

---

# Journal de l'OSGeo

Le Journal de l'Open Source Geospatial Foundation

Volume 2 / Septembre 2007

---

## Dans ce volume

Les bases de la topologie

1Spatial : *Concepts de qualité des données*

Introduction à MapWindow & GeoNetwork

LizardTech : *Pourquoi utilise-t-on des logiciels Open Source*

Rapport des Local Chapter : Taiwan, U.K., Francophone, Espagnol ...

Étude de cas : UN FAO, Traçage des navires de pêche ...

Rapport d'événements communautaires : Inde, France

Traitement distribué et GRASS

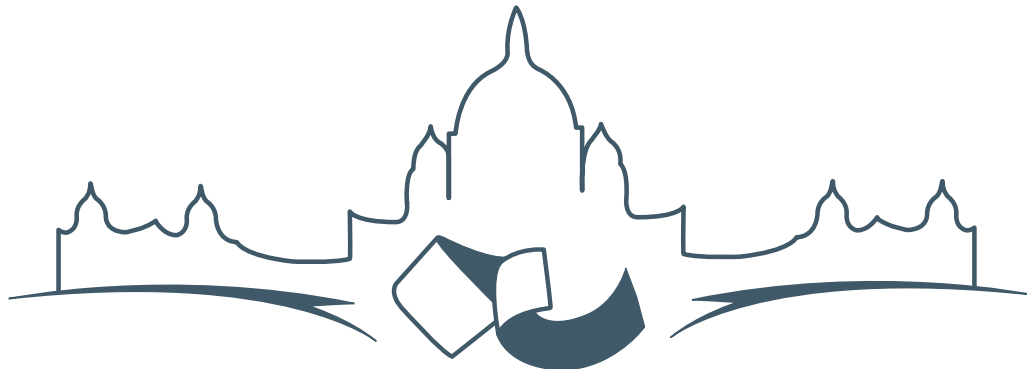
Actualités et mise à jour des logiciels ...

---

## Table des matières

|  |   |
|--|---|
| <b>Perspectives des sponsors</b>   | 2 |
| 1Spatial : Qualité de l'information spatial et communauté Open Source. . . . . | 2 |

---



**2007 FREE AND OPEN SOURCE SOFTWARE  
FOR GEOSPATIAL (FOSS4G) CONFERENCE**  
VICTORIA CANADA 🍁 SEPTEMBER 24 TO 27, 2007

---

## Perspectives des sponsors

---

# 1Spatial : Qualité de l'information spatial et communauté Open Source.

par Mike Sanderson, Graham Stickler, Steven Ramage, 1Spatial, traduit par Cyril de Runz

*Avec un intérêt special pour FDO et pour la direction prise par d'autres technologies Open Source, 1Spatial (précédemment nommée Laser-Scan) est devenu un sponsor officiel de l'OSGeo en 2006. Les domaines d'expertise de 1Spatial sont le traitement des données, la gestion de leur qualité, et aussi le produit Radius Studio. On a demandé aux auteurs de présenter quelques-unes de leurs réflexions sur ces projets, et notamment la manière dont les outils Open Source peuvent servir dans la réalisation de ces idées. – Le rédacteur en chef*

## Introduction

L'Open Source Geospatial Foundation (OSGeo) a été créé pour supporter et construire des logiciels géospatiaux open-source de la plus haute qualité. L'objectif est de créer une synergie de la communauté par la communauté pour améliorer la qualité des logiciels (1). Cet article porte sur l'élément qualité mais d'un point de vue complètement différent.

Il existe différents problèmes de gestion de la qualité des données spatiales, que l'on utilise ou non un logiciel Open Source. Ils sont toujours issus de la co-

hérence, de l'intégrité, et de la forme des données, car les données changent constamment. Ceci peut être dû aux changements du monde réel, à l'introduction de nouvelles technologies d'acquisition, ou à des mises à jour ou des changements organisationnels qui altèrent les conditions des frontières.

Depuis le milieu des années 1990, l'organisme international de standardisation (International Standards Organisation - ISO) ainsi que l'Open Geospatial Consortium (OGC) ont travaillé à l'élaboration de standards, créant et couvrant le Web Feature Service (WFS), le Web Map Service (WMS) et le Geography Markup Language (GML). Plus récemment, grâce au travail de l'OSGeo, des mécanismes d'accès aux données géospatiales ne se souciant pas des sources ont été définis en utilisant Feature Data Objects (FDO).

Un résultat fut notamment la création d'un nouvel ensemble de questions ayant pour objet la forme et la qualité des données spatiales. Désormais nous souhaitons étendre le paradigme pour inclure la qualité des informations spatiales et les problèmes de réutilisation car les données collectées pour un usage sont utilisées pour d'autres usages divers, et le plus souvent par des organisations complètement distinctes. En tant qu'industriel, Google a accru les exigences professionnelles au point que cela nous ferait

du tort à tous si la qualité des données était suspecte.

Notre objectif est de fournir un ensemble d'outils basés sur le web permettant d'évaluer la qualité des données, comme par exemple les tests de conformité. L'idéal serait d'offrir une gestion améliorée des données spatiales dans un environnement open-source. Cela comprend la volonté d'augmenter l'efficacité opérationnelle en permettant le développement d'un socle logiciel pour l'agrégation de données spatiales.

Nous souhaitons laisser la possibilité à la communauté OSGeo de contribuer à ces initiatives via notre programme dédié aux praticiens (Practitioner Program) et à travers du groupe de travail sur la qualité des données de l'Open Geospatial Consortium (OGC).

## Informations nécessaires

Dans l'optique de partager l'information, il est nécessaire de comprendre le travail effectué par l'organisation internationale de standardisation (International Standards Organisation - ISO). L'ISO fournit les lignes directrices pour mettre en commun les structures dédiées à la gestion de la qualité des données spatiales. Les normes ISO 19113 et ISO 19114 décrivent divers mécanismes déterminant et mesurant la qualité des données. Ces standards ISO présentent les principes et procédures d'évaluation de l'information géographique. Le chapitre 15 de [Spatial Data Quality](#)(2) propose un résumé utile sur ces principes et procédures. La première phrase de la norme ISO 19113 donne le ton :

Les jeux de données géographiques sont de plus en plus partagés, échangés et utilisés pour d'autres objets que ceux prévus par leurs producteurs.

Les précurseurs des standards ISO actuels furent appliqués en utilisant le projet de diagramme numérique du monde (Digital Chart of the World - DCW) comme cas d'étude. Le DCW<sup>1</sup> fut produit en 1992 et représenta un effort pour aller vers les problèmes d'"ajustement par objectif". Mais en 1995, l'initiative prit fin, car il n'était possible d'évaluer que les aspects qualitatifs (non-quantitatifs) de la qualité des jeux de données géographiques. L'évaluation quantitative était tout simplement impossible pour les grands jeux de données géospatiales, ceci étant dû au manque de puissance des processeurs de l'époque. C'est pourtant l'évaluation quantitative qui est réellement valable pour établir la cohérence logique et

la précision du positionnement. Et comme, avec tant d'éloquence, Jakobsson le dit : "Combiner les jeux de données dont on ne connaît pas la qualité peut-être vraiment difficile voir impossible." (2)

## Aspect quantitatif de la qualité des données spatiales

Un audit de la qualité des données est conçu et implémenté dans le but de déterminer la réponse à la question : dans quelle mesure vos données spatiales sont-elles bonnes (ou mauvaises) ? Il existe trois objectifs de haut niveau pour renforcer la qualité des données :

- Produire des mesures statistiquement valides de la qualité des données sources ou principales
- Chercher, identifier et documenter les preuves de la qualité de données (et les exceptions)
- Créer un rapport d'évaluation et de recommandations

On doit cependant donner quelques lignes directrices à suivre et une approche formelle dans le but d'entreprendre l'évaluation quantitative. L'organigramme de la Figure 1 décrit une approche pour passer en revue la qualité de l'information spatiale. Cette approche fut essayée, testée et utilisée par l'Spatial pour ses projets avec ses clients et ses partenaires - elle représente une voie pour entreprendre l'évaluation quantitative.

<sup>1</sup>Digital Chart of the World : <http://www.nlh.no/ikf/gis/dcw/>

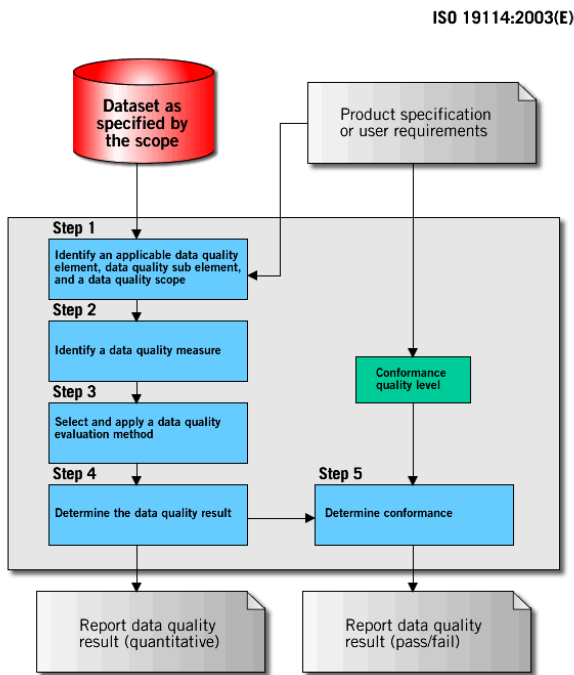


Figure 1 - Evaluating and reporting data quality results

FIG. 1 –

Le plus important dans cette approche est l'analyse contextuelle de la qualité de l'enregistrement des données ou des données spatiales possédées. Ceci implique de travailler à travers les étapes 2 à 5 à l'utilisation d'une approche basée sur un système de règles pour déterminer la conformité des données spatiales par rapport aux spécifications ou aux règles métiers. Dans un environnement open source, ceci peut-être évalué par l'intermédiaire d'une interface XML pour FDO (Feature Data Objects). Cette approche présente deux éléments clés. Premièrement, elle procure une évaluation quantitative de la qualité des données spatiales, c'est à dire la conformité des données % par rapport aux règles du métier. Deuxièmement, c'est une vérification indépendante, c'est à dire qui n'utilise pas l'outil SIG par lequel les données ont été créées ou éditées – elles restent inchangées.

Les problèmes de qualité des données sont souvent répandus par la source de données et ont leur origines dans celle-ci. Ils peuvent être géométriques, topologiques, ou basés sur les attributs. Pour combattre ces problèmes, à la fois les propriétaires, les analystes et les parties prenantes doivent comprendre la source de leurs données. Cette compréhension peut venir du profil des données ou de ce qui peut-être interprété des métriques d'autodescription.

Les analystes et gestionnaires des données doivent aussi comprendre comment le profil des données répond aux besoins métiers. C'est ici que l'audit commercial à base de règles n'est pas uniquement pratique, mais critique.

## Règles pour la qualité des données

Afin d'établir la forme des données spatiales, il est essentiel de comprendre les règles du métier relatives à ces données et de savoir les interpréter. Entreprendre un audit basé sur les règles métier peut s'avérer critique dans la lutte contre les problèmes liés à la qualité des données.

La Figure 2 propose un exemple des mesures identifiées pour l'évaluation de la qualité des données spatiales. Elle fait référence à cinq éléments qui évoluent en fonction du type des données spatiales ou de l'application. Le but de cette étape est de définir les règles du métier auxquelles les données doivent se soumettre. Il est souvent difficile d'obtenir des informations des utilisateurs. Les spécifications du modèle original des données sont souvent inaccessibles ou n'ont pas été mises à jour depuis leur conception. Plutôt que de partir de zéro, une source potentielle de règles est la donnée elle-même. Les règles peuvent être découvertes en analysant les modèles statistiques dominant dans les données.

Le défi actuel est qu'il y a des outils limités dans la communauté open source et plus largement dans le marché du géospatial en général pour mener à bien de telles tâches et mesurer quantitativement la qualité des données spatiales. Tandis que nous devons regarder vers les architectures ouvertes et les standards tels que définis par les architectures orientées service (Service-Oriented Architectures - SOA), le cadre du web sémantique du World Wide Web Consortium (W3C) et le langage pour les ontologies WEB (Web Ontology Language - OWL), il est clair que les règles du web sémantique ne conviennent pas aux besoins du géospatial. Un travail récemment effectué par l'OGC dans le cadre de leur banc d'essai OWS-4 sur un service d'évaluation quantitative topologique (Topology Quantitative Assessment Service - TQAS) appuie ce point de vue. s

ISO 19114:2003(E)

| Data quality element | Data quality subelement                | Relevant |
|----------------------|--|----------|
| Completeness         | omission                               | yes      |
|                      | commission                             | yes      |
| Logical consistency  | conceptual consistency                 | no       |
|                      | domain consistency                     | yes      |
|                      | format consistency                     | no       |
| Positional accuracy  | topological consistency                | yes      |
|                      | absolute or external accuracy          | yes      |
|                      | relative or internal accuracy          | no       |
| Temporal accuracy    | gridded data position accuracy         | no       |
|                      | accuracy of a time measurement         | no       |
| Thematic accuracy    | temporal consistency                   | no       |
|                      | temporal validity                      | no       |
|                      | classification correctness             | no       |
|                      | non-quantitative attribute correctness | no       |
|                      | quantitative attribute correctness     | no       |

Figure 2 - Summary of relevant quantitative quality information

FIG. 2 –

1Spatial s'est attaqué à ce problème et a développé un outil web, Radius Studio, basé sur ce paradigme de la qualité quantitative à base de règles. En utilisant une adaptation d'un algorithme performant d'intelligence artificielle à la fouille de données spatiales. Il fonctionne en considérant en premier un petit échantillon d'objets, issu des données stockées, pris au hasard et travaille ensuite à partir de ces objets, en considérant les objets proches. Un ensemble initial de règles spatiales est proposé et ensuite amélioré pour inclure les éléments non spatiaux à l'instar des jointures d'attributs, les égalités et inégalités, corrélées aux relations spatiales entre objets de l'échantillon. L'ensemble final de règles est converti dans un format permettant de les stocker dans un répertoire de règles qui est basé sur un langage commun d'interface incorporant les opérateurs spatiaux de l'OGC.

Il est évident qu'un tel outil quand il est combiné avec FDO et MapGuide Open Source peut permettre une évaluation des données spatiales par rapport aux règles métiers. Cette approche en test de conformité peut être d'une valeur inestimable pour fournir une évaluation de la qualité des données spatiales.

1Spatial a déjà commencé un programme utilisateur (Practitioner Program) pour permettre aux clients finaux de valider leurs données spatiales à travers le Web en lançant un audit de leurs données. Par l'accès au Radius Studio, les praticiens ont la possibilité d'utiliser le constructeur de règles et l'évaluation de la capacité des données à aller vers un audit de certification au nom des consommateurs. En retour 1Spatial reçoit le ressenti sur le produit et les outils 1Spatial sont recommandés pour n'importe quel utilisation, réutilisation ou besoin en re-ingénierie.

## Opportunité pour l'Open Source

Il y a ici une opportunité pour la communauté open source de s'engager avec Radius Studio pour

développer cette possibilité et de faire qu'elle soit largement disponible.

En hébergeant ce service d'évaluation, l'OSGeo favorise son développement, procurant ainsi la base d'une coopération entre les différentes parties et permettant ainsi l'harmonisation de l'information géographique et la construction d'une communauté. L'opportunité existe alors pour la communauté open source d'ajouter des fournisseurs FDO à ceux disponibles et d'ouvrir plus de formats à l'évaluation quantitative de la qualité des données spatiales.

Une autre raison pour placer à l'OSGeo un tel service d'évaluation est de créer une communauté qui travaillerait à la définition du langage utilisé pour définir les règles métiers qui aboutiront aux mesures de qualité. Le langage de règles web sémantiques (Semantic Web Rules Language - SWRL) n'est actuellement pas assez mature, mais l'histoire a déjà montré que nous, partie prenante de cette industrie, pouvons créer un tel standard. Une fois ce standard disponible, nous pourrions tous faire nos estimations de la validité de nos données, par rapport à leur usage. Ce doit être fait par un processus automatisé, et à la différence de 1995, les outils et la puissance de calcul sont maintenant disponibles pour cela. Cela pourrait prendre la forme d'une période d'évaluation gratuite de 14 jours pour faire cette analyse si ces données ne sont pas gratuites, et une fois l'évaluation terminée, il sera possible de décider si l'on doit payer ou non, ou combien payer pour ces données. Ainsi, le débat données libres contre propriétaires devient inutile.

Si le langage pour exprimer ces règles ne peut pas être créé assez rapidement, nous pouvons choisir la solution du géographe, "l'expression pseudo-quantitative." En suivant la tradition Amazon d'une validation par un groupe de pairs, la communauté des utilisateurs pourrait valider la valeur nominale d'un jeu de données spatiales pour sa complétude, sa constance logique, et sa précision de position, de temps et de thématique. La communauté Open Source fonctionne bien avec le même modèle de validation par les pairs, nous pensons donc qu'elle sera intéressée par cette approche.

Si vous êtes souhaitez prendre part à l'extension de l'OSGeo pour l'évaluation de la qualité de l'information spatiale ou jouer un rôle dans le développement de standards pour la qualité de l'information spatiale, nous vous prions de contacter 1Spatial directement ou à travers le groupe de travail de l'OGC sur la qualité des données :

<http://www.opengeospatial.org/projects/groups/dqwg>

## Biibliographie

1. The Cathedral & The Bazaar, Raymond, E.S., (2001) ISBN 0-596-00108-8
2. Spatial Data Quality, Wenzhong Shi, Peter Fisher, Michael Goodchild, (2002). ISBN :0415258359
3. Improving Operational Efficiency with Geographic Information, Finnish Ministry of Agriculture & Forestry, (2006). ISBN 952-453-301-4

*Mike Sanderson*

*CEO, 1Spatial*

*Graham Stickler  
Product and Marketing Director, 1Spatial*

*Steven Ramage  
Business Development Director, 1Spatial*

<http://www.osgeo.org>

[info AT 1spatial.com](mailto:info@1spatial.com)

**Rédacteur en chef :**Tyler Mitchell - [tmitchell AT osgeo.org](mailto:tmitchell@osgeo.org)**Éditeur, actualités :**

Jason Fournier

**Éditeur, Études de cas :**

Micha Silver

**Éditeur, Zoom sur un projet :**

Martin Wegmann

**Éditeur, Études d'intégration :**

Martin Wegmann

**Éditeur, Cours de programmation :**

Landon Blake

**Éditeur, Rapport d'événements :**

Jeff McKenna

**Éditeur, Études thématiques :**

Dr. Markus Lupp

**Responsable relecture :**

Daniel Ames

**Remerciements**

Divers relecteurs &amp; le projet actualités de GRASS

Le *Journal de l'OSGeo* est une publication de la *Fondation OSGeo*. La base de ce journal, le source du style  $\text{\LaTeX} 2_{\epsilon}$  a été généreusement fournie par l'équipe éditoriale de l'actualité de GRASS et R.



Ce travail est sous licence Creative Commons Paternité-Pas de Modification version 3.0. Pour voir un exemplaire de cette licence, rendez-vous sur :

<http://creativecommons.org/licenses/by-nd/3.0/deed.fr> ou envoyez une demande À Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.



the OSGeo Journal url for submitting articles, more details concerning submission instructions can be found on the OSGeo homepage. Tous les articles sous copyright par leurs auteurs respectifs. Merci d'utiliser l'URL du Journal OSGeo pour envoyer vos articles ; de plus amples détails concernant les instructions d'envoi sont disponibles sur la page d'accueil d'OSGeo.

Journal en ligne : <http://www.osgeo.org/journal>

Site Internet de l'OSGeo : <http://www.osgeo.org>

Contact mail de l'OSGeo, PO Box 4844, Williams Lake, British Columbia, Canada, V2G 2V8



ISSN 1994-1897