

# Ontologie basierte Datenintegration in der Umweltforschung

## Eine Fallstudie zur Integration von Umweltdaten

Oliver Archner<sup>1</sup>, Matthias Faerber<sup>2</sup>, Florent Jochaud<sup>2</sup>, Stefan Jablonski<sup>2</sup>, Michael Hauhs<sup>1</sup>

### Einleitung

Praktische Umweltforschung erzeugt eine große Anzahl von syntaktisch- und semantisch-heterogenen Daten.

Im Rahmen dieser Fallstudie wurde ein System aufgebaut, das Forschern erlaubt Daten kontextsensitiv zu erheben und in eine gemeinsame semantische Wissensbasis zu integrieren.

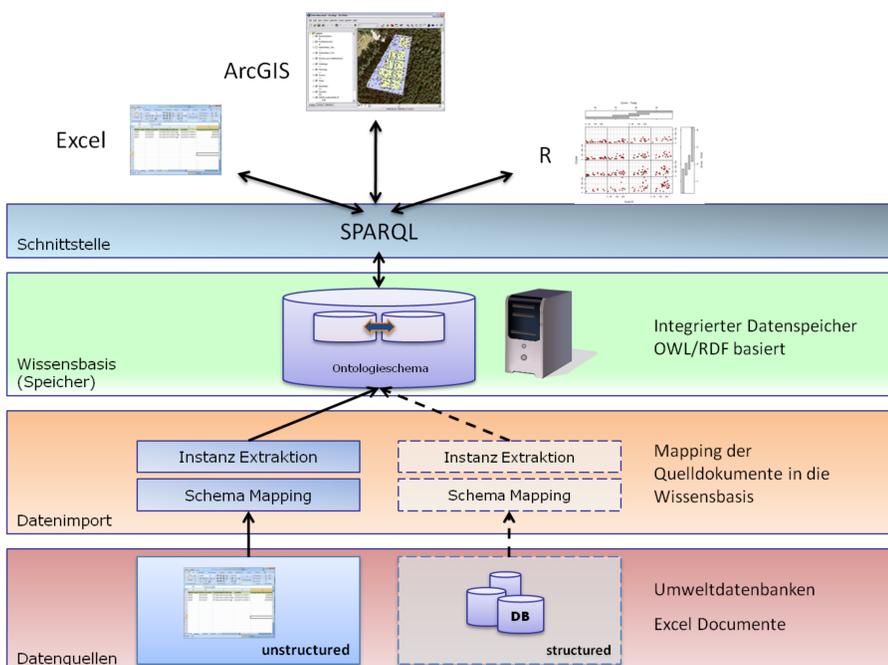
### Zielsetzung

- Integration von „spatiotemporal thematic data“
- Spezifikation von terminologischem Wissen der Anwendungsdomänen auf Basis von Ontologien
- Einsatz und Erweiterung einer globalen Observations-Ontologie
- Standardisierte Zugriffsmöglichkeit

### Systemarchitektur

Die Architektur basiert auf der Technologie des Semantic Web. Es können sowohl unstrukturierte (Excel) als auch strukturierte Datenquellen (Datenbanken) über Mapping und Instanzextraktion integriert werden.

Kern des Systems ist ein auf OWL/RDF basierender Wissenspeicher, der dem Benutzer eine integrierte Sicht auf sämtliche Informationen gestattet.



### Schichtenmodell der Systemarchitektur

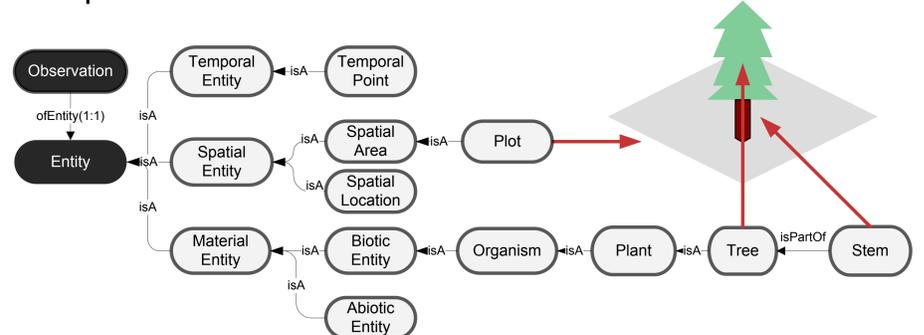
### Quellen

- Madin, Joshua; et al.: An ontology for describing and synthesizing ecological observation data, *Int. J. Ecol. Informatics* 2 (2007), pp. 279–296.
- Han, Lushan; et. al: RDF123: a mechanism to transform spreadsheets to RDF, University of Maryland, Baltimore County, 2007.
- W3C: SPARQL Query Language for RDF, Seaborne, Andy; Prud'hommeaux, Eric, Editors, W3C Recommendation, 15 January 2008, <http://www.w3.org/TR/rdf-sparql-query/>
- Berners-Lee, Tim; Hendler, Jim; Lassila, Ora: The Semantic Web, *Scientific American*, 2001.

### Observation-Ontologie

Das globale Schema der Wissensbasis fußt auf der aus dem SEEK Projekt bekannten OBOE Ontologie.

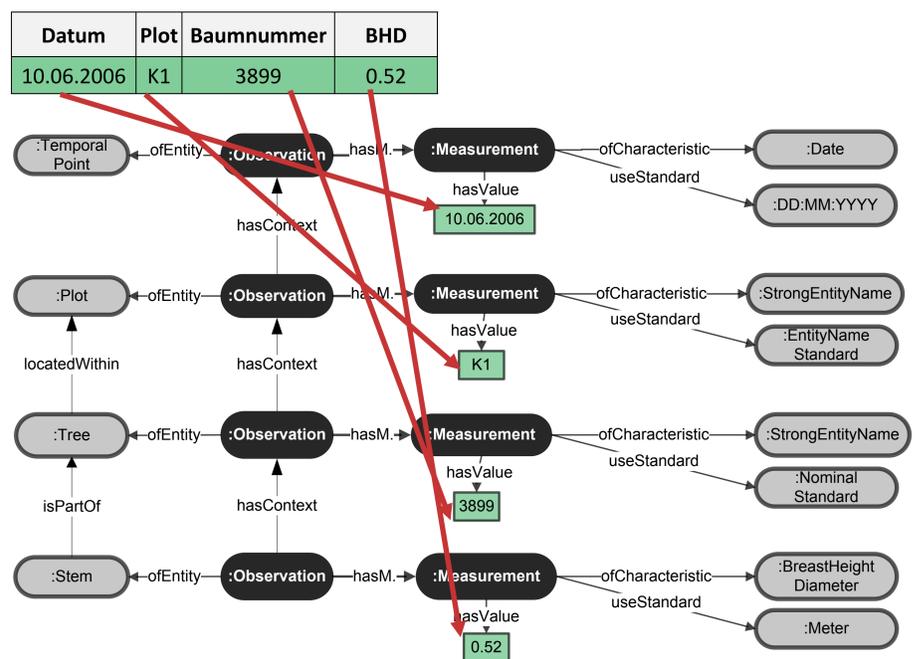
Erweiterungen der Konzepte „Entity“ und „Characteristic“ ermöglichen es uns die Wissensbasis für einzelne Fachdisziplinen aufzubauen und zu modularisieren.



### Erweiterung der Klasse „Entity“ für eine forstliche Bestandsaufnahme [nach Madin, 2007]

### Datenimport und Integrationsschicht

Über die Frameworks RDF123 und D2RQ ist ein Mapping der Datenquellen auf die Ontologie möglich.



### Beispiel für das Mappen eines Datensatzes einer forstlichen Bestandsaufnahme auf die globale Ontologie.

### Benutzerschnittstelle

- Zugriff auf den gemeinsamen Wissenspeicher über das Web-Frontend „Knowledge Navigator“ oder die SPARQL-Schnittstelle
- Schnittstellen für gängige wissenschaftliche Analyse Tools wie z. B. R, Excel oder ArcGIS sind geplant

<sup>1</sup>BayCEER, Universität Bayreuth

<sup>2</sup>Angewandte Informatik 4, Universität Bayreuth

{Oliver.Archner, Matthias.Faerber, Florent.Jochaud, Stefan.Jablonski, Michael.Hauhs}@uni-bayreuth.de