

Raum-Zeit-bezogene Agrardaten für die Anforderungen von morgen: Semantische Datenspeicherung in dezentralen, offenen Architekturen

Ansgar Bernardi, Christopher James Tuot

Deutsches Forschungszentrum für Künstliche Intelligenz – DFKI GmbH
Trippstadter Str. 122
67663 Kaiserslautern
ansgar.bernardi@dfki.de
christopher.tuot@dfki.de

Abstract: Heutige Systeme im Agrarbereich erlauben bereits, eine Fülle von orts- oder zeitbezogenen Daten zu erheben und diese in proprietären Systemen zu speichern. Der Austausch solcher Daten zwischen verschiedenen Interessenten wird jedoch durch inkompatible Datenformate, unpassende Datenstrukturen oder ungeeignete Austauschprotokolle unnötig erschwert. Mit der iGreen OnlineBox präsentieren wir eine dezentrale Plattform, die durch semantische Datenmodellierung die flexible Verwaltung anfallender Daten ermöglicht und gleich-zeitig Anforderungen nach individueller Datenhoheit erfüllt. Das vor-gestellte System nutzt etablierte Standards für Datenrepräsentation und –austausch und bietet offene Schnittstellen zum Anschluss existierender Systeme. Damit gestattet es den einfachen Aufbau zukunftssicherer Anwendungen und vernetzter Informationsstrukturen im Agrarbereich.

1 Einleitung

Im Forschungsprojekt iGreen arbeiten 24 Partner aus Wirtschaft, Wissenschaft und öffentlicher Hand an der Konzeption und Realisierung eines standortbezogenen Dienst- und Wissensnetzwerks zur Verknüpfung verteilter, verschiedener, öffentlicher, wie auch privater Informationsquellen [Be01].

Heutige Systeme im Agrarbereich erlauben bereits, eine Fülle von orts- oder zeitbezogenen Daten zu erheben und diese in proprietären Systemen zu speichern. Der Austausch solcher Daten zwischen verschiedenen Interessenten ist jedoch schwierig: Die Übermittlung von Daten eines Landwirts an seinen Berater, der Austausch mit Dienstleistern, oder die direkte Verwendung in Kommunikation mit Kunden oder Lieferanten werden durch inkompatible Datenformate, unpassende Datenstrukturen oder ungeeignete Austauschprotokolle unnötig erschwert [TS01]. Zentrale, wohlstrukturierte Datenbanken andererseits sind zwar für spezifische – auch übergreifende – Aufgaben optimiert, sind aber nur mit Aufwand an neue Anwendungen anpassbar. Ferner besteht gegenüber großen zentralen Datensammlungen unter Kontrolle von z.B. Handelsorganisationen oder

Herstellern seitens der Landwirte und Dienstleister ein erhebliches Misstrauen – welche Macht ergibt sich aus der Datenkontrolle?

Mit der iGreen OnlineBox präsentieren wir eine dezentrale Plattform, die Anforderungen nach individueller Datenhoheit erfüllt. Semantische Datenmodellierung auf Basis expliziter Vokabulare erlaubt es, jederzeit neue Datentypen und Strukturen im System abzufragen und umgekehrt neue Anwendungen automatisch an die vorhandenen Daten anzupassen.

Das vorgestellte System nutzt etablierte Standards für Datenrepräsentation und –austausch und bietet offene Schnittstellen zum Anschluss existierender Systeme. Damit gestattet es den einfachen Aufbau zukunftssicherer Anwendungen und vernetzter Informationsstrukturen im Agrarbereich.

Aus der Analyse repräsentativer Anwendungsszenarios wurden die typischen Datenquellen bestimmt, die für iGreen relevant sind. Diese sind 1. private Datenquellen der jeweiligen Landwirte, Lohnunternehmer, Berater usw. bzw. ihrer Betriebe, die unter Verantwortung ihrer Eigentümer erstellt und verwaltet werden, wie z.B. die Dokumentation durchgeführter Maßnahmen oder die Aufzeichnung von erhobenen Sensordaten der Maschinen. 2. stehen öffentliche und per Internet zugängliche Datenquellen und Dienste den jeweils berechtigten Nutzern zur Verfügung, wie etwa amtliche Geodaten oder Beratungsdienstleistungen. 3. erlauben mobile Datenquellen, wie die Terminals moderner Agrarmaschinen und die Smartphones ihrer Fahrer, die zeitnahe Erfassung konkreter Betriebsdaten und Messwerte mit aktuellem Zeit-, Raum- und Prozessbezug.

2 Architektur der iGreen Online Box – Speichern von Daten

Die OnlineBox dient der effektiven Speicherung und Erschließung der beschriebenen Daten und erlaubt, beliebige relevante Datenformate zu speichern und dabei insbesondere räumliche Daten und Fakten zu unterscheiden und jeweils adäquat zu behandeln. Konkret speichert die OnlineBox

- Binärdaten, d.h. Dateien, über deren Struktur und Inhalt nichts weiter bekannt ist. Diese Dateien werden unverändert gespeichert und wiedergegeben
- Fakten, d.h. explizit formulierte Attribute und Werte. Diese Daten werden beispielsweise aus importierten YAML-Objekten oder ISOXML-Dateien erzeugt und in Form von RDF-Triplen im TRIPLE-STORE gespeichert. Damit stehen sie für Retrieval und Reasoning zur Verfügung
- Raumbezogene Informationen, insbesondere Geodaten in allen relevanten Ausprägungen. Diese werden über den GeoServer in der darunter liegenden Datenbank abgelegt und verarbeitet.



Abbildung 1: Architektur der OnlineBox

Der OnlineBox Core ist verantwortlich für das korrekte Speichern und Retrieval dieser Datentypen, für den elementaren Datei-Import und für die Realisierung des Transaktionskonzepts. Sodann werden semantische und räumliche Abfragen ermöglicht. Eine feingranulare Zugriffssteuerung mit Nutzerauthentifizierung gewährleisten sichere Speicherung und Austausch.

Das Transaktionskonzept beruht auf dem Prinzip der Buckets. Ein Bucket wird eröffnet, schrittweise mit Daten befüllt und dann geschlossen. Erst wenn der Bucket geschlossen ist, werden die Daten als valide angesehen. Wird die Datenübertragung in einen offenen Bucket abgebrochen und nicht wieder aufgenommen, so sind die bereits im Bucket vorhandenen Daten nicht valide, die Transaktion also ungültig. Die Korrektur oder das Rollback der Transaktion bleibt dann der jeweiligen Anwendung überlassen.

Die spezifischen Importmodule erlauben die Behandlung ausgewählter Informationstypen. Zur Zeit sind die Module für YAML (zum Import beliebiger Attribut/Wert-basierter Objekte) und für ISOXML realisiert. Für den Import von raumbezogenen Informationen, Kartenmaterial und ähnlichen Geodaten sind Module für den Import des GML- und des SHAPE-Formats vorgesehen, aber noch in Arbeit.

Das Security-Modul prüft für jeden Zugriff auf die Datenbank zunächst die Nutzerauthentifizierung; sodann gleicht es die gewünschten Zugriffe gegen die datenspezifische Access Control List ab, mit der einzelnen Nutzern oder Nutzergruppen der Zugriff auf einzelne Datenelemente genehmigt oder verwehrt wird.

Schließlich stellt die OnlineBox Schnittstellen für externe Anwendungen und Bedienung bereit: Die http/html Schnittstelle erlaubt die Bedienung der OnlineBox mit normalen Webbrowsern; die http/REST API ist für die Entwicklung eigenständiger Clients vorgesehen. Die Basis des Datenaustauschs ist in beiden Fällen das http-Protokoll.

3 Exemplarische Clients im FrontEnd

Zur Bedienung der OnlineBox und zur Interaktion mit den gespeicherten Daten müssen nach Bedarf geeignete Clients implementiert werden. Im Moment werden zwei beispielhafte Anwendungen bereitgestellt: Eine Browser-Oberfläche nutzt das http/html-Interface, um Transaktion, Informations-Upload, Retrieval, Daten-Browsing und die Visualisierung einfacher Geoinformationen zu demonstrieren; während ein exemplarischer Validierungs-Client für ISOXML-Dateien die http/REST-Schnittstelle der OnlineBox verwendet.

4 Vernetzte Knoten unter Verantwortung individueller Betreiber

Die OnlineBox wird zusammen mit anderen Diensten im so genannten iGreen-Knoten realisiert und auf einem (logischen) Rechner installiert, der unter Verantwortung eines Knoten-Betreibers steht.

Beispiel: Ein Lohnunternehmer betreibt für seinen Betrieb einen iGreen-Knoten. (Es ist dabei durchaus denkbar, dass die verschiedenen Dienste real auf mehrere Rechner verteilt werden, wenn das z.B. aus Gründen der Performanz sinnvoll erscheint).

Ein Adressierungsschema erlaubt die eindeutige Bezeichnung jedes Knotens und jedes Dienstes. Über Maschinen-Connectoren werden mobile Endgeräte (Terminals der Landmaschinen oder Smartphones der Fahrer) mit den Knoten und damit mit der OnlineBox verbunden. Mehrere Knoten werden in Domänen zusammengefasst und so zu Netzen verbunden.

Literaturverzeichnis

- [Be01] A. Bernardi: iGreen: Organisationsübergreifendes Wissensmanagement in öffentlich-privater Kooperation. In: Automatisierung und Roboter in der Landwirtschaft. KTBL-Tage-2010, Erfurt, Germany.
- [TS01] C. Tuot; W. Schneider: Semantische Technologien für ein öffentlich-privates Wissensmanagement im Agrarbereich. GIL Jahrestagung. Gesellschaft für Informatik in der Land-, Forst- und Ernährungswirtschaft 2010