# RDFa:

## Embedding RDF Knowledge in HTML

# What is RDFa?

- RDF embedded in HTML or XML

  Provides set of attributes (the *a* in RDFa) to use with existing tags to carry RDF metadata

- 2004: work on developing standards began

- 2008: RDFa 1.0 a recommendation (but only in XHTML, which failed to launch)

- 2015: RDFa 1.1 recommendation (works in HTML4, HTML5)

- See http://rdfa.info/

# Principles of RDFa

- RDF content specified in XML *attributes* of tags rather than *elements*

- The XML/HTML *tree structure* is used as context, when appropriate

- Some new attributes are *introduced* and some existing ones (@href, @rel) *reused*

- When possible, HTML text content used for *literal values*

➔ Same file used by browser & RDF extractor

# Web page viewed by a person



http://www.w3.org/ns/entailment/data/RDFS.html
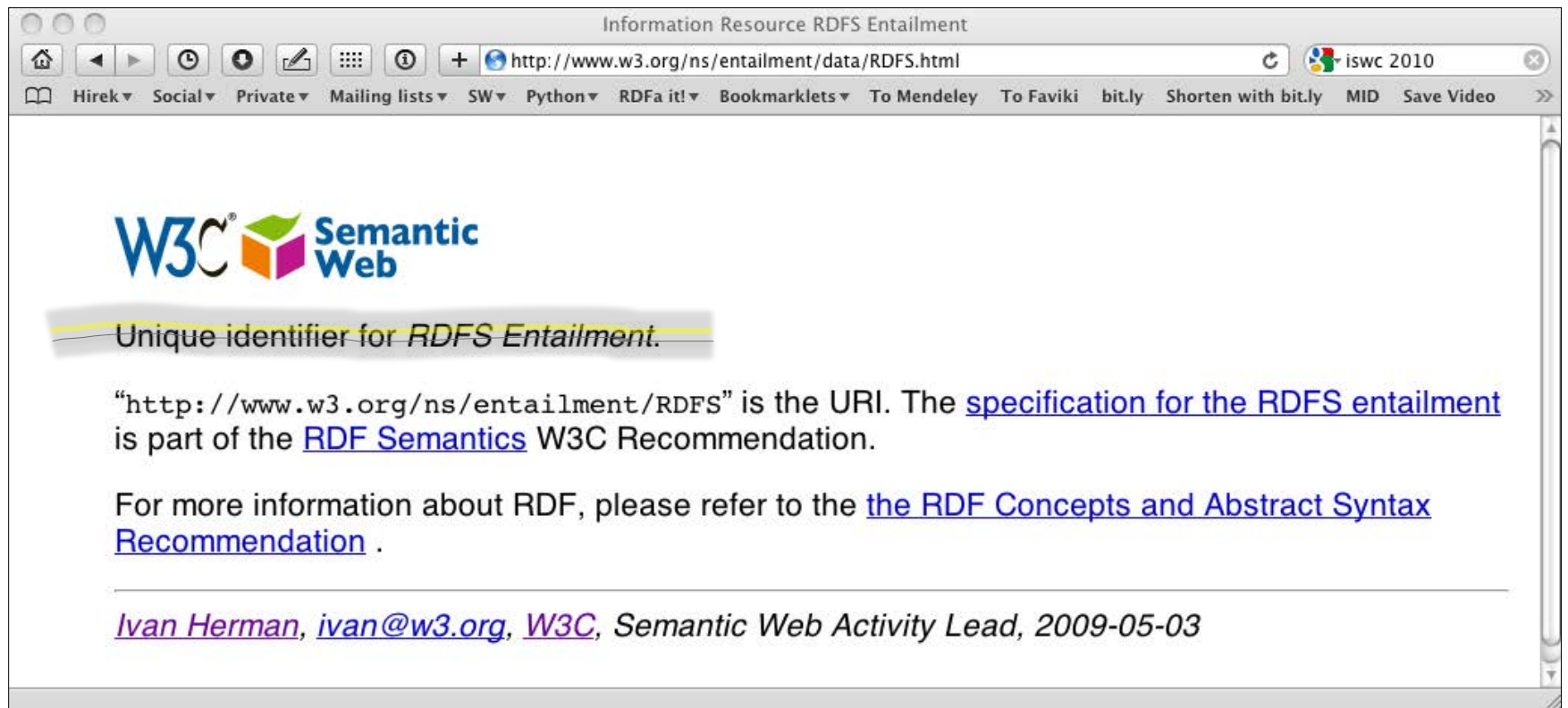
# HTML Source

```
<p about="http://www.w3.org/ns/entailment/RDFS"
   property="http://purl.org/dc/terms/description">
Unique identifier for <em>RDFS Entailment</em>.
</p>
```

# HTML Source & RDF extracted ...

```
<p about="http://www.w3.org/ns/entailment/RDFS"
    property="http://purl.org/dc/terms/description">
Unique identifier for <em>RDFS Entailment</em>.
</p>
```

```
<http://www.w3.org/ns/entailment/RDFS>
    … .
```

# HTML Source & RDF extracted ...

```
<p about="http://www.w3.org/ns/entailment/RDFS"
   property="http://purl.org/dc/terms/description">
Unique identifier for <em>RDFS Entailment</em>.
</p>
```

```
<http://www.w3.org/ns/entailment/RDFS>
    <http://purl.org/dc/terms/description>
        … .
```
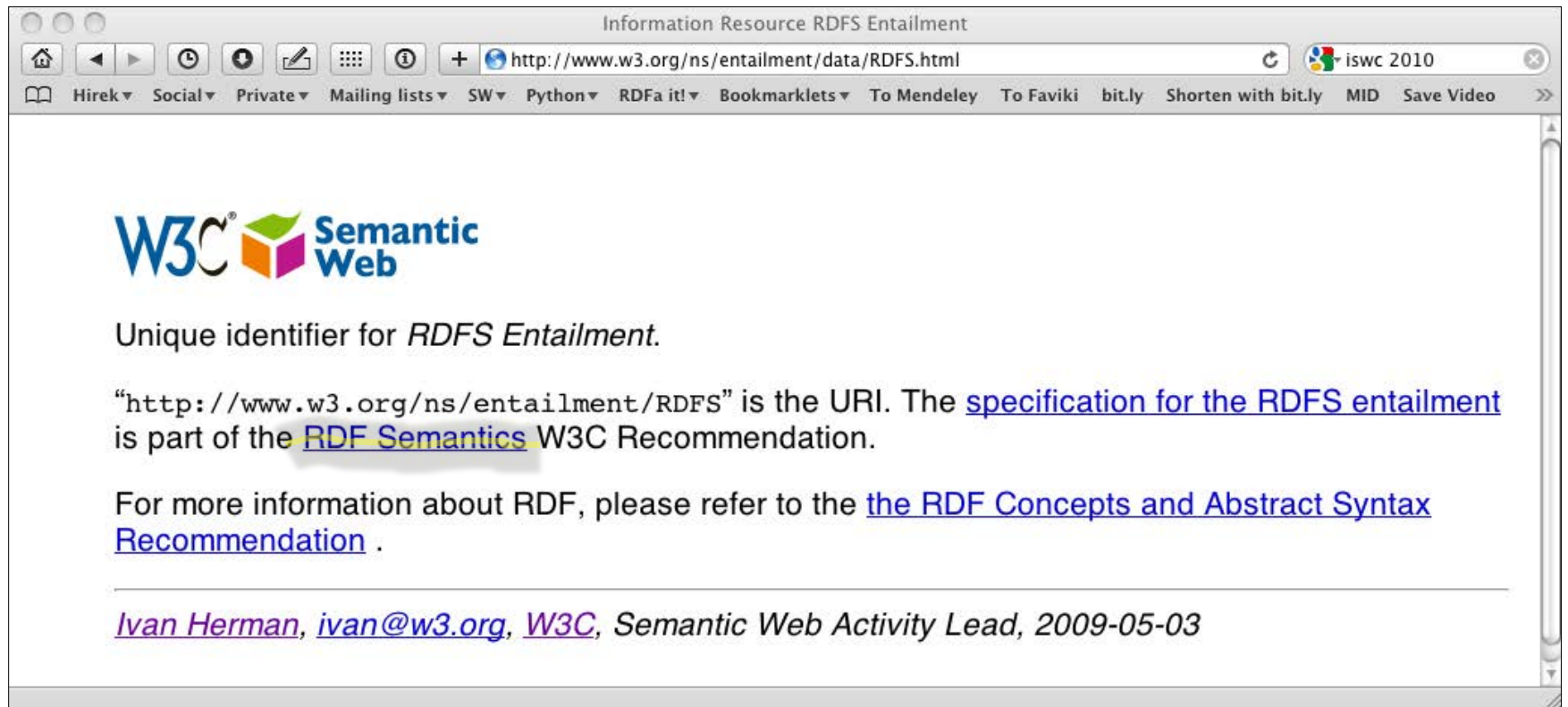
# HTML Source & RDF extracted ...

```
<p about="http://www.w3.org/ns/entailment/RDFS"
    property="http://purl.org/dc/terms/description">
Unique identifier for <em>RDFS Entailment</em>.
</p>
```

```
<http://www.w3.org/ns/entailment/RDFS>
    <http://purl.org/dc/terms/description>
        "Unique identifier for RDFS Entailment." .
```

# Web page viewed by a person

# HTML Source

```
<a about="http://www.w3.org/ns/entailment/RDFS"
   rel="http://www.w3.org/2000/01/rdf-schema#seeAlso"
   href="http://www.w3.org/TR/2004/REC-rdf-mt-20040210/">
RDF Semantics.
</a>
```

# HTML Source & RDF extracted …

```
<a about="http://www.w3.org/ns/entailment/RDFS"
    rel="http://www.w3.org/2000/01/rdf-schema#seeAlso"
    href="http://www.w3.org/TR/2004/REC-rdf-mt-20040210/">
RDF Semantics.
</a>
```

```
<http://www.w3.org/ns/entailment/RDFS>
   ….
```

# HTML Source & RDF extracted ...

```
<a about="http://www.w3.org/ns/entailment/RDFS"
    rel="http://www.w3.org/2000/01/rdf-schema#seeAlso"
    href="http://www.w3.org/TR/2004/REC-rdf-mt-20040210/">
RDF Semantics.
</a>
```

```
<http://www.w3.org/ns/entailment/RDFS>
    <http://www.w3.org/2000/01/rdf-schema#seeAlso>
        … .
```

# HTML Source & RDF extracted ...

```
<a about="http://www.w3.org/ns/entailment/RDFS"
   rel="http://www.w3.org/2000/01/rdf-schema#seeAlso"
   href="http://www.w3.org/TR/2004/REC-rdf-mt-20040210/">
RDF Semantics.
</a>
```

```
<http://www.w3.org/ns/entailment/RDFS>
   <http://www.w3.org/2000/01/rdf-schema#seeAlso>
      <http://www.w3.org/TR/2004/REC-rdf-mt-20040210/> .
```

# It's like Ntriples in HTML ☹

- **Maybe we can do better, instead of this**

```
<http://www.w3.org/ns/entailment/RDFS>
    <http://purl.org/dc/terms/description>
        "Unique identifier for RDFS Entailment." .
<http://www.w3.org/ns/entailment/RDFS>
    <http://www.w3.org/2000/01/rdf-schema#seeAlso>
        <http://www.w3.org/TR/2004/REC-rdf-mt-20040210/> .
```

- **Allow *URI* prefixes and shared subject, like**

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dcterms: <http://purl.org/dc/terms/> .

<http://www.w3.org/ns/entailment/RDFS>
    rdfs:seeAlso <http://www.w3.org/TR/2004/REC-rdf-mt-20040210/> ;
    dcterms:description "Unique identifier for RDFS Entailment." .
```

# Turtlizing  RDFa

- Turtle supports several simplifying ideas
- Use **compact URIs** (CURIE) when possible
  - URI with a prefix defined elsewhere, e.g., *foaf:mbox*
- Making use of the natural structure for
  - shared subjects
  - shared predicates
  - creating blank nodes
  - etc.

# CURIE definition and usage

```html
<html>
  …
  <p about="http://www.w3.org/ns/entailment/RDFS"
     property="http://purl.org/dc/terms/description">
       Unique identifier for <em>RDFS Entailment</em>.</p>
  …
</html>
```

- can be replaced by:

```html
<html prefix="dcterms:http://purl.org/dc/terms/">
  …
  <p about="http://www.w3.org/ns/entailment/RDFS"
     property="dcterms:description">
       Unique identifier for <em>RDFS Entailment</em>.</p>
  …
</html>
```

# Details on @prefix in RDFa

- Can be anywhere in HTML tree and holds for entire sub-tree

  - i.e., HTML element not the only place to have it

- Same @prefix attribute can hold several definitions:

  - prefix="dcterm: http://purl.org… foaf: http://…"

- CURIEs and "real" URIs can usually be mixed

- CURIEs *cannot* be used on @href

# Sharing subjects

Basic principle: @about is inherited by children nodes, so no reason to repeat it

```
<html prefix="dcterms: http://purl.org/dc/terms/
              rdfs: http://www.w3.org/2000/01/rdf-schema#">
  …
  <body about="http://www.w3.org/ns/entailment/RDFS" >

    …

    <p property="dcterms:description">
      Unique identifier for <em>RDFS Entailment</em>.</p>
    <p>…<a rel="rdfs:seeAlso"
      href="http://www.w3.org/TR/2004/REC-rdf-mt-20040210">
      RDFS Semantics</a>…</p>
```

# ... Yielding this RDF

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dcterms: <http://purl.org/dc/terms/> .

<http://www.w3.org/ns/entailment/RDFS>
    rdfs:seeAlso <http://www.w3.org/TR/2004/REC-rdf-mt-20040210/> ;
    dcterms:description "Unique identifier for RDFS Entailment." .
```

# On subjects and objects

- Here is our rule so far
  - **@about** sets the subject
  - **@href** sets the object
- But that is not always good enough
  - We may not want to introduce an active link (i.e., "a" element) on the web page
  - what about other links in HTML?

# We may not always want links...

- The RDFa **@resource attribute** is equivalent to @href

- Sets the object, just like @href but is ignored by browsers, e.g.,:

```
<span about="http://www.ivan-herman.net/foaf#me">
  <span rel="rdfs:seeAlso"
    resource="http://www.w3.org/People/Ivan/">Activity Lead</span>
</span>
```

# More features

- RDFa 1.1 has more features that make it easier to represent knowledge compactly in HTML

- These take advantage of the HTML tree context

- We'll skip the details, which you can find in

  - RDFa 1.1 Primer

  - RDFa 1.1 Core

# Authoring RDFa

- Some tools already have RDFa facilities:
  - e.g., it is possible to add the right DTD to Dreamweaver, Amaya has it at its core, etc.
- There are plugins to, e.g., WordPress, to generate RDFa markup
- CMS systems (like Drupal 7) may have RDFa built in their publication system
  - users generate RDFa whether they know about it or not…

# Consuming RDFa

- Major search engines (Google, Yahoo) process RDFa for vocabularies they understand can use

- There are libraries, distillers, etc., to extract RDFa information

  - may be part of RDF development environments like Redland, RDFLib

  - see, for further references, http://rdfa.info/wiki/Consume

- Facebook's "social graph" is based on RDFa

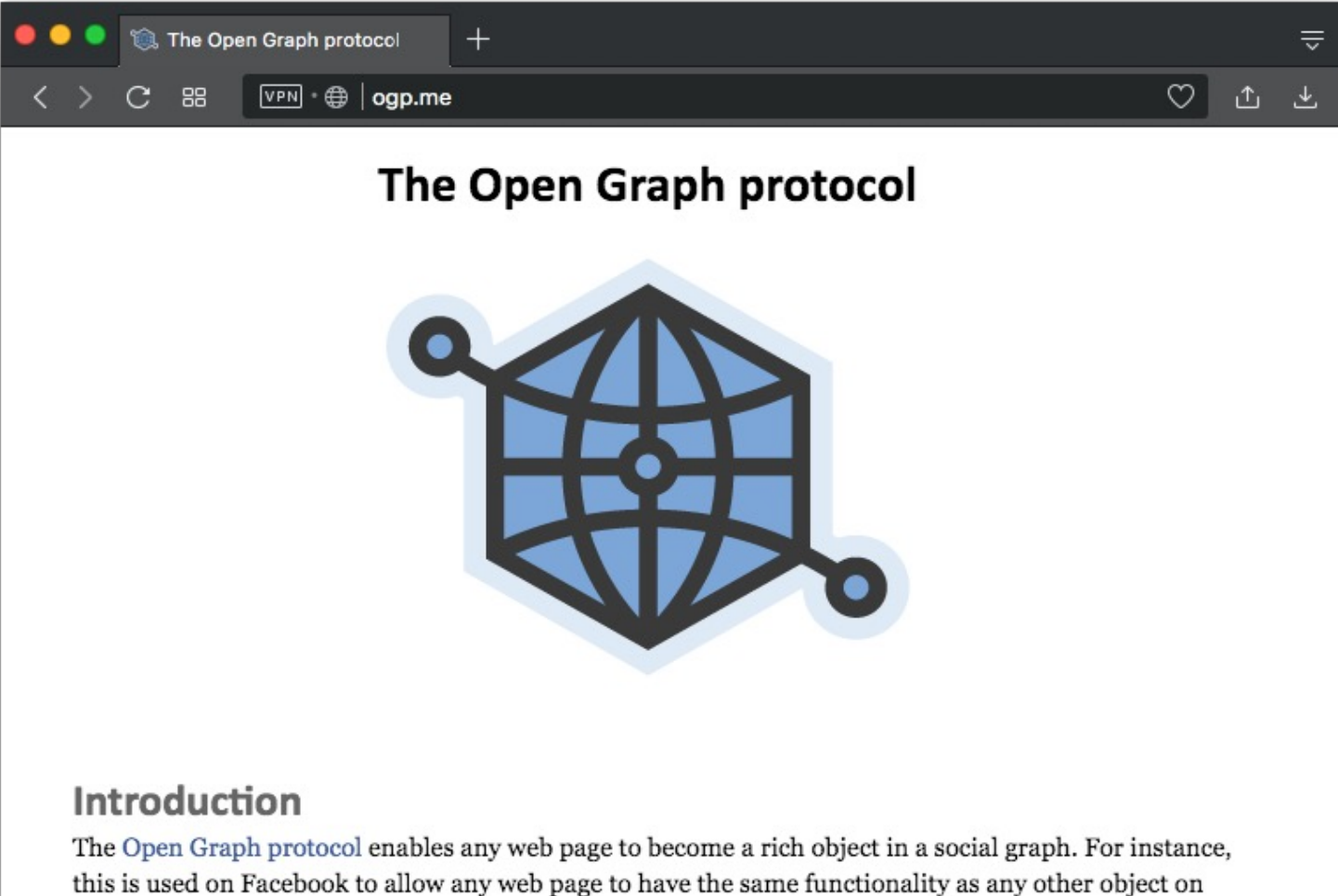Like

# Best Buy Pixel 7 page

# getdata.py is very simple*

```
import rdflib, sys
if not (1 < len(sys.argv) < 4):
    print("usage: python getdata.py url ['json-ld'|rdfa|rdfa1.1|microdata|html] ")
    sys.exit(0)

url = sys.argv[1]
format = sys.argv[2] if len(sys.argv) == 3 else 'rdfa1.1'
g = rdflib.Graph()
g.parse(url, format=format)
print g.serialize(format='n3')
```

* Currently you must use an older version of rdflib, e.g., 4.2.2

# FaceBook's Open Graph Protocol

- OGP lets web developers better present their pages on Facebook
- Twitter has a similar system for twitter cards



## The Open Graph protocol

### Introduction

The Open Graph protocol enables any web page to become a rich object in a social graph. For instance, this is used on Facebook to allow any web page to have the same functionality as any other object on

# What is RDFa Lite?

- **RDFa 1.1 Lite** is a subset of RDFa 1.1

- Five simple attributes: vocab, typeof, property, resource, and prefix

- Completely upwards compatible RDFa 1.1

- Works well with schema.org terms

- Consists of five simple attributes: vocab, typeof, property, resource and prefix

# RDFa Lite example

The *vocab* attribute sets the default vocabulary for a block, *typeof* sets the class and *property* introduces a property

```
<p vocab="http://schema.org/" typeof="Person">
 My name is
 <span property="name">Manu Sporny</span>
 and you can give me a ring via
 <span property="telephone">1-800-555-0199</span> or visit
 <a property="url" href="http://manu.sporny.org/">my homepage</a>
</p>
```

# RDFa Lite vs. Microdata

**RDFa Lite**

```
<p vocab="http://schema.org/" typeof="Person">
 My name is
 <span property="name">Manu Sporny</span>
 and you can give me a ring via
 <span property="telephone">1-800-555-0199</span> or visit
 <a property="url" href="http://manu.sporny.org/">my homepage</a>
</p>
```

**Micro-data**

```
<p itemscope itemtype="http://schema.org/Person">
 My name is
 <span itemprop="name">Manu Sporny</span>
 and you can give me a ring via
 <span itemprop="telephone">1-800-555-0199</span> or visit
 <a itemprop="url" href="http://manu.sporny.org/">my homepage</a>
</p>
```

# Microdata vs RDFa lite serialization

- The RDFa Lite serialization looks almost isomorphic to the older Microdata version

- Changes:
  - itemprop -> property
  - itemscope is dropped
  - itemtype-> typeof
  - vocab="http://schema.org/" added to the body or some other enclosing taga

# RDFa Lite example: resource

*Resource* attribute gives object value (URL) for subject and *prefix* attribute eases mixing vocabularies

```
<p vocab="http://schema.org/"
   typeof="Person"
   resource="#manu"
   prefix="ov:http://open.vocab.org/terms/" >
  My favorite animal is the
    <span property="ov:preferredAnimal">Liger</span>
</p>
```

# RDFa Lite Conclusions

- One advantage of Microdata markup was it was simpler than RDFa

- RDFa Lite offers the same simplicity

- But with two advantages:

  – You can add statements from multiple RDF vocabularies

  – You can take advantage of more complex RDFa markup features if and when needed

# Conclusions

- Web developers want content providers to add structured data to HTML pages

- Content providers are incentivized to do so because their content will be better understood, ranked higher, more useful, etc.

- RDFa is most powerful & flexible knowledge markup standard understood by search engines

- RDFa Lite is simpler and covers most use cases