



XVIIth World Congress of the International Commission of Agricultural and Biosystems Engineering (CIGR)

Hosted by the Canadian Society for Bioengineering (CSBE/SCGAB)
Québec City, Canada June 13-17, 2010



FORMAL REPRESENTATION OF AGRICULTURAL PRODUCTION STANDARDS

EDWARD NASH¹, JENS WIEBENSOHN¹, RAIMO NIKKILÄ²,
ANNA VATSANIDOU^{3,4}, SPYROS FOUNTAS^{3,4}

¹Rostock University, Faculty of Agricultural and Environmental Sciences, Chair of Geodesy and Geoinformatics, Justus-von-Liebig-Weg 6, 18059 Rostock, Germany.

¹J. Wiebensohn, jens.wiebensohn@uni-rostock.de

²Helsinki University of Technology, Department of Automation and Systems Technology, P.O. Box 5500, 02015 TKK, Finland.

³University of Thessaly, School of Ag. Science, Fytokou St. N. Ionia Magnisias, 38446 Volos, Greece.

⁴Centre for Research and Technology, Thessaly, 1st Industrial Zone, 38500 Volos, Greece

CSBE100970 – Presented at the 8th World Congress on Computers in Agriculture (WCCA) Symposium

ABSTRACT Production standards in the form of legal regulations or quality assurance labels are playing an increasingly important role in farming. Each farm must therefore gather information on all standards which apply, which may vary from field-to-field, and ensure that they are respected during operations. This information may be provided on paper or as electronic documents, by the standards publishers or by advisors. Together with the need to document compliance, the need to collect and process the requirements is becoming increasingly burdensome for farmers. In order to provide assistance, it is proposed that a machine-readable formal representation of standards be used. As an initial step, a general model of production standards was developed and applied to some common standards in European agriculture. Based on this model, separating standards into metadata and a list of individual rules (check points), a formal representation was developed. This uses elements from Dublin Core and OGC Geography Markup Language for representing the metadata, and W3C Rules Interchange Format and Web Ontology Language for representing the rules in a fully machine-readable form. Each rule is also additionally included in its original natural-language form, together with metadata indicating which field operations it regulates and what data is required for assessing compliance. In the first instance, this allows simple software to automatically generate a single checklist from all appropriate standards for manual use. More advanced software may use the machine-readable part to automate compliance checking. This paper describes the representation, its design rationale, and provides examples of its usage.

Keywords: metadata, ontology, rules, decision support, knowledge management.

INTRODUCTION Production and management standards are becoming increasingly important in agriculture: an increasing number of legal regulations to ensure food safety and agrienvironmental good practice are binding for all farmers, whilst voluntary standards and labels to demonstrate compliance to stricter requirements and ultimately

gain a higher price for agricultural produce are an important tool for farmers to market their products or enable them to sell to particular buyers and markets. Examples of legal regulations are laws affecting use of fertilisers, plant protection products, seed types, etc. Voluntary standards may be legally regulated, such as the EU Organic standard, or may be privately-run industry standard such as GlobalGAP. Adherence to particular standards may be motivated by direct financial benefits such as subsidy payments being linked to this, such as is the case with the European ‘Cross-Compliance’ regulations.

Each farm must potentially adhere to a large number of standards, and it is possible, or even likely, that different parts of the farm must be managed according to different standards, e.g. different crops being sold to different buyers who stipulate their own production standards, or parts of the farm falling in a specific area such as a water catchment protection area where tighter environmental regulations are enforced. Additionally, the laws applying to the farm vary according to the country or federal state, or in some cases even smaller administrative units. Each farm, or even each field or even partfield, must therefore be considered a potentially unique case in being managed according to a unique constellation of standards. Additionally, standards vary through time as new versions are produced.

Given this large number of different standards in use and the need for farmers to work with the correct standards, in the correct versions, active support from the farm software during the decision-making process in order to ensure that management decisions such as fertilisation and spraying plans conform to the relevant standards is desirable. The current state-of-the-art of such support is the production of individually-tailored checklists based on a number of standards, which may then be used by farmers and their advisors and consultants to perform a single manual check. Such a service is offered e.g. for common standards in use in Germany by KKL-Service (2010). This is however not integrated with the farmers’ existing software, where farm data are already held, and does not offer any automated assessment. Based on the evaluation of whether such an automated assessment is possible (Nash et al., 2009), and the general structure of an agricultural standard developed during the course of that evaluation, a formal, machine-readable XML-based encoding of such standards is proposed here.

In the next section, the general model of a standard is reviewed before the formal encoding is introduced. This follows in three sections relating to metadata, rules and vocabulary. Finally, we present some discussion regarding the work presented here and what developments may be necessary in order for such a system to be used in practice in the future.

STRUCTURE OF AGRICULTURAL STANDARDS Based on the analysis of representative agricultural production and management standards, Nash et al. (2009) present a general structural model of an agricultural standard, together with four criteria which must be met in order to enable automated compliance checking. As these form the basis for the work presented here, they will be reviewed in detail.

An agricultural standard may be considered as being composed of a set of rules together with metadata describing the intention of the standard, its validity, its publisher, etc. Additionally, each rule may have certain metadata attached to it regarding how compliance to that rule is to be assessed, and whether all rules must be complied with in

order for the whole standard to be complied with or whether only a certain percentage of individual rules must be met. Each rule is effectively a predicate (i.e. a logical statement which may be evaluated to *true* or *false*), together with a conclusion (i.e. compliance or violation of the standard). Rules may be classified as either an obligation ('the standard is complied with only if the farmer does *x*') or a prohibition ('the standard is not complied with if the farmer does *y*'). Additionally, rules may require that particular actions are documented, whilst not proscribing how they should be performed. Although these may be considered as obligations, they are treated separately as they do not directly affect the decision-making related to field operations (e.g. the volume of nitrogen fertiliser to be applied). This model is presented graphically in Figure 1.

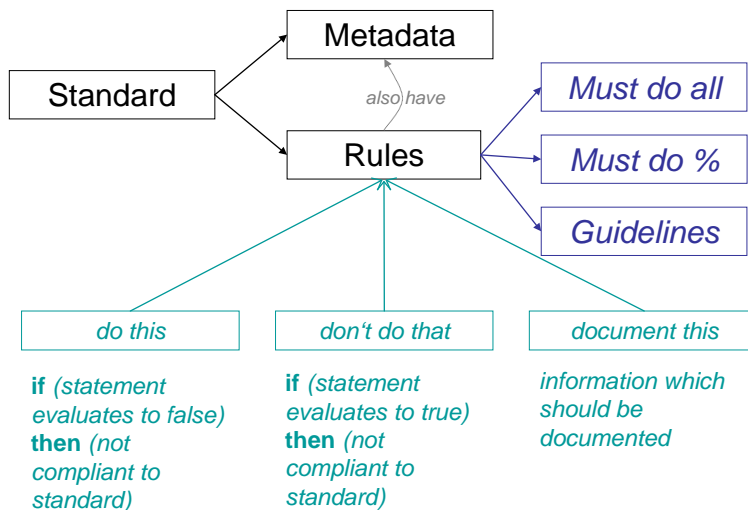


Figure 1. General structural model of an agricultural standard.

Current agricultural standards are not explicitly presented in this form; most legal regulations are presented as texts, whilst in the best case the standard may be presented as a checklist of individual rules (e.g. GlobalGAP). Any standard can however be relatively straightforwardly converted to a checklist of rules (Vatsanidou et al., 2009).

In order to enable the automated assessment of each rule, four prerequisites must be met:

1. The rule must be encoded in a machine-readable form.
2. The rule, and all terms used in defining it, must be capable of being interpreted by the software.
3. Each rule must have a discrete outcome which can be determined by a computer.
4. The required data inputs for assessment must be available in digital form at the point of assessment.

The first of these is the basic motivation for the current paper whilst the second defines the requirements of a transfer format, namely that we wish to be able to transfer computable rules and the definition of all concepts used within these rules. The former implies use of a transfer format for rules, and the latter a transfer format for ontologies where an ontology is a “formal, explicit specification of a shared conceptualisation” (Gruber, 1993) or a definition of terms used and the relationships between them. An ontology may in this case be considered as a formal specification of a vocabulary.

The basic structure for a formal machine-readable representation of agricultural standards therefore consists of three parts; metadata about the standard, a definition of the terms used in defining the rules, and the rules themselves together with the rule-specific metadata. The XML schema developed follows this structure, and its outline is shown in Figure 2.

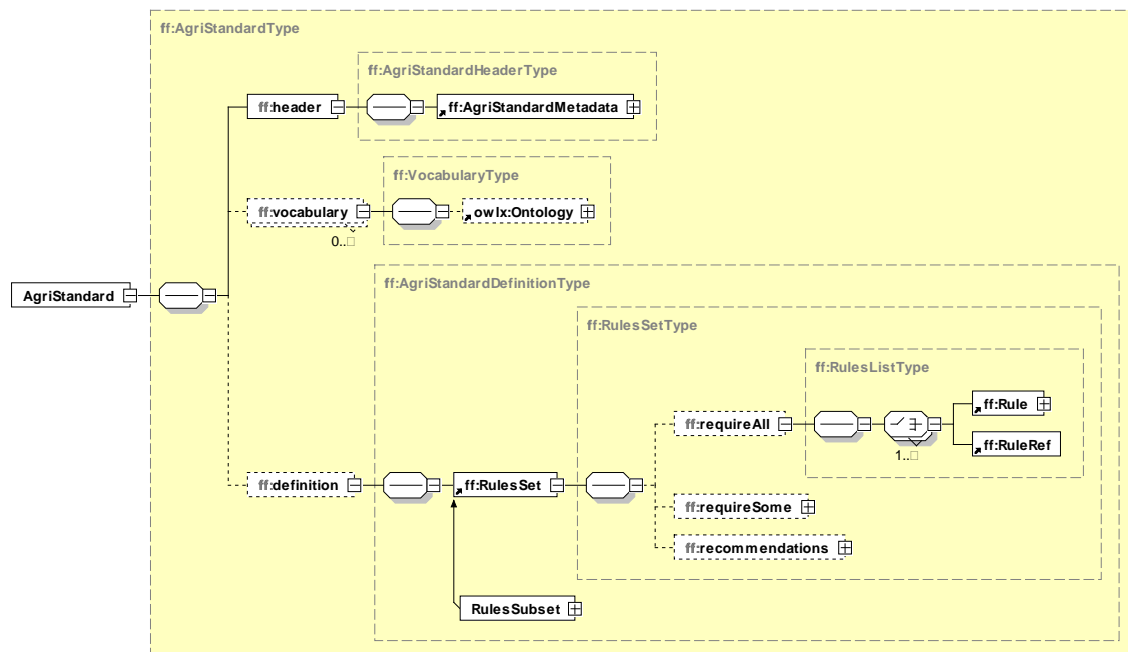


Figure 2. Top-level structure of XML schema for representing agricultural standards.

METADATA FOR AGRICULTURAL STANDARDS In this section the modelling of the metadata required for describing agricultural standards will be described. Metadata may be defined as a description of the actual contents, and thus the metadata here is a description of the standard which is necessary to understand its context and to evaluate whether it is relevant to a particular case. This includes common metadata terms such as publisher, date of publication, etc. as well as specialist items including spatiotemporal validity and relationships to other standards. There are many existing XML schema for metadata, such as Dublin Core (DCMI, 2004) and ISO19115 (ISO, 2003) but none of them cover fully the details for agricultural standards. Figure 3 illustrates the elements which are included in the agricultural standard metadata. It can be seen that elements from Dublin Core (prefix *dc*) and ISO19136 Geography Markup Language (ISO, 2007, prefix *gml*) are used for standard metadata elements and spatiotemporal representation respectively.

One complication in this case is the representation of the spatial region to which a standard applies. Two distinct categories can be identified; the production region (i.e. where the farm is located) and the end product region (i.e. where the farm products will ultimately be sold to consumers). The first of these is relevant with respect to e.g. fertiliser regulations which are valid for all farms within a particular country. The second of these is relevant for particular product standards e.g. organic production standards such as the EU organic regulations are valid for all products sold as organic in the European Union, wherever in the world they were produced. This requires the inclusion of separate meta data elements for each case. Within each element, there are multiple ways in which the

valid region may be defined: most common is likely to be the location name, e.g. the name of the country to which the law applies. Alternatively, the standard may apply to a particular class of region, e.g. nitrate vulnerable zones (NVZ). However, it is also feasible

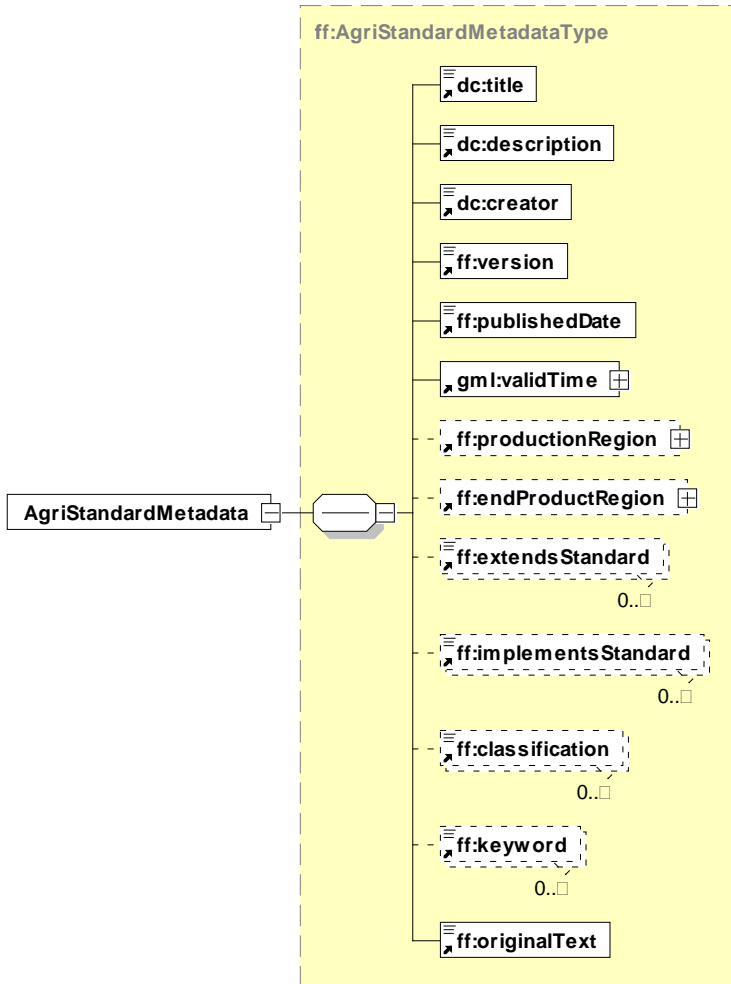


Figure 3. Elements for representing metadata for an agricultural standard

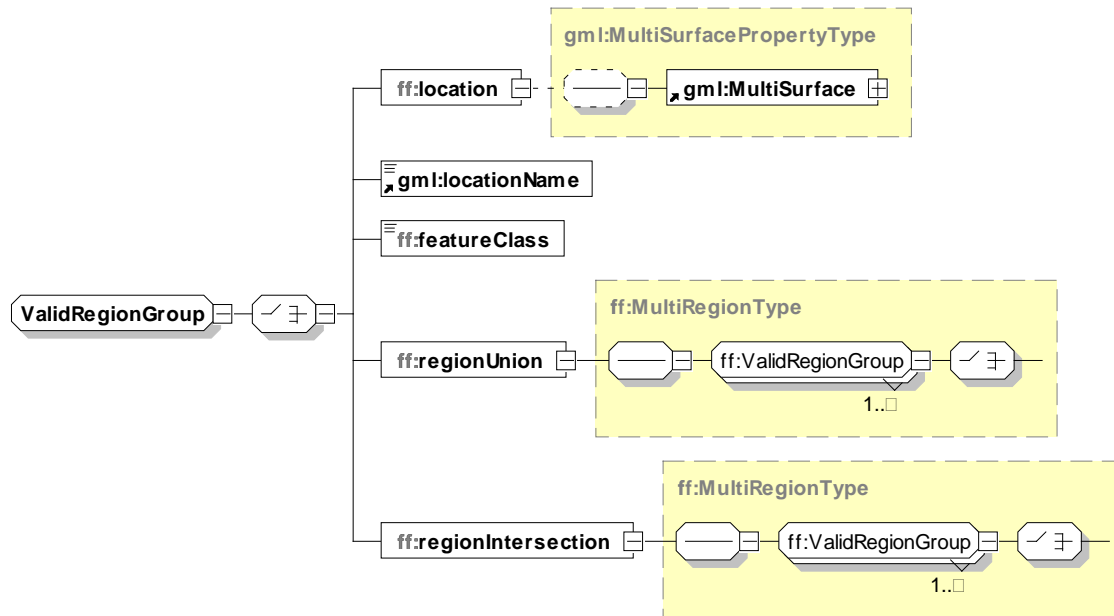


Figure 4. Structure for representing the valid region for an agricultural standard

that a geometric region is directly defined using coordinates. These individual region types may also be combined, e.g. to specify that a standard applies only to nitrate vulnerable zones in England and Wales it is necessary to specify the intersection of the feature class ‘nitrate vulnerable zone’ with the union of the location names ‘England’ and ‘Wales’. In order to allow the correct specification of regions, the structure shown in Figure 4 was developed. An example of the usage of the whole metadata encoding is given as an XML fragment below:

```

<ff:AgriStandardMetadata
  metadataId="urn:agrystandards:metadata:uk:defra:nitrate:2009-04-01">
  <dc:title>Guidance for Farmers in Nitrate Vulnerable Zones</dc:title>
  <dc:description>The Guidelines for Farmers in Nitrate Vulnerable Zones (NVZ)
    describe what farming operations are allowed in NVZ in England and
    Wales.</dc:description>
  <dc:creator>Department for Environment, Food and Rural Affairs</dc:creator>
  <ff:version>2009-04-01</ff:version>
  <ff:publishedDate>2009-04-01T00:00:00+01:00</ff:publishedDate>
  <gml:validTime>
    <gml:TimePeriod gml:id="timePeriod01">
      <gml:beginPosition>2009-04-01T00:00:00+01:00</gml:beginPosition>
      <gml:endPosition>2010-12-31T00:00:00+01:00</gml:endPosition>
    </gml:TimePeriod>
  </gml:validTime>
  <ff:productionRegion>
    <ff:regionIntersection>
      <ff:featureClass>NitrateVulnerableZone</ff:featureClass>
      <ff:regionUnion>
        <gml:locationName>England</gml:locationName>
        <gml:locationName>Wales</gml:locationName>
      </ff:regionUnion>
    </ff:regionIntersection>
  </ff:productionRegion>
  <ff:implementsStandard>
    urn:agrystandards:eu:nitrate:1991-12-12</ff:implementsStandard>
  <ff:classification>mandatory legal regulation</ff:classification>
  <ff:originalText>

```

ONTOLOGIES FOR AGRICULTURAL STANDARDS One criteria for enabling automated assessment of rules is that the terms used in defining the rules are unambiguously defined in a machine-readable form. Also for a reliable manual assessment of rules, the unambiguous definition of terms is essential in order to avoid differing interpretations between those defining the rules and those assessing compliance. Ontologies are a tool for defining concepts and the relationships and differences between them in a formal way, and have particularly risen to prominence as part of the semantic web. The most common language for modelling ontologies is the W3C Web Ontology Language (McGuinness & van Harmelen, 2004), which also provides an XML-based representation for interchange of ontologies, together with a functional, more human-readable syntax. Ontologies are widely used e.g. in the biomedical domain where large collections of orthogonal interoperable ontologies are available (e.g. at <http://www.obofoundry.org>). There has been some recent interest in ontologies and their role in data exchange in the agricultural domain (e.g. Sall et al, 2009, Maliappis, 2009, FAO, 2010), but there is no widespread acceptance of ontologies in practical use.

However, many standards for agriculture include a definition of terms in the form of a glossary or legal definitions at the start of a text. These may with a little effort be converted to a formal ontology. As an example we present a definitions of a term which is given in the German fertiliser regulations (DüVo) and how it may be represented using OWL. The DüVo defines a subclass of ‘Fertiliser’, whose use is forbidden under certain conditions, ‘Fertiliser with significant nutrient content’ is defined as a fertiliser with a nutrient content >1.5% total nitrogen or >0.5% phosphate (P₂O₅). Further subclasses are defined for fertilisers with only significant N content (>1.5%) and significant available nitrogen content (>10% soluble in CaCl). The following presents these definitions using the OWL functional syntax, where we assume that the general term ‘Fertiliser’ is defined in an Agrovoc ontology and only terms and attributes specific to the DüVo must be additionally defined. An XML version of these definitions (and more) is available at <http://schema.futurefarm.eu/ontologies/duevo.xml>.

```
Declaration(Class(duevo:FertiliserWithSignificantNutrientContent))
EquivalentClasses(
  duevo:FertiliserWithSignificantNutrientContent
  ObjectUnionOf(
    ObjectAllValuesFrom(duevo:hasPhosphateContent percentages:PercentGreater0.5)
    ObjectAllValuesFrom(duevo:hasNitrogenContent percentages:PercentGreater1.5))
SubClassOf(duevo:FertiliserWithSignificantNutrientContent agrovoc:Fertiliser)
```

```
Declaration(Class(duevo:FertiliserWithSignificantNitrogenContent))
EquivalentClasses(
  duevo:FertiliserWithSignificantNitrogenContent
  ObjectAllValuesFrom(duevo:hasNitrogenContent percentages:PercentGreater1.5))
SubClassOf(
  duevo:FertiliserWithSignificantNitrogenContent
  duevo:FertiliserWithSignificantNutrientContent)
```

```
Declaration(Class(duevo:FertiliserWithSignificantAvailableNitrogenContent))
EquivalentClasses(
  duevo:FertiliserWithSignificantAvailableNitrogenContent
  ObjectAllValuesFrom(
    duevo:hasProportionSolubleInCalciumChloride percentages:PercentGreater10))
SubClassOf(
```

```
duevo:FertiliserWithSignificantAvailableNitrogenContent
duevo:FertiliserWithSignificantNitrogenContent)
```

FORMAL REPRESENTATION OF RULES As stated previously, the automated assessment of compliance to rules requires that the rules are encoded in a machine-readable form. In the current case, this must be as XML. Although it is debatable whether all rules may be automatically assessed with any reliability, and the concept of ‘mechanical jurisprudence’ is rejected by experts in the field of law and artificial intelligence as fundamentally unrealistic (“law is more ‘rule-guided’ than ‘rule-governed’” – Gardner, 1987), this does not mean that the rules may not be formally defined and represented, only that the interpretation may not always be done automatically (Boer et al., 2007). It is therefore assumed that those rules which may be straightforwardly and unambiguously interpreted automatically may be automatically processed, but that where this is not possible the definition of the rule, together with the relevant data, may be presented to the farmer or advisor in order to manually assess compliance. Furthermore, since the conversion of rules to a formal, logical-mathematical format is time-consuming and the main benefits will not be realised until farm software is capable of reasoning with these rules, and farm systems are capable of automatically supplying all required data, it is proposed that in the initial stages, the formal representation of the rules should be optional, i.e. only the original natural-language version of each rule must be supplied. This considerably lowers the entry barrier to producing standards in a basic machine-readable way.

There have been many proposals for formal representation of rules in XML, such as RuleML (Hirtle et al., 2006), SWRL (Horrocks et al, 2004), WRL (de Bruijn, 2005) and R2ML (REWERSE, 2006). However, none of these have gained broad acceptance. A current initiative within the W3C is the creation of a Rules Interchange Format (RIF – Boley et al., 2009) which is, at the time of writing, at the ‘Candidate Recommendation’ stage. This allows the representation of rules as sentences based on the individual atoms, functions and predicates which may be identified in their natural language representation and is expected to become the future standard for representation of rules on the Internet and in XML. As well as an XML-based format, the RIF defines a human-readable ‘presentation syntax’ based on the Extended Backus-Naur Form for context-free grammars.

As an example, the German fertiliser law (DüVo) states that it is forbidden to apply fertiliser with significant available nitrogen content, other than manure, on cropland between 1 November and 31 January and on grassland between 15 November and 31 January. This could be formally expressed in the RIF presentation syntax as follows. Note that the definitions of terms defined in ontologies (agrovoc, duevo) are referenced from the RIF rule definition.

```
for all ?fertiliser_application (
violation(DüVo) :- And (
  ?fertiliser_application#agrovoc:FertiliserApplication
  ?fertiliser_application[agrovoc:fertiliser->?ft]
  ?fertiliser_application[agrovoc:application_area->?applied_to]
  ?fertiliser_application[agrovoc:date->?application_date]
  ?ft#duevo:FertiliserWithSignificantAvailableNitrogenContentWithoutManure
Or (
  And (?applied_to#agrovoc:Cropland
    External(ff:during(?application_date
```

```

External(ff:time_period(01.11. 31.01.))))))
And (?applied_to#agrovoc:Grassland
  External(ff:during(?application_date
    External(ff:time_period(15.11. 31.01.)))))))))

```

Additional to the formal definition of the rule using RIF, some metadata is required for each individual rule both to enable it to be interpreted correctly and to enable efficient searching and presentation of rules. It is assumed that the most common time for requiring the detail of a rule is whilst planning an operation which is likely to be affected by that rule, and therefore each rule may be annotated with a list of operations to which it applies, since this may not be obvious from the definition of the rule itself, in order to enable simple filtering. Similarly, the data elements which may be required in order to assess compliance to the rule may be explicitly stated: it is expected that these would be referenced to concepts in the ontologies, as is the case for the specification of concepts in the body of the rule. Although theoretically a grammatically-correct natural-language sentence could be reconstructed from the RIF representation, for simplicity and in cases where the RIF version is not available, the original text should be included, together with a clarifying description to make interpretation easier. The complete structure developed for representing an individual rule is shown in Figure 5. Due to the extreme verbosity of the RIF XML syntax, no example of its usage is given here. A full example file including usage of the RIF XML may be found at <http://schema.futurefarm.eu/agstandard/example.xml>.

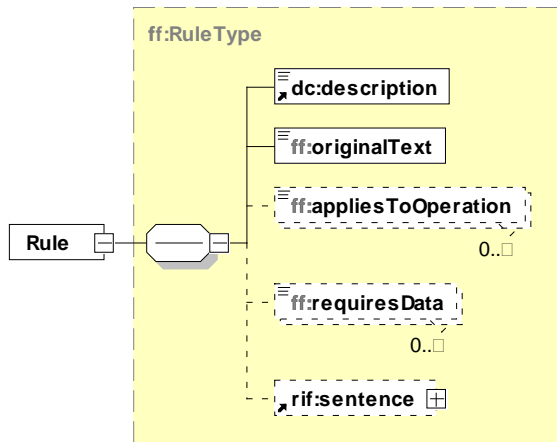


Figure 5. Structure for representation of an individual rule

CONCLUSION This paper has presented an XML-based format for formal representation of agricultural crop production and management standards. This allows the transfer of these standards between systems, which enables future service-oriented architecture (SOA) approaches to distributing knowledge on agricultural standards (described Nikkilä et al, 2010). Due to the increasing importance of such standards, and the inherent problems in incorporating such knowledge directly into software, such an approach will become increasingly important in future.

Whilst an automated internal assessment of compliance to standards is the ultimate aim, and could to a large extent be realized using the combination of ontologies and formal presentation of rules described here, assuming that the required data resources are available, even the initial step of making the standards available as a series of individual

rules which may be filtered and combined by farmers in order to inform them of requirements during decision-making and to enable them to produce individually-tailored checklists for manual compliance assessment should prove to be a great benefit.

In current and future work, the possibility for automated assessment of compliance to RIF-formulated rules for agriculture is being tested. If this may be realized, together with the SOA for agricultural standards, it will effectively enable farmer decision-support software to be self-configuring for each farm or field and to be able to respect legal and voluntary restrictions on planned operations.

Acknowledgements. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 212117.

REFERENCES

- Boer, A., Di Bello, M., van den Berg, K., Gordon, T., Föhrécz, A., Vas, R., 2007. Specification of the Legal Knowledge Interchange Format. ESTRELLA Deliverable 1.1. Available at: <http://www.estrellaproject.org/doc/D1.1-LKIF-Specification.pdf>. Accessed 16 February 2010.
- Boley, H., Hallmark, G., Kifer, M., Paschke, A., Polleres, A. and D. Reynolds. 2009. RIF Core Dialect. Available at <http://www.w3.org/TR/rif-core/>. Accessed 16 February 2010.
- DCMI (Dublin Core Metadata Initiative). 2004. Dublin Core Metadata Element Set, Version 1.1: Reference Description. Available at <http://www.dublincore.org/documents/dces>. Accessed 15 February 2010.
- De Bruijn, J. (ed). 2005. Web Rule Language (WRL). Available at <http://www.w3.org/Submission/WRL/>. Accessed 16 February 2010.
- FAO (Food and Agriculture Organization of the United Nations). 2010. Agrovoc: from a thesaurus to an ontology. Available at <http://aims.fao.org/website/AGROVOC-Concept-Server/sub>. Accessed 16 February 2010.
- Gardner, A., 1987. An artificial intelligence approach to legal reasoning. MIT Press.
- Gruber, T. 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition* 5(2): 199-220.
- Hirtle, D., Boley, H., Grosf, B., Kifer, M., Sintek, M., Tabet, S. and G. Wagner. 2006. Schema Specification of RuleML 0.91. Available at <http://ruleml.org/0.91/>. Accessed 16 February 2010.
- Horrocks, I., Patel-Schneider, P., Boley, H., Tabet, S., Grosf, B. and M. Dean. 2004. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. Available at <http://www.w3.org/Submission/SWRL/>. Accessed 16 February 2010.
- ISO (International Organisation for Standardisation). 2003. Geographic Information – Metadata. ISO19115:2003.
- ISO (International Organisation for Standardisation). 2007. Geographic information -- Geography Markup Language (GML). ISO19136:2007.
- KKL-Service. 2010. Kriterien-Kompodium Landwirtschaft (Criteria Compendium for Agriculture). Available at: www.kkl-service.info. Accessed 15 February 2010.
- Maliappis, M. 2010. Applying an agricultural ontology to web-based applications. *International Journal of Metadata, Semantics and Ontologies* 4 (1/2): 133-140.
- McGuinness, D.L. and F. van Harmelen (eds.). 2004. OWL Web Ontology Language. Available at <http://www.w3.org/TR/owl-features/>. Accessed 16 February 2010.

- Nash, E., Vatsanidou, A. and S. Fountas. 2009. Can compliance to crop production standards be automatically assessed? In: van Henten, E.J., Goense, D. and C. Lokhorst, Precision agriculture '09, Proceedings of the 7th European Conference on Precision Agriculture, Wageningen, the Netherlands, 6-8 July 2009. Wageningen Academic Publishers. 899-906.
- Nikkilä, R., Nash, E., Seilonen, I. and K. Koskinen. 2010. REST-based web services for the discovery and distribution of agricultural production standards. Proceedings, 8th World Congress on Computers in Agriculture (this volume).
- REWERSE. 2006. R2ML – The REWERSE I1 Rule Markup Language. Available at <http://rewerse.net/I1/oxygen.informatik.tu-cottbus.de/rewerse-i1/@q=r2ml.htm>. Accessed 16 February 2010.
- Sall, O., Lo, M., Gandon, F., Niang, C. and I. Diop. 2009. Using XML data integration and ontology reuse to share agricultural data. *International Journal of Metadata, Semantics and Ontologies* 4 (1/2): 93-105.
- Vatsanidou, A., Fountas, S., Aggelopoulou, K., Gemtos, F., Strobl, K., Charvat, K., Pedersen, S.M., Sørensen, C.G., Nash, E. and Blackmore, S., 2009. Compliance to standards specifications. FutureFarm Project Deliverable 2.1.1. Available at <http://www.futurefarm.eu/node/159>. Accessed 15 February 2010.