ANZ Metadata Working Group

Meeting 4 Report

13-14 June 2019, Canberra

# Background:

The outcomes for meeting were for all members to be better informed about activities and practices; confirm current priorities and activities; to endorse recommended workshop metadata elements; actions since the last meeting and to progress the best practice blueprint.

The MDWG ToR were reviewed. This discussion enable testing and alignment to the proposed ICSM ToR Template which is currently been rolled across all ICSM committees.

Meeting discussions assisted the ICSM Permanent Committee on Geodesy work on Modernising Australia’s Datum by provision of use cases and defining the FAQ. The meeting also instigated a request to partake in the ICSM reference group, so the MDWG can identify what advice is required to support the implementation of GDA2020 and other datums.

**Meeting summary:**

Day 1 of the meeting (13th June) focused predominately on metadata related presentations and updates. These consisted of overviews on National and International work surrounding Metadata Standards and Best Practices (e.g. ISO, OGC, W3C), examples on implementations of metadata schemas, transformations and tools from the Federal Government (e.g. AAD, ABARES), Jurisdictional Governments (e.g. QLD) and research organisations (e.g. AODN, TERN, ARDC). The day also provided an update of the activities which the MDWG Technical Sub group had undertaken since the last meeting.

Day 2 of the meeting (14th June) consisted of three workshops and a presentation about Australia’s Datum modernisation. Workshop one, developing the ‘Blueprint for the Australian Metadata Profile’ and workshop 2, ‘Metadata for Services’, provided direction for future work of the Technical Metadata Group and to resolve some questions. The third workshop looked at benefits and issues with implementation of the new version of GeoNetWork, a common metadata management and publication tool.

Actions from previous meetings.

* Most actions had been completed.
* **Action 30**: Evert believed there was not a good method for implementing 19115-2 standard. Carry forward to next meeting.
* Melanie asked what the reach of this group was, Andrew Whiting said other ICSM groups are interested in the work being done by MDWG, and we need to reach out as far as we can.

MDWG Terms of Reference (ToR)

The review of ToR need to be incorporated into new ICSM template now.

Melanie would like to broaden the purpose, to include others that may not use the ISO 19115 standards, and to make the elements interoperable. Byron said not just good standards, but best practice needed to be included.

Evert wished to include DCAT. A discussion was held about differences between definitions of spatial, location and geographic information. A spatial dataset will have a georeference, whereas Canberra is a geographic reference (descriptive, no coordinates). Spatial is component of geographic, and to specify spatial may be too limiting.

There is a recognition we might need to be across all geographic standards, however we are not responsible for them all.

**Action 01**: MDWG members will review the TORs, Groups and Duties and provide feedback to Andrew Whiting by 21 June.

ISO/TC Report - Chris Body

ISO and OGC update from Chris Body slides

* OGC – many ICSM relevant standards are in active developments; especially Transport and Land Administration/Tenure, big focus on developing Geospatial API
* Disruptive technologies like ML have high impact on geospatial standards and implementations – maintain awareness of those is important
* Looking at URI namespace conventions

Public release of ISO Geodetic Registry, Evert suggested that it might be better to keep using EPSG codes.

Update on DCAT2 & other Linked Data standards- Nick Car CSIRO

* Presentation directly to the W3C document - DCAT v2
* Spatial properties of DCATv2 are particularly relevant for MDWG, DCATv2 is aligned with 19115-1
* Semantic web tooling can provide more functionality than xml tooling
* The latest version includes some best practice guides for use
* DCATv2 looks at cataloguing resources - traditionally we have only considered catalogues
* OGC will allow incremental updates to standards, even after publishing: what is impact of this on tools we build? When building a tool we have to be aware that the world is moving on, have to ensure flexibility to incorporate such change, but to achieve a balance of stability as well
* Effect to Open source catalogues such as CKAN
* Nick is taking the CKAN export model to map it to DCAT 2

Evert Bleys:

* 19115-1 - amendments can occur within 12 months, major changes will take up to 3 years.
* Amendments
* INSPIRE have faced issues related to amendments
* We need to be careful with the tooling as the standards do change
* Need to ensure we are looking to toolsets which are able to change
	+ - Avoid fundamental changes to the engine, however
		- The Group needs to make a principle or recommendation related to this,
		- The change takes 1-3 years.

Blueprint for the Australian Metadata Profile Document - Byron Cochrane

Update on the progress with development of the Blueprint document and the issues

* Using Loomio to review the elements
* Each element has a vote
* Used a common software pattern to describe the elements
* i.e. is this for human reading or is it for machine
* Describes importance
* Recommended sub elements
* Will include FAQ's
* A number of issues were identified

Migration to the ISO19115-1 – Lessons learned - Dave Connell – AAD

Dave Connell demonstrated how he created bespoke XML for name spaces/ DIF code as part of 19115-1 migration. Lesions learnt from ISO 19115-1.

* Use many different formats to consume and publish
* Initial code was 4000 lines. Then there was the advice to run them as templates. This reduced further when Jonathan Kool rewrote some XML to include templates to reduce amount of code. Tested against GA validator ok.
* Template should be in the Lessons learnt, for others to use as they migrate to 19115-1.

ARDC – Melanie Barlow

The ARDC convened a [Data Services Interest Group](https://ardc.edu.au/resources/communities-of-practice/) around data service provision and consumption across the NCRIS facilities, science agencies, universities, and broader public sector, aiming to improve discovery and use of data and related services across these organisations.

* Recognised there was a disconnect between many catalogues
* Undertook a mapping audit to understand how all the elements match: looked at data, metadata and a combined metadata record
* Did a mind map of metadata concepts
* User guide of recommended elements for collecting service level metadata or data metadata
* Challenge: people have different needs, however this is just a general guide
* ARDC system,
* You can now search for data by specific data formats
* Harvests metadata and product end points
* Production agreement between data custodian and ARDC
* Goal is to ensure the provider has persistent identifiers
* Currently considering services which are a system rather than a product or service

New Metadata System in ABARES - Evert Bleys:

* + Using two instances of GeoNetwork in ABARES
	+ Spent a lot of time building the edit interface.
	+ Much customisation occurs to the editing interface to make it easier for the end user

AODN Portal and future plans for Metadata – Natalia Atkins IMOS

* + A lot of data providers and data collecting instruments.
	+ AODN Catalogue and 1-2-3 Portal are making data easily discoverable and accessible
	+ Based off the 19115 and was extended to have a series of elements which meet the needs of the marine community
	+ Moving towards 19115-1, need to upgrade geonetwork, and currently transitioning the marine community to the native 19115-1 profile
	+ Many complexities, as we manage, host and harvest, have many tools and catalogues
	+ **Need a guidelines and best practice guidelines document** for marine
	+ Question related to lineage and provenance - IMOS don’t see this as a major issue. NCI requires this.

Content Negotiation by Profile – Nick Car

* + Content Negotiation by Profile document describes how Internet clients may negotiate for content provided by servers according to profiles to assist with a standardised way to exchange information
	+ Profile is identified as base *specifications*, including the identification of any implementing subclasses of datatypes, semantic interpretations, vocabularies
	+ The returned content may be structured in a specific way to meet interoperability requirements of a community of practice
	+ Describing the parts of profiles and their relation to other profiles is the function of the Profiles Vocabulary
	+ The document is being edited by Nick & Rob Atkinson. Working drafts just published, this work is to describe a profile.

Implementation of metadata in TERN - Jenny Mahuika

* National infrastructure for collecting, storing and sharing data of Australian terrestrial ecosystems.
* 600 observation sites to provide data on biodiversity, land and terrain, carbon water
* Data integration and sharing
* Single entity to deliver data and informatics needs
* Data to be FAIR
* Scalable system and services
* Interoperability with other NCRIS systems
* Developing virtual labs to help users do data analytics
* Current state of TERN infrastructure
	+ Many systems, applications, standards, etc.
	+ In the process of fixing
* Updating the front end of the portal
* Metadata is collected through geonetwork and users will discover through the portal which enables data discovery and access
* Vocab borrowed from AODN, which in turn was taken from IMOS.

QSpatial – Ian Beitzel

* Large scope of data holdings - imagery, location, positioning, land admin, transport, water, place names, etc.
* Open data policy
* Data is delivered when and how users want it
* Expectation that ALL items have metadata and are discoverable, e.g. through AWS infrastructure.
* Many QLD departments hosting their data
* Used to charge for all data - now big change and all data is open
* 3 prong strategy for all data
	+ Download
	+ Connect
	+ View

Kinematic Datum – Nick Brown, GA

* Currently developing new technical datum's for Australia
* Challenge in how we deliver it to the end user
* Trying to get the key messages around what are the key changes, and how can the users apply
* There is still confusion in the user community between the different datums, Nick would like feedback to better communicate to all users.
* Discussion surrounding the differences ATRF and GDA2020.
* Why not go straight to ATRF instead of GDA2020; some people still need static datum, GDA2020 will be sufficient if sub-metre accuracy is not required.
* Will time element be part of dataset in ATRF? How will it be similar to WSG84? ITRF is updated each day, WSG updated only on 6 month cycle, only has 15 reference sites; Australia will have 109 reference sites.
* ITRF has time stamp for data ***and*** datum.

**Action 02**: Nick to supply link for research on WSG -> ITRF transformations, on ITRF website.

* Information is available on ICSM website, future comms will also be on website. Should there be links on Metadata pages to datum information. Develop FAQ around metadata & datums
* Users will need to understand the datum and the data time
* WGS 84 you need to ensure you are aware of which epoch you are within
* Metadata requirements
	+ - * + Need the dataset datum, height and time. Need a date of creation.

Byron asked when should users care? It relates to accuracy of data, can be left in GDA94 in some cases of low accuracy.

**Action 03**: Requirement to prepare guidelines of how store datum in metadata level – need to dive into feature level, not just product level.

**Action 04**: MDWG members to contribute datum use case stories to PCG.

Andrew Marshall} – Live Demo eCat – lunch time presentation

* The templates for GA are based on Geonetwork, running in Amazon, so there is direct publication.
* GA holds 20,000 records that are not published, either in draft, confidential or not fit to be published. Of the total of 50,000 records, 30,000 are published.
* Views to reduce the amount of content users have to put into the metadata.

**Action 05:** Workshop to be held regarding this at next meeting.

# Workshop 1 - Building the Blueprint for Australian Metadata Profile - Byron

General issues:

Parent metadata

* What does this mean, can it have more than parent
* What qualifies a series? - Similar issue to parent metadata. Time, spatial series, which do you choose?
* Evert - recently changed his mind regarding parent - each relates to each feature. Dataset which has a feature, which has an attribute - most don't deal at this level - is this true parent child relationship.
* There are also associations between datasets.
* It’s important we don't over use it, better to use series than parent.
* The role in the standard is to have one parent, but we can define different property statements.
* There are a lot of elements in the series, associated records and how we slice and dice is an ongoing discussion. For elements there is not a lot of guidance, it depends on 'who' and what domains, which organisation are used. Elements are subset of child, don't overuse.

Resource identifier = metadata identifier

Byron then worked through the spreadsheet to update the ‘star’ system.

* Default Locale - data identification
* Crosswalk issues
* When to extend a codelist, when not to,
* Is there overlap with vocab group? Understanding vocab terms.

Abstracts

* Good to have examples, totally free text. Agree on range of words, can be primary element of what people look/search for.
* Top level element that is harvested, first few words need to be descriptive, in common language if possible. ABARES has nice guidance. Think of audience, but hardest one to get consistency.
* Have examples of optimal and sub-optimal - subject experts assume everyone understands technical terms. ie Plantations of Australia, refer only to forest plantations, not a crop.
* GA focusing on keywords more in abstract - but not too many either. Guidelines would help. The abstract need to tell the story, but equally editors have to be careful not to distort/change meaning.
* Relationship between abstract and purpose.
* Have min & max word content on abstract.
* Should be human readable. User should be able to get what is needed from first paragraph.
* Have to get headline right, also have the services included.

Purpose:

* Shorter (generally) than abstract. Should be highly recommended field.
* Issues - no mapping to data.gov.au; maps to description in Dcat.
* Abstract is what is data about, purpose is what is it for, why it was created.

Topic category:

* You can't change it, not a code list. Very high level, after this go to keywords.

Spatial representation type:

* Can be optional, several types,

Spatial resolution:

* Spatial resolution (as on paper map) is becoming obsolete, map scale no longer as useful. Does it only apply to gridded data? Byron recommends resolution as solution
* Options are distance instead of scale. Needs to be adequately explained,
* Need to update ISO, extension to standard? Get taken forward during review
* If you have level of detail, can't have other elements, Rough guide as to what it is useful for.
* Recommend high level guide.

# Workshop 2 – Metadata for Services– Melanie Barlow

* The spreadsheet has been updated, but needs input from others to complete.
* Evert said that a service is not a distributable object.
* How to get from the data service to the network?

**Action 06**: Need to discuss further associated resource

* Andrew Whiting asked if there is a recommended way to enable application to consume from national map or GeoPortal?

The temporal extent of the dataset it services

**Action 07:** EMSINA to come and talk about machinery aspect

* How to represent data that is created on the fly? It is a service that creates a dataset.
* How to replicate such dynamic data sets for research purposes.
* How do we go forward:
	+ ARDC will review the content and make recommendations on elements
	+ Technical group meeting to discuss this further

**Action 08:** Investigate how to represent data created on fly and report back to wider MDWG.

# Workshop 3 – Under the hood of Geonetwork – Evert

* Suggest people concentrate on tabbed form, each agency will do this differently.
* Set of files that define the database access, schemas, templates,
* 4 ways to provide default values: template; clone; structure with values built-in; snippets.

|  |  |  |
| --- | --- | --- |
| **No.** | **Action:** | **Due** |
| 1 | MDWG members to review the TORs and provide feedback to Andrew Whiting by 21 June |  |
| 2 | Nick Brown to supply link for ITRF transformations, on ITRF website |  |
| 3 | Requirement to prepare guidelines of how store datum in metadata level – need to dive into feature level, not just product level. |  |
| 4 | MDWG members to contribute datum use case stories to PCG |  |
| 5 | Workshop to be held regarding Andy Marshall presentation at next meeting |  |
| 6 | Need to discuss further associated resource |  |
| 7 | EMSINA to come and talk about machinery aspect |  |
| 8 | Investigate how to represent data created on fly and report back to wider MDWG |  |
| 9  | Examples of abstract and purpose statements to be provided to Byron |  |
| 10 | People who wants to provide input to Blueprint discussions request access to Loomio from Byron. |  |
| 11 | All – Example of use cases for high-precision positioning requirements |  |
| 12 | ICSM MDWG Promotion Submit an abstract to eReserach 2019 |  |
| 13 | Additional ICSM MDWG members –send email to Irina |  |
| 14 | Finalise the element description and to prepare a draft of the Blueprint document by the next meeting |  |

Meeting participants:

|  |  |
| --- | --- |
| **Name** | **Organisation** |
| Jenny Mahuika | TERN |
| Karl Newport | DoEE |
| Leila Hernandez | Defence |
| Shanti Rowlinson | Defence |
| Marco Capobianco | GA |
| Andy Marshall | GA |
| Ian Beitzel | QLD Government |
| Ian Mullen | ABARES |
| Evert Bleys | Independent |
| Byron Cochrane | OpenWork Ltd, NZ |
| Angkana Whiley | ACT Government |
| Belinda Allison | ACT Government |
| Monika Kuppelwieser | DoEE |
| Aaron Sedgmen | GA |
| Genevieve Dwyer | NAA |
| Kathleen Keane | NAA |
| Kelsey Druken  | NCI |
| Melanie Barlow | ARDC |
| Jacquiline LeLievre | VIC Government |
| Natalia Atkins | IMOS |
| Dave Connell | AAD |
| Sarah Sanderson | NAA |
| Irina Bastrakova | GA |
| Shane Crossman | GA |
| Andrew Whiting | GA |
| Lesley Waterhouse | GA |
| Graham Logan | GA |
| Nick Car | CSIRO |
| Nick Brown | GA |