

Les concepts de base des systèmes d'information géographique (SIG) : les données et les fonctions générales

Sylvie Bernier¹, Sylvie Duthoit², Sylvie Ladet³, Denis Baudet⁴

Résumé. La géomatique regroupe l'ensemble des outils et méthodes permettant d'acquérir, de représenter, d'analyser et d'intégrer les informations à référence spatiale issues de sources hétérogènes. Cette discipline, déjà largement développée au sein des Unités de recherche devient de plus en plus couramment utilisée par les Unités expérimentales de l'INRA, notamment par l'utilisation des systèmes d'information géographique (SIG). Dans cet article, nous revenons sur les bases nécessaires pour comprendre et utiliser un SIG. Un SIG permet de visualiser et d'analyser des données géographiques de deux formats principaux : le format raster (images satellites, photographies aériennes, cartes scannées, etc.) et le format vecteur qui correspond à une simplification du monde réel en objets graphiques auxquels sont reliées des informations descriptives plus ou moins complexes. Afin que toutes ces données soient superposables pour être analysées ensemble, elles doivent appartenir à un système de coordonnées ; c'est un système de référence permettant d'exprimer la position d'un objet dans ses deux ou trois dimensions. Toutes les données géographiques peuvent être regroupées en base des données spatialisées contenant toutes les informations essentielles pour une bonne utilisation (types de données utilisées, structure des données descriptives, source, date de création, etc.). Les fonctionnalités des SIG permettant l'exploitation des données géolocalisées (données issues de GPS, d'imageries satellitaires, de photographies aériennes, de données d'expérimentation...) sont nombreuses. Les principales sont présentées à travers des exemples issus de problématiques rencontrées à l'INRA et sur le logiciels SIG Quantum Gis (QGis : logiciel libre et gratuit). L'utilisation des SIG ne serait pas optimisée sans une bonne qualité des données de référence. Nous dressons ainsi un panorama des données géographiques des grands producteurs de données français, accessibles gratuitement ou à faible coût au monde de la recherche publique depuis ces dernières années. La politique de l'Europe et de la France est de développer le partage de l'information géographique à l'aide des nouvelles technologies sur Internet, sur mobiles et lié au GPS. L'ensemble de ces éléments permet d'envisager une facilité d'accès et de partage à l'information spatialisée que l'on peut également mettre en place à l'INRA.

Mots clés : géomatique, fonctionnalités des SIG, couches, raster/vecteur, système de projection, données géographiques

Introduction

Le terme géomatique - issu de la contraction des termes « géographie » et « informatique » - regroupe l'ensemble des outils et méthodes permettant d'acquérir, de représenter, d'analyser et d'intégrer les informations à référence spatiale issues de sources diverses : images satellitaires, photographies aériennes ou issues de drones, données cadastrales, données topographiques, relevés effectués sur le terrain ou issus d'observations sur des sites expérimentaux...). Cette information géographique est exploitée à l'aide de l'outil informatique système d'informations géographiques (SIG) qui permet de constituer des bases de données spatialisées (Collet, 2005).

L'utilisation des SIG, déjà largement développée au sein des Unités de recherche, devient de plus en plus courante dans une très grande variété d'applications (agriculture, écologie, environnement, aménagement...) notamment grâce à l'accessibilité de plus en plus importante des données géographiques et des logiciels (facilité d'accès aux données et coût d'utilisation des logiciels) .

Dans cet article nous allons revenir sur des concepts clés pour une utilisation optimale des SIG. Les fonctionnalités des SIG permettant l'exploitation des données géolocalisées seront présentées à travers des exemples issus de problématiques rencontrées à l'INRA, et traités avec les logiciel SIG : Quantum Gis (QGis, <http://www.qgis.org/fr/site/>),

1 Institut de Développement de la Géomatique (IDGEO), 42, Avenue du Général de Crouette, F-31100 Toulouse, France ; sylvie.bernier@idgeo.fr
2 UMR 1201 DYNAFOR, Ecole d'Ingénieurs de Purpan, 75 Voie du TOEC, BP57611, F-31076 Toulouse cedex 3, France
3 UMR 1201 DYNAFOR, INRA, 24 Chemin de Borde Rouge- Auzeville, CS 52627, F-31326 Castanet-Tolosan cedex, France
4 UMR 1248 AGIR, INRA, 24 Chemin de Borde Rouge- Auzeville, CS 52627, 31326 F-Castanet-Tolosan cedex, France

logiciel libre et open source, et le logiciel ArcGis for Desktop (<http://www.esrfrance.fr/bureautique.aspx>), logiciel commercial édité par ESRI, est également communément utilisé et des tarifs préférentiels sont accordés pour l'achat de licences dans le cadre de travaux de recherche et d'enseignement.

Nous dresserons aussi un panorama des données géographiques des grands producteurs de données français, accessibles gratuitement au monde de la recherche depuis ces dernières années. La politique de l'Europe et de la France est en effet de développer le partage de l'information géographique à l'aide des nouvelles technologies sur Internet, sur des appareils mobiles en lien avec le matériel GPS (Lahaye et Ladet, 2014 ; Lahaye et Ladet, 2014a ; Lahaye et Ladet, 2014b⁵). L'ensemble de ces éléments permet d'envisager une facilité d'accès et de partage de l'information spatialisée que l'on peut également mettre en place à l'INRA.

Les données dans un SIG : organisation et formats de l'information géographique

Organisation des données spatiales

Un SIG s'appuie sur un modèle d'informations géographiques basé sur des couches organisées de façon cohérente pour caractériser et décrire le monde. Par exemple, un SIG peut contenir des couches de données pour les éléments suivants : bâtiments, réseau hydrographique, parcelles agricoles, image satellite représentant la zone d'intérêt (**Figure 1**). L'information géographique contenue dans ces couches peut provenir de sources diverses ayant des formats différents, mais qu'un SIG peut gérer ensemble. Deux formats permettent de stocker les données spatiales : le format raster et le format vecteur.

Données au format raster

Les données raster correspondent aux images issues des satellites, des photographies aériennes, des images scannées (anciennes cartes papier) ou à des données continues représentant des phénomènes tels que la température ou l'altitude. Le mode raster permet de stocker l'information géographique dans des cellules contiguës généralement carrées. Ces cellules des images numériques sont donc l'unité élémentaire appelée pixel (contraction de *picture element*). Chaque pixel possède une valeur qui traduit une couleur (photographies ou scan), ou une énergie électromagnétique captée dans un domaine de longueur d'onde à partir d'un satellite (**Figure 1**). Par exemple les images du satellite SPOT5 peuvent proposer des images dans le domaine de longueur d'onde du proche infrarouge qui permettent de bien différencier la végétation active. Pour les données raster stockant des valeurs thématiques comme l'altitude, la température, le mode d'occupation du sol, etc., on ne parlera plus de pixel mais plutôt de cellules.

Une des propriétés des données raster qu'il est particulièrement important de connaître est la résolution spatiale c'est-à-dire la taille des pixels. Elle conditionne le niveau de détail qu'il est possible de discerner sur une image. Une image satellite à haute résolution spatiale (10 m de résolution pour SPOT5) permettra de repérer par exemple les grandes classes d'occupation du sol (cultures d'été, cultures d'hiver, forêts de feuillus, etc.) alors qu'une image à très haute résolution spatiale (autour de 1 m de résolution) permettra de distinguer finement les haies, les rangs de certaines cultures et même les arbres de façon individuelle. Les photographies aériennes proposées par l'IGN (BD ORTHO[®]) d'une résolution spatiale de 50 cm sont une composante du référentiel à grande échelle (RGE[®]) accessible gratuitement pour la recherche. Ces données sont très utilisées à l'INRA comme fond de référence pour une carte thématique ou comme support de vectorisation car il est possible de créer des données au format vecteur à partir d'images raster.

5 Par ces citations, nous faisons référence à d'autres articles de ce numéro spécial



Figure 1. Photographie aérienne issue de la BD ORTHO® 2010 de l'IGN au-dessus de l'Unité expérimentale de l'INRA d'Auzeville. Sur le zoom, les pixels qui caractérisent les images raster sont bien distinguables.

Données au format vecteur

Chaque couche au format vecteur est modélisée sous forme d'une double composante (**Figure 2**).

- ✓ **une composante graphique** : chaque thème du territoire (réseau hydrographique, réseau routier, villes, parcelles agricoles, etc.) est représenté sous forme d'une couche thématique au format vecteur. Par exemple, une couche représentant des parcelles agricoles contient un ensemble d'objets graphiques qui définissent les contours de chaque parcelle. Le type de forme qui représente en général une parcelle (et tous les objets surfaciques) est un polygone, mais il est aussi possible de représenter les objets géographiques sous forme ponctuelle (points de prélèvements, arbres, bâtiments ...) ou linéaire (réseau routier, réseau hydrographique, système d'irrigation, haie, ...). Les types de représentations graphiques (point/ligne/polygone) ne sont jamais mélangés dans une même couche.

Les objets graphiques sont ainsi enregistrés sous forme de liste de sommets avec pour chacun leurs coordonnées X,Y, dans un fichier « de forme » ou « shape » (par exemple .shp pour les couches au format QGIS ou ArcGIS).

- ✓ **une composante descriptive dite attributaire** : à chaque objet graphique d'une couche thématique sont reliées des informations descriptives stockées dans une table attributaire. Pour une couche « parcelles agricoles », ces informations peuvent être par exemple : l'identifiant (unique par parcelle), la surface, le type de culture implantée, la quantité d'intrants, la date de récolte, etc. L'ajout de champs dans la table attributaire ou leur modification se fait de façon aisée. La table attributaire peut être stockée dans un simple tableur ou liée à une base de données plus ou moins complexe stockée sous un système de gestion de base de données (SGBD) liée au SIG, comme PostgreSQL ou ACCESS. Par exemple on veut pouvoir stocker dans une base de données externe toutes les interventions (itinéraires techniques, produits utilisés, temps d'intervention, ...) ou bien les mesures et observations effectuées sur chaque parcelle de la couche de parcelles sans complexifier le modèle géographique de notre domaine expérimental.



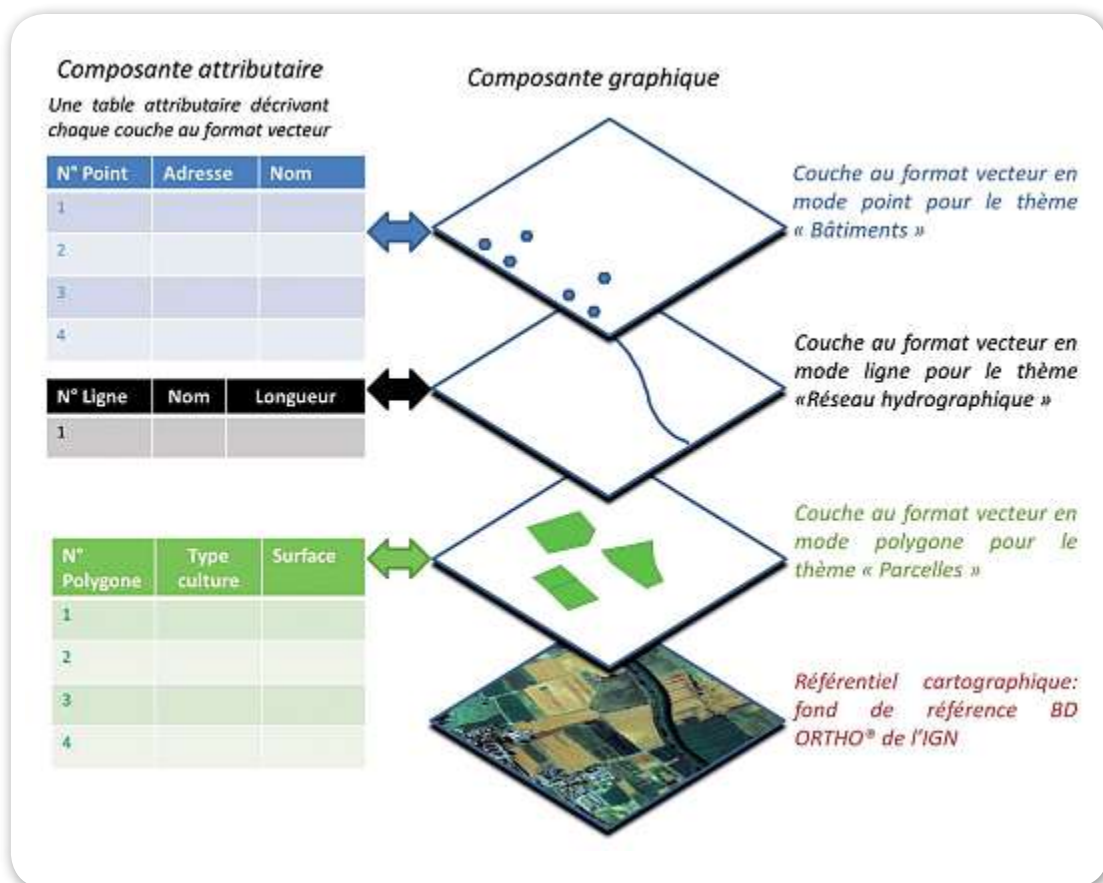


Figure 2. Organisation en couches de l'information géographique utilisée dans un système d'Information géographique. On distingue les deux principales composantes (graphique et attributaire) de l'information géographique au format vecteur.

Importance du système de coordonnées

Afin que toutes les données provenant de sources différentes puissent être superposées et exploitées dans un SIG, elles doivent appartenir à un système de coordonnées. Il existe deux types de systèmes de coordonnées utilisés dans les SIG :

- un système de coordonnées géographiques. Un point est référencé d'après ses valeurs de longitude et de latitude qui correspondent aux angles mesurés depuis le centre de la Terre vers un point de surface ; ses coordonnées sont exprimées en degrés, minutes, secondes ;
- un système de coordonnées projetées ou système de projection. Il existe diverses méthodes de projection représentant la surface sphérique de la Terre sur une surface de coordonnées cartésienne bidimensionnelle. L'EPSG (European Petroleum Survey Group) référence plus de 3 000 systèmes de projection différents variant sur leur paramétrage. Le choix du système de projection est conduit par l'usage qui sera fait de la carte mais aussi de la position de la région à cartographier sur le globe, chaque système étant adapté à une zone particulière afin d'éviter au maximum les déformations (Lahaye et al., 2014). Le système de repère cartographique officiel utilisé en France métropole est le Lambert93⁶ ; une fois appliquée, un repère cartographique XLambert93 et YLambert93 est défini et utilisé sur un plan à deux dimensions ; les coordonnées cartographiques (X, Y) d'un point sont alors exprimées en mètres et non plus en degrés comme dans le système de coordonnées géographiques.

6 Site IGN Lambert-93 : http://geodesie.ign.fr/?p=71&page=site_Lambert93

Lorsque des données de source variée sont utilisées, elles sont souvent dans des systèmes de coordonnées différents. Il faudra donc les transformer pour les rendre superposables. Les logiciels comme QGIS ou ArcGIS proposent des boîtes à outils pour ce type d'opération mais d'autres logiciels spécialisés sont également disponibles pour effectuer ces calculs lourds (citons le logiciel gratuit IGNmap : <http://ignmap.ign.fr/>).

La base de données spatialisées

La base de données (BdD) spatialisées est constituée de couches qui représentent l'information géographique sous les trois modes de représentation graphique précédemment présentés (point/ligne/polygone) et sous forme raster.

Dans une BdD spatialisée, il est primordial de suivre les deux règles suivantes :

- dans une même couche, la cohérence topologique entre les différents objets géographiques doit être respectée : par exemple, les contours de deux polygones « parcelles » adjacents ne doivent surtout pas être superposés l'un sur l'autre, leur limite doit être unique et donc partager les mêmes sommets ;
- entre deux couches, la cohérence spatiale relative entre objets géographiques doit être respectée : les systèmes d'irrigation, les lignes de semis ou de plantation d'une parcelle doivent être bien localisés à l'intérieur des parcelles. Autre exemple : les bâtiments d'exploitation sont représentés dans une couche et sont localisés dans une ou plusieurs parcelles cadastrales qui sont représentées dans une autre couche. Si la limite du bâtiment partage une partie de la limite d'une parcelle, les sommets constituant ces deux limites communes (sur ces deux couches différentes) doivent avoir les mêmes coordonnées (X, Y).

Ces deux obligations attestent d'une bonne qualité de la base de données spatiales, sachant que celle-ci peut être constituée d'une infinité de couches vecteurs sur un même territoire.

L'information géographique est décrite par son dictionnaire de données et ses métadonnées. Le dictionnaire de données fait l'inventaire des différents types de données utilisées ou la description de la structure des tables attributaires qui composent la BdD. Les métadonnées contiennent les informations sur la source, la date de création, les méthodes utilisées pour les mesures, les unités de mesure, etc. Toutes ces informations sont essentielles pour toute utilisation de cette base par quelqu'un n'ayant pas participé au travail de conception de la BdD, mais aussi pour faciliter une réutilisation ultérieure. La plupart des logiciels SIG offre des outils pour renseigner ces métadonnées.

Les principales fonctionnalités des SIG

Nous venons de voir que les SIG permettent d'intégrer des données de différents formats et provenance. Quelques-unes des autres principales fonctionnalités des SIG sont décrites ci-dessous en prenant comme exemple le cas de l'Unité expérimentale grandes cultures d'Auzeville de l'INRA de Toulouse (alias UE0802), domaine composé notamment de parcelles agricoles, de dispositifs expérimentaux, de points de prélèvement, de points d'irrigation (**Figures 3, 4, 5 et 6**). Sur les cas concrets pris en exemples, l'ensemble des objets géographiques constituant les différents thèmes sont représentés et localisés par rapport à une photographie aérienne datant de 2010.

Représentation cartographique des données attributaires et mise en page : les analyses thématiques

La **Figure 3** présente une capture d'écran du SIG QGIS dans lequel une analyse thématique sur la couche vecteur des parcelles de l'UE0802 a été effectuée. La vidéo permet de visualiser les étapes nécessaires à cette analyse thématique et à sa mise en page : au départ, toutes les parcelles sont représentées de la même couleur. En fonction du type de culture renseigné dans la table attributaire, une couleur différente est ensuite associée à chaque type de culture. La dernière étape de mise en page permet de préparer une composition cartographique comprenant



légende, titre, échelle, source. Une fois l'édition de la carte terminée on pourra l'exporter au format JPEG ou PDF par exemple.

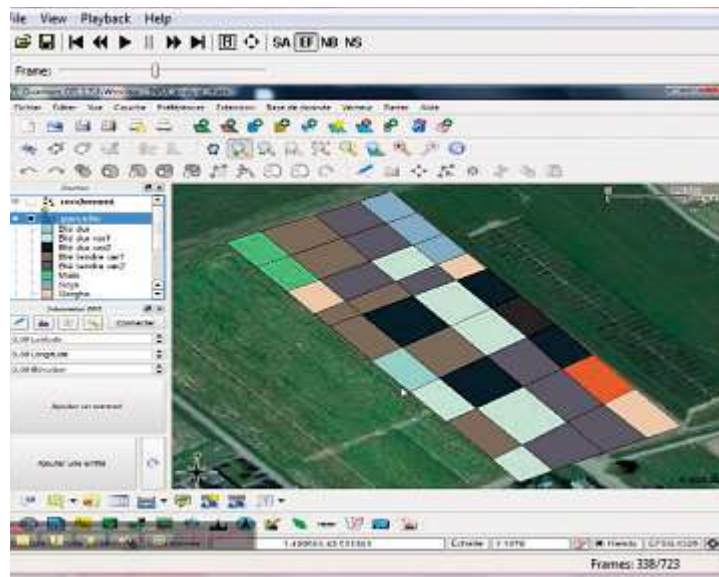


Figure 3. Capture d'écran de la vidéo (http://www.idgeo.fr/inra/demos/analyse_thematique_mise_page.swf) illustrant l'analyse thématique et la mise en page effectuées sur les parcelles de l'UE0802 dans le SIG QGis.

Requêtes attributaires : sélection des objets géographiques en fonction de critères sur les données attributaires

La Figure 4 présente un exemple de résultat de requête attributaire où les parcelles sont sélectionnées automatiquement en fonction de leur appartenance à la culture de blé tendre variété 2 et apparaissent en jaune. Une fois que les parcelles sont sélectionnées, d'autres opérations peuvent être effectuées spécifiquement sur ces parcelles. La vidéo associée montre les étapes nécessaires à cette requête attributaire.

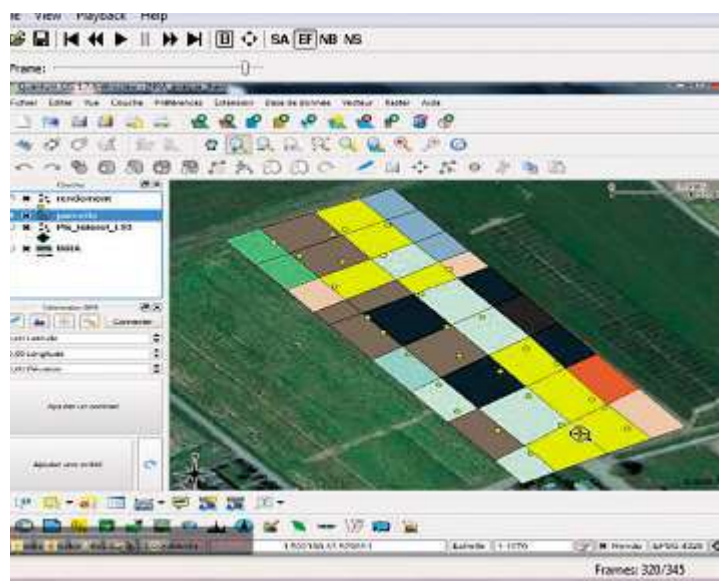


Figure 4. Capture d'écran de la vidéo (http://www.idgeo.fr/inra/demos/requete_attributaire.swf) illustrant une requête attributaire effectuée sur les parcelles de l'UE0802 dans le SIG QGis.

Requêtes spatiales : sélection des objets géographiques en fonction de leur localisation par rapport à d'autres objets graphiques (dans deux couches différentes)

La **Figure 5** présente le résultat d'une requête spatiale (étapes visibles sur la vidéo) où seules les parcelles à l'intérieur desquelles une mesure de rendement est disponible sont représentées en jaune.

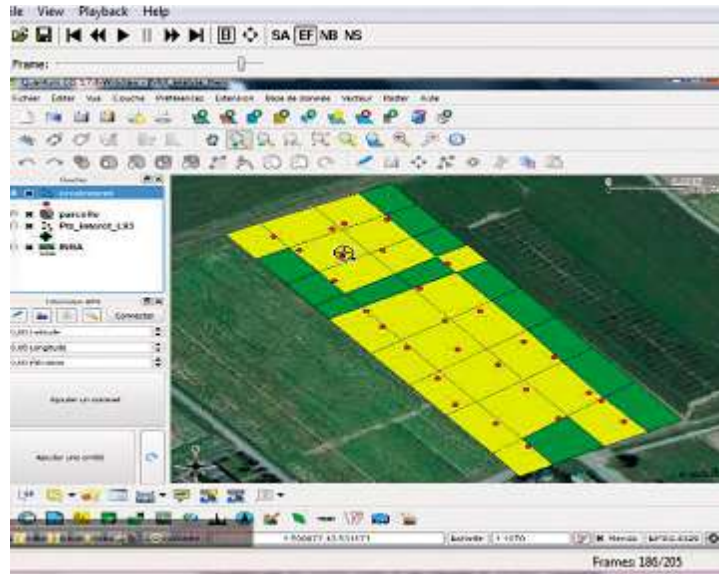


Figure 5. Capture d'écran de la vidéo (http://www.idgeo.fr/inra/demos/Requete_spatiale.swf) illustrant une requête spatiale effectuée sur les parcelles de l'UE0802 dans QGIS.

Analyse spatiale : création de zone tampon, croisement entre deux couches et calcul de superficie

Nous cherchons à connaître les superficies des zones réellement irriguées au sein de parcelles de l'UE802 en fonction des points d'eau disponibles. La capture d'écran de la **Figure 6** présente le résultat de la première étape qui a permis de créer une zone d'impact circulaire (appelée zone tampon) de 20 m autour de chaque point d'eau. La vidéo associée montre les étapes suivantes qui permettent (1) de croiser ces zones tampon avec les parcelles afin de visualiser les zones réellement irriguées au sein des parcelles (en rose), et (2) de calculer les superficies de ces zones.

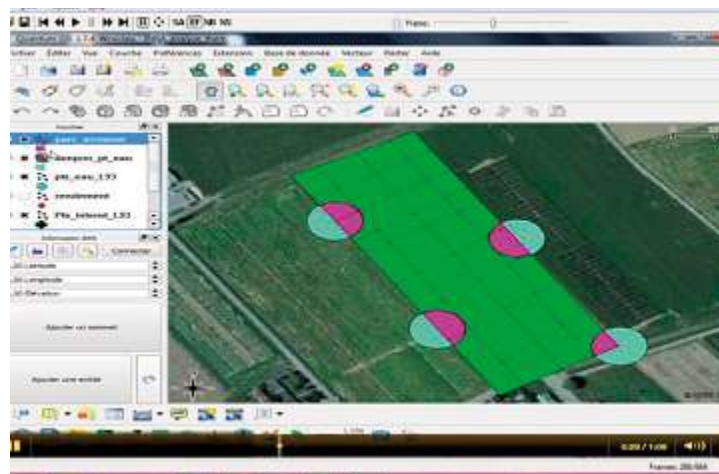


Figure 6. Capture d'écran de la vidéo (<http://www.idgeo.fr/inra/demos/Croisement.swf>) illustrant le croisement entre la couche parcelles de l'UE0802 et les zones arrosées autour des points d'eau dans QGIS.



L'ensemble des fonctionnalités des logiciels SIG décrites précédemment ne sont possibles que si l'intégration des données (données issues de GPS (Lahaye et al., 2014), images satellitaires, photographies aériennes, données de terrains géolocalisées, données descriptives, etc.) se fait de manière cohérente par une modélisation de l'information en fonction de l'objectif de l'utilisation de la base de données spatiales. Le système de repère cartographique défini doit être commun entre les différentes composantes de la base de données afin de rendre possible leur superposition.

Panel de quelques données géographiques disponibles gratuitement ou à faible coût pour la Recherche publique

Revenons sur quelques données externes souvent utilisées par les UE et UR de l'INRA :

- les données du RGE® (référentiel à grande échelle, <http://professionnels.ign.fr/rge>) fournies par l'IGN avec notamment la BD PARCELLAIRE® (information cadastrale numérique), la BD TOPO® (description des éléments du paysage), RGE®ALTI (composante altimétrique du RGE®) et la BD ORTHO® (composante orthophotographique). La BD ORTHO® à 50 cm de résolution est utilisable gratuitement pour tout usage d'enseignement initial ou de recherche publiable et celle à 5 m de résolution est maintenant utilisable gratuitement quel que soit l'usage ;
- les données du registre parcellaire graphique (RPG) administré par l'Agence de services et de paiement (ASP) fournissent des informations détaillées sur l'occupation du sol issues des déclarations politique agricole commune (PAC) (Cantelaube et Carles, 2014) ;
- certaines données satellitaires fournies par SPOTImage avec notamment le programme *ISIS* qui vise à promouvoir l'utilisation de l'imagerie spatiale SPOT auprès de la communauté scientifique en lui permettant d'acquérir ces données satellitaires à un tarif préférentiel (<http://www.isis-cnes.fr/>) ;
- les données disponibles via le projet EQUIPEX GEOSUD (<http://geosud.teledetection.fr/>) qui vise à développer une infrastructure nationale de données satellitaires accessible gratuitement par la communauté scientifique et les acteurs publics. Il permet notamment d'assurer pendant 5 années l'acquisition et la mise à disposition de couvertures satellitaires annuelles d'été, haute résolution de la France, et d'assurer la mise en réseau de la communauté scientifique et de la communauté des acteurs de la gestion autour de sa valorisation ;
- données CORINE de l'Institut français de l'environnement concernant l'occupation du sol en Europe (<http://www.ifen.fr>). La mise à jour de ces données pour 2012 sera accessible courant 2014.

Références bibliographiques

Cantelaube P, Carles M (2014) Le registre parcellaire graphique : des données géographiques pour décrire la couverture du sol agricole. *Le Cahier des Techniques de l'INRA*, N° Spécial GéoExpé. pp. 58-64.

Collet C (2005) Analyse spatiale, géomatique et systèmes d'information géographique. *Rev Int. Géomatique*, 15, n° 4, pp. 393-414.

Lahaye R, Ladet S (2014a) Les concepts de base des SIG nomades. *Le Cahier des Techniques de l'INRA*, N° spécial GéoExpé. pp 28-35.

Lahaye R, Ladet S (2014b) Les réseaux de correction différentielle. *Le Cahier des Techniques de l'INRA*, N° spécial GéoExpé. pp 36-43.

Lahaye R, Ladet S (2014) Les principes du positionnement par satellite : GNSS. *Le Cahier des Techniques de l'INRA*, N° spécial GéoExpé. pp 9-18.

Pour en savoir plus...

- Le portail francophone de la géomatique : <http://georezo.net/> C'est un portail francophone destiné à vous aider au travers de forums thématiques et techniques. Vous trouverez une foule d'informations dont des liens vers des glossaires sur le vocabulaire géomatique employé dans cet article : <http://georezo.net/annuaire/glossaires-c-72.html>.
- Liste de diffusion géomatique INRA : cette liste de diffusion par courriel, créée en 2011 à l'initiative de deux ingénieurs de l'INRA Mickael Rabotin (UMR LISAH, Montpellier) et Joël Daroussin (UR USS, Orléans), s'adresse à toute personne de l'INRA souhaitant communiquer sur tous les aspects de cette discipline : concepts, technologies, logiciels, données, etc. soit pour poser des questions, soit pour y répondre, soit pour communiquer des informations à la communauté des abonnés à la liste.

Pour s'abonner à cette liste : <https://listes.inra.fr/sympa/info/geomatique-inra>.

Pour envoyer un message à cette liste (abonnés seulement) : geomatique-inra@listes.inra.fr.

- Géoportail : vous permet de visualiser non seulement des cartes et des photographies aériennes, mais aussi de nombreuses autres données géolocalisées liées à l'environnement, à l'aménagement, au service public : <http://www.geoportail.gouv.fr/donnees>.
- Le portail d'information sur la géodésie : <http://geodesie.ign.fr/index.php>.
- Liste des références spatiales dans différents formats (EPSG, ESRI,...) : <http://spatialreference.org/>.

