABLE

• Semantic confusion
• **Significance of the scale:**
  – Representation ratio
  – Level of analysis of the studied phenomena
  – Geometric accuracy of the information

• **Level of approach of geographic space**
  – Size of a zone representable at a given scale:
    • A4 sheet: 600 km² at 1:100,000
    • … : 6 km² at 1:10,000
  – Level of analysis:
    Compartment scale, regional scale, national scale

• **Geometric accuracy of the information**
SCALE AND ACCURACY - 3

- Scale = representation ratio
  Accuracy = quality of geometric information

- 2 dimensions of accuracy:
  - resolution = size of an ‘elementary’ pixel
  - localization accuracy: localization error

- Scale/accuracy confusion: 2 causes
  - assumption of the reader: each point is significant
  - assumption of the producer: no unnecessary quality

- With ‘undesirable’ consequences
  the zoom syndrome
GEOGRAPHIC DATA QUALITY - 1

- Key-element of any GIS project
- 70 to 80% of total project costs
GEOGRAPHIC DATA QUALITY - 2

- Acquiring data from:
  - Institutional: public agencies
  - Private providers
  - External contractors
  - Internal digitalization

GIS software incorporate tools to create layer:
  - features digitalization,
  - calculation on one or many existing plans,
  - Image processing
GEOGRAPHIC DATA QUALITY - 3

• Which use?
  – To use or extract information from this document
  – As based map (illustration or filling)
  – To take measurements
  – To analyze simultaneously several information plans (spatial analysis)

• Whatever use, consider:
  – Precision of imported data
  – Level of detail
  – Reference system used (compatibility with local database)
  – Data property (copyright)
METADATA, TO GUARANTEE DATA QUALITY

• Metadata = "data about data"
  – The procedures followed to acquire the data
  – The precision and methods of measurements
  – The age of the data and update
  – The data coding
  – The geographic referencing
  – The geometry
  – The attributes

• Absence of metadata
  – False interpretation
  – Bad usage
  – Erroneous perception of precision
GIS: concepts, methods & tools

Introduction to GIS
Structuring of geographic information

Pierre BAZILE
INFORMATION SYSTEM AND ORGANIZATION

Problem -> CONTROL SYSTEM -> Decision

Requests -> Information

INFORMATION SYSTEM

STATIC

INFORMATION BASE

Data, rules, and constraints of the external universe

EXTERNAL UNIVERSE

Inputs

Outputs

DYNAMIC

INFORMATION PROCESSOR

Actions

Personnel
AUTOMATIZATION OF THE INFORMATION SYSTEM

AUTOMATIZED INFORMATION SYSTEM

<table>
<thead>
<tr>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION BASE</td>
<td>INFORMATION PROCESSOR</td>
</tr>
<tr>
<td>Data model</td>
<td>Computer</td>
</tr>
<tr>
<td></td>
<td>Software: <strong>DBMS</strong></td>
</tr>
<tr>
<td></td>
<td>(Database management system)</td>
</tr>
<tr>
<td></td>
<td>Personnel</td>
</tr>
</tbody>
</table>

**Inputs**

**Outputs**

FORMALISABLE UNIVERSE
### SPATIAL REFERENCE INFORMATION SYSTEM

#### INFORMATION SYSTEM

<table>
<thead>
<tr>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INFORMATION BASE</strong></td>
<td><strong>INFORMATION PROCESSOR</strong></td>
</tr>
<tr>
<td>Semantic data model</td>
<td>Personnel</td>
</tr>
<tr>
<td>Spatial data model</td>
<td>Software: DBMS</td>
</tr>
<tr>
<td>(Geographic information system)</td>
<td>Computer</td>
</tr>
<tr>
<td>(Geographic information system)</td>
<td>Software: GIS</td>
</tr>
</tbody>
</table>

**Inputs**

**Outputs**
GIS: WHAT EXACTLY IS IT?

The term GIS represents, in fact, 3 different concepts:

• An information system about a territory or project
• The databases describing this IS
• The IT solutions used (in particular: software)

One has to return to the basic concept!

Importance of the spatial component of the IS (localization of objects and processes, spatial interactions between elements)
REPRESENTING GEOGRAPHICAL DATA

Duality of Geographical Information:

- Graphical information: The geographical objects, their localization, their topological relationships
- Thematic information: The descriptors of these objects, of these localizations

‘semantic’ data
CODING GEOMETRY: TWO MODES

- **image vs ‘line drawings’**
- **raster vs vector**
VECTOR REPRESENTATION
VECTOR REPRESENTATION: ‘SPAGHETTI’

Point: Id, x, y

Line: Id,
   \[ x_0, y_0 \]
   \[ x_1, y_1 \]
   \[ x_2, y_2 \]
   \[ \ldots \]
   \[ x_n, y_n \]

Surface: Id,
   \[ x_0, y_0 \]
   \[ x_1, y_1 \]
   \[ x_2, y_2 \]
   \[ \ldots \]
   \[ x_n, y_n \]
   \[ (x_0, y_0) \]

Data redundancy
Undefined relationships
VECTOR REPRESENTATION: NETWORK TOPOLOGY

Point: Id, x, y

Line: Id, (L)

Node: Id, x, y

Surface: Id,

No redundancy
Management of connectedness
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

VECTOR REPRESENTATION: SURFACE TOPOLOGY

Point: Id, x, y

Line: Id, (L), P_g, P_d
   N_a
   x_1, y_1
   x_2, y_2
   ---
   N_b

Node: Id, x, y

Surface: Id,
   L_1
   L_2
   ---
   L_n

Management of connectedness
Management of contiguity

L_1   L_2   L_3
N_1   N_1   N_1
N_2   N_2   N_2
P_1   P_2   U
U     P_1   P_2

1  2  3

4  5  6

L_1   L_2   L_3
P_1   P_1   P_2
P_2   P_2   P_2

N_1   N_2
x_2, y_2  x_5, y_5
‘SPAGHETTI’ MODEL

- **Entities**
  - point: x, y coordinates
  - line: list of x,y coordinates corresponding to nodes
  - sometimes polygons: set of x,y coordinates, forming a loop: the first coordinate couple = last coordinate couple
    - The spaghetti model produces a visual effect of polygons but, most often, no polygon entity is stored

- **No spatial relationships between objects**
  - unjoined arcs – unclosed polygons
    - polygons cannot form a surface
  - intersections without nodes at the crossing of two arcs
  - Adjacent polygons that overlap or separated by ‘blanks’
TOPOLOGICAL MODEL

• Managing the connectedness and contiguity of objects

• 2 types of topology:
  – Planar topology
  – Network topology

• Topological entities:
  – With permanent storage of the topology
    • e.g.: ArcInfo coverage
  – Taking into account only when editing (creating, updating); this data is called pseudo-topological.
    • e.g.: shp (ArcView) or tab (MapInfo)
  – With defined topological rules, varying depending on context
    • e.g.: feature class within feature dataset in a geodatabase (ArcGIS)
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

DBMS LINK DESCRIPTIVE DATA

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

LINK

Weak

Strong

Hybrid/Integrated

STRUCTURING

Layer -- Table
RASTER PRESENTATION

- **N** (orientation)
- Size of the cell
- (geometrical resolution)
- dx
- dy
- Origin (X,Y)
- extent in X or columns
- extent
- in Y or rows
- Identifier
- id
- attribute

<table>
<thead>
<tr>
<th>id</th>
<th>attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wheat</td>
</tr>
<tr>
<td>2</td>
<td>Grapevine</td>
</tr>
<tr>
<td>3</td>
<td>Forest</td>
</tr>
</tbody>
</table>
RASTER REPRESENTATION: IMAGE DEPTH

✓ Number of planes

✓ Depth of each plane
  boolean = 1 bit \(2^1\)  \(\rightarrow\)  2 values: \{0,1\}
  byte = 8 bits \(2^8\)  \(\rightarrow\)  256 values: \{0,255\}
  integer = 16 bits \(2^{16}\)  \(\rightarrow\)  65536 values: \{-32767,32765\}

... Necessity of digital coding!
The value chosen for the empty cell is selected by convention.
### RASTER REPRESENTATION: CONTINUOUS DATA

<table>
<thead>
<tr>
<th></th>
<th>11.4</th>
<th>11.5</th>
<th>11.6</th>
<th>11.8</th>
<th>11.9</th>
<th>12.2</th>
<th>12.5</th>
<th>12.4</th>
<th>12.2</th>
<th>12.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.4</td>
<td>11.5</td>
<td>11.6</td>
<td>11.7</td>
<td>11.9</td>
<td>11.9</td>
<td>12.4</td>
<td>12.6</td>
<td>12.4</td>
<td>12.4</td>
<td>12.0</td>
</tr>
<tr>
<td>11.5</td>
<td>11.7</td>
<td>11.8</td>
<td>11.9</td>
<td>12.4</td>
<td>12.6</td>
<td>12.6</td>
<td>12.4</td>
<td>12.1</td>
<td>11.9</td>
<td></td>
</tr>
<tr>
<td>11.7</td>
<td>11.9</td>
<td>12.0</td>
<td>12.5</td>
<td>12.6</td>
<td>13.0</td>
<td>13.1</td>
<td>12.6</td>
<td>12.3</td>
<td>12.1</td>
<td></td>
</tr>
<tr>
<td>11.8</td>
<td>12.1</td>
<td>12.4</td>
<td>12.5</td>
<td>12.8</td>
<td>13.0</td>
<td>13.2</td>
<td>12.7</td>
<td>12.4</td>
<td>12.0</td>
<td></td>
</tr>
<tr>
<td>12.0</td>
<td>12.2</td>
<td>12.5</td>
<td>12.7</td>
<td>12.9</td>
<td>13.4</td>
<td>13.6</td>
<td>13.1</td>
<td>12.7</td>
<td>12.4</td>
<td></td>
</tr>
<tr>
<td>12.5</td>
<td>12.5</td>
<td>12.8</td>
<td>13.1</td>
<td>13.2</td>
<td>13.7</td>
<td>13.9</td>
<td>13.6</td>
<td>13.2</td>
<td>12.6</td>
<td></td>
</tr>
</tbody>
</table>

Direct representation of the variable

Associated data: statistical summary
DBMS LINK: DESCRIPTIVE DATA

- **LINK**
  - Hybrid/integrated

**STRUCTURING**

**Layer -- Table**

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

RASTER REPRESENTATION: IMAGE DATA

Map (scanned)
Orthophotography
Digital Elevation Model
VECTOR/RASTER CONVERSIONS

FROM VECTOR TO RASTER: RASTERIZATION

FROM RASTER TO VECTOR: VECTORIZATION
VECTOR AND RASTER: COMPLEMENTARITY

• **Raster:**
  – Grid of ‘independent’ cells
  – Matrix calculations
  – Redundancy – volume of information

• **Vector:**
  – Objects +/- structured
  – Information with low or no redundancy (depending on the structuring)
  – Link with external databases

• **Complementarity raster-vector**

• **Improvements in hardware architectures**
  – Processor power / disk space
  – Developments in raster solutions / image processing
SCALE AND ACCURACY - 1

• Definition:
  Constant ratio between lengths measured on the map and the corresponding lengths measured on the land.

• Expression:
  – Algebraic expression = scale ratio map at 1:25,000
  – Graphical expression by a scale bar representing the ratio

Graphical expression of scale: INDISPENSABLE

• Semantic confusion
• **Significance of the scale:**
  - Representation ratio
  - Level of analysis of the studied phenomena
  - Geometric accuracy of the information

• **Level of approach of geographic space**
  - Size of a zone representable at a given scale:
    - A4 sheet: 600 km² at 1:100,000
    - ... : 6 km² at 1:10,000
  - Level of analysis:
    Compartment scale, regional scale, national scale

• **Geometric accuracy of the information**
SCALE AND ACCURACY - 3

• Scale = representation ratio
  Accuracy = quality of geometric information

• 2 dimensions of accuracy:
  – resolution = size of an ‘elementary’ pixel
  – localization accuracy: localization error

• Scale/accuracy confusion: 2 causes
  – assumption of the reader: each point is significant
  – assumption of the producer: no unnecessary quality

• With ‘undesirable’ consequences
  
  the zoom syndrome
GEOGRAPHIC DATA QUALITY - 1

• Key-element of any GIS project
• 70 to 80% of total project costs
GEOGRAPHIC DATA QUALITY - 2

- Acquiring data from:
  - Institutional: public agencies
  - Private providers
  - External contractors
  - Internal digitalization

  GIS software incorporate tools to create layer:
  - features digitalization,
  - calculation on one or many existing plans,
  - Image processing
GEOGRAPHIC DATA QUALITY - 3

• Which use?
  – To use or extract information from this document
  – As based map (illustration or filling)
  – To take measurements
  – To analyze simultaneously several information plans (spatial analysis)

• Whatever use, consider:
  – Precision of imported data
  – Level of detail
  – Reference system used (compatibility with local database)
  – Data property (copyright)
METADATA, TO GUARANTEE DATA QUALITY

• Metadata = "data about data"
  – The procedures followed to acquire the data
  – The precision and methods of measurements
  – The age of the data and update
  – The data coding
  – The geographic referencing
  – The geometry
  – The attributes

• Absence of metadata
  – False interpretation
  – Bad usage
  – Erroneous perception of precision

Example of metadata's appearance
ArcCatalog (ESRI)
**INFORMATION SYSTEM AND ORGANIZATION**

**Problem** → **CONTROL SYSTEM** → **Decision**

**Requests** → **Information**

---

**INFORMATION SYSTEM**

<table>
<thead>
<tr>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INFORMATION BASE</strong></td>
<td><strong>INFORMATION PROCESSOR</strong></td>
</tr>
<tr>
<td>Data, rules, and constraints of the external universe</td>
<td>Actions Personnel</td>
</tr>
</tbody>
</table>

---

**EXTERNAL UNIVERSE**

**Inputs** → **Outputs**
AUTOMATIZATION OF THE INFORMATION SYSTEM

AUTOMATIZED INFORMATION SYSTEM

<table>
<thead>
<tr>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION BASE</td>
<td>INFORMATION PROCESSOR</td>
</tr>
<tr>
<td>Data model</td>
<td>Computer Software: DBMS</td>
</tr>
</tbody>
</table>

(Additional notes: Database management system, Personnel)
SPATIAL REFERENCE INFORMATION SYSTEM

INFORMATION SYSTEM

<table>
<thead>
<tr>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION BASE</td>
<td>INFORMATION PROCESSOR</td>
</tr>
<tr>
<td>Semantic data model</td>
<td>Personnel</td>
</tr>
<tr>
<td>Spatial data model</td>
<td>Software: DBMS</td>
</tr>
<tr>
<td></td>
<td>Computer</td>
</tr>
<tr>
<td></td>
<td>Software: GIS</td>
</tr>
</tbody>
</table>

(Geographic information system)
GIS: WHAT EXACTLY IS IT?

The term GIS represents, in fact, 3 different concepts:

- An information system about a territory or project
- The databases describing this IS
- The IT solutions used (in particular: software)

One has to return to the basic concept!

Importance of the spatial component of the IS (localization of objects and processes, spatial interactions between elements)
REPRESENTING GEOGRAPHICAL DATA

Duality of Geographical Information:

- Graphical information: The geographical objects, their localization, their topological relationships
- Thematic information: The descriptors of these objects, of these localizations

‘semantic’ data
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

REPRESENTING GRAPHIC DATA

Semantic database

Spatial representation

Entities

Relationship

Link

Entities

Relationships

Adjacency
Inclusion
Proximity
Path

Tables

Point

Line

Surface

Continuous

7 / 30
CODING GEOMETRY: TWO MODES

- image vs ‘line drawings’
- raster vs vector
VECTOR REPRESENTATION
VECTOR REPRESENTATION: ‘SPAGHETTI’

Point: Id,x,y

Line: Id,
\[ x_0, y_0 \]
\[ x_1, y_1 \]
\[ x_2, y_2 \]
\[ \vdots \]
\[ x_n, y_n \]

Surface: Id,
\[ x_0, y_0 \]
\[ x_1, y_1 \]
\[ x_2, y_2 \]
\[ \vdots \]
\[ x_n, y_n \]
\[ (x_0, y_0) \]

Data redundancy
Undefined relationships
VECTOR REPRESENTATION: NETWORK TOPOLOGY

Point: Id, x, y

Line: Id, (L)

Node: Id, x, y

Surface: Id,

No redundancy
Management of connectedness
VECTOR REPRESENTATION: SURFACE TOPOLOGY

Point: Id, x, y

Line: Id, (L), P_g, P_d
   N_a
   x_1, y_1
   x_2, y_2
   N_b

Node: Id, x, y

Surface: Id,
   L_1
   L_2
   ....
   L_n

Management of connectedness
Management of contiguity
'SPAGHETTI’ MODEL

• **Entities**
  - point: x, y coordinates
  - line: list of x,y coordinates corresponding to nodes
  - sometimes polygons: set of x,y coordinates, forming a loop: the first coordinate couple = last coordinate couple
    - The spaghetti mode produces a visual **effect** of polygons but, most often, no polygon entity is stored

• **No spatial relationships between objects**
  - unjoined arcs – unclosed polygons
    - polygons cannot form a surface
  - intersections without nodes at the crossing of two arcs
  - Adjacent polygons that overlap or separated by ‘blanks’
TOPOLOGICAL MODEL

• Managing the connectedness and contiguity of objects

• 2 types of topology:
  – Planar topology
  – Network topology

• Topological entities:
  – With permanent storage of the topology
    • e.g.: ArcInfo coverage
  – Taking into account only when editing (creating, updating); this data is called pseudo-topological.
    • e.g.: shp (ArcView) or tab (MapInfo)
  – With defined topological rules, varying depending on context
    • e.g.: feature class within feature dataset in a geodatabase (ArcGIS)
### DBMS LINK DESCRIPTIVE DATA

#### Identifier Attributes of the point

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

#### Identifier Attributes of the line

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

#### Identifier Attributes of the surface

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

**Link Types:**
- **Weak**
- **Strong**
- **Hybrid/Integrated**

**Structuring Options:**
- **Layer -- Table**
**RASTER PRESENTATION**

- **Origin (X,Y)**
- **Extent in X or columns**
- **Extent in Y or rows**
- **N (orientation)**
- **Size of the cell**
- **(geometrical resolution)**

<table>
<thead>
<tr>
<th>id</th>
<th>attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wheat</td>
</tr>
<tr>
<td>2</td>
<td>Grapevine</td>
</tr>
<tr>
<td>3</td>
<td>Forest</td>
</tr>
</tbody>
</table>
RASTER REPRESENTATION: IMAGE DEPTH

✓ Number of planes

✓ Depth of each plane
  boolean = 1 bit \( (2^1) \)  \( \rightarrow \) 2 values: \{0,1\}
  byte = 8 bits \( (2^8) \)  \( \rightarrow \) 256 values: \{0,255\}
  integer = 16 bits \( (2^{16}) \)  \( \rightarrow \) 65536 values: \{-32767,32765\}

Necessity of digital coding!
RASTER REPRESENTATION: LINES AND POINTS

<table>
<thead>
<tr>
<th>1</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The value chosen for the empty cell is selected by convention.
### RASTER REPRESENTATION:

**CONTINUOUS DATA**

<table>
<thead>
<tr>
<th></th>
<th>11.4</th>
<th>11.5</th>
<th>11.6</th>
<th>11.8</th>
<th>11.9</th>
<th>12.2</th>
<th>12.5</th>
<th>12.4</th>
<th>12.2</th>
<th>12.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Direct representation of the variable
- Associated data: statistical summary
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

DBMS LINK: DESCRIPTIVE DATA

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

LINK

Hybrid/integrated

STRUCTURING

Layer -- Table
RASTER REPRESENTATION: IMAGE DATA

- Map (scanned)
- Orthophotography
- Digital Elevation Model
VECTOR/RASTER CONVERSIONS

FROM VECTOR TO RASTER: RASTERIZATION

FROM RASTER TO VECTOR: VECTORIZATION
VECTOR AND RASTER: COMPLEMENTARITY

• **Raster:**
  - Grid of ‘independent’ cells
  - Matrix calculations
  - Redundancy – volume of information

• **Vector:**
  - Objects +/- structured
  - Information with low or no redundancy (depending on the structuring)
  - Link with external databases

• **Complementarity raster-vector**

• **Improvements in hardware architectures**
  - Processor power / disk space
  - Developments in raster solutions / image processing
SCALE AND ACCURACY - 1

• **Definition:**
  Constant ratio between lengths measured on the map and the corresponding lengths measured on the land.

• **Expression:**
  – Algebraic expression = scale ratio map at 1:25,000
  – Graphical expression by a scale bar representing the ratio

Graphical expression of scale: INDISPENSABLE

• **Semantic confusion**
SCALE AND ACCURACY - 2

• **Significance of the scale:**
  – Representation ratio
  – Level of analysis of the studied phenomena
  – Geometric accuracy of the information

• **Level of approach of geographic space**
  – Size of a zone representable at a given scale:
    • A4 sheet: 600 km² at 1:100,000
    • … : 6 km² at 1:10,000
  – Level of analysis:
    Compartment scale, regional scale, national scale

• **Geometric accuracy of the information**
SCALE AND ACCURACY - 3

• Scale = representation ratio
  Accuracy = quality of geometric information

• 2 dimensions of accuracy:
  – resolution = size of an ‘elementary’ pixel
  – localization accuracy: localization error

• Scale/accuracy confusion: 2 causes
  – assumption of the reader: each point is significant
  – assumption of the producer: no unnecessary quality

• With ‘undesirable’ consequences
  the zoom syndrome
GEOGRAPHIC DATA QUALITY - 1

- Key-element of any GIS project
- 70 to 80% of total project costs
GEOGRAPHIC DATA QUALITY - 2

- Acquiring data from:
  - Institutional: public agencies
  - Private providers
  - External contractors
  - Internal digitalization
    - GIS software incorporate tools to create layer:
      - features digitalization,
      - calculation on one or many existing plans,
      - Image processing
GEOGRAPHIC DATA QUALITY - 3

• Which use?
  – To use or extract information from this document
  – As based map (illustration or filling)
  – To take measurements
  – To analyze simultaneously several information plans (spatial analysis)

• Whatever use, consider:
  – Precision of imported data
  – Level of detail
  – Reference system used (compatibility with local database)
  – Data property (copyright)
METADATA, TO GUARANTEE DATA QUALITY

- **Metadata** = "data about data"
  - The procedures followed to acquire the data
  - The precision and methods of measurements
  - The age of the data and update
  - The data coding
  - The geographic referencing
  - The geometry
  - The attributes

- **Absence of metadata**
  - False interpretation
  - Bad usage
  - Erroneous perception of precision

Example of metadata's appearance
ArcCatalog (ESRI)
GIS: concepts, methods & tools

Introduction to GIS
Structuring of geographic information

Pierre BAZILE
INFORMATION SYSTEM AND ORGANIZATION

- Problem
  - CONTROL SYSTEM
    - Requests
    - Information
    - Decision
  - INFORMATION SYSTEM
    - STATIC
      - INFORMATION BASE
      - Data, rules, and constraints of the external universe
    - DYNAMIC
      - INFORMATION PROCESSOR
      - Actions Personnel

EXTERNAL UNIVERSE

Inputs

Outputs
AUTOMATIZATION OF THE INFORMATION SYSTEM

AUTOMATIZED INFORMATION SYSTEM

<table>
<thead>
<tr>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION BASE</td>
<td>INFORMATION PROCESSOR</td>
</tr>
<tr>
<td>Data model</td>
<td>(Database management system)</td>
</tr>
</tbody>
</table>

- **Inputs**
- **Outputs**
**SPATIAL REFERENCE INFORMATION SYSTEM**

<table>
<thead>
<tr>
<th>INFORMATION SYSTEM</th>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION BASE</td>
<td></td>
<td>INFORMATION PROCESSOR</td>
</tr>
<tr>
<td>Semantic data model</td>
<td></td>
<td>Personnel</td>
</tr>
<tr>
<td>Spatial data model</td>
<td></td>
<td>Software: DBMS</td>
</tr>
<tr>
<td>(Geographic information system)</td>
<td></td>
<td>Computer</td>
</tr>
<tr>
<td>Software: GIS</td>
<td></td>
<td>Software: GIS</td>
</tr>
</tbody>
</table>

**Inputs**
- Semantic data model
- Spatial data model

**Outputs**
- (Geographic information system)
The term GIS represents, in fact, 3 different concepts:

- An information system about a territory or project
- The databases describing this IS
- The IT solutions used (in particular: software)

One has to return to the basic concept!

Importance of the spatial component of the IS (localization of objects and processes, spatial interactions between elements)
REPRESENTING GEOGRAPHICAL DATA

Duality of Geographical Information:

- Graphical information: The geographical objects, their localization, their topological relationships
- Thematic information: The descriptors of these objects, of these localizations

‘semantic’ data
CODING GEOMETRY: TWO MODES

- **image vs ‘line drawings’**
- **raster vs vector**
VECTOR REPRESENTATION
VECTOR REPRESENTATION: ‘SPAGHETTI’

Point: Id, x, y

Line: Id,
    x₀, y₀
    x₁, y₁
    x₂, y₂
    ...
    xₙ, yₙ

Surface: Id,
    x₀, y₀
    x₁, y₁
    x₂, y₂
    ...
    xₙ, yₙ
    (x₀, y₀)

Data redundancy
Undefined relationships
VECTOR REPRESENTATION:
NETWORK TOPOLOGY

Point: Id, x, y

Line: Id, (L)

Node: Id, x, y

Surface: Id,

No redundancy
Management of connectedness
VECTOR REPRESENTATION:
SURFACE TOPOLOGY

Point: ld, x, y

Line: ld, (L), P_g, P_d
N_a
x_1, y_1
x_2, y_2
N_b

Node: ld, x, y

Surface: ld,
L_1
L_2
L_n

Management of connectedness
Management of contiguity
‘SPAGHETTI’ MODEL

- **Entities**
  - point: x, y coordinates
  - line: list of x,y coordinates corresponding to nodes
  - sometimes polygons: set of x,y coordinates, forming a loop: the first coordinate couple = last coordinate couple
    - The spaghetti mode produces a visual effect of polygons but, most often, no polygon entity is stored

- **No spatial relationships between objects**
  - unjoined arcs – unclosed polygons
    - polygons cannot form a surface
  - intersections without nodes at the crossing of two arcs
  - Adjacent polygons that overlap or separated by ‘blanks’
TOPOLOGICAL MODEL

• Managing the connectedness and contiguity of objects

• 2 types of topology:
  – Planar topology
  – Network topology

• Topological entities:
  – With permanent storage of the topology
    • e.g.: ArcInfo coverage
  – Taking into account only when editing (creating, updating); this data is called pseudo-topological.
    • e.g.: shp (ArcView) or tab (MapInfo)
  – With defined topological rules, varying depending on context
    • e.g.: feature class within feature dataset in a geodatabase (ArcGIS)
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

DBMS LINK DESCRIPTIVE DATA

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

LINK

Weak

Strong

Hybrid/Integrated

STRUCTURING

Layer -- Table
RASTER PRESENTATION

- **Orientation (N)**
- **Size of the cell**
- **Geometrical resolution**
- **Extent in X or columns**
- **Extent in Y or rows**
- **Identifier**
- **Origin (X,Y)**

### Grid Example

```
1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1
2 2 2 2 2 3 3 3 3
2 2 2 2 2 3 3 3 3
2 2 2 2 2 3 3 3 3
2 2 2 2 2 3 3 3 3
2 2 2 2 2 3 3 3 3
2 2 2 2 3 3 3 3 3
2 2 2 2 3 3 3 3 3
```

### Attribute Table

<table>
<thead>
<tr>
<th>id</th>
<th>attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wheat</td>
</tr>
<tr>
<td>2</td>
<td>Grapevine</td>
</tr>
<tr>
<td>3</td>
<td>Forest</td>
</tr>
</tbody>
</table>
RASTER REPRESENTATION: IMAGE DEPTH

✓ Number of planes

✓ Depth of each plane

- boolean = 1 bit ($2^1$)  ➔ 2 values: {0,1}
- byte = 8 bits ($2^8$)  ➔ 256 values: {0,255}
- integer = 16 bits ($2^{16}$)  ➔ 65536 values: {-32767,32765}

...  

Necessity of digital coding!
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

RASTER REPRESENTATION: LINES AND POINTS

The value chosen for the empty cell is selected by convention.
### RASTER REPRESENTATION: CONTINUOUS DATA

<table>
<thead>
<tr>
<th></th>
<th>11.4</th>
<th>11.5</th>
<th>11.6</th>
<th>11.8</th>
<th>11.9</th>
<th>12.2</th>
<th>12.4</th>
<th>12.6</th>
<th>12.1</th>
<th>12.2</th>
<th>12.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11.4</td>
<td>11.5</td>
<td>11.6</td>
<td>11.8</td>
<td>11.9</td>
<td>12.2</td>
<td>12.4</td>
<td>12.6</td>
<td>12.4</td>
<td>12.1</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>11.4</td>
<td>11.6</td>
<td>11.7</td>
<td>11.9</td>
<td>11.9</td>
<td>12.4</td>
<td>12.6</td>
<td>12.6</td>
<td>12.6</td>
<td>12.4</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>11.5</td>
<td>11.7</td>
<td>11.8</td>
<td>11.9</td>
<td>12.4</td>
<td>12.6</td>
<td>12.6</td>
<td>12.4</td>
<td>12.1</td>
<td>11.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.7</td>
<td>11.9</td>
<td>12.0</td>
<td>12.5</td>
<td>12.6</td>
<td>13.0</td>
<td>13.1</td>
<td>12.6</td>
<td>12.3</td>
<td>12.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.8</td>
<td>12.1</td>
<td>12.4</td>
<td>12.5</td>
<td>12.8</td>
<td>13.0</td>
<td>13.2</td>
<td>12.7</td>
<td>12.4</td>
<td>12.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.0</td>
<td>12.2</td>
<td>12.5</td>
<td>12.7</td>
<td>12.9</td>
<td>13.4</td>
<td>13.6</td>
<td>13.1</td>
<td>12.7</td>
<td>12.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.5</td>
<td>12.5</td>
<td>12.8</td>
<td>13.1</td>
<td>13.2</td>
<td>13.7</td>
<td>13.9</td>
<td>13.6</td>
<td>13.2</td>
<td>12.6</td>
<td></td>
</tr>
</tbody>
</table>

- Direct representation of the variable
- Associated data: statistical summary
DBMS LINK: DESCRIPTIVE DATA

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

LINK

Hybrid/integrated

STRUCTURING

Layer -- Table
RASTER REPRESENTATION: IMAGE DATA

Map (scanned)

Orthophotography

Digital Elevation Model
VECTOR/RASTER CONVERSIONS

FROM VECTOR TO RASTER: RASTERIZATION

FROM RASTER TO VECTOR: VECTORIZATION
VECTOR AND RASTER: COMPLEMENTARITY

- **Raster:**
  - Grid of ‘independent’ cells
  - Matrix calculations
  - Redundancy – volume of information

- **Vector:**
  - Objects ± structured
  - Information with low or no redundancy (depending on the structuring)
  - Link with external databases

- **Complementarity raster-vector**

- **Improvements in hardware architectures**
  - Processor power / disk space
  - Developments in raster solutions / image processing
**SCALE AND ACCURACY - 1**

- **Definition:**
  Constant ratio between lengths measured on the map and the corresponding lengths measured on the land.

- **Expression:**
  - Algebraic expression = scale ratio map at 1:25,000
  - Graphical expression by a scale bar representing the ratio

  Graphical expression of scale: INDISPENSABLE

- **Semantic confusion**
SCALE AND ACCURACY - 2

• **Significance of the scale:**
  – Representation ratio
  – Level of analysis of the studied phenomena
  – Geometric accuracy of the information

• **Level of approach of geographic space**
  – Size of a zone representable at a given scale:
    • A4 sheet: 600 km² at 1:100,000
    • ... : 6 km² at 1:10,000
  – Level of analysis:
    Compartment scale, regional scale, national scale

• **Geometric accuracy of the information**
SCALE AND ACCURACY - 3

- Scale = representation ratio
  Accuracy = quality of geometric information

- 2 dimensions of accuracy:
  - resolution = size of an ‘elementary’ pixel
  - localization accuracy: localization error

- Scale/accuracy confusion: 2 causes
  - assumption of the reader: each point is significant
  - assumption of the producer: no unnecessary quality

- With ‘undesirable’ consequences
  *the zoom syndrome*
GEOGRAPHIC DATA QUALITY - 1

• Key-element of any GIS project
• 70 to 80 % of total project costs
GEOGRAPHIC DATA QUALITY - 2

• Acquiring data from:
  – Institutional: public agencies
  – Private providers
  – External contractors
  – Internal digitalization
    GIS software incorporate tools to create layer:
    – features digitalization,
    – calculation on one or many existing plans,
    – Image processing
GEOGRAPHIC DATA QUALITY - 3

- **Which use?**
  - To use or extract information from this document
  - As based map (illustration or filling)
  - To take measurements
  - To analyze simultaneously several information plans (spatial analysis)

- **Whatever use, consider:**
  - Precision of imported data
  - Level of detail
  - Reference system used (compatibility with local database)
  - Data property (copyright)

Information to be stored in metadata
METADATA, TO GUARANTEE DATA QUALITY

- Metadata = "data about data"
  - The procedures followed to acquire the data
  - The precision and methods of measurements
  - The age of the data and update
  - The data coding
  - The geographic referencing
  - The geometry
  - The attributes

- Absence of metadata
  - False interpretation
  - Bad usage
  - Erroneous perception of precision
GIS: concepts, methods & tools

Introduction to GIS
Structuring of geographic information

Pierre BAZILE
INFORMATION SYSTEM AND ORGANIZATION

CONTROL SYSTEM

Problem → Requests → Control System → Information → Decision

INFORMATION SYSTEM

STATIC

INFORMATION BASE

Data, rules, and constraints of the external universe

EXTERNAL UNIVERSE

Inputs

Outputs

DYNAMIC

INFORMATION PROCESSOR

Actions Personnel

2 / 30
AUTOMATIZATION OF THE INFORMATION SYSTEM

AUTOMATIZED INFORMATION SYSTEM

<table>
<thead>
<tr>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION BASE</td>
<td>INFORMATION PROCESSOR</td>
</tr>
<tr>
<td>Data model</td>
<td>Computer</td>
</tr>
<tr>
<td></td>
<td>Software: DBMS</td>
</tr>
<tr>
<td></td>
<td>(Database management system)</td>
</tr>
<tr>
<td></td>
<td>Personnel</td>
</tr>
</tbody>
</table>
SPATIAL REFERENCE INFORMATION SYSTEM

- **INFORMATION SYSTEM**
  - **STATIC**
    - INFORMATION BASE
      - Semantic data model
  - **DYNAMIC**
    - INFORMATION PROCESSOR
      - Personnel
      - Software: DBMS
      - Computer
      - Software: GIS
      - (Geographic information system)

- **Inputs**
- **Outputs**
GIS: WHAT EXACTLY IS IT?

The term GIS represents, in fact, 3 different concepts:

• An information system about a territory or project
• The databases describing this IS
• The IT solutions used (in particular: software)

One has to return to the basic concept!

Importance of the spatial component of the IS (localization of objects and processes, spatial interactions between elements)
REPRESENTING GEOGRAPHICAL DATA

Duality of Geographical Information:

- Graphical information: The geographical objects, their localization, their topological relationships
- Thematic information: The descriptors of these objects, of these localizations

‘semantic’ data
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

REPRESENTING GRAPHIC DATA

Semantic database

Spatial representation

Tables

Entities

Relationship

Link

Entities

Relationships

Point

Line

Surface

Continuous

Adjacency
Inclusion
Proximity
Path
CODING GEOMETRY: TWO MODES

- image vs ‘line drawings’
- raster vs vector
VECTOR REPRESENTATION
VECTOR REPRESENTATION: ‘SPAGHETTI’

Point: Id,x,y
Line: Id, x₀,y₀, x₁,y₁, x₂,y₂, ..., xₙ,yₙ
Surface: Id, x₀,y₀, x₁,y₁, x₂,y₂, ..., xₙ,yₙ, (x₀,y₀)

Data redundancy
Undefined relationships
VECTOR REPRESENTATION: NETWORK TOPOLOGY

Point: Id, x, y

Line: Id, (L)

\[ N_a \]
\[ x_1, y_1 \]
\[ x_2, y_2 \]

\[ N_b \]

Node: Id, x, y

Surface: Id,

\[ L_1 \]
\[ L_2 \]
\[ L_n \]

No redundancy
Management of connectedness
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

VECTOR REPRESENTATION:
SURFACE TOPOLOGY

Point: Id, x, y

Line: Id, (L), Pg, Pd

N_a

x_1, y_1

x_2, y_2

N_b

Node: Id, x, y

Surface: Id,

L_1

L_2

----

L_n

Management of connectedness
Management of contiguity
‘SPAGHETTI’ MODEL

• **Entities**
  - point: x, y coordinates
  - line: list of x,y coordinates corresponding to nodes
  - sometimes polygons: set of x,y coordinates, forming a loop: the first coordinate couple = last coordinate couple
    - The spaghetti mode produces a visual **effect** of polygons but, most often, no polygon entity is stored

• **No spatial relationships between objects**
  - unjoined arcs – unclosed polygons
    - polygons cannot form a surface
  - intersections without nodes at the crossing of two arcs
  - Adjacent polygons that overlap or separated by ‘blanks’
TOPOLOGICAL MODEL

• Managing the connectedness and contiguity of objects

• 2 types of topology:
  – Planar topology
  – Network topology

• Topological entities:
  – With permanent storage of the topology
    • e.g.: ArcInfo coverage
  – Taking into account only when editing (creating, updating); this data is called pseudo-topological.
    • e.g.: shp (ArcView) or tab (MapInfo)
  – With defined topological rules, varying depending on context
    • e.g.: feature class within feature dataset in a geodatabase (ArcGIS)
### DBMS LINK DESCRIPTIVE DATA

#### Structuring Geographic Information

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

**LINK**

- Weak
- Strong
- Hybrid/Integrated

**STRUCTURING**

- Layer -- Table
RASTER PRESENTATION

- **N** (orientation)
- **dx** (Size of the cell)
- **dy** (geometrical resolution)
- **extent** in Y or rows
- **Origin (X,Y)**
- **extent in X or columns**

<table>
<thead>
<tr>
<th>Id</th>
<th>Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wheat</td>
</tr>
<tr>
<td>2</td>
<td>Grapevine</td>
</tr>
<tr>
<td>3</td>
<td>Forest</td>
</tr>
</tbody>
</table>
RASTER REPRESENTATION: IMAGE DEPTH

✓ Number of planes

✓ Depth of each plane
  - boolean = 1 bit \(2^1\)
  - byte = 8 bits \(2^8\)
  - integer = 16 bits \(2^{16}\)

  \[\text{Depth of each plane:} \begin{align*}
  \text{boolean} &= 1 \text{ bit} \quad (2^1) \quad 2 \text{ values: } \{0, 1\} \\
  \text{byte} &= 8 \text{ bits} \quad (2^8) \quad 256 \text{ values: } \{0, 255\} \\
  \text{integer} &= 16 \text{ bits} \quad (2^{16}) \quad 65536 \text{ values: } \{-32767, 32765\}
\end{align*}\]

Necessity of digital coding!
RASTER REPRESENTATION: LINES AND POINTS

The value chosen for the empty cell is selected by convention.
### RASTER REPRESENTATION: CONTINUOUS DATA

<table>
<thead>
<tr>
<th></th>
<th>11.4</th>
<th>11.5</th>
<th>11.6</th>
<th>11.8</th>
<th>11.9</th>
<th>12.2</th>
<th>12.4</th>
<th>12.5</th>
<th>12.6</th>
<th>12.2</th>
<th>12.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Direct representation of the variable

Associated data: statistical summary

<table>
<thead>
<tr>
<th></th>
<th>11.4</th>
<th>11.5</th>
<th>11.6</th>
<th>11.8</th>
<th>11.9</th>
<th>12.2</th>
<th>12.4</th>
<th>12.5</th>
<th>12.6</th>
<th>12.4</th>
<th>12.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>11.4</th>
<th>11.5</th>
<th>11.6</th>
<th>11.8</th>
<th>11.9</th>
<th>12.2</th>
<th>12.4</th>
<th>12.5</th>
<th>12.6</th>
<th>12.4</th>
<th>12.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Direct representation of the variable

Associated data: statistical summary
### DBMS LINK: DESCRIPTIVE DATA

#### Link

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

#### Hybrid/integrated

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

#### Structuring

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
VECTOR/RASTER CONVERSIONS

FROM VECTOR TO RASTER: RASTERIZATION

FROM RASTER TO VECTOR: VECTORIZATION
VECTOR AND RASTER: COMPLEMENTARITY

• **Raster:**
  – Grid of ‘independent’ cells
  – Matrix calculations
  – Redundancy – volume of information

• **Vector:**
  – Objects +/- structured
  – Information with low or no redundancy (depending on the structuring)
  – Link with external databases

• **Complementarity raster-vector**

• **Improvements in hardware architectures**
  – Processor power / disk space
  – Developments in raster solutions / image processing
SCALE AND ACCURACY - 1

• Definition:
  Constant ratio between lengths measured on the map and the corresponding lengths measured on the land.

• Expression:
  – Algebraic expression = scale ratio map at 1:25,000
  – Graphical expression by a scale bar representing the ratio

  Graphical expression of scale: INDISPENSABLE

• Semantic confusion
SCALE AND ACCURACY - 2

• **Significance of the scale:**
  – Representation ratio
  – Level of analysis of the studied phenomena
  – Geometric accuracy of the information

• **Level of approach of geographic space**
  – Size of a zone representable at a given scale:
    • A4 sheet: 600 km² at 1:100,000
    • ... : 6 km² at 1:10,000
  – Level of analysis:
    Compartment scale, regional scale, national scale

• **Geometric accuracy of the information**
SCALE AND ACCURACY - 3

- Scale = representation ratio
  Accuracy = quality of geometric information

- 2 dimensions of accuracy:
  - resolution = size of an ‘elementary’ pixel
  - localization accuracy: localization error

- Scale/accuracy confusion: 2 causes
  - assumption of the reader: each point is significant
  - assumption of the producer: no unnecessary quality

- With ‘undesirable’ consequences
  the zoom syndrome
GEOGRAPHIC DATA QUALITY - 1

- Key-element of any GIS project
- 70 to 80% of total project costs
GEOGRAPHIC DATA QUALITY - 2

- Acquiring data from:
  - Institutional: public agencies
  - Private providers
  - External contractors
  - Internal digitalization

GIS software incorporate tools to create layer:
  - features digitalization,
  - calculation on one or many existing plans,
  - Image processing
GEOROGRAPHIC DATA QUALITY - 3

• **Which use?**
  – To use or extract information from this document
  – As based map (illustration or filling)
  – To take measurements
  – To analyze simultaneously several information plans (spatial analysis)

• **Whatever use, consider:**
  – Precision of imported data
  – Level of detail
  – Reference system used (compatibility with local database)
  – Data property (copyright)
METADATA, TO GUARANTEE DATA QUALITY

- **Metadata** = "data about data"
  - The procedures followed to acquire the data
  - The precision and methods of measurements
  - The age of the data and update
  - The data coding
  - The geographic referencing
  - The geometry
  - The attributes

- **Absence of metadata**
  - False interpretation
  - Bad usage
  - Erroneous perception of precision
GIS: concepts, methods & tools

Introduction to GIS
Structuring of geographic information

Pierre BAZILE
AUTOMATIZATION OF THE INFORMATION SYSTEM

AUTOMATIZED INFORMATION SYSTEM

<table>
<thead>
<tr>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION BASE</td>
<td>INFORMATION PROCESSOR</td>
</tr>
<tr>
<td>Data model</td>
<td>Computer Software: DBMS</td>
</tr>
<tr>
<td>(Database management system)</td>
<td>Personnel</td>
</tr>
</tbody>
</table>

INPUTS

FORMALISABLE UNIVERSE

OUTPUTS
SPATIAL REFERENCE INFORMATION SYSTEM

INFORMATION SYSTEM

<table>
<thead>
<tr>
<th>STATIC</th>
<th>DYNAMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION BASE</td>
<td>INFORMATION PROCESSOR</td>
</tr>
<tr>
<td>Semantic data model</td>
<td>Personnel</td>
</tr>
<tr>
<td>Spatial data model</td>
<td>Software: DBMS</td>
</tr>
</tbody>
</table>

Computer

Software: GIS

(Geographic information system)
GIS: WHAT EXACTLY IS IT?

The term GIS represents, in fact, 3 different concepts:

• An information system about a territory or project
• The databases describing this IS
• The IT solutions used (in particular: software)

One has to return to the basic concept!

Importance of the spatial component of the IS (localization of objects and processes, spatial interactions between elements)
REPRESENTING GEOGRAPHICAL DATA

Duality of Geographical Information:

- Graphical information: The geographical objects, their localization, their topological relationships
- Thematic information: The descriptors of these objects, of these localizations

‘semantic’ data
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

REPRESENTING GRAPHIC DATA

Semantic database

Spatial representation

Tables

Entities

Relationship

Link

Entities

Relationships

Point

Line

Surface

Continuous

Adjacency
Inclusion
Proximity
Path
CODING GEOMETRY: TWO MODES

- image vs ‘line drawings’
- raster vs vector
VECTOR REPRESENTATION
INTRODUCTION TO GIS: STRUCTURING GEOGRAPHIC INFORMATION

VECTOR REPRESENTATION: ‘SPAGHETTI’

Point: Id, x, y

Line: Id,
\[ x_0, y_0 \]
\[ x_1, y_1 \]
\[ x_2, y_2 \]
\[ \cdots \]
\[ x_n, y_n \]

Surface: Id,
\[ x_0, y_0 \]
\[ x_1, y_1 \]
\[ x_2, y_2 \]
\[ \cdots \]
\[ x_n, y_n \]
\[ (x_0, y_0) \]

Data redundancy
Undefined relationships
VECTOR REPRESENTATION: NETWORK TOPOLOGY

Point: Id, x, y

Line: Id, (L)
    Na
    x1,y1
    x2,y2

Node: Id, x, y

Surface: Id,
    L1
    L2
    ----
    Ln

No redundancy
Management of connectedness
VECTOR REPRESENTATION: SURFACE TOPOLOGY

Point: Id, x, y

Line: Id, (L), P_{g}, P_{d}, N_{a}, x_{1}, y_{1}, x_{2}, y_{2}, N_{b}

Node: Id, x, y

Surface: Id, L_{1}, L_{2}, L_{n}

Management of connectedness
Management of contiguity
`SPAGHETTI’ MODEL

• **Entities**
  – point: x, y coordinates
  – line: list of x,y coordinates corresponding to nodes
  – sometimes polygons: set of x,y coordinates, forming a loop: the first coordinate couple = last coordinate couple
    - The spaghetti mode produces a visual **effect** of polygons but, most often, no polygon entity is stored

• **No spatial relationships between objects**
  – unjoined arcs – unclosed polygons
    - polygons cannot form a surface
  – intersections without nodes at the crossing of two arcs
  – Adjacent polygons that overlap or separated by ‘blanks’
TOPOLOGICAL MODEL

• Managing the connectedness and contiguity of objects

• 2 types of topology:
  – Planar topology
  – Network topology

• Topological entities:
  – With permanent storage of the topology
    • e.g.: ArcInfo coverage
  – Taking into account only when editing (creating, updating); this data is called pseudo-topological.
    • e.g.: shp (ArcView) or tab (MapInfo)
  – With defined topological rules, varying depending on context
    • e.g.: feature class within feature dataset in a geodatabase (ArcGIS)
### DBMS Link Descriptive Data

#### Link

- **Weak**
- **Strong**
- **Hybrid/Integrated**

#### Structuring

- **Layer -- Table**

---

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the point</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Attributes of the surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
RASTER PRESENTATION

- Orientation: N
- Extent in Y or rows
- Origin (X,Y)
- Extent in X or columns
- Size of the cell (geometrical resolution)
- Identifier

<table>
<thead>
<tr>
<th>id</th>
<th>attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wheat</td>
</tr>
<tr>
<td>2</td>
<td>Grapevine</td>
</tr>
<tr>
<td>3</td>
<td>Forest</td>
</tr>
</tbody>
</table>
RASTER REPRESENTATION: IMAGE DEPTH

✓ Number of planes

✓ Depth of each plane

boolean = 1 bit \( (2^1) \)  \( \rightarrow \) 2 values: \( \{0,1\} \)
byte = 8 bits \( (2^8) \)  \( \rightarrow \) 256 values: \( \{0,255\} \)
integer = 16 bits \( (2^{16}) \)  \( \rightarrow \) 65536 values: \( \{-32767,32765\} \)
...

Necessity of digital coding!
The value chosen for the empty cell is selected by convention.
## RASTER REPRESENTATION: CONTINUOUS DATA

<table>
<thead>
<tr>
<th></th>
<th>11.4</th>
<th>11.5</th>
<th>11.6</th>
<th>11.8</th>
<th>11.9</th>
<th>12.2</th>
<th>12.5</th>
<th>12.4</th>
<th>12.2</th>
<th>12.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.4</td>
<td>11.6</td>
<td>11.7</td>
<td>11.9</td>
<td>11.9</td>
<td>12.4</td>
<td>12.6</td>
<td>12.6</td>
<td>12.4</td>
<td>12.0</td>
<td></td>
</tr>
<tr>
<td>11.5</td>
<td>11.7</td>
<td>11.8</td>
<td>11.9</td>
<td>12.4</td>
<td>12.6</td>
<td>12.6</td>
<td>12.4</td>
<td>12.1</td>
<td>11.9</td>
<td></td>
</tr>
<tr>
<td>11.7</td>
<td>11.9</td>
<td>12.0</td>
<td>12.5</td>
<td>12.6</td>
<td>13.0</td>
<td>13.1</td>
<td>12.6</td>
<td>12.3</td>
<td>12.1</td>
<td></td>
</tr>
<tr>
<td>11.8</td>
<td>12.1</td>
<td>12.4</td>
<td>12.5</td>
<td>12.8</td>
<td>13.0</td>
<td>13.2</td>
<td>12.7</td>
<td>12.4</td>
<td>12.0</td>
<td></td>
</tr>
<tr>
<td>12.0</td>
<td>12.2</td>
<td>12.5</td>
<td>12.7</td>
<td>12.9</td>
<td>13.4</td>
<td>13.6</td>
<td>13.1</td>
<td>12.7</td>
<td>12.4</td>
<td></td>
</tr>
<tr>
<td>12.5</td>
<td>12.5</td>
<td>12.8</td>
<td>13.1</td>
<td>13.2</td>
<td>13.7</td>
<td>13.9</td>
<td>13.6</td>
<td>13.2</td>
<td>12.6</td>
<td></td>
</tr>
</tbody>
</table>

Direct representation of the variable

Associated data: statistical summary
RASTER REPRESENTATION: IMAGE DATA

Map (scanned)

Orthophotography

Digital Elevation Model
VECTOR/RASTER CONVERSIONS

FROM VECTOR TO RASTER: RASTERIZATION

FROM RASTER TO VECTOR: VECTORIZATION
VECTOR AND RASTER: COMPLEMENTARITY

- **Raster:**
  - Grid of ‘independent’ cells
  - Matrix calculations
  - Redundancy – volume of information

- **Vector:**
  - Objects +/- structured
  - Information with low or no redundancy (depending on the structuring)
  - Link with external databases

- **Complementarity raster-vector**

- **Improvements in hardware architectures**
  - Processor power / disk space
  - Developments in raster solutions / image processing
• **Definition:**
  Constant ratio between lengths measured on the map and the corresponding lengths measured on the land.

• **Expression:**
  – Algebraic expression = scale ratio map at 1:25,000
  – Graphical expression by a scale bar representing the ratio

Graphical expression of scale: INDISPENSABLE

• **Semantic confusion**
SCALE AND ACCURACY - 2

• **Significance of the scale:**
  – Representation ratio
  – Level of analysis of the studied phenomena
  – Geometric accuracy of the information

• **Level of approach of geographic space**
  – Size of a zone representable at a given scale:
    • A4 sheet: 600 km² at 1:100,000
    • ... : 6 km² at 1:10,000
  – Level of analysis:
    Compartment scale, regional scale, national scale

• **Geometric accuracy of the information**
SCALE AND ACCURACY - 3

• Scale = representation ratio
  Accuracy = quality of geometric information

• 2 dimensions of accuracy:
  – resolution = size of an ‘elementary’ pixel
  – localization accuracy: localization error

• Scale/accuracy confusion: 2 causes
  – assumption of the reader: each point is significant
  – assumption of the producer: no unnecessary quality

• With ‘undesirable’ consequences
  the zoom syndrome
GEOGRAPHIC DATA QUALITY - 1

- Key-element of any GIS project
- 70 to 80% of total project costs
• Acquiring data from:
  – Institutional: public agencies
  – Private providers
  – External contractors
  – Internal digitalization

GIS software incorporate tools to create layer:
  – features digitalization,
  – calculation on one or many existing plans,
  – Image processing
GEOGRAPHIC DATA QUALITY - 3

• **Which use?**
  - To use or extract information from this document
  - As based map (illustration or filling)
  - To take measurements
  - To analyze simultaneously several information plans (spatial analysis)

• **Whatever use, consider:**
  - Precision of imported data
  - Level of detail
  - Reference system used (compatibility with local database)
  - Data property (copyright)
METADATA, TO GUARANTEE DATA QUALITY

• Metadata = "data about data"
  – The procedures followed to acquire the data
  – The precision et methods of measurements
  – The age of the data and update
  – The data coding
  – The geographic referencing
  – The geometry
  – The attributes

• Absence of metadata
  – False interpretation
  – Bad usage
  – Erroneous perception of precision

Example of metadata's appearance
ArcCatalog (ESRI)