

# OPEN KNOWLEDGE FOR THE SOCIAL SECTOR

---

## Recommended Metadata for Knowledge Products

The following is a suggested core set of metadata points for collecting and indexing knowledge products. This core set serves as an excellent starting point for sharing data about knowledge products with external partners and interoperable platforms.

Simply put, metadata is data that describes other data. The goal of metadata is to enable end users to verify that the knowledge product they seek and the knowledge product they've discovered are one in the same. To foster discoverability and usability, every document should have at least the following four core metadata points indexed:

- Title
- Date Published (year, month, day)
- Publishing Organization(s)
- URL of downloadable file or filename if the file is to be shared directly.

The following metadata points round out the descriptive set that could be applied to a knowledge product. If you have any of these details (the more the better!), including them in the metadata for a knowledge product makes discoverability more likely:

- Author(s)
- Funder(s) - of the work itself
- Abstract/Description
- Keywords
- License/Copyright Notice - enables clear understanding of how a work may be re-used
- Language
- Document Type
- Geographic Coverage - location(s) the work focuses on
- Digital Object Identifier (DOI)
- International Standard Book Number (ISBN)

If you have additional thematic codes applied to your knowledge products consider including them as they can be useful, particularly when helping end users to filter and sort search results. They can also be useful in the development of next generation search and retrieve tools and processes.

# OPEN KNOWLEDGE FOR THE SOCIAL SECTOR

---

## Schema.org for Knowledge Products

Improving discoverability

Schema.org is a data vocabulary and microdata format that can improve the discoverability of resources that are shared online. It was created and is maintained by Google, Yahoo!, and Bing.

Essentially, Schema.org lets you turn a web page into a database. All it takes is embedding bits of special HTML code, or microdata, into the HTML that displays your web content. Once augmented with Schema.org, the information available in a web page—information that we humans can understand at a glance—becomes data that a machine can parse. Search engines and other platforms, such as IssueLab, can then use the data to do things like share content more broadly and build better search tools.

Where the knowledge produced by the social sector is concerned, Schema.org's data vocabulary lets you describe everything from the title of a report, to the name of the report's author, to the web address where the report can be downloaded. To illustrate how straightforward using Schema.org can be, let's mark-up our core set of metadata points -- title, date published, publisher(s), and web address of the resource. In this example, we'll use a recent report published by the Foundation Quarterly:

### Grey Matter(s): Embracing the Publisher Within

Date Published: 2015-06-30

Publisher(s): The Foundation Review

Download here: <http://doi.org/67j>

Here's the HTML behind this display, with Schema.org (in red) embedded in the code:

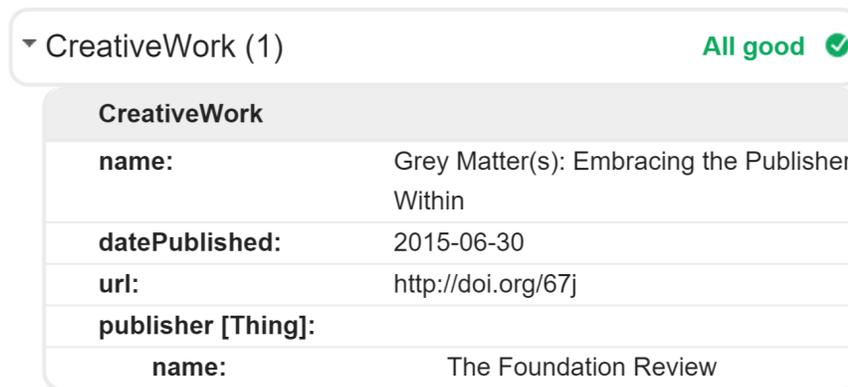
```
<div itemscope itemtype="http://schema.org/CreativeWork">  
  <h1 itemprop="name">Grey Matter(s): Embracing the Publisher Within</h1>  
  Date published: <span itemprop="datePublished">2015-06-30</span><br />  
  Publisher(s):  
  <span itemscope="organization">  
    <span itemprop="publisher">  
      <span itemprop="name">The Foundation Review</span>  
    </span>  
  </span><br />  
  Download here:  
  <a itemprop="URL" href="http://doi.org/67j">http://doi.org/67j</a>  
</div>
```

# OPEN KNOWLEDGE FOR THE SOCIAL SECTOR

---

In this code view, we can see that seven Schema.org microdata tags are embedded into our HTML. The microdata let's us describe the type of content being displayed (a "CreativeWork"), the name of the CreativeWork (itemprop="name"), the date it was published (itemprop="datePublished"), the publisher (itemtype="organization" / itemprop="publisher"), and its web address (itemprop="URL"). So what does this look like to a machine?

Google's Structured Data Testing Tool output shows below\*. This is what the machine "sees":



The screenshot shows the output of Google's Structured Data Testing Tool. At the top, it indicates 'CreativeWork (1)' with a status of 'All good' and a green checkmark. Below this, a table displays the structured data for the CreativeWork type.

CreativeWork	
<b>name:</b>	Grey Matter(s): Embracing the Publisher Within
<b>datePublished:</b>	2015-06-30
<b>url:</b>	http://doi.org/67j
<b>publisher [Thing]:</b>	
<b>name:</b>	The Foundation Review

Much more than this can be described with Schema.org, including pretty much every metadata point that can be applied to knowledge products. Have a look at the Schema.org website to get the full gist of this powerful vocabulary and microdata format: <http://www.schema.org>.

If you have questions about how to implement Schema.org in your own web pages, we can help! Just get in touch.

\* Access Google's Structured Data Testing Tool at <https://developers.google.com/structured-data/testing-tool>.