Introduction to Data Management and Spatial Database

Dealing with SpatioTemporal Data in Movement and Population Ecology - Trento, Italy - 18/22.03.2016

Ferdinando Urbano • ferdi.urbano@gmail.com

BY CHANCE DO YOU HAVE TO MANAGE DATA THAT...

- are a huge amount?
- have a complex structure?
- need quality checks?
- come in real time with high frequency?
- have spatio-temporal references?
- have multiple/distributed/different users?
- are merged with other data sources/sensors?

... OR THAT ...

- need many tools (analyze/disseminate/visualize)?
- are supposed to be used AND reused?
- will be shared (at a certain point)?
- will be connected to other information systems?
- must be preserved on the long term? [!]
- will have to be published?

... AND WITHOUT A MANAGEMENT SYSTEM IT HAPPENS THAT ...

- you continuously find and fix errors
- you spend hours to upload new data
- no one else is able to reuse the data
- you introduce new errors with data processing
- you spend ages to format your data for analysis
- there are dozens of versions of the same file
- data are “lost” when the owner changes job
- you get lost in multiple spreadsheets

SPATIAL DATABASE

There is hope!

- Storage capacity
- Retrieval performance
- Server/client structure (modular approach)
- Remote access
- Permission policy
- Concurrency control
- Data preservation
- Data formalization
- Data integrity controls

- Relational environment (data modelling)
- Easy automation of processes
- Integration in wider e-infrastructures
- Standardization
- Documentation
- Backup/recovery
- Cost effective
- Relational environment (data modelling)

DATA MANAGEMENT SKILLS ARE NEEDED!

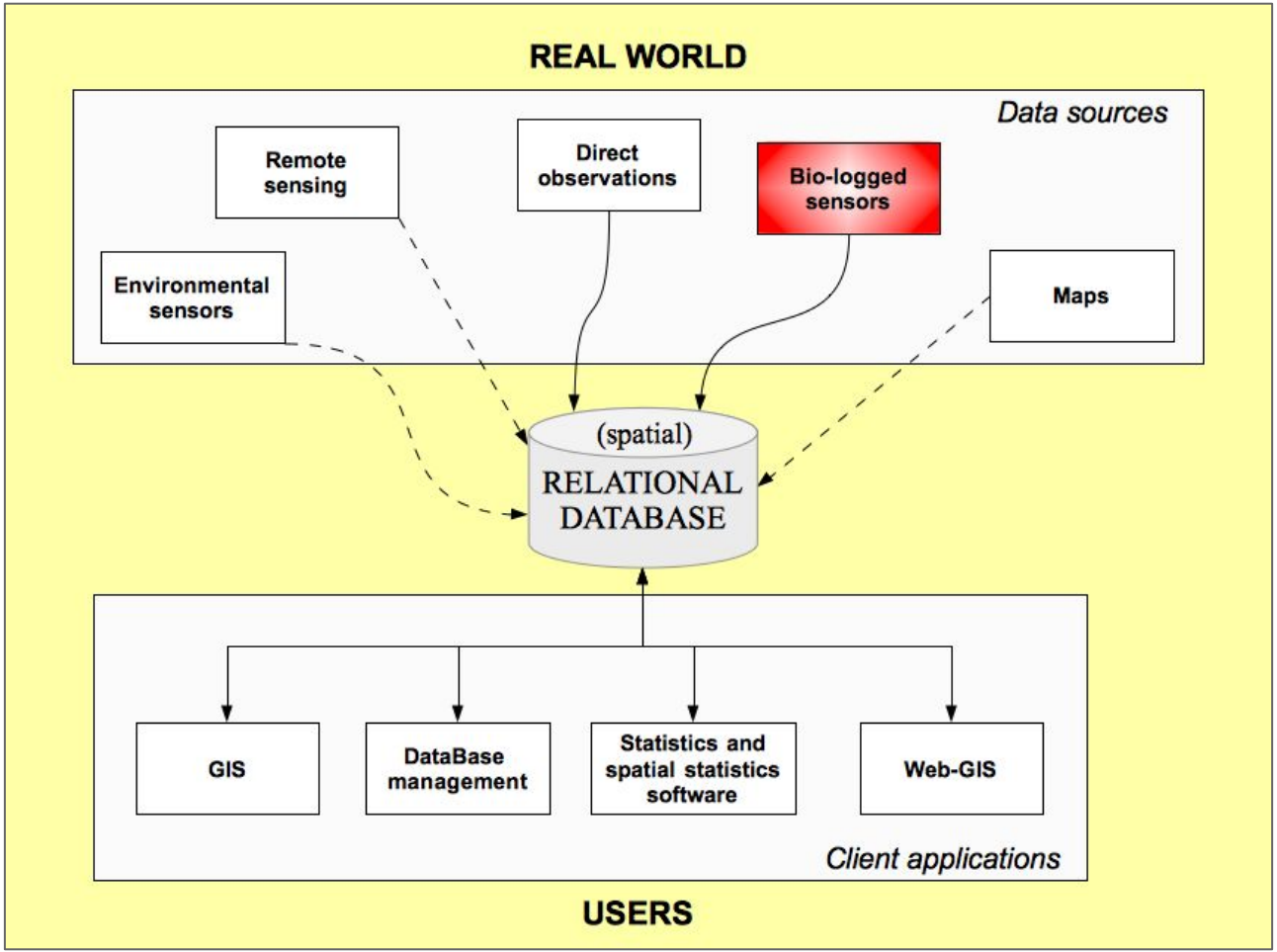
Spatial Database can make:

- **Easy** what is complex
- **Fast** what is slow
- **Automated** what is “hand (and hard) work” based
- **Permanent** what is temporary
- **Unique** what is replicated

MORE TIME FOR (BETTER) SCIENCE !

The Spatial Bit

- Spatial (and temporal) data types
- Spatial SQL
- Spatial indexes
- Animals modelled as moving object
- Integration of environmental layers
  From a geographical space to an animal's ecological space


REAL WORLD
Data sources
Remote sensing
Direct observations
Bio-logged sensors
Environmental sensors
Maps
(spatial)
RELATIONAL DATABASE
GIS
DataBase management
Statistics and spatial statistics software
Web-GIS
Client applications
USERS

Why Open Source?

- No costs for licenses
- Great spatial tools for management and analysis
- Use of standards (interoperability)
- Support of community
- Open approach to knowledge
- Why not?

Why PostgreSQL/PostGIS?

- Spatial functions
- Spatial indexes
- Geography data type, raster, topology, 3D, ...
- Supported by many software
- Active and collaborative community
- Fast development

Other options for data management: SpatiaLite

- No DBMS administration, no complex installation
- Simple with good performances
- Portable file
- Good for single users, simple applications, move data
- Implement many OGC specifications

Data Management is Important!

- ➔ SQL and Spatial SQL
- ➔ Create a database for wildlife tracking data
- ➔ R for Movement Ecology Data Analysis
- ➔ From Population Data to Spatial Modelling
- ➔ Resource Selection Analysis in Movement Ecology

- ➢ The Ecological Context Built from Satellites
- ➢ Acceleration Data
- ➢ Data sharing and Data Standards

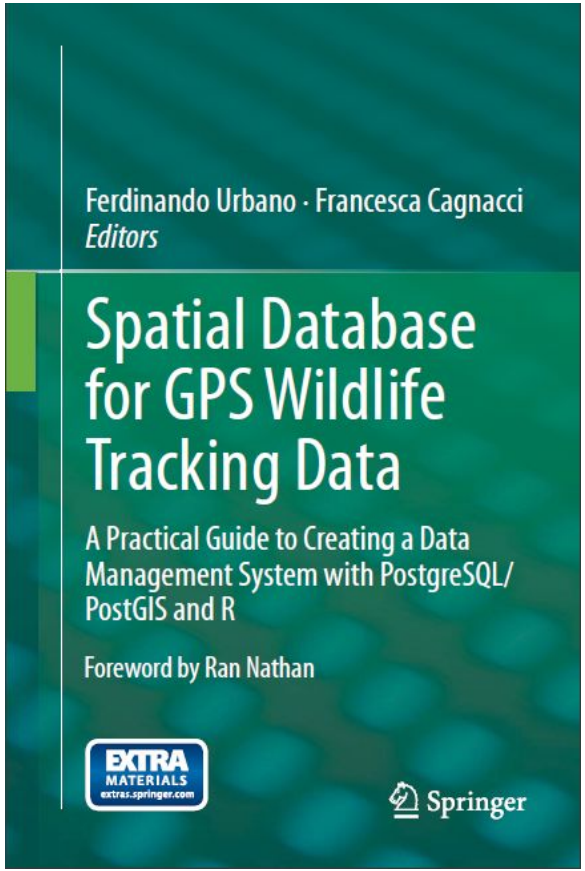
- ➢ Flash talks from students

https://github.com/feurbano/data_management_2018

PgAdmin3 [or 4] (section 1)

PosgtreSQL/PostGIS (section 2)

"Ask and it will be given to you"
(Luke, 11)



Bonus slide: Technical Issues with a Shared Data

- Quality checks
- Standardization of data structure
- Standardization of data content
- Standardization of ancillary information
- Different sampling rates
- Global spatio-temporal references
- Permission policy
- Remote access from different tools
- Increasing size of data sets (scalability)