

Australian Government Department of Health and Ageing

HealthInsite metadata specification

© Commonwealth of Australia

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

Introduction

The HealthInsite metadata specification is compliant with the AGLS metadata element set (Australian Standard AS 5044.1-2002 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 *HealthInsite metadata requirements and user guide* are online as Section 6 of the *Publishing standards for HealthInsite* at http://www.healthinsite.gov.au/content/publishing_standards.cfm. The guide provides details of metadata creation and harvesting options. Contributors to HealthInsite can choose whether to use a short version of just 8 key elements in their metadata records or the full specification as described below. Other Australian health agencies are welcome to use this specification.

Syntax for data exchange

For data exchange purposes, metadata output from HealthInsite is produced with HTML syntax as shown below. Input can be accepted in this same HTML syntax or XHTML syntax.

Generally the HealthInsite metadata harvester ignores any metadata which does not have the correct syntax. However there is some flexibility with order of elements and order of element parts.

```
<META NAME="DC.Creator" CONTENT="">
<META NAME="DC.Publisher" CONTENT="">
<META NAME="DC.Rights" CONTENT="">
<META NAME="DC.Title" CONTENT="">
<META NAME="DC.Title.Alternative" CONTENT="">
<META NAME="DC.Subject" SCHEME="Health Thesaurus" CONTENT="">
<META NAME="DC.Description" CONTENT="">
<META NAME="DC.Language" SCHEME="RFC3066" CONTENT="">
<META NAME="DC.Date.Created" SCHEME="ISO8601" CONTENT="">
<META NAME="DC.Date.Issued" SCHEME="ISO8601" CONTENT="">
<META NAME="DC.Date.Modified" SCHEME="ISO8601" CONTENT="">
<META NAME="DC.Date.Review" SCHEME="ISO8601" CONTENT="">
<META NAME="DC.Date.Reviewed" SCHEME="ISO8601" CONTENT="">
<META NAME="DC.Type" SCHEME="HI type" CONTENT="">
```

```
<META NAME="DC.Type" SCHEME="HI category" CONTENT="">
<META NAME="DC.Format" SCHEME="IMT" CONTENT="">
<META NAME="DC.Format.Extent" CONTENT="">
<META NAME="DC.Identifier" SCHEME="URI" CONTENT="">
<META NAME="AGLS.Availability" CONTENT=""
<META NAME="AGLS.Audience" SCHEME="HI age" CONTENT="">
<META NAME="HI.Complexity" CONTENT="">
<META NAME="HI.Status" CONTENT="">
```

Description of the metadata elements

The designation of an element as “required” is the requirement for the HealthInsite database. Where HealthInsite contributors provide short metadata records, the additional required elements are created by the HealthInsite Editorial Team.

For PDFs and other non-HTML resources, HealthInsite requires an HTML cover page and links to that. This is so that users can find out more information about a non-HTML resource before deciding whether to open/download it. The metadata is usually embedded in the cover page source code but the content of the metadata must match the resource itself.

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

- 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, followed by a full stop and the subdivision name if appropriate. It is not worthwhile going into the complexity of subdivisions of subdivisions – these details can be recorded in the document itself. For older resources, record the name of the organisation as it was at the time of publication.
- 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). This is often the case with reports on government sites.
- 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. Required

- Generally this will be the name of the site owner.
- For non-HTML resources (particularly PDFs), 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.

Rights A statement or pointer to a statement about the rights management information for the resource (eg a copyright statement). Optional

- AGLS recommends the use of this element and it is included in the HealthInsite database. However it is not actually being used in HealthInsite and we think it is more important to put the rights management details on the resource itself.

Title The name given to the resource. Required

- Use the title as it appears on the resource itself. This should also match the title in the <title> area of HTML.
- The AGLS element refinement Title.Alternative allows you to record an alternative title (for example, a series title, or where a resource is well known by another name).
- Titles should preferably be in lower case except for the first letter of the first word and proper names.

Subject The subject and topic of the resource that succinctly describes the content of the resource. Required

- Scheme: Health Thesaurus
- The subject element requires indexing skills. HealthInsite staff perform subject indexing for most contributors. Guidelines for indexers are in a separate document. The main tool for subject concepts is the Health Thesaurus (The health and ageing thesaurus. 7th edition. Australian Government Department of Health and Ageing, 2005). Up to 10 thesaurus terms is normal, including the check tags for age groups, gender and place. Use a semi-colon space delimiter between terms.

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

Required

- 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. Required

- 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 The date the resource was created or became available in its present form.

- AGLS element refinements: Created, Modified, Issued
- HI element refinements: Review, Reviewed, HealthInsite
- Scheme: ISO8601 (use formats YYYY or YYYY-MM or YYYY-MM-DD)

Created = first date of publication/release in any form (eg print)

Issued = date posted to your site

Modified = date content last modified

Review = reassessment due

Reviewed = reassessment done

HealthInsite = date metadata added to HealthInsite (for HealthInsite database only)

- Occurrence: There should be no more than one occurrence of any particular date type. Date.Modified is required. The Created, Issued, Review and Reviewed dates are optional but can be entered if useful for resource management on your site.
- Date.Created should be the date when the content was created/published, not the date when the content was converted to HTML or other format. 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.
- Date.Modified is required. If a resource has not been modified since it was first published, then Date.Modified should be the date of publication.
- Date.Modified must reflect the currency of the resource content. It should not be updated for trivial changes to the presentation of the resource.
- On the other hand, if the content of the resource is updated, then Date.Modified must also be updated.
- Note that for PDFs, Date.Created and Date.Modified should match the details of the source document, ie do not use the date when the PDF was created from the source document or the date when the cover page was created or modified. Generally the source document will have a creation date only, so Date.Modified=Date.Created.

Type The category or genre of the resource. Required

- Schemes: HI type, HI category. Both these schemes are used in HealthInsite to enable limiting or partitioning of search results. (Note: the AGLS recommended scheme is not suitable for HealthInsite.)
- HI type is a broad grouping related to the format. The allowed values are: document, image, video, sound, software, data, multimedia.
- HI category is a genre grouping. Values which have been used are: announcement, directory, form, guidelines, homepage, navigation, organisation, overview, personal narrative, quiz, resource, service, statistics. The list is subject to change as HealthInsite develops.
- More than one value can be assigned, with semi-colon space delimiters.

- Assigning the type is closely linked to subject indexing and full details of HI category are provided in the subject indexing guidelines.

Format The data format of the resource. Required, size optional, one entry only

- Scheme: IMT
- Commonly used values are: text/html, application/pdf, image/jpeg, image/gif.
- The full listing is available at <http://www.isi.edu/in-notes/iana/assignments/media-types/media-types>.
- For a non-HTML resource with an HTML cover page, give the format of the resource.
- For very large HTML files and for non-HTML files, provide the size of the file in Kb in a separate line using the element refinement Extent.

Identifier A unique identifier for the resource. Required, one entry only

- Scheme: URI
- Generally use the resource URL.
- For a non-HTML resource with a cover page, use the cover page URL.

Availability How the resource can be obtained or contact information.

- For a non-HTML resource with an HTML cover page, provide the URL of the resource in a note like "Available at [PDF URL]". If several files are linked from a single cover page, use a note like "Available as a set of PDF files"
- (Note: AGLS intends this element to be used more for non-Web resources.)

Audience The target audience for the resource. Required, one entry only

- Scheme: HI age (Note: The schemes recommended by AGLS do not meet the HealthInsite need for a restricted list of values.)
- Allowed values: child, youth, adult
- Use the lowest age group applicable – for example, if a resource is targeted at both youth and adults then enter “youth” as the audience. HealthInsite is particularly concerned to identify resources targeted at children and youth.

Complexity The technical complexity of the resource. Required, one entry only

Allowed values: very easy, easy, medium, difficult, very difficult

very easy = very short, very simple resources aimed at consumers eg posters

easy = short, straightforward consumer resources eg consumer leaflets

medium = resources aimed at literate adults eg plain language texts

difficult = resources aimed at professionals or service providers

very difficult = research-level resources

- Generally use “easy” for resources aimed at consumers. However, use “medium” if the resource is very lengthy or covers a complex subject.
- Generally use “medium” for homepages.
- “difficult” should be used for professional/provider resources even if they are written in plain language which could be largely understood by any literate adult. Indeed, we encourage the use of plain language in material prepared for professionals and service providers. For resources of a substantial size,

consider carefully whether some professional knowledge is required to understand them fully. Indexing resources as "difficult" simply moves them up in the complexity scale; they are still accessible to all users.

- The Audience and Complexity elements do not try to cover all the possible target audiences for a resource. For example targets might be people with disabilities, ethnic groups or professional groups. If you wish to emphasise particular target groups, use the metadata description element and/or the introductory part of the resource itself.

Status An internal rating of the resource with respect to the HealthInsite search facilities and topic navigation structure and to manage metadata workflow. Required

- This element is used in the HealthInsite database to identify resources which satisfy the HealthInsite required standards and which therefore can be selected for HealthInsite topics.
- On the contributor site, the value "registered" should be used to indicate that the resource has been identified for HealthInsite.

Additional metadata elements

HealthInsite contributors may create additional metadata elements for their own purposes or to contribute to other portals. For example, keywords and description are often used:

```
<META NAME="Keywords" CONTENT="">
```

Used to repeat and supplement selected DC.Subject and DC.Type terms for recognition by some public search engines.

```
<META NAME="Description" CONTENT="">
```

Used to repeat DC.Description content for recognition by some public search engines.

(Currently the Dublin Core metadata syntax is not recognised by most public search engines and this situation is unlikely to change in the near future. To enable public search engines to recognise keywords and description, they need to be entered using the simpler syntax. The supplementary keyword terms may include synonyms of subject terms and creator/publisher names.)

The HealthInsite metadata harvester ignores any metadata elements which are not in the HealthInsite set of elements.

Metadata example

The document is a commissioned report, available as a PDF file.

```
<META NAME="DC.Creator" CONTENT="Balmain, Antony">
<META NAME="DC.Creator" CONTENT="Chapman, Simon">
<META NAME="DC.Publisher" CONTENT="Australian Government Department
of Health and Ageing">
<META NAME="DC.Rights" CONTENT="Copyright Commonwealth of Australia
2004">
<META NAME="DC.Title" CONTENT="Reduced-ignition propensity cigarettes: a
review of policy relevant information">
<META NAME="DC.Subject" SCHEME="Health Thesaurus" CONTENT="fires;
policy; prevention and control; smoking; tobacco">
<META NAME="DC.Description" CONTENT="The report examines policy issues
regarding reduced-ignition propensity cigarettes, which are cigarettes that have the
reduced propensity to start fires, such as domestic house fires and bush fires.">
<META NAME="DC.Language" SCHEME="RFC3066" CONTENT="en">
<META NAME="DC.Date.Created" SCHEME="ISO8601" CONTENT="2004-08-
25">
<META NAME="DC.Date.Issued" SCHEME="ISO8601" CONTENT="2005-01-
19">
<META NAME="DC.Date.Modified" SCHEME="ISO8601" CONTENT="2004-08-
25">
<META NAME="DC.Date.Review" SCHEME="ISO8601" CONTENT="">
<META NAME="DC.Date.Reviewed" SCHEME="ISO8601" CONTENT="">
<META NAME="DC.Type" SCHEME="HI type" CONTENT="document">
<META NAME="DC.Type" SCHEME="HI category" CONTENT="resource">
<META NAME="DC.Format" SCHEME="IMT" CONTENT="application/pdf">
<META NAME="DC.Format.Extent" CONTENT="334 KB">
<META NAME="DC.Identifier" SCHEME="URI"
CONTENT="http://www.health.gov.au/internet/wcms/publishing.nsf/Content/health-
pubhlth-publicat-document-smoking_rip.htm">
<META NAME="AGLS.Availability" CONTENT="Available at
http://www.health.gov.au/internet/wcms/publishing.nsf/Content/health-pubhlth-
publicat-document-smoking_rip.htm/$FILE/smoking_rip.pdf">
<META NAME="AGLS.Audience" SCHEME="HI age" CONTENT="adult">
<META NAME="HI.Complexity" CONTENT="difficult">
<META NAME="HI.Status" CONTENT="registered">
```