

Ein Umweltwissenssystem zur semantischen Vernetzung forstwirtschaftlicher Datenquellen

Mareike Dornhöfer¹, Alexander Holland², Madjid Fathi³

Abstract: Der Beitrag fokussiert den Einsatz von „Intelligenten Systemen“ zur Zusammenführung von Umweltinformationen durch die Verwendung von semantischen Technologien wie etwa Linked (Open) Data hin zu Umweltwissen. Dieses Umweltwissen kann in der Konsequenz z.B. zur Entscheidungsunterstützung der forstwirtschaftlichen Nachhaltigkeit durch die betroffenen Stakeholder verwendet werden. Die zugrunde liegende Forschungsfrage liegt darin, dass Umweltinformationen heute noch autark bzgl. Themen oder von Institutionen erfasst werden, aber eine Vernetzung und Anwendbarkeit im Sinne des Wissensmanagements noch wenig stattfindet. Dieses Wissen fehlt dementsprechend für die Erkennung von Umwelttrends oder in Entscheidungssituationen. Der Beitrag betrachtet die Frage, wie eine Zusammenführung und Vernetzung von unterschiedlichen Datenquellen mittels semantischer Technologien bzw. mittels Wissensmanagementverfahren stattfinden kann. Zur Abbildung des Anwendungsszenarios wird ein Vernetzungsszenario dargestellt. Es wird erläutert, welche Möglichkeiten bestehen, aus einem reinen Informationssystem ein Umweltwissenssystem zu etablieren, um letztendlich einen Aspekt eines Grünen Wissensmanagements umzusetzen. Das vorliegende Papier fokussiert aufgrund der begrenzten Seitenzahl die logische Schicht des Umweltwissenssystems.

Keywords: Semantische Technologien, Linked Data, Umweltwissenssystem, Green Knowledge Management, Grünes Wissensmanagement

1 Ausgangssituation

Umweltinformationen (UI) für eine größere öffentliche Zielgruppe werden zumeist in öffentlichen Umweltinformationssystemen (UIS), als Datensammlungen auf Webseiten oder als Web-Services zum kontinuierlichen ad-hoc Abruf aktueller Daten bereitgestellt. Zusätzlich ist die Veröffentlichung als Umweltbilanz über einen bestimmten Zeitraum oder thematischen Umweltbericht denkbar. An der Veröffentlichung von UI sind jedoch nicht nur öffentliche Stellen beteiligt, auch Unternehmen oder andere Organisationen beteiligen sich auf freiwilliger Basis, z.B. im Rahmen von Umweltmanagementverfahren wie ISO 14001 oder EMAS-III [Br10]. Betriebliche UIS zur Unterstützung des Umweltmanagements sind ebenso im Einsatz. Nachteil an den bereits aggregierten Berichten und Auswertungen (z.B. in pdf-Format) ist die schwierige Weiterverarbeitung oder Einbindung in andere Informationssammlungen. Um jedoch Zusammenhänge zwischen

¹Universität Siegen, Institut für Wissensbasierte Systeme & Wissensmanagement (WBS), Hölderlinstr. 3, 57068 Siegen, m.dornhoefer@uni-siegen.de

²Universität Siegen, Institut für WBS, Hölderlinstr. 3, 57068 Siegen, Alexander.Holland@uni-siegen.de

³ Universität Siegen, Institut für WBS, Hölderlinstr. 3, 57068 Siegen, fathi@informatik.uni-siegen.de

UI aus unterschiedlichen Quellen zu erstellen, ist genau dies notwendig. Nur durch eine Vernetzung und Etablierung eines kontextuellen Zusammenhangs kann es als angereichertes Umweltwissen (UW) in Entscheidungen Anwendung finden. Auf diese Weise wäre eine Organisation in der Lage Handlungspotentiale (z.B. hinsichtlich der eigenen Nachhaltigkeitsstrategie) abzuleiten. Ein semantisches Verfahren zur Verknüpfung ist Linked (Open) Data, welches Daten und Informationen aus unterschiedlichen Quellsystemen miteinander in Form von Graph-Strukturen in einem einheitlichen RDF (Resource Description Framework)-Format verbindet und Abfragen via SPARQL (SPARQL Protocol and RDF Query Language) zulässt [EHR15]; zudem ist über Regeln das Auslösen von Aktionen möglich. LOD hat das inhärente Ziel, öffentlich verfügbare Datenquellen miteinander in Kontext zu setzen und so eine Erweiterung der Semantik im Internet zu schaffen. Geht man von einer unternehmensinternen Anwendung aus, so ist eine Veröffentlichung natürlich keine zwingende Vorgabe, sondern es können interne Daten mit öffentlich zur Verfügung gestellten LOD verknüpft und in eine eigene, interne Wissensbasis integriert werden [ABT14]. Das durch die Autoren bereits in einer vorherigen Veröffentlichung [DF15] vorgestellte Modell eines *Umweltwissenssystems (UmweltWiS)* basiert auf den zuvor genannten Technologien und ist Teil eines *Grünen Wissensmanagements (Green Knowledge Management)*. Es soll dabei im Kontext der Forstwirtschaft betrachtet werden, wobei im Speziellen auf den Aufbau der semantischen Logik eingegangen wird.

2 „Fortstwirtschaftliches Umweltwissenssystem“

Das generische Konzept des UmweltWiS [DF15] bildet sich schematisch aus aufeinander aufbauenden Schichten, bestehend aus der (1) Datenhaltungsschicht, welche interne und externe Umweltquellen in unterschiedlichen Formaten einbindet, einer semantischen Logik (2), einer Schicht (3) zur Analyse und Anreicherung der Daten, und einer Anwendungsschicht (4) für Endnutzer und Wissensingenieure. Diese Schichten werden von entsprechenden Geschäftsprozessen umrahmt. Eine Anwendung des UmweltWiS im Bereich der Forstwirtschaft bietet den Anwendungsnutzen unterschiedliche Statistiken und Entwicklungen von Waldbeständen, Schädlingen oder Umwelteinflüssen in eine Wissensbasis zusammenzuführen, zu analysieren und daraus entscheidungsunterstützende Erkenntnisse zu erarbeiten. Als LOD Datenlieferanten sind öffentliche LOD Services wie die LOD Cloud⁴, aber auch Geo- oder Fachinformationssysteme, welche aufzubereitende Messdaten oder Statistiken enthalten, möglich. Abbildung 1 zeigt die schematische Darstellung des UmweltWiS für den skizzierten Anwendungsfall, wobei die im Folgenden diskutierte, semantische Repräsentationsschicht umrahmt hervorgehoben wird:

⁴ Linked Open Data Cloud, <http://lod-cloud.net/>

Semantische Analyse und Erstellung eines Umweltwissenssystems 17

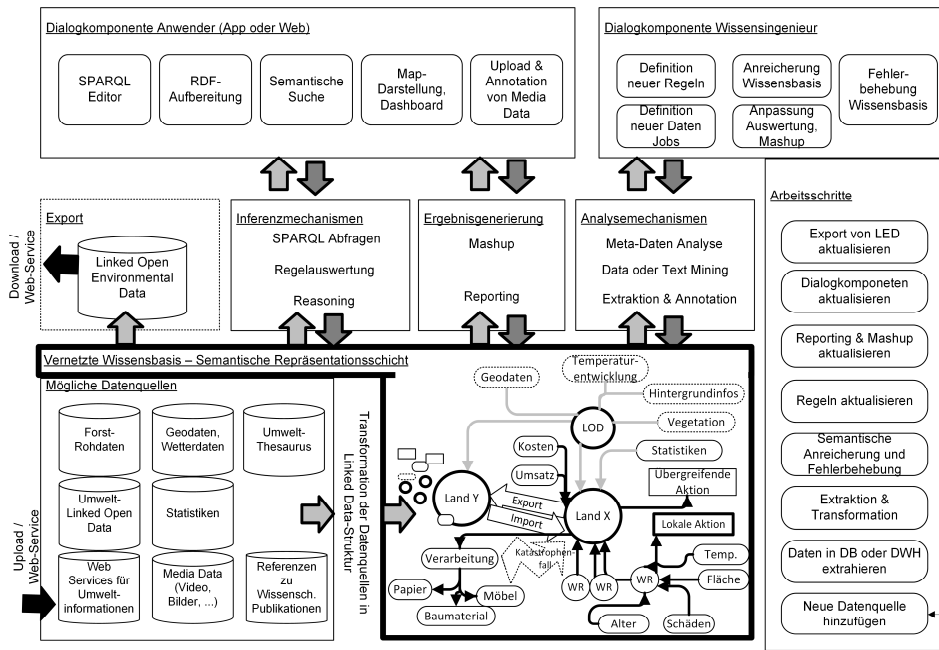


Abb. 1: Verknüpfung von semantischem Wissen im Anwendungskontext Forstwirtschaft

Betrachtet man diese Schicht eingehender, so werden ausgehend von einem hohen Abstraktionslevel auf Länderebene, alle weiteren Daten zugeordnet. Dies wird herunter gebrochen auf die einzelnen Waldregionen (WR), zu welchen der Baumbestand (inkl. Alter), Baumarten, die zugehörige Fläche, Wetterdaten zur Region, lokale Aktionen, wie etwa die Aufforstung einer bestimmten Baumart, Schädlingsbekämpfung, Durchforstung, etc. vernetzt wird. Die ermittelten Zahlen des Imports und Exports werden ebenfalls anhand anderer Quellen zugeordnet. Zusätzlich wird erfasst, für welche Anwendungszwecke das geschlagene Holz verarbeitet wurde, um direkte Abhängigkeiten zu erkennen und für Entscheidungen heranzuziehen. Interessant ist hierbei, Zusammenhänge zwischen Industriezweig und Baumart sowie Import und Export herauszuarbeiten. Hierzu ist die Einbindung von Statistiken oder anderen Hintergrundinformationen (z.B. Artenvielfalt) interessant. Durch die Zuordnung von Geodaten ergibt sich zudem die Möglichkeit der visuellen Aufbereitung in Form von interaktiven Wissenskarten. Anhand von Messreihen lassen sich zeitliche Veränderungen in den Baumbeständen oder deren Umfeld, z.B. durch Faktoren wie Erhöhung von Temperatur oder Niederschlag abbilden, ebenso wie Katastrophenfälle (z.B. Windbruch) und deren Auswirkungen. Mediendateien von Waldschäden lassen sich als interaktive Elemente in der Wissensbasis hinterlegen, um so direkten Bezug zwischen Vorfällen und daraus abgeleiteten Aktionen herzustellen. Durch eine Verlinkung zwischen verschiedenen Ländern, ist es zudem möglich weitergehende Verbindungen und Abhängigkeiten zu ermitteln.

Aufgrund der unterschiedlichen Strukturierung und Formate der Ausgangsquellen, ist eine direkte Vernetzung meist nicht möglich. Ege et al. [EHR15] sehen Herausforderungen bei der Interoperabilität, „*Redundanz, Konsistenz, Kohärenz, des Mappings*“ und der Vollständigkeit der Daten aus unterschiedlichen Quellen. Auer et al. beschreiben hierzu einen möglichen zyklischen Transformationsprozess, u.a. bestehend aus einer Klassifizierung, Qualitätsanalyse und Fehlerbehebung [ABT14]. Auch ist der zeitliche Bezug, z.B. in Form einer zyklischen oder ereignisbezogenen Zuordnung von neuen Daten zu betrachten. Der Vorteil der semantischen Vernetzung ist dabei eine modulare oder inkrementelle Entwicklung und hierdurch inhärent vorhandene Flexibilität der gewählten Wissensrepräsentationsform [PSA14]. Ebenso ist bei der Vernetzung das Recycling von bestehendem Wissen entscheidend, welches z.B. durch Referenzen auf dieselben Individuen in unterschiedlichen Inputquellen realisiert wird. Mittels Regeln lassen sich letztendlich weitergehende Auswertungen oder Aktionen ermöglichen, z.B. IF Baumart = Douglasie AND Boden = Trocken THEN Schaden [Zuwachsschwäche, Radialrisse] oder IF Boden = Trocken AND Schaden = Windbruch THEN Pflanzen [Douglasie], Die Datenhaltung und Anwendungsoberfläche sind zum Zeitpunkt der Erstellung dieses Beitrags prototypisch vorhanden, die nächsten Schritte bestehen in der Fertigstellung der Logik auf Basis der vorhandenen Datenquellen und der zugehörigen Reasoning-Abfragen und Regeln und deren interaktiven Visualisierung auf der Oberfläche.

Literaturverzeichnis

- [ABT14] Auer, S., Bryl, V., Tramp, S. (Hrsg.): Linked Open Data - Creating Knowledge Out of Interlinked Data, LNCS 8661, Springer, Heidelberg, 2014.
- [Br10] Brauweiler, J.: Umweltmanagementsysteme nach ISO 14001 und EMAS. In: Kramer, M. (Hrsg.) Integratives Umweltmanagement, Gabler, Wiesbaden, S. 279-320, 2010.
- [DF15] Dornhöfer, M; Fathi, M: UmweltWiS: Von Umweltinformationssystemen zu „Umweltwissenssystemen“? In: Umweltbundesamt (Hrsg.) Umweltinformationssysteme – Big Data – Open Data – Data Variety, Ergebnisse des 22. Workshops des Arbeitskreises „Umweltinformationssysteme“, 07-08 Mai 2015, Kassel, ISSN 2199-6571, S. 127-139, 2015,
<http://www.umweltbundesamt.de/publikationen/umweltinformationssysteme-5>
- [EHR15] Ege, B.; Humm, B.; Reibold, A. (Hrsg.): Corporate Semantic Web - Wie semantische Anwendungen in Unternehmen Nutzen stiften, Springer Vieweg, Berlin, 2015.
- [PSA14] Pellegrini, T., Sack, H., Auer, S. (Hrsg.): Linked Enterprise Data. Management und Bewirtschaftung vernetzter Unternehmensdaten mit Semantic Web. Springer Verlag, X.media.press, Berlin Heidelberg, 2014.