

Australian Government Department of Health and Ageing

Publishing standards for HealthInsite

© Commonwealth of Australia

Produced by the HealthInsite Editorial Team, Australian Government Department of Health and Ageing. Version 2, August 1999. Version 3, December 2002. Version 4, April 2005. For further information contact: healthinsite.feedback@health.gov.au, phone 02 62897505.

HealthInsite aims to provide access to up-to-date, quality information for Australian health consumers. As the information to be accessed through HealthInsite comes from the sites of the contributors, the overall quality of HealthInsite depends on the quality of the information and its presentation through those sites. These standards help to add credibility to HealthInsite and to contributor sites.

Contributor sites submitted for approval are assessed by the HealthInsite Editorial Board, a group of eminent health professionals and consumer representatives independent of the Department of Health and Ageing. Assessment is based on the criteria detailed in *Assessment of content for HealthInsite*. The HealthInsite Editorial Team supports the Board in these tasks.

In this document, the term 'resource' is used to describe an item of published information (for example, a web page or a fact sheet) or a collection of items.

1. Introduction and principles

These standards cover both general site and resource specific issues. HealthInsite aims to route users quickly to useful information - hence it will generally direct users to individual resources rather than Web sites or segments of sites. Accordingly the standards are focused at resource level rather than site level. In addition, much of the detail is focused on 'documents' and similar resources. However, in general, the same principles apply to other resource types e.g. multimedia.

The standards are aligned with the *Guidelines for Commonwealth information published in electronic format, January 2000*, available at <http://www.agimo.gov.au/information/publishing/formats>.

HealthInsite standards are updated, in consultation with contributors, as the technical environment changes and requirements vary.

The standards fall into five areas:

- quality of information - qualifications of creators, accuracy, appropriateness for target audience, documentation of source material, clearance procedures
- authentication - clear and explicit statements of title, creator, publisher and date

- usability - navigation, information architecture, markup and presentation of information
- accessibility - catering for a range of user hardware and software and for people with disabilities
- metadata requirements.

Ideally HealthInsite contributor sites will be excellent in all these areas. In reality, sites and standards have been evolving and, particularly on large older sites, not all resources meet current standards. Hence, what we are looking for from contributors is a commitment to continuous site improvement in all the listed areas.

For HealthInsite contributor sites, all selected HTML pages will be accessible via the text search engine. Metadata for each resource enables more precise sorting of search results and retrieval of relevant items for display on HealthInsite topic pages.

The role of contributors, in consultation with the HealthInsite Editorial Team, is to:

- identify and evaluate resources for possible inclusion on HealthInsite
- create and maintain resources and the corresponding metadata records
- advise the Editorial Team of the overall status of the contributor site with respect to the publishing standards
- check against the publishing standards when planning a site upgrade
- advise the Editorial Team on site improvement progress and major site changes, including URL changes.

The types of resources selected for HealthInsite are:

- homepages (representing whole sites)
- individual publications, including factsheets and brochures
- information about a health topic
- information about services
- broad level information about an organisation
- navigation pages, where these provide a useful entry point to part of a site.

Resources intended for consumer audiences or for health professionals will be considered, as will resources of general interest. To be selected, resources and sites must satisfy the 'required' standards described in this document.

2. Quality of information

Required

Sites must have a written policy and procedure that:

- includes a policy that each resource is authored by a person or group with appropriate qualifications/experience
- includes a procedure for appropriate attribution of resources

- includes a review process (with capacity for review prior to initial placement onto the site and at appropriate intervals to ensure currency). The policy needs to cite positions/qualifications/names of who reviews.
- details the final approval process (including responsibility/qualifications)
- addresses conflict of interest
- includes a policy on advertising
- includes a process for consumer consultation and/or audience testing
- ensures that sites observe the copyright restrictions on any material (including images and other media) used in resource preparation.

Recommended

- Consider applying the HONcode Principles for medical and health web sites at <http://www.hon.ch>.
- If a resource is intended for a particular audience, then make this clear as early as possible in introductory material and present information appropriately for the intended audience.
- Include references to source material used in the preparation of the resource.

3. Authentication

The aim of authentication is to ensure that a user will be able to recognise a resource on a contributor site as having authority and thus be able to trust the information in the resource. Through HealthInsite and other search engines, a user may reach a resource directly, without the context of navigating from the site's home page. Hence each HTML page, PDF document or other resource must have individual explicit authentication.

One of the benefits of the Internet is the ability to change resources quickly. However, version control is needed to keep track of changes and, for example, to synchronise print versions of documents with Internet versions. It is important to distinguish between changes in content of a resource and changes in presentation. **A user will associate a date on a resource with the date for which that resource was current - it is important that this be a realistic assumption.**

HealthInsite's success depends on providing current information for consumers. With the wide range of people who visit the site it could be misleading to include rescinded or out-of-date resources prominently. Although such resources may be useful for research purposes, our current approach is to exclude them from HealthInsite. However we encourage contributors to maintain online access to archival resources.

Required

- clear publisher identification
- a clear title
- the date of publication or, if the resource content has been updated, the date of last update. If the resource has been fully reviewed and judged to be current, the date of last review is acceptable.

- a contact point for requesting further information - this may be a standard feedback link to the site webmaster. The procedure for dealing with enquiries and feedback is entirely the responsibility of the site owner.
- clear labelling of any archival material which is retained on a site
- adherence to the copyright restrictions on any material (including images and other media) used in resource preparation.

Recommended (either at site level or resource level as appropriate)

- a full copyright statement (for example, as on print publications) or short copyright statement with a link to the full statement. Generally the site copyright notice should allow for the printing and downloading of resources for personal use.
- a written policy on links to other sites. This may include a disclaimer that links to external sites should not be taken as implying support of an external organisation or endorsement of information on other sites. It may also make special provision for links to commercial/private sector site where health advice may be combined with advertising.
- details about the ownership, management and sources of funding of the site
- a disclaimer notice. It is recommended that a disclaimer notice be placed on any resource offering health advice. It is preferred that this be a simple notice that the resource is of an educational nature rather than a long-winded legalistic disclaimer of all liability.
- details of who produced each resource (author/creator area)
- the URL of the resource
- avoid code that automatically changes the date displayed whenever the file is changed. Because this is invoked even for trivial presentation changes, such as editing a font size, it can be misleading.
- avoid multiple dates on a page - this can be confusing to users.

4. Usability

Usability is a measure of whether people manage to do what they want to on a website, and whether (consequently) the site achieves its goals.

From the user's perspective, usability is important because it can make the difference between performing a task accurately and completely or not, and between enjoying the process or being frustrated. From the developer's perspective, usability is important because it can mean the difference between the success or failure of a site.

Common factors impacting on the usability of a website include:

- ease of learning
- efficiency of use
- memorability
- error frequency and severity
- user satisfaction.

By following accessibility and standards based web design guidelines, developers can go a long way towards ensuring that their sites are usable. It is recommended that developers also consult and follow the Usability Checklist (Australian Government Information Management Office (AGIMO)) at:

<http://www.agimo.gov.au/publications/2004/06/toolkit/testing/checklist>. This checklist covers: architecture and navigation; layout and design; content; forms; platform and implementation; and accessibility.

Some particular requirements and recommendations for HealthInsite are provided below.

4.1 Site level issues

Recommended

- A distinctive style across the whole site will help users to recognise the site when they come to a resource via HealthInsite.
- Avoid frames. Although HealthInsite has a method to link to individual frame URLs, there is a loss of site context and links. The result is not entirely satisfactory for site owners or users.

4.2 Document formats

For consumer information and short documents, HTML is the preferred format, because of its linking functionality. Some HTML page styles do not print well. If this is the case, it is helpful to offer a version of the page optimised for printing.

For long documents and documents with special layouts, PDF will be a better format for many users. However, note that there are still significant accessibility issues with the PDF format. It is helpful to offer an alternative version, preferably HTML.

Document formats other than HTML and PDF are not included in HealthInsite.

4.3 General text standards for individual HTML documents/pages

A major feature of HTML design, and the operation of Web browsers, is that HTML code which is not understood by a browser is simply ignored, and does not cause errors during operation. By taking reasonable precautions, and following basic guidelines, you can be confident of successful display of documents by foreseeable web browser software.

Required

- Documents must conform to current specifications of HTML or XHTML. Note that documents conforming to earlier versions of the HTML specifications will automatically conform with later versions.
- Documents must contain informative text alternatives for all contextual images, in order to facilitate use by text-only Web browsers and text reader software. Remember that navigation buttons are images and need text alternatives (see Section 5 for further requirements on accessibility).

Recommended

- Individual pages, including images, should generally not be larger than 50Kb (i.e. total combined page source code plus images should not exceed this limit).
- Where possible, use cascading style sheets.
- Use meaningful document titles and page headings.
- Keep scripting (e.g. Javascript) to a minimum.
- Use a link (or links) to help the visitor to move into the site if desired, rather than back to HealthInsite.
- All link text should be underlined and a different colour from the body text. Default colours are preferred for link, activated link and visited link.
- Avoid using images as links. If images are to be used as links, include a text link next to the image as well.
- Users should be warned if a link is leading to a large file which might take a while to download.

4.4 Long documents

Recommended

- Very large textual documents are generally not suitable for presentation as single HTML pages. As a rule of thumb, for a document longer than about five A4 pages, consider splitting it into multiple HTML pages. Also consider creating a PDF version which can be downloaded by the reader from the site.
- A long document on a single page should have a contents block at the top of the page connecting to subheadings within the page. There should also be a link [TOP] at the bottom of the page and at intermediate points if appropriate.

4.4.1 Long documents split into parts (multiple HTML pages)

Required

- A multi-page document must contain a top page with the full identification information and a contents list. There should be a link to it from every subsequent page.
- For HealthInsite, metadata is required only for the top page.

Recommended

- Each page should contain its own identification information, since a visitor might get to it via a search engine rather than via the top page.
- [Next page] and [Previous page] links will help the user moving through the text.

4.4.2 Long documents published as PDF versions

Required

- PDF documents must contain explicit authentication as described in Section 3.
- To be included in HealthInsite, a PDF document must have a discrete HTML cover page. This should describe the document sufficiently for the user to decide whether or not to open or download it. It should provide instructions for

downloading. Make sure that the cover page has the same creator, publisher, title and date information as the PDF document.

Recommended

- Users will need a copy of the Adobe Acrobat reader software. Users of text reader software will need additional reader software from Adobe. A link should be provided to the Adobe site so that users can download the reader software.
- PDF files should generally not be larger than 1Mb because of excessive downloading time. It is better to break very long documents into two or more PDF files. HealthInsite will not link to files over 2Mb in size unless the document is also available in smaller files.
- File size should be clearly indicated with the download link.

4.5 Non-textual material

The Web is a highly visual medium and non-textual material is encouraged wherever it will add value.

Consider accessibility issues (see Section 5 below) in the use of colour and tables.

4.5.1 General image standards

Required

- Ensure that copyright requirements are adhered to and that images are in good taste.
- Unless an image is purely decorative, the document HTML code must include an alternative informative textual description of the image for use by text-only browsers and text readers.
- Long descriptions must be provided for complex images.
- Image maps must be accompanied by text equivalent menus.

Recommended

- Because images slow down the page loading process, make access to images, animations, etc, optional via links. Use thumbnails where appropriate to show the user what image they are selecting.
- Animated images are supported subject to meeting file size criteria and not requiring multiple server accesses. Use them sparingly because they can be very irritating. Use low loop counters, i.e. loop twice or three times then stop.

4.5.2 Interactivity

Web pages provide a powerful means for allowing users to interact with the site owner. For example, forms can be designed to permit visitors to input data of many types. Many issues arise - privacy, security, identification, etc. It is feasible for HealthInsite to point to interactive pages on a contributor site but management of the interactive process is entirely the responsibility of the site owner.

4.5.3 Multimedia

Multimedia resources, including sound and video files, can be selected for HealthInsite topics and be accessible via the HealthInsite search facility. They must have HTML cover pages with appropriate metadata. It is recommended that these cover pages include comprehensive descriptive information. The principles of quality, authentication, usability and accessibility must be considered.

5. Accessibility

Following standards-based web design and accessibility guidelines allows web content to be more available not only to people with disabilities, but to all users, whatever hardware and software they are using to access the Internet or constraints they may be operating under. Utilising these guidelines will also help users find information on the web more quickly.

The standard for web content accessibility is the *Web Content Accessibility Guidelines*, which were devised by the World Wide Web Consortium (W3C). The guidelines are available at: www.w3.org/tr/wai-webcontent. The W3C guidelines provide a series of checkpoints that can be used to ensure that websites are accessible. Each checkpoint has a priority level assigned by the Working Group based on the checkpoint's impact on accessibility.

The Human Rights and Equal Opportunity Commission (HREOC) has responsibility for promoting the objectives of the Disability Discrimination Act (DDA) and provides advice about the implications of the Act for website operators. HealthInsite contributors should also be familiar with the following document from HREOC: *World Wide Web Access: Disability Discrimination Act Advisory Notes*, at www.humanrights.gov.au/disability_rights/standards/www_3/www_3.html.

Other documents that will assist with ensuring that websites are accessible are *Core Techniques for Web Content Accessibility Guidelines 1.0*, available at: www.w3.org/TR/WCAG10-CORE-TECHS/.

A range of evaluation, repair and transformation tools for web content accessibility are available at the Web Accessibility Initiative website at www.w3.org/WAI/.

In addition, some organisations provide website testing and advisory services.

6. Metadata requirements and user guide

More guidance is provided in this section of the publishing standards because of the importance of metadata in HealthInsite.

6.1 Introduction

Metadata is indexing information about a resource which can be used by search facilities to make searching more precise. Metadata is crucial to HealthInsite for search and navigation functionality. A metadata record is required for each resource selected.

Many web resources are subject to change. Details such as title, creator and description may be affected. Hence it is important that contributors maintain the metadata records within their own sites and update the metadata when they update the content of any resource. The HealthInsite harvester regularly checks sites so that the HealthInsite database can be updated. More recently updated resources will have a higher ranking in HealthInsite search results.

The HealthInsite metadata specification is compliant with the AGLS metadata element set (Australian Standard AS 5004.1-2002 and AS5044.2-2002). Further details about AGLS are available on the National Archives of Australia website at http://www.naa.gov.au/recordkeeping/gov_online/agls/summary.html. The full HealthInsite specification is at <http://www.healthinsite.gov.au/metadata.cfm>. Where contributor sites are under a mandate to provide full AGLS records, the HealthInsite specification is recommended. Other contributors may also wish to provide full records; the work required to produce these is much appreciated.

In 2003, HealthInsite began an alternative approach whereby contributors maintain just 8 key elements of the metadata set. The extra elements are added by HealthInsite staff and are maintained only in the HealthInsite database. This alternative now forms the minimum requirement for contributor metadata, as described here. If you have older resources with full HealthInsite metadata records, it is not necessary to replace these with short records (though you may wish to do so at some stage).

Some contributors create additional metadata elements, either for their own purposes or for contributions to other portals. The HealthInsite harvester ignores metadata which does not match the HealthInsite syntax.

6.2 Metadata syntax

The metadata must be displayed for the HealthInsite harvester with the following HTML syntax or with XHTML syntax. It is important to use the META syntax with correct punctuation. Upper/lower case is less important. DC stands for Dublin Core, the international standard on which AGLS is based.

```
<META NAME="DC.Creator" CONTENT="">  
<META NAME="DC.Publisher" CONTENT="">  
<META NAME="DC.Title" CONTENT="">
```

```
<META NAME="DC.Description" CONTENT="">
<META NAME="DC.Language" SCHEME="RFC3066" CONTENT="">
<META NAME="DC.Date.Modified" SCHEME="ISO8601" CONTENT="">
<META NAME="DC.Format" SCHEME="IMT" CONTENT="">
<META NAME="DC.Identifier" SCHEME="URI" CONTENT="">
```

Most HealthInsite contributors have the metadata record embedded into the HTML code for the resource. With this method, any user can see the metadata by viewing the source code of the HTML page. For PDF documents, the metadata record is generally embedded into the HTML cover page that links to the resource.

Another method is to store the metadata record in a separate HTML file. With this method, HealthInsite needs to know the metadata record URL as well as the corresponding resource URL.

HealthInsite is currently working on accepting batches of metadata records as a single file, with opening `<item>` and closing `</item>` used as record separators. This could be useful for sites with content management systems - the output can come straight from the database. We are also working on accepting harvest control lists (HCLs). These lists could contain either full metadata records, or URLs which link to HTML resources with embedded metadata.

XML and XML/RDF syntax are not currently accepted. We are waiting for an Australian standard recommendation before proceeding.

6.3 Metadata creation

These are some of the ways that our current contributors create metadata records:

- Filling in the content details into a raw HTML syntax template - then embedding this block of text into the resource source code.
- Using a metadata tool which has a form interface and some menus to select from. The tool creates the correct syntax. Metadata tools are described in the MetaMatters site at <http://metabrowser.dtdns.net/dcanz2/mb.aspx>. HealthInsite has a metadata tool available to contributors although it is mostly used by HealthInsite staff.
- Creating metadata as individual database fields within a content management system. Again the system creates the correct syntax for output.
- Copying and pasting records which HealthInsite staff have created using the HealthInsite metadata tool.

It is important that the metadata match the resource and that the metadata be derived from the resource. For PDF documents, the resource, cover page and metadata must all match. Be particularly careful with titles and dates.

If you are using a metadata template with default values, always check that the defaults are appropriate for the resource you are currently indexing. If you are copying, pasting and editing metadata from a previous resource (because your new resource is similar), make sure that you do the editing promptly and completely. Proofread to make sure the metadata matches the new resource.

Who creates the metadata?

Initially, HealthInsite staff will provide assistance with metadata creation. In the longer term it is expected that contributor sites will undertake this work for their own sites but HealthInsite will continue to help if needed.

If site management is outsourced, then contributors will need to negotiate with their contractors. Some of the issues to clarify are: Which pages need metadata? Who will create the metadata and how? Does the contractor have indexing expertise? Who will check the metadata? Who will keep the metadata up-to-date and how?

Updating

If you are modifying a resource, always check the metadata and update if appropriate. If there has been any change in the real content of the resource, then, as a minimum, the metadata Date.Modified will need to be updated. However, you should also check the other fields. For example: Is there a new title? Has the resource changed so much that it needs a new description? Have you added new language versions? Has a previous PDF document now been made available as HTML?

Organisational changes

Government departments and their subdivisions are especially prone to name changes, but it can happen with other organisations as well. Generally resources should be presented with the correct authentication details as at the time of original publication and the metadata should reflect this. However, for resources which are being updated, the authentication details at the time of update are appropriate.

Metadata on higher level pages, including the home page

Sometimes it is better to route people to the top page of a set of resources rather than directly to individual items. For example, if you have a set of documentation relating to a grants scheme, it may be better to put metadata on the entry level page and not on the individual forms, advertisements, etc.

If you are putting metadata on a higher level page, then generally it should relate to the set of pages encompassed by the higher level page. For example, metadata on a site's home page should relate to the site as a whole. You also need to keep track of any changes to the set of pages encompassed by the higher level page and update the metadata accordingly. For example, in an active site, the metadata Date.Modified on the home page needs to be updated regularly.

6.4 The metadata elements

Content must be provided for each element as described below.

Creator

The name of the person or organisation primarily responsible for the content of the resource.

- For textual documents, the author is the creator.
- For personal creators, the format "Lastname, Firstname" is recommended.
- In many organisations, personal authorship is not recorded; the organisation is regarded as the author, not the person. Use the name of the organisation.
- If an organisation name has a well known acronym, add the acronym in brackets at the end of the name.
- Do not enter the name of a person or contractor who has merely converted a resource into an Internet version (for example by marking up a document with HTML coding).
- However, if the resource's content has been commissioned or produced under contract, then it may be appropriate to enter the name of the contractor (personal or company).
- You can enter more than one name for joint creators. Extra names should be added in separate lines with the same META syntax.

Publisher

The name of the entity responsible for making the resource available.

- Generally this will be the name of the site owner.
- For PDF documents, use the publisher name on the resource.
- For older resources, use the name valid for the date on which the resource was published.
- If the publisher name has a well known acronym, add the acronym in brackets at the end of the name.
- You can enter more than one name for joint publications. Extra names should be added in separate lines with the same META syntax.

Title

The name given to the resource.

- Use the title as it appears on the resource itself. Generally this would also match the title in the <title> area of HTML.
- Titles should preferably be in lower case except for the first letter of the first word and proper names.

Description

A textual description of the content and/or purpose of the resource.

- Generally the description should be one to two sentences - just enough to help a user decide whether to follow the link to the resource. Use information from the abstract or summary of the resource if available. In writing a description it is

important to step out of your organisational frame of reference and think from the user's point of view.

Language

The language of the content of the resource

- Scheme: RFC3066 - tags for the identification of language. A short list of codes is available at ftp://dkuug.dk/i18n/ISO_639.
- A more comprehensive list of codes is at <http://lcweb.loc.gov/standards/iso639-2/englangn.html>. If there is no 2-letter code for a particular language, then select a 3-letter code.
- The code for English is "en".
- More than one code can be entered. For example, the metadata might be on a cover page that links to versions of a document in different languages. Use a semi-colon space delimiter between codes.
- Scheme RFC1766 will also be accepted by HealthInsite. Use the short list of codes at ftp://dkuug.dk/i18n/ISO_639.

Date.Modified

The date when the content of the resource was last updated.

- Scheme: ISO8601 (use formats YYYY or YYYY-MM or YYYY-MM-DD).
- Date.Modified must reflect the currency of the resource content. It should not be the date when the resource was converted to HTML or other format. Nor should it be updated for trivial changes to the presentation of the resource.
- If a resource has not been modified since it was first published, then Date.Modified should be the date of publication.
- Enter the date as fully as it is displayed on the resource - for example, if the resource says September 2000, use 2000-09 not 2000-09-19 and not 2000 alone.

Format

The data format of the resource.

- Scheme: IMT
- Commonly used values are: text/html, application/pdf
- The full listing is available at <http://www.isi.edu/in-notes/iana/assignments/media-types/media-types>.
- If you have both an HTML and a PDF version of the resource, use "text/html"

Identifier

The URL for the resource.

- Scheme: URI
- For a PDF document, use the cover page URL.

6.4.1 Mock up example

```
<META NAME="DC.Creator" CONTENT="Smith, Fred">
<META NAME="DC.Creator" CONTENT="Jones, Betty">
<META NAME="DC.Publisher" CONTENT="Australian Society of Nobodies">
<META NAME="DC.Publisher" CONTENT="Australian Society of Important People
(ASIP)">
<META NAME="DC.Title" CONTENT="Are you important?">
<META NAME="DC.Description" CONTENT="Everyone is important. 10 tips on how
to improve your self esteem.">
<META NAME="DC.Language" SCHEME="RFC3066" CONTENT="en;es;it">
<META NAME="DC.Date.Modified" SCHEME="ISO8601" CONTENT="2003-07-
15">
<META NAME="DC.Format" SCHEME="IMT" CONTENT="text/html">
<META NAME="DC.Identifier" SCHEME="URI"
CONTENT="http://www.importantpeople.com.au/important.htm">
```