

Interopérabilité : état de l'art et retour d'expériences sur l'accès aux données géographiques, accéder et partager l'information géographique

Serge Margoulies

► To cite this version:

Serge Margoulies. Interopérabilité : état de l'art et retour d'expériences sur l'accès aux données géographiques, accéder et partager l'information géographique. Jun 2004. sic_00001237

HAL Id: sic_00001237

https://archivesic.ccsd.cnrs.fr/sic_00001237

Submitted on 7 Dec 2004

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Interopérabilité : état de l'art et retour d'expériences sur l'accès aux données géographiques

Accéder et partager l'information géographique

Serge MARGOULIES
Directeur Général
IONIC Software s.a.

*18, rue de Wallonie,
B-4460 Grace Hollogne
Tel. : +32 4 364 03 64
Fax : +32 4 253 47 37
sm@ionicsoft.com*

RÉSUMÉ. La présentation d'IONIC fera état de l'avancement des démarches de l'OGC et de l'ISO, en incluant les notions de conformités. Au travers de retour d'expériences vécues en 2003 dans le cadre de réalisations concrètes et déployées sur des bases d'objets riches en contenu et sur des grandes bases d'imagerie, la présentation fera un résumé des techniques mises en œuvre, les difficultés rencontrées et les avantages concrets ou espérés.

MOTS-CLÉS : interopérabilité, métamodèle, métadonnées, services d'accès, commerce électronique, catalogage, GML, WFS, WCS, WMS, Portrayal, Catalogue.

1. Introduction

Imaginez un monde où les échanges entre les systèmes hétérogènes seraient facilités, sans pertes et dégradations de l'information géographique, tout en permettant de développer des applications utiles et efficaces, plus vite et à des coûts bien inférieurs. Imaginez un monde où il serait possible d'accéder à n'importe quel type d'information (météo, démographique, environnement...) partout dans le monde, au moment où on en a besoin, et en ne payant que la donnée utilisée pour le projet en cours.

La diffusion de données sur Internet ouvre de nouvelles perspectives tant en termes de réduction des coûts à travers des architectures plus modulables, qu'en termes d'émergence et de potentiel représenté par de nouveaux marchés pour l'accès et la vente de l'information géographique. Plusieurs nouvelles techniques de diffusion sont apparues, utilisant les importants efforts de standardisation et de normalisation pour une meilleure interopérabilité des systèmes.

2. Diffusion d'information sur Internet

La diffusion d'information cartographique sur Internet a 10 ans, tout comme l'ISO tc/211. Les concepts liés à cette diffusion sont désormais bien assimilés. Depuis plusieurs années, des efforts énormes ont été réalisés en termes de recherche et de normalisation. Le but de ces travaux est d'améliorer la qualité des échanges, diminuer les efforts nécessaires pour utiliser l'information et créer une véritable plate-forme commerciale autour de la donnée géographique.

La donnée géographique est complexe et fortement diversifiée par nature. Elle n'est pas non plus gratuite, ni nécessairement ouverte à tous. Son interprétation est parfois délicate. Ces obstacles ont empêché depuis toujours que l'information soit achetée et utilisée rapidement et efficacement. Ce qui motive tous ceux qui travaillent dans l'interopérabilité, c'est le rêve qu'un jour prochain, l'utilisateur pourra accéder automatiquement aux sources de données, interroger des services et obtenir ou payer ce dont il a besoin au moment où il en a besoin.

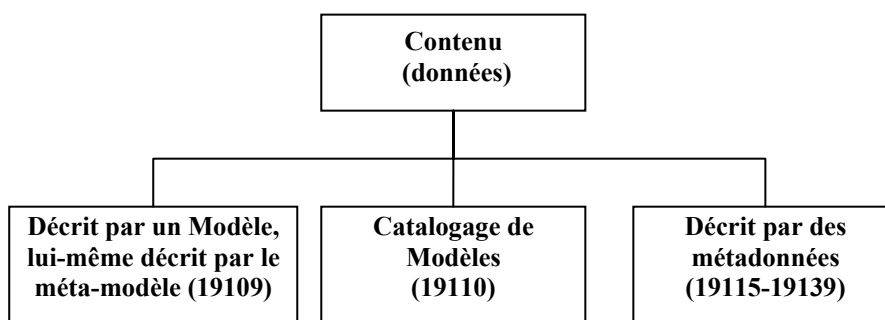
Cette motivation anime aussi les fournisseurs de données qui voient leur modèle commercial évoluer. Ils se doivent alors de rester leader dans leur métier de base, et devenir des fournisseurs en lignes, crédibles et performants pour ne pas être rattrapés par de nouveaux venus.

3. Quatre axes de normalisation

Depuis longtemps les efforts combinés de normalisation par l'ISO, l'OGC et d'autres organismes qui y contribuent consistent en quatre axes majeurs.

- a) Une définition d'un métamodèle de données. Ce mot un peu rébarbatif signifie « le canevas, la manière de décrire un modèle de données de manière à ce que d'autres puissent le comprendre. »
- b) Un modèle de métadonnées. Ce modèle permet de décrire un ou plusieurs jeux de données.
- c) Un ensemble de services d'accès. Ces services ont des interfaces bien définies, orientées métiers, permettant de mettre en ligne des données, les modèles de données et leurs métadonnées.
- d) Un environnement de commerce électronique protégeant à la fois le fournisseur de données et le client, de manière à développer l'accès et la vente des données.

Chacune de ces 4 composantes est nécessaire pour faire une infrastructure de données géospatiale en ligne vraiment optimale. Examinons ces aspects, leur normalisation et le statut de leur interopérabilité. Nous analyserons ensuite quelques cas d'applications et les leçons qu'on peut en tirer.



4. Métamodèle

Le métamodèle peut être vu comme le référentiel de base de l'interopérabilité. L'ISO a défini un modèle (ISO19109) qui indique comment définir ses propres modèles de données. Il y a donc un ensemble de règles à suivre pour définir le modèle d'un type « route », un type « parcelle » ou un type « zone protégée ». Chacun est libre de définir son propre modèle de données, pour autant qu'il le fasse dans le cadre du référentiel ISO.

Parmi les règles les plus importantes, on notera que les éléments d'un type doivent prioritairement être ceux définis par les autres normes ISO. Il est par exemple interdit de définir ses propres types de géométrie. Elles doivent nécessairement être extraites de ISO19107.

4 Modèles et Applications pour le Document Cartographique Numérique (SDN 04).

C'est ainsi que nous aurons tous, un jour, la définition de nos objets publiés et échangés, décrits dans le même métamodèle. Nos logiciels pourront automatiquement comprendre les schémas des données, sans même les connaître à l'avance. C'est un grand pas vers l'interopérabilité. Ainsi, on ne forcera personne dans un modèle de données, mais tout le monde se comprendra. Croyez-le ou non, ça marche déjà aujourd'hui !

5. GML (Geography Markup Language)

GML a été créé de manière à pouvoir définir plus facilement des schémas de données de diffusion et leur encodage en XML. Quand on définit un type en GML et qu'on suit en plus un certain nombre de règles de bon usage, notamment sur l'usage de types sémantiquement bien définis, on obtient des schémas compréhensibles dynamiquement par d'autres. GML a été adopté par l'ISO pour devenir ISO 19136.

6. Métadonnées

Communément, la majorité parle de métadonnées quand elle veut dire « métadonnées des données ». Les métadonnées sont une nécessité majeure pour la découverte et la vente des données. Le dicton dit : « Les métadonnées, on peut les encoder maintenant ou plus tard. Plus tard, ce sera plus cher ! » Il y a plusieurs types de métadonnées, et toutes sont nécessaires pour bien faire fonctionner un système. Les métadonnées les plus connues sont celles qui décrivent la donnée de base telles que qualité, date, précision, mise à jour, exhaustivité, durée de vie, propriétaire, droits. Nous verrons plus loin les métadonnées du service qui les fournit.

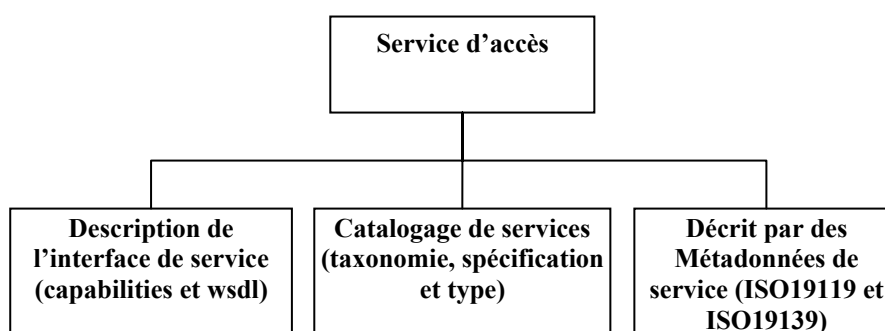
Dans ce cadre, ISO19115 est devenu omniprésent. Actuellement, un gros effort se passe sur ISO19115-2, à savoir une spécification complémentaire pour ce qui touche à l'imagerie. Dans le monde XML, c'est ISO19139 qui est applicable car il est l'encodage XML de 19115.

7. Service d'accès

Pour réaliser un accès en ligne aux données, il faut passer par des services exposés sur le web. Le Web Services le plus simple est celui qui permet de télécharger un lot de données. On est aujourd'hui capable d'offrir des services normalisés orientés métiers. Le but d'un service web est d'offrir un service métier bien défini qui donne accès à la fois au modèle de données exposé, aux métadonnées et aux données elles-mêmes. L'important est la sémantique de l'interface du service et cette définition se trouve dans une spécification. Elle est souvent indépendante du protocole utilisé. Ainsi, un serveur WFS retourne le

modèle de données dans un schéma GML et il retourne le même contenu en http get, post ou soap.

De la même manière que les métadonnées de données nous aident à décrire la donnée, les métadonnées de services permettent de décrire le service. On peut ainsi découvrir le propriétaire du service, les objets qu'il sert, la version de l'interface, les protocoles supportés, la taxonomie du service et bien d'autres informations telles que les restrictions, disponibilité, et prix du service. Pour les métadonnées de services, c'est ISO19119 dont l'encoding xml s'appelle les « capabilities »



8. Type de Services

Un service web a besoin de métadonnées pour s'auto décrire. Un service Web métier qui ne décrirait ni son service ni son contenu en termes métiers, n'aurait pas d'intérêt. Personne ne le trouverait ni ne l'utiliserait. Examinons quelques services OGC/ISO qui s'auto décrivent en utilisant les métadonnées telles que décrites :

Les services d'accès se décomposent en plusieurs catégories dont les plus connues, les plus utilisées et les plus matures aujourd'hui sont :

- a) l'interface WFS (Web Feature service): C'est l'interface majeure, celle du serveur d'entités géographiques. L'interface permet d'obtenir le modèle des données exposées, sous la forme d'un schéma GML. L'accès aux métadonnées du service et des données est direct. Il est également possible de soumettre des requêtes complexes sur les attributs des objets et leurs caractéristiques géographiques. Le résultat sera toujours une collection d'entités exprimées en GML. Cette interface, qui est le pilier de l'accès aux données géographiques, supporte les protocoles http get, http post et une version soap est finalisée et sera publiée sous peu.
- b) Une version transactionnelle de cette interface, appelée WFST, permet de faire des créations, modifications, suppressions ou lock d'objets, à condition qu'on en ait le droit. Cette fonction permet d'éditer les

6 Modèles et Applications pour le Document Cartographique Numérique (SDN 04).

objets, tant au niveau géométrique qu'attributaire ou de faire des annotations.

- c) L'interface WCS (Web Coverage Service): cette interface « Grid Coverage» permet l'accès à l'information brute ou rééchantillonnée des images. Elle propose également une description complète de l'information sous-jacente et permet l'accès aux métadonnées. On peut ainsi découvrir que tel service donne accès à 50 scènes multi-spectrales, fournit des métadonnées sur chacune des bandes des images, avec des attributs tels que le spectre, la date de saisie ou la couverture nuageuse. L'interface autorise des extractions de scènes entières, partielles voire même rassemblées, avec ou sans rééchantillonnage spectral, radiométrique ou temporel.
- d) l'interface WMS (Web Map Service): Cette interface très connue, portant le numéro ISO 19128, permet de demander une carte sous forme d'image ou de SVG. Elle permet de réaliser un rendu des informations sous jacentes (vecteurs ou images), d'une manière désirée par le « fournisseur », qui va imposer un ensemble de styles prédéfinis.
- e) L'interface Portrayal: Cette interface permet de faire le rendu de plusieurs services distribués et de les superposer en imposant un ordre et un style définis et demandés par l'utilisateur (et non plus définis par le fournisseur). Extraire des parcelles sur un serveur et les classifier, extraire de l'imagerie et appliquer une palette de couleurs et enfin superposer l'ensemble à un troisième serveur de fond de plan est possible aujourd'hui. Pour cela, il y a une interface Portrayal, extension du WMS, et un langage de style, appelé SLD.
- f) L'interface Catalogue Web (CS-W), permet d'enregistrer des ressources (services ou données) décrites par des métadonnées, de les indexer et de les rechercher. On peut ainsi demander au catalogue de retrouver toutes les images ayant telle qualité et telle couverture nuageuse sur une région. Le catalogue retournera alors l'url d'un service qui permet d'obtenir la donnée désirée. Cette interface est le profil géospatial de ebXML (le standard de l'e-business), ce qui permet d'envisager les relations commerciales réelles. Les informations découvertes peuvent être commandées dans un canevas prévu pour garantir la protection de l'acheteur et du vendeur. Notons, pour le lecteur averti, que ebXML apporte une vraie valeur ajoutée, même lors de son utilisation combinée avec un catalogue uddi.

Ces interfaces fonctionnent particulièrement bien grâce à l'amélioration constante de la sémantique des interfaces décrites dans les documents de spécification. A cet égard, l'OGC a mis en place un grand banc de tests de conformité appelé « CITE » (Compliance Interoperability Testing Evaluation) pour tester les implémentations. Ce banc, constitué de plus de 100 tests pour le WMS et

de 500 tests pour le WFS a permis de lever de nombreuses ambiguïtés de la spécification.

Nous venons de voir l'état de l'art et de la technique en termes de standardisation. Examinons quelques cas d'applications de ces technologies qui seront détaillés lors de la présentation pour en dégager des enseignements. Citons,

- Le projet PathFinder de l'Ordnance Survey : le projet consiste à créer une base de données spatiale sur l'Angleterre avec OSMasterMap et permettre son accès continu au travers d'une interface WFS avec possibilité d'extraction dynamique en GML. Les données sont livrées sous la forme d'un lot GML qui est constitué au moment de la requête. L'utilisateur accède à une interface intuitive, se positionne, sélectionne sa zone d'intérêt et le type d'objet qui l'intéresse. Il peut alors connaître le prix de son extrait et passer commande. Il peut également enregistrer sa requête pour la ré-exécuter plus tard de manière à obtenir les objets ayant varié. Enfin, il est possible de demander à être notifié régulièrement et automatiquement de modifications survenant dans la zone d'intérêt. Ce projet ambitieux passe désormais en phase 2, de tests grandeurs nature impliquant de nombreux utilisateurs/partenaires. Plusieurs enseignements seront présentés lors de l'exposé.
- Le projet « Synergy » de la Nasa. Ce grand projet a pour objectif de mettre de l'information satellitaire environnementale à disposition des agences locales et régionales en support à la prise de décision. L'ensemble des données est acquise par plusieurs senseurs et produisent au moins 3 produits différents. L'information en ligne, exprimée en terrabytes, représente une toute petite partie de l'information réellement disponible. Le but du système est de permettre à des scientifiques, des administrations et des partenaires de faire des recherches dans un catalogue spatial et temporel énorme, variant sans cesse puisque de nouvelles images arrivent en continu. Les spécifications WCS sont intensivement utilisées puisque le besoin d'accès aux données multispectrales brutes est nécessaire pour permettre l'analyse scientifique. Cependant, certains accès aux webservices se font avec traitements et réassemblage automatique multi-scènes, avec des fonctions telles que la reprojection, rééchantillonnage, filtres ou classifications. Nous examinerons ce que ces nouvelles technologies impliquent pour les utilisateurs et décideurs.
- Le projet "NatureGis" est un réseau pan Européen de protection de l'environnement. Il se positionne dans le cadre le plus avancé d'INSPIRE et préconise l'utilisation des technologies dont nous parlons. Il a pour but de réaliser un réseau distribué de participants européens qui pourront s'échanger de l'information géographique et environnementale de manière optimale. Le projet a commencé par une grande enquête du besoin et des objets à gérer. Les problèmes organisationnels ne sont pas minces, mais ce

projet permettra, grâce à l'interopérabilité mise en place, des gains au niveau des référencements des données, un accès plus rapide à l'information et surtout des possibilités infinies de croisement de l'information pour une prise de décisions rapides et efficaces.

9. Conclusion

Un monde entier d'échanges entre systèmes hétérogènes apparaît. Il permet de développer des applications plus vite et à des coûts inférieurs. On trouve désormais sur le marché des produits basés sur les standards qui respectent les règles de bon usage pour construire des solutions efficaces.

Après 10 ans de travaux et d'évolutions, l'interopérabilité est arrivée à une étape majeure de son histoire, avec des spécifications abouties, des standards ISO adoptés et un consensus global très prometteur. La diffusion de données sur Internet ouvre de nouvelles perspectives tant en termes de réduction des coûts au travers d'architectures plus modulables, que grâce au potentiel de nouveaux marchés pour l'accès et le partage de l'information géographique. L'expérience acquise permet de dégager les règles de bonne conduite et d'augmenter l'efficacité des projets suivants. L'Europe est dans le peloton de tête et peut revendiquer de très bons experts et de très bons éditeurs de logiciels dans ce domaine.

IONIC Software