

Minutes 16-11-2004

NZGLS WORKING GROUP

Time: 9:30 am
Date: Tuesday 16 November 2004

Place Conference Gallery
Level 2
Archives New Zealand
10 Mulgrave Street, Wellington

Present: John Roberts (Chair) [Archives New Zealand]
Sara Barham (NZGLS Steward Representative) [E-government Unit, SSC]
Victoria Wray (NZGLS Steward Representative) [E-government Unit, SSC]
Tracy Powell (Thesaurus Custodian), [National Library]
Michael Brownie (Chair, Thesaurus Advisory Group) [Wellington City Council]
Derek Rayner [Ministry of Social Development]
Neil Sanson [Archives New Zealand]
Keitha Booth [E-government Unit, SSC]
Gillian Ralph [Industrial Research Limited]
Karen Rollitt [National Library]
Sarah Zimmerman [Land Information New Zealand]

By Kaylene Murdoch [E-government Unit, SSC]
Invitation: Douglas Campbell [National Library]

Minutes taken by: Neil Sanson

1. Apologies

Apologies were received from Janine Delaney.

2. Review of Action Points from previous meeting

1. John to talk to Warren Tucker (Chair e-GIF Management Committee) re a local government sector replacement for Tony, ideally to sit on both e-GIF Management Committee and NZGLS Working Group.
Currently under action. Sara and John noted Michael's local government role eased the urgency. Michael advised Local Government New Zealand is currently working through a process to identify a suitable participant.
Action: To be carried over.

2. John to write to Tony accepting his resignation with regret.
Letter of thanks sent.
3. John to follow up the status of the DC translation into Maori with Karen Rollitt.
To be discussed under Item 9.
4. Everyone to carry over discussion about Contributor roles electronically.
This is being partly overtaken by a change in approach by Dublin Core Metadata Initiative, from adopting terms and incorporating them in a DC published vocabulary, to endorsing vocabularies/ encoding schemes that are developed and maintained by other agencies. For contributor roles MARC Relator codes will be maintained by the Library of Congress [USA].
5. Victoria to talk to Karen about application of suggested Contributor list to portal.
This is still being discussed within EGU.
Action: To be carried over.
6. Trudy to provide some suggestions to John about how the NZGLS WG may exchange thinking with other committees.
No suggestions received to date.
Action: Sara to encourage activity.
7. Victoria to check with Ros Coote whether shared workspace uses NZGLS for their document profiles.
The only common field is "Title", but further consideration is being given to appropriate metadata for further initiatives.
8. Trudy to invite Michael to the next xNAL (business processes) workshop.
Michael advised that he has been invited.
9. John to check what statements the Web Guidelines make about the deployment of NZGLS metadata.
Neil and John checked the web guidelines and confirm that they support the current model.
10. John to send Trudy email discussions re encoding schemes.
Sent.
11. Neil to summarise the WG discussions on encoding schemes for a submission to Trudy's consultation paper.
Submission to be discussed under Agenda Item 8.

[The order of agenda items was changed to delay discussion of some items until the arrival of Douglas Campbell.]

3. E-government Unit Update [Agenda Item 5]

John noted that Keitha has changed roles and is now the EGU Relationship and Communication Manager.

Sara distributed copies of Ferry Hendriks's "Metalogue 4 High Level Diagram" which identifies the software used for Metalogue, and "Metalogue 4 XML Export Interface" which specifies how to obtain the XML extract from Metalogue. Sara then commented that this version of Metalogue (the 3rd user implementation of the application) is much easier for E-government Unit staff to manage and for agency metadata creators to use than previous versions. Release of version 2.1 of the Portal will occur as soon as the transfer to the new server is completed satisfactorily.

[Kaylene arrived]

EGU work plan is under development.

The Government Directory project is progressing. This combines information held by Archives, Justice and SSC. John commented that establishing where these agencies sourced their information from had been a valuable exercise. Archives have signed an SLA to interface with this Directory.

Kaylene is finalising a scope for a review of the portal strategy. EGU is developing a consultation process for this review, timed for early 2005. Focus will be on how well does the Portal achieve its goals? Intention is to supply agencies with information on what the Portal does for them, to inform their responses. The review team is likely to request opinions from the NZGLS Working Group on portal strategies that have been identified as not working.

4. "NZGLS from an implementer's point of view" Kaylene Murdoch [Agenda Item 3]

Considering the issues that EGU have had in using NZGLS raises the question – has the Portal team been too literal in their interpretation of the Standard?

Metalogue is now in its 4th version (3 versions launched to users) giving a history of managing metadata to work for the Portal. EGU have developed their own data model and now treat NZGLS as an export strategy.

In the EGU data model there are three types of data : free text, CVL values, and relationships.

Relationships are with other records. *Creator*, *Availability* and *Mandate* are treated as relationships. So a service record has a 'created by' relationship with an agency. *Availability* is represented by a relationship – EGU have created a new record type "Agency Unit" – to help normalize the data.

Mandate does not use a CVL as the Portal team wanted to use URI, and as a design criteria avoid complex encoding schemes.

No record is referenced by name – references are by URN. This allows record names to be changed.

The Portal team may introduce other record types and relationship types. For example: "Minister".

Export from Metalogue is done as NZGLS metadata in XML, but concerned that there is no DTD (Document Type Definition), and therefore the XML cannot be readily validated, and there is a mix of encoding schemes.

Possible conflict with other sources of NZGLS metadata – which version is current?

Portlets will not be able to link to legislation.

Portal is likely to emphasise Portlets as the teams which support them have better subject understanding than the central Portal team can have.

John interpreted the DC view of the organization of metadata and related it to the view developed by the Portal team. John noted that the DC approach to metadata is to have simple records which list attributes for each resource. (This is essentially a ‘flat file’ database approach). However, the values of DC elements are themselves resources, and linking resources in this way (a relational database approach) is just a different way of modeling the same aspects of the real world.

Portal metadata still can be presented as NZGLS metadata indicating the abstract model of each is congruent.

John asked whether the NZGLS Working Group needs to review the obligations on elements for the different types – are they appropriate in each case? NZGLS was developed with three types of resource (Agency, Service, Document) in mind. If additional resource types are identified (e.g. legislation) the obligations and application of the elements may need to be modified for these new resource types.

Kaylene posed the question of how much of the structure of Government could the Portal usefully represent? For example – Ministers relationships with Agencies?

Derek suggested that Ministers should be described something like institutions - not as individual people who might hold the portfolio at any one time. So it could be another type of Agency record or a separate category.

John suggested that there might be needs for more types of resource. He noted that more complexity adds more power to the metadata model. The simplest approach is to have values as free text, but this can give rise to inconsistency and is difficult to maintain. Use of a CVL addresses these issues, but representing the value as a related resource provides the highest level of benefit. For example, a Minister could be free text, selected from a pick list, or from a list of related metadata records which describe Ministers.

Kaylene commented that the Portal may add more relationships. With any aspect the question is it an attribute, or is it a relationship needs to be considered. Often the aspect in question can be shown both ways, so the question is which is the better approach?

Derek asked if legislation is described. Kaylene answered that so far just links are provided (via ‘dummy’ records that do not include *Descriptions*, so are not valid NZGLS metadata).

[Douglas Campbell arrived]

Kaylene noted that the metadata is strong on functional context / relationship information, and especially for services this information cannot be found using other search tools.

Derek suggested that legislation should be described using metadata, then other records can just reference that metadata record. Kaylene thought this would mean using Metalogue to manage the legislation documents.

Kaylene summarized the Portal's XML problems for Douglas's information – primarily the lack of DTDs means the XML is fragile.

Douglas commented that the XML and RDF/XML advice from DC are very similar, but that RDF has a steep learning curve as it requires a different way of thinking about knowledge. Douglas advised to focus on creating the XML schema as that is reasonably stable.

Keitha asked about importing. Kaylene advised that importing metadata into Metalogue is not currently planned for. The Portal Team wish to limit the number of records to the most useful only.

John asked what new types of resource are being currently in place or under consideration? Kaylene noted legislation, and Ministers /MPs (Kaylene had found at least 3 lists of Ministers/MPs on the web, each presumably separately maintained.

Criteria for assessing the need for a new type are:

- public expectations of the portal
- can they be used to show relationships?
- Ministers/MPs could be described as roles held by a person, the natural persons who have those roles would not be described.
- How to describe the transition of Bills to Acts is an issue.

John asked Douglas for lessons learned on XML imports as a result of experience with Matapihi. Matapihi provides a search spanning databases of digital collections from different agencies. Examples of:

- different uses / interpretations of metadata between libraries /other institutions e.g. how to describe a fossil – no *Title* or *Creator*.
- users interpretations – a house built in the 1960s in a 1930's style is seen by architects as having a 1930s creation date (i.e. the date of the creation of the style).
- data quality / consistency problems – different date formats, different degrees of specificity i.e. 1984 vs. 1980's.

Do relationships become entities that need to be modeled/ described, or is a CVL of relationships enough?

Action: Kaylene to discuss XML schema with Douglas.

5. DC Conference report / DCMI affiliate update – John [Agenda Item 4]

John, Douglas and Karen reported back on the recent DC-2004 Meeting in Shanghai. Points noted from the conference were:

Rise of RDF and the semantic web – RDF and OWL standards stabilised by W3C, which is likely to see more extensive development of applications. For example, the “Haystack” software from IBM/ MIT – this browser tool was described as being to the Semantic Web as Mosaic was to the Internet.

Partly as a result of Semantic Web considerations, more complex understandings of metadata are becoming prevalent. As well as describing documents and their content, there is interest in describing the rich network of context in which information resources exist. Similarly, focus is shifting from values (e.g. controlled terms) to interoperability of the concepts represented by these labels (e.g. interoperability of meaning across time and across language communities). Karen commented that metadata registries are being used to cope with multiple languages and translation.

In an effort to provide scalable solutions, there is thought being given to where metadata can be found, and how it could be automatically generated. Can metadata be generated from how people interact with a system? Metadata needs to be invisible to most users, enabling functionality that people want, rather than being presented in its raw form.

Some of the key areas of work at present are:

- mapping metadata schemas;
- effective use of CVLs – referencing, maintenance, sharing.

There is a greater involvement from the recordkeeping community in DC than previously, and increased recognition in the DC world of developments occurring in archives and recordkeeping thinking.

The concept of **Application Profiles** to represent the way in which implementations may draw from multiple element sets and refine the application of terms, is well established now. A profile for libraries DC Lib is almost finalised, and one for education DC Ed is to be developed. In the DC-Government Working Group effort is currently directed to service description, and exploring differences between web-services and off-line services.

The **Affiliate** model is slowly maturing. The UK has now joined as an affiliate, along with Finland. The UK model sees affiliation by a consortium of parties. DCMI no longer seem to be looking for a single agency that will lead in a country, but for a consortium that will endorse DC to the wider community and connect to ‘flax-roots’ developments. John suggested an Archives/National Library/EGU group could fill this role in NZ. Efforts to progress this are continuing.

DC conference papers are online. [<http://dc2004.library.sh.cn/english/prog/> or <http://www.siderean.com/dcconf/search.jsp>]. Karen recommended the paper by Thomas Baker “Maintaining a Vocabulary: Practices, Policies, and Models Around Dublin Core” [http://students.washington.edu/jtennis/dcconf/Paper_01.pdf]

Dublin Core is increasingly being referred to as “integration metadata”, providing a lingua franca for exchange and interoperability. DC conference is becoming more of a meeting place for groups who want to inter-operate. People from W3C, and from SWAD-Europe were in attendance and IEEE-LOM co-located their meeting with DC for the second time, and may continue to do so.

There were around 180 delegates, including 60-70 Chinese delegates. Papers were presented in English with simultaneous translation.

6. Sector Updates

6.1. New Zealand Geospatial Metadata Standard (NZGMS) Update - Sarah Zimmerman

LINZ are seeking funding for a scoping study (timetabled for Dec 2004-June 2005) for a geospatial portal.

Geospatial Metadata Standard is to be sent to EGU for consideration as an E-GIF standard in February 2005.

Consultation will soon be undertaken to develop a geospatial strategy for presentation to Ministers in March 2005.

6.2. Education Sector – John Roberts

There are two metadata schema being developed in the Education sector – one from the chief executives’ ICT committee, the other internal to the Ministry of Education. Input to NZGLS from the education sector has been requested. This will depend on how work progresses in the Education sector initiatives.

7. Service Description

A paper was sent out to DC-Gov before the conference in Shanghai. There was a generally positive response which exposed additional relevant work from Canada and the DC Collection Description Group (around the location of the collection versus the location of the service).

Note, two versions of the document exist – one with full NZGLS context comments, and a trimmed down DC – relevant version. The DC version will be progressed through DC-Gov. The next step is for the document to be posted to DC-General for comment.

Action: Neil to copy DC version of “Using DC For Service Description” to Working Group.

8. Encoding Schemes

8.1. Controlled Value Lists (CVL)

Controlled Value Lists (CVL) were discussed with Trudy at the previous NZGLS WG meeting. Since then John has shared with Trudy the email discussion the WG has had and sent a draft response to the SSC (EGU) Discussion Paper “Standardising CVLs (Controlled Value/Code/Vocabulary Lists) for improved data exchange quality and XML schema re-use”. Considerations raised include:

- What types of encoding schemes are included in the definition – for example, syntax (e.g. language tags)
- XML
- Importance of existing CVLs / data dictionaries

- CVL criteria
- Intellectual property issues
- Process – e.g. peer review, development process
- Completeness for purpose
- Modelling
- Versioning

Action: Neil to copy the draft “Response to the SSC (EGU) Discussion Paper Standardising CVLs (Controlled Value/Code/Vocabulary Lists) for improved data exchange quality and XML schema re-use” to Working Group.

8.2. Mandate

John discussed issues raised in a draft paper provided to the WG “Mandate – encoding schemes and refinements”. The primary discussion was around the current refinements of *Mandate* (act, regulation, rule, case). John pointed out that the refinements do not cover all the types of legal instruments that provide mandates, and as refinements may not be very useful.

Kaylene commented that the refinements as such are not needed for the Portal architecture. The type of the legislation will be included in the metadata record for that resource, not in the relationship (i.e. not in *Mandate* as currently used by the Portal).

Neil clarified that that suggestion to depreciate the refinements could be seen as applying Dublin Core philosophy / guidance of avoiding using refinements where they were not required.

The Working Group agreed to depreciate the refinements in favour of using a CVL of legal instrument types and formalised naming schema.

Action: Neil to send copy draft “Mandate – encoding schemes and refinements” to Pleasance Purser (Parliamentary Library) and Michelle Edgerley for comment.

9. Maori translation of the Dublin Core Metadata Element Set – Karen Rollitt

A Maori translation was first made available in Te Kete Ipurangi. DCMI registry wanted an authoritative version and the process of getting endorsement took until September 2004. Since then the user interface has been translated and put into DC registry. This is mirrored at National Library. Users will be able to use this to select the ‘correct’ Maori term.

The next stage is qualifiers.

Subsequent work may be undertaken for Pasifika languages – particularly Cook Island Maori, but perhaps also Samoan? Samoan is less likely, in part, because DCMI prefers a distributed model – each country maintaining its own languages.

Action: Karen to forward paper for circulation to the WG.

10. Future meeting dates

Wednesday 16 February 2005 (Gallery)

Wednesday 18 May 2005 (Gallery)

Meeting concluded 12.45 pm

11. Schedule of Action Points

1. John to talk to Warren Tucker (Chair e-GIF Management Committee) re a local government sector replacement for Tony, ideally to sit on both e-GIF Management Committee and NZGLS Working Group. *[Carried over]*
2. Victoria to talk to Karen about application of suggested Contributor list to portal. *[Carried over]*
3. Trudy to provide some suggestions to John about how the NZGLS WG may exchange thinking with other committees.
No suggestions received to date. *[Carried over]* Sara to encourage activity.
4. Kaylene to discuss XML schema with Douglas.
5. Neil to copy "Using DC For Service Description" to Working Group.
6. Neil to copy the draft "Response to the SSC (EGU) Discussion Paper Standardising CVLs (Controlled Value/Code/Vocabulary Lists) for improved data exchange quality and XML schema re-use" to Working Group.
7. Neil to send copy draft "Mandate – encoding schemes and refinements" to Pleasance Purser (Parliamentary Library) and Michelle Edgerly for comment.
8. Karen to forward [Maori language] paper for circulation to the WG.