

Datenaufbereitung in der Landwirtschaft durch automatisierte semantische Annotation

Basis für vielfältige Smarte Dienste

Julian Klose¹, Markus Schröder², Silke Becker¹, Ansgar Bernardi²
und Arno Ruckelshausen¹

Abstract: Die moderne Landtechnik und zunehmende Digitalisierung landwirtschaftlicher Prozesse liefern vielfältige Daten. Deren effiziente und nutzbringende Verwendung leidet jedoch unter berechtigten Sorgen um Datenhoheit und -kontrolle, Formatbrüchen und unterschiedlichsten Interpretationen. Als Lösungsvorschlag präsentieren wir die sog. Wikinormia, eine kollaborative Plattform, in der interessierte Teilnehmer eigene, neue Datenformate beschreiben und diskutieren können. Sobald ein finalisiertes Vokabular erstellt ist, können spezifische Parser die Rohdaten in drei grundsätzliche Repräsentationen semantisch aufbereiten: Geo-Informationen, Zeitreihen und semantische Fakten (landwirtschaftlicher Wissensgraph). Dank der öffentlich zugänglichen Definitionen und Beschreibungen gewinnen Entwickler leicht Überblick über die für sie relevanten Konzepte. Vielfältige Dienste werden dann (vorbehaltlich individueller Zugriffsrechte) ihre Daten einfach über eine Query-Schnittstelle abfragen und Ergebnisse zurückspielen können. Diesen Lösungsvorschlag haben wir bereits in einem Prototyp im Projekt SDSD (Smarte Daten – Smarte Dienste) umgesetzt. Wir demonstrieren den Nutzen mit einer Reihe von repräsentativen Diensten aus der Landwirtschaft. Damit steht ein effizientes System zur kooperativen, flexiblen Digitalisierung landwirtschaftlicher Arbeitsabläufe zur Verfügung.

Keywords: Datenformate, Datenplattform, Datenverfügbarkeit

1 Einleitung

Die Landwirtschaft ist mit stetig wachsenden Anforderungen wie Dokumentation, Rückverfolgbarkeit, Prozessoptimierung und ressourcenschonendes Arbeiten konfrontiert. Zeitgleich findet eine stete Entwicklung der quantitativen und qualitativen Leistungsfähigkeit von Landmaschinen statt und führt zu einer steigenden Verfügbarkeit von Daten, die i.d.R. hersteller- bzw. maschinenspezifisch abgebildet sind. Der herstellerübergreifende Datenaustausch und die korrekte Interpretation und Nutzung der verschiedenen Datenformate und Inhalte stellen den Landwirt vor große Herausforderungen – gerade auch angesichts der dynamischen Weiterentwicklung.

¹ Hochschule Osnabrück, Fakultät IuI, Sedanstraße 26, 49076 Osnabrück, julian.klose@hs-osnabrueck.de; s.becker.1@hs-osnabrueck.de; a.ruckelshausen@hs-osnabrueck.de

² Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Trippstadter Straße 122, 67663 Kaiserslautern, markus.schroeder@dfki.de; ansgar.bernardi@dfki.de

Um diesen Anforderungen gerecht zu werden, haben wir im Forschungsprojekt *Smarte Daten – Smarte Dienste (SDSD)* ein System entwickelt, das landwirtschaftlich angereicherte Prozessdaten mit intelligenten Diensten verbindet. Unser Ansatz speichert dafür Informationen aus verschiedenen landwirtschaftlichen Quellen eines Landwirts. Durch das Aufstellen von Speicher- und Zugriffsregeln behält der Nutzer die Datenhoheit und Datenkontrolle.

Eine große Herausforderung liegt in der Erkennung und Interpretation der unterschiedlichen Datenformate und -inhalte. Standardisierung dauert oft zu lange, um sich dynamisch auf neue Anforderungen einzustellen. Hierfür bietet SDSD eine Plattform zum kollaborativen Beschreiben von Datenformaten an, die wir *Wikinormia* nennen. Durch diese offene Plattform erreichen wir eine hohe Erweiterbarkeit, die mit der laufenden Entwicklung neuer Datenformate Schritt hält. Der Einsatz von *Semantic Web Standards* macht die Verlinkung der modellierten Informationen zu einem Wissensgraphen als *Smarte Daten* möglich.

Neben der Datenaufbereitung ist SDSD auch eine Plattform zur Bereitstellung von Diensten, die aus den *Smarten Daten* Mehrwerte erzeugen, wie z. B. automatisierte Dokumentation, Berechnungen und Applikationshilfen, oder langfristige Auswertungen. Dienste können aus den verknüpften Daten gezielt die benötigten Informationen abrufen, ohne mit den unterschiedlichen Formaten der Rohdaten in Berührung zu kommen. Dieses System aus der Bereitstellung von *Smarten Daten* für *Smarte Dienste* zeigt erste Erfolge für eine nutzbringende Digitalisierung landwirtschaftlicher Prozesse.

2 Stand der Forschung

In diesem Gebiet gibt es bereits Teillösungen und verwandte Forschungsprojekte. Industrielle Lösungen wie *MyJohnDeere* [My19] und *365FarmNet* [Di19] sind Plattformen zur Übermittlung und Speicherung von Maschinendaten und Auftragsdaten. Diese Plattformen sind allerdings rein proprietär nur mit den jeweiligen eigenen Maschinen kompatibel. Bei dem diversen Einsatz von Maschinen unterschiedlicher Hersteller ist ein herstellerübergreifendes System unumgänglich.

Der *agrirouter* [DK19] von DKE bietet eine offene Datenübertragungsplattform, wodurch es möglich ist, Auftragsdaten und Echtzeit-Telemetrie-Daten von den Maschinen an angeschlossene Applikationen zu übertragen bzw. Daten an die Maschinen zu senden. Die Datenpakete werden aber vom *agrirouter* nicht geöffnet und bleiben auch nur zu Übertragungszwecken kurzfristig gespeichert. Mit der DKE als Partner in SDSD haben wir erfolgreich den *agrirouter* genutzt, um die Kommunikation zu und vom SDSD System zu vereinfachen.

Andere Forschungsprojekte, wie *GeoBox* [GE19], *ODiL* [OD16] und *OPeRate* [OP19] arbeiten an ähnlichen Themen zur landwirtschaftlichen Datenhaltung und Bereitstellung.

SDSD unterscheidet sich von diesen Projekten unter anderem durch die offene erweiterbare Formatbeschreibung mittels Wikinormia. Des Weiteren kooperiert SDSD zusammen mit dem kürzlich gestarteten ATLAS-Projekt [AT19].

3 Ansatz und Demo

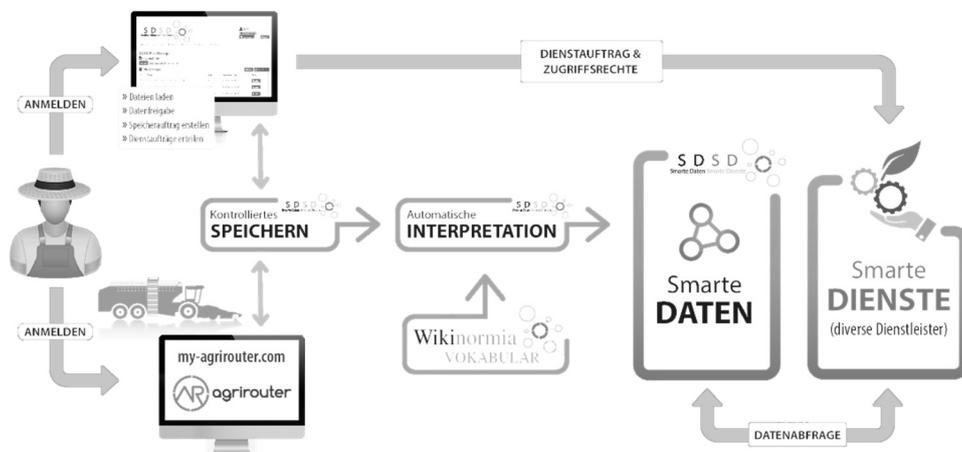


Abb. 1: Übersicht der SDSD Komponenten: Über agrirouter oder unser Webinterface ist ein kontrolliertes Speichern von Daten möglich, die durch Wikinormia Vokabular automatisch zu Smarten Daten interpretiert werden. Smarte Dienste fragen die aufbereiteten Daten ab

Unser Hauptansatz ist die Erzeugung von Smarten Daten sowie deren Bereitstellung für Smarte Dienste. Um die eingegangenen Rohdaten in Smarte Daten umzuwandeln, muss zunächst das Format beschrieben werden. Hierzu bietet SDSD die Wikinormia als Wiki-ähnliche Plattform: Ein Format wird hierbei durch das Resource Description Framework [RD14] (RDF) in Form eines Graphen definiert. Diese Definition sollte durch die Datenprovider erfolgen. Bei öffentlichen zugänglichen Formaten kann das auch von Dienste-Entwicklern übernommen werden. Damit möchten wir keine Standardisierung ersetzen, sondern ausschließlich den Entwicklungsprozess deutlich beschleunigen. In unserer prototypischen Wikinormia wurden bereits das weit verbreitete ISOXML [AE15], aber auch spezifischere Datenformate wie NRW Agrarantragsdaten (siehe Abb. 2), beschrieben.

Teilschlag NRW

Teilschlag aus einem NRW Flächenantrag (ELAN).

Identifizier	Part of
teilschlag	Flächenantrag NRW (antragNRW)
Format	Subclass of
Flächenantrag NRW (antragNRW)	Field (field)
Name	Type
Year of application (antragsjahr)	Integer

Abb. 2: Auszug aus Wikinormia-Eintrag für NRW Agrarantrag

Beim Einfügen eines neuen Datenformats in der Wikinormia muss zusätzlich ein Parser übergeben werden, der das Datenformat aus den Rohdaten auslesen kann. Dies ist wegen der Vielzahl möglicher Übertragungsformate (z. B. JSON, XML, Binär) nötig. In ausgewählten Fällen könnte basierend auf der Wikinormia-Beschreibung ein Parser automatisch generiert werden.

SDSD unterscheidet zwischen drei Datenmodellen und nutzt hierfür spezialisierte Datenbanken (Polyglotter Speicher [FS12]):

1. Geodaten umfassen positionierte Geometrien, wie Fahrspuren oder Feldgrenzen. Sie werden in einer MongoDB [Mo19] mit raumbezogener Indexierung abgelegt. Dies ermöglicht die Suche nach Überschneidungen oder Elementen in räumlicher Nähe.
2. Zeitreihen enthalten zeit- und positionsbezogene Sensordaten, wie Maschinendrehzahlen oder spezifische Ausbringmengen. Sie werden aufgrund ihrer Menge in einer Cassandra Datenbank [Ap19] gespeichert, die für das Ablegen großer Tabellen mit unterschiedlichen Spalten optimiert ist.
3. Semantische Graphen enthalten verknüpfte Informationen, wie Maschinenbeschreibungen oder Aufträge. Sie werden in einem Stardog Triplestore [St19] abgelegt. Dadurch können Beziehungen zwischen Informationen strukturiert abgefragt werden, wie beispielsweise „Alle Aufträge, in denen eine Sämaschine genutzt wurde“.

Instanzen (Maschinen, Aufträge, Felder, Personen, etc.) bekommen im RDF-Graphen eindeutige Adressen (URIs). Dadurch ist es möglich, Geodaten und Zeitreihen mit den

Instanzen zu verbinden sowie mit der entsprechenden Beschreibung in der Wikinormia zu verknüpfen. Die Verlinkung macht es leicht, Zusatzinformationen direkt an eine Instanz zu annotieren, z. B. Wetterinformationen an ein Feld. Auch die Deduplikation von gleichen Instanzen aus verschiedenen Quellen wird vereinfacht. Wenn beispielsweise ein Feld, welches bereits aus einem Flächenantrag bekannt ist, in einer Auftragsdokumentation (ISOXML TaskData) erneut auftritt, kann das System die beiden Feld-Instanzen mit einer SameAs-Beziehung verlinken. Die Gleichheit basiert hierbei auf den Geodaten.

Die erzeugten Smarten Daten können von Smarten Diensten abgerufen werden. Entwickler von Smarten Diensten müssen sich nicht mehr mit der Integration verschiedener Datenformate oder den Besonderheiten einzelner Datenquellen auseinandersetzen. Über die SDSD-Schnittstelle können ausgewählte Informationen abgerufen werden, die der Dienst benötigt. Dies ermöglicht es auch kleinen Anbietern, kostengünstig ihre Dienste bereitzustellen. Zusätzlich kann der Landwirt genau spezifizieren, auf welche Informationen der Dienst zugreifen darf, und kann damit vertrauliche Daten schützen.

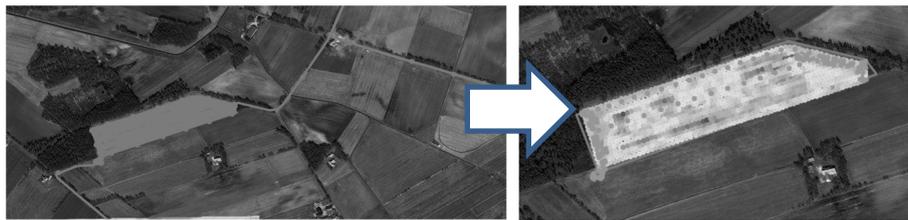


Abb. 3: Zeitreihen-Trennungs-Dienst: links eine zusammenhängende Zeitreihe, rechts ein automatisch separiertes Feld mit einer Ertragskartierung

Der Zeitreihen-Trennungs-Dienst in Abb. 3 löst ein typisches Problem bei aufgezeichneten Maschinendaten: Neben dem eigentlichen Feld sind noch die Straßenfahrten und die Bearbeitung anderer Felder im Auftrag abgebildet. Eine solche Aufzeichnung entsteht, wenn die Aufträge vom Fahrer nicht ordnungsgemäß gestartet und beendet werden. Der Beispieldienst liest solche Daten aus und überlagert sie mit bereits bekannten Feldgrenzen, um die relevanten Informationen zu separieren. Das SDSD-System kombiniert dabei Feld-Informationen aus verschiedenen Quellen und stellt sie zur Abfrage über eine geografische Bereichssuche zur Verfügung. Falls die Felder bisher noch nicht bekannt sind, werden noch Feldgrenzen aus OpenStreetMaps [Ov19] berücksichtigt.

Ein Prototyp steht unter <https://app.sdsd-projekt.de> mit dem Account „sdsd“ und Passwort „sdsd“ zum Testen zur Verfügung.

4 Schluss und Ausblick

Um die Prozesse in der Landwirtschaft zunehmend digital abzubilden, brauchen wir ein gemeinsames Vorgehen aller Beteiligten. SDS D bietet eine Plattform zur Kommunikation zwischen Datenerzeugern, Landwirten und Anbietern von Diensten.

Ein Lösungsansatz für die anstehenden Herausforderungen der Digitalisierung in der Landwirtschaft kann die oben beschriebene SDS D-Plattform sein. Diese beinhaltet a) die Beschreibung von Datenformaten in der Wikinormia, b) das Auslesen durch Parser, c) die semantische Aufbereitung und d) die kontrollierte Bereitstellung der Informationen für Dienste.

Durch die zielgerichtete Datennutzung und Datenintegration werden kurz- und mittelfristig Bewirtschaftungsstrategien optimiert. Das kann sich positiv auf Prozessketten und Betriebsmitteleinsatz auswirken. Langfristig betrachtet baut sich der Landwirt durch die Verwendung der SDS D-Plattform eine Anbauhistorie auf.

Literaturverzeichnis

- [RD14] RDF Working Group, „Resource Description Framework (RDF)“, 2014.
- [AE15] AEF, „ISO 11783“, 2015.
- [FS12] Fowler, M.; Sadalage, P. J.: „Introduction to Polyglot Persistence“, informIT, 2012.
- [My19] MyJohnDeere Anmeldung, <https://myjohndeere.deere.com/>, 01.12.2019.
- [Di19] Digitalisierung der Landwirtschaft 4.0, <https://www.365farmnet.com/de/>, 01.12.2019.
- [DK19] DKE agrirouter | Your Farming Network, <https://my-agrirouter.com/>, 01.12.2019.
- [GE18] GEOBOX - Standardisierung der Geobox-Infrastruktur, <https://www.dfki.de/web/forschung/projekte-publikationen/projekte/projekt/geobox/>, 01.12.2019.
- [OD16] ODiL - Projekte, <https://saat.dfki.de/de/projekte/odil.html>, 01.12.2019.
- [OP19] OPeRAte, <http://operate.edvsz.hs-osnabrueck.de/>, 01.12.2019.
- [AT19] ATLAS Project, <https://cordis.europa.eu/project/rcn/223982/factsheet/en>, 01.12.2019.
- [Mo19] MongoDB, <https://www.mongodb.com/>, 01.12.2019.
- [Ap19] Apache Cassandra, <https://cassandra.apache.org/>, 01.12.2019.
- [St19] Stardog: The Enterprise Knowledge Graph, <https://www.stardog.com/>, 01.12.2019.
- [Ov19] Overpass API, https://wiki.openstreetmap.org/wiki/Overpass_API, 01.12.2019.