

# Datenverknüpfung und (Meta)Datenmigration: Verfahren und Technologien

Günther Neher / Felix Sasaki

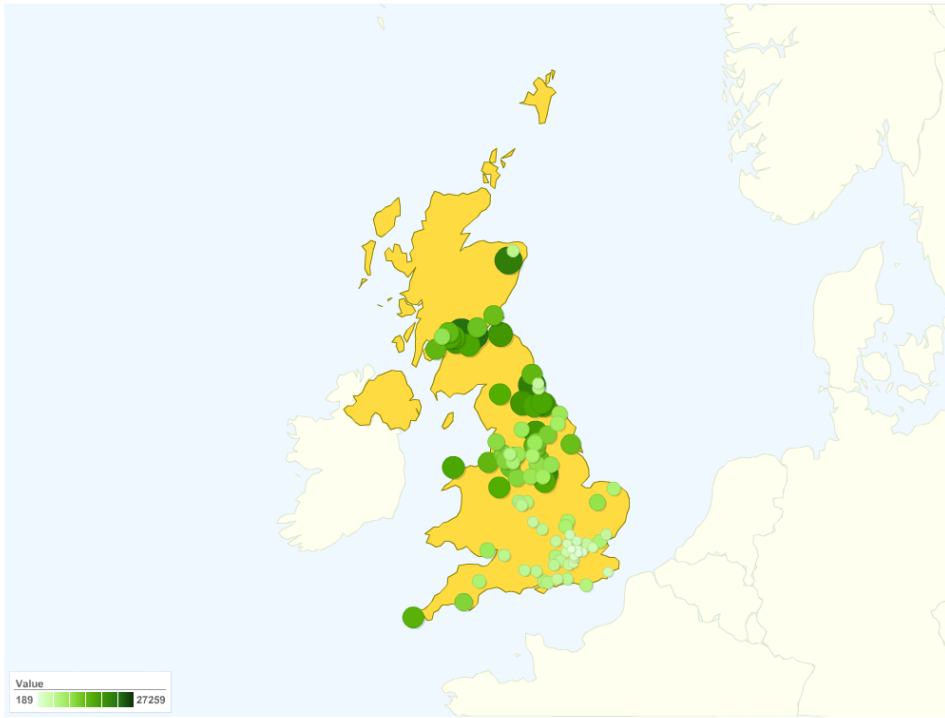
FH Potsdam, Fachbereich  
Informationswissenschaften

Demo von Jeni Tennison: Reisekosten  
von Parlamentariern in Großbritannien  
dynamisch aus diversen Datenquellen  
generiert & visualisiert

<http://tinyurl.com/jt-mp-expenses>

## MPs Travel Expenses

This map shows the relative total travel expenses of the 100 MPs with the lowest majorities. The larger/darker circles indicate higher expenses.



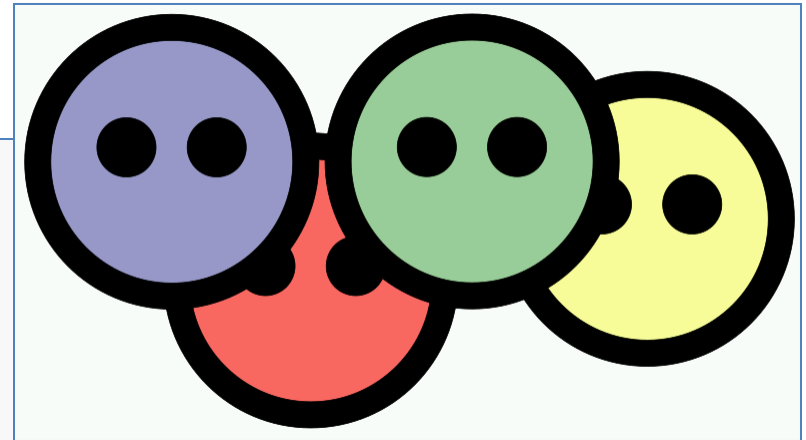
These charts show the highest 25 expense claims in the three main parties.

DE Deutsch (Deutschland)

**WAS MUSS MAN WISSEN UM SO  
ETWAS UMZUSETZEN?**

# Nötige Informationen: Allgemeine Datenquellen

- Beispiel DBpedia & FOAF
  - Repräsentation von Wikipedia in maschinenlesbarer Form
  - Information zu Personen



# Nötige Informationen: Allgemeine Datenquellen

- Beispiel WGS84 Geo Positioning Vocabulary
  - Web-Format zur Repräsentation von Geo-Koordinaten

## WGS84 Geo Positioning: an RDF vocabulary

A vocabulary for representing latitude, longitude and altitude information in the WGS84 geodetic reference datum. Version ID: wgs84\_pos.rdf.v 1.22 2009/04/20 15:00:30 timbl Exp \$. See <http://www.w3.org/2003/01/geo/> for more details.

date: \$Date: 2009/04/20 15:00:30 \$

label: geo

comment: Recent changes to this namespace: \$Log: wgs84\_pos.rdf.v \$ Revision 1.22 2009/04/20 15:00:30 timbl Remove the time bits which have been deal with elsewhere eg in iCal. Revision 1.21 2009/04/20 12:52:47 timbl try again Revision 1.20 2009/04/20 12:42:11 timbl Add Event (edited ages ago and never checked in), and location (following discussion <http://chatlogs.planetrd.com/swig/2009-04-20#T12-36-09>) Revision 1.19 2009/04/20 12:36:31 timbl Add Event (edited ages ago and never checked in), and location (following discussion <http://chatlogs.planetrd.com/swig/2009-04-20#T12-36-09>) Revision 1.18 2006/02/01 22:01:04 danbri Clarified that lat and long are decimal degrees, and that alt is decimal metres about local reference ellipsoid Revision 1.17 2004/02/06 17:38:12 danbri Fixed a bad commit screwup Revision 1.15 2003/04/19 11:24:08 danbri Fixed the typo even more. Revision 1.14 2003/04/19 11:16:56 danbri fixed a typo Revision 1.13 2003/02/19 22:27:27 comolly relaxed domain constraints on lat/long/alt from Point to SpatialThing Revision 1.12 2003/01/12 01:41:41 danbri Trying local copy of XSLT doc. Revision 1.11 2003/01/12 01:20:18 danbri added a link to morten's xslt rdfs viewer. Revision 1.10 2003/01/11 18:56:49 danbri Removed datatype range from lat and long properties, since they would have required each occurrence of the property to mention the datatype. Revision 1.9 2003/01/11 11:41:31 danbri Another typo, repaired rdfs:Property to rdfs:Property r4 Revision 1.8 2003/01/11 11:05:02 danbri Added an rdfs:range for each lat/long/alt property, <http://www.w3.org/2001/XMLSchema#float> Revision 1.7 2003/01/10 20:25:16 danbri Longer rdfs:comment for Point, trying to be Earth-centric and neutral about coordinate system(s) at the same time. Feedback welcomed. Revision 1.6 2003/01/10 20:18:30 danbri Added CVS log comments into the RDF/XML as an rdfs:comment property of the vocabulary. Note that this is not common practice (but seems both harmless and potentially useful). revision 1.5 date: 2003/01/10 20:14:31, author: danbri, state: Exp, lines: +16 -5 Updated schema: Added a dc:date, added url for more info. Changed the rdfs:label of the namespace from gp to geo. Added a class Point, set as the rdfs:domain of each property. Added XML comment on the lat\_long property suggesting that we might not need it (based on #rdfig commentary from implementors) revision 1.4 date: 2003/01/10 20:01:07, author: danbri, state: Exp, lines: +6 -5 Fixed typo; several rdfs:about attributes are now rdfs:about. Thanks to MortenF in #rdfig for catching this error. revision 1.3 date: 2003/01/10 11:59:03, author: danbri, state: Exp, lines: +4 -3 fixed buglet in vocab, added more wgs links revision 1.2 date: 2003/01/10 11:01:11, author: danbri, state: Exp, lines: +4 -4 Removed alt from the as-a-flat-string property, and switched from space separated to comma separated revision 1.1 date: 2003/01/10 10:53:23, author: danbri, state: Exp, basic geo vocab

Namespace: [http://www.w3.org/2003/01/geo/wgs84\\_pos#](http://www.w3.org/2003/01/geo/wgs84_pos#)



### Classes

**SpatialThing** Anything with spatial extent, i.e. size, shape, or position. e.g. people, places, bowling balls, as well as abstract areas like cubes.

[SpatialThing]

**point** A point, typically described using a coordinate system relative to Earth, such as WGS84. Uniquely identified by lat/long/alt. i.e. spatiallyIntersects(P1, P2) - lat(P1, LAT), long(P1, LONG), alt(P1, ALT), lat(P2, LAT), long(P2, LONG), alt(P2, ALT) sameThing(P1, P2) :- type(P1, Point), type(P2, Point), spatiallyIntersects(P1, P2).

[Point] subClassOf: [http://www.w3.org/2003/01/geo/wgs84\\_pos#SpatialThing](http://www.w3.org/2003/01/geo/wgs84_pos#SpatialThing)

### Properties

**latitude** The WGS84 latitude of a SpatialThing (decimal degrees).

[lat] domain: [http://www.w3.org/2003/01/geo/wgs84\\_pos#SpatialThing](http://www.w3.org/2003/01/geo/wgs84_pos#SpatialThing)

**location** The relation between something and the point, or other geometrical thing in space, where it is. For example, the relationship between a radio tower and a Point with a given lat and long. Or a relationship between a park and its outline as a closed arc of points, or a road and its location as a arc (a sequence of points). Clearly in practice there will be limit to the accuracy of any such statement, but one would expect an accuracy appropriate for the size of the object and uses such as mapping.

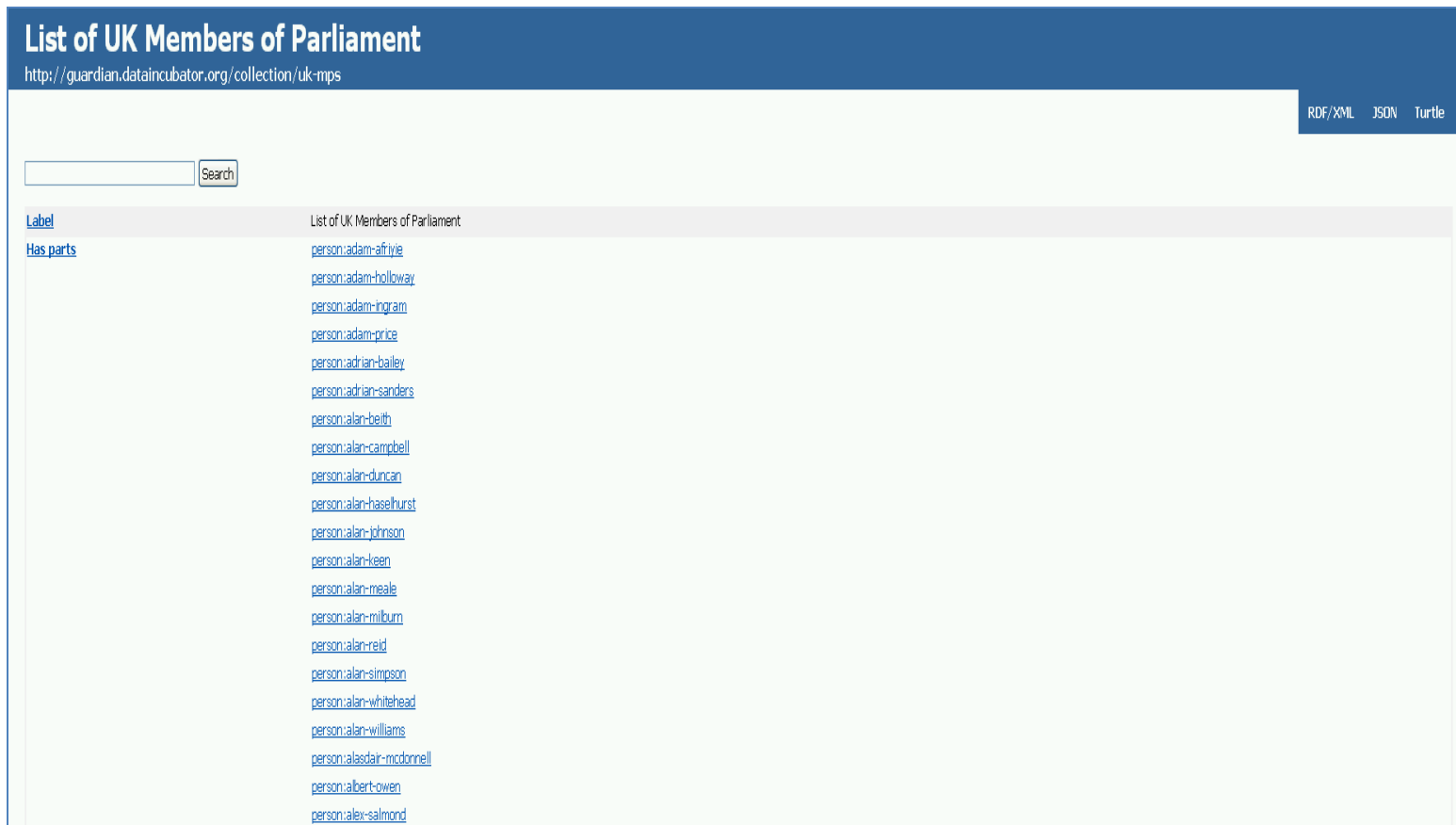
[location] range: [http://www.w3.org/2003/01/geo/wgs84\\_pos#SpatialThing](http://www.w3.org/2003/01/geo/wgs84_pos#SpatialThing)

[location] subPropertyOf: [http://xmlns.com/foaf/0.1/based\\_near](http://xmlns.com/foaf/0.1/based_near)

**longitude** The WGS84 longitude of a SpatialThing (decimal degrees).

# Nötige Informationen: Spezielle Datenquellen

- Beispiel „Guardian MP Dataset“ (1)
  - Daten zu „Member of Parliament“ in Großbritannien




The screenshot displays the 'List of UK Members of Parliament' dataset page. At the top, there is a blue header with the title 'List of UK Members of Parliament' and the URL 'http://guardian.dataincubator.org/collection/uk-mps'. To the right of the header are links for 'RDF/XML', 'JSON', and 'Turtle'. Below the header is a search bar with a 'Search' button. The main content area shows a table with two columns: 'Label' and 'List of UK Members of Parliament'. The 'Label' column contains the text 'Has parts'. The 'List of UK Members of Parliament' column contains a list of 20 URIs, each representing a member of parliament, such as 'person:adam-afrivie', 'person:adam-holloway', 'person:adam-ingram', 'person:adam-price', 'person:adrian-bailey', 'person:adrian-sanders', 'person:alan-beith', 'person:alan-campbell', 'person:alan-duncan', 'person:alan-hasehurst', 'person:alan-johnson', 'person:alan-keen', 'person:alan-meale', 'person:alan-milburn', 'person:alan-reid', 'person:alan-simpson', 'person:alan-whitehead', 'person:alan-williams', 'person:alasdair-mcdonnell', 'person:abert-owen', and 'person:alex-salmond'.

Label	List of UK Members of Parliament
Has parts	<a href="#">person:adam-afrivie</a>
	<a href="#">person:adam-holloway</a>
	<a href="#">person:adam-ingram</a>
	<a href="#">person:adam-price</a>
	<a href="#">person:adrian-bailey</a>
	<a href="#">person:adrian-sanders</a>
	<a href="#">person:alan-beith</a>
	<a href="#">person:alan-campbell</a>
	<a href="#">person:alan-duncan</a>
	<a href="#">person:alan-hasehurst</a>
	<a href="#">person:alan-johnson</a>
	<a href="#">person:alan-keen</a>
	<a href="#">person:alan-meale</a>
	<a href="#">person:alan-milburn</a>
	<a href="#">person:alan-reid</a>
	<a href="#">person:alan-simpson</a>
	<a href="#">person:alan-whitehead</a>
	<a href="#">person:alan-williams</a>
	<a href="#">person:alasdair-mcdonnell</a>
	<a href="#">person:abert-owen</a>
	<a href="#">person:alex-salmond</a>

# Nötige Informationen: Spezielle Datenquellen

- Beispiel „Guardian MP Dataset“ (2)
  - Umfasst auch Reisekosten der Parlamentarier

<a href="#">Name</a>	Alex Salmond
<a href="#">Types</a>	<a href="#">msg0:MemberOfParliament</a> <a href="#">foaf:Person</a>
<a href="#">Picture</a>	
<a href="#">mp-expenses:pdfs</a>	<a href="http://mps-expenses.guardian.co.uk/mp/579/">http://mps-expenses.guardian.co.uk/mp/579/</a>
<a href="#">mp-expenses:mp-rail</a>	0 [xsd:int]
<a href="#">mp-expenses:mp-other-mileage</a>	0 [xsd:int]
<a href="#">mp-expenses:mp-family-journeys</a>	0 [xsd:int]
<a href="#">mp-expenses:total-travel</a>	6170 [xsd:int]
<a href="#">mp-expenses:mp-misc</a>	0 [xsd:int]
<a href="#">mp-expenses:stationery-associated-postage-costs</a>	4768 [xsd:int]
<a href="#">mp-expenses:total-claim</a>	139233
<a href="#">mp-expenses:office-running-costs</a>	19808 [xsd:int]
<a href="#">Is referenced by</a>	<a href="http://www.guardian.co.uk/commentisfree/2007/feb/04/comment.politics">http://www.guardian.co.uk/commentisfree/2007/feb/04/comment.politics</a>

# Möglichkeit der Verarbeitung: Abfrage mit SPARQL

- Abfragesprache für Datenquellen im RDF-Format
- Ergebnis einer Abfrage im JSON-Format, zur folgenden Visualisierung (Ausschnitt)

```
...n.dataincubator.org/person/stephen-williams'),(v:'Stephen Williams'),(v:5128),(v:'http://guardian.dataincubator.org/constituency/bristol-west'),(v:'Bristol West'),(v:'http://dbpedia.org/resource/Liberal_Democrats'),(v:51.4751007652),(v:-2.609785629314),(v:139393),(v:6192),(v:20721),(v:145585)],(c:[(v:'http://guardian.dataincubator.org/person/alex-salmond'),(v:'Alex Salmond'),(v:11837),(v:'http://guardian.dataincubator.org/constituency/banff-and-buchan'),(v:'Banff and Buchan'),(v:'http://dbpedia.org/resource/Scottish_National_Party'),(v:57.5345041469),(v:-2.286256783714),(v:133063),(v:6170),(v:14130),(v:139233)],(c:[(v:'http://guardian.dataincubator.org/person/bob-russell'),(v:'Bob Russell'),(v:6277),(v:'http://guardian.dataincubator.org/constituency/colchester'),(v:'Colchester'),(v:'http://dbpedia.org/resource/Liberal_Democrats'),(v:51.8875986796),(v:0.875784880164),(v:138217),(v:6156),(v:10963),(v:144103)],(c:[(v:'http://guardian.dataincubator.org/person/bob-russell'),(v:'Bob Russell'),(v:6277),(v:'http://guardian.dataincubator.org/constituency/colchester'),(v:'Colchester'),(v:'http://dbpedia.org/resource/Liberal_Democrats'),(v:51.8875986796),(v:0.875784880164),(v:138217),(v:6156),(v:10963),(v:144373)]],(c:[(v:'http://guardian.dataincubator.org/person/huw-irranca-
```

# Möglichkeit der Darstellung: Visualisierungs-API

- Beispiel Google Visualization API
- Ausgabe der SPARQL-Abfrage wird dem API übergeben

The screenshot shows the Google Visualization API Gallery page. At the top, there is a search bar with the text "Google code" and a search button. Below the search bar, the page title is "Google Chart Tools / Interactive Charts (aka Visualization API)". The page is divided into a left sidebar and a main content area. The sidebar contains links for "Visualization API", "Galleries", "Google Spreadsheets", "Submit", and "Google Chart Tools". The main content area is titled "Google Visualization API Gallery" and contains a description of the gallery, a list of visualization types, and a list of specific visualization examples. The examples include "Annotated Time Line", "Area Chart (Image)", "Area Chart", "Bar Chart (Image)", "Bar Chart", and "Bars of Stuff". Each example includes a small thumbnail image of the chart, a title, a description, and a link to the documentation or source code.

Google code  Search  
e.g. "ajax apis" or "open source"

Google Chart Tools / Interactive Charts (aka Visualization API) [Home](#) [Docs](#) [FAQ](#) [Forum](#) [Terms](#)

Visualization API  
[What's New?](#)  
[Release Notes](#)  
[Issues & Feature Requests](#)  
④ [Using Visualizations](#)  
④ [Creating Visualizations](#)  
④ [Implementing a Data Source](#)  
④ [Code Samples](#)  
④ [Reference](#)

Galleries  
Visualization Gallery  
[Gadget Gallery](#)  
[Data Sources and Tools Gallery](#)

Google Spreadsheets  
[Spreadsheet Integration](#)

Submit  
[Submit a Visualization](#)  
[Submit a Gadget](#)

Google Chart Tools  
[What are the Chart Tools?](#)  
[Choosing a Chart Tool](#)

## Google Visualization API Gallery

This gallery lists JavaScript visualizations built on the Google Visualization API. Some of these have been written by Google, and some have been written by third parties. Links below point to instructions for and demonstrations of each visualization. (Links to third-party visualizations will take you off the Google site.) To learn how to use JavaScript visualizations, read [Using Visualizations](#).

To see a list of visualization gadgets, see the [Google Visualization API Gadget Gallery](#).

If you have developed your own visualization, we encourage you to [submit](#) it to this gallery. Note that you should test any visualizations you use on your target browsers to ensure that they work as expected. Older browsers may not support all visualizations.

*Click a visualization below to see documentation and examples.*

Filter Chart Types:

	<b>Annotated Time Line</b> An animated time series chart. By: Google	<a href="#">GWT Integrated</a>
	<b>Area Chart (Image)</b> Area chart using <a href="#">Google Chart API</a> . By: Google	
	<b>Area Chart</b> Interactive area chart. By: Google	<a href="#">GWT Integrated</a>
	<b>Bar Chart (Image)</b> Bar chart using <a href="#">Google Chart API</a> . By: Google	
	<b>Bar Chart</b> Interactive bar chart. By: Google	<a href="#">GWT Integrated</a>
	<b>Bars of Stuff</b> Fun bar charts using images of trains, chocolate, worms, and more. By: The visapi project	

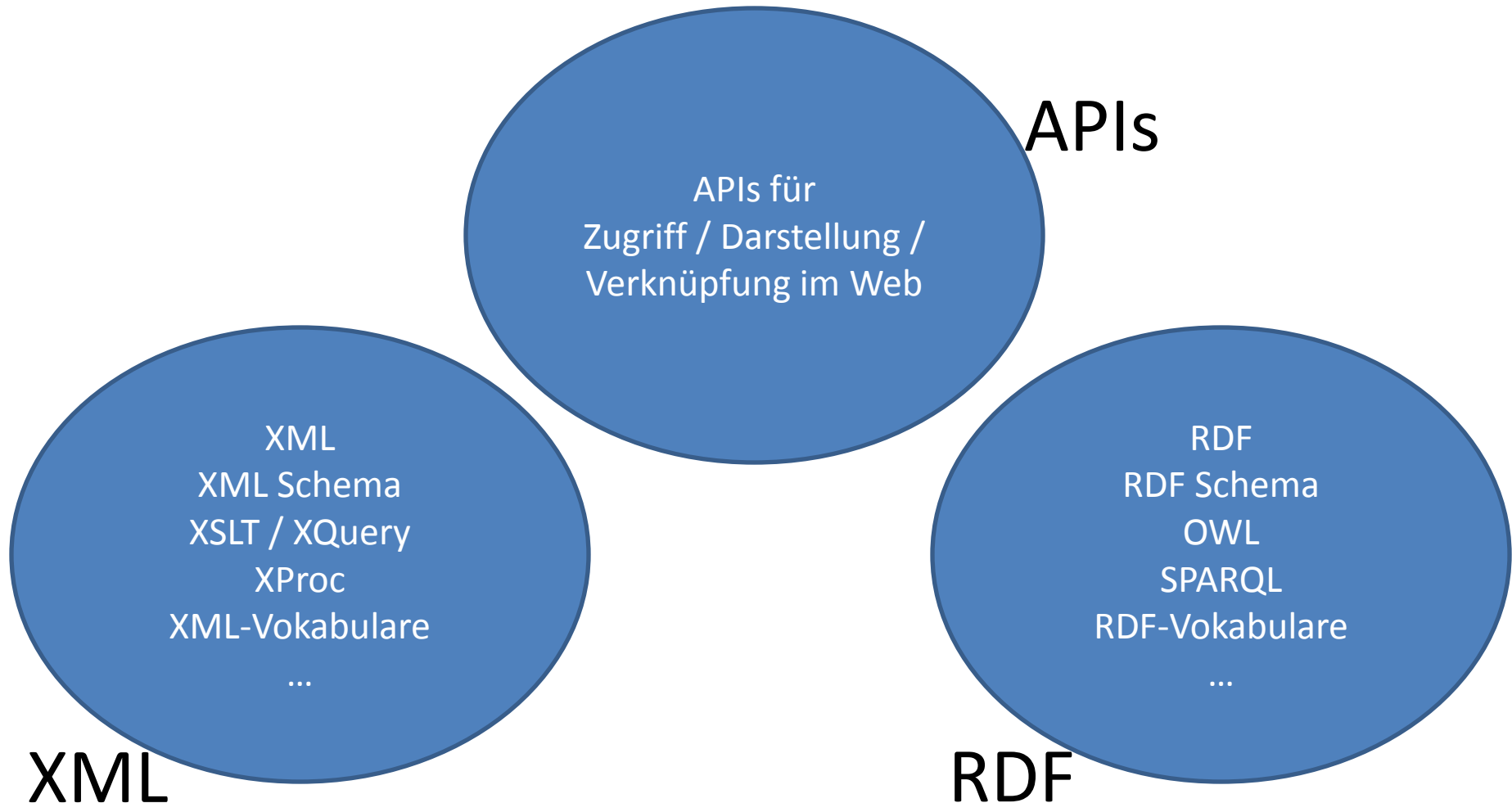
# Häufiger Verarbeitungsschritt: (Meta)datenmigration (Beispiel)

- „Konvertiere mal diese Datenbank in RDF ...“
- „Bei meinen 500+ (Word / PDF / Format XYZ)-Dokumenten muss ein Metadatenfeld hinzugefügt werden ...“
- „Die Bildmetadaten liegen in Excel vor ... konvertiere sie nach XMP / füge sie in die Bilder ein ...“

→ Häufig auf XML-basierende

Metadaten-Workflows

# Basistechnologien für Datenverknüpfung und –migration



# Wir ...

- bieten
  - Kenntnisse zu Verfahren und Technologien für Datenverknüpfung und (Meta)Datenmigration
  - Nähe zu relevanten technologischen Neuentwicklungen im Web:



- suchen
  - Anwendungsszenarien die von Datenverknüpfung und -migration profitieren

**Vielen Dank für Ihre Aufmerksamkeit!**