Microdata for Dallas County Historical & Genealogical Cemetery Data

Tony Hanson
Webmaster
Presentation & Related Info

http://dallasgenealogy.org/DigitalFrontiers
What You Will Learn

- **What** is Microdata
- **How** Microdata is used
- **Why** Microdata is significant
The Problem

• The goal of a well designed web page is to make information useful to natural language interpreters
  – a.k.a. ‘People’

• This was the original goal/purpose of HTML
There Are Several Classes of Users

• **Browsers**

• **Web Crawlers**
  – Where our site appears on the results list depends a lot on what the web crawlers find
  – Organizing and tagging data in a way that is meaningful to the search engine is very important...

• **Web based applications** that use your data
HTML As Obfuscator

• ‘Reverse Engineering’ the structure and relationships on the original data from HTML is difficult...
  – Maybe even impossible
Data Source: Relational Database Table

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>cemetery_name</td>
<td>varchar(60)</td>
</tr>
<tr>
<td>first_name</td>
<td>varchar(30)</td>
</tr>
<tr>
<td>last_name</td>
<td>varchar(30)</td>
</tr>
<tr>
<td>spouse_name</td>
<td>varchar(60)</td>
</tr>
<tr>
<td>birth_date</td>
<td>varchar(25)</td>
</tr>
<tr>
<td>death_date</td>
<td>varchar(25)</td>
</tr>
<tr>
<td>burial_date</td>
<td>varchar(25)</td>
</tr>
<tr>
<td>section</td>
<td>varchar(20)</td>
</tr>
<tr>
<td>grave</td>
<td>varchar(15)</td>
</tr>
<tr>
<td>stone_type</td>
<td>varchar(15)</td>
</tr>
</tbody>
</table>
Josephine **Meyer** Munn

**Date Born:** 12 Nov 1858  
**Date Died:** 24 Feb 1936

**Cemetery:** Oakland  
**Stone Type:** Single  
**Spouse:** [Thomas J. Munn]  
**Date Buried:** 25 Feb 1936
<h2>Josephine <b><u>Meyer</u></b> MUNN</h2>
<table>
  <tr>
    <td align=right><b>Date Born:</b></td>
    <td align=left>12 Nov 1858</td>
  </tr>
  <tr>
    <td align=right><b>Date Died:</b></td>
    <td>24 Feb 1936</td>
  </tr>
  <tr>
    <td align=right><b>Cemetery:</b></td>
    <td>Oak Cliff</td>
  </tr>
  <tr>
    <td align=right><b>Stone Type:</b></td>
    <td>Single</td>
  </tr>
  <tr>
    <td align=right><b>Spouse:</b></td>
    <td>[Thomas J. Munn]</td>
  </tr>
  <tr>
    <td align=right><b>Date Buried:</b></td>
    <td>25 Feb 1936</td>
  </tr>
</table>
Data Structure: Lost In Translation...

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>cemetery_name</td>
<td>varchar(60)</td>
</tr>
<tr>
<td>first_name</td>
<td>varchar(30)</td>
</tr>
<tr>
<td>last_name</td>
<td>varchar(30)</td>
</tr>
<tr>
<td>spouse_name</td>
<td>varchar(60)</td>
</tr>
<tr>
<td>birth_date</td>
<td>varchar(25)</td>
</tr>
<tr>
<td>death_date</td>
<td>varchar(25)</td>
</tr>
<tr>
<td>burial_date</td>
<td>varchar(25)</td>
</tr>
<tr>
<td>section</td>
<td>varchar(20)</td>
</tr>
<tr>
<td>grave</td>
<td>varchar(15)</td>
</tr>
<tr>
<td>stone_type</td>
<td>varchar(15)</td>
</tr>
</tbody>
</table>

Josephine <b><u>Meyer</u></b> MUNN

<table>
  <tr>
    <td align=right><b>Date Born:</b></td>
    <td align=left>12 Nov 1858</td>
  </tr>
  <tr>
    <td align=right><b>Date Died:</b></td>
    <td>24 Feb 1936</td>
  </tr>
  <tr>
    <td align=right><b>Cemetery:</b></td>
    <td>Oak Cliff</td>
  </tr>
  <tr>
    <td align=right><b>Stone Type:</b></td>
    <td>Single</td>
  </tr>
  <tr>
    <td align=right><b>Spouse:</b></td>
    <td>[Thomas J. Munn]</td>
  </tr>
  <tr>
    <td align=right><b>Date Buried:</b></td>
    <td>25 Feb 1936</td>
  </tr>
</table>
Sharing Data & Structure

• Possible using a variety of methods
• Most are programmer intensive to create
• Each one is unique
• May expose your system to security risks
• Has little (or nothing) to do with the actual web page
Two Views Of The Data

API View

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>cemetery_name</td>
<td>varchar(60)</td>
</tr>
<tr>
<td>first_name</td>
<td>varchar(30)</td>
</tr>
<tr>
<td>last_name</td>
<td>varchar(30)</td>
</tr>
<tr>
<td>spouse_name</td>
<td>varchar(60)</td>
</tr>
<tr>
<td>birth_date</td>
<td>varchar(25)</td>
</tr>
<tr>
<td>death_date</td>
<td>varchar(25)</td>
</tr>
<tr>
<td>burial_date</td>
<td>varchar(25)</td>
</tr>
<tr>
<td>section</td>
<td>varchar(20)</td>
</tr>
<tr>
<td>grave</td>
<td>varchar(15)</td>
</tr>
<tr>
<td>stone_type</td>
<td>varchar(15)</td>
</tr>
</tbody>
</table>

Browser/Crawler View

```html
<h2>Josephine <u>Meyer</u> MUNN</h2>
<table>
<tr>
<td align=right><b>Date Born:</b></td>
<td align=left>12 Nov 1858</td>
</tr>
<tr>
<td align=right><b>Date Died:</b></td>
<td>24 Feb 1936</td>
</tr>
<tr>
<td align=right><b>Cemetery:</b></td>
<td>Oak Cliff</td>
</tr>
<tr>
<td align=right><b>Stone Type:</b></td>
<td>Single</td>
</tr>
<tr>
<td align=right><b>Spouse:</b></td>
<td>[Thomas J. Munn]</td>
</tr>
<tr>
<td align=right><b>Date Buried:</b></td>
<td>25 Feb 1936</td>
</tr>
</table>
```
Microdata

A vocabulary and syntax used to extend HTML with additional machine-readable semantics
Microdata Attributes

• Extension to existing HTML & CSS markup techniques
• Machine readable
  – Usually not seen by the Browser user
• Controlled Vocabulary (defines the scope)
• Information identified as Property-Value pairs
How Does Microdata Help?

• Through the use of additional tags, it is possible to identify the specific meaning for each item of information
  – These tags do not alter the way the information is displayed by the browser
  – However, they do provide a great deal additional information that can be used by the browser, the web crawlers and other applications
• Launched in June 2011 by Bing, Google and Yahoo
  • Yandex – Russia’s largest search engine – has since signed onto the effort
• Goal: To create and support a common set of schemas for structured data markup on web pages
What is Schema.org?

This site provides a collection of schemas, i.e., html tags, that webmasters can use to markup their pages in ways recognized by major search providers. Search engines including Bing, Google, Yahoo! and Yandex rely on this markup to improve the display of search results, making it easier for people to find the right web pages.

Many sites are generated from structured data, which is often stored in databases. When this data is formatted into HTML, it becomes very difficult to recover the original structured data. Many applications, especially search engines, can benefit greatly from direct access to this structured data. On-page markup enables search engines to understand the information on web pages and provide richer search results in order to make it easier for users to find relevant information on the web. Markup can also enable new tools and applications that make use of the structure.

A shared markup vocabulary makes it easier for webmasters to decide on a markup schema and get the maximum benefit for their efforts. So, in the spirit of sitemaps.org, search engines have come together to provide a shared collection of schemas that webmasters can use.

We invite you to get started!

View our blog at blog.schema.org.
The Type Hierarchy

• Thing
  – CreativeWork
  – Event
  – MedicalEntity
  – Organization
  – Person
  – Place
  – Product
**Thing > Event**

An event happening at a certain time at a certain location.

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>additionalType</td>
<td>URL</td>
<td>An additional type for the item, typically used for adding more specific types from external vocabularies in microdata syntax. This is a relationship between something and a class that the thing is in. In RDFa syntax, it is better to use the native RDFa syntax – the 'typeof' attribute – for multiple types. Schema.org tools may have only weaker understanding of extra types, in particular those defined externally.</td>
</tr>
<tr>
<td>description</td>
<td>Text</td>
<td>A short description of the item.</td>
</tr>
<tr>
<td>image</td>
<td>URL</td>
<td>URL of an image of the item.</td>
</tr>
<tr>
<td>name</td>
<td>Text</td>
<td>The name of the item.</td>
</tr>
<tr>
<td>sameAs</td>
<td>URL</td>
<td>URL of a reference Web page that unambiguously indicates the item's identity. E.g. the URL of the item's Wikipedia page, Freebase page, or official website.</td>
</tr>
<tr>
<td>url</td>
<td>URL</td>
<td>URL of the item.</td>
</tr>
<tr>
<td>Property</td>
<td>Type</td>
<td>Description</td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td>-------------</td>
</tr>
<tr>
<td>attendee</td>
<td>Organization or Person</td>
<td>A person or organization attending the event.</td>
</tr>
<tr>
<td>attendees</td>
<td>Organization or Person</td>
<td>A person attending the event (legacy spelling; see singular form, attendee).</td>
</tr>
<tr>
<td>duration</td>
<td>Duration</td>
<td>The duration of the item (movie, audio recording, event, etc.) in ISO 8601 date format.</td>
</tr>
<tr>
<td>endDate</td>
<td>Date</td>
<td>The end date and time of the event (in ISO 8601 date format).</td>
</tr>
<tr>
<td>location</td>
<td>Place or PostalAddress</td>
<td>The location of the event, organization or action.</td>
</tr>
<tr>
<td>offers</td>
<td>Offer</td>
<td>An offer to sell this item—for example, an offer to sell a product, the DVD of a movie, or tickets to an event.</td>
</tr>
<tr>
<td>performer</td>
<td>Organization or Person</td>
<td>A performer at the event—for example, a presenter, musician, musical group or actor.</td>
</tr>
<tr>
<td>performers</td>
<td>Organization or Person</td>
<td>The main performer or performers of the event—for example, a presenter, musician, or actor (legacy spelling; see singular form, performer).</td>
</tr>
<tr>
<td>startDate</td>
<td>Date</td>
<td>The start date and time of the event (in ISO 8601 date format).</td>
</tr>
<tr>
<td>subEvent</td>
<td>Event</td>
<td>An Event that is part of this event. For example, a conference event includes many presentations, each are a subEvent of the conference.</td>
</tr>
<tr>
<td>subEvents</td>
<td>Event</td>
<td>Events that are a part of this event. For example, a conference event includes many presentations, each are subEvents of the conference (legacy spelling; see singular form, subEvent).</td>
</tr>
<tr>
<td>superEvent</td>
<td>Event</td>
<td>An event that this event is a part of. For example, a collection of individual music performances might each have a music festival as their superEvent.</td>
</tr>
</tbody>
</table>
Example 1

Original HTML:

```
<a href="nba-miami-philidelphia-game3.html">
NBA Eastern Conference First Round Playoff Tickets:
  Miami Heat at Philadelphia 76ers - Game 3 (Home Game 1)
</a>

Thu, 04/21/16
8:00 p.m.

<a href="wells-fargo-center.html">
Wells Fargo Center
</a>
Philadelphia, PA

Priced from: $35
1938 tickets left
```
With Schema.org:

```
<div itemscope itemtype="http://schema.org/Event">

<a itemprop="url" href="nba-miami-philadelphia-game3.html">
NBA Eastern Conference First Round Playoff Tickets:
</a>

<span itemprop="name">Miami Heat at Philadelphia 76ers - Game 3 (Home Game 1)</span>

<meta itemprop="startDate" content="2016-04-21T20:00">
Thu, 04/21/16
8:00 p.m.
```

This is an Event
With Schema.org:

```
<div itemscope itemtype="http://schema.org/Event">
  <a itemprop="url" href="nba-miami-philadelphia-game3.html">
    NBA Eastern Conference First Round Playoff Tickets:
  </a>
  <span itemprop="name">Miami Heat at Philadelphia 76ers - Game 3 (Home Game 1) </span>
</div>

<meta itemprop="startDate" content="2016-04-21T20:00">
  Thu, 04/21/16
  8:00 p.m.

<div itemprop="location" itemscope itemtype="http://schema.org/Place">
  <a itemprop="url" href="wells-fargo-center.html">
    Wells Fargo Center
  </a>
  <div itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">
    <span itemprop="addressLocality">Philadelphia</span>,
    <span itemprop="addressRegion">PA</span>
  </div>
</div>
```
With Schema.org:

```xml
<div itemscope itemtype="http://schema.org/Event">
  <a itemprop="url" href="nba-miami-philadelphia-game3.html">
    NBA Eastern Conference First Round Playoff Tickets:
    <span itemprop="name">Miami Heat at Philadelphia 76ers - Game 3 (Home Game 1)</span>
  </a>

  <meta itemprop="startDate" content="2016-04-21T20:00">
  Thu, 04/21/16
  8:00 p.m.

  <div itemprop="location" itemscope itemtype="http://schema.org/Place">
    <a itemprop="url" href="wells-fargo-center.html">
      Wells Fargo Center
    </a>
    <div itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">
      <span itemprop="addressLocality">Philadelphia</span>,
      <span itemprop="addressRegion">PA</span>
    </div>
  </div>
</div>
```

US Mail Address
For Sale

For Sale:

With Schema.org:

```html
<div itemscope itemtype="http://schema.org/Event">
  <a itemprop="url" href="nba-miami-philadelphia-game3.html">
    NBA Eastern Conference First Round Playoff Tickets:
  </a>
  <span itemprop="name">Miami Heat at Philadelphia 76ers - Game 3 (Home Game 1) </span>
</div>

<meta itemprop="startDate" content="2016-04-21T20:00"/>
Thu, 04/21/16
8:00 p.m.

<div itemprop="location" itemscope itemtype="http://schema.org/Place">
  <a itemprop="url" href="wells-fargo-center.html">
    Wells Fargo Center
  </a>
  <span itemprop="addressLocality">Philadelphia</span>,
  <span itemprop="addressRegion">PA</span>
</div>

<div itemprop="offers" itemscope itemtype="http://schema.org/AggregateOffer">
  Priced from: <span itemprop="lowPrice">$35</span>
  <span itemprop="offerCount">1938</span> tickets left
</div>
```
Metadata Is Available To All Users

• Web Crawlers have more meaningful data
• Web based applications can understand the structure and interrelationships of your data
• Browsers can do some useful things with it too...
Chrome

• Google’s Chrome Browser supports extensions
• People Inspector
  – Senses when a page has information tagged using the schema.org People Schema
  – Allows the user to launch queries to several web sites
DGS Cemetery Database

• Data gathered by volunteers for 13 years
• Sources:
  – Cemetery Records
  – Tomb Stones
  – Other Records
• Nearly 35,000 records
• Has been tagged with schema.org schema’s
William Henry ABBOTT

Date Born: 29 Nov 1847
Date Died: 08 Jan 1912

Cemetery: Oak Cliff
Stone Type: Single
Spouse:
Date Buried: 09 Jan 1912
Age:
Section: Old 1
Subsection: Lot
Subsection Number: 22a
Part of Lot:
Grave #: 
Lot Part:
Lot Owner:
Funeral Home:
Tombstone Inscription:
Notes:

Military Service: 
<table>
<thead>
<tr>
<th>Name</th>
<th>Birth</th>
<th>Death</th>
<th>Copy</th>
<th>Search</th>
</tr>
</thead>
<tbody>
<tr>
<td>William Henry ABBOTT</td>
<td>11/29/1847</td>
<td>1/8/1912</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Date Born:** 29 Nov 1847  
**Date Died:** 08 Jan 1912

**Notes:**  
**Military Service:**
Refine your search

First Names
William

Last Names
ABBOTT

Restrict records by:
Location
Type
Batch Number
Film Number

Search with a life event:
Birthplace

Birth Year (Range)
1847 1847

Marriage
Residence

Death Place

Death Year (Range)
1912 1912
<table>
<thead>
<tr>
<th>Name</th>
<th>William Henry Abbott</th>
</tr>
</thead>
<tbody>
<tr>
<td>Titles and Terms</td>
<td></td>
</tr>
<tr>
<td>Event Type</td>
<td>Death</td>
</tr>
<tr>
<td>Event Date</td>
<td>08 Jan 1912</td>
</tr>
<tr>
<td>Event Place</td>
<td>Dallas, Texas</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
</tr>
<tr>
<td>Certificate Number</td>
<td>562</td>
</tr>
<tr>
<td>Death Date</td>
<td>8 January 1912</td>
</tr>
<tr>
<td>Death Place</td>
<td>Dallas, Texas</td>
</tr>
</tbody>
</table>

| John William Abbott   | Christening: 20 November 1847     |
|                       | Tydd Saint Mary, Lincoln, England  |
|                       |                                    |
| William Henry Abbott  | Christening: 27 June 1847         |
|                       | St. Thomas the Apostle, Cornwall,  |
|                       | England                            |

| Death Year (Range)    | 1912                               |
| Any                   |                                    |

| Death Year (Range)    | 1912                               |
| Any                   |                                    |
TALKS | IN LESS THAN 6 MINUTES
Tim Berners-Lee: The year open data went worldwide
FILMED FEB 2010 • POSTED MAR 2010 • TED2010
Summary

• Microdata provides semantic meaning for your web data
• This makes your data more meaningful to crawlers, browsers and other web services
• schema.org appears to be the dominant standard for the major search engines
Thank You!

http://dallascgenealogy.org/DigitalFrontiers

Tony Hanson
Webmaster