A machine-to-machine architecture to merge semantic sensor measurements

Amelie Gyrard, Christian Bonnet, Karima Boudaoud

A Machine-to-Machine Architecture to Merge Semantic Sensor Measurements

Amelie Gyrard

Supervised by: Christian Bonnet[*] and Karima Boudaoud[†]

Eurecom

Campus SOPHIATECH

450 Route des Chappes, 06410 BIOT

+ 33 (0) 4.93.00.81.92

ameliegyrard@gmail.com

ABSTRACT

The emerging field Machine-to-Machine (M2M) enables machines to communicate with each other without human intervention. Existing semantic sensor networks are domain-specific and add semantics to the context. We design a Machine-to-Machine (M2M) architecture to merge heterogeneous sensor networks and we propose to add semantics to the measured data rather than to the context. This architecture enables to: (1) get sensor measurements, (2) enrich sensor measurements with semantic web technologies, domain ontologies and the Link Open Data, and (3) reason on these semantic measurements with semantic tools, machine learning algorithms and recommender systems to provide promising applications.

Categories and Subject Descriptors

C.2.0 [**Computer-Communication Networks**]: General—*data communications*; C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*distributed networks, network communications, wireless communication*; C.2.4 [**Computer-Communication Networks**]: Distributed Systems—*distributed applications, distributed databases*

Keywords

Machine-to-Machine (M2M); M2M gateways; M2M applications; Semantic Sensor Networks; Semantic Web technologies; Resource Description Framework (RDF); Ontology

1. PROBLEM

Machine-to-Machine (M2M) applications are more and more popular due to the availability of smart M2M devices (sensors). M2M devices are used in a great deal of realms such as home monitoring [19], vehicular networks [3], environmental monitoring (weather forecasting), health monitoring (pacemaker, blood glucose level, heartbeat, brain waves). M2M area networks (sensor networks) gather heterogeneous data from M2M devices. Each application focuses on a specific M2M area network. We would like to link these existing heterogeneous M2M area networks to provide promising M2M applications (see the Figure 1).

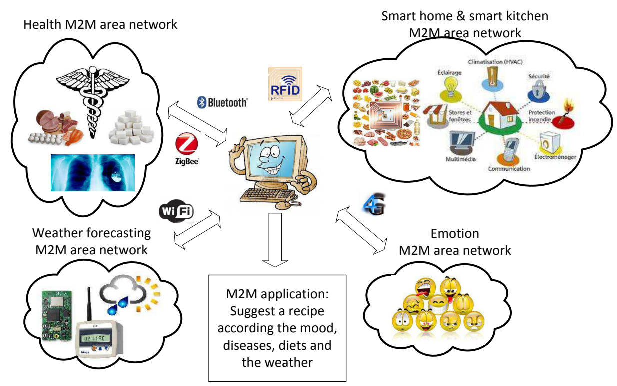

Figure 1: Merge heterogeneous M2M area networks.

We assume four M2M area networks: the smart kitchen, the weather forecasting, brain waves and the health. The weather forecasting is composed of thermometer, barometer and atmospheric pressure devices. The health M2M area network is composed of pacemaker, thermometer, blood glucose/cholesterol level and blood pressure devices. RFID tags are already found on CDS, DVDs; due to the Moore's law[1], RFID tags or EPC (Electronic Product Code) will be on each products including food. We propose to merge these M2M area networks to provide M2M applications such as suggest a recipe according to food available in the kitchen, adapted to the weather and tailored to diseases/diets/allergies or the mood of the person. Such applications do not exist yet because merging M2M area networks is a difficult task due to:

- Various protocols used (Zigbee, Bluetooth, CoAP, RFID, 3G, 4G, Wi-Fi, etc.)

[*]Eurecom. Campus SOPHIATECH, 450 Route des Chappes, 06410 BIOT. + 33 (0) 4.93.00.81.08. christian.bonnet@eurecom.fr

[†]Laboratoire I3S-CNRS/UNSA. 930 Route des Colles, BP 145, 06903 Sophia Antipolis Cedex. + 33 (0) 4.93.95.51.87. karima@polytech.unice.fr

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

[1]http://en.wikipedia.org/wiki/Moore%27s_law

- Heterogeneous data formats.
- The lack of description of measurements.

Most of the time, the semantics of data is implicit. For more advanced scenarios, combining data from various M2M area networks, there is a need for explicit semantic descriptions of the data. For example, there are temperature measurements both in the health and the weather forecasting M2M area networks, with a body temperature we deduce if the person is sick whereas with an outside temperature we deduce the season.

The main challenges of this motivating scenario are: (1) manage heterogeneous data from M2M area networks, (2) convert sensor measurements into semantic measurements using semantic web technologies and (4) reason on these semantic data.

2. STATE OF THE ART

The emerging field Machine-to-Machine (M2M) enables machines to communicate with each other without human intervention. The ETSI M2M architecture [4] is split into several main components: (1) M2M area networks are composed of: M2M devices (sensors, RFID tags or EPC) and M2M gateways aggregating heterogeneous data from M2M devices, (2) M2M applications are accessible through RESTful web services and (3) M2M service capabilities enable users through devices (PC, laptop, mobile phone, PDA) to access to applications. Starsinic [16] proposes a home M2M network which focuses on protocols and communications supported by M2M gateways but not mentions how to manage data. Lee et al. [11] review key enabling technologies in M2M architecture such as SWE (Sensor Web Enablement) and semantic web technologies (RDF, OWL). However, they do not indicate how these standards and technologies are integrated in the M2M architecture. Chen et al. [7] introduce some M2M networks like the fridge and the supermarket, but do not mention the need of semantics.

SWE (Sensor Web Enablement) [5] [6] or SenML [1] are used to retrieve sensor measurements. SWE propose services to interact with sensors: to be alerted when a specific event occurs and be notified by email. SWE is difficult to set up, configure and deploy. SenML is a lightweight protocol to get simple measurements but does not provide services such as proposed by SWE. Sheth et al. [15] have defined the Semantic Sensor Web concept to annotate sensor measurements, they use the SWE standards to retrieve sensor measurements and focus on weather data. Patni et al. [13] use SWE to convert measured data into RDF data, and publish them on the Web. Zafeiropoulos et al. [18] propose a large-scale architecture using both SWE and semantic web tecnologies. Le Phuoc et al. [14] propose a user friendly interface and manage environmental semantic sensor networks. SemSOS [10] adds semantics essentially to weather and environmental measurements. Coyle et al. [9] add semantics to smart home. For example, they control the heating system in order that is does not fall below freezing in winter.

Existing works focus on a specific sensor network: smart home, smart kitchen, weather forecasting or environmental monitoring. They design a domain ontology without be linked to the existing ones and add semantics to the context.

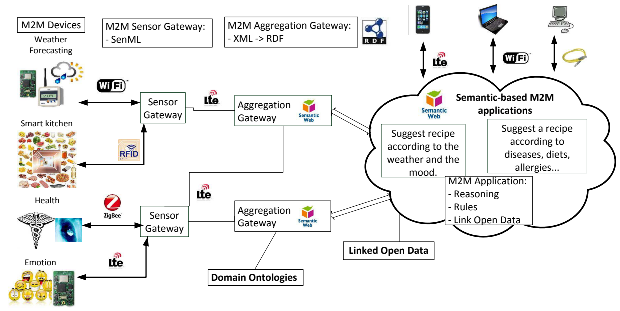

Figure 2: Our M2M architecture overview.

3. PROPOSED APPROACH

We present in this section an architecture to merge heterogeneous sensor networks, an ontology to convert automatically the sensed measurements into semantic measurements and a generic algorithm to reason on these semantic data. Existing works are domain-specific and add semantics to the context.

3.1 Architecture

We design an architecture to merge heterogeneous sensor networks and convert them into semantic sensor networks. We plan to add semantics to the sensed measurements and reason on these semantics data. To handle heterogeneous sensors, protocols and formats, our architecture is inspired by the M2M architecture. In our M2M-based architecture (see the Figure 2), we integrate semantic web technologies both in the M2M gateways and the M2M applications. We propose two kinds of M2M gateways due to various treatments:

- The **M2M sensor gateways**, use a protocol (senML or SWE) to retrieve heterogeneous data from M2M devices and convert them into XML to provide an interoperable measurement. The sensor gateways forward these XML measurements to the aggregation gateways.
- The **M2M aggregation gateways** add semantics to XML sensor measurements. It is achieved using semantic web languages (RDF, RDFS, OWL) and domain ontologies.

Sophisticated semantic treatments are performed in M2M applications through semantic-based reasoning tools (machine learning, recommender system).

In our scenario, we have two aggregation gateways, the former manages semantic data related to food and the weather forecasting, the latter stores semantic data related to health and the brain waves to detect emotions. The M2M applications queries both aggregation gateways and provides sophisticated web services such as suggesting the menu for dinner adapted to the weather, the season, the user's health, the mood and according to food available in the kitchen. We will describe in more details the M2M gateways and the M2M applications in the following subsections.

3.1.1 *Heterogeneous sensor measurements*

The sensor gateways retrieve sensor measurements and include the acquisition interface to support heterogeneous protocols such as RFID, Bluetooth, 6LowPan, CoAP and Zigbee. Several protocols can be used such as SenML or SWE to get measurements. We use the lightweight senML protocol to retrieve heterogeneous sensor measurements from M2M devices for a first and quick implementation. SenML provides simple sensor measurements: the name, the value, the unit and the date (e.g., 5 °C). SenML or SWE bridges the gap of interoperability of heterogeneous sensor data but does not provide descriptions such as the temperature measurement corresponds to the body temperature or the milk is produced by cows and contains lactose. For these reasons, we propose to enrich sensor data with semantic web technologies.

3.1.2 Add semantics to measurements

There is a need to add an explicit description on measurements: a temperature provided by heterogeneous sensors can be a body temperature, an outside temperature or an inside temperature. A RFID tag on a bottle of milk provide few information (the name, the expiry date), we want additional information (e.g., the milk contains lactose). We integrate semantic technologies to sensor measurements: RDF[2], RDFS (RDF Schema)[3], OWL (Ontology Web Language)[4] and domain ontologies to convert sensor measurements into semantic measurements. RDF is based on triplets, a triplet is like a sentence (subject, verb and complement) called "subject, predicate, object". For example, "Amelie likes milk". "Amelie" is the subject, "likes" the predicate, and "milk" the object. We can describe a large number of triplets such as: the milk is produce by cows, it contains lactose, Alice is allergic to lactose, etc. RDFS provides the property "label" to describe measured data in various language and the property to define hierarchies (e.g., milk is an ingredient and a dairy product). OWL is a language to create your own ontology: the descriptions of concepts and the relationships between them (e.g., ingredients are used in recipe, persons are allergic to ingredients).

3.1.3 Reasoning on measurements

M2M applications aim at reasoning on heterogeneous semantic measurements. An example is to suggest recipes according to the mood, diets, diseases, ingredients available in the kitchen, according to the season, etc. This example shows that four sensor networks have been merged: health, smart kitchen, weather forecasting and emotion sensor networks. Existing semantic-based machine learning and recommender systems are domain-specific, we intent to provide a generic algorithm to reason on heterogenous semantic measurements.

3.1.4 Distributed and cloud computing

We have in mind a distributed and mesh-based architecture for interesting features such as resilience, heterogeneous communications, flexibility, self-maintenance, scalability and low cost deployment. Another aspect is to propose treatments of gateways or applications on the cloud computing when they are high energy consuming.

[2]http://www.w3.org/TR/REC-rdf-syntax/
[3]http://www.w3.org/TR/rdf-schema/
[4]http://www.w3.org/TR/owl-ref/

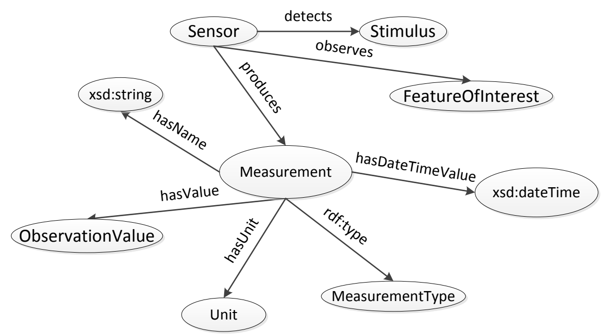

Figure 3: The SenMESO ontology.

3.2 An Ontology to Link Domain Ontologies

The SenMESO ontology (Sensor Measurements Ontology) depicted in the Figure 3 acts as a hub to merge heterogeneous domain ontologies. This ontology defines that sensors produce measurements (a name, a value, a unit, and a date) and observes a feature of interest. We classify heterogeneous sensor measurements (health, food, temperature, etc.) and link similar concepts to numerous domain ontologies to obtain additional information:

- Sensor ontologies: SSN (Semantic Sensor Network) ontology [8] for the observation value concept. CSIRO and SWEET to describe numerous units.
- Environmental ontologies. The senMESO ontology defines several kind of sensors (e.g., temperature, atmospheric pressure, radiation, precipitation sensors). By linking the temperature sensor concept to the meteo ontology, we obtain all kind of temperature sensors (Thermistor, ChilledMirrorSystem, etc.).
- Smart home ontologies. Our ontology linked the room concept to the one defined in the dogont ontology, to obtain descriptions of numerous rooms such as the kitchen, the bedroom, etc.
- Health ontologies (hospital, ontoreachir) to define health measurements (blood glucose level, heartbeat, etc.) and describe that a patient has diseases, allergies (e.g. diabetes) or diets.
- Emotion ontologies to deduce from M2M devices the mood or the emotion of the person (joy, fear, etc.).

We are enriching this ontology with other domain ontologies (military, pollution, agriculture, etc.). We intent to create a tool to link automatically our ontology to other domain ontologies.

3.3 A Generic Algorithm to Reason on Semantic Measurements

We intent to integrate semantic-based reasoning tools to reason on heterogeneous semantic measurements: machine learning algorithms and recommender systems. For example, we retrieve temperature measurements from three temperature sensor: the body temperature, the outside temperature, and the temperature inside. We want to deduce by using the machine learning algorithms that the sensor is a body temperature because it always send the (37 °C) measurement and deduce when the person is sick (39 °C). We will integrate a knowledge-based recommender system (constraint-based) to adapt M2M devices according to user's profile (e.g., women prefer the heating higher than men).

Semantic tools are used to query, link or manage semantic measurements: the SPARQL language, the Silk platform, the Jena framework, the reasoner Pellet, the rule language SWRL and the Linked Open Data. To obtain additional information, we link our semantic measurements to the Linked Open Data (existing semantic datasets). The M2M aggregation gateway stores semantic measurements (the cholesterol, butter in the fridge). We need additional information (e.g. the butter is a fat ingredient, semantic-based recipes). To achieve these requirements, the M2M application links our sensed measurements to the Linked Open Data [2]. The Linked Recipe Schema[5] describes concepts such as foods, ingredients, recipes, diets, menus. The wine ontology[6] provides the flavor of wines (dry, sweet), the color (red, white or pink), region where they are produced and describes that sweet wine are suitable for dessert (e.g., Sauternes). The SmartProducts project[7] for recipes, foods. Geonames[8] for describing locations (country, cities). SWEET[9] (Semantic Web for Earth and Environmental Terminoly) that describes 6000 concepts in 200 separate ontologies like time concepts (Evening, Sunrise, Dawn), seasons. FOAF[10] for describing persons (name, interests). The medical domain is a great example of ontologies which will provide descriptions about patients, their allergies (peanut, lactose, gluten) and their diets (vegetarian, vegan).

The Silk platform [17] can be used to perform and improve the matching of our sensed measurements to the Linked Open Data. We also aim to construct SPARQL endpoints to access remotely to the semantic sensed data.

4. METHODOLOGY

We designed the architecture at the beginning of the thesis and an ontology to convert heterogeneous measurements into semantic data. We evaluate our ontology by using it in the prototype implementation.

Current steps are to work on the refinement of this architecture and the ontology. We are working on updating automatically this ontology with new domain ontologies. We are implementing a prototype to evaluate the components of our architecture (sensor gateway, aggregation gateway) and the M2M applications.

Future steps are to integrate semantic-based recommender system on semantic measurements to propose applications as presented in the problem description. Our prototype will be integrated to the Com4Innov platform deploying a real M2M architecture. Finally, we will evaluate the performance of the prototype and the real architecture; more precisely, algorithms implemented to aggregate, convert sensed data and reason on them.

5. RESULTS

At the end of this thesis, we expect to have a generic algorithm to merge heterogeneous semantic sensor networks to propose promising applications.

5.1 A First Prototype to Validate the Architecture

We propose a first prototype to validate each component (sensor gateway, aggregation gateway) and the M2M applications of the architecture. The simulation of this components are available online[11].

We are implementing the SenMESO ontology with RDF, RDFS and OWL languages. This ontology acts as a hub to merge domain ontologies and heterogeneous semantic sensor networks. We are developing the prototype in Java and use the following technologies: Google Application Engine (GAE), RESTful Web services, the Jena framework [12] to manage semantic data and the SPARQL language[12] to perform the queries. Jena supports the reasoning tool Pellet[13] and SWRL (Semantic Web Rule Language)[14] a semantic language for describing rules (if the temperature is under 10 °C, we are in winter). The user interface is implemented with HTML5, CSS3, JavaScript and AJAX technologies.

The **sensor gateway** simulate sensor measurements according to the SenML protocol has been implemented by Eurecom and is accessible online via the XML/JSON API[15]. We simulate measurements for heterogeneous area networks: health measurements, food measurements, environmental measurements (humidity, temperature), home building measurements (light, temperature), location measurements (longitude, latitude).

The implemented **aggregation gateway** converts sensor measurements into semantic measurements using the senMeso ontology. The conversion is achieved by the JAXB API[16] transforming XML data into Java object, and then to convert Java object into RDF data. For the aggregation gateway, we intent to convert JSON sensed measurements to semantic data by using the JACKSON API[17] that enables to convert Java object to/from JSON.

The **M2M application** links our semantic measurements to existing semantic datasets by comparing the label (e.g., milk, macadamia, etc.). We propose a M2M application to merge the health and the smart home area networks to suggest a recipe according to food available in the kitchen and adapted to diseases (diabetes, hypertension) or diets (cholesterol). Another M2M application merges weather forecasting measurements with the food measurements to suggest a recipe according to the weather and the season. We intent to create other kind of M2M applications: (1) use location measurements to suggest the nearest restaurant according to user's tastes and (2) use magnetic fields measurements to alerts persons wearing a pacemaker.

[5]http://linkedrecipes.org/schema#
[6]http://www.w3.org/TR/owl-guide/wine.rdf
[7]http://projects.kmi.open.ac.uk/smartproducts/ontology.html
[8]http://www.geonames.org/ontology/
[9]http://sweet.jpl.nasa.gov/ontology/
[10]http://www.foaf-project.org/
[11]http://sensormeasurement.appspot.com/
[12]http://www.w3.org/TR/rdf-sparql-query/
[13]http://clarkparsia.com/pellet
[14]http://www.w3.org/Submission/SWRL/
[15]http://goo.gl/OIeMF
[16]http://jaxb.java.net/
[17]jacksonapi

5.2 A Real Implementation for the Com4Innov Project

The final version of the prototype will be integrated into the Com4Innov project[18]: a decentralized architecture for M2M sensors network in an LTE (4G) environment to test with real sensors and heterogeneous communications.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a M2M-based architecture to merge heterogeneous sensor networks and we enriched sensor measurements with semantics to provide promising applications. To the best of our knowledge, our work is the first M2M architecture integrating semantics to the measurements.

Currently, we are working on the refinement of the architecture, the distributed aspect and the implementation. Furthermore, we are interesting in the security aspects of this architecture. We are developing a semantic-based application available online[19] helping the non-expert security developers to secure an application by suggesting the right cryptographic algorithms, security protocols, etc. We have already created our own upper security ontology which defines relationships between attacks, countermeasures (cryptographic concepts, security protocols, security tools) and security properties (i.e., authentication). We also classify attacks and countermeasures by domain: web applications, sensor networks, cellular networks (2G, 3G, 4G), wireless networks (Wi-Fi) and network management. This application will be used to secure our architecture.

7. ACKNOWLEDGMENTS

Authors would like thanks the Wimmics Team (F. Gandon), the Eurecom's Semantic Web Team (R. Troncy), colleagues (I. Dalmasso) and friends for fruitful discussions and help for the implementation. This work is supported by the Com4Innov Platform of Pole SCS[20].

8. REFERENCES

[1] J. Arkko. Network working group c. jennings internet-draft cisco intended status: Standards track z. shelby expires: January 18, 2013 sensinode. 2012. Media Type for Sensor Markup Language (SENML), draft-jennings-senml-09 (work in progress).

[2] C. Bizer, T. Heath, and T. Berners-Lee. Linked data-the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3):1–22, 2009. http://www.w3.org/DesignIssues/LinkedData.html.

[3] M. Booysen, J. Gilmore, S. Zeadally, and G. Van Rooyen. Machine-to-machine (m2m) communications in vehicular networks. In *Article*. Korea Society of Internet Information (KSII), 2012.

[4] D. Boswarthick. M2m activities in etsi. *Presentation Report*, 2009.

[5] M. Botts, G. Percivall, C. Reed, and J. Davidson. Ogc® sensor web enablement: Overview and high level architecture. *GeoSensor networks*, pages 175–190, 2008. http://www.opengeospatial.org/projects/groups/sensorweb.

[6] A. Bröring, J. Echterhoff, S. Jirka, I. Simonis, T. Everding, C. Stasch, S. Liang, and R. Lemmens. New generation sensor web enablement. *Sensors*, 11(3):2652–2699, 2011.

[7] M. Chen, J. Wan, and F. Li. Machine-to-machine communications: Architectures, standards and applications. 2012.

[8] M. Compton, P. Barnaghi, L. Bermudez, R. Garcia-Castro, O. Corcho, S. Cox, J. Graybeal, M. Hauswirth, C. Henson, A. Herzog, et al. The ssn ontology of the w3c semantic sensor network incubator group. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2012. http://www.w3.org/2005/Incubator/ssn/ssnx/ssn.

[9] L. Coyle, S. Neely, G. Stevenson, M. Sullivan, S. Dobson, and P. Nixon. Sensor fusion-based middleware for smart homes. *International Journal of Assistive Robotics and Mechatronics*, 8(2):53–60, 2007.

[10] C. Henson, J. Pschorr, A. Sheth, and K. Thirunarayan. Semsos: Semantic sensor observation service. In *Collaborative Technologies and Systems, 2009. CTS'09. International Symposium on*, pages 44–53. IEEE, 2009.

[11] U. Lee, H. Lee, B. C. Jung, and J. Song. Emerging ubiquitous knowledge services: From mobile sensing to ubiquitous crowdsourcing and beyond. In *Knowledge service engineering handbook*. CRC Press, 2012.

[12] B. McBride. Jena: A semantic web toolkit. *Internet Computing, IEEE*, 6(6):55–59, 2002.

[13] H. Patni, C. Henson, and A. Sheth. Linked sensor data. In *Collaborative Technologies and Systems (CTS), 2010 International Symposium on*, pages 362–370. IEEE, 2010.

[14] D. Phuoc and M. Hauswirth. Linked open data in sensor data mashups. 2009.

[15] A. Sheth, C. Henson, and S. Sahoo. Semantic sensor web. *Internet Computing, IEEE*, 12(4):78–83, 2008.

[16] M. Starsinic. System architecture challenges in the home m2m network. In *Applications and Technology Conference (LISAT), 2010 Long Island Systems*, pages 1–7. IEEE, 2010.

[17] J. Volz, C. Bizer, M. Gaedke, and G. Kobilarov. Discovering and maintaining links on the web of data. *The Semantic Web-ISWC 2009*, pages 650–665, 2009. http://www4.wiwiss.fu-berlin.de/bizer/silk/.

[18] A. Zafeiropoulos, N. Konstantinou, S. Arkoulis, D. Spanos, and N. Mitrou. A semantic-based architecture for sensor data fusion. In *Mobile Ubiquitous Computing, Systems, Services and Technologies, 2008. UBICOMM'08. The Second International Conference on*, pages 116–121. IEEE, 2008.

[19] Y. Zhang, R. Yu, S. Xie, W. Yao, Y. Xiao, and M. Guizani. Home m2m networks: Architectures, standards, and qos improvement. *Communications Magazine, IEEE*, 49(4):44–52, 2011.

[18] http://www.com4innov.com/platforms_presentation.en.htm
[19] http://securitytoolbox.appspot.com/
[20] http://www.pole-scs.org/