

# Datenintegration zwischen Standards in der Landwirtschaft auf Basis semantischer Technologien

Daniel Martini, Mario Schmitz, Martin Kunisch

Bereich Datenmanagement  
Kuratorium für Technik und Bauwesen in der Landwirtschaft e. V. (KTBL)  
Bartningstraße 49  
64289 Darmstadt  
d.martini@ktbl.de  
m.schmitz@ktbl.de  
m.kunisch@ktbl.de

**Abstract:** Obwohl die Standardisierung von Datenformaten und Austauschmechanismen für Informationen im Agrarsektor und angrenzenden Bereichen inzwischen gute Fortschritte gemacht hat, ist nach wie vor eine mangelhafte Integration von Daten aus verschiedenen Quellen zu beobachten. Semantische Technologien können dabei helfen, diesen Missstand zu beheben indem sie Methoden bereitstellen, mit verschiedenen syntaktischen Formaten umzugehen. Der Beitrag zeigt Vorgehensweise und Methodik bei der Erstellung eines semantischen Modells auf Basis von agroXML für diesen Zweck auf.

## 1 Einleitung

Die Menge an für Entscheidungsunterstützung auf dem landwirtschaftlichen Betrieb nützlichen und verfügbaren Daten steigt stetig. Im Bereich der Fernerkundung werden weitere Parameter über neue Sensoren meßbar. Maschinen erlauben die elektronische Erfassung einer Vielzahl von Größen während des Betriebs. Stallnetze ermöglichen das Sammeln von Daten zu Stallklima, Fütterung, Energie- oder Wasserverbrauch. Auch über das Internet werden zunehmend Informationen mit landwirtschaftlichem Bezug – z. B. zu Sorten, Pflanzenschutz- oder Tierarzneimitteln – zugänglich.

Die betriebszweigübergreifende Nutzbarkeit all dieser Daten blieb aber bislang hinter den Erwartungen zurück. Zu Anlagen und Maschinen lieferten die Hersteller in der Vergangenheit in aller Regel proprietäre Softwarelösungen, die dann jeweils nur in beschränktem Maße mit Daten aus anderen Quellen umgehen konnten. Daten im Internet müssen meist umständlich interaktiv gesucht und bei Notwendigkeit einer weitergehenden Auswertung manuell kopiert werden. Insgesamt ist also immer noch eine unzureichende Datenintegration im landwirtschaftlichen Bereich zu beobachten.

Bisher wurde versucht, diesem Problem durch Standardisierung zu begegnen. Dabei konnten für Teilbereiche der Domäne Landwirtschaft beachtliche Erfolge erzielt werden. Die Kompatibilität von ISOBUS-fähigen Maschinen nimmt stetig zu. agroXML ist auf dem Weg, den Datenaustausch mit Internettechnologien zu ermöglichen. In angrenzenden Domänen, wie dem Geodaten- oder dem Finanzbereich, stehen zwischenzeitlich eine Reihe von standardisierten Diensten mit zugehörigen Austauschformaten wie die Geography Markup Language (GML, [CDLPW04]) bzw. die Universal Business Language (UBL, [BM06]) oder die eXtensible Business Reporting Language (XBRL, [EHSKW03]) zur Verfügung. All diese Standards stellen aber Daten mit unterschiedlichen Protokollen und in unterschiedlicher Syntax zur Verfügung. Inhaltlich bestehen teilweise Überlappungen. Der Aufwand, der für eine erfolgreiche Datenintegration zu betreiben ist, ist derzeit deshalb hoch. Unterschiede zwischen Standards sind häufig das Ergebnis einer Optimierung für den jeweiligen Anwendungsbereich, d. h. selbst wenn versucht würde, sich auf einen Standard zu einigen, bliebe das Ergebnis voraussichtlich insgesamt hinter den Erwartungen zurück.

## **2 Material und Methoden**

Im Zusammenhang mit den Arbeiten an agroXML wird derzeit ein Lösungsansatz für das beschriebene Problem auf Basis von semantischen Technologien entwickelt. Hierbei wird auf vier Ebenen gearbeitet. Ebene eins bildet die eigentliche syntaktische Definition eines Datenstandards. Derzeit wird als Beispiel hierfür agroXML herangezogen, die Einbindung weiterer Standards ist möglich (s. Abschnitte 3 und 4). Hierüber wird eine Ebene gelegt, die über einen Satz von Aussagen Zusammenhänge zwischen Datenelementen beschreibt und somit ein semantisches Modell des Datenformates liefert. Als Technologie kommt hier der Resource Description Framework (RDF, [KC04]) sowie RDF(S) [BG04] der W3C zum Einsatz. Die Verknüpfung zwischen semantischem und syntaktischem Modell wird mit Hilfe des W3C Standards SAWSDL [FL07] hergestellt. Als dritte Ebene wird ein Mapping in eine allgemein verständliche Begriffssammlung erstellt. Dabei werden zunächst nur exakte Übereinstimmungen zwischen Bedeutungen von Konzepten betrachtet, so dass hierfür aus der Web Ontology Language (OWL, [PHH04]) die Eigenschaft ‚sameAs‘ herangezogen werden kann. Die vierte Ebene bildet das Vokabular, in das verknüpft wird. Im bearbeiteten Beispiel ist dies der AGROVOC Thesaurus der FAO (s. <http://aims.fao.org>), der derzeit weit über 25000 Begriffe aus dem Agrarbereich in formalisierter, multilingualer Weise vorhält. Auch für diese Ebene ist ein Austausch denkbar, sodass z. B. Mappings in andere Vokabularien und Modelle möglich wären.

### 3 Ergebnisse und Diskussion

Abbildung 1 zeigt einen stark vereinfachten, kurzen Ausschnitt aus dem auf Basis von agroXML erstellten, prototypischen semantischen Modell. Kommentare (rdf:comment) sowie Label (rdf:label) wurden der Einfachheit halber weggelassen. In der Abbildung gezeigt sind nur die meistverwendeten Konstrukte zur Beschreibung von Relationen zwischen und Eigenschaften von Ressourcen: rdfs:Class als Klassentyp, rdfs:subClassOf für die Erstellung von Unterklassen, rdf:Property zur Definition von Eigenschaften/Attributen und rdf:subPropertyOf als Möglichkeit zur Definition von abgeleiteten Eigenschaften. rdfs:range und rdfs:domain geben an, in welchem Zusammenhang Eigenschaften und Klassen stehen können (s. a. [BG04]).

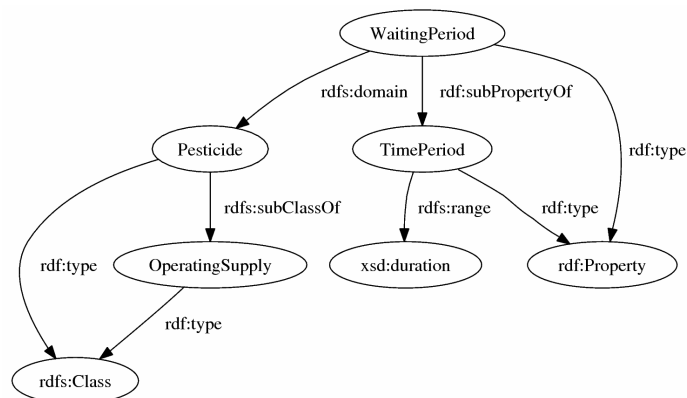


Abbildung 1: Stark vereinfachter Ausschnitt aus der RDF(S)-Repräsentation von agroXML.

Diese graphenorientierte Vorgehensweise bei der Erstellung des semantischen Modells weist eine Reihe von Vorteilen auf. Zum Einen ist sie von verschiedenen Datenmodellierungsparadigmen unabhängig, d. h. tabellenorientierte Strukturen lassen sich ebenso wie objekt-orientierte oder hierarchische Modelle bearbeiten. Zum Anderen sind graphenorientierte Modelle leichter zur Laufzeit eines Systems noch erweiterbar und flexibel an neue Anforderungen anpassbar.

Als ein Nachteil von RDF hat sich erwiesen, dass n-äre Relationen nur mit dem Hilfskonstrukt von sogenannten blank nodes abbildbar sind. So lässt sich beispielsweise die enge Bindung zwischen Zahl und Einheit, die in physikalischen Größen besteht, nur unzureichend explizit modellieren. Andere semantische Technologien wie z. B. ISO Topic Maps (ISO 13250) bieten diese Möglichkeit. Diese können aber wegen der noch nicht in ausreichendem Maße gegebenen Unterstützung durch Bibliotheken in verschiedenen Programmiersprachen im Projekt nicht zum Einsatz kommen.

Unter Zuhilfenahme der geschilderten Ebenen ist eine modellgestützte Konvertierung zwischen verteilten Daten in verschiedenen syntaktischen Formaten möglich, sodass diese für eine einheitliche Suche und Verarbeitung zugänglich werden. Semantische Modelle sind eine Grundvoraussetzung für die Vernetzung von Anwendungen, die verschiedene Standards nutzen und vereinfachen somit die Entwicklung von hersteller-, anwendungs- und standardübergreifenden Informationssystemen für die Landwirtschaft.

## 4 Ausblick und Danksagung

Das Mapping in den AGROVOC Thesaurus ist derzeit noch unvollständig und wird weiter ausgearbeitet. Methoden, die erlauben, in dieses Mapping auch Relationen einzufügen, die nicht durch exakte Übereinstimmung (owl:sameAs) repräsentiert werden können, sind bislang noch nicht untersucht und eingesetzt, werden in Zukunft aber auch betrachtet.

Ähnliche semantische Ebenen sind für weitere potenziell nutzbare Datenquellen neben agroXML in Planung. Der Aufwand einer Umsetzung für das Agricultural Data Element Dictionary (ADED) aus dem Tierhaltungsbereich beispielsweise würde sich nach heutigem Kenntnisstand in überschaubarem Rahmen halten, da wichtige Bausteine (Identifikation der Datenelemente, Label) bereits vorhanden sind. Im Bereich der Norm ISO 11783 (Tractors and machinery for agriculture and forestry – Serial control and communications data network) wäre eine zusätzliche, semantische Modellierung ebenso denkbar, aufgrund der dort gegebenen Auftrennung in Schema und Data Dictionary aber voraussichtlich im Detail anders umzusetzen.

Prototypen und Demonstratoren für die modellgestützte Konvertierung von Daten aus agroXML-Instanzen in die vernetzte, graphenbasierte Struktur und zurück, mit deren Hilfe dann auch Navigation und Darstellung der Daten und die Informationsgewinnung für den Nutzer verbessert werden können, sind aktuell in Arbeit.

Wir danken dem Bundesministerium für Bildung und Forschung für die Förderung der Arbeiten unter dem Förderkennzeichen 01IA08005X.

## Literaturverzeichnis

- [BG04] Brickley, D.; Guha, R. V.: RDF Vocabulary Description Language 1.0: RDF Schema. World Wide Web Consortium, 2004. <http://www.w3.org/TR/rdf-schema/>.
- [BM06] Bosak, J.; McGrath, T.: Universal Business Language 2.0. Organization for the Advancement of Structured Information Standards (OASIS), 2006. <http://docs.oasis-open.org/ubl/os-UBL-2.0/>.
- [CDLPW04] Cox, S.; Daisey, P.; Lake, R.; Portele, C.; Whiteside, A.: OpenGIS Geography Markup Language (GML) Implementation Specification. Open GIS Consortium Inc., 2004.
- [EHKW03] Engel, P.; Hamscher, W.; Shuetrim, G.; von Kannon, D.; Wallis, H.: Extensible Business Reporting Language (XBRL) 2.1. XBRL International, 2003. <http://www.xbrl.org/Specification/XBRL-RECOMMENDATION-2003-12-31+Corrected-Errata-2008-07-02.rtf>.
- [FL07] Farrell, J.; Lausen, H.: Semantic Annotations for WSDL and XML Schema. World Wide Web Consortium, 2007. <http://www.w3.org/TR/sawSDL/>.
- [KC04] Klyne, G.; Carroll, J. J.: Resource Description Framework (RDF): Concepts and Abstract Syntax. World Wide Web Consortium, 2004. <http://www.w3.org/TR/rdf-concepts/>.
- [PHH04] Patel-Schneider, P. F.; Hayes, P.; Horrocks, I.: OWL Web Ontology Language – Semantics and Abstract Syntax. World Wide Web Consortium, 2004. <http://www.w3.org/TR/owl-semantics/>.