From raw data to web services: the French soil information system

Benoît Toutain, Nicolas Saby, Manuel Martin, Florent Millet, Jean-Philippe Chenu, Jean-Baptiste J.-B. Paroissien, Marion Bardy

From raw data to web services
The french soil information system

B. Toutain 1 , N. Saby 1 , M. Martin 1 , F. Millet 1 , J.-P. Chenu 1 , J.-B. Paroissien 1 , M. Bardy 1

Introduction

At the French national level, soil data are collected according to harmonized protocols and formats in the frame of the French Soil Scientific Interest Group (Gis Sol)1, created in 2001 in response to a need to inventory and monitor French soils. Data collection is based on a network of partners with a national coverage coordinated by the Inra InfoSol unit that ensures the homogeneity and quality of data. Data are stored in harmonized formats in the French soil information system.

The first ten years of Gis Sol were dedicated to the amplification of existing programmes of soil survey and soil test collection, as well as the set-up of a national soil monitoring programme. France now has:

- an operational soil monitoring network, whose second sampling campaign that has just started
- a national soil map coverage at 1/250 000 almost complete
- an ongoing process for soil test collection

At the end of this first period, large dataset covering the whole national territory had been produced. Together with an increased awareness on soil and a context of data opening, this generated a growing interest in these datasets, and a higher number of requests for data, treated on an individual basis. Furthermore, data users reported difficulties in the appropriation of the data, in relation with the relational structure of databases, imposed by the need for data integrity and unicity. In 2012, the Gis Sol endorsed a new agreement, encompassing an objective of increasing data dissemination.

To achieve this new objective, the French national soil information system was totally redesigned, based on business intelligence models developed by Inmon2. The French soil information system now has three components (figure 1).

Architecture of the French soil information system

Operational information systems (A on figure 1): Soil data are stored in databases using harmonized national formats. Data are inserted either directly by partners in charge of soil description and sampling using user-friendly web interfaces, or inserted by the Inra InfoSol Unit. Databases are coupled with a tool, called Sivercoh, used for performing quality checks on the data. This tool is configured with sets of requests aiming at checking the consistency of databases, which are applied to every new database, with a system to track inconsistencies and any corrections of the data.

Decision support information system (B on figure 1): Soil data as well as external multi-sources data used as covariables in soil data processing are inserted in a data warehouse (figure 2), where they undergo a first step called “data staging”. In this step, multi-source data with various formats are first homogenized into a common database system. They are then transformed using Extract, Transform and Load (ETL) processes and linked to international reference data in order to make datasets interoperable. Every data transformation is traced and repeatable. Data are then filtered, aggregated and made available *via* datamarts designed to meet the needs of specific users.

Statistical information system (C on figure 1): Scripts used for processing data in order to produce maps of soil properties or indicators are stored in the statistical information system. They are linked on the one end to datamarts stored in the decision support information system, and on the other end to a Geoserver that is then used to produce web services.

In every component, data are described by extensive metadata meeting international standards.

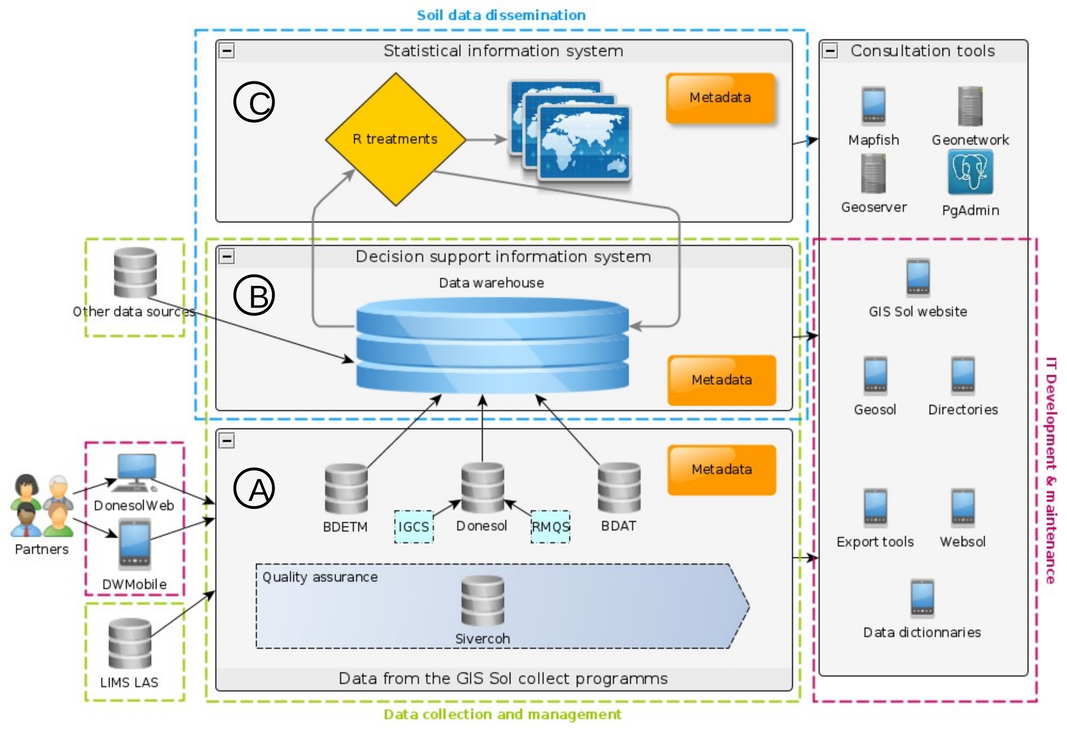

Figure 1 : *Architecture of the French soil information system*

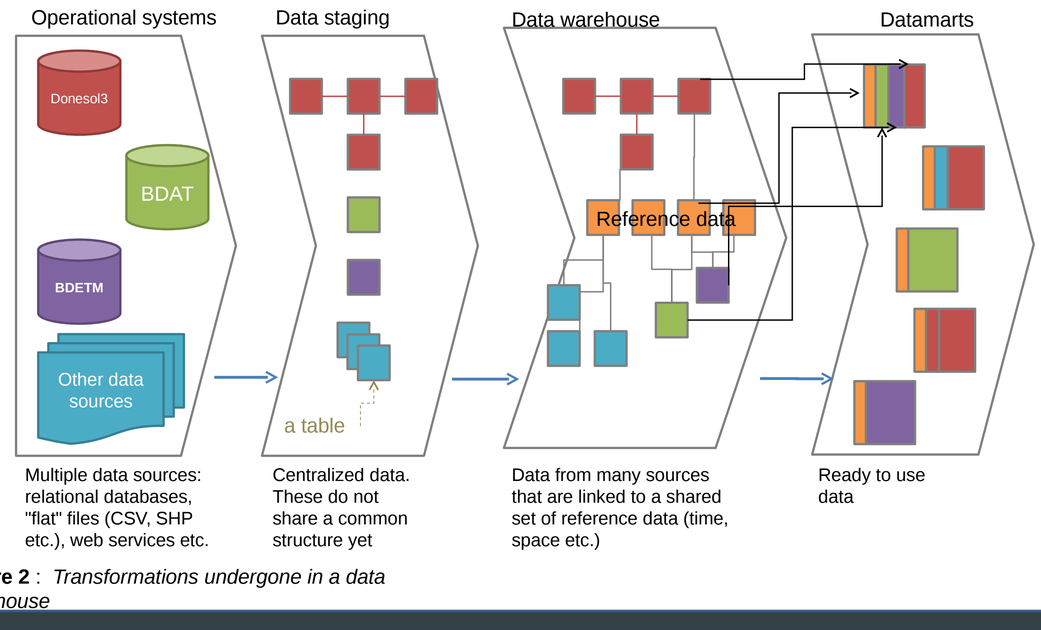

Figure 2 : *Transformations undergone in a data warehouse*

Conclusion

Redesigning the French soil information system thus improves the efficiency of data validation, storage, processing and dissemination. It enhances its quality by avoiding as much as possible errors and inconsistencies and by assuring the traceability of the transformations performed. Furthermore, transformation steps used to integrate data in the datawarehouse facilitate data appropriation, since they are available in tables rather than in relational databases. It thus highly improves data accessibility and visibility, and enables easier and more frequent updates of available data.

Références
1. https://www.gissol.fr/
2. W.H. Inmon, John Wiley & Sons; 576 p. (2005)
3. A.C. Richer de Forges et al., *Etude et Gestion des Sols* **21**, pp. 25-36 (2014)
4. http://annuaires.gissol.fr/applicasol

1 INRA US 1106 INFOSOL

Centre Val de Loire,
2163 Avenue de la Pomme de Pin, CS 40001 Ardon, 45075
Orléans