

Wibaklidama Sommerworkshop - 22.06.2010

Semantic-Web-Infrastruktur für Geoscience-Data – ein Werkstattbericht

Bernd Ritsdchel, Günther Neher

GFZ/Fachhochschule Potsdam, FB Informationswissenschaften

Vorspann

Semantic Web/Linked Data
in 5 Minuten

Ausgangspunkt: HTML - Web der Dokumente

HTML

[Tim Berners-Lee](#) ist der **Gründer** des [W3C](#)

[Tim Berners-Lee](#) **gründete** das [W3C im Jahr 1994](#)

[Tim Berners-Lee](#) ist **Autor des Artikels** "[Tabulator: Exploring and Analyzing Linked Data on the Semantic Web](#)"

Der **Artikel** "[Tabulator: Exploring and Analyzing Linked Data on the Semantic Web](#)" **stammt von** [Tim Berners-Lee](#)

[Tim Berners-Lee](#) **kennt** [Dan Brickley](#)

... aus Rechnersicht:

```
<a href="http://www.w3.org/People/Berners-Lee/">Tim Berners-Lee</a>
```

```
???
```

```
<a href="http://www.w3.org">W3C</a>
```

```
...
```

```
<a href="http://www.w3.org/People/Berners-Lee/">Tim Berners-Lee </a>
```

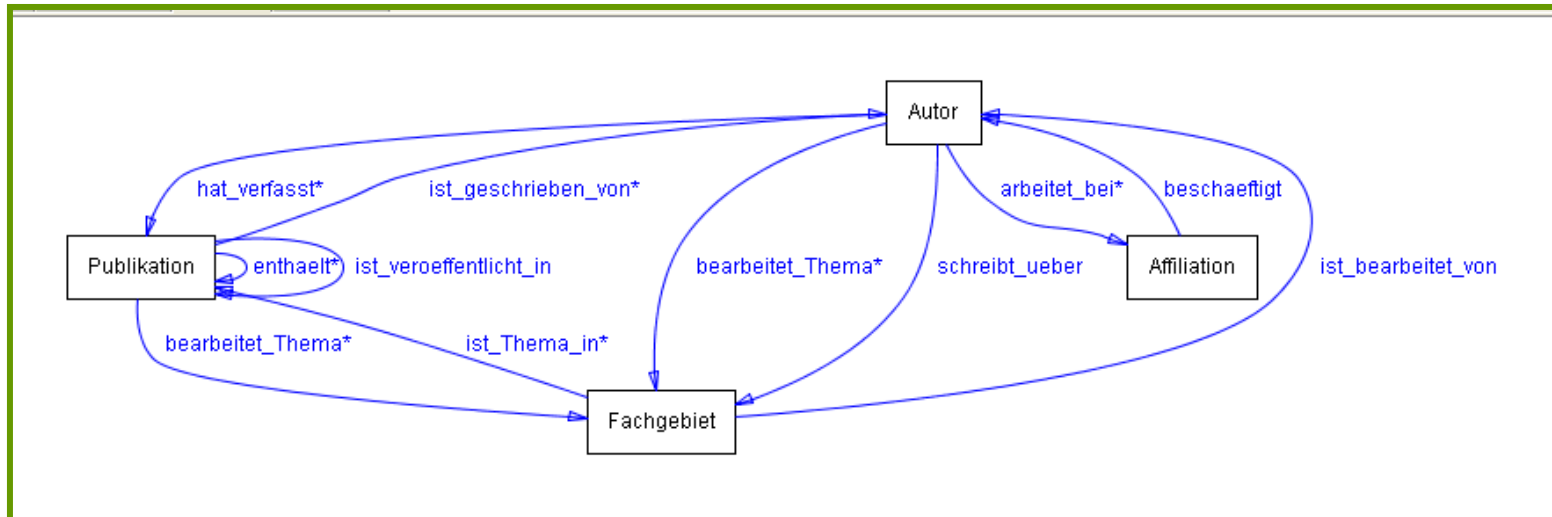
```
???
```

```
<a href="http://danbri.org/">Dan Brickley</a>
```

Ausgangspunkt: HTML - Web der Dokumente

Voraussetzung für ein „Web des Wissens“:

Möglichkeit der maschineninterpretierbaren eindeutigen Benennung von Konzepten und Beziehungen

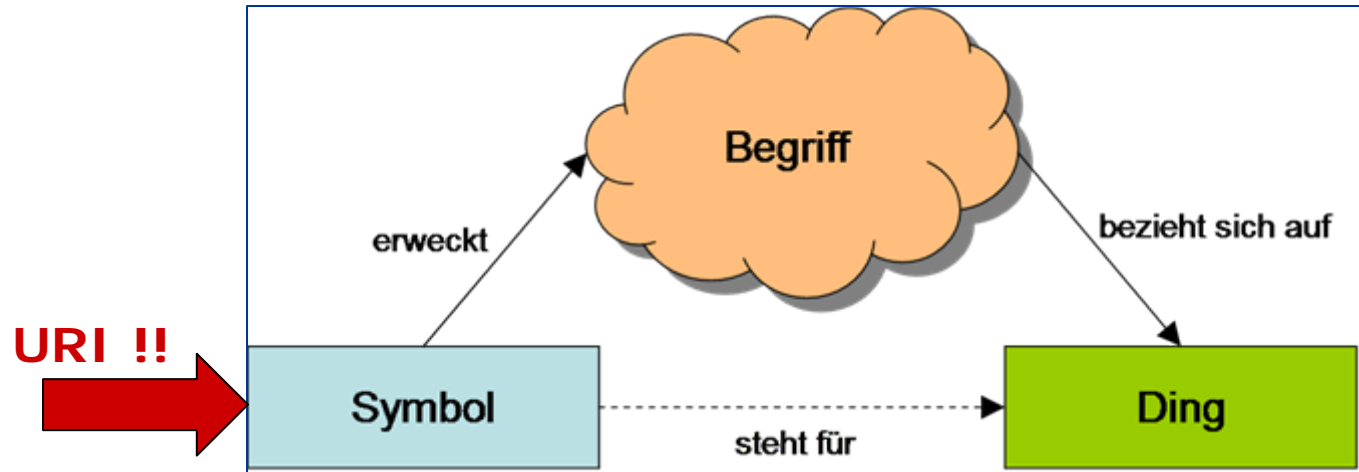


"In the Semantic Web, it is not the Semantic which is new, it is the Web which is new".

[Chris Welty/IBM](#)

Semantic Web: Basiskonzepte

Herausforderung: „Semiotisches Dreieck“



Lösung: Eindeutige Festlegung von **Vokabular** auf Basis von URIs

SW: Everything has a URI

Don't say "colour" say `<http://example.com/2002/std6#col>`

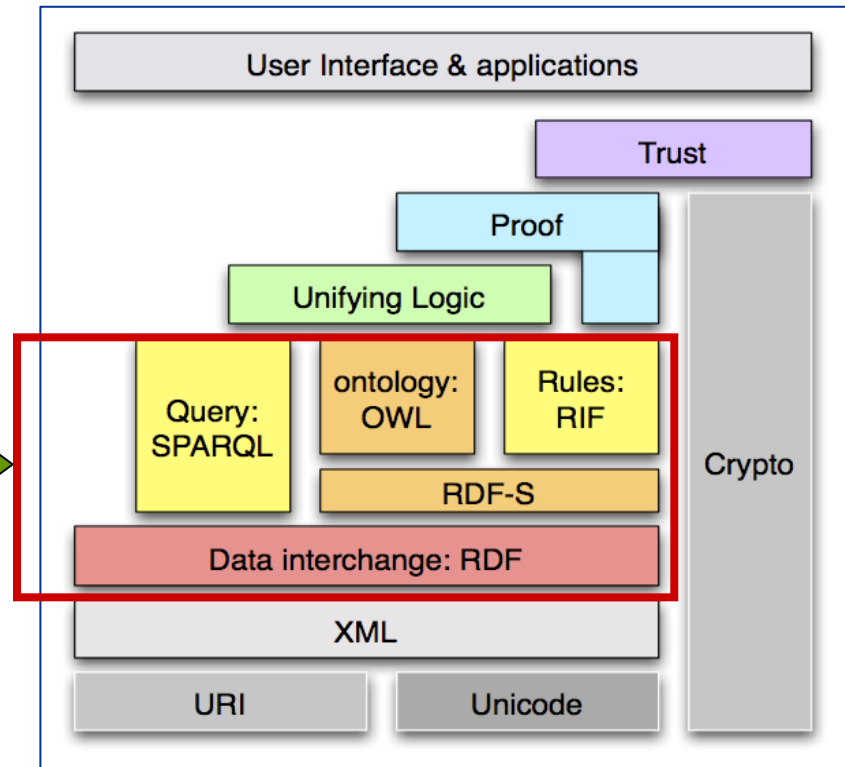
Semantic Web: Basiskonzepte

Herausforderung:

Form der Wissensrepräsentation, welches Netzwerkeffekte auf globaler Ebene ermöglicht (d.h. offen, interoperabel, erweiterbar, verknüpfbar, wiederverwendbar, ...)

Ansatz: „Semantic Web Layer Cake“:

Familie von Formatstandards mit **RDF** („Resource Description Framework“) als grundlegendem Datenmodell



RDF: Datenmodell

Elementare Einheit des RDF-Datenmodells ist ein „**Statement**“ in Form eines Tripels
„**Subjekt-Prädikat-Objekt**“

Subjekt und Prädikat MÜSSEN, Objekt KANN durch einen **URI** repräsentiert werden.
Jedes Objekt eines Statements KANN Subjekt eines neuen Statements sein.

RDF-Statement in „Notation 3“

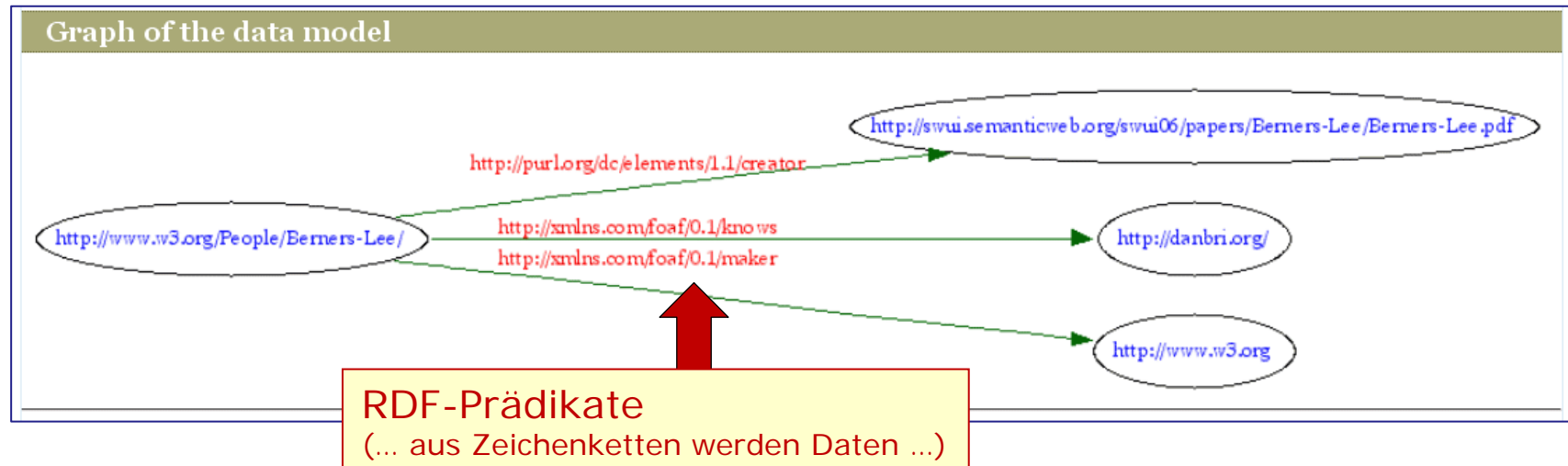
`<http://www.w3.org/>` `<http://purl.org/dc/elements/1.1/title>` `"World Wide Web Consortium"`

(„Subjekt“)

(„Prädikat“)

(„Objekt“)

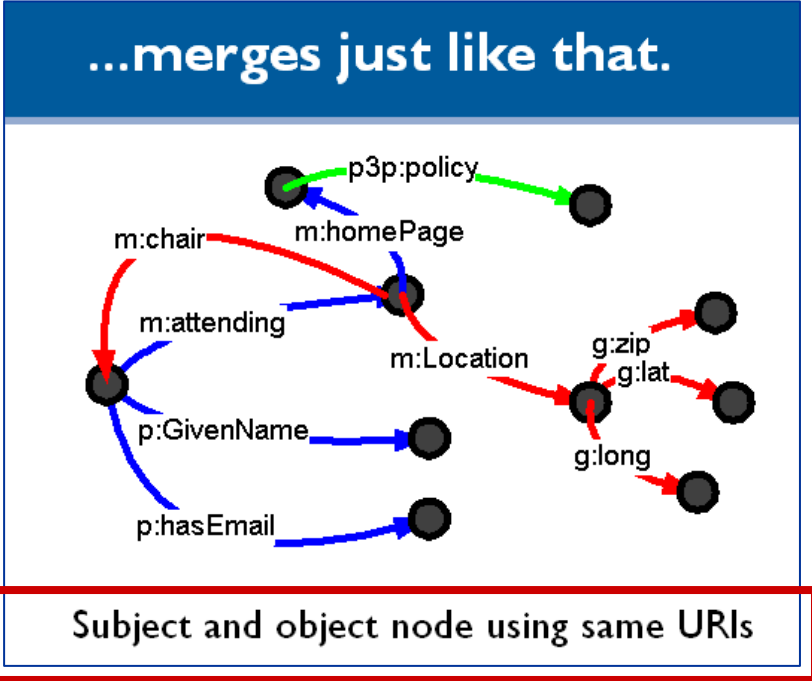
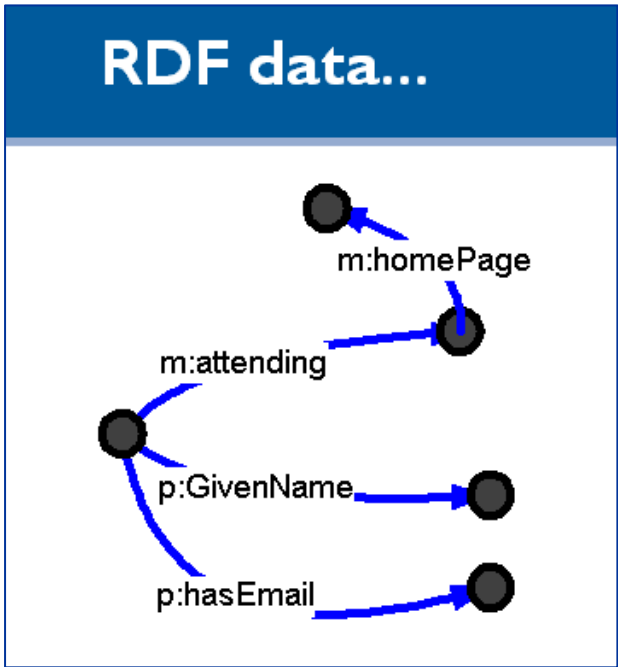
RDF-Statement als gerichteter Graph



Semantic Web: Basiskonzepte

RDF-Datmodell:
Das „Objekt“ jedes Statements kann „Subjekt“ eines neuen Statements sein

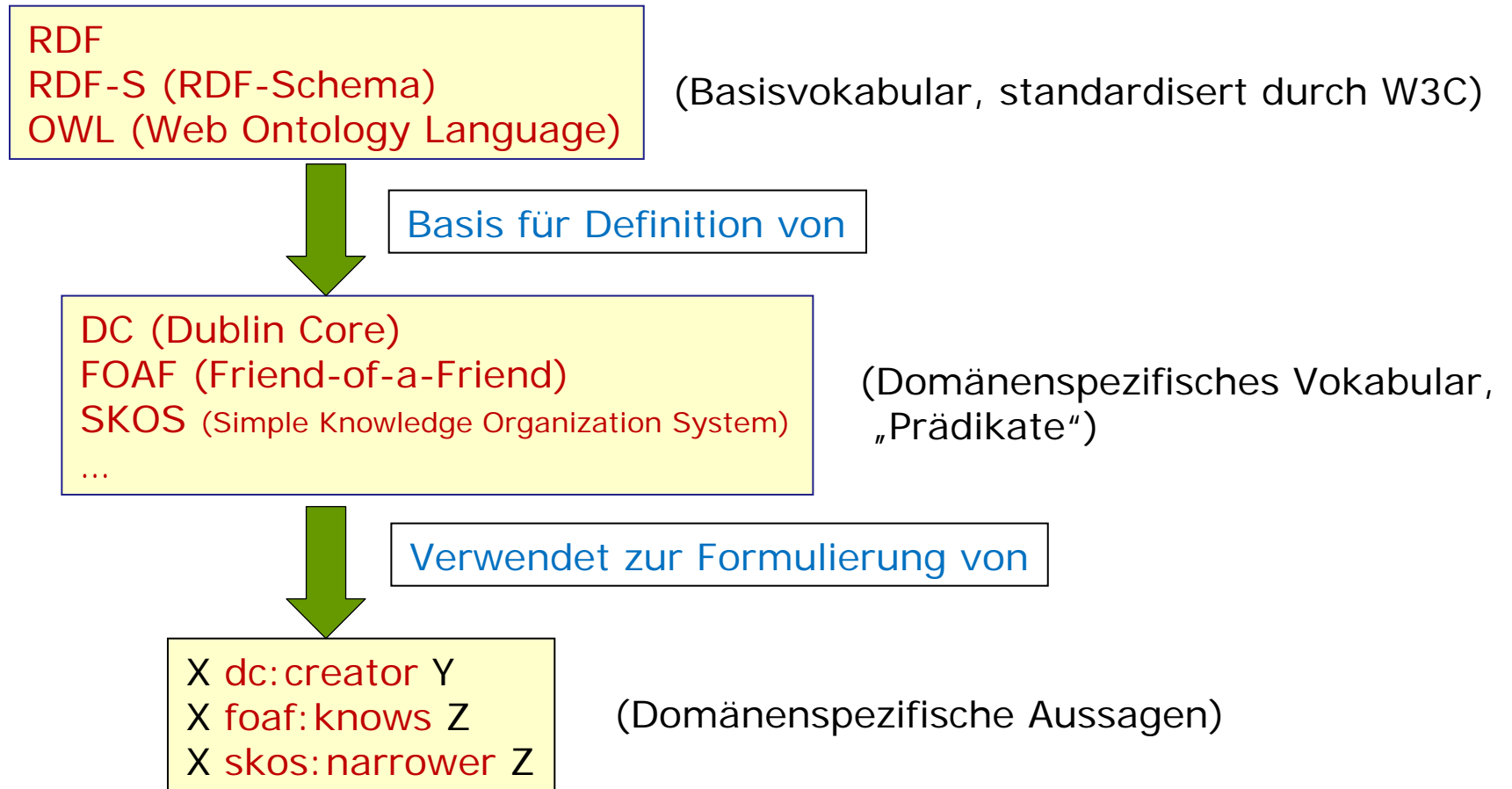
➔ ermöglicht: „Linked Data“



FHP:-)

(Quelle: Tim Berners-Lee: „Semantic Web: Building On What Exists“, 2006)

Themenbereich: Formulierung domänenspezifischer Aussagen



Themenbereich: Domänenspezifisches Vokabular, Namensräume

These namespaces are used to describe entities in X number of documents

Namespaces (814 know namespaces)	Number of documents
http://xmlns.com/foaf/0.1/	1,261,382
http://purl.org/goodrelations/v1#	645,201
http://blogs.yandex.ru/schema/foaf/	584,781
http://www.w3.org/2002/07/owl#	451,101
http://sites.wiwiwiss.fu-berlin.de/suhl/bi...	252,361
http://rdfs.org/sioc/ns#	183,227
http://www.w3.org/2003/01/geo/wgs84_pos#	157,659
http://rdfs.org/sioc/types#	117,599
http://semantic-mediawiki.org/swivt/1.0#	64,281
http://xmlns.com/foaf/spec/	50,514
http://auswiki.org/index.php/Special:URI...	25,887
http://usefulinc.com/ns/doap#	19,127
http://purl.org/dc/dcmitype/	17,766
http://www.radarnetworks.com/shazam#	8,061
http://wecowi.org/view/Spezial:URIResolv...	6,967
http://www.radarnetworks.com/2007/09/12/...	6,113
http://www.geonames.org/ontology#	5,474
http://purl.org/rss/1.0/	4,606
http://smw.ontoware.org/2005/smw#	4,459
http://www.w3.org/1999/02/22-rdf-syntax-...	4,346

„Nutzung von Vokabularen“

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:openid="http://xmlns.openid.net/auth#"
  xmlns:bio="http://vocab.org/bio/0.1/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:wot="http://xmlns.com/wot/0.1/"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:dc="http://purl.org/dc/terms/"
  xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xhv="http://www.w3.org/1999/xhtml/vocab#"
  xmlns:rel="http://vocab.org/relationship/"
  xmlns:sioc="http://rdfs.org/sioc/ns#"
  xmlns:air="http://www.daml.org/2001/10/html/airport-ont#"
  xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#"
  xmlns:rss="http://web.resource.org/rss/1.0/"
>
  <foaf:PersonalProfileDocument
rdf:about="http://www.ivan-herman.net/foaf.rdf">
    <foaf:primaryTopic>
      <dc:Agent rdf:about="http://www.ivan-herman.net/me">
        <bio:event>
          <bio:Birth>
            <bio:date
```

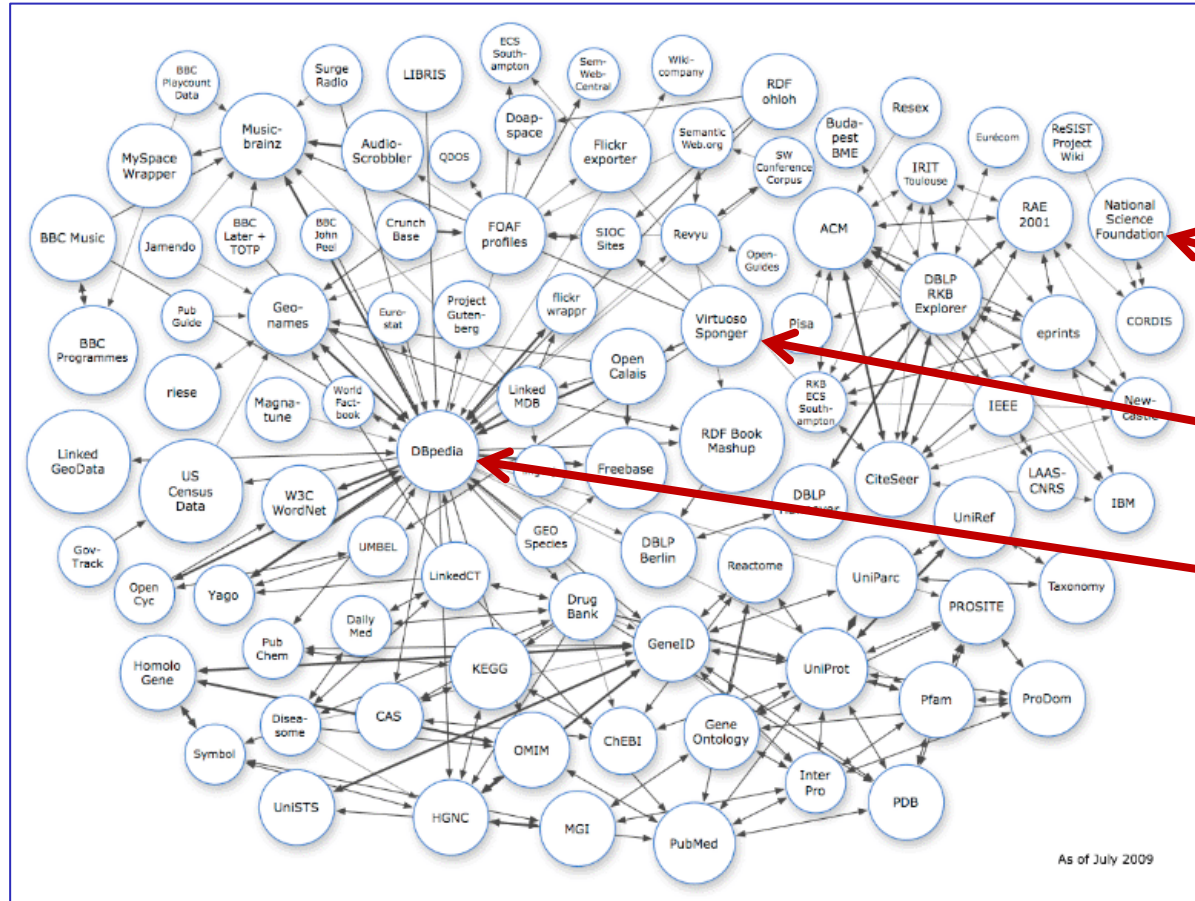
<http://pingthesemanticweb.com/stats/namespaces.php>

RDF(S)/OWL-Vokabular/Bezeichner: Übersicht

RDF (25)	RDFS (16)	OWL (Auswahl)
rdf: RDF	rdfs: Class	owl: AllDifferent
rdf: about	rdfs: Resource	owl: Class
rdf: Description	rdfs: Datatype	owl: FunctionalProperty
rdf: Property	rdfs: domain	owl: InverseFunctionalProperty
rdf: type	rdfs: range	owl: Restriction
rdf: resource	rdfs: Container	owl: SymmetricProperty
rdf: Statement	rdfs: isDefinedBy	owl: Thing
rdf: subject	rdfs: equivalentClass	owl: TransitiveProperty
rdf: predicate	rdfs: subClassOf	owl: allValuesFrom
rdf: object	rdfs: subPropertyOf	owl: cardinality
rdf: value	rdfs: seeAlso	owl: complementOf
rdf: Bag	rdfs: ContainerMembershipProperty	owl: disjointWith
rdf: Seq	rdfs: member	owl: equivalentClass
rdf: List	rdfs: comment	owl: equivalentProperty
rdf: Alt	rdfs: label	owl: hasValue
rdf: XMLLiteral	rdfs: Literal	owl: intersectionOf
rdf: ID , rdf: nodeID		owl: inverseOf
rdf: parseType		owl: sameAs
rdf: datatype		owl: unionOf

Projekt

Zielsetzung: DIF-Data: „Linked-Data ready“



?

**DIF-
Geosci
ence
Data**

➔ <http://linkeddata.org/>

2 Arbeitsgruppen: Teilaufgaben

Transformation nach RDF(S)/OWL

Teilaufgaben

- Analyse der DIF-Schemata (XSD) und ausgewählter Beispieldatensätze (XML)
- Auswahl und Test geeigneter Transformationswerkzeuge ("RDFizers")
- Durchführen der Transformation und Evaluation der Ergebnisse
- Beispielabfragen mit SPARQL
- Ggf. Nachbearbeitung/Modellierung in RDFS/OWL
- Dokumentation und Bewertung des Transformationsprozesses

Verknüpfung mit Linked Data Cloud

Teilaufgaben

- **Grobanalyse DIF-Schemata und Beispieldatensätze?**
- [Bestandsübersicht Linked-Data-Cloud](#)
- **Auswahl potentiell geeigneter Datenbestände der Linked-Data-Cloud auf Basis von Nutzungsszenarios / Use Cases)?**
- **Analyse der dort verwendeten RDF-Vokabulare und potentiellen Verknüpfungspunkte?**
- **Modellierung der Verknüpfungen mit RDFS/OWL (inkl. Test der Funktionalität)?**
- **Beispielabfragen mit SPARQL?**
- **Dokumentation und Bewertung der Ergebnisse?**

Arbeitspaket: Analyse DIF-Datenmodell/Schema

Directory Interchange Format (DIF) Writer's Guide

[View PDF version](#)

About DIFs

[What is a DIF?](#)

[XML DIF Schema](#) [XML DIF Template](#)

[Some Tips on Writing DIFs](#)

[Procedure for Suggesting Changes](#)

DIF Fields **Note:** All fields denoted as either: **Required**, **Highly Recommended**, **Recommended**.

[Entry ID](#)

[Quality](#)

[Entry Title](#)

[Access Constraints](#)

[Parameters \(Science Keywords\)](#)

[Use Constraints](#)

[ISO Topic Category](#)

[Distribution](#)

[Data Center](#)

[Data Set Language](#)

[Summary](#)

[Data Set Progress](#)

[Metadata Name](#)

[Related URL](#)

[Metadata Version](#)

[DIF Revision History](#)

[Data Set Citation](#)

[Keyword \(Ancillary Keyword\)](#)

[Personnel](#)

[Originating Center](#)

[Instrument](#)

[Multimedia Sample](#)

<http://gcmd.nasa.gov/User/difguide/difman.html>

Arbeitspaket: Transformation nach RDF/S

[Home](#) » [Wiki](#)

[Article](#) | [Discussion](#) | [Edit](#) | [History](#)

RDFizers

glossary definition:=The RDFizer project is directory of tools for converting various data formats into RDF. I
some of these tools.

What can I do with these?

You can have a computer generate the RDF representation of your data for you, instead of you doing it by h

Why were they built?

Writing RDF by hand can be a very time consuming and error prone experience, but the Semantic Web suff
that no killer app will be written without enough data and no data will be exposed without the benefit of a ki

This is one of our solutions to unlocking the catch-22: identify existing datasets that could be of potential int
capture at least a little bit of those data structures.

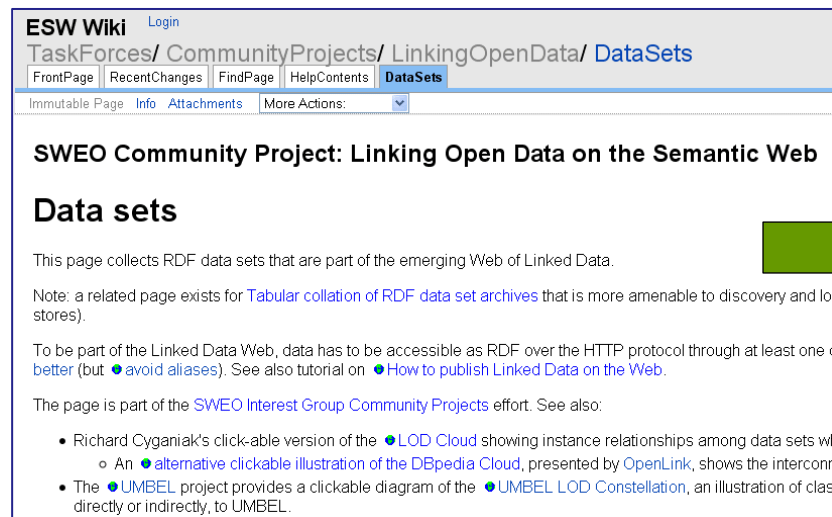
<http://simile.mit.edu/wiki/RDFizers>

Arbeitspaket: Analyse von LOD-Beständen

Um LOD-Kandidaten für eine spätere SINNVOLLE und nachnutzbare Verknüpfung zu identifizieren, muss ein relativ DETAILLIERTER Überblick über die aktuell verfügbaren LOD-Bestände vorhanden sein, insbesondere über die jeweils verwendeten Vokabulare.

Aufgabe:

Analysieren einen LOD-Bestand Ihrer Wahl und dokumentieren Sie diesen im Kurs-Glossar (als Glossareintrag verwenden Sie bitte den **Namen** des Bestandes). Falls möglich, „bewerten“ Sie das verwendete Vokabular auf der Basis der „Popularität“ (s. nachfolgende Folie)



ESW Wiki Login
TaskForces/ CommunityProjects/ LinkingOpenData/ DataSets
FrontPage RecentChanges FindPage HelpContents DataSets
Immutable Page Info Attachments More Actions: [v]

SWEOW Community Project: Linking Open Data on the Semantic Web

Data sets

This page collects RDF data sets that are part of the emerging Web of Linked Data.

Note: a related page exists for [Tabular collation of RDF data set archives](#) that is more amenable to discovery and load stores).

To be part of the Linked Data Web, data has to be accessible as RDF over the HTTP protocol through at least one of [better](#) (but [avoid aliases](#)). See also tutorial on [How to publish Linked Data on the Web](#).

The page is part of the [SWEOW Interest Group Community Projects](#) effort. See also:

- Richard Cyganiak's clickable version of the [LOD Cloud](#) showing instance relationships among data sets which
o An [alternative clickable illustration of the DBpedia Cloud](#), presented by [OpenLink](#), shows the interconnections
- The [UMBEL](#) project provides a clickable diagram of the [UMBEL LOD Constellation](#), an illustration of class relationships, directly or indirectly, to UMBEL.

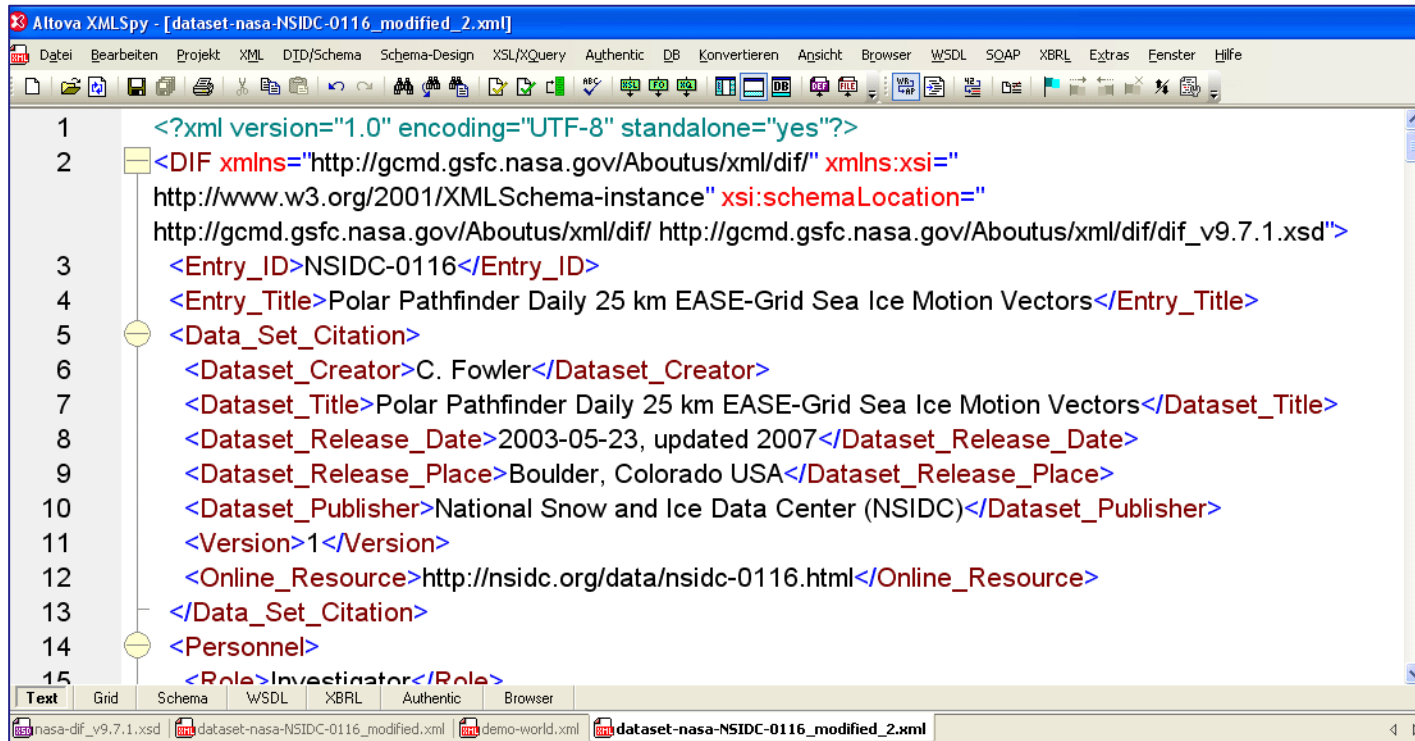
- Name und URL
- Urheber
- Themenbereich
- Umfang
- Verwendetes Vokabular
 - Namensräume
 - Terme (Erläuterungen ?)
- Aktuelle LOD-Verknüpfungen
- Zugangsart
 - SWBrowser (z.B. „Tabulator“)
 - „SPARQL-Endpoint“

<http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/DataSets>

Vorabexperimente

Ausgangspunkt: **Beispieldatensatz NSIDC-0116 im XML-Format**

<http://nsidc.org/data/nsidc-0116.html>



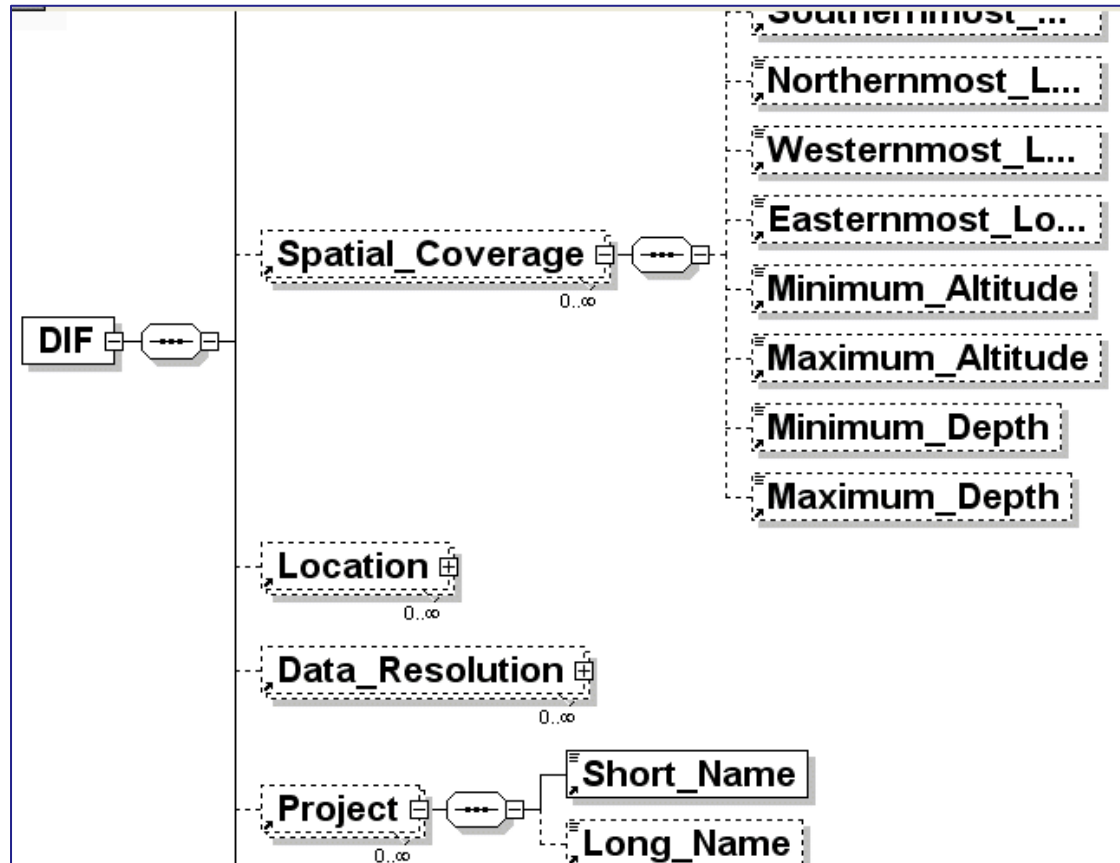
The screenshot shows the Altova XMLSpy interface with the following XML content:

```
1 <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2 <DIF xmlns="http://gcmd.gsfc.nasa.gov/Aboutus/xml/dif/" xmlns:xsi="
  http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="
  http://gcmd.gsfc.nasa.gov/Aboutus/xml/dif/ http://gcmd.gsfc.nasa.gov/Aboutus/xml/dif/dif_v9.7.1.xsd">
3   <Entry_ID>NSIDC-0116</Entry_ID>
4   <Entry_Title>Polar Pathfinder Daily 25 km EASE-Grid Sea Ice Motion Vectors</Entry_Title>
5   <Data_Set_Citation>
6     <Dataset_Creator>C. Fowler</Dataset_Creator>
7     <Dataset_Title>Polar Pathfinder Daily 25 km EASE-Grid Sea Ice Motion Vectors</Dataset_Title>
8     <Dataset_Release_Date>2003-05-23, updated 2007</Dataset_Release_Date>
9     <Dataset_Release_Place>Boulder, Colorado USA</Dataset_Release_Place>
10    <Dataset_Publisher>National Snow and Ice Data Center (NSIDC)</Dataset_Publisher>
11    <Version>1</Version>
12    <Online_Resource>http://nsidc.org/data/nsidc-0116.html</Online_Resource>
13  </Data_Set_Citation>
14  <Personnel>
15    <Role>Investigator</Role>
```

http://nsidc.org/cgi-bin/get_metadata.pl?id=NSIDC-0116&format=DIF&style=XML

[http://gcmd.nasa.gov/KeywordSearch/Metadata.do?Portal=GCMD&KeywordPath=\[Sensor_Name%3A+Short_Name%3D%27SMR%27\]&OrigMetadataNode=GCMD&EntryId=NSIDC-0116&MetadataView=Full&MetadataType=0&Ibnode=mdlb2](http://gcmd.nasa.gov/KeywordSearch/Metadata.do?Portal=GCMD&KeywordPath=[Sensor_Name%3A+Short_Name%3D%27SMR%27]&OrigMetadataNode=GCMD&EntryId=NSIDC-0116&MetadataView=Full&MetadataType=0&Ibnode=mdlb2)

Ausgangspunkt: DIF XML-Schema: dif_v9.7.1.xsd



http://gcmd.gsfc.nasa.gov/Aboutus/xml/dif/dif_v9.7.1.xsd

XML-Datensatz in Relationalem DB-Schema

The screenshot shows the phpMyAdmin interface for a MySQL database. The left sidebar shows the database 'pe25demo-nsidc-0116' and a list of tables, with 'dif' selected. The main area displays the table structure for 'dif'.

	Feld	Typ	Kollation	Attribute	Null	Standard
<input type="checkbox"/>	Primärschlüssel	varchar(255)	utf8_unicode_ci		Ja	NULL
<input type="checkbox"/>	xmlns	varchar(255)	utf8_unicode_ci		Ja	NULL
<input type="checkbox"/>	xsi	varchar(255)	utf8_unicode_ci		Ja	NULL
<input type="checkbox"/>	schemaLocation	varchar(255)	utf8_unicode_ci		Ja	NULL
<input type="checkbox"/>	Entry_ID	varchar(255)	utf8_unicode_ci		Ja	NULL
<input type="checkbox"/>	Entry_Title	varchar(255)	utf8_unicode_ci		Ja	NULL
<input type="checkbox"/>	ISO_Topic_Category	varchar(255)	utf8_unicode_ci		Ja	NULL
<input type="checkbox"/>	Keyword	varchar(255)	utf8_unicode_ci		Ja	NULL
<input type="checkbox"/>	Data_Set_Progress	varchar(255)	utf8_unicode_ci		Ja	NULL
<input type="checkbox"/>	Quality	varchar(255)	utf8_unicode_ci		Ja	NULL
<input type="checkbox"/>	Data_Set_Language	varchar(255)	utf8_unicode_ci		Ja	NULL
<input type="checkbox"/>	Reference	varchar(255)	utf8_unicode_ci		Ja	NULL
<input type="checkbox"/>	Summary	varchar(255)	utf8_unicode_ci		Ja	NULL

XML-Datensatz in Relationalem DB-Schema:

Beispiel: contact_address

Primärschlüssel	Address	City	Province_or_State	Postal_Code	Country	FK_Person
1	Colorado Center for Astrodynamics Research	Boulder	CO	80309-0431	NULL	1
2	National Snow and Ice Data Center	Boulder	CO	80309-0449	USA	2
3	National Snow and Ice Data Center	Boulder	CO	80309-0449	USA	3
4	National Snow and Ice Data Center	Boulder	CO	80309-0449	USA	4


personnel

Primärschlüssel	Role	First_Name	Last_Name	Email	Phone	Fax	FK_DIF
1	Investigator	Chuck	Fowler	NULL	NULL	NULL	1
2	Technical Contact	NSIDC	USER SERVICES	nsidc@nsidc.org	+1 (303) 492-6199	+1 (303) 492-2468	1
3	Data Center Contact	NSIDC	User Services	nsidc@nsidc.org	+1 (303) 492-6199	+1 (303) 492-2468	NULL
4	Data Center Contact	NSIDC	User Services	nsidc@nsidc.org	+1 (303) 492-6199	+1 (303) 492-2468	NULL

3. Konversion DB => RDF: Tool D2R (FU-Berlin)

D2R Server

Publishing Relational Databases on the Semantic Web

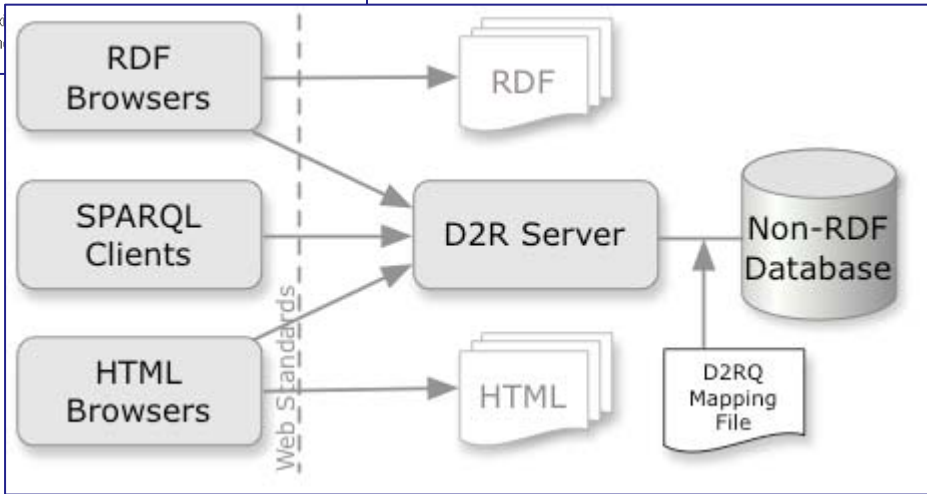

[Chris Bizer](#)
[Richard Cyganiak](#)

D2R Server is a tool for publishing relational databases on the Semantic Web. It enables RDF and HTML browsers to navigate the content of the database, and allows applications to query the database using the SPARQL query language.

[Download D2R Server](#) [Live Demo](#)
v0.7 (alpha), released 2009-08-10 Web-based Systems Group database

News

- **2009-08-10: Version 0.7 released.** This version provides several bugfixes, better dump performance, several new features as well as new optimizations that must be enabled using the new `--fast` switch.
- **2009-02-19: Version 0.6 released.** This version introduces serving of vocabulary data and includes D2RQ 0.6, which provides significantly improved performance and memory usage, new features and several bugfixes.
- **2007-11-03: Version 0.4 released.** This version can be run as a J2EE web application inside ex
- **2007-02-13: Version 0.3.2 released,** featuring more configurability, easy installation as a Win



Zwischenergebnis 1: Datensatz NSIDC-0116 in lokalem D2R-Server

<http://localhost:2020/>

D2R Server
Running at <http://localhost:2020/>

Home | [contact](#) | [address](#) | [data](#) | [center](#) | [data](#) | [center](#) | [name](#) | [data](#) | [resolution](#) | [data](#) | [set](#) | [citation](#) | [dif](#) | [distribution](#) | [idn](#) | [node](#) | [location](#)
[parameters](#) | [personnel](#) | [project](#) | [sensor](#) | [name](#) | [source](#) | [name](#) | [spatial](#) | [coverage](#) | [temporal](#) | [coverage](#)

This is a database published with D2R Server. It can be accessed using

1. your plain old web browser
2. Semantic Web browsers
3. SPARQL clients.

1. HTML View
You can use the navigation links at the top of this page

2. RDF View

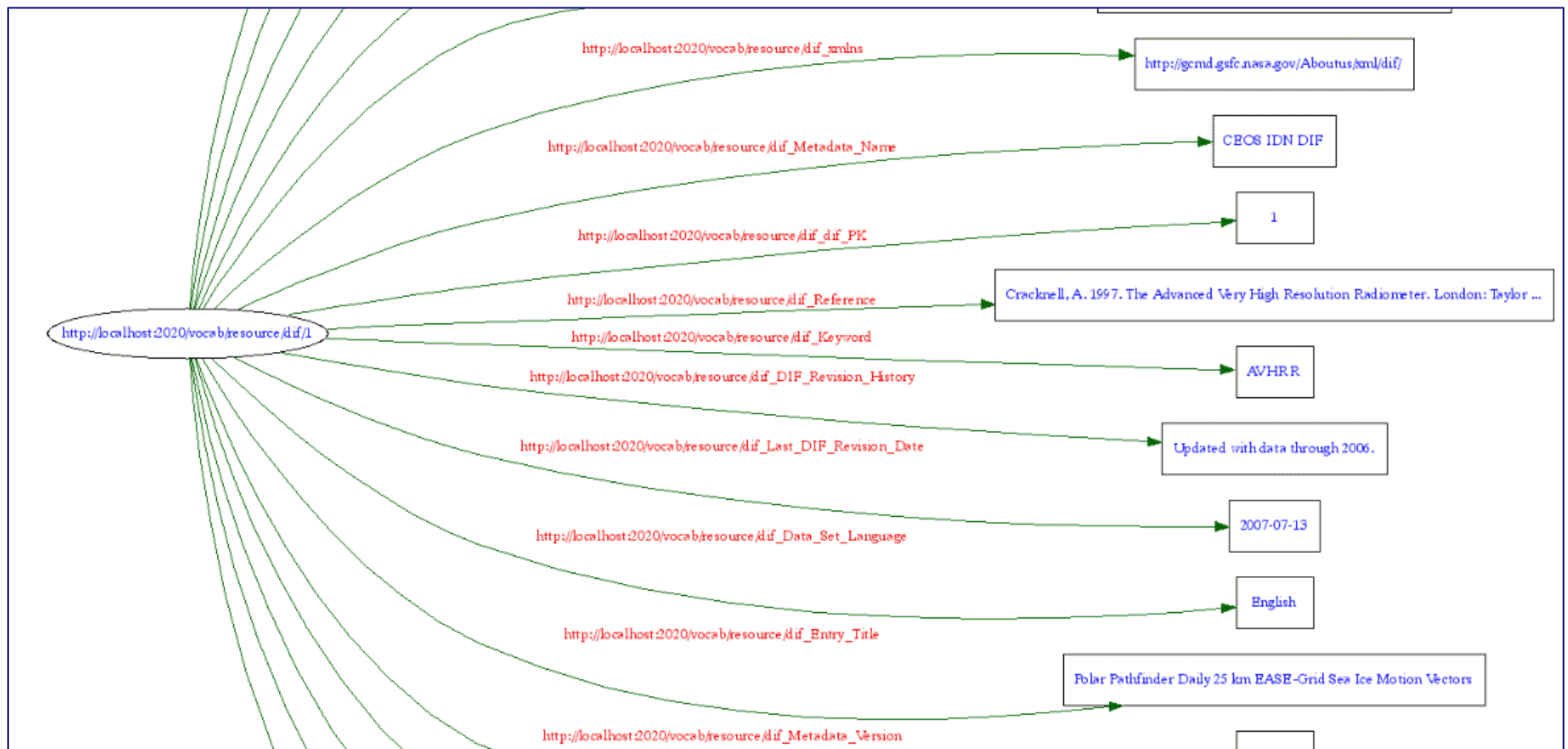
dif #1
Resource URI: <http://localhost:2020/resource/dif/1>

Home | [All dif](#)

Property	Value
vocab:dif_DIF_Creation_Date	2003-05-23
vocab:dif_DIF_Revision_History	Updated with data through 2006.
vocab:dif_Data_Set_Language	English
vocab:dif_Data_Set_Progress	COMPLETE
vocab:dif_Entry_ID	NSIDC-0116
vocab:dif_Entry_Title	Polar Pathfinder Daily 25 km EASE-Grid Sea Ice Motion Vectors
vocab:dif_ISO_Topic_Category	Climatology/Meteorology/Atmosphere
vocab:dif_Keyword	AVHRR
vocab:dif_Last_DIF_Revision_Date	2007-07-13
vocab:dif_Metadata_Name	CEOS IDN DIF
vocab:dif_Metadata_Version	9.7
vocab:dif_Quality	Direct comparisons between the more accurate buoy...

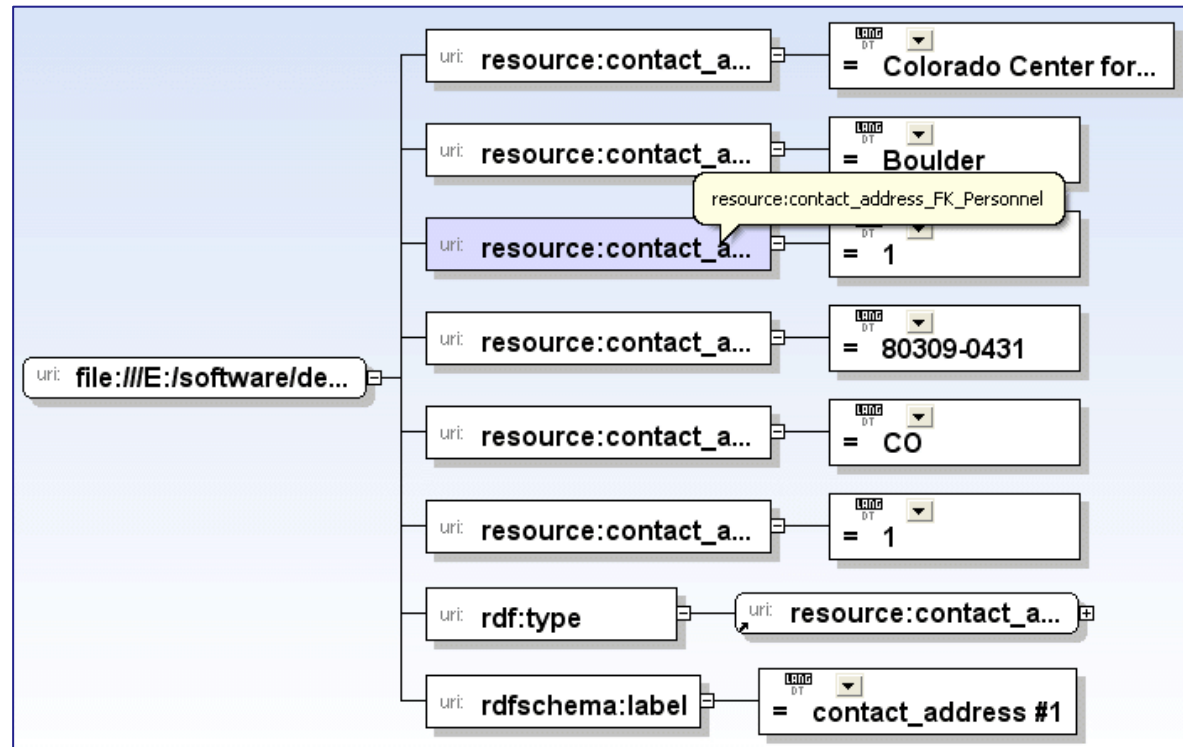
Rohergebnis 3: Datensatz NSIDC-0116 als Graph

Visualisierung mit W3C-RDF-Validator: <http://www.w3.org/RDF/Validator>

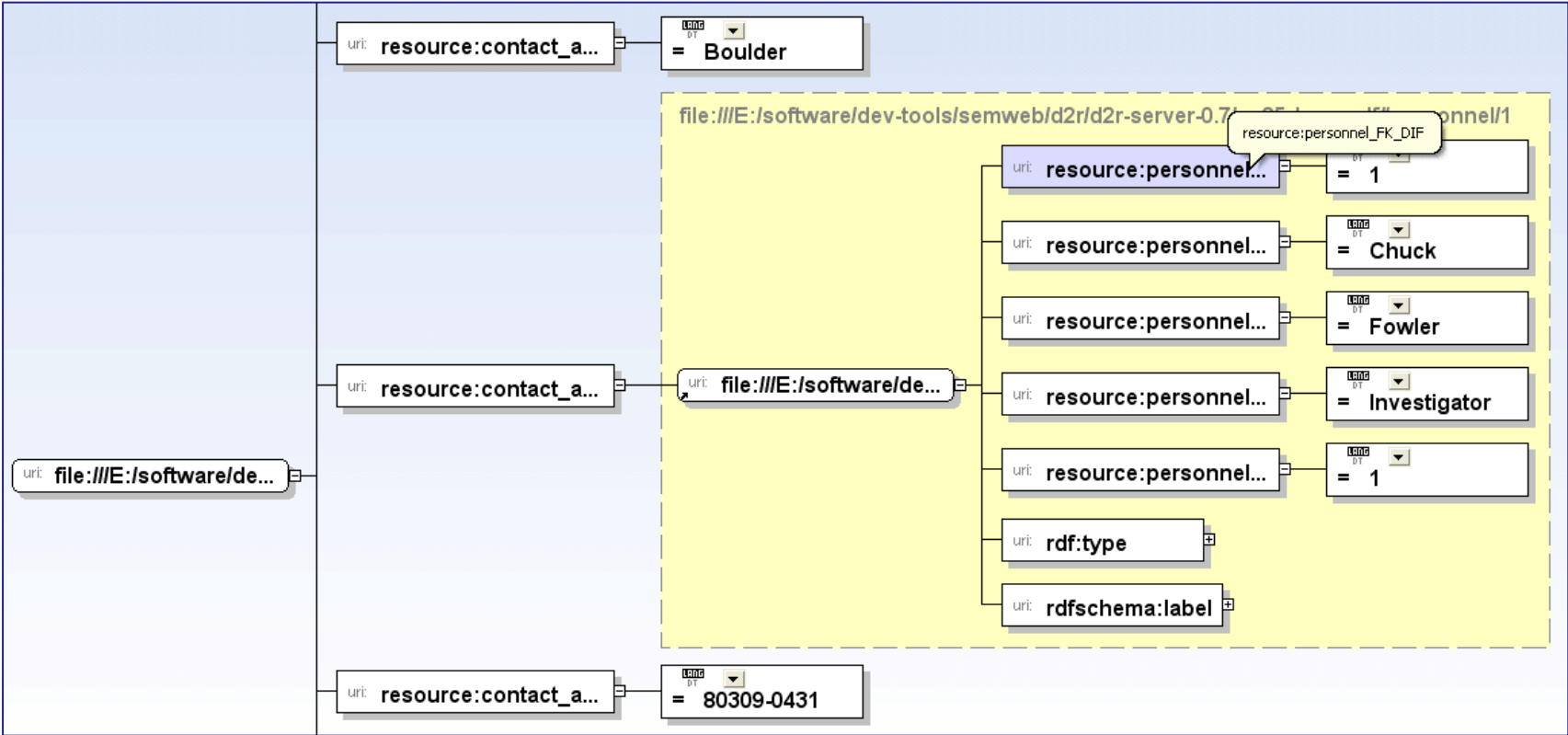


Weiterführend: Nachbearbeitung des RDF-Datensatzes

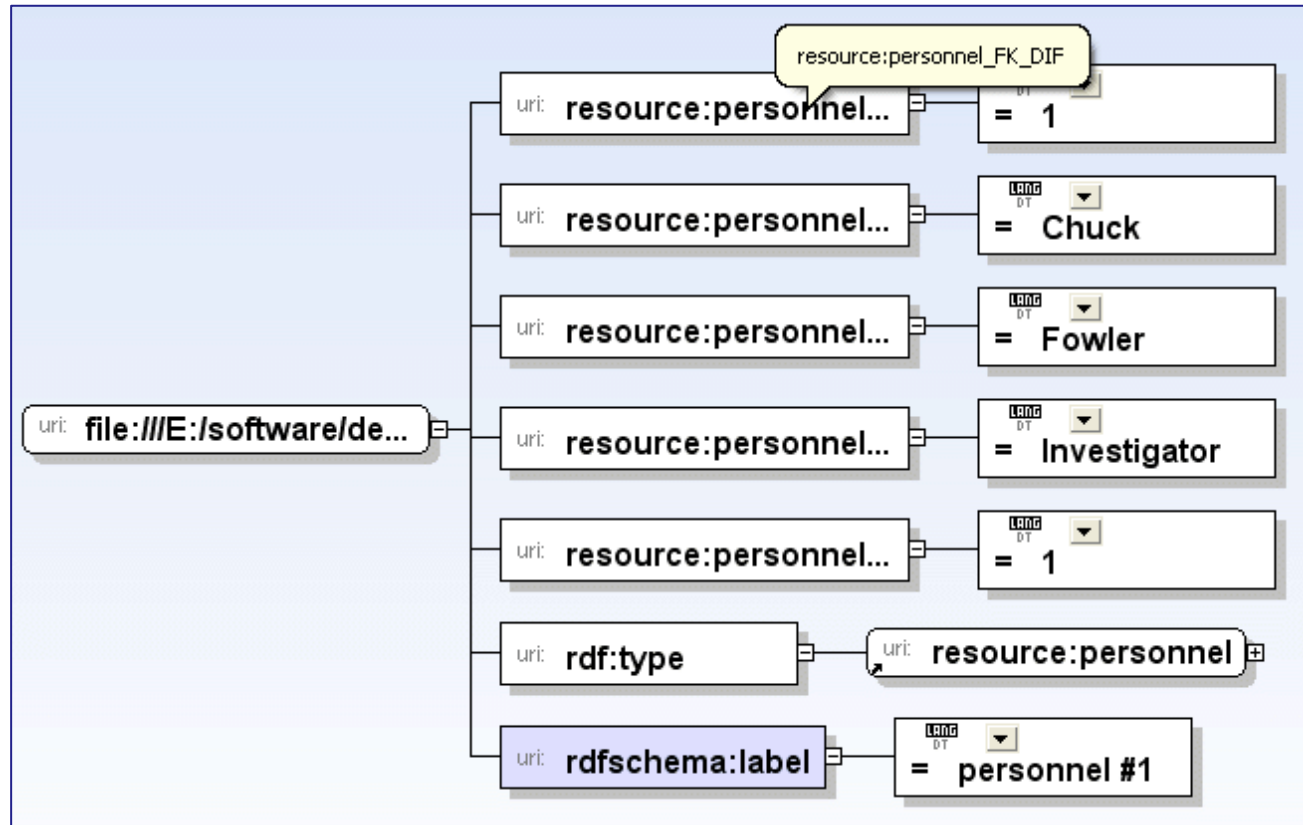
Graphisch z.B. mit **ALTOVA SemanticWorks**



Ersetze Literal „1“ bei „contact_address_FK_Personnel“ durch „ressource“
„personnel/1“



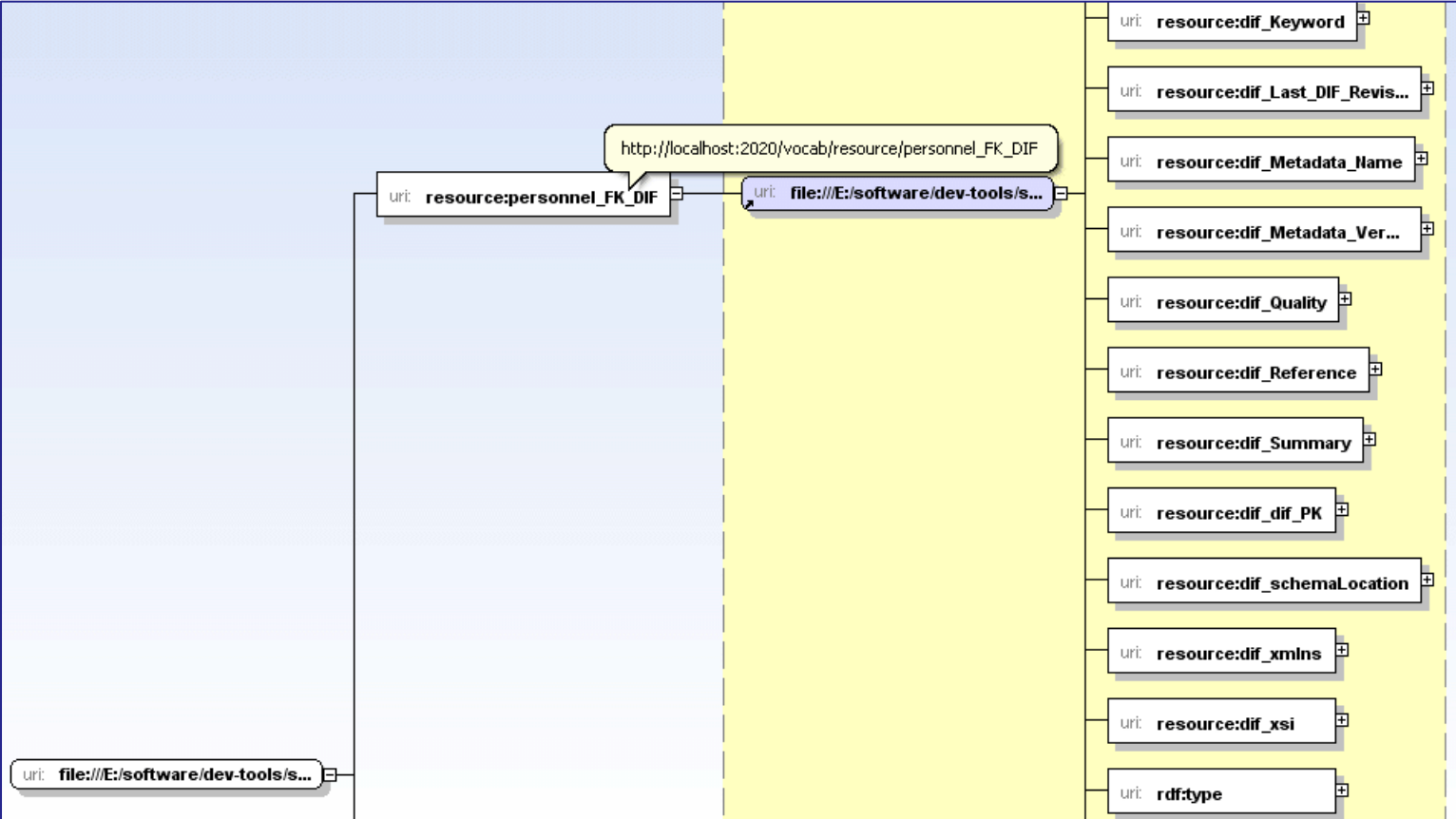
Weiterführend: Nachbearbeitung des RDF-Datensatzes



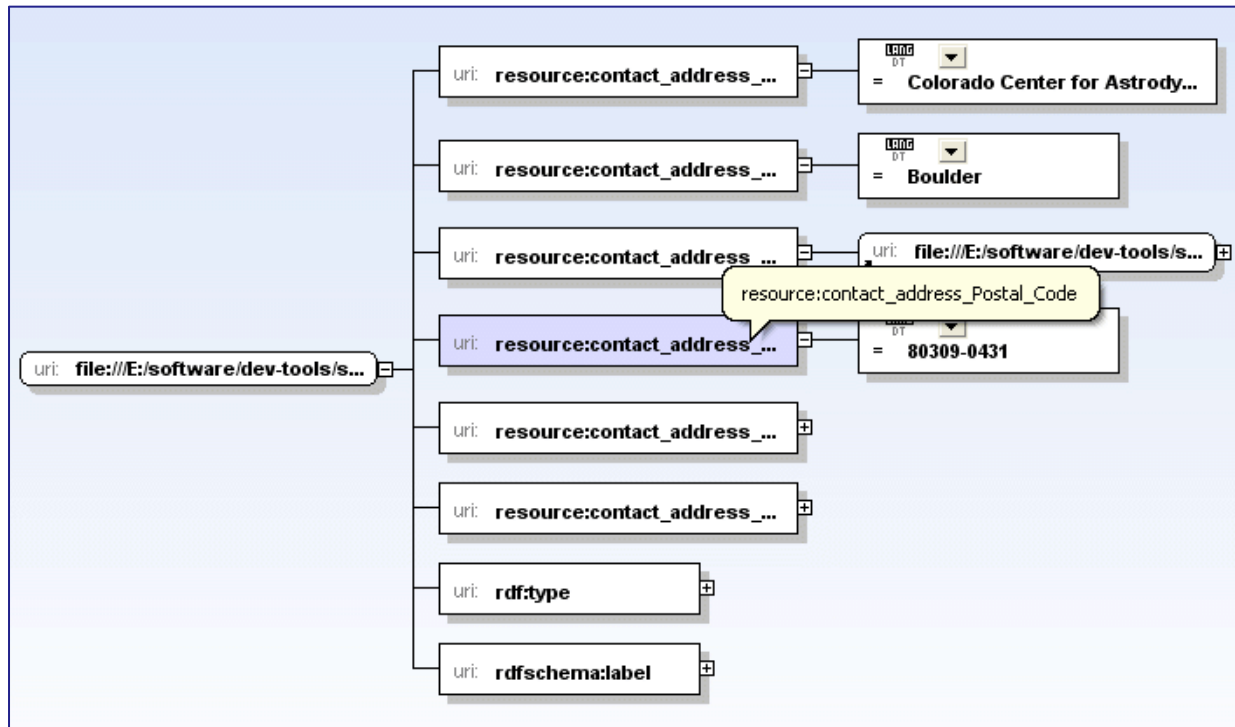
Ersetze Literal „1“ bei „personnel_FK_DIF“ durch „ressource“ „dif/1“

Weiterführend: Nachbearbeitung des RDF-Datensatzes

Ergebnis:



Nachbearbeitung des RDF-Datensatzes: **Externe Verknüpfungen**



z.B. DIF: **contact_address_Postal_Code:**

Möglicherweise sinnvolle Verknüpfung mit Geonames-Vokabular:

<http://www.schemaweb.info/schema/SchemaInfo.aspx?id=296>

z.B. Postal Code: <http://www.geonames.org/ontology#postalCode>

<http://www.geonames.org/export/ws-overview.html>

Erstes Zwischenergebnis

The screenshot shows an RDF document viewer interface. At the top, the file path is `file:///F:/FHPotsdam/lehrveranstaltungen/lv-ss2010/pe-semwebss10/TB2/experimente/results/pe25demo-xml.V3.rdf`. The document type is identified as an RDFDocument. The main content is a tree structure under the label 'Mentions'. The root node is `contact_address`. It has a child node `contact_address #1`. This node has three properties: `Contact_address Address` with value 'Colorado Center for Astrodynamics Research', `Contact_address City` with value 'Boulder', and `Contact_address FK_Personnel` with value `personnel #1`. The `personnel #1` node has several properties: `Personnel FK_DIF` (value `dif #1`), `Personnel First_Name` (value 'Chuck'), `Personnel Last_Name` (value 'Fowler'), `Personnel Role` (value 'Investigator'), `Personnel personnel_PK` (value '1'), `Type` (value `personnel`), and `Label` (value 'personnel #1'). Below these, there is a property `Is contact_address FK_Personnel of` with value `contact_address #1`. Further down, the `contact_address #1` node has `Contact_address Postal_Code` (value '80309-0431') and `Contact_address Province_or_State` (value 'CO'). A `Postal Code` property has a value of `http://ws.geonames.org/postalCodeSearch?postalcode=80309&placename=Boulder`. The `Type` is `contact_address` and the `Label` is `contact_address #1`. At the bottom, there are four more `contact_address` nodes labeled #2, #3, and #4.