



Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in:

<http://oatao.univ-toulouse.fr/22699>

Official URL :

http://agritrop.cirad.fr/589351/1/Actes_Sageo2018.pdf

To cite this version: Sakka, Amir and Bimonte, Sandro and Sautot, Lucile and Camilleri, Guy and Zaraté, Pascale and Besnard, Aurelien *Résolution collaborative des conflits des besoins d'analyse OLAP Spatial des données issues des observatoires citoyens*. (2018) In: Conférence internationale francophone Spatial Analysis and GEomatics (SAGEO 2018), 6 November 2018 - 9 November 2018 (Montpellier, France).

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

Résolution des conflits lors de la conception collaborative de cubes OLAP pour des observatoires citoyens

Amir Sakka², Sandro Bimonte¹, Lucile Sautot⁴, Guy Camilleri²,
Pascale Zaraté², Aurelien Besnard³

1. UR TSCF, IRSTEA de Clermont-Ferrand, 9 Avenue Blaise Pascal, 63170 Aubière, France, sandro.bimonte@irstea.fr

2. IRIT, Université Paul Sabatier, Toulouse, France

3. LPO Aquitaine, 433 Chemin de Leysotte, 33140 Villenave-d'Ornon, France

4. UMR TETIS, AgroParisTech, CIRAD, CNRS, IRSTEA 500 rue Breton, 34093 Montpellier Cedex 5, France, lucile.sautot@agroparistech.fr

RÉSUMÉ. DANS LE CONTEXTE DE L'INFORMATION GÉOGRAPHIQUE VOLONTAIRE (VGI), LES VOLONTAIRES NE SONT PAS IMPLIQUÉS DANS LES PROCESSUS DÉCISIONNELS. DE PLUS, LES SYSTÈMES VGI N'OFFRENT PAS D'OUTILS PUISSANTS POUR MENER DES ANALYSES TEMPORELLES. C'EST POURQUOI, DANS CET ARTICLE, NOUS PROPOSONS D'UTILISER LES SYSTÈMES D'INFORMATION DÉCISIONNELS POUR ANALYSER LES DONNÉES VGI, ET NOUS PROPOSONS LA DÉFINITION UNE NOUVELLE MÉTHODOLOGIE DE CONCEPTION DES ENTREPÔTS DE DONNÉES, QUI PERMET L'IMPLICATION DES VOLONTAIRES DANS LA DÉFINITIONS DES BESOINS ANALYTIQUES SUR LES DONNÉES VGI. NOS PROPOSITIONS ONT ÉTÉ TESTÉES SUR UN CAS D'ÉTUDE RÉEL CONCERNANT LA BIODIVERSITÉ.

MOTS-CLÉS : OLAP, Entrepôt de données, Information géographique volontaire (VGI), Système de prise de décision en groupe

ABSTRACT. IN THE CONTEXT OF VOLUNTEERED GEOGRAPHIC INFORMATION (VGI), VOLUNTEERS ARE NOT INVOLVED IN DECISION-MAKING PROCESSES. IN ADDITION, VGI SYSTEMS DO NOT OFFER POWERFUL TOOLS FOR TEMPORAL ANALYSES. THEREFORE, IN THIS ARTICLE WE PROPOSE TO USE BUSINESS INTELLIGENCE SYSTEMS TO ANALYZE VGI DATA, AND WE DEFINE A NEW DATA WAREHOUSE DESIGN METHODOLOGY, WHICH ALLOWS THE INVOLVEMENT OF VOLUNTEERS IN THE DEFINITIONS OF ANALYTICAL NEEDS ON VGI DATA. OUR PROPOSALS HAVE BEEN TESTED ON A REAL CASE STUDY ON BIODIVERSITY.

KEYWORDS: OLAP, Data Warehouse, Voluntary Geographic Information, Groupe Decision Making System

1. Introduction

Le VGI (Information géographique volontaire) est la mobilisation d'outils pour créer, assembler, et disséminer les données géographiques fournies par des

volontaires [8]. Classiquement, les volontaires sont des producteurs de données actifs, et des consommateurs passifs des analyses de données fournies par les organismes ou les entreprises concernées. Ce paradigme représente une barrière importante pour le développement d'observatoires basés sur des contributions volontaires, car les producteurs de données peuvent se sentir exclus du processus de prise de décision [5]. De plus, comme souligné dans [1], la VGI ne propose pas de fonctionnalités analytiques pour de grands volumes de données spatiales. En revanche, les systèmes OLAP Spatial (SOLAP) sont pertinents pour les analyses basées sur l'exploration de jeux de données géographiques massifs stockés dans un entrepôt de données spatiales [7, 4].

Plus le modèle de l'entrepôt de données reflète les besoins des décideurs, plus ces décideurs utilisent les données [4,6]. Cet article a donc pour objectif de "transformer" les volontaires producteurs de données en analystes de données. Plusieurs méthodologies de conception d'entrepôts de données ont été proposées dans la littérature [2,6]. Néanmoins, quand les décideurs sont des volontaires, ils peuvent avoir des parcours et des fonctions variés (scientifiques, citoyens, etc.), ce qui peut conduire à de multiples interprétations contradictoires des mêmes besoins. Quand les décideurs ont des buts différents, il devient difficile de maintenir un consensus entre eux, comme le montrent les travaux en ingénierie des besoins [3,11]. Ainsi, les besoins définis par ces "décideurs volontaires" présentent *des définitions d'éléments multidimensionnels qui sont soit erronées, soit non pertinentes* [11]. De fait, impliquer ces utilisateurs particuliers dans les méthodologies de conception d'entrepôt de données existantes n'est pas possible car ces méthodologies nécessitent une connaissance avancée des principaux concepts de la modélisation multidimensionnelle, et font l'hypothèse que les utilisateurs sont effectivement impliqués dans l'ensemble du projet.

Nous proposons donc une méthodologie innovante de conception collaborative d'entrepôt de données, utilisant un système d'aide à la décision en groupe (GDSS), qui aide les volontaires à choisir les besoins les plus pertinents parmi les besoins exprimés. En effet, les outils de GDSS sont conçus pour aider un groupe engagé dans un processus de décision collectif et collaboratif. Ce type de système n'a pas encore été utilisé pour la conception d'entrepôt de données [9].

2. Cas d'étude

Dans le cadre du projet ANR VGI4Bio, nous avons mobilisé deux bases de données VGI (Visionature et l'Observatoire Agricole de la Biodiversité - OAB) afin de construire une application SOLAP, qui permettra l'analyse d'indicateurs de biodiversité dans les milieux agricoles. Visionature et l'OAB impliquent respectivement environ 7700 et 1500 volontaires, qui produisent les données. A cette étape du projet, nous avons identifié 12 volontaires pouvant se prêter à une expérimentation de la méthodologie proposée.

La Figure 1 montre un modèle multidimensionnel défini par un volontaire concernant l'analyse de l'abondance de plusieurs espèces animales. Ce modèle permet de formuler des requêtes SOLAP comme : “*Quel est l'abondance totale des oiseaux selon l'altitude, l'espèce et la semaine ?*”.

L'objectif de notre proposition étant la résolution de conflits au sein des besoins définis par des volontaires, nous proposons d'illustrer la notion de conflit à partir de ce cas d'étude. Par exemple, pour la mesure de l'abondance de certaines espèces (“abondance”) comme les papillons, le protocole d'acquisition prévoit que l'opérateur observe les animaux en se déplaçant le long d'une ligne, d'une distance donnée, ou pendant une durée définie (environ 100 mètres pour les papillons). Ainsi, la mesure de l'abondance des papillons n'a de sens pour l'analyse biologique, que si elle est accompagnée de la durée ou de la longueur de l'observation.

Cet exemple nous montre que certains éléments multidimensionnels, considérés comme nécessaires par les volontaires lors de la définition du modèle, ne sont pas cohérents vis à vis du domaine d'application. Ces conflits ne sont pas causés par les données sources, mais sont le résultat de disparités de connaissance et d'expertise dans le domaine d'application du volontaire. Ainsi, ces conflits ne peuvent pas être résolus par un outil automatique, mais seulement par des experts du domaines.

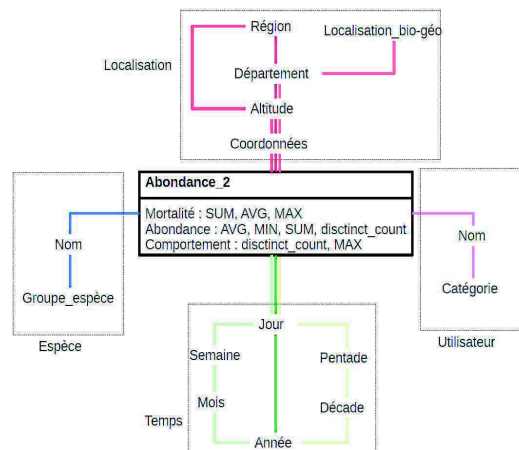


Figure 1. Exemple de modèle multidimensionnel conçu par un volontaire.

3. Méthodologie proposée

L'objectif de méthodologie est de résoudre les conflits engendrés par la définition des modèles multidimensionnels par les volontaires : “*Les éléments multidimensionnels (mesures, dimensions, hiérarchies) proposés par les volontaires sont-ils tous nécessaires ?*”.

Comme nous l'avons illustré avec le cas d'étude, les conflits engendrés par la définition des besoins par les volontaires doivent être tranchés au regard du domaine

d'application. De plus, comme indiqué précédemment, les volontaires participent ponctuellement à la collecte des données et ne sont pas impliqués en permanence dans le projet VGI. Pour pallier à ces limites dans notre approche, les modèles multidimensionnels sont proposés à un ensemble de volontaires particuliers appelés "committers", qui sont complètement impliqués dans le projet. Les committers décident si les besoins recueillis auprès des volontaires (sous forme de modèles multidimensionnels) seront implémentés ou non, en jugeant de la pertinence de ces besoins, grâce à leur expertise dans le domaine d'application.

Dans cette section, nous présentons la méthode générale de résolution des conflits. Cette méthode présente les différentes étapes que doivent suivre les committers pour statuer sur un modèle proposé.

Les committers définissent leur niveau d'expertise pour le cube selon leurs compétences dans le domaine d'application. Concrètement, les committers s'évaluent eux-mêmes. Ce niveau d'expertise permet de rendre prioritaires les choix des committers avec les compétences les plus appropriées vis-à-vis de l'évaluation du cube. Les committers évaluent successivement les indicateurs, les dimensions, les hiérarchies de chaque dimension, les niveaux de chaque hiérarchie. Chaque élément qui n'est pas jugé assez utile par le groupe est éliminé du modèle. Notons qu'un élément (cube, dimension, hiérarchie) dont les sous-éléments (indicateurs, hiérarchie, niveau) ont tous été éliminés, est lui même supprimé du modèle. Après cela, les committers doivent évaluer l'utilisabilité du cube avec chaque dimension retenue [10], car le nombre de dimensions utilisées affecte l'utilisabilité du cube, et le processus de prise de décision. Dans cette optique, on ajoute, en commençant par les dimensions les mieux notées par les committers, les dimensions au cube consécutivement. Après chaque ajout de dimension, le cube résultant est présenté aux committers. Par ce moyen, les committers peuvent explorer le cube avec la dimension ajoutée, décider de l'utilisabilité du cube avec la nouvelle dimension, et choisir de la conserver ou pas. Enfin, les committers votent l'implémentation du cube résultant.

4. Expérimentation

Notre approche a été implémentée dans une architecture ROLAP, utilisant Postgresql comme base de données, Mondrian comme serveur OLAP et JRubik comme client OLAP. Les méthodes de l'algorithme ont été implémentés dans le GDSS GRUS [9]. GRUS est un GDSS web, qui permet d'organiser le processus de choix des éléments à conserver ou à retirer des modèles. Il permet aux committers de voter pour les différents éléments de manière indépendante.

Dans cette section nous présentons les résultats d'une première expérimentation menée sur notre cas d'étude. Pour cette expérimentation, qui aura vocation à être complétée dans des travaux futurs, les committers sont trois experts appartenant à la LPO, la DREAL et AgroParisTech. Ils ont des profils différents (écologie,

gestionnaire environnementale et ingénieur en agronomie) et ils connaissent très bien les données sources.

Pour cette première expérience, notre méthodologie a été appliquée au cube de Figure 2a, qui est un cube qui a été conçu par un volontaire. Le cube résultat est montré en Figure 2b.

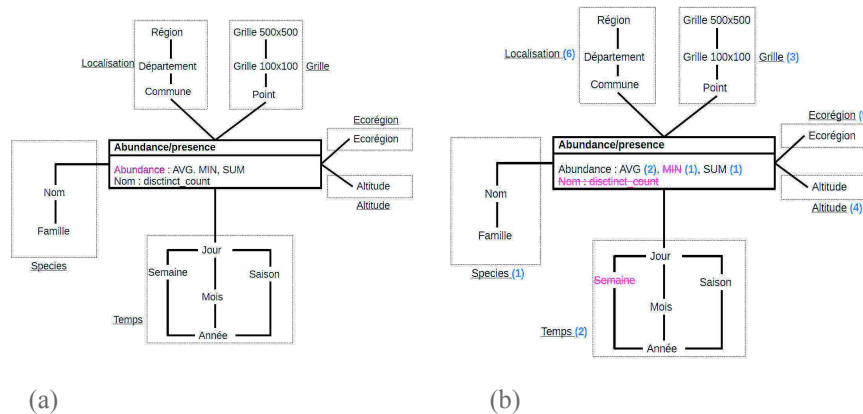


FIGURE 2. *Modèle multidimensionnel pour la biodiversité défini par un volontaire*
 Les indicateurs “Abondance+Min” et “nom+Distinct Count” et le niveau “Semaine” ont été éliminés du cube de la Figure 2b. Les chiffres notés à côté des noms des dimensions indiquent l’ordre d’évaluation des dimensions. Les committers ont ensuite voté le cube de la Figure 2b en décidant de l’implémenter.

Le processus d’évaluation a duré 2 heures, et a été réalisé par visioconférence. Nous avons ensuite questionné les committers sur la méthodologie. Ils estiment utile l’organisation et donc l’évaluation des composantes du cube comme nous l’avons définie. Ils ont aussi apprécié le système collaboratif fourni par GRUS. Par contre, ils ont fait émerger la nécessité de pouvoir apporter des modifications au cube proposé au début de l’évaluation, ce qui représente une future amélioration de notre algorithme.

5. Conclusions et travaux futurs

Dans cet article, nous proposons une nouvelle méthode de résolution de conflits, adaptée à la conception collaborative d’entrepôt de données. Cette méthode est une étape importante pour l’implication des volontaires dans la définition des besoins analytiques concernant des données VGI. La méthode proposée permet à des volontaires sans aucune compétence en système OLAP de participer au processus de conception de ce type d’outil. Cette méthode a été implémentée, ce qui a permis de mener les premières expérimentations. Ces premières expérimentations, bien que nécessitant d’être complétées, ont permis de valider la méthode sur un cas d’étude réel concernant la biodiversité dans les milieux agricoles. Ainsi, à l’avenir, nous proposerons une extension des critères utilisés par notre approche collaborative en se

basant sur des métriques qualitatives définies pour la satisfaction des utilisateurs d'entrepôts de données.

Remerciements. Ces travaux ont reçus le soutien financier du projet ANR-17-CE04-0012 VGI4Bio. Nous remercions tous les volontaires qui participent au projet VGI4Bio.

Références

1. Bimonte, S., Boucelma, O., Machabert, O., Sellami, S.: A new Spatial OLAP approach for the analysis of Volunteered Geographic Information. *Computers, Environment and Urban Systems*. 48, 111–123 (2014).
2. Cravero, A., Sepúlveda, S.: Multidimensional Design Paradigms for Data Warehouses: A Systematic Mapping Study. *Journal of Software Engineering and Applications*. 07, 53 (2013).
3. Egyed, A., Grunbacher, P.: Identifying requirements conflicts and cooperation: how quality attributes and automated traceability can help. *IEEE Software*. 21, 50–58 (2004).
4. Kimball, R., Ross, M.: *The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence Remastered Collection*. John Wiley & Sons (2016)
5. Levrel, H., Fontaine, B., Henry, P.-Y., Jiguet, F., Julliard, R., Kerbirou, C., Couvet, D.: Balancing state and volunteer investment in biodiversity monitoring for the implementation of CBD indicators: A French example. *Ecological Economics*. 69, 1580–1586 (2010).
6. Romero, O., Abelló, A.: A Survey of Multidimensional Modeling Methodologies. *IJDWM*. 5, 1–23 (2009).
7. Stefanovic, N., Han, J., Koperski, K.: Object-based selective materialization for efficient implementation of spatial data cubes. *IEEE Transactions on Knowledge and Data Engineering*. 12, 938–958 (2000).
8. Sui, D.Z., Elwood, S., Goodchild, M. eds: *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*. Springer Netherlands (2013)
9. Zaraté, P.: *Tools for Collaborative Decision-Making: Zaraté/Tools for Collaborative Decision-Making*. John Wiley & Sons, Inc. (2013)
10. Golfarelli, M., Rizzi, S.: Data warehouse testing: A prototype-based methodology. *Information & Software Technology* 53(11): 1183-1198 (2011)
11. Pohl, K.: *Requirements Engineering: Fundamentals, Principles, and Techniques*. Springer Publishing Company, Incorporated (2010)