The need of meta-database for storing and managing large amount of soil biological data: EcoBioSoil

Daniel Cluzeau, Laurence Rougé, Mario Cannavacciuolo, Anne Bellido,
Claudy C. Jolivet, Muriel Guernion, Charlène Briard, Denis Piron, Guenola
Pérès

The need of meta-database for storing and managing large amount of soil biological data: *EcoBioSoil* ®

UNIVERSITÉ DE RENNES 1

CNRS INEE

Terre, Ecosystèmes et Sociétés
Observatoire de Rennes

ECOBIO
UMR 6553

D. Cluzeau[1], L. Rougé[1], M. Cannavacciuolo[2], A. Bellido[1], C. Jolivet[3], M. Guernion[1], C. Briard[1], D. Piron[1], G. Pérès[1]

(1) University Rennes 1 - Lab. UMR CNRS 6553 *EcoBio* - Station Biologique, Paimpont (France) daniel.cluzeau@univ-rennes1.fr
(2) ESA - Angers (France)
(3) INRA US InfoSol - Orléans (France)

Context

Currently, among the international strategies for biodiversity conservation (2010-Millennium), the European Union wants to develop a unified approach of biodiversity's analysis[1,2]. During three years (2005-2007), ENVASSO program had investigated the pan-European strategies for soil sampling and data management, and had proposed selected recommendations, including for soil biodiversity.

In the 1960s, as soil biodiversity inventories begin in European countries[3,4], the same researches begin in France for earthworm groups[5]. These data have been implemented in a first database by Bouché and Soto: **Lombricien2000***.

The EcoBio's research team of the Rennes 1 University (France) had continued these inventories on several monitoring sites, at different scales (local, regional, national and european) and on various ecosystems and agrosystems. All these studies were completed by soil parameters and agricultural practices, and sometimes related to the other biological soil organisms (from microbiology to macrofauna).
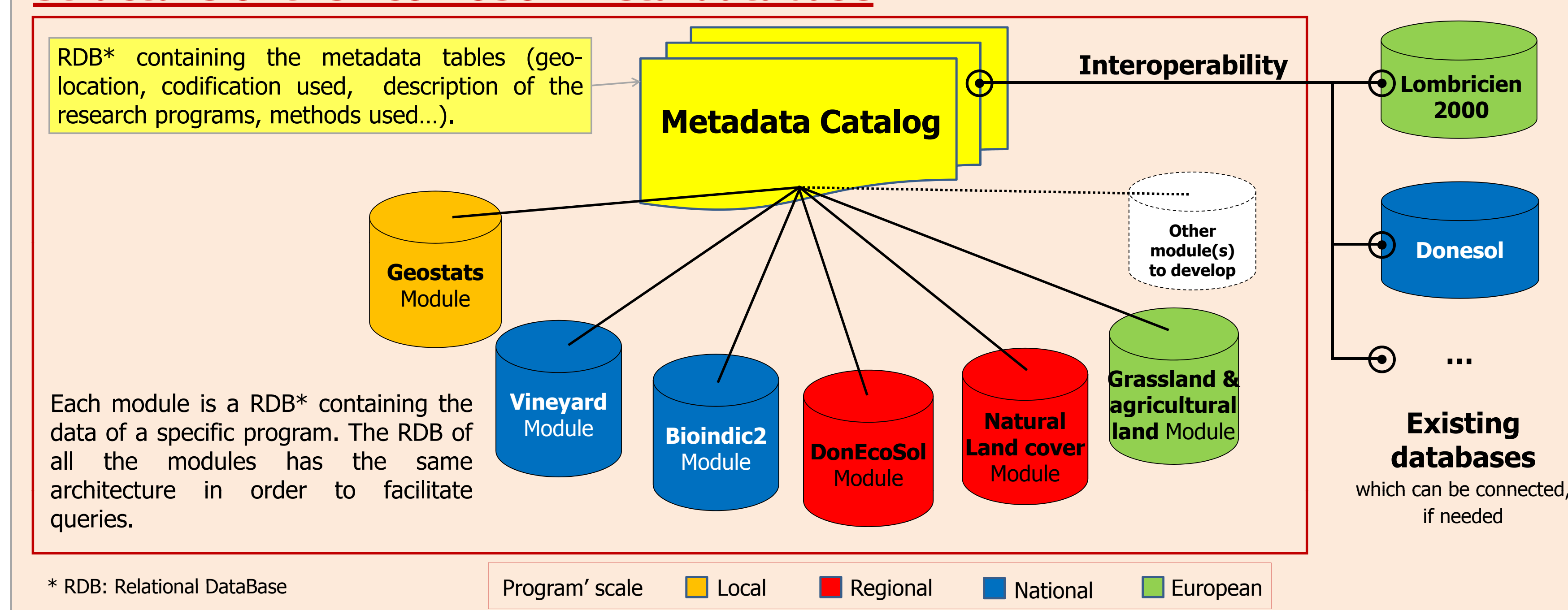
It rapidly became evident that a database was needed to integrate and coordinate amount earthworm French datasets, coming from different providers. Furthermore, it was essential to ensure the interoperability with others existing soil biodiversity databases (e.g. international "Macrofauna database"[6] hosted by FAO) and with national (**Donesol****), European (ESDB[7]) and international (ISRIC[8]) soil databases.

* **Lombricien2000**, the database developed by the CNUSC (INRA / CNRS, Montpellier, France), and hosted on the local server of EcoBio team, which contains all the data generated by the work of M.B. Bouché : taxonomic descriptions of earthworm species and their related anatomo-morphologic characteristics, as well as the location and the species lists of Bouché's samplings in Europe.

** **Donesol**, the French Soil Quality Monitoring Network database, developed by INRA Infosol (Orléans, France) , accessible via internet (http://www.gissol.fr/outil/donesol/donesol.php)

Objectives

In order to make the soil biological informations easily accessible to potential users in France, it is essential to create a tool allowing the **harmonisation and pertinent structuring of data' sets**.
It has to be **evolutive**, **interoperable** with existing databases and has to ensure **data's quality** for a **long-term conservation**.
Our aim is to develop a **relational meta-database** "*EcoBioSoil*" and, for underlying objective, to create an **integrated environment** for improving the **information sharing** and the **multidisciplinary exploration** of the soil biota data.

Material and methods

Database design has been conducted using the **entity-relationship model**[9] and the **relational theory**[10,11].

At present, the data model is implemented in the relational database management system (RDBMS) Microsoft **Access©**, which is particularly suitable for an internal use. The team is now working on a migration in another RDBMS more adapted for web use, such as **MySQL** (Oracle) or **PostgreSQL**.

Results

Structure of the *EcoBioSoil* meta-database

RDB* containing the metadata tables (geo-location, codification used, description of the research programs, methods used...).

Metadata Catalog

Interoperability

Lombricien 2000

Donesol

...

Existing databases which can be connected, if needed

Other module(s) to develop

Geostats Module

Vineyard Module

Bioindic2 Module

DonEcoSol Module

Natural Land cover Module

Grassland & agricultural land Module

Each module is a RDB* containing the data of a specific program. The RDB of all the modules has the same architecture in order to facilitate queries.

* RDB: Relational DataBase

Program' scale: Local | Regional | National | European

Glossary

« **Metadata** » is a common term referring to "data about data".
If the data are the results of the analysis of samples collected on field (taxonomic determination, countings, individual weighing, soil physico-chemical measurements,...), the **metadata are all the other informations allowing to explain these data** (geo-location, sampling date, land use, land management, methods used, climatic conditions,...). Metadata are thus the descriptive variables which are going to allow to group together or, on the contrary, to separate the datasets for datamining. The types of metadata required for ecological databases have been reviewed in numerous articles[12,13]. The Biological Data Profile[14] provides standards for the different types of metadata associated with biological data. In Europe, a working group is dedicated to give recommendations for metadata's standards within the framework of the **Inspire Directive**, establishing an infrastructure for spatial information in Europe[2].

In the literature, the term « **meta-database** » can have two meanings:
- A database dedicated to the metadata management (also called **metadata catalog**),
- A database of databases : "Strictly speaking a **meta-database** can be considered a database of databases, rather than any one integration project or technology. It collects data from different sources and usually makes them available in new and more convenient form, or with an emphasis on a particular (...) organism."[15] **This is the definition used for this term in this poster.**

Applications: some illustrations

1. *EcoBioSoil* can be connected to ArcGis© (ESRI) to generate maps with the geolocated sampling sites.

Key
Crop
Meadow
Forest
Other (dune, fallow land)
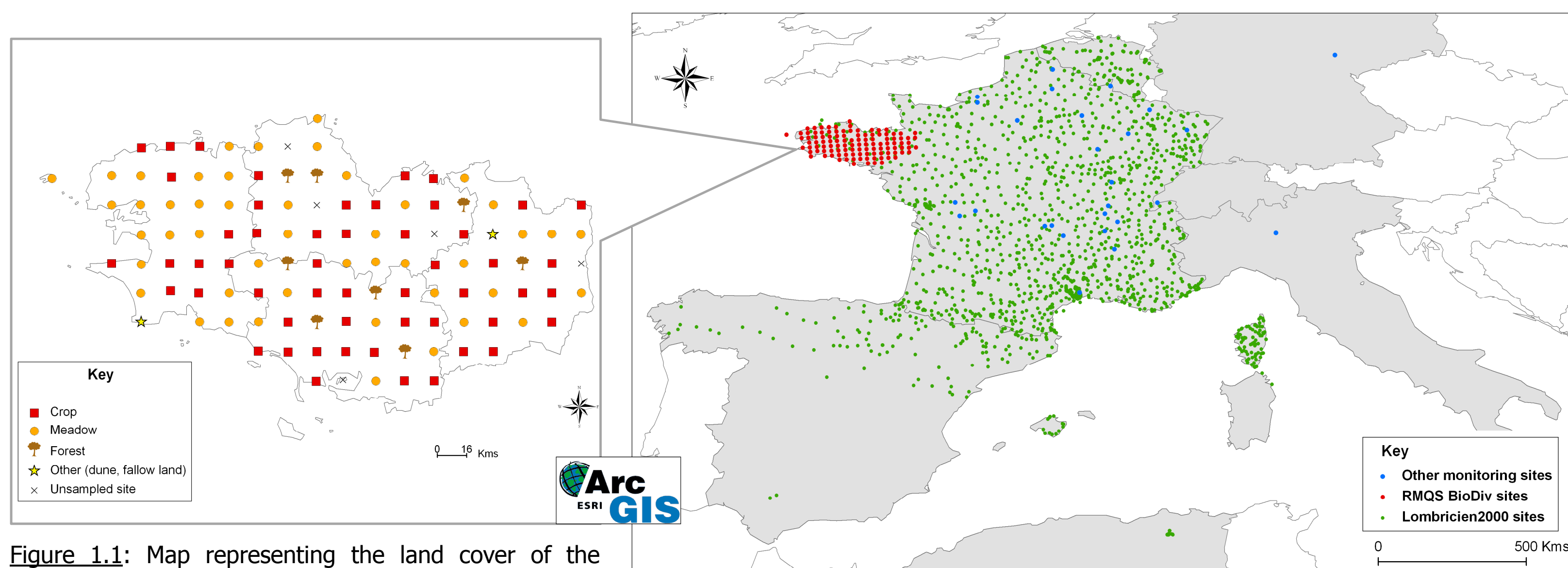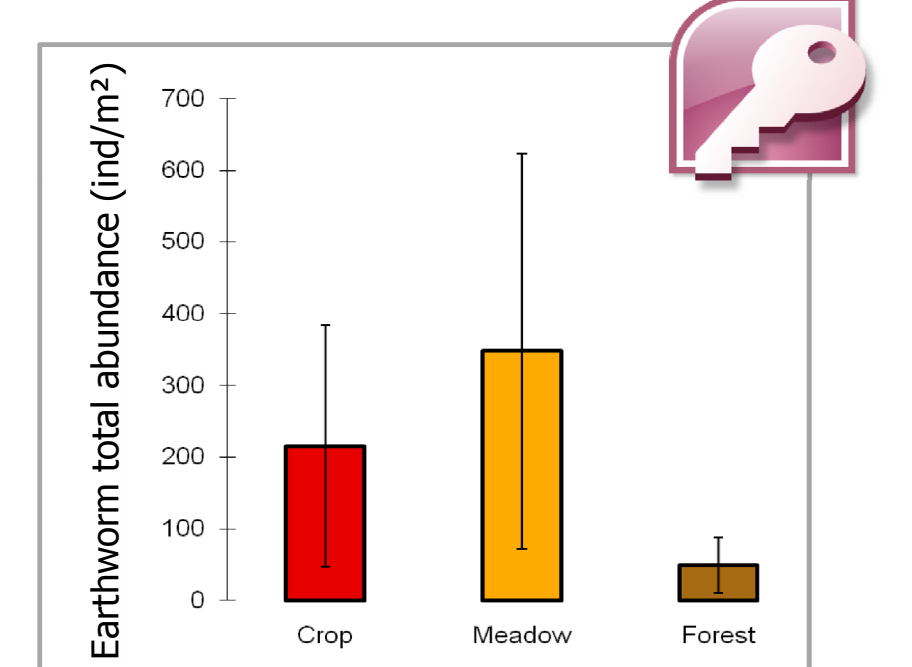Unsampled site

ArcGIS ESRI

Figure 1.1: Map representing the land cover of the geolocated sampling sites of the DonEcoSol module, containing the data of the regional Soil Quality Monitoring Network for BioDiversity (RMQS *BioDiv*, n=109).

Key
Other monitoring sites
RMQS BioDiv sites
Lombricien2000 sites

Figure 1.2: Map of the whole geolocated sampling sites of *EcoBioSoil*: DonEcoSol module (109 RMQS *BioDiv* sites), and all the other modules (27 monitoring sites), connected with the **Lombricien2000 database**, representing 2000 sampling sites in France, Europe and also Algeria.

2. *EcoBioSoil* is implemented in Access© (Microsoft). Directly in this software, simple operations of descriptive statistic can be programmed to analyse datasets.

Figure 2: Histogram representing the earthworm's total abundance of the RMQS *BioDiv* sampling sites (DonEcoSol module; n=109), separated by the land cover (see figure 1.1)
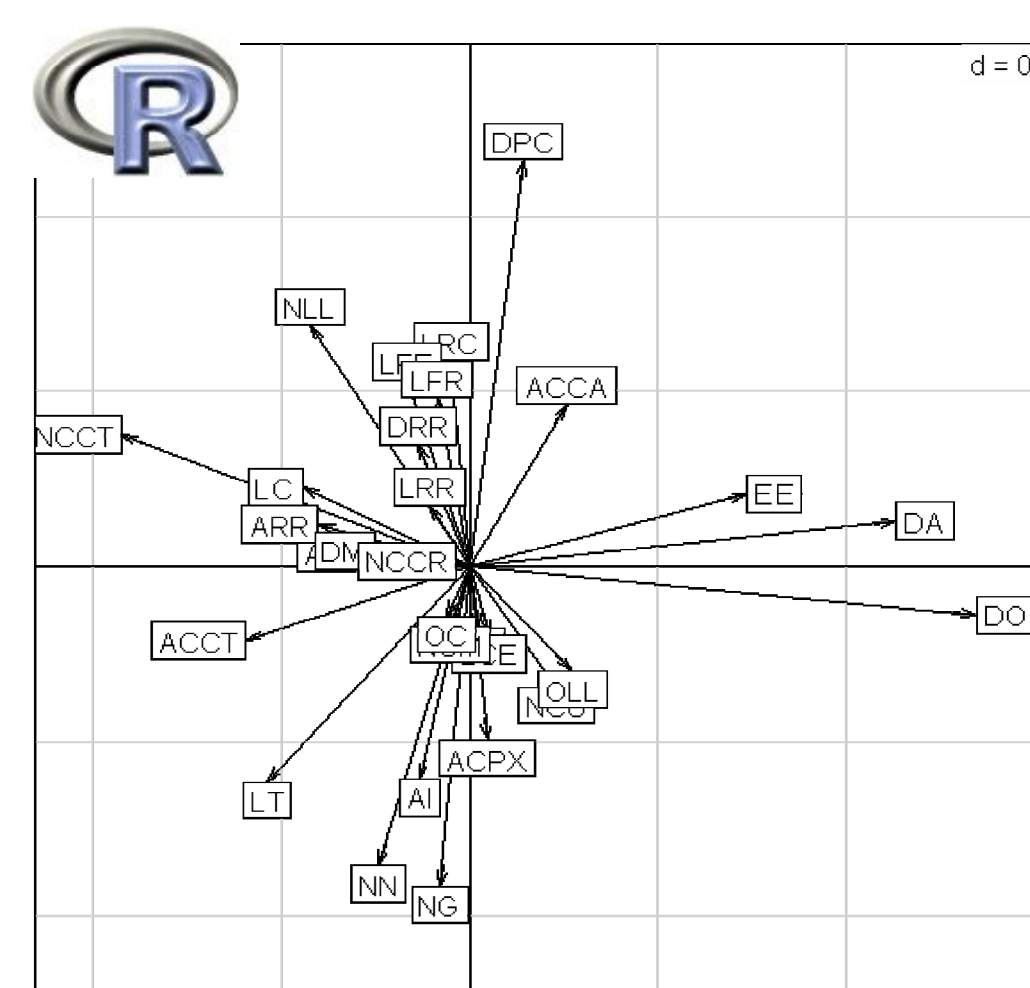

3. *EcoBioSoil* can be connected to statistic softwares, like R (open source), to make more advanced analysis, as geostatistical or multivariate analysis.

Codes used for earthworm's species are detailed in the poster : "Integration of biodiversity in soil quality monitoring: earthworm result of RMQS *BioDiv* program".

Figure 3: Graphical display of a multivariate analysis (PCA) discriminating the earthworm's species of the RMQS *BioDiv* samplings (DonEcoSol module).

Conclusion - Perspectives

For the moment, *EcoBioSoil* is reserved to an internal use, allowing the EcoBio team and scientific partners to examine and analyze amount datasets generated either by local studies or multi-partnership large-scale programs. Currently, the trend in the biological research community is for information sharing. The scientists and politics (regional, national and European) are now convinced that such an approach is necessary to build real biodiversity observation systems. So, an additional level of the *EcoBioSoil* interoperability will be achieved with the open source website project development, as it will allow many bridges with other databases (abiotic, taxonomic...). In order to meet requirements of pan-European strategies (included in 2010 - Millennium) for developing a soil biodiversity analysis unified approach at European level, databases interconnectivity will have to be enhanced, for instance by using the new possibilities offered by the web services. Moreover, the development of the open-source website will also focus on data final representation, as **communication tools** for public, or eco-informatic products to support ecological and environmental **decision makers**.

Bibliography

1: **ENVASSO** (ENVironmental ASseSsment Of Soil for mOnitoring) EU 6th Framework Research Program Priority SSP-4 Policies-1.5 Task 6. http://www.envasso.com/home.htm
2: **Directive 2007/2/EC** of the European Parliament and the Council (14/03/07) establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) http://inspire.jrc.ec.europa.eu/
3: **Rutgers M.**, Schouten A. J., Bloem J., Van Eekeren N., De Goede R. G. M., Jagersop Akkerhuis G. A. J. M., Van der Wal A., Mulder C., Brussaard L. and Breure A. M. (2009). Biological measurements in a nationwide soil monitoring network. European Journal of Soil Science, 60: 820–832. doi: 10.1111/j.1365-2389.2009.01163.x
4: **Graefe U.**, Beylich A. (2002). The German long-term soil monitoring program and its implications for the knowledge of Enchytraeidae. Christensen, B., Standen, V. (eds.): Proceedings of the 4th International Symposium on Enchytraeidae, Mols Laboratory, Denmark, 2-4 June 2000 (Newsletter on Enchytraeidae No. 7). Natura Jutlandica, Occasional papers No. 2, 2002.
5: **Bouché, M.B.** (1972). Lombriciens de France. Ecologie et systématique. Ann. Soc. Ecol. Anim. 72,1671.
6: **Lavelle P.** and Fragoso C. (Eds.). 2000. The IBOY-MACROFAUNA project: Report of an international workshop held at Bondy (France), 19-23 June 2000. IRD, Bondy, France.
7: **European Soil DataBase** http://eusoils.jrc.ec.europa.eu/ESDB_Archive/ESDB/index.htm
8: **World Soil Information Database** http://www.isric.org/
9: **Chen P.**, Pin-Shan P., 1976. The Entity-Relationship Model - Toward a Unified View of Data. ACM Transactions on Database Systems 1 (1): 9–36.
10: **Codd E. F.**, 1970. A Relational Model of Data for Large Shared Data Banks, CACM 13, No. 6.
11: **Tardieu H.**, Rochfeld A., 1983. MERISE: An information system design and development methodology. Information & Management, Volume 6, Issue 3, p. 143-159.
12: **Michener W.K.**, Brunt J.W., Helly J.J., Kirchner T.B., Stafford S.G., 1997. Non geospatial metadata for the ecological sciences. Ecological Applications 7: 330-342.
13: **Hale S.S.**, 2000. How to manage data badly (part 2). Bulletin of the Ecological Society of America 81:101-103.
14: **FGDC** Biological Data Working Group and USGS Biological Resources Division, 1999. Content standard for digital geospatial metadata - biological data profile, FGDCSTD- 001.1-1999. Federal Geographic Data Committee. Washington, D.C., U.S.A.
15: http://en.wikipedia.org/wiki/Biological_database#Meta-databases