



Potentiale der Semantic-Web-Technologien RDF und OWL zur semantischen Beschreibung von Data-Warehouse-Metadaten

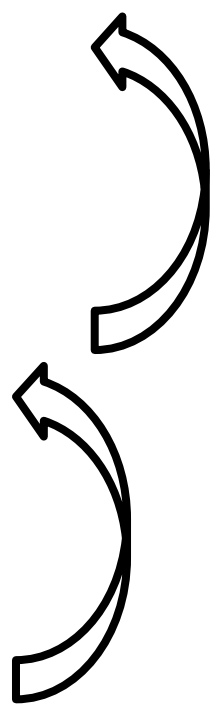
Gemeinsames Treffen der GI-Fachgruppen
Datenbanksysteme und Information Retrieval
08. Mai 2008

Dipl.-Wirtsch.Inf. Stefan Hartmann

1. Problemstellung
2. Metadaten in Data-Warehouse-Systemen
3. Potentiale des Resource Description Frameworks (RDF) zur semantischen Beschreibung von Data-Warehouse-Metadaten
4. Potentiale der Web Ontology Language (OWL) zur semantischen Beschreibung von Data-Warehouse-Metadaten
5. Schlussbemerkung und Fazit

1 Problemstellung

Ausgangssituation

- Heterogene Data-Warehouse-Systeme in betrieblichen Organisationen
 - Komplexitäts-, Kostengründe, gewachsene Strukturen etc.
 - Heterogenität auf Ebene der Datenschemata
 - Kein einheitliches Begriffsverständnis
 - Bestehende Metadaten-Standards
 - Aufgrund ihrer Komplexität wenn überhaupt nur partiell realisiert
 - Kaum semantische Konzepte
 - Semantic Web
 - Dokumente im WWW semantisch reichhaltig beschreiben
 - Standardisierung neuer Technologien mit hoher semantischer Mächtigkeit
 - **Ziel:** Semantische Beschreibung von Data-Warehouse-Metadaten
- 

1. Problemstellung
2. Metadaten in Data-Warehouse-Systemen
3. Potentiale des Resource Description Frameworks (RDF) zur semantischen Beschreibung von Data-Warehouse-Metadaten
4. Potentiale der Web Ontology Language (OWL) zur semantischen Beschreibung von Data-Warehouse-Metadaten
5. Schlussbemerkung und Fazit

2 Metadaten in Data-Warehouse-Systemen

Metadaten in DWH-Systemen

- Voraussetzung zur korrekten Interpretation der Daten eines DWH
- Eingehende Metadaten-Beschreibung anhand der *verwendungszweckgerichteten Metadatenkategorisierung* [Auth04, S.44ff.]

Kategorie	Elemente (Auswahl)
Terminologie	Begriffsbenennung, -identität, -beziehungen
Datenstruktur / Datenbedeutung	Name, Beschreibung, Typ und Elemente der Datenstruktur
Organisationsbezug	Dateneigentümer, -verwender, Berechtigungen
Datentransformation	Datenquelle, -ziel, Transformationsschritte
Datenanalyse	Cube, Dimension, Kennzahl
Datenqualität	Vollständigkeit, Semantik, Identifizierbarkeit
Metadatenhistorie	Version, Änderung, Änderungsgrund
Systembezug	Softwarekomponente, Hersteller

1. Problemstellung
2. Metadaten in Data-Warehouse-Systemen
3. Potentiale des Resource Description Frameworks (RDF) zur semantischen Beschreibung von Data-Warehouse-Metadaten
4. Potentiale der Web Ontology Language (OWL) zur semantischen Beschreibung von Data-Warehouse-Metadaten
5. Schlussbemerkung und Fazit

3 Potentiale des RDF zur semantischen Beschreibung von Data-Warehouse-Metadaten

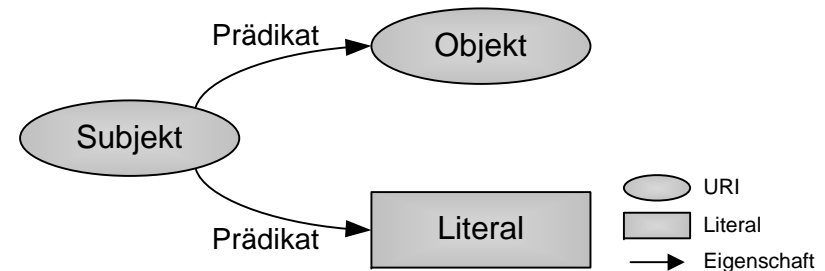
RDF/RDFS

Resource Description Framework (RDF)

Anyone can say anything about anything [KICa02]

- Metadaten-Rahmenwerk, um Ressourcen einheitlich zu annotieren
- Formale Semantik zur korrekten maschinellen Interpretation (RDF/XML)

- Aussagen in Form von RDF-Tripeln



- RDF: Fachlich neutrales Datenmodell
- RDF-Schema: Erstellung fachspezifischer Vokabulare

3 Potentiale des RDF zur semantischen Beschreibung von Data-Warehouse-Metadaten

RDF/RDFS

Metadatenkategorie Terminologie

Metadatum	RDF	RDFS
Begriffsidentität	ID/URI	ID/URI
Begriffsbenennung	rdfs:label	rdfs:label
Begriffsdefinition	rdfs:comment	rdfs:comment
Begriffsbeziehungen	Property	rdf:Property, rdfs:subClassOf

3 Potentiale des RDF zur semantischen Beschreibung von Data-Warehouse-Metadaten

RDF/RDFS

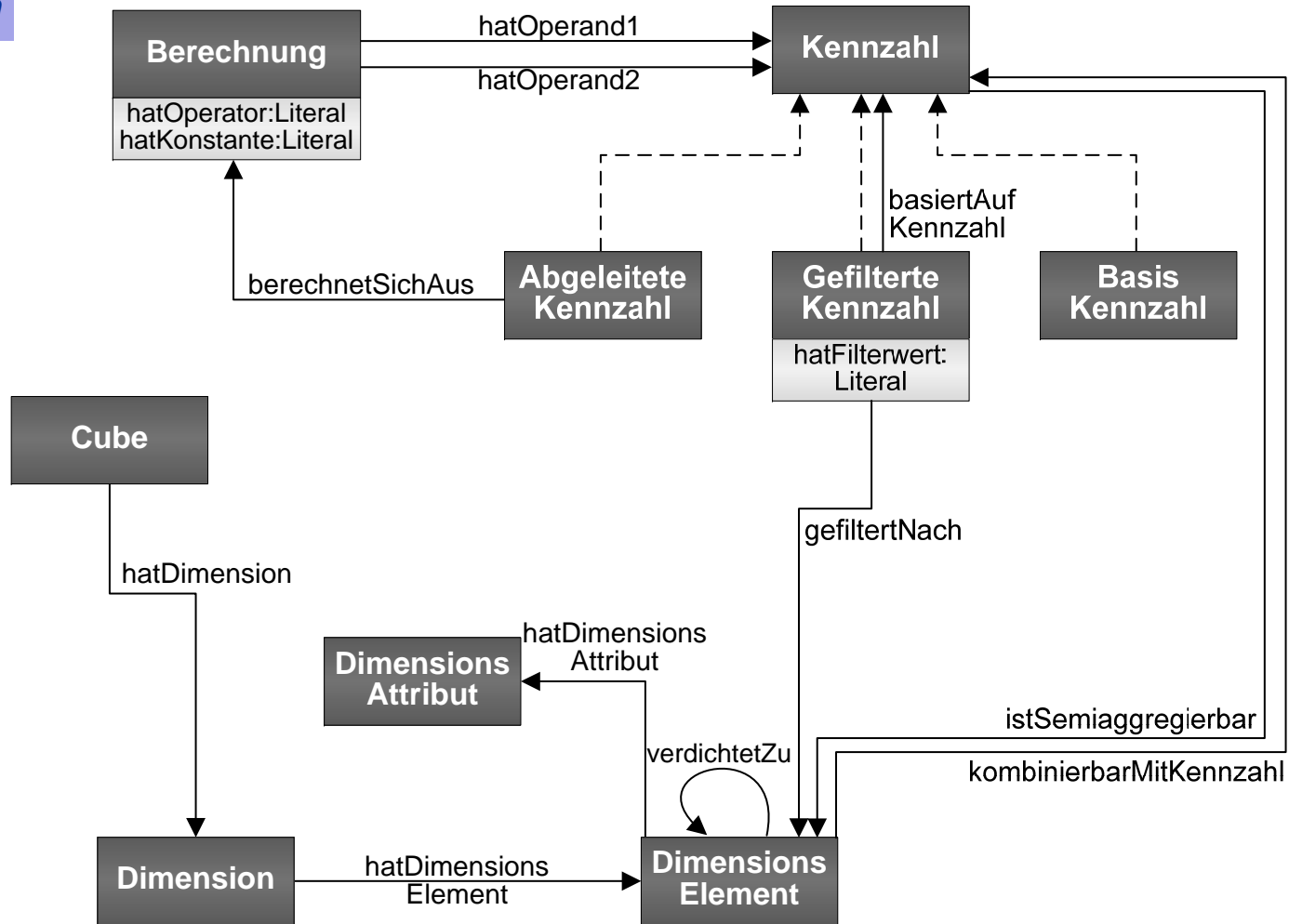
Metadatenkategorie Datenanalyse

Metadatum	RDF	RDFS
Cube	dwhmd:Cube	rdfs:Class
Dimension	dwhmd:Dimension	rdfs:Class
Dimensionselement	dwhmd:DimensionsElement	rdfs:Class
Dimensionsattribut	dwhmd:DimensionsAttribut	rdfs:Class
Kennzahl	dwhmd:Kennzahl	rdfs:Class
Kennzahlenbeziehungen	Property	rdfs:subClassOf, rdf:Property, rdfs:Class

Analog übertragbar auf Metadaten-Elemente der weiteren genannten Metadaten-Kategorien (vgl. [HaWe08]).

3 Potentiale des RDF zur semantischen Beschreibung von Data-Warehouse-Metadaten

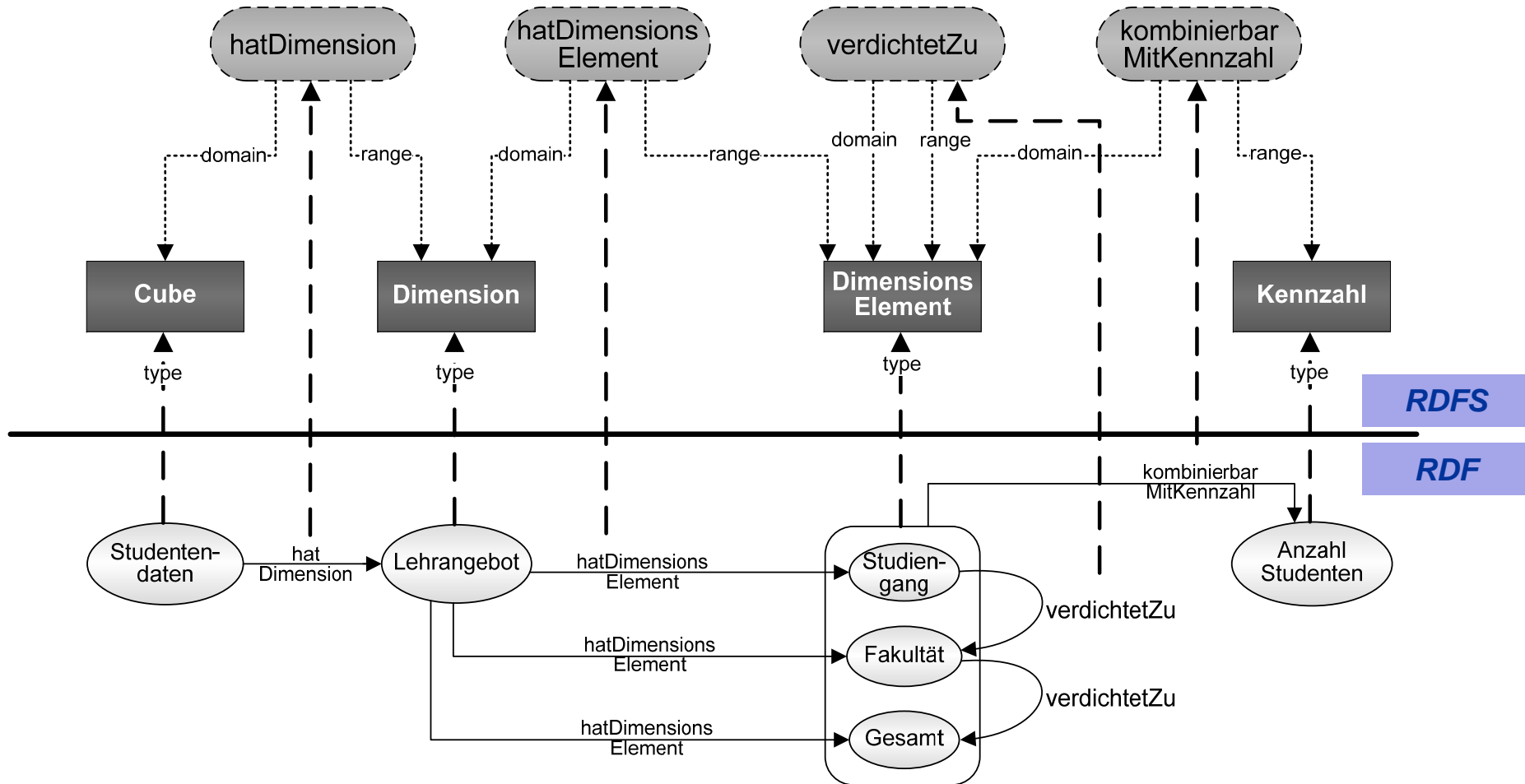
RDF-Schema



Class Klasse
Property Eigenschaftstyp (Wertebereich: Literal)
 —Property—> Eigenschaftstyp
 - -subClassOf- -> Unterklasse

3 Potentiale des RDF zur semantischen Beschreibung von Data-Warehouse-Metadaten

RDF-Schema



Legende

- Klasse
- Eigenschaftstyp
- Eigenschaft (Instanz)—>
-type, range, domain.....>
- Gruppierung
- Subjekt / Objekt

3 Potentiale des RDF zur semantischen Beschreibung von Data-Warehouse-Metadaten

RDF-Schema

```
<?xml version="1.0" ?>
```

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
```

Namespace-
Deklaration

```
<rdf:Description
  rdf:about="http://www.semantic-metadata.com/dwhmd/elements/1.0/dwh-schema-simple#">
  <dc:title xml:lang="de">Data-Warehouse-Metadaten</dc:title>
  <dc:description xml:lang="de">Ein RDF-Schema fuer DWH-Metadaten.</dc:description>
  <dc:creator>Stefan Hartmann</dc:creator>
  <dc:contributor>M. Weber</dc:contributor>
</rdf:Description>
```

Datenstruktur-
Beschreibung

```
<rdfs:Class rdf:ID="Cube">
  <rdfs:label>Cube</rdfs:label>
  <rdfs:comment>Diese Klasse dient der Beschreibung eines Cubes.</rdfs:comment>
</rdfs:Class>
<rdfs:Class rdf:ID="Dimension">
  <rdfs:label>Dimension</rdfs:label>
  <rdfs:comment>Diese Klasse definiert die Dimensionen eines Cubes.</rdfs:comment>
</rdfs:Class>
```

Klassen-
Definition

```
<rdf:Property rdf:ID="hatDimension">
  <rdfs:label>hatDimension</rdfs:label>
  <rdfs:comment>Verknüpfung eines Cubes mit einer Dimension.</rdfs:comment>
  <rdfs:domain rdf:resource="#Cube" />
  <rdfs:range rdf:resource="#Dimension" />
</rdf:Property>
```

Eigenschaftstyp-
Definition

```
</rdf:RDF>
```

1. Problemstellung
2. Metadaten in Data-Warehouse-Systemen
3. Potentiale des Resource Description Frameworks (RDF) zur semantischen Beschreibung von Data-Warehouse-Metadaten
4. Potentiale der Web Ontology Language (OWL) zur semantischen Beschreibung von Data-Warehouse-Metadaten
5. Schlussbemerkung und Fazit

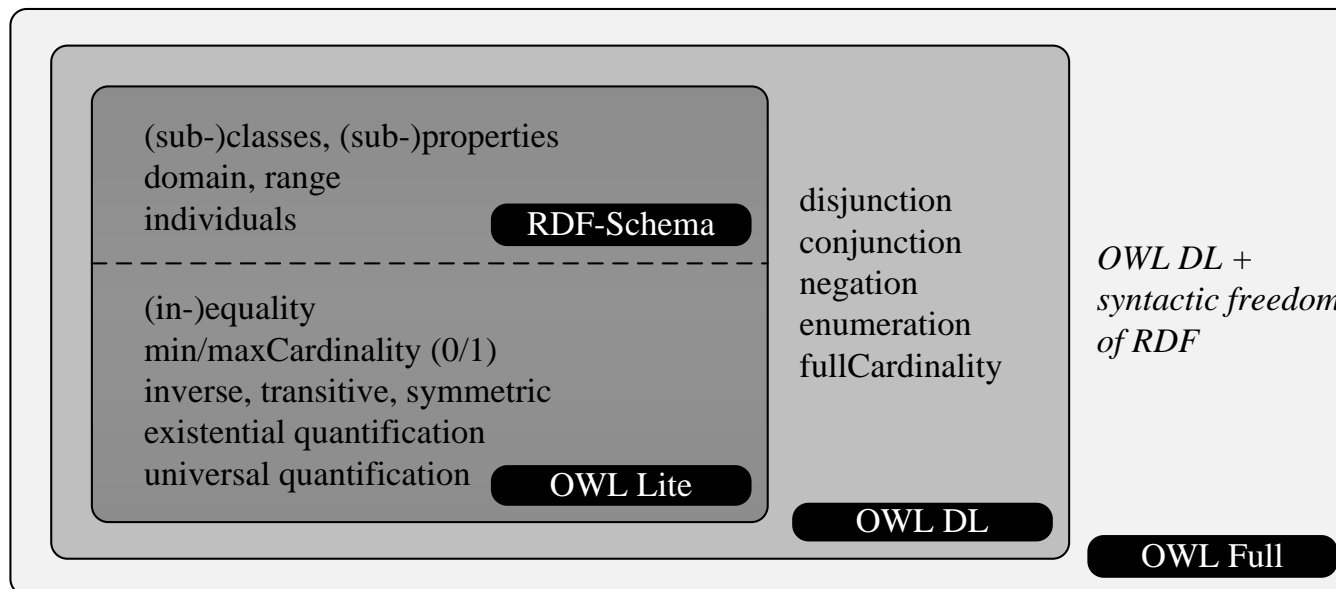
4 Potentiale der OWL zur semantischen Beschreibung von Data-Warehouse-Metadaten

OWL

Web Ontology Language (OWL)

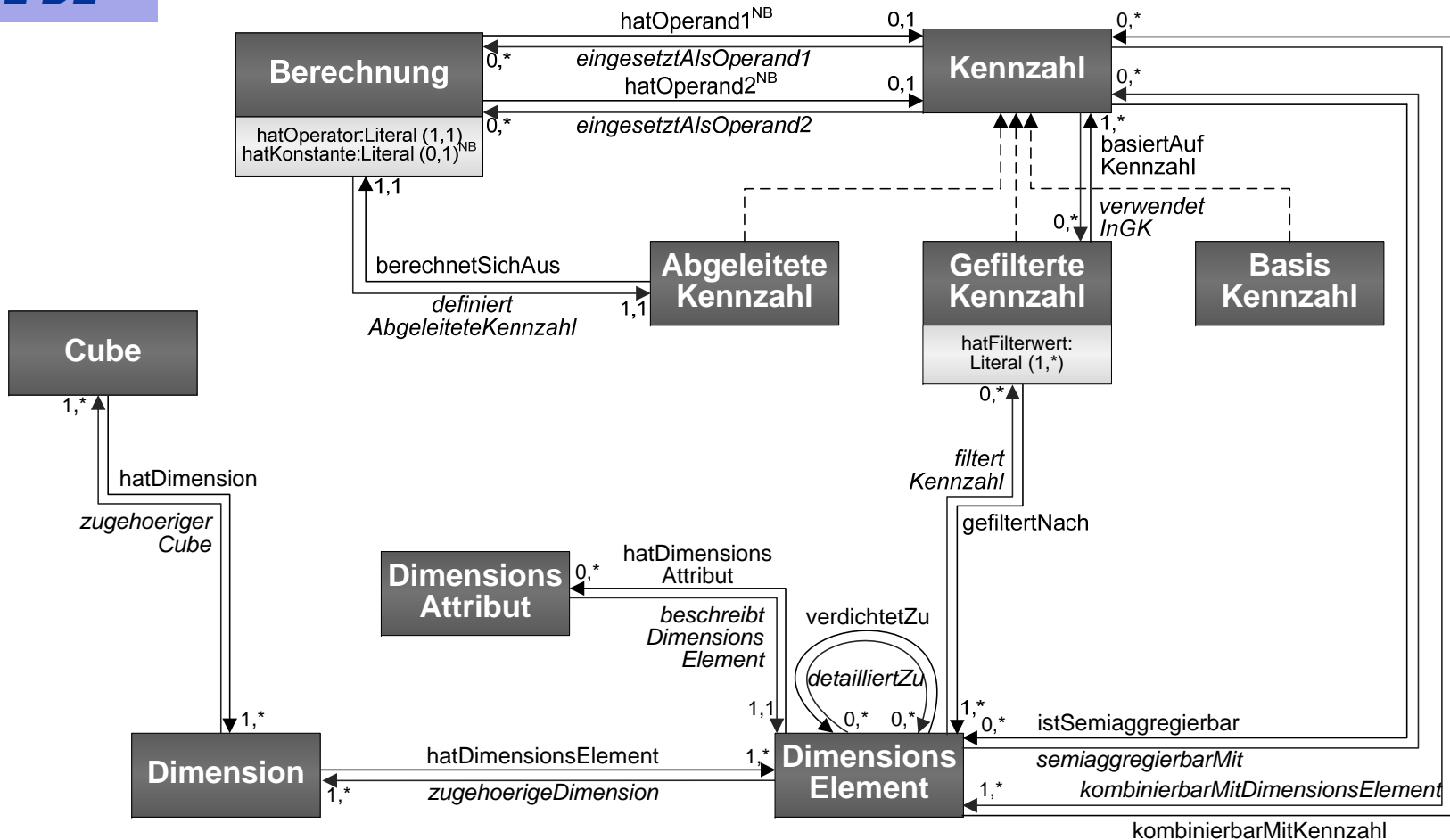
An ontology is an explicit specification of a conceptualization. [Grub93]

- W3C standardisierte Ontologiesprache [BHH+04]
- Vokabular-Erweiterung von RDF
- Möglichkeit der Inferenz (z. B. durch Reasoner wie Pellet, FaCT++)



4 Potentiale der OWL zur semantischen Beschreibung von Data-Warehouse-Metadaten

OWL DL



Class Klasse

Property Rolle (Wertebereich: Literal)

Property → Rolle (inverse Rolle)

-- subClassOf → Unterklasse

NB: exactly \equiv owl:cardinality
 ((hatOperand1 exactly 1) AND (hatOperand2 exactly 1)) OR
 ((hatOperand1 exactly 1) AND (hatKonstante exactly 1)) OR
 ((hatKonstante exactly 1) AND (hatOperand2 exactly 1))

1. Problemstellung
2. Metadaten in Data-Warehouse-Systemen
3. Potentiale des Resource Description Frameworks (RDF) zur semantischen Beschreibung von Data-Warehouse-Metadaten
4. Potentiale der Web Ontology Language (OWL) zur semantischen Beschreibung von Data-Warehouse-Metadaten
5. Schlussbemerkung und Fazit

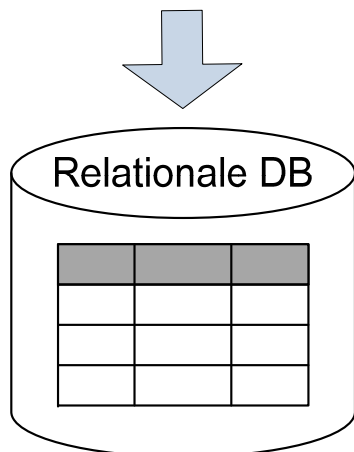
5 Schlussbemerkung und Fazit

RDF-Anfragesprachen

- SPARQL (W3C-Standard, <http://www.w3.org/TR/rdf-sparql-query/>)
- RDQL, SeRQL etc.

SQL

```
SELECT *
FROM   tabelle3
WHERE  NAME=' SH'
```



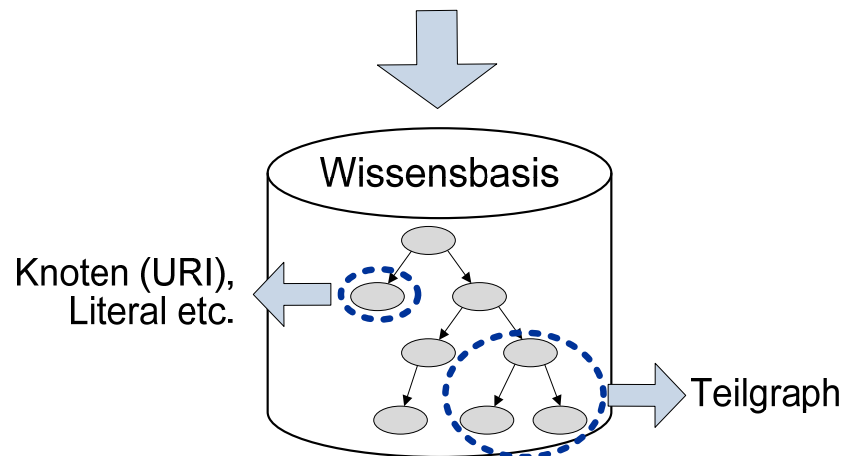
SPARQL

PREFIX

```
dwhmd: <http://www.semantic-metadata.com/...>
```

```
SELECT ?Kennzahl
```

```
WHERE { dwhmd:DimensionsElement_X
        dwhmd:kombinierbarMitKennzahl
        ?Kennzahl. }
```



5 Schlussbemerkung und Fazit

Zusammenfassung

- Mit RDF ist eine formale und semantische Beschreibung von DWH-Metadaten möglich
 - Klar spezifizierte Bedeutung der Klassen und Eigenschaftstypen.
 - Semantisch definierte Beziehung zwischen Metadaten-Elementen.
 - Semantisch einheitliche Annotation.
 - Sprachliche Defekte können aufgedeckt werden.
- Der Einsatz von OWL ermöglicht zudem eine weitere Detaillierung der semantischen Auszeichnung sowie Inferenz.
- Werkzeugunterstützung (Editoren, Visualisierungen, XML-Parser etc.)

Einschränkung und Forschungsbedarf

- (Fachliche) Abstimmungen zur Standardisierung eines RDF-Vokabulars bzw. einer OWL-Ontologie bleiben nach wie vor notwendig.
- Fehlen einer ausgereiften Schnittstelle, um Metadaten kommerzieller DWH-Systeme in RDF bzw. OWL zu transformieren.

Dipl.-Wirtsch.Inf. **Stefan Hartmann**

Wissenschaftliches Institut für Hochschulsoftware
der Universität Bamberg (ihb)

Feldkirchenstr. 21
96045 Bamberg

✉ stefan.hartmann @ ihb.uni-bamberg.de

🌐 <http://www.ihb.uni-bamberg.de>

🌐 <http://www.ceushb.de>

Literatur

- Auth04 Auth, G.: Prozessorientierte Organisation des Metadatenmanagements für Data-Warehouse-Systeme. Books on Demand, Norderstedt 2004.
- BHH+04 Bechhofer, S.; van Harmelen, F.; Hendler, J.; Horrocks, I. et. al.: OWL Web Ontology Language Reference. W3C Recommendation 10 February 2004. (<http://www.w3.org/TR/owl-ref>).
- Grub93 Gruber, T. R.: A Translation Approach to Portable Ontology Specifications. In: Knowledge Acquisition 5 (1993) 2, S.199-220.
- HaWe08 Hartmann, S.; Weber, M.: Semantisch homogene Beschreibung von Data-Warehouse-Metadaten mit RDF. In: Bichler, M. et. al. (Hrsg.): Multikonferenz Wirtschaftsinformatik 2008, GITO-Verlag, Berlin 2008, S.117-128. (http://ibis.in.tum.de/mkwi08/04_Business_Intelligence/01_Hartmann.pdf).
- KICa04 Klyne, G.; Carrol, J.: Resource Description Framework (RDF): Concepts and Abstract Syntax – W3C Recommendation 10 February 2004. (<http://www.w3.org/TR/rdf-concepts/>).