Exploiting Aggregated Open Data from Smart Cities for the Future Internet Society

D3.1: SMART-FI Urban Data Characterization

Authors	Ehsan Valizadeh, Omer Ozdemir, Peter Nilsson, Stefan Nastic
Institution lead	SAMPAS
Version	1.0
Reviewers	Javier Cubo(UMA)
Work package	WP3
Task	T3.1
Due date	31/3/2017
Submission date	25/09/2017
Distribution level (CO, PU):	Public
Abstract	This document describes the urban data sets and data characterization aspects in smart cities for the technical work packages in SMART-FI.
Keywords	Urban Data Characterization, FIWARE Data models, Orion Context Broker, Smart Cities Ontologies
License information	This work is licensed under Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0) http://creativecommons.org/licenses/by-sa/3.0/

Document Description

Document Revision History

Version	*Date*	*Modifications Introduced*	
		Description of change	*Modified by*
v0.1	20/09/16	First draft version and TOC	Name (institution)
v0.2	08/03/17	Malaga Dataset definitions added	Omer Ozdemir (ATOS)
v0.3	31/03/17	Karlshamn Dataset definitions added	Peter Nilsson (NetPort)
V0.4	04/04/17	Linked data	Stefan Nastic (TUW)
V0.5	15/04/17	Malatya Dataset definitions added	Ehsan Valizadeh (SAMPAS)
V0.6	28/04/17	Sent to reviewer	Javier Cubo(UMA)
V0.7	20/06/17	Check the status and contributions	Ehsan Valizadeh (SAMPAS)
V0.8	15/07/17	Check the status and contributions	Ehsan Valizadeh (SAMPAS)
V0.9	20/08/17	Check the status and contributions	Ehsan Valizadeh (SAMPAS)
V1.0	25/09/17	Overall revision of document and prepare final version to release	Ehsan Valizadeh (SAMPAS)
V1.1	25/09/17	Final version	Malena Donato (ATOS)

Table of Contents

Table of Figures

Table of Tables

Terms and abbreviations

FIWARE	Middleware platform, driven by the European Union, for the development and global deployment of applications for Future Internet
SMART-FI	Provides services for smart city applications, creates business opportunities and improves public's quality of life using Open Data
ORION CONTEXT BROKER	The reference implementation of the Publish/Subscribe Context Broker GE.
Generic Enabler(GE)	Software tools offered by FIWARE
CKAN	Open-source DMS (data management system) for powering data hubs and data portals
WP	Work Package

Executive Summary

This document addresses the first point, describing the technical features of the legacy data and examines the legal aspects of ownership and possible privacy constraints on such data. The first section of the document describes the reference framework for data assessment, with a specific focus on Public Sector Information. The following sections describe Smart Cities Common Ontologies and the datasets to be made available by each partner.

Considering the dimension of many of the datasets, this document only covers the most relevant entities of the stored information. However, where available, Entity-Relationship (ER) diagrams describing the datasets as a whole may be presented.

In particular, the data assessment – carried out mainly by submitting checklists to the partners and collecting relevant information – was focused on:

- describing structure, fields and other features of the datasets;
- mapping the possible constraints (in terms of legal restrictions)
- collecting relevant information from the data holders;
- finding right-holders;
- detecting IPRs (including database rights)
- Managing the existence of personal data in the datasets

As the main next step in this thread, further effort is to be produced by the partners in order to identify the main drivers of technical interoperability and to harmonize, wherever possible, vocabularies, metadata and DBs structure.

1 Introduction

This document supposes the starting point for defining and analyzing characterized data sets for the case studies conducted by WP6 aligned with the task 3.1 objectives and FIWARE based Context Management solutions. We will discuss Smart Cities Common Ontologies and finally will examine FIWARE Components for Urban Data Characterization. Context Management solutions will be investigated and presented such as Orion Context Broker and NGSIv2 and some data models guidelines in FIWARE will be listed.

1.1 About this deliverable

This is a report to describe common smart city ontologies and linked to open data sets to describe the urban environment with focus on initial SMART-FI pilot's data sets definitions.

1.2 Document structure

In Section 1 we present an introduction on document. Section 2 is about reference framework for data assessment, with a specific focus on Public Sector Information. In Section 3, we will present smart city common ontologies and state of the art. Section 4 is about Linked Data Technologies and section 5 contains initial SMART-FI Pilot Data Sets is going to be used within use cases implementations and Section 6 is going to cover FIWARE Components for Urban Data Characterization. Then, Section 7 presents the conclusion and future work.

2 Reference Framework for Urban Data Characterization

2.1 Public Sector Information

Public Sector Information (hereinafter PSI) encompasses all information held by a public sector organization, e.g. a government department or, more in general, any entity which is majority owned and/or controlled by a Public Administration. Datasets held and managed by public sector bodies may be the direct outcome of their institutional mandate (e.g. the records of a firm registry) or a 'by-product' of the public task (like, for instance, the opening hours of the shops in a given area). In its wider extent, PSI covers all domains describing the actors that populate society, their interaction, the environment (e.g. meteorological data or geodata) and the functioning of the public sector (its budget for instance) with frequent or even constant updates.

The awareness on the importance of PSI dissemination and reuse as a driver for transparency, democratic participation and economic opportunities is indeed increasing within Public Administrations and the society as a whole. National, regional and local portals for PSI reuse are being implemented in Europe, supplying datasets covering relevant sectors. In some cases, datasets are made available in digital formats (machine readable and, in the best cases, semantically linkable with each other) through web catalogues. In this case, information is somehow 'new', never released before, and its reuse enables the creation of innovative services (such as apps and the like). This fosters 'low-end' markets of applications, some of which are able generate a broad demand. On the other hand, Public Sector bodies hold and manage huge amounts of information ('Big Data') already feeding mature markets. Some key examples, just to mention a few, are land registries and firm registries. In those cases, downstream markets are populated by a limited number of medium-high sized players.

The main features of PSI in a digital environment imply a high potential for reuse, both for commercial and non-commercial purposes. In fact

- making PSI available in open formats is supposed to foster both simultaneous innovation, i.e. 'forking' projects or services and cumulative innovation / 'inventing around';
- an open approach for PSI panders the high degree of componentization of applications, products and services supplied downstream, which are likely to be grounded on multiple and heterogeneous sources, including but not limited to PSI;
- supply- and demand-side economies of scale and scope may indeed represent a tangible driver both for PSI holders and re-users. Once adopted a proper technical interface, which may partially overlap with the internal interchange infrastructure, the former are able to open new datasets with small incremental effort.

The market value of PSI reuse is substantial. According to a recent study by the European Commission (‘Vickery report’) the aggregate direct and indirect economic impacts from PSI applications and use across the whole EU28 economy are estimated to be around 140 billion euros annually. Further analysis suggests that if PSI policies were open, with easy access for free or marginal cost of distribution, direct PSI use and re-use activities could increase by up to 40 billion euros for the EU27.

The main aspects of PSI dissemination and reuse are covered by the Directive 2013/37/EU, (hereinafter ‘PSI Directive), adopted by the Member States in the following years and currently under review in light of a public consultation. The PSI Directive provides key definitions and prescriptions on data formats, charging principles, transparency, licensing and practical arrangements related with PSI, with the aim to set the stage for a pan-European, transparent and non-discriminatory framework for PSI reuse.

2.2 Open Data & Standardization

A broad set of analyses show that in order to reach the objectives of enabling and improving democratic participation, transparency and new opportunities to create value-added services grounded on it, PSI has to be disseminated as Open Data. Quoting the Open Definition provided by the Open Knowledge Foundation, we assume that "A piece of content or data is open if anyone is free to use, reuse, and redistribute it - subject only, at most, to the requirement to attribute and share-alike". This definition implicitly encompasses:

- technological aspects , meaning that datasets have to be machine readable and made available in open formats;
- legal aspects, meaning that datasets have not subject to intellectual property rights or other legal constraints preventing copy, redistribution and reuse.

This paragraph focuses on the latter, and in particular on two key issues: Intellectual Property Rights (IPRs) and privacy / sensitive data.

2.3 Licensing Models & Intellectual Property Rights

In fact, in most cases, it is reasonable to assume that certain IPRs (copyright and/or sui generis rights on databases) apply to the datasets held by Public Sector bodies, even if no legal entity or individual has explicitly claimed any rights on them. In those cases, from the viewpoint of a reuser, it is required to ask for permission to the copyright owner to achieve all basic activities related with reuse, such as data extraction, reproduction, update, mash-up with other sources, etc., even for noncommercial purposes. Thus, as a default rule and unless explicitly stated, datasets held by Public Sector bodies fall under the "all rights reserved" framework. A data owner has therefore to address two subsequent sets of issues:

(i) whether the public entity can dispose of such data and allow reuse;
(ii) (ii) if so, how to manage those rights.

The first point recalls the notion of 'rights clearance', i.e. the analysis of the legal status of a database in order to verify the actual existence of IPRs related to it (e.g. from a third party) and assess whether and how data can be opened. The second point relates with the licensing schemes adopted by the data holder, where a license is a legal instrument explaining the terms of (re)use of the datasets. If the 'rights clearance' activity has been accurate, applying a license taking into account the legal status of the databases can be trivial. However, the possibility of legally combining data from different sources and/or making derivative works from the original source to create new services is indeed a key driver of reuse. Being legal interoperability crucial, using standard licenses such as Creative Commons (with several combinations of attributes), Open Data Commons or other National Licenses specifically designed for PSI is highly recommended. The choice of the license is grounded, as a baseline, on the outcome of the 'rights clearance' analysis and, subsequently, on the 'openness degree' the PSI holder intends to apply to the datasets, also considering legal interoperability with datasets made available by others.

2.3.1 Open Data Licensing

A broad set of analyses show that in order to reach the objectives of enabling and improving democratic participation, transparency and new opportunities to create value-added services grounded on it, PSI has to be disseminated as Open Data. Quoting the Open Definition provided by the Open Knowledge Foundation, we assume that "A piece of content or data is open if anyone is free to use, reuse, and redistribute it - subject only, at most, to the requirement to attribute and share-alike". This definition implicitly encompasses:

- Technological aspects , meaning that datasets have to be machine readable and made available in open formats;
- Legal aspects, meaning that datasets have not subject to intellectual property rights or other legal constraints preventing copy, redistribution and reuse.

The overall assessment of the level of Open Data Maturity for each European country is divided into two key indicators: Open Data Readiness and Portal Maturity.

The first key indicator – Open Data Readiness – assesses to what extent countries have an Open Data policy in place, licensing norms and the extent of national coordination regarding guidelines and setting common approaches. In addition, the transposition of the revised PSI Directive is taken into account. Besides the presence of an Open Data policy, the use made of the Open Data available and the estimated political, social and economic impact of Open Data are assessed. The combination of these three sub-indicators provides a good overview of how ready a country is in terms of its Open Data policy, thus called Open Data Readiness.

The second key indicator – Portal Maturity – assesses the usability of the portal regarding the availability of functionalities, the overall re-usability of data such as machine readability and accessibility of data sets, for example, as well as the spread of data across domains.

Databases and Creative Commons

Much of the potential value of data is to society at large — more data has the potential to facilitate enhanced scientific collaboration and reproducibility, more efficient markets, increased government and corporate transparency, and overall to speed discovery and understanding of solutions to planetary and societal needs.

A big part of the potential value of data, in particular its society-wide value, is realized by use across organizational boundaries. How does this occur legally? Many sites give narrow permission to use data via terms of service. Much ad hoc data sharing also occurs among researchers. And increasingly, sharing of data is facilitated by distribution under standard, public legal tools used to manage copyright and similar restrictions that might otherwise limit dissemination or reuse of data, e.g. CC licenses or the CC0 public domain dedication.

Many organizations, institutions, and governments are using CC tools for data. For case studies about how these tools are applied, see:

Uses of CC Licenses with Data and Databases

Uses of CC0 with Data and Databases

The Open Database License (ODbL) and the Open Data Commons Attribution License (ODC-BY) are licenses designed specifically for use on databases and not on other types of material. There are many differences between those licenses and CC licenses, but the most important to be aware of relate to license scope and operation. The ODC licenses apply only to sui generis database rights and any copyright in the database structure, they do not apply to the individual contents of the database. The latest version of the CC licenses on the other hand apply to sui generis database rights and all copyright and neighboring rights in the database structure as well as the contents.

This paragraph focuses on the latter, and in particular on two key issues: Intellectual Property Rights (IPRs) and privacy / sensitive data.

2.4 Privacy and Personal Data

The decision to make available certain public data should not (and cannot, by law) put into question the right of citizens to protect their personal data: the regulations on the reuse of data is totally affect the rules on personal data. Personal data is defined by Directive 95/46/EC (on the protection of individuals with regard to the processing of personal data and on the free movement of such data), Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity". More in general, the PSI Directive is fully compliant with the data protection rules contained in Directive 95/46/EC. Thus, when a dataset contains personal data, the most plausible hypothesis is that the access and reuse are to be discarded. Indeed, a broad set of PSI potentially useful to reuse, contain personal data. Sensitive personal data (revealing racial or ethnic, religious, philosophical or other beliefs, political opinions, membership of political parties, unions, associations or organizations of a religious, philosophical, political union, as well as personal data disclosing health and sex life) cannot be disclosed in any circumstance. However, under certain conditions, other kinds of personal data can be made accessible and their reuse can still be allowed. In those cases, it is important that the scope of the PSI reuse is consistent (or, at least, is not inconsistent) with the legislation and the original purposes under which the collection of personal data by the PSI holder has been made; this information has therefore to be made explicit through a public disclaimer addressed to reusers. Upstream, a clear privacy disclaimer addressed to the persons providing personal data, mentioning that such data will be made available via Internet and potentially reused, is moreover suitable.

A more 'radical' (since it entails the loss of possibly useful information) way to address the case of personal data is a technical solution that makes possible the publication of data sets (allowing reuse) by eliminating all references to individuals, so that all related data cannot be retrieved. The anonymisation of datasets encompasses several possible techniques, including the aggregation of information.

In June 2016, the European Data Portal published a report[1] on Open Data and privacy where guidelines are provided to promote the utility of data while ensuring data controllers' obligation to respect the right of data subjects to personal data protection, this report has developed a series of 8 recommendations, as follows:

[1] Open Data Maturity in Europe 2016: https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n2_2016.pdf

1. Understand the data. Consider potential use cases, the value of the data and potential risks.
2. Consult. Engage stakeholders about the publication programme, be mindful of additional risks that are identified.
3. Remember the three pillars of privacy, data protection and public confidence.
4. Be very sure of the grounds for publishing personal data.
5. Anonymise well and thoroughly. Follow guidelines for anonymising personal data.
6. Remember utility. There is no point publishing data which has been denuded of serious content.
7. Don’t release and forget. Anonymisation and Open Data are not cheap options.
8. Have a plan in place in the event of a problem. Be not only transparent, but also transparent about your transparency

3 Smart Cities & Open Data

One major opportunity when building smart cities is creating an interactive data model to engage citizens and businesses. Through collaboration initiatives, such as hackathons, new ideas for the city can be developed. To support citizen entrepreneurship, from residential to business users, CDOs must build an open innovation and developer forum tailored to their city's requirements. As a result, data scientists and application developers can use data from open-city data portals to create insights and services.

"API-driven access like this can also ensure an enforcement point for data governance, which should be in place at the start," Ms. Tratz-Ryan said. "CitySDK, a project by the European Commission, and the Smart Nation API coLab from Singapore are two examples already in progress."

Open-data governance is paramount to building a smart city. Preparing for privacy issues, accommodating informative visualization and fostering a culture of open innovation will help CDOs, CIOs and urban planners to succeed in this quest.

3.1 Smart Cities Common Ontologies

The "Smart City Ontology" or "Intelligent City Ontology" describes cities and city districts that have adopted and implemented the intelligent city planning paradigm.

3.1.1 Ontologies overview

Basically, an ontology is a philosophical term, dealing with the description of entities, their grouping or hierarchy, their characteristics and their interrelationships. An entity, in this context, is something that exists, physically or abstract, and can be used to understand and describe the world or a particular domain of interest.

In computer science, and to a developer or engineer, the definition of ontology is more understandable and useful, since this resembles the art of data modelling. An ontology, in this context, is still a formal description and naming of entities and their grouping/hierarchy, properties and interrelationships. However, the mapping to well-known modelling concepts in computer/information science, such as object oriented modelling using classes with inheritance and relations, or Entity-Relationships modelling is obvious.

The ontology model should be implementation independent, since the main objective with the ontology is to capture the terminology and vocabulary used in a certain context or domain.

Since a large domain can often be divided into several related domains/subdomains the same applies to ontologies. There are often several more or less related ontologies describing a large domain.

A simple approach to implement an ontology could be to define an API with a protocol, for example based on REST/HTTP, together with corresponding data models for the payload (requests/responses) using for example JSON. This approach is more static and doesn't for example easily adapt to new and changed data types and properties.

In the context of the Semantic Web, the implementation of an ontology involves describing the resources, their properties and relationships with RDF using for example OWL or RDFS as an authoring language and SPARQL for queries to the RDF data. This approach exploits the full potential of ontology techniques.

3.1.2 Ontology examples

At present, large-scale urban data are produced from diverse sources and increasingly become publicly available from governmental authorities. Each source reflects a specific aspect of the urban environment and the generated datasets vary in scale, speed, quality, format, and, most importantly, semantics. In principle, municipal data are mostly static or semi-static, in the form of demographic records, such as census statistics on population, age and gender distribution, average income, and land uses. Besides static records, planning stakeholders gradually utilize dynamic streams of information stemming from sensor resources, in relation to transportation flows, weather status, and environmental conditions. However, in the current context, city related organizations create and operate on datasets based on each sector's particular purposes and problems at hand. Any correlation of information across different departments is presently performed in a manual fashion, hence requiring great amounts of time and effort.

Some examples of public ontologies are:

- BabelNet (http://babelnet.org), a very large multilingual semantic network and ontology, lexicalized in many languages
- Dublin Core (http://dublincore.org), a simple ontology for documents and publishing
- Friend of a Friend (http://www.foaf-project.org), an ontology for describing persons, their activities and their relations to other people and objects

3.1.3 Ontologies for Smart Cities - initiatives and standards

There are many initiatives and standard attempts to ontologies related to Smart Cities, some examples are:

- **City Protocol Society** – E.g. "City Anatomy Ontology", http://cityprotocol.org/publications/
- **The British Standards Intitution** - "PAS 182 Smart City Concept Model", https://www.bsigroup.com/en-GB/smart-cities/Smart-Cities-Standards-and-Publication/PAS-182-smart-cities-data-concept-model/
- **ETSI** – "SmartM2M, Smart Applicances, Reference Ontology and oneM2M mapping", http://www.etsi.org/deliver/etsi_ts/103200_103299/103264/01.01.01_60/ts_103264v010101p.pdf
- **ISO 30145** – **ongoing** standardization work on "Smart City ICT Reference Framework", https://standards.incits.org/apps/group_public/download.php/78819/ISO-IECJTC1-WG11_N0048_Initial_Working_Draft_of_30145_Part_2.pdf
- **KM4CITY:** Km4City platform is a Smart City tools for implementing the city vision, monitoring the city evolution, diffusely providing new services for improving quality of life of city users, for the city economic grow; stimulating city users; since an attractive city is a `city that produces` in which users are happy and proud, http://www.disit.org/drupal/?q=node/6056
  http://www.km4city.org

3.2 SMART-FI Smart Cities Ontology

SMART-FI will produce a set of facilities aligned to FIWARE platform. As an open source platform, supported by companies, universities and research institutions, the FIWARE platform will play a key role in 21 the cities of the future . Its massive adoption may help to speed up the replica- tion of key components for setting up and consolidating the smart city ecosystem. The FIWARE cloud and software platform is a good catalyst for an open ecosystem of entrepreneurs aiming at developing state-of-the-art data-driven applications. This ecosystem is formed by application developers, technology and infrastructure providers and entities who aim to leverage the impact of developing new applications based on the data they produce and publish. In this context cities will play a unique role, especially those implementing Smart City strategy, which open up their data to facilitate the creation of applications built by developers that form part of this ecosystem. FIWARE enables quick and easy application development because they make use of prefabricated components known as generic enablers in its cloud, sharing their own data as well as accessing open data from cities. In SMART-FI platform will uses several FIWARE generic enablers, such as CKAN,COSMOS,IDM, Cygnus and Orion Context Broker. For example, CKAN is intended for data publishers, e.g., national and regional governments, companies and organizations that want to make their data open and available. IDM covers a number of aspects involving users' access to networks, services and applications, includ- ing secure and private authentication from users to devices, networks and services, authorization and trust management, user profile management, privacy-preserving disposition of personal data, Single Sign-On (SSO) to service domains and Identity Federation towards applications. IDAS enable connecting IoT devices to FIWARE-based solutions, by translating IoT-specific protocols into a standard- ized NGSI context information protocol. Finally, Wirecloud GE builds on cutting- edge end-user development, RIA and semantic technologies in order to offer a next- generation end-user centred web application mashup platform aimed at leveraging the long tail of the Internet of Services.

4 Linked Data Technologies

4.1 Linked Data

Linked Data is a term coined by Tim Berners-Lee in 2006 in the context of the Semantic Web2. The Semantic Web is an ongoing effort to extend the Web with standards for integration and combination of heterogeneous Web data sources, and for representing semantic knowledge of how data relates to real world objects3. In this context, Linked Data refers to a set of best practices for publishing data on the Web. Specifically, it defines rules that foster the creation of a Web of data: a machine-readable browsable graph of things (concepts or objects), their properties, and their relations to other things4.

This Web of data enables completely new types of applications. For example, Smart City or Open Government Data and applications can greatly benefit from Linked Open Data (LOD) which is Linked Data released under an open license.

4.2 Linked Data Technologies

To make linked data available to an even broader range of applications, different linked data technologies and extensions have emerged. In the following, we focus on two technologies, which are particularly relevant in the context of SMART-FI project, namely Linked USDL and the FIWARE support for Linked data.

4.2.1 Linked USDL

With the advent of service-oriented architecture came the need for formal service descriptions, which the Web Service Description Language (WSDL) initially satisfied. As the amount and variety of services grew, so did the need for automated discovery and composition of services. The Universal Service Description Language (USDL)5 facilitate this by enriching WSDL documents with semantic information. This and other research on service descriptions has focused mainly on software-based services, and has disregarded the domain of real-world services and eCommerce.

Linked USDL is an effort to support Web-scale automated trading of cross-domain services6. It builds on existing research on Semantic Web technologies, and defines a new service description vocabulary based on USDL and Linked Data concepts. Specifically, it provides a comprehensive ontology that allows concise modeling of services, the types of entities involved, the entities' business role, etc. Linked USDL aims to provide means of comparing, discovering, and trading of services, and has found use in many research projects, among them FIWARE.

[2]Berners-Lee, T. (2006). Linked Data, *Design Issues. W3C. (Online) https://www.w3.org/DesignIssues/LinkedData.html*, Accessed 2017-04-03.
[3]"W3C Semantic Web Activity". World Wide Web Consortium (W3C). (Online) https://www.w3.org/2001/sw/, Accessed 2017-04-03.
[4]Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *Semantic services, interoperability and web applications: emerging concepts*, 205-227.
[5]Simon, L., Mallya, A., Bansal, A., Gupta, G. and Hite, T.D., 2005, July. A universal service description language. In Web Services, 2005. ICWS 2005. Proceedings. 2005 IEEE International Conference on. IEEE.
[6]Pedrinaci, C., Cardoso, J. and Leidig, T., 2014, May. Linked USDL: a vocabulary for web-scale service trading. In European Semantic Web Conference (pp. 68-82). Springer International Publishing.

4.3 Linked Data & FIWARE

Generally, the FIWARE is a Cloud-based middleware platform that provides a set of open standard APIs aimed at Future Internet applications7. The FIWARE Catalogue8 contains a rich component library that allows developers to create applications for, e.g., IoT scenarios or Big Data analytics, using the FIWARE platform. The platform integrates Linked Data technologies, such as Linked USDL, as open standards. An example is the IoT Discovery GE specification, which allows the registration and discovery of IoT devices using Linked Data technologies. In particular, it exposes the Sense2Web9 platform that provides a repository to register and manage semantic descriptions of IoT objects. Providers can link descriptors to other LOD resources on the Web, allowing context consumers to discover IoT objects using Semantic Web search techniques.

[7] Glikson, A., 2011. Fi-ware: Core platform for future internet applications. In Proceedings of the 4th Annual International Conference on Systems and Storage.
[8] "The FIWARE Catalogue" FIWARE. (Online) https://catalogue.fiware.org/, Accessed 2017-04-02
[9] Barnaghia, P., Ganza, F., Abangara, H., Presserb, M. and Moessnera, K., 2011. Sense2web: A linked data platform for semantic sensor networks.

5 Heterogeneous SMART-FI Pilot Data Sets

This section describes the data sets will be used in SMART-FI pilot scenarios.

5.1 MALAGA Data Sets

The data set which we will be explaining next is being used as the main entity among flows between different components in Malaga city CityGO application.

This final data set is built up by using legacy data sets which are exposed by Malaga Municipality.

5.1.1 Legacy Data Environment

We are using 3 main data sets which are exposed by Malaga Municipality for traffic management:

- Bicycle stations data
  - GPS location/address of the stations and the real time data of free/occupied bicycle slots
- Bus lines and stations data
  - Each bus line data(name of the line and number of the bus) for the bus routes in the city, GPS location/address of the bus stations(There is no dynamic data provided by the municipality for this section)
- Vehicle parking stations data
  - GPS location/address of the vehicle station with the real time data of the free/occupied slots

5.1.2 Data Set Descriptions, Schema & ER Diagrams

We are using FIWARE Data Models repository[10]. to generate data sets for CityGO application environment.

This repository contains:

- code that allows to expose different harmonized datasets useful for different applications. Such datasets are exposed through the FIWARE NGSI version[11] 2 API (query).
- JSON Schemas and documentation on harmonized datamodels for smart cities, developed jointly with OASC[12], and other domains.

We have transformed the 3 datasets mentioned in the last section (*5.1.1 Legacy Data Environment*) to generate a single dataset that is built on top of FIWARE open data model standards. Determine the output formats for each single data set and designing the transformation needed

[10] https://github.com/Fiware/dataModels

[11] http://fiware.github.io/specifications/ngsiv2/stable

[12] http://oascities.org/

5.1.2.1 CityGO Dataset

We will now introduce our data set which is the main source for the intercommunication between CityGO mobile app, backend server and the dashboard to visualize the statics based on citizen movement(by bus, bicycle or vehicle) in the city. It is mostly built upon by GPS coordinates of the person using mobile application, the route and the transportation type that the user takes.

Here is the list of attributes that define the data set:

Table 1 CityGO Data Set

	Name	Description	Type
Attribute 1	id	Entity's unique identifier	INTEGER
Attribute 2	type	Entity type	COMMUTE R
Attribute 3	username	User name for the commuter	STRING
Attribute 4	location	Commuter's last known location represented by a GeoJSON Point	GEO:JSON
Attribute 5	route	Structured attribute that includes all the attributes for the suggested route.	GEO:JSON
Attribute 6	route:start	Starting point of the suggested route, represented by a GeoJSON Point.	GEO:JSON
Attribute 7	route:end	End point of the suggested route, represented by a GeoJSON Point.	GEO:JSON
Attribute 8	route:transportation	Structured attribute that includes all the attributes for the suggested route.	STRUCTU RED
Attribute 9	onType	Name of the transportation type	STRING
Attribute 10	route:transportation :busRouteNumbers	Array of strings which represents the bus line numbers	STRING ARRAY
Attribute 11	route:estimatedDist ance	Denotes the estimated distance for the suggested route and is specified in Kilometers. If provided, the value of the distance attribute must be a non-negative real number. null MAY be used if distance is transiently unknown for some reason.	NUMBER
Attribute 12	route:estimatedTim e	Denotes the estimated duration for the suggested route and is specified in minutes. If provided, the value of the timeestimated attribute must be a non-negative real number. null MAY be used if timestimated is transiently unknown for some reason.	NUMBER
Attribute 13	route:confirmed	Denotes whether the commuter has confirmed in their mobile app that they have taken the suggested route and is specified as strings Yes or No. If provided, the value of the confirmed attribute must be the strings Yes or No.	NUMBER
Attribute 14	dateModified	Last update timestamp of this entity.	DATETIME
Attribute 15	dateCreated	Creation timestamp of this entity	DATETIME

5.1.3 Data Ownership & Legal Considerations

The ownership of the data is Malaga municipality. Data collecting by municipality also. This datasets includes real time bicycle station information, vehicle parking areas and bus lines, social event and other sources of data from open source portal which are managed by municipality under the terms of the Creative Commons Attribution-ShareAlike 3.0 Generic License and we do not foresee any legality problems.

Weather forecast service is provided by OpenWeatherMap and free weather API is provided under the terms of the Creative Commons Attribution-ShareAlike 4.0 Generic License.

5.2 MALATYA Data Sets

Pilot datasets are provided by Malatya municipality databases, data services and open data published on the web.

5.2.1 Legacy Data Environment

Datasets for Malatya pilot consist of different categories as:

- POIs: Data services for all important places information for Malatya including districts, touristic places, on duty pharmacies, museums, education zones, religious facilities, cultural zones, government building, health zones, transportation routes, accommodation, business districts and green areas in spatial format,
- Business Information: retail industry information for all businesses (restaurants, shops, entertainment places, accommodation, etc.) and enables citizens to share their comments, feedbacks and get recommendations based on time of the day, location, type, ratings, their personalized historical records, interested keywords and using semantics based search capabilities,
- Request and Complaint: Inputs of users such as scorings, comments, and feedbacks for important places will be recorded via MalatyaInsight application in a way that is accessible to other users. In addition, complaints and demands of citizens will be reported to Municipality directly.
- Other sources of data: The accessible information, scores and comments gathered by other social media tools can be transferred to application.
    - Malatya Metropolitan Municipality Web Site http://www.malatya.bel.tr
    - Investment Projects in Malatya: http://www.malatyagelisiyor.com/
    - E-Municipality Services: https://webportal.malatya.bel.tr/web/guest/2

5.2.2 Data Set Descriptions, Schema & ER Diagrams

Here is the list of data sets and attributes defines the data sets:

Table 2 POI Categories Data Set

Data Set 1	**POI Categories**		
Social Institutions, Cultural Institutions, Government Institutions, Educational Institutions, etc.)			
	Name	**Description**	**Type**
Attribute 1	categoryId	Category Id	string
Attribute 2	categoryName	Category Name	string
Attribute 3	description	Description	string

Table 3 POI List Data Set

Data Set 2	POI List		
Important Places in Malatya such as districts, touristic places, museums, hospitals, schools, pharmacies, religious facilities, recreation areas, parks, education zones, cultural zones, government buildings, transportation routes, accommodation, business districts etc.			
	Name	**Description**	**Type**
Attribute 1	poiId	POI ID	string
Attribute 2	name	Name	string
Attribute 3	description	Description	string
Attribute 4	coordinate	Latitude + Longtitude	CoordinateType
Attribute 5	phoneNumber	PhoneCode + PhoneNumber	string
Attribute 6	subDoorNumber	Sub Door Number	string
Attribute 7	doorNumber	Door Number	string

Attribute 8	street	Street	ComponentType
Attribute 10	district	District	ComponentType
Attribute 11	categoryId	Category Id	string

Table 4 Business Office Information Data Set

Data Set 3	**Business Office Information**		
Services for all businesses in Malatya including retail products and services of these businesses (restaurants, shops, entertainment places, accommodation, etc.)			
	Name	**Description**	**Type**
	businessId	Business ID	integer
Attribute 1	title	title	string
Attribute 2	email	Email	string
Attribute 3	businessOfficeNumber	Business Office Number	string
Attribute 4	registerNumber	Register Number	string
Attribute 5	countOfEmployee	Employees Count	string
Attribute 6	phone	Phone	string
Attribute 7	name	Name	string
Attribute 8	surname	Surname	string
Attribute 10	identificationNumber	Identification Number	string
Attribute 11	signTitle	Sign Title	string
Attribute 12	commercialTitle	Commercial Title	string
Attribute 13	taxOffice(CT)	Tax Office	string
Attribute 14	taxIdentityNumber	Tax Identity Number	string
Attribute 16	socialSecurityNumber	Social Security Number	string

Table 5 Request and Complaint Data Set

Data Set 4	**Request and Complaint**		
Service Request and Complaint Data			
	Name	**Description**	**Type**
Attribute 0	name	Name	string
Attribute 1	surname	Surname	string
Attribute 2	applyNumber	Ref. Number	string
Attribute 3	phoneCode	Phone Code	string
Attribute 4	phoneNumber	Telefon Number	string
Attribute 5	subjectCode	Subject Code	string
Attribute 6	subjectDescription	Subject Description	string
Attribute 7	applyDate	Record Date	DateTime
Attribute 8	applyKind	Record Type	string
Attribute 10	departmentCode	Department Code	string
Attribute 11	departmentName	Department Name	string
Attribute 12	littleDescription	Brief Description	string
Attribute 13	doorNumber	Door Number	string
Attribute 14	subDoorNumber	Sub Door Number	string
Attribute 16	Street (CT)	Street	ComponentType
Attribute 17	District (CT)	District	ComponentType

5.2.3 Data Ownership & Legal Considerations

The ownership of the data is Malatya municipality. Data collecting by municipality also. This datasets includes POIs, Business Information, Request and Complaint and other sources of data from municipality web site and we do not foresee any legality problems and no license is attached to the published data.

5.3 KARLSHAMN Data Sets

Karlshamn is running two different use cases, based on different data sets.

- Transportation use case – Data from Blekingetrafiken (the bus company)

- Energy use case – Data from Karlshamnsfastigheter/Raybased (the building where the energy pilot is implemented)

5.3.1 Legacy Data Environment

For the transportation use case, we are using different open datasets/services to be exposed by Blekingetrafiken, providing for example:

- Bus data
  - Bus number
  - Other static vehicle data
- Real time bus data
  - GPS location of the buses
  - Speed
  - Route
- Bus stations data
  - GPS location of the bus stations
- Bus routes data
  - Route number
  - Departure information
  - Arrival information
- Time tables data
  - Route
  - Direction
  - Time period

For the energy use case, we are using another set of open datasets/services to be exposed by Karlshamnsfastigheter/Raybased, providing for example:

- Nodes
  - Name
  - Address
  - Status
  - Power consumption
- Sensor measurements, such as:
  - Precence
  - Temperature
  - Illuminance
  - Humidity
- Light monitoring and control
- Power Outlet monitoring and control

5.3.2 Data Set Descriptions, Schema & ER Diagrams

Here is the list of data sets and attributes defines the data sets:

5.3.2.1 Transportation Data Sets from Blekingetrafiken

Table 6 RealTimeVehicleData Data Set

Dataset 1	**RealTimeVehicleData**		
Real time data from buses			
	Name	Description	Type
Attribute 1	Latitude	Latitude	DECIMAL

Attribute 2	Longitude	Longitude	DECIMAL
Attribute 3	Speed	Speed	DECIMAL
Attribute 4	DoorOpen	Door open	BOOLEAN
Attribute 5	DistanceToNextStop	Distance in meter to next stop	DECIMAL
Attribute 6	Deviation	Deviation	DECIMAL
Attribute 7	Route	Route Number	STRING
Attribute 8	InventoryNumber	Inventory number	STRING
Attribute 9	ConnectingRoutes	Connecting routes/lines	STRING

Table 7 RouteInformation Data Set

Dataset 2	**RouteInformation**		
Route information from traffic planning system REBUS3			
	Name	Description	Type
Attribute 1	Route	Route number	STRING
Attribute 2	Variantnr	Routes and tours relations	INTEGER
Attribute 3	Left	Drive variant identifier	INTEGER
Attribute 4	LeftText	Drive variant description	STRING
Attribute 5	StartDestinationId	Destination at the departure	INTEGER
Attribute 6	StartDestinationDesciption	Destination description	STRING
Attribute 7	DepartureTime	Departure time	STRING
Attribute 8	DepartureStop	Departure stop	STRING
Attribute 10	ArrivalTime	Arrival time	STRING
Attribute 11	ArrivalStop	Arrival Stop	STRING

Table 8 Stop Data Set

Dataset 3	**Stop**		
Stop information from traffic planning system REBUS3			
	Name	**Description**	**Type**
Attribute 1	StopId	Stop Id	STRING
Attribute 2	Name	Stop name	STRING
Attribute 3	DistanceToNextStop	Distance in meter to next stop	DECIMAL
Attribute 4	Location	Stop location	STRING
Attribute 5	Speed	If exist - affects interpolation	INTEGER
Attribute 6	Rules	Rules exist	BOOLEAN
Attribute 7	VSec	Variant sequence	INTEGER
Attribute 8	Disembarking	Disembarking	BOOLEAN
Attribute 9	ExclamationNumber	Exclamation number	INTEGER
Attribute 10	ExclamationName	Exclamation name id	INTEGER
Attribute 11	Designation	Designation	STRING
Attribute 12	NorthingEasting	Coordinates for the designation	STRING

Table 9 TimeTable Data Set

Dataset 4	**TimeTable**	
Time Table information from traffic planning system REBUS3		

	Name	Description	Type
Attribute 1	TablePeriod	Time table period	STRING
Attribute 2	TableIndex	Time Table period index	INTEGER
Attribute 3	Route	Route number	STRING
Attribute 4	RadioCode	RadioCode	INTEGER
Attribute 5	Alias	External name	STRING
Attribute 6	RouteType	Route category	INTEGER
Attribute 7	RouteHeading	Description	STRING
Attribute 8	TransportAuthority	Transport authority	STRING
Attribute 10	TransportType	Transport Type/Mode	INTEGER
Attribute 11	TransportTypeDescription	Transport Type/Mode description	STRING
Attribute 12	Direction	Direction	STRING

5.3.2.2 Energy Data Sets from Karlshamnsfastigheter/Raybased

Table 10 Nodes Data Set

Dataset 5	**Nodes**		
Node descriptions			
	Name	**Description**	**Type**
Attribute 1	NodeID	Primary Key	STRING
Attribute 2	Description	Textual description of the node	STRING
Attribute 3	Address	Address of the node (text in hexadecimal format)	STRING
Attribute 4	Active	If set to one, the node is monitored	NUMBER
Attribute 5	Status	Status of the node ("OK" or "Timeout")	STRING
Attribute 6	RelayStatus	Relay status (0 – relay off, 1 – relay off)	NUMBER
Attribute 7	ButtonStatus	Button status (bit 0 - button 1, bit 1 – button 2, etc) 0: button not pressed, 1: button pressed	NUMBER
Attribute 8	PowerConsumtion	Power consumption of the node (in Watts)	NUMBER
Attribute 9	LastEnergy	Last energy value reported from node (in Wattseconds) Used to see how much additional energy that has been consumed	NUMBER

Table 11 Lights Data Set

Dataset 6	**Lights**		
All lights in the system			
	Name	**Description**	**Type**
Attribute 1	LightID	Primary Key	STRING
Attribute 2	Address	Textual decimal light address. 0: relay on the puck. 1-255: DALI address	STRING
Attribute 3	Name	Name displayed in the App for this light	STRING
Attribute 4	DimLevel	Light level of light range 0 to 100 percent. Only used for dimmable lights	NUMBER
Attribute 5	ColorLevel	Light color range 0 to 8	NUMBER

Attribute 6	NodeID	The light is connected to this node References to NodeID in the nodes table	STRING
Attribute 7	Digital	0: Dimmable 1: Not dimmable	NUMBER
Attribute 8	Value	For digital lights 0 is off and 100 is ON otherwise it is the same as DimLevel	NUMBER

Table 12 Measurements Data Set

Dataset 7	**Measurements**		
Measurements from sensors			
	Name	**Description**	**Type**
Attribute 1	MeasurementID	Primary Key	STRING
Attribute 2	TimeStamp	Timestamp when the measurement was received	TIMESTAMP
Attribute 3	Temperature	Sensor temperature (in Celsius with at most two decimals)	NUMBER
Attribute 4	Humidity	Sensor relative humidity (in percent)	NUMBER
Attribute 5	Presence	Sensor presence (0 – no presence, 1 – presence)	NUMBER
Attribute 6	Illuminance	Sensor Illuminance (in lux)	NUMBER
Attribute 7	NodeID	The measurement is connected to this node. References to NodeID in the nodes table	STRING

5.3.3 Data Ownership & Legal Considerations

- The ownership of the data for the two different use cases are Karlshamn municipality and the local bus company Blekingetrafiken, respectively. We do not foresee any legality problems and no license is attached to the published data.

6 Data Characterization & Homogenization of SMART-FI Data sets

Data Characterization provides a concise and succinct summarization of the given collection of data. Generally SMART-FI considers cities data streams including Open Data Portals, Open Data Sets and Services, Internet of Things (IoT) and Sensor data sets and silos of legacy databases as data input streams. The main components of the Data Normalization layer include: i) Data Handling, ii) Data Processing and iii) Data Presentation components. Next, we discuss the first two component in more details.

Data Handling layer will receive heterogeneous data sets via the Input Handler and apply pre-processing to produce a list of data sets with access rules and a schema. The data will reside in semantic data stores. During the data processing period, Metadata processing, SPARQL based Query processing and Linked Open and Linked Sensor Data processing techniques will be applied using several different Urban Ontologies to create Urban Environment Ontology and the Data Presentation layer will provide Data Services to both Analytics components and Service Orchestration components. Data processing layer will homogenize the data using urban ontology definitions and will provide a uniform data format.

SMART-FI Data Normalization components will ease the discovery, the standards based normalized data sharing and reduce redundancy by also adding value to build ecosystems around cities data and contents. This component also includes Linked Data processing capabilities which is one of the best means for publishing and interlinking structured data for access by both humans and machines via the use of the RDF (Resource Description Framework) family of standards for data interchange and SPARQL for query. Here SPARQL- based solutions will be used to facilitate the discovery compared to conventional search mechanisms.

6.1 FIWARE Components for Urban Data Characterization

Smart applications and services for cities do need information about everything happening at every moment. FIWARE provides a mechanism to generate, collect, publish or query massive context information and use it for applications to react to their context. This is a complex process, as this information may come from different sources: systems, mobile apps' users, sensor networks, etc. It is FIWARE **Context Broker**, through a REST implementation of API OMA NGSI, which allows to shape and access it, whatever the source is.

The use and management from data coming from "Things" (i.e. sensors, actuators and other devices) is also a complex process, as there are many different protocols in the IoT sphere, but **FIWARE provides a set of GEs allowing to access the relevant information through only one API (NGSI)**. It not only allows to read this sensor information, but also to act on some elements. Therefore, **Context Broker is an essential part of the architecture to collect data, analyse them on real time, consult archives and their analysis, as well as to publish them as open data from a city**. On the other hand, other functionalities such as business intelligence, web interfaces and advanced interfaces allow the creation of very powerful applications and solutions.

FIWARE becomes then a fundamental pillar in the infrastructures of Smart Cities, as the different GEs build an architecture that can serve most of their needs. More than 15 European cities and some Spanish cities such as Málaga, Valencia, Seville, Santander or Lleida are already working and experimenting with this platform.

Among other initiatives to adopt FIWARE as Smart City platform, **31 cities from six European countries (Spain, Finland, Denmark, Italy, Portugal and Belgium) and from Brazil have recently announced at CeBIT 2015 (Hannover) the initiative** "Open & Agile Smart Cities", which is based on four main strategic elements: adopting common APIS accessing data from cities through NGSI, adopting a common platform for the publication of data such as CKAN, working on a driven-by-implementation basis and on building applications allowing to reuse, improve and define data models for different vertical markets in cities and to prove the interoperability in a city and among different cities. To that purpose, existent initiatives and models like **CitySDK** will be used. CitySDK is a "service development kit" for cities and developers that aims at harmonizing application programming interfaces (APIs) across cities. CitySDK APIs enable new services and applications to be rapidly developed, scaled and reused through providing a range of tools and information for both cities and developers. In the CitySDK project that started in January 2012, 8 cities across Europe have worked together to create some re-usable interfaces and processes. CitySDK toolkit provides all the information you will need to begin developing applications in your city.

6.1.1 Orion Context Broker

The following figure shows a logical architecture of the Publish/Subscribe Context Broker GE in FIWARE with its main components and interactions with other actors.

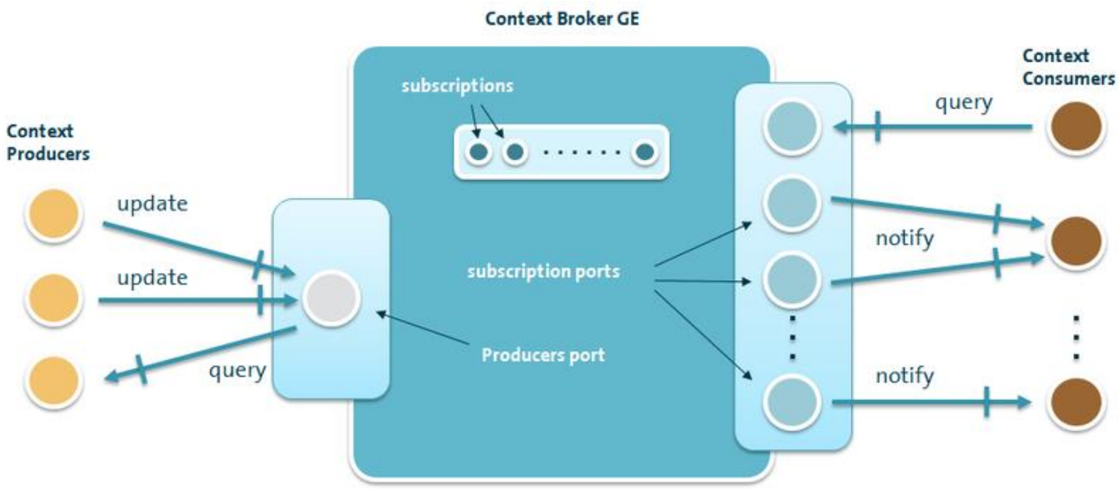

Figure 1 Logical architecture of the Publish/Subscribe Context Broker GE

The Orion Context Broker is an implementation of the Publish/Subscribe Context Broker GE, providing the NGSI9 and NGSI10 interfaces. As a part of the FIWARE platform Orion Context Broker provides facilities to manage all the whole lifecycle of context information including updates, queries, registrations and subscriptions and also register context elements and manage them through updates and queries.

Using NGSI9 and NGSI10 interfaces, clients can do several operations:

- Register context producer applications
- Update context information
- Being notified when changes on context information take place or with a given frequency
- Query context information. The Orion Context Broker stores context information updated from applications, so queries are resolved based on that information.

Apart from Orion Context Broker, there are other related components such as Cygnus or Steelskin PEP. Cygnus implements a connector for context data coming from Orion Context Broker and aimed to be stored in a specific persistent storage, such as HDFS, CKAN or MySQL. Steelskin PEP is a proxy meant to secure Orion Context Broker, by intercepting every request sent to the Orion, validating it against the Access Control component.

To develop Smart-FI Data/Context scenarios, a broker like the Orion Context Broker is a must. We would need a component in the architecture able to mediate between consumer producers (e.g. sensors) and the context consumer applications (Smart-FI smartphone applications: CityGo, MalatyaInsight; taking advantage of the context information provided by the sensors or other data sources).

6.1.2 FIWARE Data Models

FIWARE Data Models have been harmonized to enable data portability for different applications including, but not limited, to Smart Cities. They are intended to be used together with **FIWARE NGSI version 2.**

- Alarms: Events related to risk or warning conditions which require action taking.
- Parks & Gardens: Data models intended to make an efficient, effective and sustainable management of green areas.
- Environment: Enable to monitor air quality and other environmental conditions for a healthier living.
- Point of Interest: Specific point locations that someone may find useful or interesting. For instance, weather stations, touristic landmarks, etc.
- Civic Issue tracking: Data models for civic issue tracking interoperable with the de-facto standard Open311.
- Street Lighting: Modeling street lights and all their controlling equipment towards energy-efficient and effective urban illuminance.
- Device: IoT devices (sensors, actuators, wearables, etc.) with their characteristics and dynamic status.
- Transportation: Transportation data models for smart mobility and efficient management of municipal services.
- Indicators: Key performance indicators intended to measure the success of an organization or of a particular activity in which it engages.
- Waste Management: Enable efficient, recycling friendly, municipal or industrial waste management using containers, litters, etc.
- Parking: Real time and static parking data (on street and off street) interoperable with the EU standard DATEX II.
- Weather: Weather observed, weather forecasted or warnings about potential extreme weather conditions.

This repository contains:

- code that allows to expose different harmonized datasets useful for different applications. Such datasets are exposed through the FIWARE NGSI version 2 API (query).
- JSON Schemas and documentation on harmonized datamodels for smart cities, developed jointly with OASC, and other domains.

This work is aligned with the results of the GSMA IoT Big Data Project. Such project is working on the harmonization of APIs and data models for fueling IoT and Big Data Ecosystems. In fact the FIWARE datamodels are a superset of the GSMA Data Models.

All the code in this repository is licensed under the MIT License. However each original data source may have a different license. So before using harmonized data please check carefully each data license.

All the datamodels documented here are offered under a Creative Commons by Attribution 4.0 License.

6.1.2.1 JSON Schemas

FIWARE intend to provide a JSON Schema for every harmonized data model. In the future all the documentation could be generated from a JSON Schema, as it is part of our roadmap. The different JSON Schemas usually depend on common JSON Schema definitions found at the root directory of this repository.

There are different online JSON Schema Validators, for instance: http://jsonschemalint.com/. For the development of these schemas the AJV JSON Schema Validator is being used. For using it just install it through npm:

```
npm install ajv
npm install ajv-cli
```

A `validate.sh` script is provided for convenience.
Note: JSON Schemas only capture the NGSI simplified representation (`options=keyValues`)

6.1.2.2 How to contribute

Contributions should be sent in the form of pull requests.

6.1.3 Data models guidelines

FIWARE Data Models repository contains code that allows to expose different harmonized datasets useful for different applications. Such datasets are exposed through the FIWARE NGSI version 2 API (query) and JSON Schemas and documentation on harmonized data models for smart cities, developed jointly with OASC, and other domains. This is a set of guidelines for defining new data models.

6.1.3.1 Syntax

- Use English terms, preferably American English.
- Use camel case syntax for attribute names (`camelCase`).
- Entity type names must start with a Capital letter, for instance, `WasteContainer`.
- Use names and not verbs for attribute names, ex. `name`, qualifying it when necessary, ex. `totalSpotNumber` or `dateCreated`.
- Avoid plurals in attribute names, but state clearly when a list of items fits. Ex. `category`.

6.1.3.2 Reuse

- Check for the existence of the same attribute on any of the other models and reuse it, if pertinent.

- Have a look at schema.org trying to find a similar term with the same semantics.
- Try to find common used ontologies or existing standards well accepted by the Community, or by goverments, agencies, etc. For instance, Open311 for civic issue tracking or Datex II for transport systems.

6.1.3.3 Data types

- When possible reuse schema.org data types (`Text`, `Number`, `DateTime`, `StructuredValue`, etc.).

6.1.3.4 Attribute definition

- Enumerate the allowed values for each attribute. Generally speaking it is a good idea to leave it open for applications to extend the list, provided the new value is not semantically covered by any of the existing ones.
- State clearly what attributes are mandatory and what are optional. If needed state clearly what is the meaning of a `null` value.

6.1.3.5 Units

- Define a default unit for magnitudes. Normally it will be the unit as stated by the international system of units.
- If a quantity is expressed in a different unit than the default one, use the unitCode metadata attribute.

6.1.3.6 Relative values

- Use values between `0` and `1` for relative quantities, which represent attribute values such as `relativeHumidity`, `precipitationProbability`, etc.

6.1.3.7 Modelling location

- Use `address` attribute for civic locations as per schema.org
- Use `location` attribute for geographical coordinates. Ideally use GeoJSON for codifying geospatial properties. That works from Orion 1.2 on. If not use, old NGSI version 1 type `coords`.

6.1.3.8 Modelling linked data

- When an entity attribute is used as a link to other entities name it with the prefix `ref` plus the name of the target (linked) entity type. For instance `refStreetlightModel`, represents an attribute which contains a reference to an entity of type `StreetlightModel`.

6.1.3.9 Date Attributes

- Attribute type must be `DateTime`.
- Use the `date` prefix for naming entity attributes representing dates (or complete timestamps). Ex. `dateLastEmptying`.
- `dateCreated` must be used to denote the (digital) entity's creation date.
- `dateModified` must be used to denote the (digital) entity's last update date.
- `dateCreated` and `dateModified` are special entity attributes provided off-the-shelf by NGSIv2 implementations. Be careful because they can be different than the actual creation or update date of the real world entity represented by its corresponding digital entity.
- When necessary define additional attributes to capture precisely all the details about dates. For instance, to denote the date at which a weather forecast was delivered an attribute named `dateIssued` can be used. In that particular case just reusing `dateCreated` would be incorrect because the latter would be the creation date of the (digital) entity representing the weather forecast which typically might have a delay.

6.1.3.10 Dynamic attributes

- Use a metadata attribute named `timestamp` for capturing the last update timestamp of a dynamic attribute. Please note that this is the actual date at which the measured value was obtained (from a sensor, by visual observation, etc.), and that date might be different than the date (metadata attribute named `dateModified` as per NGSIv2) at which the attribute of the digital entity was updated, as typically there might be delay, specially on IoT networks which deliver data only at specific timeslots.

6.1.3.11 Some of the most used attributes

In case of doubt check other existing models!

- `name`
- `alternateName`
- `description`
- `serialNumber`
- `category`
- `features`
- `source`
- `relativeHumidity`
- `temperature`

6.1.4 Point of interest data model Samples

These data models allow to model points of interest and related entity types:

- **PointOfInterest**: A harmonised geographic description of a point of interest. According to Wikipedia a point of interest, or POI, is a specific point location that someone may find useful or interesting.
- **Beach**: A harmonised description of a beach. OpenStreetMap defines it as a loose geological landform along the coast or along another body of water consisting of sand, gravel, shingle, pebbles, cobblestones or sometimes shell fragments, etc.
- **Museum**: A harmonised description of a museum. OpenStreetMap defines it as an institution which has exhibitions on scientific, historical, cultural topics. Typically open to the public as a tourist attraction. May be more heavily involved in acquiring, conserving or researching such topics.
- **TouristInformationCenter**: A tourist information center which serves as an information source for tourists, travellers and visitors. It can be represented by an entity of type `PointOfInterest` which category is equal to `439`. Another option is to use the schema.org `TouristInformationCenter` entity type and include those properties which domain is `PointOfInterest` and/or properties which domain is `http://schema.org/TouristInformationCenter`.

6.1.4.1 Point of interest

6.1.4.1.1 Description

This entity contains a harmonised geographic description of a Point of Interest. This entity is used in applications that use spatial data and is applicable to Automotive, Environment, Industry and Smart City vertical segments and related IoT applications. This data model has been created in cooperation with the GSMA and the members of the IoT Big Data Project.

6.1.4.1.2 Data Model

A JSON Schema corresponding to this data model is as:


```
{
  "$schema": "http://json-schema.org/schema#",
  "id":
"https://fiware.github.io/dataModels/PointOfInterest/PointOfInterest/schema.json",
  "title": "GSMA / FIWARE - Point of Interest schema",
  "description": "A point of interest",
  "type": "object",
  "allOf": [
    { "$ref": "https://fiware.github.io/dataModels/common-schema.json#/definitions/GSMA-
Commons" },
```

```
    { "$ref": "https://fiware.github.io/dataModels/common-
schema.json#/definitions/Location-Commons" },
    {
      "properties": {
        "type": {
          "type": "string",
          "enum": [
            "PointOfInterest"
          ],
          "description": "NGSI Entity type"
        },
        "category": {
          "type": "array",
          "items": {
            "type": "string"
          },
          "minItems": 1
        },
        "refSeeAlso": {
          "type": "array",
          "items": {
            "anyOf": [{
              "$ref": "https://fiware.github.io/dataModels/common-
schema.json#/definitions/EntityIdentifierType"
            }]
          }
        }
      }
    }
  ],
  "required": [
    "id",
    "type",
    "category",
    "location",
    "name"
  ]
}
```


- `id` : Unique identifier.
- `type` : Entity type. It must be equal to `PointOfInterest`.
- `dateModified` : Last update timestamp of this entity.
  - Attribute type: DateTime
  - Optional
- `dateCreated` : Entity's creation timestamp.
  - Attribute type: DateTime
  - Optional
- `source` : A sequence of characters giving the source of the entity data.
  - Attribute type: Text or URL
  - Optional
- `name` : Name of this point of interest.
  - Normative References: https://schema.org/name
  - Mandatory
- `alternateName` : Alternative name for this point of interest.
  - Normative References: https://schema.org/alternateName

- Optional
- `description` : Description of this point of interest.
  - Normative References: https://schema.org/description
  - Optional
- `location` : Location of the point of interest represented by a GeoJSON geometry, usually a `Point`.
  - Attribute type: `geo:json`.
  - Normative References: https://tools.ietf.org/html/rfc7946
  - Mandatory if `address` is not defined.
- `address` : Civic address of this point of interest.
  - Normative References: https://schema.org/address
  - Mandatory if `location` is not present.
- `category` : Category of this point of interest.
  - Attribute type: List of text Text
  - Allowed values: Those defined by the Factual taxonomy. For instance the value `113` corresponds to beaches, and the value `311` corresponds to museums.
  - Mandatory
- `refSeeAlso` : Reference to one or more related entities that may provide extra, specific information about this point of interest.
  - Attribute type: List of References
  - Optional

Examples of use:

```
{
    "id": "PointOfInterest-A-Concha-123456",
    "type": "PointOfInterest",
    "name": "Playa de a Concha",
    "description": "La Playa de A Concha se presenta como una continuación de la Playa
de Compostela, una de las más frecuentadas de Vilagarcía.",
    "address": {
      "addressCountry": "ES",
      "addressLocality": "Vilagarcía de Arousa"
    },
    "category": ["113"],
    "location": {
      "type": "Point",
      "coordinates": [
        -8.768460000000001,
        42.60214472222222
      ]
    },
    "source": "http://www.tourspain.es",
    "refSeeAlso": ["Beach-A-Concha-123456"]
}
```

Get POIs around Fira Montjuic (2 Kms around)

```
curl -S --header 'fiware-service:poi' --header 'fiware-servicepath:/Spain' --header 'x-auth-token:dd3a796995294dc5b6d7bf2cdb36e15f'
"http://130.206.118.244:1027/v2/entities?type=PointOfInterest&attrs=location,name&orderBy=geo:distance&georel=near;maxDistance:2000&geometry=point&coords=41.3730751,2.1482742&options=keyValues"
```

6.2 SMART-FI Data Models

SMART-FI consortium planned to contribute and extend the FIWARE data models in order to use as initial data models for pilots in cases they are compatible and also future development. FIWARE Data Models are available on github[13]and are open for contribution based on data model development guidelines[14] provided.

[13] https://github.com/fiware/dataModels

[14] http://fiware-datamodels.readthedocs.io/en/latest/guidelines/index.html

7 Conclusion

Assessing the nature, technical feature, legal ownership and other related aspects of the datasets that each partner is making available in the framework of the project activities represents a fundamental activity in order to progress towards the SMART-FI objectives.

This review performed by the SMART-FI consortium served as:

- an internal reference in order to assess whether and how technical interoperability between datasets and DBs can be achieved;
- the starting point for the creation of pilots grounded on such datasets;

In this document we have described the FIWARE Data Models repository to be considered as SMART-FI harmonized data portability for different applications including Smart Cities.

As the main next step in this thread, further effort is to be produced by the partners in order to identify the main drivers of technical interoperability and to harmonize, wherever possible, vocabularies, metadata and DBs structure.

The SMART-FI consortium may plans to release updated versions of this document, mainly to take into account inclusions and if new datasets made available by the partners, also in light of the scopes and the user requirements that will emerge within the activities of WP6 (pilots implementation).

8 References

- SMART-FI Proposal 2016.
- SMART-FI Project Consortium Agreement
- Project website: http://www.smart-fi.eu/
- SMART-FI: Exploiting Open IoT Data from Smart Cities in the Future Internet Society. Stefan Nastic, Javier Cubo, Malena Donato, Schahram Dustdar, Orjan Guthu. Mats Jonsson, Omer Ozdemir, Ernesto Pimentel, M. Serdar Yumlu, 2017. Springer Series, Internet of Things Technology, Communication and Computing.
- Open-DAI; Opening Data Architectures and Infrastructures" for European Public Administrations website: http://www.open-dai.eu/
- FIWARE Data Models: https://www.fiware.org/data-models/, https://github.com/Fiware/dataModels
- FIWARE Catalogue: https://catalogue.fiware.org
- EU Internet Handbook – Keep it simple http://ec.europa.eu/ipg/content/tips/volume_en.htm