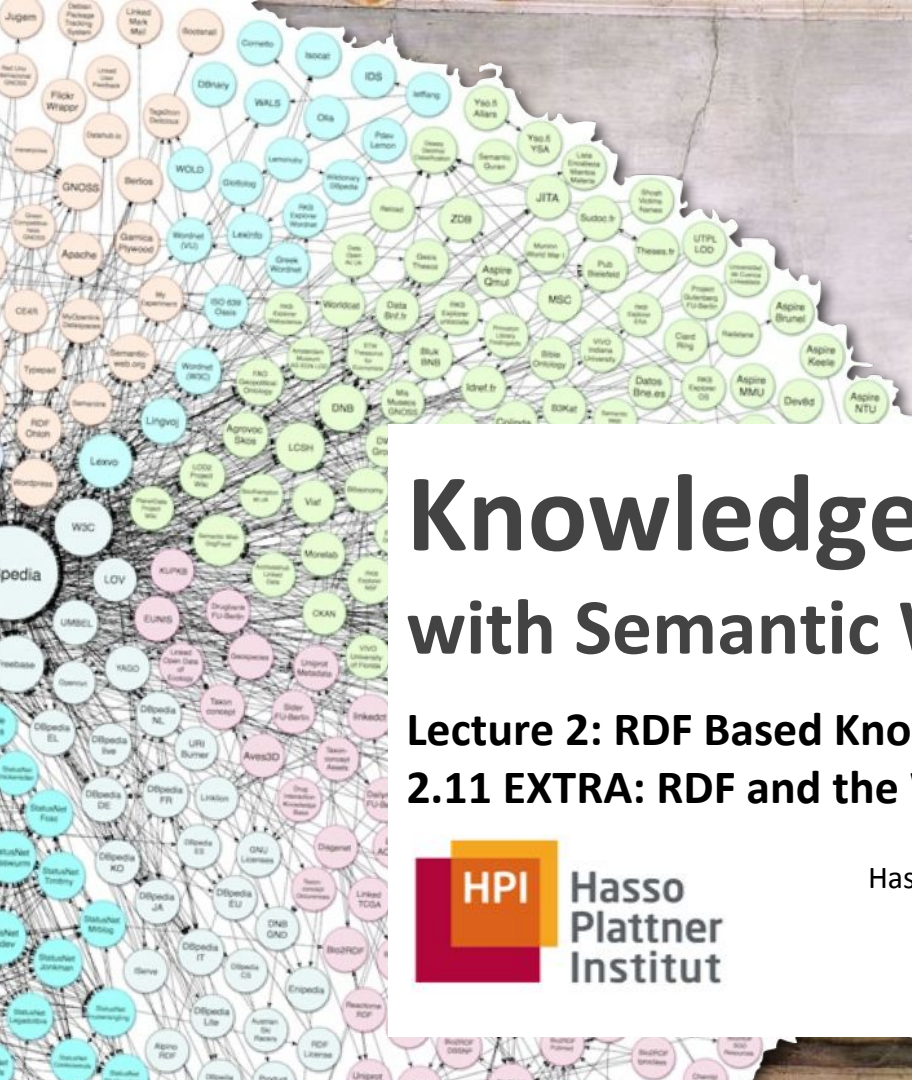


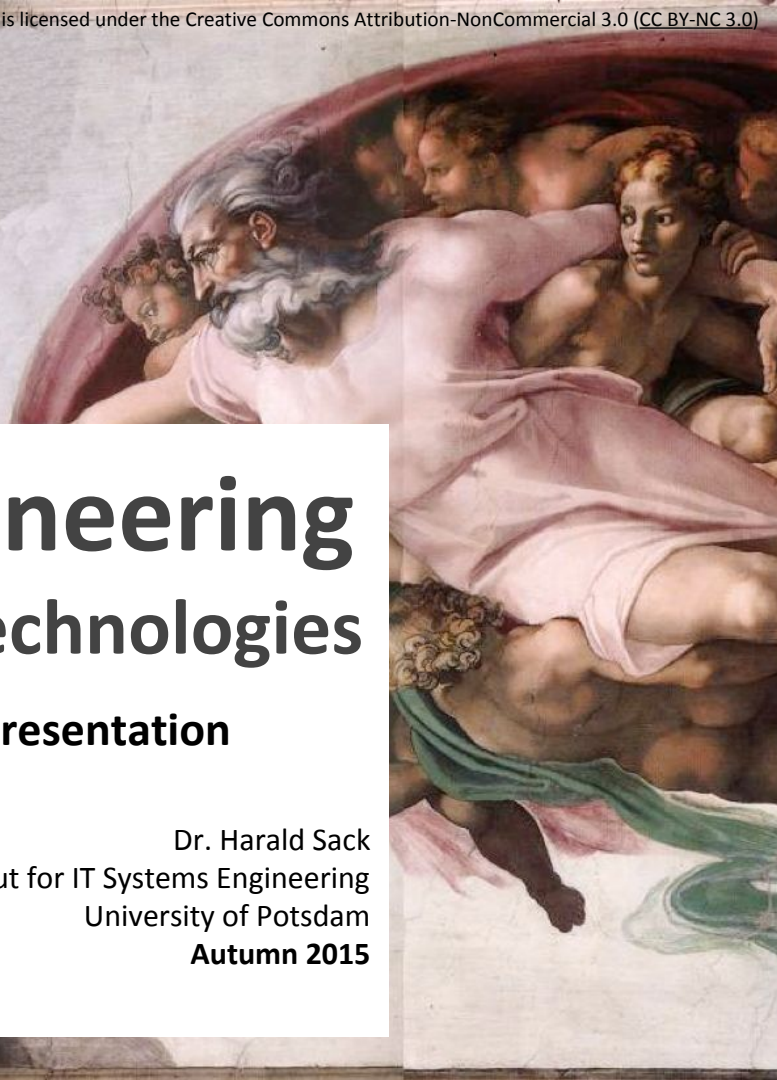


Knowledge Engineering with Semantic Web Technologies

Lecture 2: RDF Based Knowledge Representation 2.11 EXTRA: RDF and the Web

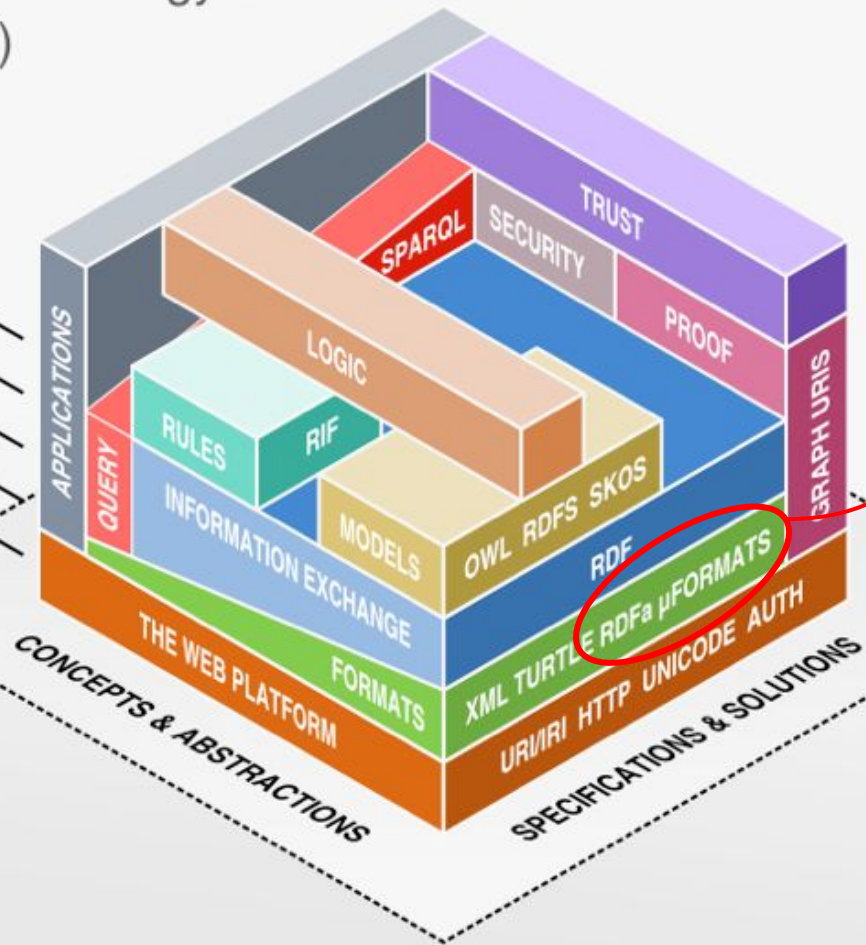
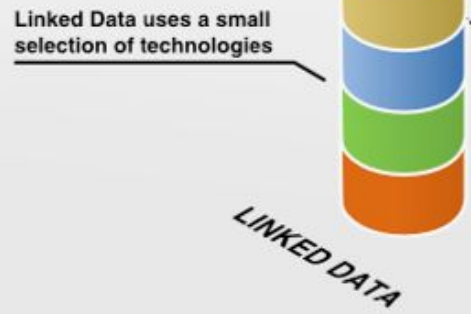


Dr. Harald Sack
Hasso-Plattner-Institut for IT Systems Engineering
University of Potsdam
Autumn 2015



The Semantic Web Technology Stack (not a piece of cake...)

- Most apps use only a subset of the stack
- Querying allows fine-grained data access
- Standardized information exchange is key
- Formats are necessary, but not too important
- The Semantic Web is based on the Web



RDFa and
μFormats

RDF and the Web

- RDF/XML can directly be embedded in an HTML document via `<head>` and `<script>`
- RFC 3870 defines Mime-Type **application/rdf+xml**

```
<head>
<title>My Document</title>
<script type="application/rdf+xml">
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
           xmlns:dc="http://purl.org/dc/elements/1.1/">
    <rdf:Description rdf:about="http://www.w3.org/" dc:title="W3C Homepage"/>
  </rdf:RDF>
</script>
</head>
```

RDF and the Web

- W3C recommended linking of external RDF documents via HTML `<link>` element in HTML `<head>`

```
<link rel="meta" type="application/rdf+xml"
      href="Meta-Data-for-Web-Page.rdf"/>
```

- RDF documents can also be embedded via `<a href=...>` and additional markup (e.g. RDF logo) within HTML `<body>`

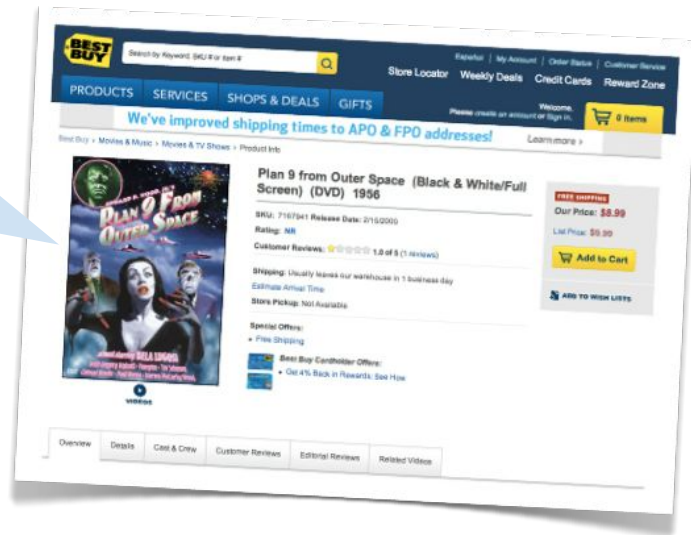
RDF and the Web

- Problem: HTML and RDF data are separate

```
<http://www.bestbuy.com/site/Plan+9+from+Outer+Space+-+DVD/7167941.p?id=23883&skuId=7167941&st=plan%209%20from%20outer%20space&lp=6&cp=1> <http://ogp.me/ns#site_name> "Best Buy" .
<http://www.bestbuy.com/site/Plan+9+from+Outer+Space+-+DVD/7167941.p?id=23883&skuId=7167941&st=plan%209%20from%20outer%20space&lp=6&cp=1> <http://www.facebook.com/2008/fbmlapp_id> "125188000891129" .
<http://www.bestbuy.com/site/Plan+9+from+Outer+Space+-+DVD/7167941.p?id=23883&skuId=7167941&st=plan%209%20from%20outer%20space&lp=6&cp=1> <http://ogp.me/ns#description> "Plan 9 from Outer Space - DVD" .
<http://www.bestbuy.com/site/Plan+9+from+Outer+Space+-+DVD/7167941.p?id=23883&skuId=7167941&st=plan%209%20from%20outer%20space&lp=6&cp=1> <http://ogp.me/ns#upc> "025493070026" .
<http://www.bestbuy.com/site/Plan+9+from+Outer+Space+-+DVD/7167941.p?id=23883&skuId=7167941&st=plan%209%20from%20outer%20space&lp=6&cp=1> <http://www.w3.org/1999/xhtml/vocab#stylesheet>
<http://images.bestbuy.com:80/BestBuy_US/store/styles/global-min.css> .
<http://www.bestbuy.com/site/Plan+9+from+Outer+Space+-+DVD/7167941.p?id=23883&skuId=7167941&st=plan%209%20from%20outer%20space&lp=6&cp=1> <http://www.w3.org/1999/xhtml/vocab#stylesheet>
<http://images.bestbuy.com:80/BestBuy_US/store/styles/pdp-min.css> .
<http://www.bestbuy.com/site/Plan+9+from+Outer+Space+-+DVD/7167941.p?id=23883&skuId=7167941&st=plan%209%20from%20outer%20space&lp=6&cp=1> <http://www.w3.org/1999/xhtml/vocab#stylesheet>
<http://images.bestbuy.com:80/BestBuy_US/store/styles/product_image_headers.css> .
<http://images.bestbuy.com:80/BestBuy_US/store/styles/tabs-min.css> .
<http://www.bestbuy.com/site/Plan+9+from+Outer+Space+-+DVD/7167941.p?id=23883&skuId=7167941&st=plan%209%20from%20outer%20space&lp=6&cp=1> <http://www.w3.org/1999/xhtml/vocab#stylesheet>
<http://images.bestbuy.com:80/BestBuy_US/store/styles/buttons-min.css> .
```

web page
HTML

RDF data
Turtle



- Maintenance
- Verification
- Consistency

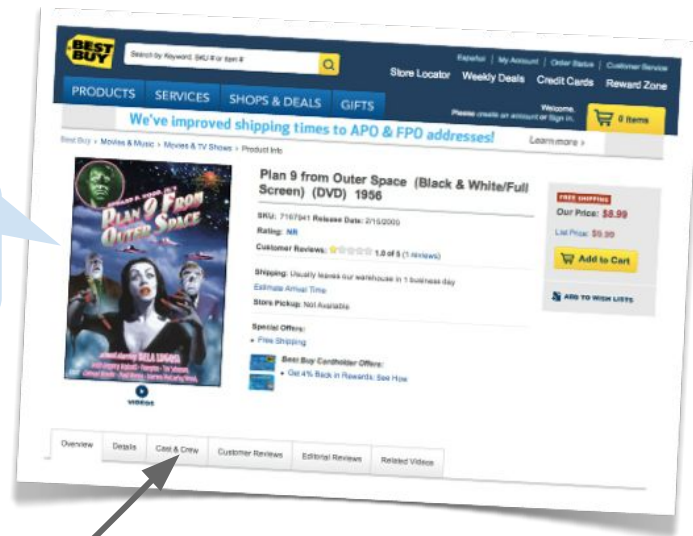
RDF and the Web

- embed RDF directly into HTML page

```
<http://www.bestbuy.com/site/Plan+9+from+Outer+Space+-+DVD/7167941.p?id=23883&skuId=7167941&st=plan%209%20from%20outer%20space&lp=6&cp=1> <http://ogp.me/ns#site_name> "Best Buy" .
<http://www.bestbuy.com/site/Plan+9+from+Outer+Space+-+DVD/7167941.p?id=23883&skuId=7167941&st=plan%209%20from%20outer%20space&lp=6&cp=1> <http://www.facebook.com/2008/fbmlapp_id> "125188000891129" .
<http://www.bestbuy.com/site/Plan+9+from+Outer+Space+-+DVD/7167941.p?id=23883&skuId=7167941&st=plan%209%20from%20outer%20space&lp=6&cp=1> <http://ogp.me/ns#description> "Plan 9 from Outer Space - DVD" .
<http://www.bestbuy.com/site/Plan+9+from+Outer+Space+-+DVD/7167941.p?id=23883&skuId=7167941&st=plan%209%20from%20outer%20space&lp=6&cp=1> <http://ogp.me/ns#upc> "025493070026" .
<http://www.bestbuy.com/site/Plan+9+from+Outer+Space+-+DVD/7167941.p?id=23883&skuId=7167941&st=plan%209%20from%20outer%20space&lp=6&cp=1> <http://www.w3.org/1999/xhtml/vocab#stylesheet>
<http://images.bestbuy.com:80/BestBuy_US/store/styles/global-min.css> .
<http://www.bestbuy.com/site/Plan+9+from+Outer+Space+-+DVD/7167941.p?id=23883&skuId=7167941&st=plan%209%20from%20outer%20space&lp=6&cp=1> <http://www.w3.org/1999/xhtml/vocab#stylesheet>
<http://images.bestbuy.com:80/BestBuy_US/store/styles/pdp-min.css> .
<http://www.bestbuy.com/site/Plan+9+from+Outer+Space+-+DVD/7167941.p?id=23883&skuId=7167941&st=plan%209%20from%20outer%20space&lp=6&cp=1> <http://www.w3.org/1999/xhtml/vocab#stylesheet>
<http://images.bestbuy.com:80/BestBuy_US/store/styles/product_image_headers.css> .
<http://images.bestbuy.com:80/BestBuy_US/store/styles/tabs-min.css> .
<http://www.bestbuy.com/site/Plan+9+from+Outer+Space+-+DVD/7167941.p?id=23883&skuId=7167941&st=plan%209%20from%20outer%20space&lp=6&cp=1> <http://www.w3.org/1999/xhtml/vocab#stylesheet>
<http://images.bestbuy.com:80/BestBuy_US/store/styles/buttons-min.css> .
```

web page
HTML

RDF data
Turtle



Semantic Annotation in the Web

- In principle there are three ways to embed structured data with explicit semantic annotations within HTML documents



Domain specific microformats (μ Format)



Generic RDFa

- HTML5 Microdata (including schema.org)



Microformats

- Microformats (μformats) emerged about 2005
- (X)HTML Markup to express (limited) semantics in an HTML document
 - designed to solve simple, specific problems
 - designed for humans first, machines second
 - used in web pages to describe a specific type of information, as e.g. a person, an event, a product, a review, etc.
- Applications can easily extract data from HTML documents
- In general, Microformats use the `class` attribute in **HTML tags** (most times `<span>` or `<div>` tags) and assign brief and descriptive names to entities and their properties



Microformats

- Microformats reuse the following (X)HTML tag **attributes**:
 - `class`
 - `rel`
 - `rev`
- Predefined standard microformats:
 - hCard - personal data (vCard, RFC2426)
 - hCalendar – calendars and events
 - rel-Tag – tags, keywords, categories
 - XFN – XHTML friends network
 - hReview, VoteLinks -- opinions, ratings, and reviews
 - XOXO – lists and outlines
- hcard, hReview are used by Google (2009) and Facebook (2011)

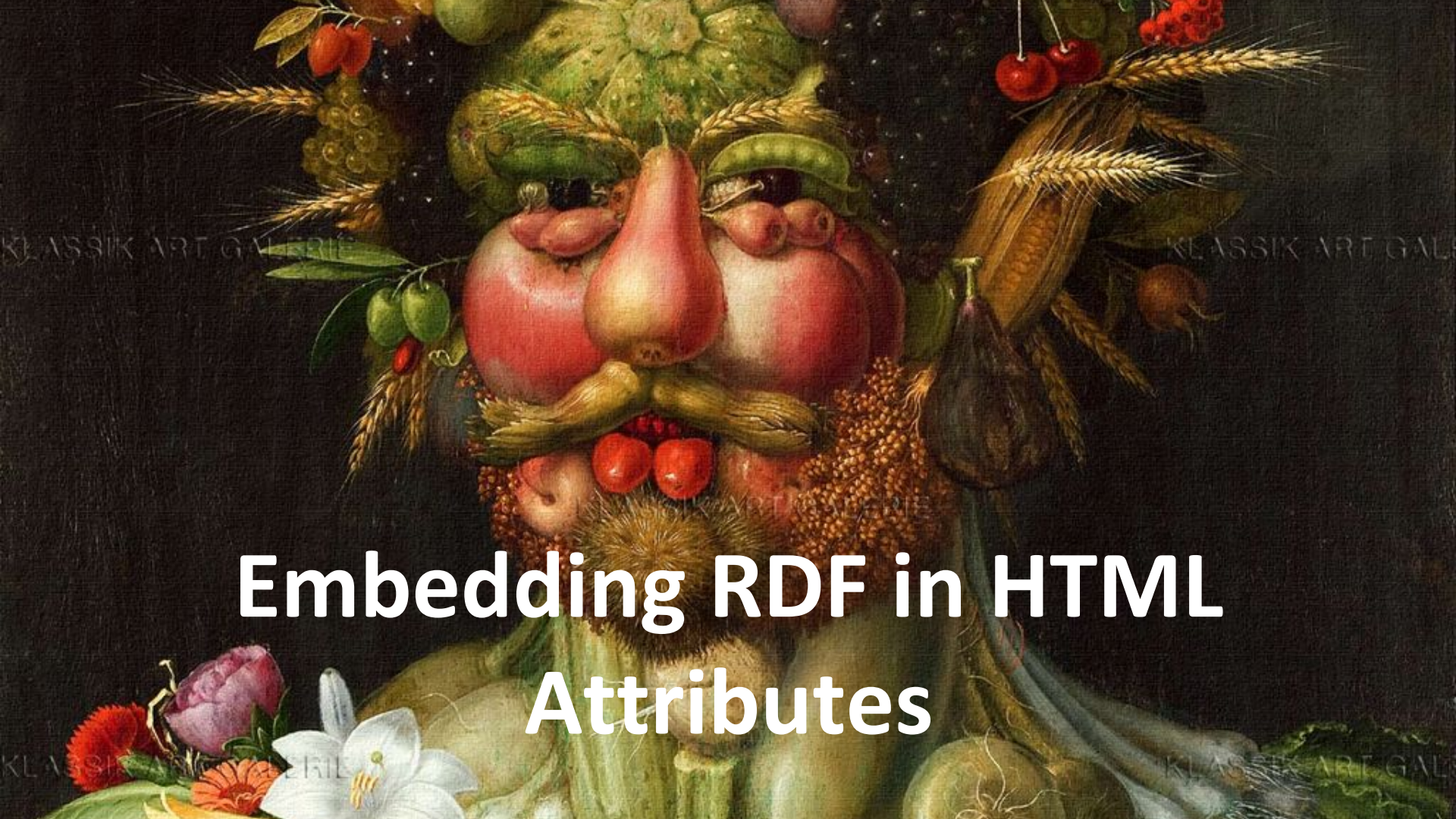
- Simple Example: HTML marked up with **hCard** microformat

```
<div class="vcard">
  <span class="fn">Joe Blow</span>
  <span class="title">Senior Manager</span>
  <span class="org">The Example Company</span>
  <span class="adr">
    <span class="street-address">Hauptstr. 123, </span>
    <span class="postal-code">14482</span>
    <span class="locality">Potsdam</span>
  </span>
  Tel.<span class="tel">604-555-1234</span>
  
</div>
```



Microformats - Pros and Cons

- Microformats can **easily be transcoded to RDF** via XSLT
- New microformat vocabularies **first must be consolidated** by the community, while always a new XSLT stylesheet has to be developed for extraction
- By **using more than one** microformat vocabulary in a single (X)HTML document the processing **complexity increases** rapidly
- **Conflicts** with used (X)HTML attributes might be possible

A detailed still life painting of various fruits and vegetables, including pears, apples, corn, and flowers, set against a dark background. The composition is dense and rich in color, with a variety of produce arranged in a central cluster. The lighting highlights the textures and colors of the fruits, creating a sense of depth and realism. The background is dark, making the vibrant colors of the produce stand out. The overall style is characteristic of 17th-century Dutch still life painting.

Embedding RDF in HTML Attributes

- RDFa = **RDF** in HTML **a**tributes
- enables generic RDF annotation in (X)HTML documents by reusing existing (X)HTML attributes
- RDFa 1.0 based on XHTML (W3C Recommendation 2008)
- RDFa 1.1 based on HTML5 (W3C Recommendation June 2012)
 - **RDFa Lite 1.1**
 - RDFa 1.1



RDFa Lite 1.1

- RDFa reuses existing (X)HTML attributes (e.g. `href`, `src`) and introduces new HTML attributes
 - `vocab`,
 - `typeof`,
 - `property`,
 - `resource`,
 - `prefix`



RDFa Lite 1.1

- First we need a **vocabulary** to talk about things

```
<p vocab="http://xmlns.com/foaf/0.1/">  
  My name is Harald Sack and you can  
  give me a ring via 1-800-555-0527.  
</p>
```



RDFa Lite 1.1

- Then we have to define the **type of thing** we are talking about

```
<p vocab="http://xmlns.com/foaf/0.1/"  
  typeof="Person">  
  My name is Harald Sack and you can  
  give me a ring via 1-800-555-0527.  
</p>
```



RDFa Lite 1.1

- Now we can define all **properties** of the thing we are talking about

```
<p vocab="http://xmlns.com/foaf/0.1/" typeof="Person">
  My name is
  <span property="name">Harald Sack</span>
  and you can give me a ring via
  <span property="phone">1-800-555-0527</span>.
  
</p>
```



RDFa Lite 1.1

- We can create **identifiers** for the things we are talking about

```
<p vocab="http://xmlns.com/foaf/0.1/"  
  resource="#harald" typeof="Person">  
  My name is  
  <span property="name">Harald Sack</span>  
  and you can give me a ring via  
  <span property="phone">1-800-555-0527</span>.  
    
</p>
```

- `resource` refers to the base URI of the web page



RDFa Lite 1.1

- And if the vocabulary is not sufficient to describe all properties, we can use additional vocabularies by using prefixes

```
<p vocab="http://xmlns.com/foaf/0.1/"  
  prefix="ov: http://open.vocab.org/terms/"  
  resource="#harald" typeof="Person">  
  My name is  
  <span property="name">Harald Sack</span>  
  and you can give me a ring via  
  <span property="phone">1-800-555-0527</span>.  
  My favorite beverage is  
  <span property="ov:preferredBeverage">Green Tea</span>.  
</p>
```



RDFa 1.1

- With full RDFa 1.1 you have can add additional functionality
 - separate **content** from presentation

```
<p vocab="http://purl.org/dc/terms/" >  
  <h2 property="title">Syllabus SWT 2012/13</h2>  
  <h3 property="creator" resource="#me">Harald</h3>  
  Creation Date:  
  <span property="created" content="2012-10-28"> 28.10.2012 </span>  
</p>
```

content

presentation



RDFa 1.1

- With full RDFa 1.1 you have can add additional functionality
 - use **datatypes** from XML Schema Definition

```
<p vocab="http://purl.org/dc/terms/" >  
  <h2 property="title">Syllabus SWT 2012/13</h2>  
  <h3 property="creator" resource="#me">Harald</h3>  
  Creation Date:  
  <span property="created" datatype="xsd:gYear"> 2012 </span>  
</p>
```



- Difference between **resource** and **about**

```
<div resource="/alice/posts/trouble_with_bob">  
  <h2 property="title">The trouble with Bob</h2>  
  <h3 property="creator" resource="#me">Alice</h3>  
  ...  
</div>
```

subject

object

- The **resource** attribute in the `<div>` element sets the context, i.e. the **subject** for all subsequent statements.
- Also, when combined with the **property** attribute, **resource** can be used to set the **object** for the statement



- Difference between **resource** and **about**

```
<div resource="/alice/posts/trouble_with_bob">  
  <h2 property="title">The trouble with Bob</h2>  
  <h3 property="creator" resource="#me">Alice</h3>  
  ...  
</div>
```

subject

object

- The **resource** attribute in the `<div>` element sets the context, i.e. the **subject** for all subsequent statements.
- Also, when combined with the **property** attribute, **resource** can be used to set the **object** for the statement



RDFa 1.1

S

P

O

- Difference between **resource** and **about**

```
<ul resource="/alice/posts/trouble_with_bob">  
  ...  
  ...  
  <li resource="/alice/posts/trouble_with_bob">  
    <span property="title">The trouble with Bob</span>  
  </li>  
  <li resource="/alice/posts/jos_barbecue">  
    <span property="title">Jo's Barbecue</span>  
  </li>  
  ...  
</ul>
```

- ...works, but it's a bit convoluted



S

P

O

- Difference between **resource** and **about**

```
<ul resource="/alice/posts/trouble_with_bob">  
  ...  
  <li resource="/alice/posts/trouble_with_bob"  
    property="title">The trouble with Bob</li>  
  <li resource="/alice/posts/jos_barbecue"  
    property="title">Jo's Barbecue</li>  
  ...  
</ul>
```

- not correct; the combination of **property** and **resource** would generate a different statement than originally intended.



RDFa 1.1

- Difference between **resource** and **about**

S

P

O

```
<ul resource="/alice/posts/trouble_with_bob">  
  ...  
  <li about="/alice/posts/trouble_with_bob"  
    property="title">The trouble with Bob</li>  
  <li about="/alice/posts/jos_barbecue"  
    property="title">Jo's Barbecue</li>  
  ...  
</ul>
```

- **about** is only used to set the context (subject)



RDFa 1.1

- Distinguish two different sorts of RDF triples
 - Triple with resource as object
 - Triple with literal as object

	Subject	Property	Object
Object is Literal	<code>resource / about</code>	<code>property</code>	<code>content</code> or <code>#PCDATA</code>
Object is Resource (URI)	<code>resource / about</code>	<code>property/rel</code>	<code>href</code> or <code>resource</code>

- `about` / `rel` only for compatibility with RDFa 1.0



Differences with RDFa 1.0

```
<html>
...
<div xmlns:dc="http://purl.org/dc/elements/1.1/" >
  I'm currently reading
  <span about="urn:ISBN:0-596-00027-8" >
    <span property="dc:title" >
      Programming Perl
    </span>
    by
    <span property="dc:creator" >Larry Wall</span>
  </span>.
</div>
...
</html>
```

Namespace

Subject

Property

Object



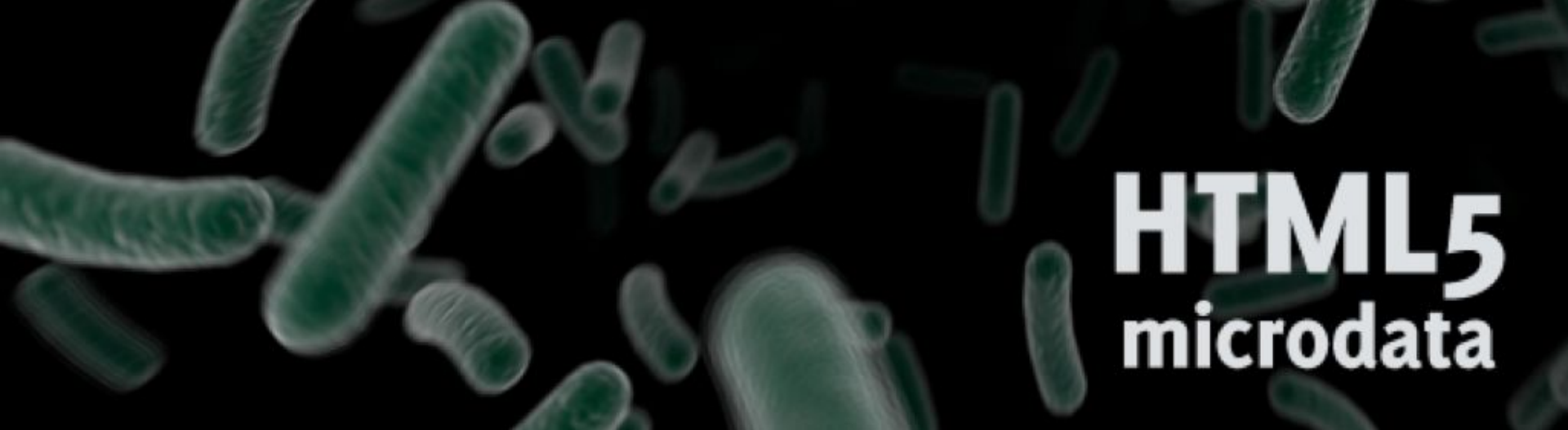
Differences with RDFa 1.0

```
<?xml version="1.0" encoding="UTF-8"?>
<html xmlns="http://www.w3.org/1999/xhtml" version="XHTML+RDFa 1.0"
  xml:lang="en">
  <head>...</head>
  <body>...</body>
</html>
</xml>
```

XHTML

```
<!DOCTYPE html>
<html version="HTML+RDFa 1.1" lang="en"
  xmlns="http://www.w3.org/1999/xhtml"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:cc="http://creativecommons.org/ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">
```

HTML5



HTML5 microdata

- **HTML5 Microdata** extends microformats and addresses its shortcomings
 - **schema.org** is the most prominent microdata vocabulary
- Microdata annotates the DOM with
 - **scoped name/value pairs** from **custom vocabularies**



HTML5 microdata

HTML5 microdata specification comprises:

- **microdata vocabularies** (<http://data-vocabulary.org/>)
 - Persons, Events, Organisations, Products, Reviews, Breadcrumbs, Ads, aggregated Reviews/Ads
- **microdata Global Attributes**
 - **itemscope** (creates resource)
 - **itemtype** (applied vocabulary)
 - **itemid** (identifies resource)
 - **itemprop** (property of a resource)
 - **itemref** (references on other resources)

HTML5 microdata

```
<section itemscope itemtype="http://data-vocabulary.org/Person">
  Hello, my name is
  <span itemprop="name">John Doe</span>,
  I am a
  <span itemprop="title">graduate research assistant</span>
  at the
  <span itemprop="affiliation">University of Dreams</span>.
  My friends call me
  <span itemprop="nickname">Johnny</span>.
  You can visit my homepage at
  <a href="http://www.JohnnyD.com" itemprop="url">www.JohnnyD.com</a>.
  <section itemprop="address" itemscope
    itemtype="http://data-vocabulary.org/Address">
    I live at
    <span itemprop="street-address">1234 Peach Drive</span>
    <span itemprop="locality">Warner Robins</span>
    ,
    <span itemprop="region">Georgia</span>.
  </section>
</section>
```

<http://yovisto.com/labs/swt1112/microdata01.html>

HTML5 microdata



Structured Data Testing Tool

Fetch URL

Examples ▾

VALIDATE

Results - All data ▾

```
1 <section itemscope itemtype="http://data-vocabulary.org/Person">
2   Hello, my name is
3   <span itemprop="name">John Doe</span>,
4   I am a
5   <span itemprop="title">graduate research assistant</span>
6   at the
7   <span itemprop="affiliation">University of Dreams</span>.
8   My friends call me
9   <span itemprop="nickname">Johnny</span>.
10  You can visit my homepage at
11  <a href="http://www.JohnnyD.com" itemprop="url">www.JohnnyD.com</a>.
12  <section itemprop="address" itemscope
13    itemtype="http://data-vocabulary.org/Address">
14    I live at
15    <span itemprop="street-address">1234 Peach Drive</span>
16    <span itemprop="locality">Warner Robins</span>
17    ,
18    <span itemprop="region">Georgia</span>.
19  </section>
20 </section>
```

▾ Person (1)

All good ✓

Person

name:	John Doe
title:	graduate research assistant
affiliation:	University of Dreams
nickname:	Johnny
url:	http://www.johnnyd.com/
address [Address]:	
street-address:	1234 Peach Drive
locality:	Warner Robins
region:	Georgia

▸ Custom Search Result Filters

schema.org

- **What is schema.org?**
 - Initiative started by Bing, Google, Yahoo! on June 2, 2011.
- **Goals:**
 - Semantic markup for web page content based on **proprietary HTML5 microdata schemata**
- **Search engines, web crawler, and browsers** recognize and understand schema.org markup
- small number of formats (extensible)
- RDF mapping available

schema.org and Rich Snippets

Google



Digital Communication - Buch - Harald Sack, Christoph Meinel

€85.59 from beck-shop.de ★★★★★ 1,979 seller reviews

Free shipping

Von Harald Sack, Christoph Meinel: Communication, Multimedia, SecurityThe authors giv

Baidu

《数字通信技术 Christoph Meinel, Harald Sack, 季松 978730... 京东



2015年5月2日 - 数字通信技术 Christoph Meinel, Harald Sack, 季松
9787302371700 京东价: ¥34.00 客户端首单 满79送79 加载中,请稍候... 店铺: 思诺华教图书专营店 出版社...

item.jd.com/15017272..... - 百度快照 - 70%好评

- **Snippets** = the few lines of text that appear under every search result
 - Designed to give users a sense for what's on the web page and why it's relevant to their query
- If the search engine understands the content on the web page, it can create **“Rich” Snippets**
 - = detailed information intended to help users with specific queries
 - You can help the search engine find this relevant information by adding additional microdata as HTML markup

Microdata Usage

- **26.3% of all HTML pages contain structured data**
 - Source: WebDataCommons.org, Oct 2014
 - 2.3 billion HTML pages parsed (Common Crawl corpus)
 - >585 million pages contained RDFa, Microformat, Microdata data, etc.
- **30% of all HTML pages contain structured data**
 - Source: Yahoo! Research, Jan 2012
 - 3.2 billion HTML pages parsed (Bing Crawl corpus)
 - >973 million pages contained RDFa, Microformat, Microdata data
- **Structured Data on the Top K Alexa Websites**

First x in AL	PLDs in CC		% containing structured data			
	#	% AL	overall	RDFa	Microdata	Microformats
100	99	99.00	74.75	34.34	55.56	68.69
1k	963	96.30	62.62	40.08	31.67	46.11
10k	9,294	92.94	47.34	30.47	15.55	29.75
100k	85,058	85.01	31.94	16.46	7.20	20.07
1m	734,882	73.49	20.56	7.55	3.04	14.18

Bizer et al: Deployment of RDFa, Microdata, and Microformats on the Web – A Quantitative Analysis, (ISWC 2013)



01: RDF on the Web

OpenHPI - Course Knowledge Engineering with Semantic Web Technologies

Lecture 3: Ontologies and Logic