Metadata Standard for online content

Version 2

Date published: August 2016

This document forms part of The Open University’s suite of web standards, available online at www.open.ac.uk/about/web-standards.
The Open University metadata standard for online content

If this is the first time you have referred to this standard we recommend that you read through the whole document to ensure you fully understand the importance of adding the correct metadata to your web pages and linked documents.

Summary of key points

- You **MUST** include title metadata, description metadata, keywords metadata and a default language attribute on all web pages as a minimum.
- Metadata on externally-facing websites **MUST** be utilised to ensure the visibility of OU web content.
- Make sure your web pages are indexed appropriately using the robots meta tag - you may or may not want search engines to find, index, and follow links from your pages.
- You **MUST** include metadata on documents linked to from web pages. Generally, the metadata is specified in a document's properties fields. In brief:
  - Write titles that are meaningful out of context so people can understand what the documents are about from a search results page.
  - Include at least 3 keywords that summarise the document's subject matter.
  - Specify an owner for each document (either an individual or a team).
  - Make sure you update the metadata of any document that you are reusing.

1. Overview

   1.1. This standard describes the minimum amount of metadata to be supplied on The Open University websites and any documents linked to from web pages.

   1.2. This standard applies to both public-facing and intranet sites.

2. An introduction to metadata

- Metadata is the term given to data that describes the content (or subject matter) of web sites, web pages and documents, which allows it to be indexed appropriately by the University’s search engine, as well as other search engines such as Google and Bing. All the metadata tags specified in this standard are used by OU Search but may not necessarily be used by an external search engine.

- It might be helpful to think of metadata as describing the contents of a page or document (what it is about), which is not necessarily what you see on that page For example, how do you ensure a Google search using the terms 'OU course funding' will return the page entitled 'Ways to pay for a qualification', which does not include this phrase within the body of its text? Summarising your page succinctly and using synonyms and alternate phrases to describe your content in your metadata will help ensure its visibility and inclusion within relevant search results.

- The presence of metadata ensures that search results are meaningful to users. It also improves the chances of a document or web page being given a high relevancy ranking, and therefore appearing higher in the list of search results. In 2015 Library Services carried out a
case study demonstrating the huge positive impact metadata has on the ranking of web pages in search results. It showed that 81% of web pages moved up in search results when metadata was added - and the figure was 92% when looking just at keywords. Documents fared just as well: adding relevant titles increased their rankings in 75% of tests. Please see the full case study for further details on the testing and results.

- Using inconsistent or inappropriate metadata, or omitting it altogether, can prevent users finding your content and can cause frustration.
- Careful consideration ought to be given to metadata early in the authoring/production process. The Library can offer advice on appropriate taxonomies (or controlled vocabularies) to use in your metadata. Email library-metadata@open.ac.uk for details.

3. Metadata on web pages

- This section describes the metadata that MUST be included on web pages.
- As a minimum, you MUST include the following on each page: title, description, and keywords.
- If you are publishing pages on an externally-facing website you MUST also refer to the information in the SEO guidelines to ensure your site is fully optimised for external search engines. Please consult the Digital Engagement team (email digital-engagement@open.ac.uk) for further information.

3.1. Title tag

3.1.1. Background information

- The University’s search engine uses the title tag to index and weight the contents of a page.
- The words specified in the title tag are displayed in the bar at the top of the browser window or, if the browser uses tabs to organise different pages, on the tab itself.
- Note that the title tag is not the same as the title a web author might give a page using the <h1> element.
- The title tag is the most important element for ranking pages higher in the University’s internal search engine.
- All search engines use title tags to gather information about your website. The words in the title tag are what appear in the clickable link on the search engine results page.
- What you put in this tag is a key factor in determining how high you will rank in search results with the major search engines.

3.1.2. The title tag MUST be the first metadata element of the head section of a page.

3.1.3. Each page SHOULD have a unique title tag up to 80 characters long.

- A title tag can be longer than 80 characters, but we recommend keeping it to within this length as search engine results pages tend to display only the first 80 characters.
- You may want to structure your title so that the most relevant words are at the beginning, and therefore unlikely to be cut off on search engine results pages.
- For advice on optimising your pages for external search engines, refer to the SEO Guidelines.
3.2. Description tag

3.2.1. Background information
- The University’s search engine uses the description tag to index and weight the contents of a page.
- Depending on the search engine, the words in the description tag are generally what appear below the clickable link title on the search engine results page. The description should be an accurate summary of the contents of a particular page.

3.2.2. Home pages **MUST** have a unique description specific to the website.
3.2.3. Content pages **MUST** have a unique description specific to that particular page.
3.2.4. Descriptions **SHOULD** be limited to approximately 25 words as search engines often truncate them in search results.

3.3. Keywords tag

3.3.1. Background information
- The University’s search engine uses keywords to index and weight the contents of a page.
- Most importantly it’s quality not quantity.

3.3.2. You **SHOULD** use words that you think people might use in searches to find your page. Consider what attributes the page has that make it unique.
3.3.3. You **SHOULD** use words that will distinguish it from other similar resources.
3.3.4. If possible, you **SHOULD** use a minimum of 3 keywords per resource; however if you are not able to specify 3 keywords then enter as many that are relevant.
3.3.5. You **SHOULD NOT** exceed 250 characters.

- Many search engines do not like pages to have too many keywords and will push them further down the search results page. Excessive keyword repetition is known as ‘keyword stuffing’ and will adversely affect search results.

3.3.6. You **SHOULD** use lowercase.

- All search engines will find keywords in lowercase, but only some search engines will find keywords that are written in capitals; therefore, to obtain the optimum results you should always write keywords in lowercase, including proper nouns and acronyms.

3.3.7. If you are using acronyms you **SHOULD** include both the acronym itself and its full name; for example, “ou” and “the open university”.
3.3.8. Individual terms from keyword phrases **SHOULD NOT** be separately entered; for example, “equality and diversity training” is sufficient, “equality training”, “diversity training”, “equality” and “diversity” do not need to be separately recorded.

3.3.9. Home pages **SHOULD** have a set of keywords relevant to the whole website.
3.3.10. Content pages **SHOULD** have unique keywords specific to each page.
3.3.11. You **SHOULD NOT** repeat keywords across pages as this reduces the effectiveness of indexing and the relevance of search results.
The following is an example metadata record from the University’s Human Resources web site (this is what you would see if you right click on the ‘About Human Resources Development’ and choose to view source):

```
<title>Human Resources Development: About HRD</title>
<meta name="description" content="Website providing a learning and development resource area for all staff at The Open University." />
<meta name="keywords" content="staff training, leadership, inclusion, compliance, diversity in the workplace, equality, valuing diversity, management development, learning, organisational development, change"/>
```

3.4. Language tag

3.4.1. Background information

- The language tag indicates the main language of the page for the screen readers and other text processors.
- It also allows us to differentiate between pages we have alternate versions e.g. pages made available in both English and Welsh

3.4.2. The default language of the page **MUST** be encoded as an attribute of the html:

```
<html lang="en">
```

3.4.3. You **MUST NOT** use the meta element in the document head, as this is no longer accepted practice and has been classed as ‘non-conforming in html’ by the W3C

3.4.4. When there is content within the page that is different to the predominant language, you **SHOULD** add a language attribute to an element surrounding that content:

```
<p lang="fr"> would signify that the text within the paragraph is French.
```

Here the attribute is added to a span to show that part of a sentence isn’t in the default language:

```
<p>In Spanish that would be: <span lang="es">la biblioteca es a través del campus</span>.</p>
```

3.5. Robots tag

3.5.1. The robots meta tag allows you to control how a webpage is indexed.

Mostly, indexing choices will be made automatically by the content management system being used. However, sometimes these choices should be made manually. For example:

- You might want a search engine to follow all links but not index the content, such as in the case of pages that are merely a collection of links.
- You might want the content to be indexed, but none of the links followed.
- You might not want a page indexed or any of its links followed, in effect excluding this page and pages accessed through it from the search index. This is required if a site needs
to be made live before it is made public; for example, to allow people to add document links.

3.5.2. You **MUST NOT** index a page if it does not have standard toolbar navigation; for example, a page that will appear in a pop-up window.

- This is because users who click on links to pop-ups from a search results page have no way to get back to the page they were previously on.

3.5.3. You **MUST NOT** index a page if it is not meaningful out of context; for example, a page presented in response to a form submission, or a transactions page.

Options:

```xml
<meta name="robots" content="noindex" />
This tells search engines to follow the links, but don't index the content. In our case, this will mean that the content is not indexed either internally or externally.

<meta name="robots" content="nofollow" />
Index the content but don't follow the links.

<meta name="robots" content="noindex, nofollow" />
Don't index the content (internally or externally) and don't follow the links.
```

For more complex indexing requirements contact the Digital Engagement team (email digital-engagement@open.ac.uk) who will forward on your request to the relevant team in IT.

4. **Document metadata**

Document metadata is generally contained in a document's properties. For instructions on finding and completing document properties refer to [Adding metadata to document properties](#).

- This section describes the metadata that **MUST** be added to the properties of documents included on OU websites.

4.1. As a minimum, you **MUST** include the following on each page: title, keywords, date and owner.

4.2. You **MUST** ensure that you provide a meaningful title for your document with sufficient information to enable user to understand the context.

- The document’s title is what is displayed in search results. Remember that each item in the list will be displayed out of context of the structure of your site.
- For example "Financial templates and guidance notes for xxx 2010" is much more descriptive than "Financial templates" or "Guidance notes" and helps users make informed choices.

4.3. You **MUST** include a list of comma-separated keywords that summarise the contents of the document.

- Remember to try and identify at least 3 keywords.
- Use words or phrases that other people might use to search for the document.

4.3.1. Use terms that specify the general content of a document. These should be included when they have been omitted from the full text of the document.
• For example, an article that discusses the Shard building in London at length but the word 'architecture' is never used within the text.

4.3.2. Use synonyms, including:
  • True synonyms; for example, "leprosy" and "Hansen's disease" or "tutors" and "Associate Lecturers".
  • Variant spellings; for example, "donut" and "doughnut".
  • Abbreviations, acronyms and initialisms; for example, "T-rex" and "Tyrannosaurus rex".
  • Variance across languages or dialects; for example, "soccer" and "football".

4.3.3. These keywords SHOULD NOT repeat words in the document title; however, you MAY choose to put in the full text if the title contains an acronym (for example, if the title is CDSA you can use "Career Development" and "Staff Appraisal" as keywords).

4.4. You MUST ensure that each document has a clearly identified owner, either an individual or a team.

4.5. You MUST include a date on the document, preferably at the end of the text.

4.6. You MAY include the date in the document's footer; however, be aware that this is only viewable if the document is being displayed in page layout view.

4.7. You MUST ensure that you update a document's properties if you are reusing it (copying an existing document) otherwise the old metadata will remain and return inaccurate search results.