



HAL
open science

Modélisation générique de données capteurs hétérogènes : De la captation de données à la compréhension de phénomènes

Sylvie Servigne, Yann Gripay, Jean-Michel Deleuil, Jacques Jay, Radouane
Mebrouk

► To cite this version:

Sylvie Servigne, Yann Gripay, Jean-Michel Deleuil, Jacques Jay, Radouane Mebrouk. Modélisation générique de données capteurs hétérogènes : De la captation de données à la compréhension de phénomènes. SAGEO'2014: Conférence Internationale en Géomatique et Analyse Spatiale, Nov 2014, Grenoble -, France. pp.1-14. hal-01313189

HAL Id: hal-01313189

<https://hal.science/hal-01313189>

Submitted on 12 Sep 2017

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.

Modélisation générique de données capteurs hétérogènes

De la captation de données à la compréhension de phénomènes

Servigne Sylvie¹, Gripay Yann¹, Deleuil Jean-Michel ², Jay Jacques³, Mebrouk Radouane¹

1. Université de Lyon, CNRS, INSA-Lyon, LIRIS, UMR5205, Labex IMU
Laboratoire d'Informatique en Image et Systèmes d'Information
prenom.nom@insa-lyon.fr
2. Université de Lyon, CNRS, INSA-Lyon, EVS, UMR5600, Labex IMU
Laboratoire Environnement, Ville et Société
prenom.nom@insa-lyon.fr
3. Université de Lyon, CNRS, INSA-Lyon, CETHIL, UMR500
Centre de Thermique de Lyon
prenom.nom@insa-lyon.fr

RESUME. Un modèle générique capteur données-centré est présenté. Il est le résultat d'une fertilisation croisée pluridisciplinaire en Science des Données : Informatique, Thermique, urbanisme et sociologie. L'objectif est d'offrir dans un premier temps une modélisation homogène des systèmes de captation hétérogène afin de permettre l'exploration de données multisources et multidimensionnelles facilitant ainsi la compréhension de phénomènes complexes liés à des objets d'études instrumentés. Une expérimentation a été réalisée in vivo dans le cadre du bâtiment intelligent notamment pour comprendre les phénomènes énergétiques de l'habitat en intégrant des mesures physiques réelles et des données empiriques issues d'enquêtes usager.

MOTS-CLES : données capteurs – BD – enquête - données spatio-temporelles – modèle – bâtiment intelligent

1. Introduction

La ville est de plus en plus instrumentée et l'information circulera partout et conditionnera la qualité de la vie et la qualité de la ville. L'objectif de cette instrumentation peut être l'optimisation de la consommation énergétique, le monitoring urbain (gestion de l'eau, de la circulation ou suivi de trajectoire (Mefeth *et al* 2012), de la voirie,...), la maîtrise de risques naturels (inondations, incendies...) ou écologiques, technologiques ou urbains (Davoine *et al.* 2013, Servigne *et al* 2009), l'amélioration de la qualité de vie. Dans ce contexte, des masses de données sont ou vont être disponibles qu'elles soient sous formes de flux, structurées dans des bases de données ou moins structurées au sein de nombreux corpus. Ces données multisources et multiformats peuvent être issues de capteurs matériels (prises de mesures, de photos, de films, de sons,...), d'enquêtes terrains (questionnaires, sondages, inventaires...), d'analyses statistiques mais aussi de réseaux sociaux ou plus communément d'annuaires ou d'archives. Une approche originale est de considérer tous ces systèmes comme des « capteurs », au sens captation d'information.

Par ailleurs, la compréhension de phénomènes environnementaux ou urbains complexes nous confronte aujourd'hui aux limites des modélisations et méthodes disciplinaires nous conduisant à renforcer le dialogue aux interfaces des disciplines jusqu'à concevoir des approches pleinement pluridisciplinaires notamment en Science des Données. Prenons un exemple dans le cadre des écotecnologies : l'évaluation de l'efficacité énergétique des bâtiments est actuellement basée sur la consommation énergétique estimée en amont au travers de modèles physiques prenant en compte les matériaux, les équipements et le type d'occupation envisagé. Selon les thermiciens, force est de constater que l'occupation effective des bâtiments conduit à une grande variabilité de la consommation énergétique et globalement à une baisse significative des performances annoncées du bâtiment. Pourtant, la performance énergétique réelle des bâtiments n'est toujours pas effective en France (Le Moniteur 2014). Il s'agit alors de comprendre l'impact des usages sur l'efficacité énergétique et ensuite de définir de nouveaux modèles d'évaluation à partir de données issues de **capteurs physiques** et de données recueillies auprès d'utilisateurs via des **enquêtes**.

Un de nos premiers objectifs est d'offrir une représentation générique et homogène des données permettant de confronter ces données multisources et pluridisciplinaires donc hétérogènes et multidimensionnelles. Notre hypothèse est qu'un modèle homogène et générique facilitera ainsi des analyses métier pluridisciplinaires au sein d'un outil mutualisé. De cette approche « modèle de données générique » découle un schéma de données opérationnel, intégrant des données statiques correspondant à la description du système et des données dynamiques correspondant à la production du système. Notre approche vise une mise en œuvre exploitable basée sur des standards. Notre objectif est également d'offrir des fonctionnalités interactives à des experts scientifiques de différentes

disciplines et de permettre une communication ou intégration facilitées avec leurs propres outils et « data-process » d'exploration, d'analyse, etc.

D'autre part, parmi les écueils actuels de la recherche en science des données, l'un concerne le temps incontournable mais assurément indispensable à investir dans « l'appropriation mutuel » et compréhension des concepts et démarches de chacune des autres disciplines ; une autre difficulté est de disposer de jeux de données réels avec des questionnements identifiés sur ces données. En effet, si beaucoup de données comme les mesures capteur deviennent disponibles au travers de l'Open Data, des informations sur les systèmes et processus de collecte sont rarement disponibles ou souvent incomplètes. C'est pourquoi notre consortium pluridisciplinaire a élaboré une plateforme capteurs in vivo sur un bâtiment tertiaire afin de maîtriser le processus de collecte et d'exploitation des données ainsi que les objets d'étude et le type de phénomène à observer. Dans cet article, nous présentons le modèle capteur générique VGS (Virtual Generic Sensor) et son expérimentation dans le contexte de monitoring d'un bâtiment d'enseignement et de recherche en exploitant de façon homogène des données mesures et des données empiriques d'enquêtes usagers.

2. Etat de l'art

2.1 Autour des capteurs

La modélisation autour des capteurs a récemment engendré nombre de travaux de recherche. Plusieurs ontologies capteurs ont été proposées depuis 2005 (table 1). Elles diffèrent par leur objectif de représentation en étant souvent orientées Observation ou Réseaux de capteurs ou Capteur physique sur des domaines d'application ciblés.

<i>Nom</i>	<i>Orientation</i>	<i>Cible</i>	<i>Année</i>
SSN	Capteurs/Obs	Modélisation générique capteurs (Compton <i>et al.</i> 2011)	2011
DUL	Concepts	Concepts génériques capteurs (Gangemi 2010)	2010
CESN	Réseau	Réseaux côtiers (Calder <i>et al.</i> 2010)	
CSIRO	Observation	Description sémantique (Neuhaus <i>et al.</i> 2009)	2009
Stimuli-centered	Observation	Description de processus (Stasch <i>et al.</i> 2009)	
OOSThesis	Observation	Milieu marin (Bermudez <i>et al.</i> 2009)	
A3ME	Réseau	Aspects techniques et contraintes (Herzog <i>et al.</i> 2008)	2008
MMI	Observation	Domaine marin (MMI 2008)	

SensorML	Observation	Description de processus (SensorML 2007)	2007
SDO	Réseau	Recherche dans des réseaux de capteurs (Eid et al. 2007)	
WISNO	Réseau	Réseau sans fil (Hu et al. 2007)	
SEEK OBOE	Observation	Observations scientifiques (Madina et al. 2007)	
OntoSensor	Capteurs	Spécifications techniques (Russomanno et al. 2005)	2005

Table 1. Une synthèse des ontologies capteurs

Dans un objectif de standardisation, le W3C Incubator Group (SSN XG3), basé sur SensorML (2007), Sensor Web Enablement (SWE) standards (Reed *et al.* 2005) et Dolce Ultra Lite (Gangemi 2010), propose Semantic Sensor Network ontology (Compton *et al.* 2011) en 2011. SSN est une ontologie qui permet de décrire les capteurs et les réseaux de capteurs. Elle propose un ensemble de modèles (Stimulus-Sensor-Observation Pattern, Figure 1) pour représenter les principaux concepts liés au contexte des capteurs et définir les relations entre ces concepts. Comparé à SensorML, SSN propose trois modèles "Process", "Skeleton" et "Device" pour séparer l'aspect matériel et processus dans un capteur. Pour représenter les mesures, SSN définit les "Data" appelées valeur d'observation "ObservationValue". SSN offre ainsi une ontologie adaptée à la description d'un système de captage dont nous nous sommes inspirés pour proposer un modèle générique de données capteurs. Toutefois, les descriptions SSN sont abstraites et centrées capteur et système de captage. La spécification du concept abstrait de données mesurées (Observation Value dans SSN) est peu détaillée. Une structuration plus précise tout en restant générique est nécessaire pour faciliter le développement d'applications métier d'éco-monitoring environnemental ou urbain. Une structuration plus précise simplifierait le "requêtage" des données (expression et optimisation des requêtes); une structuration générique permettrait d'homogénéiser des données issues de systèmes de captage hétérogènes et favoriserait la mise en oeuvre d'analyses croisées multidimensionnelles (spatiales, temporelles, sémantiques).

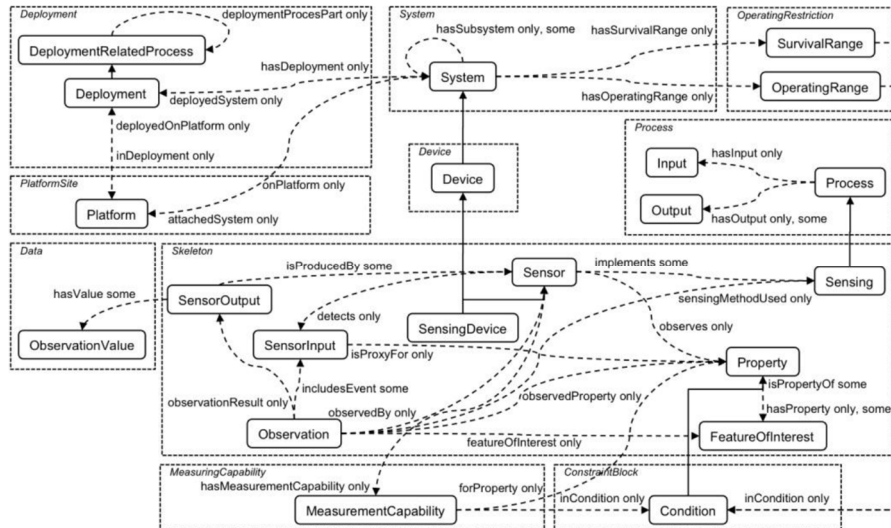


Figure 1. SSN, (Compton et al. 2011)

D'autres travaux sur les données capteurs ont été réalisés. Dans (Noël et al. 2005) les auteurs proposent un modèle de données servant de base à une proposition d'indexation spatio-temporelle des données capteurs valorisant les données les plus récentes. Le modèle a ensuite été repris dans (Gutierrez et al. 2009, 2013) et amélioré comme support à une méthodologie d'évaluation de la qualité des données capteurs pour les systèmes de surveillance de phénomènes environnementaux. En s'inspirant de ce modèle et en se basant sur SensorML et SSN, nous nous proposons de concevoir un modèle plus ouvert à tout contexte de monitoring environnemental et urbain, en nous concentrant sur les concepts ["Sensor", "MeasureType", "Measure"] et en détaillant la partie Data de SSN pour aboutir à une représentation générique des données capteur.

2.2 Autour des enquêtes

Pour analyser un phénomène sociologique, un des outils adapté est l'enquête. Une enquête correspond au recueil et à l'analyse de réponses à des questionnaires, ces questions pouvant être plus ou moins complexes. Plusieurs éditeurs proposent des logiciels de sondage pour la création des enquêtes et l'exploitation des données collectées. LimeSurvey est un progiciel libre qui permet de publier un questionnaire en ligne et d'en collecter les réponses. Nous avons étudié de près le schéma de la base de données de cet outil pour proposer un modèle conceptuel.

SPHINX iQ est un logiciel commercial qui permet de gérer des enquêtes et analyser les données, quelle que soit leur nature, quantitative ou qualitative. SPHINX permet de créer des questionnaires composés de plusieurs types de

questions : questions fermées (uniques, multiples ou ordonnées), ouvertes, échelles graduées, numériques, etc. La documentation sur le logiciel SPHINX nous a permis d'approfondir les spécificités liées à la conception des questionnaires à savoir les différents types de questions réponses. De plus nous nous sommes aperçus de l'importance de la phase d'analyse des données d'enquêtes qui constitue un des objectifs importants du recueil de données. Toutefois de nombreux utilisateurs d'enquêtes se contentent d'un tableur pour gérer les questionnaires (Deleuil 2010). En effet, les tableurs offrent une structuration basique accompagnée d'outils relativement accessibles. Pourtant, ces outils ne sont pas destinés à la gestion et l'exploitation de masses de données hétérogènes contrairement aux bases de données et aux entrepôts. Nous nous sommes donc inspirés d'une partie des concepts sous-jacents aux produits du marché pour concevoir un modèle conceptuel essentiellement centré sur les questionnaires et les données collectées.

3. Des concepts au modèle générique

3.1. Enquête

Analyse métier

Une enquête est une démarche intellectuelle qui vise à découvrir des faits. "Réaliser une enquête, c'est interroger un certain nombre d'individus en vue d'une généralisation" (Ghiglione *et al.* 1998). Premièrement, un échantillonnage permet de préciser la population cible de l'enquête. Ensuite, les méthodes d'interrogation sont définies ; celles-ci diffèrent selon le thème de l'enquête et les informations à extraire. La méthode d'interrogation ne se résume pas seulement à la manière d'écrire les questions mais elle définit aussi une relation de communication à établir entre l'enquêteur et l'enquêté. Vient ensuite l'étape de construction du questionnaire, instrument rigoureusement élaboré de collecte et d'évaluation des données pendant une enquête. Les questions doivent suivre un ordre cohérent et sont souvent regroupées en thèmes. Il existe généralement deux types de questions, les questions de faits et les questions d'opinion. Les questions de faits portent sur des données contestables et peuvent être connues autrement que par l'enquête. Les questions d'opinions, d'attitudes, de motivations ou de préférences sont des questions auxquelles seul l'enquêté peut fournir la réponse. On peut aussi classer les questions selon leurs formats : les questions ouvertes auxquelles les individus enquêtés peuvent répondre librement, et les questions fermées pour lesquelles le sujet choisit la réponse la plus adéquate parmi une liste de réponses prédéfinies. La liste de réponses prédéfinies peut se limiter à oui/non ou comporter plusieurs réponses à choix multiples ; il est parfois utile d'ajouter "autre réponse" si la liste proposée ne correspond pas tout à fait à la réponse. Outre la forme, la différence entre ces deux types de questions réside dans l'analyse statistique des réponses. Dans le cas des questions fermées, il est possible d'analyser automatiquement les réponses. Quant aux questions ouvertes, il faut regrouper les réponses en catégories et ensuite coder les réponses selon leurs catégories ce qui peut soulever des

problèmes d'interprétation de contenu. L'importance de structurer les questions d'un questionnaire entraîne une certaine complexité des concepts.

Définition des concepts et objets métier

Nous définissons une enquête *Survey* comme une précision d'un thème ou d'un phénomène ou d'un objet d'études à étudier. L'enquête est élaborée à travers des campagnes *Campaign* dans des périodes de temps et pour différentes localisations. Plusieurs enquêteurs *Investigator* participent à une campagne d'enquête ; ils se chargent de la passation des questionnaires et du recueil des réponses. Un exemplaire *Copy* est un questionnaire rempli. Nous le définissons ainsi pour différencier un questionnaire vide *Questionnaire* qui représente l'ensemble des questions, d'un exemplaire « valué » qui regroupe les réponses à ces questions. Les questions sont ordonnées et regroupées en thèmes *QuestionGroup* pour assurer une meilleure cohérence, l'ordre des questions influençant parfois les réponses. Une question est une demande d'informations et peut être décomposée en sous-questions. Nous différencions alors les questions simples ("Moyen de transport préféré ?") des questions composées ("Précisez votre moyen de transport préféré pour chaque activité : emploi, course, sorties en famille ?"). Il existe des questions dont la réponse est multivaluée; ces questions sont alors structurées en attributs (e. g. Quels sont les moyens de transports que vous utilisez et à quelle fréquence par semaine ?). La réponse *Answer* à une question multivaluée est l'ensemble des valeurs *Value* associées chacune à un attribut de la question. Nous précisons deux catégories de réponses, les réponses prédéfinies pour les questions fermées et les valeurs réponses libres pour les questions libres. Les valeurs d'une réponse peuvent avoir plusieurs types (numériques, chaînes de caractères, donnée géographique. etc.). Une question pouvant avoir des réponses multiples, nous attribuons plusieurs réponses à une seule question dans un même exemplaire. En synthèse une question peut être simple ou composée, monovaluée ou multivaluée, à réponse unique ou multiple.

3.2. Modèle générique VGS (Virtual Generic Sensor)

L'ontologie définie par SSN (Figure 1) permet de formaliser les concepts autour des réseaux de capteurs mais est peu détaillée sur la formalisation des données. En s'inspirant du modèle (Noël et Servigne 2005), en alignement avec SSN, et suite à l'analyse des concepts issus des enquêtes, nous avons conçu un modèle générique (Figure 2) inspiré du standard MOF (MetaObject Facility (OMG 2000)) pour structurer le modèle en deux parties. Nous présentons la deuxième couche du modèle MOF, correspondant au modèle métier VGS lui-même. La partie statique (à gauche de la Figure 2) décrit le système et la partie dynamique (à droite de la Figure 2) accueille les données qui sont produites par le système. Les entités "Class" (à gauche) décrivent les propriétés statiques du système qui produit les données. Les entités "Instance" (à droite) contiennent des données dynamiques qui évoluent dans le cycle de vie du système. La relation "Classifier" correspond à une relation d'instanciation entre une entité statique et dynamique.

La classe *Hypothesis* concerne l'Observation. Toutefois dans SSN, l'observation désignant le fait d'observer un phénomène en vue de récupérer des mesures, dans

notre cas, la classe a davantage un rôle de métadonnée; elle indique ce que l'on veut observer avec une description textuelle. Pour mettre en place des capteurs en vue de valider une hypothèse, on dispose de la classe *Deployment*. Après avoir étudié les modèles capteurs existants, il a été observé que le terme *Deployment* est utilisé (notamment dans SSN) et concorde avec notre définition. Les sources de données et les données sont modélisées par *Sensor*, le capteur, et *Detector* la sonde. *Sensor* est une terminologie utilisée dans SSN et *Detector* est utilisé dans SensorML. Nous avons également ajouté *MeasureAttribute* qui définit l'attribut de type de mesure. Il peut y avoir un ou plusieurs attributs pour un type de mesure ce qui permet d'avoir des structures complexes adaptées à des sources diverses. Par exemple une mesure de type localisation peut avoir 3 attributs: les positions GPS en X, Y et Z. Pour la partie dynamique, une entité *Sample* désigne l'échantillon qui est produit à un instant donné et qui peut contenir une ou plusieurs mesures (si le capteur a plusieurs sondes). Ces mesures peuvent avoir une ou plusieurs valeurs.

Pour vérifier la généralité de notre modèle, nous avons défini un concept pivot dans chaque contexte : capteur et enquête. Dans le domaine Enquête l'information que nous cherchons à obtenir est la "réponse", cette entité correspond à l'entité "mesure" dans le modèle capteur. En effet un système de captage produit des mesures pour l'observation d'un phénomène donné, de même une enquête a pour but de recueillir les réponses autour d'un sujet précis.

Une mesure est captée par une sonde *Detector* et un *Sensor* capte plusieurs types de mesures à travers les sondes. Au moment de l'enquête, les enquêteurs obtiennent des réponses à des questions qui ensemble forment un questionnaire. Le questionnaire est donc un capteur qui collecte les réponses aux questions. Nous relevons donc deux correspondances : *Detector/Question* et *Sensor/Questionnaire*. Toutefois, la structure de la question est différente de celle de la sonde : une question peut être définie par plusieurs attributs et en l'occurrence la réponse est l'ensemble des valeurs correspondant à ces attributs. Quant à la sonde, elle ne produit qu'une seule valeur d'un type donné à la fois, la mesure. Nous nous sommes inspirés de la structuration de la question pour mieux spécifier le concept de sonde en adoptant un nouveau point de vue sur la représentation des sondes qui peuvent fournir plusieurs paramètres sur leur environnement. Un exemple concerne les sondes multi-paramètres d'analyse des milieux aquatiques (Oxygène dissous, pH, Conductivité. etc.). Une sonde a donc plusieurs attributs, chacun correspondant à un type de mesure. La sonde fournit des valeurs pour ces attributs constituant ainsi une mesure. Enfin, dans une enquête, les réponses sont collectées dans des exemplaires qui correspondent donc au *Sample* puisqu'il s'agit de regrouper les données « mesure » ou « réponse » d'une même source (capteur ou questionnaire).

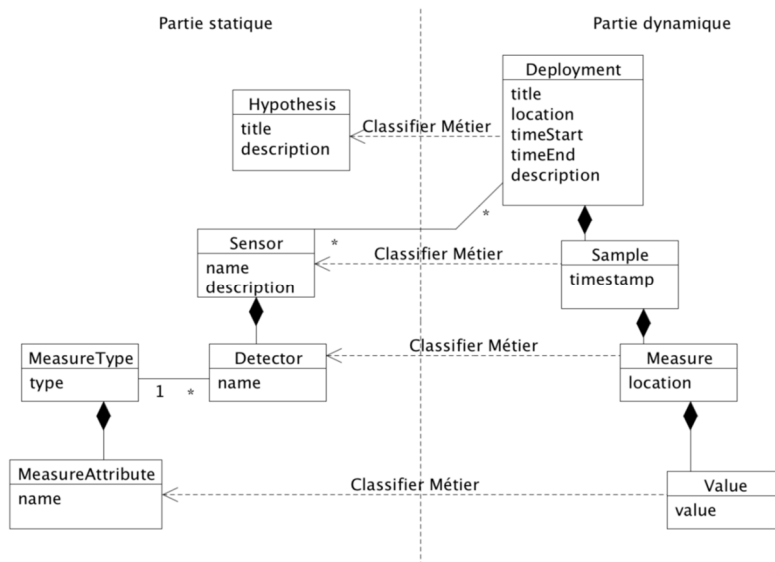


Figure 2. Modèle VGS

3.3. Discussion et caractérisation de l'aspect spatial

Dans un objectif à terme d'exploration des données selon différentes dimensions (Gomez *et al.* 2012), nous nous intéressons notamment à la dimension spatiale. Dans notre modèle générique, l'aspect spatial se trouve pour l'instant à minima identifié au travers d'un attribut des deux entités *Deployment* et *Measure*. Concernant la mesure, l'attribut *location* indique la localisation des valeurs mesurées. Si dans le cadre de capteurs et sondes physiques, l'attribut de localisation est souvent identifié par collecte automatique, il est nécessaire de clarifier cet attribut pour des données issues d'une enquête afin de le rendre explicite. Il s'agira alors d'identifier les questions localisées et ainsi de spécifier cette localisation en la prédéfinissant à la construction du questionnaire ou en la collectant. L'information de localisation ou de spatialité peut ou pourrait ainsi être présente dans le modèle à plusieurs niveaux :

1. Au niveau instance : entité Valeur (Value). La valeur est elle-même une localisation ou une coordonnée de localisation. Par exemple, dans un questionnaire, lorsque la réponse sollicitée par une question est une donnée géographique de localisation (e. g. Où travaillez-vous ?).
2. Au niveau attribut. D'autres attributs de localisation pourraient être ajoutés par exemple dans l'Entité *Sample*. Il s'agirait ainsi d'identifier la localisation à laquelle un exemplaire du questionnaire a été rempli.
3. Au niveau entité. Des entités spatiales pourraient être ajoutées au modèle par exemple concernant la zone d'observation que couvre une campagne. Il s'agit de la zone géographique où la campagne de mesure/l'enquête est menée. La zone d'observation peut porter sur un quartier, ville, région ou pays. Etc.

4. Au niveau métadonnées. Lors d'enquêtes ou de campagnes de mesures, des informations contextuelles pourraient être relevées.

Les données spatiales ou de localisation conjuguées avec des données d'enquêtes et mesures peuvent permettre d'enrichir l'analyse et de répondre à des questions concernant par exemple, la moyenne des réponses/*mesures* pour une localisation donnée ou encore la provenance des exemplaires/*samples* et son impact sur les réponses/*mesures*. Les dimensions d'analyse spatiale et temporelle requièrent donc une étude plus approfondie notamment grâce à différents cas d'utilisation.

4. Mise en œuvre expérimentale

Afin de mettre en œuvre notre proposition, nous avons implémenté le modèle VGS dans une BD relationnelle. Nous avons intégré cette base au sein de la plateforme expérimentale de bâtiment intelligent SoCQ4Home (2012) : la BD VGS est ainsi alimentée par les mesures des capteurs. Nous avons également instancié un questionnaire dans cette même base et simulé les résultats d'une enquête auprès d'usagers du bâtiment. Enfin, nous avons exploité cette base de données pour illustrer l'intérêt du modèle commun facilitant les analyses croisées de données pluridisciplinaires multi-sources.

4.1. Plate-forme SoCQ4Home

La plate-forme expérimentale de bâtiment intelligent SoCQ4Home (2012) est une instrumentation du 3ème étage du bâtiment Blaise Pascal du Campus LyonTech, hébergeant une partie du Département Informatique de l'INSA de Lyon et du Laboratoire LIRIS. Environ 70 capteurs ont été installés dans une vingtaine de salles : des bureaux, une salle de repos, des salles d'enseignement (salles de TD, amphi). Ces capteurs utilisent la technologie EnOcean. 5 types de capteurs ont été déployés : Température (36), Température & Humidité (2), Température & Humidité & CO2 (7), Contact de porte ou fenêtre (11), Luminosité & Présence (10). Les données sont produites environ toutes les 15 minutes ou ponctuellement sur un évènement particulier (grande variation de température, présence détectée, etc.). Une infrastructure matérielle (passerelles capteurs EnOcean) et logicielle (basée sur le principe du framework SoCQ (Gripay *et al.* 2010)) est également déployée dans le bâtiment, pour acquérir, interpréter et finalement historiser ces données capteurs. Des outils Web d'exploitation des données capteurs ont été développés sur cette infrastructure : un outil de tableau de bord (Figure 3) permet de visualiser l'évolution des phénomènes physiques mesurés ; une visualisation du bâtiment en 3D permet de projeter les résultats de requêtes d'exploration ou d'analyse sur la géographie réelle du bâtiment (exemple: la moyenne de température par pièce et l'état courant des portes/ fenêtres).



Figure 3. DashBoard et projection 3D de la température

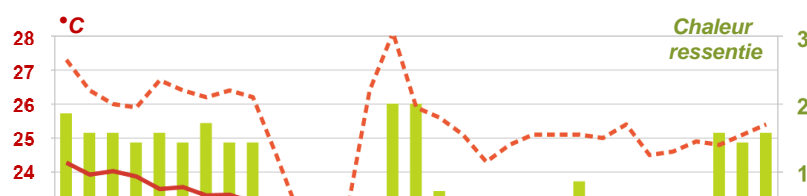
4.2. Implémentation : base de données VGS et visualisation corrélée

Données capteurs. Nous avons implémenté une base de données suivant le modèle VGS dans le SGBD MySQL. Chaque classe du cœur du modèle conceptuel VGS se transpose logiquement en table : *Hypothesis*, *Sensor*, *Detector*, *MeasureType* et *MeasureAttribute* pour la description du système ; *Deployment*, *Sample*, *Measure* et *Value* pour les données produites par le système. Afin de construire un référentiel commun, d'autres tables sont ajoutées (sur la partie droite de la figure) concernant la localisation. Dans notre cas, capteurs, sondes et mesures sont localisés par rapport à la géométrie du bâtiment (pièces du bâtiment, et éventuellement coordonnées précises). Une fois la base de données VGS en place, nous avons importé l'inventaire des capteurs de la plate-forme SoCQ4Home.

Instanciation d'un questionnaire. Dans la base de données VGS, nous avons instancié un questionnaire concernant le ressenti des usagers du bâtiment. Ce questionnaire est composé de 5 questions :

1. Comment qualifiez-vous votre forme aujourd'hui ? Précisez une note de 0 à 10 (0 : pas du tout en forme, 10 : tout à fait en forme)
2. Quel niveau de température ressentez-vous dans votre bureau ? (-2 : trop basse, -1 : un peu basse, 0 : confortable, 1 : un peu élevée, 2 : trop élevée)
3. Quel niveau d'humidité ressentez-vous dans votre bureau ? (1 : faible, 2 : moyen, 3 : élevé)
4. Quel niveau de luminosité pensez-vous avoir dans votre bureau ? (1 : faible, 2 : moyen, 3 : élevé)
5. Vous sentez-vous de bonne humeur ? Précisez une note de 0 à 10 (0 : de très mauvaise humeur, 10 : de très bonne humeur)

Un *Sensor* de type Questionnaire a donc été créé, associé à 5 *Detectors* représentant les questions. Ces *Detectors* ont un type de mesure « Indicateur subjectif » associé à un attribut de type entier. Nous avons ensuite implémenté un générateur de réponses pour ce questionnaire, guidé par notre connaissance fine du comportement thermique de notre bâtiment pendant les mois froids de l'année et en



extrapolant l'impact sur le ressenti des utilisateurs de 7 bureaux de la partie Nord du bâtiment. Nous nous sommes concentrés sur le mois de Janvier 2014, où plusieurs épisodes thermiques significatifs ont eu lieu. Nous avons ainsi généré, pour chaque bureau et chaque jour du mois de Janvier 2014, une réponse au questionnaire, c'est-à-dire un *Sample* daté du jour concerné, contenant 5 *Measures* (1 par question), chacune composée d'une *Value* contenant une valeur entière. Par exemple, les valeurs [8;1;2;1;9] indiquant une vitalité de 8/10, un ressenti de température un peu élevé, un ressenti d'humidité moyenne, etc.

Figure 4. Corrélation de mesures et données d'enquêtes

Exploitation. Afin d'illustrer l'intérêt du modèle commun VGS, nous avons réalisé une analyse croisée entre données capteurs (réelles) et données d'enquête (simulées). Le résultat est illustré dans la Figure 4. Sur le premier graphique, pour chaque jour de la période concernée, nous avons représenté la température (moyenne, maximum, minimum) sous forme de courbes, et le ressenti moyen de la température sous forme d'histogrammes (échelle de droite entre -3 et 3). Cette représentation permet d'explorer visuellement les données et d'identifier des corrélations potentielles entre les différentes mesures. Les données concernant 7 bureaux ont été ici agrégées par jour (agrégation selon la dimension spatiale et temporelle), et visualisées sur un mois (visualisation selon la dimension temporelle). Le même type de visualisation peut se réaliser à une autre échelle spatiale (par bureau, par étage) ou temporelle (par mois, par année). Par exemple, la Figure 5 présente une visualisation homogène par pièce du bâtiment, de mesures de température et humidité agrégées par heure avec une courbe par capteur (cinq capteurs dans la première pièce et un dans les suivantes).

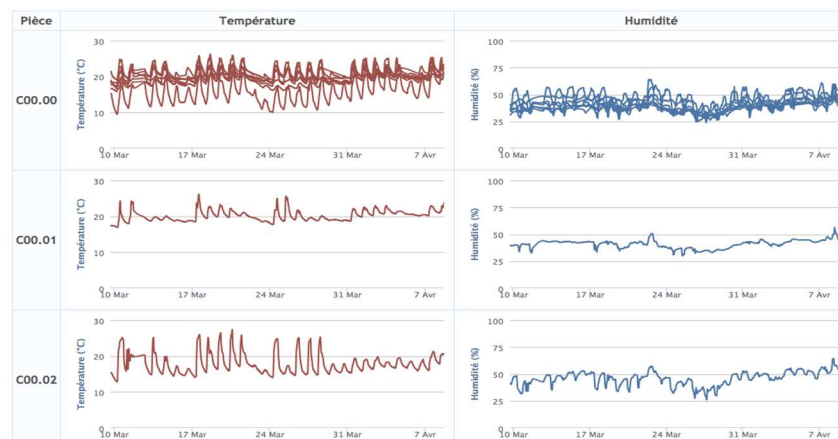


Figure 5. Visualisation homogène de données agrégées localisées

5. Conclusion et perspectives

Dans cet article, nous avons présenté le modèle générique VGS offrant une représentation homogène de données hétérogènes pour le domaine du monitoring afin d'en permettre d'une part le stockage dans une base de données et d'autre part l'analyse pluridisciplinaire. Le modèle VGS est le fruit d'un processus incrémental pluridisciplinaire basé sur la définition d'un modèle pour les capteurs et d'un modèle pour les enquêtes (Mebrouk 2012). Par sa genericité, le modèle VGS permet de représenter des systèmes de captage hétérogènes ou par ailleurs peut servir de modèle pivot pour aligner des modèles existants de monitoring. Nous avons implémenté un modèle physique de données VGS pour expérimenter le stockage et l'analyse de données hétérogènes. Les données capteurs (réelles) et enquête (synthétiques) ont été parfaitement intégrées. Pour la partie analyse nous avons effectué des agrégations sur les données enregistrées. Grâce à la genericité du modèle, la captation et l'interrogation des données sont facilitées favorisant l'agrégation et la corrélation multidimensionnelles et hétérogènes.

Dans le cadre de la plateforme SoCQ4Home, nous avons réalisé une implémentation dans une base de données relationnelle pour profiter dans un premier temps de la puissance du langage SQL. Nous souhaitons également appliquer notre approche sur d'autres jeux de données-questionnements réels à plus grande échelle. Nous envisageons donc une deuxième implémentation de VGS dans une base de données orientée BigData comme MongoDB ou Cassandra. Par ailleurs, nous avons lancé une campagne de mesures (2014–2019) comprenant d'une part l'instrumentation d'un bâtiment tertiaire sur plusieurs étages (300 capteurs) et d'autre part la passation de questionnaires dans des enquêtes auprès des usagers à des temporalités différentes. Cette plateforme MARBRE a été financée par le CNRS au travers d'investissements mi-lourds et du projet AMADOUER du programme

MASTODONS de la mission interdisciplinaire de l'INS2I, et soutenue via un BQR de l'INSA de Lyon.

Nous nous intéressons aujourd'hui à la phase d'exploration. En effet, l'exploration de données par les utilisateurs peut conduire à l'émergence de faits ou hypothèses conduisant à l'expression d'éléments de réponses à des questionnements ou à la compréhension de phénomènes complexes. L'exploration requiert un langage de requête efficient et la définition de dimensions pertinentes fonction de l'objet d'étude et des questionnements soulevés. Par exemple, l'aspect spatio-temporel est au cœur des systèmes de monitoring. Dans le modèle proposé, l'information temporelle, de localisation et de spatialité est notamment présente à plusieurs niveaux de granularité et de sémantique et nécessite une analyse approfondie des opérateurs et dimensions pour son exploitation corrélée avec d'autres dimensions.

Bibliographie et références

- Arnaud-Fassetta G., Astrade L., Bardou E., Corbonnois J., Delahaye D., Fort M., Gautier E., Jacob N., Peiry J.L., Piégay H., Penven M.J. (2009). Fluvial geomorphology and flood-risk management. *Géomorphologie : relief, processus, environnement* 2 : 109-128.
- Bermudez L., Delory E., O'Reilly T., Del Rio Fernandez J. (2009). Ocean observing systems demystified. In *Proceedings of OCEANS 2009*, Biloxi, MS, USA, October 2009; IEEE: New York, NY, USA, 2009; pp. 1-7.
- Calder M., Morris R.A., and Peri F. (2010). Machine reasoning about anomalous sensor data. *Ecological Informatics*. Volume 5, Issue 1, January 2010, Pages 9-18
- Compton M. et al. (2011). The SSN Ontology of the W3C Semantic Sensor Network Incubator Group. 7p.
- Davoine P-A., Poulenard L., Gueguen P., Gensel J., Bringel J. (2013). Dispositif mobile et géoweb pour la collecte de données d'inventaire sismique du bâti en milieu urbain : L'application Isibat. In *proceeding of SAGEO'2013*. Pp. 227-241.
- Deleuil J-M. (2010) : Les plans lumière, stratégies et méthodes, ouvrage bilingue, Ed. Luci.
- Diallo O., Rodrigues J., Sene M. (2012). Real-time data management on wireless sensor networks: A survey. *Journal of Network and Computer Applications*. Vol.35/3. p.1013-21
- Eid M., Liscano R., El Saddik A. (2006). A novel ontology for sensor networks data. In *Proceedings of the IEEE International Conference on Computational Intelligence for Measurement Systems and Applications*. Pp. 75-79
- Gangemi A. (2010). Dolce+dns ultralite.
http://ontologydesignpatterns.org/wiki/Ontology:DOLCE+DnS_Ultralite, 2010.
- Ghiglione R., Matalon B. (1998). Les enquêtes sociologiques : théories et pratique. Decitre : Paris. 301 p.
- Gómez L.I., Gómez S.A., Vaisman A.A. (2012). A generic data model and query language for spatiotemporal OLAP cube analysis. *Proceedings of the 15th EDBT-ICDT*. Pp. 300-311. ACM, New York, NY, USA.
- Gripay Y., Laforest F., Petit J-M (2010). A Simple (yet Powerful) Algebra for Pervasive Environments. Dans *EDBT 2010, EDBT-ICDT*, Lausanne, Switzerland. Pp. 1-12.
- Gutierrez C., Servigne S. (2013). Managing Sensor Data Uncertainty: a data quality approach. *International Journal of Agricultural and Environmental Information Systems (IJAEIS)* 4(1):35-54, IGI Global.
- Gutierrez C., Servigne S. (2009). Métadonnées et Qualité pour les Systèmes de Surveillance en Temps-Réel. *Revue Internationale de Géomatique*. Lavoisier. Vol. 19/2. Pp. 151-158
- Herzog A., Jacobi D., Buchmann A. (2008). A3ME - an Agent-Based middleware approach for mixed mode environments. In *UBICOMM 2008*. Valencia, Spain: IEEE Computer Society Press, October 2008, pp. 191-196.

- Hu Y., Wu Z., Guo M. (2007). Ontology driven adaptive data processing in wireless sensor networks. InfoScale '07 Proceedings of the 2nd INFOSCALE. Suzhou, China, June 2007.
- Le Moniteur. (2014). Performance énergétique réelle des bâtiments . 22 mai 2014.
<http://www.lemoniteur.fr/150-performance-energetique/article/actualite/24558593-il-est-temps-de-mesurer-la-performance-energetique-reelle-des-batiments>
- Madina J., Bowersb S., Schildhauera M., Krivovc S., Penningtond D., Villa F. (2007). An ontology for describing and synthesizing ecological observation data. *Ecological Informatics*. Volume 2, Issue 3, October 2007. Pp. 279–296
- Mebrouk R. (2012). Modélisation conceptuelle générique pour un monitoring éco-informatique durable. Mémoire de Master Informatique. Insa Lyon. Travail réalisé grâce au soutien financier du LABEX IMU (ANR-10-LABX-0088) de l'Université de Lyon, dans le cadre du programme "Investissements d'Avenir" (ANR-11-IDEX-0007) géré par l'Agence Nationale de la Recherche (ANR).IMU Investissement d'Avenir, 2012.
- Mefteh W., Bouju A., Malki J. (2012). Une approche ontologique pour la structuration de données spatio-temporelles de trajectoires. *Revue Internationale de Géomatique - Hermes-Lavoisier*, vol 22/1-2012, pp 55-75
- MMI (2008). Marine Metadata Interoperability device ontology: A community development project. <https://marinemetadata.org/references/sensorsontmmi>
- Neuhaus H., Compton M. (2009). The semantic sensor network ontology: A generic language to describe sensor assets. *AGILE Workshop: Challenges in Geospatial Data Harmonisation*. 7 p.
- Noel G., Servigne S. (2005). Indexation multidimensionnelle de bases de données capteur temps-réel et spatiotemporelles. *Ingénierie des Systèmes d'Information* 10(4):pp.59-88, Lavoisier, ISBN 2-7462-1267-6, ISSN 1633-1311.
- OMG (2000). Meta Object Facility Specifications.
<http://www.laputan.org/pub/books/MOF%2000-04-03.pdf>
- Prévost A., Rodrigues D., Molines N., Beckers B., Caudron I. (2010). Optimiser les performances énergétiques du bâti et de la ville grâce à la modélisation 3D. In *Proceedings of SAGEO'2010*. Toulouse. Novembre 2010. Pp. 89-105.
- Reed C., Botts M., Davidson J., Percivall G. (2007) : Ogc® sensor web enablement: overview and high level achitecture. *Autotestcon, 2007 IEEE*. pp. 372–380
- Russomanno D., Kothari C., Thomas O. (2005). Building a sensor ontology : A practical approach leveraging. ISO and OGC Models. In *proceedings of the 2005 International Conference on Artificial Intelligence*, Las Vegas, NV.
- SensorML (2007). OpenGIS® Sensor Model Language (SensorML) Implementation Specification. Open Geospatial Consortium. pp. 1–87.
<http://www.w3.org/2005/Incubator/ssn/XGR-ssn-20110628/.SensorML>
- Servigne S., Devogele T., Bouju A., Bertrand F., Gutierrez C., Laucius S, Noël G., Ray C. (2009). Gestion de masses de données au sein de bases de données capteurs. *Revue Internationale de Géomatique*. Lavoisier. Vol. 19/2. Pp. 133-150
- SoCQ4Home (2012). SoCQ4Home Project, <http://liris.cnrs.fr/socq4home/>
- Stasch C., Janowicz K., Bröring A., Reis I., Kuhn W. (2009). A stimulus-centric algebraic approach to sensors and observations. *GeoSensor Networks. LNCS*. Vol.5659, p.169-179.