Enabling Semantic Web for Precision Agriculture: a Showcase of the Project agriOpenLink

Dana Tomic
FTW
Donau-City Strasse 1
Vienna, Austria
tomic@ieee.org

Domagoj Drenjanac
FTW
Donau-City Strasse 1
Vienna, Austria
drenjanac@ftw.at

Wilifred Wöber
Universität für Bodenkultur
Gregor-Mendel-Strasse 33
Vienna, Austria
wilfried.woeber@boku.ac.at

Sandra Hörmann
Josephinum Research
Rottenhauser Straße 1
Wieselburg, Austria
sandra.hoermann@josephinum.at

Wolfgang Auer
MKW electronics GmbH
Jutogasse 3
Weibern, Austria
wolfgang.auer@mkwe.com

ABSTRACT

This paper describes the agriOpenLink approach towards triplification of the production data generated by heterogeneous agricultural equipment, for their easy integration and querying within a common information context of various decision support applications. The presented approach has been developed within the running project agriOpenLink, which aims at improving agricultural production processes. Triplification is performed by equipment-specific model adapters (plugins), which translate the equipment data into RDF triples. The plugins are realized by means of semantic REST services, and triplification is a result of a workflow orchestrated "on-demand" by chaining services of appropriate data sources. In the project, we have also conceptualized, and are currently implementing a platform and tools that facilitate easy plugin development and deployment. Dealing with the problem of creating a common information context, the project designed a domain ontology taking into account existing work, experimented with triplification of existing domain knowledge, and proposed an approach for the ontology maintenance. The agriOpenLink platform is being tested and demonstrated in the precision irrigation use case and the precision dairy farming use case.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Design, Experimentation

Keywords

Triplification, ontology, semantic services

1. INTRODUCTION

1.1 Precision Farming

Precision agriculture stands for management practices and tools that leverage advanced sensor, actuator and decision support technology to aid in optimization of the production processes [1][2].

The market for the precision agriculture equipment, both for the arable farming and the livestock farming is rapidly growing, and the robotic and sensor systems are already indispensable at larger farms. These are expected to be ubiquitously used in the years to come. Today, modern agricultural equipment such as autonomously driving tractors with their smart implements, milking and feeding robots or animal monitoring systems, aggregate many different types of sensors and collect many different sensor data. In many cases, these data are stored in local databases, e.g., imbedded within robots or accompanying controllers (PCs) that provide user applications. The raw sensor data are often processed locally so as to provide users with a meaningful aggregated information for the decision support. Such interpreted data is often presented to users in a tabular of graphical form. Cloud-based solutions are being increasingly offered, however the issues of data ownership or reuse by 3rd parties are still subject of many controversies hampering the take-off of these solutions.

Nevertheless, the equipment is just one source of the production data. The farmers need to interact with a multitude of other external information systems with the weather data, soil, weed, crops, pest, animal breeds, feed, and other information. Although the integration of data from different sources is a paramount enabler for the decision support applications and for the production process optimization, it still remains a grand challenge. Today it is often a farmer who has to collect data from different sources e.g., in form of comma separated value (CSV) files, and link them, and process them together. This is particularly noticeable in the domain of precision livestock farming where proprietary data models are often used, partially because of a low acceptance of existing standards (e.g., ISOagriNET framework, www.isoagrinet.org) and their often criticized inflexibility. As an answer to strong user needs, emerging data *integration* solutions [3] typically also fall back to using proprietary data models, without attempting to solve the problem of data model interoperability. These approaches focus more on providing a single integrated user interface for the farmers, and less on establishing open interfaces for any new system that can be meaningfully connected. Therefore they often lead to non-scalable solutions sensitive to changes in data formats that are used on the interfaces of heterogeneous integrated sources.

1.2 Semantic Web for Precision Farming

The goal of Semantic Web is to use Web as an information management platform [4]. The standards of the Semantic Web enable unified modelling of data and metadata, so that they could be incrementally integrated within transparent knowledge structures within which concepts can be disambiguated and interrelated, and data easily linked to their defining schemas. The standards, tools and practices of Semantic Web have reached considerable maturity, and are used for the management of both open data (freely accessible) and enterprise data (accessible to registered users) in many different sectors and applications [5][6]. These tools support the whole lifecycle of creating and maintaining linked data, which includes 1) extraction of structured data from other sources, 2) data authoring or creation via “triplification”, enrichment, interlinking and fusing and 3) data maintenance in a repository. As a result of the Linked Data community effort the Open Linked Data Cloud today include huge number of data sets (ontologies, vocabularies, … and real data) with geographic information, publications, user generated content, government, life sciences, and cross-domain. The enterprise systems can re-use established semantic descriptions and uniquely defined resources and benefit from the linked data concepts and the shared semantics.

The languages and tools for translation of data from different formats into RDF are important components of existing Semantic Web tool chains. Several approaches have already been proposed, demonstrated and offered as standards. For the translation of relational databases the W3C defined R2RML [7]. The RML approach [8] defines a language that extends the R2RML standard particularly addressing concurrent translation of many related sources. These approaches are very well suited to be used by owners of large databases that have already accumulated huge amounts of data.

Within the agricultural domain the Semantic Web approach already inspired a number of solutions in the research spectrum, e.g., [9-13] and is also embraced by the Food and Agriculture Organization of the United Nations (FAO; http://aims.fao.org) in their global initiatives for agricultural information management systems (AIMS), AGROVOC vocabulary, and agricultural ontology service [14]. Also in the agricultural domain there are many different sources of information, and actors who currently maintain these data for the farmers and are opening them via APIs, e.g. as described in [15], or can be found on [16].

2. THE AGRIOPENLINK APPROACH

2.1 The Motivation

Motivated with the shortcomings of the existing data integration approaches in commercial precision farming solutions and the promises of the Semantic Web, the project agriOpenLink[1] developed a semantic approach with a goal to establish a common context for the integration of heterogeneous interfaces and data, and to develop tools to facilitate in better decision making and process optimization. This approach integrates three tasks: 1) establishing a platform in which a common ontology-based model can be created and maintained to reflect the common data integration possibilities and needs, 2) designing tools to ease the development of the model adapters – the equipment-specific or data-source-specific triplifiers – that translates process data into RDF triples based on the concepts of the common ontology, and 3) designing a run time environment in which data translation (tripligication) is performed on demand as a result of execution of queries. As these RDF data relate to the common ontology (the knowledge context) they can be easily integrated and queried together.

While agreeing on a common model (a defacto standard) is a prominent requirements in the agricultural sector, it is not possible to aspire model completeness. Therefore, the model extension or change has to be technologically easy to implement and propagate. This requirement makes semantic technology the technology of choice for creating and maintaining a standard model. Ontology-based domain knowledge is easy to extend and interlink with already established concepts. That is why agriOpenLink focus on tools for the ontology-based lifecycle management of the schema and data. Accordingly adding a new data source or changing formats does not result in a detrimental complexity or a high implementation cost.

The agriOpenLink approach also aims to account for specificities of data sources in agricultural production environment and special requirements on structuring them. While the applicability of the relational database schema translation approaches is high for the owners of large databases, this does not apply for the equipment as a source of production data, because data stored within the equipment database first needs to be processed (e.g., averaged within a specific interval, etc.) to be useful for sharing. Therefore, agriOpenLink uses plugin approach and application specific translation of data into RDF by means of semantic REST services.

2.2 The Realization

2.2.1 Plugins as Triplifiers

The agriOpenLink platform offers tools and procedures to translate the agricultural data from the precision agriculture equipment or other agricultural information sources into their semantic RDF-based form, in which they can be interlinked and jointly queried. We adopted a plugin-based approach where each source of a relevant information is wrapped within a so called plugin acting as a model adapter - a triplefier. The triplification is realized as service-based business logic. More specifically a plugin is a software component which publishes semantic REST services, similar to the SADI Web Services [17] that consume and produce RDF triples. A plugin acts as a model mediator – it translates data from the internal model of the device into instances of particular class of the ontology decorated with specific properties. Accordingly its function is to open and provide for further linking RDF data from the devices. An important component of our plugin approach is the development

[1] www.agriopenlink.com

environment which offers plugin skeleton code and basic functionality to aid in the plugin implementation.

The decision to create a specific plugin platform and not use existing Internet of things (IOT) frameworks such as [18], and to encapsulate translation of the data into RDF within semantic services is based on specific requirements and current constraints existing within the agriculture sector. First of all, we aimed at a compact solution with a minimum set of functionality. The goal is to bring the benefits of the semantics and ontology to the involved actors (equipment vendors/ 3rd party application providers), but hide as much as possible the complexity of the semantic representation. To this aim the ontology is maintained via the platform, and the plugin development environment automates some of the complexity of semantic data processing and semantic REST service creation, for the plugin developers. Accordingly, agriOpenLink aims at scalable integration of any data interface and data format within a common information and knowledge context in which data can easily be related to and combined with other data.

While many agricultural equipment use relational databases to organize its internal data, our triplification approach is not based on the approaches and tools for data model translation such as [7][8]. The reasons for this are twofold. Firstly, the raw data generated by the equipment is very often quite sensitive and must first be processed into interpreted data that can be offered to farmers. The raw data is often considered as being owned by the equipment vendors. Accordingly, the amount of interpreted information is lower as compared to the raw data. The information offered to the farmer is sometimes not in the database, but is calculated and only presented on the user interface. Secondly, the engineers implementing the internal logic of this equipment, and accordingly the plugins, do not have to learn semantic-heavy tools but can program the business logic of translation completing the plugin skeleton.

The project agriOpenLink is demonstrating the use of the platform in two scenarios: (1) in the precision irrigation scenario we focus on triplifying weather, soil and sensor data and expressing the transpiration model knowledge in form of semantic services [19]. In the precision dairy farming scenario we triplify data from the farm robots and farm systems and experiment with expert knowledge for animal state diagnosis expressed as SPARQL queries [20]. Figure 1 illustrates the platform architecture in a precision dairy farming. A plugin server that implements HTTP REST interfaces is a central component of the local system that initiates and run different plugins on the farm. The agriOpenLink platform also includes a plugin development environment in which plugins can be designed and built into dynamically loadable components.

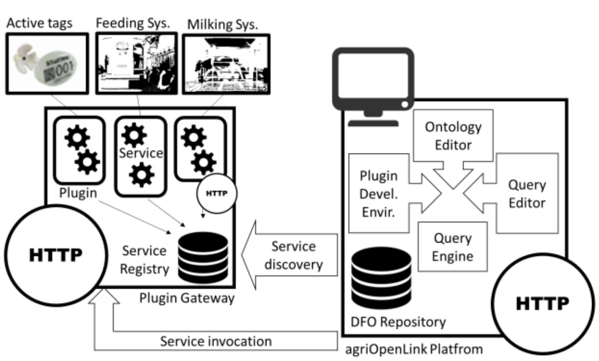

Figure 1. The agriOpenLink Platform (dairy farming scenario)

We experimented with the ways of encoding the expert knowledge in form of defined classes, SPARQL queries and semantic services. To this aim the Ontology Editor offers a user interface to create defined classes with restrictions on properties as a basis for classification. An example is a class that restricts the activity property to define an animal which may be potentially lame. The Query Editor offers a user interface for the creation of SPARQL queries. We are currently implementing a solution with which any query can be also translated into 1) a dynamically deployed service (which controls invocation of other services) and 2) an API which can be programmatically included in any 3rd party application that wants to start such query.

2.2.2 The Domain Ontology

Particularly the creation of the domain ontology for the dairy farming scenario was challenging, due to not much of existing work in this domain. In the process of DFO engineering for the domains of milk quality, feeding, breeding, fertility, we analyzed both the ISOagriNET standard dictionaries [16] and the schemas of the milking robots, concentrate feeder, and heat and activity monitoring equipment. We created the first version of the ontology in the OWL ontology editor Protégé. It addresses the scenario in which a linked data set is created for a herd at the farm as a basis for herd management applications. The herd linked data set includes the farm core data, i.e., the animal registration information, and is enriched with the data continuously coming from different farm equipment and external sources, including the milk quality, feed, activity, fertility, and health information. We started with a small set of OWL classes including: Animal, Farm, Farmer, Equipment, Organization and the object properties to reflect the relationships between instances of these classes such as parent-child relationship between Animals, similar to the ontologies [11][21]. While these existing ontologies provided an important input, only the latter one has been recently published.

During the ontology creation it became clear that standard ISOagriNET dictionaries need to be available in the linked data format in order to reference them in a formal way. Consequently, we defined their "triplification" as a goal for the 2nd phase of our DFO engineering task. In finalizing the 1st phase we revisited the requirements of our initial herd linked data scenario and re-focused the DFO engineering on the animal state diagnoses scenario, in which WC3 SPARQL queries coming from Decision Support Systems trigger search, access, interlinking and filtering of semantic data available via Plugin Services. We reduced our DFO to the properties that belong to a core data set, and that can be extracted from equipment and external sources. DFO further includes defined classes that are used by Plugin Services. They specify what properties can be extracted from each specific Plugin, i.e., specific type of the equipment.

2.2.3 Translating ISOagriNET Data Dictionaries

To obtain an RDF/OWL model of the standard dictionaries for livestock farming we developed a triplifyer tool that follows the best practices for creating the linked data [6]. Our triplifier tool can also transform standard exchange data files into linked data. The triplification of the standard schema was described in [22].

To test the procedure we triplified all dictionaries available via http://ian.lkv-nrw.de as well as exemplary data files. The translated dictionary for the year 2003 is available at www.agriopenlink.com/ADR2003. We are currently working towards completing and opening the ontology platform with ontology browsing and editing GUI and a SPARQL interface.

3. CONCLUSIONS

The presented agriOpenLink platform is a work-in-progress solution that was designed by strongly focusing on data created within the agricultural production environment, on the data formats that are currently used in these environments, and on the benefits of translating production data into RDF for their easy integration and integrated querying.

Today farmers need to use data from many different data sources and to integrate them in a meaningful way. Very often relevant data sources, such as agricultural equipment, do not offer open data interfaces, so farmers need to either manually verify data or dump data into csv files and process them in some common purpose tools. Also, some data interfaces do not comply with the existing ISOagriNET standard, and often it takes a long time to introduce new data properties pertaining to innovative new sensors and systems into this standard.

The contribution of the agriOpenLink solution is in offering an integrated approach and a platform which can support 1) maintenance of the domain knowledge (ontology) in the repository with the SPARQL and user interfaces for ontology community-based editing, 2) adding of new devices by means of plugins that publish RDF data complying to the ontology, and which can be easily created in a plugin development environment, and, 3) querying of RDF data on-demand, where the resulting data can be stored for further publishing and querying. We have implemented and demonstrated core functionalities of the platform in a demonstrator and are currently focusing on deployment on farms, as well as on completing and opening the agriOpenLink ontology maintaining platform for experimentation.

4. ACKNOWLEDGMENTS

The work presented here is partially funded by the Austrian Research Agency (FFG) within the project agriOpenLink (research fund 898398).

5. REFERENCES

[1] McBratney, A., Whelan, B., Ancev, T., 2005. Future Directions of Precision Agriculture. Precision Agriculture, 6, 7-23.

[2] Bewley, J.M. 2013. Exciting Dairy Breakthroughs: Science Fiction or Precision Dairy Farming? Proceedings of the Precision Dairy Conf., Rochester, MI, June 26-27, 2013.

[3] Elite, 2014 'Drei Programme für die Organisation Ihrer Herde', in German, Elite Magazin 6/14, 2014.

[4] W3C Semantic Web, http://www.w3.org/standards/semanticweb/, last accessed: Aug 2015.

[5] Auer S., 2014. Introduction to LOD2, in Linked Open Data – Creating Knowledge Out of Interlinked Data, 1-17, Springer, 2014.

[6] Mezaour, A-D., Van Nuffelen, B., and C. Blaschke. 2014. Building Enterprise Ready Applications Using Linked Open Data, in Linked Open Data – Creating Knowledge Out of Interlinked Data, Springer, 2014.

[7] W3C Semantic Web, R2RML: RDB to RDF Mapping Language, (2012) http://www.w3.org/TR/r2rml/, last accessed Aug 2015.

[8] Dimou, A., Vander Sande, M, Colpaert, P, Verborgh, R, Mannens, E. and R. Van de Walle. RML: A Generic Language for Integrated RDF Mappings of Heterogeneous Data. Proceedings of the 7th Workshop on Linked Data on the Web, 2014.

[9] Eckartz, S, Verhoosel, J, Folmer E and E Somers 2013. Semantics for Smart Dairy Farming: a milk production registration standard v1.0, Report, available at http://www.smartdairyfarming.nl, last accessed Aug 2015.

[10] Chaplinskyy, Y. and O. Subbotina, 2013. Ontology-driven advice and decision-making within knowledge, dissemination for extension, HAICTA 2013.

[11] Gao Yongchun, 2005. The application of Web Ontology Language for information sharing in the dairy industry, PhD Thesis, McGill Univ. Montreal. 2005.

[12] Athanasiadis IN, Rizzoli AE, Janssen S, Andersen E, Villa F. 2009. Ontology for Seamless Integration of Agricultural Data and Models. MTSR 2009; Milan, Italy, October 1-2, 2009, 282-293.

[13] Grimnes MK, Abufouda, M and A Schröder. 2012, Semantic Integration through Linked Data in the iGreen project. GIL Jahrestagung, volume 194 of LNI, page 107-110.GI,(2012)

[14] Lauser, B., Sini, M., Liang, A., Keizer, J. and S. Katz, Stephen, 2006. From AGROVOC to the Agricultural Ontology Service / Concept Server, available at http://aims.fao.org

[15] Haezebrouck, T-P, Emeric, E, Sine M, API-AGRO: An Open Data and Open API platform to promote interoperability standards for Farm Services and Ag Web Applications, EFITA 2015, June 29-July 2, Poznan, 2015.

[16] LKV, 2015 'ISOagriNet, http://ian.lkv-nrw.de/index.php, http://ian.lkv-nrw.de, http://webapp.lkv-nrw.de/AdedDataDictionary, last accessed Aug 2015

[17] Wilkinson, M.D., McCarthy, E.L., Vandervalk, B.P., Withers, D., Kawas, E.A., Samadian, 2010. S.: SADI, SHARE, and the in silico scientific method. ;BMC Bioinformatics(2010)7-7

[18] FIWARE Backend Device Management – IDAS, http://catalogue.fiware.org/enablers/backend-device-management-idas, last accessed Aug 2015

[19] Wöber, W.; Supper, G.; Aschauer, C.; Gronauer, A.; Tomic, D.; Hörmann, S., Entwicklung eines auf semantischer Technologie basierenden Analysesystems zur Überwachung der Wasserversorgung von landwirtschaftlichen Nutzflächen, GIL Tagung 2015, Available at http://www.gil-net.de/Publikationen/27_201.pdf, Last accessed June 2015

[20] Tomic SDK, Hoermann S., Handler F., Wöber W., Otte, M., and W. Auer, 2014. agriOpenLink: Semantic Services for Adaptive Processes in Livestock Farming, AgEng 2014.

[21] Seamless , 2015. http://ontologies.seamless-ip.org/livestock.owl, last accessed Aug 2015

[22] Tomic, D., Drenjanac, D., Hoermann S., Auer, W, Experience with Creating and Maintaining Ontology for Precision Dairy Farming: A Showcase of agriOpenLink, EFITA 2015, June 29-July 2, Poznan, 2015.