Victorian Government Data Directory Metadata

Standard

This document describes the metadata standard used by the Victorian Government Data Directory.

Document Control

Applies to	Victorian Government Data Directory	**Authority**	Director – Digital Policy, Governance and Assurance
Period	2017-2020	**Advised by**	Digital Engagement and the Victorian Centre for Data Insights
Issue Date	10/09/2018	**Document ID**	IM-STD-09
Review Date	10/09/2020	**Version**	1.0

Standard

The Victorian Government Data Directory (data directory) aligns with the Australian Government Locator Service (AGLS) metadata standard as follows:

Term name	Field name	Term definition	Obligation
Title	Title	The title or name given to the data asset (or the series of data assets).	Mandatory
Description	Description	A description of the content and/or components of the data asset.	Mandatory
Subject	Keywords	Keywords or search terms which describe the topic or content of the data asset.	Mandatory
Publisher	Data Owner	The publisher is the agency; division; department or 3rd party authorised to release and/or share the data asset.	Mandatory
Creator	Data Custodian	A resource that acts or has the power to act on behalf of the Data Owner.	Mandatory
Email	Email	Address of the electronic mailbox of the responsible party.	Mandatory
License	License	License details giving official permission to use the data asset/s.	Mandatory
Protective Marking	Security Classification	The protective marking applied to describe the data asset.	Mandatory
Access Rights	Access	Access Rights may include information regarding access controls or restrictions based on privacy, security, or other policies.	Conditional
Date Created	Created	Date the data asset was created or acquired (not the date the metadata record was created).	Mandatory
URL	Data Source	The online location of the data asset.	Optional
Format	Format	The format of the data asset	Mandatory

For a more detailed description of each term and how it is used in the data directory please see Appendix 1 – Metadata Standard.

Please note this standard is not intended to imply a metadata standard for department or agency information asset registers.

Overview

The purpose of the Victorian Government Data Directory Metadata Standard (the standard) is to define a metadata schema for registering data assets in the Victorian Government Data Directory.

Rationale

Defining and communicating a metadata standard helps to establish a common meaning for the metadata used in the data directory, making it easier for government to upload or feed their information asset register metadata.

Supporting Information

Victorian Government Data Directory

The Victorian Government Data Directory[1] will provide a catalogue of metadata records describing data assets available from across government. The aim of the data directory will be to provide a means to share data underlined internally to the Victorian Government. It will enable discovery and maximise the reuse, repurpose and integration of the government's data. The outcome will be improved data insight that informs policy decisions and service delivery.

The data directory will be published via a CKAN open source repository based on AGLS compliant metadata. It will contain functionality to create and manage metadata records, manage workflows and user accounts and govern security and visibility of metadata records.

Metadata records about data assets will be able to be created/published directly in the data directory or uploaded via feeds.

Metadata

Much like a library catalogue, metadata is structured information that describes, explains, locates or otherwise makes it easier to discover, retrieve, use or manage a data asset.

The AGLS metadata standard, with a Victorian Government extension and Dublin Core inclusions, will be used in the data directory to establish a common meaning for the metadata record terms. The data directory metadata standard will be able to be mapped to via the Victorian Government API Gateway[2].

Data assets versus metadata record

A data asset is a defined body of information that is practically managed so it can be understood, shared, protected and used to its full potential. It may be an application, database, spreadsheet,

---

[1] The Victorian Government Data Directory is scheduled for delivery in 2018/19.

[2] For access to the Victorian Government's API Gateway please see the API Gateway Developer's Portal.

output of a business process, a dataset or image etc. It may be stored in a variety of locations, medias and formats. Data assets have a recognisable and manageable value, risk, content and lifecycle.

Much like a card in a library catalogue describes a book; a metadata record describes a data asset. It helps the user to understand the scope and intent of the data asset and determine its applicability for their intended use.

Geospatial data assets

Data assets that describe geospatial information should be published via Vicmap's Spatial Datamart; the authoritative spatial data for the State of Victoria.

Vicmap published metadata records are then ingested into the data directory and displayed according to their security classification and access rights.

Derivation and scope

Derivation

This standard is derived from the AGLS Metadata Standard and the DataVic Access Policy.

Scope

The Department of Premier and Cabinet and the Victorian Government Data Directory.

Related documents, tools and references

- DataVic Access Policy
- Victorian Protective Data Security Framework (VPDSF)
- WoVG Information Management Framework
- WoVG Information Management Policy
- WOVG Information Management Standard

Further information

For further information regarding this standard, please contact Digital Strategy and Transformation, Department of Premier and Cabinet, at: digital.transformation@dpc.vic.gov.au.

For further information regarding the Victorian Government Data Directory, please contact Digital Design and Innovation, Department of Premier and Cabinet, at: https://data.vic.gov.au/contact-us .

For further information regarding the Victorian Government API Gateway, please contact the API Gateway team, Department of Premier and Cabinet, at: apiteam@dpc.vic.gov.au.

Appendix 1 - Metadata Standard

Key attributes of each metadata term have been described using the following table structure:

Term attribute	Description / detail
Term name	The metadata term name as displayed in the data directory.
Term definition	Definition, intent or rationale for the term.
Field name	The human readable name provided in the interface of the directory.
Guidance	Advice and direction for the use and implementation of the term.
Obligation	Denotes whether the term is **mandatory**, **conditional** (including business rules) or **optional**.
Default value	Default value to be used if applicable (e.g. where mandatory and no other value is entered).
Encoding scheme	Reference to the defined scheme, code list or reference data to be used, if applicable.
Prescribed list of values	Specific or custom list of values to be used in the absence of an external encoding scheme.
Example	An example of how data should appear for this term.
AGLS, DC or VICGOV	XML schema metadata term

Title

Term name	Title
Term definition	The title or name given to the data asset (or the series of data assets).
Field name	Title
Guidance	The Title field provides a clear and descriptive title for the data asset that describes its business use/intent. Any alternative names for the data asset may be captured in the description field.

Where the metadata record describes a series of data assets, the title should be descriptive of the series.

The author of the metadata record should make sure the title is meaningful to the user and is keyword rich to support search. |
Obligation	Mandatory
Default value	N/A
Encoding scheme	N/A
Prescribed list of values	N/A
Example	Major Projects, Victoria Projects Register
AGLS, DC or VICGOV	dcterms:Title

Description

Term name	Description
Term definition	A description of the content and/or components of the data asset.
Field name	Description
Guidance	The Description field should describe the nature of the data asset, the program (if applicable), the data type, any relevant legislation, how the data was collected and how it is intended to be used. It should help the user to determine if the data asset is fit for their intended purpose.

Ideally commence the description with the abstract (see the Abstract |

term) and ensure it is rich in keywords to aid indexing and discovery via the search engine.

The more detailed the description the better an understanding of the data the user gains. It allows the user to know what they are getting before making the decision to access the dataset. There is no text limit in this field so the more detailed description the better.

This field may be the combined Asset Description and Purpose under the Victorian Protective Data Security Framework.

Obligation	Mandatory
Default value	N/A
Encoding scheme	N/A
Prescribed list of values	N/A
Example	Major Projects Victoria is the Victorian Government's in house project delivery agency. This register lists all current and completed Major Projects Victoria projects including address and location details.
AGLS, DC or VICGOV	dcterms:Description

Keywords

Term name	Subject
Term definition	Keywords or search terms which describe the topic or content of the data asset.
Field name	Keywords
Guidance	Along with the Description field, the Keywords field helps to describe the data asset to the user. The aim should be to include a minimum of two or three keywords (i.e. search terms or key phases) to assist users in locating the data asset and understanding its intend. Keywords drive better search discovery and also support the ability to link datasets.
Obligation	Mandatory
Default value	N/A
Encoding scheme	N/A
Prescribed list of values	N/A
Example	Major projects, Register, Project management
AGLS, DC or VICGOV	dcterms:Subject

Data Owner (Organisation)

Term name	Publisher
Term definition	The publisher is the agency; division; department or 3rd party authorised to release and/or share the data asset.
Field name	Data Owner (Organisation)
Guidance	The Data Owner field is the organisation that is the legal owner of the data asset (as defined in IM STD 03 Information Management Governance) and is responsible for publishing the asset via the data directory.

The Data Owner may choose to delegate ownership through relevant legal or regulatory instruments to an authorised role, group or organisation (the Data Custodian).

The Data Owner may be the parent organisation, but may also be the same as the Data Custodian (the sub-organisation agency). |
| Obligation | Mandatory |
| Default value | N/A |

Encoding scheme	Based on CKAN Account Permissions
Prescribed list of values	N/A
Example	Department of Health and Human Services
AGLS, DC or VICGOV	dcterms:Publisher

Data Custodian (Sub-organisation)

Term name	Creator
Term definition	A resource that acts or has the power to act on behalf of the Data Owner.
Field name	Data Custodian (Sub-organisation)
Guidance	The Data Custodian field is the organisation who has responsibility (delegated or otherwise) over the day to day management of the data asset (as defined in IM STD 03 Information Management Governance).

The Data Custodian may be the Data Owner (the parent organisation), or an agency that sits under the parent organisation.

The Data Custodian is the initial contact point for gaining access to the data asset. |
Obligation	Mandatory
Default value	Data Owner
Encoding scheme	
Prescribed list of values	
Example	Family Safety Victoria
AGLS, DC or VICGOV	agentterms:corporateName

Email

Term name	Email
Term definition	Address of the electronic mailbox of the responsible party.
Field name	Email
Guidance	The Data Custodian is responsible for the day to day management of the data asset and the contact for gaining access to the data asset/s. The Data Custodian may be the Data Owner (see Data Custodian above).

Data Custodians should ensure that email address provided is received by a role responsible for managing data asset sharing and release. |
Obligation	Mandatory - data directory only
Default value	N/A
Encoding scheme	AS 4590.1:2017 Interchange of client information, Part 1: Data elements and interchange formats
Prescribed list of values	N/A
Example	
AGLS, DC or VICGOV	agentterms:email

License

Term name	License
Term definition	License details giving official permission to use the data asset/s.
Field name	License
Guidance	The License field allows the Data Owner to select the relevant license from the Open Definition – Conformant Licenses list to be added to

	the metadata record. This list includes the suite of Creative Commons Licences (CC) as well as a number of other licenses relevant for releasing or sharing data.
	Licensing your data for sharing, reuse and repurpose (even internally to government) helps to ensure there is appropriate attribution and conditions of use are applied. Licensing your data, along with applying a protective marking, helps to ensure appropriate access permissions are met.
Obligation	Mandatory
Default value	Creative Commons Attribution 4.0 International
Encoding scheme	Open Definition – Conformant Licenses
Prescribed list of values	Creative Commons Attribution 4.0 International Creative Commons Attribution no Derivative Works 4.0 International Creative Commons Attribution Share-Alike 4.0 International Creative Commons Non-Commercial 4.0 International Creative Commons Non-Commercial No Derivatives 4.0 International Creative Commons Non-Commercial Share Alike 4.0 International GNU Free Documentation License 1.3 with no cover tests and no variant sections License Not Specified MIT license (MIT) Other (Copyright/Closed) Other (Open) The BSD License
Example	GNU Free Documentation License 1.3 with no cover tests and no variant sections
AGLS, DC or VICGOV	dcterms:License

Security Classification

Term name	Protective Marking
Term definition	The protective marking applied to describe the data asset.
Field name	Security Classification
Guidance	The Security Classification field is used to apply the protective marking to the metadata record that is linked to the business impact level (see the Victorian Protective Data Security Framework (VPDSF) for guidance) for the potential compromise of the confidentiality of the data asset.
	Metadata records describing data assets as 'Protected' or 'Unclassified: Sensitive: Vic Cabinet" are not suitable to be published on the data directory, shared with government or released to the public.
	Metadata records describing data assets as 'Unclassified' or 'Unclassified: Sensitive:..' may be published on the data directory, but access to the data asset will require approval from the Data Owner.
	Metadata records describing data assets as 'Unclassified: For Official Use Only can be published on the data directory and the data asset shared with Victorian Public Sector employees.
	Metadata records describing data assets as 'Public Domain' can be published on the data directory and released or shared
Obligation	Mandatory
Default value	Unclassified: For Office Use Only
Encoding scheme	N/A

Prescribed list of values	PROTECTIVE MARKING	RELEASE or SHARE
	Protected	Do not release/Do not share
	Unclassified : Sensitive: Vic Cabinet	Do not release/Do not share
	Unclassified: Sensitive	Approval Required
	Unclassified: Sensitive: Personal	Approval Required
	Unclassified: Sensitive: Legal	Approval Required
	Unclassified: For Official Use Only	Share
	Unclassified	Approval Required
	Public Domain	Release/Share
Example	Unclassified: Sensitive: Personal	
AGLS, DC or VICGOV	aglsterms:protectiveMarking	

Access

Term name	**Access Rights**
Term definition	Access Rights may include information regarding access controls or restrictions based on privacy, security, or other policies.
Field name	Access
Guidance	Based on the protective marking applied use the Access field helps to describe describe who can access the data asset, what purpose it can be used and what may need to occur before the information is made available.

For example, in the case of Unclassified: Sensitive: Personal the data owner may need to undertake de-identification. |
Obligation	Conditional: Unclassified: For Official Use Only
Default value	N/A
Encoding scheme	N/A
Prescribed list of values	N/A
AGLS, DC or VICGOV	dcterms:accessRights

Created (Data Asset)

Term name	**Date Created**
Term definition	Date the data asset was created or acquired (not the date the metadata record was created).
Field name	Created
Guidance	The Created (Data Asset) field is the date the data asset (rather than the metadata record) was created or acquired or first released to the public.
Obligation	Mandatory
Default value	N/A
Encoding scheme	ISO 8601 i.e. YYYY-MM-DD
Prescribed list of values	N/A
Example	2017-04-20
AGLS, DC or VICGOV	dcterms:Created

Data Source

Term name	URL
Term definition	The online location of the data asset.
Field name	Data Source
Guidance	The Data Source field provides a direct link to the data source

location. This field is only required if the data is published/available online.

Where two or more versions of the same data asset are uploaded (e.g. a time series or different formats) then the Name, Data Source and Format fields should be repeated per version.

Obligation	Optional
Default value	N/A
Encoding scheme	HTMLUniform Resource Locators
Prescribed list of values	N/A
Example	http://www.premier.vic.gov.au/images/stories/datasets/brigades.csv
AGLS, DC or VICGOV	vicgov:URL

Format

Term name	Format
Term definition	The format of the data asset.
Field name	Format
Guidance	The Format field allows you to describe the format of the data asset. The asset may be provided in multiple formats. Where two or more versions of the same data asset are uploaded (e.g. a time series or different formats) then the Name, Data Source and Format fields should be repeated per version.
Obligation	Mandatory
Default value	N/A
Encoding scheme	N/A
Prescribed list of values	*(see table below)*

FORMAT	DESCRIPTION
CSV (comma separated values)	for simple spreadsheets and simple databases. (Note: CSV files can be previewed within the Data Directory without the need to download the file, allowing end users to decide it the file is suitable for its purposes)
XML (extensible markup language)	a general purpose markup language for complex datasets, standardised by the main international standards organisation for the World Wide Web.
XBRL (extensible business reporting language)	a freely available global standard, standards based way to communicate and exchange business information between business systems.
KML (formerly Keyhole Markup Language)	an XML language focused on geographic visualization, including annotation of maps and images;
WMS (Web Map Service)	a protocol that allows georeferenced map images to be served over the web;
WFS (Web Feature Service)	allows requests for geographical features to be drawn across the web;
WCS (Web Coverage Service Interface Standard)	provides access to coverage data in forms that are useful for client side rendering, as input into

		scientific models, and for other clients; and
	ESRI Shapefile	geospatial vector data format for geographic information systems software.
	XLS and XLSX (Excel Workbook)	main spreadsheet format which holds data in worksheets, charts, and macros; and
	GTFS (General Transit Feed Specification)	defines a common format for public transportation schedules and associated geographic information.
Example	CSV (comma separated values)	
AGLS, DC or VICGOV	dcterms:format	

Document Control

Approval

This document was approved by the Director - Government ICT Strategy & Policy on 10/09/2018 and applies from the date of issue (see first page).

Version history

Version	Date	Comments
0.1	07/08/2018	First full draft
0.2	29/08/2018	Feedback received from Digital Engagement and Victorian Centre for Data Insights
0.3	30/08/2018	Further feedback received from the Chief Data Officer
1.0	10/09/2018	Final version