

La Géodatabase sous ArcGIS, des fondements conceptuels à l'implémentation logicielle

Françoise PIROT – Ingénieur de recherche CNRS – CNRS-CEIAS-UMR8564
Centre de Compétence Thématique CNRS « Modélisation, Analyse spatiale, SIG », UMR8564, Paris
pirot@msh-paris.fr

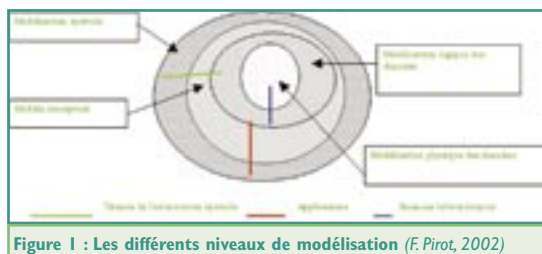
Thierry SAINT GERAND – Professeur – Université de CAEN – GEOSYSCOM – UMR6063 IDEES -CNRS
saint-gerand@mrsh.unicean.fr

Rendre intelligible la complexité spatiale

Parmi la gamme des systèmes d'information existants, les SIG sont probablement ceux qui s'attaquent à l'un des problèmes méthodologiques les plus ardues : comment appréhender le monde réel qui est à la fois complexe et multidimensionnel ? Parmi ces dimensions, trois au moins concourent ensemble à une certaine compréhension et à une

sont l'espace de la réalité observable (thématique, sémantique), l'espace « spatial » ou géographique, l'espace temporel. Il s'agit des espaces dans lesquels s'inscrivent les phénomènes

hension des phénomènes complexes exprimant une partie du monde réel étudié, d'autre part à réaliser un modèle conceptuel de données. L'initiateur de cette méthode



connue sous le nom HBDS (Hypergraph Based Data Structure) est F. Bouillé. La méthode hypergraphique permet

qui participent à la construction de tout ou d'une partie du monde réel en fonction d'une problématique, d'une thématique ou d'une application donnée.

d'élaborer des MCD sous forme de diagrammes de classes à partir desquels est construit le modèle logique puis physique.

Les étapes de la modélisation des données

1. La modélisation géographique du monde réel réalisée en fonction d'une problématique donnée, d'une thématique donnée, d'une application donnée à l'aide de la méthode HBDS. Cette modélisation doit être réalisée par le thématicien du domaine ;
2. Le modèle conceptuel de données ;
3. Le modèle logique ;
4. Le modèle physique.

Afin d'exprimer et communiquer la structure interne des phénomènes ainsi que des liens existants entre et à l'intérieur de ceux-ci, une modélisation préalable est à réaliser. Au sein des méthodes proposées pour traiter les phénomènes spatiaux, celles d'inspiration hypergraphique et ensembliste apparaissent aujourd'hui comme des plus efficaces. Elles permettent de représenter sous forme de graphes et d'hypergraphes, de classes et d'hyperclasses, de liens et d'hyperliens la structure interne des phénomènes simples. Les phénomènes simples liés entre eux vont contribuer d'une part à la compré-

certaine identification des phénomènes qui composent celui-ci. Ces dimensions ou ensembles ou espaces

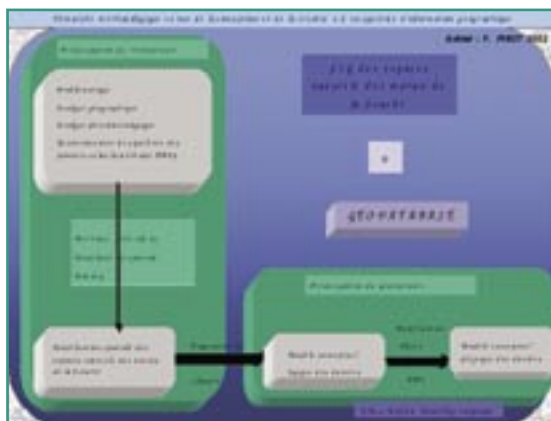


Figure 2 : Démarche méthodologique en vue de la conception et de la création d'un système d'information géographique (F. Pirot, 2002)

- Premier niveau de modélisation spatiale : Analyse de la thématique dont relève la problématique à résoudre ;
- Détermination de la structure des données spatiales et aspatiales (thématiques) selon la méthode H.B.D.S. ;
- Conception et rédaction du modèle conceptuel de données ;
- Création de la geodatabase ;
- Création des données spatiales et aspatiales.

Des concepts topologiques d'HBDS à leur implémentation logicielle : 25 ans

À la fin des années 1970, en parallèle avec le développement de la pensée « systémique », sont apparues de nouvelles approches, de nouveaux concepts et principes de formalisation concernant d'une part la structure des données, leur gestion, leur archivage, leur mise à jour, et d'autre part la structuration de l'information spatiale et aspatiale (thématique). C'est François Bouillé, géologue et informaticien, qui proposa dans le cadre de sa thèse d'État l'application et la mise en œuvre de la théorie des graphes et des hypergraphes, de la topologie, de la théorie des ensembles pour créer « un modèle universel de banques de données, portable et simultanément partageable » (titre de la thèse de Bouillé). Ces nouvelles approches sont apparues suite à un constat que les méthodes de type relationnel mises au point et utilisées dans le domaine de la gestion ne permettaient pas de prendre en compte d'une façon satisfaisante l'aspect géographique des « objets » c'est à dire, entre autres, la référence spatiale (longitude, latitude, type d'emprise, forme, voisinage...) comme composante intégrante de l'objet et non comme de simples attributs de l'objet.

Mais il aura fallu 25 ans pour que ces concepts, ces réflexions, ces méthodes trouvent leur implé-

mentation complète dans un logiciel, jusqu'à l'intégration de la topologie aussi bien dans la structure interne de l'information spatiale que dans la structure externe des données constituant des bases de données spatiales et les systèmes d'information géographique.

En effet, c'est déjà en 1977, que François Bouillé, géologue et informaticien, proposait la modélisation HBDS. Il introduisait les concepts de la théorie des graphes et des ensembles pour appréhender d'une part les structures internes de l'information spatiale et thématique, d'autre part pour manipuler, gérer, organiser, archiver les données dans les « banques de données ». Les concepts d'hypergraphe, de graphe furent alors utilisés dans la conception et la création des bases de données spatialisées qui vont être alors structurées topologiquement. À l'époque, F. Bouillé écrivait à propos du modèle mathématique de la structure des données : « Un ensemble de graphes et hypergraphes constitue l'armature. Cette armature est en quelque sorte un espace topologique supportant les données, sans le secours d'aucun espace métrique. »

Dans le second chapitre de sa thèse, Bouillé traite d'un point fondamental qui est la reconnaissance de la structure des phénomènes guidant la construction de la SD (Structure de Données). Il présente une construction peu traditionnelle et à l'encontre des idées communément admises jusqu'alors : « la SD est déduite des phénomènes non des problèmes » : il s'écarte ainsi résolument des démarches habituellement empiristes dans le domaine, pour se placer dans une optique beaucoup plus hypothético-déductive.

D'autre part, la structuration de l'information spatiale est pensée selon un modèle topologique c'est à dire qu'un objet géographique est assimilé à un graphe planaire particulier. Au début des années 1980, est conçu et créé le logiciel ArcInfo de type géorelationnel.

Les objets géographiques ont une structure interne possédant une topologie en terme de proximité, de contiguïté, de continuité. Ce sont des graphes planaires topologiques sans isthme auxquels sont associés des graphes duals. Durant la décennie 1980-1990, sont développées des méthodes algorithmiques prenant en compte le concept de topologie lors de la modélisation des objets, méthode qu'on va appeler « orientée objet ». De 1990 à 1995 surgissent plusieurs langages de modélisation qui vont donner naissance au langage de modélisation qui porte le nom de UML c'est-à-dire *Unified Modelling Language*. En 1999-2000, apparaît clairement la philosophie orientée objet dans une nouvelle gamme de produits ESRI à savoir ArcGis 8.0. Au fur et à mesure des versions successives se déploie le concept de geodatabase et son expression à travers le langage UML. En 2003, les principes de la topologie commandent la structuration interne de la geodatabase avec l'avènement d'ArcGis 8.3.

ArcGis/ArcEditor est le module de la gamme ArcGis qui est entièrement dédié à la création de la geodatabase topologique. Le module ArcEditor ne travaille qu'à partir d'un modèle conceptuel de données créé à partir d'une modélisation géographique. Une fois le modèle conceptuel rédigé selon la méthode hypergraphique HBDS, on peut créer la geodatabase qui va être une image physique du modèle conceptuel de données. En d'autres termes, la geodatabase sous sa forme concrète n'est autre que le modèle conceptuel de données (MCD) créé physiquement en « dur » mais vide d'objets. Par ailleurs, chaque case de la geodatabase c'est à dire du diagramme de classe correspond à une classe (ou sous-classe) du MCD. On peut faire le parallèle entre les concepts développés dans la méthode hypergraphique HBDS de F. Bouillé et les concepts développés dans ArcGis et plus particulièrement dans ArcGis/ArcEditor (ESRI). La correspondance est explicitée dans le tableau suivant :

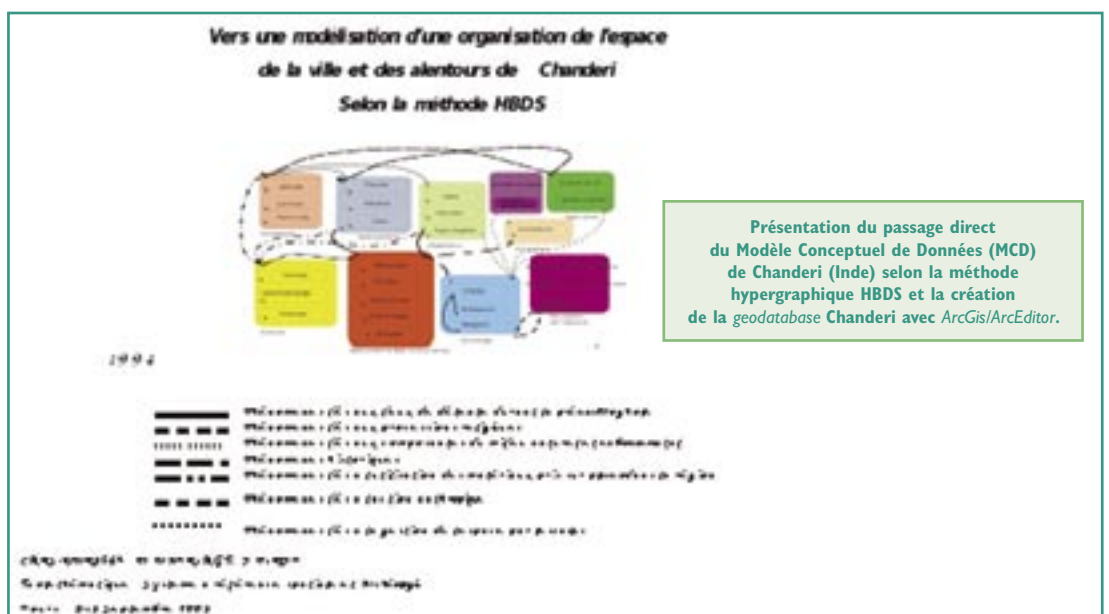
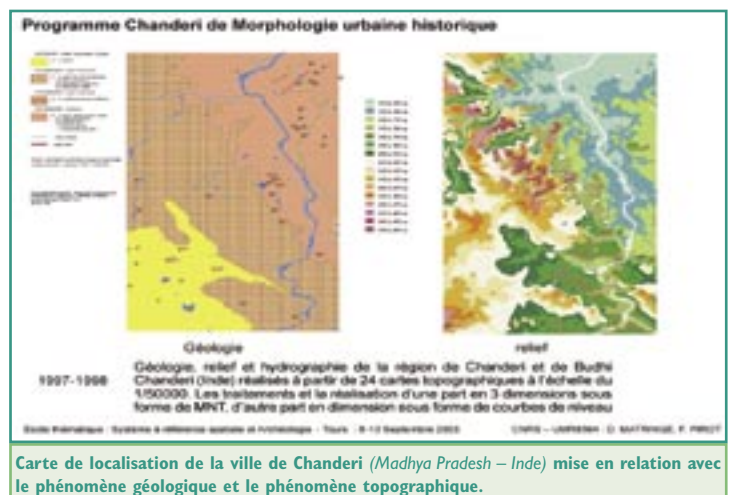
Terminologie ArcGis (ESRI)	Terminologie HBDS (F. Bouillé)	Remarques
Geodatabase	Hypergraphe, « forêt »	Définitions de la théories des graphes et des hypergraphes de C. Berge
Feature dataset	Hyperclasse, Jeu de classe d'entités (traduction française ESRI France)	La <i>feature dataset</i> doit avoir en théorie les mêmes référentiels spatiaux mais ce n'est pas obligatoire
Subtype	Hyperclasse	
Feature class	Classe	
Relationship class	Lien (HBDS), Relation topologique	
Objet	Objet HBDS	
Topology	Topologie	Entre et à l'intérieur de données
Domain	Domaine	Intervalle des valeurs prises par les attributs

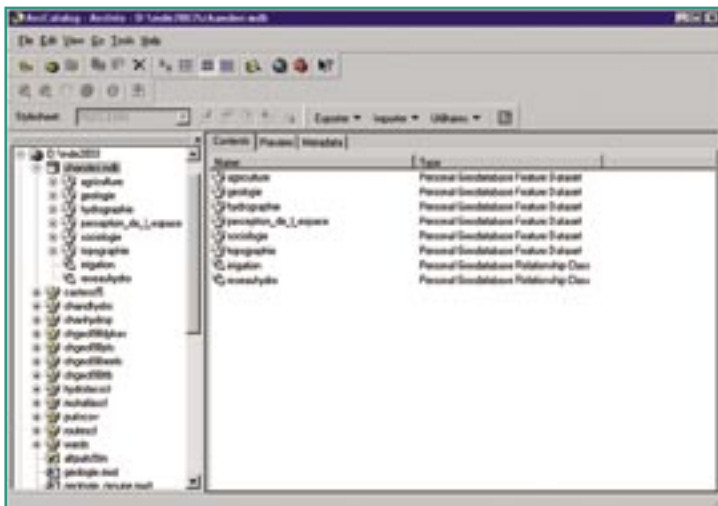
Étude de cas : Projet Chanderi (Inde)

Le projet « Chanderi », relevant du domaine de la sociologie urbaine historique, a été initié par le Professeur G. Fussman du Collège de France, spécialiste de la civilisation indienne. Chanderi est une ville moyenne de 20 000 habitants située au sud-ouest de Delhi, dans l'état du Madhya-Pradesh, en bordure du plateau du Malwa (cf. carte n° 1). Le projet s'appuyait sur un SIG pour étudier la ville du 11^{ème} au 20^{ème} siècle et comprendre globalement les interrelations entre les différents phénomènes historiques, sociologiques, culturels, spatiaux, naturels, économiques, etc... Un modèle conceptuel a été conçu

et créé selon la méthode HBDS en 1994. Dans un premier temps, les informations spatiales et/ou thématiques ont été créées intégralement (rien n'existant alors) à partir de sources très variées, hétérogènes, de qualité et de

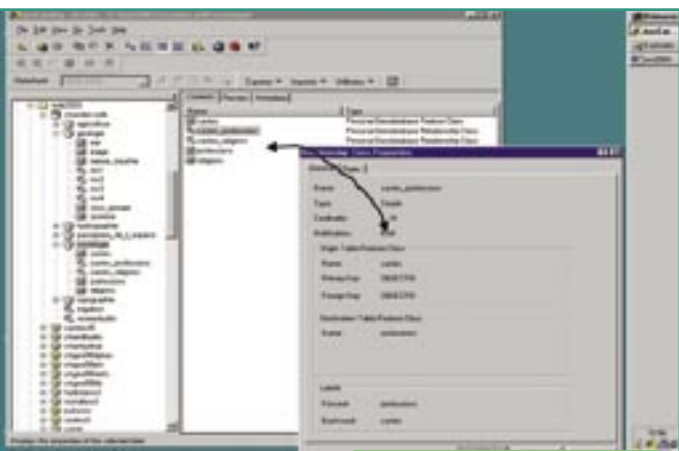
précision différentes comme les photographies aériennes, l'information satellitaire, les cartes topographiques, géologiques, les plans cadastraux, les enquêtes sociologiques, le suivi des parcours des processions religieuses, etc.



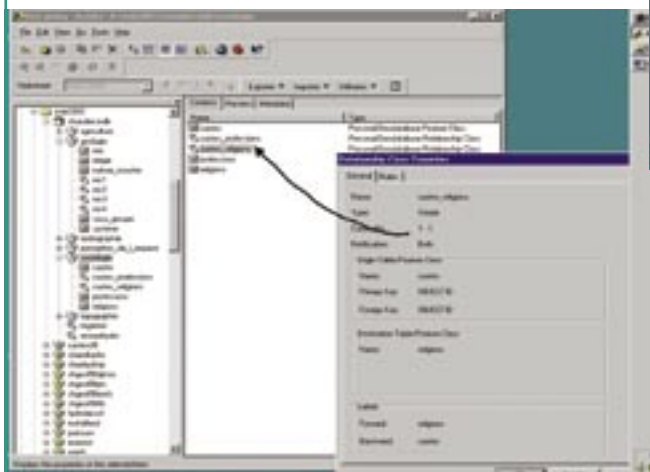


Modélisation de la réalité observable sous forme de Modèle Conceptuel de Données (MCD) selon la méthode hypergraphique HBDS pour le projet CHANDERI

Création de la geodatabase ou hypergraphe Chanderi à partir du modèle conceptuel précédent composé de classes, d'hyperclasses et de lien : création des hyperclasses (feature dataset) : agriculture, géologie, hydrographie, perception de l'espace, sociologie, topographie.
Création des liens entre les classes avec relationship class.



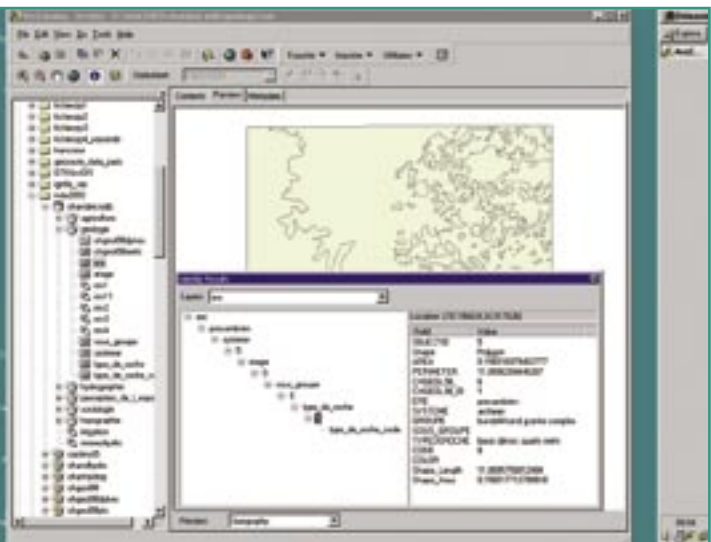
Classe (featureclass) professions



Création du lien castes_religions (relationship class) entre la classe castes et la classe religions

Interrogation de l'information spatiale et aspatiale contenue dans l'hyperclasse (featuredataset) géologie

classes (featureclass) ère, système, groupe, sous_groupe, type_de_roche de l'hyperclasse (featuredataset) géologie



Conclusion

Ces quelques développements ont permis de résumer l'historique de l'apparition de la modélisation hypergraphique, son développement logiciel, et de les illustrer via l'exemple thématique d'application fourni par le cas de la ville de Chandéri. L'avancée métho-

dologique majeure à en retenir est que, sous réserve d'un MCD suffisamment approfondi, la philosophie HBDS, telle que le concept de geodatabase l'introduit dans l'encapsulation objet permet de rendre informatiquement opérable sur la base d'une formalisation universelle, l'univers des requêtes de tous types (logiques, spatiales,

temporelles, fonctionnelles...) nécessaires à la compréhension globale (éléments constitutifs, structure et processus évolutif éventuellement) d'un phénomène géographique. Elle constitue ainsi le socle conceptuel le plus opérationnel à ce jour pour étayer l'usage des SIG comme langages d'interrogation de l'espace. ■

Références Bibliographiques

Berge C., 1970 : *Théorie des graphes et des hypergraphes*. Dunod.

Bouillé François, 1975 : *Structuration et saisie des données cartographiques*. Journée d'étude: *Acquisition et structuration de l'information graphique*, Comité Français de Cartographie (CFC), Paris.

Bouillé François, 1977 : *Un modèle universel de banque de données simultanément portable, répartie*, Thèse d'État ès sciences (spécialité : mathématiques, mention : informatique) Paris : Université Pierre et Marie Curie-Paris VI.

Bouillé François, 1978 : *Structuring cartographic data and spatial processes with the hypergraph-based data structure*. In Geoffrey Dutton (ed.) *First International Symposium on Topological Data Structures for GIS*. Cambridge, Massachusetts: *Laboratory for Computer Graphics and Spatial Analysis*, Harvard University.

Denègre J. et Salgé F., 1997 : *Système d'information géographique*. Que sais-je n°3122 – PUF

Gabay J., 2001 : *Merise et UML pour la modélisation des systèmes d'information*. Editeur : Dunod

Longley Paul A., Goodchild Michael F., Maguire David J., Rhind David W., 1999 : *Geographical Information Systems (2 volumes), Second Edition*. John Wiley and sons.

Longley Paul A., Goodchild Michael F., Maguire David J., Rhind David W., 2001 : *Geographic Information Systems and Science*. John Wiley and sons.

Longley Paul A., Brooks Sue M., McDonnell Rachael, MacMillan Bill, 1998 : *Geocomputation, A primer*. John Wiley and sons.

Molenaar Martien, 1998 : *An introduction to the theory of spatial object modelling for GIS*. Taylor and Francis.

Raper J. 2000 : *Multidimensional geographic information science*. Taylor and Francis.

Fotheringham S. and Rogerson P., 1994 : *Spatial Analysis and GIS*. Taylor and Francis.

Fussman G., Sharma K., 1996 : Chandéri 1993-1996. Collège de France, publications de l'Institut de civilisation indienne, édition-diffusion de Boccard.

Fussman G; Matringe D., E. Ollivier, Pirot F., 2003 : *Naissance et déclin d'une qasba: Chandéri du X^e au XVIII^e siècle*. Collège de France, publications de l'Institut de civilisation indienne, édition-diffusion de Boccard.

Keith C. Clarke, 1990 : *Analytical and computer cartography*. Hunter College - City University of New York. Prentice Hall, Englewood Cliff, New Jersey.

Haggett P., 1970 : *Analyse spatiale en géographie humaine*. Traduction de l'anglais. Collection U2-Dunod.

Matringe D., Pirot F., Saint Gérard Th., Fussman G., Ollivier E., 1996 : *Un système d'information géographique sous ArcInfo appliqué à la recherche en sciences humaines : le programme « Chandéri » de*

morphologie urbaine historique. Mémoire Vive, Bulletin de l'association française pour l'histoire et l'informatique, n°15-16.

Pirot F., Ernould J.C., Fournet F., 2002 : *Les disparités spatiales de santé dans la ville de Ouagadougou (Burkina Faso), Une modélisation d'une genèse d'espaces à risques sanitaires selon la méthode hypergraphique H.B.D.S.*, Colloque francophone ESRI 2002, Issy-les-Moulineaux, Octobre 2002.

Pirot F., Saint-gérard Th., 2003 *ArcInfo : un logiciel pour thématiciens. Bilan de 10 ans de formation par la recherche dans le secteur des Sciences de l'homme et de la société*, Colloque francophone ESRI 2003, Issy-les-Moulineaux, Octobre 2003.

Saint-Gérard T., 2002, « *SIG : Structures conceptuelles pour l'analyse spatiale* ». Thèse d'HDR, Univ. De Rouen.

Pirot F., Saint-gérard Th., 2003 : *ArcInfo : un logiciel pour thématiciens. Bilan de 10 ans de formation par la recherche dans le secteur des Sciences de l'homme et de la société*, Colloque francophone ESRI 2003, Issy-les-Moulineaux, Octobre 2003.

Zeiler Michael, 1999 : *Modelling our World, The ESRI Guide to Geodatabase Design*, ESRI Press.

Zeiler Michael, David Arctur, 2004 : *Designing Geodatabases, Case Studies in GIS Data Modeling*, ESRI Press.

MacDonald A. 1999 : *Building a geodatabase, Arcinfo8*, ESRI Press.