
Une méthode de conception collaborative d'entrepôt de données

Application à la biodiversité

**Amir Sakka^{1,2}, Sandro Bimonte¹, Lucile Sautot³, Guy Camilleri²,
Pascale Zaraté², Aurélien Besnard⁴**

1. UR TSCF, IRSTEA

9 avenue Blaise Pascal, 63178 Aubière, France
prenom.nom@irstea.fr

2. IRIT, Université de Toulouse

prenom.nom@irit.fr

3. UMR TETIS, AgroParisTech

M. de la Télédétection, 500 rue Breton, 34000 Montpellier, France
lucile.sautot@agroparistech.fr

4. LPO Aquitaine

433 chemin de Leysotte, 33140 Villenave-d'Ornon, France
aurelien.besnard@lpo.fr

RÉSUMÉ. Dans le contexte de l'information géographique volontaire (VGI), les bénévoles ne participent pas aux processus décisionnels. De plus, les systèmes VGI n'offrent pas d'outils d'analyse avancés. C'est pourquoi, dans cet article, nous proposons d'utiliser les systèmes d'information décisionnels pour analyser les données VGI, et nous définissons une nouvelle méthode de conception de ces systèmes, qui permet d'impliquer les volontaires dans la définition des besoins analytiques. Nous le validons à l'aide d'une étude de cas sur la biodiversité.

Mots-clés : OLAP, Entrepôt de données, Information géographique volontaire, GDSS

Introduction

La VGI est "*la mobilisation d'outils pour créer, rassembler et diffuser des données géographiques fournies par des volontaires*" [18]. Elle permet de gérer des données géolocalisées (par exemple Openstreetmap¹), et est largement utilisée dans différents domaines d'application : urbanisme, suivi de la biodiversité, caractérisation risques, etc. Habituellement, les volontaires sont des producteurs de données et des consommateurs passifs d'analyses de données, fournies par des organismes/entreprises. Ce paradigme "collecte des données *bottom-up* et analyse des données *top-down*" représente un obstacle important pour le développement des observatoires volontaires, car les producteurs de données se sentent exclus du processus décisionnel [13]. De plus, comme nous l'avons souligné dans [2], la VGI ne présente pas de fonctionnalités permettant d'analyser d'importants volumes de données géospatiales.

Les systèmes VGI sont conçus pour les tâches opérationnelles et l'analyse complexe de petits volumes de données spatiales, tandis que les systèmes SOLAP (Spatial On-Line Analytical Processing) sont plus pertinents pour l'analyse basée sur l'exploration d'ensembles massifs de données spatiales stockées dans un entrepôt de données [17, 12]. Puisque les entrepôts de données sont conçus en fonction des sources de données et des besoins des utilisateurs, plus le modèle de l'entrepôt de données reflète les besoins des intervenants, plus ces derniers utiliseront leurs données [12, 15]. La mise à disposition d'applications décisionnelles adaptées aux besoins analytiques de la communauté VGI représentera un progrès important, puisque : (1) de nouvelles possibilités d'analyses efficaces sur des données volumineuses provenant d'observatoires différents seront possibles et (2) les volontaires seront de plus en plus motivés à collecter des données.

1. Méthode proposée

Nous proposons un nouveau système OLAP (OLAP 2.0). L'idée principale est de permettre aux volontaires d'exprimer séparément leurs besoins en matière d'analyse OLAP, dans un premier temps. Ces exigences seront ensuite traduites en modèles multidimensionnels (c.-à-d. en modèle d'entrepôt de données). Ces modèles multidimensionnels sont transformés automatiquement en prototypes grâce à la méthode ProtOLAP, afin de permettre aux volontaires une meilleure appropriation des concepts d'informatique décisionnelle. Ensuite, ces modèles multidimensionnels sont soumis à un ensemble de volontaires particuliers appelés *committers*, qui sont pleinement impliqués dans le projet et qui ont de l'expérience vis-à-vis des données collectées par les volontaires. Certains auteurs mettent l'accent sur la nécessité de l'intendance des données (menée par les *committers* dans notre approche) pour résoudre les problèmes liés au manque d'expérience des utilisateurs dans la spécification des requêtes et aux problèmes de propriété/sensibilité des données

¹ <https://www.openstreetmap.org>

rencontrés par les organisations au cours du processus d'implémentation des entrepôts de données. Ainsi, les *committers* décident de mettre en œuvre ou non des besoins exprimés (c'est-à-dire des modèles multidimensionnels) par les volontaires, en fonction de leur expertise pour juger de la pertinence des besoins. Ensuite, des informaticiens spécialisés en SI décisionnels sont chargés de la mise en œuvre des modèles choisis par les *committers*. Enfin, les nouveaux modèles multidimensionnels sont implémentés et mis à la disposition de tous les utilisateurs qui peuvent visualiser, explorer et analyser les données.

2. Expérimentation

Dans le cadre du projet VGI4Bio², nous mobilisons deux bases de données VGI (Visionature et l'Observatoire Agricole de la Biodiversité - OAB) pour construire des applications SOLAP d'analyse des indicateurs de biodiversité des terres agricoles. Visionature et OAB comptent respectivement 7682 et 1500 volontaires qui produisent des données. Parmi les utilisateurs potentiels intéressés par l'analyse de ces données, nous avons identifié un grand nombre d'utilisateurs appartenant à des catégories diverses telles que : les volontaires intéressés pour améliorer la qualité de leur production de données, leurs pratiques quotidiennes, etc., et des organismes publics et privés (DREAL, Chambre d'Agriculture, etc.).

Pour la validation expérimentale de notre proposition, nous avons engagé quatre volontaires avec des compétences différentes, et nous avons identifié quatre *committers*. Pour la validation de la première étape à l'aide de ProtOLAP, nous avons compté le nombre de rencontres entre volontaires et informaticiens et leur durée. Le temps d'implémentation d'un prototype d'entrepôt de données avec ProtOLAP est négligeable, puisqu'il ne prend que quelques minutes. En moyenne, il y a trois réunions par bénévole et chacune dure une heure. Par conséquent, nous pouvons conclure que ce n'est que lorsque le nombre de volontaires est faible que l'utilisation de la méthodologie ProtOLAP est possible. Lorsque le nombre de bénévoles devient important, une nouvelle méthodologie doit être fournie pour permettre aux bénévoles de définir eux-mêmes leurs modèles OLAP sans l'intervention des informaticiens.

Conclusion

Dans cet article, nous proposons une nouvelle méthode de conception collaborative qui permet d'impliquer les volontaires dans la définition des besoins d'analyse par rapport aux données VGI. Notre méthode permet aux volontaires non qualifiés en informatique de participer au processus de conception.

Nos travaux futurs porteront sur l'automatisation de la collecte des besoins auprès des volontaires, sans quoi notre méthode ne sera pas utilisable sur de larges communautés.

² www.vgi4bio.fr