



Search...



WHITE PAPER

EVERYTHING YOU NEED TO KNOW ABOUT SEMANTIC SEO

BY GENNARO CUOFANO, ANDREA VOLPINI,
AND MARIA SILVIA SANNA

TORQUE®

 WPengine®

The Semantic Web is here. Those that are taking advantage of Semantic Technologies to build a Semantic SEO strategy are benefiting from staggering results. From a research paper put together with the team at [WordLift](#), presented at [SEMANTICS 2017](#), we documented that structured data is compelling from the digital marketing standpoint.

For instance, on the analysis of the design-focused website [freeyork.org](#), **after three months of using structured data in their WordPress website** we saw the following improvements:

- +12.13% new users
- +18.47% increase in organic traffic
- +2.4 times increase in page views
- +13.75% of sessions duration

In other words, many still think of Semantic Technologies belonging to the future, when in reality quite a few players in the digital marketing space are taking advantage of them already.

Semantic SEO is a new and powerful way to make your content strategy more effective. In this article/guide I will explain from scratch what Semantic SEO is and why it's important.

Why Semantic SEO?

In a nutshell, search engines need context to understand a query properly and to fetch relevant results for it. Contexts are built using words, expressions, and other combinations of words and links as they appear in bodies of knowledge such as encyclopedias and large *corpora* of text.

Semantic SEO is a marketing technique that improves the traffic of a website by **providing meaningful data that can unambiguously answer a specific search intent**. It is also a way to create clusters of content that are semantically grouped into topics rather than keywords. In a [famous Google patent on context-vectors](#), an example with the word "horse" is provided.

The document looked at how the same word can have different meanings: a "horse" is an animal for an equestrian, a working tool for a carpenter, and sport equipment for a gymnast. In Semantic SEO, much like Wikipedia does, content is cataloged and organized around each context in such a way that machines can understand and value its uniqueness.

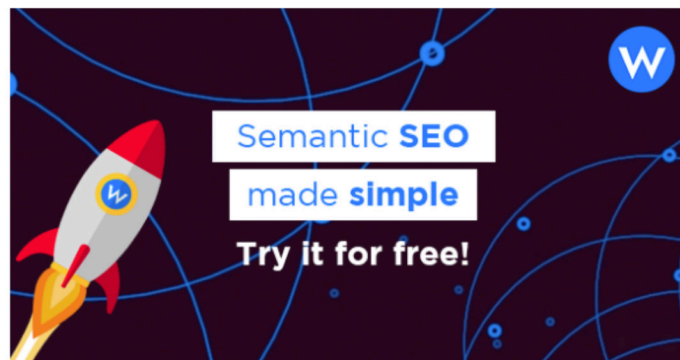


Table Of Contents

The Beginning.....	4
The Old Web.....	4
The New Web.....	4
The Power Of Natural Language Processing	4
What Is An Entity?.....	5
Metadata: Data About Data	5
Schema Markup: The Gold Standard Of The Semantic Web	6
What Is Structured Data?.....	6
What Is a Triple?.....	7
Why Is JSON-LD The Best Format For Structured Data?.....	7
Why Is Linked Data So Important	8
How Can I Link Entities With One Another	8
Why Is It Important To Publish 5-Stars Linked Data?	8
How Can you Link Entites From Your WP Site With The Linked Open Data Cloud	9
Putting It All Together With A Knowledge Graph.....	10
Summary And Conclusions	10
About The Authors.....	11
About Torque.....	12
About WP Engine.....	13

The Beginning

https://www.ted.com/talks/tim_berniers_lee_on_the_next_web

Before we dive into how to use Semantic SEO, we need to talk about how important it is, and where it originated.

On Feb. 2009, Sir Tim Berners-Lee, the founding father of the web was performing a TED talk in Long Beach, Calif. He spoke of the the formation of a new web, built over the past net. A Semantic Web based on open linked data.

Now the **Semantic Web** is here, and its technologies are available to digital marketers to make their SEO strategy more effective. How did we get there? Let's take a few steps back.

The Old Web

Today well over a billion websites comprise the web.



Tim Berners-Lee ✓
@timberners_lee

Follow

internetlivestats.com/watch/websites/
recently passed a billion websites by their count....

2:20 PM - 16 Sep 2014

In less than a decade the number of websites exploded. It comprised millions of pages. Berners-Lee figured out he could connect web pages with what we all know today as hypertext. However, surfing the web was still limited because you could just go from one page to the next through links: the effort it took to find what you were looking for was massive.

That is why many ventured out in finding a way to search through those pages to find specific content to queries.

This idea led to the creation of **PageRank**, which was the foundation of Google, an algorithm that could rank pages on the web based on the popularity of each page.

The more quality backlinks a website received the more it could rank higher in the **SERP**. Backlinks are still the backbone of the web. However on that backbone a new web blossomed.

The New Web

In 2012, futurist Ray Kurzweil arrived at Google, with one mission: make search engines understand human language. From that quest Google updated its algorithm in 2013, with Hummingbird and later on in 2015, AI became a major factor for search **RankBrain**.

It was a revolution. In fact, even though Google looks at more than 200 factors to assess the ranking of a page, it also uses Artificial Intelligence to rank those pages. In other words, Google looks more and more at the intention behind a user query **based on the context** rather than keywords.

For instance, if I type in the search box "french fries" I may be looking for something to eat or just the story behind the name. Of course, if I do this search at 8 a.m., I'm probably more interested in learning the story behind the food.

If I do the same search at 8 p.m., I may be looking for something to eat for dinner. But how does the search engine know what is the context? It reads human language, through **Natural Language Processing** (NLP). Before we dive into it, how does NLP work?

The Power Of Natural Language Processing

When I type in Google's search box "moon distance," that is what I get:

Moon / Distance to Earth

384,400 km



People also search for



Sun
149.6 milli...



Sirius
8.611 light
years



Europa
628.3 milli...

You may think this is simple keyword matching, but it is not.

In fact, if I ask "How far is the moon?"

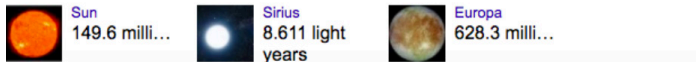
I get the same answer:

Moon / Distance to Earth

384,400 km



People also search for



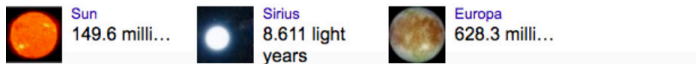
Google's ability to understand language goes further. If I search "moon distance in meters" that is what I get:

Moon / Distance to Earth

384,400 km



People also search for



In short, Google knows I'm referring to the same thing and gives me the proper answer.

But if Google isn't looking anymore at keywords (at least for specific queries) where is it getting the results? Google learns through topics. In the context of the Semantic Web those concepts are called **entities**. Those entities are organised in Giant Graphs. In fact, on May 16, 2012 Google announced the use of a massive Knowledge Graph.

In other words, it is a knowledge base to provide more useful and relevant results to searches using a semantic-search technique.

What Is An Entity?

According to Wikipedia:

An entity is something that exists as itself, as a subject or as an object, actually or potentially, concretely or abstractly, physically or not.

In the context of Semantic Web an entity is much more than that:

In the Semantic Web an entity is the "thing" described in a document. An entity helps computers understand everything you know about a person, an organization or a place mentioned in a document. All these facts are organized in statements known as triples that are expressed in the form of subject, predicate, and object.

I'M GOING TO RECONSTRUCT HOW A KNOWLEDGE GRAPH WORKS, BY STARTING FROM AN ENTITY. BUT REALLY, WHAT IS AN ENTITY?

Source: *What is an entity in the Semantic Web?*

Why Are Entities Way More Effective than Keywords?

For three simple reasons. Entities are:

- connected
- unambiguous
- contextual

Through entities you create meaningful relationships that can be read, understood, and interpreted by search engines. That is what Semantic SEO allows you to achieve. **Entities, in the context of the Semantic Web**, are really data points that computers can use to analyze and interpret the human language.

Let's take this a step at the time.

How do entities gain context?

Metadata: Data About Data

In its most basic definition, metadata is just data about data.

The concept of metadata is not new. In fact, librarians have been using it for a long time to discover, and manage documents. Imagine, that for each document you're specifying the author, date, book length, and so on. That is all metadata that helps classify a book. Therefore, it makes it easier to find it later on.

To work appropriately, metadata has to follow a logic of classification that everyone understands. In short, there must be a set of rules that everyone can follow to make the system

work. Like in a vocabulary where grammatical rules are arbitrarily selected to create a standard language. Ontologies are the foundation of metadata.

The simplest form of an ontology is a vocabulary. The vocabulary that today makes **Semantic SEO** possible is called Schema.org.



Schema Markup: The Gold Standard Of The Semantic Web

At the question "What is Schema.org?" that is what Google says:

Schema.org is a joint effort, in the spirit of sitemaps.org, to improve the web by creating a structured data markup **schema** supported by major search engines. On-page markup helps search engines understand the information on web pages and provide richer search results.

[FAQ - schema.org](#)
schema.org/docs/faq.html

What is Schema.org?

Schema.org is a joint effort by Google, Microsoft, Yahoo!, and Amazon to create a common vocabulary of terms to be used in web pages to help search engines understand the content of the pages. The goal is to make it easier for search engines to understand the content of the pages and provide richer search results.

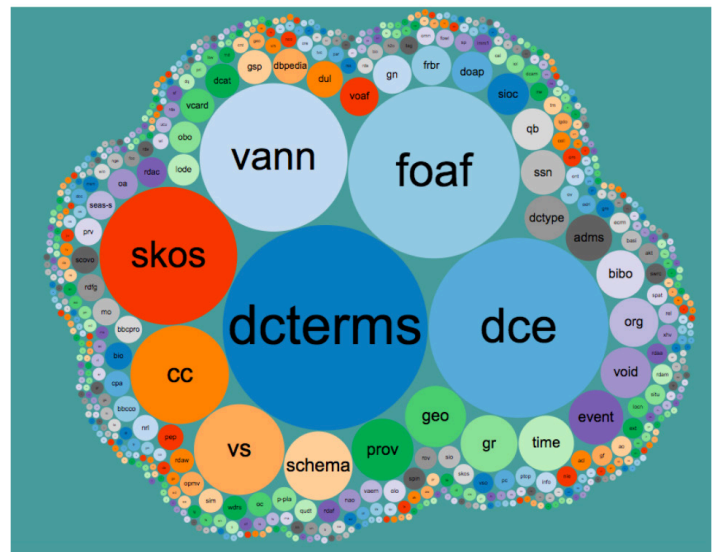
Schema.org is a project to create a common vocabulary of terms to be used in web pages to help search engines understand the content of the pages. The goal is to make it easier for search engines to understand the content of the pages and provide richer search results.

Schema.org is a project to create a common vocabulary of terms to be used in web pages to help search engines understand the content of the pages. The goal is to make it easier for search engines to understand the content of the pages and provide richer search results.

blog.kissmetrics.com

As we saw, the pillar of the Semantic Web is a linked open vocabulary shared across different websites, driven by an open community and usable alongside other open vocabularies and ontologies. Like in language, where the lack of standard grammatical rules makes it hard for the same language to exist. So the Semantic Web wouldn't be here without a gold standard. Schema.org is that gold standard.

In fact, out of all the competing standards that existed, [Schema.org](#) was the first linked open vocabulary that was introduced for a business-driven purpose (helping search engines organize the web and improve the quality of their results).



617 vocabularies in LOV

There are today 617 open vocabularies in the linked data world and they can be combined to organise and structure different knowledge domain.

In terms of SEO, Schema, being created by the search engine themselves is the most useful.

By adding schema markup to web pages, content is interlinked with data using standard linked vocabularies like schema.org and becomes more accessible.

What Is Structured Data?

Structured data is a standardized format for providing information about a page and classifying that content on the page; for example, on a recipe page, what are the ingredients, the cooking time, the temperature, the calories, and so on.

Source: [DevelopersGoogle.com](https://developers.google.com/ai)

Structured data that uses Schema.org as a reference vocabulary and can be embedded in web pages using three formats:

- JSON-LD
- Microdata
- RDEa

Format	Description and Placement
JSON-LD* (Recommended)	JavaScript notation embedded in a <script> tag in the page head or body. The markup is not interleaved with the user-visible text, which makes nested data items easier to express, such as the Country of a PostalAddress of a MusicVenue of an Event. Also, Google can read JSON-LD data when it is dynamically injected into the page's contents, such as by JavaScript code or embedded widgets in your content management system.
Microdata	An open-community HTML specification used to nest structured data within HTML content. Like RDFa, it uses HTML tag attributes to name the properties you want to expose as structured data. It is typically used in the page body, but can be used in the head.
RDFa	An HTML5 extension that supports linked data by introducing HTML tag attributes that correspond to the user-visible content that you want to describe for search engines. RDFa is commonly used in both the head and body sections of the HTML page.

Source: [DevelopersGoogle.com](https://developers.google.com/structured-data/rdfa/)

Imagine a book supported in three different formats: ebook, paperback, and hardcover. Each has different weights, sizes and so on. So does [Schema.org](https://schema.org).

JSON-LD is the preferred format by Google. In fact, that is a JavaScript embedded in a <script> tag in the page or head or body. The code encapsulates useful and contextual information regarding the article written using linked data standards.

Information, in the semantic web, is written using *subject-predicate-object* predicates called triples.

What Is A Triple?

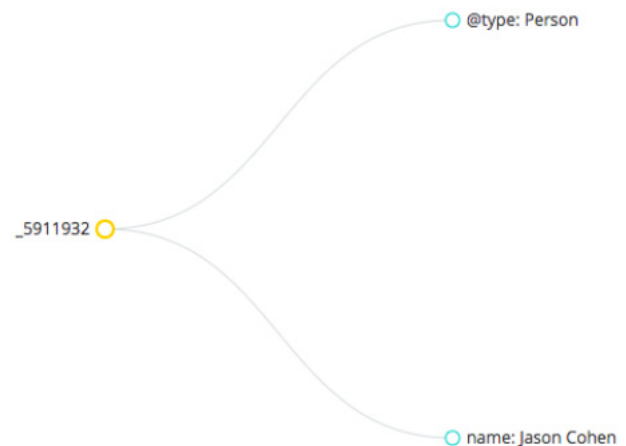
Like an atom in the material world, a triple is the most fundamental part of how information is encoded in a knowledge graph using semantic web technologies. A triple comprises a *subject-predicate-object* expression (Jason is 40 or Jason knows David).

Each entity gets connected to the next through relationships that are expressed with simple triples. Like in reality where from simple local rules, arise complexity. By mixing up billions of triples, you get the Semantic Web.

For instance, those below are two triples expressed in JSON-LD:

```
{
  "@type": "Person",
  "name": "Jason Cohen"
}
```

"This entity is a person" and "The name of this entity is Jason Cohen." Which, visually (using an online tool called the [JSON-LD Playground](https://playground.googleapis.com/jsonld)) will look like the following:



Within each triple we can add more information in the form of subject-predicate-object and using a particular Schema type and its associated properties, so that we can make each piece of information truly accessible to machines.

Why Is JSON-LD The Best Format For Structured Data?

The advantages of using JSON-LD are the following:

- It does not affect the performance of the page because it can be loaded asynchronously
- It is injected in the page without impacting on the existing HTML structure and tags
- It is easy to be reused by web developers as it follows the JSON syntax
- It leverages on linked data

That is also why Google incentives websites owners to use structured data on a web page in the form of **JSON-LD**. This format allows entities, therefore concepts to become linked.

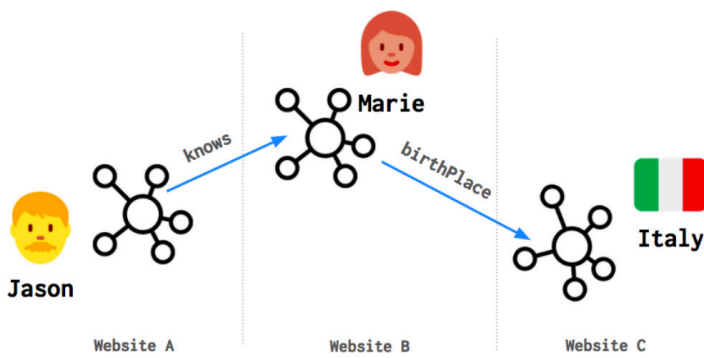
However, that information can still be ambiguous. For instance, in the example above Jason Cohen can be either *the founder of WP Engine* or *the American filmmaker*. I'm referring to the former obviously but how do I make a search engine understand that?

Once again, linked data (a semantic web standard for publishing data) is the answer.

Why Is Linked Data So Important?

Schema is the vocabulary used to help search engine understand the content of webpages. Linked data is a method for publishing structured data using vocabularies like Schema that can be connected together and interpreted by machines.

Using linked data, statements encoded in triples can be spread across different websites. On Website A we can present the entity *Jason and the fact that he knows Marie*. On website B we can provide all the information about Marie and on the Website C we can find information about Marie's birthplace.



Each page contains the structured data to describe an entity and the *link* to the entity that could be described on a different website.

Using on WordPress, semantic plugins like WordLift, you can link one entity with another much like you do with web pages. But how can you make sure that search engine understands unambiguously that you are referring to the CEO of WP Engine and not to the movie director?

There are really two ways:

- **Implicitly**, the search engine could grasp that on Torque we are more likely to refer to the CEO of WP Engine (and this is what *context-vectors* are used for)
- **Explicitly**:
 - by reading a linked data attribute called @id (a unique identifier for the entity that might be already known by the search engine) or
 - by reading any sameAs link for the entity that points to a known dataset. The purpose of these links is to tell machine that the entity is equivalent of another entity in an open encyclopedia like Wikipedia or directly into the Google Knowledge Graph

How Can I Link Entities With One Another?

According to [Schema.org](https://schema.org) the sameAs property is:

URL of a reference Web page that unambiguously indicates the item's identity. E.g., the URL of the item's Wikipedia page, Wikidata entry, or official website.

It is like you're saying to the search engine "this is the same thing as the one you find at this address." Today only between 10,000 to 50,000 domains use this property. That is also why you can make a difference for your SEO strategy by using it.

However the sameAs property alone might not be enough if you need to query the data that you're publishing (or simply if you want others to query the data that you're publishing across multiple datasets).

You need something more. You need to publish data following the so-called **five-stars open data scheme** introduced by Berners-Lee that requires you to link every piece of data with other data.

Here is where the *owl:sameAs* property comes into play.

Why Is It Important To Publish 5-Stars Linked Data?

There are four simple principles to follow when publishing data on the web (and yes, structured data is open data as it is fully accessible).

1. Use URIs as names for things (this is the unique identifier that we introduced above)
2. Use HTTP URIs so that people can look up those names (this means that the ID of every entity shall be accessible via HTTP URI)
3. When someone looks up a URI, provide useful information, using the standards (Behind these URIs we need to publish data using a linked data standard called RDF)
4. Include links to other URIs. so that they can discover more things (and here is where we need to add *owl:sameAs* property).

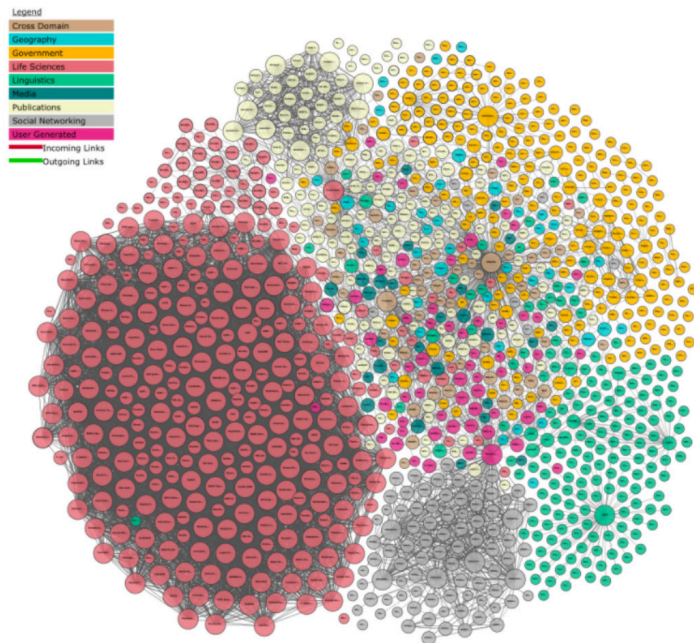
To simplify a 5-stars open dataset is simply a way to publish metadata (using **linked data** standard) that makes it readable and accessible by machines.

That is why by connecting a piece of text, written by a human with an open dataset created by a machine, our content becomes fully accessible and machine-friendly.

Once again Semantic SEO is about helping machine understand our content by using open web standard to describe it.

Some of the primary datasets that implement the 5-stars linked data schema are foundational for the machine learning algorithms behind semantic search engines like Google and Bing as well as digital personal assistants like Alexa, Cortana, and the Google Assistant.

These datasets (like DBpedia, Wikidata, Geonames just to name it a few) are all interlinked together to form a Linked Open Data Cloud.



Source: lod-cloud.net

When you add structured data to your WordPress website using a plugin like **WordLift**, that structured data gets published as open linked data. In short, your WordPress website, and the metadata of your own content becomes part of that Linked Open Data Cloud. Thus by adding an additional layer to the Semantic Web your WordPress website also becomes part of it!

How Can You Link Entities From Your WordPress Site To The Linked Open Data Cloud?

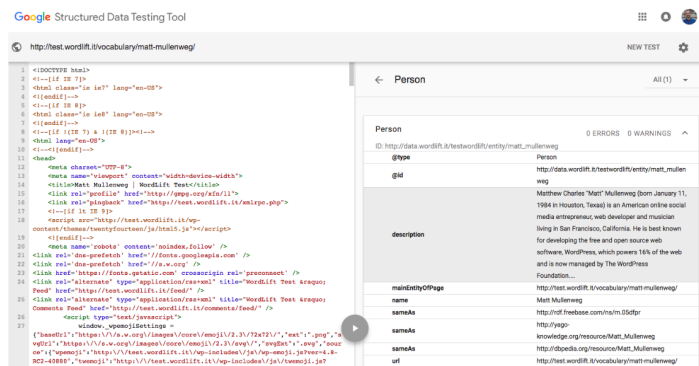
Imagine we want to explain to a search engine Matt Mullenweg is and link the page I have for him on my blog with entities in the LOD cloud. How do I do that on my WordPress website?



As you can see above, I used WordLift within my WordPress to create a page about Matt Mullenweg. That page is set up as a Schema Entity Type "person". To make it clear who I am talking about I run a search using **WordLift** that taps into giant graphs published in LOD and in a snap I can get the reference to the entity of Matt Mullenweg on Freebase, Wikidata, and DBpedia.

Once I update the page, the Schema sameAs, and the owl:sameAs properties are automatically added by WordLift and made available to search engines.

We can now use the **Structured Data Testing Tool** of Google to see how the search engine sees the page:



The structured data has been created without writing a single line of code and now Google can crawl and index that page way more efficiently. However, there is also more to it.

As we saw earlier by using 5-stars linked data now my data is interoperable with other datasets and published also off-page in an RDF-based knowledge graph.

Putting It All Together With A Knowledge Graph

In the context of Semantic Web, a knowledge graph is a way of representing knowledge. In short, you start from a few triples and those triples are put in relationship to build a graph. For instance, let's have a closer look - using Semantic Web technologies - at the Matt Mullenweg entities on my blog:



Source: [LodLive](#)

As you can see we have a set of triples that tell us a story: Matt Mullenweg, also named Matthew Charlie, was born in Houston and attended the University of Houston. He is an entrepreneur.

A knowledge graph doesn't speak any particular language. Language is human; a knowledge graph gets expressed in open linked data, which is the language of machines.

Imagine your entire website built upon a large **knowledge graph** made of all the metadata that describes the thing that you write about. That knowledge graph becomes part of a larger graph that comprises the new web. That is the power of Semantic Web.

Summary And Conclusions

Throughout this article, we saw that the Semantic Web is here. A few players already are using semantic technologies in the digital marketing space. They are getting tremendous results from SEO and findability standpoint. That is the birth of a new field, within the SEO industry, commonly referred as Semantic SEO.

Although, backlinks are still the foundation of the web. Keywords are losing relevance, and they will lose it even further over time. Semantic SEO allows you to structure your content around triples and topics and to gain control over context-vectors that search engine use to rank your content against a particular search intent.

Data is encoded into properties defined within a set of linked vocabulary like Schema.org.

When you add linked data to your WordPress website using a plugin like **WordLift**, you add structured data that can be quickly read and understood by search engines. From that linked data, machines can extrapolate relationships and context, in a word: topics and meanings.

When that structured data is published in a knowledge graph using semantic web standards, your WordPress website "joins" the Linked Open Cloud. Being part of the Semantic Web means making data accessible to machines and allow them to compute it for providing better answer and for improving the quality of the traffic that they will bring you.

It also means that you're back in control of your content: you're now publishing content but also the metadata that is required to market and to monetize this content. Your SEO strategy is now Semantic and all you need to do is just a simple plugin that you will find featured in the WP Engine solution center.

About The Authors



Gennaro Cuofano

Part of the *growth team* at **WordLift**. Gennaro helps the company accelerate its growth while focusing on business development and innovative content strategies. He is also creator of the **FourWeekMBA.com**.



Andrea Volpini

CEO of **WordLift** with 20+ years of experience in content management, SEO and web publishing. In 2017, after many years of research, Andrea decided to follow his passion for AI and co-founded WordLift, a first-mover in the emerging Semantic SEO industry.



Maria Silvia Sanna

Digital Communication specialist at **WordLift**, Maria Silvia is the heart and soul of the content marketing for the *blue* plugin. A passionate blogger (**cupofbrain.it**) and esteemed marketer she is in deep love with movies, great content and Italian food.



About Torque

Torque is a news site featuring all things WordPress. We are dedicated to informing new and advanced WordPress professionals, users, and enthusiasts about the industry. Torque focuses primarily on WordPress News, Business, and Development, but also covers topics relating to open source and breakthrough technology. Torque made its debut in July 2013, at WordCamp San Francisco, and has since produced valuable content that reflects the evolution of WordPress, both as a platform and a community. Torque is a WP Engine publication, though maintains complete editorial independence. torquemag.io



About WP Engine

WP Engine powers amazing digital experiences for websites and applications built on WordPress. The company's premium managed hosting platform provides the performance, reliability and security required by the biggest brands in the world, while remaining affordable and intuitive enough for smaller businesses and individuals. Companies of all sizes rely on WP Engine's award-winning customer service team to quickly solve technical problems and create a world-class customer experience. Founded in 2010, WP Engine is headquartered in Austin, Texas and has offices in Limerick, Ireland, San Francisco, California, San Antonio, Texas, and London, England. wpengine.com

TORQUE®

