MDWG Definitions – Report

OpenWork Ltd

16 January 2019

Author – Byron Cochrane

Attachments – MDWG element definitions.xlsx,

Combined Metadata between profiles.xlsx tween profiles

# What we proposed:

Objective:

Provide substantive information to use for future evidence base analysis as to the differences between existing applications of the ISO 19115-1 metadata standard in organisations. This will be used to resolve differences in application of this standard and provide guidance to future use.

Tasks:

1. Insert and Populate an ISO 19115-1 definitions column on the “Metadata Mappings Between Profiles” table (as currently compiled by Irina)
2. Insert and Populate an CKAN (data.gov.au) definitions column on the “Metadata Mappings Between Profiles” table for those elements that have identified data.gov.au mappings
3. Insert and Populate an DCAT definitions column on the “Metadata Mappings Between Profiles” table for those elements that have identified DCAT mappings
4. In parallel, ask participants for best and most complete examples of metadata records from each of the 3 organisations that have submitted ISO 19115-1 metadata to the “Metadata Mappings Between Profiles” table. Confirm that these are fit for purpose of this exercise.
5. For each of the 58 elements, we will create a table containing:
   1. ISO 19115-1 element definition
   2. real examples from each of the 3 above mentioned organisations of how the element is populated
   3. a cursory “stoplight” level assessment of the level agreement between organisations on use of each element
6. Results will be compiled and shared with all participants for review and further analysis.

As our focus was on alignment of definitions to gain mutual agreement on terminology and use, we have chosen not to address obligation and cardinality issues. We have limited in this report to those elements cited in “Metadata Mappings Between Profiles\_main 2.xlsx”

## Notes and Variance from proposal

1. The ISO Definitions consist of 3 columns – one that contains the path, one that contains the ISO definitions for each node in the path, and one that creates a summary definition of the element based on its location in the element path. This was done in order to add context to the element definitions in a transparent way. These definitions are sourced from the official AS/NZS standard document “ASNZSISO19115.1-2015+A1.pdf”
2. The data.gov.au elements (CKAN) are defined by two columns sourced from the “Description” and “Vocab Control” columns in the tables at <https://toolkit.data.gov.au/index.php/Discovering_Metadata>.
3. DCAT definitions are described using the latest documentation from the W3C Dataset Exchange Working Group (DXWG) with two columns added “DCAT def” and “Notes”. These are sourced from <https://w3c.github.io/dxwg/dcat/> and <https://github.com/w3c/dxwg/blob/gh-pages/DCAT-ISO19115-mapping.xlsx> respectively
4. Due to lack of access to needed participants over the holidays, we used two rather than three ISO 19115-3 examples. (Also, it seemed unclear that AADC has yet developed records sufficient for reference.) From ABARES we used the “test.xml” that Evert supplied with his earlier documentation. For GA, we randomly chose a ISO19115-3 record from their catalogue, “Geomorphic features of the Antarctic and Southern Ocean 2012” - <http://www.ga.gov.au/metadata-gateway/metadata/record/102441>. We did, however, include example metadata from data.gov.au which provides insight to how the metadata field mappings align between the two standards.
5. In the MDWG Definitions report spreadsheet, we also included definitions supplied by ABARES and GA. These are source from documents previously supplied to MDWG, “draftDataMetadataReqV0.4.xlsx” and “GA Profile 19115\_1 0.2 draft.pdf”. **Note** - Prior to commencement of this work, Irina Bastrakova rightly pointed out that there are somewhat less that 58 named elements in the proposed profile. However, because of nested elements and multiplicity, the true number is somewhat higher than a straight count of elements.

# MDWG element definitions.xlsx Column Names

|  |  |
| --- | --- |
| Package | Copied from “Metadata Mappings Between Profiles\_main 2.xlsx” |
| Element | Copied from “Metadata Mappings Between Profiles\_main 2.xlsx” |
| Path | The full ISO nested path to the element, expressed in CSS fashion using > for descendant elements, and + for sibling elements |
| ISO definitions | String of linked definitions in alignment with the ISO nested path |
| ISO definition summary | A more human readable definition derived from the string of definitions in the ISO Definition field |
| ABARES Definition | Copied from “draftDataMetadataReqV0.4.xlsx” supplied by Evert Bleys |
| ABARES help | Copied from “draftDataMetadataReqV0.4.xlsx” supplied by Evert Bleys |
| GA Attribute Justification | Copied from “GA Profile 19115\_1 0.2 draft.pdf” |
| GA Other Uses | Copied from “GA Profile 19115\_1 0.2 draft.pdf” |
| Data.gov.au refs | Sourced from https://toolkit.data.gov.au/index.php/Discovering\_Metadata |
| Data.gov.au notes | Sourced from https://toolkit.data.gov.au/index.php/Discovering\_Metadata |
| DCAT Defs | Sourced from https://w3c.github.io/dxwg/dcat/#Property:record\_update\_date |
| DCAT Notes | Sourced from the spreadsheet DCAT-ISO19115-mapping.xlsx supplied by Nick Car |
| ABARES Example | Sourced from “test.xml” ISO 19115-3 record supplied for example by by Evert |
| GA Example | Sourced from“Geomorphic features of the Antarctic and Southern Ocean 2012” - <http://www.ga.gov.au/metadata-gateway/metadata/record/102441> |
| Data.gov.au example | Sourced from https://w3c.github.io/dxwg/dcat/#Property:record\_update\_date |
| OWL Notes | Our comments on identified issues |

Yellow highlights areas of concern

Orange highlight focus areas unaddressed in this report - keywords

# Identified issues

## **Metadata Capture – Resource information**

* Resource Status not captured by GA
* What advice (if any) should we give about browse graphics?
* Resource Point of Contact differences
  + Need best practice for use of xlinks for this
    - and other sections where xlinks are used
    - especially in relation to DCAT and linked data support and advice
  + Organisation contact, individual in org contact, individual contact – should we develop best practice recommendations?
  + GA use of ’contactinstructions = “1”’ ?
    - What does this mean?
    - Is it only internal?
    - (applies to all GA CI\_ResponsibleParty reccords in example metadata)
* Keyword section – needs further discussion and recommendations developed
  + Use Thesauri!
  + Standard and recommended thesauri – how to address in best practice
    - Which ones?
    - Common location?
    - How to use in schema?
  + GA use of “Published\_External” as a general keyword
    - What is this? Looks like a kludge to support identification of externally published records. If so, there are other tools available – especially in GeoNetwork, but also in the standard
    - Should we encourage this type of use of keywords or other fields?
    - Is there a best practice way to separate Internal data management metadata from external metadata? GeoNetwork provides options.
* Distribution
  + GA incorrectly cites “Geoscience Australia” as format distributor.
    - Should be mrd:MD\_Distributor/mrd:distributorContact/cit:CI\_Responsibility/cit:party/cit:CI\_Organisation/cit:name
    - Digital transfer options undefined in GA example metadata
* Constraints
  + ABARES cites “Creative Commons” under “mco:otherContraints” instead of mco:MD\_LegalContraints
  + A general discussion is needed about copyright, license and other legal contraints and where they apply is needed. Special guidance with the help of a lawyer to draft is recommended.
* Extents
  + ABARES record lacks gml namespace on date information on sample record
* Linage, Usage. Associated resource, Spatial Representation, Data Attribute definition
  + Not populated in either example metadata record
  + Need to review what we want to do with these in Best Practice terms

**Metadata Capture – Metadata**

* Best practice needed for identifiers and URL.
  + Should we consider recommending point of truth dereferencable URI’s
* Metadata constraints
  + Thin on metadata use constraints – is this okay?
  + ABARES -
    - Heavy reliance on generic MD\_Constraints/mco:useLimitations
    - Citation of “General Public” and “AnyPosition” under individual within organisation contact information
    - Cryptic ‘X3 “Licence type: Copyright Commonwealth of Australia 2018…”’ entry under /mco:MD\_LegalConstraints/mco:otherConstraints
    - Vague reference ‘codeListValue="copyright"’ under mco:MD\_LegalConstraints/mco:releasability/mco:MD\_Releasability/mco:disseminationConstraints
  + GA - No metadata constraint info captured in example record

**General issues**

* Confusion often occurs in what is being described
  + The Resource?
  + A Distribution of the resource?
  + The Metadata describing the resource?
* Confusion as to what the metadata describes
  + A data resource?
  + A distribution of a data resource?
* Copyrights
  + Are primarily for Distributions or Collections
  + Different distributions can have different copyright licence
  + Facts cannot be copyrighted
  + There is a difference between copyright and copyright licence
  + Creative commons is a licence on a copyright
  + End user licence agreements are not licence in themselves but are a contract containing a grant a licence (which allows more restrictions than otherwise available)
  + There are important copyright considerations inherent in provenance/lineage
  + Legal advice would be nice to gain authoritative clarity

# Also included:

Combined Metadata between profiles.xlsx t

This file addresses mapping between ABARES use of 19115-3 and GA’s use of 19115-3.

Included for reference purposes.