

Linked Open Data im Pflanzenschutz

Maschinenlesbare Bereitstellung der notwendigen Informationen zu Indikationen und Abstandsauflagen

Katharina Albrecht¹, Daniel Martini¹ und Mario Schmitz¹

Abstract: Die Berücksichtigung von Abstandsauflagen ist essenziell für die richtige Anwendung von Pflanzenschutzmitteln in der guten fachlichen Praxis. Allerdings verlangt die Einhaltung dieser Auflagen ein erhöhtes Maß an Sorgfalt und eine regelmäßige Kontrolle der Auflagentexte. Um den Landwirt bei der sachgerechten Anwendung von Pflanzenschutzmitteln zu unterstützen, wurde im PAM-Projekt das Entscheidungshilfesystem für den Landwirt entwickelt. Besonders die Extraktion der Abstandsaufgaben litt unter einer hohen Anzahl von Medienbrüchen, die in der vorliegenden Überarbeitung größtenteils entfernt werden konnten. Um den Abstandsservice produktiv bereitzustellen, wurden standardisierte Tools und Standardvokabular verwendet. Für den bereitgestellten Abstandsservice wurde das Gesamtdatenmodell überarbeitet und der Webservice angepasst.

Keywords: Semantic Web, SPARQL, Linked Data, Pflanzenschutz

1 Einleitung

Im Kontext der Digitalisierung in der Landwirtschaft werden Ansätze des Semantic Web und Linked Open Data (LOD) immer wichtiger. Sie ermöglichen maschinen-, aber auch menschenlesbare Repräsentationen von Daten und unterstützen das Verständnis der Informationen auf der Anwender- sowie die Kommunikation auf der technischen Seite. Durch die Bereitstellung sogenannter Triplestores, also Graphdatenbanken, welche über Webservices abgefragt werden können, wird eine nicht-proprietäre, flexibel erweiterbare Datenhaltung ermöglicht. Dies vereinfacht den Zugriff auf die notwendigen Informationen und erlaubt den Datenbestand durch Integration weiterer Daten für andere Anwendungsfälle zu nutzen.

Für jede vom Bundesamt für Verbraucherschutz und Lebensmittelsicherheit (BVL) freigegebene Pflanzenschutzanwendung existieren Abstandsaufgaben zu Saumstrukturen, Oberflächengewässern und Siedlungen. Jeder Landwirt muss für jeden zu bearbeitenden Schlag diese Abstandsaufgaben ermitteln und in seiner Applikation berücksichtigen. Dies geschieht händisch und ist damit nicht nur aufwendig, sondern auch fehleranfällig, und verlangt damit ein erhöhtes Maß an Sorgfalt und regelmäßiger Überprüfung hinsichtlich Änderungen bei Ablauf des Genehmigungszeitraums, Neuzulassungen oder Änderungen

¹ Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V., Datenbanken und Wissenstechnologien, Bartningstraße 49, 64298 Darmstadt, k.albrecht@ktbl.de

der Gesetzesentwürfe. Die indikations- und/oder mittelbezogenen Abstandsaufgaben finden sich in der vom BVL bereitgestellten Datenbank.

Für den hier präsentierten Anwendungsfall des Abstandsservice wird diese Datenbank aufbereitet und hierfür das Resource Description Framework (RDF) verwendet, welches mit Hilfe sogenannter Triple bestehend aus Subjekt, Prädikat und Objekt die verschiedensten Datensätze beschreiben kann [Sp15]. RDF eignet sich hervorragend, um relationale Datensätze in ein Semantic-Web-Datenmodell zu übertragen, da hier ein direkter Zusammenhang existiert. [Be09] beschreibt das wie folgt: Das Subjekt ist der Zeileninhalt einer Tabelle der Datenbank, das Prädikat ist abgeleitet vom Spaltennamen und das Objekt der Zellinhalt, allerdings ermöglicht RDF nicht nur eine 1:1-Übertragung der Tabelleninhalte, sondern auch die Definition oder Ableitung von neuen, komplexeren Zusammenhängen, die in Relationalen Datenbanken nicht ohne Weiteres abgebildet werden können.

Das Ergebnis wird dann mithilfe eines Webservices in verschiedenen Formaten entweder weiteren Diensten in maschinenlesbarer Form – beispielsweise zur Berechnung von Applikationskarten – oder dem Anwender über den Webbrowser bereitgestellt, sodass sich auch Landwirte hierüber über Abstandsaufgaben informieren können.

2 Eine Graphdatenbank für Pflanzenschutzmittel

Ein wichtiger Bestandteil des Pesticide Application Managers (PAM), des Entscheidungshilfesystems für Pflanzenschutzmittel-Applikationen [Sc16], ist der vom KTBL bereitgestellte Abstandsservice, welcher die Informationen zu Indikationen, Mitteln, Abständen und Auflagen aus der Pflanzenschutzmittel-Datenbank extrahiert und maschinenlesbar und plattformunabhängig bereitstellt. In [Ma15] und [Ma18] werden die verschiedenen Ansätze und Konzepte dieses Dienstes beschrieben. Besonders im Bereich des Mappings von relationaler Datenbank zur Semantischen Repräsentation gab es in [Ma15] jedoch noch Limitierungen hinsichtlich der vorhandenen und nutzbaren Werkzeuge, sodass das zum R2RML-Standard des W3C [DSC12] konforme Programm db2triples nicht genutzt werden konnte und extra für die Anwendung ein über mehrere Zwischenschritte laufender Konvertierungsablauf implementiert werden musste. Der im Vorgängerprojekt prototypisch programmierte Abstandsservice wurde daher mit dem Ziel, Prozesse zu vereinfachen und die Umsetzung robuster gegenüber Fehlersituationen zu machen, komplett überarbeitet.

2.1 Datenaufbereitung

Als beste Wahl für die Umsetzung des Mappings von relationaler Datenbank zu RDF-Modell wurde in einer am KTBL durchgeführten Masterarbeit das Werkzeug db2triples identifiziert. Allerdings war db2triples bisher nicht mit der verwendeten Datenbank koppelbar, sodass zunächst diese Anbindungen in den Quellcode integriert werden mussten.

Für die neue Umsetzung wurde db2triples nun um die notwendigen JDBC-Schnittstellen für weitere Datenbanken erweitert. Hierdurch können die Daten direkt in RDF-Graphen überführt werden, es muss lediglich eine Mappingbeschreibung gemäß R2RML-Standard erstellt werden. Eine zusätzliche Programmierung ist nicht mehr nötig.

2.2 Datenstruktur

Von der Relationalen Datenstruktur wurde die Graphenstruktur, welche in Abb. 1 dargestellt ist, abgeleitet.

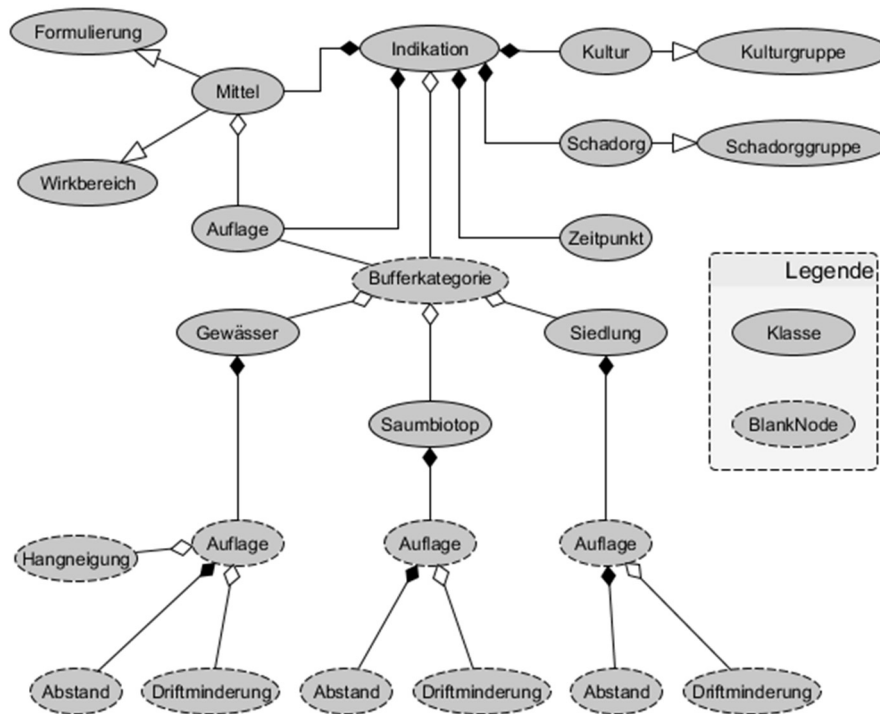


Abb. 1: Übersicht der Graphstruktur, Wurzelknoten Indikation

Die Indikation ist der Wurzelknoten der vorliegenden Datenstruktur, welche die verschiedenen Informationen besitzt. Falls vorhanden, sind zudem die Abstandsauflagen über den Knoten „Bufferkategorie“ mit der Indikation verbunden. Diese teilt sich auf in die drei Kategorien Gewässer, Saumbiotop und Siedlung, die die jeweiligen Abstände beinhalten. Die Knoten „Abstand“ und „Driftminderung“ enthalten ein menschenlesbares Label z. B. „5 m“, die maschinenlesbare Zahl „5“, die Typbezeichnung „xsd:Integer“ sowie die Ein-

heit „unit:Meter“. Die Informationen über das Vorhandensein im Kleinstrukturenverzeichnis und der Anwendung in einem Gemisch, sowie die verschiedenen Pflanzenaufwuchshöhen der Kulturen werden im jeweiligen Auflageknoten berücksichtigt.

Die Darstellung des Graphen zeigt zwar eine gerichtete Baumstruktur, da der Graph jedoch aus einzelnen Tripeln, also Aussagen besteht, werden „Rückwärts“-Abfragen ermöglicht.

2.3 SPARQL-Endpoint

Für den Webservice wird zunächst ein SPARQL-Endpoint, welcher für Abfragen auf den Triplestore genutzt wird, bereitgestellt. Hierfür wird der Fuseki Service von Apache Jena verwendet [Ap14].

Über die Abfragesprache SPARQL [SP13] können verschiedene Abfragen durchgeführt werden. Zum einen gibt es die Möglichkeit, alle direkten Nachbarinformationen einer Instanz zu ermitteln, dazu wird über das Stichwort „DESCRIBE“ eine Abfrage über den Fuseki auf alle Tripel, welche im Zusammenhang des gesuchten Subjekts existieren, durchgeführt, allerdings kann die Rückgabe je nach SPARQL-Endpoint variieren und ist somit nicht allgemeingültig. Dennoch können über den bereitgestellten Fuseki mit Hilfe der DESCRIBE-Abfrage alle notwendigen Informationen zu einer gegebenen Indikation ermittelt werden.

Allerdings ist es nicht möglich, mit der DESCRIBE-Abfrage den Graphen rückwärts zu durchlaufen. Dazu muss die CONSTRUCT-Abfrage genutzt werden. Mit deren Hilfe wird ein der Anwendung angepasster Graph bzw. Subgraph erstellt und mit diesen Informationen weitergearbeitet. SPARQL ermöglicht eine vielseitige Weiterverwendung der Ressourcen des Triplestores, für eine Vielzahl von Anwendungsfällen. Der beschriebene SPARQL-Endpoint ist aktuell unter <https://daten.ktbl.de/test/pam/sparql> erreichbar. Dort können Abfragen gemäß SPARQL 1.1 Protocol [FW13] mittels HTTP POST- oder GET-Aufrufen abgesetzt werden.

2.4 Webservice

Der Webservice bietet die neue Datenstruktur nicht nur menschenlesbar, sondern auch maschinenlesbar in verschiedenen Datenformaten an. Der Vorgänger-Webservice wurde in [Ma18] vorgestellt. Bisherige Erweiterungen sind eine dateibasierte Konfiguration, um den bisherigen Service leichter erweiterbar für weitere Anwendungsfälle zu machen – wie beispielsweise für den beschriebenen Abstandsservice. Dafür wurde zusätzlich eine vom jeweiligen Datenbestand unabhängige Template-Generierung für die HTML-Seiten der Daten erstellt.

3 Ergebnisse und Diskussion

Die Anzahl der Medienbrüche bei der Vorverarbeitung und Bereitstellung der Daten konnte durch Einsatz eines standardkonformen Mappingwerkzeugs deutlich verringert werden. Neue Datenbankanbindungen wurden durch Umsetzung der Schnittstellen in db2triples ermöglicht.

Die Datenstruktur für den LOD-Service beinhaltet auch einige inhaltliche Neuerungen und Verbesserungen: Durch eine Subklassen-Beziehung (`rdfs:subClassOf`) [BG04] der Kulturen und Schadorganismen können nun Hierarchien bidirektional abgeleitet werden. Hiermit ist es möglich, für den Fall, dass ein Mittel oder eine Indikation für die geforderte Kultur nicht zugänglich sind, über den Hierarchiebaum nach Alternativen zu suchen. So kann es beispielsweise sein, dass für den Apfel für den vorhandenen Fall keine Indikation bereitsteht. Nun kann der Anwender eine Hierarchieebene weiter oben, also beim Kernobst überprüfen, ob hier eine passende Indikation bereitgestellt wird, und daraufhin die dortigen abstandsrelevanten Auflagen berücksichtigen.

Die im Vorgängerprojekt nicht berücksichtigten Pflanzenaufwuchshöhen sind ebenfalls abgebildet. Des Weiteren ist es möglich, dem Mittel sowohl eine Formulierung als auch einen Wirkungsbereich zuzuweisen, die ebenfalls in eine Klassenhierarchie eingeordnet sind. Dadurch wird es möglich, in Abfragen gezielt Mittel bestimmter Formulierungs- oder Wirkungsbereichsklassen auszuwählen. Bisher waren diese nur durch ein Label gekennzeichnet.

Auf dieser inhaltlichen und technischen Basis wurde ein RESTful Service aufgesetzt, welcher auf einen SPARQL Endpoint zugreift. Die Graphenstruktur ermöglicht eine freie Abfragestruktur und kann gleichzeitig durch die SPARQL-seitigen vordefinierten oder selbst konstruierten Rückgabestrukturen beliebig genutzt werden. Der Webservice ermöglicht die Abfrage nach Indikationen und deren notwendigen Informationen zur Berücksichtigung der Abstandsauflagen. Bereitgestellt werden die Daten menschenlesbar als HTML oder in maschinenlesbaren Formaten wie Turtle oder Ntriples. Dadurch ist es möglich, die Abstandsauflagen automatisiert und standardkonform auszulesen, was eine leicht umsetzbare Übertragung in Applikationskarten und auf dem Schlepper genutzte Systeme möglich macht. Selbst bei Nichtbestehen der automatischen Übertragungsmöglichkeiten können die Abstandsauflagen leicht extrahiert und weiterverarbeitet werden.

4 Ausblick und Danksagung

Aktuell können Informationen zu Mitteln in der Aufbrauchfrist abgerufen werden, die Erstellung eines Mittelarchivs ist in Bearbeitung. Eine Abbildung und Integration von Hangabstandsauflagen wird berücksichtigt und ist Teil des aktuell durchgeführten Nachfolgeprojektes PAM3D, um durch die Geländetopographie gegebene Bedingungen in Applikationskarten für Pflanzenschutzmaßnahmen berücksichtigen zu können.

Die Weiterentwicklung des Webservice beinhaltet eine angepasste visuelle Darstellung, die Umsetzung von JSON-LD als weiteres Exportformat zur einfacheren Verarbeitung mit gängigen JSON-Bibliotheken oder in JavaScript-basierten Clients. Zudem sind weitere vordefinierte SPARQL-Queries für den Webdienst in Bearbeitung, die jeweils spezifische Datenausschnitte für unterschiedliche Anwendungsfälle liefern sollen. Zudem ist die vollständige Umsetzung der Linked Data Platform Spezifikation des W3C [SAM15] geplant.

Die Förderung des Vorhabens erfolgt (bzw. erfolgte) aus Mitteln des Bundesministeriums für Ernährung und Landwirtschaft (BMEL) aufgrund eines Beschlusses des deutschen Bundestages. Die Projektträgerschaft erfolgt (bzw. erfolgte) über die Bundesanstalt für Landwirtschaft und Ernährung (BLE) im Rahmen des Programms zur Innovationsförderung.

Literaturverzeichnis

- [Ap18] Apache Jena - Fuseki, <https://jena.apache.org/documentation/fuseki2/>, Stand: 25.10.2018.
- [Bc09] Berners-Lee, T.: Relational Databases on the Semantic Web. World Wide Web Consortium, 2009. <http://www.w3.org/DesignIssues/RDB-RDF.html>, Stand: 19.10.2018.
- [BG04] Brickley, D.; Guha, R. V.: RDF Vocabulary Description Language 1.0. World Wide Web Consortium, 2004. <http://www.w3.org/TR/rdf-schema/>, Stand: 24.10.2018.
- [DSC12] Das, S.; Sundara, S.; Cyganiak, R. (Hrsg.): R2RML: RDB to RDF Mapping Language. World Wide Web Consortium, 2012. <http://www.w3.org/TR/r2rml/>, Stand: 23.10.2018.
- [FW13] Feigenbaum, L.; Williams, G. T. (Hrsg.): SPARQL 1.1 Protocol. World Wide Web Consortium, 2013. <http://www.w3.org/TR/sparql11-protocol/>, Stand: 01.11.2018.
- [HS13] Harris, S.; Seaborne, A. (Hrsg.): SPARQL 1.1 Query Language. World Wide Web Consortium, 2013. <http://www.w3.org/TR/sparql11-query/>, Stand: 11.10.2018.
- [Ma15] Martini, D. et al.: KTBL-Planungsdaten auf dem Weg in die Zukunft – Bereitstellung über Linked Open Data. In (Ruckelshausen, A.; Schwarz, H.-P.; Theuvsen, B. Hrsg.): Informatik in der Land-, Forst- und Ernährungswirtschaft 2015, Geisenheim. GI-Edition Lecture Notes in Informatics Bd. 238, Bonn, S. 105-108, 2015.
- [Ma18] Martini, D.: Webservices auf heterogenen Datenbeständen – Methoden der Umsetzung am Beispiel der KTBL-Planungsdaten. In (Ruckelshausen, A. et al. Hrsg.): Informatik in der Land-, Forst- und Ernährungswirtschaft 2018, Kiel. GI-Edition Lecture Notes in Informatics, Bd. 278, Bonn, S. 155-158, 2018.
- [Sc16] Scheiber, M. et al.: Pflanzenschutz-Anwendungs-Manager (PAM): Automatisierte Berücksichtigung von Abstandsauflagen. Praktische Vorführung und Feldtestergebnisse. In (Ruckelshausen, A. et al. Hrsg.): Informatik in der Land-, Forst- und Ernährungswirtschaft 2016, Osnabrück. GI-Edition Lecture Notes in Informatics Bd. 253, Bonn, S. 177-180, 2016.
- [SAM15] Speicher, S.; Arwe, J.; Malhotra, A. (Hrsg.): Linked Data Platform 1.0. W3C Recommendation 26 February 2015. <https://www.w3.org/TR/ldp/>, Stand: 23.10.2018.