CERISE

Combining Energy and Spatial Information Standards as Enabler for Smart Grids

TKI Smart Grid Project: TKISG01010

D4.1 Semantic Mappings to Harmonize Energy, Geo and Government-related Information Models
Work package – 40

Lead partner: TUDelft

23 September 2015

Version 1.0 - Final

DOCUMENT INFORMATION	
ID	D4.1 Semantic Mappings to Harmonize Energy-related Information Models
Work package	WP40
Type	Report
Dissemination	Public
Version	1.0 - Final
Date	23 September 2015
Authors	Maarten Steen (TNO), Frans Knibbe (Geodan), Wilko Quak (TU Delft), Paul Janssen (Geonovum), Roel Stap (Alliander), Laura Daniele (TNO)
Reviewer(s)	Jasper Roes (TNO)

Table of contents

Summary

The CERISE-SG project (Combining Energy and Geo information standards as enabler for Smart Grids) focuses on interoperability with a special interest in the information exchanges between smart grids and their surroundings. We hereby focus on the exchange of information to and from smart grids, the government domain and the geo domain. Within the fast changing smart grid world acquiring reliable information from different sources is invaluable. The information required comes from different sources that all use their own (often different) definitions for the data they control. The national registration of buildings for instance contains different data with different definitions from the data source of energy consumption. To exchange information between these sources connections need to be made between the different areas that make sure that correct and reliable data is available. In this document we describe a process to define these mappings.

First an identification of the playing field of smart grids is given and the three major domains that are involved are described i.e.
- (1) the utilities domain,
- (2) the geographic domain; and
- (3) the government domain.

Because there are already many standardisation processing running within each domain, with different set-ups the cross domain harmonization is extra challenging. In this document we work out two solutions to the harmonization problems:
- Define relationships between elements from different models
- Express model elements from different models in a common model

In this document we propose 'linked data' and semantic web technology as a solution for this challenge, we explain Linked Data and explain how Linked Data can be used to semantically map data models to each other. By actively harmonizing a collection of datasets that are relevant for the smart grid we prove the viability for using 'linked data' as a solution for interoperability challenges of the smart grid. We did this with the following datasets:
- Liander Open Data set
- Zonnedeal smart meter readings
- Mpare smart meter readings

1   Introduction

The overall goal of CERISE-SG is to support future-proof information exchange between the energy, e-government and geography domains in order to enable the realization and management of smart energy grids. Each domain has its own set of standards and information models for exchanging information within that domain, and even within domains there still are interoperability problems. This document describes possible solutions for those interoperability problems, by using common standards and harmonized information models.

This document contains a description of the harmonization problem (chapter 2) and an overview of different approaches to solving that problem (chapter 3). Within the project, we opted for one particular approach: Linked Data. It is explained in chapter 4. We applied this approach on several datasets using various ontologies. The relevant datasets are described in chapter 5, the ontologies used to represent the data in chapter 6. The technical details of the mapping are not part of this document but are given in the Cerise Cookbook for Standardization and Harmonization (Cerise D5.1, 2015) of this project. The result of the mapping is presented in chapter 7 and also its formalization is given. Finally, conclusions and recommendations are given in chapter 8.

2   The harmonization challenge

Smart energy grids are a relatively recent phenomenon. Enabling them requires data, and those data cannot be found solely in one domain. Instead, the required data should come from different domains, i.e. also from outside the grid. The operation of smart grids depends on a combination utilities data, governmental data and spatial data.

Although existing structures often are based on standards, they are not always set up with external interoperability in mind as a primary design objective. And existing structures often stem from before the web era, the first time in history when things could be interconnected on a global scale. In the following paragraphs we describe how the three major data domains for CERISE-SG are set up.

2.1   Governmental data

Many data that are important to make Smart Grids work are government data, although the situation will be different in different countries. In the Netherlands, the government recognizes the need for making governmental data interoperable, as a means of providing better service to citizens and as a means of improving efficiency within the government itself. Also the Dutch government is aware of growing needs for automation and digitization of information flows, eGovernment. There are several initiatives for standardizing data exchange and information gathering within the Dutch government. An overarching framework is the Dutch Government Reference Architecture NORA[1] (*Nederlandse Overheid Referentie Architectuur*). It mostly describes quality criteria. For specific levels of government (e.g. provincial, municipal) more specific frameworks have been derived from NORA (e.g. EAR, GEMMA, PETRA). Most, probably all data exchange specifications are based on XML.

A national model for facilitating data exchange not only serves as a basis for regional models, it also has to comply with international models. These primarily come from the European Union, with its strong drive towards cooperation between its member states. Notable European frameworks for data exchange are the European Information Framework (EIF) and the Infrastructure for Spatial Information in the European Community (INSPIRE). In INSPIRE one of the data themes is 'Energy resources'. This theme mainly deals with data on primary energy sources like hydrocarbons, wind and solar irradiation.

A national framework that is of particular importance to Smart Grids is the System of Base Registries (*Stelsel van Basisregistraties*)[2]. An outline of the framework is given in figure 1. Base Registries are important mainly because they contain data for many relevant topics such as persons, buildings, vehicles, addresses and topography. Those kinds of data are essential for many applications of governmental data. Work to harmonize the base registries is ongoing. A result of that work is a common catalogue of definitions, the *Stelselcatalogus* (system catalogue).

---

[1] http://www.noraonline.nl/wiki/NORA_online
[2] http://www.digitaleoverheid.nl/onderwerpen/stelselinformatiepunt
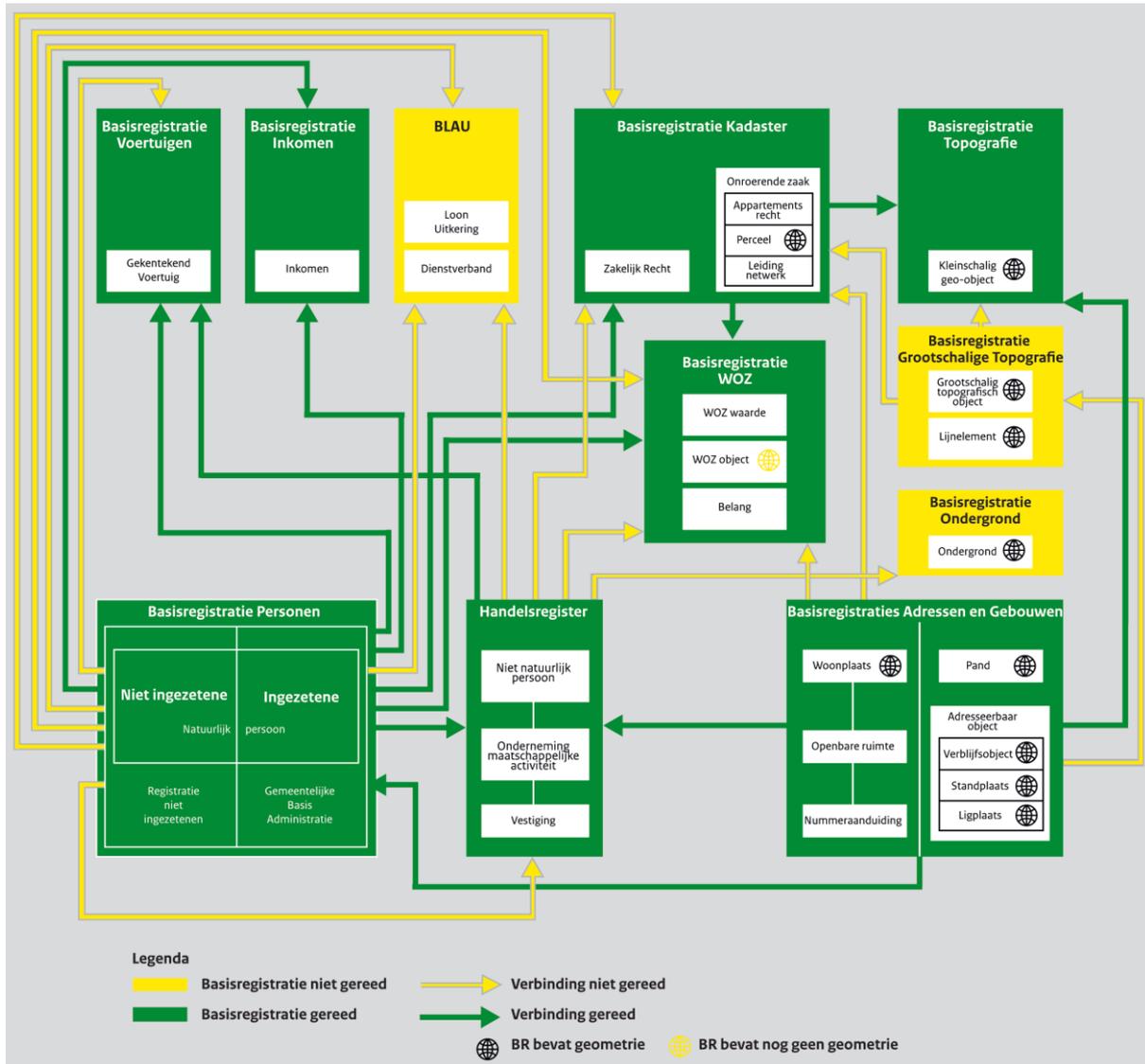
Figure 1 Diagram showing the different base registries and their relationships (in Dutch). Arrows indicate links between registries. The yellow part of the diagram is under construction, the green part has already been established.

2.2 Utilities data

The intrinsic drive for ITC standardization of data exchange in the utilities sector traditionally has been less than in the public sector. Nonetheless, recent global and national developments have caused increased activity in this area. This has resulted in the development of the Common Information Model (CIM)[3], a global standard adopted by the International Electrotechnical Commission (IEC). CIM information model is developed as a UML model for among other transmission and distribution of electric power[4]. It is foreseen that electric power companies will make increased use of CIM for exchanging information between applications with other parties, or within their own organisations.

In light of harmonization is should be mentioned that the CIM model is self-contained: It does not reuse elements from other models but has its own definitions of model elements (e.g. classes,

---

[3] http://www.dmtf.org/standards/cim
[4] IEC is currently working on an extension of CIM for natural gas and water

properties, relationships). The CIM model is an extensive model that has a lot of detail. For specific applications it is possible to derive and use simpler subsets of the model. These subsets are called CIM profiles.

2.3 Geographic data

The domain of geography is somewhat different than the government and utilities domains. Rather than being concerned with an aspect of society, it is concerned with a special kind of data: geographical data. Like the two domains described above, it also has its heritage of standardisation, and it too has its own way of achieving standardisation.

The important standards body in the domain of geographic data is the Open Geospatial Consortium. It has developed a general model for geographic features, on which various standards are based. Information models are described in UML and encoded in XML. There are standards for various web services for the exchange of geographic data or geographic information. The OGC has a Domain Working Group for the energy and utilities domain (see http://www.opengeospatial.org/projects/groups/energyutilities), but work in that group has not lead to standards or recommendations yet.

Like CIM for the utilities domain, OGC standards are also largely self-contained. Other than basic XML data types (e.g. text string, date, number) no external data definitions are used.

2.4 Conclusions

The descriptions above show that within the three domains that CERISE-SG is concerned with there is a clear drive towards standardization, but the resulting standards are mainly useable within their respective domains, not so much outside of it. With existing domains being set up in different ways, efficiently combining data, for example to enable Smart Grids, requires a harmonization effort.

A benefit of current domain standards being based on common practices for information[5] modelling and information exchange like UML and XML, is that syntactic interoperability is not much of a problem. Semantic interoperability becomes a problem, as soon as information concepts surpass basic XSD datatypes. Something has to be done to make information from domains with different designs interoperable. Possible harmonization strategies will be discussed in the next section.

---

[5] Information is that which informs, i.e. an answer to a question, as well as that from which knowledge and data can be derived (as data represents values attributed to parameters, and knowledge signifies understanding of real things or abstract concepts) (https://en.wikipedia.org/wiki/Information)

3 Possibilities for semantic interoperability between domains

As explained in chapter 2, interoperability is needed between different domains that each have their own way of describing the world, or that part of the world that is of interest to the domain. This is a general problem, for which an optimal general solution should be found. The problem is mainly one of semantic interoperability. Should it become possible for a party with an interest in obtaining data from multiple domains to express a data query using a single semantic model, then actually performing the query and getting a meaningful set of data in response should be straightforward.

Semantic harmonization involves two basic types of problem. The first is the case of the same concepts being defined in different ways in domain models. This happens for common concepts, like 'person', 'address' or 'location'. For example, the way a utilities information model defines a person should be interoperable with the way a governmental model defines a person. The second type of problem is definitions of specialized concepts that only exist in one of the domain models. It should be noted that the second case occurs less often than one might expect, because in most domain models class hierarchies are used, in which specialized concept definitions are derived from more abstract definitions. The more abstract a concept, the higher the likelihood of it having some semantic overlap with a concept from another model.

In the following sections two different methods for achieving semantic interoperability are described.

3.1 Define relationships between elements from different models

One way of achieving semantic interoperability is to define mappings between entities in the domain models. This should only have to happen for those concepts that are shared between models. Concepts that are uniquely defined within a single domain model do not have to be mapped to another model, their original definitions can be used.
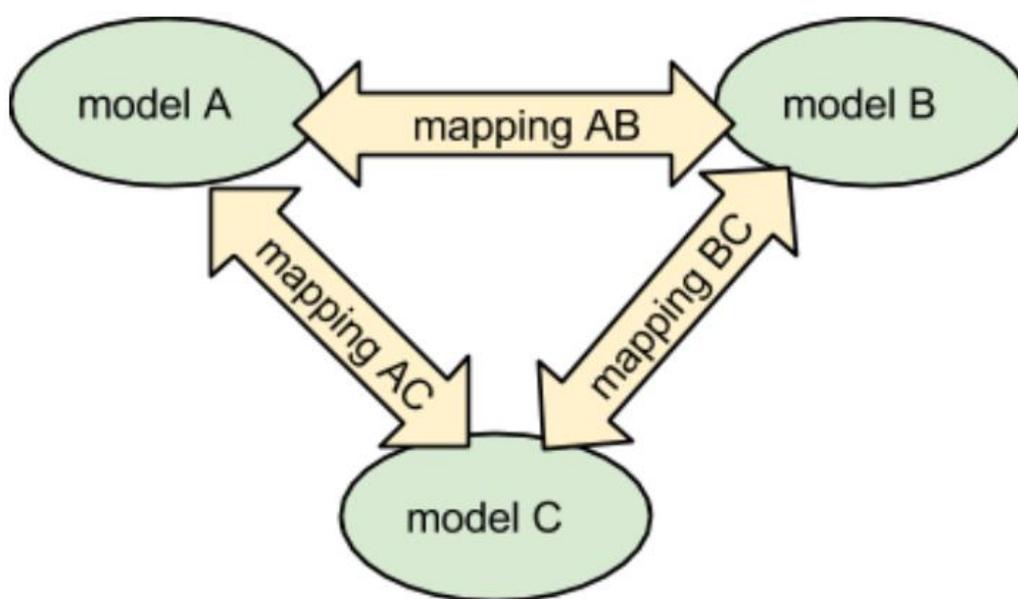
Figure 2 Establishing semantic interoperarability by defining mappings between the various domains.

This method of semantic mapping is especially suited for use from within one of the domains. For example, someone working in the utilities domain could make use of the utilities-government mapping to obtain data from a data source that uses a governmental information model. For use outside of the domains this method seems less suited. Within the context of CERISE-SG there are examples of this kind of use: a neighbourhood energy collective does not have its own domain model, but will need to obtain data from different domains with existing models.

An important disadvantage of this approach is that when the number of domain models to map increases, the number of required mappings increases drastically. For instance, interoperability between three domain models requires three mappings, but with five domain models twenty domain-domain mappings can be made. Complexity increases even more when domain models change over time, which means that multiple mappings will have to be updated.

3.2   Express model elements from different models in a common model

A different approach is to map concepts from a domain model to concepts from a shared information model. The general model can then be used to express all domain data.
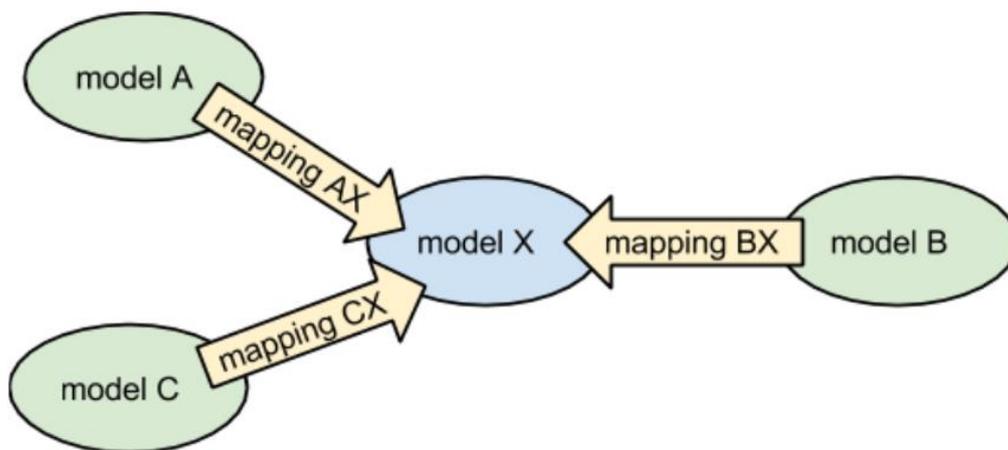
Figure 3 Establishing interoperability by defining mappings to a common shared model.

Care should be taken to make this approach extensible. When new system requirements call for interoperability with yet another domain model, it has to be possible to add another mapping without having to change existing mappings. This means that the shared model should be sufficiently general.

A useful property of this method is that a data consumer only needs to know the general model in order to make sense of data from the domain models. Such use would require *all* concepts in the domain models to be mapped to the general model. For specialized concepts this means that mapping should take place at a sufficiently high abstraction level (e.g. parent class), which in turn means that data consumers could incur a loss of semantic accuracy.

3.3   Conclusion

To make domain data interoperable some sort of semantic mapping needs to be done. Such a mapping can be expressed in a modelling language like OWL or a rule language like SPIN. From these formal mappings automatic transformation procedures can be derived.

4 Practical interoperability

4.1 Linked Data

Given the project mission - achieve interoperability between different domains for users within and outside those domains - we have the Linked Data paradigm[6] as offering the required capabilities for investigating the problem.

In short, Linked Data is a way of sharing raw data on the world wide web. Linked Data is strongly related to the Semantic Web, the idea of annotating information on web pages with semantic tags so that those web pages can easily be interpreted by automatic procedures (machines versus humans). Those automatic procedures could improve the information available to humans, for example by creating and maintaining indexes that make data discoverable, or by enriching data with derived data.

Key principles of Linked Data are:
1. All data (including metadata and semantics) are web resources, identified by HTTP(S) URIs (Uniform Resource Identifiers).
2. Looking up a URI returns data describing the resource.
3. The Semantic Web stack family of standards (RDF, RDFS, SPARQL, OWL, SKOS, SPIN, ...) is used to model and query data.
4. Content negotiation is used to request data in a specified format (e.g. HTML is nice for humans, XML is nice for processing, JSON-LD is nice for web developers).
5. Data providers are encouraged to link their data to other data sources on the web. This way, all data on the web become interconnected and form one global database (or one global graph, since RDF models data as graphs).

Fully explaining the concept of Linked Data falls outside the scope of this document, but there is much information available on the web. A starting point could be http://www.w3.org/standards/semanticweb/data. Nevertheless, some aspects of Linked Data that make it especially suitable for CERISE-SG can be listed:

- It uses modular semantics - small data sets with data definitions (called vocabularies or ontologies) are published on the web and can be used, mixed and matched by data providers.
- It is adopted by many different domains, especially those that want to achieve better inter domain and cross-domain interoperability. Domains working with Linked Data include the three domains that CERISE-SG is concerned with: geography, energy and government.
- It builds on existing web and existing web architecture: much of the system and infrastructure for data exchange is already in place.
- It allows advanced data analysis, e.g. reasoning/inference (see http://www.w3.org/standards/semanticweb/inference)

4.2 Existing Linked Data semantics in the geography, government and utilities domains

Linked Data principles have found their way into the three domains that CERISE-SG is concerned with, to different extents. This section describes the existing semantics in the three domains.

4.2.1 Geography

Semantics for geography in the Semantic Web come from both the web and the geography communities. Recognition of geographic geometry as a basic type of data took place early in the development of the Semantic Web. A notable result was the publication of the Basic Geo vocabulary[7] in 2003. It provides semantics for expressing point geometry as latitude/longitude

---

[6] http://linkeddata.org/

[7] http://www.w3.org/2003/01/geo/

coordinates. A more recent specification is schema.org, a vocabulary developed by major web search engines. Among other things it contains classes for expressing geography, e.g. http://schema.org/GeoShape. Many more vocabularies that can be used to express geographic data have come into existence, both from communities with a web background and with a geographic background. A vocabulary that is recognized as authoritative by people and organizations with a geographic background is GeoSPARQL[8] . It is a standard from the Open Geospatial Consortium (OGC), the main standards body for the geography domain. The GeoSPARQL specification is based on the foundations of the OGC (or ISO/TC 211) theoretical framework for geography that is documented as UML class diagrams. Next to definitions of geometry in RDF, GeoSPARQL defines topological functions for SPARQL (an RDF query language).

Lastly, a specification that is worth mentioning is the ISA Programme Location Core Vocabulary[9] , which is a product of INSPIRE-related research. The vocabulary defines concepts for locations and addresses in a general way, making it easy to apply these semantics as umbrella terms.

4.2.2 Government

The Dutch government, like most national governments, is a large and heavily segmented organization. For that reason it fully understands the need for frictionless data exchange within and between its many subdivisions, and that is why it is looking at what Linked Data can offer. Also there is the understanding that opening up governmental data to the general public has important societal and economic benefits, something that other countries have also realized and have adjusted their policies to that effect (e.g. the UK and the USA). When looking at the best way to provide open data, Linked Data is a consideration (see the five star open data concept: http://5stardata.info/).

Recently a member of the RDF family of standards, SKOS (Simple Knowledge Organisation System) has been put on the comply-or-explain list of the Dutch Standardisation Forum. But with a national government being big and complex, it is understandable that most changes in data exchange techniques and procedures can't be made overnight. At the moment, experiments and pilots are undertaken to get an idea of costs and benefits.

Of immediate interest to CERISE-SG is governmental participation the Platform Linked Open Data Nederland (PLDN), a continuation of the Pilot Linked Open Data Nederland (PiLOD). The system of base registrations plays an important role there, as well as the Dutch Cadastre, an important source of national geographic data. Among the results of the platform is a national strategy for minting URIs, and experimental publication of two important datasets as Linked Data, the BAG (buildings and addresses) and the BGT (large scale base topography).

4.2.3 Utilities

Of the three domains under consideration, until now the utilities domain shows the least inclination to move towards web based data exchange. It is likely that this is because the utilities world I(before the emergence of smart grids was more self-contained than the other domains and had less need for sharing data with the outside world. Nevertheless, the global standard for energy data exchange, CIM, is published in RDF, although not with persistent URIs.

4.3 Possibilities for Semantic mapping in Linked Data

The Linked Data paradigm, with its emphasis on linking data (resources) to common semantics is well suited for solving the problem of semantic interoperability.

A general truth is that the more a data set is linked, the more usable it is. That goes especially for links to vocabularies, which provide meaning to data, and provide automated procedures with the means to combine data from different sources. The key to improving semantic interoperability is to provide linkage to common semantics. That way, a data consumer does not need to know about

---

[8] http://www.opengeospatial.org/standards/geosparql
[9] http://www.w3.org/ns/locn

domain specific semantics, but can use general expressions to discover and filter data. For example, a data set containing some address data (e.g. street name and number, postal code and city) could be published on the web using semantic annotation that is specific to the data set. The semantics could be derived from the names of the columns that are used in the relational database where the data are internally stored. According to the Five Star Data scheme, this would count as four star data. To make the data more useful for consumers, the address data could be linked to additional semantics from a general domain model. Both the utilities domain and the national government domain have their own semantics for address data. That would make that part of the data set interoperable with other data sets from the same domain. A further improvement can be made if the address data are also linked to global semantics (for example the Location Core Vocabulary). When that happens, the data are usable by user agents from any domain.

It should be stressed that common semantics do not need to replace local semantics. In an RDF dataset it is possible and perfectly acceptable to model data using different models (vocabularies). It will be up to the requirements of the data consumers which semantics that are provided will be used. Semantics with a narrow scope could carry over details that have been abstracted away in more general models, while general semantics provide the means of data harmonization and interoperability.

Two different strategies for providing access to common semantics (and through that, achieving semantic interoperability) can be distinguished. These two strategies could be viewed as exclusive ways of achieving interoperability, but they can also be applied both, mutually supporting each other.

4.3.1 Use an external mapping and reasoner

In this strategy, links from locally defined concepts to more general concepts are not included in the published data set, but are defined externally and optionally augmented with a smart reasoner. A mapping between the local semantics and the more general semantics can be published as a separate dataset, where local semantic resources (identified by URIs) are related to general semantic resources (also identified by URIs). This mapping could be straightforward (stating that two classes are equivalent), or more complex, using rules. Formalizations like OWL and SPIN are well suited for expressing the latter kind of mapping.

One can imagine a specialized service on the web that contains these mappings, and also provides the means to use the mappings to infer implicit relationships. Because of the way RDF is set up, automated procedures can use deductive reasoning to find relationships between resources that have not been explicitly published. Servers with high reasoning capabilities are not required by any standard and are not common when RDF data are published, so this would be an added value.

This method could relieve data publishers of having to add extra semantics to the data they publish, but more is required from data consumers. For one thing, they need to know that there is an external mapping and reasoner available somewhere. So at least there should be a standardised way of linking from the data set to the reasoner. Secondly, the data consumer will have to interact with two web servers to work with the data, instead of one. Thirdly, a single reasoner could be considered a potential single point of failure in an otherwise distributed architecture.

Drawbacks of this strategy could largely be negated if the output of the central service is fed back to the source dataset. The service housing the mappings and reasoner could be made to produce RDF data that could be added to the source data as an enrichment.

4.3.2 Provide general semantics at the source

Instead of having an external service provide the data needed for semantic interoperability, those data can be added to a data set by the data provider. For instance, a data set that is based on CIM could have addresses that are stated to be instance of a CIM address class. A data provider could add extra data to the dataset for the addresses, stating that the addresses are also instances of the

address class that is defined in the Location Core Vocabulary. That way the address data would be discoverable and queryable by consumers that only know about general web semantics.

This strategy is more demanding on the data publisher, he or she has to understand common web semantics in order to link to them in the right way. Moreover, sometimes complex rule-based relationships need to be defined if there is no one on one match between local concepts and general concepts.

An important advantage of this method is that interoperability does not rely on the functioning and availability of a single network node (the server that has the mapping and the reasoning capabilities).

4.4 Conclusion

In this chapter it was argued that Linked Data is a very suitable paradigm for achieving the kind of data harmonization that is sought after in CERISE-SG

Two different strategies for adding semantics to data were described. Both these strategies need the same groundwork to be done: mappings between information models need to be made, in order to make the data available with common semantics on the worldwide web of data.

5 Datasets

In this chapter we identify a number of energy-related datasets that are of relevant for the smart grid and were used during the project to create and experiment with mappings.

In the project several data sets have been published and have been used. Each dataset has associated metadata that describe the dataset and provide handles to interact with the data set. For example, data set metadata can describe the topics of the dataset and point to a SPARQL endpoint that can be used to query the data.

CERISE-SG has made a data catalogue available that serves as a single entry point for accessing data set metadata and, through those, the data themselves. Like the data and metadata, the catalogue is published as Linked Data. The URI of the catalogue is http://lod.geodan.nl/cerisesg/datasets/.

The CERISE-SG data catalogue is based on the W3C Data Catalog Vocabulary (DCAT). This vocabulary can be used to describe any data set, not just Linked Data data sets. Data sets that are not directly accessible on the web can still be described by metadata that are directly accessible on the web, and can be crawled and indexed. Establishing data catalogues can help in grouping together data sets for specific purposes and can help making data sets discoverable on the web. Data catalogues that are set up according to DCAT semantics can themselves be referenced by other catalogues, making it possible to have a federated system of data catalogues.

The data catalogue provides a single access point (URI) to all relevant data sets and their metadata. This means that the web application does not need have a priori knowledge of individual data sets. Reading data sets from the catalogue should provide sufficient descriptions of available data sets and provide the means to further exploit those data sets.

In the following sections we give an overview of some of the datasets relevant for the rest of this document.

5.1 Liander Open Data

Liander manages the energy distribution network in a large part of The Netherlands. They transport gas and electricity from energy producers to households and other users. In order to support their operations, Liander has many datasets, but is not allowed to use this data in applications due to legal limitations. Nevertheless, this data could be used by third parties in applications, for instance to facilitate the transition towards a more sustainable energy future. Therefore, Liander is providing some of their data as open data (see https://www.liander.nl/over-liander/innovatie/open-data). Liander collects data on energy consumption and (local) production, e.g. through the use of energy meters. Liander would like to publish this data in order to:

- Be transparent as a public utility company
- Stimulate open innovation
- Gain insight into data needs
- Improve data quality by receiving feedback

Although this data can be interesting for data consumers, e.g., to visualize energy consumption for individual households at different periods during a day, there are several issues with this data and publication will have to be restricted in a number of ways. Firstly, this data is subject to data protection laws. It is personal data and publication will violate the privacy of the households concerned. Therefore, it cannot be published as is. Secondly, the data may be of commercial interests for Liander. Publishing the data as open data now would make it harder to create their own revenue streams later on. Thirdly, the quality of the data varies a lot. Households with smart meters

may provide measurements with a 15 minute interval, but not all households have smart meter. In the worst case, for households without smart meters, meter readings are only validated once every three years. And even for households with smart meters readings are sometimes received only once a quarter.

In order to deal with these issues, the data is restricted in the following ways:

- The data quality is standardized. Rather than publishing actual meter readings at regular intervals, Liander only publishes the estimated, standardized annual usage. This value is recalibrated once a quarter using recent readings, but will be published only once a year.
- Commercially sensitive data is removed from the dataset, i.e., only energy usage of private households, the so called small users, is published.
- The data is anonymized. Rather than publishing the annual usage for each individual household, the annual usage is aggregated for all households in the geographical area determined by the 6-digit postcode. If there are less than ten households in one postcode area, the annual usage of two or more consecutive postcode areas are aggregated.

The result is a set of (spreadsheet) tables, one for each year, that contain the standardized average energy usage for the households in a certain postcode area. This data is published on the Liander website at www.liander.nl/opendata. Below a small excerpt of such a table.

STRAATNAAM	POSTCODE VAN	POSTCODE TOT	WOONPLAATS	LAND	PRODUCT	Aantal	%Richting	%TYPE	TYPE	SJV	%Laag	%Slimme Meter
Rijksweg A44	1000AA	1011AB	NIEUW VENNEP	NL	ELK	31	100	29	3x25	16245	38,71	16,13
De Ruyterkade	1011AC	1011AC	AMSTERDAM	NL	ELK	32	100	31	3x25	11433	28,13	15,63
't Spiker	7231JS	7231JT	WARNSVELD	NL	ELK	24	75	54	3x25	3764	41,67	0
't Spiker	7231JS	7231JT	WARNSVELD	NL	GAS	20	100	100	G4	2615	0	0
't Spiker	7231JV	7231JV	WARNSVELD	NL	ELK	16	100	88	1x25	2425	0	0
't Spiker	7231JV	7231JV	WARNSVELD	NL	GAS	16	100	100	G4	1626	0	0

The tables, i.e., the column names, are documented in a document that is published alongside the dataset itself. Unfortunately, this is only available in Dutch. Below we copied it verbatim.

Variabele/veld	Mogelijke waarden	Toelichting
[A.04] Meetverantwoordelijke	Naam	naam van het bedrijf dat de meterstanden van de elektriciteitsmeter of gasmeter opneemt, verifieert, registreert en doorgeeft aan relevante marktpartijen
[A.02] Netbeheerder	Naam	Naam van de regionale netbeheerder
[A.06] Netgebied	Naam	Aanduiding van het gesloten inkoopgebied waar een aansluiting onder valt.
[A.07] Postcode	4 cijfers 2 letters	Twee kolommen: van en naar
[A.10] Straatnaam	Naam	Bij verschillende postcodes bij "van en" naar": de straatnaam van de eerste postcode
[A.11] Woonplaats	Naam	Naam van de woonplaats
[A.60] Landcode	Nl	Afkorting van de landsnaam conform GBA-landentabel (Tabel 34)
[A.17] Productsoort	ELK of GAS	De energiesoort waarover het SJV gegeven wordt: ELK= elektriciteit, GAS = aardgas

[A.18] Verbruikssegment	KVB	Kleinverbruik, dat wil zeggen dat de aansluitwaarde niet groter is dan 3x 80 Ampère (elektra) of G25 (gas).
[A.19] Leveringsrichting	%	Percentage van de aansluitingen dat netto elektriciteits- of gasverbruik heeft. Dit percentage wordt lager naarmate er meer teruglevering plaatsvindt (bv vanwege zonnepanelen)
Aantal aansluitingen	getal	Het aantal aansluitingen in het betreffende postcodegebied voor het betreffende product.
[A.21] Fysieke status	%	Het percentage van de aansluitingen dat in bedrijf is. De overige aansluitingen zijn in aanleg, uit bedrijf of gesloopt.
[A.25] Type aansluiting	%	Het percentage definitieve aansluitingen. De overige aansluitingen zijn tijdelijke of bouwaansluitingen.
[A.29] Soort aansluiting	%	Bij gas en stroom wordt de zwaarte van de aansluiting aangeduid met aansluitcapaciteit, ook wel aansluitwaarde genoemd. De aansluitwaarde geeft aan hoeveel Ampère elektriciteit of m³ /uur gas er door de aansluiting kan. Hier wordt weergegeven het percentage van de meest voorkomende aansluitwaarde.
A.29-b Soort aansluiting Naam	codering	Naam van de meest voorkomende, waarop het percentage in de voorgaande kolom betrekking heeft. Opties: "#zekeringen x # ampère" (elektra), G4 tot G25 (gas).
[A.33] SJV	kWh of m3	Gemiddeld Standaardjaarverbruik, waarbij voor aansluitingen met een normaal én laagtarief SJV het totale SJV wordt meegenomen in de middeling, zonder decimalen. Het standaardjaarverbruik is het verwachte jaarverbruik van een afnemer op een netaansluiting bij gestandaardiseerde condities en op basis van een genormaliseerd jaar.
		Wanneer een aansluiting uit bedrijf is blijft het laatste SJV staan tot het moment waarop de aansluiting weer in bedrijf genomen wordt.
[A.34] SJV laag tarief	%	Percentage van de aansluitingen dat een laagtarief SJV heeft, oftewel een dag/nachttarief geactiveerd heeft.

[M.102] Type meter	%	Percentage slimme meters. Het betreft alle typen slimme meters, zowel de op afstand schakelbare als de niet op afstand schakelbare meters (resp codes DUS en DUN)
[M.115] Aantal telwielen	getal	Gemiddeld aantal telwielen van de meters

As this data contains only aggregate data it does not fit directly in the IMSG model, but it describes aggregates of MeasurementValues and PostalAdresses. The question whether aggregates of existing types should also be part of an information model is not answered yet.

5.2 Zonnedael smart meter readings

The 'Zonnedael' data set is provided by Alliander as open data (see the section 'Slimme meter' on web page https://www.liander.nl/over-liander/innovatie/open-data/data). The data set contains readings from smart energy meters, collected in 2013 by some eighty households. For reasons of privacy, the readings have been given a slight random alteration and the addresses have been changed to comparable addresses in Julianadorp. Actual readings were taken at an undisclosed location elsewhere in the Netherlands. The data set contains readings from both gas meters and electricity meters. Some households in the dataset have the means to generate electricity themselves. In those cases, the electricity meter has recorded both consumption and production.

This data set was published as Linked Data in the project. The URI of the data set is http://lod.geodan.nl/cerisesg/ebif/julianadorp/. The data from this data set were used in the first demonstrator (web application) that was developed in CERISE-SG, to give an idea of possibilities that could arise if household level meter readings were to become available on the web.

This data covers the MeasurementValue, the BagAdres and MeasuredObject parts of the IMSG model.

5.3 Mpare Smart Meter Readings

MPARE offers an open platform for sharing data about energy generations and consumption. The MPARE system comprises three main parts. The Energy Data Box (EDB) is a small device that is installed in a household and collects energy data. The box is connected to the internet and the data is sent to the online ENDAX platform where you can manage your data and share it with third parties. Third parties, such as housing corporations, energy collectives, project developers and equipment manufacturers can created applications or apps that use your data. Via ENDAX (Energy DAta eXchange) users can control who has access to their data.

We describe the data that can be accessed via the ENDAX interface of which the API is defined in MPare (2014). The web interface of the EDB that can be accessed via a local connection the user can discover the data sources that are connected to the EDB and label them with a meaningful name. This data that is described here is transmitted to the ENDAX system.

Application builders that wish to use the data from a specific user request permission to the user via the ENDAX system: The application subscribes to the data stream from the ENDAX server and data is provided to the app in near-time via http post messages. When a user runs an app the user is identified in de Service Provider API (SPA) by a unique identifier. The SPA request data sources and when the user agrees a subscription identifier is returned. Each data source has a resolution (update frequency) (Between 1s and 1d), a description of the quantity (water, gas, electricity, power) and a sample type (instantaneous, minimum, maximum, average, cumulative). So the ENDAX system gives the user control over which data is provided: when an application is run, the SPA only knows the subscription identifier of the user and it is up to the user to reveal his or her location to the service provider. As a result this dataset covers the MeasurementValue and MeasurementInstrument of the IMSG model; there is no direct link to the Location of the user.

6 Available Vocabularies, Ontologies and Information Models

In this paragraph we describe vocabularies that are have used in our harmonization process. The vocabularies of models are all relevant to the CERISE-SG project as they are used to describe the data of chapter 5. For each vocabulary we give a reference to the datasets that use the vocabulary. In chapter 7 we harmonize the datasets of chapter 5 by defining mappings between the vocabularies in which they are defined.

6.1 CERISE Smart Grid Information Model (SGIM)

The SGIM is developed to facilitate information exchange in the context of smart grids. Strong emphasis is on spatial information as related to the energy domain and specifically smart grid applications. Smart grid in this context is defined as data related management of the grid by dynamic data capture and exchange through a network of sensors using sensor web technology.

he application domain is primarily energy balancing using smart grid technology but is extensible to other specific use. The approach is in line with object-oriented information modelling using UML class diagrams; the resulting diagram is given in Figure 4. The model defines semantics without implementation specifications. It is developed in CERISE-SG WP30 and fully described in (CERISE-SG, D3., 2014).
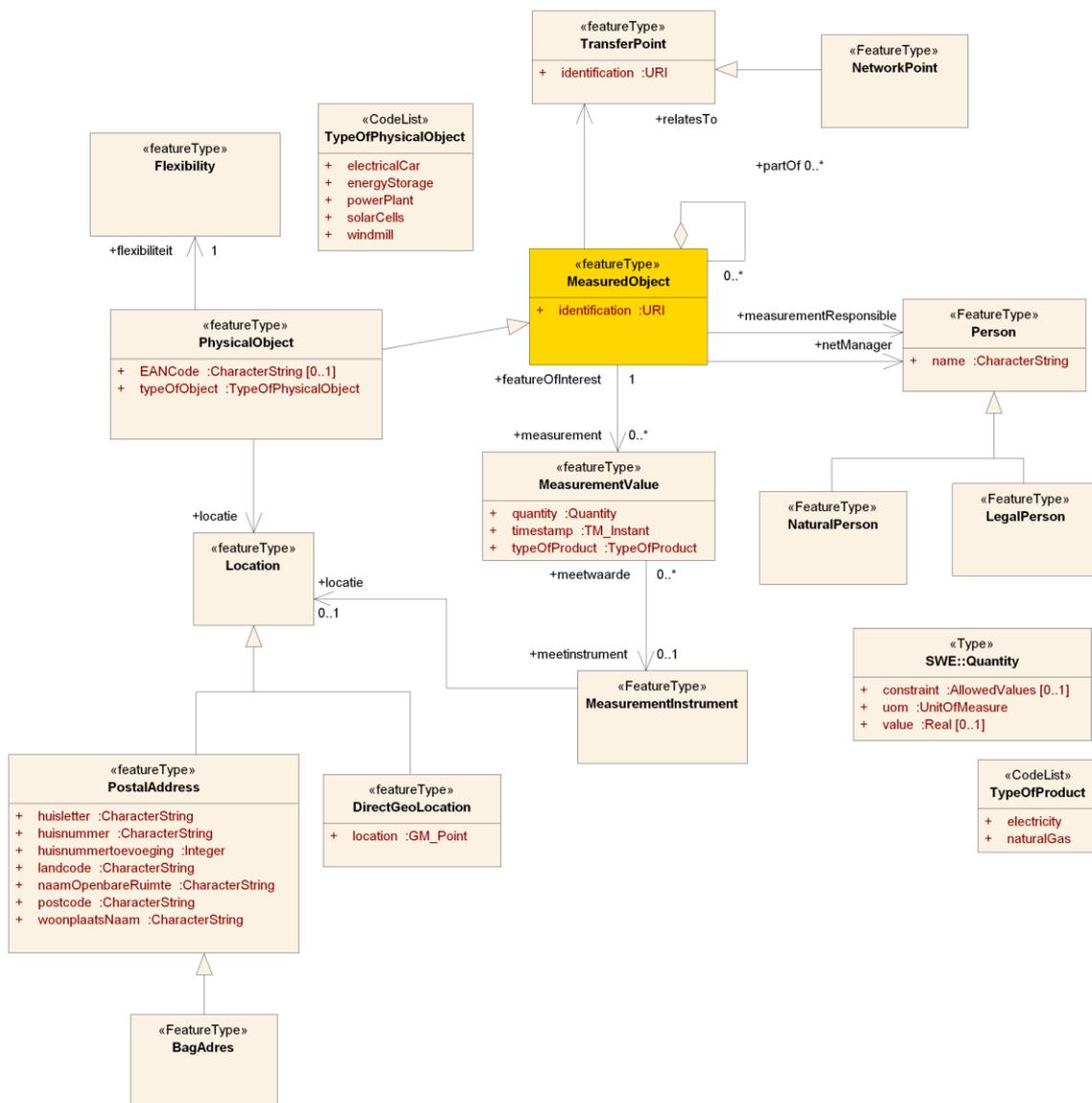
Figure 4 Information Model Smart Grid energy (IMSG) balancing model as defined in D3.1

The model relates location data to energy consumption and production data. Basically the energy data is brought into the context of a spatial data infrastructure. The model defines a conceptual information unit named Information Point. This point represents any point in a grid of which energy data are needed or wanted. The Information Point contains metering values that are provided by (smart)meters or sensors. Sensors have a location and are related to physical objects like buildings, rooms, street furniture, wind turbines etc. Information Points also are related to authorities with specific roles like grid manager. The physical objects again also relate to physical or legal persons and have direct location coordinates or are referenced by address.

The sensors provide data to the information point about energy consumption and energy production. By applying a temporal model energy data can be accessed through time and aggregations can be computed. Together with characteristics of energy production, consumption and storage at the level of the Information Point detailed balancing related calculations can be made. Since the information is spatially referenced spatial analysis and aggregations are possible.

The Information Point is furthermore associated to a Transfer Point in a grid. The information then becomes part of the grid topology or connectivity.

As a result energy data about production, consumption and storage are related to topographical objects, grid infrastructure, persons and legal entities and geography in general, providing a basis

for standardized data exchange or any kind of spatial related analysis or data publication. Depending on data availability data exploitation can be real-time.

During the project the focus from using the CERISE SGIM to map standards was changed to using CIM for the power system related information. We still consider the CERISE SGIM to be a good step to be able to map information from a broader domain but more works needs to be done to verify whether the SGIM model is complete or needs to be modified.

As the SGIM vocabulary touches all part of the smart grid all datasets can use this vocabulary to map their data.

6.2   Common Information Model (CIM)

The Common Information Model (CIM) is a standard developed by the electric power industry under the auspices of the International Electrotechnical Commission (IEC) that aims to allow application software to exchange information about an electrical network.

The CIM standard is currently maintained as a UML class model. It defines a common vocabulary and basic ontology for aspects of the electric power system. The model was created in different phases and consists of different parts. We have used the following versions of the IEC standards as a starting point.

- IEC 61970 cim16v17 (Core CIM, transmission focus)
- IEC 61968 cim12v06 (Distribution extensions)
- IEC 62325 cim02v07 (Energy Market Extensions)

The IEC CIM standard defines about 1500 different concepts (classes) and a multitude of properties divided over some 100 packages. However, for the purpose of the work presented here, we are not interested in the entire CIM, but only on these packages, classes and properties that are relevant to express metering- and location related information. Therefore, we present here the specific metering snapshot of the CIM, from which we selected only concepts and properties that are relevant to the Cerise metering domain. This selection is specified in the cim-metering profile.

This Cerise cim-metering profile contains core information classes representing metering, end-devices and other concepts necessary to cover the variety of metering datasets in the Cerise project. While we do not support specific metering functions yet, we concentrate on the information stated in the current datasets. Based on these datasets we are interested in the next IEC CIM classes (main classes):

- *UsagePoint:* a specialised subclass of IdentifiedObject
- *UsagePointLocation:* a specialised subclass of Location
- *Meter:* a specialised subclass of EndDevice
- *MeterReading:* Set of values obtained from a meter *EndDevice*.
- *Reading:* Specific value measured by a meter or other asset, or calculated by a system
- *Channel*: a single path for the collection or reporting of register values over a period of time.
- *Register*: a device that indicates or records units of the commodity or other quantity measured

A *UsagePoint* is a representation logical or physical point in the network to which readings may be attributed. Used at the place where a physical- or virtual meter may be located; however, it is not required that a meter device is present.

The figure below shows the *UsagePoint* class and its properties that are relevant for the purpose of this work.
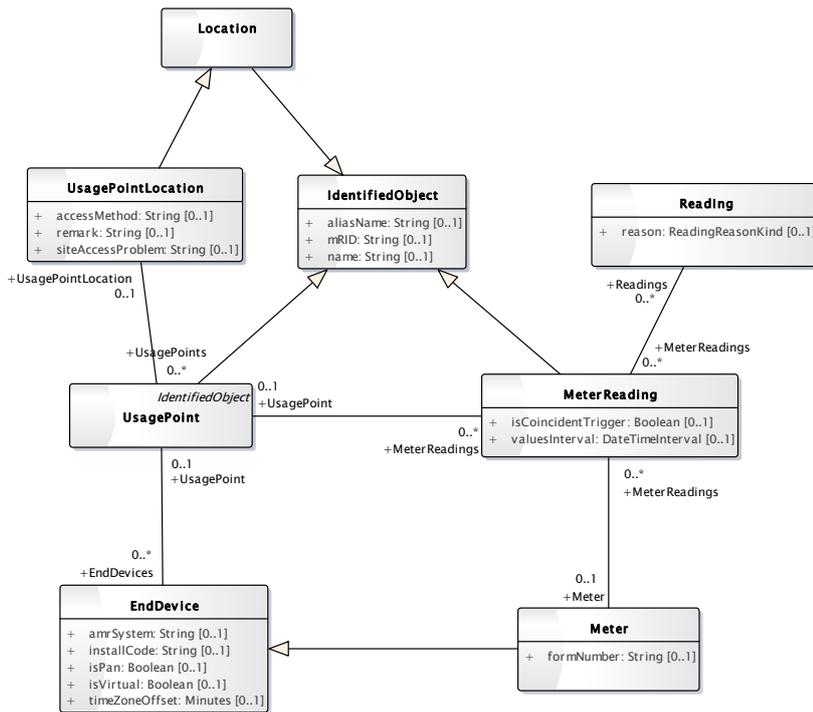
Figure 2 IEC CIM based metering model (part of Cerise cim-metering profile)

The location of an individual usage point is modelled as a *UsagePointLocation*, which is a specific type of *Location* in CIM. A *UsagePoint* is the place, scene, or point of something (e..g. address-, electronic- of GPS-location are examples of locations). Example of a UsagePoint is the place where a physical or virtual peter is located in the power system. It can be defined with one or more position points (coordinates) in a given coordinate system. The figure below shows the *UsagePointLocation* and *Location* classes and their attributes and relations with other classes that are relevant for Cerise metering domain.
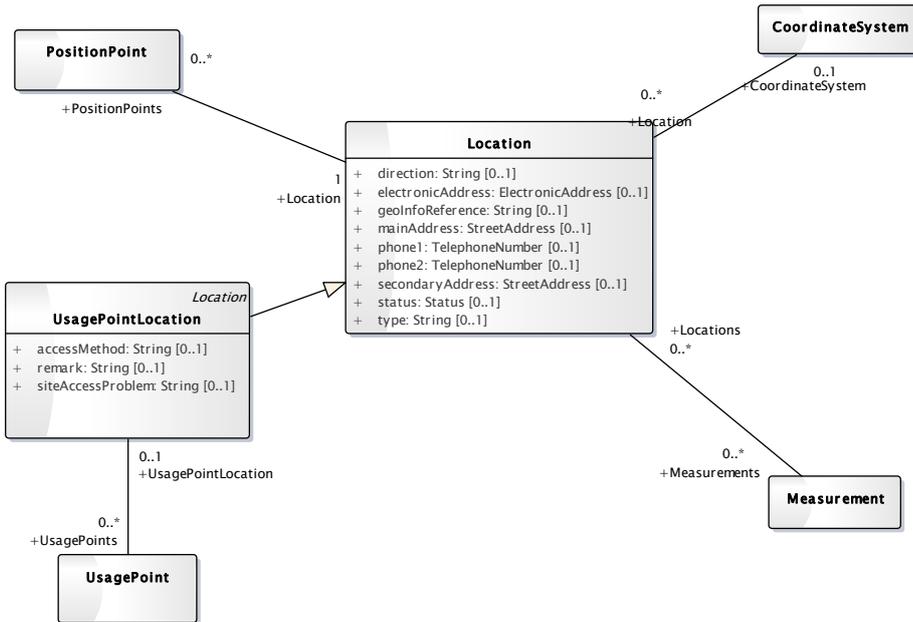
Figure 5 IEC CIM based location and usage point location model (part of Cerise cim-metering profile)

A *Meter* is defined as a physical asset that represents the physical attributes of a meter *EndDevice* and enables the implementation of *MeterReading* at a specific *UsagePoint*, used for measuring consumption and detection of EndDeviceEvent.

A *Meter* specializes the concept of *EndDevice*, which is an *AssetContainer* that performs one or more end device functions. One specific type of *EndDevice* is a *Meter* which can perform metering and optional other *EndDeviceFunction* like, load management, connect/disconnect, accounting functions, etc. Other type of *EndDevice*, such as monitoring and controlling air conditioner, refrigerator and pool pumps may be connected to a *Meter* as well. All *EndDevice* may have communication capability defined by the associated communication function(s).

The figure below shows the Meter class and its properties and relations with other classes in the CIM smart grid profile that is relevant for this work.

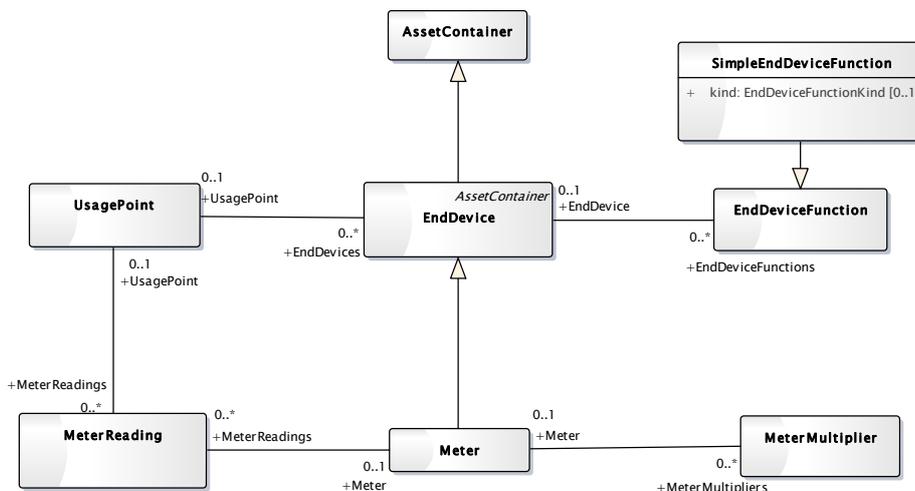**Figure 6 IEC CIM based Cerise end device asset model (part of Cerise cim-metering profile)**

A *MeterReading* is a set of values obtained via the *Meter*. Each *Reading* in the set of *MeterReading* is a specific value measured by a *Meter*, or calculated by a system. Each *Reading* is associated with a specific *ReadingType*. An example of a *ReadingType* is a individual- or regular 15-minutes interval reading. The figure below shows the *MeterReading* class and its properties and relations with other classes that are relevant for the purpose of the IEC CIM smart grid profile used in this work.
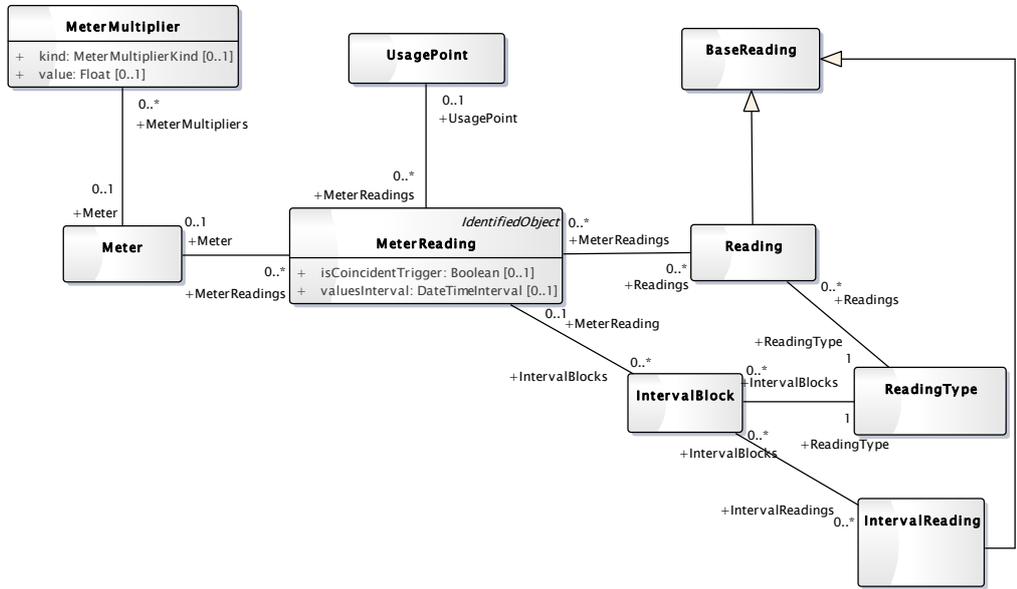

Figure 7 IEC CIM based MeterReading model (part of Cerise cim-metering profile)

The combination of diagrams presented in this section is used to construct a Cerise cim-metering profile, an IEC CIM method to specify an extract based on the overall model. The Cerise cim-metering profile contains the IEC CIM classes mentioned for the specification of terms relevant for the expression (ontology) of Cerise open dataset information.

The IEC CIM model methodology is based on the model driven principles of the OMG Model Drive Architecture. The computational independent model is translated to platform independent models for message-, storage- and ontology implementation schema's. For our purpose we are interested in the translation of the UML model specifications into ontology specifications. *CIMtool* is an open source tool that supports this translation into ontology schema's in different formats (.ttl, OWL/XML and .nt). However, there are more (open) tools available which can be used to make this type of profile specifications.

CIMtool is responsible for the interpretation of the UML model (.xmi- and/or eap based), and is doing this with great difficulty[10]. Due to the object oriented orientation of UML and the ambiguous interpretation of the UML-XMI and OWL interchange format, CIMtool produces several interpretations of OWL profiles. For more information about the profile interpretation and generation see the CIMtool website[11].

Based on the concepts and properties of interest of the CIM that are described above, we created a CERISE cim-metering profile using the CIMtool and generated an OWL version of this profile (turtle format). This ontology is published and available at http://ontology.tno.nl/cerise/cim-profile.ttl, while the profile documentation is available at http://ontology.tno.nl/cerise/cim-profile.

---

[10] Explanation of the CIMtool translation from UML to OWL http://wiki.cimtool.org/UMLOWL.html
[11] http://wiki.cimtool.org/CIMTool_Topics.html

The CIM vocabulary is used in semantic mappings for all the dataset of chapter 5 that contain Meters or Meter readings.

6.3   EBIF Vocabulary

The EBIF vocabulary (http://lod.geodan.nl/vocab/ebif) is a vocabulary that was developed and published early in the project. It was created to be able to publish the 'Zonnedael' data set (see section 6.2) as Linked Data, to be used in the first use case, the Energy Broker Information Facility (EBIF). This vocabulary can be used to express energy meter readings. As such, it purpose is comparable to the CIM-metering vocabulary (see section 7.2). The EBIF vocabulary was published before the project identified CIM as the main information model for energy data and before the CIM-Metering vocabulary was made.

Notable differences between the EBIF vocabulary and CIM-Metering is that the EBIF vocabulary is less extensive and reuses externally defined concepts. For example, an energy meter is defined to be a subclass of the general Location class as defined in the Dublin Core Metadata vocabulary. This indicates that a meter has a physical location which could be encoded by an address, a geometry, or something else.

6.4   BAG vocabulary

The BAG vocabulary was made to accommodate publication of the base registry for buildings and addresses (Basisregistratie Adressen en Gebouwen) as Linked Data. At the time, common semantics for Dutch base registries were available: the System Catalogue (Stelselcatalogus, see http://www.digitaleoverheid.nl/onderwerpen/stelselinformatiepunt/stelsel-van-basisregistraties/stelselvoorzieningen/stelselcatalogus). However, the RDF representation of the System Catalogue was considered insufficient for publication of Linked Data. For that reason, only the semantics of the BAG were published as a vocabulary. Were possible, links to the System Catalogue are made.

To aid publication of interoperable data, the BAG vocabulary was published in English next to Dutch. Also, it contains links to common concepts from, for instance, the Location Core Vocabulary (http://www.w3.org/ns/locn), the Dublin Core Metadata vocabulary (http://purl.org/dc/terms) and GeoSPARQL (http://www.opengis.net/ont/geosparql) . Those measures should make it possible for international users or users without intimate knowledge of the BAG information model to make use of data that are published with semantics from the BAG vocabulary.

6.5   Location Core Vocabulary

The ISA Programme Location Core Vocabulary (http://www.w3.org/ns/locn) is published by the W3C and mainly a product of INSPIRE R&D. The vocabulary can be used to describe locations by means of address or geometry. IT is a relatively simple vocabulary, which makes it a low-threshold way of adding general location semantics to data. As such, it can play an important role in data harmonization between domains in which location plays a role. In the specific case of CERISE-SG, geography is a type of data that is present in the energy domain and the government domain. Should data publishers in both domains publish location data with semantics from the Location Core Vocabulary, this will establish semantic interoperability.

It should be noted that it is possible and likely expressions of locations according to semantics defined in the Location Core Vocabulary will be accompanied by semantics from other sources. For instance, next to Location Core semantics locational semantics from the CIM or governmental models can be used. This enables discovery and selection of data by either worldwide general semantics and specialized national or domain specific semantics.

With the Location Core vocabulary being a simple and extensible model that can provide semantic interoperability for location data, it is recommendable for domain models like CIM or the Dutch System Catalogue to incorporate these semantics in future versions.

7 Mapping by ontologies

This chapter describes the implementation of a semantic mapping by mapping the relevant concepts to a shared vocabulary. First in

7.1 Conceptual mapping

As the shared vocabulary the (subset of the) IEC CIM (Common Information Model) described in section 6.2 is used. The concepts of the EBIF vocabulary as described in section 6.4 are mapped to the CIM model. We have chosen this mapping for demonstrating how to integrate different datasets based on different ontologies/information models for the following two main reasons:

- The IEC CIM is in an international information standard that defines common concepts for the electric power industry. As explained in (Crapo, 2010) this standard represents "the most complete and widely accepted model for generation, transmission, and distribution of electrical energy". Therefore, by creating mappings to IEC CIM, we enable smart grid datasets build on different vocabularies/models to be interoperable with the international community of industrial practitioners that is already using IEC CIM.
- The EBIF vocabulary is used as basis to build two of the datasets considered in this work, namely the Julianadorp- and Mpare datasets described in section 5.2 and 5.3, respectively. By mapping EBIF to IEC CIM we do not only allow these datasets to be integrated via the IEC CIM ontology (in other words, we can query EBIF concepts using the IEC CIM terminology), but from a practical point of view we can use the existing Julianadorp and Mpare datasets to test our mappings with SPARQL queries.

Notice that also other datasets based on a different vocabulary than EBIF, for example the Liander Open dataset (see section 5.1) can be made interoperable with the Julianadorp and Mpare datasets by creating an additional mapping to IEC CIM. The figure below shows the benefits of using IEC CIM as a basis for translation between different datasets and vocabularies for smart grid. These benefits are especially evident when the number of datasets/vocabularies (==number of interfaces) to be integrated becomes large, since instead of a set of dedicated mappings for each pair of vocabularies we would only need one set of mappings between the dedicated ontology to IEC CIM.
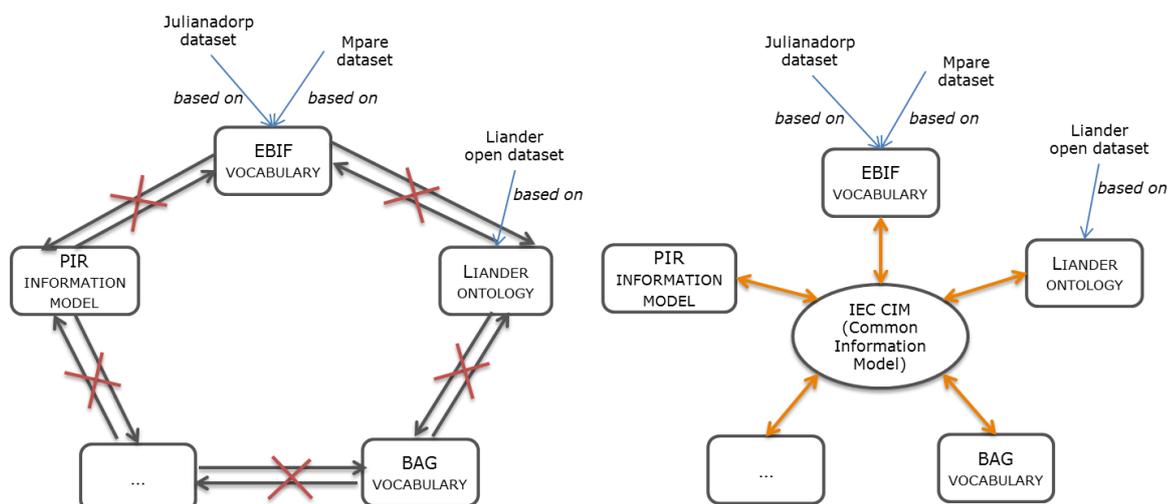
We acknowledge that using the IEC CIM as the reference model for translation among different datasets and vocabularies also has some drawbacks. For example, the IEC CIM is a large and complex conceptual model that requires extensive knowledge of the IEC standards. In contrast, lots of datasets are more practical and built using simple vocabularies to be easily used in real applications. However, we argue that this does not prevent one to consider only the parts of the CIM that are relevant for the application at hand. Towards this aim, the IEC CIM community has

developed the *CIMTool*, which is a tool that can be used to create subsets of the IEC CIM, or so called *profiles*, for specific purposes. In this way, it is possible to create profiles of the IEC CIM that can be used in different context and situations when only parts of the model are relevant. In section 8.1 we describe how we have created a CIM metering profile for the CERISE project and translated it into an OWL ontology that can be used as semantic reference model for the mappings shown in the figure above.

The following table presents the conceptual mappings between the EBIF vocabulary (see section 6.4) and the CIM profile (see section 6.2). These conceptual mappings provide a correspondence between concepts of the two models regardless to the language in which these models are expressed, specifying whether the two considered concepts are overlapping, or one of them is broader or narrower. The next section provides a formalization of these mappings taking into account the specific constructs of the language in which the source and target models are expressed.

EBIF concept	EBIF definition	EBIF Visualization	Mapping	CIM concept	CIM definition	CIM Visualization
Meter	A device that can measure the throughput of a energy medium		OVERLAPPING	Meter	A Meter is a physical asset that performs the metering role of the usage point. Used for measuring consumption and detection of events. It specializes the concept of EndDevice, which is an asset container that performs one or more end device functions.	
Electricity Meter	A device that can measure the throughput of electricity		OVERLAPPING (not 1 to 1 mapping; several concepts in CIM corerspond to only one concept in EBIF)	Meter, SimpleEndDeviceFunction of kind "electricMetering"	An EndDevice may have a related end device function that identifies a sensor or control point within a metering application or communications systems (e.g., water, gas, electricity).	
Gas Meter	A device that can measure the throughput of natural gas		OVERLAPPING (not 1 to 1 mapping; several concepts in CIM corerspond to only one concept in EBIF)	Meter, SimpleEndDeviceFunction of kind "gasMetering"	An EndDevice may have a related end device function that identifies a sensor or control point within a metering application or communications systems (e.g., water, gas, electricity).	
Measurement	A determination of energy througput in a certain time, on a certain moment		OVERLAPPING (not 1 to 1 mapping; several concepts in CIM corerspond to only one concept in EBIF)	Reading, ReadingType, MeterReading, DateTimeInterval	A MeterReading is a set of values obtained from the meter. Each Reading in the set is a specific value measured by a meter or other asset, or calculated by a system. Each Reading is associated with a specific ReadingType.	

7.2 Preparing the Formalization of the mappings

The EBIF vocabulary is expressed in RDF and the IEC CIM model in UML. In order to formalize the EBIF2CIM conceptual mappings, such that they could be used for automating data integration, it was necessary to generate a RDF/OWL version of the IEC CIM model that was originally expressed in UML. Since we were interested only in some parts of the IEC CIM model, we first specified a IEC CIM-based CERISE profile for metering and location using the CIMTool, and then transformed this profile from UML/XMI to RDF/OWL, also using CIMTool. The actual mapping is described in Deliverable 5.1. Section 8.3 presents the formalization of the EBIF2CIM conceptual mappings using RDF/OWL. Finally, Section 7.4 proposes some sample SPARQL queries that can be used to test the mappings.

7.3 The formal Mappings

The table in section 7.2 showed the (high-level) conceptual mapping between the EBIF vocabulary and our CIM profile. The following table presents the formalization of this mapping using RDF/OWL. Then left side of the table shows the concepts under consideration for the mapping, namely *meter*, *electricity meter*, *gas meter*, *measurement*, *measurement value*, *measurement time*, and *meter address*. The green columns show the representation of these concepts in RDF triples (subject, predicate, object) using the EBIF vocabulary. The columns in red show the corresponding representation in RDF triples using the CIM profile. For example, the table should be read as follows:

- **Meter**: the concept of meter is expressed in the EBIF vocabulary using `ebif:Meter`, which is a subclass of `rdfs:Resource`. The `ebif:Meter` is also a subclass of `locn:Location` and `geosparql:Feature`. The concept of meter is expressed in the EBIF vocabulary not only with the `ebif:Meter` resource, but also using the `ebif:meter` property, which relates a meter (`ebif:Meter`) to its measurement (`ebif:Measurement`). On the right part of the table (in red), it is shown the `ebif:Meter` concept it is mapped onto the `cim:Meter` class, which is of type `owl:class`. Moreover, the `ebif:meter` property is mapped onto the `cim:MeterReading.Meter` property in the CIM profile, which relates the class `cim:MeterReading` to the class `cim:Meter`."

- **Gas Meter**: the concept of gas meter is expressed in the EBIF vocabulary using the `ebif:GasMeter` resource, which is a subclass of `ebif:Meter`. The mapping to the CIM profile is not straightforward, but includes a number of CIM classes and properties. In fact, a gas meter in the CIM profile is a `cim:Meter` class with a property `cim:EndDeviceFunctions`, which relates to a `cim:SimpleEndDeviceFunction` class with a property `cim:SimpleEndDeviceFunction.kind`, which has value `cim:EndDeviceFunctionKind.gasMetering`.

	EBIF vocabulary			CIM profile		
	ebif:subject	ebif:predicate	ebif:object	cim:subject	cim:predicate	cim:object
meter	ebif:Meter	rdfs:subClassOf	rdfs:Resource	cim:Meter	rdf:type	owl:Class
		rdfs:subClassOf	locn:Location			
		rdfs:subClassOf	geosparql:Feature			
	ebif:meter	rdf:type	rdf:Property	cim:MeterReading.Meter	rdf:type	rdf:Property
		rdfs:domain	ebif:Measurement		rdfs:domain	cim:MeterReading
		rdfs:range	ebif:Meter		rdfs:range	cim:Meter
electricity meter	ebif:ElectricityMeter	rdfs:subClassOf	rdfs:Resource	cim:Meter	rdf:type	owl:Class
		rdfs:subClassOf	ebif:Meter	cim:EndDevice.EndDeviceFunctions	rdf:type	rdf:Property
					rdfs:domain	cim:Meter
					rdfs:range	cim:EndDeviceFunction
				cim:SimpleEndDeviceFunction.kind	rdf:type	rdf:Property
					rdfs:domain	cim:SimpleEndDeviceFunction
					rdfs:range	cim:EndDeviceFunctionKind
				cim:EndDeviceFunctionKind.electricityMetering	rdf:type	cim:EndDeviceFunctionKind
gas meter	ebif:GasMeter	rdfs:subClassOf	rdfs:Resource	cim:Meter	rdf:type	owl:Class
		rdfs:subClassOf	ebif:Meter	cim:EndDevice.EndDeviceFunctions	rdf:type	rdf:Property
					rdfs:domain	cim:Meter
					rdfs:range	cim:EndDeviceFunction
				cim:SimpleEndDeviceFunction.kind	rdf:type	rdf:Property
					rdfs:domain	cim:SimpleEndDeviceFunction
					rdfs:range	cim:EndDeviceFunctionKin
				cim:EndDeviceFunctionKind.gasMetering	rdf:type	cim:EndDeviceFunctionKind
measure	ebif:Measurement	rdfs:SubClassOf	rdfs:Resource	cim:Reading	rdf:type	owl:Class
					rdfs:subClassOf	cim:BaseReading
	ebif:measurement	rdf:type	rdf:Property	cim:MeterReading.Readings	rdf:type	rdf:Property
		rdfs:domain	ebif:Meter		rdfs:domain	cim:MeterReading
					rdfs:range	cim:Readings
value	ebif:measuredValue	rdf:type	rdf:Property	cim:BaseReading.value	rdf:type	rdf:Property
		rdfs:domain	ebif:Measurement		rdfs:domain	cim:BaseReading
		rdfs:range	xsd:decimal		rdfs:range	xsd:string
time	ebif:time	rdf:type	rdf:Property	cim:DateTimeInterval	rdf:type	owl:Class
		rdfs:domain	ebif:Measurement	cim:DateTimeInterval.end	rdf:type	rdf:Property
		rdfs:range	xsd:dateTime		rdfs:domain	cim:DateTimeInterval
					rdfs:range	xsd:dateTime
				cim:DateTimeInterval.start	rdf:type	rdf:Property
					rdfs:domain	cim:DateTimeInterval
					rdfs:range	xsd:dateTime
address	locn:address	rdf:type	rdf:Property	cim:UsagePoint.UsagePointLocation	rdf:type	rdf:Property
		rdfs:domain	ebif:Meter		rdfs:domain	cim:UsagePoint
		rdfs:range	locn:Address		rdfs:range	cim:UsagePointLocation
				cim:Location.mainAddress	rdf:type	rdf:Property
					rdfs:domain	cim:Location
					rdfs:range	cim:StreetAddress
	locn:thoroughfare	rdf:type	rdf:Property	cim:StreetAddress.streetDetail	rdf:type	rdf:Property
		rdfs:domain	locn:Address		rdfs:domain	cim:StreetAddress
		rdfs:range	rdfs:Literal		rdfs:range	cim:StreetDetail
				cim:StreetDetail.name	rdf:type	rdf:Property
					rdfs:domain	cim:StreetDetail
					rdfs:range	xsd:string
	locn:postCode	rdf:type	rdf:Property	cim:StreetAddress.streetDetail	rdf:type	rdf:Property
		rdfs:domain	locn:Address		rdfs:domain	cim:StreetAddress
		rdfs:range	rdfs:Literal		rdfs:range	cim:StreetDetail
				cim:StreetDetail.code	rdf:type	rdf:Property
					rdfs:domain	cim:StreetDetail
					rdfs:range	xsd:string
	bag:huisnummer	rdf:type	rdf:Property	cim:StreetAddress.streetDetail	rdf:type	rdf:Property
		rdfs:domain	bag:Nummeraandui dingmutatie		rdfs:domain	cim:StreetAddress
		rdfs:range	xsd:integer		rdfs:range	cim:StreetDetail
				cim:StreetDetail.number	rdf:type	rdf:Property
					rdfs:domain	cim:StreetDetail
					rdfs:range	xsd:string
	locn:postName	rdf:type	rdf:Property	cim:StreetAddress.townDetail	rdf:type	rdf:Property
		rdfs:domain	locn:Address		rdfs:domain	cim:StreetAddress
		rdfs:range	rdfs:Literal		rdfs:range	cim:TownDetail
				cim:TownDetail.name	rdf:type	rdf:Property
					rdfs:domain	cim:TownDetail
					rdfs:range	xsd:string

The mappings above can be implemented using OWL constructs, such as `owl:equivalentClass`, `owl:equivalentProperty`, and `owl:sameAs`. The same mappings can also be implemented using rule languages, such as SWRL - the Semantic Web Rule Language (http://www.w3.org/Submission/SWRL/), which is based on OWL and RuleML, or SPIN - the SPARQL Inference Notation (http://spinrdf.org/), which is based on SPARQL . SPIN is the rule and constraint language used by Topbraid Composer.

7.4 Queries

The following table shows some sample queries that can be used to retrieve information from the Julianadorp dataset (EBIF vocabulary column), and the corresponding queries expressed using the CIM profile terminology. If the mappings EBIF2CIM presented in section 8.3 are implemented, the queries on the CIM profile column can also be used to retrieve values from the Julianadorp dataset.

Query	EBIF vocabulary	CIM profile

retrieve all meters	# select all meters
	SELECT *
	FROM <http://lod.geodan.nl/cerisesg/ebif/julianadorp/>
	WHERE {
	service <http://lod.geodan.nl/sparql> { ?meter a ebif:Meter . }
	}
	ORDER BY ?meter
		# C1 select all meters
		SELECT *
		WHERE {
		?meter a cim:Meter .
		}

retrieve all gas meters	# select all gas meters
	SELECT ?meter
	FROM <http://lod.geodan.nl/cerisesg/ebif/julianadorp/>
	WHERE {
	service <http://lod.geodan.nl/sparql> {
	?meter a ebif:GasMeter . }
	}
	ORDER BY ?meter
		# select all gas meters
		SELECT ?meter ?meterType
		WHERE {
		?meter a cim:Meter .
		?meter cim:EndDevice.EndDeviceFunctions ?function .
		?function cim:SimpleEndDeviceFunction.kind ?meterType .
		FILTER (?meterType = cim:EndDeviceFunctionKind.gasMetering)
		}

retrieve all electricity meters	# select all electricity meters
	SELECT ?meter
	FROM <http://lod.geodan.nl/cerisesg/ebif/julianadorp/>
	WHERE {
	service <http://lod.geodan.nl/sparql> {
	?meter a ebif:ElectricityMeter . }
	}
	ORDER BY ?meter
		# select all electricity meters
		SELECT ?meter ?meterType
		WHERE {
		?meter a cim:Meter .
		?meter cim:EndDevice.EndDeviceFunctions ?function .
		?function cim:SimpleEndDeviceFunction.kind ?meterType .
		FILTER (?meterType = cim:EndDeviceFunctionKind.electricMetering)
		}

retrieve all meters with their associated info	# select all meters and their info (measurement, unit of measure, time interval)
	SELECT ?meter ?measurement ?unit ?timeInterval
	FROM <http://lod.geodan.nl/cerisesg/ebif/julianadorp/>
	WHERE {
	service <http://lod.geodan.nl/sparql> {
	?meter a ebif:Meter .
	?meter ebif:timeInterval ?timeInterval .
	?meter ebif:unit ?unit .
	OPTIONAL {?meter ebif:measurement ?measurement . }
	}
	}
	ORDER BY ?meter
		# select all meters and their info (meter type, reading, time Interval)
		SELECT ?meter ?meterType ?reading ?timeIntervalStart ?timeIntervalEnd ?intervalDuration
		WHERE {
		?meter a cim:Meter .
		?meter cim:EndDevice.EndDeviceFunctions ?function .
		?function cim:SimpleEndDeviceFunction.kind ?meterType .
		?meter cim:EndDevice.UsagePoint ?usagePoint .
		?meter cim:Meter.MeterReadings ?reading .
		?reading cim:MeterReading.valuesInterval ?interval .
		?timeInterval cim:DateTimeInterval.end ?timeIntervalEnd .
		?timeInterval cim:DateTimeInterval.start ?timeIntervalStart .
		BIND (?timeIntervalEnd - ?timeIntervalStart AS ?intervalDuration)
		}

retrive all measurements, showing the corresponding meter, the measured value and the time it is measured	# retrive all measurements for a certain meter, showing the measured value and the time it is measured
	SELECT ?measurement ?meter ?value ?time
	FROM <http://lod.geodan.nl/cerisesg/ebif/julianadorp/>
	WHERE {
	service <http://lod.geodan.nl/sparql>
	{
	?measurement a ebif:Measurement .
	?measurement ebif:meter ?meter .
	?measurement ebif:meter
	<http://lod.geodan.nl/cerisesg/ebif/julianadorp/gmeter_klant43> .
	?measurement ebif:measuredValue ?value .
	?measurement ebif:time ?time }
		# select a reading, its meter, the measured value and time of measurement
		SELECT ?basicReading ?meter ?value ?timeIntervalStart ?timeIntervalEnd
		WHERE {
		?meter a cim:Meter .
		?meter cim:Meter.MeterReadings ?meterReading .
		?meterReading cim:MeterReading.Readings ?basicReading .
		?basicReading cim:BaseReading.value ?value .
		?basicReading cim:BaseReading.timePeriod ?timeInterval .
		?timeInterval cim:DateTimeInterval.end ?timeIntervalEnd .
		?timeInterval cim:DateTimeInterval.start ?timeIntervalStart .
		}

retrieve all meters with their address	# select all meters with their address
	SELECT ?meter ?woonplaats ?straat ?nummer ?letter ?postcode
	FROM <http://lod.geodan.nl/cerisesg/ebif/julianadorp/>
	WHERE {
	service <http://lod.geodan.nl/sparql> {
	?meter a ebif:Meter .
	?meter locn:address ?adres .
	?adres locn:postName ?woonplaats .
	?adres locn:thoroughfare ?straat .
	?adres locn:postCode ?postcode .
	?adres bag:huisnummer ?nummer .
	OPTIONAL {?adres bag:huisletter ?letter .}
	}
	}
	ORDER BY ?woonplaats ?straat ?nummer
		# select all meters and their address (street, number, postcode, town and country)
		SELECT ?meter ?streetName ?streetNumber ?postCode ?townName ?countryName
		WHERE {
		?meter a cim:Meter .
		?meter cim:EndDevice.UsagePoint ?usagePoint .
		?usagePoint cim:UsagePoint.UsagePointLocation ?location .
		?location cim:Location.mainAddress ?address .
		?address cim:StreetAddress.streetDetail ?street .
		?address cim:StreetAddress.townDetail ?town .
		?street cim:StreetDetail.name ?streetName .
		?street cim:StreetDetail.number ?streetNumber .
		?street cim:StreetDetail.code ?postCode .
		?town cim:TownDetail.name ?townName .
		?town cim:TownDetail.country ?countryName .
		}

8 Evaluation and recommendations

In this document we describe how the information from the three domains that are of interest to the CERISE-SG project can be harmonized by first mapping the information of the domains to semantic web standards and then harmonizing the models by defining semantic mappings between the models. We discuss two mapping strategies. In the first strategy mappings are made between the different domains models. This has been executed in a separate research (see document D4.2). In the second strategy mappings from the domains are made to a shared model. Chapter 7 describes our experiments with this mapping and it can be seen that it is technically possible to formally define mappings between the various smart grid datasets that we collected and harmonized. When the mapping is done, it is possible to query a dataset using the centralized vocabulary. The query is then automatically transformed to a query on the domain dataset where it is executed and the result is translated back to the centralized vocabulary. So harmonization is achieved because the translation is done transparent to the user.

One of the things that is not easy to solve in a harmonized dataset is the level of detail needed when a shared central model is used for harmonization. For example in our experiment we used CIM as a central model. In CIM the link between a ElectricityMeter and its Readings goes via several intermediate classes whereas in the EBIF model the relationship between a meter and its readings is more straightforward. Obviously the complexity of a model depends on it application domain and making a shared model for different domains will result in a very complex model.

The decision whether to use a shared model or define direct mappings between the different domains depends on the use case. Semantic web technologies can be used for both strategies. One of the advantages of using semantic web strategies is that the process of harmonization is separate from the process of data conversion, so it possible to first convert your data to semantic web standards (i.e. triples) and then performing the semantic mappings.

9 References

Cerise D3.1: Informatiemodel Smart Grid - Energy Balancing Information Facility (IMSG-ebif) (2014)

Cerise D4.2 Linked Data for Crisis management (2015)

Cerise D5.1: Cookbook for Standardization and Harmonization (2015)

Crapo, A., et al.: Overcoming Challenges Using the CIM as a Semantic Model for Energy Applications. Grid-Interop Forum, GridWise Architecture Council (2010)

MPare. Service Provider Applications (0.5) (2014)