

## **Structuring information on the Web for re-usability: Consultations and Job Vacancies**

Version:	1.4
Date created:	15/07/2010
Policy official:	Adam Bailin
Date last updated:	19/07/2010
Date issued:	July 2010
Lead official:	David Pullinger
Guidance number:	TG124

### **Purpose**

Government is committed to making its public information and data as widely available as possible. The best way to make structured information available online is to publish it as Linked Data. Linked Data makes the information easier to cut and combine in ways that are relevant to citizens. This document describes how to put government consultations information into Linked Data using RDFa.

## Table of contents

<b>Introduction .....</b>	<b>2</b>
Purpose .....	2
What are the benefits for users? .....	2
Background .....	2
Audience .....	4
<b>Introduction to linked data .....</b>	<b>5</b>
How does Linked Data work? .....	5
What Linked Data standard should I use? .....	6
Why not just use HTML + RDF/XML? .....	6
<b>Putting consultations into Linked Data .....</b>	<b>7</b>
a. One consultation description per page.....	7
b. Exposed to web search engines .....	8
c. Common structure .....	8
d. How to mark-up consultation descriptions using RDFa.....	8
<b>Putting job vacancies into Linked Data .....</b>	<b>15</b>

## Introduction

### Purpose

Government is committed to making its public information and data as widely available as possible. The best way to make structured information available online is to publish it as Linked Data. Linked Data makes the information easier to cut and combine in ways that are relevant to citizens. This document describes how to put government consultations and job vacancy information into Linked Data using a format that allows machines to access and re-use the information: RDFa.

### What are the benefits for users?

Consistently structured online information means:

- **End users** know what to expect, find content easier to read and experience reduced time/effort to complete tasks.
- **Centralised services** aggregate accurate, up to date and comprehensive information.
- **Data re-users** create new cuts and combinations of data that are interesting and relevant to end users.

### Background

Public data should be made available to the public. For example, data on crime, health and education should be published online so that is easily accessible to the public. This may require publishing the data in a particular way so that it can be more easily combined by third parties in ways that the Government may not have thought of.

Some progress has been made towards this goal, for example the [Public Sector Information Unlocking Service](#)<sup>1</sup> allows people to request that specific government data stores are made available.

The [Show Us a Better Way](#)<sup>2</sup> competition asked people for ideas about how to combine public data in useful ways.

Government must always be mindful of data privacy, protection of personal data and secure data handling. However, *public* information and data should be published routinely, not just made available on request via the Freedom of Information Act.

---

<sup>1</sup> <http://www.opsi.gov.uk/unlocking-service/>

<sup>2</sup> <http://www.showusabetterway.com/>

## Making consultations available in one place

One of the key drivers for this work is the [consultations code of practice](#)<sup>3</sup> which requires public sector organisations to make its consultations easily accessible. This guidance supports the ongoing work to make central government consultations automatically available in one place. The ability to browse and find consultations will appear on Directgov.

In addition, the [Power of Information \(POI\) Taskforce Report](#)<sup>4</sup> recommends 'clear and mandatory standards on accurate tagging and metadata which would allow consultations to be found by the subjects, interests and places they affect as well as by the policy issue.'

A [vocabulary for marking up consultations](#)<sup>5</sup> and an example implementation is available online. This guidance provides supporting documentation to help developers integrate the appropriate code into their web pages.

## Making job vacancy descriptions available online in a re-usable form

Another key driver comes from Ministerial announcements for all public sector jobs to be advertised online and to do so in a re-usable form.

A vocabulary for marking up jobs and an example [implementation](#)<sup>6</sup> is available online. This guidance provides supporting documentation to help developers integrate the appropriate code into their web pages.

## Looking ahead

The Public Sector Transparency Board is leading the work to release key government datasets. The role of the Transparency Board, the definition of public data and the [draft principles by which public data should be released](#)<sup>7</sup> are published on the data.gov.uk website.

Tim Berners-Lee, a member of the Transparency Board, sets out the case for [putting government data online](#)<sup>8</sup> using Linked Data techniques. A key message is to put data online *where it is* and in the most convenient format. Linked Data is the preferred format but may come later as needs dictate.

The following types of structured information on the Web should be put into Linked Data using RDFa:

- Public sector jobs
- Central Government consultation descriptions

---

<sup>3</sup> <http://www.berr.gov.uk/whatwedo/bre/consultation-guidance/page44448.html>

<sup>4</sup> <http://poit.cabinetoffice.gov.uk/poit/>

<sup>5</sup> <http://code.google.com/p/argot-hub/wiki/ArgotConsultation>

<sup>6</sup> <http://code.google.com/p/argot-hub/wiki/GettingStartedEmbeddedHtml>

<sup>7</sup> <http://data.gov.uk/blog/new-public-sector-transparency-board-and-public-data-transparency-principles>

<sup>8</sup> <http://www.w3.org/DesignIssues/GovData.html>

This list may grow in the future.

## **Audience**

This document provides an introduction to Linked Data aimed at Central Government Heads of e-Communications.

It also provides help for web developers who need to add RDFa data to consultations and job vacancy web pages.

## Introduction to linked data

1. The Web, in its most basic sense, is a collection of pages linked together via hyperlinks. These documents can contain any number of paragraphs, headings, lists, images and so on. These page elements - defined in an HTML document - determine what the page looks like. However, they don't reveal anything about the *meaning* of the page.
2. It is only when you read a paragraph or view an image that you begin to understand what the page is about. This may sound trivial and in many cases it is a perfectly adequate way to interact with the Web. However, when a page contains structured information such as a job vacancy or a government consultation, there is a way to describe what the various elements mean or, to put it another way, to describe the *semantics*.
3. The Semantic Web is an evolution of the World Wide Web which links *data* together, not just *pages*. Adding semantic meaning to the information on the Web allows it to be linked, searched and combined in interesting ways perhaps more relevant to the user, leading to an enhanced user experience.
4. Users want to be able to search for information by subject, interest or geographical region and don't necessarily know which government organisation is responsible. Publishing information as *Linked Data* means that it can be combined (or mashed up) with other government data without advance planning. For example health information may be combined with information about schools based on geographical area. Using Linked Data standards means the information can be combined more easily and efficiently.
5. Putting government data online as Linked Data makes it easier for citizens to find relevant information which promotes public engagement and increases accountability.

### How does Linked Data work?

6. Publishing unstructured information on the Web is fairly straightforward, for example posting a blog article containing text and images. Publishing structured information on the Web, for example a job vacancy or a consultation, requires additional thought.
7. Linked Data adds meaning to structured information on the Web. For example, a job vacancy becomes more than just a page of text. Meaning is added through descriptions of the job based on a published vocabulary – or set of agreed terms – used to describe the job title, salary, start date and so on.
8. The terms in Linked Data vocabularies are publicly available and referenced via URIs – similar to URLs. These are used to uniquely identify a particular term and ascribe an agreed meaning to it. For example the Linked Data term

for describing a job is specified in the [ArgotJob](http://code.google.com/p/argot-hub/wiki/ArgotJob)<sup>9</sup> vocabulary and referenced as <http://purl.oclc.org/argot/Job>.

9. Linked Data vocabularies mix terms from other common vocabularies wherever possible. This enables the data to be combined with other Linked Data at a later date. For example, the common term for describing the title of a resource is specified in the [Dublin Core](http://dublincore.org/)<sup>10</sup> vocabulary, a standard developed by information experts, librarians and publishers to describe publications, and referenced as <http://purl.org/dc/terms/title>. Hence ArgotJob uses the Dublin Core term for job title rather than inventing its own term.
10. All Civil Service jobs are published as Linked Data. [Find a job](http://www.civilservice.gov.uk/jobs/)<sup>11</sup> and then view the page source: you will see references to the Linked Data terms described earlier: *Job* and *title* (search for <div id="jobAdvert">). See Mark Birbeck's [blog article](http://webbackplane.com/mark-birbeck/blog/2009/04/23/more-rdfa-goodness-from-uk-government-web-sites)<sup>12</sup> for details.

### What Linked Data standard should I use?

11. For information already published as web pages, the best Linked Data standard to use is [RDFa](http://www.w3.org/TR/rdfa-syntax/)<sup>13</sup>. Adding RDFa to existing web pages involves adding code to the HTML code that sits behind it. This is often easier to do than generating additional web documents.
12. RDFa is a good choice where you have similar types of information in lots of different places. Adding extra HTML markup to the page allows information from a variety of sources to be extracted and pulled together in one place, making it easier for the user – who may not know where the information is – to find information relevant to them.
13. The best way to add RDFa to a web page is to make changes to the scripts which generate a page so that the data is embedded automatically when the page is published (e.g. via a content management system).

### Why not just use HTML + RDF/XML?

14. RDFa is easy to implement, particularly where you already have the information published in web pages. Making changes to existing pages rather than creating new ones is a pragmatic route to Linked Data.

---

<sup>9</sup> <http://code.google.com/p/argot-hub/wiki/ArgotJob>

<sup>10</sup> <http://dublincore.org/>

<sup>11</sup> <http://www.civilservice.gov.uk/jobs/>

<sup>12</sup> <http://webbackplane.com/mark-birbeck/blog/2009/04/23/more-rdfa-goodness-from-uk-government-web-sites>

<sup>13</sup> <http://www.w3.org/TR/rdfa-syntax/>

## Putting consultations into Linked Data

15. Adding RDFa data to consultation descriptions means the information will be picked up automatically by Directgov so that there is a single point of access for all central government consultations, making it easier for citizens to find out about and engage with Government. This will be at:  
[direct.gov.uk/consultations](http://direct.gov.uk/consultations)
16. All central government consultation descriptions must be put into Linked Data using RDFa, starting from 1 January 2010.
17. Ideally, all existing web published consultations (past and present) should be searchable through the single point of access. As a minimum, all consultations open on 1 December 2009 onwards should be marked up with RDFa.
18. The following four measures are required for compliance:
  - a. As a rule, you must publish one consultation description per page.
  - b. Ensure that all consultation descriptions are exposed to web search engines, for example via an XML Sitemap.
  - c. Ensure that all consultation descriptions are structured in the standard way, according to [ArgotConsultation](#).
  - d. Ensure that all consultation descriptions are marked up with RDFa using the [ArgotConsultation](#)<sup>14</sup> vocabulary. Please note the [minimum requirement for identification](#) as a consultation.
19. You may not need to change the appearance of consultations. The emphasis is on adding HTML markup to augment what is already there.
  - a. One consultation description per page**
20. Each consultation should have a separate webpage including a short description of what the consultation is about and who should be interested in it. The short description may need input from the policy lead and should be three to five sentences in length.

---

<sup>14</sup> <http://code.google.com/p/argot-hub/wiki/ArgotConsultation>

## b. Exposed to web search engines

21. Ensure that all consultations are searchable by Web search engines by adding into your XML Sitemap and if necessary permitting the Directgov search spider to index. Guidance on [exposing your website to search engines](#)<sup>15</sup> is available on the COI website.

## c. Common structure

22. The common structure required is in order that consultations from across central government are quickly understood by end users and have the same information when aggregated. The exact details are explained below.

## d. How to mark-up consultation descriptions using RDFa

23. The following tutorial will take you through the steps required to publish basic index information about consultations in RDFa. Since RDFa allows us to add this information into a page by reusing information that already exists, we'll show how a typical web page can be augmented.
24. The following examples are taken from the Ministry of Justice's [Future of the Parole Board consultation](#). They are illustrative only and should not be used as compliance measures. Compliance is determined by the four measures listed previously.

## DOCTYPE and Validation

25. Use the XHTML+RDFa DOCTYPE at the top of your document:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
"http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
```

26. You will need to validate your documents to ensure they are picked up by Directgov. Use the [W3C Validator](#) to check your documents.
27. The [RDFa Distiller](#) is also a useful tool for developing and checking RDFa markup.

## Prefix mappings

28. RDFa makes extensive use of URIs which can get a little unwieldy. RDFa supports a mechanism for shortening URIs (called Compact URIs, or CURIEs), which involves using a prefix to replace part of the URI. We usually declare these prefixes at the top of the document, on the html element:

---

<sup>15</sup> <http://coi.gov.uk/guidance.php?page=246>

```

<html
  xmlns="http://www.w3.org/1999/xhtml"
  xmlns:argot="http://purl.oclc.org/argot/"
  xmlns:dc="http://purl.org/dc/terms/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:v="http://www.w3.org/2006/vcard/ns#"
  xmlns:dbp="http://dbpedia.org/resource/"
>
  ...
</html>

```

29. However, as long as these prefixes appear above, or on, the elements that make use of them, then they can appear anywhere in the document. This means that if you don't have full control over the page that is publishing the consultation, but you do have control over a step in the workflow where data is output, you can place the prefixes onto some suitable `div`, that will in turn contain the consultation:

```

<div
  xmlns:argot="http://purl.oclc.org/argot/"
  xmlns:dc="http://purl.org/dc/terms/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:v="http://www.w3.org/2006/vcard/ns#"
  xmlns:dbp="http://dbpedia.org/resource/"
>
  ...
</div>

```

## How to get a consultation identified by Directgov

30. Two minimum properties are required for a document to be identified by Directgov as a consultation and therefore made available for searching. The first property indicates that a consultation is a type of document, by using the `Document` class from the Friend of a Friend (FOAF) vocabulary as follows:

```

<div about="#this" typeof="foaf:Document">
  ...
</div>

```

The FOAF vocabulary provides some basic terms relating to people, organisations, projects, and so on, and we'll be using a few more of these terms to mark-up our consultation.

31. The second property indicates not only that a consultation is a type of document, but also its document type, via the Dublin Core `type` property:

```

<div about="#this" typeof="foaf:Document">
  <span rel="dc:type" resource="[argot:Consultation]"></span>
  ...
</div>

```

The Dublin Core vocabulary provides us with a number of terms relating to documents. We'll see more of these below.

These two properties allow the consultation to be picked up by Directgov. You'll also need to mark up information about the consultation so people can search and filter the results. These mark ups are explained below.

Directgov displays the publisher, title, abstract, open and closing dates and reference for each consultation. These are the essential mark ups you should apply. However it is recommended that all properties are marked up in case the remit of the tool widens.

## Publisher

32. It's important to know where a consultation originates from, and we indicate the publisher of the consultation using the Dublin Core `publisher` property. Many government sites will already have Dublin Core information in the head of the document, like this:

```
<meta name="DC.publisher" content="Ministry of Justice" />
```

This is the same information that we need for RDFa, expressed in a slightly different way. However, RDFa gives us a number of different ways that we can add this information.

33. If the information is to also be human readable, then it can be placed into a `span` or other element, and the appropriate RDFa property added:

```
Published by <span property="dc:publisher">Ministry of  
Justice</span>
```

34. However, if we don't want the information to be visible anywhere, then it can be hidden by using the RDFa `@content` attribute. In this case the value can be placed on its own element:

```
<div about="#this" typeof="foaf:Document">  
  <span rel="dc:type" resource="[argot:Consultation]"></span>  
  <span property="dc:publisher" content="Ministry of  
Justice"></span>  
  ...  
</div>
```

or even merged onto the first element:

```
<div about="#this" typeof="foaf:Document"  
  property="dc:publisher" content="Ministry of Justice">
```

```
<span rel="dc:type" resource="[argot:Consultation]"></span>
...
</div>
```

35. Whichever method is chosen, the publisher should be your department or organisation name in full. Avoid using abbreviations or alternative names (e.g MOJ or Justice).
36. The publisher attribute will be displayed in the Directgov search results, therefore it is not appropriate to include full address details.

### The source of the consultation

37. To indicate the source of the consultation (i.e. where it can be located) we use the Dublin Core property `source`:

```
<span rel="dc:source" resource=""></span>
```

By leaving the `@resource` value blank, we're saying that 'this document' is the source of the information.

### Language

38. The language that the consultation is in can be indicated with the Dublin Core's language property:

```
<span property="dc:language" content="en"></span>
```

39. Note that there is nothing to say that the span can't contain human-readable information, and indeed it is good practice:

```
This consultation will be conducted in <span property="dc:language"
content="en">English</span>
```

### Title

40. The consultation's title is indicated using the Dublin Core's `title` property. Since this will usually be available on the web-page anyway, it's a good idea to simply add the RDFa property to the existing title:

```
<h1 property="dc:title">Future of the Parole Board</h1>
```

### Reference number

41. Consultations will invariably have a reference number, that is unique within the publishing organisation. This can be indicated using the Dublin Core `identifier` property, and again, we would probably reuse existing mark-up:

Reference Number: `<span property="dc:identifier">CP14/09</span>`

## Dates

42. The consultation will have a date published (or 'issued'), a start date for receiving comments (or 'available') and a closing date (or a date through which the consultation is 'valid'). These values can be set using the Dublin Core `issued`, `available` and `valid` properties. (Note the `available` and `issued` properties will often have the same value, but you will need to include both mark ups or the Directgov tool will not return your dates properly.) The human-readable version of a date is not always appropriate for processing, so `@content` can be used to specify a precise value, which must be of the form YYYY-MM-DD:

Issued: `<span property="dc:issued" content="2009-07-20">20 July 2009</span>`

Open date: `<span property="dc:available" content="2009-07-20">20 July 2009</span>`

Close date: `<span property="dc:valid" content="2009-11-20">20 November 2009</span>`

## Abstract

43. The Dublin Core abstract property is used to provide a summary of the consultation:

```
<div property="dc:abstract">
  <p>Since its creation in 1968, the role of the Parole Board has
  changed significantly, from an advisory body to a court-like
  decision-making body.&nbsp; It has evolved in light of legislative
  changes, court judgments and changing caseloads, but its functions,
  status and resources have not been systematically considered in
  light of these changes.</p>
  <p>This consultation paper represents an opportunity to address
  this, and consider whether any changes to its sponsorship and role
  would support the Parole Board in carrying out its functions
  more effectively and efficiently.</p>
</div>
```

44. You should write approximately 3-5 sentences in plain English, avoiding the use of specialist language.
45. Note that any formatting will be carried through to the Directgov search results. For example, any `<b>bold</b>` text will appear with that formatting applied within Directgov's search results which may or may not be the intention.

## Audience

46. The Dublin Core `audience` property is used to give some indication of the kinds of people the consultation is aimed at:

```
<div property="dc:audience">
  The consultation is aimed at other criminal justice departments
  and agencies, legal practitioners and their representative
  organisations, prisoners and their representatives and bodies
  representing prisoners and victims in England and Wales.
</div>
```

## Related documents

47. There will be some documents that form part of the consultation itself, and these can be referred to using the Dublin Core's `hasPart` property:

```
<a rel="dc:hasPart" href="/consultations/docs/future-parole-board-
consultation.pdf">Future of the Parole Board - a consultation
paper</a>
<a rel="dc:hasPart" href="/consultations/docs/parole-board-
consultation-questionnaire.doc">Future of the Parole Board - list of
questions for response</a>
```

48. Other documents may be related to this consultation, but may not actually be a formal part of it. For these documents the Dublin Core `relation` property can be used. For example, to link to the Hansard entry where this consultation was announced by the Ministry of Justice, we might do this:

```
<a rel="dc:relation"
href="http://www.publications.parliament.uk/pa/ld200809/ldhansrd/tex
t/90720-wms0003.htm#wms_st_22">Announcement by Lord Bach, the The
Parliamentary Under-Secretary of State, Ministry of Justice.</a>
```

49. Another category of related documents is documents referred to; these can be indicated with the Dublin Core `references` property.

## How to respond

50. A key part of a consultation is the public's response, so the consultation also needs to contain details about where to reply to. The first step is to create a relationship between the consultation and the 'reply to' information:

```
<div rel="argot:replyTo">
  ...
</div>
```

51. The information about where to reply to is provided via the vCard vocabulary, so the next step is to create a container for this information:

```
<div rel="argot:replyTo">
```

```

    <div typeof="v:VCard">
      ...
    </div>
  </div>

```

52. Now we can add any properties we like to the vCard, such as phone and fax numbers, email addresses, postal addresses, and so on:

```

<div rel="argot:replyTo">
  <div typeof="v:VCard">
    <h2>Contact us</h2>
    Tel: <span property="v:tel">020 3334 5473</span>
    Fax: <span property="v:fax">0870 739 5917</span>
    <a rel="v:email"
href="mailto:paroleboardconsultation@justice.gsi.gov.uk">email</a>
    <h2>Post:</h2>
    <span property="v:fn">David Liddemore</span>
    <div rev="foaf:member">
      <div typeof="foaf:Organization">
        <span>Access to Justice Sponsorship and Performance
Unit</span>
        <span property="foaf:name">Ministry of Justice</span>
      <div rel="v:adr">
        <div property="v:street-address">3.18, 102 Petty
France</div>
        <div property="v:locality">London</div>
        <div property="v:postal-code">SW1H 9AJ</div>
      </div>
    </div>
  </div>
</div>

```

More information on the [Consultation argot](#) is available online.

## Putting job vacancies into Linked Data

53. All public sector jobs should be advertised online in a form that allows them to be re-usable.
54. All external, and all 'Civil Service only' vacancies must be advertised on the [Civil Service website](#).
55. You must [register online](#) using a .gov.uk email address: to be able to view this. You will be sent an email to activate your account. Copy and paste your password to avoid error.
56. You then need to contact [civilservicecareers@parity.net](mailto:civilservicecareers@parity.net) for authorisation to load jobs. Read the [Job Form guidance page](#) for more details.
57. Further guidance on [how to mark up a job vacancy](#)<sup>16</sup> is available as a tutorial on the argot-hub. The tutorial covers the following concepts and elements:
  - Loading vocabularies
  - Identifying a vacancy
  - Joining the Linked Data cloud
  - Identifying the publisher
  - Indicating the source of the vacancy
  - The title of the job
  - Providing a reference
  - Closing date
  - Interview start date
  - Number of jobs available
  - Short description
  - Long description
  - References to supporting documents
  - Indicating a job
  - Salary range
  - Location
  - Type of job
  - Working arrangements
  - Specific requirements

---

<sup>16</sup> <http://code.google.com/p/argot-hub/wiki/GettingStartedEmbeddedHtml>