

Le système d'information national sur les sols :

DONESOL et les outils associés

E. Grolleau⁽¹⁾, L. Bargeot⁽²⁾, A. Chafchafi⁽³⁾, R. Hardy⁽¹⁾, J. Doux⁽¹⁾, A. Beaudou⁽⁴⁾, H. Le Martret⁽⁴⁾, J.-Cl. Lacassin⁽⁵⁾, J.-L. Fort⁽⁶⁾, P. Falipou⁽⁷⁾, D. Arrouays⁽¹⁾

- (1) INRA CR d'Orléans, Unité Infosol, avenue de la Pomme de Pin, BP 20619, 45166 Olivet Cedex
- (2) CNERTA-ENESAD, 4 rue Champs-Prevois, bâtiment Grand-Champs, 21000 Dijon
- (3) Sol Info Rhône-Alpes, 5, rue Hermann Frenkel, 69364 Lyon Cedex 07.
- (4) IRD Montpellier, S018 VALPEDO, BP 64501, 34394 Montpellier
- (5) Société du Canal de Provence, Le Tholonet, BP 100 13603, Aix-en-Provence Cedex 1
- (6) Chambre Régionale d'Agriculture Poitou, Charentes, 86550 Mignaloux-Beauvoir
- (7) INRA, UMR1221 Études des interactions entre sols, agrosystèmes et hydrosystèmes LISAH, 2, place Viala, 34060 Montpellier Cedex 1

RÉSUMÉ

Ces dernières années le modèle de données « DONESOL » s'est imposé, au niveau national comme en région, pour le stockage des données pédologiques spatialisées. Parallèlement, de nombreux outils ont été conçus autour de ce modèle. Cet article a pour but de dresser un premier bilan de ces développements.

Dans un souci de clarté de l'exposé, nous avons classé les outils suivant quatre fonctionnalités : (1) les outils d'aide à la saisie, qui peuvent être fixes ou portables sur le terrain tels que ceux développés à Montpellier ou à Dijon ; (2) les outils de validation des données, essentiellement conçus à l'INRA d'Orléans, ils permettent la validation, par le ministère en charge de l'agriculture, d'un niveau de qualité donné pour une étude du programme Référentiel Régional Pédologique ; (3) les outils de consultation des données qui vont de la simple mise en ligne d'un dictionnaire de méta-données à la création de sites Internet interactifs autorisant des requêtes cartographiques ou sémantiques (Rhône-Alpes et Bourgogne) ; (4) et les outils d'échange de données conçus pour le transfert de données pédologiques entre DONESOL et d'autres systèmes d'information tels que VALSOL, développé à l'IRD de Montpellier. Il convient de garder à l'esprit que cette classification est conceptuelle et que bien souvent une application informatique se situera à l'interface de plusieurs fonctionnalités.

Aujourd'hui de nombreux outils réalisés en régions ont atteint un stade de maturité suffisant pour que l'on puisse largement envisager une véritable mutualisation des développements. Pour mener à bien cette mission, les membres du programme Inventaire Gestion et Conservation des Sols ont décidé de mettre en place un groupe de réflexion sur la coordination trans-régionale et la diffusion des différentes applications informatiques.

Mots clés

Sol, système d'information, DONESOL, base de données, cartographie pédologique, surveillance, Internet, VALSOL.

SUMMARY**THE FRENCH NATIONAL INFORMATION SYSTEM FOR SPATIAL SOIL DATA : DONESOL**

In recent years, the French database system "DONESOL" has been used both at national and regional levels, for the storage of spatially-referenced pedological data. In parallel, many tools have been developed for this system. The purpose of this article is to give a first assessment of these developments.

In the interests of clarity, we can classify the tools according to four functionalities: (i) tools that assist data capture, which can be fixed or portable in the field, such as those developed at Montpellier and Dijon; (ii) tools for validation of the data, primarily conceived at the INRA of Orleans, which allow the ministry responsible for agriculture to validate the level of quality of a study of the Référentiel Régional Pédologique program; (iii) tools for consultation of the data, which range from simple on-line publishing of a dictionary of metadata to the creation of interactive Internet sites authorising cartographic or semantic requests (e.g. the Rhone-Alpes and Bourgogne); and (iv) tools for data exchange, conceived for the transfer of pedological data between DONESOL and other information systems such as VALSOL, developed by the IRD of Montpellier. It is advisable to keep in mind that this classification is conceptual and that very often a data-processing application will be located at the interface of several functionalities.

Today, many of the tools produced by the different regional partners have reached a sufficient stage of maturity to be used by all partners. To expedite the utilisation of these tools, the members of the Inventory Management and Conservation of Soils (IGCS) program have decided to set up a group responsible for the trans-regional coordination and dissemination of these various data-processing applications.

Key-words

Soil, information system, DONESOL, database, soil mapping, soil inventory, soil monitoring, Internet, VALSOL.

RESUMEN**EL SISTEMA DE INFORMACIÓN NACIONAL SOBRE LOS DATOS PEDOLÓGICOS ESPACIALIZADOS : DONESOL y sus herramientas asociadas.**

Estos últimos años el modelo de datos "DONESOL" se impone al nivel nacional como al nivel regional para el almacenamiento de los datos pedológicos espacializados. Paralelamente, numerosas herramientas fueron concebidas en relación con este modelo. Este artículo tiene como objetivo de hacer un primero balance de estos desenvolvimientos.

Con una preocupación de claridad del artículo, hemos clasificado las herramientas según cuatro funcionalidades: (1) las herramientas de ayuda a la captura de datos, que pueden ser fijos o llevados al campo como los desarrollados a Montpellier o a Dijon; (2) las herramientas de validación de los datos, esencialmente concebidos por INRA Orleans, que permiten la validación por el ministerio en carga de la agricultura, con un nivel de buena calidad para un estudio del programa Referencial Regional Pedológico; (3) las herramientas de consultación de los datos que van de la simple puesta en línea de un diccionario de meta-datos hasta la creación de sitios interactivos que autorizan las peticiones cartográficas o semánticas (Ródano-Alpes y Borgoña); (4) y las herramientas de intercambio de datos concebidas para la transferencia entre DONESOL y otros sistemas de información como VALSOL desarrollado por IRD de Montpellier. Conviene guardar en mente que esta clasificación es conceptual y que a menudo una aplicación informática se localiza al interfaces de varias funcionalidades.

Hoy, numerosas herramientas realizadas en regiones han adquirido un estado de madurez suficiente para poder ampliamente prever una verdadera mutualización de las desenvolvimientos. Para llevar a cabo esta misión, los miembros del programa Inventario gestión y conservación de los suelos decidieron establecer un grupo de reflexión sobre la coordinación trans-regional y la difusión de las diferentes aplicaciones informáticas.

Palabras claves

Suelo, sistema de información, DONESOL, banco de datos, cartografía pedológica.

En France, depuis une dizaine d'années, les préoccupations sur la fonction productive des sols ont peu à peu laissé place à des interrogations d'ordre environnemental sur la qualité des sols et sur leur rôle d'interface vis-à-vis de l'environnement (Bornand, 1997 ; Boiffin *et al.*, 2000 ; King *et al.*, 1999). Dans ce contexte, il est rapidement apparu nécessaire d'élaborer un système d'information à vocation nationale harmonisant les données relevant de la cartographie pédologique et celles issues des programmes d'observation de la qualité des sols.

Harmoniser des données environnementales est un exercice difficile, voire périlleux. En effet, de par leur hétérogénéité spatiale et, pour la plupart, leur instabilité temporelle ces données se prêtent mal à une gestion informatique rigide et standardisée. Cette spécificité doit pourtant être surmontée afin d'aboutir à un système stable et cohérent permettant de répondre tout autant à des questions de synthèse qu'à des interrogations précises dans l'espace ou dans le temps.

L'idée de devoir faire entrer des informations pédologiques dans des cases informatiques structurées et prédéfinies est souvent déroutante voire repoussante pour le pédologue expert dont les méthodes de travail sont souvent bien rodées. Il est donc nécessaire de communiquer largement sur le sujet pour expliquer tous les avantages générés par l'utilisation d'une base de données structurée. L'un des bénéfices les plus convaincants est sans doute la possibilité pour le pédologue de puiser dans la base pour retrouver d'anciens profils existants sur sa zone d'étude (Bui *et al.*, 2003). Cette fonction de stockage de la base de données permet une économie de temps puisqu'il suffit alors d'intégrer ces anciens profils dans la nouvelle étude. Un autre avantage est celui qu'engendre l'utilisation des outils de requête permettant une vision synoptique rapide des caractéristiques des sols. Couplé aux puissants Systèmes d'Informations Géographiques le résultat de ces requêtes peut être croisé avec d'autres couches d'informations (végétation, géologie, météorologie, etc.) afin de répondre à des thématiques complexes (réserve utile en eau, aléa érosif, limite de zones humides, etc.).

Ainsi, petit à petit, les maîtres d'ouvrage et les maîtres d'œuvre d'études cartographiques de sols ont pris conscience de l'importance de la création d'un système national de gestion des données pédologiques. Aujourd'hui, pour la majorité des études à petites échelles, les maîtres d'ouvrage imposent généralement dans leur cahier des charges le stockage des données sur support informatique selon le modèle de données en vigueur « DONESOL » (Gaultier *et al.*, 1993). Parallèlement, on assiste en région au développement de nombreuses applications tournant autour de ce standard. L'objet de cet article est de dresser un état d'avancement des outils développés autour du système d'information des données pédologiques spatialisées en France. Nous verrons que l'on peut distinguer quatre types d'outils : les outils d'aide à la saisie ; les outils de validation des données ; les outils de consultation et les outils d'échange de données.

LE SYSTÈME D'INFORMATION NATIONAL SUR LES DONNÉES PÉDOLOGIQUES SPATIALISÉES

Les bases de données géographiques issues de l'informatisation de l'inventaire cartographique des sols comprennent deux ensembles distincts reliés entre eux : la base géographique et la base sémantique (King *et al.*, 1994).

- La base géographique permet la localisation spatiale des objets cartographiés. Elle est gérée au sein d'un système d'information géographique, et les entités manipulées sont des polygones (contours des plages cartographiques), des lignes (limites entre plages cartographiques) et des points (emplacements des observations ponctuelles).

- Les objets figurés dans la base géographique sont décrits dans la base sémantique. Cette dernière est gérée au sein d'un Système de Gestion de Base de Données Relationnelle. La spécificité de la couverture pédologique a conduit à concevoir un modèle de données original dénommé DONESOL.

Historique

DONESOL est la structure commune permettant l'harmonisation de l'ensemble des bases de données géographiques des programmes d'inventaire (Legros, 1996). Ce modèle est né en 1992 de la fusion d'un premier prototype de description des unités cartographiques de sols avec le système de description ponctuelle des sols et des résultats d'analyses de laboratoire appelé STIPA « Système de Transfert de l'Information Pédologique et Agronomique » (Bertrand *et al.*, 1979). Il est à noter qu'historiquement, d'un point de vue terminologique, « DONESOL » désigne tout autant un modèle de données que le système d'information national basé à Orléans.

Sous la responsabilité de Hardy, des améliorations successives ont abouti, en 2001, à une nouvelle version appelée DONESOL 2. Elle s'est traduite par l'introduction de très nombreuses variables pour décrire les Unités Typologiques de Sol et leurs strates. Elle a également permis une plus grande souplesse dans l'affectation des profils pédologiques à différentes études cartographiques possédant une zone commune. Parallèlement à l'évolution du modèle, des interfaces de saisie, consultation et modification des données ont été développées à l'INRA et dans les régions partenaires. La qualité sémantique de la base a pu être assurée grâce à l'implémentation d'une multitude de contraintes d'intégrité au sein des différents composants applicatifs. Ces contraintes, si nécessaires mais si difficiles à élaborer dans le cadre d'une modélisation d'objets environnementaux, ont contribué à la création d'une véritable « logique métier » au cœur du système d'information.

Dans un souci d'ergonomie et de standardisation, la base de surveillance des sols stockant les données du Réseau National de Mesures de la Qualité des Sols (Arrouays *et al.*, 2002) a été « greffée » au modèle DONESOL en 2002, finalisant ainsi un système d'in-

formation sur les sols capable de répondre aux interrogations d'ordre qualitatif, quantitatif, spatial ou temporel.

Le modèle de données DONESOL

Comment identifier un type de sol de manière unique dans une base de données sémantique ? Cette interrogation sous-entend en réalité une question de terminologie : qu'appelle-t-on un sol ? La formalisation du concept de type de sol trouve une réponse dans DONESOL au sein de l'entité appelée « Unité Typologique de Sol (UTS) » (Gaultier *et al.*, 1993). Cette entité regroupe les caractéristiques générales d'un type de sol telles que son nom, son organisation géologique, sa différenciation, etc. Toutefois cette description se heurte à un problème majeur, celui de l'échelle à laquelle est réalisée la caractérisation. En effet, on ne décrit pas une couverture pédologique sur une carte au 1/250 000^e comme sur une carte au 1/10 000^e. Pourtant DONESOL se veut une base multi-échelle permettant de comparer les types de sol entre eux. La notion de « Strates » permet de résoudre ce problème en décrivant de manière analytique et statistique un ou plusieurs horizons d'un type de sol (Gaultier *et al.*, 1993). En effet, la comparaison de variables peut s'affranchir de l'échelle d'étude sous réserve que les statistiques sommaires (moyenne, variance, distribution) des différents échantillons concernés soient comparables.

Une des spécificités de DONESOL est de permettre la gestion d'unités cartographiques complexes. Une Unité Cartographique de Sol donnée peut être constituée d'une ou plusieurs Unités Typologiques de Sol, dont on décrit les pourcentages et le mode de répartition, qui sont eux-mêmes décrits dans des tables spécifiques. L'emboîtement et les liaisons des tables permettent de décrire l'organisation et la complexité de la couverture pédologique, depuis l'unité pédopaysagère, jusqu'aux caractéristiques morphologiques et analytiques des horizons. Le modèle de données DONESOL 2 est présenté sous forme d'un diagramme de classes UML en *figure 1*.

Le programme Inventaire Gestion et Conservation des Sols (IGCS) est le programme cadre des études de cartographie pédologique en France à petites et moyennes échelles.

Le Réseau de Mesures de la Qualité des Sols (RMQS) correspond à la mise en place d'un réseau national de surveillance de la qualité des sols.

L'ajout du modèle de données du RMQS à DONESOL (Thorette *et al.*, 2001) a constitué une petite révolution au sein du système d'information dans la mesure où l'aspect temporel est venu se greffer à la dimension spatiale existante.

Au final le système d'information national sur les données pédologiques spatialisées a pour objectif d'assurer les fonctions suivantes :

- fonction de stockage des données de cartographie pédologique variant dans le temps et/ou dans l'espace ;
- fonction d'interrogation et de comparaison de données brutes ;
- fonction d'interrogation et de comparaison de données synthétiques (agrégation dans le temps ou dans l'espace) ;

- fonction de modélisation (possibilités d'appliquer un modèle ou une fonction de pédotransfert sur tout ou partie de la base de données).

Perspectives et évolution

L'évolution spectaculaire des nouvelles technologies autorise maintenant le stockage d'énormes quantités d'informations sous forme numérique. Les données stockées peuvent être textuelles ou codifiées, voire plus complexes (photographie, entité géo-référencée). La cartographie des sols, à l'image de la cartographie géologique, se doit d'évoluer afin de tirer profit de ces nouvelles techniques.

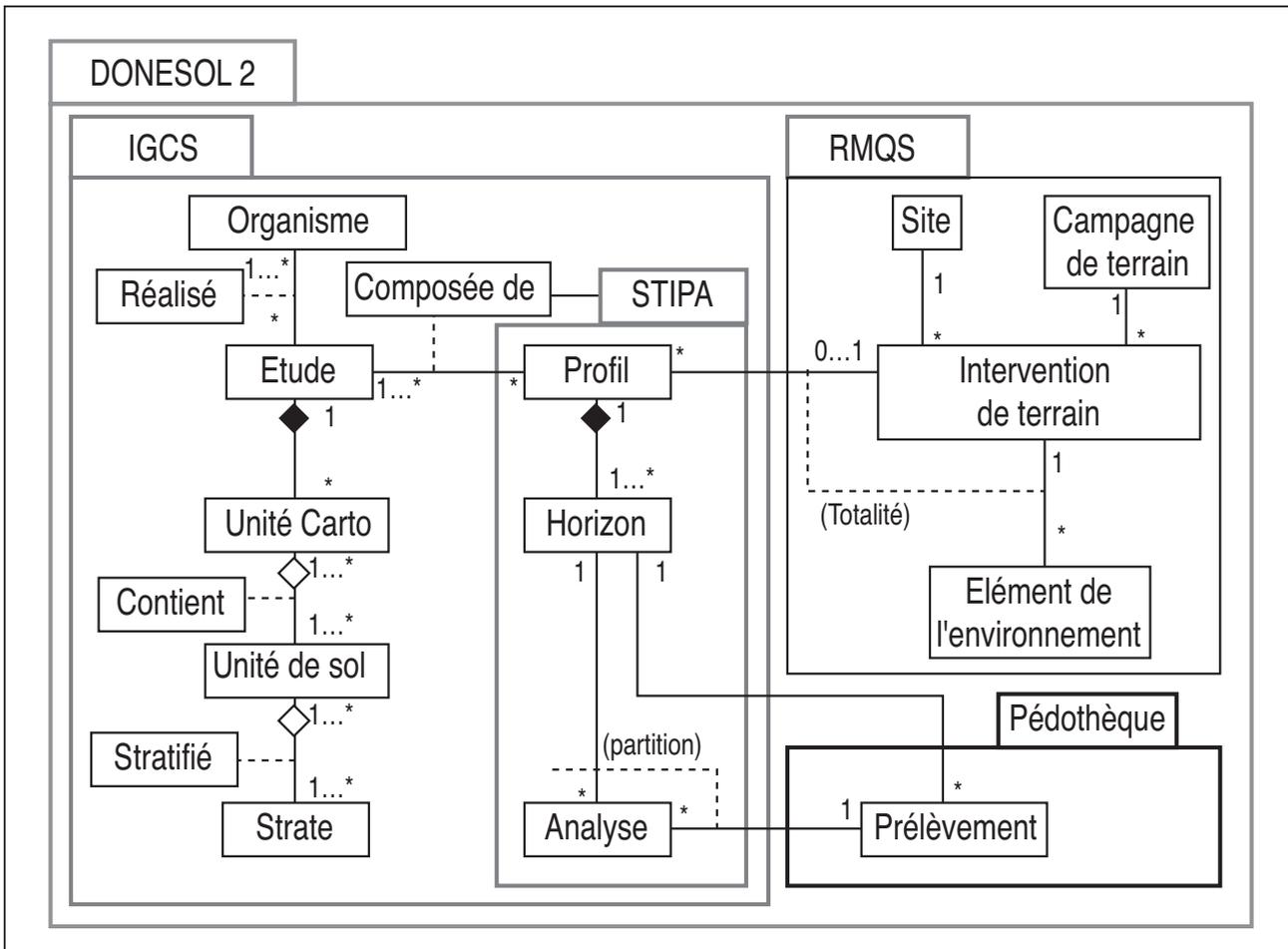
Actuellement, la connaissance de la répartition des sols sur le territoire national est nettement en retard par rapport à d'autres domaines tels que la géologie. Cette situation est handicapante pour l'amélioration des systèmes d'information sur les sols dans la mesure où la grande majorité des crédits alloués est avant tout destinée à l'acquisition de nouvelles données pédologiques. Mais ce contexte historique constitue aussi un avantage car les systèmes informatiques existants sont encore à des stades de développement leur permettant aisément d'évoluer vers une harmonisation nationale.

Au cours des cinq dernières années, de nombreuses applications fonctionnant autour du modèle DONESOL ont vu le jour en région. Le développement de ces outils a été progressif mais il a maintenant atteint un niveau suffisant pour que l'on s'intéresse de près au retour de ces expériences. Chaque application a nécessité de longues heures de codage et il apparaît donc nécessaire d'envisager une mutualisation des développements. C'est dans ce but qu'un groupe de travail transversal a été créé au sein du programme Inventaire Gestion et Conservation des Sols. Ce groupe de travail doit réunir autour d'une même table tous les acteurs de la conception d'outils dédiés à DONESOL, qu'ils proviennent de l'INRA ou des organismes régionaux partenaires. La mise en commun des outils réalisés doit respecter le travail de chacun tout en tirant avantage des innovations apportées. Cet objectif passe d'abord par un premier bilan de l'existant. C'est ce que nous allons tenter de réaliser ci-dessous en proposant une classification simple des différents outils développés autour de DONESOL.

LES OUTILS

Les outils d'aide à la saisie

La base de données DONESOL a une vocation nationale. Toutefois, les programmes d'inventaire et d'observation de la qualité des sols en France font appel à de multiples partenaires (chambres d'agriculture, écoles nationales supérieures d'agriculture, conseils régionaux ou généraux, etc.). Aussi, la nécessité de disposer de bases de données délocalisées en région est rapidement apparue. Dans un souci d'harmonisation l'INRA d'Orléans, sous l'impulsion de Hardy, a décidé de mettre à disposition de ses partenaires

Figure 1 - Diagramme de classes UML du modèle de données DONESOL 2**Figure 1** - UML class diagram of the DONESOL 2 data model

Le programme Inventaire Gestion et Conservation des Sols (IGCS) est le programme cadre des études de cartographie pédologique en France à petites et moyennes échelles.

Le Réseau de Mesures de la Qualité des Sols (RMQS) correspond à la mise en place d'un réseau national de surveillance de la qualité des sols.

régionaux une interface de saisie pour poste PC appelée « DONESOL PC ». Ce logiciel a l'avantage d'avoir été entièrement développé au sein de l'unité INFOSOL de l'INRA ce qui le rend entièrement compatible avec le modèle de données DONESOL 2. Le logiciel évolue de manière transparente pour les partenaires en fonction des variations du modèle de données.

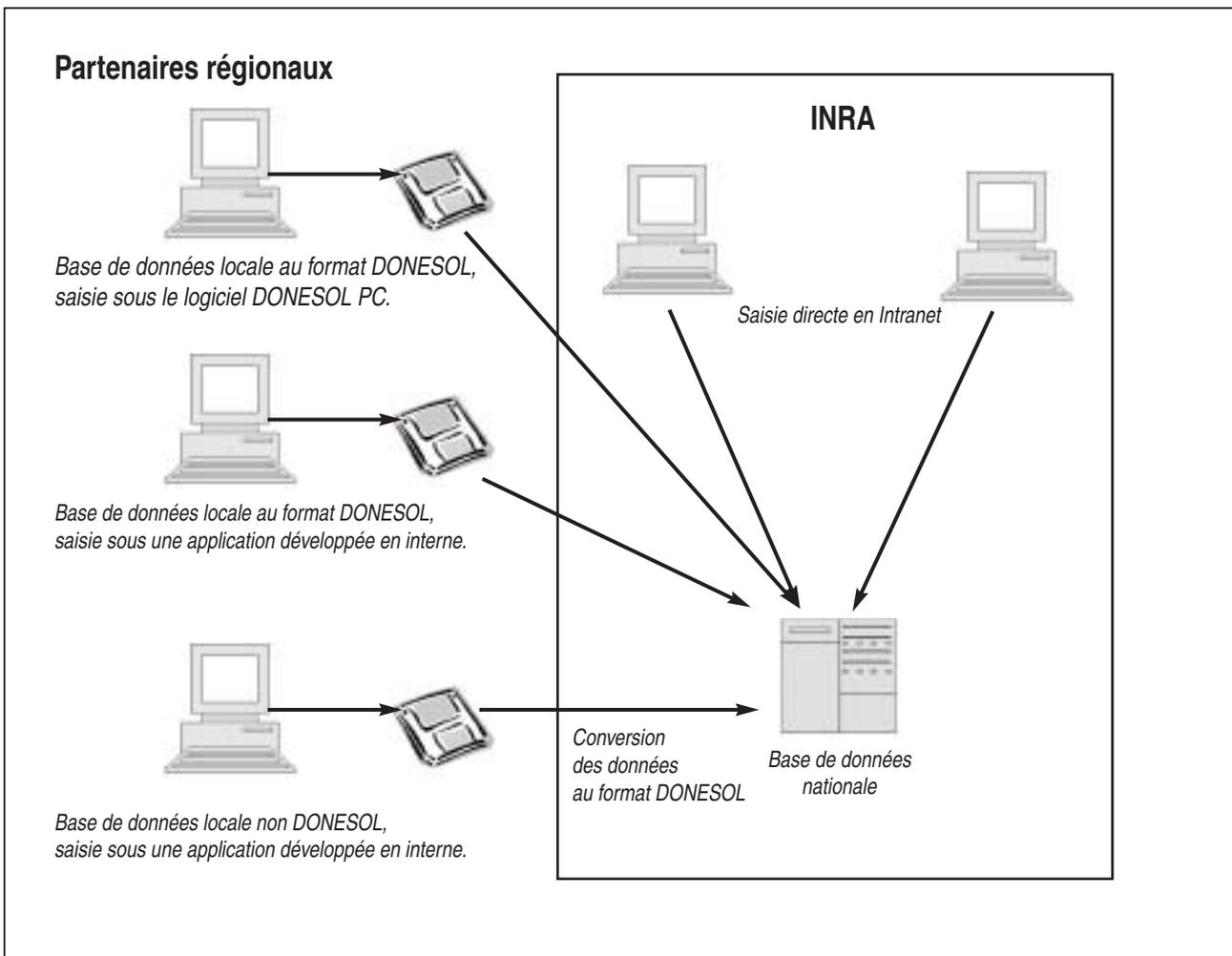
Ainsi, chaque partenaire saisit ses propres données pédologiques en local à l'aide du logiciel et fournit régulièrement un fichier à l'INRA d'Orléans afin d'alimenter la base nationale.

Au sein de l'INRA, une interface Intranet permet la saisie des données directement dans la base de données nationale. Les différentes possibilités de saisie sont résumées en *figure 2*.

Afin de permettre une saisie plus conviviale pour le pédologue, l'INRA de Montpellier a développé en 2000 une application de sai-

sie des entités profils, horizons, analyses (Falipou *et al.*, 2002). Ces objets correspondent au modèle STIPA intégré dans DONESOL en 1992. L'application, entièrement compatible DONESOL 2, possède également une interface spécialement conçue pour la saisie sur le terrain au moyen d'un PC portable. Cet outil a été baptisé « STIPA 2000 » et fonctionne sous Access.

Certains partenaires régionaux ont également développé leurs propres applications de saisie. C'est notamment le cas du CNER-TA de Dijon (Centre National d'Études et de Ressources en Technologie Avancée). Depuis trois années, ce département de l'ENESAD (Établissement National d'Enseignement Supérieur Agronomique de Dijon), maître d'oeuvre dans la réalisation du programme IGCS en Bourgogne, a développé un outil de saisie embarqué de description des fosses et sondages pédologiques (*figu-*

Figure 2 - Les différentes possibilités de saisie de données dans la base de données nationale**Figure 2** - Pathways for data input into the national database

re 3). Cette solution contribue notamment, à la diminution des erreurs de saisie, à la diminution du temps de saisie, et globalement à la rationalisation des flux de données entre le terrain et la base de données régionale.

Cet outil permet au pédologue, lors de la description des sondages et des fosses, de saisir l'information sur un mini-ordinateur de terrain. Les avantages de cette solution sont multiples :

- une localisation au GPS systématique. En effet, le mini-ordinateur est relié à un GPS qui permet une localisation du point de mesure avec une précision de l'ordre du décimètre. On notera que la simplicité de la mesure permet de saisir l'ensemble des points de sondage à la tarière ;
- une normalisation du processus de saisie. L'ensemble des modalités apparaît dans des listes déroulantes, le pédologue n'a ainsi plus à les mémoriser (figure 4).

- une phase de saisie plus courte. Même si sur le terrain l'utilisation de l'appareil augmente légèrement la durée de la description, si on tient compte de la durée totale du processus (description + saisie), le temps gagné est important. La collecte des données issues d'une campagne de terrain, opération qui consiste à effectuer le transfert des données du mini-ordinateur vers la base de données stockée sur un serveur, est effectuée par un programme informatique. Pendant cette opération, le programme vérifie la cohérence des données selon certains critères : types (entier, caractère, date), ou longueur (nombre de caractères, taille des nombres...).

La réalisation de cet outil n'a pas été possible sans tenir compte d'un certain nombre de contraintes, matérielles et logicielles. Le mini-ordinateur de terrain doit être suffisamment autonome pour une journée complète, robuste et étanche pour fonctionner dans les conditions d'observation des fosses (humidité, boue, chocs), et avoir un

Figure 3 - Description d'une fosse pédologique au moyen du système de saisie embarqué.

Figure 3 - Describing a soil profile with a portable computer-based data capture system.



écran suffisamment confortable (luminosité et taille). Le logiciel retenu est Arcpad 6, de la société ESRI, capable d'une part de gérer des objets géographiques dans un système de projection adapté à la zone d'étude, d'autre part d'associer à la capture d'un point via le GPS, un formulaire de saisie. Enfin, la taille des formulaires, le nombre de variables, leur codage et la structure des données dans Arcpad 6 ont nécessité un important travail de préparation informatique.

Même si l'état actuel des développements apporte de nombreuses améliorations dans la phase de collecte de la donnée pédologique, plusieurs évolutions sont envisagées :

- la saisie des horizons aux besoins du pédologue en affichant uniquement les variables souhaitées dans l'ordre souhaité ;
- une meilleure exploitation des propriétés du logiciel Arcpad (ou d'un autre logiciel) afin d'étendre la qualité des contrôles effectués sur les valeurs saisies et diminuer encore les erreurs de saisie ;
- l'embarquement des fonds de cartes utiles à la détermination des types de sol (photographies aériennes, cartes géologiques, "scans 25", esquisses cartographiques), sous réserve de l'investissement dans un mini-ordinateur de terrain plus puissant.

Avec cet outil, et au prix d'un effort d'adaptation certain, le pédologue s'affranchit ensuite de la saisie de ces données tout en conservant le maximum d'informations spatiales. Les procédures d'importation associées à l'outil de saisie génèrent un système fonctionnel qui peut être transféré sur d'autres problématiques et d'autres projets. Enfin, dans le cas d'approches du type "retour à la parcelle", ces outils permettraient de confronter directement les observations de terrain avec le contenu de la base de données régionale.

Les outils de contrôle des données

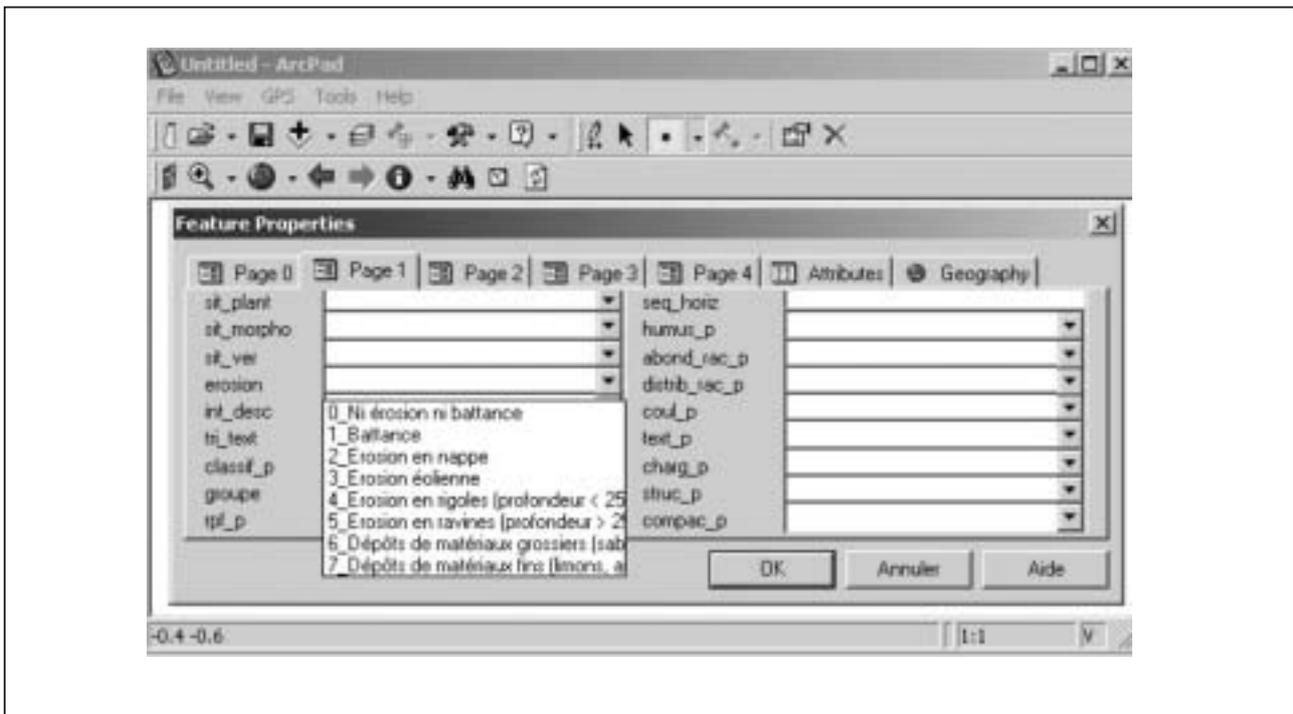
La saisie informatique des informations pédologiques peut être réalisée soit par le pédologue lui-même, soit par des opérateurs de saisie ayant peu ou pas de connaissances en sciences du sol. Le premier cas est assez rare car, bien souvent, cette tâche apparaît comme très contraignante au pédologue dont les compétences l'amènent à préférer les longues heures de terrain plutôt que les longues heures devant un écran d'ordinateur. Les pédologues nouvellement diplômés ont cependant bien intégré cet aspect fondamental de la constitution d'une base de données et semblent plus enclins à réaliser cette tâche fastidieuse.

Dans l'hypothèse d'une saisie par un opérateur extérieur, on conçoit aisément que le nombre d'erreurs de saisie puisse être non négligeable. Il convient donc de contrôler scrupuleusement les données.

On distingue plusieurs types de vérification :

- la vérification visuelle par le pédologue en charge de l'étude. Le pédologue réalise une sortie papier des valeurs saisies afin de vérifier la cohérence des données. Cette méthode, bien que très efficace, reste extrêmement coûteuse en temps ;
- l'automatisation des vérifications par la création de scripts informatiques de contrôle. Les scripts peuvent être implémentés directement au sein de l'application de saisie « validation *a priori* » (cas du logiciel DONESOL PC) ou exécutés « *a posteriori* » sur la base de données.

Dans le cadre des programmes d'inventaire pédologique, des contrôles ont été mis en place à différents niveaux de la chaîne de réalisation des bases de données cartographiques pour finalement aboutir à une véritable procédure de validation (conçue à l'initiative du secrétariat technique IGCS). Celle-ci permet au ministère en charge de l'agriculture d'attribuer à la base de données cartographiques (base sémantique et base graphique) un niveau de qualité allant de 1 à 3. Actuellement, cette procédure concerne uniquement les Référentiels Régionaux Pédologiques (RRP) mais elle

Figure 4 - Formulaire de saisie embarquée*Figure 4* - A data capture form for soil description

devrait prochainement être étendue au programme Connaissance Pédologique de France (figure 5).

Le contrôle de la qualité des bases de données est réalisé lors de l'étape 4. Les procédures de vérification sont décrites dans le tableau 1.

Jusqu'à présent tous les outils de validation des données ont été développés au sein de l'unité INFOSOL de l'INRA. En effet, en région, les maîtres d'ouvrage n'ont pas vocation à financer ce type de développements qui nécessitent à la fois une expertise en pédologie et de bonnes connaissances en programmation et bases de données. L'INRA doit donc veiller à mettre à disposition de ses partenaires le résultat de ses travaux afin de ne pas assister à la prolifération de bases de données non validées en région.

Les outils de consultation

L'un des objectifs principaux de la mise en place d'un système d'information est de permettre la consultation des données à un grand nombre d'utilisateurs afin de répondre à leurs diverses interrogations. Une fois la base de données constituée, il peut sembler trivial de développer la couche logiciel de consultation, pourtant il n'en est rien. En effet, l'expérience nous a montré que chaque utilisateur potentiel de la base souhaite interroger celle-ci de manière très spécifique. Il est donc très difficile de concevoir des applications de consultation génériques répondant à toutes les demandes. De plus, les connais-

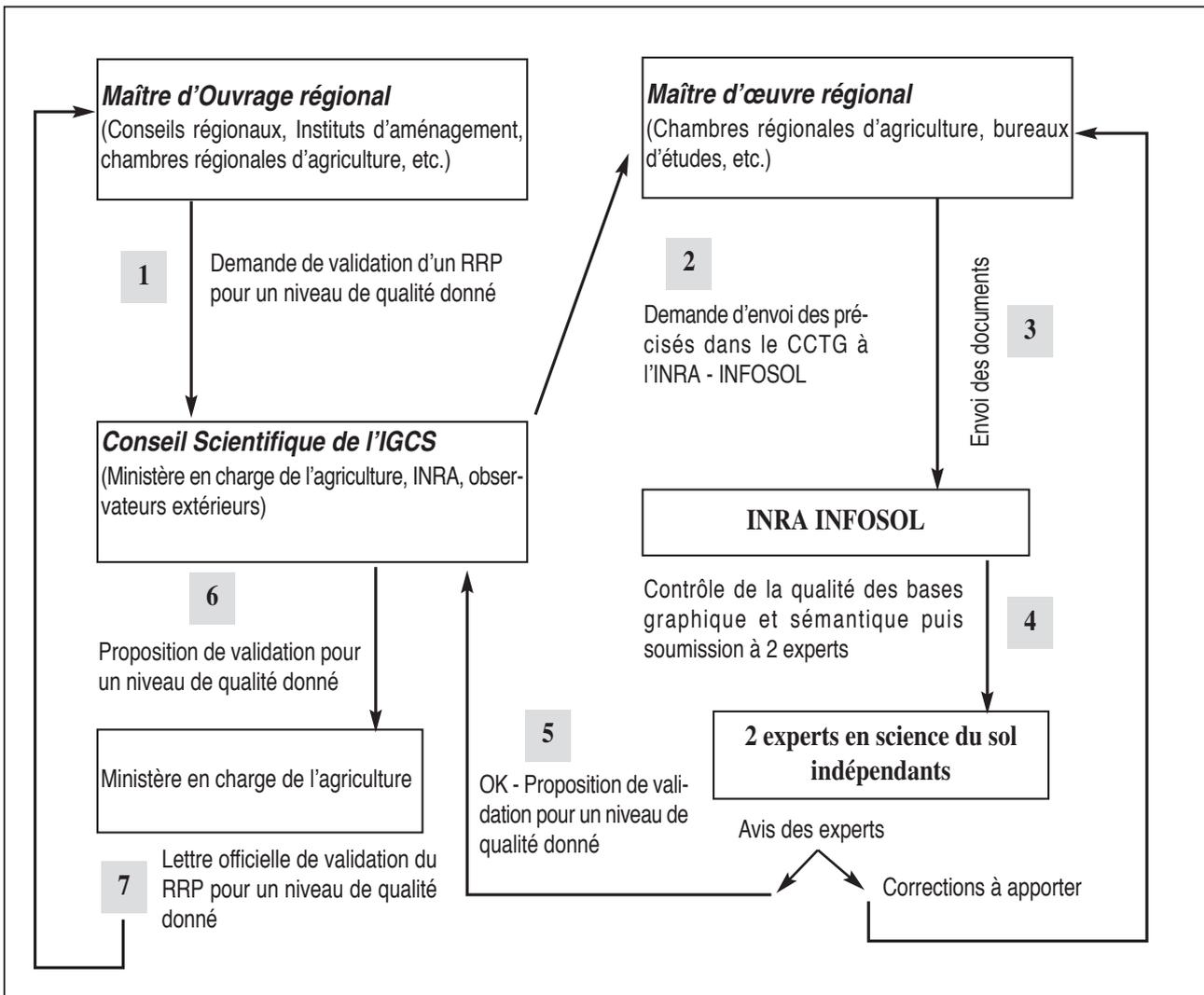
sances informatiques des utilisateurs finaux sont souvent très hétérogènes. Dans ce contexte, il convient donc de formaliser très clairement l'objectif et la cible de chaque outil avant la phase de développement (figure 6). Sans cette précaution on risque fort d'aboutir à des applications peu utilisées.

REPEDO et REFERSOLS

Internet étant, par essence, le réseau où l'hétérogénéité en matière de connaissances informatiques des utilisateurs est la plus forte, un outil reposant sur ce média se situera invariablement dans les cas (1) ou (2). L'INRA d'Orléans a ainsi développé un site Internet de consultation des coordonnées des organismes intervenant en pédologie « REPEDO » ainsi qu'un répertoire des études de sol existantes sur le territoire « REFERSOLS » (figure 7).

REPEDO fait apparaître des données brutes puisqu'elles sont stockées telles quelles dans le système d'information. REPEDO se situe donc dans le cas (1). En revanche, REFERSOLS fait appel à des requêtes plus complexes. Cet outil se situe donc plutôt dans le cas (2). De manière générale, les outils de consultation grand public en cartographie des sols se positionnent presque exclusivement dans le cas (2). En effet, ceux-ci utilisent des techniques d'affichage de contours cartographiques nécessitant la combinaison entre des informations issues d'une base géographique et des informations issues d'une base sémantique.

Figure 5 - Procédure pour l'obtention de la validation d'un Référentiel Pédologique pour un niveau de qualité donné
Figure 5 - The procedure for assessing the quality of regional pedological surveys



Applications développées en région

En région plusieurs applications d'interrogation de base de données reposant sur le modèle DONESOL ont été développées ou sont en projet. Ainsi, le Système d'Information Régional Poitou-Charentes (<http://www.sir-poitou-charentes.org/>), à l'initiative du conseil régional et de l'Etat et grâce aux développements réalisés conjointement par l'Institut Atlantique d'Aménagement des Territoires (IAAT) et la Chambre Régionale d'Agriculture, offre la possibilité de consulter des données sols à partir d'une thématique « milieux physiques ».

En premier lieu, un répertoire des études de sols en Poitou-Charentes est accessible via une recherche cartographique ou par mots-clés. Il est ainsi possible de visualiser l'ensemble des études réalisées sur le territoire à différentes échelles et d'accéder aux prin-

cipales « meta-données ». Des informations précises peuvent ensuite être obtenues auprès de la Chambre Régionale d'Agriculture. Un second outil met à disposition, à partir d'une base de données géographiques découpée en bassins versants, quelques informations sur les différentes unités cartographiques répondant à une thématique « infiltration et transfert de l'eau dans le sol ». Les données sont issues en partie de la base de données DONESOL mais également de données « à dire d'expert ». Il est par ailleurs envisagé de rendre accessible une version allégée de DONESOL dans le cadre d'un projet régional de plateforme informatique. Cet outil devrait permettre à un nombre restreint d'utilisateurs de consulter des informations concernant les Unités Cartographiques et Typologiques de Sols de la région, en complément de la mise à disposition de la base de données géographiques.

Tableau 1 - Procédures de contrôle de la qualité des bases de données d'un Référentiel Régional Pédologique.**Table 1** - The quality control procedures for regional pedological survey databases.

Contrôle de la base graphique par SIG et par des experts pédologues
<p><u>Contrôle du respect des critères de qualité en adéquation avec le niveau de qualité visé.</u></p> <p>La localisation géographique des sites décrits et des sites décrits et analysés permet de calculer leur nombre/ha et de contrôler et visualiser leur répartition spatiale.</p> <p><u>Contrôle des Unités Cartographiques de Sols (UCS)</u></p> <ul style="list-style-type: none"> - vérification du calage géographique des limites des UCS par comparaison avec les bases de données de l'IGN (topographique et hydrographique) ; - calcul de la taille des UCS (superficie et forme). Elle doit être compatible avec l'échelle du 1/250 000 ème ; - appréciation du tracé des limites des UCS : il doit être homogène sur toute la carte ; - vérification du nombre total d'UCS (à titre indicatif il est en moyenne compris entre 60 et 100 pour un département de 600 000 ha).
Contrôle de la base sémantique par traitements informatiques et par des experts pédologues
<p><u>Contrôle de la présence de données pour les variables requises par le CCTG (pour chaque table de la base de données DONESOL)</u></p> <ul style="list-style-type: none"> - variables obligatoires (toujours présentes quel que soit le type de sol) ; - variables conjoncturelles (à renseigner pour certains types de sol). <p><u>Contrôle de la qualité des données</u></p> <ul style="list-style-type: none"> - programmes informatiques permettant d'interdire, dès la mise en base, la saisie de données erronées ou dépendantes de variables « en amont » (exemple : la somme des pourcentages Argiles + Limons + Sables ne doit pas être supérieure à 100%) ; - programmes informatiques permettant de vérifier l'adéquation de la donnée avec les « normes » imposées par le manuel de classification des sols, (exemple : pour le Référentiel Pédologique Français un Calcosol hypo-calcaire doit avoir moins de 15% de calcaire total dans l'horizon Sca, etc.) ; - actuellement l'étude de la cohérence des données entre elles et donc de leur qualité est essentiellement réalisée par les experts pédologues.

Figure 6 - Positionnement d'un outil de consultation des données en fonction de la complexité de la requête et des connaissances informatiques de l'utilisateur

Figure 6 - Positioning of a data request tool as a function of the complexity of the request and the pre-existing informatic knowledge of the user

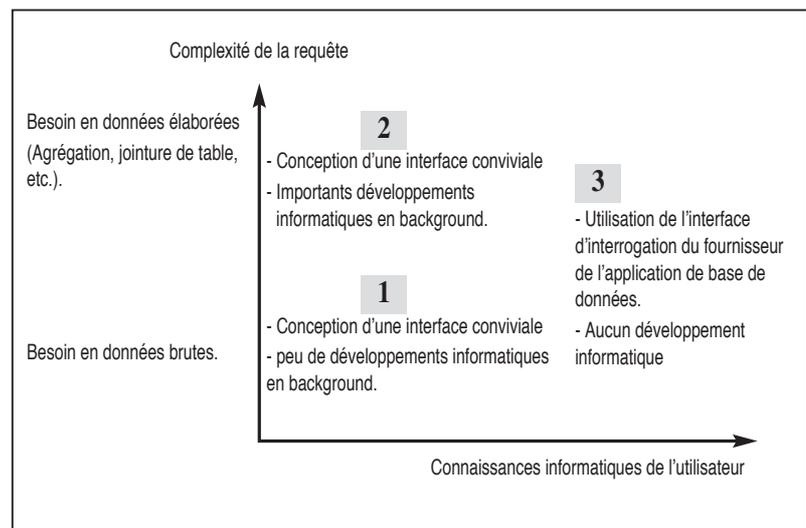


Figure 7 - Page de recherche d'un organisme sur le site Internet REPEDO

Figure 7 - Company search form on the web site REPEDO

Dans un même esprit, la région Provence-Alpes-Côte d'Azur par l'intermédiaire de son Centre Régional de l'Information Géographique (CRIGE-PACA) a mis en place un portail Internet permettant l'accès aux différentes bases de données cartographiques de la région (<http://www.crig-paca.org>). Dans ce cadre les données pédologiques du programme IGCS qui ont été mises à disposition sont :

- l'inventaire des études pédologiques de la région PACA au format MapInfo® et au format image (atlas cartographique) ;
- la carte des pédopaysages de la région PACA au format image ainsi que la description des UCS associées au format Microsoft Word® ;
- la base de données DONESOL du département des Bouches-du-Rhône au format Microsoft Access® et MapInfo (partie réservée aux membres du CRIGE).

Ce type de portails, ou de plateformes informatiques, tend à se développer largement car il reste relativement peu coûteux comparé à des développements plus lourds tels que ceux entrepris par la région Rhône-Alpes ou la région Bourgogne. Ces deux régions ont, en effet, souhaité mettre à la disposition des professionnels et du grand public des applications de consultation des données IGCS répondant aux requêtes complexes et variées des utilisateurs.

L'application SIRA

L'association des Chambres d'Agriculture de Rhône-Alpes (SIRA : Sols Info Rhône-Alpes) a mis en ligne en février 2004 un site Internet de consultation des Unités Cartographiques et des Unités Typologiques de Sols issues des cartes pédologiques du pro-

gramme IGCS sur la région (<http://www.rhone-alpes.chambagri.fr/sira/>). Elle a souhaité exploiter les possibilités qu'apportent l'utilisation des nouvelles technologies et des serveurs cartographiques : le Webmapping. Le choix s'est préférentiellement tourné vers des moyens « Open-source », tels que MapServer ou PostgreSQL, qui semblent posséder des capacités identiques aux solutions commerciales pour un coût de mise en place réduit.

L'application SIRA : principe de fonctionnement

La configuration de MapServer est malléable. Cet outil peut en effet être intégré sur différentes plates-formes, fonctionner comme un moteur cartographique pour des serveurs divers et interagir avec des langages variés. La configuration adoptée pour l'application cartographique SIRA est la suivante :

- « Mapserver », produit libre de visualisation de cartes interactives sur Internet, développé par l'Université du Minnesota, dont les bibliothèques additionnelles permettent de lire certains formats de fichiers (notamment les couvertures géographiques) ou de fabriquer des images qui pourront être renvoyées au navigateur Internet client ;
- une plate-forme Linux (Debian) ;
- un serveur Web « Apache » ;
- un serveur d'application « PHP » pour générer des pages Web dynamiques ;
- un module « MapScript » pour construire et manipuler des objets cartographiques ;
- et un logiciel SGBD (Système de Gestion de Bases de Données) de type relationnel-objet : « PostgreSQL » renforcé par le module « PostGIS » pour le stockage des objets géographiques 2D ou 3D.

Tous ces produits sont libres et largement reconnus, c'est pourquoi l'INRA d'Orléans, forte de l'expérience du SIRA en Rhône-Alpes et du CNERTA de Dijon, s'oriente également vers ce type de configuration.

L'application SIRA : interfaces de consultation

Pour son serveur cartographique SIRA propose deux interfaces : une interface principale développée autour de Macromedia Flash® ; une interface en HTML proposée comme solution de replis en cas d'absence du plug-in Flash. Ces deux interfaces proposent à l'utilisateur les mêmes fonctionnalités. L'utilisateur aura la possibilité de (figure 8) :

- visualiser ou non certaines couches ;
- savoir si les couches sont visibles à l'échelle du zoom actuel ;
- effectuer des opérations de zoom classique (agrandir, éloigner, déplacer) ;
- se focaliser directement sur une commune ;
- visualiser l'échelle de la carte ;
- visualiser une carte de référence afin de se situer plus facilement ;
- afficher la légende complète des couches ;
- effectuer des requêtes spatiales simples en cliquant sur la carte

des sols pour interroger les unités cartographiques, renvoyer le nombre de résultats, et obtenir plus de détail sous forme de « fiches sols » créées à la volée ;

- élaborer une mise en page de la zone visualisée à l'écran afin de l'imprimer ou d'en faire une sauvegarde au format PDF.

L'application SIRA : les données utilisées

Enfin, certaines couvertures ou données géographiques sont utiles pour aider l'utilisateur à se situer sur la carte. Il s'agit de points de repère tels que :

- les limites de départements extraits de la base IGN GéoFla® ;
- les fonds cartographiques à 1/100 000 ème de l'IGN ;
- le modèle numérique de terrain radar au pas de 80 m (libre de droit) pour l'aspect du relief.

Les données pédologiques SIRA sont issues du programme IGCS. Certaines sont directement disponibles dans les bases de données SIRA, d'autres nécessitent quelques traitements supplémentaires.

L'application du CNERTA

L'application de consultation la plus élaborée autour de DONE-SOL est certainement celle développée par le CNERTA de Dijon (Bargeot *et al.*, 2003). En région Bourgogne, la donnée pédologique est acquise sur des fonds publics. Les choix politiques de diffusion de cette donnée sont clairs. La donnée publique doit rester accessible au plus grand nombre et à un coût raisonnable, lié à sa diffusion. De ces choix découlent des orientations techniques comme la diffusion des bases de données IGCS régionales par Internet. L'ENESAD, dans un souci de souplesse des développements, de maîtrise des dépenses, et d'interopérabilité du système d'information géographique Sols et Territoire de Bourgogne avec d'autres systèmes d'information, a choisi l'alternative des logiciels libres.

L'application du CNERTA : les caractéristiques techniques

Comme pour la région Rhône-Alpes, la base de données sémantique est gérée par le système de gestion de base de données PostgreSQL. Les couches graphiques, issues de la digitalisation des pédopaysages à l'aide de SIG du marché, sont importées dans la cartouche spatiale du SGBD : PostGIS. L'utilisation de cette cartouche spatiale permet une interaction avec le module de représentation cartographique du serveur Web : Mapserver. Des modules de schématisation des strates à partir du contenu de la base de données (épaisseur, couleur) ont également été développés.

L'application du CNERTA : les fonctionnalités

Après avoir consulté quelques utilisateurs, un ensemble de fonctionnalités a été retenu pour le développement. De cette analyse des besoins, découlent deux approches distinctes.

Une approche exploratoire

Un premier module, baptisé CARTOSOL, permet la navigation dans un département et l'interrogation de la base de données. Les fonctions de navigation autorisent l'utilisateur à se déplacer et à « zoo-

mer » dans les limites acceptables de l'échelle de cette base de données.

Les fonctions d'interrogation permettent la sélection des pédopaysages, pour accéder dans un premier temps à une description succincte (*figure 9*), puis à toutes les données relatives à l'objet sélectionné (*figure 10*).

Cette partie a fait l'objet d'un travail important. Elle a notamment nécessité la mobilisation de méta-données pour la mise en forme (regroupement, décodage, unités de mesures) de toutes les informations contenues dans la base (UCS, UTS, Strates, Profils, Horizons, Analyses). Dans ce module sont incluses la représentation schématique des profils, la représentation schématique des strates, et l'extraction des photographies si elles existent (profils ou UCS). Au final, ces informations sont également disponibles au format PDF pour faciliter l'impression. L'outil d'interrogation cartographique permet d'explorer la base de données et de prendre connaissance de la diversité des sols de la région. Il a également un rôle pédagogique puisqu'il facilite la compréhension du modèle de représentation de la couverture pédologique à cette échelle. Mais il est insuffisant pour répondre à des demandes spécifiques, par exemple, dans le cas où l'utilisateur doit insérer la donnée pédologique dans un modèle de traitement nécessitant la comparaison avec d'autres données dans un SIG.

Interrogation multicritère

Ce module permet une interrogation multicritère pour agréger l'information à l'échelle du pédopaysage. Les méthodes d'agrégation retenues sont celles de la dominante ou d'une requête précise sur plusieurs variables pédologiques. On peut par exemple extraire :

- la modalité dominante par pédopaysage du type de structures observées dans les strates superficielles (une information sur la pureté du pédopaysage est également disponible) ;
- le pourcentage de types de sol par pédopaysage qui ont une strate superficielle avec une effervescence moyenne et un pourcentage d'éléments grossiers élevé.

Les résultats sont alors disponibles, soit sous forme de données attributaires, l'utilisateur peut alors les télécharger pour les intégrer

Figure 10 - Description d'une Unité Typologique de Sol

Figure 10 - Description of a Soil Typological Unit

Type de forme morphologique Terrasse	
Pente minimale	0 %
Pente moyenne	1 %
Pente maximale	3 %
Type d'occupation 1	Champs : plantes herbacées, plantes cultivées, légumes
Type d'occupation 2	Forêts semi-naturelles, forêts secondaires
Type d'occupation 2	Cultures annuelles

dans son SIG, soit sous forme de cartes au format PDF pour les imprimer.

Perspectives

Le vecteur de diffusion qu'est Internet a de nombreux atouts. Il permet de rendre la donnée accessible, et il oblige à une vulgarisation des outils de consultation. Dans le cadre d'une mutualisation des efforts sur ces outils, d'autres développements sont envisageables. L'exploitation des fonctions de pédotransfert et certains modèles de traitements arborescents, pourraient être interfacés simplement pour donner aux utilisateurs le choix dans le paramétrage de ces modèles. Néanmoins, la mise à disposition de ces contenus publics sur le réseau devra s'accompagner de mécanismes d'authentification et de gestion des droits d'accès, en fonction des diverses politiques de diffusion régionales.

Les outils d'échange de données

Le modèle DONESOL se veut générique pour le stockage des données pédologiques spatialisées ; néanmoins, il possède des limites. C'est notamment le cas pour le stockage des études pédologiques réalisées hors des frontières de la métropole, le plus souvent par l'Institut de Recherche pour le Développement (IRD). L'approche de l'IRD, pour la description du milieu naturel est sensiblement différente. Elle s'appuie sur une organisation de ce milieu en grands-paysages, paysages, segments et pédon et ne prend pas en compte les parcelles et leur suivi agronomique. Le dictionnaire de données est également adapté aux régions intertropicales et à la description de ces types de sols et de leur environnement. Devant ce constat l'INRA d'Orléans et l'Unité de Service Valpédo de l'IRD, plus impliquée dans les bases de données pédologiques des territoires d'outre-mer, ont entrepris en 2004 un travail de conception d'un outil d'échange de données entre DONESOL et la base de données pédologiques de l'IRD appelée « VALSOL ». Une première version de cet outil, ne prenant en compte que les profils réels géoréférencés, a été développée et devrait être testée pour les sols de la Martinique. L'application finale, reposant sur des séquences de script, devrait permettre, à terme, un échange de données bidirectionnel entre DONESOL et VALSOL.

Ces dernières années, l'interconnexion des systèmes informatiques est devenue incontournable dans la plupart des organisations. Ainsi, un nouveau langage d'échange de données est apparu, le langage XML. Il s'agit du format universel d'échange de données qui s'impose lorsque des systèmes, hétérogènes ou non, doivent communiquer entre eux. Ce format a l'avantage de séparer la forme du contenu des documents. Dans le futur, il serait intéressant d'utiliser cette technique au sein du système d'information national sur les données pédologiques spatialisées. Ce projet nécessitera la conception d'un modèle de document DONESOL, de type DTD ou XML Schéma, puis la création de feuilles de style XSL pour le transfert des données entre DONESOL et d'autres systèmes d'informations.

CONCLUSION

Au cours des cinq dernières années, de nombreux outils ont été développés autour du système d'information national sur les données pédologiques spatialisées. Dans un premier temps, l'INRA d'Orléans s'est consacré à la conception d'applications de saisie afin de promouvoir le modèle de données DONESOL. Puis, quelques partenaires régionaux du programme IGCS (Bourgogne, Rhône-Alpes, PACA, Poitou-Charentes, etc.), grâce aux financements des conseils régionaux et de l'Etat, ont entrepris l'élaboration d'outils de consultation des données. Ces réalisations informatiques, qu'il s'agisse de sites Internet conviviaux et ergonomiques ou de plateformes informatiques, ont permis aux régions d'afficher une vitrine de leurs acquis en matière de données environnementales. Actuellement l'INRA d'Orléans se consacre essentiellement à la conception d'outils de validation et de transfert des données. Cette évolution sensible du nombre de réalisations techniques trouve sa source dans plusieurs facteurs. Tout d'abord, une prise de conscience croissante, de la part des maîtres d'ouvrages régionaux, du rôle fondamental du sol comme interface vis-à-vis de l'environnement et donc de la nécessité de mieux connaître ce milieu tant d'un point de vue quantitatif que qualitatif. Le second facteur clef fut sans doute la conception et la diffusion d'un modèle de données standard de description des données pédologiques spatialisées : DONESOL. Enfin la création en 2001 du Groupement d'Intérêt Scientifique SOL, regroupant les grands acteurs institutionnels nationaux (INRA, ADEME, IFEN, ministères chargés de l'agriculture et de l'environnement) a permis d'apporter une structure stable et une assise financière non négligeable pour la réalisation des différents programmes.

Devant l'ampleur des travaux réalisés dans les différentes régions, il apparaît maintenant nécessaire, voire fondamental, de fédérer tous ces efforts en favorisant largement une mutualisation des développements. C'est dans cet objectif qu'a été mis en place un groupe de réflexion sur les outils associés au modèle DONESOL. Ce groupe, réunissant les principaux acteurs régionaux et nationaux, devrait permettre un rassemblement des énergies et une meilleure coordination dans la conception des outils afin de relever le formidable challenge des années futures, à savoir la création d'un véritable système d'information trans-régional sur les sols.

BIBLIOGRAPHIE

- Arrouays D., Jolivet C., Boulonne L., Bodineau G., Saby N. et Grolleau E., 2002 - Une initiative nouvelle en France : la mise en place d'un réseau multi-institutionnel de mesures de la qualité des sols (RMQS). - Comptes Rendus de l'Académie d'Agriculture de France, 2002, 88, n° 5, pp 93-103. Séance du 5 juin 2002. France
- Bargeot L., Darracq S., 2003 - Comment améliorer l'utilisation des bases de données géographiques sur les sols : SIG et Internet. - Abstracts of The 3rd International Symposium on Digital Earth, page 92 - Czech Republic - September pp. 21-25, 2003
- Bertrand R., Falipou P., Legros J.P., 1979 - Notice pour l'entrée des descriptions et analyses de sols en banque de données. STIPA 1979. INRA - IRAT, Montpellier. 119 p.
- Boiffin J. et Stengel P., 2000 - Réapprendre le sol : nouvel enjeu pour l'agriculture et l'espace rural. - Déméter, Economie et stratégies agricoles. Armand Colin. pp 148-199.
- Bornand M., 1997 - Connaissance et suivi de la qualité des sols en France. Etat des lieux, enjeux, besoins en données, proposition pour une gestion raisonnée de la ressource en sol. Rapport d'expertise du ministère de l'agriculture, du ministère de l'environnement et de l'INRA. Montpellier 176 p.
- Bui E.N., Moran C.J., Henderson B.L., 2003 - Synthesis studies : how to make the most of existing soil data - Australian Soil and Land Survey Handbook - Volume 2 (Guidelines for Conducting Surveys - Eds AJ Ringrose-Voase, NJ McKenzie and M Grundy) - CSIRO Publishing : Melbourne - Australia
- Falipou P. et Legros J.P., 2002 - Le système STIPA-2000 d'entrée et d'édition des données pour la base nationale des sols DONESOL II - Etude et Gestion des Sols, Volume 9, 1, 2002, pp. 55 à 70 - France
- Gaultier J.P., Legros J.P., Bornand M., King D., Favrot J.C. et Hardy R., 1993 - L'organisation et la gestion des données pédologiques spatialisées : le projet DONESOL. Revue de Géomatique, 3, pp. 235-253 - France
- King D., Daroussin J. and Jamagne M., 1994 - Proposal for a Spatial Organization Model in Soil Science (The Example of the European Communities Soil Map) - Journal of the American Society for Information Science. Volume 45 (Issue 9) : pages 705-717, 1994 - John Wiley & Sons, Inc - New York, NY, USA
- King D., Jamagne M., Arrouays D., Bornand M., Favrot J.C., Hardy R., Le Bas C., Stengel P., 1999 - Inventaire cartographique et surveillance des sols en France. Etat d'avancement et exemples d'utilisation. Etude et Gestion des Sols 6, 4, pp. 215-228.
- Legros J.P., 1996 - Cartographies des sols - collection « Gérer l'environnement » - Presses Polytechniques et Universitaires Romandes - Suisse
- Thorette J., Grolleau E., Boulonne L. et Jolivet C., 2001 - Rapport pour le Ministère de l'Aménagement du Territoire et de l'Environnement - Etude de la structure d'une base de données pour le stockage des informations du Réseau de Mesures de la Qualité des sols (RMQS) - 32 p. - France

