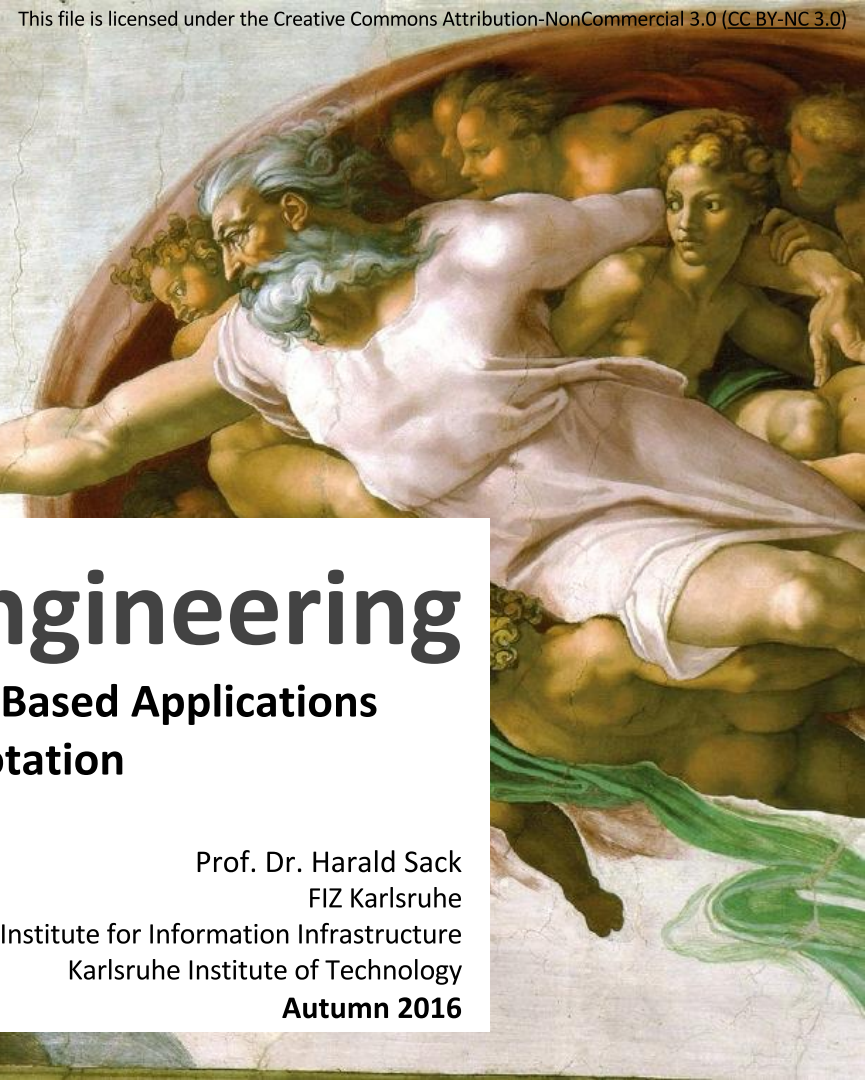




Lecture 6: Advanced Linked Data Based Applications

6.2 Metadata and Semantic Annotation

 **FIZ Karlsruhe**
Leibniz Institute for Information Infrastructure

Leibniz Institute for Information Infrastructure
Karlsruhe Institute of Technology
Autumn 2016



By I ♥ happyH from NY, NY (Flickr) [CC BY 2.0 (<http://creativecommons.org/licenses/by/2.0>)], via Wikimedia Commons from Wikimedia Commons

How do you find something (again)...

Resources are described via Metadata

- „Metadata is data that **contains information about other data**“ and „Metadata is required to **fulfill a specific purpose** or to **achieve a specific result**“ (informal definition, Wikipedia)
- „Metadata is **structured, encoded data** that **describes the characteristics** of information bearing entities for the purpose of **identification, retrieval, evaluation** and **administration** of the described entities.“ (W.R. Durell, 1985)
- „Metadata is machine **understandable information about web resources or other things.**“ (W3C)

Structured Metadata

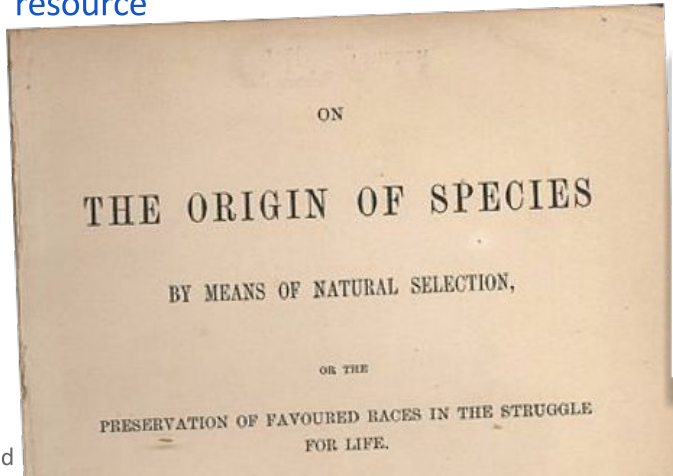
- usually consist of **Name-Value pairs** (*Author = "Ernest Hemingway"*)
- are **typed** (*Author is of type String*)
- **Semantics** (Meaning) of structured data relies on **common agreement**
- e.g. via standard vocabularies such as Dublin Core

```
<head profile="http://dublincore.org/documents/dcq-html/">
  <title>Dublin Core</title>
  <link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
  <link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
  <meta name="DC.format" scheme="DCTERMS.IMT" content="text/html" />
  <meta name="DC.type" scheme="DCTERMS.DCMIType" content="Text" />
  <meta name="DC.publisher" content="Jimmy Wales" />
  <meta name="DC.subject" content="Dublin Core Metadaten-Elemente, Anwendungen" />
  <meta name="DC.creator" content="Björn G. Kulms" />
  <meta name="DCTERMS.license" scheme="DCTERMS.URI" content="http://www.gnu.org/copyleft/fdl.html" />
  <meta name="DCTERMS.rightsHolder" content="Wikimedia Foundation Inc." />
  <meta name="DCTERMS.modified" scheme="DCTERMS.W3CDTF" content="2006-03-08" />
</head>
```

Unstructured Metadata

- (mostly) **text-based** metadata
- **Semantics** is determined via **natural language explanations** (not necessarily via agreement)
- e.g. description or abstract

resource



On the Origin of Species, published on 24 November 1859, is a work of [scientific literature](#) by Charles Darwin which is considered to be the foundation of [evolutionary biology](#). Its full title was *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. For the sixth edition of 1872, the short title was changed to *The Origin of Species*. Darwin's book introduced the [scientific theory](#) that populations [evolve](#) over the course of generations through a process of [natural selection](#). It presented a body of evidence that [the diversity of life](#) arose by [common descent](#) through a [branching pattern of evolution](#). Darwin included evidence that he had gathered on the *Beagle* expedition in the 1830s and his subsequent findings from research, correspondence, and experimentation.

unstructured metadata

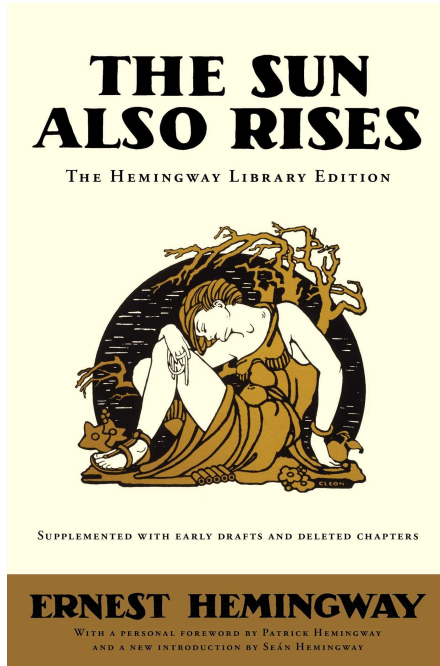


“Herr Lehrer Lämpel” by Wilhelm Busch, Max und Moritz (1865)

- **Authoritative Metadata**
have been created by a trusted
(=authoritative) source, as e.g.
 - the author of the described resource
 - a trusted expert

HPI Hasso Plattner Institut



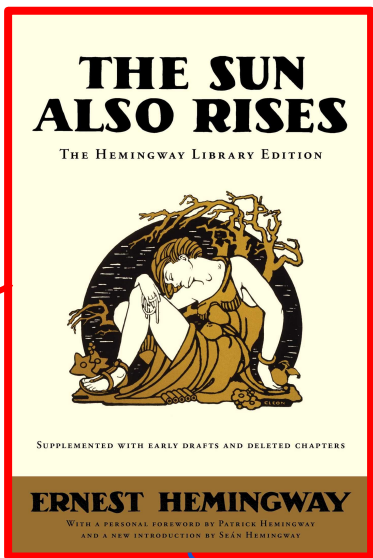


- The **semantics** of semantic metadata is explicitly and formally defined (Ontology) and therefore **machine understandable** (i.e. can be automatically read and correctly interpreted by a machine)

```
PREFIX dbr: <http://dbpedia.org/resource/> .  
PREFIX dbo: <http://dbpedia.org/ontology/> .  
  
:9787532717071 dbo:author dbr:Ernest_Hemingway .
```

Semantic Annotation

- The term “**annotation**” implies to **attach data to some other piece of data**. It establishes a **(typed) relation between the annotated data and the annotating data**
- The term “**annotation**” can denote both the **process of annotating** and the **result of that process**
- **Formal Model:**
 - An **annotation** A is a tuple (a_s, a_p, a_o, a_c) , where
 - a_s is the **subject** of the annotation (the annotated data)
 - a_o is the **object** of the annotation (the annotating data)
 - a_p is the **predicate** (the annotation relation) that defines the type of relationship between a_s and a_o , and
 - a_c is the **context** in which the annotation is made.

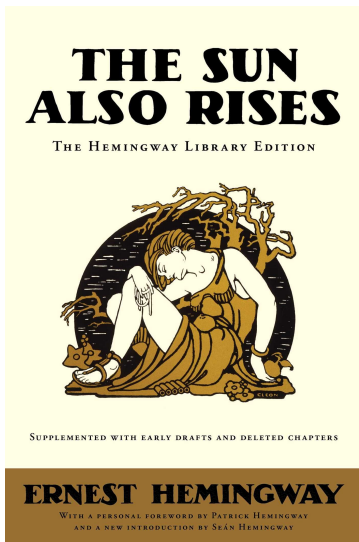


- **Formal Model:**

- An **annotation** A is a tuple (a_s, a_p, a_o, a_c) , where
 - a_s is the **subject** of the annotation (the annotated data)
 - a_o is the **object** of the annotation (the annotating data)
 - a_p is the **predicate** (the annotation relation) that defines the type of relationship between a_s and a_o , and
 - a_c is the **context** in which the annotation is made.

```
PREFIX dbr: <http://dbpedia.org/resource/> .  
PREFIX dbo: <http://dbpedia.org/ontology/> .  
  
:9787532717071 dbo:author dbr:Ernest_Hemingway .
```

Semantic Annotation

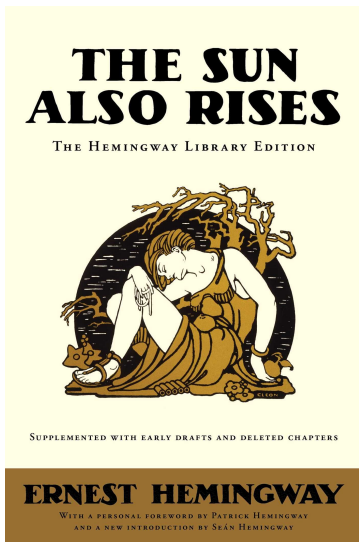


- **Formal Model:**

- An **annotation** A is a tuple (a_s, a_p, a_o, a_c) , where
 - a_s is the **subject** of the annotation (the annotated data)
 - a_o is the **object** of the annotation (the annotating data)
 - a_p is the **predicate** (the annotation relation) that defines the type of relationship between a_s and a_o , and
 - a_c is the **context** in which the annotation is made.

```
PREFIX dbr: <http://dbpedia.org/resource/> .
PREFIX dbo: <http://dbpedia.org/ontology/> .

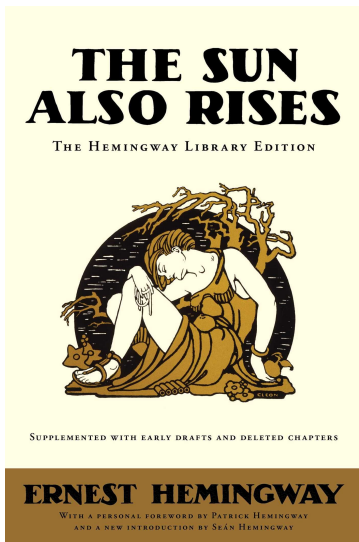
::9787532717071 dbo:author dbr:Ernest_Hemingway .
```



- **Formal Model:**

- An **annotation** A is a tuple (a_s, a_p, a_o, a_c) , where
 - a_s is the **subject** of the annotation (the annotated data)
 - a_o is the **object** of the annotation (the annotating data)
 - a_p is the **predicate** (the annotation relation) that defines the type of relationship between a_s and a_o , and
 - a_c is the **context** in which the annotation is made.

```
PREFIX dbr: <http://dbpedia.org/resource/> .  
PREFIX dbo: <http://dbpedia.org/ontology/> .  
  
::9787532717071 dbo:author dbr:Ernest_Hemingway .
```



- **Formal Model:**

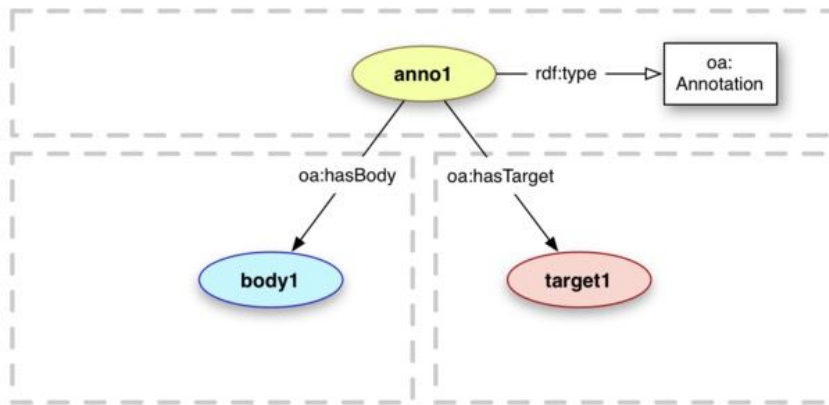
- An **annotation** A is a tuple (a_s, a_p, a_o, a_c) , where
 - a_s is the **subject** of the annotation (the annotated data)
 - a_o is the **object** of the annotation (the annotating data)
 - a_p is the **predicate** (the annotation relation) that defines the type of relationship between a_s and a_o , and
 - a_c is the **context** in which the annotation is made.

```
PREFIX dbr: <http://dbpedia.org/resource/> .  
PREFIX dbo: <http://dbpedia.org/ontology/> .  
  
::9787532717071 dbo:author dbr:Ernest_Hemingway .
```

- created at Dec 11, 2013
- created by Lysander07
- ...

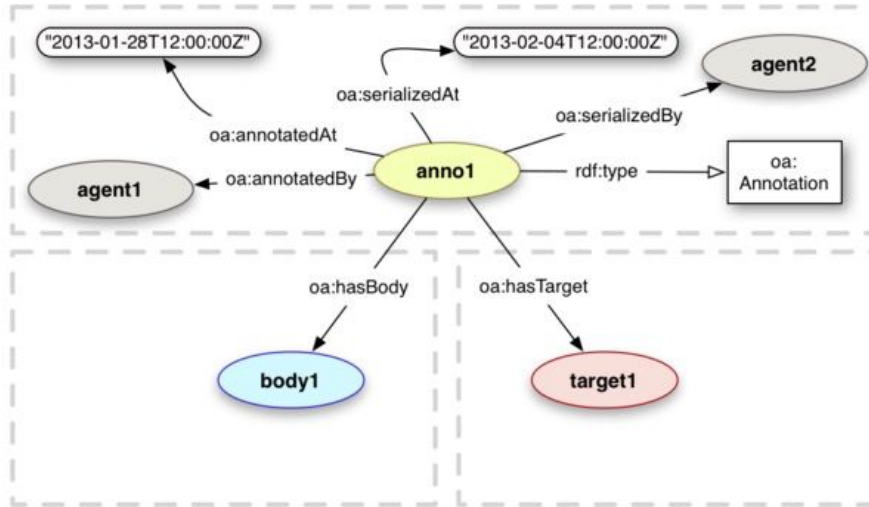
Ontologies for Semantic Annotation

- The **Open Annotation Ontology** is a W3C Draft (2013)
 - “...specifies an interoperable framework for creating associations between related resources, annotations, using a methodology that conforms to the Architecture of the World Wide Web.”



- The **Body** provides the information which is annotating the **Target**

Ontologies for Semantic Annotation

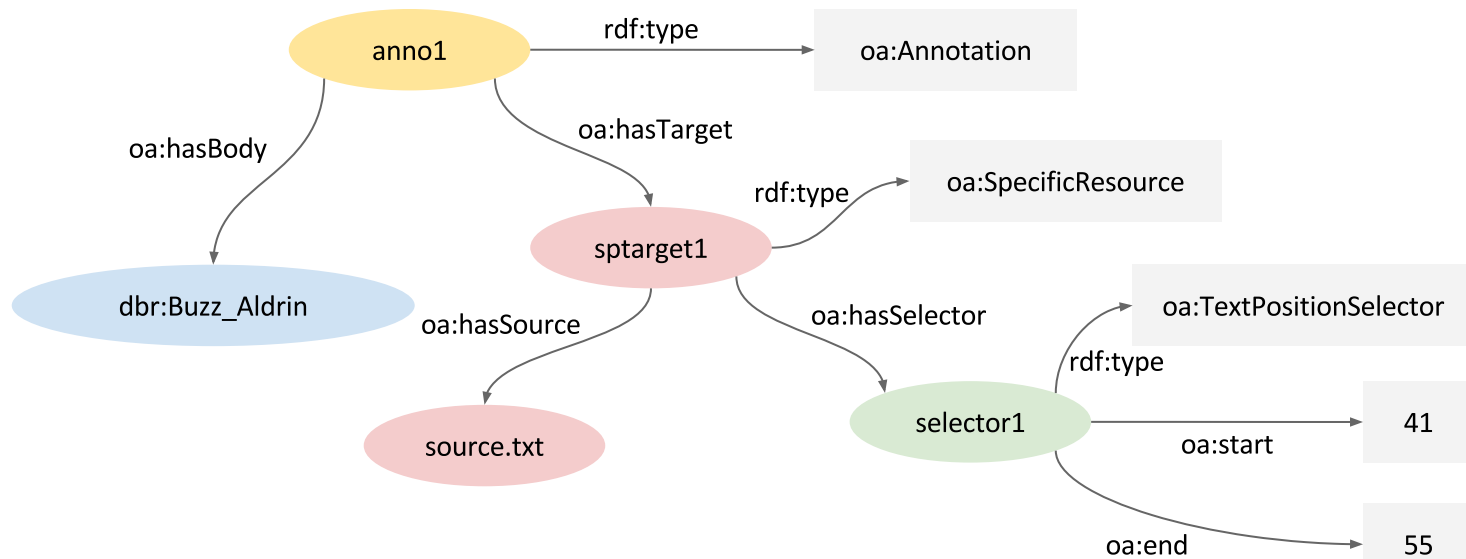


- Open Annotation provides means to describe the **annotation context**
- **Provenance information** can be attached to the Annotation, Body, Target or any other resource in the Annotation graph

Semantic Text Annotation Example

source.txt

On July 16, 1969, Armstrong, along with **Edwin E. Aldrin, Jr.**, and Michael Collins, blasted off in the Apollo 11 vehicle toward the Moon.



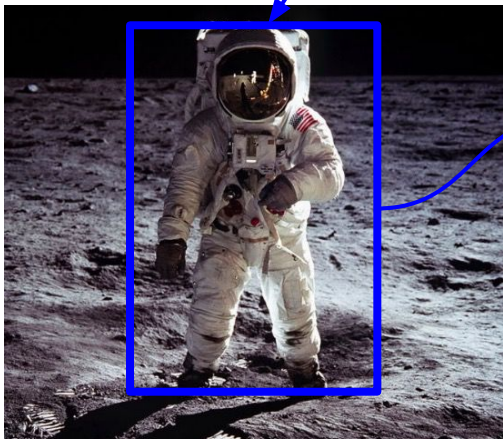
Semantic Multimedia Annotation Example



media fragment
<http://example.com/apollo11.ogv#t=10,20>

media fragment
<http://example.com/apollo11.ogv#t=20,30>

media fragment
<http://example.com/apollo11.ogv#t=30,44>



```
...
<div vocab="http://www.w3.org/ns/oa#"
      prefix="dctypes: http://purl.org/dc/dcmitype/
             foaf: http://xmlns.com/foaf/0.1/"
      typeof="Annotation"
      resource="#contentAnnotation-001">
  <div property="hasTarget"
        resource="http://example.com/apollo11.ogv#t=20,30&xywh=480,150,140,330"
        typeof="dctypes:video">
  </div>
  <div property="hasBody" typeof="SemanticTag">
    <a property="foaf:page" href="http://dbpedia.org/resource/Buzz_Aldrin">
      Buzz Aldrin
    </a>
  </div>
</div>
...
```

HTML with RDFa



Next: 03 Named Entity Resolution

OpenHPI - Course Linked Data Engineering - **Lecture 6: Advanced Linked Data Based Applications**