

Développement d'une application d'exploration de données géospatiales comme support à la fouille archéologique

INTRODUCTION

La fouille archéologique est le moyen d'acquisition des données archéologiques et le regroupement de ceux-ci en vue de l'interprétation. La donnée archéologique peut être un artefact (ie. un objet qui a été façonné par un humain), un écofact (ie. un reste naturel - non travaillé - d'un organisme végétal ou animal), un lot (ie. une entité volumétrique de terre arbitrairement délimités par les fouilleurs suivant sa couleur, sa consistance, sa granulométrie et formant l'ensemble d'un site archéologique). C'est avec ce dernier type de données que l'on va travailler. Or, ces données archéologiques comportent de multiples composantes spatiales que les archéologues tentent de mieux comprendre. Jusqu'à maintenant cette tentative de compréhension des comportements spatio-temporels des données archéologiques de terrain ne s'est faite qu'intuitivement. Le présent projet va donc être de trouver une nouvelle manière d'analyser, de structurer, d'intégrer à un modèle interprétatif des données générées par l'archéologie.

RÉSUMÉ

TRANSACTION VS ANALYSE

La transaction

Avant une géomatization du processus d'interprétation, la géomatization de la fouille archéologique va d'abord consister à une géomatization du processus d'acquisition (ie. la cueillette des données archéologiques). Ce processus consiste à développer :

- l'acquisition des données spatiales avec un GPS
- l'acquisition des données dans une base de données dite transactionnelle

Les outils traditionnels de gestion et d'exploitation des données spatiales sont du type transactionnel ou OLTP (On-Line Transaction Processing). Ce sont par exemple :

- un SGBD (Système de gestion de base de données)
- un SIG (Système d'information géographique)

Le transactionnel réfère à un mode d'exploitation de données tourné vers la saisie, le stockage, la mise à jour, la sécurité et l'intégrité des données.

L'analyse

La structure transactionnelle n'est pas faite pour une aide à la décision. Les systèmes d'analyses servant aux processus de décision des organisations nécessitent entre autres des données sommaires (agrégées ou résumées) sur l'ensemble de l'organisation et aussi des réponses rapides à l'aide de requêtes de type agrégatif avec une interface à l'utilisateur faciles à utiliser.

Les nouveaux outils d'exploitation des données sont du type décisionnel ou OLAP (On-Line Analytical Processing). Ces outils OLAP visent à assister l'utilisateur dans son analyse en lui facilitant l'exploration de ses données et en lui donnant la possibilité de le faire rapidement. Pour cela, OLAP exploite une dénormalisation maximale des données, sous la forme d'une pré-agrégation stockée dans une base de données multidimensionnelle. Ces données peuvent être groupées à différents niveaux de granularité (niveau de détail des données emmagasinées dans une base de données)

Les bases de données relationnelles destinées aux opérations transactionnelles sont normalisées de telle sorte à ce que la redondance des données soit à son minimum. A l'intérieur d'une base de données relationnelle destinée à l'analyse, la dénormalisation des données est utilisée afin de limiter le nombre de tables et le nombre de joints nécessaire entre les différentes tables dans le but d'améliorer la rapidité d'exécution des requêtes.

Cette approche multidimensionnelle structure un jeu de données détaillées de façon à agréger selon un ensemble de possibilités d'agrégation. On appelle « agrégation » un regroupement de données. Par exemple, « Pays » est une agrégation de « régions ». Les données non agrégées sont nommées « Mesure ». Il s'agit de la variable qui fait l'objet de l'analyse (dans l'exemple choisit, la mesure pourrait-être 22). Finalement, les possibilités d'agrégation sont nommées « dimension ».

LES ATTENTES DES ARCHÉOLOGUES

Bien qu'il soit de plus en plus fréquent d'utiliser des nouvelles technologies pour faciliter le processus de fouilles, à ce jour, aucun système n'a mis l'accent sur l'interprétation et l'analyse faciles et rapides des données archéologiques numériques en intégrant la géomatique avec les systèmes de découverte des connaissances (data mining et OLAP – On-Line Analytical Processing).

Les archéologues souhaitent rendre à l'interprétation sa place entière dans le processus de compréhension des fouilles. Pour se faire, il faut qu'ils puissent interroger la base de données transactionnelle générée par le processus d'acquisition, il faut qu'ils puissent visualiser les « lots de fouille » pour faciliter leur interprétation grâce à une représentation graphique.

Leurs attentes sur le prototype vont être une rapidité et une convivialité de l'outil, mais aussi une compatibilité cognitive entre le processus analytique des archéologues et le logiciel. Le principe d'une compatibilité cognitive est de permettre à l'utilisateur, en l'occurrence l'archéologue, d'avoir une réflexion poussée dans son interprétation et ne pas se trouver limité par l'obstacle temporel de l'informatique (longueur temporelle des requêtes par exemple). Cette « lenteur » va stopper leur démarche intellectuelle.

LE SYSTÈME SOLAP (Spatial On-Line Analytical Processing)

Principe

Le data mining ou la fouille de données est une « technique de recherche et d'analyse de données qui permet de dénicher des tendances ou des corrélations cachées parmi des masses de données, de détecter des informations ou encore de découvrir de nouvelles connaissances en s'appuyant sur des méthodes de traitement statistique » [OLF, 2001]. Le présent projet vise à intégrer l'approche des bases de données multidimensionnelles (tel que défini en data mining) avec les SIG et les logiciels OLAP afin de permettre aux archéologues d'interpréter et analyser plus facilement leurs données directement sur le chantier de fouille. Cette approche constitue le prototype SOLAP (Spatiale On-Line Analytical Processing).

Les SIG seuls ne présentent pas l'efficacité requise par les applications d'analyses (langages d'interrogation, interfaces complexes, temps de traitement longs). Il a été démontré que l'OLAP possède un potentiel réel pour supporter l'analyse spatio-temporelle, cependant, sans volet cartographique, il est impossible de visualiser la composante géométrique nécessaire pour répondre aux attentes des archéologues. La solution SOLAP va donc consister à combiner des technologies spatiales (SIG) et non spatiales (OLAP). (Figure 1)

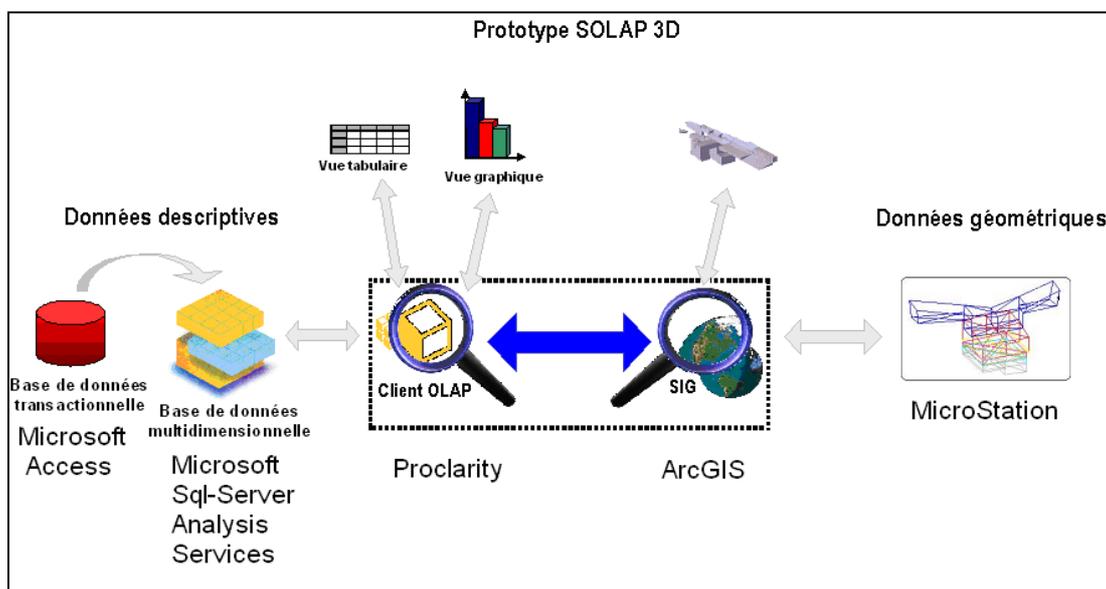


Figure 1: Présentation du prototype SOLAP 3D en fonction des différents logiciels

Résultats

CONCLUSION ET PERSPECTIVES
