



ANDS Guide



Vocabularies and research data

Level: Awareness

Last updated: 3 February 2017

Web link: www.ands.org.au/guides/vocabularies-and-research-data

Overview

This guide explains what vocabularies are and how they are useful for supporting research. A brief discussion of vocabulary services is included.

What is a vocabulary?

A vocabulary sets out the common language a discipline has agreed to use to refer to concepts of interest in that discipline. It is a kind of model of the concepts in a discipline, with labels applied to the concepts and some kind of structure relating the concepts to each other. Vocabularies take many forms. They include authority files, glossaries, dictionaries, gazetteers, code lists, taxonomies, subject headings, thesauri, semantic networks and ontologies. More technically (per <http://marinemetadata.org/guides/vocabs/vocdef>), a vocabulary is a set of terms or labels (words, codes, icons) that are used in a specific community to represent concepts.

How do vocabularies support research?

Data specification and description

When sharing data or combining data from different sources, there is a need for an agreed language to make sure the meaning of data is clear and explicit.

Researchers planning observation or surveys need to define their data items clearly. In formal system development environments this is done using metadata registries, data dictionaries, or data modelling software to define the permissible values/codes for data.

An agreed vocabulary (a standard) makes a good starting point for translating concepts into other vocabularies so that collaboration can occur.

Examples of vocabularies used to specify data values:

- [Marine science vocabularies](#)
- [Darwin Core: An Evolving Community-Developed Biodiversity Data Standard](#)
- Health and welfare statistics values are defined in AIHW's [METeOR Metadata Online Registry](#)
- ABS 2011 Census Data Dictionary [Example](#)

[2011 Census Dictionary](#) >> [2011 Census Classifications](#) >> MSTP

MSTP	Registered Marital Status	First release
	Records a person's formal registered marital status. If registered marital status is not stated it is imputed. See also Imputation Flag for Registered Marital Status (IFMSTP), Social Marital Status (MDCP).	
Applicable to:	Persons aged 15 years and over	
Categories:	1 Never married 2 Widowed 3 Divorced 4 Separated 5 Married @ Not applicable	
Number of categories:	6	
Not applicable (@) category comprises:	<ul style="list-style-type: none"> • Persons aged under 15 years 	

Data analysis

Ontology-mediated data integration

In this process scientists annotate data sets with semantically precise terms from an ontology, enabling reasoning across the data and transformations of the data for further analysis.

- Case study from genomics: [Ontologies: Scientific Data Sharing Made Easy](#)
- Case study from ecoinformatics: [Ecoinformatics: supporting ecology as a data-intensive science](#)

Statistical analysis

Statistical analysis involves aggregating data and applying statistical analytical techniques. Use of standard classification schemes (a kind of vocabulary) means that data from different sources can be compared. If standard classifications are not used, it is difficult to aggregate data from different sources with a high degree of confidence.

Examples of statistical vocabularies

- [International Classification of Diseases \(ICD\)](#), used for national mortality and morbidity statistics
- [Australian and New Zealand Standard Research Classification \(ANZSRC\)](#), used for measuring and analysing research and experimental development (R&D) in Australia and New Zealand

Data retrieval

Indexing vocabularies are used to tag items in library catalogues and search portals and to provide keywords for academic journal articles. Without indexing vocabularies search precision is reduced and valuable relevant research may not be retrieved. Indexing vocabularies are most effective when they mirror the searcher's terminology and conceptual perspective.

Examples of indexing vocabularies:

- [Medical Subject Headings \(MeSH\)](#) used in the [PubMed](#) biomedical literature portal
- [Powerhouse Museum Object Name Thesaurus](#) used for indexing museum collections

Example of journal article with keywords: [Example](#)



ANIMAL GENETICS
Immunogenetics, Molecular Genetics and Functional Genomics

White spotting in the domestic cat (*Felis catus*) maps near *KIT* on feline chromosome B1

M. P. Cooper¹, N. Fretwell², S. J. Bailey² and L. A. Lyons¹

Article first published online: 23 NOV 2005
DOI: 10.1111/j.1365-2052.2005.01389.x

Issue: **Animal Genetics**
Volume 37, Issue 2, pages 163–165, April 2006

Am score 3

Additional Information (Show All)
How to Cite | Author Information | Publication History

Abstract | Article | References | Supporting Information | Cited By

View Full Article with Supporting Information (HTML) | Enhanced Article (HTML) | Get PDF (218K) | Find Text Here

Keywords:
cat; feline; *KIT*; *PDGFRA*; spotting; white

SEARCH
In this issue
Advanced > Saved Searches >

ARTICLE TOOLS
Get PDF (218K)
Save to My Profile
E-mail Link to this Article
Export Citation for this Article
Get Citation Alerts
Request Permissions
Share | Facebook | LinkedIn | Twitter

Vocabulary services

Traditionally most vocabularies were managed in custom software, and either printed or published as read-only web pages or downloadable documents (for example, see the [APAIS Thesaurus](#)).

A vocabulary service is a machine-to-machine service that can support activities such as creating, managing and querying vocabularies.

Examples of vocabulary services:

- [Linked Open Vocabularies \(LOV\)](#)
- [Library of Congress Authorities and Vocabularies service](#)
- [Research Vocabularies Australia \(RVA\)](#)

SKOS

Knowledge organisation systems such as thesauri or any other type of structured controlled vocabulary can be represented using SKOS (Simple Knowledge Organization System). SKOS provides a standard way to represent knowledge organisation systems using the Resource Description Framework (RDF). This means that vocabulary information can be passed between computer applications in an interoperable way.

Find out more

Introduction to vocabularies:

- [Marine Metadata Initiative \(MMI\)](#) — a comprehensive explanation of vocabularies and their use

Standards:

- [ANSI/NISO Z39.19](#) - Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies 2005 (revised 2010)
- ISO 25964 -1:2011 Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 1: Thesauri for information retrieval
- [SKOS](#) Simple Knowledge Organisation System
- [Resource Description Framework \(RDF\)](#)

ANDS Controlled Vocabulary service

- [Research Vocabularies Australia \(RVA\)](#)

Feedback?

We welcome your feedback on this guide. Please email contact@ands.org.au with any comments or questions.

About ANDS

The Australian National Data Service (ANDS) makes Australia’s research data assets more valuable for researchers, research institutions and the nation.

ANDS is a partnership led by Monash University in collaboration with the Australian National University (ANU) and the Commonwealth Scientific and Industrial Research Organisation (CSIRO). It is funded by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS).

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/). You are free to reuse and republish this work, or any part of it, with attribution to the Australian National Data Service (ANDS).

NCRIS
National Research
Infrastructure for Australia
An Australian Government Initiative

