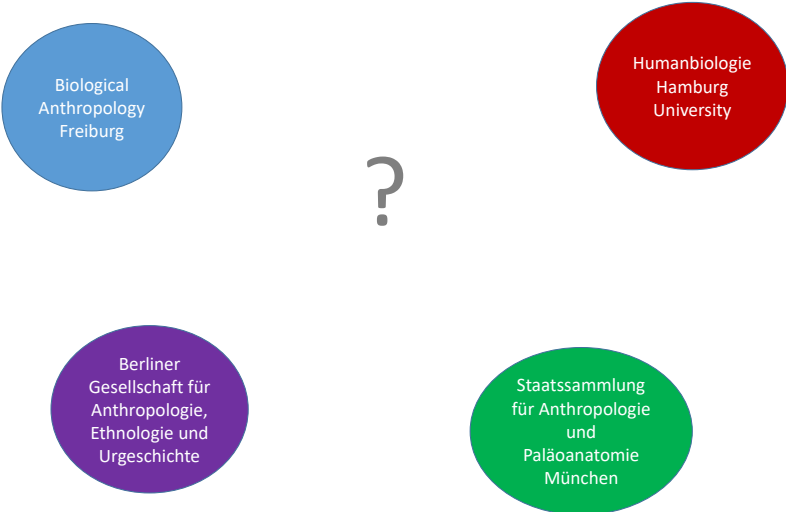


# Data Integration and RDF

Georg Lausen  
Department of Computer Science  
Faculty of Engineering  
Albert-Ludwigs-University of Freiburg

# 1. Motivation



Biological  
Anthropology  
Freiburg

Humanbiologie  
Hamburg  
University

?

Berliner  
Gesellschaft für  
Anthropologie,  
Ethnologie und  
Urgeschichte

Staatsammlung  
für Anthropologie  
und  
Paläoanatomie  
München

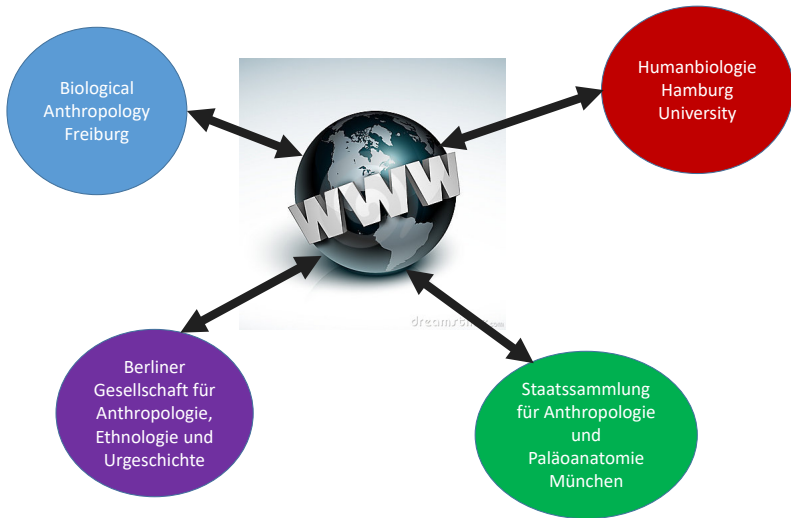
Hallo  
KollegIn!

Biological  
Anthropology  
Freiburg

Humanbiologie  
Hamburg  
University

Berliner  
Gesellschaft für  
Anthropologie,  
Ethnologie und  
Urgeschichte

Staatssammlung  
für Anthropologie  
und  
Paläoanatomie  
München



## 2. Data Integration

- ▶ The Web is the common platform which can be used by nearly everybody to publish and access data.
- ▶ Data from different sources needs to be integrated. The goal is to achieve interoperability based on URIs.
- ▶ Data integration does not come for free!
- ▶ ... its one of the oldest still not sufficiently solved problems of Computer Science.
- ▶ ... but fortunately, today we can build on many useful tools and technologies.
- ▶ Data integration will remain a complicated task because the application problem behind usually is complicated!

### 3. Data Representation

#### Tabular Representation: One institute

Epoch	Collection	
	Region	Curator
Prehistory	Greece	Adam
Antiquity	Greece	Mary
Prehistory	Rome	Peter
Antiquity	Rome	Mary

## Tabular Representation: Two institutes

Epoch	Collection	
	Region	Curator
Prehistory	Greece	Adam
Antiquity	Greece	Mary
Prehistory	Rome	Peter
Antiquity	Rome	Mary

Epoch	Collection	
	Greece	Rome
Prehistory	Adam	Peter
Antiquity	Mary	Mary

## Tabular Representation: Three institutes

Epoch	Collection Region	Curator
Prehistory	Greece	Adam
Antiquity	Greece	Mary
Prehistory	Rome	Peter
Antiquity	Rome	Mary

Epoch	Collection Greece	Rome
Prehistory	Adam	Peter
Antiquity	Mary	Mary

	Collection	
Region	Prehistory	Antiquity
Greece	Adam	Mary
Rome	Peter	Mary



## Tabular Representation: Four institutes

Epoch	Collection	
	Region	Curator
Prehistory	Greece	Adam
Antiquity	Greece	Mary
Prehistory	Rome	Peter
Antiquity	Rome	Mary

Epoch	Collection	
	Greece	Rome
Prehistory	Adam	Peter
Antiquity	Mary	Mary

Region	Collection	
	Prehistory	Antiquity
Greece	Adam	Mary
Rome	Peter	Mary

Collection			
Greece		Rome	
Epoch	Curator	Epoch	Curator
Prehistory	Adam	Prehistory	Peter
Antiquity	Mary	Antiquity	Mary

Data integration is a difficult task!

Must be directed by an agreed upon ontology.

## 4. Ontology

- ▶ Thomas R. Gruber (1993):

*An ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an ontology is a systematic account of Existence. For knowledge-based systems, what "exist" is exactly that which can be represented.*

- ▶ [https://en.wikipedia.org/wiki/Conceptualization\\_\(information\\_science\)](https://en.wikipedia.org/wiki/Conceptualization_(information_science)):  
*In information science a conceptualization is an abstract simplified view of some selected part of the world, . . .*

*An explicit specification of a conceptualization is an ontology, and it may occur that a conceptualization can be realized by several distinct ontologies.*

## why not use Relational Databases?

- ▶ Relational Databases are a great well established technology. They are based on tables, which everybody can understand.
- ▶ Limitations:
  - ▶ Missing information (*null values*) complicates querying.
  - ▶ Not flexible, hard to extend once defined.
  - ▶ Integrated structure- and content-based querying is difficult.
  - ▶ Querying databases at different locations requires extra sophisticated technology.
- ▶ Not designed for achieving interoperability on the Web!

## 5. Resource Description Framework (RDF)

RDF W3C Recommendation; 2004;

RDF 1.1 W3C Recommendation; 2014

- ▶ Everything is a *resource*; resources are either atomic values or identified by URIs.
- ▶ Information is represented by triples: (subject, predicate, object).

Assume a social network,  
e.g. Facebook or Twitter.

Users A, B, C, D may have  
a follower relationship.

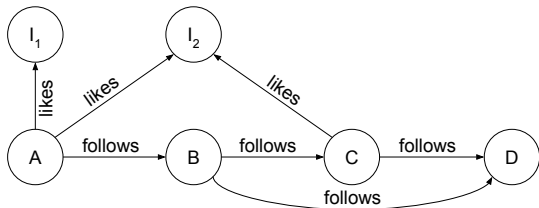
Users may like certain items  
I1, I2 and I3.

A	follows	B
B	follows	C
C	follows	D
B	follows	D
A	likes	I1
A	likes	I2
C	likes	I2

- ▶ Triples have a natural representation by directed, labelled graphs.

## RDF-Triples and RDF-Graph

A	follows	B
B	follows	C
C	follows	D
B	follows	D
A	likes	I1
A	likes	I2
C	likes	I2



## 6. RDF in the Real World

### ... from Louvre

```
<http://dbpedia.org/resource/Venus_de_Milo>
```

```
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/ontology/Artwork> ;  
<http://www.w3.org/2000/01/rdf-schema#label> "Venus de Milo"@en ;  
<http://dbpedia.org/ontology/wikiPageExternalLink> <http://www.venusdemilo.gr/> ;  
<http://dbpedia.org/property/artist> <http://dbpedia.org/resource/Alexandros_of_Antioch> ;  
<http://dbpedia.org/property/type> <http://dbpedia.org/resource/Marble> ;  
<http://dbpedia.org/property/year> "Between 130 and 100 BC"@en ;  
<http://dbpedia.org/ontology/museum> <http://dbpedia.org/resource/The_Louvre> ;  
...
```

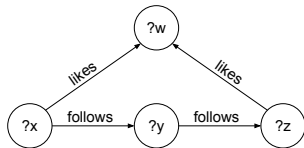
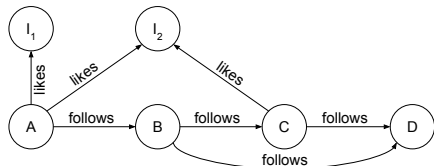
### Where to start reading?

RDF 1.1 Primer; W3C Working Group Note 24 June 2014;

<https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>

## 7. SPARQL: RDF's SQL

### RDF Graph and Query Pattern



### SPARQL basic construct: Triple Pattern

```
SELECT * FROM {  
  ?x likes ?w.  
  ?x follows ?y.  
  ?y follows ?z.  
  ?z likes ?w }
```

## 8. Finally, more on RDF

### RDF Stores - either native or based on a Relational Database

Triple Store:

Subject	Predicate	Object
A	follows	B
...		
C	likes	I2

Vertical Partitioning:

	follows		likes	
	Subject	Object	Subject	Object
	A	B	A	I1
	B	C	A	I2
	C	D	C	I2
	B	D		

Property Table:

	Subject	follows	likes
	A	B	I1
	A	B	I2
	B	C	
	B	D	
	C	D	I2



## RDF and the Semantic Web

- ▶ RDF is the *lingua franca* for the Semantic Web.
- ▶ W3C: *The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries*".

Tim Berners-Lee: *The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.*

- ▶ W3C Web Ontology Language (OWL): *OWL facilitates greater machine interpretability of Web content than that supported by RDF by providing additional vocabulary along with a formal semantics.*

OWL adds formal (logical) reasoning to ontologies.