

Le petit livre sur la

Qualité des Données Spatiales



\$3,1

MILLIARDS

est combien IBM estime que
les données de mauvaise
qualité coûte à l'économie
américaine chaque année

L'importance de la qualité des données

Les données géospatiales sont de plus en plus au cœur des prises de décision dans tous les domaines. En fait, ces informations pourraient générer jusqu'à 700 milliards \$ de valeur pour l'utilisateur final.¹

Néanmoins, toutes les données ne sont pas prêtes à l'emploi, et le coût occasionné par des données incorrectes peut s'avérer considérable. D'après de récentes études, IBM estime que la piètre qualité des données (de quelque type que ce soit) coûte 3,1 milliards \$ à l'économie des États-Unis chaque année.²

Avec cet accroissement de l'utilisation et du partage des données géospatiales, la précision et la fiabilité de celles-ci acquièrent une importance capitale. Une erreur de quelques mètres sur un plan de rues peut envoyer un service d'urgence vitale à des kilomètres de la bonne adresse et lui faire perdre de précieuses minutes. La mauvaise représentation d'une ligne électrique peut mettre en péril la vie des travailleurs et occasionner des pannes de courant en zone urbaine. Toute opération basée sur des données géospatiales inadéquates peut entraîner d'importantes pertes financières et entacher la bonne réputation du service à la clientèle.

La gestion de la qualité des informations spatiales peut sembler extrêmement fastidieuse. En général, les données d'une entreprise proviennent de différentes

sources. Elles ont été saisies avec des objectifs distincts, sur des périodes de temps différentes, à des fréquences irrégulières et avec des niveaux de précision variables. Ces informations sont souvent conservées dans plusieurs formats, et présentent des degrés de qualité et de complétude incertains. Leur intégration avant une prise de décision unique et éclairée constitue un défi majeur. La gestion et la maintenance des données pour un usage continu sont plus laborieuses encore.

Plutôt que de considérer la gestion de la qualité des données comme une série de tâches ponctuelles (et coûteuses), de plus en plus d'entreprises proactives adhèrent à une approche plus holistique et prônent une stratégie de vérification de la qualité des données alignée sur leurs exigences et besoins organisationnels. Ces entreprises favorisent des outils d'automatisation des procédures routinières de vérification de la qualité des données, réduisant ainsi le coût et le temps investis pour maintenir leurs informations prêtes à l'emploi.

Cette approche garantit que les données sont disponibles, à jour et précises.

Ce petit livre explique comment une vérification de la qualité des données automatisée et basée sur des règles permet aux entreprises de profiter de données plus intelligentes pour prendre des décisions plus intelligentes.

Le coût de la mauvaise qualité

Les dépenses occasionnées par des données de mauvaise qualité sont énormes (à savoir 3,1 milliard \$ de frais annuels pour l'économie des États-Unis ⁴), mais elles sont la plupart du temps occultées.

Le coût ne résulte pas seulement de la gestion des données, mais également des conséquences de l'utilisation d'informations erronées. L'effet réel de ces données inexactes peut s'avérer considérable et, parfois même, catastrophique :

- ▶ Le temps perdu par les utilisateurs pour valider et corriger manuellement les données, souvent sans justifier ces corrections ;
- ▶ Les coûts d'opportunité découlant de la perte de revenu causée par des ventes, de la promotion et des services mal orientés ;
- ▶ Les ressources gâchées suite aux décisions prises sur base de données erronées ;
- ▶ Les vies mises en danger par le tracé inexact des réseaux d'utilité publique ou la déviation de services d'urgence.

L'Administration for Children's Services, New York City, comme tant d'autres grandes agences, était aux prises avec des sources multiples de données similaires, stockées dans des formats distincts et avec des niveaux de qualité variables. Se basant sur des sources différentes, les analystes obtenaient des résultats contradictoires. Le Département estimait que ceux-ci ne passaient que 20 % de leur temps à réellement analyser les données et à créer des rapports. Le reste du temps, 80 % donc, était consacré à la recherche, à la

validation et au formatage des données afin de les rendre utilisables. ⁵

L'État du Victoria, en Australie, a réalisé que, chaque année, l'imprécision de ses données spatiales lui coûtait des millions de dollars. Pour un seul projet de développement immobilier, des erreurs géospatiales de maximum 20 mètres sur le site en question ont occasionné des retards et des travaux de remaniement coûteux estimés à 1 million AUS\$.⁶

Réseaux d'utilité publique souterrains.

Au Royaume-Uni, la Dre Nicole Metje, de la School of Engineering ⁷ de l'université de Birmingham et Directrice Adjointe de l'UKCRIC National Buried Infrastructure Facility ⁸, a réalisé une enquête mettant en lumière les coûts et les causes des dommages infligés aux réseaux d'utilité publique souterrains par les ouvriers de la construction. L'étude ⁹ a révélé les faits suivants :

- ▶ Dans 52% des incidents où les plans avaient été examinés préalablement, le réseau n'apparaissait pas sur les plans ;
- ▶ Dans 84% des cas où le réseau apparaissait sur les plans, celui-ci était représenté de façon inexacte.

Dre Metje a calculé que le coût direct pour chaque erreur oscille entre 300 £ (eau) et 2 800 £ (fibre optique). Cependant, le coût total réel par erreur (y compris les frais d'interruption) est environ 30 fois supérieur au coût direct. ¹⁰

Sans compter, évidemment, les risques pour la sécurité encourus par un ouvrier qui viendrait à toucher un câble électrique souterrain.



Différents services de l'entreprise avaient partiellement implémenté leurs propres solutions SIG ; dès lors, nous nous retrouvions avec une demi-douzaine de systèmes différents et contradictoires. Il était indispensable d'évoluer vers un seul SIG central, une seule version de la réalité.

John Daniels, Responsable de la Livraison des Données,
United Utilities plc



Données traditionnelles et incomplètes

Certains secteurs, tels que l'industrie de l'eau, doivent composer avec des données traditionnelles de piètre qualité. Généralement, ces compagnies héritent d'une base de données provenant d'une autre entreprise ou d'enregistrements historiques créés plus d'un siècle auparavant. Ces informations qui étaient déjà très souvent incomplètes et imprécises n'ont cessé de perdre de leur pertinence depuis lors.

En outre, lorsqu'une nouvelle norme gouvernementale a transféré davantage de responsabilités à des compagnies d'utilité publique privées, le patrimoine supplémentaire était assorti d'entrées risquant de polluer, ou du moins de dégrader, toute procédure de vérification de la qualité des données précédemment réalisée dans ces compagnies.

Des données de piètre qualité compliquent la tâche de localisation des problèmes et, de ce fait, portent préjudice au service à la clientèle.

De surcroît, elles entravent toute gestion efficace des actifs. Les maintenances prévisionnelles et les plans d'investissement souffrent de lacunes au niveau des connaissances relatives à l'emplacement et à la nature de ces actifs.

Silos départementaux

Nombre d'entreprises possèdent déjà une vaste quantité de données géospatiales qui traînent dans leurs bureaux, inexploitées, souvent confinées dans des silos départementaux.

Il se peut que l'entreprise ait dépensé une somme colossale pour récolter ces données en vue d'une prise de décision capitale ; mais les données ont été saisies de façon indépendante

et sont ensuite laissées à l'abandon dans un dossier compressé ou dans une base de données. Déconnectées des autres données de l'entreprise, ces informations sont souvent copiées, mais leur qualité est douteuse.

Multiples sources de données

Certaines entreprises se basent sur des mises à jour constantes d'informations provenant de sources différentes. Les Départements fédéraux des Transports (DoT) aux États-Unis en sont une bonne illustration : chaque État doit rassembler les informations relatives au réseau routier provenant de plusieurs comtés, puis y ajouter des données liées à la densité du trafic, aux limitations de vitesse, etc. Le Département des Transports de l'Arizona, par exemple, combine des données provenant de 15 comtés et de 17 centres de gestion des appels d'urgence.

Le défi est analogue pour d'autres agences, tel le Bureau du Recensement des États-Unis ou le Département de la Défense No.1 AIDU, qui fait partie du « Joint Forces Intelligence Group ». AIDU fournit cartes, graphiques et données aéronautiques aux équipes aériennes de la Défense, aux contrôleurs du trafic aérien et aux dirigeants des forces aérospatiales militaires du Royaume-Uni et de ses alliés.

De la même manière, les agences cadastrales doivent constamment mettre à jour les limites de propriété par des procédures automatisées pour conserver des données cohérentes et valides.

S'assurer que les nouvelles informations enrichissent les bases de données centrales plutôt que de les polluer, est un challenge perpétuel pour ces organismes.

Ce défi ne peut être relevé que grâce à un traitement continu de la qualité des données.

Saisir l'opportunité des données spatiales

La profusion et la variété des données spatiales mises à la disposition des entreprises ne vont cesser de croître de façon spectaculaire avec l'évolution de l'Internet des Objets. Un nombre grandissant de sources (employés, clients, articles en stock et actifs) vont produire des données géospatiales de plus en plus abondantes.

La précision des informations disponibles va aussi s'améliorer considérablement. Les avancées enregistrées chez Google Maps, par exemple, au cours des cinq dernières années ne sont rien comparées à celles qui se produiront dans les cinq ans à venir ! Les consommateurs et les entreprises profiteront bientôt des niveaux de précision en temps réel jadis réservés aux agences nationales de cartographie, qui ne diffusaient leurs données qu'avec parcimonie.

Les données sont une chose, mais la vue d'ensemble en est une autre.

Une vue d'ensemble naît de la combinaison et de l'exploitation de données fiables et officielles.

Les données ne sont utiles que si leur degré de précision est à la hauteur de l'objectif fixé. Elles ne sont précises que si elles sont gérées, et la gestion de données peut s'avérer une procédure onéreuse.

Des données de mauvaise qualité ne peuvent engendrer que des prises de décision lentes et sujettes aux erreurs. Toute vue d'ensemble potentielle est donc compromise.

Lorsque l'on combine des données d'origines différentes, leur valeur peut diminuer plutôt qu'augmenter. Une piètre qualité de données pourrait ruiner des efforts (et faire imploser des

initiatives jusque-là excellentes).

La valeur d'une vue d'ensemble géographique procède d'une base de données solide, fixe et entièrement fiable.

Dès sa concrétisation, la valeur potentielle de cette vue se voit décuplée.

D'après McKinsey & Company, un cabinet de conseil international, les données géospatiales pourraient représenter un chiffre d'affaires de plus de 100 milliards \$ pour les fournisseurs de services et générer jusqu'à 700 milliards \$ de valeur pour les utilisateurs finaux. Cependant, dans son rapport de décembre 2016, McKinsey a estimé que seulement 50 à 60 % de ce potentiel était exploité à ce jour.

McKinsey a également souligné l'émergence de nouvelles opportunités :

« Aujourd'hui, les entreprises de tous les domaines bénéficient d'un nombre grandissant de nouvelles opportunités d'utilisation des données géospatiales pour localiser les actifs, les équipes et les clients à divers endroits, afin d'acquérir de nouvelles perspectives et de gagner en efficacité. » ¹²

L'entreprise de conseil Gartner estime que 6,4 milliards « d'objets » composaient l'Internet des Objets en 2016, c'est-à-dire 30 % de plus qu'en 2015. Gartner prédit également que l'IdO permettra aux consommateurs et aux entreprises d'économiser 1 milliard \$ par année en maintenance, services et fournitures d'ici 2022.¹⁴

Effectuant une enquête pour Google, Oxera a estimé que les « services géographiques » ont permis d'économiser 1,1 milliard d'heures de déplacements et 3,5 milliards de litres d'essence chaque année. L'étude a révélé que les améliorations pour le secteur de l'agriculture ont permis une économie allant de 8 à 22 milliards \$ à l'échelle mondiale, et que les services géographiques ont généré une valeur ajoutée mondiale de 100 milliards \$ par an.¹⁵

Les données de l'organisme Ordnance Survey, réputé pour leur fiabilité, ont été objectivement évalué à 100 milliards £.¹⁶

La valeur des données géospatiales en tant qu'actifs de l'entreprise est indéniable. Et pourtant, même si leur valeur est incontestable, leur coût peut parfois sembler insurmontable. La gestion de données provenant de sources

multiples et disparates est une procédure complexe qui semble aussi onéreuse que chronophage.

Le nettoyage des données avant une prise de décision unique constitue un défi majeur ; la maintenance pour un usage continu peut paraître impossible.

Cependant, plus les entreprises se rendent compte de la valeur stratégique des données géospatiales, plus l'importance de celles-ci grandit au sein de l'organisation. De nos jours, les entreprises de premier plan considèrent les données comme une priorité, et réservent aux informations géospatiales une place de choix au cœur de cette masse d'informations.

Pour que les données géospatiales tiennent leurs promesses, leur gestion doit être fiable et économique.



Les informations [géospatiales] ne sont plus simplement un fond de plan, elles sont maintenant un élément fondamental des solutions métier de nos utilisateurs.

Colin Bray, CEO, Ordnance Survey Ireland



Gestion de la qualité des données

Adaptez vos données à votre objectif

Pour être fiables, efficaces et économiques, les données spatiales doivent être en adéquation avec leur objectif.

Pour y parvenir, vous devez tout d'abord bien comprendre le sens du mot « qualité » dans votre contexte précis.

Chez 1Spatial, notre approche est conçue afin de vous aider à découvrir et définir avec précision vos besoins, à vérifier si vos données sont conformes à ces exigences, puis à les standardiser de façon efficace et économique.

Qu'entendons-nous par « qualité des données » ?

Aucune donnée ne peut refléter fidèlement le monde réel. Les entreprises doivent donc décider du niveau de qualité qui est considéré comme recevable. Par exemple, quelles règles sont obligatoires ou facultatives, et quels sont les niveaux de conformité acceptables pour chacune.

Il est primordial de bien comprendre vos besoins. Lorsque nous travaillons avec des clients, nous les aidons généralement à déterminer les exigences exactes de leur entreprise afin de leur permettre de définir une échelle qualitative et quantitative.

La compréhension de l'adéquation des données commence par l'appréhension de l'objectif qu'elles vont servir. Quel est le niveau de qualité requis ? Quels sont les possibles manquements ? Et comment pallier ces manquements pour atteindre le niveau de qualité requis ?

Il existe deux points de références essentiels :

- ▷ La qualité requise (l'état souhaité de vos données)
- ▷ Le niveau actuel (l'état réel de vos données)

C'est seulement après avoir clairement défini ces deux stades qu'une gestion efficace de la qualité des données peut être lancée.

La qualité requise

Avant de viser une certaine excellence en matière de données, il est essentiel de bien comprendre l'objectif final. À quoi vont servir ces données ? Quel niveau de qualité est requis pour répondre à ces besoins ? En effet, une donnée adaptée à un objectif peut se révéler totalement impertinente dans un autre contexte.

Le choix de la « perfection » n'est pas non plus sans conséquence. Une gestion excessive de données a un coût.

Atteindre l'excellence dans la qualité des données, c'est disposer de données qui cadrent parfaitement avec l'objectif poursuivi.

Les décisions prises dans les premières phases sont capitales ; il est vital d'avoir une vision claire de la qualité requise pour les données. Des données de piètre qualité peuvent être dangereuses, mais une gestion excessive peut s'avérer coûteuse. Pour être menés à bien, les projets requièrent une vue à court et à moyen termes des usages qui seront faits des données.

Il peut être bénéfique d'appréhender la qualité des données dans une perspective d'utilité. Est-ce bien utile de disposer de données d'un certain niveau de qualité (à mettre en balance avec le coût généré par la tâche qui en découle) ? Cet aspect est à considérer sous plusieurs angles :



Coût

augmentation des frais et des retards causés par la piètre qualité des données, opportunités manquées ou pénalités appliquées.



Productivité

augmentation de la charge de travail et diminution de la productivité, retards, produits finis médiocres, etc.



Fiabilité

les conséquences de la mauvaise qualité des données sur les clients, fournisseurs, employés et autres acteurs clés ; doute lors des prises de décision ou prévisions basées sur de mauvaises données ; perte de la confiance dans les produits (des cartes, par exemple) réalisés sur base de données médiocres.



Risque et conformité

le risque d'amendes ou autres pénalités liées à la non-conformité aux normes en vigueur ou à l'incapacité de satisfaire aux exigences réglementaires, augmentation du risque lié aux investissements ou mauvais aperçu de la concurrence.

Les données géospatiales souffrent particulièrement des variations de précision et de complétude en matière de position, topologie, temps et thématique.

Le niveau actuel

À quel point vos données sont-elles bonnes ou mauvaises ? Étonnamment, peu d'entreprises connaissent réellement l'état actuel de leurs données géospatiales. D'après une étude menée par l'OGC (Open Geospatial Consortium) il y a quelques années, seuls 57,5 % des entreprises interrogées estimaient leurs données adaptées à l'objectif poursuivi.¹⁷

Cette même étude a également révélé que 39,6 % des entreprises participantes ne menaient aucun projet en lien avec la qualité des données géospatiales. De plus, seulement 42,5 % d'entre elles pensaient avoir les moyens de mesurer quantitativement la qualité de leurs données spatiales.

Évidemment, la qualité des données ne peut être évaluée que dans le cadre de l'objectif poursuivi.

Procédure d'amélioration des données

Chez 1Spatial, nous utilisons la méthode suivante qui nous permet d'apprécier rapidement l'état actuel de vos données. Nous désirons vous aider à comprendre à quel point vous pouvez avoir réellement confiance en vos données.

La procédure débute par un atelier d'exploration des données dans le but d'acquiescer une compréhension initiale de celles-ci, de leur structure ainsi que de la quantité de travail nécessaire. Nous invitons quelques personnes maîtrisant les données et l'usage qui leur est réservé à se joindre à l'atelier. Ensemble, nous décidons quels sont les éléments des données qui méritent une attention particulière.

La découverte des données nous renseigne sur la présence éventuelle d'un problème

de qualité. Ensuite, nous en étudions les conséquences pour votre entreprise, et décidons ensemble s'il y a lieu d'améliorer vos données. Dans ce cas, nous passons aux étapes suivantes de la procédure :

1. L'atelier consacré aux critères d'amélioration des données
2. Le Sprint
3. La Rétrospective

Les workshops, ou ateliers, qui sont normalement organisés en début de processus, rassemblent les acteurs clefs de votre entreprise. Ces sessions permettent de définir les tâches ou ensembles de tâches à réaliser afin d'explorer les données de manière plus approfondie, de résoudre les problèmes de qualité, d'établir une liste des priorités et de cerner le travail restant à réaliser.

Les tâches sont alors assignées à un « Sprint », qui s'étend habituellement sur une période de une à deux semaines.

Lors de chaque Sprint, nous travaillons sur les tâches assignées, testons les résultats et réalisons des améliorations mesurables.

Quand les changements requis ont été effectués et les résultats testés, nous passons au stade de la Rétrospective. À cette étape, nous faisons un compte rendu pour tous, nous tirons des leçons de notre expérience et décidons des actions qui permettront l'amélioration continue lors de la prochaine itération.

C'est aussi à ce stade que nous commençons à établir la liste des priorités pour le travail de l'atelier qui suivra, qui, avec les retours des différents intervenants, constitue la base du prochain Sprint.

Chaque Sprint entraîne habituellement une nouvelle amélioration de la qualité des données.

Le processus dans son entièreté est un effort collaboratif entre les experts de votre équipe et les nôtres.

Nous adoptons une méthodologie de consultance « Agile » pour ce travail sur les données, qui s'est développée grâce à l'expérience acquise dans des projets similaires. Cette méthode renforce la collaboration et fournit un feedback aussi rapide que possible tout en éliminant les pertes de temps.

Nous avons acquis une expertise précieuse pour aider nos clients à définir et à exécuter leur stratégie relative aux données. Nous leur permettons de cerner leurs objectifs en matière de qualité des données ; nous auditions et évaluons ensuite leur adéquation actuelle par rapport aux objectifs définis. À l'aide de nos outils de gestion des données, nous appliquons des algorithmes avancés et des techniques de prédiction pour améliorer la qualité de vos données et vous permettre d'atteindre ces objectifs.

Atelier

Retours des utilisations finaux, des clients, de l'équipe et des autres acteurs clés



Backlog

Sprint

Sprint de données



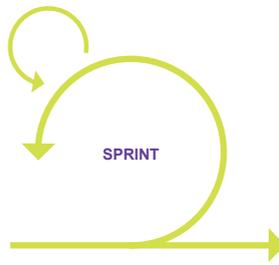
Backlog



Réunion de définition du sprint



Sprint Backlog



Rétrospective



Amélioration des données
Potentiellement livrables

Six principes d'excellence.

01. Automatiser

Favoriser l'automatisation quand c'est possible

02. Fiabiliser

Assurer la répétabilité et la traçabilité

03. Concevoir

Elaborer des solutions simples pour les scénarios difficiles et éviter toute complexité technique inutile

04. Cibler

Cibler ce qui est habituel, et non ce qui est exceptionnel, afin de maximiser la valeur

05. Expérimenter

Opter pour un processus de prise de décision fondé sur des preuves afin d'inspirer la confiance des entreprises quant au résultat

06. Collaborer

Echanger avec les experts pour identifier les problématiques et trouver les solutions

Outils de gestion de données

Les entreprises qui possèdent de vastes bases de données géospatiales déploient toute une variété d'outils qui utilisent l'automatisation basée sur des règles pour garantir la conformité des données aux standards de qualité en vigueur, et ce de façon économique et rapide. Ces outils basés sur des règles permettent d'extraire des sous-ensembles de données géospatiales, de les mettre à jour, de les valider, puis de les réintégrer à la base de données sans interrompre la production.

Ces systèmes sont parfaits pour la réalisation d'itérations rapides afin de tester des idées et des hypothèses. Ainsi, les équipes peuvent rapidement déceler les problèmes qui doivent être résolus. De tels systèmes garantissent une gestion de données économique et rapide, de telle sorte que le jeu de données principal soit toujours aussi fiable et précis que possible pour tous ses utilisateurs.

Pour être efficaces, il est capital que ces outils interagissent harmonieusement avec d'autres SIG ou systèmes d'information géospatiale également exploités au sein de l'entreprise, tel le système ArcGIS d'Esri ou Elyx de 1Spatial.

1Spatial s'engage à soutenir les normes en matière de données géospatiales et s'efforce de travailler de façon harmonieuse avec les principaux Systèmes d'Information Géographique.

Nous sommes également les partenaires de grands groupes technologiques tels qu'Oracle, Esri, HERE, SAP, Safe Software et Latitude Geographics.

Principes d'excellence des données

Les procédures de mise en adéquation des données peuvent s'avérer coûteuses et chronophages, d'autant plus lorsqu'elles sont appliquées en l'absence d'objectif clairement établi. Trop d'initiatives sont planifiées et gérées comme si elles étaient des exercices ponctuels. Et nombre d'entre elles échouent parce qu'elles ciblent les problèmes les plus complexes plutôt que de viser le retour le plus important.

Les projets performants sont gérés conformément à six principes d'excellence des données :

1. Automatiser
2. Assurer la répétabilité et la traçabilité
3. Élaborer des solutions simples pour les scénarios difficiles et éviter toute complexité technique inutile
4. Cibler ce qui est habituel, et non ce qui est exceptionnel, afin de maximiser la valeur
5. Opter pour un processus de prise de décision fondé sur des preuves afin d'inspirer la confiance des entreprises quant au résultat
6. Collaborer pour identifier les problématiques et trouver les solutions

Gestion des données

La procédure d'amélioration de qualité des données ne peut être considérée comme un exercice ponctuel. En effet, lorsqu'un jeu de données est entretenu en vue d'un usage régulier, les données gagnent en valeur et le retour sur investissement s'accroît.

Les compagnies qui tirent le plus de profit de leurs données géospatiales possèdent une stratégie clairement définie et reconnaissent l'importance de la gestion des données : la gestion et l'entretien de certains jeux de données à un degré de qualité défini pour un objectif fixé.

Une bonne gestion de données utilise des approches telles que l'amélioration des données et les principes d'excellence des données pour parvenir à une gestion efficace et économique de ces données. La gestion de données définit le degré de qualité requis et exploite ensuite des procédures fiables pour garantir la correction de tout problème à la source afin que les erreurs ne réapparaissent pas à chaque mise à jour.

Cette procédure soulève les problèmes de propriété des données et permet de définir des standards. Le gestionnaire sélectionne également des outils capables d'interagir harmonieusement, réduisant de ce fait l'éventualité d'un travail supplémentaire et le risque d'erreurs manuelles lorsque les données sont transférées d'un système à un autre.

Question de qualité de données

Chez 1Spatial, nous rencontrons généralement des entreprises qui organisent la qualité de leurs données autour de trois points essentiels :

- ▷ La validation des données
- ▷ Le nettoyage des données
- ▷ L'enrichissement des données

Validation des données

La validation des données (vérification de la conformité des données existantes et des nouvelles données entrantes aux exigences en

vigueur) peut être une vraie corvée. Comment gardez-vous vos données à jour sans une vérification régulière ? Comment validez-vous vos données mises à jour ? Enfin, comment préservez-vous la fiabilité et l'accessibilité de vos données de base en tout temps ?

Notre approche automatisée et basée sur des règles valide les données au moment de la saisie, sur le terrain via un appareil mobile, ou avant leur entrée dans votre base de données.

Elle empêche les données non pertinentes de polluer votre base et d'infléchir négativement vos décisions.

Ayez confiance en vos données

Les organisations comptent de plus en plus sur les données spatiales et s'y réfèrent souvent en temps réel lors de la prise de décision, tant en matière d'itinéraires de livraison que de projets d'investissements majeurs.

Lorsque votre entreprise dépend de la fiabilité de vos données, vous avez besoin d'un système capable de garantir, de façon économique et rapide, la conformité des données qui entrent dans votre base par rapport à vos exigences. Sont-elles précises, cohérentes, exactes, actuelles et complètes ?

Chez 1Spatial, notre approche permet à l'utilisateur de définir et de gérer les règles qui testent toutes les données. Ces règles sont reprises dans un entrepôt unique, central et technologiquement neutre, et peuvent être appliquées à toute nouvelle donnée sur demande. Les règles peuvent aussi être partagées entre différents systèmes, différents individus ou différentes entreprises.

Elles peuvent être exécutées en arrière-plan, tandis qu'un géomètre recueille de nouvelles

informations sur le terrain, et signaler la non-conformité d'une donnée face aux exigences afin de permettre une vérification ponctuelle, et d'éviter des vérifications coûteuses.

Les règles peuvent également être appliquées sur des jeux de données lors de leur saisie, avant qu'elles ne soient intégrées à vos bases de données.

Par ailleurs, les données peuvent être corrigées automatiquement, à la volée, sur

base de ces mêmes règles, afin d'intégrer rapidement les données exactes et de signaler les exceptions en vue d'une correction manuelle.

Notre système automatisé, global et multiplateforme, protège votre base de données de façon efficace, économique et rapide, afin de rendre vos données plus intelligentes.





En utilisant la solution Cloud de 1Spatial... nous avons pu réduire d'au moins un quart le temps nécessaire au contrôle de la qualité du projet, et ceci principalement grâce à la flexibilité de ce programme, à sa facilité d'utilisation, à la vitesse avec laquelle l'équipe a pu être mise en place et rendue opérationnelle, et du fait que le service était toujours disponible en ligne.

Jimena Martinez, Project Manager, Sinfogeo



Exemple : Sinfgeo

Entreprise de consultance en SIG, Sinfgeo a remporté un contrat portant sur la production d'une nouvelle série de cartes reprenant la totalité du territoire espagnol. Dans le cadre de ce contrat, la société a dû consolider, valider, nettoyer et harmoniser des données récoltées par le gouvernement et l'armée espagnols.

Un des principaux défis était de garantir la qualité et la précision des données tout au long du projet. Avec des données provenant des quatre coins du pays, ce programme s'annonçait coûteux et fastidieux.

Cependant, par l'utilisation de la plateforme 1Spatial Cloud, Sinfgeo a pu automatiser les procédures de validation et de nettoyage des données grâce à une approche basée sur des règles. Jimena Martinez, Chef de Projet chez Sinfgeo, l'explique en ces termes :

« Nous savions que nous avions besoin de règles pour mener à bien notre projet, mais nous ne savions pas comment les élaborer. La collaboration avec les équipes de 1Spatial a été très fructueuse. Ensemble, nous avons rapidement créé et testé toutes les règles dont le gouvernement et l'armée avaient besoin. Notre société a ainsi échappé à l'obligation d'engager et de rémunérer du personnel afin de développer de nouveaux concepts de notre propre côté. »

Dès la finalisation des règles de validation, celles-ci ont simplement été téléchargées dans le cloud afin que tous les contributeurs puissent y accéder.

En utilisant des règles parfaitement identiques, Sinfgeo pouvait garantir que la qualité de toutes les données soumises était contrôlée et que ces dernières correspondaient aux

mêmes normes. Les contributeurs soumettaient simplement leurs données à un service unique et recevaient un rapport de validation en quelques minutes seulement. Ce rapport leur donnait un aperçu général de la qualité de leurs données et localisait précisément toute erreur, de sorte qu'il était ensuite facile et rapide de corriger les irrégularités avant de soumettre la version finale à Sinfgeo.¹⁸

Exemple : Département des Transports de l'Arizona

Aux États-Unis, le Département des Transports de chaque État s'est vu confier la responsabilité de regrouper les données relatives au réseau routier qui étaient compilées séparément dans les différents comtés, avant de les transférer à la Federal Highway Administration (agence du Département des Transports des États-Unis). Généralement, les nombreux comtés stockent leurs informations dans des systèmes variés, avec des niveaux de précision incertains et des dates de saisie différentes. La consolidation et la validation de ces données représentaient donc un projet coûteux et chronophage.

Le Département des Transports de l'Arizona a donc acquis la solution 1Integrate développée par 1Spatial pour valider le réseau routier de tout l'État. Désormais, le Département validera et intégrera ses propres informations sur base des mises à jour régulières reçues de plusieurs autres organismes gouvernementaux.

La technologie 1Integrate de 1Spatial permet au Département des Transports de garantir automatiquement la qualité des informations reçues des 15 comtés et 17 centres de gestion des appels d'urgence (PSAP) de l'État en les validant grâce à un ensemble de règles métier prédéfinies. De plus, les données nouvellement collectées s'agrégeront directement pour être compatibles avec la version actuelle du réseau routier afin d'identifier et d'appliquer les changements de géométrie et d'attributs. En passant d'un processus manuel à un processus automatique maîtrisé, une quantité importante de temps et d'argent est économisée et la précision des données sur la carte à l'échelle de l'État s'en trouve améliorée.

Nettoyage des données

La compréhension de l'état de vos données est importante, mais la correction est capitale en vue d'une amélioration. Des données de piètre qualité peuvent entraîner de mauvaises décisions opérationnelles, et leur rectification nécessite une intervention souvent manuelle, coûteuse et chronophage.

La technologie 1Spatial dresse un portrait de l'état actuel de vos données, mais elle peut également les corriger. Mieux encore, nous pouvons mettre sur pied une procédure d'amélioration automatisée et continue afin de nettoyer vos données, de les actualiser régulièrement et de les garder fiables en permanence.

Automatisez des procédures traditionnellement manuelles, complexes et chronophages

Nous travaillons de concert pour définir le niveau de qualité de données dont vous avez besoin pour atteindre vos objectifs. Ensuite, nous vous aidons à développer des règles définies et gérées par les utilisateurs afin d'identifier et de résoudre les problèmes de qualité.

Une fois créées, ces règles seront appliquées à vos données afin de corriger toutes les erreurs fréquentes et d'identifier les exceptions plus complexes qui requièrent une intervention manuelle.

Conservées dans un entrepôt unique et central, les règles deviennent alors une procédure automatique et reproductible capable de nettoyer et de corriger toute donnée nouvellement saisie pour éviter toute pollution de votre base par des données inexacts. Les niveaux de qualité peuvent être enregistrés au fil du temps, afin de produire des indicateurs importants et de permettre des mesures d'amélioration continue.

Nécessitant moins de temps pour obtenir des données utilisables et rencontrant plus rapidement l'approbation de l'utilisateur, la technologie 1Spatial rend vos données plus propres et plus intelligentes.

Exemple : Bureau du Recensement des États-Unis

Le Bureau du Recensement des États-Unis (US Census Bureau) réalise tous les dix ans la plus importante activité civile de la nation. Le projet comprend le comptage et le profilage d'une population de plus de 318 millions d'individus, ainsi que la vérification de 135 millions d'adresses. Il permet d'offrir une représentation politique précise et correcte à la Chambre des représentants, ainsi que d'allouer des fonds fédéraux équivalant à 400 milliards \$ par an.

Pour créer la plateforme géographique de chaque recensement, le Bureau doit combiner des données provenant de 3 200 comtés et autres organisations. L'identification des changements, l'intégration des nouvelles données et la maintenance des bonnes relations entre les différents jeux de données représentent un défi majeur pour le Bureau.

Comme l'explique Tim Trainor, Chef de la Division de Géographie du Bureau, « L'intégration de données était une procédure essentiellement manuelle. La gestion d'une telle abondance de données était extrêmement chronophage. »

Lors du recensement de 2010, le Bureau avait engagé 140 000 personnes pour parcourir à pied ou en voiture toutes les rues du pays afin de valider les rapports compilant les adresses. Plus tard, il avait encore fait appel à 600 000 individus pour effectuer un suivi auprès des non-répondants et identifier les logements

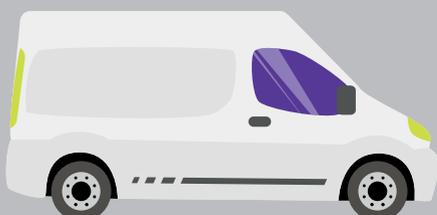
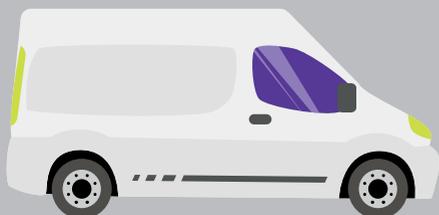


140 000



140 000 personnes embauchées pour se déplacer à pied ou en voiture pour valider les enregistrements d'adresse du bureau





25%

Seulement 25% des 140 000 agents sur le terrain seraient nécessaires par rapport à 2010



vides. Globalement, le recensement avait coûté 12 milliards \$ en 2010. Si le Bureau ne changeait pas de méthode, le coût de celui de 2020 aurait atteint 17 milliards \$.

1Spatial et le Bureau ont travaillé de concert dans le but d'automatiser la procédure de gestion de données basée sur des règles définies et gérées par l'utilisateur. Désormais, un processus automatisé de combinaison des données se charge de la validation et de l'intégration des nouvelles entrées provenant des 3 200 partenaires.

Ce haut degré d'automatisation permet au Bureau de traiter plus rapidement plus de dossiers des partenaires, ainsi que d'actualiser et d'affiner sa base de données centrale.

Grâce à cette approche, le recensement des États-Unis de 2020 requerra moins d'agents sur le terrain. Trainor estime que seulement 25 % des 140 000 agents sur le terrain en 2010 seront nécessaires. Cette capacité à intégrer des données provenant d'autres agences telles que la US Postal Service, l'agence postale des États-Unis, a également permis de réduire les coûts.

« Nous attendons une économie [...] d'un peu plus de 5 milliards \$ », explique Trainor. « C'est-à-dire que le coût sera presque équivalent à celui du recensement de 2010. C'est une belle réussite. »¹⁹



Nous estimons que les économies réalisées grâce à ces quatre innovations s'élèveront à un peu plus de 5 milliards \$.

Tim Trainor, Geography Chef de la Division de Géographie du Bureau, US Census Bureau





La différence de coût est assez significative. Au départ, notre idée était de cartographier physiquement un tiers du réseau transféré au cours d'une période de gestion des équipements de cinq ans. Cette approche nous aurait coûté 10 millions £. La collaboration avec 1Spatial nous a permis de créer une carte de l'ensemble du réseau transféré en à peine deux ans, pour 1,25 million £.

Mike Madine, Directeur des Réseaux d'Assainissement et des Services de Développement, Northumbrian Water



Enrichissement des données

L'enrichissement des données, par la combinaison du meilleur des différentes bases de données et la création de nouveaux jeux de données, vous permet de retirer un bénéfice plus important des investissements en données existants. Vous pouvez également utiliser des données publiques ou provenant d'un tiers dans le but d'augmenter votre base d'informations et ainsi d'obtenir plus rapidement des données fiables et solides.

Vos données sont-elles prêtes à l'emploi ?

Garder vos données prêtes à l'emploi peut être un défi de tous les instants : les exigences de l'entreprise évoluent, les standards de données changent et, bien sûr, les données perdent en pertinence avec le temps.

Une stratégie d'enrichissement des données vous donne la garantie de tirer le meilleur profit de vos données existantes, afin qu'elles soient toujours à jour et précises.

Il se peut que les adresses de vos clients encodées dans votre système de facturation soient les plus exactes de votre entreprise, mais votre base de données des actifs détient la meilleure vue d'ensemble de votre réseau. En adoptant une approche automatisée et basée sur des règles, vous pouvez combiner le meilleur des deux ensembles de données, même si ceux-ci se présentent sous différents formats, dans des systèmes distincts et dans plusieurs silos. De plus, vous pouvez pallier tout manquement grâce à des données achetées à un fournisseur tiers.

Nos solutions vous aident à réduire le rapport « valeur-temps » grâce à l'obtention et au maintien de données utilisables et fiables en un minimum de temps et à moindre coût.

Exemple : Northumbrian Water

La société Northumbrian Water (NWL) a endossé la responsabilité d'environ 13 500 km de canalisations et égouts lors du changement de la loi, qui en transférait la propriété des entités privées aux compagnies des eaux locales. Cependant, seulement 5 % du réseau transféré était cartographié.

Le plan initial de NWL était d'arpenter manuellement un tiers du réseau pendant la période de planification et de gestion des actifs. Les connaissances ainsi accumulées devaient alors être appliquées aux tronçons non cartographiés du réseau. Cependant, le coût de ce projet avait été estimé à 10 millions £, et sa mise en œuvre devait s'étaler sur cinq ans.

La société avait donc besoin d'une solution plus rapide et moins coûteuse.

1Spatial a proposé une approche basée sur sa solution 1Integrate. Nous avons collaboré avec des experts en la matière chez NWL, afin de développer une série de règles. L'objectif était d'inférer l'information manquante à partir des données partielles accessibles et de l'expertise de l'équipe de NWL.

Cette approche innovante et itérative n'a coûté que 1,25 million £ pour cartographier l'ensemble du réseau. Cela représente une économie de 8,75 millions £ (c.-à-d. 87 %) par rapport au budget des cinq premières années seulement. Le projet a également été mis en œuvre en deux ans, soit 1/8 du temps prévu initialement.

NWL a donc acquis plus rapidement l'information nécessaire à un meilleur service clientèle et à l'élaboration de programmes plus efficaces de gestion et d'investissements.²⁰

L'approche 1Spatial

1Spatial le leader mondial en matière de gestion des données géospatiales.

En tant qu'intégrateur de systèmes géospatiaux, nous collaborons avec les utilisateurs et les créateurs de bases de données géospatiales parmi les plus grandes au monde.

Nous construisons des partenariats à long terme avec nos clients et les aidons à gérer leurs données tout au long de leur cycle de vie, c'est-à-dire à collecter, stocker, gérer et interpréter les informations géospatiales.

En rendant les informations actuelles, automatiques, complètes et cohérentes, nous en faisons des données intelligentes. Nous veillons à ce que vos décisions soient toujours basées sur des informations de qualité supérieure.

Notre approche unique, basée sur des règles, offre, à l'échelle de l'entreprise, une automatisation multiplateforme de toutes les étapes du cycle de vie des données. La fiabilité des données s'en trouve renforcée, tandis que le temps et le coût d'entretien sont réduits.

Nous collaborons avec des entreprises du monde entier, dont des agences nationales de cartographie et du cadastre, des compagnies d'utilité publique, des Départements de la Défense et autres organismes gouvernementaux de tous niveaux. Parmi nos clients, nous comptons Ordnance Survey, l'Environment Agency, le Bureau

de Recensement des États-Unis, l'armée brésilienne, United Utilities plc, le groupe Northumbrian Water et le gouvernement du Queensland.

Chef de file dans notre domaine et présents dans le monde entier, nous bénéficions de plus de 45 ans d'expérience, d'innovation permanente et de développement constant. 1Spatial s'engage à respecter les normes en vigueur en matière de données géospatiales et à créer des outils qui interagissent harmonieusement avec les principaux SIG.

Nous sommes également les partenaires des grands fournisseurs de technologie dans le domaine tels qu'Oracle, Esri, HERE et SAP.

Aujourd'hui, le monde étant de plus en plus dépendant de données spatiales précises, l'expertise de 1Spatial n'a jamais été aussi indispensable.

1Spatial doit son succès à la fiabilité constante de ses conseils prodigués aux entreprises responsables de big data géospatial. Nous œuvrons à vos côtés, en partenaires, afin d'atteindre des objectifs sur le long terme.

Notre but est, simplement, de rendre vos données plus intelligentes.

Pour plus d'informations sur notre capacité à rendre vos données plus intelligentes, rendez-vous sur notre site web www.1spatial.com.



Ingénierie



Réseaux d'utilité publique



Agences Nationales de
Cartographie



Transport & Infrastructure



Gestion territoriale



Gouvernement



Conclusion

À données plus intelligentes,
monde plus intelligent.

Les données géospatiales sont de plus en plus au cœur du processus décisionnel des entreprises et des particuliers. Comme le dit ce vieil adage populaire dans notre secteur, « tout arrive quelque part », et les informations spatiales sont souvent l'unique point commun de jeux de données disparates.

À l'échelle mondiale, la valeur ajoutée des données géospatiales est estimée à 100 milliards \$ par an. Plus les entreprises se rendent compte de la valeur stratégique de ces données, plus l'importance de celles-ci grandit au sein de l'organisation. De nos jours, les entreprises de premier plan considèrent ces informations comme une priorité.

Cependant, la valeur potentielle des données géospatiales dépend entièrement de leur qualité. Les frais et les conséquences procédant de données médiocres peuvent s'avérer énormes, et parfois même catastrophiques : retards, pertes de revenu, mauvaises décisions et vies en danger.

La qualité des données est d'importance capitale, et les entreprises considèrent de plus en plus la maintenance de données comme une procédure continue plutôt que comme un projet ponctuel. Ces entreprises déploient des solutions capables d'automatiser les procédures de gestion et de vérification de l'adéquation des données, afin de :

- ▷ Réduire considérablement le coût de la qualité ;
- ▷ Gagner beaucoup de temps ;

- ▷ Permettre aux experts à l'interne de se concentrer sur le développement et l'innovation ;
- ▷ Diminuer le risque d'erreurs manuelles.

1Spatial bénéficie de 45 ans d'expérience avec les plus vastes bases de données géospatiales au monde. Notre approche unique et basée sur des règles offre une automatisation multiplateforme de toutes les étapes du cycle de vie des données, à l'échelle de l'entreprise. La fiabilité des données s'en trouve renforcée, tandis que le temps et le coût de maintenance sont réduits.

Nous collaborons avec des agences nationales de cartographie, des compagnies d'utilité publique, des ministères de la Défense, ainsi que des gouvernements fédéraux, régionaux et municipaux, afin de les aider à récolter, stocker, gérer et interpréter des données géospatiales. Notre approche ouverte d'excellence en matière d'information garantit que nos outils, procédures et données fonctionnent parfaitement avec des systèmes provenant d'autres acteurs majeurs du secteur.

1Spatial compte des clients dans le monde entier, notamment Ordnance Survey, l'Environment Agency, le Bureau de Recensement des États-Unis, l'armée brésilienne, United Utilities plc, le groupe Northumbrian Water et le gouvernement du Queensland.

Pour plus d'informations sur notre capacité à rendre les données plus intelligentes pour un monde plus intelligent, rendez-vous sur notre site web www.1spatial.com.

Références bibliographiques

¹McKinsey Global Institute, The Age of Analytics: Competing in a Data-Driven World (2016), <http://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world>

²IBM, Extracting business value from the 4 V's of big data (2016), <http://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data>

³Citation tirée du cas pratique réalisé par 1Spatial. Pour lire ce cas pratique, rendez-vous sur : <https://1spatial.com/customers/united-utilities/>

⁴IBM, Extracting business value from the 4 V's of big data (2016), <http://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data>

⁵KPMG, Better data, better government (2016), <https://assets.kpmg.com/content/dam/kpmg/pdf/2016/06/co-gv-6-better-data,-better-government.pdf>

⁶CIO.com.au, Victoria in need of a more accurate map database (2013), <http://www.cio.com.au/mediareleases/17192/victoria-in-need-of-a-more-accurate-map-database/>

⁷Voir <http://www.birmingham.ac.uk/staff/profiles/civil/metje-nicole.aspx>

⁸Voir <http://www.birmingham.ac.uk/research/activity/ukcric/national-buried-infrastructure-facility.aspx>

⁹Metje, N, Ahmad, B & Crossland, SM 2015, 'Causes, impacts and costs of strikes on buried utility assets', Institution of Civil Engineers. Compte rendu. Municipal Engineer, vol 168, no 3, pp. 165-174. DOI :10.1680/muen.14.00035, http://pure-oai.bham.ac.uk/ws/files/24091427/Metje_Ahmad_Crossland_2015_Causes_impacts_costs_ICE_Proceedings.pdf

¹⁰Between the Poles, New research on the cost of hitting underground utilities (2016), <http://geospatial.blogs.com/geospatial/2016/11/new-research-on-the-cost-of-hitting-underground-utilities-in-the-uk.html>

¹¹McKinsey (2016)

¹²McKinsey (2016), p.31

¹³Gartner, Gartner Says 6.4 Billion Connected "Things" Will Be In Use in 2016 (2015), <http://www.gartner.com/newsroom/id/3165317>

¹⁴Gartner, Top Strategic Predictions for 2017 and Beyond (2016), <https://www.gartner.com/doc/3471568>

¹⁵Oxera (pour Google), What is the economic value of Geo services ? (2013), http://www.oxera.com/Oxera/media/Oxera/downloads/reports/What-is-the-economic-impact-of-Geo-services_1.pdf

¹⁶Voir http://1spatial.com/wp-content/uploads/2012/10/1Spatial_CaseStudy_OrdnanceSurvey1.pdf

¹⁷Open Geospatial Consortium Data Quality Working Group, Geospatial Data Quality Survey (2008). Enquête réalisée sur base de 750 réponses provenant d'entreprises de toutes tailles réparties dans 107 pays différents. Q14 : Les données de votre entreprise sont-elles utilisées aux fins prévues par votre utilisateur final ?

¹⁸Pour plus d'informations sur l'approche de Sinfogeo, rendez-vous sur le cas pratique réalisé par 1Spatial sur https://1spatial.com/fr/wp-content/uploads/sites/2/2017/02/SINFOGEO_CS_FR.pdf

¹⁹Pour plus d'informations sur le Bureau de recensement des États-Unis, rendez-vous sur le cas pratique réalisé par 1Spatial sur <https://1spatial.com/fr/client/u-s-census-bureau/>



