agroRDF as a Semantic Overlay to agroXML: a General Model for Enhancing Interoperability in Agrifood Data Standards

Daniel Martini [1], Mario Schmitz [1], Esther Mietzsch [1]
[1]Kuratorium für Technik und Bauwesen in der Landwirtschaft e. V.
Bartningstraße 49, 64289 Darmstadt, Germany
d.martini@ktbl.de

ABSTRACT

Although standardisation of data exchange formats and protocols in agriculture has made considerable progress during the last few years, achieving interoperability for information integration in IT systems in the agrifood sector is still seen as a challenge. Traditional ways of agreeing on interfaces by long and elaborate discussions in standardisation committees probably will not be able to cope anymore with demands from a more and more networked world and faster changes to requirements in the future. Novel ways are needed to enable at least to a certain degree automatic interface negotiations. Semantic technologies are one of the building blocks within such a framework. The authors present methods used, potential benefits in facilitating interoperability and experiences gained in implementing a semantic layer on top of the agroXML standard.

Keywords: Semantic web, standardisation, interoperability, web services, ontologies, germany.

1. INTRODUCTION

The amount of data available for decision support, quality and process management in the agricultural sector is increasing constantly. New devices allow for more or other parameters to be measured in remote sensing. Agricultural machinery can capture lots of different quantities during the process. Networks in stables in livestock farming enable the collection of data on temperature, humidity, feeding, water consumption, etc. Also, web pages provide information on e. g. varieties, pesticides or veterinary drugs. Nevertheless, usability of these data lagged behind expectations. Proprietary solutions delivered in combination with machinery often allowed only in a very restricted manner to deal with data from other sources and files on the internet have to be retrieved interactively and data be copied/pasted manually.

D. Martini. "agroRDF as a Semantic Overlay to agroXML: a General Model for Enhancing Interoperability in Agrifood Data Standards". EFITA-WCCA-CIGR Conference "Sustainable Agriculture through ICT Innovation", Turin, Italy, 24-27 June 2013. The authors are solely responsible for the content of this technical presentation. The technical presentation does not necessarily reflect the official position of the Internation Commission of Agricultural and Biosystems Engineering (CIGR) and of the EFITA association, and its printing and distribution does not constitute an endorsement of views which may be expressed. Technical presentations are not subject to the formal peer review process by CIGR editorial committees; therefore, they are not to be presented as refereed publications.

1.1 Standards in the Agrifood Sector

Barriers in the – for integration necessary – electronic communication in the agricultural sector, e. g. between farmer and machinery cooperative or into the food chain, have in part been attributed to a lack of IT standardisation.
And yet, analysis of use cases as e. g. conducted in the agriXchange project showed that there are lots of standards that can be used in agriculture to represent certain sets of data or information on-the-wire. The aXtool (Pesonen et al., 2012) available at http://www.agrixchange.org/page-area/axtool provides an overview on currently nine use cases with ten standards attached to them to implement certain interfaces. Twenty-five more standards have been found to be of relevance when extending the analysis to the broader scope of the food sector (Mietzsch et al., 2013).
With regard to certain standards, also impressive progress in certain subareas of the agrifood domain has been made in recent years. Compatibility of ISOBUS-capable equipment being able to log and transfer data according to the norm ISO 11783 is increasing. agroXML is on its way to enable information exchange between farm management systems and external stakeholders and services in the cloud using internet protocols (Doluschitz et al., 2005; Martini et al., 2009). Regarding spatial data, a number of standardized service interfaces like the Web Feature Service (WFS; Vretanos, 2005) and accompanying data formats like the Geography Markup Language (GML; Cox et al. 2004) are available from the Open Geospatial Consortium and are backed by supporting proprietary and open source implementations. Also, the financial sector has seen considerable standardisation efforts leading to e. g. the Universal Business Language (UBL; Bosak and McGrath, 2006) or the eXtensible Business Reporting Language (XBRL; Engel et al., 2003).
Based on differing requirements, standards available for use within the agrifood domain use distinct technologies for exchange protocols and syntactical representation of data. All of them however support only part of the domain and/or are specialized to certain processes. Therefore, overlaps in content exist and necessities for transferring data from one representation format into another come up. For these reasons, on a technical level, different existing standards are barely interoperable and most of them lack extension mechanisms flexible enough to be able to integrate arbitrary data from other, not yet foreseen process stages. Systems that communicate with a set of different stakeholders that each insist on their domain industry standard format to be used are difficult to implement as developers have to extract the meaning of data items from long and often differently structured specifications.

1.2 Semantic Technologies

Semantic technologies in computer science date back as far as into the seventies of the last century (see e. g. Simmons, 1973 in Schank and Colby, 1973). Traditionally, they have been concerned with methods to convey meaning of concepts within a language to computer programs. Recently, the term has been used in a broader sense and (re)gained increased attention by the movement on the world wide web around the so-called web of data or the semantic web (Berners-Lee, 2001). Vocabularies and ontologies are – apart from the data to be provided – at the heart of the semantic web. According to Gruber (1993) an ontology is an "explicit specification of a conceptualization" of a domain. It thus provides in a formal, machine-readable representation a list of concepts (classes, terms, objects, entities etc.) used within a domain and sets them in relation to each other. That way, it enables assigning meaning to items given in information resources like data sets or texts by providing "pre-set" associations. That again facilitates sorting out identical objects, interpreting properties, generating translation tables etc. The basic model used in today's standards for ontology definition for the semantic web (some of them described in section 2) is commonly very simple and relies on graph data structures. As a graph is a very general structure, it can be applied upon any existing data model no matter what paradigm has originally been used during initial creation – be it object-oriented, relational or hierarchical.

2. DEVELOPMENT OF THE SEMANTIC OVERLAY

By building a layer providing the formalized semantics described in 1.2 on top of existing standards it should be possible to reduce the grieve caused by the problems described in section 1.1. It is supposed to enhance interoperability in systems by providing a framework that facilitates mapping and linking between different models and that is flexible and extensible for dealing with new requirements. The authors tested that approach upon the agroXML standard. As a technological background, a short description of agroXML is given followed by the methods used for creating the semantic model.

2.1 agroXML

agroXML is an XML dialect for representing and describing farm work. It provides elements and XML data types for representing data on work processes on the farm including accompanying operating supplies like fertilizers, pesticides, crops etc. It can be used within FMISs as a file format for documentation purposes but also within web services and interfaces between the farm and external stakeholders as a means to exchange data in a structured, standardized and easy to use way.
Among its use cases are for example:

- Extensive documentation of crop growing: Demand for documentation of activities during cultivation of crops like seeding, fertilization, harvesting etc. is constantly rising. The same goes for livestock farming, where relevant events to document may include birth of animals, veterinary treatments, feeding etc. On the one hand, information like amounts of fertilizers applied, variety used, veterinary drugs handed out etc. is requested by buyers of agricultural goods like mills or slaughterhouses. On the other hand, in certain settings, information like this may be used within web platforms or quality programs to achieve a certain marketing advantage by transparency on production processes. agroXML provides the necessary elements to construct files or web services that provide the necessary information. Data can be reused for other purposes as well, while control on content handed out to third parties stays within applications and therefore, if necessary in the hands of the farmer.

- Data sharing within cooperatives: in software applications for cooperative farming or for enterprises that manage several farms at different places, a need for data sharing among different production sites or components applications may arise. agroXML allows to build a web services based peer-to-peer network to exchange data on sites, fields, crops grown, animals kept etc. Data items can be picked from the schemas and provided within a service oriented architecture either within a full-blown SOAP (Mitra and Lafon, 2007) messaging stack or using lightweight, flexible approaches like ReSTful web services (Richardson and Ruby, 2007).
- Providing data on farming products and supplies: Providers of agricultural supplies and goods or an extension service or governmental organization may want to offer an additional service to clients by delivering extended information on the things they sell or have knowledge on. agroXML enables machine-readable, electronic provision of data sheets and specification documents. The information can then be imported into management systems relieving your customers from cumbersome manual data entry.

agroXML has been developed by a team consisting of members from makers of agricultural software systems, machinery companies, service providers and research organisations. Its origins date back around the year 2000. Currently, agroXML is at version 1.5 aiming for publication of 1.6 in 2013. The latter will include the overlay semantic model described below as an integral part of the release.
Structures are defined using W3C's XML Schema (Fallside and Walmsley, 2004). Schema files of the most recent released version are currently available at http://www.agroxml.de/schema/agroxml_1_5. Essential modules with definitions concerning the farm and plant production are currently availabe, modules for livestock farming are in development. These can be used as a whole, but also independently from each other.

2.2 agroRDF

Using the resource description framework (RDF; Klyne and Carroll, 2004) and RDF Schema (RDFS; Brickley and Guha, 2004), in a first step, a complete 1:1 representation of the agroXML schemas has been created. Although this can be done, it led in part to rather complex constructs and proved to be difficult to maintain. The work was thus reorganized into a set of small, modular ontologies. They provide a semantic model of the data format by capturing in a set of statements relations between individual data elements. For formulating the statements, a small subset from the OWL (Motik et al., 2012) vocabulary – basically providing sameAs-functionality to express equivalence – have also been introduced. Figure 1 shows a strongly reduced excerpt of the RDF model for Pesticides.

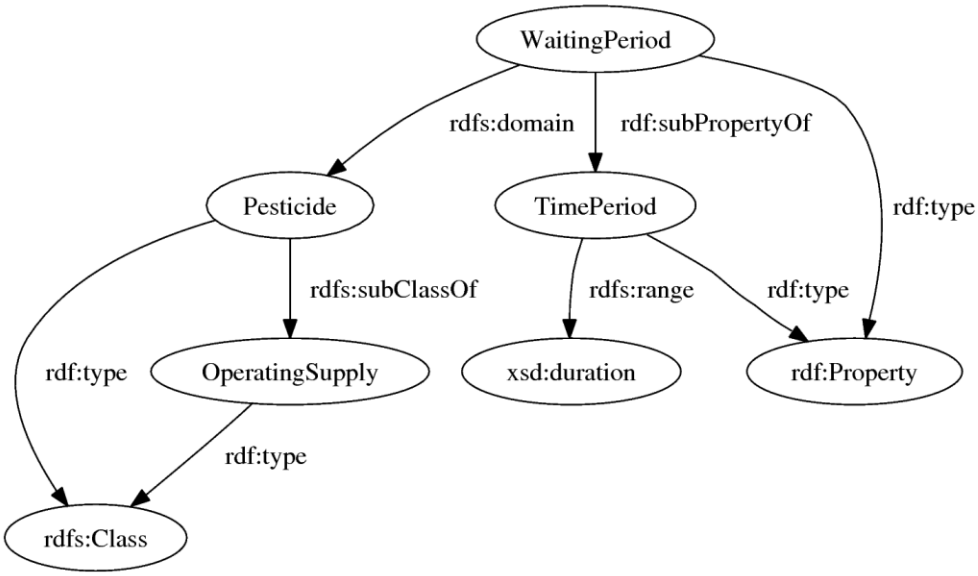

Figure 1. Reduced excerpt of the RDF model on Pesticides.

A semantic model in the context of the semantic web and focusing on interoperability issues, where data is supposed to be exchanged between anonymous partners makes most sense if it builds as much as possible on already published ontologies made available by domain experts. That way, data can easily be put into context with other data sets. During the development process, therefore certain classes and properties in agroRDF have step by step been replaced by references to existing ontologies like the QUDT ontology published by the NASA jet propulsion laboratories for quantities and dimensions and the vCard and FOAF ontologies for data on people. Further candidates for reuse include the semantic sensor network (SSN) ontology by W3C for sensor data and measurements, the Core Business Location vocabulary for organization's data and the RDF vocabulary defined in the GeoSPARQL specification for spatial data. That way, the model can be simplified significantly by leaving model details up to the external ontologies. An application can then pull in further information just as needed. An especially prominent example is the provision of human readable unit labels and conversion multipliers by the QUDT ontology offering means for human-readable presentation of quantities in GUIs and conversion between different units of measure. An important step is the establishment of links between the XML schemas providing a syntax specification for documents and data packages and the RDF model providing the formal description of semantics. This has been done in this case using additional schema attributes from the semantic annotations to WSDL and XML schema

(SAWSDL; Farrell and Lausen, 2007) recommendation. Element and data type definitions in the schema are set into relation with their counterpart in the semantic model by a reference to class/property URLs used in the RDF model. As an example of how to link against a multilingual thesaurus, a mapping into AGROVOC of the Food and Agriculture Organization of the United Nations (FAO) has been created. The AGROVOC has recently been provided using semantic web technologies as well, assigning Uniform Resource Locators (URLs) to each concept and thus allowing references from other ontologies. The mapping allows e. g. multilingual search in agroXML data sets.

3. CONCLUSIONS AND OUTLOOK

Coming back to the solution approach for the issues described in section 1, principles and mechanisms for achieving system-independent, inter-standards, cross-domain interoperability can be derived as follows from the exercise in data modeling described above:

A first step is assigning URLs to concepts within a domain. This is a natural step within development of any RDF model, the only task to be actively done by the developer is choosing an appropriate convention. Within our work, the choice was made to rely on existing human readable element names. Standards that provide a data dictionary like the ISOBUS mentioned in 1.1 may e. g. choose existing data dictionary entity numbers and build up an URL by adding a prefix. In fact, a (not yet published) RDF representation of the ISOBUS data dictionary has been worked out within the iGreen project relying on such a mechanism.

As a second step, a mapping to an available syntax specification has to be established. Data can then be used and semantically annotated "as-is" by simply parsing and attaching syntactical tokens found to a semantic description using the pre-existing mapping. Standards relying on XML and providing their specification in a schema can use the facilities of SAWSDL described above. For other serializations, similar mechanisms have to be developed.

Third and last, reuse of existing ontologies is crucial for developing references to a common ground. Step by step, linkages can be established leading to improvement by allowing developers to easily extract identical concepts from data sets represented in different standard formats or to set data into context by e. g. interpreting subclass-/superclass relationships. In the future, large networks of semantic standards descriptions can evolve leading to a constant growth of understanding.

Another benefit in the context of using XML schema together with RDF was, that the schema can be relieved from also having to carry model information. Although there are mechanisms to model superclass/subclass and other relationships like a feature-property model in XML, it usually leads to very clumsy, hard to understand schemas. By pushing that information into the semantic model layer, the schema can thus be flattened out and simplified leading to a clean separation of concerns and using the right technology in the right place.

A prototypical web service showing application of these semantic web technologies to agricultural machinery data has been created showing flexibility in extensions however covering only a single data source. Recently, work is conducted to build a set of web services for livestock farming that demonstrate integration of different data sources and standards using agroRDF as a facilitator.

4. REFERENCES

Berners-Lee, T., J. Hendler, O. Lassila. 2001. The Semantic Web. In *Scientific American Magazine*, May 2001.

Brickley, D., R. V. Guha. 2004. RDF Vocabulary Description Language 1.0: RDF Schema. World Wide Consortium. http://www.w3.org/TR/rdf-schema/.

Bosak, J., T. McGrath. 2006. Universal Business Language 2.0. Organization for the Advancement of Structured Information Standards (OASIS). http://docs.oasis-open.org/ubl/os-UBL-2.0/.

Cox, S., P. Daisey, R. Lake, C. Portele, A. Whiteside. 2004. OpenGIS Geography Markup Language (GML) Implementation Specification. Open GIS Consortium Inc.

Doluschitz, R., M. Kunisch, T. Jungbluth, C. Eider. 2005. agroXML – A Standardized Data Format for Information Flow in Agriculture. In *EFITA/WCCA 2005, Portugal, Book of Abstracts, July 2005*, pp. 26.

Engel, P., W. Hamscher, G. Shuetrim, D. vun Kannon, H. Wallis. 2003. Extensible Business Reporting Language (XBRL) 2.1. XBRL International.

Fallside, D. C., P. Walmsley. 2004. XML Schema Part 0: Primer. World Wide Web Consortium. http://www.w3.org/TR/xmlschema-0/.

Farrell, J., H. Lausen. 2007. Semantic Annotations for WSDL and XML Schema. World Wide Web Consortium. http://www.w3.org/TR/sawsdl/.

Gruber, T. R. 1993. A Translation Approach to Portable Ontology Specifications. In *Knowledge Acquisition*, 5(2). 199-220.

Klyne, G., J. J. Carroll. 2004. Resource Description Framework (RDF): Concepts and Abstract Syntax. World Wide Web Consortium. http://www.w3.org/TR/rdf-concepts/.

Martini, D., M. Schmitz, J. Frisch, M. Kunisch. 2009. Enabling integration of distributed data for agricultural software applications using agroXML. In *Bregt, A., S. Wolfert, J. E. Wien, C. Lokhorst. EFITA Conference ’09. Wageningen Academic Publishers, Wageningen, the Netherlands*. 145-150.

Mitra, N., Y. Lafon. 2007. SOAP Version 1.2 Part 0: Primer (Second Edition). World Wide Web Consortium. http://www.w3.org/TR/soap12-part0/.

Motik, B., P. F. Patel-Schneider, B. Parsia. 2012. OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax (Second Edition). World Wide Web Consortium. http://www.w3.org/TR/owl-syntax/.

Vretanos, P. A. 2005. Web Feature Service Implementation Specification. Open GIS Consortium Inc.

C0222
D. Martini. “agroRDF as a Semantic Overlay to agroXML: a General Model for Enhancing Interoperability in Agrifood Data Standards”. EFITA-WCCA-CIGR Conference “Sustainable Agriculture through ICT Innovation”, Turin, Italy, 24-27 June 2013.

Pesonen, L., B. Fusai, M. Koistinen, R. Lokers, E. Mietzsch, E. Rehben, A. Ronkainen, M. Schmitz, A. Turchi. 2012. D4.7 Final report of the achieved results in data exchange harmonizing. agriXchange project deliverable 4.7.

Richardson, L., S. Ruby. 2007. RESTful Web Services. O'Reilly, Sebastopol.

Schank, R., K. Colby. 1973. Computer Models of Thought and Language. Freeman, San Francisco.

Simmons, R. F. 1973. Semantic networks: Their Computation and use for Understanding English Sentences. In *Schank, R., K. Colby. Computer Models of Thought and Language. Freeman, San Francisco*.