SEMIC.EU Semantic Interoperability Centre Europe

# Semantic Interoperability in Agriculture

[Daniel Martini](), DE

## A Language for Agriculture

Documentation of agricultural practices is becoming more and more of an issue for farmers. On the one hand, they are increasingly obliged to it by legislation, on the other hand, integrative planning of agricultural production requires thorough information about measures and events in the past. In many cases, documentation has to be handed on to external partners, like e. g. government agencies or agricultural service providers. The demand for appropriate technical solutions for this purpose has become obvious. Up to now only individual interfaces between different communication partners in agriculture were available. Even if the farmer had electronic systems to record production data, the required data had to be transferred by hand from one software to another or from screen into paper forms. A standardised system for electronic data exchange offers new possibilities for information-directed agricultural production increasing sustainability and keeping adversary effects to the environment at a minimum. By allowing for an integrated view of farm production data and other data like e. g. climate or geographic data, measures can be adapted to different conditions, optionally leveraging algorithms or expert systems provided by third parties.

agroXML provides the necessary standardised language for these purposes. It is an XML dialect to describe the production processes on the farm and the real-world objects needed to conduct them. The released versions of the agroXML schema are available at http://www.agroxml.de/schema under the W3C open source licence. In the past, development of agroXML concentrated mostly on plant production. The content of agroXML instances can basically be classified into five categories: A block providing information about the farm in general like e. g. address, name of farm manager etc., a block of data about the fields, like e. g. area and geographic coordinates, a further block of data about the cultivation on different fields, like e. g.

the plant species, catch crops etc., then data about the individual measures carried out: fertilisation, seeding, pest control, tillage etc. and finally a block of data about supply items like fertilisers, pesticides, machinery etc. On the one hand, agroXML can be used to generate consistent stand-alone XML documents containing each of the five parts. But following the extensibility paradigm of XML, it also offers a collection of data types and elements reuseable and embeddable in other documents.

To facilitate integration with geographic services, spatial vector data are modeled in agroXML reusing constructs from the Geography Markup Language (GML) from the Open Geospatial Consortium (OGC). Technically, the reuse of GML datatypes and elements is achieved by creating a profile (a subset of necessary datatypes and elements) of GML and importing this profile together with the GML-namespace into agroXML. This approach has the advantage of allowing for a very lightweight implementation in programme code. Practical feasibility in different computing environments is an important factor while developing a data exchange standard for agriculture, especially, when it comes to integration of other XML vocabularies. Farm management information systems are written in different programming languages. Components of these management systems providing certain functionality run on a variety of hardware platforms from handhelds to powerful servers for web applications. While in theory the combination of different XML vocabularies seems desirable, in practice it often leads to large, bulky constructs unmanageable by common XML tools. However, simplicity, clarity and generality are key properties of well engineered IT systems. Therefore, one of the challenges in the future will be to allow for integration and extensibility of XML dialects while keeping unnecessary overhead at a minimum.

Besides the schema, agroXML also provides content lists. They offer the functionality of XML Schema eNumerations, however the mechanism of how they are included in the schema allows to add to their content dynamically without effecting a change in the schema itself. In addition, they not only contain the enumeration values themselves but also a name and a description of the item at hand. The lists conform to a unified schema and can be downloaded at http://www.agroxml.de/content. Several lists exist containing e. g. soil types, machine types, fertiliser types, pesticides and plant variety names. Where possible, content for these lists is obtained from the respective official agencies, like e. g. the plant variety offices. Software systems implementing agroXML can either use a local copy of the content lists for filling instances or use the version on the web. Different caching strategies are possible to ensure a recent data pool even if the internet connection is only intermittent.

It is important to note, that due to the dynamic integration into an XML instance (lists are referenced by their Uniform Resource Locator), it is possible to include different lists than the ones provided at http://www.agroxml.de/content for special purposes or containing language or country-specific content.

For real world applications, the transfer of agroXML instances on the internet can be conducted using standard protocols like the hypertext transfer protocol (HTTP), the file transfer protocol (ftp) or the simple mail transfer protocol (SMTP). Exchange is best done in a document- or resource-oriented manner as opposed to message-oriented systems: either a complete agroXML document or a document describing a certain object or process is transferred in a single file. This allows for a very simple setup of web services following the paradigm of so called Representational State Transfer (ReST). But it is also possible to embed agroXML content into messaging systems like for example ones based on SOAP.

Most of the development work of agroXML is done at the KTBL in Germany. However, agroXML is open to contributions from other stakeholders. The KTBL is providing and maintaining an infrastructure consisting of a source code management system and documentation. Coordination is done in a working group made up of the major producers of farm management information systems in Germany. Recently, an effort is going on to lift the developed XML technology onto a broader international level. There are different requirements concerning data exchange in agriculture in the different countries. This is mostly due to distinct regional agricultural practices but also due to different legislation. Integrating these different requirements will lead to challenging tasks especially concerning semantics of the data.

Further upgrading of geo-data (also raster data) functionality as well as addition of elements for livestock farming and cultivation of vegetables and fruit are currently worked on. This increasing demand from agricultural sectors other than only plant production leads to technological issues which have to be dealt with.

Especially, a schema architecture and design to allow modularisation and extensibility, while at the same time keeping internal consistency, is needed. The goal of the work is to provide a schema which can be used only in part to implement the datatypes needed for a specific application, while not breaking application interoperability. This basically calls for mechanisms to allow for more dynamic relations between data objects. In the (currently experimental) livestock farming extensions, an approach is worked out to relate basic building blocks to each other using XML linking technologies like XLink or XInclude.

At the moment, data about operating supply items, like e. g. fertilisers or pesticides are copied directly into XML instances by the farm management information systems. Suppose now, a company changes the nitrogen content of a certain fertiliser. With the current model of information integration, the farmer has to update this value in his farm management information system. If he is not aware of the change, he will transmit incorrect data in further transactions. So, in most cases, a better model would be to follow the paradigm of distributed storage and to leave this information at the place where it is produced, i.e. in the example above on a web server at the fertiliser producer, and use generic link mechanisms like the XLink Standard, to only reference the information. URIs offer an excellent system to provide globally unique identifiers. Resources like fertilisers would then be described using agroXML element hierarchies. Their relations could be modelled using the Resource Description Framework (RDF). Not only would this enable real distributed data storage and ensure recent information, but it would also enhance the possibilities to be able to build real knowledge bases for data mining and harvesting. Key factor for the success of such an architecture is a simple and easily adoptable standard.

However, especially with regard to documentation of processes, there are still some problems to solve in such a scenario of distributed storage. An infrastructure for reliable statements in RDF using strong cryptographic mechanisms to allow for identification of sources of such statements or to provide non-repudiation is still missing although recently, there are approaches described in computer science literature.

Future work on agroXML will try to address the issues described in the last two paragraphs. For agriculture as a whole, global linking of data sources would bring significant advantages. Farmers could derive useful information for planning measures on the field and in the barn. More appropriate reactions to current conditions become possible. Solving the semantic challenges in this area will thus finally lead to more food safety and a sustainable and environmentally friendly agricultural production.

***Daniel Martini, DE,*** *studied Agricultural Sciences at the University of Hohenheim. He specialised in soil science and got deeper involved with a broad range of information technologies while working with geographic information systems and doing modelling of water and solutes flow in soils. Since 2005, he has been engaged at the KTBL in the development of the agroXML data exchange standard for farm management information systems.*