

# Microdata and schema.org

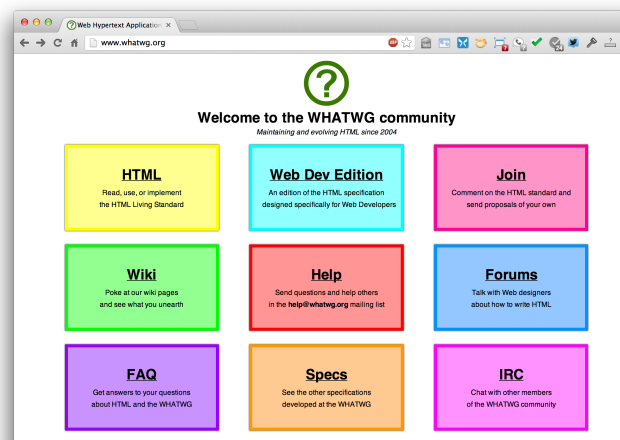
## Basics

- [Microdata](#) is a simple semantic markup scheme that's an alternative to RDFa
- Developed by WHATWG and supported by major search companies (Google, Microsoft, Yahoo, Yandex)
- Like RDFa, it uses HTML tag attributes to host metadata
- Vocabularies are controlled and hosted at [schema.org](http://schema.org)

## What is WHATWG?

- [Web Hypertext Application Technology Working Group](#)
  - Community interested in evolving the Web with focus on HTML and Web API development
  - [Ian Hickson](#) is a key person, now at Google
- Founded in 2004 by individuals from Apple, Mozilla and Opera after a W3C workshop
  - Concern about W3C's embrace of XHTML
- Current work on [HTML5](#)
- Developed [Microdata](#) spec

## <http://whatwg.org/>



## HTML5



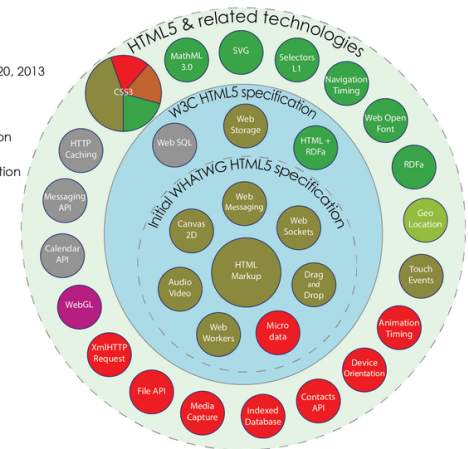
- Started by WHATWG as an alternative to XHTML, joined by W3C
  - A W3C candidate recommendation in 2012 ([draft](#))
  - WHATWG will evolve it as a “living standard”
- HTML5 ≈ HTML + CSS + js
- Native support for graphics, video, audio, speech, semantic markup, ...
- Partial support in current browsers & extensions

## HTML taxonomy and status

### HTML5

Taxonomy & Status on January 20, 2013

- W3C Recommendation
- Proposed Recommendation
- Candidate Recommendation
- Last Call
- Working Draft
- Non-W3C Specifications
- Deprecated



by Sergey Mavrody © BY-SA

## Microdata

- The microdata effort has two parts:
  - markup and
  - a set of vocabularies
- The markup is similar to RDFa in providing ways to identify subjects, types, properties & objects
  - There’s also a standard way to encode microdata as RDFa
- The sanctioned vocabularies are found at [schema.org](http://schema.org) and include a small number of very useful ones: people, movies, etc.

## An example

```
<div>
  <h1>Avatar</h1>
  <span>Director: James Cameron (born 1954) </span>
  <span>Science fiction</span>
  <a href="avatar-trailer.html">Trailer</a>
</div>
```

## An example: itemscope

- An *itemscope* attribute identifies a content subtree that is the subject about which we want to say something

```
<div itemscope >
  <h1>Avatar</h1>
  <span>Director: James Cameron (born 1954) </span>
  <span>Science fiction</span>
  <a href="avatar-trailer.html">Trailer</a>
</div>
```

## An example: itemtype

- An *itemscope* attribute identifies a content subtree that is the subject about which we want to say something
- The *itemtype* attribute specifies the subject's type

```
<div itemscope itemtype="http://schema.org/Movie">
  <h1>Avatar</h1>
  <span>Director: James Cameron (born 1954) </span>
  <span>Science fiction</span>
  <a href="avatar-trailer.html">Trailer</a>
</div>
```

## An example: itemprop

- An *itemscope* attribute identifies a content subtree that is the subject about which we want to say something
- The *itemtype* attribute specifies the subject's type
- An *itemprop* attribute gives a property of that type

```
<div itemscope itemtype="http://schema.org/Movie">
  <h1 itemprop="name">Avatar</h1>
  <span>Director: James Cameron (born 1954) </span>
  <span itemprop="genre">Science fiction</span>
  <a href="avatar-trailer.html" itemprop="trailer">Trailer</a>
</div>
```

## An example: embedded items

- An *itemprop* immediately followed by another *itemscope* makes the value an object

```
<div itemscope itemtype="http://schema.org/Movie">
  <h1 itemprop="name">Avatar</h1>
  <div itemprop="director"
    itemscope itemtype="http://schema.org/Person">
    Director: <span itemprop="name">James Cameron</span> (born
  <span itemprop="birthDate">1954</span>) </div>
  <span itemprop="genre">Science fiction</span>
  <a href="avatar-trailer.html" itemprop="trailer">Trailer</a>
</div>
```



## Mixing markup from other vocabularies

- Microdata is intended to work with just one vocabulary – the one at schema.org
- Advantages
  - Simple, organized, well designed
  - Controlled by the schema.org people
- Disadvantages: too simple, controlled
  - Too simple, narrow, mono-lingual
  - Controlled by the schema.org people

## Extending the schema.org ontology

- <http://www.schema.org/docs/extension.html>
- You can subclass existing classes
  - Person/Engineer
  - Person/Engineer/ElectricalEngineer
- Subclass existing properties
  - musicGroupMember/leadVocalist
  - musicGroupMember/leadGuitar1
  - musicGroupMember/leadGuitar2

## Extension Problems

- Do agreed upon meaning
  - Through axioms supported by the language (e.g., equivalence, disjointness, etc.)
  - No place for documentation (annotations, labels, comments)
- Without a namespace mechanism, your Person/Engineer and mine can be confused and might mean different things

## Conclusions

- Microdata is a good effort by the search companies to use a simple semantic language
- RDFa has a more powerful encoding and works with the RDF stack
- There's a bit of infighting in the WEB community
- RDFa Lite is maybe a good solution