



Le Système d'Information Décisionnel Sol : une architecture informatique décisionnelle au service de la diffusion des données du GISSOL

Benoit F.P. Toutain, Nicolas P.A. Saby, Jean-Philippe Chenu, Manuel P Martin, Florent Millet, Jean-Baptiste J.-B. Paroissien, Marion Bardy

► To cite this version:

Benoit F.P. Toutain, Nicolas P.A. Saby, Jean-Philippe Chenu, Manuel P Martin, Florent Millet, et al.. Le Système d'Information Décisionnel Sol : une architecture informatique décisionnelle au service de la diffusion des données du GISSOL. 12. Journées d'Etude des Sols (JES), Association Française pour l'Etude du Sol (AFES). FRA., Jun 2014, Le Bourget du Lac, France. pp.67-68. hal-02740616

HAL Id: hal-02740616

<https://hal.inrae.fr/hal-02740616>

Submitted on 2 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le Système d'Information Décisionnel Sol : une architecture informatique décisionnelle au service de la diffusion des données du GISSOL.

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

¹ : INRA, Unité Infosol, US1106, Orléans, France

Le cadre réglementaire de la directive INSPIRE et la montée en puissance des données libres conduisent les producteurs de données à réfléchir à des politiques de diffusion. Avec la prise de conscience de l'importance des services éco systémiques rendus par les sols, les données produites par le GIS Sol sont au cœur de cet enjeu. Pourtant, la diffusion de la donnée sur les sols se heurte à deux obstacles de nature informatique et thématique. Premièrement, les utilisateurs sont en effet généralement à la recherche de données faciles à appréhender, directement extractibles et aisément exploitables. Hors, les bases de données sont structurées de manière à assurer l'intégrité et l'unicité des données. L'utilisation généralisée des formes normales, qui sont des techniques de modélisation destinées à optimiser le stockage des informations, a pour conséquence de complexifier fortement les opérations d'extraction et contraint les utilisateurs des différentes communautés scientifiques à un apprentissage long pour la mise en forme des données. Deuxièmement, le sol étant à l'interface entre la lithosphère, la biosphère et l'atmosphère, l'interprétation des données sur les sols implique l'utilisation conjointe des données émanant de ces trois domaines. L'expérience montre que la préparation des données en vue de traitements statistiques ou de l'utilisation de modèles prend un temps non négligeable.

L'architecture décisionnelle est de nature à lever ces verrous. Il s'agit d'une organisation de composants informatiques qui permet de mobiliser des données brutes de différentes sources, de leur appliquer des règles métiers de transformation et de les restituer sous une forme d'information exploitable. Elle permet ainsi de déporter la mise en forme des données du côté du producteur et d'obtenir directement des données multi-sources. Cette architecture s'inscrit pleinement dans une démarche qualité car les informations ont subi des contrôles et sont reproductibles, historisables et pleinement décrites par des métadonnées.

Au sein du GIS Sol, l'architecture décisionnelle est décrite dans l'illustration 1.

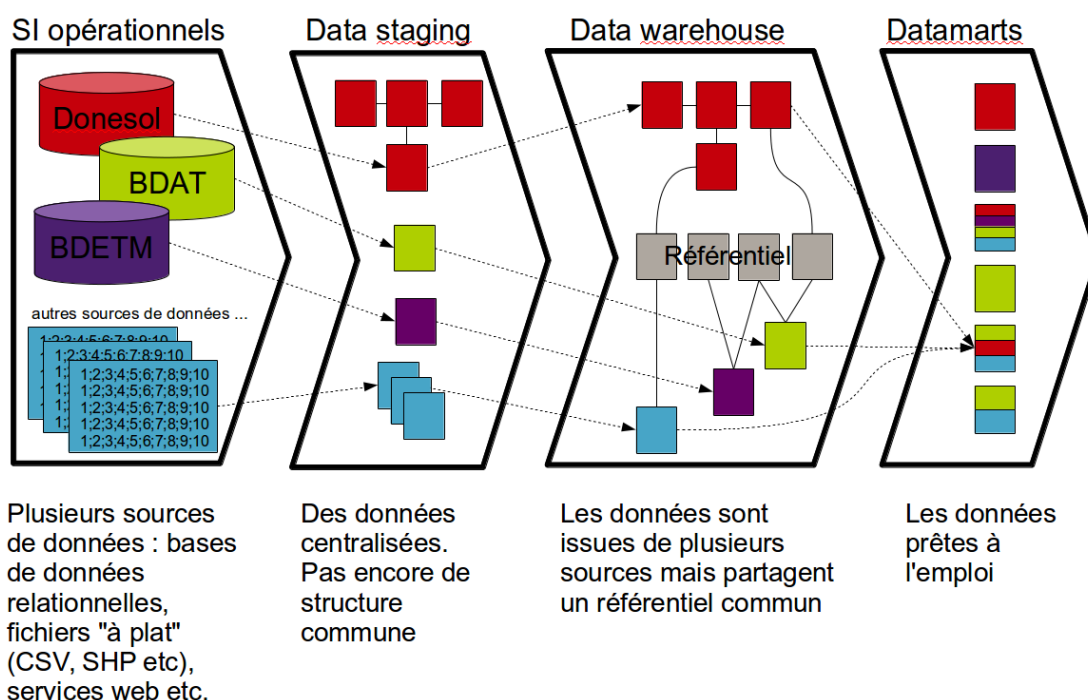


Illustration 1: l'architecture décisionnelle d'Infosol

Les systèmes d'information opérationnels contiennent à la fois les bases de données servant à collecter et stocker les données obtenues dans le cadre des programmes d'acquisition du GIS Sol (IGCS, RMQS, BDAT, BDETM) et d'autres sources de données telles que l'occupation du sol Corine Land Cover, des séries de données climatiques, des données administratives etc.

Ces données sont placées dans une structure intermédiaire appelée data staging qui permet de les centraliser et de ne pas travailler directement sur les bases de données en production.

L'entrepôt de données ou data warehouse relie à un référentiel commun les données transférées depuis le data staging. Il est structuré en faits correspondant aux observations sur les sols et en dimensions qui sont les axes d'analyse retenus. Pour les sols, les observations sont rattachées d'une part à un référentiel en quatre dimensions (un référentiel géographique x,y,z et le temps) et d'autre part à des données de référence de nature thématique (type de sol et d'horizon, entité administrative, occupation du sol).

Une fois mises en relation, les données sont filtrées, traitées et agrégées dans des magasins de données ou datamarts. Chacun de ces datamarts répond à un besoin exprimé par les utilisateurs. On trouve par exemple parmi ces besoins diffusion des résultats de la première campagne du RMQS, des statistiques issues de la base de données d'analyse de terre et également des estimations statistiques des teneurs habituelles en éléments trace métalliques. Les datamarts sont alors directement interrogeables à travers des services web (geoserver.gissol.fr), des logiciels client de base de données (e.g. pgAdmin3, ODBC), ou des applications web (e.g. bdat.gissol.fr). Ils peuvent également être publiés sur des portails de diffusion généralistes tels que pangaea.de.

Cette architecture a fait ses preuves au niveau local puisqu'elle permet le partage des données au sein de l'unité Infosol. Elle est amenée à se développer avec l'augmentation du nombre de datamarts disponibles et du nombre d'utilisateurs extérieurs venant consulter les

données. Elle sert de support à la production en masse de données statistiques et pourrait permettre l'interopérabilité des données Sol avec les plateformes de modélisation de l'INRA.