

# Semantic Annotation Engine for Smart Grid Applications

Ralf Mosshammer, Alfred Einfalt and Andreas Lugmaier  
Siemens AG Österreich  
Corporate Technology  
Siemensstraße 90, 1210 Vienna, Austria  
{ralf.mosshammer, alfred.einfalt, andreas.lugmaier}@siemens.com

Jack Hodges, Florian Michahelles  
Siemens Corporation  
Corporate Technology  
1995 University Avenue, Berkeley, CA, USA  
{jack.hodges, florian.michahelles}@siemens.com

**Abstract**—The advance of Smart Grid concepts towards lower-voltage layers of the electricity grid connects an enormous number of devices, many of them through legacy automation protocols, to back-end IT systems. With many of these devices giving little or no indication as to the semantics of their data, it is customary to manually engineer this information, which is cumbersome, costly and prone to errors. In this paper, we follow the web of things approach to introduce a smart aggregation layer device which allows for easy annotation of data with URIs tying into well-established ontologies, thus abstracting from technical addressing schemes. We present an annotation engine that can interlink actors of different sophistication levels in a light-weight messaging scheme, distributing data between operative algorithms, communication applications and operator interfaces. We present results of an ongoing field deployment of this approach in Austrian Smart Grids model regions.

## I. INTRODUCTION

The electrical grid has traditionally been a domain associated with robust, engineering-heavy solutions, with planning and operation dependent on the experience of seasoned professionals. The reason for this, aside from long-term investments, was the lack of observability and automation, especially in the low-voltage sections that are directly connecting households and small businesses to the grid.

With the advent of the Internet of Things paradigm [?], the Internet – originally limited to virtual data – now starts reaching out into the real world and providing access to real-world phenomena in real time.

The Smart Grid follows a different approach: instead of analog meters, a new class of digital measurement devices (“Smart Meters”) is currently being introduced, to some extent even as part of overall Smart Building management concepts. This approach allows for progression from the simple reporting of voltage measurements to complex predictions of overall consumption, and further enables control algorithms, optimizers, trading systems, and end-user feedback interfaces. There are several ongoing research projects addressing these topics [11], [8], [1].

From an engineering perspective, this vision has much in common with the idea of interconnecting everyday objects in a Web of Things [5]: a massive number of devices needs to be integrated; the devices are often limited in processing power and general communication capabilities; and heterogeneity such as conflicting protocols and data formats must be dealt

with. In addition, the electrical grid imposes further restrictions that are tightly associated with the fact that it represents critical infrastructure backed by mostly long-term investments: a large number of devices that are not eligible for replacement for years to come is already deployed in the field, and these devices are often bound to legacy protocols that lack the ability to structure data, thus hindering configuration and engineering.

In this paper, we present the development of an embedded data annotation engine for Smart Grid applications that provides monitoring and control in low-voltage grids in a decentralized, distributed way. We provide first results from a field test conducted in Austrian Smart Grid model regions. We present the installation on a device that is able to deal with legacy automation protocols (IEC 60870-5-104, ModBus) and bridges the gap between automation- and IT domains by providing a consistent communication framework, application management, and a dashboard driven by RESTful Web services.

The work that we describe in this paper is inspired by [10] which discusses a natural language processing approach to extract information from tweets and transform it into a semantic knowledge base. While the underlying data sources of this approach and ours are completely different, the idea of auto-annotation is similar. Likewise, our work is based on [13]. Other research like [14] focuses on demand-response load optimization. Our approach is much further ahead, as it already has been installed and implemented in a Smart Grid field test. We see our approach as a building block for Smart Grid architectures [12], as it supports the annotation of Smart Grid components based on data without requiring operator intervention.

## II. CONCEPT

Our goal in this paper is to demonstrate the advantages of annotating data with Uniform Resource Identifiers (URIs) – this is a well-established technique in the semantic modeling community, e.g. [3] and facilitates the storage, search, transport, and processing of raw automation data, while at the same time serving as a point of reference for comparison to other information sources. Additionally, we describe how basic annotation information can automatically be attached to data without the intervention of a human engineer.

While remote meter reading is a first step of forming a Smart Grid, the power grid can further be optimized by

establishing bidirectional communication to actor components:

- In rural areas, voltage control algorithms can be introduced to maintain power quality in the presence of distributed loads (e.g., electric vehicles) and generators (e.g., solar panels).
- Household-level generators can be federated in virtual power plants to partake in market operations.
- Smart Building controllers can be integrated into the Smart Grid to fully harness their flexibility.
- Aggregation-level devices can intelligently filter data to provide semantically rich information to the back end.

The following sections will outline problems arising in the implementation of this ‘second level’ Smart Grid architecture, and present a solution on the aggregation level, implementing a voltage control and grid monitoring application.

### A. Requirements

The traditional low-voltage grid is built as a hierarchical architecture (see Fig. 1). A large number of devices is distributed in the field and communicates with several Data Concentrators (DCs) that represent the meter communications network endpoint. This communication link allows for low-bandwidth, possibly unreliable communication, where common options include meshed radio, cellular link (in rural areas mostly limited to GPRS), and Power/Distribution Line Communication (P/DLC). The latter option enjoys widespread acceptance due to the fact that no extra wiring is necessary, thus curtailing costs. From the DCs, data is transported to an IT back-end system via high-speed (TCP/IP) links.

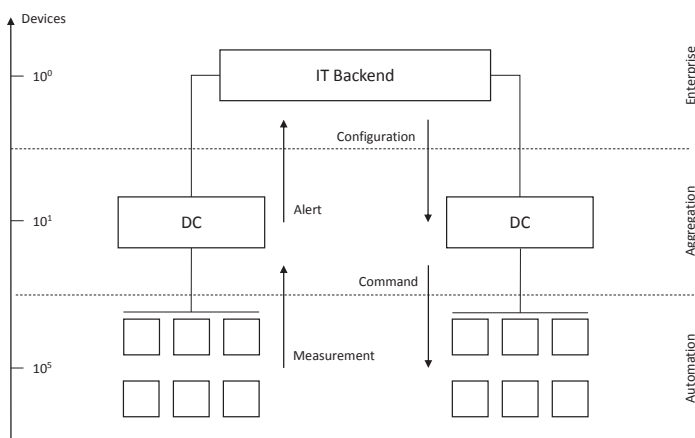


Fig. 1. Classical hierarchy of a Smart Grid automation architecture

As shown in Fig. 1, the low bandwidth of the sensor-to-DC connection inhibits the usage of protocols with large overhead. Because additional restrictions come from requirements that pertain to process safety, common protocol choices originate from the equipment automation domain. Smart Meters are designed as simple measurement devices that currently are not capable of providing structured data. These systems usually run proprietary event loops instead of operating systems.

The typical use case that drives the communication requirements is the monthly transmission of billing information based on daily user consumption data. Thus, the communication mechanism is characterized by low transmission frequencies (once per month), high tolerance for latency (the interrogation is triggered with a large temporal safety margin to account for slow connections) and unidirectionality.

The principal data source is the DC which provides structured data (i.e., unique address per data point time stamp) following IEC 60870-5-104 [2]. However, the only additional information about these data points is their *type* (e.g., floating point value, single point, double point).

Thus, manual effort is required to map system addresses to human-readable descriptions that consist of a *device* identifier (such as ‘meter23’, ‘transformer’, ‘inverter3’, etc.) and a *data point* description that usually identifies the actual type and unit of the measurement (or command): ‘voltageAn’, ‘TapPosition’, etc.

This approach makes routing information to relevant actors much more manageable and abstracts from the protocol details, but also comes with two drawbacks:

- 1) Manual inspection of the actual data is required to perform the routing.
- 2) The complexity of routing instructions increases strongly with the number of data points.

Since practical use cases comprise queries such as “all meters in a specific region”, “sensors that exhibit irregularities”, or “data that needs to be persisted”, a routing algorithm that simply reads the header (which contains the tags) is much more efficient, compared to algorithms that perform deep inspection of individual packets.

### B. Problem statement: Grid Monitoring and Control

Due to the emergence of distributed power generation (e.g., solar panels) and dynamic consumption (e.g., electric vehicles), implementing the regulatory requirement of staying within a stable voltage band becomes a major challenge for utilities. The basic approach to achieve that is to regularly collect voltage readings from the Smart Meters – together with checks for impending violations of the voltage band at the transformer, this allows for according and timely reactions (see [4] for details). Fig. 2 shows an exemplary setup, consisting of a DC that collects data from sensors and acts upon a voltage-controllable transformer, and a platform extension with the actual application software, including a lightweight message bus with routing capabilities, a Web server for providing REST services, and a storage solution for persisting measurement time series data.

In this simplified setup, the only actor available to the control algorithm is the transformer, which can be stepped up or down to influence the voltage. More sophisticated setups allow for regulation of electric vehicle charging current, setting of inverter characteristics, or sending arbitrarily complex commands to Smart Building controllers.

To provide semantic context (classification categories, in this case), data points must be individually *engineered*, i.e., augmented with semantic meta-data by a human operator, a

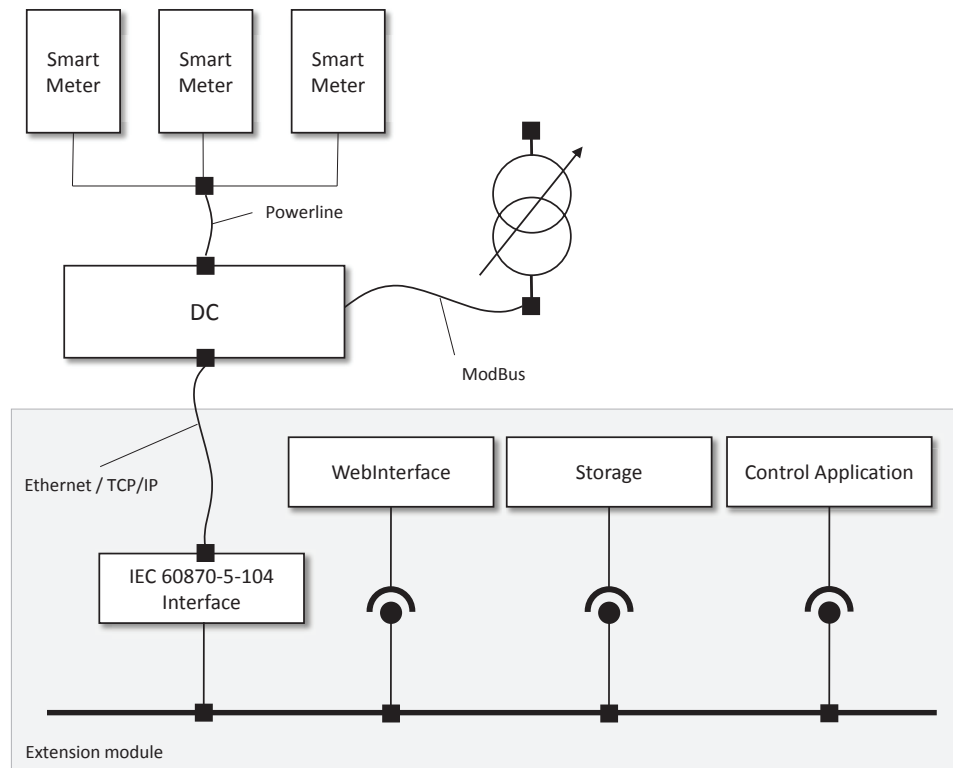


Fig. 2. Voltage control application installed on a DC extension module

process that is suspected to incur prohibitive efforts (and costs) with widespread adoption of Smart Grid technology.

To allow leveraging the benefits of semantic annotations without incurring many of these costs, we propose to assist the engineering process by providing automatic annotations based on data that is collected from the network.

### C. Approach

For adding semantic description to the data handled in the grid, we decided to introduce an abstraction layer atop the DC component. This allowed us to implement our solution independently of the DC hardware.

Communication of this platform with the DC is established via Ethernet link transporting IEC 60870-5-104, a well-established automation protocol providing client-server setups, interrogations, automation-domain specific, process-safe data types (such as double point command) and safe handshake procedures.

A REST Webservice provides connectivity to components making use of the collected data from the grid.

Our approach comprises the following steps:

- 1) Map the IEC 60870-5-104 address of a data point to a corresponding device/datapoint identifier to facilitate readability and routing engineering. Principal devices in this application included *meter* (Smart Meter), *trafo* (transformer) and *inverter* (photovoltaic inverter). Measurements from field devices included three-phase voltages ( $V_{An}$ ,  $V_{Bn}$ ,  $V_{Cn}$ ) and, for the transformer, active and reactive power ( $W$ ,  $VAR$ ).

- 2) Assign internal data points according to the same scheme, including commands to set the transformer tap position, the state of the algorithm (offline, enabled) and heartbeat signals to and from the DC.
- 3) Manually attach URIs[3] to the data points relating to their *identity*, e.g.:

`http://qudt.org/schema/qudt#QuantityValue`  
for every data point representing a field measurement

`http://qudt.org/vocab/quantity#Voltage`  
for every data point representing a voltage measurement, etc.

- 4) Set up mechanisms to automatically attach URIs relating to the *behavioral aspects* of a data point, e.g.:

for data points failing basic filtering tests  
`http://purl.oclc.org/NET/ssnx/ssn\#OperatingRange`

for data which has to be brought to immediate attention of an operator (e.g. because of out-of-bounds voltage)  
`http://purl.oclc.org/NET/ssnx/ssn\#SurvivalRange`

for all data which needs to be persisted, based on the change frequency value  
`http://voag.linkedmodel.org/voag/ChangeFrequency`

for data coming in through the web interface from the logged-in user admin

based on a common governance model  
<http://voag.linkedmodel.org/voag/hasOwner>

- 5) Set up routing tables for the entire system based mostly on URIs and, if necessary, device/data point identity, e.g.:
  - a) <http://purl.oclc.org/NET/ssnx/ssn\#OperatingRange> → logging subsystem
  - b) <http://purl.oclc.org/NET/ssnx/ssn\#SurvivalRange> → web dashboard
  - c) <http://qudt.org/schema/qudt\#QuantityValue> → algorithm
  - d) <http://qudt.org/schema/qudt\#Quantity> → algorithm

These semantic categories are used in the classification process of incoming data (see Fig. 3, at 2), but they are also used more broadly in the semantic modeling communities, and they are related to one another. This combination makes it possible to relate and reason about incoming data in ways that key word classifications wouldn't support outside an enterprise model.

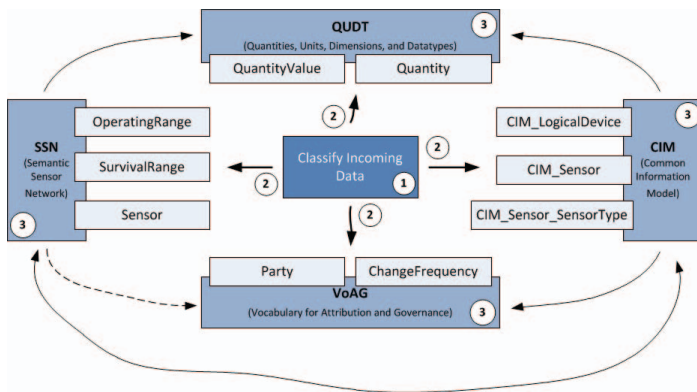


Fig. 3. Semantic models for smart grid data categorization, classification, and assignment

From Fig. 3, incoming data (from the device, at 1) can be classified along the dimensions mentioned above (e.g., QuantityValue for measurements or ChangeFrequency to determine if data needs to be persisted, at 2). These categories are themselves part of a broader set of semantic models (i.e., CIM for smart grid devices [15], QUDT for quantities and units [7], SSN for devices and their quantification [9], and VoAG for governance [6]) which are, themselves interrelated (at 3).

The same models can also be used by algorithms to change behavior through attribute value assignment, also illustrated in Fig. 3.

We evaluate the performance of this control algorithm by a proof of concept installation as described in the following section.

### III. PROOF OF CONCEPT IMPLEMENTATION

The application setup presented in section II-B was deployed as a field test in Austrian model regions

Salzburg/Koestendorf and Eberstallzell/Littring, with only slight differences in configuration. The principal goal of these deployments was the assessment of control algorithm performance, with the following additional requirements:

- Enabling fast turnaround from algorithm development in a simulation environment to field deployment.
- Providing an open-loop mode.
- Offering operators and scientists a comprehensive interface for operations monitoring.
- Offering the option to manually control operations from a remote location.

The system was installed inside the secondary substation or, in one case, remotely in a server cloud. Fig. 4 shows a screen shot of a web dashboard driven by the RESTful WebService interface on the extension module, which not only gives an overview of the state of the grid in various graphs, but also allows direct operator control of device operations, e.g. switching the transformer to a defined position or opening the control loop to manually confirm algorithm commands before they are sent to the DC.

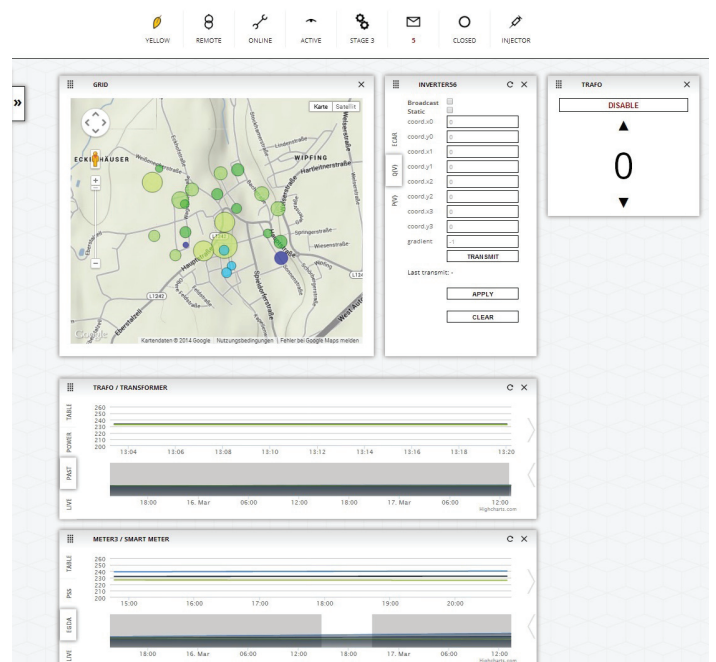


Fig. 4. Screenshot of the REST-driven web dashboard, showing the status of known grid elements projected onto a Google map view; a control element for setting inverter characteristics; the transformer tap position control; and two historical plots of voltages on the transformer and a selected Smart Meter.

## IV. DISCUSSION

The management of emerging Smart Grid networks based on current technology may become a tedious task. In the rural grid areas where this system was deployed, large amounts of PV and, in the case of Salzburg, e-car charging infrastructure were installed. Traditional engineering methods correspond to hard-wiring data points from source to sink. Even small changes to the network, e.g., an additional household or

charging pole, would require re-engineering and re-deploying the automation system, a process which is not only tedious but also critical, given that the DC primarily handles the task of collecting billing data from customers.

In this work, we outlined a possible migration path for moving legacy electrical grid automation installations towards the Smart Grid by introducing an abstraction layer on the aggregation level which allows for the annotation of data with URIs. This approach improves upon existing structures by:

- hiding details of the automation protocol addressing scheme from engineers.
- facilitating the handling of dynamics of the network structure (addition/removal of nodes).
- adding human- and machine-readable annotation and semantics to data.
- hiding details of semantic models from engineers.
- providing a way to distribute data between applications in a way natural to the application purposes and use cases.
- allowing for exposition of annotated data in a rich web dashboard.

Overall, this establishes a redistributable, modular Smart Grid application which can sit atop arbitrary automation protocol stacks.

#### A. Heuristic Auto-annotation

The existence of a decentralized repository of structured data finally allows to automatically add URIs to data. Heuristic assumptions can be applied as follows:

- Calculating statistical properties such as mean value and variance allows for the estimation of physical unit (e.g.,  $\mu \approx 230 \rightarrow \text{http://qudt.org/vocab/quantity\#Voltage}$ )
- Observing discrete value changes within defined limits (e.g.,  $[0,1]$ ) can indicate a <http://purl.oclc.org/NET/ssnx/ssn/SensingDevice\#Switch>.
- Observation of data frequency (from timestamps) can give valuable insights into the equipment status, e.g., whether it is `#offline`.
- Arbitrarily complex relations between data points (e.g., switch movement corresponds to voltage jump) can indicate their co-location in a specific <http://www.w3.org/ns/prov/Location\#GridSection>.

While heuristics are not yet yielding descriptions by themselves, they still reveal major events happening in the grid and can alert an operator of a web dashboard accordingly, e.g. *type* of device actually causing a voltage violation.

#### B. Annotation Wiki

Data which is partially machine-annotated can be offered to web dashboard operators for further annotation in a wiki scheme. To this end, data as displayed in the web dashboard would indicate the confidence of the existing annotation, either with a percentage or some simpler abstraction. Human operators could then be allowed to add annotations to increase the confidence level.

For example, a device might already be known to behave as a <http://purl.oclc.org/NET/ssnx/ssn/SensingDevice\#Switch>, be [http://vs.cs.hs-rm.de/CIM\\\_EnabledLogicalElement\\\_\\\_AvailableRequestedStates\\\_Value\\\_Enabled](http://vs.cs.hs-rm.de/CIM\_EnabledLogicalElement\_\_AvailableRequestedStates\_Value\_Enabled), emitting measurements at a <http://voag.linkedmodel.org/voag/High> and being located in <http://www.w3.org/ns/prov/Location/GridSection\#GS-A> together with a lot of elements producing <http://qudt.org/vocab/quantity\#Voltage> data exhibiting sudden value changes as the element itself experiences value transitions (switching actions). There is a high likelihood that the device producing this data is the transformer: it has a switching element (for controlling the tap positions), is connected to the DC with a comparatively stable and fast connection (ModBus), accounting for the high transmission frequency, and influencing elements in its grid section. Accordingly, this data might be assigned the <http://purl.oclc.org/NET/ssnx/ssn/SensingDevice\#Transformer> tag with likelihood 'high'. An operator on the dashboard can confirm this assessment with a single action, thus setting this datapoint to 'confirmed', which in turn can have a cascading influence on the identity estimation of other data points in the system.

#### C. Dynamic clients and Subscription

Semantic tagging opens the system for the possibility of a publish/subscribe routing scheme. Thus, instead of specifying at where to route data at the source, clients on the extension module could express interest in certain (composite) topics, such as meter voltages from a specific region (<http://purl.oclc.org/NET/ssnx/ssn/SensingDevice/Meter\#SmartGridMeter>, <http://qudt.org/vocab/quantity\#Voltage>, <http://qudt.org/vocab/unit\#VOLT>, <http://www.w3.org/ns/prov/Location/GridSection\#GS-A>) or commands issued from a specific operator (<http://voag.linkedmodel.org/voag/hasOwner\#MaintenanceBot>).

Hybridizing this approach with the existing source-based routing tables unifies the process security of the latter with the dynamics of the former approach: critical, monolithic clients such as the DC are engineered as unchanging parts of the system, which can be directly addressed with commands, while transient clients such as building controllers can freely connect to the system and subscribe interesting topics.

Using URIs, which can be transported via any mechanism, instead of more sophisticated data structuring opens the system to clients with limited capabilities. Furthermore, the set of URIs recognizable system-wide constitutes an ad-hoc data

structuring, which can be as sophisticated or simple as the application requires.

## V. CONCLUSION

In this paper we have outlined the challenges of managing the emerging smart grid with today's technologies. As supply and demand become more dynamic novel mechanisms of engineering such systems are demanded. We proposed the creation of an abstraction layer on the aggregation level which allows for the annotation of data with URIs tied to well-documented existing ontologies. We presented the results of a pilot installation in Austria and discussed heuristic auto-annotation, annotation wiki and dynamic client subscriptions as possible further research directions. Our research underscores the usability of semantic technologies to allow brownfield data from legacy devices to operate within state-of-the-art Internet of Things installations.

## ACKNOWLEDGMENT

The results presented in this work were obtained as part of the research project *DG DemoNet – Smart LV Grid*, funded within the program *Neue Energien 2020* by the Austrian *Klima- und Energiefonds* under project number 829867.

## REFERENCES

- [1] P. Birkner, I. Jeromin, M. Zdrallek, C. Oertner, and N. Neusel-Lange. Integrated smart grid concept — experience in a german distribution grid. In *Proceedings of the 22nd CIREN Conference*, Stockholm, 2013.
- [2] International Electrotechnical Commission. 60870-5-104, 2006. [Online; accessed 2014-04-17].
- [3] Dan Connolly. Naming and addressing: Uris, urls, ..., 2006. [Online; accessed 2015-05-07].
- [4] A. Einfalt, F. Zeilinger, R. Schwalbe, B. Bletterie, and S. Kadam. Controlling active low voltage distribution grids with minimum efforts on costs and engineering. In *Proceedings of the 39th Annual Conference on IEEE Industrial Electronics Society (IECON)*, Vienna, 2013.
- [5] Dominique Guinard and Vlad Trifa. Towards the web of things: Web mashups for embedded devices. In *Workshop on Mashups, Enterprise Mashups and Lightweight Composition on the Web (MEM 2009)*, in proceedings of WWW (International World Wide Web Conferences), Madrid, Spain, page 15, 2009.
- [6] R. Hodgson. Voag: Vocabulary of attribution and governance (voag) schema version 2.0, 2015. [Online; accessed 2015-05-19].
- [7] R. Hodgson, P. Keller, J. Hodges, and J. Spivak. Qudt: Quantities, units, dimensions, and datatypes ontologies, 2014. [Online; accessed 2015-05-07].
- [8] S. Lu, M. Pikkarainen, S. Repo, and F. Alvarez-Cuevas Figuerola. Utilizing scada and iec 61850 for real-time mv/lv network monitoring. In *Proceedings of the 4th IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, Copenhagen, 2013.
- [9] et al M. Compton. The ssn ontology of the w3c semantic sensor network incubator group. *Journal of Web Semantics*, January 2012.
- [10] Sascha Narr, Ernesto William De Luca, and Sahin Albayrak. Extracting semantic annotations from twitter. In *Proceedings of the Fourth Workshop on Exploiting Semantic Annotations in Information Retrieval, ESAIR '11*, pages 15–16, New York, NY, USA, 2011. ACM.
- [11] C. Oerter, N. Neusel-Lange, M. Zdrallek, P. Sahm, W. Friedrich, and M. Stiegler. Experience with first smart autonomous lv-grids in germany. In *Proceedings of the 22nd CIREN Conference*, Stockholm, 2013.
- [12] S. Rohjans, M. Uslar, and H.J. Appelrath. Opc ua and cim: Semantics for the smart grid. In *Transmission and Distribution Conference and Exposition, 2010 IEEE PES*, pages 1–8, April 2010.
- [13] Amit Sheth, Christopher Thomas, and Pankaj Mehra. Continuous semantics to analyze real-time data. *IEEE Internet Computing*, 14(6):84–89, November 2010.
- [14] Yogesh Simmhan, Qunzhi Zhou, and Viktor Prasanna. Semantic information integration for smart grid applications. In JaeH. Kim and MyungJ. Lee, editors, *Green IT: Technologies and Applications*, pages 361–380. Springer Berlin Heidelberg, 2011.
- [15] A. Textor, J. Stynes, and R. Kroeger. Transformation of the common information model to owl. In F. Daniel and F.M. Facca, editors, *Current Trends in Web Engineering - 10th International Conference on Web Engineering - ICWE 2010 Workshops*, volume 6385 of *Lecture Notes in Computer Science*, pages 163–174. Springer Verlag, July 2010.