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

Field	Type
cemetery_name	varchar(60)
first_name	varchar(30)
last_name	varchar(30)
spouse_name	varchar(60)
birth_date	varchar(25)
death_date	varchar(25)
burial_date	varchar(25)
section	varchar(20)
grave	varchar(15)
stone_type	varchar(15)

Resulting Page

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

HTML

```
<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…

Field	Type
cemetery_name	varchar(60)
first_name	varchar(30)
last_name	varchar(30)
spouse_name	varchar(60)
birth_date	varchar(25)
death_date	varchar(25)
burial_date	varchar(25)
section	varchar(20)
grave	varchar(15)
stone_type	varchar(15)

```
<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>
```

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

Field	Type
cemetery_name	varchar(60)
first_name	varchar(30)
last_name	varchar(30)
spouse_name	varchar(60)
birth_date	varchar(25)
death_date	varchar(25)
burial_date	varchar(25)
section	varchar(20)
grave	varchar(15)
stone_type	varchar(15)

API View

```
<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>
```

Browser/Crawler View

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

schema.org

- 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.

Property	Expected Type	Description
Properties from Thing		
additionalType	URL	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.
description	Text	A short description of the item.
image	URL	URL of an image of the item.
name	Text	The name of the item.
sameAs	URL	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.
url	URL	URL of the item.

Property Type Description

Properties from Event		
attendee	Organization or Person	A person or organization attending the event.
attendees	Organization or Person	A person attending the event (legacy spelling; see singular form, attendee).
duration	Duration	The duration of the item (movie, audio recording, event, etc.) in ISO 8601 date format.
endDate	Date	The end date and time of the event (in ISO 8601 date format).
location	Place or PostalAddress	The location of the event, organization or action.
offers	Offer	An offer to sell this item—for example, an offer to sell a product, the DVD of a movie, or tickets to an event.
performer	Organization or Person	A performer at the event—for example, a presenter, musician, musical group or actor.
performers	Organization or Person	The main performer or performers of the event—for example, a presenter, musician, or actor (legacy spelling; see singular form, performer).
startDate	Date	The start date and time of the event (in ISO 8601 date format).
subEvent	Event	An Event that is part of this event. For example, a conference event includes many presentations, each are a subEvent of the conference.
subEvents	Event	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).
superEvent	Event	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.

Property Type Description

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
```

This is an Event

With Schema.org:

```
<div itemscope itemtype="http://schema.org/Event">
  <a itemprop="url" href="nba-miami-philidelphia-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.
```

With Schema.org:

```
<div itemscope itemtype="http://schema.org/Event">
  <a itemprop="url" href="nba-miami-philidelphia-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>
```

Location

With Schema.org:

```
<div itemscope itemtype="http://schema.org/Event">
  <a itemprop="url" href="nba-miami-philidelphia-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>
```

US Mail Address

With Schema.org:

```
<div itemscope itemtype="http://schema.org/Event">
  <a itemprop="url" href="nba-miami-philidelphia-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 itemprop="offers" itemscope itemtype="http://schema.org/AggregateOffer">
    Priced from: <span itemprop="lowPrice">$35</span>
    <span itemprop="offerCount">1938</span> tickets left
  </div>
</div>
```

For Sale

Metadata Is Available To All Users

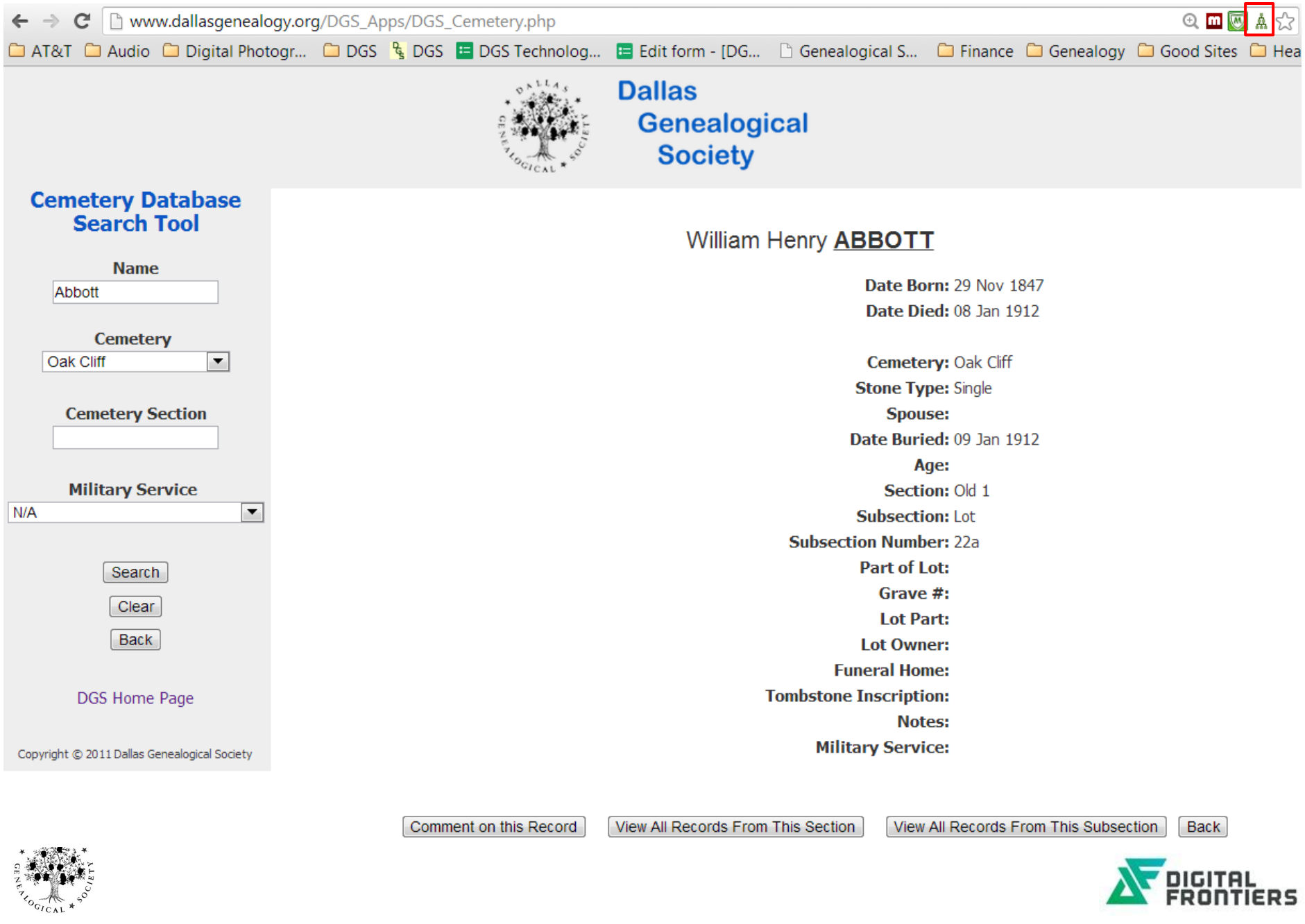
- Web Crawlers have more meaningful data
- Web based applications can userstand the structure and interrelationships of your data
- Browsers can do some useful things with it too…

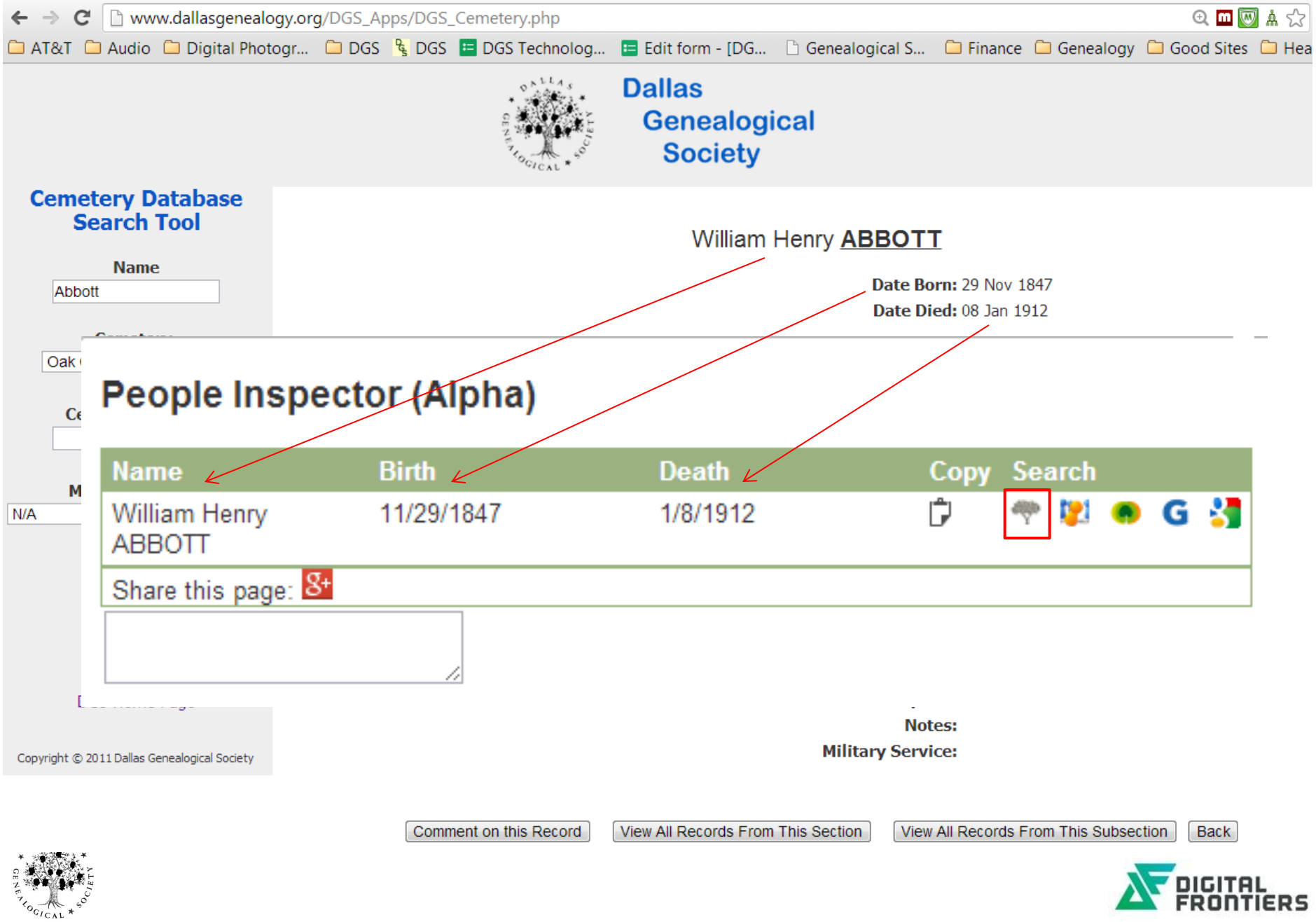
Chrome

- 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


www.dallasgenealogy.org/DGS_Apps/DGS_Cemetery.php
AT&T
Audio
Digital Photogr...
DGS
DGS
DGS Technolog...
Edit form - [DG...
Genealogical S...
Finance
Genealogy
Good Sites
Hea
Dallas Genealogical Society
Cemetery Database Search Tool
Name
Abbott
Cemetery
Oak Cliff
Cemetery Section
Military Service
N/A
Search
Clear
Back
DGS Home Page
Copyright © 2011 Dallas Genealogical Society
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:
Comment on this Record
View All Records From This Section
View All Records From This Subsection
Back
DIGITAL FRONTIERS


www.dallasgenealogy.org/DGS_Apps/DGS_Cemetery.php
AT&T
Audio
Digital Photogr...
DGS
DGS
DGS Technolog...
Edit form - [DG...
Genealogical S...
Finance
Genealogy
Good Sites
Hea
Dallas Genealogical Society
Cemetery Database Search Tool
Name
Abbott
William Henry ABBOTT
Date Born: 29 Nov 1847
Date Died: 08 Jan 1912
People Inspector (Alpha)
Name
Birth
Death
Copy
Search
William Henry ABBOTT
11/29/1847
1/8/1912
Share this page:
Notes:
Military Service:
Copyright © 2011 Dallas Genealogical Society
Comment on this Record
View All Records From This Section
View All Records From This Subsection
Back

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

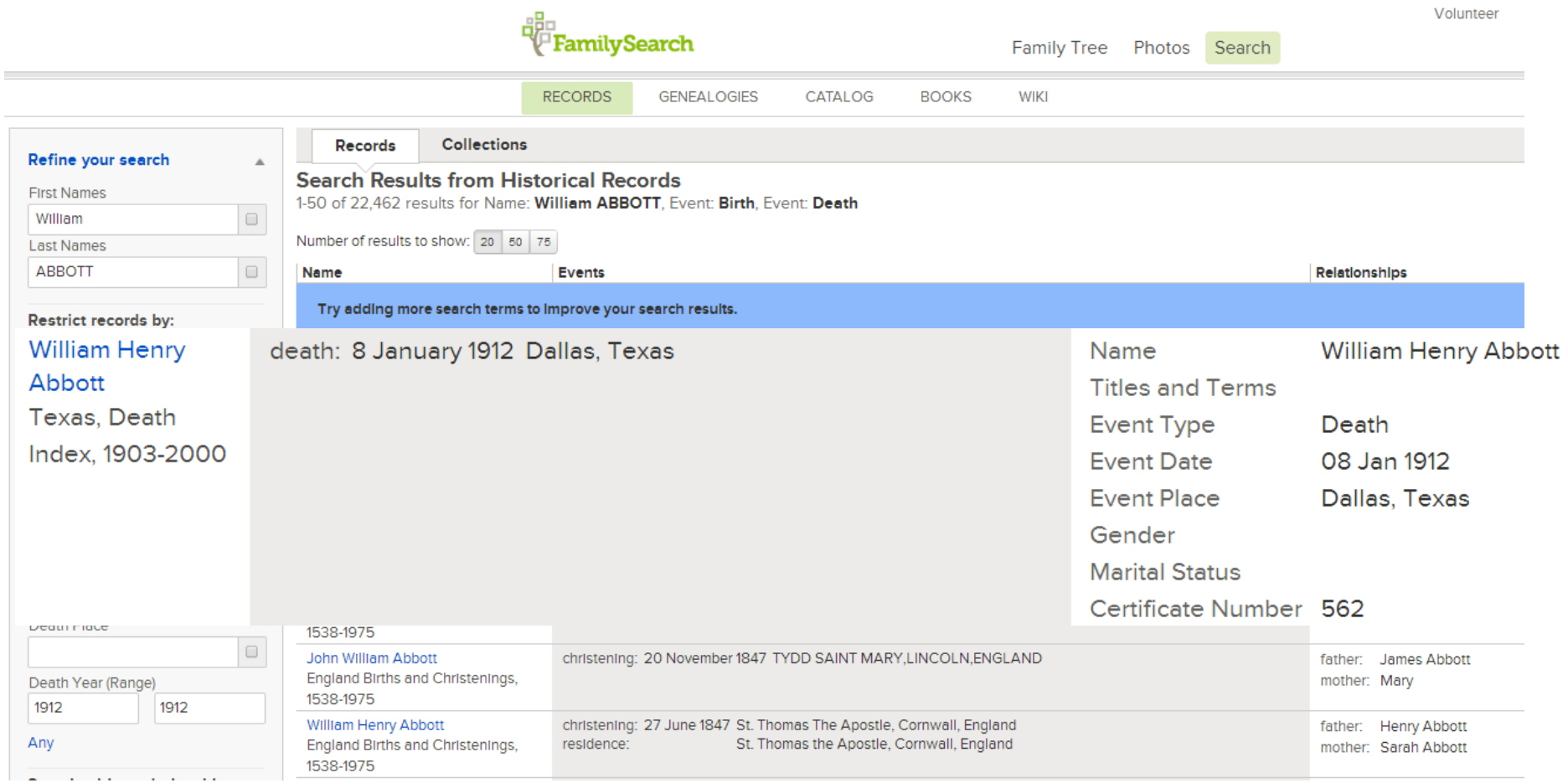
Marriage
Residence

Death Place

Death Year (Range)

1912 1912

FAMILYSEARCH



DALLAS GENEALOGICAL SOCIETY

DIGITAL FRONTIERS

TED Ideas worth spreading

TALKS | IN LESS THAN 6 MINUTES

Tim Berners-Lee: The year open data went worldwide

FILMED FEB 2010 • POSTED MAR 2010 • TED2010

Embed | Download | Favorite | Rate | Show transcript

Get this talk on DVD

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://dallasgenealogy.org/DigitalFrontiers

Tony Hanson
Webmaster