HAL
open science

Heterogeneous Sensor Data Exploration and Sustainable Declarative Monitoring Architecture: Application to Smart Building

Sylvie Servigne, Yann Gripay, Ozgun Pinarer, John Samuel Samuel, Atay Ozgovde, Jacques Jay

To cite this version:

Sylvie Servigne, Yann Gripay, Ozgun Pinarer, John Samuel Samuel, Atay Ozgovde, et al.. Heterogeneous Sensor Data Exploration and Sustainable Declarative Monitoring Architecture: Application to Smart Building. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2016, III-4/W1, pp.97-104. 10.5194/isprs-annals-IV-4-W1-97-2016 . hal-01386110

HAL Id: hal-01386110

https://hal.science/hal-01386110v1

Submitted on 22 Oct 2016

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.

HETEROGENEOUS SENSOR DATA EXPLORATION AND SUSTAINABLE DECLARATIVE MONITORING ARCHITECTURE: APPLICATION TO SMART BUILDING

Sylvie Servigne[a], Yann Gripay[a], Ozgun Pinarer[ab] John Samuel[a], Atay Ozgovde[b], Jacques Jay[c]

[a]Université de Lyon, CNRS, INSA-Lyon, LIRIS,
UMR CNRS 5205, F-69621, France – (Sylvie.Servigne, Yann.Gripay)@insa-lyon.fr – John.Samuel@liris.cnrs.fr
[b]Galatasaray University, Dept. of Computer Engineering
Ciragan Cad. No:36, 34349, Istanbul, Turkey – (fopinarer, aozgovdeg)@gsu.edu.tr
[c]Université de Lyon, CNRS, INSA-Lyon, CETHIL,
UMR CNRS 5008, F-69621, France – Jacques.Jay@insa-lyon.fr

KEY WORDS: Sensor data model, Heterogeneous data integration, Data Cross-Analysis, Data exploration, Monitoring architecture, Energy consumption, Smart Building

ABSTRACT:
Concerning energy consumption and monitoring architectures, our goal is to develop a sustainable declarative monitoring architecture for lower energy consumption taking into account the monitoring system itself. Our second is to develop theoretical and practical tools to model, explore and exploit heterogeneous data from various sources in order to understand a phenomenon like energy consumption of smart building vs inhabitants' social behaviours. We focus on a generic model for data acquisition campaigns based on the concept of generic sensor. The concept of generic sensor is centered on acquired data and on their inherent multi-dimensional structure, to support complex domain-specific or field-oriented analysis processes. We consider that a methodological breakthrough may pave the way to deep understanding of voluminous and heterogeneous scientific data sets. Our use case concerns energy efficiency of buildings to understand relationship between physical phenomena and user behaviors. The aim of this paper is to give a presentation of our methodology and results concerning architecture and user-centric tools.

1. INTRODUCTION

Over the past decade, popular attention to smart building has increased. Smart building and sustainability are intertwined. 2011 Annual Energy Review of the World Business Council for Sustainable Development announced that in United States, buildings are responsible for 41% of energy consumption of the country. Moreover, in 2008 energy consumption of buildings in European Union (EU) is 37% of the total energy budget. Surprisingly this consumption is higher than the energy consumption of industry and transport which are respectively 28% and 32%. These results should be considered as a warning for the future and show the necessity of monitoring energy consumption of the buildings (EIA, 2011) and (Eichholtz, 2013).

Usage of smart buildings is not only for monitoring energy consumption but also to provide useful services for the occupants such as illumination, thermal comfort, air quality, physical security, sanitation etc. Since nature of buildings are changing, inhabitants prefer more dynamic work and living environments to be actively supported and assisted by smart building management system. Each smart building is a pervasive environment that covers wireless sensor network (WSN) environment and management of sensor data streams. WSN environment consists of small, light weight computational devices that are able to communicate over wireless connection channels. These devices are equipped with sensing, processing and communication facilities.

The major challenge in wireless sensor devices is the limited energy and lifetime. Since wireless sensor devices are autonomous in terms of energy and their energy consumption determines the lifetime of the monitoring system, energy consumption of WSN environment is considered as a key metric for the system. In WSN applications, sensor devices sample physical quantity measures and transmit them with defined acquisition and transmission frequencies. These application requirements make sensor device consume energy. Thus, energy should be monitored during the execution and it is necessary to monitor energy consumption of the monitoring architecture itself.

Our main intention of our works is to build a monitoring system that supports multi-applications in real time in a pervasive environment. Multiple application system requires handling several data stream requests with different data acquisition/transmission frequencies for the same wireless sensor device and supporting dynamic requirements of applications (e.g. high transmission frequencies for occupied rooms, lower frequency during night). A common static configuration does not optimize energy consumption of the monitoring system. Thus, we focus on interaction between application requirements and wireless sensor devices and we propose a Smart- Service Stream-oriented Sensor Management, an approach to optimize interactions between application requirements and wireless sensor environment in real-time. A Smart- Service Stream-oriented Sensor Management system performs energy-aware dynamic sensor device configuration to lower energy consumption while fulfilling real-time application requirements.

Furthermore, another subject of major interest concerning smart building is user behaviours. Nowadays, energy efficiency of a building is given by theoretical estimations based on its initial design and on typical usage scenarios. Indeed the practical efficiency is usually lower due to the complexity of the building

process and to the real behaviour of occupants. If energy consumption and environmental conditions can be directly measured through instrumentation, understanding the practical energy efficiency of a building requires cross-analyses of "instrumentation data" and "survey data" (and other studies data). These data are thus necessary to fully discover and then understand complex correlations between physical and human parameters. Data are heterogeneous and complex, multi-sources, multi-dimensional (e.g., spatiotemporal data), and multimedia (e.g., numbers, texts, images, sounds, videos).
Those data need to be deeply analyzed and advanced skills are required first to understand raw data, then to discover their multi-scale properties, and finally to perform relevant aggregations and cross-source comparisons.

Our objective is to develop theoretical and practical tools to model, explore and exploit heterogeneous data from various sources in order to understand a phenomenon. We focus on a generic model for data acquisition campaigns based on the concept of generic sensor. This concept allows integration of heterogeneous captured data. The concept of generic sensor data is centred on captured data and on their inherent multi-dimensional structure. This multi-dimensional structure is then a support for complex domain-specific or field-oriented analysis processes.

In this paper, in Section 2, we focus on our methodology for integration of heterogeneous data and on our methodology for cross-analysis of integrated heterogeneous data in Section 3. Then we present briefly our approach concerning a Smart-Service Stream-oriented Sensor Management in Section 4 and then in Section 5, related works concerning sensor data and survey data modelling are explained.

2. METHODOLOGY FOR HETEROGENEOUS DATA INTEGRATION

To develop theoretical and practical tools to model, explore and exploit heterogeneous data from various sources in order to understand a phenomenon, our approach consists in a generic conceptual model of collected data. A generic model would permit the representation of heterogeneous data through only a few generic concepts to facilitate future cross-analysis.

In order to build this final model, we first took a bottom-up approach. We designed a preliminary model for heterogeneous physical sensors, based on a real experimental platform in occupied buildings. We designed another preliminary model for sociological surveys, to take into account opinions and feelings of occupants "measured" by a questionnaire.

We then took a top-down strategy. Inspired by existing abstract ontologies that describe sensor systems (Compton, 2011) and (EIA, 2011) and (Eichholtz, 2013), we designed the Virtual Generic Sensor model (VGS model) that can homogeneously represents data from heterogeneous physical sensors, data from sociological surveys, and data from other kinds of sources. This model is also based on a previous sensor model designed for natural risk monitoring (Gutierrez 2007) and (Noel, 2005) and (Laurini, 2005) and (Rodriguez, 2013). The VGS model focuses on data produced by generic sensors linked to a common multi-dimensional structure. This multi-dimensional structure describes time, location and source of measured data, and is designed to support additional specific field-oriented dimensions.

Our first result is the VGS (Virtual Generic Sensor) model. Figure 1 shows the UML class diagram of this model. It describes the static structure of a generic acquisition system, with a **Sensor** composed of several **Detectors** (further detailed by **MeasureAttributes**, via the **MeasureType**), and a dynamic structure with **Samples**, produced by a Sensor, composed of **Measures** (further detailed as **Values**). **Deployments** correspond to campaign of measures that take place to validate **Hypothesis**.

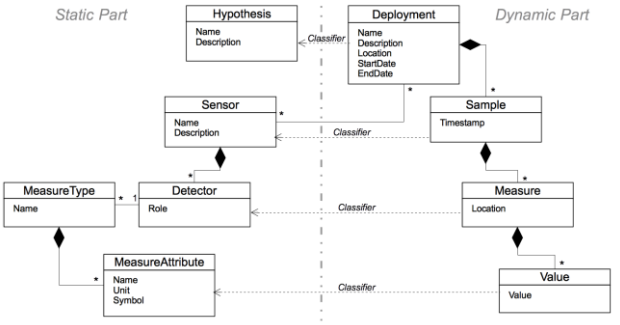

Figure 1. VGS Model with UML representation

The process to create the VGS model was the following (Figure 2) : analysis and matching of existing sensor ontologies and, analysis of survey tools, design of a sensor data model (M1), design of a survey data model (M2), alignement of sensor data (M1) and survey data models (M2). The VGS model allows managing data issued from physical sensors and survey data with the same concepts. This is the base to allow then the user centric multidimensional exploration of heterogeneous data.

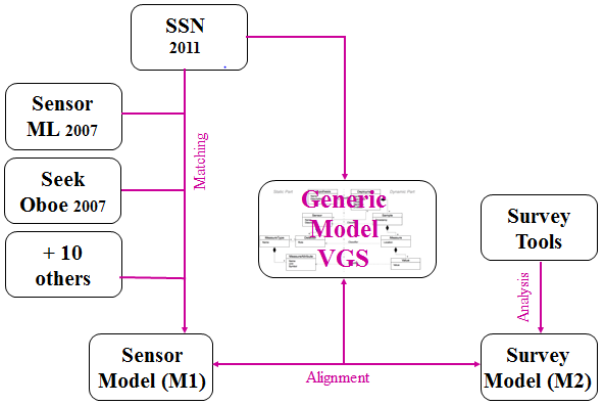

Figure 2. VGS Model design process

This model has been implemented for two Smart Building experimentation platforms managed by our project team: SoCQ4Home (illustrated with Figure 3), since October 2012; and MARBRE, since February 2014.

A total of around 400 heterogeneous physical sensors have been deployed to measure: temperature, humidity, CO2/VOC, contact (for doors/windows), electricity consumption, weather conditions... Production: about 5000 measure per day. Physical sensor devices are modeled as VGS sensors, with detectors depending on the actual sensor type. A survey concerning fifty occupants of one of the buildings has also been realized. Results of questionnaires are going to be integrated: questionnaires are modeled as VGS sensors, and answers to questions are modeled as measures. The current implementation of the VGS model is

based on a MySQL database in particular to benefit from the expressiveness of the "golden standard" SQL language; moreover MySQL is a free open source tool.

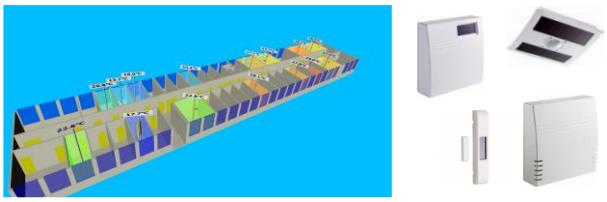

Figure 3. SoCQ4Home Platform: Monitoring Building with Sensors

3. METHODOLOGY FOR HETEROGENEOUS DATA CROSS-ANALYSIS

We also designed a methodology for an agile multi-dimensional exploration of those data. Based on the VGS model and its multi-dimensional structure, we propose a language to finely define domain-specific or field-oriented indicators through successive aggregations along dimensions (Patil, 2011) (in a similar way to Data Warehouses). We designed a visualization framework linked to those dimensions that enables users to visually explore indicators using graphs. And in order to visually compare indicators, we propose an interactive "matrix layout" for those graphs. The common multidimensional structure of data is then exploited at three levels: to structure data, to define indicators, and to explore data with these indicators.

Our agile approach allows incremental and iterative data processes and analysis. Users can start to explore raw data with some predefined dimensions for time, source and location of measured data, and with basic aggregated indicators like MIN, MAX, AVG. Exploration can then be incrementally and iteratively enriched by users themselves.

At the data level, we consider incremental data sets, e.g., when raw data are still being captured and appended to the data set. The data set generation is also iterative: data can be progressively enriched with new interpreted data that may then be used in the same fashion as raw data.

At the analysis level, we consider an incremental exploration process: new (aggregated) indicators can be added on-the-fly, when needed by users. The exploration process is also iterative when knowledge from past explorations is used to refine current and future explorations: existing analysis dimensions can be refined, with more precise levels of granularity in their level hierarchy, and new dimensions can even be added, in order to offer new points of view on data.

This approach is designed to support and enrich current domain-specific approaches for complex and/or scientific data analysis. In particular, each data visualisation is precisely and concisely described by its "lineage", i.e. by the link to the data subset, data aggregation definitions, and visual projection parameters that lead to this data visualization.

3.1 User centric multidimensional exploration of heterogeneous data

3.1.1 User centric definition of multidimensional space

One of the major challenges involved in multidimensional data analysis (Kimball, 2011) is to identify and define the dimensions and dimension levels of analysis. For example, Figure 4 shows a time dimension with one hierarchy. Three default dimensions are attached to the core VGS model: Time, with the timestamp attribute of Samples; Location, with the location attribute of measures; and Source, that represents the hierarchy of VGS concepts (Detector, Sensor, and Deployment). User-defined additional dimensions can be added to further describe measures from generic sensors. When dimensions and dimension levels are defined by users, these dimensions and dimension levels built a multidimensional space. These dimensions are then used to specify data aggregation and data visualization levels. For example, Figure 5 shows the definition by a user of an aggregation level of data (hour level into the Time dimension) and a visualization level of data (month level into the Time dimension).

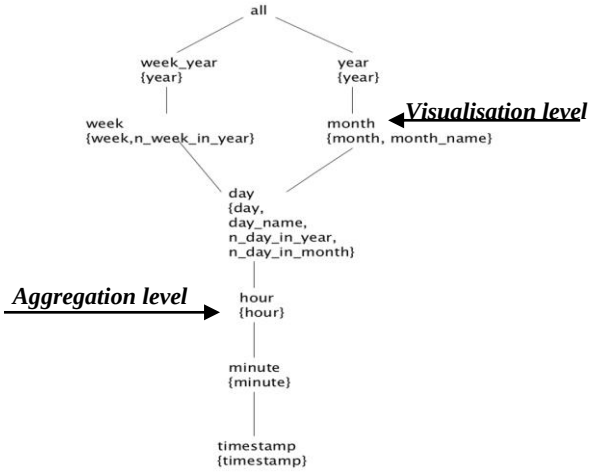

Figure 4. Dimension hierarchy: Aggregation level and Visualization level

3.1.2 Multidimensional space concepts

A temporal series consists of various observations or measurements at different points of time. Spatiotemporal series are temporal series used in conjunction with the location of the observation or measurement. Sensor data series are multidimensional. While time and location gives the information concerning when and where an event or measurement was recorded, these two dimensions are not enough for heterogeneous sensors. We require other various information's concerning the type of measurement (what) like temperature, humidity and the manner (how) by which the measure was generated (especially in cases of interpolated data).

In multidimensional data series, a relation of interest (here, the sensor measure) is analysed along different dimensions. A relation of interest (also called fact table in the literature) contains several data tuples. A dimension is an hierarchical organization of levels that allows to create a partition of these tuples at different levels of granularity. For example, the time dimension describing the temporal aspect of the tuples can contain the following levels: timestamp, hour, day, week, month, year. The hierarchy of a dimension starts with a unique base level (also known as the leaf and is highly granular like a timestamp) and ends with a unique level root (also referred to as

ALL in the literature), that partition the tuples in one single group containing all the tuples.

If there is a relationship between two levels, the lower level (closer to the leaf) is called the child level and the higher level is called the parent level. Every level has a set of descriptor attributes. Take for example; the level “day” can have various descriptor attributes like “date”, “day of week”, “day number in year”. Each member of a level is described by a value for each descriptor attribute of its level. “February 1, 2014” is a member associated to “2014-02-01, Saturday, 32” for the level “day”. Every member of a child level is linked to only one single member of the parent level in the hierarchy. For example, for the levels “day” and “month”, the members of “day”: “January 1, 2014” and “January 2, 2014” are linked to “January 2014”, a member of “month” and “February 1, 2014” is linked to “February 2014”.

A set of tuples in a relation of interest can be partitioned according to any one dimension. Every tuple must be linked to one unique member of the leaf level. A partition is thus formed by the subsets of tuples linked to each member of this leaf level. Recursively, partition at level N + 1 (e.g. level “month”) is formed by the subsets of tuples linked to each member of the level N + 1 (e.g. “January 2014”), those subsets each being constituted by the union of subsets of tuples linked to members of level N (e.g. “January 1st, 2014” of level “day”) themselves linked to this member of level N + 1 (e.g. “January 2014”) in the hierarchy.

A multidimensional space is a set of dimensions with their associated levels of hierarchies. A multidimensional series in a multidimensional space is a series of values (of the relation of interest) from the domain of possible values, attached to dimensions of this space. Finally users are able to build multiple multidimensional spaces to better explore data step by step. According to the various analyses, new spaces can be refined or defined that allow an agile and incremental way of data exploration.

3.2 Declarative language for indicators definition

Our second result is a formal model and a declarative language to finely define indicators as aggregations along dimensions for VGS data. It is based on the relational algebra (a foundation concept for relational databases, like SQL databases). In our current prototype, we implemented this language by automatically translating it to complex nested SQL aggregation queries. It enables to easily integrate new domain-specific dimensions and/or adapted existing dimensions (like Time and Location).

Due to space limitations, we do not describe the formal definition of this language. We sketch its expressiveness with an example: a user can define an indicator as a 2-step aggregation along the Time dimension for temperature sensor data, with an average at the minute level, and then a formula like “(MAX+MIN)/2” at the hour level, applied to the previous average at the minute level. An indicator definition can also span over multiple dimensions, like Time and Source.

3.3 Visualization with a Web Interface

A web dashboard for the sensor data has been developed to visualize the various measures and indicators that describe a phenomenon. Figure 5 shows the SoCQ4Home dashboard of the administrator. It shows the temperature recorded during the last 48 hours and average temperature recorded during last 30 days, in user’s office. The dashboard also shows the current temperature in some other representative rooms of the building. A visualization of building in 3D permits to project the results of exploration queries over the real geography of the building (see Figure 3).

Figure 5. Dashboard for Smart Building

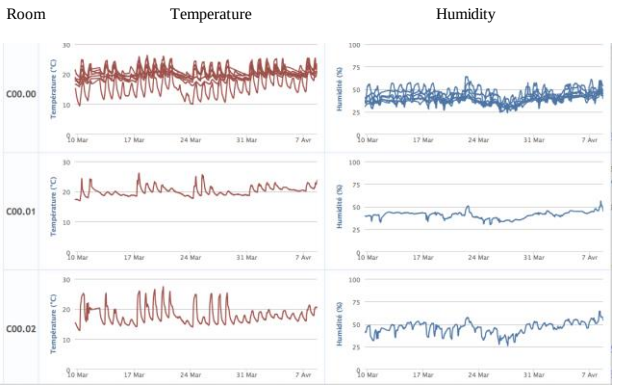

Figure 6. Web user interface to visualize raw sensor data (temperature and humidity) as a matrix of graphs (actual data from MARBRE platform) by room and by month.

Our third result is a proof-of-concept Web user interface to visualize data and/or indicators as a matrix of graphs, according to the aggregation level of indicators (e.g., hour level in the Time dimension) and to the navigation level interactively defined by the user (e.g., month level in the Time dimension). The style of visualization is illustrated in Figure 6: one line per sensor in a graph (with one data point for each hour), and one graph per room (matrix rows) for Temperature and Humidity indicators (matrix columns).

In order to facilitate visual correlations, graph scales are identical within a column, as well as the time axis between the 2 indicators. In this example, we can visually compare Temperature evolution between rooms, and visually search for potential correlations between Temperature and Humidity in each room. Aggregation level and visualisation level are defined by users using the web interface showed in Figure 7. Development of a full exploration Web interface, including definition of dimensions, indicators, and the dynamic navigation along dimensions is a work in progress.

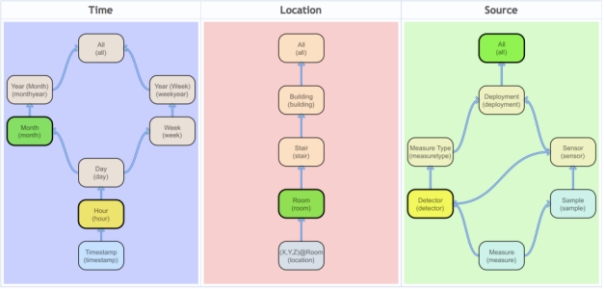

Figure 7. Multidimensional space corresponding to visual results showed in Figure 8. Aggregation levels selected by users: hour in Time hierarchy, and detector in Source hierarchy. Visualization levels selected by users: month in Time hierarchy, room in Location hierarchy, All sensors in Source hierarchy

4. SUSTAINABLE DECLARATIVE MONITORING ARCHITECTURE AND SMART-SERVICE STREAM-ORIENTED SENSOR MANAGEMENT

We focus on monitoring system of pervasive environment especially smart buildings. More precisely we focus on interactions between application service oriented queries and wireless sensor devices. We propose a sustainable declarative monitoring architecture and energy aware dynamic sensor stream management system for smart building environment. High level declarative monitoring architecture is given in Figure 8.

Proposed architecture covers Pervasive Environment Management System and WSN environments. In one side, Pervasive Environment Management System handles application service oriented queries and deals with heterogeneous sensor data streaming. On the other side, WSN covers wireless sensor devices (either physical real sensor devices or virtual ones).

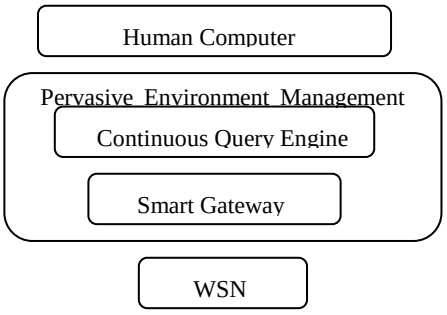

Figure 8. High level declarative monitoring architecture

From our perspective, sensor management consists of dynamic sensor configuration in terms of sensor data acquisition and transmission frequencies. Basically, real-time sensor device configuration can be realized with these two parameters. Real-time configuration profile for sensor device can be determined dynamically. This configuration can be considered as acquisition/transmission scheduling time pattern from sensor device side. This schedule is a set of timestamps that indicate moments for data acquisition and transmission. This mechanism lets a sensor device configure itself in real-time to avoid unnecessary data measurements and to promote data transmission shorter/compressed.

Since our intention is to support multiple applications, with configurable sensors, it is possible to have several subscriptions to the same sensor device with different frequency parameters. Without our approach, system considers highest acquisition and transmission frequency among the parametrized subscriptions. This will cause high energy consumption during the execution. However, with our proposed solution, parametrized subscriptions are managed, frequencies of subscription form acquisition/transmission schedule time pattern dedicated to a sensor device. Moreover, schedule time pattern algorithm is updated in real-time when a new subscription or non-subscription occurs.

In summary, existing smart building energy management systems adopt static sensor device configuration and fitted to a single monitoring application. However, sensor devices can be configurable more precisely than a duty-cycle in real time. We propose a new generation Smart Service-Stream Oriented Sensor Management. Our proposition, application requirement based energy-aware stream management and dynamic sensor configuration mechanism, are perform in the smart gateway layer in order to optimize consumed energy by a sensor device, independent from application layer and/or query.

5. RELATED WORKS

5.1 Sensor data model

Sensors have started becoming an integral part of our personal lives, for example in the form of temperature and humidity sensors, or smoke and fire detectors. Examples of some wireless sensors are shown in Figure 3. Sensors can send periodical measurements for long periods, with only very little human intervention. Depending on the chosen periodicity of data acquisition, these sensors can produce a large amount of data in a very short amount of time.

Sensor data modelling has recently generated a number of research works. Many sensor ontologies have been proposed since 2005 (see Table 1). They may focus on the description of the observation process, on the structure of the sensor network, or on the description of the physical sensors. They often contain specificities from a targeted application domain.

Name	Orientation	Aim	Year
SSN	Sensors/Observation	Generic sensor modelization [9]	2012
DUL	Concepts	Generic Sensor Concepts [14]	2010
CESN	Network	Coastal Networks [8]	2010
CSIRO	Observation	Semantic description [32]	2009
Stimuli-centered	Observation	Description of processes [51]	2009
OOSThesys	Observation	Marine Environments [6]	2009
A3ME	Network	Technical aspects and constraints [19]	2008
MMI	Observation	Marine Domain [31]	2008
SensorML	Observation	Description of processes [7]	2007
SDO	Network	Research in Sensor Networks [12]	2007
WISNO	Network	Wireless networks [23]	2007
SEEK OBOE	Observation	Scientific observations [27]	2007
OntoSensor	Sensors	Technical specifications [44]	2005

Table 1. Sensor data Models and Ontologies since 2005

Towards an objective of standardization, the W3C incubator group SSN XG3 proposed the Semantic Sensor Network (SSN) ontology based on SensorML (SensorML, 2007) and Dolce UltraLite (DUL, 2010). SSN ontology describes key concepts related to sensors and sensor networks. It proposes a collection of models for representing the main concepts linked to the context of sensors and defining the relations between these concepts, like Process, Stimulus-Sensor-Observation Pattern, Device, Data. It allows a separation between hardware aspects and abstract processes of a sensor. SSN therefore offers ontology adapted to the description of a sensing system. However SSN descriptions are abstract and focus on physical sensors and sensing systems. The specification of abstract concepts for data measures is less detailed, with only one

concept of Observation Value in the Data model (at the bottom left of SSN Ontology Diagram in Figure 10).

A more precise, but still generic structure is necessary to facilitate the development of applications managing heterogeneous sensor data, like environmental or urban monitoring. A more precise structure simplifies the querying of data (easiness of query expression and query optimization). A generic structure allows a homogeneous management of data obtained from heterogeneous sensing systems.

A lot of existing works focused on physical sensor and real-time systems (Bonnet, 2001) and (Diallo, 2012). In our project, we rather focus on issues with heterogeneous data sources (Noel, 2005). Query optimization and data indexing of sensor data are also issues in this context. In (Noel, 2010), the authors proposed a data model based on a proposition of spatiotemporal indexation of sensor data considering the most recent data. The model was also later used in (Rodriguez, 2013) and (Gutierrez, 2007) and improved to support a methodology for evaluation of the quality of sensor data for the environmental phenomena monitoring systems. These ontologies and models inspired our proposed data-centric VGS (Virtual Generic Sensor) model detailed in Section 2.1.

For analyzing sociological phenomenon, one of the commonly adapted tools is the survey. A survey corresponds to collection and analysis of answers to questionnaires. Several survey softwares have been proposed for the creation of surveys and the exploration of collected data. LimeSurvey (LimeSurvey 2015) is a free software package that permits to publish a questionnaire online and to collect their answers. SPHINX iQ (Sphinx. 2015) is a commercial software that permits to create surveys and analyze the data and their nature (quantitative or qualitative). Another group of users of surveys use spreadsheets for managing the questionnaires. We studied closely the database schema of LimeSurvey to obtain a conceptual model of survey data. The documentation of the SPHINX software permits us to extend the specifications linked to the conception of questionnaires to know the different types of answers (to questions). We preview the importance of the analysis phase of questionnaire data that constitute an important objective of this data collection. However, these tools are not destined towards the management, exploration and cross-analysis of voluminous heterogeneous data.

Concerning data management, the use of data warehouses for sensor data storage and analysis can be seen in various works like monitoring of pollinators (Da Costa, 2010), building energy and maintenance (Gökçe, 2009) and (Stack, 2012), soil ecosystem (Szlavecz, 2007). The difficulties are that they only deal with physical sensor data and not with other heterogeneous data.

5.2 User-centric Multidimensional data analysis

Sensor and survey data are historical and multidimensional. In a conceptual multidimensional design phase (Kimbal, 1996) and (Kimbal, 2011), dimensions and facts are decided along with the appropriate schema (star schema, snowflake schema, star cluster schema etc.) by taking into account all the user requirements.

Graphical conceptual modelling is particularly interesting since it helps the designers to visualize the multidimensional data modelling. These models can also be used during visual interactive exploration of data. Automated conceptual multidimensional modelling can be classified into three approaches based on how they are obtained: supply-driven, demand-driven and hybrid (a mix of both supply-driven and demand driven) approaches. Supply-driven approach considers only the schema and the ontology (Romero, 2010) of the data sources to obtain a multidimensional data model, whereas demand-driven (Romero, 2006) and (Romero, 2008) and (Romero, 2010) takes into consideration the user queries to obtain the final model.

But the major limitation in most of the works discussed above is that they do not take into consideration the possibility of various categories of end users during the conceptual modelling phase. The designer entrusted with the conceptual modelling proposes a model based on the initial user requirements or by using automated techniques discussed above. For projects involving multidisciplinary teams and dealing with heterogeneous data sources, the final conceptual model may have a large number of dimensions, levels and associated hierarchies. We believe that it is a shortcoming of the current conceptual modelling approach since it overwhelms the end user with a number of irrelevant fact tables, dimensions and the associated hierarchies.

Much recently focus on conceptual design has turned user-centric to facilitate flexible analysis of multidimensional data. Research works like (Ahmed, 2011) and (De Aguiar Ciferri, 2013) and (Viswanathan, 2011) discuss the need for user-centric data analysis. Cube Algebra proposed by (De Aguiar Ciferri, 2013) takes the cube metaphor seriously (likewise BigCube proposed by (Viswanathan, 2011)) and allow the end user to create, manipulate and query the data taking into account only the notion of cube (or n-cube). But the proposed conceptual model is not graphical and is not an extension to any existing models like E/R or UML.

The approach proposed by us in this article also reflects the user-centric approach, but goes one step further by allowing users to create and manage their own spaces to analyze the multidimensional data. Our proposal is in the form of Multidimensional Space that provides every user a space to manage dimensions and hierarchies relevant to her own requirements. We also propose a conceptual graphical model based on ME/R and MultiDimER for multidimensional spaces.

5.3 Monitoring architectures for smart building

Furthermore, concerning monitoring systems for smart building, most of existing studies focus on design and architecture side of smart buildings (Byun, 2011) and (Chen, 2009) and (Doukas, 2007). (Doukas, 2007) proposes a model for the intelligent building monitoring in which a real-time decision unit interacts with sensors for diagnosis of the building's state and with the building's controllers to select the appropriate interventions. (Chen, 2009) focuses on design of an intelligent building system that covers monitoring of energy consumption of the system, integrated building operations, occupant-aware building control. In these works, even though application parameters (user preferences) are processed and decision model is used, deployed wireless sensor devices are fixed and have static configuration during the execution.

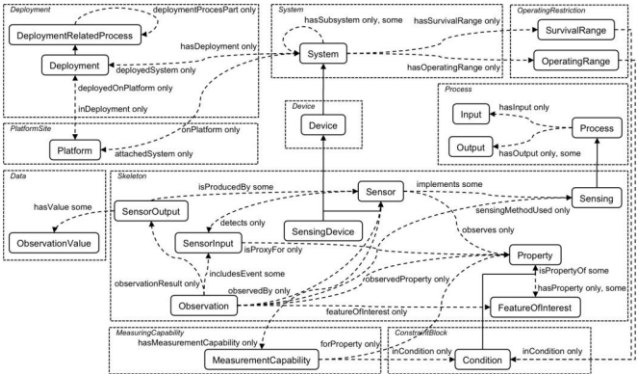

Figure 9. SSN ontology for Sensor Network

Moreover, high energy problem is not considered as a major issue. Most relevant study to our approach is (Byun, 2011). Thus, author, similar to ours, propose an approach based on self-adapting intelligent gateway mechanism for service, service management and provision of energy consumption. However, their approach does not benefit from potential reconfiguration of acquisition and transmission frequencies: sensor configuration stays static during the system lifetime. In our Smart-Service Stream-oriented Sensor Management approach, we intend to introduce an energy-aware dynamic sensor reconfiguration process while fully fulfilling application requirements.

6. CONCLUSION

Our proposed methodology for heterogeneous data integration and cross-analysis to understand complex phenomena contributes to a better understanding of urban-related phenomena through cross-disciplinary analyses of large amount of data coming from phenomenon observations issued from multiple sources (sensors, surveys, various studies). Our case of application is smart building and energy consumption.

This methodology aims at being well integrated with user-specific and domain-specific analyses processes, in particular for scientific data analyses. Analyses may be multi-dimensional through the definition of dimensions: spatial dimensions, temporal dimensions, and field-specific dimensions. Dimensions dedicated to a given phenomenon have to be identified and co-built by computer scientists and scientists from other urban-related disciplines. Our methodology is agile, incremental, iterative, and interactive to allow knowledge discovery along the way by users.

Our VGS model is generic as it enables heterogeneous data handling. The VGS model is semantically compatible with standard sensor ontologies defined by standardization organizations like Open Geospatial Consortium standards (DUL, 2010) and (Reed , 2007).

Our conceptual VGS model has been built from a real multi-disciplinary approach and its generic design makes it easy to apply to other phenomena observation. This model, as well as our agile exploration approach, is moreover independent from a specific data management technology. Although it is currently implemented on a SQL database, we aim at implementing it also on Big-Data-oriented databases like MongoDB or Cassandra.

Furthermore, concerning energy consumption and monitoring systems, we briefly point at a major challenge of pervasive environments: to take into account energy consumption of the monitoring architecture itself. In the context of smart building, we focus on the lifetime of a monitoring system based on wireless sensor devices. Most existing studies on smart building and/or pervasive environment systems do not tackle energy consumption of wireless devices and lifetime of the whole system. They adopt static configurations for wireless devices.

Our approach concerns new mechanisms for dynamic sensor reconfiguration in terms of sensor data acquisition/transmission frequencies. In this paper, we present a sustainable declarative monitoring architecture for pervasive environment. We introduce our approach of Smart-Service Stream-oriented Sensor Management based on data acquisition/transmission scheduling time patterns. The next step is to design, specify and implement time patterns to support multiple measure types on a same device and then to optimize the time pattern fusion process.

ACKNOWLEDGEMENTS

This work was supported by the LABEX IMU (ANR-10-LABX-0088) of University of Lyon, within the program “Investissements d’Avenir” (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

REFERENCES

Ahmed, E. B., Nabli, A., Gargouri, F., 2011. A survey of user-centric data warehouses: From personalization to recommendation. CoRR, abs/1107.1779.

Byun J., Park S., 2011. Development of a self-adapting intelligent system for building energy saving and context-aware smart services. *IEEE Transactions on Consumer Electronics*, vol. 57, no. 1, pp. 90–98.

Bonnet P., Gehrke J., Seshadri P., 2001. “owards Sensor Database Systems”. In: *Mobile Data Management, Conference*, LNCS Springer, pp.3-14.

Chen H., Chou P., Duri S., Lei H., Reason J., 2009. “The design and implementation of a smart building control system”. In: *ICEBE’09*. IEEE, 2009, pp. 255–262.

Compton, M. & al., 2011. The SSN Ontology of the W3C Semantic Sensor Network Incubator Group, 7p. "http://www.ict.csiro.au/staff/kerry.taylor/SSN-XG_SensorOntology.pdf"http://www.ict.csiro.au/staff/kerry.taylor/SSN-XG_SensorOntology.pdf

Da Costa R. A. G, Cugnasca C. E., 2010. Use of data warehouse to manage data from wireless sensors networks that monitor pollinators. *Eleventh International Conference on Mobile Data Management, MDM 2010*, Kanas City, Missouri, USA, 23-26 May 2010, pp. 402–406. IEEE Computer Society

De Aguiar Ciferri, C. D., Ciferri, R. R., G´omez, L. I., Schneider, M., Vaisman, A. A., Zimanyi, E., 2013. Cube algebra: A generic user-centric model and query language for OLAP cubes. *IJDWM*, 9(2):39–65

Diallo O. & al., 2012. Real-time data management on wireless sensor networks: A survey. *Journal of Network and Computer Applications*, vol. 35, no3, pp. 1013-1021.

Doukas H., Patlitzianas K. D., Iatropoulos K., Psarras J., 2007. Intelligent building energy management system using rule sets. *Journal of Building and Environment*, vol. 42, no. 10, pp. 3562–3569.

DUL: The DOLCE+DnS Ultralite ontology, 2010, "http://ontologydesignpatterns.org/wiki/Ontology:DOLCE+DnS_Ultralite"http://ontologydesignpatterns.org/wiki/Ontology:DOLCE+DnS_Ultralite

EIA, The world business council for sustainable development. *Annual Energy Review* 2010, 2011.

Eichholtz, P., Kok,N., and Quigley J. M., 2013. The economics of green building. *Review of Economics and Statistics*, vol. 95, no. 1, pp. 50–63.

Gökçe H-Y., Wang Y., Gökçe K., Menzel K. 2009. A data-warehouse architecture supporting energy management of buildings, CIB W78.

Gripay Y., Laforest F., Petit J-M., 2010. A Simple (yet Powerful) Algebra for Pervasive Environments. In: *EDBT 2010, 13th International Conference on Extending Database Technology*, Lausanne, Switzerland, pp. 1-12.

Gutierrez Rodriguez, C., Servigne, S., Laurini R.., 2007. Towards real time metadata for networked based geographic database. In: *ISSDQ2007, 5th International Symposium of Spatial Data Quality, Enschelde*, 8p.

Kimball R., Ross M., 2011. *The data warehouse toolkit: the complete guide to dimensional modeling*, John Wiley & Sons.

Kimball, R. and Ross, M. 2011. *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons.

Kimball, R. 1996. *The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*. John Wiley.

Laurini, R., Servigne, S., Noel, G. 2005, Soft Real-Time GIS for Disaster Monitoring, ed. by P. Oosterom, S. Zlatanova, E.M. Fendel, In: GI4DM. Springer Berlin. pp. 465–479

LimeSurvey. 2015. Limesurvey. https://www.limesurvey.org/en/.

Lumineau N., Laforest F., Gripay Y., Petit J-M., 2012. Extending Conceptual Data Model for Dynamic Environment. In: *31st International Conference on Conceptual Modeling (ER*), Florence, Italy.

Noel, G., Servigne, S., 2005. Indexation multidimensionnelle de bases de données capteur temps-réel et spatio-temporelle. *Ingénierie des Systèmes d'Information*. Vol.10, n°4, pp.59-88.

Noel G., Servigne S., Laurini R. 2010. Spatial and temporal information structuring for natural risk monitoring. *GISPlanet'2005, International Conference and Exhibition on Geographic Information*, pp. 1-10

Patil N-S & al., 2011. Data aggregation in wireless sensor network. *International Journal of Service Computing and Computational Intelligence*, vol. 1, no 1, p. 7-10.

Reed C., Botts M., Davidson J., Percivall G. 2007. Ogc® sensor web enablement: overview and high level architecture, Autotestcon, IEEE, pp. 372–380.

Rodriguez, CCG, Servigne, S., 2013. Managing Sensor Data Uncertainty: A Data Quality Approach. *International Journal of Agricultural and Environmental Information Systems*, pp. 35-54.

Romero, O., Abell´o, A., 2006. A Multidimensional design by examples. In Tjoa, A. M. and Trujillo, J., editors, In: *8th International Conference, DaWaK 2006, Krakow, Poland, September 4-8, 2006*, Lecture Notes in Computer Science, Springer, volume 408, pp. 85–94.

Romero, O., Abell´o, A., 2008. A. MDBE: automatic multidimensional modeling. In: *Conceptual Modeling - ER 2008, 27th International Conference on Conceptual Modeling, Barcelona, Spain, October 20-24, 2008*. Lecture Notes in Computer Science, Volume 231, pages 534–535.

Romero, O. and Abell´o, A., 2010. A. Automatic validation of requirements to support multidimensional design. *Data Knowledge*. Eng., 69(9):917–942.

Romero, O. and Abell´o, A. 2010. A framework for multidimensional design of data warehouses from ontologies. *Data Knowledge*. Eng., 69(11):1138–1157.

Servigne S., Gripay Y., Deleuil J-M., Nguyen.C., Jay J., and al. 2015. Data science approach for a cross-disciplinary understanding of urban phenomena: Application to energy efficiency of buildings. *Procedia Engineering*, Elsevier, vol. 115, pp. 45–52.

SensorML, OpenGIS® Sensor Model Language (SensorML) Implementation Specification. Open Geospatial Consortium, 2007. pp. 1–87. http://www.w3.org/2005/Incubator/ssn/XGR-ssn-20110628/.SensorML

SoCQ4Home, 2012. SoCQ4Home Project, HYPERLINK "http://liris.cnrs.fr/socq4home/"http://liris.cnrs.fr/socq4home/

Stack, P., Menzel K., Flynn D., Ahmed A., McCormac S., 2012. A cloud-based platform for energy and maintenance management. In: *4th International Conference on Computing in Civil and Building Engineering (14th ICCCBE)*, Volume 27, pp. 1-8.

Sphinx. 2015. sphinx iq. http://www.sphinxsurvey.com/en.

Szlavecz, K., Terzis A., Ozer S., Musaloiu-Elefteri R., Cogan R., Small S., Burns R. C., Gray J., Szalay A. S. 2007. Life under your feet: An end-to-end soil ecology sensor network, database, web server, and analysis service. CoRR abs/cs/0701170.

Viswanathan, G., Schneider, M. 2011. "On the requirements for user-centric spatial data warehousing and SOLAP". In: *Database Systems for Adanced Applications - 16th International Conference, DASFAA 2011, International Workshops: GDB, SIM3, FlashDB, SNSMW, DaMEN, DQIS*. Lecture Notes in Computer Science, Springer, Volume 6637, pages 144–155.