



CONSERVATOIRE NATIONAL DES ARTS ET METIERS

CENTRE REGIONAL RHÔNE-ALPES

CENTRE D'ENSEIGNEMENT DE GRENOBLE

MEMOIRE

présenté par **Raphaël THOMAS**

en vue d'obtenir

LE DIPLOME D'INGENIEUR C.N.A.M.

en INFORMATIQUE

**Evolutions d'outils dédiés à l'analyse territoriale et
à l'analyse spatiale dans le cadre du projet
HyperCarte**

Soutenu le 20 juin 2008

JURY

Présidente : Mme Véronique Donzeau-Gouge

Membres : M. Eric Gressier-Soudan

M. Jean-Pierre Giraudin

M. André Plisson

M. Mathias Voisin-Fradin

M. Jérôme Gensel

Mme Paule-Annick Davoine



CONSERVATOIRE NATIONAL DES ARTS ET METIERS

CENTRE REGIONAL RHÔNE-ALPES

CENTRE D'ENSEIGNEMENT DE GRENOBLE

MEMOIRE

présenté par **Raphaël THOMAS**

en vue d'obtenir

LE DIPLÔME D'INGENIEUR C.N.A.M.

en INFORMATIQUE

**Evolutions d'outils dédiés à l'analyse territoriale et
à l'analyse spatiale dans le cadre du projet
HyperCarte**

Soutenu le 20 juin 2008

Les travaux relatifs à ce mémoire ont été effectués au sein de l'équipe Steamer du Laboratoire d'Informatique de Grenoble (LIG) sous la direction de Jérôme Gensel.

Remerciements

Je tiens à exprimer mes remerciements aux membres du Jury : Véronique Donzeau-Gouge et Eric Gressier-Soudan, professeurs au CNAM Paris ; à Jean-Pierre Giraudin, professeur à l'Université Pierre Mendès France de Grenoble ; à André Plisson, directeur du centre d'enseignement du CNAM de Grenoble ; à Mathias Voisin-Fradin, directeur des études au CNAM de Grenoble ainsi qu'à Paule-Annick Davoine, maître de conférences à l'INPG.

Je remercie également vivement mon tuteur, Jérôme Gensel, professeur à l'Université Pierre Mendès France de Grenoble, pour son accueil, ses conseils et son soutien. Il m'a permis de travailler sur ce passionnant sujet.

Je souhaite également remercier les permanents de l'équipe Steamer. Merci donc à Hervé Martin, professeur à l'Université Pierre Mendès France de Grenoble, de m'avoir accueilli dans son équipe et à Marlène Villanova-Oliver.

Je tiens également à remercier l'ensemble des thésards et des stagiaires de l'équipe Steamer pour leur accueil et leur bonne humeur. Merci donc à Céline, Christine, Dia, Bogdan, José, Laurent, Marius et Windson.

Pour finir, je souhaite remercier l'ensemble des personnes qui m'ont soutenu pendant cette période. Je pense notamment à ma famille et à mes amis. J'ai une pensée particulière pour mon papy.

Remerciements	5
Sommaire.....	7
Table des illustrations.....	9
Liste des tableaux et des extraits de code.....	11
Notations.....	13
Introduction.....	15
1.1. Contexte.....	15
1.2. Le projet HyperCarte	15
1.3. Cahier des charges.....	17
1.4. Organisation du mémoire	18
Chapitre 1 - Etat de l'art	19
1. Préambule.....	20
2. Les Systèmes d'Information Géographique.....	21
2.1. L'acquisition des données.....	23
2.2. L'archivage des données géographiques	26
2.3. L'affichage et la mise en page des données géographiques	26
2.4. L'affichage et la mise en page des données statistiques	28
3. Quelques outils cartographiques existants.....	31
3.1. Un outil cartographique intuitif : Google Maps	31
3.2. Les outils cartographiques dédiés à l'analyse statistique	33
3.3. Synthèse des différents outils présentés.....	39
4. Synthèse	40
Chapitre 2 - Analyse de l'existant	41
1. Un projet européen pour la cartographie interactive	43
1.1. Le programme ESPON	43
1.2. Définition des termes du domaine métier	44
2. Un module de visualisation : HyperAtlas	47
2.1. L'interface utilisateur	47
3. Un module d'administration : HyperAdmin	55
3.1. Problématique	55
3.2. L'interface utilisateur	55
3.3. Les fichiers de données.....	57
3.4. Fonctionnalités	62

4. Un module d'analyse spatiale : HyperSmooth	65
4.1. Principes.....	65
4.2. HyperSmooth	66
5. Architecture et choix technologiques	68
5.1. Le modèle MVC.....	68
5.2. Choix technologiques.....	70
6. Conclusion	72
Chapitre 3 - Maintenance évolutive et évolutions	73
1. Maintenance évolutive sur HyperAtlas	75
2. Réingénierie d'HyperAdmin	82
2.1. Problèmes rencontrés	82
2.2. Solution proposée et mise en œuvre	83
3. Evolutions d'HyperAdmin.....	91
3.1. Les ratios pertinents.....	91
3.2. Les hiérarchies multiples.....	94
4. Livraisons	96
4.1. EEA-HyperAtlas	96
4.2. EP-HyperAtlas	97
4.3. Nordic HyperAtlas	97
4.4. ESPON HyperAtlas.....	98
5. Conclusion	100
Conclusion	101
1.1. Objectifs.....	101
1.2. Bilan des réalisations.....	101
1.3. Perspectives.....	102
1.4. Bilan personnel.....	102
Glossaire	105
Bibliographie	109

Table des illustrations

Figure 1 : L'information géographique selon J. Dangermond.....	22
Figure 2 : Exemple d'orthoimage : la place Verdun à Grenoble.	23
Figure 3 : Région agricole de Kozloduy (Bulgarie).	24
Figure 4 : La Nouvelle Orléans avant et après le passage de Katrina.	25
Figure 5 : Une image raster et son équivalent en image vectorielle.	27
Figure 6 : Exemple d'affichage combiné : la place Verdun de Grenoble.	27
Figure 7 : La population au Viêt-Nam en 1989.	28
Figure 8 : Exemple d'une carte choroplèthe.	29
Figure 9 : L'onglet "Mes Cartes" de Google Maps.	32
Figure 10 : Une carte personnalisée sur Google Maps.....	32
Figure 11 : Copie d'écran de GeoClip.	34
Figure 12 : Une carte générée par le Cartographeur.....	35
Figure 13 : Choix des paramètres de l'analyse.	37
Figure 14 : Choix de la méthode de discrétisation.	38
Figure 15 : Affichage d'une carte et de sa légende sous PhilCarto.	38
Figure 16 : L'affichage des zones survolées dans la barre de titre.	39
Figure 17 : Diagramme de classes : hiérarchie des unités.	44
Figure 18 : Diagramme de classes : relation entre l'aire d'étude et les unités.	44
Figure 19 : Les différents maillages NUTS.....	45
Figure 20 : Diagramme de classes : relation entre les unités territoriales et le maillage administratif.	46
Figure 21 : Diagramme de classes : relation entre les unités et les indicateurs.	46
Figure 22 : L'interface utilisateur d'HyperAtlas.	47
Figure 23 : Sélection des paramètres "Area and Zoning".	48
Figure 24 : Influence du choix de l'espace d'étude.	48
Figure 25 : Le choix des indicateurs.....	49
Figure 26 : le paramétrage des déviations.	49
Figure 27 : Les différents composants d'une carte.	50
Figure 28 : Les cartes à disques sous HyperAtlas.....	51
Figure 29 : Une carte de ratio.....	52
Figure 30 : Les cartes de déviation.	52
Figure 31 : La carte de synthèse.	53
Figure 32 : Histogramme récapitulatif présentant la position de la Région Rhône-Alpes face au taux de chômage selon différentes échelles.	54

Figure 33 : Image de chargement d'HyperAdmin.....	56
Figure 34 : L'interface utilisateur d'HyperAdmin.	56
Figure 35 : Création d'un projet.	63
Figure 36 : Création de matrices de voisinages.	64
Figure 37 : Analyse territoriale.	65
Figure 38 : Analyse spatiale.	65
Figure 39 : Architecture d'HyperSmooth.....	66
Figure 40 : Image Raster construite par HyperSmooth (non plaquée).....	67
Figure 41 : Image raster plaquée sur le fond de carte.	67
Figure 42 : Le modèle MVC.	68
Figure 43 : la nouvelle zone de paramétrage des indicateurs.....	76
Figure 44 : Diagramme de classes : composition d'un ratio.	76
Figure 45 : La classe HCRatio.	76
Figure 46 : Classe centrale mal positionnée sur l'ancienne légende.	77
Figure 47 : Nouvelles palettes de couleurs pour les distributions divergentes.....	79
Figure 48 : Exemple de légende avec la nouvelle méthode de discrétisation dans le cas d'une distribution non symétrique.	79
Figure 49 : Illustration du changement de la méthode de discrétisation.	80
Figure 50 : Visualisation du taux de chômage à Paris.	81
Figure 51 : Mauvais message d'avertissement.....	83
Figure 52 : Enchaînement des différents scénarii du <i>Wizard</i>	84
Figure 53 : Fenêtre type du <i>Wizard</i>	85
Figure 54 : Présentation de différents écrans du <i>Wizard</i>	86
Figure 55 : Ecran de progression pendant la création d'un projet.	88
Figure 56 : Image de chargement d'HyperAdmin.....	90
Figure 57 : Gestion des ratios depuis le <i>Wizard</i>	92
Figure 58 : Diagramme de classes : les différentes entités au sein du projet HyperCarte.....	93
Figure 59 : Exemple de maillages non compatibles.	94
Figure 60 : Incidence sur le calcul des stocks sur une hiérarchie non compatible. ...	95
Figure 61 : Les différents logos d'HyperAtlas.....	96
Figure 62 : Comparaison du niveau de détail des pays nordiques.	98

Liste des tableaux et des extraits de code

Tableaux

Tableau 1 : Synthèse des fonctionnalités.....	39
Tableau 2 : Le paradigme MVC dans Swing.....	69
Tableau 3 : Caractéristiques de l'EEA-HyperAtlas.....	96
Tableau 4 : Caractéristiques de l'EP-HyperAtlas.....	97
Tableau 5 : Caractéristiques du Nordic HyperAtlas.....	97
Tableau 6 : Caractéristiques de l'ESPON HyperAtlas.....	98

Extraits de code

Code 1 : Extrait de fichiers Mif-Mid.....	57
Code 2 : Syntaxe du fichier .mif.....	58
Code 3 : Extrait des fichiers de données d'HyperAdmin : Area.....	58
Code 4 : Extrait des fichiers de données d'HyperAdmin : AreaLanguage.....	59
Code 5 : Extrait des fichiers de données d'HyperAdmin : Zoning.....	59
Code 6 : Extrait des fichiers de données d'HyperAdmin : ZoningLanguage.....	59
Code 7 : Extrait des fichiers de données d'HyperAdmin : Unit.....	59
Code 8 : Extrait des fichiers de données d'HyperAdmin : UnitLanguage.....	60
Code 9 : Extrait des fichiers de données d'HyperAdmin : UnitSup.....	60
Code 10 : Extrait des fichiers de données d'HyperAdmin : UnitArea.....	60
Code 11 : Extrait des fichiers de données d'HyperAdmin : UnitZoning.....	60
Code 12 : Extrait des fichiers de données d'HyperAdmin : Contiguity.....	60
Code 13 : Extrait des fichiers de données d'HyperAdmin : ContiguityLanguage.....	61
Code 14 : Extrait des fichiers de données d'HyperAdmin : UnitContiguity.....	61
Code 15 : Extrait des fichiers de données d'HyperAdmin : Neighbourhood.....	61
Code 16 : Extrait des fichiers de données d'HyperAdmin : NeighbourhoodLanguage.....	61
Code 17 : Extrait des fichiers de données d'HyperAdmin : StockLanguage.....	62
Code 18 : Extrait des fichiers de données d'HyperAdmin : Stock.....	62
Code 19 : Algorithme d'affectation des classes pour la nouvelle méthode de discrétisation.....	78
Code 20 : Les erreurs de topologies.....	89
Code 21 : Extrait des fichiers de données d'HyperAdmin : RatioStock.....	91
Code 22 : Extrait du fichier CreateRatios.sql.....	93

Notations

Les termes écrits dans le style suivant : « **EXEMPLE** » seront explicités dans le glossaire à la fin de ce document. Seule la première occurrence des mots listés dans ce glossaire sera mise en évidence par ce style.

1.1. Contexte

Ce document s'inscrit dans le cadre de mon mémoire d'ingénieur CNAM en informatique. A la suite de mon probatoire, soutenu en juin 2006 et intitulé « La persistance et le paradigme MVC dans les applications Web », M. Jérôme Gensel m'a accueilli au sein de l'équipe STEAMER afin de travailler sur un projet de **CARTOGRAPHIE** avec le groupe de recherche HyperCarte.

On pourrait définir la cartographie comme étant une science dédiée à la réalisation et à l'analyse de **CARTES**. De nos jours, avec l'essor des différentes technologies liées à Internet, certains aspects de la cartographie sont accessibles au plus grand nombre. Citons ainsi les sites permettant de calculer son itinéraire routier (Mappy¹, ViaMichelin² ou encore Google Maps³), de visualiser notre planète (GeoPortail⁴, Google Earth⁵). Cependant les **SYSTEMES D'INFORMATION GEOGRAPHIQUE** (SIG) sont utilisés dans de nombreux domaines différents. Ainsi, certains peuvent être dédiés à la surveillance et à la localisation de phénomènes naturels comme les séismes, alors que d'autres permettent de visualiser des données socio-économiques. La représentation spatiale de ces données statistiques est l'un des objectifs du groupe de recherche HyperCarte.

1.2. Le projet HyperCarte

Le contexte de ce mémoire est axé sur le projet HyperCarte. Ce groupe de recherche a été fondé en 1996 et s'est forgé une reconnaissance aussi bien au niveau national qu'au niveau international dans le domaine de la cartographie interactive. Plus précisément, ses thèmes de recherche s'articulent autour de l'analyse territoriale **MULTISCALEAIRE** et de **L'ANALYSE SPATIALE**, afin de mettre au point une méthodologie de cartographie interactive permettant d'analyser et de visualiser les multiples représentations d'un même phénomène. Ainsi, le groupe de recherche HyperCarte a développé des outils permettant d'analyser des données socio-économiques dans un espace d'étude (un district scolaire, un quartier, un département, une région, un pays) par rapport à un espace de référence (comme l'Europe des 27). Ces analyses peuvent utiliser le **MAILLAGE** administratif dans le cadre de l'analyse territoriale mais elles peuvent également en faire abstraction : c'est le rôle de l'analyse spatiale. Dans le cadre du projet HyperCarte, ces deux types d'analyse sont supportés et implémentés dans deux logiciels distincts. On retrouve donc le logiciel **HyperAtlas** pour l'analyse territoriale multiscalaire et **HyperSmooth** pour l'analyse spatiale.

Les données traitées par ces deux outils se classent dans différentes catégories : des données statistiques (par exemple : démographiques, environnementales) et des données géographiques (localisation et description des formes géométriques).

¹ <http://www.mappy.fr>

² <http://www.viamichelin.com>

³ <http://maps.google.com>

⁴ <http://www.geoportail.fr>

⁵ <http://earth.google.fr>

L'ensemble de ces données est transmis à un troisième logiciel : **HyperAdmin** qui construit un jeu de données directement exploitables par les autres programmes développés par le groupe de recherche HyperCarte.

1.2.1. Les objectifs du projet

Les objectifs du projet HyperCarte consistent à concevoir et développer des applications dédiées à la cartographie interactive permettant l'analyse de phénomènes sociaux ou encore environnementaux.

Les « atlas papier » permettant seulement la représentation statique de phénomènes et les SIG ne possédant pas de capacités d'analyse spatiale évoluées, le groupe de recherche HyperCarte s'est lancé dans la cartographie interactive en partant du constat qu'un phénomène spatial peut faire l'objet d'un nombre infini de représentations. Le logiciel HyperAtlas remplit ces objectifs en étant un SIG, c'est-à-dire un outil d'aide à la décision pour ses utilisateurs qui peuvent produire des cartes en changeant à la volée les paramètres de la carte. Cet outil se présente comme étant un module d'analyse territoriale multiscalaire car il permet l'analyse d'un même phénomène spatial à différentes échelles.

Outre ce module d'analyse territoriale multiscalaire, le projet HyperCarte a souhaité également se doter d'un module d'analyse spatiale multiscalaire. Le logiciel HyperSmooth en est le résultat.

Un objectif du projet HyperCarte est également de se faire reconnaître par la communauté internationale pour avoir davantage de moyens pour continuer son développement. C'est pourquoi, HyperAtlas peut être personnalisé pour répondre aux demandes des clients notamment de la part des institutions qu'elles soient locales, nationales, européennes ou internationales. Citons ainsi quatre déclinaisons d'HyperAtlas :

- **ESPON**-HyperAtlas dans le cadre de l'observation de l'aménagement du territoire européen ;
- EP-HyperAtlas : version livrée au Parlement Européen ;
- EEA-HyperAtlas pour la version remise à l'Agence Européenne de l'Environnement ;
- Nordic-HyperAtlas : version spécifique aux pays nordiques (Finlande, Norvège, Suède) réalisée pour NordRegio.

Le groupe de recherche HyperCarte s'appuie sur des équipes provenant de domaines différents (informatique, géographie, sciences sociales). Il s'agit donc d'un projet pluridisciplinaire. La partie suivante décrit ces équipes et leur rôle dans ce projet.

1.2.2. Les acteurs du projet

Le groupe de recherche HyperCarte est composé de quatre équipes. Chacune apportant une compétence précise autour du thème de l'analyse territoriale et spatiale multiscalaire.

- L'équipe PARIS de l'U.M.R. Géographie-Cités ;
- L'U.M.S. RIATE ;
- L'équipe MESCAL du Laboratoire d'Informatique de Grenoble (LIG) ;
- L'équipe STEAMER du Laboratoire d'Informatique de Grenoble (LIG).

Les principaux objectifs de l'équipe PARIS (acronyme de « Pour l'Avancement des Recherches en Interaction Spatiale ») sont de décrire, représenter et de modéliser l'organisation de l'espace géographique. Dans le cadre d'HyperCarte, cette équipe apporte son savoir-faire dans les domaines de la géographie et de la statistique (sciences sociales).

La mission de l'UMS RIATE est le recensement et la valorisation des compétences scientifiques françaises en matière d'aménagement du territoire européen. Elle est le point focal français pour l'Observatoire en Réseau de l'Aménagement du Territoire Européen (**ORATE**).

La problématique de l'équipe MESCAL, quant à elle, a trait aux calculs distribués. Dans le cadre du projet HyperCarte, l'équipe MESCAL s'occupe de la parallélisation sur grappes de PC afin de pouvoir calculer en temps réel des cartes. Cette question de performance est primordiale dans le cadre d'HyperSmooth.

Enfin, l'équipe STEAMER apporte à ce groupe de recherche ses connaissances dans les systèmes d'information et le multimédia. Cette équipe a pour rôle de concevoir et développer les outils cartographiques. C'est au sein de cette équipe, que j'ai effectué mon stage et mon mémoire a été alimenté par les différents travaux et améliorations réalisés pendant cette année sur les différents outils du projet HyperCarte. Le paragraphe suivant présente mon cahier des charges.

1.3. Cahier des charges

Le projet HyperCarte a défini une architecture dédiée à l'analyse territoriale de divers espaces d'étude. L'application **HyperAtlas** permet de générer à la volée et de visualiser un ensemble de cartes rendant compte de la distribution de différents indicateurs socio-économiques (population, PNB, taux de chômage) dans les limites du ou des maillages territoriaux choisis pour l'étude. Ces cartes, validées par des experts géographes, sont jugées pertinentes par rapport aux analyses attendues par les utilisateurs (géographes, politiques et grand public). Cette interface permet la visualisation de six cartes spécialisés : deux cartes à disques proportionnels des deux indicateurs étudiés, une carte choroplèthe du ratio de ces indicateurs, trois cartes de déviation par rapport aux unités territoriales de même niveau hiérarchique (commune, département, région ou pays), de niveau hiérarchique immédiatement supérieur (comparaison commune/département, département/région, région/état, etc.), ou aux unités voisines (partage d'une frontière administrative), et d'un tableau de synthèse ainsi que des histogrammes des trois déviations considérées. La création et l'intégration de nouveaux jeux de données sont effectuées par **HyperAdmin** : le module d'administration permettant de créer des fichiers directement exploitables par HyperAtlas.

Durant cette année, ma mission est de faire évoluer ces deux outils. D'une part, de nouvelles fonctionnalités doivent être développées pour HyperAtlas afin de pouvoir satisfaire aux exigences des différents partenaires. D'autre part, l'acquisition de données avec HyperAdmin n'est pas jugée convaincante en raison de problèmes d'ergonomie. Une de mes tâches consiste donc à créer un assistant pour faciliter l'intégration de ces données.

1.4. Organisation du mémoire

Ce mémoire est composé de quatre parties présentant :

- Un état de l'art sur le domaine de la cartographie ;
- L'étude de l'existant autour des applications HyperAtlas et HyperAdmin ;
- L'ensemble des améliorations apportées à ces logiciels ;
- Une conclusion.

Plus précisément, suite à cette introduction, la prochaine partie présente la cartographie interactive et les systèmes d'information géographique. Nous décrivons comment sont récupérées, stockées, traitées et représentées les données dans un logiciel de cartographie. Cette partie présente quelques SIG dédiés à l'analyse spatiale de données statistiques.

Le deuxième chapitre présente les différents modules du projet HyperCarte. Nous abordons ainsi les modules HyperAtlas et HyperAdmin. Nous définissons l'architecture qui a été retenue par les précédents concepteurs. Enfin, nous évoquons HyperSmooth : le module dédié à l'analyse spatiale.

Les évolutions apportées au cours de cette année sont rapportées dans le troisième chapitre. Ce chapitre relate les différents aspects de mon travail : de la maintenance évolutive sur HyperAtlas à la réalisation d'un assistant à l'intégration de données géographiques pour HyperAdmin. Enfin, nous retraçons également un historique des différentes livraisons d'HyperAtlas.

La conclusion porte sur le bilan du travail réalisé pendant cette année au sein de l'équipe STEAMER et du groupe de recherche HyperCarte. Elle est également accompagnée d'un bilan personnel.

CHAPITRE 1

Etat de l'art

L'état de l'art de ce mémoire porte sur la cartographie interactive sur Internet. De nombreux outils permettent aujourd'hui de visualiser des cartes mais également, pour certains, d'en créer. Dans un premier temps, nous étudions les Systèmes d'Information Géographique en définissant le vocabulaire essentiel de ce domaine métier. Nous présentons ensuite les différents outils disponibles en nous intéressant aussi bien à leurs fonctionnalités qu'aux publics visés.

Sommaire du chapitre :

Chapitre 1 - Etat de l'art	19
1. Préambule.....	20
2. Les Systèmes d'Information Géographique.....	21
2.1. L'acquisition des données.....	23
2.2. L'archivage des données géographiques	26
2.3. L'affichage et la mise en page des données géographiques	26
2.4. L'affichage et la mise en page des données statistiques	28
3. Quelques outils cartographiques existants.....	31
3.1. Un outil cartographique intuitif : Google Maps	31
3.2. Les outils cartographiques dédiés à l'analyse statistique	33
3.2.1. GeoClip	33
3.2.2. Cartes & Données.....	35
3.2.3. PhilCarto	36
3.3. Synthèse des différents outils présentés.....	39
4. Synthèse	40

1. *Préambule*

La cartographie est la science dédiée à la réalisation ainsi qu'à l'étude des cartes. Une carte met en valeur un espace géographique et la localisation des éléments (villes, routes, hôtels) dans cet espace. Ces cartes sont ensuite utilisées dans des applications et des domaines très variés : géographie, météorologie, démographie, etc. Traditionnellement, elles sont représentées par un planisphère ou un globe terrestre.

A la fin des années 60, la cartographie commence à s'intéresser à l'outil informatique. Les premiers outils de production automatique de cartes apparaissent. Les représentations cartographiques sont d'abord au format vectoriel. Les images rasters seront créées quelques années plus tard par l'**ESRI**⁶. Ces deux formats d'images sont décrits dans les paragraphes suivants. Les domaines d'application sont d'abord liés à des applications militaires, l'étude des ressources naturelles ou de l'urbanisme. Ce début de la cartographie a été rendu possible par la montée en puissance des ordinateurs. Il faut noter également qu'à cette époque, la télédétection spatiale a permis d'obtenir des images satellites de la Terre qui ont ainsi pu être peu à peu intégrées dans les premiers outils de cartographie automatique.

Durant les années 80, le développement des outils cartographiques automatiques a continué. Durant cette période, l'interactivité des interfaces graphiques des stations de travail a été nettement améliorée et le travail autour du thème des bases de données relationnelles a permis un essor d'outils au service du raisonnement cartographique : les systèmes d'information géographique.

A la fin du XX^{ème} siècle, la diffusion des systèmes d'information géographique s'est accrue. En effet, l'informatique a encore beaucoup évolué. L'intégration des données est également possible grâce à la télédétection (spatiale et aérienne) ainsi que par les capteurs GPS. Ces nouveaux moyens ont permis aux SIG de se faire connaître du grand public.

Ces dernières années, avec l'essor des technologies liées à Internet, il est désormais possible pour n'importe quel internaute de créer et de visualiser des cartes directement sur le Web.

Si l'informatique a apporté à la cartographie un nouveau mode de représentation, on remarquera également que cela a permis de nouvelles applications. Ainsi, la **GEOLOCALISATION** est devenue un service incontournable aujourd'hui.

Dans la partie suivante, nous allons définir plus précisément cette notion de Système d'Information Géographique

⁶ ESRI : *Environmental Systems Research Institute*

2. *Les Systèmes d'Information Géographique*

Tous les outils de cartographie que nous présentons dans ce chapitre ont un point commun : ils reposent tous sur un Système d'Information Géographique. Un système d'information géographique, communément appelé SIG, est un système informatique qui permet à partir de différentes sources de rassembler, d'organiser, de gérer, d'analyser et de combiner mais aussi de présenter des informations localisées géographiquement contribuant notamment à la gestion de l'espace [DIERCKX]. Un SIG combine la puissance d'une base de données (les requêtes, les analyses statistiques) et la représentation de l'information sur des cartes. Ces capacités d'analyse géographique permettent au SIG de répondre à de nombreux besoins.

De nombreux domaines utilisent les SIG pour répondre à leurs problématiques. Ainsi, on peut visualiser des cartes interactives lorsque l'on cherche un itinéraire, les résultats d'une élection ou encore la répartition du taux de chômage en France. De plus, la carte représente un support de communication attractif et synthétique qui permet de diffuser un message, une information, facilement compréhensibles.

Cette définition apporte quelques indications sur le rôle et les fonctionnalités d'un SIG. Ces fonctionnalités sont souvent nommées les « 5A » :

- Acquisition des données sous forme numérique ;
- Archivage des données de manière persistante ;
- Analyse des données (manipulations et requêtes sur les données) ;
- Affichage et mise en page des données ;
- Abstraction du monde réel.

Les données jouent donc un rôle primordial dans les outils de cartographie interactive. Jack Dangermond, fondateur et président de l'ESRI, estime que l'information géographique est composée de données géographiques, statistiques et historiques (cf. Figure 1) [Dangermond, 1990]. Ces trois composants renseignent sur le type de l'information à représenter, sa localisation mais aussi son époque. Cependant, les données historiques ne sont pas faciles à gérer pour un SIG car, si au cours des années les données statistiques évoluent, cela peut également être le cas pour les données géographiques. En effet, les frontières de certaines unités (pays, villes) peuvent être différentes entre deux périodes.

Information géographique					Données historiques	Instant 1
						Instant 2
						Instant 3
Données géographiques						
Données de localisation			Données statistiques			
Localisation (x,y)			Localisation topographique			
Point	Ligne	Polygone	Grille	Réseau	variables	valeurs
					classes	noms

Figure 1 : L'information géographique selon J. Dangermond.

Etant donné que peu de SIG prennent en compte ces données temporelles en raison de la complexité due aux transformations du **MAILLAGE**, on ne distingue, dans ce document, que deux types de données :

- Les données géographiques ;
- Les données tabulaires.

Tout d'abord, on trouve des informations relatives à la géométrie. Cette catégorie contient les données relatives aux formes ainsi que les **COORDONNEES** géographiques des objets. Par exemple, la liste des **POLYGONES** qui représentent la France. Ce sont des données de localisation. Ces coordonnées peuvent être explicites ou implicites. Ainsi, des informations comme la **LATITUDE**, la **LONGITUDE** ou encore une grille de coordonnées nationales sont des références géographiques explicites. Par contre, des informations comme une adresse, un code postal, sont des données implicites. Pour localiser des données implicites afin de les analyser, il faut, au préalable, les convertir en références explicites. Ce processus s'appelle le géocodage.

On recense également des données attributaires associées à un objet qui nous renseignent sur un phénomène (exemple : la population française). Ce sont les données statistiques. On peut également trouver des données historiques qui permettent de rendre compte de l'évolution d'un phénomène au cours du temps. Comme nous l'avons déjà vu, ce genre de données peut engendrer quelques problèmes de cohérence des données lorsque le maillage a également changé au cours des années.

Citons enfin les paramètres d'affichage des objets qui sont davantage liés à l'outil informatique et aux préférences de l'utilisateur. Ces données graphiques peuvent ainsi être : l'épaisseur du contour de la France, la couleur des contours, la couleur de remplissage, etc.

2.1. L'acquisition des données

La collecte de données géographiques est l'une des activités de la **GEOMATIQUE**. Il s'agit d'une discipline qui combine la géographie et l'informatique.

Cette collecte peut s'effectuer par différents moyens :

- La photographie aérienne ;
- Les images satellites ;
- La numérisation de documents existants ;
- Le relevé direct.

Les données issues de la photographie aérienne doivent d'abord être traitées avant d'être utilisées. Ce traitement permet de corriger les aberrations optiques. Les images sont ensuite égalisées radiométriquement. Les images alors obtenues sont appelées **ORTHOIMAGES** (ou encore **ORTHOPHOTOGRAPHIES**). Les aberrations optiques sont dues à l'inclinaison de l'appareil photographique lors de la prise de vue ou encore à la distorsion causant la déformation de l'image. Sur l'image résultante, on a donc l'impression que la photographie a été prise à la verticale de tous les points la composant. Il faut ensuite assembler les images de façon à construire un diaporama ou une mosaïque. A cette étape, il faut encore procéder à des retouches afin d'obtenir une image homogène afin de masquer les raccords et de corriger les différences de couleurs. Ce processus d'équilibre chromatique est appelé processus de correction radiométrique.

Le logiciel Google Earth [Google Earth] permet de visualiser des orthoimages. Voici, un exemple d'orthoimage avec la prise de vue de la place Verdun à Grenoble.

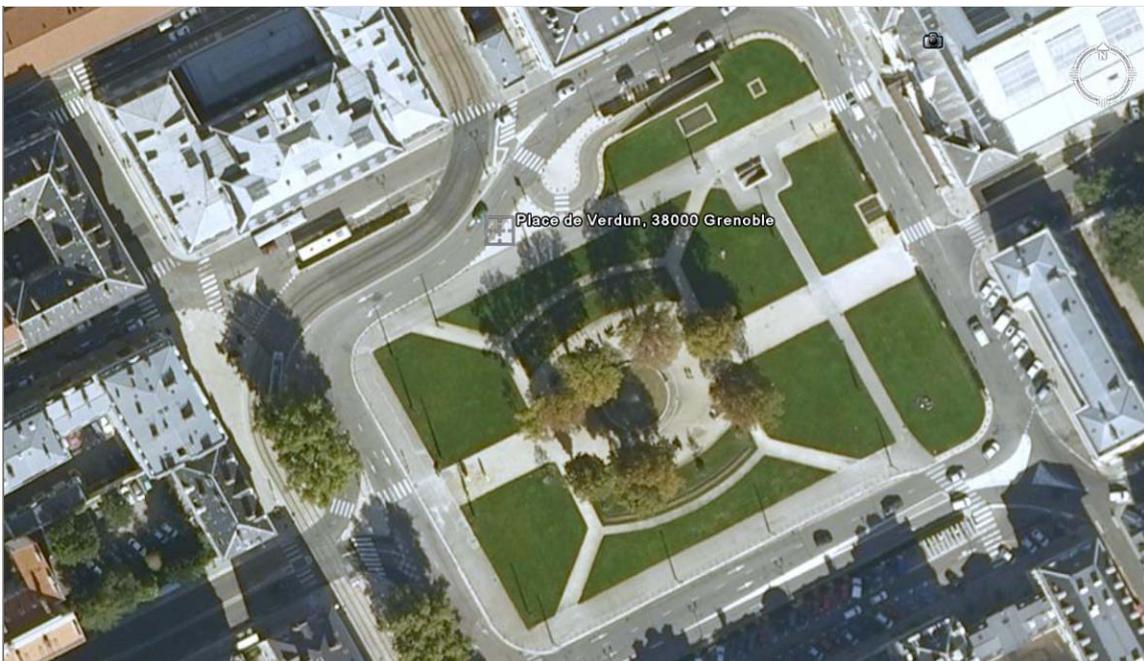


Figure 2 : Exemple d'orthoimage : la place Verdun à Grenoble.

Les satellites permettent également d'obtenir des données et des images de la Terre. Parmi les différents programmes spatiaux dédiés à l'observation de notre planète, nous citerons les programmes SPOT et LANDSAT.

Les satellites SPOT, acronyme de « Satellites Pour l'Observation de la Terre », sont issus d'une collaboration européenne menée par la France à la fin des années 1970. L'utilisation des images provenant de ces satellites (le plus récent, SPOT5, a été lancé en 2002) permet de créer des cartes au 1:50000 mais donne également des évaluations sur les dégâts liés à des intempéries (le vent par exemple). La Communauté Européenne utilise également des images issues de SPOT dans le cadre de sa politique agricole commune (PAC) afin de contrôler les plantations et leur stade de maturation. Le programme SPOT est sur le point d'être remplacé par le projet PLEIADES issu d'une collaboration entre la France (le CNES⁷) et l'Italie (l'ASI⁸). La vocation de ce programme est double puisque les données seront utilisées à des fins civiles mais aussi militaires. [©CNES]



Figure 3 : Région agricole de Kozloduy (Bulgarie).

La Figure 3 présente une région agricole en Bulgarie. A partir de cette image, il est possible de déterminer la nature des plantations. En la comparant avec d'autres images de cette zone prises à d'autres instants, il est également possible de quantifier le rendement agricole de chaque parcelle.

Notons également que depuis le début de l'année 2007, Google Earth et SPOT Image se sont entendus pour la diffusion d'images provenant de SPOT sur le logiciel de cartographie de Google.

⁷ CNES : Centre National d'Etudes Spatiales, l'agence française de l'Espace

⁸ ASI : *Agenzia Spaziale Italiana*, l'agence italienne de l'Espace

Le programme LANDSTAT est, quant à lui, l'un des programmes d'observation de la surface terrestre le plus ancien. Ce projet américain (NASA⁹) produit des images servant à la cartographie, l'étude des changements climatiques ainsi que pour déterminer l'occupation des sols (zones agricoles, zones boisées).



Figure 4 : La Nouvelle Orléans avant et après le passage de Katrina.

La Figure 4 montre ainsi l'effet dévastateur de l'ouragan Katrina. L'image de gauche a été prise en 2000, tandis que celle de droite date de 2005, quelques jours après le passage de la tempête. Ces deux photos renseignent sur l'occupation des sols et notamment ici sur les zones inondées et l'accroissement du lac Pontchartrain. [Landstat]

Deux autres méthodes permettant d'obtenir des données géographiques sont la numérisation de documents existants comme le cadastre par exemple et le relevé direct sur le terrain par des géomètres à l'aide de GPS.

Ces différentes méthodes nous donnent en premier lieu des données géographiques : coordonnées d'une parcelle agricole à Kozloduy par exemple. S'il est possible d'obtenir de ces images des données d'ordre statistique (comme la part de zones boisées dans une région), il faut la plupart du temps avoir recours à d'autres moyens pour obtenir des indications précises sur certains phénomènes. Ainsi, pour les phénomènes socio-économiques en France, on pourra s'aider des données issues de l'INSEE¹⁰ pour connaître la population active de la ville de Grenoble, par exemple, ou encore des données issues des différents ministères. Ainsi, le site du Ministère de l'Intérieur¹¹ peut être consulté pour obtenir les résultats du second tour des dernières élections présidentielles pour l'Isère, tandis que le site du Ministère de l'Education Nationale¹² propose des statistiques sur les examens nationaux. Pour des statistiques au niveau régional et supra-régional, le site de l'EuroStat¹³ propose des données pour les niveaux NUTS1 (les pays) au NUTS3 (l'équivalent des régions en France) (cf. **NUTS**) sur de nombreux domaines (éducation, développement durable). On recense également des méthodes de collecte de données statistiques par le système du « porte à porte » notamment pour la réalisation de sondages ou encore le recensement de la population.

⁹ NASA : *National Aeronautics and Space Administration*, l'agence gouvernementale américaine responsable du programme spatial des Etats-Unis

¹⁰ <http://www.insee.fr>

¹¹ <http://www.interieur.gouv.fr>

¹² <http://www.education.gouv.fr>

¹³ <http://epp.eurostat.ec.europa.eu>, service statistique de la Commission Européenne

2.2. L'archivage des données géographiques

L'archivage des données consiste à rendre persistantes les données récoltées en les transférant sur un espace d'archivage, par exemple un disque dur. On précise que l'on compte plusieurs moyens de stocker ces données.

Premièrement, un Système de Gestion de Base de Données (SGBD) peut permettre d'archiver des données géographiques. Certains SGBD proposent en effet des extensions pour gérer les données spatiales. Citons notamment PostGreSQL¹⁴ et son extension postGIS¹⁵, mais aussi MySQL¹⁶ et son extension openGIS¹⁷ ou encore Oracle¹⁸.

Ces données peuvent également être enregistrées dans des fichiers propres aux SIG. Notamment les fichiers MIF-MID de MapInfo mais aussi les Shapefiles propres à ArcInfo.

2.3. L'affichage et la mise en page des données géographiques

Après avoir récolté, stocké et interrogé les données, un SIG doit pouvoir afficher une représentation de l'espace d'étude. On distingue deux modes de représentation pour l'information spatiale :

- Le mode *raster* ;
- Le mode vectoriel.

Le mode *raster* est également appelé mode « point », image matricielle ou bien encore *bitmap*. Une image matricielle est constituée d'un tableau rectangulaire d'éléments appelés *pixels* (abréviation de l'anglais *picture element*). Chaque *pixel* représente une couleur qui est définie par ses composantes RVB (Rouge Vert Bleu). Les défauts de ce mode de représentation sont :

- La taille importante du fichier : la taille d'une image matricielle augmente considérablement avec la résolution et sa définition ;
- La qualité de l'image : plus on *zoome* sur l'image, plus la qualité diminue.

Cependant, ce genre d'images à l'avantage de pouvoir être visionné nativement, c'est-à-dire sans avoir besoin d'installer un logiciel spécifique, par un ordinateur. Les principaux formats d'image *raster* sont : JPEG, PNG, BMP et GIF. Suivant le format du fichier, des algorithmes de compression permettent de réduire le volume du fichier. Ainsi, le format JPEG supporte des taux de compression élevés mais cela engendre une certaine dégradation de l'image, alors que le format PNG utilise un mode de compression sans perte d'informations. L'obtention de ce type d'images est facile et peu onéreuse. Les documents (photographie par exemple) ou des scènes du monde réel sont numérisés à l'aide d'appareils tels un scanner ou un appareil photo numérique.

Le mode vectoriel ne décrit pas des images comme étant un canevas de point mais donne un ensemble de coordonnées mathématiques. Ainsi, pour décrire une ligne seuls deux points sont requis : le point de départ et le point d'arrivée. En plus des informations géographiques, les fichiers vectoriels contiennent des indications sur le tracé : épaisseur, couleur de remplissage, couleur du trait. Dans le cas de la cartographie, où les objets ont des formes plus ou moins géométriques, la taille des fichiers est considérablement amoindrie. Le grossissement de l'image n'en altère pas la qualité. En effet, l'agrandissement ne fait pas apparaître un crénelage sur l'image comme c'est le cas pour

¹⁴ <http://www.postgresql.org>

¹⁵ <http://www.postgis.org>

¹⁶ <http://www.mysql.com>

¹⁷ <http://www.opengis.org>

¹⁸ <http://www.oracle.com>

les images en mode *raster*. L'inconvénient de ce format est qu'un ordinateur ne peut le lire de façon native et qu'il faut compenser par l'installation d'un logiciel ou *plugin*. Les principaux formats d'images vectorielles sont : SVG, Macromedia Flash, Adobe PDF, Adobe Illustrator ou encore les fichiers EPS (Encapsuled PostScript). Les formats prédominants étant SVG, format libre, et Flash, format propriétaire.

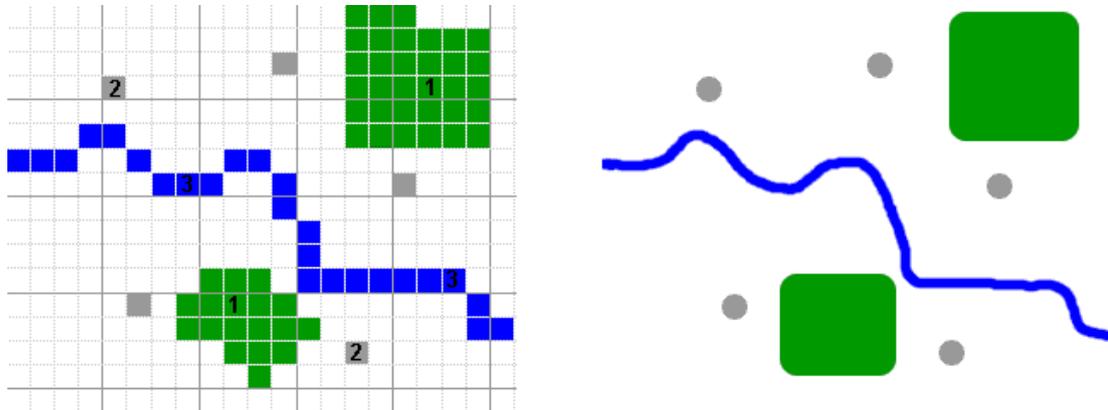


Figure 5 : Une image raster et son équivalent en image vectorielle.

Figure 5 est composée de deux figures. Celle de gauche montre un exemple d'image raster. Les données géographiques de cette image sont stockées dans chaque pixel qui peut prendre différentes valeurs. Celle de droite est l'équivalent en mode vectoriel.

Certaines applications combinent les deux modes d'affichage. C'est le cas, par exemple, de Google Maps. Cette application permet trois types d'affichage :

- Un mode « plan » pour visualiser une carte : image vectorielle ;
- Un mode « satellite » qui affiche des orthoimages sur le fond de carte : image raster ;
- Un mode « mixte » pour afficher des orthoimages en fond de carte sur lesquelles sont placés des tracés vectoriels pour la représentation des routes (cf. Figure 6).



Figure 6 : Exemple d'affichage combiné : la place Verdun de Grenoble.

2.4. L'affichage et la mise en page des données statistiques

Après avoir vu les différentes représentations possibles pour les données géographiques, nous présentons les différentes manières de présenter des données statistiques sur une carte.

Tout d'abord, les données peuvent être issues de domaines différents comme la démographie (population, taux de natalité), l'économie (nombre de chômeurs, taux de chômage) et des indicateurs environnementaux comme la superficie de terres agricoles sur une région donnée. Ce qui est le plus important dans un premier temps est de distinguer deux catégories : les variables statistiques résultant d'un comptage et les variables statistiques non additives. Ainsi, on utilisera des symboles de tailles proportionnelles (généralement des cercles) pour les données additives et des couleurs dans le cas contraire. Parmi les différents types de cartes existants, nous n'en présentons que quelques-uns. Cette sélection n'est pas arbitraire, car devant la grande quantité de cartes (et de leurs variantes) nous avons du faire un choix pour ne montrer que les plus importantes avec quelques variantes.

Les cartes en cercles proportionnels servent ainsi à représenter des quantités ou des effectifs. La taille du disque est proportionnelle à la valeur représentée et les contours de ces disques sont généralement colorés en blanc afin de mieux distinguer les petits cercles placés sur les grands. L'inconvénient de ce genre de représentation est qu'elle est davantage adaptée aux espaces où les unités sont relativement disposées de manière homogène mais également en nombre peu élevé. Philippe Waniez estime que ce genre de carte n'est pas adapté si on recense plus de quelques centaines d'unités [Waniez].

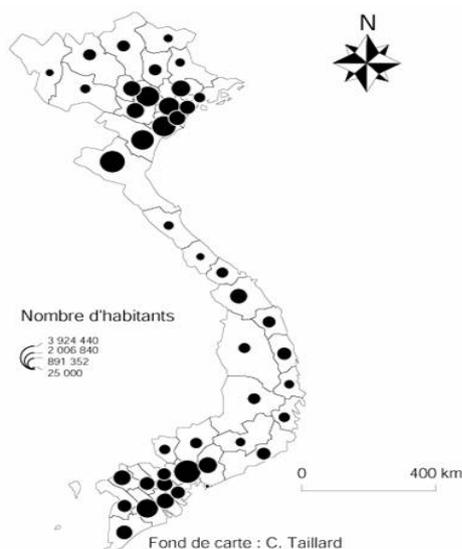


Figure 7 : La population au Viêt-Nam en 1989.

La Figure 7 montre un exemple de carte en cercles proportionnels et présente la population du Viêt-Nam en 1989. A chaque unité administrative est associée un disque. Il n'y a pas de règles précises quant à la situation de ce disque par rapport à l'unité qu'il représente. Cela peut correspondre à la localisation d'un chef-lieu ou peut-être également le centre du polygone représentant l'unité territoriale. Cette figure montre que les cercles les plus petits se situent au dessus des cercles plus grands comme précisé auparavant.

Ce genre de représentation autorise quelques variantes. Les cartes dites en cocardes (ou demi-cercles affrontés) sont également dédiées aux variables issues d'un comptage mais elles permettent une comparaison directe entre deux valeurs. Pour cela, à chaque unité est associée deux demi-cercles accolés et de couleurs distinctes. Ce type de carte permet donc de comparer deux catégories d'indicateurs

différents sur la même carte ou bien d'analyser le même phénomène à des périodes différentes.

Lorsque l'on veut représenter une variable numérique continue ou discrète sur une carte, on utilise une carte choroplèthe. Il s'agit d'une représentation de quantités (en grec : *plethos*) relatives à des espaces ou aires géographiques (en grec : *chorè*) par le moyen d'une échelle de tons gradués. Une carte choroplèthe utilise le maillage (territorial ou administratif) de la surface étudiée. Les polygones représentant les unités de ce maillage sont coloriés afin de traduire les variations du phénomène étudié. Le découpage en classes et l'affectation des unités dans ces classes se nomme la **DISCRETISATION**. Il existe plusieurs formes de discrétisations qui peuvent être classées en trois catégories :

- Les méthodes mathématiques s'appuyant sur les valeurs des données ;

- Les méthodes statistiques reposant sur les fréquences ;
- Les méthodes graphiques demandant la création de diagrammes ou de courbes auxiliaires.

Citons quelques méthodes de discrétisation : la méthode des amplitudes égales, la méthode des effectifs égaux et la méthode de la progression géométrique. Les choix de la méthode de discrétisation dépendent de plusieurs paramètres. Il faut notamment prendre en compte les propriétés de la distribution et le message que l'on souhaite transmettre au public visé. Le choix des couleurs n'est également pas anodin. Si l'on s'oriente de plus en plus vers des palettes colorées, il faut trouver la bonne gamme et le bon enchaînement des couleurs dans cette palette. Ainsi, Cindy Brewer propose plusieurs palettes et les a classées dans différentes catégories comme les palettes simples pour les séries séquentielles et les palettes doubles pour les distributions divergentes. La Figure 8 présente une copie d'écran de ColorBrewer¹⁹, le site internet de Cindy Brewer. La carte affichée est un exemple de carte choroplèthe avec une distribution divergente nécessitant la présence de deux couleurs. Les valeurs se trouvant dans la classe centrale étant colorées en blanc. Nous revenons sur cet outil dans la partie liée à HyperAtlas.

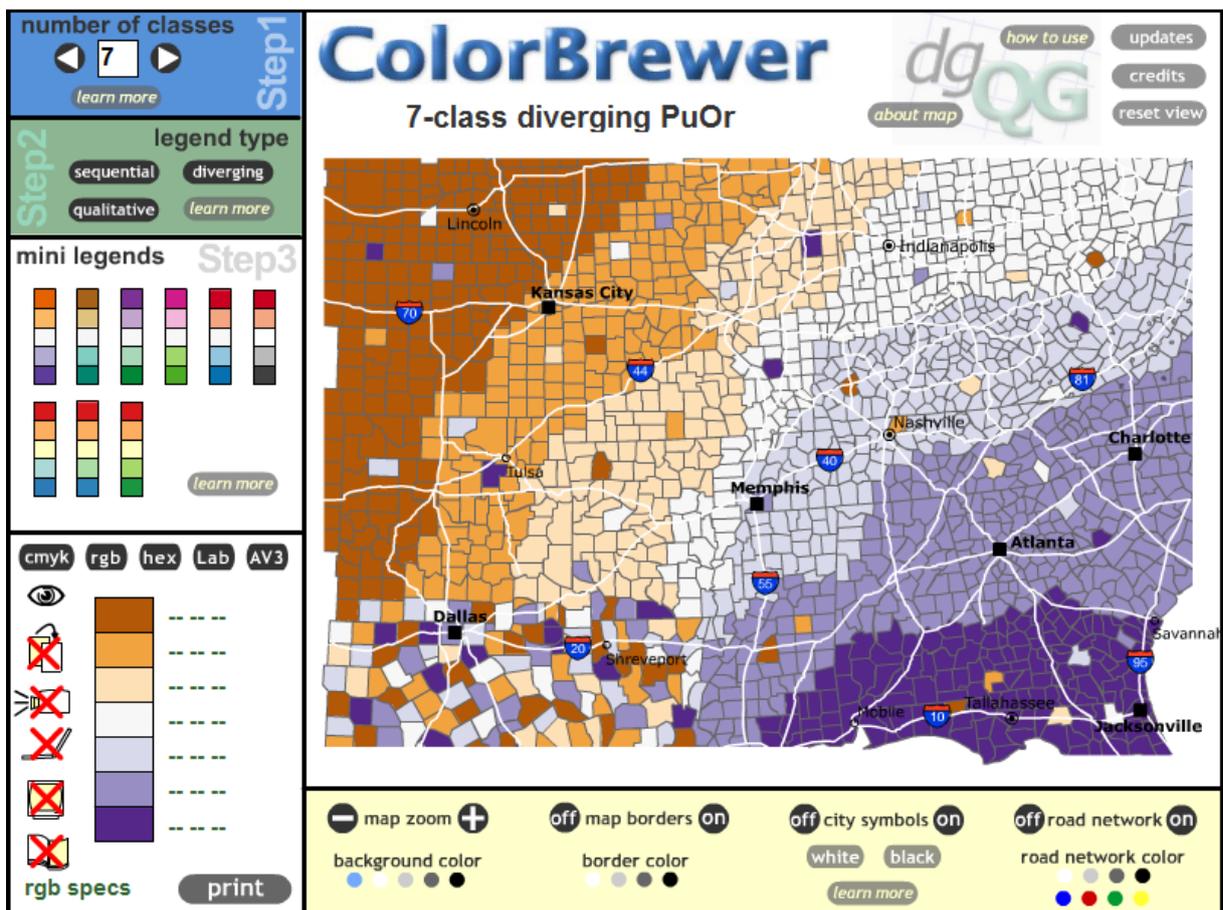


Figure 8 : Exemple d'une carte choroplèthe.

Lorsque l'on souhaite afficher sur la même carte des données non additives et des effectifs, on peut combiner les deux types de carte (choroplèthe et à disques). Nous présentons deux de ces types : les cartes en cercles proportionnels colorés et les cartes en cercle proportionnels sur surfaces colorées. Dans les deux cas, les disques sont réalisés de la même façon que sur la Figure 7, la différence provenant du fait que dans le premier cas on colorie uniquement le cercle, tandis que dans l'autre, on colorie

¹⁹ <http://www.personal.psu.edu/cab38/ColorBrewer/ColorBrewer.html>

l'ensemble du polygone représenté par le cercle. Un exemple d'application pourrait être l'étude de la population hispanique aux Etats-Unis. Les disques représenteraient le nombre d'habitants pour chaque état, et la couleur nous renseignerait sur le taux d'hispaniques vivant dans cet état.

3. Quelques outils cartographiques existants

De nombreuses applications permettent aujourd'hui de réaliser et de consulter des cartes. Parmi ces logiciels, certains permettent l'analyse spatiale et/ou territoriale de phénomènes sociaux. Nous en présentons quelques-uns mais avant cela nous analysons un des outils de cartographie les plus prometteurs et qui est très intéressant par sa simplicité d'utilisation. Il s'agit de Google Maps.

3.1. Un outil cartographique intuitif : Google Maps

Google Maps²⁰ est une application gratuite de cartographie interactive. Lancé par Google en 2004 aux Etats-Unis, il a fallu attendre deux ans pour voir ce service arriver en France.

Google Maps permet à l'utilisateur de se déplacer aisément sur les cartes et de pouvoir également zoomer sur la zone que l'on souhaite étudier. Trois types de vue sont disponibles : « plan », « satellite » et « mixte » comme sur la Figure 6 dans le paragraphe intitulé « L'affichage et la mise en page des données géographiques » (cf. page 27). Hormis, la visualisation de la planète, ce service permet également de calculer des itinéraires routiers ou encore de rechercher des commerces comme des hôtels ou des restaurants dans une ville.

Mais l'intérêt de Google Maps et ce qui le distingue par rapport aux autres solutions, réside dans la possibilité de créer ses propres cartes. Pour cela, l'utilisateur a deux moyens de procéder :

- En programmant et en utilisant l'API Google Maps ;
- Interactivement en utilisant l'interface Web.

L'intérêt de personnaliser des cartes est double : l'utilisateur peut ainsi créer une nouvelle carte pour afficher par exemple l'itinéraire de ses dernières vacances en proposant des photos aux niveaux des points touristiques ; ou bien il peut combiner les informations provenant d'une carte créée par un autre utilisateur avec ses propres données. La cartographie interactive prend tout son sens lorsque l'on essaye l'interface graphique de cet outil. Pour ajouter du contenu, l'utilisateur peut donc placer des repères pour marquer des lieux précis mais également tracer des lignes et des polygones. Il peut également utiliser des modules pour afficher du contenu plus complexe comme les conditions météorologiques ou de circulation. Ces modules sont comparables à des mini-applications et sont nommés « *mapplets* » (contraction de *map* et *widget*).

Ce contenu peut alors être publié afin de le rendre visible aux autres utilisateurs qui pourront alors réutiliser cette carte, mais aussi l'exporter pour être consultable dans Google Earth. Grâce à l'API Google Maps, l'utilisateur peut également afficher sa carte sur un site Web distinct de Google.

Nous nous intéressons essentiellement à la partie interactive plutôt qu'à l'utilisation de cette API. Cette fonctionnalité est récente puisqu'elle est activée depuis l'été 2007 et se présente sous la forme d'un onglet intitulé « mes cartes ».

²⁰ <http://maps.google.com>



Figure 9 : L'onglet "Mes Cartes" de Google Maps.

La Figure 9 présente cet onglet. Il récapitule la liste des cartes créées par l'utilisateur, affiche une liste de cartes produites par d'autres internautes et permet de créer une nouvelle carte ainsi que d'ajouter du contenu à la carte en cours.

L'ajout de contenu se fait de façon triviale. L'utilisateur marque des emplacements sur la carte à l'aide d'une collection d'icônes. Sur ces marqueurs, il peut ajouter une description dans un format texte brut ou enrichi. Dans ce cas, la description est au format HTML qui autorise une certaine mise en page (gras, souligné, liens, photos, vidéos) pour améliorer la présentation de sa carte. Il peut également tracer des lignes et dessiner des formes pour mettre en relief certaines routes et autres zones géographiques.

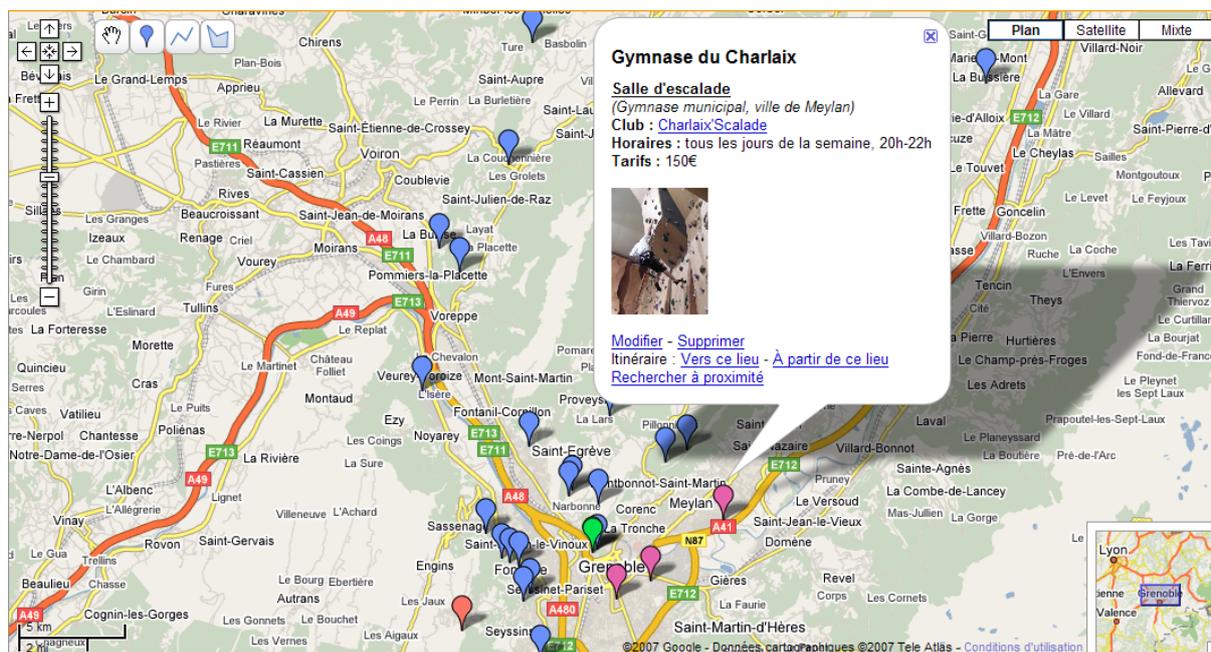


Figure 10 : Une carte personnalisée sur Google Maps.

La Figure 10 nous montre ainsi un exemple de carte personnalisée réalisée uniquement avec l'interface graphique de Google Maps. Cette carte recense les différents sites d'escalade dans la région grenobloise. Les marques de couleurs différentes permettant de différencier les différentes catégories d'escalade (écoles d'escalade, grandes voies, sites de bloc, structures artificielles d'escalade).

3.2. Les outils cartographiques dédiés à l'analyse statistique

Après avoir présenté le logiciel de cartographie de Google et constaté sa facilité d'utilisation et son interactivité, nous nous intéressons plus particulièrement aux SIG dédiés à l'analyse statistique.

3.2.1. GeoClip

Nous débutons la présentation des logiciels d'analyse spatiale territoriale par un outil payant nommé GeoClip²¹, qui est développé par la société « E=MC3 ». Ce produit existe en deux gammes : une version solo et une version serveur.

La version solo peut être considérée comme étant un générateur d'applications cartographiques autonomes. Le terme « autonome » signifie que l'application est packagée avec les données statistiques et le fond de carte. Les applications conçues, nommées GeoClips, sont constituées d'un ensemble de fichiers regroupés dans un même répertoire. Parmi ces fichiers, on trouve un fichier HTML qui lance l'application. Elles sont donc facilement publiables sur Internet mais aussi elles peuvent être distribuées sur des supports comme des CD-ROM. Le seul pré-requis étant que le lecteur Flash de MacroMedia soit installé sur le poste client.

La version serveur, quant à elle, apporte quelques suppléments par rapport à la version de base et notamment un support des bases de données pour le stockage et l'utilisation de données (géographiques et statistiques).

Différents exemples sont visualisables sur le site internet du produit. Ainsi, au niveau des fonctionnalités, on constate que cet outil permet d'afficher sur une même carte plusieurs analyses. On peut donc faire soit une analyse choroplèthe, soit une analyse par symboles mais aussi coupler ces deux types de représentation pour avoir une carte en cercles proportionnels sur surfaces colorées. Il est également intéressant de pouvoir afficher ou de masquer certaines informations complémentaires comme le réseau routier, les fleuves mais aussi d'afficher le nom des villes. Cette visualisation en couches permet une meilleure représentation de l'espace.

²¹ <http://www.geoclip.fr/fr/>

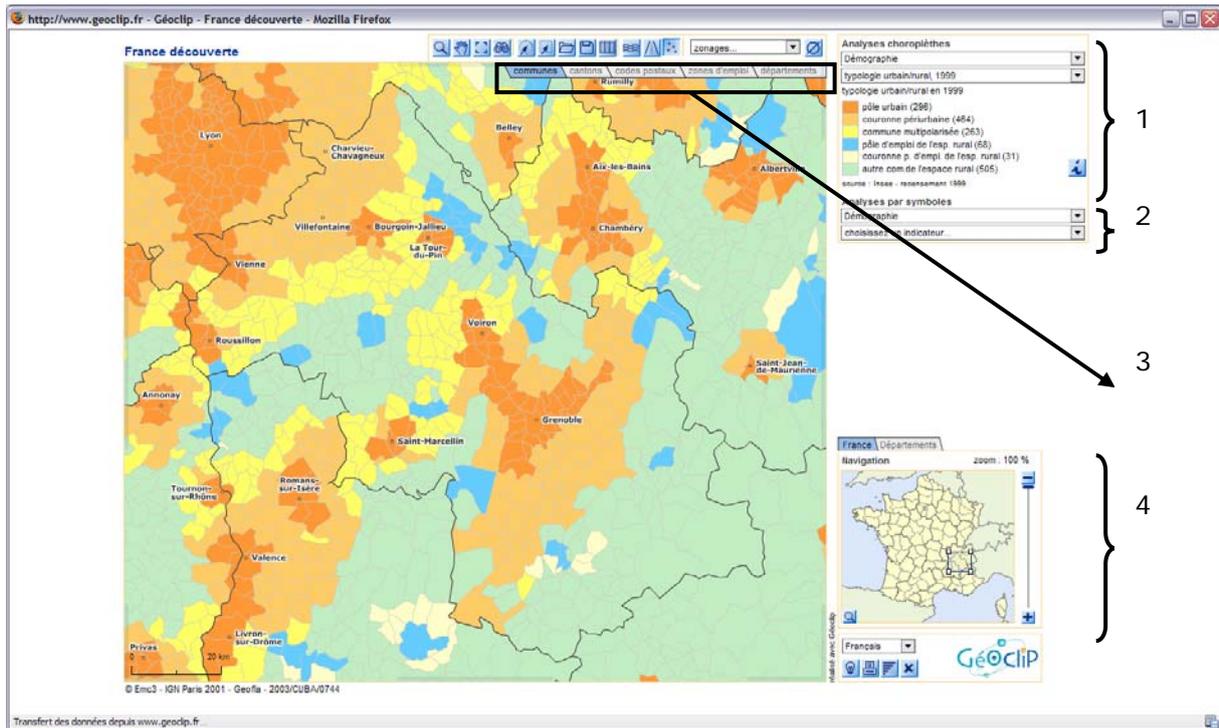


Figure 11 : Copie d'écran de GeoClip.

La Figure 11 présente une carte choroplèthe montrant la typologie urbain/rural en 1999 en France (en l'occurrence en Isère sur l'image). Sur cette copie d'écran, nous avons mis en évidence quatre zones :

- La zone n°1 est consacrée à l'analyse choroplèthe. L'utilisateur peut choisir un indicateur à afficher et visualise ensuite la légende expliquée de la carte générée. Ces indicateurs sont regroupés par thème ;
- La deuxième partie mise en évidence est consacrée à l'analyse par symboles. Ici aussi, l'utilisateur peut choisir un indicateur et visualise également une légende détaillée ;
- Les onglets (3) permettent de changer de maillage administratif. L'application nous propose cinq vues différentes :
 - Communes ;
 - Cantons ;
 - Codes postaux ;
 - Zones d'emplois ;
 - Départements.
- La mini carte (4) est un système de navigation qui permet de zoomer et de se déplacer sur la carte.

Au niveau de la convivialité et de l'ergonomie, on appréciera le classement thématique des indicateurs (démographie, environnement, élections). Néanmoins, le changement de maillage administratif réinitialise les paramètres de l'analyse mais également le niveau de zoom. De même, cet outil ne permet pas l'affichage de plus de 6000 objets sur une carte. Pour pallier ce problème, la carte n'est donc pas visualisable en totalité mais en portions. Un changement d'un paramètre d'analyse, un déplacement sur la carte mais également une modification du niveau de zoom, implique un recalcul ce qui se traduit par un temps de chargement conséquent.

3.2.2. Cartes & Données

Il s'agit d'un produit développé et commercialisé par la société Arctique experte en systèmes d'analyse géographique et en géomarketing. Cette société produit et vend des cartes statistiques mais elle a également commercialisé un puissant outil pour la génération de cartes. Il s'agit du logiciel Cartes & Données²², anciennement nommé le Cartographeur.

Malgré la puissance et les fonctionnalités offertes par ce logiciel, l'utilisateur novice rencontrera certainement quelques difficultés à visualiser sa première carte. En effet, il doit choisir un fond de carte, importer des données statistiques, choisir son mode de discrétisation, choisir un mode d'affichage pour les contours et pour le remplissage, et enfin, choisir sa représentation. Chacune de ces opérations est réalisée à l'aide d'un module spécifique. L'interface graphique de l'outil présente notamment une hiérarchie pour la visualisation de ces modules. Si la convivialité n'est pas optimale pour un utilisateur débutant, il s'avère que ce logiciel laisse une grande autonomie à l'utilisateur plus confirmé. Des modules complémentaires payants peuvent également être ajoutés dans l'application.

Un autre reproche à l'adresse de cet outil est qu'il est davantage destiné à la visualisation statique d'un phénomène sur une carte plutôt qu'à l'analyse de phénomènes statistiques. En effet, si l'utilisateur désire changer l'indicateur, il doit régénérer l'intégralité de la carte. Cela se traduit également par l'impossibilité de se déplacer et de zoomer. La carte générée est plus à considérer comme une image statique d'un phénomène. De ce fait, il est possible d'exporter sa carte dans un grand nombre de formats comme : PNG, GIF, JPEG, PDF, Flash, HTML.

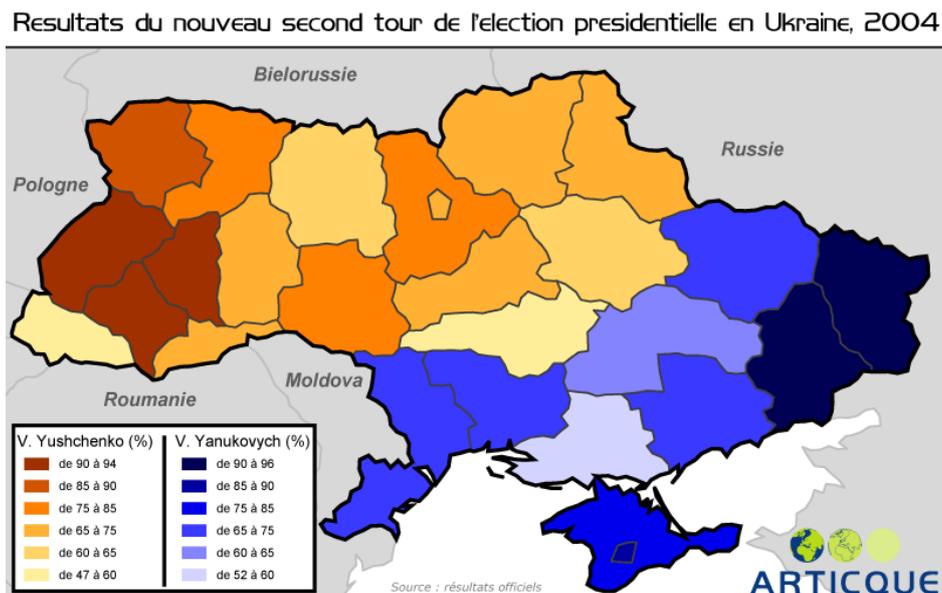


Figure 12 : Une carte générée par le Cartographeur.

La Figure 12 présente un exemple de carte générée par cet outil. Il s'agit d'une carte choroplèthe utilisant une palette de couleurs divergente afin de représenter les résultats de l'élection présidentielle en Ukraine en 2004.

²² <http://www.cartesetdonnees.com/>

Le SIG d'Arctique est une solution complète et efficace pour la génération de cartes. En effet, cet outil s'inscrit dans une logique de production de cartes que l'on peut ensuite présenter dans un document pour présenter un phénomène donné.

3.2.3. PhilCarto

Nous décrivons ici un autre SIG : PhilCarto²³. Contrairement aux précédents outils que nous venons de présenter, PhilCarto n'est pas une solution commerciale mais est un logiciel issu de la recherche universitaire, conçu et développé par un géographe Philippe Waniez, directeur de recherche à l'Institut de Recherche pour le Développement en poste à l'Université Catholique de Rio de Janeiro (Brésil). Ce logiciel est ainsi disponible gratuitement à l'adresse suivante <http://philgeo.club.fr>.

Dans la dernière version, PhilCarto V5.0, il est désormais possible de procéder à des analyses spatiales multiscalaires. Ce logiciel se distingue donc des précédents, car il semble offrir des fonctionnalités similaires à celles d'HyperAtlas.

Au démarrage de l'application, l'utilisateur doit renseigner les différentes données à utiliser. Les données géographiques doivent être au format « ai ». Il s'agit d'un format propriétaire d'images vectorielles propre à Adobe Illustrator. Les données attributaires peuvent être stockées dans un fichier texte ou dans un classeur Excel. Une fois ces fichiers renseignés, l'utilisateur peut alors commencer son analyse.

On constate rapidement qu'il s'agit d'un outil très riche. En effet, on dénombre une grande variété de cartes disponibles : cartes choroplèthes, cartes à disques mais également carte en oursins, carroyage, vecteurs de flux, isolignes, surface de tendances, etc. Il est également possible de représenter trois types de déviations :

- Macro déviation, que l'on nomme déviation globale dans le projet HyperCarte ;
- Meso déviation, que l'on nomme déviation moyenne dans le projet HyperCarte ;
- Micro déviation, qui est l'équivalent de la déviation locale dans le projet HyperCarte.

Pour analyser un phénomène statistique, l'utilisateur doit sélectionner deux indicateurs appelés « numérateur » et « dénominateur ». La Figure 13 montre cette fenêtre de sélection.

²³ <http://philgeo.club.fr>

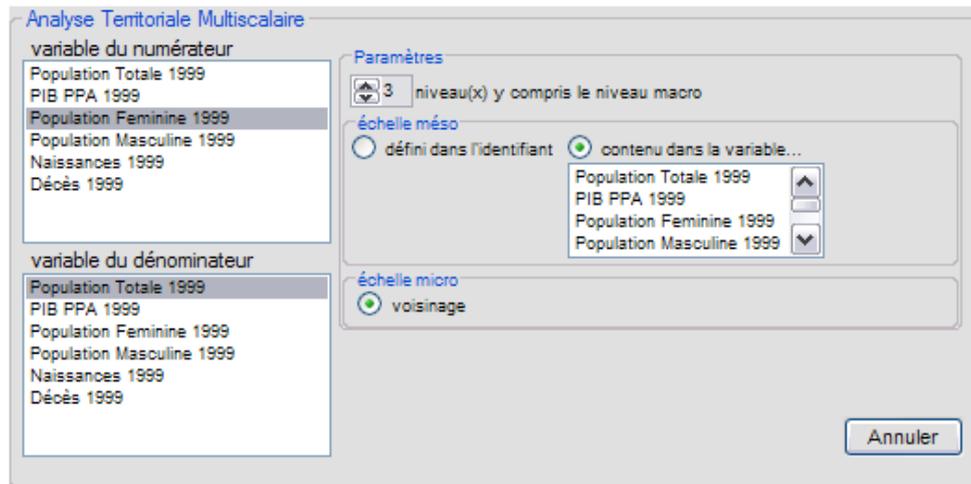


Figure 13 : Choix des paramètres de l'analyse.

Pour le calcul de ces déviations, l'utilisateur a le choix entre six méthodes de discrétisation :

- Q6 : il s'agit d'une discrétisation selon les quartiles avec isolement des queues de la distribution par définition de deux classes extrêmes. Les six classes sont définies avec les bornes suivantes: minimum, percentile 5, 1^{er} quartile, médiane, 3^{ème} quartile, percentile 95, maximum ;
- S5 : Discrétisation sur une variable standardisée en cinq classes. Les bornes des classes sont calculées de la manière suivante :
 - Minimum ;
 - Moyenne \pm 0,5 écart-type ;
 - Moyenne \pm 1 écart-type ;
 - Maximum.
- AM : Discrétisation en N classes d'égales amplitudes. Le nombre de classes, N, est fixé à six par défaut, mais peut varier grâce à un curseur ;
- EF : Discrétisation en N classes d'effectifs égaux (méthode des quantiles). Le nombre de classes, N est fixé à cinq par défaut, mais peut également être modifié ;
- Jenks : Discrétisation proposée par Jenks en 1977 qui repose sur la notion de variance, c'est-à-dire sur la dispersion des valeurs de la donnée autour de la moyenne. Elle vise à minimiser la variance intra-classes (donc à maximiser l'homogénéité des classes), et à maximiser la variance inter-classes (donc à accroître les différences entre classes). Le nombre de classes est fixé à cinq, mais l'utilisateur peut également sélectionner une autre valeur ;
- 1, 2, 3 : Discrétisation selon des bornes définies par l'utilisateur.

Si PhilCarto est très complet, il est également complexe d'utilisation pour un utilisateur néophyte. En effet, devant un tel choix de méthodes de discrétisation, comment choisir celle qui correspond le mieux à l'analyse voulue ? Comme le montre la Figure 14, aucune explication n'est associée à une méthode de discrétisation. De plus, en laissant la possibilité à l'utilisateur de choisir librement le numérateur et le dénominateur, il n'est pas exclu de faire des non-sens ou bien des erreurs d'analyse.

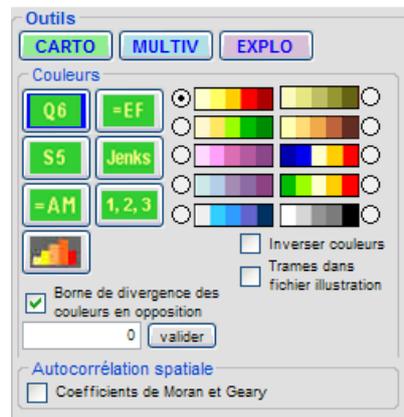


Figure 14 : Choix de la méthode de discrétisation.

La carte résultante n'est pas très interactive. Il est en effet impossible de se déplacer ou de zoomer. Il n'est également pas possible de changer le maillage administratif afin de pouvoir comparer des indicateurs à différentes échelles. Cependant, il est possible de changer les palettes de couleurs. La légende, située en haut à droite, est composée de plusieurs caissons, chaque caisson représentant une classe. Cette légende est complète. En effet, pour chaque classe sont renseignés les valeurs seuils, le nombre d'unités comprises dans cette classe, la moyenne et son écart-type. A cette légende, s'ajoute un histogramme qui permet de présenter les classes vis-à-vis de la distribution statistique. La Figure 15 est un exemple de carte visualisable avec PhilCarto.

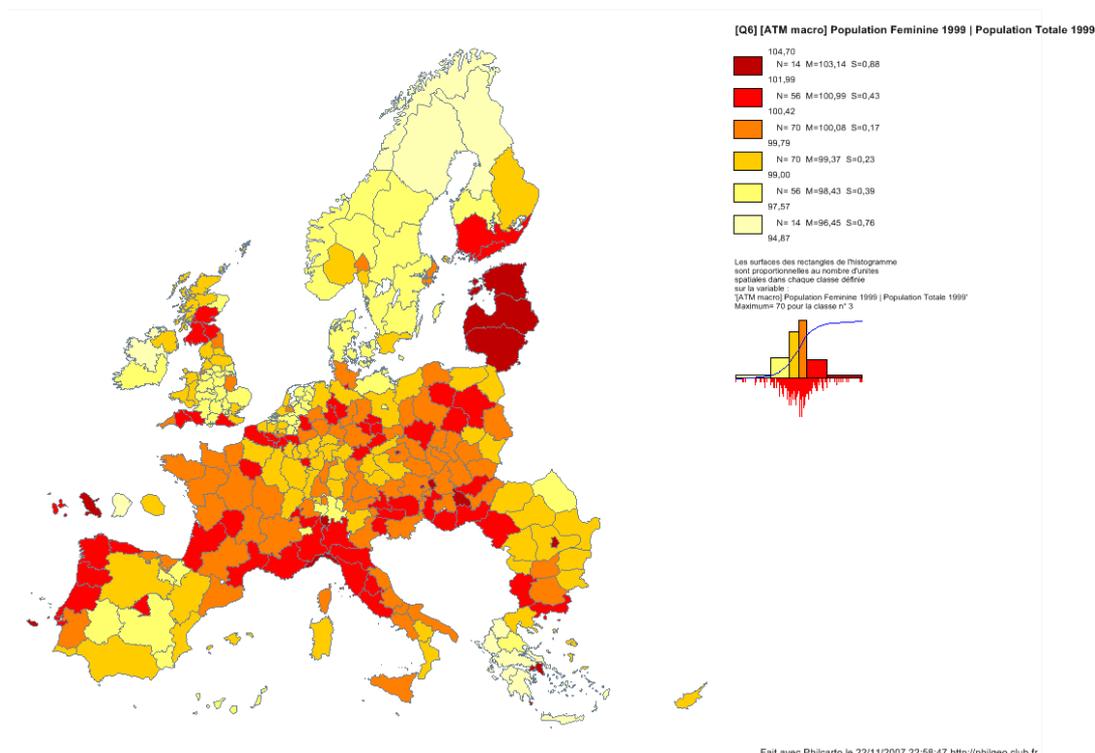


Figure 15 : Affichage d'une carte et de sa légende sous PhilCarto.

On regrettera en revanche, le manque de lisibilité pour l'affichage des zones survolées sur la carte. En effet, lorsque l'utilisateur sélectionne une partie de la carte, les informations liées à l'unité géographique sont affichées dans la barre de titre de

l'application. La Figure 16 présente la barre de titre de PhilCarto avec l'affichage des données statistiques de la zone survolée (ici, la région Rhône-Alpes).



Figure 16 : L'affichage des zones survolées dans la barre de titre.

Il est également possible de visualiser une carte de synthèse qui est paramétrable car l'utilisateur peut choisir et sélectionner les différents résultats de son analyse.

Pour conclure, on retiendra que PhilCarto est un logiciel très complet qui offre à l'utilisateur initié un ensemble d'outils pour réaliser son analyse. Cet outil est plutôt un logiciel didactique utilisé par des étudiants en géographie. On trouve d'ailleurs sur Internet de nombreux documents de cours et TD sur PhilCarto.

3.3. Synthèse des différents outils présentés

Le tableau ci-dessous récapitule les différentes fonctionnalités de chaque outils que nous venons de présenter.

Tableau 1 : Synthèse des fonctionnalités.

Fonctionnalités	GéoClip	Cartes & Données	PhilCarto
Affichage d'une carte de contexte	Oui	Oui	Oui
Affichage d'une carte présentant le numérateur	Oui	Oui	Oui
Affichage d'une carte présentant le dénominateur	Oui	-	Oui
Affichage d'une carte présentant des ratios	Oui	Oui	Oui
Affichage d'une carte des déviations	-	-	Oui
Zoom	Oui	-	Oui
Déplacement	Oui	-	Oui
Choix de la méthode de discrétisation	-	-	Oui
Export possible	-	Oui	Oui
Licence	Commercial	Commercial	Gratuit

4. *Synthèse*

Dans cet état de l'art, nous avons tout d'abord défini la notion de Système d'Information Géographique. Nous nous sommes ensuite essentiellement intéressés aux données utilisées dans ce type de logiciel. Nous avons ainsi expliqué le rôle de ces données via une sorte de cycle de vie : de l'acquisition de ces informations à leur visualisation.

Nous avons ensuite présenté quatre logiciels de cartographie qui n'ont pas tous la même vocation. En effet, certains proposent davantage de cartes imprimables tandis que d'autres mettent l'accent sur la navigation. Cependant, cette comparaison nous permet de dresser un parallèle entre ergonomie, confort d'utilisation et complexité d'utilisation. Il semble en effet que plus le logiciel s'adresse à un public d'experts moins il est ergonomique. D'un autre côté, plus le logiciel s'adresse à un large public, plus il est ergonomique et simple d'utilisation mais il offre également moins de possibilités. Le Tableau 1 montre bien que PhilCarto est l'outil offrant le plus de fonctionnalités, cependant comme nous l'avons vu, il n'est pas aussi simple à utiliser que les autres. L'ergonomie, ou utilisabilité, désigne la capacité d'un programme à être facilement utilisé par une personne donnée pour réaliser la tâche pour laquelle il a été conçu. On s'efforcera donc à respecter, voire dans certains cas, à renforcer cette qualité.

Dans le prochain chapitre, nous présentons les différents modules du projet HyperCarte notamment HyperAtlas et HyperAdmin.

CHAPITRE 2

Analyse de l'existant

Comme nous l'avons vu, le groupe de recherche HyperCarte a été fondé en 1996. Ce projet existe donc depuis plus de dix ans et nous allons, dans ce chapitre, présenter les différentes applications du projet HyperCarte. Nous commençons par présenter les principes généraux sur lesquels sont fondées ces applications.

Sommaire du chapitre :

Chapitre 2 - Analyse de l'existant	41
1. Un projet européen pour la cartographie interactive	43
1.1. Le programme ESPON	43
1.2. Définition des termes du domaine métier	44
1.2.1. Unité territoriale.....	44
1.2.2. Espace.....	44
1.2.3. Maillage	45
1.2.4. Les indicateurs.....	46
2. Un module de visualisation : HyperAtlas	47
2.1. L'interface utilisateur	47
2.1.1. La zone de paramétrage	47
2.1.1.1. Les paramètres « <i>Area and Zoning</i> »	48
2.1.1.2. Les paramètres « <i>Indicators</i> »	48
2.1.1.3. Les paramètres liés aux déviations	49
2.1.2. Les différents types de carte	49
2.1.2.1. La carte « <i>Area and Zoning</i> ».....	50
2.1.2.2. Les cartes d'indicateurs.....	50
2.1.2.3. La carte de ratio	51
2.1.2.4. Les cartes de déviation	52
2.1.2.5. La carte de synthèse	53
2.1.3. Sauvegarde et export de l'analyse	54
3. Un module d'administration : HyperAdmin	55
3.1. Problématique	55
3.2. L'interface utilisateur	55
3.3. Les fichiers de données.....	57

3.3.1. Les données géographiques	57
3.3.2. Les données structurelles	58
3.3.3. Les données statistiques.....	61
3.4. Fonctionnalités	62
3.4.1. Création d'un nouveau jeu de données	62
3.4.2. Calcul de la contiguïté	63
3.4.3. Création de relations de voisinage.....	63
3.4.4. Sauvegarde et export.....	64
4. Un module d'analyse spatiale : HyperSmooth	65
4.1. Principes.....	65
4.2. HyperSmooth	66
5. Architecture et choix technologiques	68
5.1. Le modèle MVC.....	68
5.1.1. Le modèle	68
5.1.2. La vue	69
5.1.2.1. <i>Look and Feel</i>	69
5.1.2.2. SwingX.....	70
5.1.3. Le contrôleur	70
5.2. Choix technologiques.....	70
6. Conclusion	72

1. Un projet européen pour la cartographie interactive

1.1. Le programme ESPON

HyperAtlas est un outil d'analyse multiscalaire territorial développé dans le cadre du projet de recherche **ESPON** 3.1. ESPON est l'acronyme de l'anglais « *European Spatial Planning Observation Network* » ; en français, on utilise davantage le terme **ORATE** pour désigner cet « Observatoire en Réseau de l'Aménagement du Territoire Européen ». Cet organisme a été créé en 1997 par les différents ministres européens en charge de l'aménagement du territoire. La phase 1 de ce projet, qui a débuté en 2002 et qui a pris fin en 2006, a financé 34 projets de recherche. Parmi ces projets, on distingue 4 thèmes :

- Projets de priorité 1 : études thématiques analysant les effets spatiaux de différents phénomènes ayant des répercussions sur l'espace (villes, démographie, transports, énergie) ;
- Projets de priorité 2 : études ayant pour objet les répercussions de différentes politiques communautaires ou politiques spatiales des Etats-Membres sur les différents types de régions ;
- Projets de priorité 3 : études transversales, capitalisant les études des priorités 1 et 2 pour établir, à partir d'une analyse des résultats des autres projets, des résultats intégrés en matière par exemple d'indicateurs spatiaux, des typologies spatiales, des scénarii d'évolution ou des banques de données. Parmi les différents projets de cette catégorie, on trouve HyperAtlas qui appartient à un ensemble de projets dénommé : *integrated tools for the European spatial development* ;
- Projets de priorité 4 : activités de soutien scientifique et de mise en réseau pour établir des synergies entre les ressources scientifiques nationales et celles de l'Union Européenne.

Le projet ESPON 3.1 consiste à réaliser des outils intégrés pour l'aménagement du territoire européen. L'objectif d'HyperAtlas est donc d'intégrer des données statistiques sur l'ensemble de l'Union Européenne et de les mettre en valeur grâce à différentes cartes (affichage des indicateurs, ratios, écarts par rapports à différentes moyennes).

La seconde phase du programme ORATE a débuté en 2007 et prendra fin en 2013. Cinq priorités ont été définies :

- La recherche appliquée sur le développement territorial, la compétitivité et la cohésion ;
- Des recherches ciblées répondant aux besoins des praticiens ;
- La construction d'une plate-forme scientifique pour la production de données comparables et d'outils d'analyses ;
- Capitalisation, approbation, diffusion et mise en réseau des travaux d'ORATE ;
- Les actions de communication et d'assistance technique.

Parmi les principaux partenaires de ce programme, on compte les 27 états membres de l'Union Européenne ainsi que des pays invités [*@Riate*], [*@Espon*], [*@Espon Luxembourg*].

Nous nous consacrons désormais à l'étude des différents modules de ce projet. Cependant, avant d'en découvrir les fonctionnalités, nous commençons par définir les notions utilisées dans ce domaine métier.

1.2. Définition des termes du domaine métier

Les entités traitées sont des unités territoriales, qui s'organisent en niveaux. Chaque niveau forme une partition de l'espace. Ce découpage est également appelé maillage administratif.

1.2.1. Unité territoriale

Une unité territoriale, notée UT, est une zone géographique quel que soit son niveau hiérarchique. Ainsi, un département, une région et un pays sont des unités territoriales. Certaines unités peuvent être agrégées et forment alors une unité territoriale supérieure (UTS). C'est le cas par exemple des régions qui associées, forment un pays. Les unités qui ne sont pas issues d'une combinaison sont appelées unités territoriales élémentaires (UTE). Elles composent généralement le niveau le plus bas. Ainsi, l'Isère, la région Rhône-Alpes et la France sont des exemples d'unités territoriales. L'Isère étant une UTE tandis que la France est une UTS, car elle est obtenue par la réunion des régions françaises.

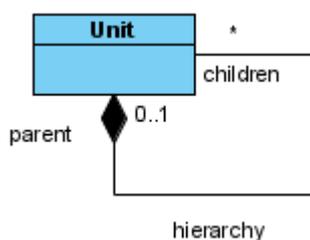


Figure 17 : Diagramme de classes : hiérarchie des unités.

1.2.2. Espace

L'espace peut être considéré comme étant une zone d'étude. Cette zone d'étude répond à des enjeux économiques ou politiques sur lesquels porte l'analyse. Un espace est composé d'un ensemble d'unités territoriales.

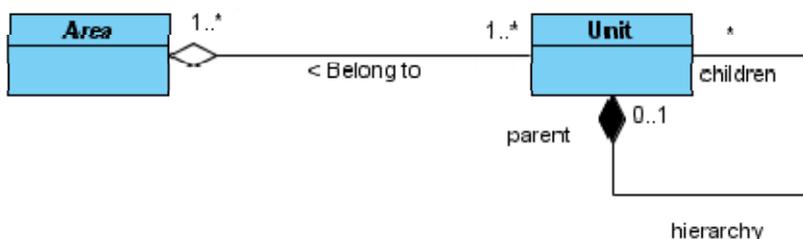


Figure 18 : Diagramme de classes : relation entre l'aire d'étude et les unités.

1.2.3. Maillage

Le maillage est un découpage de l'espace. Chaque découpage correspond à un niveau. Il est constitué par un ensemble d'unités territoriales. Par exemple, on peut citer le partitionnement en quartiers ou en régions. On notera que plus la maille administrative est fine, plus le niveau est élevé.

Deux découpages sont dits compatibles si l'on peut passer de l'un à l'autre par agrégation des unités du niveau le plus fin pour obtenir le niveau le plus élevé. Les différents découpages s'organisent alors selon une structure hiérarchique emboîtée. On peut donc passer d'un niveau à l'autre par simple agrégation (réunion) des unités territoriales.

Au niveau européen, la Commission Européenne a subdivisé les pays en régions dénommées NUTS (Nomenclature des Unités Territoriales Statistiques). Il existe 6 niveaux :

- NUTS 0 : ce niveau correspond aux pays européens (ex : France, Allemagne) ;
- NUTS 1 : ce niveau correspond aux grandes régions (ex : ZEAT²⁴ pour la France, *Länders* pour l'Allemagne)
- NUTS 2 : ce niveau regroupe les régions européennes (ex : les régions françaises, les districts italiens et les communautés autonomes espagnoles)
- NUTS 3 : ce niveau comporte l'équivalent des départements français (ex : les départements français, les arrondissements en Allemagne, les provinces en Espagne ou en Italie)
- NUTS 4 : ce niveau est désormais dénommé LAU 1. Les LAU (acronyme de « *Local Administrative Unit* ») sont les composants primaires des régions NUTS. Le niveau LAU 1 n'est pas défini pour l'ensemble des pays européens. Les districts anglais sont un exemple de LAU 1 ;
- NUTS 5 : ce niveau est maintenant noté LAU 2. Ce rang correspond aux communes européennes.

Dans HyperAtlas, un niveau intermédiaire a été ajouté afin d'obtenir des ensembles d'unités d'une taille homogène. Ce niveau supplémentaire est noté NUTS 2-3.

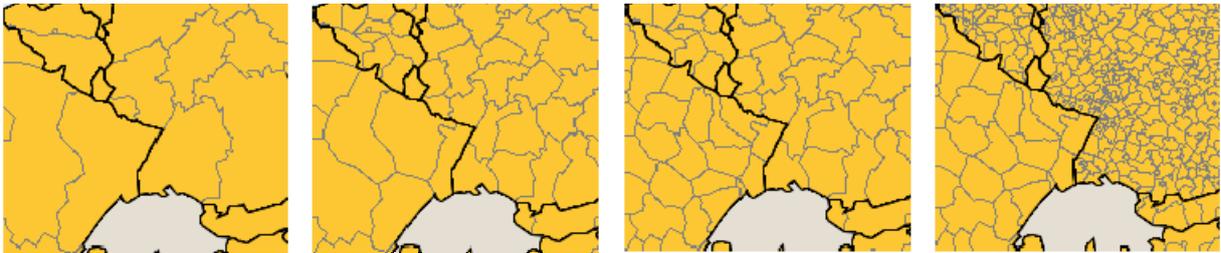


Figure 19 : Les différents maillages NUTS.

Les différentes captures d'écran de la Figure 19 présentent les différences entre les niveaux NUTS. De gauche à droite, on visualise le niveau NUTS1, le niveau NUTS2, le niveau intermédiaire NUTS2-3 et enfin le niveau NUTS3. On remarque donc l'intérêt de ce maillage « NUTS2-3 » qui permet de faire des observations et des analyses sur des unités ayant une taille équivalente.

Le niveau 0 de la hiérarchie (NUTS0 dans le cas de l'Europe) est un niveau privilégié car il s'agit du niveau de collecte du point de vue statistique.

²⁴ ZEAT : Zone d'Etudes et d'Aménagement du Territoire

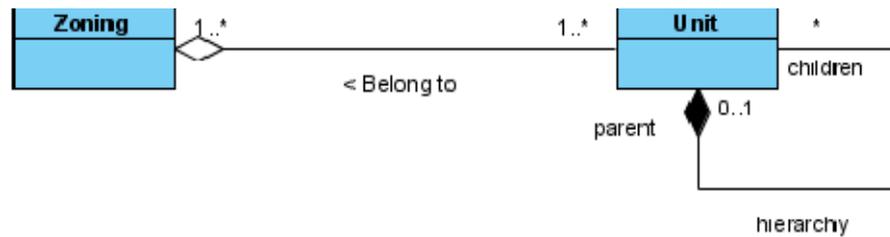


Figure 20 : Diagramme de classes : relation entre les unités territoriales et le maillage administratif.

1.2.4. Les indicateurs

Les indicateurs sont des données statistiques. Ces données peuvent être de plusieurs sortes : démographiques, économiques, environnementales. Une valeur d'indicateur est associée à chaque unité territoriale élémentaire. HyperAtlas associe ensuite par une opération d'agrégation une valeur d'indicateur à l'ensemble des unités territoriales supérieures.

Dans HyperAtlas, une analyse fine nécessite la présence de deux indicateurs, que l'on nommera numérateur pour le premier et dénominateur pour le second, à partir desquels il est possible de calculer :

- Un ratio comme étant le rapport du numérateur et du dénominateur ;
- Un ensemble de moyennes (ou déviations) comme étant l'expression du ratio par rapport à une moyenne de référence :
 - La déviation globale se base par rapport à la moyenne de l'aire d'étude ;
 - La déviation moyenne se base par rapport à la moyenne d'une unité englobante (le pays, par exemple) ;
 - La déviation locale se calcule par rapport à la moyenne des unités voisines.

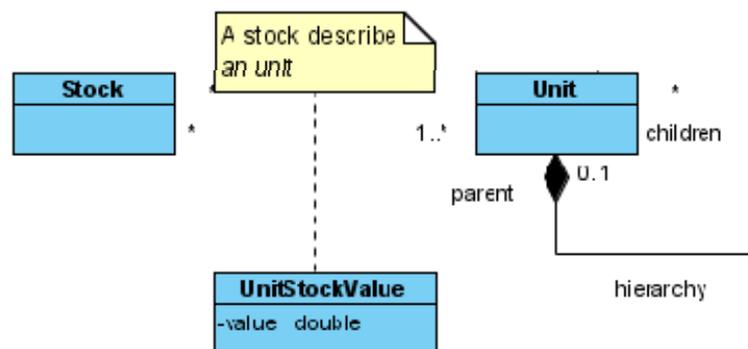


Figure 21 : Diagramme de classes : relation entre les unités et les indicateurs.

2. Un module de visualisation : HyperAtlas

2.1. L'interface utilisateur

Nous commençons cette partie en présentant l'interface d'HyperAtlas.

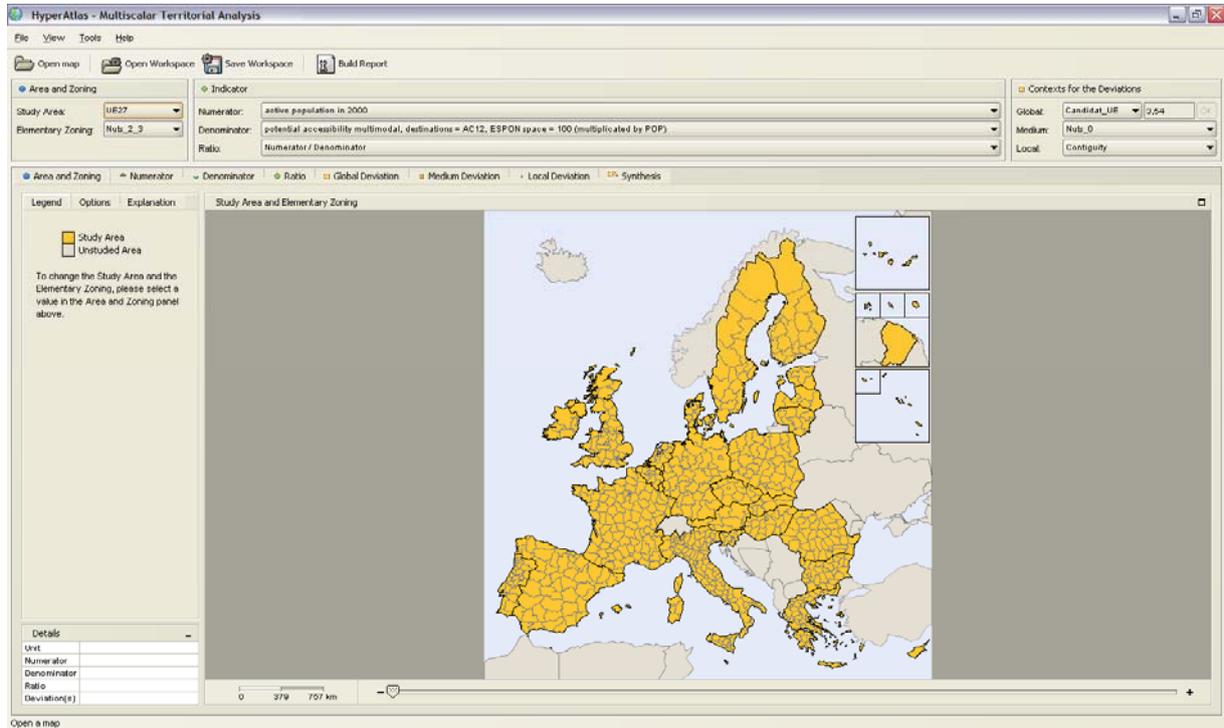


Figure 22 : L'interface utilisateur d'HyperAtlas.

Ainsi, la Figure 22 présente l'interface de l'application. Elle est composée de plusieurs parties :

- La barre de menus : liste des actions possibles sur l'application ;
- La barre d'outils : raccourci vers les actions les plus courantes ;
- La zone de paramétrage : zone regroupant les différents réglages que l'on peut effectuer sur la carte. Cette zone est composée de 3 éléments :
 - Le choix de l'aire d'étude et du maillage ;
 - Le choix des indicateurs ;
 - Les options sur les calculs des déviations.
- Les onglets : navigation et changement de carte.

Les barres d'outils et de menus sont des éléments classiques d'une Interface Homme-Machine (IHM), nous ne les détaillerons donc pas. Par contre, les deux derniers éléments sont spécifiques à HyperAtlas et méritent donc notre attention.

2.1.1. La zone de paramétrage

Cette partie permet à l'utilisateur de choisir les paramètres de son analyse spatiale multiscalaire. Comme nous l'avons vu précédemment, ces paramètres ont été classés en trois familles que nous précisons.

2.1.1.1. Les paramètres « *Area and Zoning* »

Cette zone permet de choisir l'espace d'étude (*Area*) ainsi que le niveau de maillage (*Zoning*).

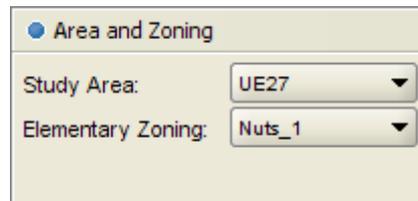


Figure 23 : Sélection des paramètres "Area and Zoning".

La Figure 23 montre un exemple de paramétrage possible. La modification de ces paramètres induit un réaffichage des différentes cartes.

L'aire d'étude (*Study Area*) permet de sélectionner les unités sur lesquelles va porter l'analyse. Dans cet exemple, le jeu de données est basé sur l'Europe et l'utilisateur peut choisir comme espace de référence l'UE27 (L'Union Européenne composée de ses 27 membres) mais il pourrait également choisir des aires d'étude comme l'Arc Atlantique, qui rassemble une trentaine de régions s'étendant de l'Andalousie à l'Ecosse, ou encore le regroupement PECO (Pays d'Europe Centrale et Orientale), qui contient l'ensemble des anciens pays communistes du centre et de l'est de l'Europe.



Figure 24 : Influence du choix de l'espace d'étude.

Ainsi, la figure ci-dessus présente différents espaces d'étude dans HyperAtlas. La zone de référence étant alors mise en évidence par rapport au reste de la carte grâce à une couleur de remplissage spécifique.

Le choix du découpage administratif (*Elementary zoning*) permet de définir le niveau de maillage sur lequel est basée l'analyse. Dans le cas de notre exemple sur le jeu de données européen, l'utilisateur pourra choisir un maillage allant du niveau NUTS0 au niveau NUTS3. La Figure 19 permet de visualiser l'effet de différents maillages sur une zone couvrant l'est de la France et le sud-ouest de l'Allemagne.

2.1.1.2. Les paramètres « *Indicators* »

Il est possible dans cette zone de choisir les indicateurs que l'on souhaite analyser. L'utilisateur choisit un couple d'indicateurs que l'on nomme numérateur et dénominateur.

Figure 25 : Le choix des indicateurs.

L'utilisateur choisit donc librement un numérateur parmi la liste des indicateurs, que l'on nomme également *stocks*, puis un dénominateur dans cette même liste. Il peut donc choisir en toute liberté les phénomènes de son choix pour son analyse.

2.1.1.3. Les paramètres liés aux déviations

Il s'agit du troisième et dernier panneau de configuration de cette application. Il permet à l'utilisateur de choisir aisément le contexte de calcul des déviations.

Figure 26 : le paramétrage des déviations.

Etant donné qu'il existe trois types différents de déviation dans HyperAtlas, la zone de paramétrage est composée de trois lignes qui permettent d'agir sur ces moyennes. La déviation globale permet de comparer la valeur du ratio d'une unité par rapport à la moyenne de l'espace d'étude. L'utilisateur peut choisir un des espaces d'étude proposés par l'application auquel cas la moyenne du ratio est affichée à côté de cette liste déroulante. Mais il peut également saisir une valeur qui est le seuil utilisé pour calculer l'ensemble des déviations globales.

De la même façon, l'utilisateur peut choisir le maillage de référence pour le calcul de la déviation moyenne et aussi estimer la déviation locale. Concernant ce dernier type de déviation, l'utilisateur peut choisir l'étendue du voisinage d'une unité. Il peut en exister plusieurs types :

- La contiguïté simple qui désigne comme unité voisine toute unité ayant une frontière commune avec l'unité courante ;
- Les matrices de contiguïté qui permettent d'étendre la notion de voisinage. Deux unités peuvent être voisines à différents niveaux en utilisant différents types de distance comme notamment les distances « temps voiture » et « temps camion ».

2.1.2. Les différents types de carte

Après avoir présenté l'ensemble des différents paramétrages possibles pour choisir le phénomène à analyser sur HyperAtlas, nous nous intéressons à sa visualisation.

Il existe huit types de cartes dans cet outil d'analyse territoriale multiscalaire. L'utilisateur peut naviguer de l'une à l'autre grâce à un système d'onglets. Un onglet permet d'accéder rapidement et directement à l'information souhaitée mais il permet aussi d'enrichir l'interface graphique. Notons également que le changement d'onglet

engendre un réaffichage de la carte mais également de l'ensemble du contexte lié à la carte. Nous commençons par présenter ces éléments.

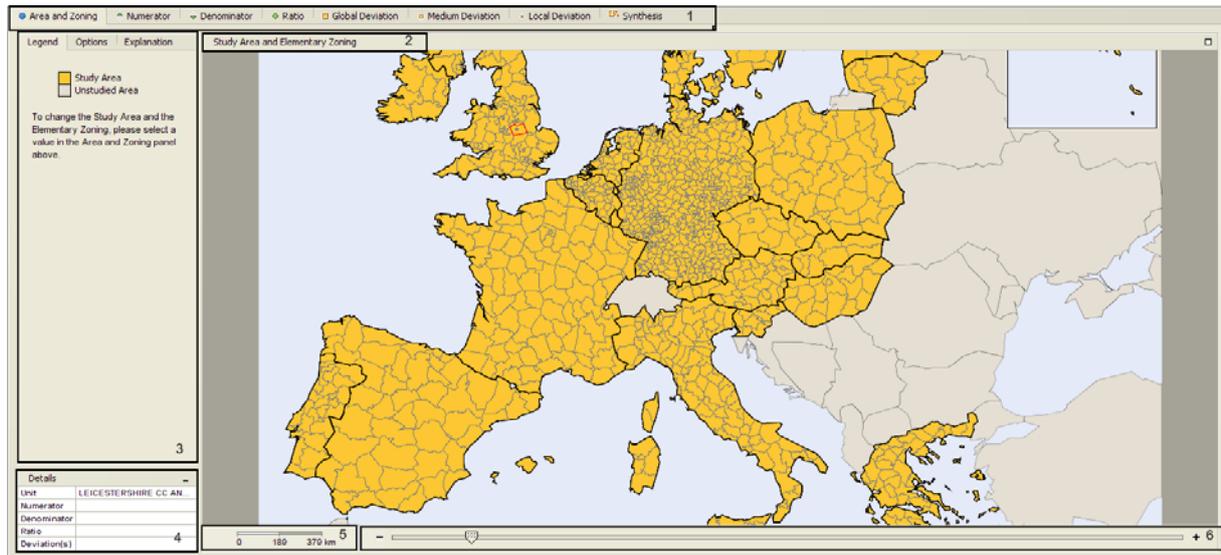


Figure 27 : Les différents composants d'une carte.

Sur la Figure 27, mise à part la carte, on dénombre six éléments annexes qui ajoutent du sens et de l'information à cette carte.

- L'élément n°1 est le système de navigation par onglets ;
- La zone n°2 correspond au titre de la carte visualisée ;
- La zone n°3 est un panneau d'information regroupant :
 - La légende de la carte ;
 - Les options de visualisation (choix des couleurs, taille des disques) ;
 - Un texte d'explication.
- Le composant n°4 est un tableau affichant des informations détaillées sur l'unité survolée par le curseur de la souris ;
- L'élément n°5 permet l'affichage de l'échelle ;
- Le curseur correspondant à la zone n°6 permet d'afficher et de régler le niveau de zoom.

2.1.2.1. La carte « Area and Zoning »

La carte « Area and Zoning » est la carte de base de l'application. Elle permet de visualiser l'espace d'étude et le maillage sélectionné. Cette carte est vierge de toute représentation statistique. La Figure 27 en est un exemple.

Les couleurs associées à l'espace d'étude et à l'espace hors du champ d'étude sont paramétrables dans l'onglet « Options ».

2.1.2.2. Les cartes d'indicateurs

Il existe deux cartes d'indicateurs dans HyperAtlas. La première permet la visualisation du numérateur tandis que la seconde est dédiée à l'affichage du dénominateur. Ainsi, à chaque unité est associé un disque proportionnel à la valeur de l'indicateur.

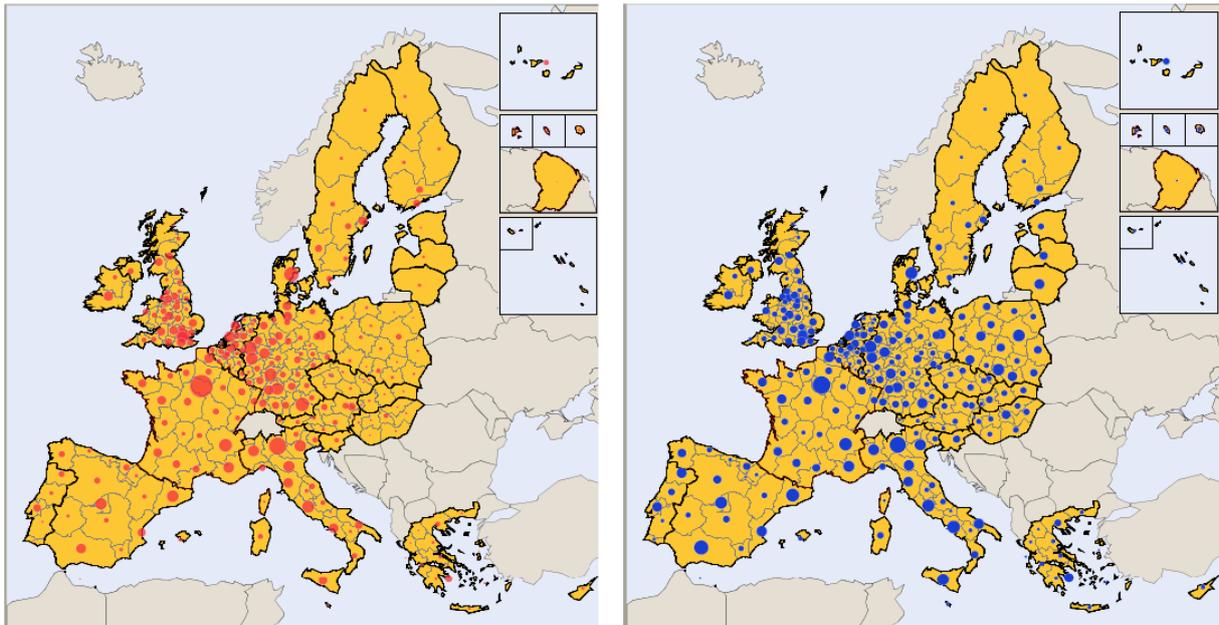


Figure 28 : Les cartes à disques sous HyperAtlas.

Les cartes de la Figure 28 présentent d'une part le produit intérieur à brut (image de gauche) et d'autre part la population totale en 1999 (image de droite).

La couleur, la taille et le degré de transparence des disques peuvent être modifiés par l'utilisateur.

2.1.2.3. La carte de ratio

Cette carte est de type choroplèthe. Elle permet donc la visualisation de quantités grâce à une échelle de couleurs. La carte de ratio est ainsi la représentation du phénomène correspondant à la division du numérateur par le dénominateur.

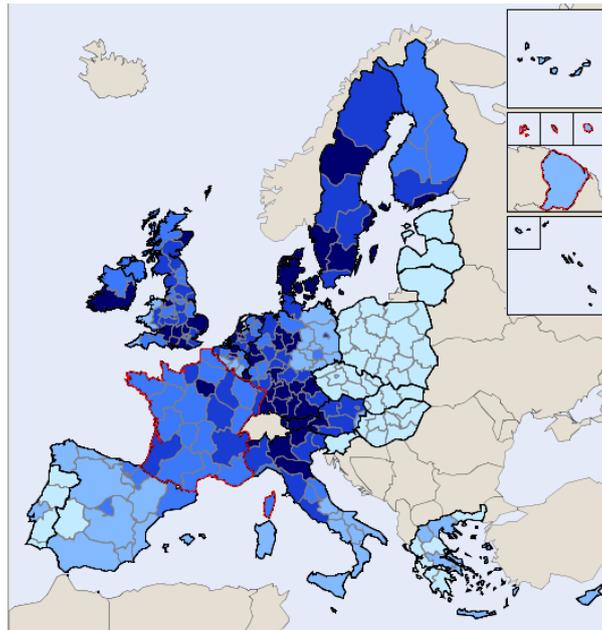


Figure 29 : Une carte de ratio.

La carte ci-dessus correspond au ratio des indicateurs cités dans le paragraphe 2.1.2.2 de ce chapitre, elle permet donc d'analyser la valeur du PIB par habitant dans l'Union Européenne en 1999.

Les options paramétrables permettent notamment de modifier le choix de la palette de couleurs et le nombre de classes.

2.1.2.4. Les cartes de déviation

Comme nous l'avons déjà vu, il existe trois types de déviation ; on distingue donc autant de cartes pour les représenter. Comme dans le cas du ratio, ces cartes sont dites choroplèthes. Elles permettent de comparer pour chaque unité la valeur du ratio par rapport à une moyenne (globale, moyenne ou locale) de ce phénomène. Ces déviations s'expriment en pourcentage.

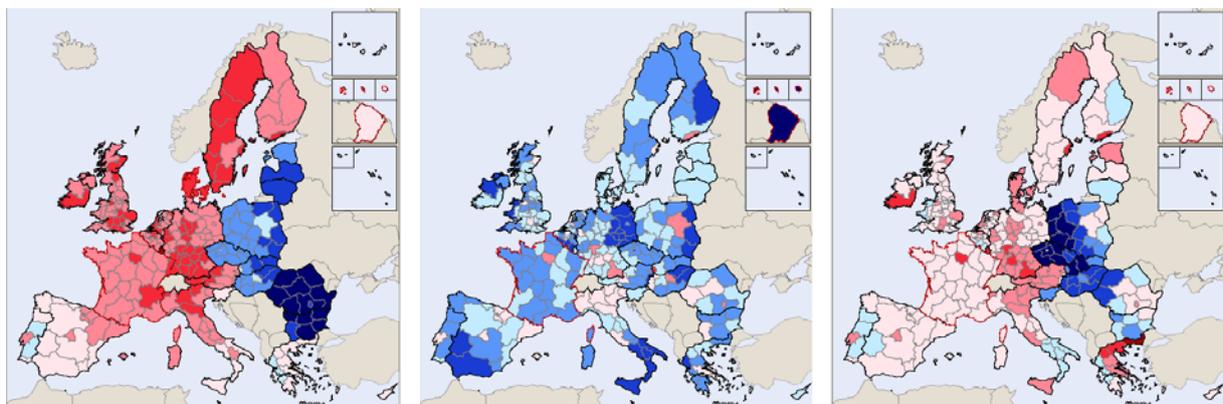


Figure 30 : Les cartes de déviation.

Les différentes cartes de la Figure 30 montrent de gauche à droite les déviations globale, moyenne et locale pour la représentation du PIB par habitant en 1999.

L'utilisateur peut spécifier la palette de couleurs à utiliser et agir également sur la méthode de discrétisation en choisissant le mode de progression (arithmétique ou géométrique) ainsi que le nombre de classes.

2.1.2.5. La carte de synthèse

Le dernier onglet permet d'afficher une carte de synthèse qui récapitule les résultats obtenus sur les différentes cartes de déviation. Ainsi, à chaque unité est associée une couleur qui dépend des résultats obtenus sur les trois cartes de déviation. D'un coup d'œil, l'utilisateur peut ainsi voir quelles sont les régions les plus favorisées et celles qui le sont moins bien. La Figure 31 présente la synthèse de notre analyse sur le PIB par personne en Europe et en 1999.

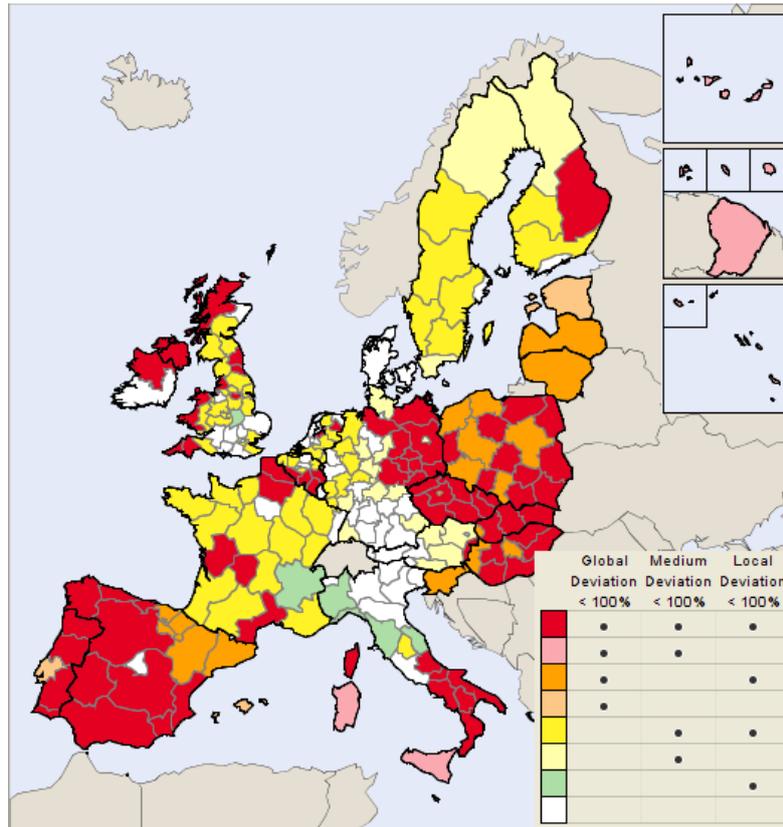


Figure 31 : La carte de synthèse.

Les couleurs ne sont pas paramétrables, seuls les critères de la synthèse peuvent être modifiés par l'utilisateur. Il pourra ainsi choisir la valeur seuil (fixée à 100% par défaut) mais aussi le sens de la requête « plus grand que » ou « plus petit que ».

Au niveau de l'interactivité, l'utilisateur peut désormais sélectionner chaque unité pour afficher un diagramme en bâtons (cf. Figure 32) permettant de comparer plus facilement les résultats des trois déviations pour une unité donnée.

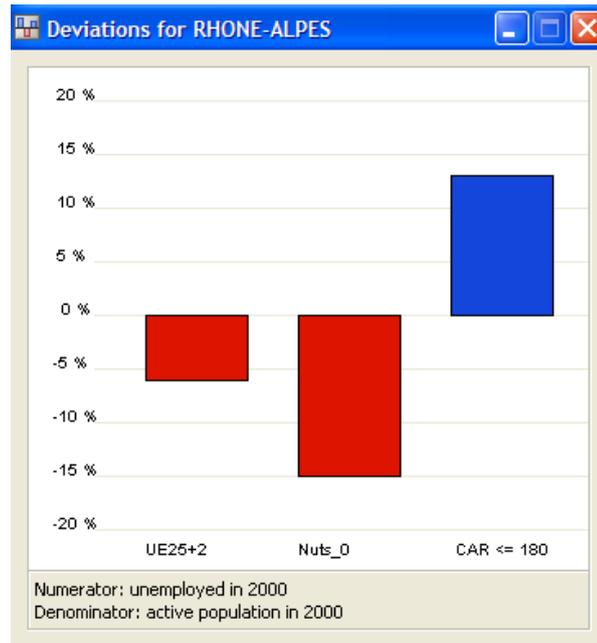


Figure 32 : Histogramme récapitulatif présentant la position de la Région Rhône-Alpes face au taux de chômage selon différentes échelles.

2.1.3. Sauvegarde et export de l'analyse

Une fois l'analyse terminée, l'utilisateur a la possibilité de sauvegarder le contexte de l'analyse ou de générer un rapport.

La sauvegarde du contexte de l'analyse enregistre l'ensemble des paramètres dans un fichier au format XML. L'utilisateur pourra alors recharger ces paramètres lors d'une prochaine session de travail afin de poursuivre son analyse. Tous les paramètres de l'analyse seront restaurés : choix des indicateurs, niveau de zoom, contexte des déviations.

La génération d'un rapport permet d'obtenir une version imprimable ou diffusable via Internet d'une analyse menée avec HyperAtlas. Le rapport ainsi généré contient un récapitulatif des paramètres de l'analyse, les 7 cartes ainsi qu'un tableau récapitulatif des stocks et des résultats obtenus (ratio et déviations). Le format du rapport est XHTML²⁵, standard du W3C²⁶ permettant de créer facilement des fichiers consultables en ligne.

²⁵ XHTML : *eXtensible HyperText Markup Language*

²⁶ W3C : *World Wide Web Consortium* ; consortium établissant des recommandations pour les technologies liées à Internet.

3. Un module d'administration : HyperAdmin

3.1. Problématique

Comme nous l'avons vu, HyperAtlas permet d'analyser des phénomènes statistiques en utilisant des fichiers sérialisés : les fichiers « .hyp ». Ce sont des fichiers binaires qui regroupent l'ensemble des données nécessaires à la visualisation de ces cartes. Ces différentes données regroupent :

- Les données géographiques : coordonnées des différentes unités ;
- Les données de structure : liste des unités, des maillages et des aires d'études ainsi que la hiérarchie entre ces différents éléments ;
- Les données statistiques : liste des stocks et valeur de chaque stock pour chaque unité.

Ces informations sont regroupées dans un ensemble de fichiers textuels : fichiers textes, Excel et Mif/Mid. La structure de ces fichiers a été définie au début du projet HyperCarte par les géographes. Nous présentons la structure de ces fichiers dans le paragraphe suivant.

Ainsi, pour obtenir un nouveau jeu de données, il suffit d'acquérir les informations stockées dans ces fichiers, de les traiter et de fournir un fichier sérialisé comprenant l'ensemble de ces informations et qui est directement utilisable par HyperAtlas. Le programme spécifique en charge de cette intégration s'appelle HyperAdmin.

Ce traitement est essentiellement réalisé au sein de l'équipe Steamer à Grenoble. Les fichiers de données sont transmis par les géographes, puis l'ingénieur informaticien utilise HyperAdmin pour créer un nouveau jeu de données qui est ensuite de nouveau envoyé aux collègues géographes. Cette phase d'intégration peut se révéler fastidieuse. En effet, en cas d'erreurs sur le jeu de données, des échanges sont alors effectués par les deux équipes afin de les corriger. Pour palier cette difficulté mais aussi pour donner plus de liberté et d'autonomie aux géographes, le logiciel HyperAdmin a été livré aux géographes début 2007. Avant cette date, ils n'avaient en leur possession qu'un prototype qu'ils jugeaient complexe d'utilisation. Mon prédécesseur avait alors diagnostiqué les défauts de ce prototype afin de fournir une version répondant davantage aux attentes des géographes.

Dans les paragraphes suivants, nous présentons la version actuelle d'HyperAdmin, en nous intéressant aussi bien à ses fonctionnalités qu'à la structure des fichiers fournis par les géographes.

3.2. L'interface utilisateur

L'interface utilisateur d'HyperAdmin est issue de celle d'HyperAtlas. A vrai dire, il s'agit de l'interface d'HyperAtlas enrichie d'un menu supplémentaire. En effet, HyperAdmin hérite du code d'HyperAtlas. L'interface graphique est donc identique et les fonctionnalités d'HyperAtlas sont incluses dans HyperAdmin. On retrouve donc la zone de paramétrage mais aussi les sept différentes cartes d'HyperAtlas.

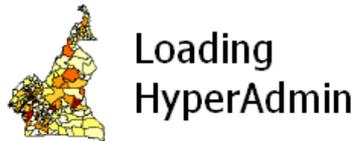


Figure 33 : Image de chargement d'HyperAdmin.

L'image ci-dessus présente l'image de chargement d'HyperAdmin. La copie d'écran ci-dessous montre l'application :

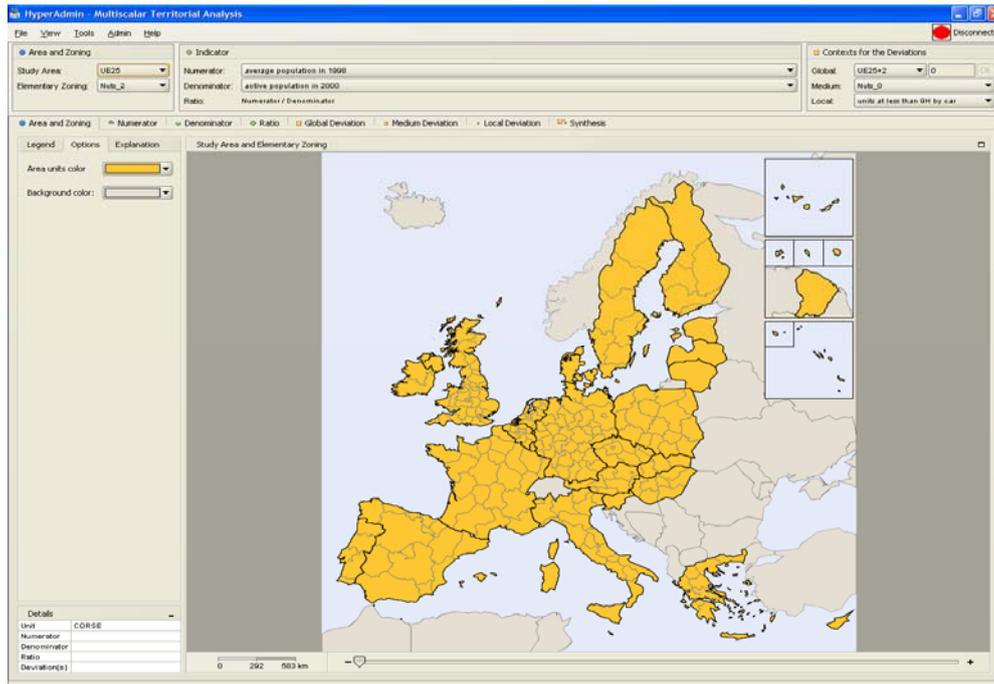


Figure 34 : L'interface utilisateur d'HyperAdmin.

Comme nous l'avons déjà dit, la différence entre ces deux applications est le menu « Admin ». Ce menu se présente de la façon suivante :

- *Database connection* – Connexion à la base de données
- *Define a connection* – Définir une connexion
 - *Connect*– Connexion
 - *Disconnect* – Déconnexion
- *Open project* – Ouvrir un projet
 - *Import project from database* – Charger un projet depuis la base de données
 - *Read project from files* – Créer un projet à partir des fichiers
- *Neighbourhood* – Voisinage
 - *Create a neighbourhood* – Créer une relation de voisinage
- *Export project in serial file* – Exporter le projet dans un fichier sérialisé
- *Commit* – « Sauvegarder en base de données »

Ce menu nous indique les différentes actions que l'utilisateur pourra effectuer. Il nous renseigne donc sur les différentes fonctionnalités d'HyperAdmin. Cependant, avant de nous intéresser à ces fonctionnalités, nous commençons par présenter la structure et le

format des fichiers de données qui seront utilisés lors du processus de création d'un nouveau jeu de données.

3.3. Les fichiers de données

Pour produire un jeu de données avec HyperAdmin, il faut fournir différents fichiers qui contiennent des données géographiques au format Mif-Mid, des données statistiques ainsi que des données structurées au format Excel ou texte.

3.3.1. Les données géographiques

Les informations géographiques sont stockées dans des fichiers Mif-Mid. Il s'agit d'un format d'exportation issu du logiciel MapInfo. Le fichier ayant pour extension .mif contient la structure de la table ainsi que la géométrie (ensemble de coordonnées) tandis que le fichier .mid regroupe les informations attributaires.

Voici un exemple de fichiers Mif-Mid :

Fichier .mid	Fichier .mif
"FR711"	Region 1
"FR712"	29
"FR713"	-724490.002 -438732.516
"FR714"	...
"FR715"	-724490.002 -438732.516
"FR716"	Pen (1,2,0)
"FR717"	Brush (2,16777215,16777215)
"FR718"	Center -742036.417 -468693.351
...	...

Code 1 : Extrait de fichiers Mif-Mid.

Le fichier .mid recense donc la liste des unités. Etant donné qu'HyperAtlas ne nécessite que les unités de plus bas niveau (les autres étant construites par agrégation), seules les unités de plus bas niveau sont représentées dans ces fichiers Mif-Mid. L'extrait de ce fichier .mif liste les départements de la région Rhône-Alpes.

Le fichier .mif contient davantage d'informations. Chaque unité est définie dans ce fichier de la manière suivante :

```
Region 1
  Nombre de points
  Liste des coordonnées de chaque point
  Options de tracé
```

Code 2 : Syntaxe du fichier .mif.

Ainsi, pour trouver les informations correspondant à l'Isère (code : FR714), qui se trouve à la ligne 487 du fichier .mid, il faut trouver la 487^{ème} occurrence de « Region 1 ». On apprend alors que pour tracer le contour de cette unité, il faut 29 points et les 29 prochaines lignes du fichier .Mif recensent les coordonnées de ces points. Les options d'affichage ne sont pas traitées par HyperAtlas. Cependant les coordonnées du **CENTROÏDE** (Center) pourraient être utiles pour placer les disques proportionnels des cartes d'analyse sur ce point. En effet, actuellement, le centre d'une unité est calculé. Pour cela, on construit un rectangle autour de l'unité et on en déduit le centre. Cette méthode a quelques inconvénients. En effet, pour certaines unités, notamment en Grèce, le centroïde ainsi déterminé, peut se retrouver en pleine mer.

A partir de ce couple de fichiers, il est donc possible de tracer les contours d'une carte. Si la lecture et la compréhension de ce format sont fastidieuses pour un être humain, il est très facile de créer un programme (un *parser*) pour en extraire les informations. Un *parser* est en fait un analyseur syntaxique qui permet de « décoder » un texte répondant à une grammaire formelle grâce à des règles syntaxiques. HyperAdmin est donc doté d'un analyseur syntaxique afin de traiter les données géographiques.

3.3.2. Les données structurelles

Les données structurelles vont permettre d'organiser les unités géographiques que nous venons d'extraire avec le *parser*.

Ces données peuvent être stockées aussi bien dans des fichiers textes que des classeurs Excel. Ce dernier format est le plus fréquemment utilisé, car il permet de regrouper toutes les informations structurelles dans un même fichier. Chaque type d'information étant stocké dans une feuille de calculs (onglet) différente.

Ces fichiers (ou feuilles de calculs) permettent de spécifier :

- Les aires d'études
 - *Area* : liste les différentes aires d'études qui seront disponibles

```
Area_ID
UE15
UE25
PECO
Arc_Atlantique
Candidat_UE
...
```

Code 3 : Extrait des fichiers de données d'HyperAdmin : Area.

- *AreaLanguage* : permet d'associer à chaque aire d'étude un label

Area_ID	Language_ID	Area_NAME
UE15	FR	Union européenne des 15
UE25	FR	Union européenne des 25
PECO	FR	Pays d'Europe Centrale et Orientale
Arc_Atlantique	FR	Arc Atlantique
Candidat_UE	FR	Candidats de l'Union européenne
...		

Code 4 : Extrait des fichiers de données d'HyperAdmin : AreaLanguage.

- Les maillages
 - *Zoning* : liste les différents maillages disponibles ainsi que le niveau de chaque maillage

Zoning_ID	Rank
Nuts_0	1
Nuts_1	2
Nuts_2	3
Nuts_3	5
Nuts_2_3	4

Code 5 : Extrait des fichiers de données d'HyperAdmin : Zoning.

- *ZoningLanguage* : permet d'associer un libellé à chaque maillage

Zoning_ID	Language_ID	Zoning_NAME
Nuts_0	FR	Nomenclature des UT de niveau 0
Nuts_1	FR	Nomenclature des UT de niveau 1
Nuts_2	FR	Nomenclature des UT de niveau 2
Nuts_3	FR	Nomenclature des UT de niveau 3

Code 6 : Extrait des fichiers de données d'HyperAdmin : ZoningLanguage.

- Les unités
 - *Unit* : liste les différentes unités

UT_ID
...
FR711
FR712
FR713
FR714
FR715
FR716
FR717
FR718
...

Code 7 : Extrait des fichiers de données d'HyperAdmin : Unit.

- *UnitLanguage* : permet de donner un nom à chaque unité

UT_ID	Language_ID	UT_NAME
...		
FR711	FR	AIN
FR712	FR	ARDECHE
FR713	FR	DROME
FR714	FR	ISERE
FR715	FR	LOIRE
FR716	FR	RHONE

FR717	FR	SAVOIE
FR718	FR	HAUTE-SAVOIE
...		

Code 8 : Extrait des fichiers de données d'HyperAdmin : UnitLanguage.

- *UnitSup* : permet de définir des relations entre les unités (quelle est l'unité qui englobe l'Isère ?)

UT_ID	UTSup_ID
...	
FR711	FR71
FR712	FR71
FR713	FR71
FR714	FR71
FR715	FR71
FR716	FR71
FR717	FR71
FR718	FR71
...	

Code 9 : Extrait des fichiers de données d'HyperAdmin : UnitSup.

- *UnitArea* : permet de définir quelles sont les unités disponibles pour une aire d'étude

UT_ID	Area_ID
AL	PECO
BA	PECO
BG	PECO
EE	PECO
MK	PECO
HU	PECO
LT	PECO
LV	PECO
PL	PECO
RO	PECO
CZ	PECO
SI	PECO
SK	PECO

Code 10 : Extrait des fichiers de données d'HyperAdmin : UnitArea.

- *UnitZoning* : permet d'associer une unité à un maillage administratif

UT_ID	Zoning_ID
FR712	Nuts_3
FR713	Nuts_3
FR714	Nuts_3
FR715	Nuts_3
FR716	Nuts_3
FR717	Nuts_3
FR718	Nuts_3

Code 11 : Extrait des fichiers de données d'HyperAdmin : UnitZoning.

- Les relations de voisinage (optionnel)
 - *Contiguity* : liste les relations de voisinage disponibles

ID
CAR

Code 12 : Extrait des fichiers de données d'HyperAdmin : Contiguity.

- o *ContiguityLanguage* : permet d'associer un libellé à chaque contiguïté

Contiguity_ID	Language_ID	Contiguity_NAME
CAR	EN	car time between units (min)
CAR	FR	distance en voiture (min)

Code 13 : Extrait des fichiers de données d'HyperAdmin : ContiguityLanguage.

- o *UnitContiguity* : permet d'affecter une distance entre deux unités pour chaque contiguïté

UT_ID1	UT_ID2	CAR
FR714	AT33	452,8670
FR714	AT34	403,1563
FR714	BE10	465,5800
FR714	BE21	477,6551
FR714	BE22	451,7852
FR714	BE23	477,7868

Code 14 : Extrait des fichiers de données d'HyperAdmin : UnitContiguity.

- o *Neighbourhood* : permet d'affiner les contiguïtés en y ajoutant un comparateur (<, >, <=, >=) et une distance seuil

Neighbourhood_ID	Contiguity_ID	distance	comparator
CAR <= 480	CAR	480	<=
CAR <= 360	CAR	360	<=
CAR <= 240	CAR	240	<=
CAR <= 120	CAR	120	<=

Code 15 : Extrait des fichiers de données d'HyperAdmin : Neighbourhood.

- o *NeighbourhoodLanguage* : permet d'associer un libellé à chacune des relations de voisinages

Neighbourhood_ID	Language_ID	Neighbourhood_NAME
CAR <= 480	EN	less than 8H by car
CAR <= 360	EN	less than 6H by car
CAR <= 240	EN	less than 4H by car
CAR <= 120	EN	less than 2H by car

Code 16 : Extrait des fichiers de données d'HyperAdmin : NeighbourhoodLanguage.

3.3.3. Les données statistiques

Tout comme les données de structures, ces données sont stockées dans des fichiers textes ou des fichiers Excel.

Pour spécifier ce type d'information, on utilise deux fichiers (ou feuilles de calculs Excel) qui regroupent, d'une part la définition des stocks (code et labels), et, d'autre part, l'initialisation pour chaque unité élémentaire d'une valeur pour l'ensemble des indicateurs. La contrainte forte de ce fichier réside dans l'association d'une valeur de stock à chaque indicateur. La présence d'une valeur non renseignée n'est en effet pas supportée par l'application.

Le tableau suivant présente la définition des indicateurs statistiques. On constate que chaque indicateur est défini par une clé (colonne stock) et un ensemble de libellés (colonne « code-pays », ex : FR).

stock	EN	FR
AREA_99E	area in km ² in 1999	superficie de l'unité territoriale en km ²
POPMT98E	average population in 1998	population totale moyenne en 1998
POPMT99E	average population in 1999	population totale moyenne en 1999
GDPEU99E	gross domestic product in euros in 1999	PIB en 1999 en euros
GDPPS99E	gross domestic product in pps in 1999	PIB en 1999 en parité de pouvoir d'achat
...		

Code 17 : Extrait des fichiers de données d'HyperAdmin : StockLanguage.

L'extrait de fichier ci-dessous montre l'affectation de valeurs statistiques pour chaque couple (unité, stock).

UT_ID	GDPEU99E	GDPPS99E	AREA_99E	POPMT99E	POPMH99E
FR712	4773000000	4502600000	5528,6	286000	141000
FR713	9504200000	8965800000	6529,9	439000	214000
FR714	25099800000	23677900000	7431,4	1098000	540000
FR715	13882300000	13095800000	4780,6	729000	351000
FR716	43663900000	41190200000	3249,1	1582000	760000
FR717	9018100000	8507200000	6028,3	374000	184000
FR718	14482200000	13661700000	4387,8	634000	313000
...					

Code 18 : Extrait des fichiers de données d'HyperAdmin : Stock.

Si l'ensemble de ces fichiers est bien structuré, ces fichiers ne peuvent être construits que par des experts en géographie. Ces données sont généralement fournies par l'UMS-Riate, mais d'autres points focaux du projet peuvent également intervenir pour transmettre ces fichiers. Ces différents points focaux ont permis l'élaboration de jeux de données pour la Roumanie, le Cameroun, la Tunisie ou encore la région Rhône-Alpes.

La présentation de la structure des fichiers de données et la complexité qui en résulte, permet de présenter une deuxième catégorie d'utilisateurs : les experts. En effet, HyperAtlas est une application prête à être utilisée par des analystes ou des politiques. HyperAdmin ne s'adresse pas au même public mais davantage à des experts capables d'administrer des données statistiques et géographiques.

3.4. Fonctionnalités

Le menu « Admin » nous renseigne sur les différentes possibilités pour la création d'un jeu de données. On s'aperçoit, en effet, que l'on peut soit créer un projet à partir des fichiers précédemment vus, soit charger un projet déjà sauvegardé en base de données. Etant donné, que pour charger un projet existant, il faut en premier lieu le créer à partir des fichiers de données, nous commençons par présenter cette étape.

3.4.1. Création d'un nouveau jeu de données

La création d'un nouveau jeu de données s'effectue à travers plusieurs écrans. Le premier d'entre eux permet de nommer le projet, de lui donner une description et de fournir l'emplacement des fichiers de données. Il faut également préciser le type de données.



Figure 35 : Création d'un projet.

Suite à la validation de cet écran, HyperAdmin parcourt les fichiers contenus dans le répertoire saisi par l'utilisateur à la recherche des fichiers de données. Il va ensuite les *parser* c'est-à-dire en extraire les informations afin de créer ce nouveau projet. Cette étape peut prendre quelques minutes selon la taille des fichiers.

3.4.2. Calcul de la contiguïté

Deux unités sont dites contiguës si elles sont situées dans un voisinage immédiat. La contiguïté permet donc de déterminer quelles unités géographiques sont voisines. Cette information est alors ensuite utilisée pour le calcul des déviations locales.

HyperAdmin permet de calculer la contiguïté. Deux méthodes sont implémentées dans ce logiciel pour effectuer ce calcul. La première repose sur l'utilisation d'une API Java nommée JTS (acronyme de *Java Topology Suite*), tandis que la seconde utilise les fonctionnalités spatiales de la base de données PostGreSQL avec son extension PostGIS.

Etant donné que les données sont localisées en mémoire, l'utilisation de JTS permet de travailler directement avec ces dernières. L'utilisation de PostGIS implique par contre que les données soient stockées en base de données. Mon prédécesseur avait alors effectué des tests de performance afin de choisir la solution la plus adaptée.

Ces tests n'ont pas écarté l'une des deux technologies. Au contraire, ils ont permis de montrer que JTS est plus performant que PostGIS quand le nombre d'unités à traiter n'est pas trop élevé. En effet, à partir de 1000 unités l'utilisation de PostGIS engendre de meilleurs résultats. De plus, en calculant les contiguïtés avec l'extension spatiale de PostGreSQL, on s'affranchit des possibles limites liées à la machine virtuelle de Java (JVM).

La contiguïté ainsi calculée se présente sous la forme d'une matrice, où pour chaque unité on est capable de déterminer si elle est voisine ou non d'une autre unité.

3.4.3. Création de relations de voisinage

Lorsqu'un projet possède une contiguïté, il est possible d'ajouter la notion de voisinage. Cela permet d'enrichir le jeu de données et de proposer à l'utilisateur de nouveaux contextes pour l'étude d'un phénomène via les déviations locales. La Figure 36 présente l'interface graphique permettant de réaliser cette opération.

Les relations de voisinage permettent de déterminer si deux unités sont voisines. Si la notion de contiguïté entre deux unités est vérifiée lorsque ces deux dernières ont une frontière en commun, la notion de voisinage est plus complexe. En effet, désormais deux

unités peuvent être considérées comme voisines si elles sont situées à moins de 2h en voiture. La notion de distance utilisée ne respecte pas les distances mathématiques. En effet, selon la définition d'une distance en mathématiques, l'une des propriétés précise que la distance d'un point A à un point B est la même que la distance entre le point B et le point A.

$$d(A, B) = d(B, A)$$

Il s'agit de la propriété de symétrie d'une distance. Or en géographie, la distance en temps entre deux unités n'est pas forcément symétrique. L'espace géographique n'étant pas isotrope, un véhicule peut parcourir la distance entre le point A et le point B plus rapidement que lors du voyage retour à cause notamment du relief et de l'état des routes. Dans les applications du projet HyperCarte, la notion de voisinage est donc un triplet composé d'une matrice de distance, d'un seuil numérique et d'un opérateur de comparaison. Par exemple, si on estime que deux unités sont voisines si elles sont situées à moins de 3h en train, on peut en conclure que les villes de Paris et Grenoble sont voisines pour cette relation de voisinage.

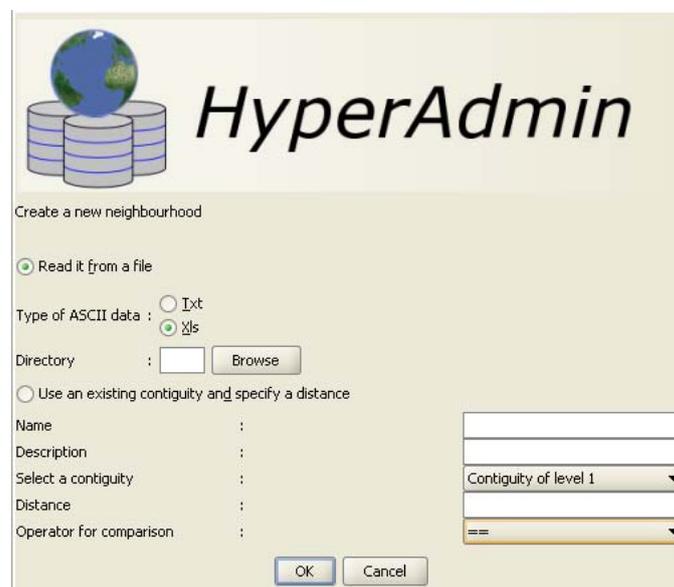


Figure 36 : Création de matrices de voisinages.

3.4.4. Sauvegarde et export

Une fois le jeu de données correctement initialisé, l'utilisateur peut le stocker en base de données. Cela permet de retravailler par la suite ce projet en l'enrichissant avec, par exemple, de nouvelles matrices de voisinages.

Cependant, si la sauvegarde en base n'est que facultative, la possibilité d'exporter le projet dans un fichier sérialisé : le fichier « .hyp » est vivement recommandé. La sérialisation est un processus permettant d'encoder des objets en mémoire sous la forme d'une suite d'octets. Le fait de sauvegarder le résultat de cette manipulation de données dans un fichier .hyp permet à l'utilisateur de créer un jeu de données qui sera ensuite directement exploitable par HyperAtlas.

4. Un module d'analyse spatiale : HyperSmooth

Le projet HyperCarte comporte un module d'analyse spatiale multiscalaire : HyperSmooth. HyperSmooth est le dernier né du projet HyperCarte. Ce nouvel outil cartographique pour l'analyse spatiale de phénomènes statistiques met en œuvre une méthode dite de potentiel. L'objectif de ce programme est de représenter de façon continue une information statistique échantillonnée sur des maillages administratifs.

Le développement de ce logiciel n'en est qu'à son début et je n'ai pas beaucoup travaillé sur HyperSmooth. Cette partie ne présente que les notions nécessaires à la bonne compréhension de la problématique de ce prototype.

4.1. Principes

L'analyse spatiale s'intéresse aux propriétés des objets géographiques, qu'ils soient physiques (une zone agricole, une forêt, une zone industrielle), ou humains (noyau de peuplement, réseau de transport). L'objectif de l'analyse spatiale est de montrer l'importance de la prise en compte des localisations pour la connaissance de ces objets afin d'en expliquer des caractéristiques et à en comprendre les dynamiques. Les localisations sont donc appréhendées comme un facteur susceptible d'intervenir dans l'explication de la diversité des lieux et partant, des modes d'organisation de l'espace qui y sont associés. Désormais au cœur des usages qui sont faits des Systèmes d'Information Géographique, l'analyse spatiale donne une grande variété d'outils au géographe. Elle apprend à résumer une information géographique, telle que celle contenue dans une carte. Pour un espace donné, elle aide à reconnaître des structures spatiales, par exemple celle d'un réseau de villes. Enfin, elle propose à la fois des outils de modélisation des structures spatiales comme, par exemple, le modèle centre-périphérie, les champs d'interaction de type gravitaire, l'effet de la distance sur une influence ou sur un flux d'échanges. [@Hypergeo]

L'approche spatiale permet de s'affranchir des maillages administratifs. En effet, si pour l'analyse territoriale, un indicateur est associé à chaque unité (cf. Figure 37), ce n'est pas le cas pour l'analyse spatiale où le résultat ressemble davantage à des courbes de tendance (cf. Figure 38).

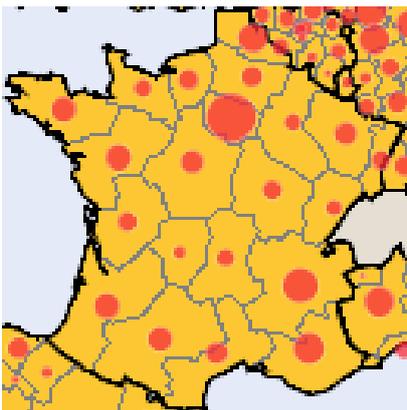


Figure 37 : Analyse territoriale.



Figure 38 : Analyse spatiale.

Elle analyse des processus qui sont à l'origine de ces structures, à travers des concepts comme ceux de distance, d'interaction spatiale, de portée spatiale, de polarisation, de centralité, de stratégie ou choix spatial, de territorialité... Des lois de la

spatialité relie ces formes et ces processus, et sont intégrées dans des théories et des modèles du fonctionnement et de l'évolution des systèmes spatiaux.

De nombreuses méthodes peuvent être utilisées dans ce contexte. Dans le cadre du projet HyperSmooth, la méthode retenue est la méthode de transformation par potentiel [Plumejeaud, 2007], [Plumejeaud et al., 2007]. Il en existe cependant de nombreuses autres. Certaines sont ainsi basées sur l'interpolation spatiale (par exemple : triangulation, Krigeage, méthode de Shepard), d'autres sont fondées sur la géostatistique (Krigeage par bloc) [Grasland et al., 2000].

Les calculs basés sur ces méthodes sont très coûteux, c'est pourquoi une architecture client/serveur a été adoptée pour HyperSmooth. L'ensemble des calculs est en fait effectué par un serveur de calculs lui-même relié à une grappe de PC. Les échanges entre ce serveur et le poste client s'effectuent à l'aide du protocole SOAP. Il s'agit d'un protocole RPC²⁷ basé sur XML permettant donc l'invocation de méthodes distantes. L'ensemble des messages transite via le protocole http et afin de garantir la sécurité des données transitant sur le réseau, les informations sont codées avec le protocole TLS²⁸ anciennement dénommé SSL²⁹. La Figure 39 présente cette architecture.

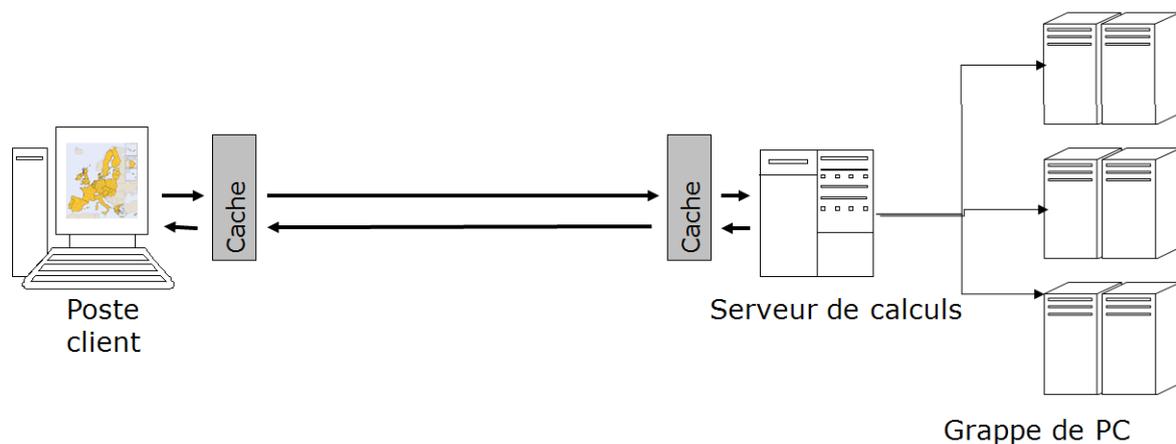


Figure 39 : Architecture d'HyperSmooth.

4.2. HyperSmooth

Comme HyperAdmin, HyperSmooth hérite des fonctionnalités et de l'aspect graphique d'HyperAtlas. Nous ne revenons donc pas sur la présentation de ces aspects.

Pour lancer une requête d'analyse spatiale, l'utilisateur sélectionne l'espace d'étude à l'aide de la souris (sélection rectangulaire) ou bien à l'aide du menu contextuel pour se focaliser davantage sur une unité particulière. Il choisit ensuite les indicateurs, la résolution, la fonction d'interaction ainsi que deux portées. En choisissant deux portées, il est en effet possible d'étudier le même phénomène à un niveau local et à un niveau global [Grasland, 2003].

²⁷ RPC : *Remote Procedure Call* ; protocole permettant l'appel de procédures distantes.

²⁸ TLS : *Transport Layer Security* ; protocole de sécurisation des échanges sur Internet.

²⁹ SSL : *Secure Socket Layer*

Une fois ces paramètres choisis, l'utilisateur doit valider son choix afin de transmettre via le protocole SOAP la requête au serveur de calculs. Lorsque les calculs sont terminés, le serveur renvoie à l'application cliente des tableaux de données. HyperSmooth doit alors les convertir en images raster (cf. Figure 40) qui sont alors calées sur le fond de carte (cf. Figure 41).

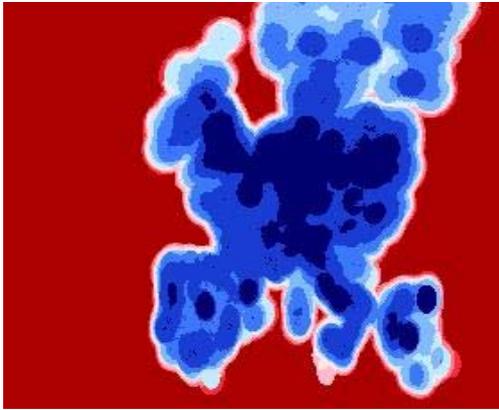


Figure 40 : Image Raster construite par HyperSmooth (non plaquée).

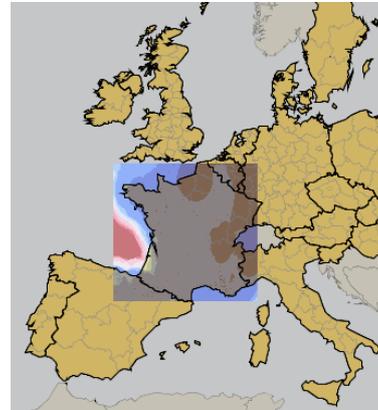


Figure 41 : Image raster plaquée sur le fond de carte.

5. Architecture et choix technologiques

Dans cette partie, nous nous intéressons à l'architecture logicielle retenue pour ces deux applications mais nous décrivons également l'environnement de développement.

5.1. Le modèle MVC

HyperAtlas est une application écrite entièrement en Java. Respectant le paradigme MVC, on retrouve donc les 3 couches :

- Le modèle ;
- La vue ;
- Le contrôleur.

Ce modèle MVC est un *design pattern* c'est-à-dire un patron de conception. Un *design pattern* est un concept dont l'objectif est d'aider à la résolution de problèmes récurrents. Ce modèle sépare les données (issues du modèle) de l'interface (la vue) et des contrôles. Il s'agit donc d'une architecture multi-tiers [Shin, 2003], [Thomas, 2006].

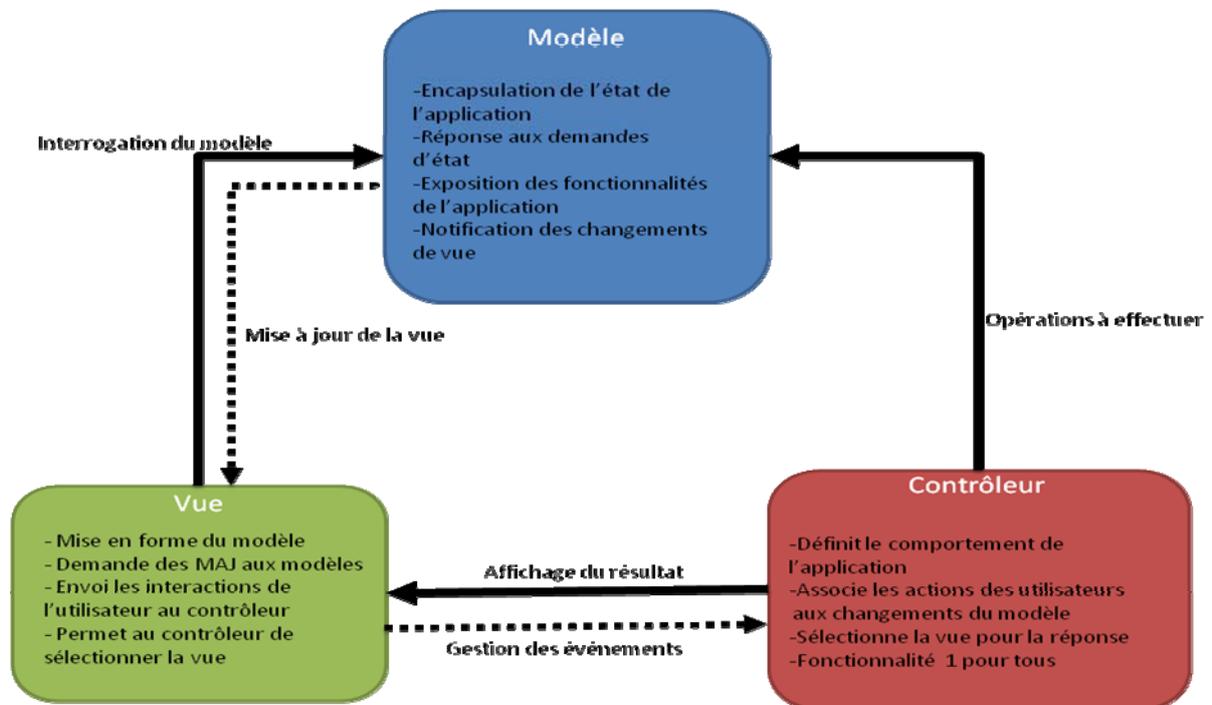


Figure 42 : Le modèle MVC.

La figure ci-dessus, est une francisation d'un schéma de Sang Shin [Shin, 2003]. Ce schéma met en évidence les différents rôles des trois couches ainsi que leurs interactions. Nous allons désormais détailler chaque couche.

5.1.1. Le modèle

Le modèle représente la classe métier de l'application. Une classe métier permet de créer des objets métiers à partir d'un entrepôt de données (fichier, bases de données). Le modèle contient donc les données manipulées par le programme. Il en assure la gestion et garantit leur intégrité. [Thomas, 2006]

Dans le cas d'HyperAtlas, les objets métiers existent déjà sous forme de fichiers sérialisés. La sérialisation permet de stocker des données dans un fichier pour les restituer ultérieurement. La création de ces fichiers sérialisés est effectuée par HyperAdmin et nous nous intéressons à ce processus dans le prochain chapitre.

Ainsi, HyperAtlas crée son modèle de données en lisant le fichier sérialisé contenant toutes les informations nécessaires à l'initialisation d'un jeu de données. On retrouve donc l'ensemble des unités territoriales, des maillages administratifs, des espaces d'étude ainsi que les indicateurs.

Le modèle permet également d'effectuer des traitements sur ces objets. L'ensemble de ces traitements est rassemblé dans une même classe intitulée `Logic`. Parmi ces traitements, on trouve :

- Le calcul d'un ratio ;
- Le calcul des différentes déviations.

5.1.2. La vue

La vue fait l'interface avec l'utilisateur. Sa première tâche est d'afficher les données qu'elle a récupérées auprès du modèle. Sa seconde tâche est de recevoir toutes les actions de l'utilisateur (clic, sélection, boutons). Ces différents événements sont envoyés au contrôleur. [Thomas, 2006]

L'application ayant été développée en Java, la couche graphique repose sur Java Swing. Swing est une bibliothèque graphique pour le langage Java.

Swing respecte également le modèle MVC [Shin, 2003]. Ainsi, on notera que la plupart des composants Swing utilisent une classe spécifique pour stocker les données. Les actions sur certains de ces composants sont également gérées par une autre classe. Le tableau suivant montre ce découpage.

Tableau 2 : Le paradigme MVC dans Swing.

Composant (Vue)	Données (Modèle)	Sélection (Contrôleur)
JTable	TableModel	ListSelectionModel
JList	ListModel	ListSelectionModel
JTree	TreeModel	TreeSelectionModel

5.1.2.1. Look and Feel

L'architecture de Swing est conçue pour que l'utilisateur puisse changer le *Look and Feel* de l'interface de son application. Le *look and feel* souvent noté « laf » ou encore « l&f » est l'ensemble des spécificités et des caractéristiques d'une interface qui lui donnent une identité et qui peut être perçu de manières différentes selon les utilisateurs.

Le changement du *Look and Feel* d'une application a une incidence sur les polices de caractères, les couleurs, les formes et la disposition des composants graphiques (par exemple : liste déroulante, bouton).

Parmi les apparences fournies par défaut dans la JRE de Sun, on trouve le `SystemLookAndFeel`. Avec ce thème, l'application utilise le même look que le système d'exploitation. Si ces thèmes ne conviennent pas, il est possible de créer son propre *Look and Feel* en surchargeant un style existant. Modifier l'apparence de chaque composant représente un travail conséquent, c'est pourquoi il est possible de télécharger d'autres

Look and Feel. Parmi les différentes solutions existantes, mes prédécesseurs ont décidé d'opter pour les JGoodies Looks. Si HyperAtlas bénéficie du rendu des JGoodies Looks depuis longtemps, la version de cette librairie a été mise à jour. Ce changement a amélioré légèrement l'interface graphique notamment en éclaircissant la couleur de fond de l'application. [@JGoodies]

5.1.2.2. SwingX

SwingX, abréviation de Swing eXtension, est un projet de SwingLabs. L'objectif de ce projet est de concevoir de nouveaux composants Swing et d'améliorer les composants existants. Sun souhaitant travailler sur l'avenir des interfaces graphiques en Java, il soutient largement ce projet et participe même au financement. Si SwingX est un projet récent, il évolue très rapidement et certaines fonctionnalités sont peu à peu implémentées dans les nouvelles versions de Java. [@SwingX]

Parmi les composants de ce **FRAMEWORK** que j'ai utilisé au cours de cette année, on trouve `JXTable` qui implémente des fonctionnalités supplémentaires au `JTable` comme le tri des colonnes. Cette implémentation a d'ailleurs été intégrée dans Java 1.6. Ainsi, en intégrant cette bibliothèque, j'ai pu rendre certains objets graphiques plus interactifs.

5.1.3. Le contrôleur

Le contrôleur est chargé de la synchronisation du modèle et de la vue. Il reçoit tous les événements de l'utilisateur et enclenche les actions à effectuer. Si une action nécessite un changement des données, le contrôleur demande la modification des données au modèle et avertit ensuite la vue que les données ont changé pour que celle-ci se mette à jour. Certains événements de l'utilisateur ne concerne pas les données mais la vue. Dans ce cas, le contrôleur demande à la vue de se modifier. [Thomas, 2006]

Dans HyperAtlas, le contrôleur est centralisé dans la classe Dispatcher. Cette classe émet et reçoit des événements. Il s'agit donc de programmation événementielle. Le fonctionnement de ce composant est largement décrit dans les mémoires précédents. [Bissler, 2004], [Cuenot, 2005], [Chabert, 2007], [Plumejeaud, 2007]

La séparation des couches et le respect du paradigme MVC pour le logiciel HyperAtlas a permis de pouvoir réutiliser facilement certaines fonctionnalités (affichage, gestion des cartes) dans HyperAdmin puis HyperSmooth. Il n'y a ainsi pas de duplication de code ce qui facilite la maintenance de ces applications.

5.2. Choix technologiques

Ce projet a déjà un fort vécu. Mes prédécesseurs avaient déjà mis au point une architecture et un environnement de travail. Ces choix sont décrits dans leurs mémoires respectifs. [Cuenot, 2005], [Chabert, 2007], [Plumejeaud, 2007]

Cependant, pour faciliter les nombreuses livraisons survenues durant cette année, j'ai néanmoins utilisé un nouvel outil. Il s'agit de Launch4J [@Launch4j], un logiciel libre permettant de créer des fichiers exécutables à partir d'une archive Java (.jar). Ce petit outil est très pratique, car il est très paramétrable. Parmi ces fonctionnalités, on trouve :

- Création du fichier .exe ;
- Paramétrage de l'application : nom, icône, version ;
- Paramétrage de Java : vérification de la version de la JVM du client. Si cette version est inférieure (voire supérieure) à celle indiquée, l'utilisateur est redirigé vers la page de téléchargement de la version de Java requise. Il est également possible d'embarquer la JVM dans l'application ;
- Passage d'arguments : utile pour déterminer le look graphique selon le client (ESPON, EEA) ;

L'intérêt d'utiliser un logiciel de ce type, est de faciliter le travail pour l'utilisateur. En effet, en livrant une archive Java exécutable comme un jar (Java Archive), on ne

sait pas comment l'utilisateur va réagir. Il peut en effet double-cliquer directement sur ce fichier et selon la configuration de son environnement cette action peut soit lancer l'application (sans paramètres néanmoins), lister le contenu de l'archive ou encore le système d'exploitation peut demander à l'utilisateur de choisir le programme permettant d'ouvrir ces fichiers. Cette solution n'étant pas satisfaisante, on peut alors également livrer un script qui lancera l'application directement avec les bons paramètres. Jusqu'à présent, chaque livraison était accompagnée d'un script permettant de l'exécuter. L'inconvénient de cette méthode est qu'elle laisse libre accès au `.jar`. On se retrouve donc avec les défauts que nous venons de décrire (possibilité de lancer l'application sans paramètres).

Launch4J est un *wrapper*. Il s'agit donc d'un programme enveloppant un autre programme pour lui préparer un environnement particulier [`@Jargonf`]. On contrôle ainsi beaucoup plus facilement les interactions du logiciel avec l'environnement de l'utilisateur. Cependant, étant donné que le type de fichier `.exe` est associé à Windows, il est préférable de laisser un script qui permette à un utilisateur sous MacOS ou Linux d'utiliser le programme. L'archive Java n'étant qu'enveloppée dans le fichier exécutable, il reste possible de la lancer avec la commande traditionnelle `java -jar`.

6. Conclusion

Comme nous l'avons vu, HyperAtlas est un logiciel de cartographie que l'on peut considérer en version stable et diffusable. Cependant, de nouvelles demandes ont été formulées afin de faire évoluer ce module. Le prochain chapitre dresse la liste des évolutions apportées à HyperAtlas.

HyperAdmin est, quant à lui, le logiciel permettant « d'alimenter » HyperAtlas. A partir de fichiers de données comprenant des informations géographiques mais aussi statistiques, HyperAdmin génère un fichier binaire que l'on nomme « fichier Hyp ». Ce processus de création d'un nouveau jeu de données fonctionne mais n'est pas jugé convaincant par les utilisateurs en raison d'un enchaînement complexe d'actions et d'erreurs d'exécution fréquentes. Ce mécanisme a donc été revu et nous le présentons dans le chapitre suivant.

CHAPITRE 3

Maintenance évolutive et évolutions

Dans ce chapitre, nous présentons le travail réalisé durant cette année de stage. Nous commençons par évoquer la maintenance évolutive sur HyperAtlas. Nous décrivons notamment les nouvelles fonctionnalités ajoutées à l'application existante.

Dans un second temps, nous nous consacrons à HyperAdmin et notamment nous nous intéressons à l'assistant permettant de créer de nouveaux jeux de données.

Sommaire du chapitre :

Chapitre 3 - Maintenance évolutive et évolutions	73
1. Maintenance évolutive sur HyperAtlas	75
1.1.1. Les ratios pertinents	75
1.1.1.1. Contexte.....	75
1.1.1.2. Mise en place	75
1.1.2. Discrétisation.....	77
1.1.2.1. Les écarts relatifs.....	77
1.1.2.2. La nouvelle méthode de discrétisation	77
1.1.2.3. Palettes de couleurs	78
1.1.3. Les indicateurs inconnus.....	80
2. Réingénierie d'HyperAdmin	82
2.1. Problèmes rencontrés	82
2.1.1. Faits techniques	82
2.1.2. Ergonomie.....	82
2.2. Solution proposée et mise en œuvre	83
2.2.1. Le <i>Wizard</i>	83
2.2.2. Gestion des erreurs	86
2.2.2.1. Les erreurs d'intention	87
2.2.2.2. Les erreurs d'exécution.....	89
2.2.3. Affichage.....	90
3. Evolutions d'HyperAdmin.....	91
3.1. Les ratios pertinents.....	91
3.2. Les hiérarchies multiples.....	94
4. Livraisons	96

Evolutions d'outils dédiés à l'analyse territoriale et à l'analyse spatiale

4.1. EEA-HyperAtlas	96
4.2. EP-HyperAtlas	97
4.3. Nordic HyperAtlas	97
4.4. ESPON HyperAtlas.....	98
5. Conclusion	100

1. Maintenance évolutive sur HyperAtlas

La version d'HyperAtlas développée par un stagiaire précédent est une version stable et qui a déjà été livrée. Cependant, de nouvelles fonctionnalités ont été demandées. Il s'agit notamment de mettre en place un système de ratios prédéfinis mais également de changer la méthode de discrétisation. Ces modifications vont permettre à HyperAtlas d'aboutir à une version plus élaborée sans remettre en cause l'existant. Mettre à jour un logiciel en le rendant plus performant et en garantissant la non-régression du logiciel revient à en assurer la maintenance évolutive.

1.1.1. Les ratios pertinents

1.1.1.1. Contexte

Actuellement, dans les outils du projet HyperCarte, l'utilisateur peut choisir librement le numérateur et le dénominateur. Le ratio obtenu peut donc ne pas avoir de sens. Claude Grasland, membre du projet HyperCarte et professeur de géographie à Paris VII, a souvent remarqué ce genre de confusions notamment avec le taux de chômage. Ainsi, l'erreur commune est de penser que le taux de chômage est obtenu en divisant le nombre de chômeurs par la population totale³⁰.

Pour éviter ce genre d'erreurs d'analyse qui conduisent à des non-sens, il a été décidé d'ajouter la notion de « ratio pertinent ». Cette évolution impacte HyperAtlas mais également HyperAdmin qui aura la charge d'intégrer ces nouvelles informations à partir des fichiers de données. Nous nous intéressons dans cette partie uniquement à HyperAtlas.

1.1.1.2. Mise en place

Comme nous l'avons vu dans la première partie de ce chapitre, le panneau de configuration lié aux indicateurs est composé de trois éléments (cf. Figure 25, page 49) :

- Deux listes déroulantes pour choisir le numérateur et le dénominateur ;
- Un simple libellé pour indiquer le ratio (ratio = numérateur / dénominateur).

La mise en place des ratios pertinents sur HyperAtlas a d'abord été implémentée avec des MultiMap. Une Map est un groupe de paires contenant une clé et une valeur associée à cette clé. Une MultiMap en est une extension permettant de faire correspondre plusieurs valeurs à une même clé. Si cette collection n'existe pas dans la version actuelle de Java, on peut soit l'implémenter, soit la retrouver dans des bibliothèques spécialisées. S'il n'est pas compliqué de développer sa propre interface MultiMap (chaque clé pointe vers une collection de valeurs), nous avons opté pour une solution existante : celle implémentée dans la bibliothèque Commons Collections du projet Apache Commons. [Apache]

Ainsi, dans cette association, chaque numérateur était une clé à laquelle on associait un ensemble de dénominateurs. Avec cette implémentation, l'utilisateur choisissait donc un numérateur, puis un dénominateur parmi les indicateurs autorisés. L'inconvénient de cette méthode réside dans le fait que la construction du ratio n'est pas très aisée pour l'utilisateur. De plus, cette première implémentation n'indique pas le nom du ratio ainsi calculé. Si cette version permet de ne plus faire de non-sens dans une analyse, elle n'est pas convaincante du point de vue ergonomique. Elle n'a donc finalement pas été retenue.

Dans la seconde implémentation, le comportement des listes déroulantes pour le choix du numérateur et du dénominateur n'a pas été modifié. Les deux listes contiennent donc l'intégralité des indicateurs. Le changement intervient au niveau du libellé du ratio. Cette zone, jusque là statique, va devenir dynamique. La zone de texte indiquant

³⁰ Le taux de chômage est en fait obtenu en divisant le nombre de chômeurs par la population active.

« numérateur / dénominateur » a été remplacée par une liste déroulante contenant le nom des ratios prédéfinis plus le libellé « numérateur / dénominateur ». L'utilisateur bénéficie alors de plusieurs moyens pour mener à bien son analyse. Il peut choisir librement ses deux stocks, ou bien directement choisir un ratio prédéfini. La saisie d'un ratio pertinent sélectionne automatiquement le bon numérateur et le bon dénominateur. L'intérêt de cette implémentation est que le fonctionnement d'HyperAtlas reste inchangé pour le choix des indicateurs. La seule évolution réside dans la possibilité de sélectionner un ratio prédéfini.

Figure 43 : la nouvelle zone de paramétrage des indicateurs.

La Figure 43 présente cette évolution. A titre de comparaison, l'ancienne zone de paramétrage dédiée aux choix des indicateurs est visualisable sur la Figure 25.

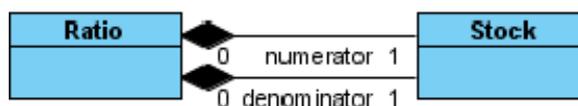


Figure 44 : Diagramme de classes : composition d'un ratio.

La Figure 44 montre la relation de composition entre un ratio (composé) et les indicateurs (composants).

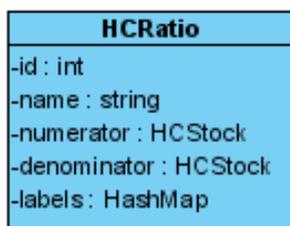


Figure 45 : La classe HCRatio.

Comme le montre la Figure 45, l'objet HCRatio est très simple. Il contient trois éléments importants :

- L'identifiant du numérateur ;
- L'identifiant du dénominateur ;
- Un libellé.

Le libellé principal est stocké dans l'attribut name, tandis que la HashMap labels contient une liste de labels gérée de la façon suivante (Code langue, label).

1.1.2. Discrétisation

Il a été décidé de revoir le calcul et l’affichage des déviations. La méthode de discrétisation a donc évolué. Au niveau de l’affichage, les palettes de couleurs ont été revues non seulement au niveau des couleurs mais aussi de l’ajustage de la classe centrale.

1.1.2.1. Les écarts relatifs

Les déviations seront désormais calculées en écart relatif. Ce passage a impacté essentiellement la classe métier d’HyperAtlas : la classe `Logic`. Auparavant, les déviations étaient calculées à partir de ratio.

$$(Valeur\ étudiée) / (Valeur\ de\ base)$$

Les résultats de ces calculs étaient alors toujours positifs. La valeur centrale est égale à 1.

La nouvelle formule pour calculer les écarts relatifs est la suivante :

$$(Valeur\ étudiée - Valeur\ de\ base) / (Valeur\ de\ base)$$

Les nouveaux résultats seront désormais centrés sur 0 et pourront donc être négatifs.

1.1.2.2. La nouvelle méthode de discrétisation

L’inconvénient majeur des anciennes méthodes de discrétisation est que la classe centrale est mal placée. En effet, la valeur moyenne (1 dans le cas du ratio, 0 dans le cas des écarts relatifs) doit être facilement visualisable. Ainsi, si le nombre de classe est pair, cette valeur moyenne doit être borne de classe. Dans le cas contraire, elle doit être au centre d’une classe.

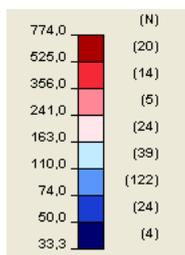


Figure 46 : Classe centrale mal positionnée sur l’ancienne légende.

Comme le montre la Figure 46, cette règle n’est pas respectée. La valeur moyenne (100, exprimée en %) n’est en effet pas une borne de classe. De plus, le changement de couleurs au niveau de la palette semble indiquer la frontière entre les valeurs supérieures à la moyenne et celles qui sont inférieures. Cependant, étant donné que la valeur moyenne n’est située que dans la sixième classe (sur 8), on peut conclure que cette légende est faussée. Après avoir implémenté la nouvelle méthode de discrétisation, il faut donc corriger cette erreur sémantique. De plus, il a été remarqué que les distributions étaient moins symétriques lorsque l’on passe à une échelle plus petite. En outre, les valeurs extrêmes peuvent créer des valeurs anormales.

C’est pourquoi, les géographes partenaires du projet ont proposé une nouvelle méthode de discrétisation. Ainsi, le nouvel algorithme permet d’éliminer les valeurs les plus extrêmes. Par défaut, sont éliminés 2.5% des valeurs les plus extrêmes de part et d’autre de la distribution. Puis, on détermine le minimum (`min2`) et le maximum de cette nouvelle distribution. A partir de ses valeurs et du nombre de classes (`k`), on calcule l’étendue notée `e`.

$$e = (\max2 - \min2) / k$$

Si le nombre de classes (k) est pair, 0 doit être borne de classe. On partage donc la distribution en 2 et on essaye de créer k/2 classes à gauche de 0 et k/2 classes à droite. Ce découpage permet d'obtenir un nombre égal de classes de part et d'autre de 0. Cependant, si la distribution n'est pas symétrique autour de 0, on risque d'avoir des classes vides (certainement en haut) et au contraire une dernière classe bien trop pleine de l'autre côté (sûrement en bas). Il faut donc travailler de proche en proche afin d'affiner les bornes des différentes classes.

On note NGj le nombre d'unités comprises entre 0 et Min2 non classées à l'étape j. Soit NDj le nombre d'unités comprises entre 0 et Max2 non classées à l'étape j et NCj le nombre d'unités classées dans la classe Cj à l'étape j.

L'algorithme suivant montre la construction des classes à gauche de 0.

```
Tant que NGj>0 et borne Cj<min2 faire
  Cj=]-(j-1)*e ; -j*e]
  Si NCj>0 alors
    NGj=NGj-NCj
  Fin si
Fin faire
```

Code 19 : Algorithme d'affectation des classes pour la nouvelle méthode de discrétisation.

On rajoute également dans la dernière classe la queue de la distribution c'est-à-dire l'ensemble des unités comprises entre min2 et min.

On reprend la même démarche pour construire les (k-j) classes restantes.

Si le nombre de classes (k) est impair, 0 doit être au centre d'une classe. Il faut donc commencer par construire la classe centrale]-e/2 ; e/2]. On peut ensuite reprendre la même logique.

Si cette nouvelle méthode de discrétisation basée sur les écarts relatifs est désormais activée par défaut dans HyperAtlas, il est toujours possible d'utiliser les anciennes méthodes. Pour cela, il suffit de renseigner le paramètre `ratio_deviation` dans le fichier de configuration de l'application. De même, le seuil d'élimination des valeurs extrêmes peut être ajusté en modifiant la valeur du paramètre `threshold_elim`.

A la fin de la partie intitulée « Discrétisation », deux copies d'écrans illustrent le changement de méthode.

1.1.2.3. Palettes de couleurs

Les palettes de couleur utilisées dans les cartes de déviation ont également été revues. Les nouvelles palettes sont issues du site de Cindy Brewer. Ainsi, on compte neuf palettes doubles pour représenter les déviations.

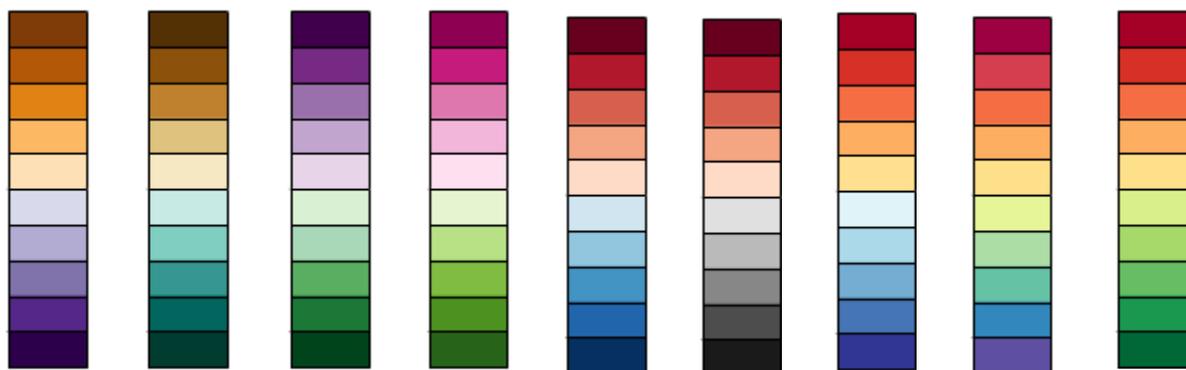


Figure 47 : Nouvelles palettes de couleurs pour les distributions divergentes.

Sachant que la classe centrale doit être la classe d'intensité la plus faible, quelques modifications ont été effectuées sur les palettes. La figure suivante présente l'une des nouvelles palettes intégrées dans HyperAtlas. Etant donné que la distribution n'est pas symétrique, on constate que la borne 0 n'est pas située au niveau de la classe centrale. On remarquera aussi, que cette borne montre bien le changement de gamme de couleurs.

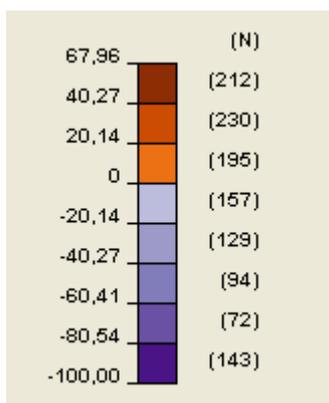


Figure 48 : Exemple de légende avec la nouvelle méthode de discrétisation dans le cas d'une distribution non symétrique.

Les copies d'écran ci-dessous illustrent le changement de la méthode de discrétisation mais aussi le changement des palettes de couleurs. Ces deux images représentent le même phénomène étudié avec les mêmes paramètres. Il s'agit du taux de chômage en 2000 et la valeur de référence est la moyenne européenne de cet indice (déviation globale).

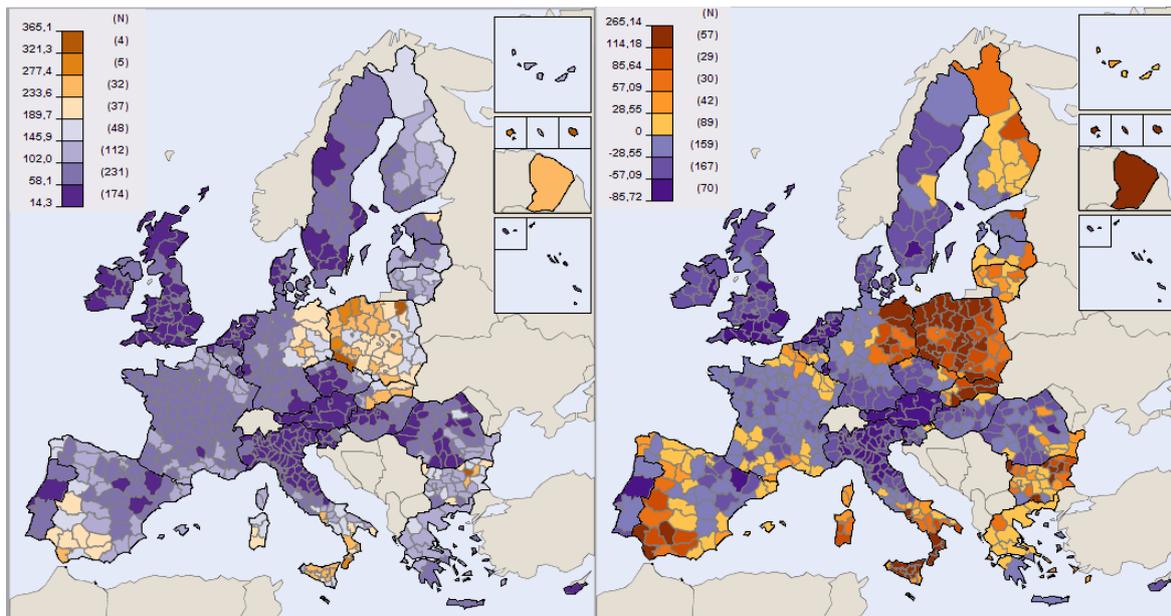


Figure 49 : Illustration du changement de la méthode de discrétisation.

L'ancienne méthode de discrétisation est représentée à gauche. On constate que la classe centrale (100) n'est pas borne de classe et que la répartition des classes est très inégale. On remarque en effet que deux classes sur huit comptent moins de cinq unités territoriales. Cette aire d'études comprenant 643 unités territoriales, chacune de ces deux classes n'est représentative que de 0.6% de l'aire d'étude.

La nouvelle méthode de discrétisation est quant à elle située à droite de la Figure 49. La classe centrale (0) est borne de classe et l'inversion de couleurs s'effectue à ce niveau. Ainsi, d'un coup d'œil, l'utilisateur est capable de déterminer quelles sont les unités ayant un taux de chômage inférieur à la moyenne européenne. La répartition entre les différentes classes est également plus homogène.

1.1.3. Les indicateurs inconnus

Jusqu'à présent, il fallait définir l'intégralité des indicateurs statistiques pour l'ensemble des unités territoriales du plus bas niveau. Ainsi, si une variable statistique était inconnue pour une région, il fallait quand même lui affecter une valeur. Cette façon de procéder pouvant conduire à des erreurs d'analyse, HyperAtlas a été modifié pour pouvoir traiter ces cas particuliers.

Ainsi, désormais, lorsque l'on visualise une carte avec le maillage élémentaire, les unités pour lesquelles le stock n'est pas exprimé sont coloriées en gris clair. Lorsque l'on passe à un maillage supérieur, l'unité supérieure n'est pas coloriée uniformément. La partie correspondant à l'unité pour laquelle le stock n'est pas renseigné reste en gris clair, alors que le reste de l'unité est correctement coloriée. Cela montre que le résultat ne prend pas en compte l'unité en question.

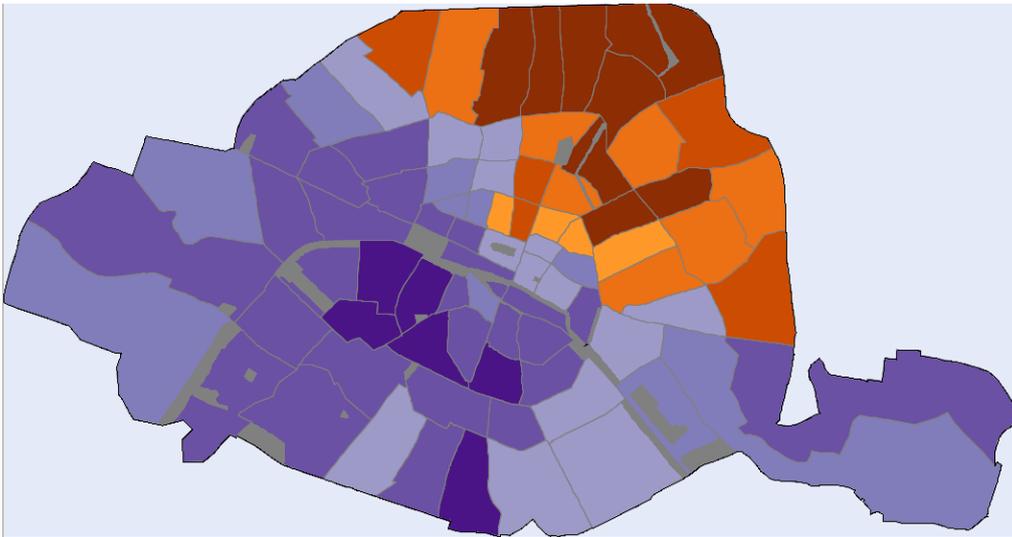


Figure 50 : Visualisation du taux de chômage à Paris.

La Figure 50 montre le taux de chômage à Paris. Cette visualisation de Paris s'effectue par quartier. Le maillage élémentaire est celui des IRIS³¹. L'INSEE³² définit les IRIS comme étant les briques de base en matière de diffusion de données locales. On en dénombre environ 50800 en France. Les IRIS correspondent soit à la commune dans le cas de petites communes non découpées soit à des quartiers (nommés IRIS 2000) pour toutes les communes urbaines d'au moins 10000 habitants [INSEE]. On distingue bien sur cette représentation, les zones pour lesquelles on ne possède pas d'information statistique. On reconnaît ainsi les berges parisiennes, les Tuileries ou encore le Champ de Mars.

³¹ IRIS : Ilots Regroupés pour l'Information Statistique

³² INSEE : Institut National de la Statistique et des Etudes Economiques

2. Réingénierie d'HyperAdmin

Cette version d'HyperAdmin bien que fonctionnelle présente quelques défaillances. Nous allons, dans un premier temps, décrire les problèmes de cette version actuelle, puis nous présentons les solutions apportées afin d'améliorer ce logiciel.

2.1. Problèmes rencontrés

Les défauts de cette version d'HyperAdmin se situent à plusieurs niveaux. On évoque ainsi les faits techniques (ou *bugs*) liés au fonctionnement de l'application mais aussi des problèmes liés à l'ergonomie du logiciel.

2.1.1. Faits techniques

Le plus important des faits techniques est la gestion des erreurs. Cette dernière n'est pas au point. Ainsi, par exemple, en cas d'erreurs lors de la création d'un nouveau jeu de données, l'application se ferme et aucun message d'erreur n'avertit l'utilisateur sur les origines de cet ennui.

Les erreurs ne sont pas toujours remontées au niveau de l'interface graphique. Même si elles sont parfois renseignées dans le journal d'événements de l'application, l'utilisateur non informaticien n'a pas toujours le réflexe de consulter ce journal et peut également ne pas comprendre le message d'erreur souvent trop général. Au début de mon travail sur HyperAdmin, j'utilisais le debugger intégré dans l'environnement de développement Eclipse pour trouver les causes de ces dysfonctionnements me permettant de corriger les jeux de données. Ce travail d'investigation m'a permis de bien prendre en main l'application et de maîtriser assez rapidement toutes les opérations permettant de créer un jeu de données. Les résultats de ce travail permettent de séparer les causes de ces erreurs en 3 catégories :

- Erreur dans l'algorithme. Dans ce cas, la solution apportée était la correction de ce bug ;
- Erreur due à des circonstances exceptionnelles et qui peut compromettre la poursuite de l'exécution du programme. Dans ces cas là, ce problème était résolu avec une gestion des exceptions ;
- Limitations du modèle de données. Dans ces cas là, la découverte de l'erreur peut provoquer soit une évolution d'HyperAtlas et d'HyperAdmin soit une correction avec une exception montrant que cette opération n'est pas autorisée.

2.1.2. Ergonomie

Parmi les erreurs liées à l'ergonomie et à l'interface graphique, on citera des erreurs comme l'affichage d'un message d'erreur pour avertir l'utilisateur que certains fichiers optionnels (notamment pour la gestion des contiguités et des voisinages) n'existent pas. Comme le montre la Figure 51, ce message s'avère mal choisi pour préciser à l'utilisateur que les fonctionnalités sur la déviation locale ne seront pas prises en charge.



Figure 51 : Mauvais message d'avertissement.

On trouve également des problèmes de rafraîchissement de l'interface graphique. Certains traitements sont longs et pendant l'exécution de ces tâches, l'interface graphique n'est plus mise à jour. Ainsi, si l'utilisateur change de fenêtre, lorsqu'il revient sur HyperAdmin, il ne visualise qu'un écran blanc tant que le traitement n'est pas terminé. De plus, sur chacun de ces traitements, l'utilisateur ne peut pas percevoir l'avancement de la tâche. Aucun indicateur, aucune barre de progression ne permet en effet à l'utilisateur de situer l'avancement. Cette remarque n'est cependant pas valable pour l'écran de création des voisinages.

On regrettera également le cheminement nécessaire à la réalisation d'un jeu de données. En effet, le processus de création (ou d'enrichissement) d'un jeu de données n'est pas trivial. Chaque opération s'effectue de manière indépendante : c'est donc à l'utilisateur de trouver le bon enchaînement pour créer son projet.

2.2. Solution proposée et mise en œuvre

Afin de répondre aux problèmes soulevés dans le précédent paragraphe, les solutions suivantes ont été mises en œuvre.

2.2.1. Le *Wizard*

Un *wizard* est en fait un assistant qui permet de guider l'utilisateur, étape par étape, tout au long d'un processus. Afin de simplifier la création d'un jeu de données sous HyperAdmin, le menu « Admin » a été totalement revu. Tous les précédents items ont disparu au profit d'un unique item : le *Wizard* (l'assistant).

L'intérêt de cet outil est qu'il permet de simplifier le travail de l'utilisateur, puisque les tâches les plus complexes sont désormais effectuées par cet outil. Le cheminement permettant la création et l'édition d'un jeu de données est donc désormais calculé automatiquement. Les traductions françaises pour le terme *wizard* existent (assistant, génie) mais sont peu utilisées. On utilisera donc le terme anglais.

La première étape de cette mise en œuvre a été de prévoir les différents scénarii possibles. La Figure 52 présente donc la succession des différentes étapes.

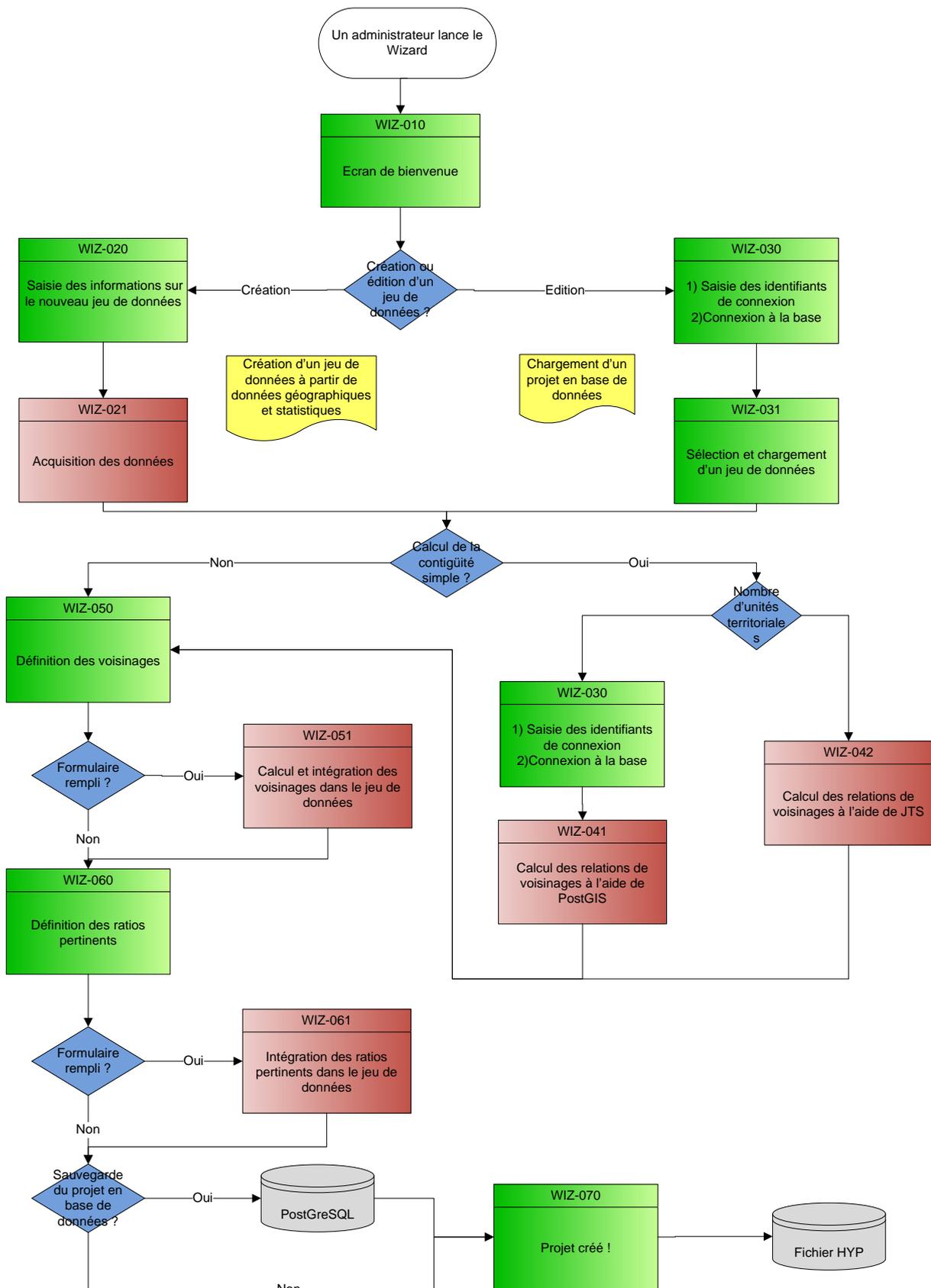


Figure 52 : Enchaînement des différents scénarii du Wizard.

Commençons par présenter les différents symboles de ce diagramme. Les rectangles verts symbolisent les écrans nécessitant une intervention de l'utilisateur : saisie d'informations, sélection d'un projet. Les rectangles rouges représentent les étapes de calculs ne demandant pas d'interaction avec l'administrateur. Enfin, les losanges bleus symbolisent un point de décision.

On s'aperçoit donc qu'il existe deux scénarii principaux. Le premier permet de créer un nouveau projet à partir des fichiers de données (scénario WIZ-020) tandis que le second permet d'enrichir un jeu de données préalablement sauvegardé en base de données (WIZ-030). Si l'administrateur choisit de calculer la contiguïté simple pour son projet, le système choisit la méthode la plus optimisée pour la bonne réalisation de cette tâche. Ainsi, en dessous de 1000 unités territoriales, le système va opter pour la bibliothèque JTS (WIZ-042). Dans le cas contraire, l'utilisateur devra d'abord se connecter à la base de données (WIZ-030) afin d'utiliser l'extension PostGIS (WIZ-041) pour calculer cette contiguïté. L'administrateur a ensuite la possibilité d'ajouter des relations de voisinage supplémentaires (WIZ-050), de définir des ratios dits « pertinents » (WIZ-060). Il peut ensuite enregistrer son projet en base de données en vue d'un éventuel futur enrichissement de données. Le dernier écran indique à l'utilisateur que son projet a été correctement initialisé et qu'il peut en générer le fichier « Hyp ».

Après avoir analysé les différentes étapes nécessaires, la mise en place du *Wizard* a pu commencer. En s'appuyant sur un exemple tiré du site de SUN [Sun] ainsi qu'en suivant quelques règles sur l'ergonomie et le design [Nogier, 2006], la maquette de l'assistant a pu être créée. La Figure 53 présente la maquette d'une fenêtre type du *Wizard*.

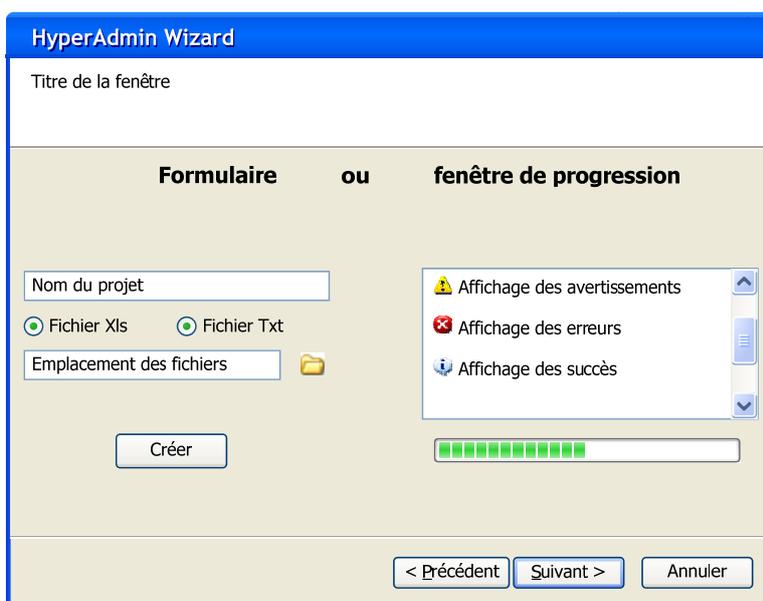


Figure 53 : Fenêtre type du *Wizard*.

Cette fenêtre type permet de distinguer deux utilisations possibles des différents écrans composant ce *Wizard*. Ainsi, la zone principale peut permettre l'affichage d'un formulaire permettant à l'utilisateur de saisir des informations pour la création d'un nouveau jeu de données, de sélectionner un projet sauvegardé en base de données, etc. Cette zone principale permet également d'afficher des écrans de progression renseignant ainsi l'utilisateur sur l'avancement de l'exécution de la tâche (barre de progression). Il peut également vérifier la bonne exécution de la tâche en consultant le journal d'évènements retraçant l'ensemble des messages (succès, erreurs, avertissements).

Chaque fenêtre du *Wizard* possède également deux composantes :

- le titre de la fenêtre ;
- les boutons de navigation qui permettent à l'utilisateur de revenir à l'écran précédent, de passer à l'étape suivante mais aussi d'annuler le processus.

La mise en place de cet assistant offre une certaine unité entre les différents écrans. En effet, auparavant, chaque écran étant indépendant, on ne retrouvait pas forcément le même aspect entre chaque fenêtre.

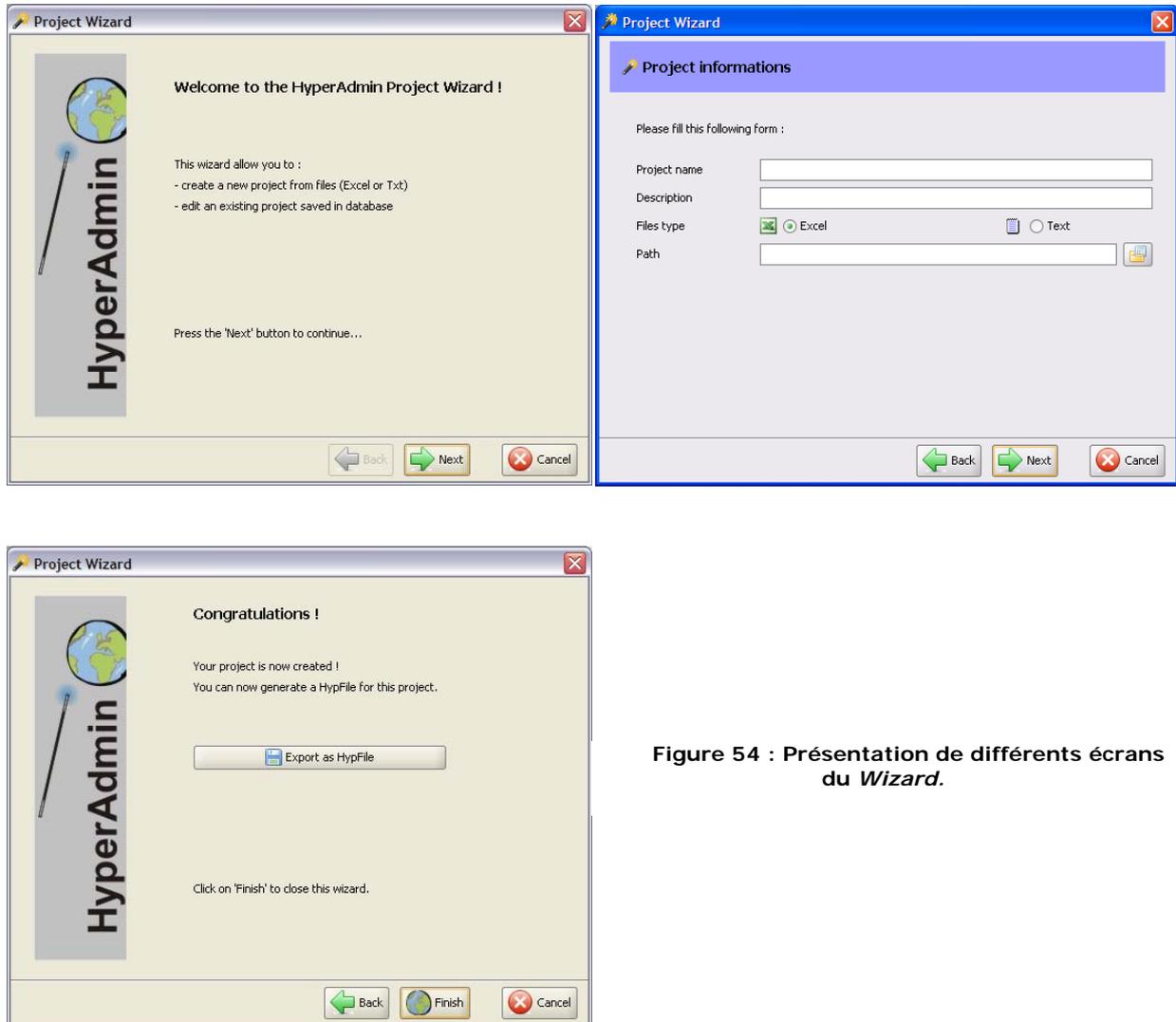


Figure 54 : Présentation de différents écrans du Wizard.

La figure ci-dessus, présente trois écrans de cet assistant. Le premier (en haut à gauche) est l'écran de bienvenue (WIZ-010) qui explique à l'utilisateur le rôle de ce Wizard. La copie d'écran suivante montre l'étape de saisie des informations du nouveau projet (WIZ-020) (cf. Figure 35). Enfin, la dernière capture permet d'annoncer à l'utilisateur que le jeu de données a été correctement créé et qu'il peut le sauvegarder (WIZ-070).

2.2.2. Gestion des erreurs

Les erreurs sont généralement mal acceptées par l'utilisateur car elles perturbent le déroulement de l'application et allongent le temps de réalisation. Minimiser le risque d'erreur améliore de façon notable l'utilisabilité d'un logiciel [Nogier, 2006]. Comme nous l'avons déjà dit, le traitement des erreurs nécessite d'une part, en amont, de prévenir les erreurs par une analyse des causes, et d'autre part, de faciliter la correction en assistant davantage l'utilisateur. On distingue deux catégories d'erreurs :

- Les erreurs d'intention ;
- Les erreurs d'exécution.

2.2.2.1. Les erreurs d'intention

Les erreurs d'intention résultent d'une mauvaise compréhension du système par l'utilisateur. Ce genre d'erreur peut être du notamment à un mauvais enchaînement des commandes. La mise en place du *Wizard* permet de diminuer considérablement ce genre d'erreurs car l'utilisateur est assisté tout au long du processus. Les étapes s'enchaînent dans un ordre correct selon les différentes actions de l'utilisateur. Il est encore possible de diminuer ce genre d'erreurs en fournissant une meilleure visibilité sur le fonctionnement du système. C'est pour cela, que les messages ont été revus. En effet, en affichant les résultats des différentes actions au fur et à mesure de l'avancement du processus, l'utilisateur comprend mieux le fonctionnement du logiciel et commet ainsi moins d'erreurs.

Les messages permettant au système de communiquer explicitement avec l'utilisateur sont soit pour le questionner afin de déterminer les prochaines actions, soit pour lui expliquer un fonctionnement. Il convient de mettre en place une typologie claire pour chaque type de messages (erreur, avertissement, information, question). Afin de garantir la cohérence de l'interface utilisateur, il faut respecter cette typologie. Ainsi, lorsque l'utilisateur reçoit un message dont la forme ne correspond pas au contenu, ce message peut semer le doute dans l'esprit de l'utilisateur [Nogier, 2006]. Pour éviter ces amalgames (cf. Figure 51), la gestion des messages a été revue lors de l'implémentation du *Wizard*.

La plupart des messages sont désormais affichés dans une zone dédiée contenant l'intégralité des messages obtenus pendant une tâche. Chaque message est accompagné d'une icône. Ainsi, la forme et la couleur de l'icône présente l'intérêt de l'utilisateur sur le type de donnée affichée avant qu'il ne la lise effectivement. Il s'agit donc d'un moyen de mise en évidence riche car l'icône apporte une connotation additionnelle au texte.

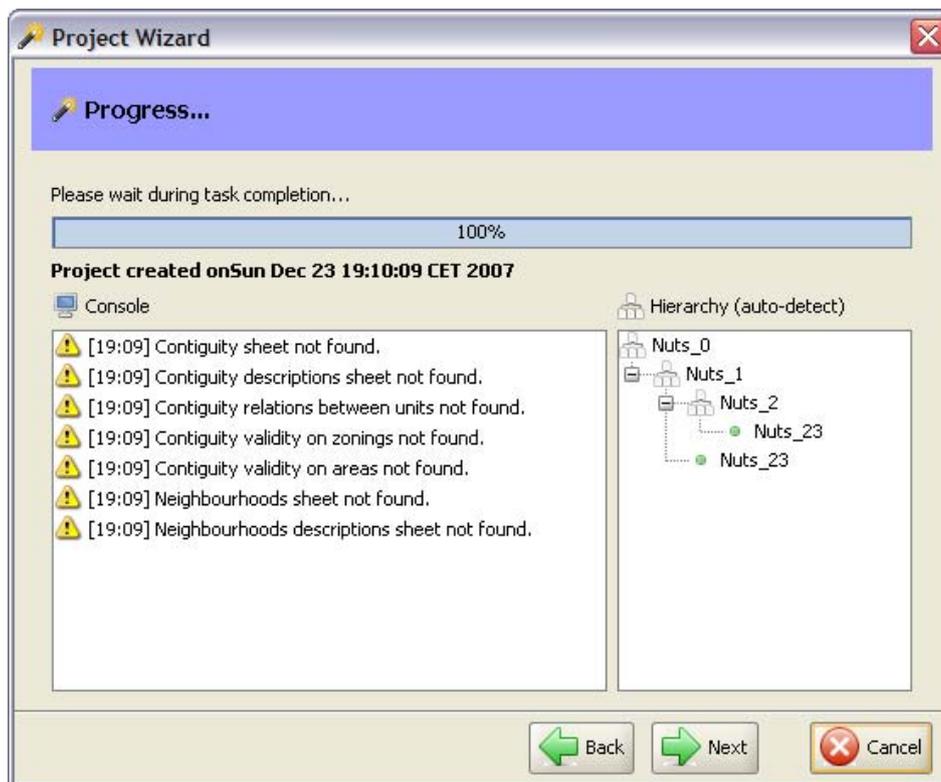


Figure 55 : Ecran de progression pendant la création d'un projet.

La Figure 55 présente une des étapes du *Wizard*. Cet écran correspond au scénario WIZ-021 du diagramme représenté sur la Figure 52. On constate notamment que les messages d'avertissement que nous avons vus avec la Figure 51 ne sont plus en contradiction avec le style de l'icône.

L'utilisateur doit également avoir une vision de l'avancement du processus. En effet, sur la version actuelle, la grande majorité des traitements s'effectue sans que l'utilisateur ait conscience de l'avancement de la tâche. En plus, de cette attente, nous avons énoncé dans le paragraphe 2.1.2 (cf. page 82), que lorsque de longs traitements étaient effectués, l'interface graphique n'était plus mise à jour. Cet inconvénient graphique était causé par le fait que toute l'application reposait sur un et un seul processus. Il faut savoir que sur les machines monoprocesseur, la simultanéité qui permet d'exécuter plusieurs programmes en même temps n'est en fait qu'une illusion. En effet, le système passe la main d'un programme à un autre à des intervalles de temps suffisamment courts pour donner l'impression de la simultanéité. Le langage Java permet d'appliquer cette possibilité de multiprogrammation au sein d'un même programme grâce à des *threads* (ou processus) indépendants [Delannoy, 2004].

Ainsi, en laissant l'ensemble des calculs être réalisé par le thread principal qui est déjà en charge du rafraîchissement de l'interface graphique, l'ensemble des opérations est effectué séquentiellement. Tant qu'une tâche n'est pas finie, les suivantes ne sont pas traitées mais mises dans une file d'attente. Par exemple, lors du traitement des fichiers de données, si l'utilisateur demande un rafraîchissement de la fenêtre, celui-ci ne peut être effectué qu'à la fin de l'opération en cours.

En mettant l'ensemble des calculs potentiellement longs dans un *thread* séparé, l'application gagne en interactivité car les tâches de rafraîchissement et de calculs sont effectuées à tour de rôle. Ce problème d'affichage étant désormais réglé, il a été également ajouté une barre de progression qui indique la tâche en cours et le pourcentage d'avancement. Grâce à ces deux évolutions (les *threads* séparés et la barre de progression), l'utilisateur perçoit donc que la tâche est en cours de réalisation et cela

lui évite de penser que l'application s'est brusquement interrompue (affichage non mis à jour) et qu'il doit y mettre fin.

La Figure 55 montre également cette barre de progression. On s'aperçoit qu'elle est à la fois graphique et textuelle : la barre se remplissant au fur et à mesure du déroulement de la tâche et le libellé (en gras) situé sous la barre permet d'explicitier la tâche en cours.

2.2.2.2. Les erreurs d'exécution

Les erreurs d'exécution ne sont pas causées par une mauvaise compréhension de l'application par l'utilisateur mais plutôt par des erreurs de traitement dans le code.

Certains traitements pouvaient en effet causer des erreurs dans certaines conditions. Ces erreurs ont donc été traitées par des exceptions. Parmi ces erreurs, on citera une erreur pouvant être due à des données incorrectes dans les fichiers Mif-mid. Il s'agit d'une erreur liée à la topologie qui peut se déclencher par exemple lors de l'agrégation de deux unités ou du calcul de la contiguïté. Dans la version précédente d'HyperAdmin, cette erreur n'était pas détectée et cela pouvait causer des erreurs d'analyse par la suite. Au début, cette erreur était gérée de façon classique, ainsi dès la première erreur de topologie l'exception `HATopologyException` était levée et le processus de création était alors stoppé. Le message indiquait alors les deux unités pour lesquelles cette exception avait été détectée. Cependant, ce fonctionnement n'était pas très efficace. En effet, comment savoir quelle unité sur les deux était mal définie. Après ce premier travail d'investigation et de correction du jeu de données par les géographes, on pouvait encore retomber sur une erreur du même type. Le processus de détection et de correction de l'anomalie pouvait donc être assez long. Le traitement de l'exception `HATopologyException` a donc été revu. Ainsi, la découverte d'une erreur de ce type ne provoque plus l'arrêt du programme mais seulement l'affichage d'un message. A la fin du processus de création d'un jeu de données, on s'aperçoit donc que ce nouveau jeu a été créé avec des erreurs et la liste des unités impactées, sont listés.

L'extrait suivant issu des traces d'HyperAdmin liste une partie des messages dus à une erreur de topologie. En recoupant ces informations, on s'aperçoit que l'unité qui pose problème est NIBR104.

```
Got an exception to compute simple contiguity between unit1 [NIBR101] and
unit2 [NIBR104]
Got an exception to compute simple contiguity between unit1 [NIBR102] and
unit2 [NIBR104]
Got an exception to compute simple contiguity between unit1 [NIBR104] and
unit2 [NIBR101]
Got an exception to compute simple contiguity between unit1 [NIBR104] and
unit2 [NIBR102]
Got an exception to compute simple contiguity between unit1 [NIBR104] and
unit2 [NIBR104]
Got an exception to compute simple contiguity between unit1 [NIBR104] and
unit2 [NIBR105]
Got an exception to compute simple contiguity between unit1 [NIBR104] and
unit2 [NIBR108]
Got an exception to compute simple contiguity between unit1 [NIBR104] and
unit2 [NIBR96]
Got an exception to compute simple contiguity between unit1 [NIBR105] and
unit2 [NIBR104]
Got an exception to compute simple contiguity between unit1 [NIBR108] and
unit2 [NIBR104]
```

Code 20 : Les erreurs de topologies.

En traitant de cette façon ce genre d'erreurs, il est nettement plus aisé pour les géographes de l'équipe de corriger le jeu de données. Comme le montre cet exemple, le travail d'investigation pour trouver quelle unité pose problème se limite désormais à un simple recouplement ce qui permet un gain de temps considérable.

2.2.3. Affichage

De plus, afin d'égayer l'interface graphique, des icônes ont été placées sur chaque bouton ce qui permet de rendre l'application moins austère. [Nogier, 2006]

L'image de chargement d'HyperAdmin a également été revue. En effet, dans l'optique d'obtenir une unité de style au sein des applications du projet HyperCarte, l'image a été totalement modifiée. Elle est désormais de la même taille que les différents logos d'HyperCarte (cf. page 96) et les positions des éléments (logo, texte) au sein des différentes images sont respectées.



Figure 56 : Image de chargement d'HyperAdmin.

La Figure 56 présente ce nouveau logo. Le globe terrestre en forme de puzzle permet de rappeler que la création d'un jeu de données (le globe) n'est que l'assemblage de différentes informations (les pièces).

3. Evolutions d'HyperAdmin

Ce paragraphe décrit comment les nouvelles fonctionnalités que nous avons décrites dans HyperAtlas ont été mises en œuvre dans HyperAdmin.

3.1. Les ratios pertinents

Les ratios pertinents permettent de spécifier et de donner à l'utilisateur des phénomènes statistiques prêts à être analysés. Nous avons vu dans le chapitre précédent, comment ils sont intégrés dans HyperAtlas, nous nous intéressons désormais à leur mise en place dans HyperAdmin.

Dans un premier temps, il a fallu étendre les fichiers de données et notamment celui contenant les informations sur les indicateurs statistiques. La modification apportée consiste en un onglet (ou fichier) qui est facultatif et dans lequel on structure les nouvelles données de la façon suivante :

NUM_ID	DEN_ID	EN	FR
pop_2003	sup_2003	Density	Densité
chom_2003	act_2003	Unemployment rate in 2003	Taux de chômage
gdp_eur_2003	pop_2003	PIB / Inhabitant (euro)	PIB / Habitant (euro)
gdp_pps_2003	pop_2003	PIB / inhabitant (PPS)	PIB / Habitant (PPA)
gdp_eur_2003	act_2003	Productivity (euro)	Productivité (euro)
gdp_pps_2003	act_2003	Productivity (PPS)	Productivité (PPA)

Code 21 : Extrait des fichiers de données d'HyperAdmin : RatioStock.

Un ratio pertinent est donc composé des identifiants du numérateur et du dénominateur et d'une série de labels.

Ces informations sont traitées (*parsées*) en même temps que les autres données (fond de carte, structure et données statistiques). Cependant, une étape du *Wizard* permet de créer des ratios pertinents à l'aide d'une interface graphique (scénario WIZ-060).

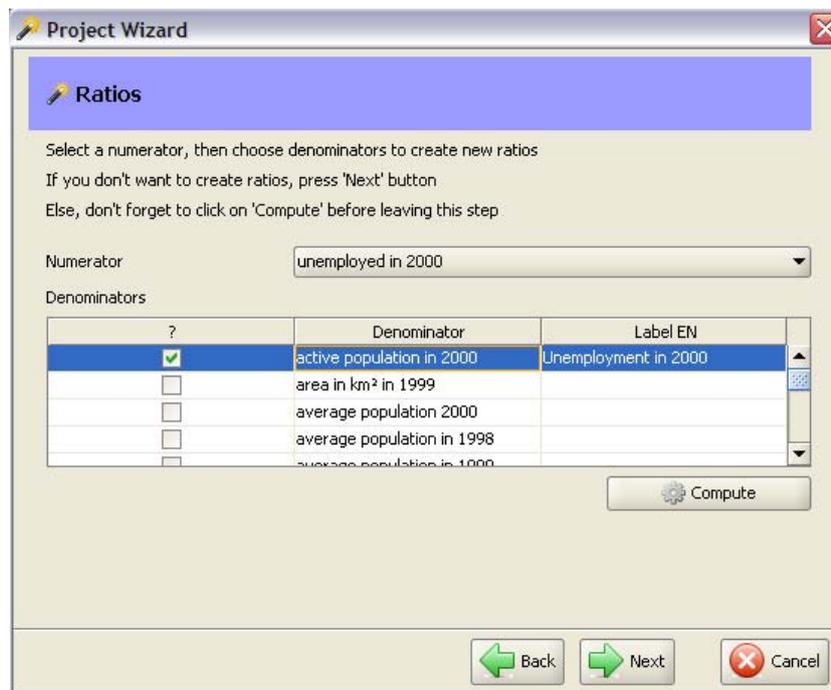


Figure 57 : Gestion des ratios depuis le *Wizard*.

Comme le montre la Figure 57, cette interface permet de gérer les ratios d'une manière assez intuitive. Pour cela, l'utilisateur choisit un numérateur puis lui affecte un ou plusieurs dénominateurs. Il doit également préciser le nom de chaque ratio ainsi créé. Les changements ne sont pris en compte qu'après la validation du formulaire (bouton « *Compute* »)

La gestion des ratios pertinents a également engendré une modification de la structure de la base de données. Ainsi, deux tables ont été ajoutées pour traiter et stocker ces nouvelles informations.

L'extrait suivant propose le script SQL permettant la mise à jour de la base de données :

```
-- *** RATIO ***
CREATE SEQUENCE ratio_id_seq INCREMENT BY 1 NO MAXVALUE NO MINVALUE
CACHE 1;

CREATE TABLE ratio (
    numerator_id text NOT NULL,
    denominator_id text NOT NULL,
    project_id integer NOT NULL,
    id integer DEFAULT nextval('ratio_id_seq'::regclass) NOT NULL
);

ALTER TABLE ONLY ratio ADD CONSTRAINT ratio_pkey PRIMARY KEY (id);

ALTER TABLE ONLY ratio ADD CONSTRAINT project_id FOREIGN KEY
(project_id) REFERENCES project(id) ON DELETE CASCADE;
```

```

-- *** RATIO_DESCRIPTION ***
CREATE TABLE ratio_description (
    ratio_id integer NOT NULL,
    language_id integer NOT NULL,
    name text NOT NULL
);

ALTER TABLE ONLY ratio_description ADD CONSTRAINT
ratio_description_pkey PRIMARY KEY (ratio_id, language_id);

ALTER TABLE ONLY ratio_description ADD CONSTRAINT
ratio_description_language_id_fkey FOREIGN KEY (language_id) REFERENCES
"language"(id) ON UPDATE CASCADE ON DELETE CASCADE;

ALTER TABLE ONLY ratio_description ADD CONSTRAINT
ratio_description_ratio_id_fkey FOREIGN KEY (ratio_id) REFERENCES
ratio(id) ON UPDATE CASCADE ON DELETE CASCADE;

```

Code 22 : Extrait du fichier CreateRatios.sql.

Suite à l'intégration des ratios pertinents, le diagramme de classes suivant présente les relations entre les différentes entités dans HyperAtlas. Avec ce schéma général, on s'aperçoit que les unités territoriales représentent la partie essentielle de la structure d'HyperAtlas (et par extension d'HyperAdmin).

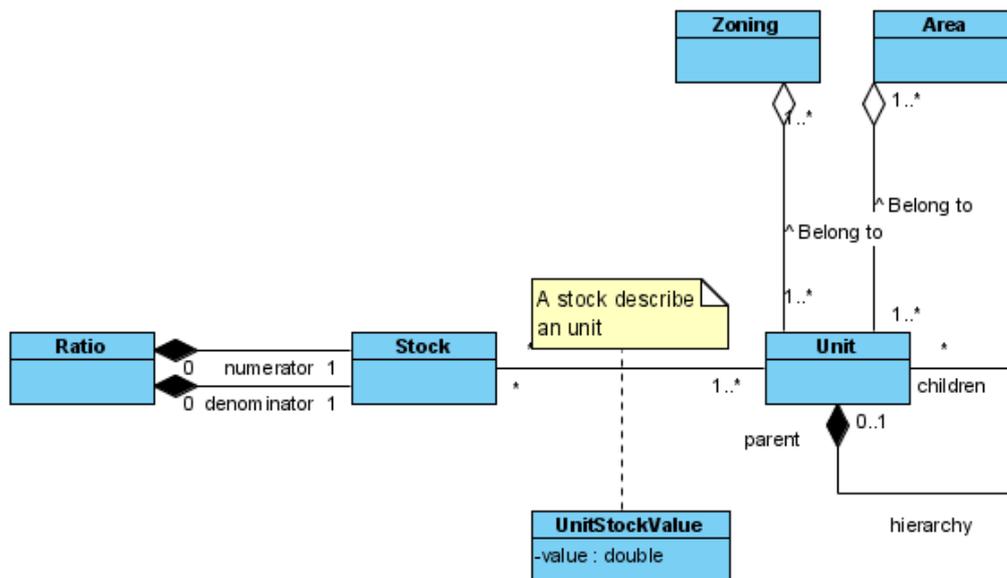


Figure 58 : Diagramme de classes : les différentes entités au sein du projet HyperCarte.

3.2. Les hiérarchies multiples

La gestion des hiérarchies multiples permet de gérer des maillages non compatibles. Ainsi, jusqu'à présent les différents maillages que l'on peut sélectionner dans HyperCarte sont rigoureusement emboîtés et forment une hiérarchie d'appartenances de type « arbre » (dessin de gauche). Cependant, les géographes de l'équipe ont souhaité qu'un même niveau élémentaire donne naissance à plusieurs hiérarchies de maillages non compatibles entre eux.

Ainsi, par exemple, les 36000 communes françaises peuvent engendrer plusieurs arbres hiérarchiques.

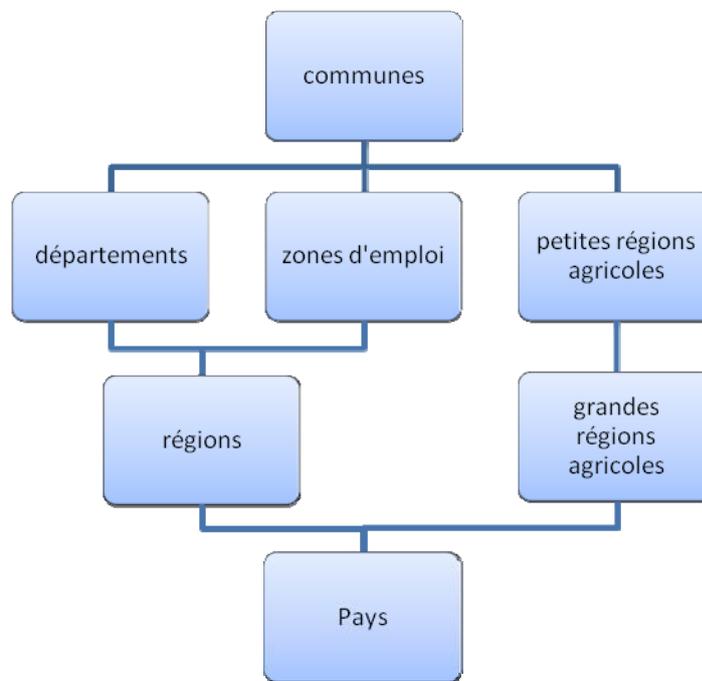


Figure 59 : Exemple de maillages non compatibles.

La Figure 59 présente un exemple de hiérarchies non compatibles. Cet arbre contient trois hiérarchies différentes. On trouve ainsi l'arbre « administration politique générale », un maillage consacré à des analyses sur l'emploi et enfin une hiérarchie dédiée à l'agriculture.

Pour gérer cette nouvelle fonctionnalité, aucun changement n'a été apporté dans les fichiers de structure. En effet, lors du traitement des fichiers de données, HyperAdmin détecte automatiquement la hiérarchie du nouveau projet. Pour cela, il parcourt la hiérarchie entre chaque unité définie dans le fichier *UnitSup*. Pour chaque unité, il vérifie dans quel maillage elle se trouve et construit petit à petit un arbre retraçant cette hiérarchie de maillages. Pour information, elle est affichée lors du traitement des fichiers de structure. En effet, la visualisation de cet arbre permet à l'utilisateur de se rendre compte de l'exactitude de ses données. La Figure 55 (cf. page 88) montre cet arbre.

Cependant, en cas de présence d'une hiérarchie multiple et non compatible, il faut recalculer les valeurs des stocks pour chaque unité. En effet, dans le cas général d'une hiérarchie simple, les valeurs de stocks sont obtenues en additionnant les valeurs de ces stocks au niveau inférieur. Dans le cas de hiérarchie multiple, il y a un risque de double compte (voire même davantage).

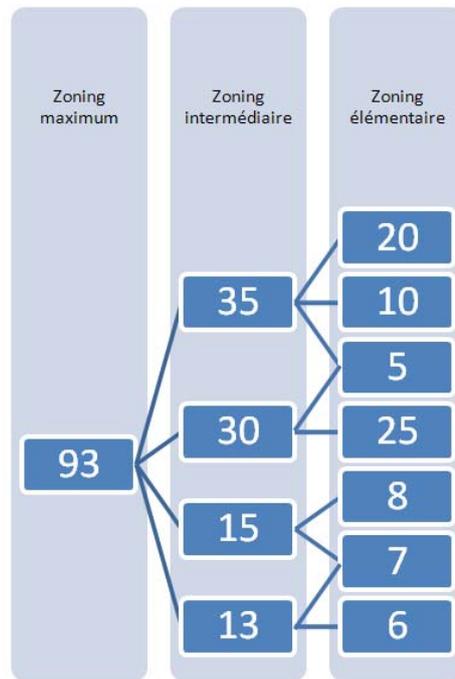


Figure 60 : Incidence sur le calcul des stocks sur une hiérarchie non compatible.

La Figure 60 présente une hiérarchie fictive où seules les valeurs des indicateurs sont représentées. On estime que les valeurs des unités sont renseignées dans les fichiers de structure. On s'aperçoit que le calcul est bien adapté pour le zoning intermédiaire. Cependant, le résultat est faux pour le niveau supérieur. En effet, si la somme des valeurs des stocks au niveau précédent est bien égale à 93, il faut noter que certaines unités sont comptées plusieurs fois. Il faut donc rechercher l'ensemble des feuilles (correspondant donc au zoning élémentaire) et d'additionner leurs différentes valeurs en prenant soin de ne pas compter plusieurs fois une même feuille. Il faut donc affecter 81 comme valeur de stock pour l'unité située au *zoning* maximum.

Hormis cet ajustement des stocks, une autre modification a été apportée afin de gérer convenablement les maillages non compatibles. Cette dernière intervient au niveau de la liste déroulante dédiée à la déviation moyenne. Etant donné que le maillage est dit incompatible, il est préférable de n'afficher dans cette liste que les maillages ayant un lien de descendance avec le zoning sélectionné.

Cette évolution n'a pas encore été incluse dans une livraison d'HyperAtlas. Elle a cependant été testée sur un jeu de données sur la région Ile-de-France.

4. Livraisons

Durant cette année de stage, le logiciel HyperAtlas a été livré à plusieurs reprises. Outre les versions destinées aux membres de l'équipe, on comptabilise quatre livraisons importantes. Ces livraisons comprennent les nouvelles fonctionnalités que nous venons de décrire.

Chaque livraison est composée du logiciel HyperAtlas avec un jeu de données spécifique. Selon la livraison, la version d'HyperAtlas est plus ou moins personnalisée en fonction du client. Ainsi, il est possible de paramétrer le nom de l'application, le logo de chargement, la licence mais également le panneau d'affichage de la carte.



Figure 61 : Les différents logos d'HyperAtlas.

Voici une liste des principales livraisons réalisées durant cette année. Ces livraisons sont classées de la plus ancienne à la plus récente et sont accompagnées d'un tableau récapitulatif sur le jeu de données.

Une livraison contient généralement le logiciel HyperAtlas ainsi que le jeu de données pré-chargé mais aussi une licence au format HTML ainsi que le manuel utilisateur d'HyperAtlas.

4.1. EEA-HyperAtlas

L'EEA-HyperAtlas, pour *European Environment Agency HyperAtlas*, est la version livrée à l'Agence Européenne de l'Environnement.

Tableau 3 : Caractéristiques de l'EEA-HyperAtlas.

Total					Taille du Mif-Mid (Ko)	
espace d'études	maillages	unités	stocks	ratios	.mif	.mid
1	5	1614	41	0	1111	12

Le Tableau 3 montre que cette première livraison ne bénéficie pas encore de la fonctionnalité des ratios pertinents.

4.2. EP-HyperAtlas

L'EP-HyperAtlas, pour *European Parliament HyperAtlas*, est la version conçue pour le Parlement Européen.

Cette version illustre une étude [Dubois et *al.*, 2007] qui a pour but d'évaluer l'ampleur des disparités régionales au sein de l'Union Européenne. Elle a été présentée devant le Parlement Européen à Bruxelles en juin 2007 lors d'une audition sur les disparités régionales et la cohésion.

Tableau 4 : Caractéristiques de l'EP-HyperAtlas.

Total					Taille du Mif-Mid (Ko)	
espace d'études	maillages	unités	stocks	ratios	.mif	.mid
3	4	899	6	6	735	6

Le Tableau 3 indique que cette livraison comporte les ratios pertinents. Il s'agit de la première version livrée contenant cette évolution.

4.3. Nordic HyperAtlas

Le Nordic HyperAtlas est la version d'HyperAtlas dédiée aux pays nordiques et livrée à NordRegio.

L'aire d'étude se limite aux pays nordiques (Norvège, Suède, Finlande et Danemark). Cet espace a beau être plus petit que l'Union Européenne, on constatera néanmoins que ce jeu de données est plus lourd que les autres. Cela est dû à la plus grande précision des tracés des frontières. Cependant si l'on gagne en précision, on perd néanmoins en interactivité. En effet, les temps de chargement et de rafraîchissement sont plus longs que sur les autres versions. Ces délais restent néanmoins acceptables.

Tableau 5 : Caractéristiques du Nordic HyperAtlas.

Total					Taille du Mif-Mid (Ko)	
espace d'études	maillages	unités	stocks	ratios	.mif	.mid
1	2	482	4	2	2398	5

Le tableau ci-dessus montre que ce jeu de données contient 3 fois moins d'unités territoriales que l'EEA HyperAtlas. Cependant, le fichier .Mif est deux fois supérieur à celui de l'Agence Européenne de l'Environnement. Cette différence montre que le jeu de données pour l'agence nordique est très précis. Les contours des pays sont en effet très détaillés comme le montre la Figure 62 qui permet de comparer la finesse des contours entre ces deux versions. L'image de gauche provient des jeux de données que l'on retrouve sur les versions de l'EEA ou du Parlement Européen tandis que l'image de droite provient du Nordic HyperAtlas. On perçoit que la première carte est bien plus **GENERALISEE** que la seconde. En effet, la plupart des polygones représentant ces données sont composés d'environ 100 points. Certaines régions sont même définies par plus de 360 points.

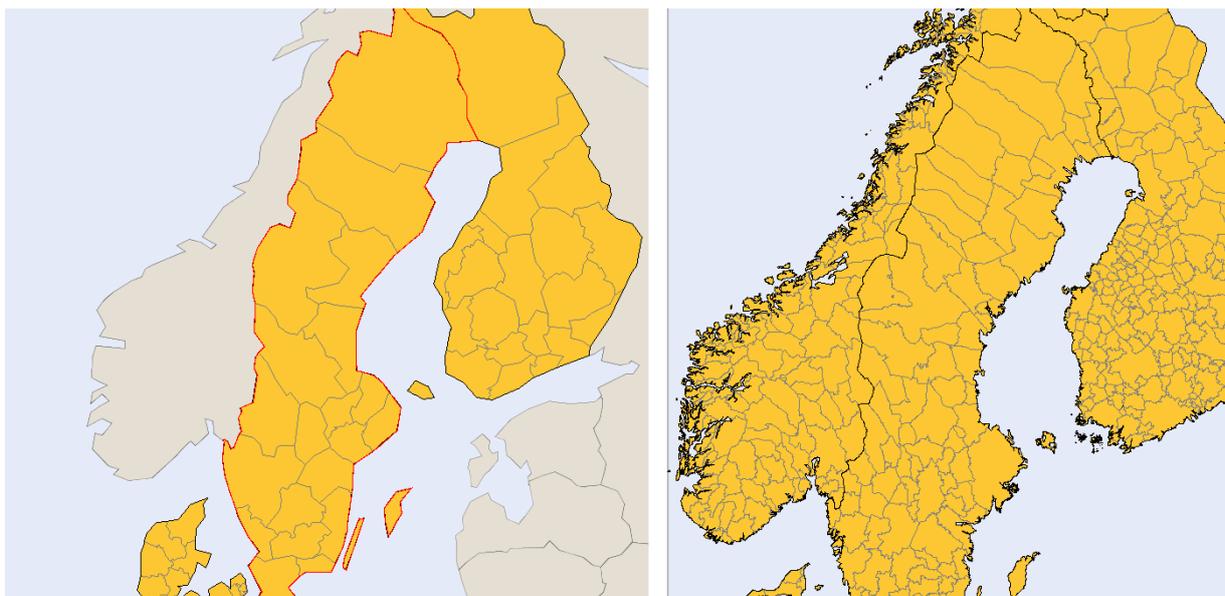


Figure 62 : Comparaison du niveau de détail des pays nordiques.

4.4. ESPON HyperAtlas

Il s'agit de la dernière version livrée durant cette année. Cette version est accompagnée de quatre fichiers « .hyp », les fichiers sérialisés contenant chacun un jeu de données. Un de ces jeux de données est packagé dans l'application et se lance donc à l'ouverture de l'ESPON HyperAtlas ; les autres pouvant être chargés avec le menu « Open Map ». Chacun de ces fichiers est consacré à un thème : démographie, statistiques sur l'emploi.

Cette version réalisée pour ESPON est la plus personnalisée. En effet, en plus de l'image de chargement, on constate que l'entête du rapport contient le logo de l'observatoire du territoire européen mais surtout un affichage particulier autour de la carte qui permet de rappeler l'origine de cette version.

Tableau 6 : Caractéristiques de l'ESPON HyperAtlas.

Total				
espace d'études	maillages	unités	stocks	ratios
5	3	n.c. ³³	27	15
5	3	n.c.	15	12
3	4	n.c.	6	7
5	5	n.c.	19	12

³³ N.c. : Non Communiqué

On constate que pour le Tableau 6 certaines données sont manquantes. En fait, l'ensemble des jeux de données pour cette version de l'ESPON HyperAtlas a été réalisé par les géographes de l'équipe. Si la version délivrée au Parlement Européen avait montré la première mise en place des ratios pertinents, cette version pour ESPON montre la prise en main par les géographes de l'outil d'administration des jeux de données : HyperAdmin et plus particulièrement de son assistant.

5. Conclusion

Comme nous l'avons vu dans ce chapitre, de nombreuses livraisons ont été réalisées au cours de cette année. La maintenance évolutive effectuée sur HyperAtlas a permis à ce logiciel de s'étoffer encore un peu plus avec de nouvelles fonctionnalités.

On peut encore imaginer d'autres développements autour de ce logiciel. Notamment, il faudrait travailler sur la gestion des indicateurs. Il serait en effet intéressant de pouvoir proposer des thèmes d'indicateurs. Cela permettrait donc de proposer sur une seule version d'HyperAtlas plusieurs domaines d'analyse. De plus, certains indicateurs sont présents dans l'application afin de pouvoir calculer des ratios. Seuls, ces indicateurs n'ont aucun sens et n'ont donc aucun intérêt à être visualisés. Il serait donc intéressant de pouvoir les cacher.

Pour HyperAdmin, les processus de création et d'enrichissement étaient déjà existants, mais le travail effectué pendant cette année a permis de rendre cet outil plus robuste, plus convivial mais aussi plus facile d'utilisation. La version livrée aux géographes début octobre leur a permis de créer de façon autonome les différents jeux de données pour le projet ESPON. Les préparations des différentes livraisons d'HyperAtlas m'a également permis de tester l'assistant d'HyperAdmin avant de le livrer. Chaque livraison d'HyperAtlas m'a donc permis de tester la nouvelle version d'HyperAdmin qui était encore en cours de développement pendant ces différentes périodes et de lister les défaillances afin de les corriger.

Les échos sur cette livraison sont plutôt bons voire encourageants. En effet, quelques fiches d'anomalies ont été postées sur le site du projet. Si cela montre que le logiciel connaît encore quelques imperfections, cela prouve également que la version livrée est utilisée par les membres du projet HyperCarte.

Cette application peut encore évoluer. Les évolutions d'HyperAtlas, citées plus haut, impliquent déjà des modifications dans HyperAdmin. Mais on peut également imaginer d'autres développements. Ainsi, si les géographes peuvent désormais créer plus facilement des jeux de données, ils sont encore tributaires de l'équipe STEAMER pour intégrer, *packager*, le jeu de données dans HyperAtlas. Cette opération nécessite en effet de recompiler HyperAtlas avec le jeu de données à intégrer. On pourrait alors imaginer qu'HyperAdmin devienne non plus seulement un atelier de gestion de jeux de données mais un générateur d'applications autonomes. Ainsi, le résultat ne serait plus un simple fichier « .hyp » mais un HyperAtlas *packagé* avec le jeu de données nouvellement créé.

Nous concluons en rappelant d'une part les objectifs fixés mais aussi les différents résultats obtenus. Nous étudions également quelles perspectives méritent d'être approfondies.

1.1. Objectifs

Le projet HyperCarte, groupe de recherche pluridisciplinaire spécialisé dans le développement de SIG dédiés à l'analyse statistique, a réalisé trois modules. Le premier **HyperAtlas** est le plus abouti. Plusieurs stagiaires ont déjà travaillé sur ce projet et le résultat est aujourd'hui satisfaisant. Certaines évolutions sont néanmoins souhaitées par les différents membres du groupe. Afin de combler certaines lacunes d'HyperAtlas, deux autres applications ont été créées. La première **HyperAdmin** est un module permettant de créer des jeux de données directement utilisables par HyperAtlas. Le second **HyperSmooth** doit apporter la notion d'analyse spatiale au projet HyperCarte. Pour cela, il s'appuie sur une méthode dite de potentiel que l'on peut apparenter (grossièrement) à du lissage. Ces deux applications sont encore en cours de développement. Une première version d'HyperAdmin a cependant été livrée, tandis qu'HyperSmooth n'est encore qu'un prototype.

Les objectifs de ce stage sont donc de poursuivre le développement des applications du projet HyperCarte, d'assurer la maintenance d'HyperAtlas mais aussi de fournir et de livrer des versions d'HyperAtlas.

1.2. Bilan des réalisations

Au cours de cette année, j'ai ajouté de nouvelles fonctionnalités à HyperAtlas. La gestion des ratios pertinents, le changement de la méthode de discrétisation mais aussi d'autres développements sont désormais intégrés dans HyperAtlas. Ce logiciel a donc évolué au cours de cette année tout en assurant une compatibilité avec les précédents jeux de données. Les ratios pertinents ne sont effet qu'optionnels, les projets qui n'en bénéficient pas sont donc toujours supportés par HyperAtlas. De même, il est possible de revenir très facilement sur l'ancienne méthode de discrétisation. Les livraisons d'HyperAtlas ont été nombreuses. HyperAtlas a en effet été livré à l'Agence Européenne de l'Environnement, au Parlement Européen, à ESPON et à NordRegio. Chacune de ses versions comporte les évolutions que nous venons de citer. La grande majorité des évolutions d'HyperAtlas a pu être testée par les membres du projet avec les différentes versions préliminaires à la version finale de l'EP HyperAtlas.

Concernant HyperAdmin, j'ai soigné particulièrement l'ergonomie de l'interface utilisateur permettant de créer des jeux de données. Le résultat de la refonte de cette interface donne une application plus claire et plus simple d'utilisation. De plus, si HyperAtlas est désormais doté de nouvelles fonctionnalités, il a fallu au préalable les penser puis les réaliser au niveau d'HyperAdmin. La gestion des ratios pertinents et des hiérarchies multiples a donc débuté sur ce module. La livraison début octobre 2007 de cette application semble satisfaisante. Si quelques anomalies ont été transmises, des

évolutions ont été demandées, cela montre que les utilisateurs utilisent l'application et ne sont donc plus (trop ?) dépendants des informaticiens du projet. L'assistant à la création de projets d'HyperAdmin a été développé en parallèle pendant les différentes livraisons d'HyperAtlas. Chacune de ces livraisons me permettait donc de faire le point sur l'état d'avancement du *Wizard* et de lister les points à améliorer. Le développement de cet assistant a donc suivi un cycle de vie itératif.

HyperSmooth n'a cependant pas aussi bien évolué que les autres modules. Il reste encore au stade de prototype. Si au début de mon stage, j'ai été affecté à son développement, les premiers temps ont été consacrés à la découverte de mon nouvel environnement : nouveau domaine métier, nouvelles technologies, formation aux applications du projet HyperCarte. Au départ de mon prédécesseur, j'ai ensuite basculé sur le développement d'HyperAdmin et la maintenance d'HyperAtlas. De par ce fait, je n'ai pas trop participé au développement d'HyperSmooth. Si j'ai ajouté la gestion des connexions à un « serveur HyperSmooth », la possibilité de sélectionner des zones à analyser (sélection d'une unité via le menu contextuel ou d'une aire d'étude avec un rectangle de sélection), cela reste des développements mineurs. De plus, l'image résultante n'est toujours pas parfaitement cadrée.

Il est cependant difficile de chiffrer véritablement mon travail. La maintenance évolutive sur HyperAtlas m'a fait impacter un nombre relativement important de classes tandis que rien que la conception de l'assistant a dû demander la création d'une trentaine de classes.

1.3. Perspectives

De nombreuses évolutions sont encore possibles pour l'ensemble de ces projets. La possibilité de classer les indicateurs mais aussi les ratios selon une thématique (démographie, environnement, politique) permettrait de gagner en lisibilité.

HyperAtlas est une application de type « client lourd » mais peut également être visualisée via Internet grâce à une applet. Cette technologie étant vieillissante, il serait peut-être judicieux d'utiliser le format SVG afin de fournir une application Web plus interactive et plus moderne.

Hormis les quelques retouches demandées suite à la dernière livraison d'HyperAdmin, on pourrait également envisager de nouveaux développements. Actuellement, on peut créer un nouveau projet à partir des fichiers de structure ou bien charger un jeu de données précédemment sauvegardé en base de données. Il serait intéressant de pouvoir enrichir directement un fichier sérialisé. Cela permettrait de rajouter à ce jeu de données les nouvelles fonctionnalités (ratio, voisinage) mais on peut également penser à corriger les valeurs des indicateurs. Initialement prévu dans le cahier des charges, la modification des valeurs de stocks n'a pas encore été réalisée.

1.4. Bilan personnel

Travailler durant une année dans un laboratoire de recherche aura été une expérience enrichissante.

Tout d'abord, sur le plan technique. Il faut savoir que j'ai une expérience assez réduite en développement logiciel. J'ai en effet commencé ma carrière professionnelle en tant qu'administrateur système et j'ai rapidement voulu devenir développeur. Débutant par des projets internes, j'ai participé dès mars 2006 à mes premières grosses réalisations de système informatique tout en continuant l'administration du parc

informatique de ma société. Durant, mon stage j'ai pu acquérir de nouvelles compétences notamment en conception d'interfaces graphiques avec Java Swing, mais aussi approfondir mes connaissances dans le langage Java. De plus, je me suis fortement intéressé à l'ergonomie logicielle. A mon retour dans mon entreprise, les compétences acquises durant mon stage m'ont permis de me consacrer pleinement au métier de développeur. De ce point de vue, j'estime que ma formation est bénéfique car elle m'a permis d'évoluer au sein de ma société.

J'ai également énormément apprécié de travailler dans un domaine métier différent. Le monde de la cartographie est en effet très intéressant. De même, le projet HyperAtlas m'a permis de travailler avec des personnes venant d'horizons différents : informaticiens, géographes mais également politiques. J'ai en effet eu la chance de participer à une réunion de travail avec l'Agence Européenne de l'Environnement.

Sur le plan humain, cette année aura également été très enrichissante. L'ambiance qui règne dans l'équipe Steamer est en effet excellente. Grâce aux origines variées des thésards de l'équipe (Brésil, Colombie et Roumanie), j'ai également pu découvrir de nouvelles cultures. J'ai gardé de nombreux contacts avec les membres de Steamer.

J'estime donc que cette formation et que cette expérience m'ont été bénéfiques sur tous les plans.

A

Analyse Spatiale

Etude formalisée de la configuration et des propriétés de l'espace produit et vécu par les sociétés humaines.

C

Carte

En géographie, une carte est une représentation d'un espace géographique. Elle met en valeur l'étendue de cet espace, sa localisation relative par rapport aux espaces voisins, ainsi que la localisation des éléments qu'il contient.

Cartographie

Science désignant l'étude et la réalisation de cartes.

Centroïde

Point fictif, situé à l'intérieur d'un polygone, dont les coordonnées correspondent approximativement à celles du centre de ce polygone

Coordonnées

En géographie, un système de coordonnées est un référentiel dans lequel on peut représenter des éléments dans l'espace. Ce système permet de se positionner sur l'ensemble du globe terrestre grâce à un couple de coordonnées géographiques.

D

Discrétisation

La discrétisation consiste en un découpage d'une série statistique en "classes" dans le but de la représenter en plages de couleurs sur une carte.

E

ESPON – *European Spatial Planning Observation Network*

Voir **ORATE**

ESRI – *Environmental Systems Research Institute*

Société informatique spécialisée dans la production de SIG. Fondé en 1969 par Jack Dangermond (Président), ESRI a été le premier éditeur de logiciel à concevoir des SIG. Aujourd'hui, ESRI propose la gamme la plus complète et la plus utilisée d'outils dédiés à la cartographie (le produit le plus populaire étant ArcGIS).

F

Framework

Un *framework* représente un cadre d'applications. Le terme francophone est cadriciel. Il s'agit d'un ensemble de bibliothèques fournissant des briques logicielles permettant le développement d'applications.

G

Généralisation

La généralisation est l'opération permettant de réduire la quantité d'information lors de la réduction de l'échelle. Parmi les techniques de généralisation, on trouve la simplification qui permet, par exemple, de réduire le nombre de points composant une courbe.

Géolocalisation

La géolocalisation est la localisation d'un objet sur une carte à l'aide de positions géographiques.

Géomatique

La géomatique regroupe l'ensemble des outils et méthodes permettant de représenter, d'analyser, et d'intégrer des données géographiques. La géomatique consiste donc en au moins trois activités distinctes : collecte, traitement et diffusion des données.

L

Latitude

La latitude est une valeur angulaire, expression du positionnement nord-sud d'un point sur la Terre, au nord ou au sud de l'équateur.

Longitude

La longitude est une valeur angulaire, expression du positionnement est-ouest d'un point sur la Terre à partir du méridien de Greenwich.

M

Multiscale

Qui peut se repérer à différentes échelles géographiques.

N

NUTS - Nomenclature des Unités Territoriales pour les Statistiques

Subdivision de l'espace économique européen permettant un recueil homogène des données statistiques afin de faciliter les comparaisons d'un stock entre pays, ou entre régions.

O

ORATE – Observatoire en Réseau de l'Aménagement du Territoire Européen

Il s'agit d'un programme de recherche sur le développement territorial européen.

⇒ <http://www.espon.eu/> (site officiel)

⇒ <http://www.ums-riate.fr> (point focal français)

Orthoimage, Orthophotographie

Une orthoimage est obtenue à partir de photographies satellitaires rectifiées géométriquement et radiométriquement. Une orthophotographie provient, quant à elle, de photographies aériennes. L'orthophotographie et l'orthoimage sont des mosaïques numériques alliant la précision géométrique de la carte à la richesse de la photographie. Elles peuvent être géoréférencées dans n'importe quel système de coordonnées. [@IGNFI]

P

Polygone

Un polygone (du grec *polus*, nombreux, et *gônia*, angle) est une figure géométrique formée d'une suite de segments. Chaque segment partage une extrémité avec le précédent et le suivant, délimitant ainsi un contour fermé. En cartographie, les unités territoriales sont représentées par des polygones.

S

SIG - Système d'Information Géographique

Outil informatique permettant le stockage, la gestion, l'analyse et la visualisation de données géographiques.

Bibliographie

Livres, Rapports, Mémoires, Articles

[Bissler, 2004]

Bissler T., *Conception et développement d'une plate-forme pour la génération de Systèmes d'Information Spatio-Temporelle et Multimédia dédiés aux Risques Naturels*, mémoire CNAM soutenu en juillet 2004.

[Chabert, 2007]

Chabert C., *HyperAtlas et HyperAdmin : des outils cartographiques pour l'analyse de phénomènes sociaux*, mémoire CNAM soutenu en mars 2007.

[Cuenot, 2005]

Cuenot O., *Modélisation spatiale multiscalaire de phénomènes sociaux : réalisation du logiciel HyperCarte*, mémoire CNAM soutenu en mars 2005.

[Dangermond, 2000]

Dangermond J., Article « A classification of software components commonly used in geographic information systems » publié dans *Introductory readings in Geographic Information Systems*, Taylor & Francis.

[Dubois et al., 2007]

Dubois A., Gensel J., Hanell T., Schürmann C., Lambert N., Zanin C., Ysebaert R., Grasland C., Damsgaard O., Lähteenmäki-Smith K., Gloersen E., Thomas R., *Les disparités régionales et la cohésion : quelles stratégies pour l'avenir ?* (Etude pour le Parlement Européen), Disponible : http://www.europarl.europa.eu/meetdocs/2004_2009/documents/dv/200/200705/20070530intraregionaldisparitiesfr.pdf

[Grasland et al., 2000]

Grasland C., Mathian H., Vincent J.M., « Multiscalar Analysis and Map Generalization of Discrete Social Phenomena : Statistical Problems and Political Consequences. » *Statistical journal of the European Community*, 2000.

[Grasland, 2003]

Grasland C., « Richesse et population dans le monde : une représentation multiscalaire des inégalités », *Mappemonde*, n°69, p20-25, 2003

[Grasland et al., 2003]

Grasland C., Lizzi L., Martin H., Mathian H., Vincent J.M., « *Hypercarte : un outil d'analyse spatiale multiscalaire des inégalités régionales en Europe* », 2003

[Nogier, 2006]

Nogier J.-F., *Ergonomie du logiciel et design Web, le manuel des interfaces utilisateur*, Dunod, 2006 (3^{ème} édition)

[Delannoy, 2004]

Delannoy C., *Programmer en Java*, Eyrolles, 2004 (3^{ème} édition)

[Plumejeaud, 2007]

Plumejeaud C., *Acquisition de données et cartes de potentiel pour l'analyse spatiale*, Mémoire CNAM soutenu en juillet 2007

[Plumejeaud et al., 2007]

Plumejeaud C., Vincent J.-M., Grasland C., Gensel J., Mathian H., Guelton S., *HyperSmooth : calcul et visualisation de cartes de potentiel interactives*, Sageo, Clermont-Ferrand, 2007

[Shin, 2003]

Shin S., *MVC Pattern(MVC Framework)*, 2003, PDF disponible sur <http://www.javapassion.org>

[Thomas, 2006]

Thomas R., *La persistance et le paradigme MVC dans les applications Web*, Probatoire CNAM, soutenu en juin 2006

Sites Internet

[@Apache]

Site internet du projet Apache Commons, Consulté en avril 2007, Disponible sur : <http://commons.apache.org>

[@CNES]

Site internet du Centre National d'Etudes Spatial (CNES), Consulté en août 2007, Disponible sur : <http://www.cnes.fr>

[@Espon]

Site internet d'ESPON, Consulté en novembre 2007, Disponible sur : <http://www.espon.eu>

[@Espon Luxembourg]

Site internet du point focal d'ESPON du Grand Duché du Luxembourg, Consulté en novembre 2007, Disponible sur : <http://www.espon.public.lu/fr/>

[@Google Earth]

Google Earth, Version 4.1, Consulté en juillet 2007, Disponible sur : <http://earth.google.fr>

[@HyperGeo]

Site internet HyperGeo : l'encyclopédie consacrée à l'épistémologie de la géographie, Consulté en août 2007, Disponible sur : <http://www.hypergeo.eu>

[@IGNFI]

Site internet de l'IGN France International, Article sur l'orthophotographie et l'orthoimage, Consulté en décembre 2007, disponible sur : <http://www.ignfi.fr/france/expertiseddonnees02.htm>

[@INSEE]

Site internet de l'INSEE, Rubrique « Nomenclature, Définitions, Méthodes », Article « IRIS », Consulté en décembre 2007, Disponible sur : http://www.insee.fr/fr/nom_def_met/definitions/html/ilots-regr-pour-inf-stat.htm

[@IRD]

Document pédagogique « *Qu'est ce qu'un Système d'Information Géographique ?* » disponible sur le Site Web de l'Institut de Recherche pour le Développement, Consulté en août 2007, PDF disponible sur <http://www.cartographie.ird.fr/publi/documents/sig1.pdf>

[@Jargonf]

Dictionnaire informatique francophone en ligne, Consulté en décembre 2007, Disponible sur : <http://jargonf.org/wiki/wrapper>

[@Jgoodies]

Site internet de Jgoodies, version 2.1.2 de Looks, consulté en mars 2007, disponible sur : <http://www.jgoodies.com>

[@Landstat]

Site internet du programme LANDSTAT, Consulté en août 2007, Disponible sur : <http://landsat.gsfc.nasa.gov>

[@Launch4j]

Site internet de Launch4J, version 2.1.5, consulté en mai et décembre 2007, disponible sur : <http://launch4j.sourceforge.net/>

[@Riate]

Site internet de l'UMS Riate, point focal français d'ESPO

[@Sun]

Site internet de Java SUN, Article : « *Creating Wizard dialogs with Java Swing* », Consulté en mai 2007, Disponible sur : <http://java.sun.com/developer/technicalArticles/GUI/swing/wizard/>

[@SwingX]

Site internet de SwingX, version 0.8.0, consulté en septembre et décembre 2007, disponible sur : <http://swingx.dev.java.net>

[@Waniez]

Site internet de P. Waniez, Article sur les différents types de cartes, consulté en décembre 2007, disponible sur :

<http://philgeo.club.fr/Carto/CarteThematiques.htm>

MEMOIRE D'INGENIEUR C.N.A.M. en INFORMATIQUE
Evolutions d'outils dédiés à l'analyse territoriale et à l'analyse spatiale dans le cadre du projet HyperCarte

Raphaël THOMAS.

Grenoble, le 20 juin 2008

Résumé

Les avancées dans le domaine du Web, de l'informatique et du multimédia ont permis de nouvelles perspectives dans de nombreux de domaines. Ainsi, dorénavant la cartographie est de plus en plus interactive. Les Systèmes d'Information Géographique permettent en effet d'accéder, de visualiser et d'analyser d'importantes quantités de données.

Dans ce contexte, le projet de recherche HyperCarte propose différents outils cartographiques dédiés à l'analyse de phénomènes statistiques. Chacune de ses applications a un rôle mais aussi un état d'avancement différent. Ainsi, HyperAtlas est le module d'analyse territoriale multiscalaire qui permet la visualisation de données statistiques selon plusieurs échelles : communales, départementales, régionales, nationales mais aussi continentales. HyperAdmin se positionne comme un outil permettant la création et la gestion des jeux de données directement utilisable par HyperAtlas. Le dernier né de ce projet de recherche est HyperSmooth. Son objectif est de représenter sur un maillage administratif une information statistique échantillonnée de manière continue. Plus particulièrement, mon travail a consisté à maintenir et à faire évoluer HyperAtlas afin de garantir les différentes livraisons réalisées durant cette année. La réingénierie d'HyperAdmin et notamment la création de l'assistant a permis de faciliter la création et la gestion de données exploitables par HyperAtlas.

Mots-clés : Système d'information géographique - Cartographie - Analyse territoriale multiscalaire - Analyse spatiale multiscalaire - Interface graphique utilisateur - Ergonomie logicielle - Java

Summary

Advances in the field of Web, information technology and multimedia have allowed new opportunities in many areas. Thus, cartography is now more and more interactive. The Geographic Information Systems makes it possible to access, view and analyze large amounts of data.

In this context, the research project HyperCarte offers different cartographic tools dedicated to the analysis of statistical phenomena. Each of these applications has a role but also a different state of progress. Thus, HyperAtlas is the multiscalar territorial analysis tool which allows the display of statistical data, according to several levels: municipal, departmental, regional, national, but also continental. HyperAdmin positions itself as a tool for creating and managing data files directly usable by HyperAtlas. The latest addition to this research project is HyperSmooth. Its aim is to represent a mesh administrative statistical information sampled on a continuous basis. More specifically, my job was to maintain and to evolve HyperAtlas software to ensure the deliveries made during this year. The reengineering of HyperAdmin including the wizard has facilitated the creation and management of HyperAtlas data.

Keywords : Geographical Information System - Cartography - Multiscalar territorial analysis - Multiscalar spatial analysis - Graphical user interface - Ergonomics software - Java