

SÉMINAIRE ESF SUR LES BASES DE DONNÉES GÉOGRAPHIQUES

à Aix-en-Provence du 11 au 15 mars 1992

par Benoît DAVID, François SALGÉ

◆ 1. Introduction

17 experts¹ provenant de 9 nations se sont réunis du 11 au 15 mars 1992 à Aix-en-Provence pour définir un agenda de recherche européen sur les Bases de Données Géographiques (BDG). Pendant le séminaire, une table ronde a été organisée pour rencontrer les utilisateurs et les chercheurs locaux. Le séminaire était organisé par F. Salgé à la demande de l'ESF. Deux autres séminaires ESF sont prévus, sur l'utilisation des données géographiques, l'un dans l'environnement et l'autre en socio-économie en mai et juin 1992.

◆ 2. Sujets principaux de la réunion

F. Salgé a résumé dans son article de fond les besoins d'un programme européen de recherche sur les SIG et a proposé cinq domaines de recherche :

■ Définition du contenu des données

Les modèles de données actuels ne permettent pas de modéliser correctement les informations géographiques. De nouvelles fonctionnalités seront essentielles dans le futur comme la gestion du vrai 3D, la gestion de données sans échelle et sans couture² et la gestion d'une localisation multiple.

Comme le contexte international suscite des échanges de données entre nations, un schéma conceptuel international et multilingue de données devrait être rapidement discuté et standardisé.

■ Recueil de données

Dans ce domaine, le défi consiste à extraire automatiquement de l'information à partir de photographies aériennes stéréoscopiques, de données de télédétection et de cartes scannées. L'utilisation de récepteurs GPS embarqués pourrait constituer une autre solution pour recueillir et mettre à jour des données géographiques.

Des procédures d'assurance qualité doivent être définies pour chaque étape des processus de saisie et un contrôle qualité final doit être effectué.

■ Manipulation et gestion des données

En raison de leurs caractéristiques, les BDG soulèvent des problèmes de gestion qui ne sont pas encore résolus. L'énorme volume de données fournit des sujets de recherche que les informaticiens doivent examiner. De plus, la prochaine génération de SGBD devra être capable de prendre en compte plus efficacement les évolutions du paysage.

■ Post-traitement des données

Les données provenant d'une BDG peuvent devoir être intégrées verticalement et horizontalement : des couches distinctes peuvent être combinées en respectant leur structure profonde. Des données externes peuvent aussi être fusionnées avec les données d'une BDG.

■ Exploitation des données

En général, une BDG est conçue pour répondre à des besoins à une échelle spécifique. Ici l'échelle ne se rapporte pas à un rapport graphique mais exprime le niveau d'abstraction auquel le monde est décrit. Il est nécessaire de changer ce niveau d'abstraction pour synthétiser l'information et produire une carte à une plus petite échelle. Actuellement cette opération, appelée généralisation, est réalisée manuellement en utilisant un ensemble de règles qui ne sont pas totalement objectives car elles ont été transmises oralement par des générations de cartographes et font partie de l'habileté du cartographe.

La visualisation sur une console graphique soulève des problèmes similaires car la possibilité de faire un zoom peut être comparée à la technique de généralisa-

NOTES

¹ A.M. Arnaud (Faculdade de Ciencias e Tecnologia - Portugal), H.-P. Bähr (Universität Karlsruhe - Allemagne), A. Bianchin (Centro di Servizio Interdipartimentale di Cartografia «NANI VALLE» - Italie), M. Brand (Ordnance Survey - Irlande du Nord), H. Bruggemann (LVA-NRW - Allemagne), I. Campari (CNR - Italie), B. David (Institut Géographique National - France), A. Frank (Technical University of Wien - Autriche), R.G. Henriques (CNIG - Portugal), M. Goodchild (NCGIA - USA), L. Koen (Kadaster - Pays-Bas), D. Mark (NCGIA - USA), I. Masser (University of Sheffield - Grande-Bretagne), S. Openshaw (University of Newcastle - Grande-Bretagne), F. Salgé (CERCO - France), H. Scholten (Rijksinstituut voor Volksgezondheid en Milieuhygiene - Pays-Bas), N. Smith (CERCO - Grande-Bretagne)

² la gestion de données sans couture correspond à une gestion dans laquelle les limites entre les différentes feuilles (ou tuiles) sont invisibles pour l'utilisateur ; on parle aussi de données continues.

tion. Malheureusement les progrès techniques en matériel graphique dépassent de loin les développements de la compréhension des mécanismes cognitifs et de perception au moyen desquels les formes spatiales sont identifiées et interprétées.

Les experts invités au séminaire pour discuter de ces questions ont été choisis afin de représenter trois groupes : les fournisseurs de données provenant principalement des agences cartographiques européennes, des utilisateurs de données et des spécialistes en méthodologie et analyse spatiales. Deux experts appartenant au NCGIA (USA) ont été aussi invités pour donner leur opinion sur l'organisation d'un agenda de recherche européen.

◆ 3. Présentations

Les questions soulevées dans les présentations des experts sont résumées et organisées en cinq groupes :

■ Concepts / Langages

Ce groupe de questions traite des langages formels et des concepts pour l'utilisation d'information géographique, principalement orientés vers l'analyse spatiale. Les problèmes ont été principalement soulevés par A.M. Arnaud, I. Campari, A. Frank, R.G. Henriques et S. Openshaw. Aujourd'hui, les concepts utilisés pour décrire les données cadastrales et topographiques sont bien définis mais ne s'appliquent pas de manière satisfaisante à l'analyse spatiale et aux besoins de l'environnement et de la socio-économie.

En Europe, il existe actuellement une diversité culturelle importante et la plupart des concepts de base utilisés dans les bases de données géographiques proviennent des Etats-Unis. Afin d'être capable d'échanger des connaissances ou des informations, soit entre différentes disciplines soit entre pays, il est nécessaire de définir des concepts communs à toutes les applications utilisant de l'information géographique.

■ Qualité

La question de la qualité a été soulevée par tous les participants mais plus particulièrement par H.-P. Bähr, M. Brand et L. Koen. H.-P. Bähr a soulevé le problème de la propagation des erreurs dans une base de données géographiques ; M. Brand et L. Koen discutèrent plutôt la standardisation de la description de la qualité afin que les utilisateurs puissent se rendre compte de l'adéquation d'un jeu de données à leur application.

La gestion de la qualité est un point sensible de l'intégration européenne ; en effet dans des jeux de données pan-européens la qualité sera différente pour chaque producteur et les utilisateurs devront gérer des jeux de données ayant une qualité hétérogène.

■ Acquisition et mise à jour

Le problème de l'acquisition et de la mise à jour est principalement le problème des producteurs de don-

nées. Dans quelques années, nous aurons une couverture topographique numérique complète de la plupart des pays européens et le problème principal sera de tenir à jour cette information (S. Openshaw, M. Brand, H. Bruggemann and H.-P. Bähr). L'information topographique et cadastrale change lentement mais d'autres informations telles les données environnementales varient beaucoup plus rapidement. H. Bruggemann a proposé d'étudier les techniques de comparaison entre le contenu de bases de données géographiques existantes et des images satellitaires ou aériennes.

Le problème de mise à jour sera particulièrement crucial en Europe à cause de l'utilisation grandissante de données à grande échelle. Il devrait être abordé à l'échelle européenne car il est nécessaire de disposer de suffisamment de personnes compétentes.

■ Système de Gestion de Bases de Données (SGBD)

Le problème de gestion de bases de données géographiques doit être abordé par des informaticiens, mais les spécifications doivent être précisées par des spécialistes SIG en collaboration avec les informaticiens. Cette collaboration peut être plus facile en Europe qu'aux Etats-Unis car il existe un groupe d'informaticiens connaissant déjà les problèmes de gestion de données géographiques et disponible pour une telle collaboration (A. Frank).

Les applications futures nécessiteront une interopérabilité entre les différents SIG et les SGBD géographiques, soit localement, soit entre systèmes informatiques situés à des endroits différents et reliés par des réseaux (M. Brand, N. Smith)

■ Généralisation

Pour H. Bruggemann et F. Salgé, la généralisation est une opération qui permet de changer le niveau d'abstraction de données et de synthétiser l'information. Si l'information est sous forme analogique la généralisation est perçue comme une opération cartographique : comment dessiner une carte à petite échelle en utilisant les informations issues d'une carte à plus grande échelle. Avec des données numériques, pour J.P. Cheylan le problème de la généralisation peut se décomposer en trois sous-problèmes :

- comment choisir le bon niveau d'abstraction pour un problème donné ?
- comment définir les nouveaux objets à partir des objets existants ? Parfois il s'agit d'une simple agrégation mais cela peut aussi mettre en oeuvre des opérations complexes ou nécessiter un apport d'information externe.
- comment agréger la géométrie ? C'est la partie cartographique du problème.

Le second point devrait être abordé dans le cadre d'une coopération européenne car le niveau d'abstraction fait partie des spécifications des bases de données et doit être défini et caractérisé afin de faciliter la gestion et l'échange de données à travers les frontières nationales.

◆ 4. Recommandations

Pour chacun des cinq groupes de questions, ce chapitre donne une action recommandée en termes concrets et suggère des projets de collaboration possibles :

■ Concepts / Langages

Les concepts existants utilisés pour décrire, gérer et manipuler des données géographiques tels que ceux discutés par le comité technique 287 du CEN ont été largement influencés par les besoins des applications topographiques et cadastrales et devraient être évalués dans différents domaines, tels que l'environnement ou la socio-économie. Deux extensions principales sont envisagées, premièrement la possibilité de décrire et d'analyser un phénomène qui varie continûment dans l'espace et dans le temps et pour lequel seules des mesures discrètes sont connues. La seconde extension concerne la possibilité de prendre en compte la qualité et le flou. Le problème d'interpolation dans de tels modèles devra être étudié attentivement.

Les besoins en formalisation géographique ne sont probablement pas les mêmes pour toutes les applications et devraient être classifiés en fonction de la manière dont une application utilise l'information géographique. Cette classification fournira des classes de problèmes qui permettront de formaliser des solutions génériques.

Dans un premier temps, une analyse critique et culturelle des résultats obtenus par le NCGIA sera fructueuse puis, dans un second temps, le travail pourrait être réalisé en collaboration avec le comité technique 287 du CEN.

■ Qualité

Afin de définir plus précisément la notion de qualité, il est nécessaire de comparer «l'adéquation à l'utilisation» de différents jeux de données existants entre pays et pour différentes applications. Ceci pourrait être réalisé par un audit de jeux de données existants à travers l'Europe, par exemple les données CORINE pourraient être de bons candidats pour un tel audit. Cet audit devrait aboutir à la définition d'un ensemble de paramètres qui caractériseraient la qualité pour une classe donnée d'applications et de méthodes pour mesurer et utiliser ces paramètres.

Le travail déjà réalisé par le NCGIA devrait être critiqué en tenant compte des spécificités culturelles européennes afin de fournir des lignes de conduite pour de futurs utilisateurs.

■ Acquisition et mise à jour

Un travail important de recherche a déjà été réalisé sur les techniques d'acquisition et sera encore amélioré dans un futur proche. Il n'y a pas de besoins pour un programme de recherche majeur dans ce domaine. En revanche, les recherches devraient se centrer sur les techniques de mise à jour de bases de données géographiques existantes et particulièrement sur la comparaison entre le contenu de bases de données géographiques avec des images numériques. C'est un domaine

favorable à une collaboration industrielle.

Le développement de nouveaux systèmes de gestion de bases de données géographiques sera réalisé par des informaticiens mais les spécifications de ces systèmes (incluant la gestion de la qualité, les représentations multiples, la gestion du temps et de l'espace et des versions successives) devrait être réalisée en collaboration avec des experts SIG. Les nouveaux systèmes devront être capable de gérer de l'hypertexte et du multimédia.

Une attention particulière devra être apportée à l'interopérabilité de ces systèmes entre eux et avec les systèmes existants. Un modèle d'interopérabilité facilitant un échange de données simple et automatique devrait être développé avec des spécifications similaires à celles du paragraphe précédent et devra être standardisé.

Les spécifications pourraient se centrer, dans un premier temps, sur la gestion d'une base de données routière européenne dans laquelle les routes seraient définies par leur axe et qui sera disponible sur toute la CEE d'ici quelques années.

Une collaboration étroite avec la DG XII de la CEE pour le troisième et le quatrième PCRD est nécessaire. Une collaboration avec les organisations internationales promouvant l'interopérabilité dans les technologies de l'information faciliterait le travail.

■ Généralisation

Afin de comprendre comment fonctionne la généralisation, il est nécessaire dans un premier temps d'étudier le niveau d'abstraction des bases de données géographiques et des cartes existantes pour différentes applications. Ceci peut être réalisé en auditant et en évaluant le niveau d'abstraction défini par les spécifications réelles de certains lots de données. Cette comparaison conduira à la définition de différents mécanismes de généralisation qui montreront comment les procédures de rédaction cartographique ont influencé la généralisation. Le principal but de cette étude est de caractériser un modèle géographique et cartographique européen de niveaux d'abstraction en le comparant aux modèles américains, chinois ou japonais et de prédéfinir quelques niveaux d'abstraction pour les pays européens. Pour ce travail, une comparaison avec les modèles américains peut être réalisée en utilisant les travaux du NCGIA qui ont débuté en utilisant des cartes européennes. Ce travail devra prendre en compte la linguistique afin de définir ce que nous pourrions appeler une «grammaire spatiale» ou une «grammaire de la généralisation».

■ Acronymes utilisés dans le texte

CEN	Comité Européen de Normalisation
PCRD	Programme Coordonné de Recherche et de Développement
NCGIA	National Center for Geographical Information Analysis (USA)