

DynaforNet, un système d'information pour un site de recherche à long terme

Exemple de la gestion de données sur la biodiversité

Sylvie Ladet¹, Wilfried Heintz¹, Valéry Rasplus¹

Résumé : cet article décrit la chaîne de traitement informatique de données diverses acquises sur le terrain dans un espace géographique délimité au cours du temps et qui comprend la structuration de type relationnelle (SGBD) et spatialisée (SIG), le stockage (serveur et sauvegarde) et la valorisation (métadonnées et éventuellement diffusion vers l'extérieur).

Mots clés : SGBD, SIG, métadonnées, IDS, IDG, catalogage, stockage, sauvegarde, écologie du paysage, relevés de terrain, avifaune, modélisation.

Introduction

Les activités de l'unité mixte de recherche Dynafor du centre Inra de Toulouse sont centrées sur la gestion durable des ressources forestières et de l'espace rural dans le cadre de l'écologie du paysage. L'objectif principal de Dynafor est de comprendre et de modéliser les relations entre des processus écologiques, des processus techniques et des processus socio-économiques dans la gestion de ressources naturelles renouvelables (cf. **encadré n°1**). Par construction Dynafor est un projet de recherche interdisciplinaire qui mobilise des disciplines biologiques, techniques et socio-économiques. Cette unité travaille depuis plus de 30 ans sur deux sites d'études localisés dans le Sud-Ouest de la France. Nous allons ici aborder le site atelier des Vallées et Coteaux de Gascogne, situé au sud de Toulouse qui a été labellisé LTSER (Long Term Sociological and Ecological Research) en 2007. La thématique centrale des recherches effectuées concerne le suivi à long terme (i) des services écosystémiques rendus par la biodiversité dans les paysages agricoles, (ii) des facteurs qui les influencent (composition et structure du paysage, pratiques agricoles) et (iii) de leur évolution dans le contexte des changements globaux (changement d'utilisation des terres, changement climatique, changements dans l'environnement économique et réglementaire). Aussi chaque site est caractérisé par un périmètre géographique délimité au sein duquel seront mesurées in situ conjointement des données biologiques, des données sur les pratiques de gestion (parcelle et exploitation), des données sur les structures paysagères (mosaïque des occupations du sol, infrastructures écologiques, etc.). L'ensemble de ces données sont structurées numériquement au sein d'un système d'information pour faciliter l'intégration et la mise à disposition de ces données aux chercheurs de Dynafor. Pour illustrer concrètement nos propos, nous avons trouvé intéressant de faire écho à l'article de Laurent Raison (2007) qui porte sur l'inventaire des oiseaux nicheurs par la méthode des points d'écoute dans les paysages

¹ UMR1201 DYNAFOR – Dynamiques forestières dans l'espace rural - INRA - F-31326 Castanet Tolosan
☎ 05 61 28 52 55 ✉ sylvie.ladet@toulouse.inra.fr ; Wilfried.Heintz@toulouse.inra.fr ;
Valéry.Rasplus@toulouse.inra.fr

agricoles. Il décrit la méthodologie de collecte de données sur le terrain des Coteaux de Gascogne avec la description précise de la phase de planification. Plusieurs fois il fait référence au système d'information en citant les mots-clés suivants SIG, bases de données... que nous gérons.

Nous allons décrire, au travers de ses différentes parties, la chaîne de traitement informatique de ces données acquises sur le terrain au cours du temps qui comprend la structuration de type relationnelle (SGBD) et spatialisée (SIG), le stockage (serveur et sauvegarde) et la valorisation (métadonnées et éventuellement diffusion vers l'extérieur).

Encadré n°1 : Présentation de l'unité Dynafor

L'unité Dynafor est une unité mixte de recherche INRA- INP/ENSAT regroupant chercheurs de l'INRA et enseignants-chercheurs de l'école d'agronomie de Toulouse y compris du personnel technique aussi bien en informatique qu'en observation de terrain. Elle s'est engagée dans le développement de recherches qui s'inscrivent résolument dans le champ d'une écologie du paysage « pour l'action ». L'écologie du paysage est vue comme science intégratrice des relations homme-nature et une science utile pour l'action. Dans nos travaux, nous considérons le paysage comme le fruit d'une coévolution entre les systèmes sociaux et les systèmes écologiques à des échelles spatiales et temporelles multiples qui correspondent aux différents niveaux d'organisation des processus étudiés (Allen et Star, 1982). Les processus étudiés concernent globalement la durabilité des entités de production agricole et forestière, le changement d'utilisation des terres dans les paysages, la dynamique de colonisation des arbres dans les paysages en déprise agricole, l'influence des pratiques de gestion agricole et forestière sur la biodiversité des paysages à différentes échelles, l'étude spécifique du rôle des lisières agriculture-forêt sur la biodiversité et les services rendus par la biodiversité à l'agriculture. Il s'agit de faire face aux enjeux actuels dans les espaces ruraux et forestiers, induits par les changements globaux qui concernent conjointement le climat, l'occupation des terres, la biodiversité et les activités humaines.

A Dynafor, nous sommes 3 responsables¹ techniques du systèmes d'information qui se construit « chemin faisant » sur ce terrain d'étude et qui est vu comme un ensemble de moyens humains, techniques, informationnels et méthodologiques, permettant d'acquérir, mémoriser, structurer, traiter, interpréter et diffuser entre les membres d'une organisation (ici, l'UMR), les informations recueillies par les disciplines en place. L'objectif est que chacun puisse consulter les informations produites par les uns et les autres, que chacun puisse utiliser et croiser les informations recueillies en y apportant des données produites dans sa discipline, afin de créer un « capital » de connaissances communes, matière première de la démarche interdisciplinaire de l'unité. Nous gérons donc un gros volume de données qu'il s'agit d'organiser et de mettre à disposition en respectant les standards informatiques.

1

Valery Rasplus (gestionnaire de bases de données) a la responsabilité technique de la conception et la maintenance des bases de données relationnelles.

Sylvie Ladet (géomaticienne généraliste) a la responsabilité technique de la conception et la maintenance des systèmes d'information géographique et veille à la cohérence globale du système d'information.

Wilfried Heintz (géomaticien spécialisé sur la mise en réseau des SI) intervient sur la métadonnée pour mettre en œuvre la traçabilité.

1. Contexte informatique à Dynafor

1.1 Vallées et Coteaux de Gascogne

La plateforme LTSER de Dynafor est un paysage rural à vocation agricole situé à 80 kilomètres de Toulouse dans la région du Bas-Comminges (Deconchat et al, 2007). Il s'agit d'une zone vallonnée d'une altitude comprise entre 200 et 400m couvrant 250 km² avec une matrice agricole de polyculture-élevage marquée par la présence de nombreux bois fragmentés et d'éléments linéaires boisés de type haie, alignement d'arbres et un habitat dispersé. Ce terrain est l'objet, depuis les années 80 de nombreuses études menées *in situ* et répétée (avec une fréquence variable, par exemple tous les 10 ans par exemple pour un dispositif de suivi de communautés d'oiseaux ou tous les 20 ans pour le suivi des exploitations agricoles) sur l'évolution de la biodiversité des communautés végétales herbacées et animales (insectes, oiseaux), sur l'évolution des pratiques de gestion des exploitations agricoles et forestières. Aussi en 2007, Dynafor a obtenu la labellisation LTSER de ce territoire (numéro d'enregistrement LTER_EU_FR_003). C'est un label lancé en 1980 par la *National Science Fondation* aux Etats-Unis. Le but de ce label, maintenant international, est de mettre en réseau des sites de recherche à long terme, dans des écosystèmes importants et sensibles à travers le monde. Il en résulte une grande quantité de données de nature diverses qu'il convient de structurer informatiquement. Une des caractéristiques les plus importantes de ce label est le développement de bases de données, disponible pour tous les chercheurs, contenant des informations sur des données de types écologiques, socio-techniques et sur les conditions environnementales (occupation des sols, climat ...) sur les sites étudiés. Nous identifions l'importance de la profondeur historique des données recueillies pour comprendre ces systèmes socio-écologiques complexes.

1.2 Les 4 composantes du système d'information à Dynafor

Nous avons mis en place progressivement un système d'information pour nous aider dans cet objectif de mutualisation et de partage. Ce système mis en place dans l'unité est un outil scientifique et technique fédérateur de type plateforme qui participe à établir un lien tangible entre les différents compartiments des systèmes étudiés. Cette plateforme est « chemin faisant » c'est-à-dire évolutive, modulaire. Il comporte 4 composantes qui interagissent : les outils (matériels et logiciels dédiés), les ressources humaines (disciplines des sciences et technologies de l'information et de la communication allant de la géomatique, à la modélisation en passant par la gestion de bases de données), les données (cartographiques et tabulaires) et les procédures (méthodologies, protocoles). Un effort particulier a été placé à Dynafor dans son développement et sa structuration au cours des dernières années. Il s'intègre peu à peu dans des réseaux d'échange de données (réseau LTER) qui nécessitent une organisation claire de son fonctionnement.

1.3 Exemple de la structure de la plateforme sur les données Oiseaux

Depuis 1981, chercheurs et techniciens de l'unité Dynafor effectuent des inventaires de communautés d'oiseaux en utilisant une méthode standardisée des points d'écoute (Blondel, J. *et al*, 1970) en réalisant concrètement des relevés de terrains comme l'explique Raison (2007).

Les recensements ornithologiques sont notés dans un premier temps sur une fiche de terrain construite préalablement (support papier) afin de recueillir des données sur l'abondance des oiseaux et les variables du milieu susceptibles d'expliquer cette abondance.

Analysons plus précisément le contenu de ces fiches qui indiquent le type de données à stocker numériquement. Un point d'écoute consiste à réaliser une observation stationnaire, géo-localisée, pendant une période totale de 20 minutes. Pour un point d'écoute donné, le temps global est divisé en quatre parties de cinq minutes (Raison L., 2007). On note sur la fiche les espèces d'oiseaux, vues, entendues, posées ou en vol dans un rayon préalablement défini (en moyenne 250 mètres). Des conditions environnementales sont également prises en compte, comme le vent, la pluie, le soleil, les nuages. Ces données couplées à des données d'habitat (forêt, jardin, verger, haie, lisière, pâture, type de culture), de préférence alimentaire (insecte, grain, etc.) permettent par la suite de modéliser les relations entre les oiseaux et les différents habitats. Pour le chercheur, l'évolution des communautés, l'apparition ou la disparition d'espèces et leur distribution sont des indicateurs de changements des structures paysagères (Balent G. et Courtiade B., 1992 ; Monteil et al, 2005). Les campagnes d'inventaire des populations d'oiseaux par la méthode des points d'écoute sur le terrain délimité des Coteaux de Gascogne, répétées dans le temps, ont exigé d'acquérir une compétence de gestion informatique fine afin de répondre au double objectifs de conservation et de structuration des données.

2. Structuration des données

2.1 SGBD : intégration des données de terrain

Ces informations manuscrites sont ensuite enregistrées dans un système informatique de gestion de données afin d'assurer à la fois une restitution rapide des données scientifiques pour les chercheurs intéressés et une réponse au besoin de sauvegarde sécurisée de données numériques. La collecte de données de terrain exige que celles-ci soient conservées et structurées dans un système de gestion approprié afin de répondre aux exigences de traitements et de diffusion de données volumineuses pour les recherches.

À l'unité Dynafor, l'ensemble des données ornithologiques recueillies sur le terrain sont intégrées dans une base de données relationnelle (Access©) ; elle a été déclarée dans l'inventaire des bases de données scientifiques structurées à l'Inra en 2009 sous le nom de *DynaBird*. Elle se démarque des bases de données traditionnelles en prenant en compte le facteur temps - dans l'intégration des données de campagne - et l'intégration de données spatiales via la localisation géographique des points d'écoute en X et Y.

Cette base de données géo-référencée comprend 13 tables et 130 champs. Elle comprend deux types de catégories de variables : les variables à expliquer comme l'abondance des espèces (137)², l'évolution de la répartition ; les variables explicatives comme les caractéristiques de terrain, les traits d'oiseaux (137), le nombre de campagnes (27), le nombre de communes (36), le nombre de stations (2981), le nombre de points d'écoute (5644), le nombre d'oiseaux contactés (167743). La volumétrie totale de cette base de données est de 113 546 lignes d'enregistrements.

² Entre parenthèses figure l'effectif rencontré pour donner une idée de la quantité de données manipulées

La **figure 1** donne une vision d'ensemble des tables physiques et des contraintes d'intégrités qui garantissent la cohérence des données lors des mises à jour de la base (Gardarin G., 2005).

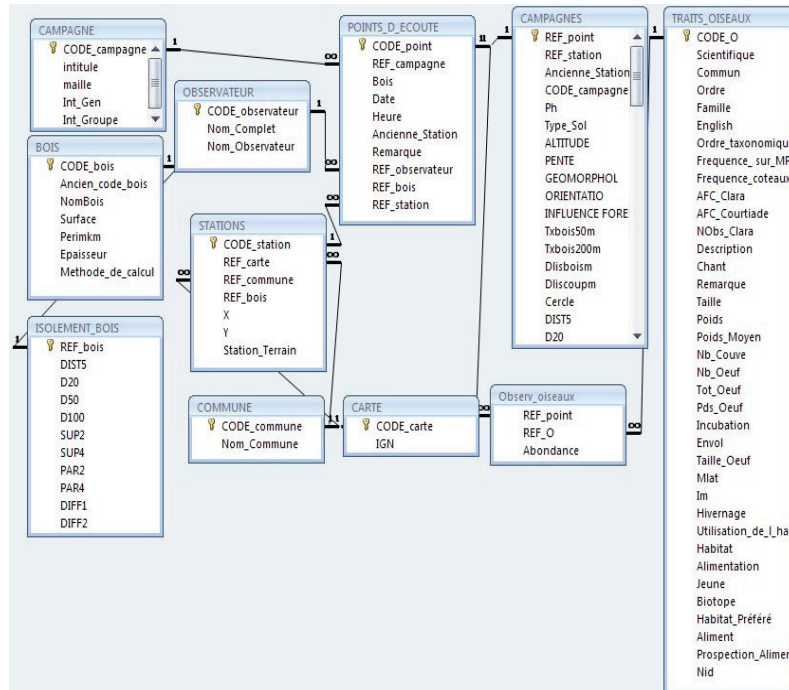


Figure 1 : modèle physique des données de DynaBird

2.2 Interfaçage des bases pour utilisation par les chercheurs

Un point important du cahier des charges était de construire un outil de gestion informatique facile à prendre en main afin de privilégier l'autonomie de ses utilisateurs. Son utilisation ne nécessite pas de connaissance particulière du logiciel Access®, ne demande pas de voir et de différencier les objets Access® (tables, formulaires, requêtes, états, macro, modules), ni de saisir directement dans les tables et aucune compréhension du langage de requête SQL (*Structured Query Language*) n'est exigée. Pour ce faire, un ensemble de formulaires, servant d'interface entre les données intégrées dans les tables et l'utilisateur, ont été créés en fonction des besoins et des pratiques de terrain, facilitant la rapidité de la saisie des données. L'utilisateur peut réaliser des opérations de consultation, de modification, de suppression, de mise à jour, de sélection et d'extraction de données croisées ou non.

La **figure 2** montre une partie du formulaire de saisie des campagnes d'écoute d'oiseaux.

Figure 2 : extrait du formulaire de saisie de points d'écoute

En mode saisie, la tabulation permet de passer automatiquement d'un objet à un autre (zone de texte, case d'option, zone de liste déroulante). La quasi-totalité des données subissent un contrôle de validité. En cas de doute, une consultation visuelle de n'importe quelle campagne ou point d'écoute permet de modifier rapidement les données éventuellement erronées, comme le montre la **figure 3**.

Figure 3 : extrait du formulaire de consultation et modification de points d'écoute

A tout moment, on peut consulter l'ensemble des données oiseaux point d'écoute par point d'écoute et ainsi suivre l'évolution des effectifs et des répartitions à un moment donné sur un point géo-localisé donné

Cette base de données construite sous Access© est interconnectée au logiciel de système d'information géographique (SIG) utilisé à Dynafor (ArcGis©) via une connexion ODBC (*Open DataBase Connectivity*). C'est une technologie permettant d'interfacer de façon standard une application à n'importe quel serveur de bases de données, pour peu que celui-ci possède un driver ODBC (la quasi-totalité des SGBD possèdent un tel pilote).

Pour plus d'informations sur ce point particulier, se reporter à la note technique « *Comment connecter ArcGIS à des fichiers Microsoft Access 2007 (ACCDB)* » :

<http://support.esri.fr/index.asp?page=/articles/arcgis/connexions%20aux%20bases%20de%20donnees/32976.htm> (consulté le 12/03/2010). Dans le sens base vers SIG, ceci permet de spatialiser le résultat de requêtes effectuées dans la base de données. Dans le sens SIG vers base, ceci permet d'enrichir la base avec des données saisies dans le SIG (comme le pourcentage des occupations du sol autour des points d'écoute).

3. Spatialisation

Une partie des données acquises sur le site des Coteaux de Gascogne se présente sous la forme spatialisable (coordonnées X et Y des points d'écoute) ou cartographique (données brutes comme une photographie aérienne ou élaborées comme une carte d'occupation des sols). Nous touchons ici à la dimension spatiale qui est un des principaux facteurs des données manipulées par Dynafor. Le système d'information construit comprend donc un module géographique conséquent car la dimension spatiale est cruciale pour comprendre et modéliser les relations entre des processus écologiques, techniques et sociotechniques dans la gestion des ressources naturelles. Nous sommes alors dans le monde des SIG.

Mais au fait qu'est ce qu'un SIG ? Il existe des réponses multiples dans la littérature dédiée à la géomatique et la recherche d'une définition acceptée par tous est quasiment impossible tant elle varie en fonction des auteurs. Nous avons repris celle de l'université de géographie de Reno à savoir que le SIG est « *un outil unique intégrant des données diverses mais localisées dans le même espace géographique, relatives à la fois à la terre et à l'homme, à leurs interactions et à leurs évolutions respectives, quels que soient les domaines concernés: physiques, sociaux, économiques, écologiques, culturels, etc. Ce rassemblement permet d'élaborer les synthèses indispensables à la prise de décision dans tous les domaines aussi bien dans les situations de crise que dans le suivi des évolutions à long terme* ». Ce qui nous intéresse dans cette définition est le fait que l'utilisation des SIG peut faire émerger les liens dans une perspective synchronique et diachronique dans l'organisation du territoire et synthétiser des informations provenant de sources et de domaines divers. À Dynafor, le SIG sert à plusieurs niveaux d'intervention en amont et en aval après la récolte des données sur le terrain.

3.1 En amont, la dimension spatiale des données collectées

Le SIG à Dynafor est un moyen efficace pour :

- aider à la planification des dispositifs d'observations de formes variées : au bureau nous construisons avec le chercheur responsable les plans d'échantillonnage régulier sur un maillage défini (campagne d'écoute sur plusieurs communes contigües) *versus* échantillonnage stratifié selon la nature d'occupation des sols (campagne d'écoute uniquement dans des bois). Pour se faire, il existe des scripts et d'autres extensions permettant de construire ces plans d'échantillonnage. Citons par exemple les outils gratuits proposés par l'extension *Hawth's Analysis Tools* (Beyer, 2004) qui sont développés pour le logiciel ArcGis© (figure 4).

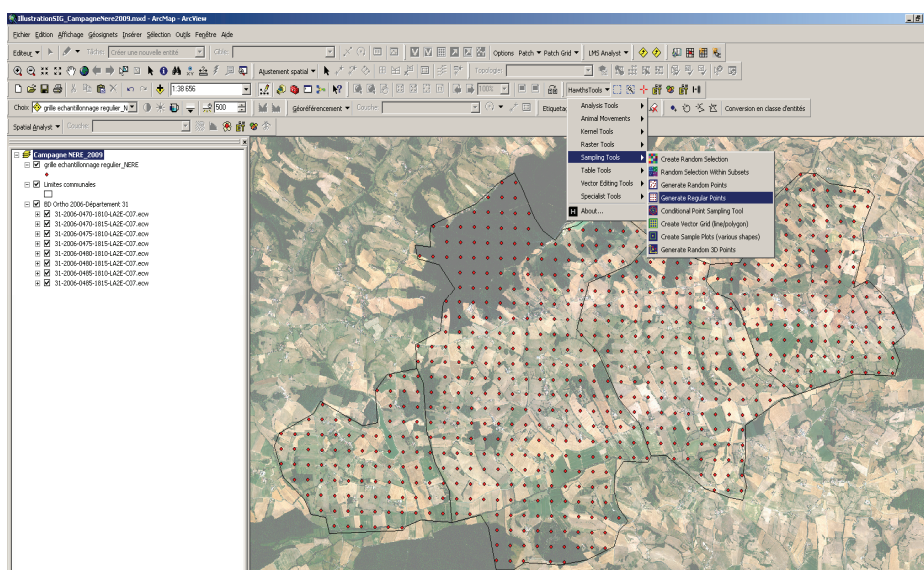


Figure 4 : grille (676 points rouges) d'échantillonnage régulier d'un pas de 250 mètres sur 4 communes contigües construite sous ArcGis© avec la fonctionnalité « Generate regular points tool » de l'extension Hawth's Analysis Tools

- aider à la coordination sur le terrain. Raison et al (2007) ont présenté dans leur article les documents cartographiques élaborées en amont de la campagne de terrain d'écoute des oiseaux (carte générale du dispositif sur fond de scan topographique de l'IGN, fiche de terrain recto avec fond photographie aérienne en couleur avec les points d'écoute géo-référencés). Effectivement, ces supports cartographiques peuvent aider à la coordination du dispositif sur le terrain en facilitant le repérage sur le terrain. Il est important pour l'ornithologue par exemple de se positionner précisément sur le terrain, de bien visualiser et de délimiter la zone dans laquelle il va dénombrer les oiseaux et décrire le milieu (zone tampon de type carré centré sur le point). C'est d'autant plus essentiel que les points d'écoute sont réalisés au même endroit à des dates différentes pour pouvoir comparer la composition de l'avifaune. La précision de la localisation géographique est donc primordiale. Pour se faire, nous recherchons des outils complémentaires pour personnaliser Arcgis© en rajoutant des extensions pour combler des manques de la version par défaut du logiciel.

Citons par exemple l'extension « *Créer un atlas cartographique* » disponible par téléchargement : <http://support.esri.fr/index.asp?page=/outilsscripts/arcgis/arcmap/miseenpage/atlas/atlas.html>. (Consulté le 12/03/2010).

Cet outil permet de créer un document ArcMap³ contenant plusieurs mises en page cartographiques avec la même architecture. Dans une chaîne de production cartographique, l'objectif est d'automatiser la production de cartes (produire les 676 fiches recto de terrain automatiquement) en générant une grille d'index qui correspond à une classe d'entités utilisée pour diviser vos données en tuiles. En ce qui concerne les campagnes d'écoute d'oiseau, la grille d'index est la classe d'entités contenant les zones tampons autour de chaque point d'écoute. Dans l'atlas, nous pouvons ajouter des éléments (titre, échelle, carroyage, fichier tableur avec numéro des points, nom de communes...), qui seront pré-remplis de manière dynamique (**figure 5**).

The screenshot shows the 'Atlas' software interface. On the left, a window titled 'Atlas' contains icons for map, data, and other functions. Below it, text reads 'Capture d'écran de la barre d'outil de l'extension « Atlas »'. The main window is titled 'Fiche recto N°1 de la campagne NERE 2009'. It contains a data entry form with fields for Station (1), Carré manip 3 sites, Commune (Peyrissas), Observateur, Date, Heure, Ciel (Soleil), and Couvert (Couvert). Below the form is a table with columns for 'Mètres' (0-1m, 1m-4m, 4m-8m, 8-16m, >16m) and rows for 'Esp. Dominante' and '%'. Below the table is a map showing a landscape with a red square indicating the 'Emprise géographique de la zone à relever'. To the right of the map is a legend titled 'Légende d'occupation des sols définie' with a list of soil types and their corresponding codes: Avoine (Cs), Blé (Cb), Colza (Cc), Labour (Cl), Maïs (Cm), Orge (Co), Soja (Cs), Sorgho (Cso), Tourmesol (Cl), Trèdèrenné (Cl), Prairies temporaires (PTg), Prairies permanentes (PPg), Fiches (Fg), Forêts (Fp), and Bandes enherbées (B).

Figure 5 : fiche recto du point d'écoute de la campagne Nère en 2009 (1^{ère} page de l'atlas cartographique) réalisée sous ArcGis avec l'extension Atlas. L'ornithologue est assisté pour relever l'occupation des sols autour d'u point d'écoute. Il doit reporter sur l'image, le sigle de l'occupation du sol qu'il observe.

Ces méthodes actuelles servent à prendre en compte la dimension spatiale sur le terrain via l'utilisation de fonds cartographiques annotés. D'autres solutions plus abouties pourront être utilisés dans un futur proche comme les GPS embarqués ou le développement d'une application SIG nomade. Installé sur un récepteur PDA (*Personal Digital Assistant*), une tablette PC ou un

³ ArcMap est l'application centrale d'ArcGis pour la visualisation de cartes, l'édition, l'analyse et la mise en page

micro-ordinateur portable, cela permet aux utilisateurs d'accéder aux bases de données (relationnelles et/ou géographiques) directement sur le terrain et d'utiliser des fonctions d'assistance à la localisation avec la possibilité d'intégrer un système GPS.

3.2 En aval : l'organisation des données géographiques

Dans le logiciel de SIG, l'intégration de données hétérogènes qui proviennent de sources de données variées, va nous aider : (1) à formaliser ces données à la fois dans leur géométrie, dans leur thématique et leur agencement temporel, (2) à combiner ces données entre elles grâce aux outils d'analyse spatiale, (3) à répondre aux interrogations des thématiciens par requêtes multi-critères et (4) à quantifier et modéliser les dynamiques spatio-temporelles de processus complexes (Collet C, 2005).

À Dynafor, nous disposons d'une base de données géographiques déjà structurée contenant des données dites brutes (comme les photographies aériennes de l'IGN, les images satellites et autres données images du Référentiel à grande échelle (cf. protocole de diffusion IGN/MAP/MEDAD de diffusion de données aux organismes de recherche ...)) ou des données au format vecteur (cadastre, limites administratives, réseaux routiers...) et des données dites élaborées (comme des cartes d'occupations des sols, carte des réseaux de haies, carte des parcellaires d'exploitation agricole...). Mais notre objectif de suivi demande de mettre à jour cette base en intégrant des données plus récentes (carte d'occupation du sol de l'année en cours) ou des données récemment numérisées (cartes anciennes). Nous sommes alors confrontés à des problèmes spécifiques, à savoir le géoréférencement de cartes anciennes, la digitalisation de notations relevées sur des fiches de terrain (Joliveau T., 2004). Le traitement des données est ensuite relativement aisé ; les couches d'information étant superposables, à la manière de calques, les possibilités de croisements des données et de requêtes thématiques sont très grandes dans le logiciel SIG et permettent l'élaboration de cartes recoupant différents thèmes rendant l'analyse des données (**figure 6**) et de leur évolution spatio-temporelle (**figure 7**) à la fois souple et rapide.

L'atout majeur du logiciel SIG est sa capacité de navigation dans la base de données et son interrogation à la fois dans sa dimension spatiale (cartes) et dans sa dimension attributaire (tables). On peut donc établir des requêtes attributaires, fondées sur des relations logiques et des requêtes spatiales ou sur une combinaison des deux pour sélectionner des objets ayant une caractéristique particulière. Par exemple, nous pouvons analyser l'évolution des linéaires de haies au cours du temps. Pour rappel, les haies sont un objet important du paysage pour les oiseaux car ils assurent la connectivité des éléments boisés et sont donc essentiels à la circulation des oiseaux forestiers.

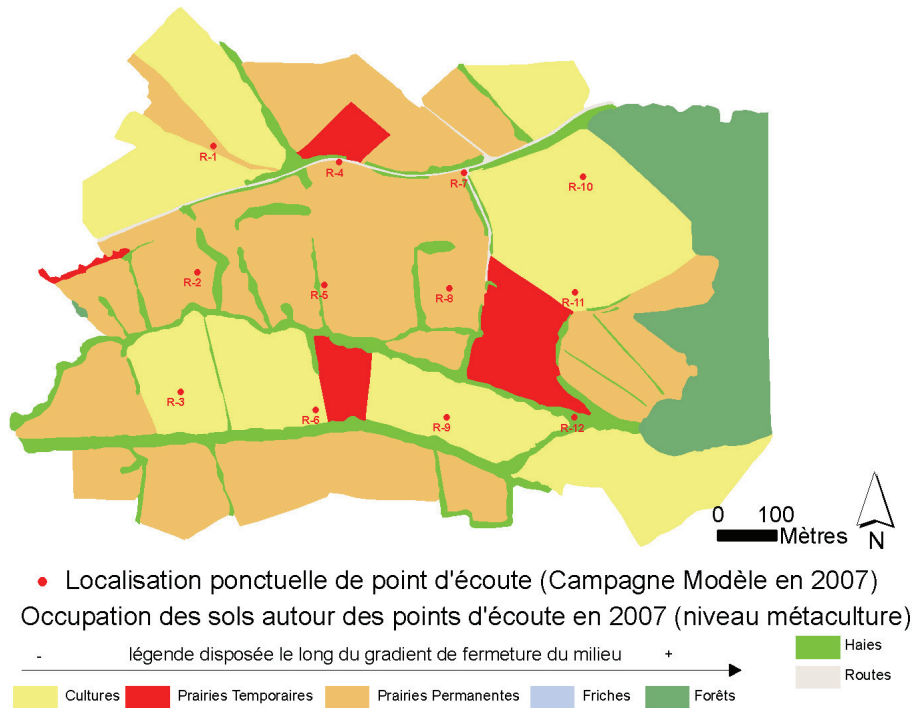


Figure 6 : saisie sous ArcGis sous forme de polygone de l'occupation des sols relevée autour du point d'écoute de la campagne Modèle en 2007.

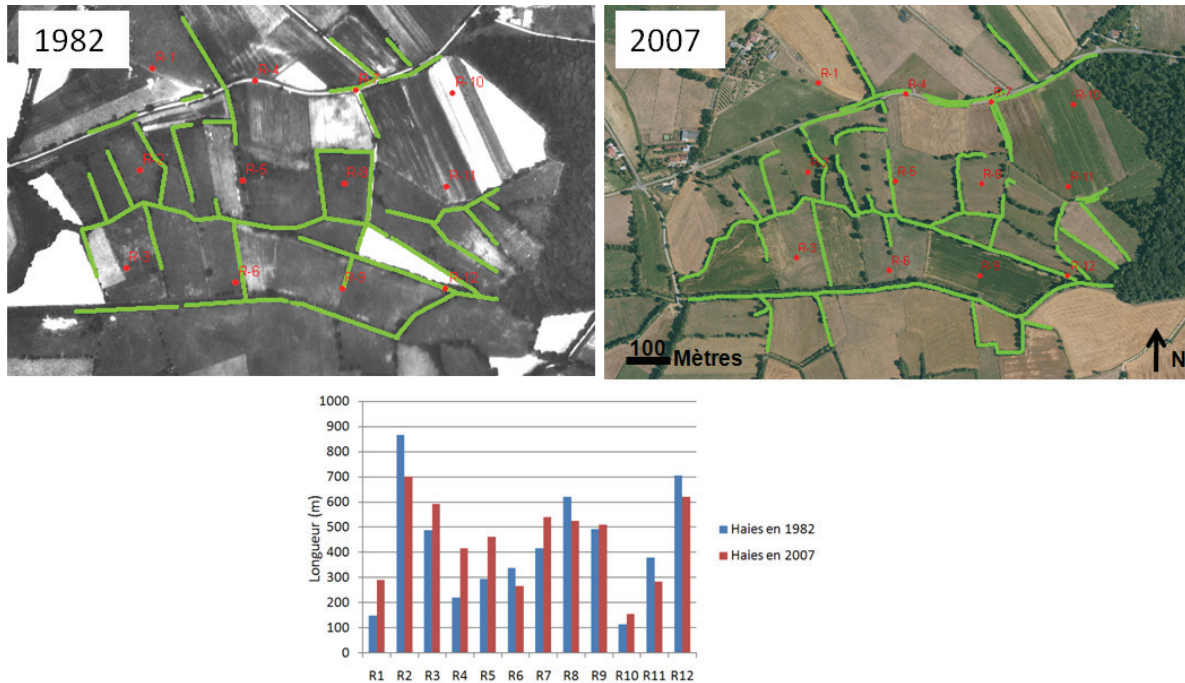


Figure 7 : évolution des haies en 1979 et en 2007 sous ArcGis de la campagne Modèle représentée sous forme de cartes statiques en haut ou de graphique en bas.

Le SIG apparaît comme un moyen d'organiser le traitement des données issues du dépouillement de données de terrain, de les analyser, de les confronter à d'autres données et de formaliser le tout. Il rend aussi possible le traitement des données à des échelles précises (résolution spatiale de l'ordre du mètre et résolution temporelle de l'ordre de l'année et résolution thématique de plus en plus précise). Ces données stockées atteignent alors des volumétries lourdes mais ces systèmes informatiques peuvent les gérer et les combiner sans limite de capacité liée à leur stockage. Une nouvelle question apparaît alors, la description des données de plus en plus nombreuses.

4. Stockage et sauvegarde

4.1 Nécessité d'un serveur dédié

Le traitement de données volumineuses requiert avant toute chose un espace adéquat pour les stocker. Malgré les considérables évolutions technologiques dans le domaine du stockage de données numériques et l'augmentation permanente de la taille des disques durs, il n'en reste pas moins que les données géographiques telles que les images satellites ou aériennes voient leur taille grandir proportionnellement à leur résolution. Les données issues des nouvelles technologies comme le LIDAR (*Light Detection and Ranging*) sont extrêmement volumineuses. La notion de stockage suggère aussi implicitement celle de l'accès à la donnée. S'il est important de disposer d'un espace suffisant pour stocker la donnée, il est également déterminant de pouvoir la récupérer simplement et rapidement, depuis différents postes de traitement.

De nombreuses solutions techniques, parmi lesquelles les disques durs portables disposent désormais de grandes capacités. Néanmoins, un système optimisé tel qu'un serveur de données s'avère rapidement indispensable pour un fonctionnement mutualisé et durable. On peut se contenter d'un serveur de stockage tel un système NAS (*Network Attach Storage*) où les données brutes sont déposées et où chaque agent peut les consulter et manipuler. Une solution plus évoluée consiste à mettre en place un serveur où les données sont stockées dans une base de données, permettant une consultation et une manipulation plus poussée. Il est actuellement possible de stocker des données géographiques vectorielles sous forme d'enregistrements à l'intérieur d'une base de données. Ceci est permis par la cartouche spatiale PostGIS⁴ du système de gestion de base de données (SGBD) PostgreSQL. Le requêtage s'en trouve considérablement optimisé. Le stockage des données de type matriciel n'est pas compatible avec un SGBD mais de nouveaux logiciels tels que PostGRID permettent d'indexer ces données pour un traitement optimisé. Ainsi, la solution retenue à Dynafor est la mise en place d'un serveur dédié au stockage et à la diffusion des données dites volumineuses.

4.2 Processus de sauvegarde des données

Un aspect primordial du stockage de données est d'en garantir la pérennité au sens « durabilité ». En clair, comment se prémunir d'une défaillance matérielle qui mettrait en péril la restitution des données numériques stockées sur le serveur ? Il s'agit ici du processus de sauvegarde des données, étape indispensable d'une chaîne de stockage correctement mise en œuvre. En effet,

⁴ PostGIS est une extension du système de gestion de bases de données relationnel à objets PostgreSQL qui permet de stocker des objets SIG dans une base. L'ensemble des logiciels PostGIS et PostgreSQL est libre de droit.

une véritable sauvegarde se termine sur un support durable, à l'abri des aléas extérieurs. La solution retenue dans la plupart des cas consiste à écrire les données du serveur sur bande magnétique, elle-même stockée dans une armoire blindée et ignifugée.

L'équipe informatique du centre Inra de Toulouse utilise un robot de sauvegarde mutualisé qui réplique quotidiennement le contenu du serveur sur bande magnétique ; cette solution de sauvegarde reste onéreuse (de l'ordre de 400€ pour une licence annuelle).

5. Valorisation des données

5.1 Catalogage

Le préambule à la diffusion de données est la structuration et la description de ces données. En effet, il a été maintes fois constaté que beaucoup de données (géographiques, statistiques, etc.) étaient inutilisables car leurs méthodes d'acquisition ou certaines de leurs caractéristiques intrinsèques n'avaient pas été notées ou conservées. La traçabilité d'une ressource est donc avant tout une garantie de pérennité (Michener, 2005).

Pour mettre en œuvre cette traçabilité, il existe la métadonnée et c'est là qu'intervient le géomaticien spécialisé sur la mise en réseau de SI (Wilfried Heintz pour notre unité). La métadonnée est une « information sur l'information », composée d'un ensemble de champs descriptifs normalisés pour caractériser une ressource. S'il est vrai que ce concept est étroitement lié aux données de nature géographique, il reste néanmoins applicable à toute information quelle que soit sa nature. C'est toutefois dans le domaine de l'information géographique que les innovations sont les plus importantes en terme de gestion et valorisation des données. Ceci est particulièrement dû à la profusion de ce type d'information, elle-même permise par les avancées technologiques en matière de systèmes d'informations géographiques (Heintz et Guéro, 2005). L'usage des SIG s'est démocratisé ; les logiciels cartographiques ne sont plus l'apanage des spécialistes, justifiant notre démarche de diffusion des données auprès des chercheurs de l'unité. Néanmoins, face à cette déferlante de données aussi nombreuses que diversifiées, il est nécessaire de créer des outils de gestion, de diffusion et de valorisation. Ainsi, les outils de gestion de métadonnées - également appelés (géo)catalogues - sont devenus des outils de diffusion et de partage de l'information géographique (CNIG, 2007). Ils sont même devenus la pierre angulaire des infrastructures de données spatiales et d'autres géoportails. Mais le développement de ces derniers ne peut se faire qu'en respectant les lois sur la diffusion des données. Le contexte législatif actuel justifie donc pleinement la démarche de catalogage des données produites et utilisées dans le fonctionnement de l'unité. Ce catalogue de données doit se construire conformément aux recommandations techniques de la directive européenne Inspire ratifiée par la France (cf. **encadré n°2**).

Plusieurs outils de catalogage existent, dont deux implémentent le profil français de la norme ISO (*International Organization for Standardization*). Ces logiciels sont Geosource, version « française » de l'outil Geonetwork, et MDWeb, développé par l'IRD⁵. Tous deux permettent de

⁵ IRD = Institut de recherche pour le développement ayant pour vocation de mener des recherches pour l'hémisphère Sud.

générer des fiches de métadonnées 100% compatibles avec les directives et préconisations en vigueur. Le choix de l'un ou de l'autre dépendra des caractéristiques du serveur hébergeant l'outil et des compétences du futur administrateur du logiciel sur des langages de programmation comme le XML (*eXtensible Markup Language*) /XSL (*eXtensible StyleSheet Language*) ou sur la configuration du conteneur libre de servlets Tomcat.

Encadré n°2 : [catalogage et législation européenne](#)

La directive européenne *Inspire (Infrastructure for Spatial Information in Europe)* incite fortement les acteurs du service public à se doter d'un système de gestion de métadonnées. Elle vise à mettre en place une infrastructure de données éographiques (IDG) européenne pour favoriser l'harmonisation des pratiques, la mise en réseaux et l'élaboration d'accords de partage et d'accès aux données. Cette directive, approuvée au conseil des ministres de l'Union européenne et par le Parlement européen en avril 2007, a été transposée dans le droit français et mise en application en avril 2009 (CNIG, 2007).

La mise en œuvre d'un catalogue de données doit se faire conformément aux recommandations techniques de la directive qui portent essentiellement sur le choix de la norme de métadonnées. Dans cette directive, la norme ISO 19115 a été retenue, garantissant une interopérabilité totale avec les autres systèmes de catalogage. La norme ISO 19115:2003 a le statut de norme internationale depuis 2003. Cette norme abstraite de contenu définit en les organisant par classes toutes les informations que l'on peut mettre à dispositions pour décrire la donnée. Cette norme s'est affirmée comme une référence pour l'information géographique dans le domaine des métadonnées. Outre le fait qu'elle est enfin disponible après une longue gestation, ses atouts principaux pour la communauté résident dans son caractère modulaire et extensible qui la rend aisément adaptable. Son pendant « technique » est la norme ISO19139 qui définit un schéma XML et les balises correspondant aux différents champs descriptifs de la norme.

Bien que les intérêts de la traçabilité des données soient indéniables, il n'en demeure pas moins qu'une telle entreprise apporte une surcharge de travail. Outre les aspects de mise en place des outils qui nécessitent diverses compétences, c'est le remplissage des métadonnées qui s'avère être la tâche la plus lourde. En effet, ce travail incombe en toute logique à la personne « responsable » de la donnée décrite. Or le passif des métadonnées à remplir est tel que les personnes concernées se voient contraintes d'y consacrer beaucoup de temps. Une vision à long terme de cette tâche justifie d'y passer du temps, mais en pratique, cet aspect se révèle être le plus contraignant pour la mise en place des métadonnées.

5.2 Mise à disposition et diffusion externe

Les données décrites dans un catalogue de métadonnées sont facilement identifiables, il s'agit alors de les rendre consultables. En effet, la directive *Inspire* préconise également la mise à

disposition des données cartographiques via des protocoles spécifiques : la finalité étant la mise en œuvre d'une infrastructure de données spatiales.

Le principe d'une infrastructure de données spatiales est d'offrir la visualisation sur une plateforme commune de données hébergées sur des serveurs physiquement distants les uns des autres.

Ceci est possible grâce à des services web normalisés tels que le Web Map Service (WMS) ou le Web Feature Service (WFS). Ils en existent d'autres, tous définis par l'Open GIS Consortium (OGC). Ces services ont été créés pour faciliter le partage distant de données géographiques par serveurs interposés, en allant jusqu'à la manipulation et à l'édition des données. On mesure tout l'intérêt de pouvoir accéder à tout moment à des données stockées physiquement chez le producteur de ces données : pas de duplicata obsolètes ni de mises à jour régulières des données. Sur une telle plateforme, accessible à l'aide d'un simple navigateur web, tout membre de l'unité peut accéder à un ensemble de données géographiques jusqu'ici réservées à la seule utilisation des experts SIG.

Si l'on se replace dans notre chaîne de traitement, nous voici donc à l'ultime étape. Les données collectées sur le terrain et traitées par le responsable SIG ont été structurées et stockées sur le serveur dédié. Elles sont identifiables facilement via le catalogue de métadonnées, et chacun peut les manipuler à l'aide d'outils simples disponibles sur une interface intégrée (**figure 8**).

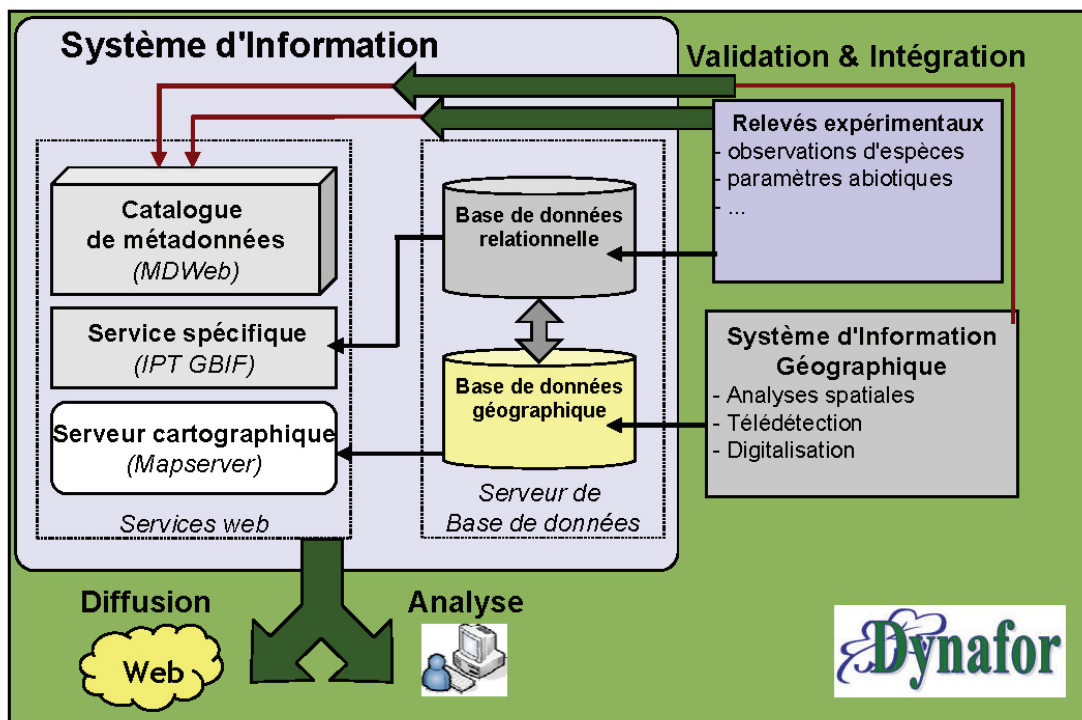


Figure 8 : architecture du système d'information complet de Dynafor

Des cas concrets de systèmes similaires existent déjà comme l'Observatoire du Développement Rural⁶ (ODR) géré par une unité de service de l'Inra de Toulouse et Dynafor est membre de l'Observatoire Spatial Régional (OSR). L'OSR est aujourd'hui un service d'observation labellisé par l'Institut national des sciences de l'univers (INSU). C'est un dispositif d'organisation des systèmes d'observation dévolus aux suivis de long terme sur le fonctionnement et l'évolution des surfaces continentales aux échelles du paysage et de la région, et de la valorisation des données et informations à des fins scientifiques (fonctionnement des surfaces continentales) et appliquées (gestion durable et intégrée des territoires). Les principales fonctions et moyens de l'OSR sont :

- d'assurer la collecte et le traitement d'un ensemble de mesures et d'informations issues de mesures *in situ*, d'enquêtes et de données de télédétection ;
- d'organiser et de gérer ces données dans un système d'information ;
- de diffuser ces données, associées à des outils d'analyse, au moyen de l'Internet.

Conclusion

Le système d'information géré par Dynafor coordonne ainsi, grâce à l'information, les activités de recherche menée sur son site d'étude à long terme et lui permet ainsi d'atteindre ses objectifs de gestion durable des ressources naturelles et de l'espace rural dans le cadre de l'écologie du paysage. La constitution d'un SI mobilise des compétences variées et complémentaires dans le domaine de l'informatique. Le système d'information se construit autour de processus "métier" et leurs interactions, et non simplement autour de bases de données ou de logiciels informatiques.

Ce SI garantit une pérennité des recherches passées et présentes en facilitant la transparence et la qualité des données recueillies *in situ*. Il doit être chemin faisant, c'est-à-dire évolutif, modulaire pour appuyer les recherches scientifiques futures. C'est un vrai outil d'intégration scientifique. Sa mise en place requiert des compétences informatiques et techniques spécifiques qui mobilisent des métiers différents ainsi qu'un matériel informatique performant. Les métadonnées ont une fonction déterminante dans ce système d'information complet car elles servent de relais permettant l'utilisation adéquate des données par les autres utilisateurs. Il est important de noter qu'actuellement, la difficulté ne réside plus véritablement dans l'acquisition et l'analyse de données volumineuses, mais bien dans leur valorisation et leur traçabilité. Les performances techniques sont peu voire plus du tout restrictive, en revanche, la connaissance des données accumulées pose un véritable problème de gouvernance de l'information.

Remerciements : Ce travail a été financé par 2 projets ANR : POPULAR (ANR-06-PADD-014-004 ; 2006-2010) et BiodivAgriM (ANR-07-BDIV-002-03 ; 2008-2011) pour le recueil de données et par le département SAD pour l'achat du serveur de stockage.

⁶ Il s'agit, pour l'essentiel, d'une plate-forme informatique associant une base de données et des fonctionnalités permettant d'une part de réaliser des traitements statistiques et cartographiques et d'autre part de gérer les modalités d'accès des différentes catégories d'utilisateurs aux informations contenues dans cette base. Ces dernières sont relatives à la mise en oeuvre des différents dispositifs du second pilier, aux caractéristiques des exploitations et des territoires.

Bibliographie

- Allen TFH., Starr TB. (1982) Hierarchy Perspectives for ecological complexity. *The University of Chicago Press*, Chicago.
- Balent G., Courtiade B. (1992) Modelling bird communities/landscape patterns relationships in a rural area of South-Western France, *Landscape Ecology*, 6.3, p. 195-211.
- Beyer H. L. (2004) Hawth's Analysis Tools for ArcGIS. Disponible sur <http://www.spatial ecology.com/htools>. Consulté le 12/03/10.
- Blondel, J., Ferry C., Frochot B. (1970) Méthode des indices ponctuels d'abondance (I.P.A.) ou des relevés d'avifaune par « stations d'écoute », *Alauda*, Vol. 37, n°1, ,57-71.
- Bucher B. (2002) *L'aide à l'accès à l'information géographique : un environnement de conception coopérative d'utilisations de données géographiques*. Thèse. Université Paris VI. p. 207.
- Collet C. (2005) Analyse spatiale, géomatique et systèmes d'information géographique, *Revue Internationale de Géomatique*, 15 (4), 393-414.
- Craglia M. (2003) *Towards a European Approach to Metadata for Geographic Information*. In European Commission GI&GIS Portal.
- Deconchat M., Gibon A. et coll. (2007) How to set up research framework to analyse socio-ecological interactive processes in a rural landscape. *Ecology and Society*, 12, 15. [online] URL: <http://www.ecologyandsociety.org/vol12/iss1/art15>
- Fraisse S., Pornon H. (2008) Les métadonnées : corvée ou nécessité ? SIG expert - juin/juillet 2008, n°63, 29-35.
- Gardarin G. (2005) *Bases de données*, Eyrolles, sixième édition.
- Heintz W. et Guéro M.-C. (2005) *Projet SInPa : Systèmes d'Informations Partagées pour la gestion forestière régionale*. De l'observation des écosystèmes forestiers à l'information sur la forêt – 2 et 3 fév. 2005, Paris, France.
- Joliveau T. (2004) Géomatique et gestion environnementale du territoire. Recherches sur un usage géographique des SIG, *Mémoire d'Habilitation à Diriger des Recherches en Sciences Humaines*, Rouen, Université de Rouen, 2 vol.
- Lembo A., Wagenet L., Schusler T., De Gloria S. (2007) *Creating affordable Internet map server applications for regional scale applications*. *Journal of Environmental Management* 85, 1120-1131
- Michener W.K. (2005) Meta-information concepts for ecological data management. *Ecological Informatics* 1
- Monteil C, Deconchat M, Balent G. (2005) Simple neural network reveals unexpected patterns of bird species richness in forest fragments. *Landscape Ecology* 20, 513-527.
- Najar C. *et al.* (2006) Spatial Data and Metadata Integration for SDI Interoperability. Article under Review for the International Journal of Spatial Data Infrastructures

- Noucher M., De Sede-Marceau MH, Golay F. et Pornon H. (2007) *Contributions socio-cognitives aux dynamiques de coopération inter-organisationnelle autour de la donnée géographique.*, In GéoCongrès, Québec, octobre 2007.
- Noucher M. (2007) *Coopérer autour des SIG : typologie et exemples en France et au Canada*, Atelier SIG Pyrénées (en visioconférence depuis le GéoCongrès de Québec), St Gaudens, octobre 2007.
- Raison L. (2007) L'inventaire des oiseaux nicheurs par la méthode des points d'écoute dans les paysages agricoles – Regard d'un ornithologue - *in* Techniques et pratiques de recueil de données *in situ* - *Le Cahier des Techniques de l'Inra*, numéro spécial, 79-86.