

Daten- und Systemintegration im Precision Livestock Farming mit Serviceorientierten Architekturen und Semantischen Technologien

Daniel Martini, Jochen Traunecker, Mario Schmitz, Eva Gallmann

Kuratorium für Technik und Bauwesen in der Landwirtschaft e. V. (KTBL)

Bartningstraße 49

64289 Darmstadt

d.martini@ktbl.de

jochen.traunecker@gridsolut.de

m.schmitz@ktbl.de

eva.gallmann@uni-hohenheim.de

Abstract: Innerhalb der Wertschöpfungskette Ferkelerzeugung – Mast – Schlachthof existieren mehrere Informationsquellen und -senken die bislang technisch nur eingeschränkt vernetzt sind. Derzeit werden im InfrAgrar-Projekt Methoden und Ansätze entwickelt, die Verknüpfung bestehender Systeme auf Basis von semantischen Technologien und serviceorientierten Architekturen zu vereinfachen. Kern hierbei ist der „virtuelle Bauernhof“, der in Komponenten an verschiedenen Stellen verteilt lauffähig ist und Kommunikation über ontologiegesteuerte Schnittstellen realisiert.

1 Einleitung und Problemstellung

Neben den klassischen Produktionsfaktoren Arbeit, Boden, Kapital und Rechten gewinnt Information in der Landwirtschaft zunehmend an Bedeutung als Faktor, der den unternehmerischen Erfolg eines Betriebes bestimmt. Daten werden zwar an einer Vielzahl von Stellen – z. B. durch Sensoren oder durch Anbieter im Internet – bereitgestellt, dennoch handelt es sich dabei um Informationsinseln, die bislang in nur geringem Maß sinnvoll zur Ableitung von Handlungsempfehlungen verknüpft werden können. Im Rahmen des Projektes InfrAgrar werden derzeit Methoden und Technologien entwickelt, die am Beispiel der Domäne Mastschweinehaltung die Verknüpfung von im Stall erhobenen Daten mit weiteren Daten, die von außerhalb des Betriebes oder aus anderen Betriebszweigen stammen und derzeit in verschiedenen Standards, unstrukturierten Formaten oder Dokumenten geliefert werden, ermöglicht. Ziel dabei ist es, Daten anwendungsfallunabhängig zugänglich zu machen. Es wird also nicht lediglich ein einziges Anwendungsszenario betrachtet sondern versucht, grundlegende Spezifikationen zu schaffen, die es erlauben, an verschiedenen Stellen in vernetzten Systemen gespeicherte Daten in unterschiedlichem Kontexten und für individuelle Fragestellungen zu nutzen.

2 Material und Methoden

Eine Hauptkomponente der Entwicklungen ist eine Ontologie der Domäne Mastschweinehaltung. Eine Ontologie ist nach Gruber eine explizite Spezifikation einer Konzeptualisierung eines Anwendungsbereichs [Gr92]. Wichtige Konzepte – wie beispielsweise Begriffe einer Domäne, Objekte oder weitere Datenentitäten – und deren Relationen werden dabei identifiziert und in formaler Form, d. h. maschinenlesbar, festgehalten. Diese explizite Formulierung geht hierbei über das Spezifizieren von Syntax oder Datenstrukturen – wie es z. B. aus der Erstellung formaler Sprachen (z. B. Programmiersprachen oder Markup Languages) oder der Modellierung von Datenbanken und Anwendungen (objektorientierte Modellierung, Entity-Relationship-Diagramme) bekannt ist – hinaus und versucht durch graphenbasierte Technologien auch die Semantik, also die Bedeutung der Konzepte zugänglich zu machen. Betrachtet wird die Wertschöpfungskette von der Ferkelerzeugung über die Mast bis zum Schlachthof. Technologien, die hierbei eingesetzt werden sind sowohl der Resource Description Framework (RDF, [KC04]) des World Wide Web Consortium als auch Topic Maps, die als ISO-Norm 13250 von der International Standards Organization gepflegt werden [ISO03]. Dabei werden sogenannte RDF Schemas [BG04] entwickelt, die sich mit bestehenden RDF-Modellen für die Landwirtschaft – insbesondere agroRDF [Ma10, MSK11] – integrieren lassen. Mittels Topic Maps wird versucht, spezielle Funktionalitäten zu realisieren, die sich in RDF nur über zusätzliche Hilfskonstrukte oder weniger intuitiv implementieren lassen (z. B. Typisierung von Ressourcen).

Bei der Modellierung wird kein vordefinierter Anwendungsfall mit bekannten Akteuren und Datenflüssen herangezogen, sondern alle relevanten Objekte sowie deren Eigenschaften und Beziehungen innerhalb der Domäne berücksichtigt. Dies schließt beispielsweise Tiere (in verschiedenen Stufen der Entwicklung), Gebäude und Anlagen, Futter-, Arznei- und Betriebsmittel etc. mit ein.

Um die Kernkonzepte und Relationen zu identifizieren, wird zurückgegriffen auf eine Reihe von implizit in Dokumenten, IT-Standards und Anwendungen existierenden Modellen. Relevante Klimaparameter oder einige Stammdaten des Tieres werden beispielsweise aus dem Agricultural Data Element Dictionary des ISOagriNet-Standards [ISO07], der zunehmend Verbreitung in der Vernetzung von Anlagen im Stall findet, abgeleitet. Außerdem werden öffentlich zugängliche Dokumente (mit teilweise gesetzlich geregelter Informationsinhalt) der Registrierungsbehörden für Tierarzneimittel und Futtermittel und Laborberichte zu Futtermittelanalysen o.ä. sowie Schlachtberichte berücksichtigt. Auf Seite existierender Anwendungen in Versuchsställen stehen Datenbankexporte zur Verfügung.

Als experimentelle Infrastruktur wird der Versuchsstall der Universität Hohenheim auf dem Unteren Lindenhof genutzt. Außerdem stehen Ergebnisse aus anderen Projekten zur Verfügung, die in die Modellierung einfließen. Gegebene Modelle werden über die übergreifende Domänenontologie verknüpft, wo notwendig werden entsprechende, formale Mappings erstellt. Im Anschluss werden Demonstratoren erstellt, die Kontextunabhängigkeit und Flexibilität des ontologiebasierten Ansatzes aufzeigen (s. Abschnitt „4 Ausblick“).

3 Ergebnisse und Diskussion

Aus Sicht der Datenmodellierung mussten sehr unterschiedliche Repräsentationen von Informationen in der Ontologie berücksichtigt werden. Einerseits müssen komplexe Dokumente wie Arzneimittelbeipackzettel und andere zusammenfassende Berichte sowie deren Inhalte in der Ontologie eingeordnet sein. Diese sollen in Dienstinfrastrukturen später ad hoc mit semantischen Strukturen angereichert werden. Auf der anderen Seite des Spektrums stehen Messdaten von Sensoren in den unterschiedlichsten Ausprägungen, die meist fest definiert und immer in denselben Strukturen ausgeliefert werden. Diese bilden z. B. die Grundlage für Aussagen über das Stallklima und Futterkonsum der Tiere. Die Beziehung zu den Tieren wird in der Regel über den Standort abgeleitet, weil man derzeit noch nicht von einer Einzeltiererkennung bei Mastschweinen ausgehen darf. Eine lückenlose Dokumentation von Zu- und Abgängen in einzelnen Stufen ist daher für die Zuordnung unabdingbar.

Ein Schwerpunkt ist der Umgang mit unterschiedlichen Methoden der Datenbereitstellung aus dem Bereich des Semantic Web. Um Wiederverwendbarkeit von allen Dienstleistungsanbietern ohne Einschränkung gewährleisten zu können, mussten bestehende Formate und Datenstandards – die gegebenenfalls verlustfrei ineinander umgewandelt werden müssen – in das Konzept miteingebunden werden. Einige Lösungen zu Detailfragen stehen zum jetzigen Zeitpunkt noch aus. So gibt es für die Repräsentation physikalischer Größen als Produkt von Wert und Einheit mehrere Möglichkeiten, die für eine automatisierte Weiterverarbeitung jeweils eigene Vor- und Nachteile aufweisen. Auch stößt der Einsatz von Hilfskonstrukten wie z.B. "Blank Nodes" in RDF bei der Konvertierung und komplexeren Datenbankabfragen auf Grenzen.

4 Ausblick

Anhand einer Beispielanwendung wird aufgezeigt, wie Daten aus verschiedenen Quellen mit unterschiedlichem Grad an fester Struktur - z. B. aus der Datenerfassung im Stall, aus Webservices, aus Webseiten etc. - zusammengeführt werden können. Die erstellten Ontologien steuern später Schnittstellen zwischen verteilten Systemen. Dabei werden Architekturmuster aus dem Umfeld des serviceorientierten Computings übernommen, die offen für neue Dienste sind, die erst zu einem späteren Zeitpunkt – vielleicht auch im Zusammenhang mit neuen Anlagen, Anwendungen oder gesetzlichen Vorschriften – verfügbar sind und die es erlauben, Dienste flexibel und verteilt zu betreiben. Nicht-funktionale Aspekte wie zum Beispiel Autorisierung und Authentifizierung, Service Discovery und Governance werden durch eine Opensource Middlewarerlösung bereitgestellt, sodass sich das Projekt auf die fachlichen Aspekte konzentrieren kann. Abbildung 1 zeigt vereinfacht und schematisch auf, wie Komponenten dabei an verschiedenen Stellen im Netz betrieben werden können. Hieraus ergeben sich besondere Anforderungen an Kontextunabhängigkeit, Flexibilität und Erweiterbarkeit der eingesetzten Datenmodelle, die nur über die oben beschriebenen Ansätze mit Hilfe von semantischen Technologien erfüllt werden können.

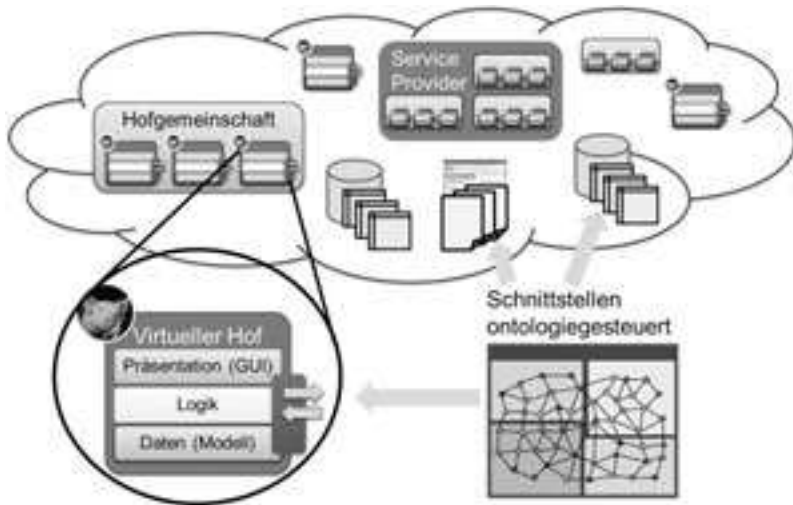


Abbildung 1: Dienste zur Datenerfassung und Auswertung können verteilt im Netz und mit verschiedenem Grad an Bündelung betrieben werden. Schnittstellen und Einlesen von Daten aus semistrukturierten/unstrukturierten Quellen werden über die entwickelte Ontologie gesteuert.

Danksagung: Die Förderung des Vorhabens erfolgt aus Mitteln des Bundesministeriums für Ernährung, Landwirtschaft und Verbraucherschutz (BMELV) aufgrund eines Beschlusses des deutschen Bundestages. Die Projektträgerschaft erfolgt über die Bundesanstalt für Landwirtschaft und Ernährung (BLE) im Rahmen des Programms zur Innovationsförderung.

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/>.
- [Gr93] Gruber, T. R.: A Translation Approach to Portable Ontology Specifications. In: Knowledge Acquisition, 5(2), 1993. S. 199-220.
- [ISO03] International Standards Organization: ISO/IEC 13250:2003, Topic Maps, 2003.
- [ISO07] International Standards Organization: ISO/IEC 17532:2007, Stationary equipment for agriculture -- Data communications network for livestock farming, 2007.
- [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/>.
- [Ma10] Martini, D.; Schmitz, M.; Kullick, R.; Kunisch, M.: Fitting information systems to the requirements of agricultural processes: a flexible approach using agroXML and linked data technologies. In: Proceedings AgEng 2010 – International Conference on Agricultural Engineering – Towards Environmental Technologies, Clermont Ferrand September 6-8, 2010.
- [MSK11] Martini, D., Schmitz, M., Kunisch, M.: Datenintegration zwischen Standards in der Landwirtschaft auf Basis saemantischer Technologien. In (Clasen, M.; Schätzel, O.; Theuvsen, B., Hrsg.): Qualität und Effizienz durch informationsgestützte Landwirtschaft – Referate der 31. GIL-Jahrestagung 24.-25. Februar 2011, Oppenheim. GI-Edition Lecture Notes in Informatics, 181, 2011. S. 133-136.