

# Planungsdaten schnell finden und einfach nutzen: Linked Open Data und semantische Suche im Einsatz für das KTBL-Datenangebot

<sup>1)</sup>Daniel Martini, <sup>1)</sup>Martin Kunisch, Daniel Herzig, Günter Ladwig

<sup>1)</sup>Kuratorium für Technik und Bauwesen  
in der Landwirtschaft e. V. (KTBL)  
Bartningstraße 49  
64289 Darmstadt  
d.martini, m.kunisch@ktbl.de

**Abstract:** Um Entscheidungen fundiert treffen zu können, sind Daten als Grundlage unerlässlich. Der Aufwand für die Recherche nach relevanten Daten sowie die anschließende Aufbereitung der Daten für die Verwendung in Kalkulationswerkzeugen stellen die wesentlichen Herausforderungen für datenbasierte Entscheidungen dar. In diesem Beitrag wird eine Lösung zur Unterstützung von datengestützten Entscheidungen in der Agrarwirtschaft vorgestellt, die die Recherche nach Planungsdaten aus dem KTBL-Datenangebot durch den Einsatz einer semantischen Suchmaschine vereinfacht und durch die Bereitstellung der KTBL-Daten gemäß der Linked Open Data Prinzipien eine leichte Weiterverarbeitung ermöglicht.

## 1 Einleitung

In der Landwirtschaft werden täglich Entscheidungen über Investitionen, Arbeitsplanungen und Produktionsabläufe anhand von Daten und darauf aufsetzenden Berechnungen gefällt. Wenn beispielsweise ein Landwirt überlegt, in den Anbau von Weizen einzusteigen und dafür die Investition für neue Maschinen abwägen möchte, konnte er in der Vergangenheit auf umfangreiche Datensammlungen des Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V. in gedruckter Form zurückgreifen, wie beispielsweise die Faustzahlen Landwirtschaft [K09] oder die Datensammlung Betriebsplanung [K12]. Bereits seit einigen Jahren spielt das Internet auch im Agrarbereich eine zunehmende Rolle bei der Veröffentlichung und Verbreitung von Daten. Bislang waren online angebotene Daten eng integriert mit ihren Anwendungen und dem Nutzer lediglich über vorgefertigte, von den jeweiligen Entwicklern in den Bedienoberflächen festgelegte Interaktionswege zugänglich. Eine Weiternutzung und Verarbeitung der Daten in anderen Systemen wie z. B. Farmmanagement-Informationssystemen oder Beratungswerkzeugen oder ein Zusammenführen von Datenbeständen verschiedener Organisationen zur Beantwortung neuer Fragestellungen war mithin nur erschwert möglich und erforderte aufwändige Arbeiten für Einpflegen oder Import.

Technologien im Umfeld des Semantic Web erlauben inzwischen für die Datennutzer elegantere und flexiblere Lösungen. So wird im Rahmen der Linked Open Data (LOD) Initiative des World Wide Web Consortium versucht, Daten von Auswertungslogik und Benutzerschnittstelle zu trennen und mit einfachen Internet-Protokollen (in erster Linie HTTP) zugänglich zu machen [Be06, HB11]. Insbesondere werden die Daten hier auch maschinenlesbar in standardisierten Formaten bereitgestellt, sodass diese über einfache URL-Aufrufe abgerufen und automatisiert in Anwendungen eingelesen werden können.

Für die Repräsentation der Daten setzt man auf das Resource Description Framework (RDF), sowie Vokabulare, die z. B. in RDF Schema [BG04] erstellt werden können. Kernaspekt ist dabei die Modellierung von Daten als gerichteter Graph. In diese allgemeine und generische Datenstruktur können Datensätze jeglicher Art meist einfach überführt werden. Außerdem werden durch diese Art der Repräsentation Beziehungen zwischen Daten explizit modelliert und spezifiziert, sodass Datensätze besonders zugänglich werden für Suche, Navigation und die Beantwortung komplexer Fragestellungen, die die Einbeziehung einer Reihe von Bezügen erfordern.

Im Rahmen des iGreen-Projektes und einer langjährigen Kooperation mit der FAO konnten die Autoren grundlegende Entwicklungen in diesem Bereich vorantreiben. Um die Daten des KTBL gemäß dieser Prinzipien zugänglich zu machen, wurden das Vokabular agroXML<sup>1</sup> sowie Ergebnisse aus dem iGreen-Projekt<sup>2</sup> als Grundlage zur Beschreibung verwendet.

Aufbauend auf diesen Vorarbeiten hat das KTBL nun auf Basis der vorhandenen KTBL-Datenbank mit Informationen zu Arbeitsverfahren und Maschinenkosten einen Linked Open Data Service aufgesetzt, der den maschinellen Abruf dieser Daten erlaubt. Oben auf wurde von SearchHaus, einem Spin-Off des Karlsruher Institut für Technologie (KIT), eine intelligente semantische Suche eingerichtet, die ein gezieltes Auffinden von Datensätzen ermöglicht. Wird beispielsweise die Frage untersucht, wie hoch die Investitionskosten für Maschinen sind, die zur Aussaat von Weizen verwendet werden, so kann mit einfach Stichwortanfragen, z.B. „anschaffungspreis maschinen weizen saat“, der KTBL-Datenbestand durchsucht werden. Die Anfragen können frei formuliert werden und es sind keine speziellen Kenntnisse über Aufbau oder Abfrage der Datenbank notwendig. Für mögliche Treffer werden verschiedene Interpretationen, die anhand der in den Daten vorhandenen Beziehungen ermittelt werden, angezeigt, Ergebnisse können anschließend über dynamisch berechnete Facetten eingeschränkt werden.

## 2 Material und Methoden

Der Linked Open Data Service wurde mit Hilfe des Open Source Werkzeugs d2rq (<http://d2rq.org>) auf die bestehende Oracle 11 Datenbank aufgesetzt. Hierfür wurde mit der d2rq mapping language ein Mapping erstellt, das spezifiziert, wie relationale Datenbanktabellen in ein RDF Graphen-Modell zu überführen sind. Als Ergebnis steht sowohl

---

<sup>1</sup> <http://www.agroxml.de>, abgerufen am 14.10.2013

<sup>2</sup> <http://www.igreen-projekt.de>, abgerufen am 14.10.2013

ein http-Service, der Daten sowohl im HTML-Format für den Browser als auch im maschinenlesbaren Turtle-Format [PC13] für RDF ausliefern kann, sowie ein SPARQL-Endpoint [HS13] für formal spezifizierte Abfragen zur Verfügung. Neben den in der lokalen Datenbank vorhandenen Daten wurden in den Datensatz auch Verknüpfungen zu dem multilingualen AGROVOC Thesaurus der FAO<sup>1</sup> eingebaut, sodass für einige Maschinenarten die entsprechenden Konzepte dort sowie die Übersetzungen der Bezeichnung in eine Vielzahl von Sprachen abgerufen werden können. Die semantisch aufbereiteten Daten bilden die Grundlage für die Suche.

Mit der semantischen Suche von SearchHaus können die Benutzer einfach in den KTBL-Daten mit Schlüsselwörtern suchen, so wie sie es von Websuchmaschinen gewohnt sind. Die Suchmaschine nimmt Schlüsselwortanfragen entgegen, interpretiert die Anfragen mit Hilfe des Datenbestands und liefert dem Benutzer die passenden Ergebnisse aus den KTBL-Daten. Intern aggregiert und gruppiert die Suchmaschine die komplexen, strukturierten Daten, so dass eine schnelle Beantwortung der Anfragen auch über große Datenbestände möglich ist. Auf dieser kompakten Datenrepräsentation werden dann mittels Explorationsalgorithmen und Rankingverfahren die Suchergebnisse ermittelt [LT10]. In einem zweiten Schritt kann der Benutzer das Suchergebnis interaktiv mittels automatisch berechneter Facetten einschränken und so genau an sein Informationsbedürfnis anpassen. Der Vorteil der Suche gegenüber bestehenden Lösungen ist, dass die Suche die Zusammenhänge in den Daten erkennt und für die Berechnung der Ergebnisse ausnutzen kann. Abbildung 1 zeigt einen Ausschnitt aus den Ergebnissen für die Suchanfrage nach Anschaffungspreisen von Maschinen, die für die Aussaat von Weizen geeignet sind (als Stichwortanfrage: “maschine preis weizen saat”).

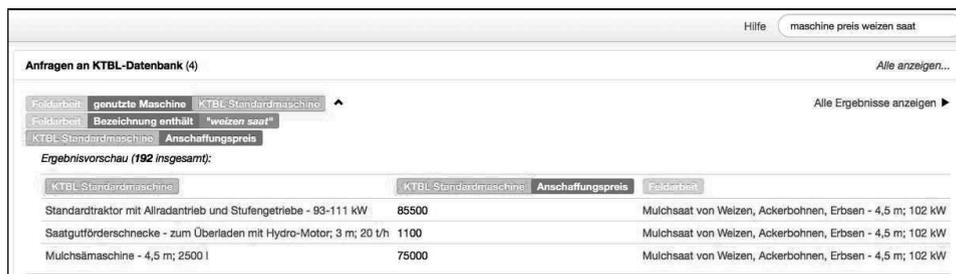


Abbildung 1: Ausschnitt eines Screenshots. Zu sehen sind Ergebnisse für die Suchanfrage nach Maschinen und deren Anschaffungspreis, die für die Aussaat von Weizen geeignet sind.

### 3 Ergebnisse

Sowohl der Linked Open Data Service als auch die semantische Suche ließen sich sehr einfach in die bestehende Infrastruktur integrieren. Auf Datenbankebene waren keine Anpassungen an bestehenden Zugangsmechanismen, Datenbankschemas oder Tabellen notwendig. Mittelfristig kann die Implementation durch die Erstellung zusätzlicher Views optimiert werden, ein zwingender Bedarf ergab sich hierfür jedoch im prototypi-

<sup>1</sup> <http://aims.fao.org/standards/agrovoc/about>, abgerufen am 14.10.2013

schen Betrieb nicht. Mit überschaubarem Aufwand konnte innerhalb eines Zeitraumes von etwa drei Monaten unter Aufwendung von etwa 2-3 Personenmonaten eine vollständig spezifikationsgemäß funktionierende Implementation erzielt werden.

Derzeit startet der interne Probetrieb und die ersten Indikatoren lassen darauf schließen, dass die KTBL-Mitarbeiter von den Möglichkeiten, sich schnell einen Überblick über vorhandene Daten verschaffen zu können, profitieren. Es ist zu erwarten, dass Berater und Landwirte innerhalb des KTBL-Datenangebotes schneller und zielgerichteter Anwendungen, Daten und Veröffentlichungen finden, die für ihre Fragestellungen Antworten liefern können. Zusätzlich bekommen sie bei weiterer Verbreitung der Technik weitere Informationen anderer Anbieter, die zu der Anfrage passen, angezeigt. Die maschinenlesbare Schnittstelle kann von Entwicklern bei Herstellern von Farmmanagement-Informationssystemen und Entscheidungsunterstützungssystemen genutzt werden, um neue Funktionalitäten zu implementieren. Auch seitens des KTBL können so neue Anwendungen durch Daten externer Einrichtungen angereichert werden.

In Detailpunkten wurden auch Limitationen gefunden. So ist es beispielsweise in der d2rq mapping language nicht möglich, in RDF übliche Angaben zur Sprache an Bezeichnungen, die aus Inhalten mehrerer Datenbanktabellen zusammengesetzt wurden, anzuhängen. Auch die Abbildung, Interpretation und Sortierung von physikalischen Einheiten hat noch Raum für Verbesserungen.

Insgesamt stellen die Arbeiten einen zukunftsweisenden Ansatz zur Etablierung von innovativen und flexibleren Lösungen zur Nutzung des KTBL-Datenangebots und insbesondere zur Beantwortung von Fachfragen dar, die langfristig die integrierte Nutzung von Daten, die in verschiedenen Organisationen erhoben wurden, erlauben.

## Literaturverzeichnis

- [Be09] Berners-Lee, T.: Linked Data. World Wide Web Consortium, 2009. <http://www.w3.org/DesignIssues/LinkedData.html>, aufgerufen am 10.10.2013.
- [BG04] Brickley, D., Guha, R. V.: RDF Vocabulary Description Language 1.0. World Wide Web Consortium, 2004. <http://www.w3.org/TR/rdf-schema/>, aufgerufen am 10.10.2013.
- [HB11] Heath, T., Bizer, C.: Linked Data – Evolving the Web into a Global Data Space. Morgan & Claypool Publishers, 2011.
- [HS13] SPARQL 1.1 Query Language. World Wide Web Consortium, 2013. <http://www.w3.org/TR/sparql11-query/>, aufgerufen am 10.10.2013.
- [K09] Kuratorium für Technik und Bauwesen in der Landwirtschaft (Hrsg.): Faustzahlen für die Landwirtschaft (14. Auflage). KTBL e. V., Darmstadt 2009.
- [K12] Kuratorium für Technik und Bauwesen in der Landwirtschaft (Hrsg.): Betriebsplanung Landwirtschaft 2012/13 – Daten für die Betriebsplanung in der Landwirtschaft (23. Auflage). KTBL e. V., Darmstadt, 2012.
- [LT10] Ladwig G., Tran T.: Combining Keyword Translation with Structured Query Answering for Efficient Keyword Search, Proceedings of the 7th Extended Semantic Web Conference (ESWC '10), Springer, 2010.
- [PC13] Prud'hommeaux, E., Carothers, G.: Turtle – Terse RDF Triple Language. World Wide Web Consortium, 2013. <http://www.w3.org/TR/turtle/>, aufgerufen am 10.10.2013.