VCOI Call Notes

*23 November 2020*

# Attendees

* Mike
* Rob
* Claude

# Agenda

As usual we will pick up where we left off. In addition, we need to think about what and where to deliver and publicize our output, including whether to approach one or another of the candidate TFs or SIGs (collectively, 'Sub-groups') about presenting at the December Quarterly Meeting.

The next stuff, picking up where we left off, is more detail on the metadata, along with an understanding of the ways we would need to cite the origin or a term, definition, reference or other ARIAS, so we define the metadata to support that. Again we are using recent Blockchain PSIG work as input to this process.

# Meeting Notes

## Where do we Go from Here?

Some open issues from last week e.g.

* How to reference to 2 kinds of Context
* Detail on the metadata

### To get there:

Identify what we were trying to do and how we extract from this body of theory, what we want people to do in practice in the appropriate sub-groups (‘SG’s i.e. Task Force or SIG).

For example, if you are in a given community and you have a certain vocab and a case for defining a vocab e.g. for ambiguity, overlaps, different constituencies etc., then you have to start explaining why they should be interested in this work.

Then pragmatically, outline the steps they need to follow in order to arrive at the desired situation.

That is whereby they have:

* Rigorously defined vocab
	+ covering the work in that sub-group
* Is human and / or machine readable
	+ Hence link to MVF for machine readable vocabulary
	+ Also machine readable Taxonomy or Ontology formal artifacts
	+ Other things you might want to draw into a KG
* You have certain constructs for articulate the ARIAS, namely
	+ Word link to Concept (per usage context)
		- This Sub-group
		- This Document (controlled by this Sub-group)
	+ Concept library
		- Definitions (per Concept)
	+ Relationships
		- IsA
		- Other (broader / narrower) e.g. aggregations (mereology)
* Determining what is suitable for that sub-group’s purposes
	+ Menu including
		- SKOS
		- OWL
		- MVF
		- Other artifacts
* Method for gaining consensus on:
	+ Concepts to use for Words
		- Including definitions
	+ Sources for concepts (local versus other SGs or CoPs or government etc. e.g. Fed Register)

### To achieve this:

Some kind of ‘cook book’

* Not needing understanding of the theory
* Here’s what you do and in what order

The above set of bullets sets out the scope of the Cook Book.

#### Also:

Short list of references to the applicable standards

* SKOS
* ISO 1087
* ISO other one
* OWL

Further Reading

* The Accidental Taxonomist (Heather Heddon)
	+ Recommendation

We set out a set of recommendations that are:

* consistent with what we are propagating
* We explain how to use what’s in that reference in line with this overall process

### Some Choices

**Option 1:** Give them a range of possible tools and techniques

e.g.

* They might just want to use an indented spreadsheet
* We might show how to ingest and emit via a given tool, for a given standards-based serialization

Identify gotchas, selection criteria for using a given tool v a given technique

Need to segregate *Tools* v *Techniques* v standard *Formats*

*Format* – standard e.g. SKOS, RDF.

*Tool* – something that can ingest and emit some standard format (e.g. Protégé for OWL)

**Option 2:** Define the overall architecture and define how it’s used

e.g. with a SKOS template they can re-use

**Option 3:** Define an end point based on a single architecture but also define a path from here (e.g. spreadsheets) to there.

Then people can climb up that path in their own time, and not go off on a tangent.

The spreadsheet idea is v portable.

We can do what MB did with the DebtTerms06.xls that was done for the FISD.

This uses indentation with specific semantics.

On this: We would define a precise format of spreadsheet (template) for this.

CB has also done this successfully in Oil and Gas.

Were then able to load that spreadsheet into a suitable tool (SharePoint) for more sophisticated tool support

Note that most tools support CSV input e.g. Solidatus support. As does ANZO - has a native spreadsheet format for formal OWL based ontology.

Everyone uses these in a coherent way. We can define a format, but would also need to have transformations for this for the precise CSV format for a given toolset.

Also includes columns for e.g. synonyms.

Some tools would even aim to recognize the intended semantics of columns in existing CSV.

People have to be disciplined when they create their spreadsheet.

**Proposal:** We do that discipline for them. We define a specific Template that embodies all the requirements so that these can easily be reordered for specific tools and formats (e.g. SKOS).

**Consensus:** Agreed

Also the Cook Book has as an appendix, an example of the usage of the spreadsheet Template.

* Agreed we must have an example

Use a simple example like the Wine ontology. Show how that would be ingested into a variety of tools (including as a minimum, RDF/OWL and SKOS

Important consideration reuse by one SG of stuff from another SG. Need to have a ‘doability index’ – is it doable, is it consumable by the other SIGs and TFs that might pick up material in that resource?

Examples exist in the XBRL world, where content is exchanged among groups using CSV as a medium of exchange for content for XBRL Taxonomies.

### Next Steps

How to proceed – what’s next?

Do a prototype based on the 2 things we have:

* GovDTF Pilot XLS
* Blockchain PSIG example

Then do a separate Wine one to see how it stacks up.

#### Suggestions

Using the BC-PSIG stuff might be a better motivator for others to understand what we are doing (as well as having stuff they can already use). Closer to home than the Wine one. Will help in adoption.

Hold off the wine one for later once this is more locked in, as a back-up example.

Another quick win: FIBO

* Transform from the existing XLS format (with ‘isa’ as a term type)
* Transform into the multi-indentation format (MIF).

### Indented Spreadsheet Issues

MIF weakness is you have a fixed number of indentations (e.g. 7).

Within the context of particular usage (SG or Document) you don’t need a full TLO Taxonomy even if one exists. So 6 or 7 should be fine.

(some debt instruments in FIBO might go deeper – can do a quick check on that).

e.g. CDO Squared. Whatever’s needed for that should be enough.

If e.g. you had a term like Contract as the ancestor of e.g. Debt Instrument, then Contract would have its own entry (possibly from a CoP like legal / contractual, or an ontology like LKIF). That would have its own entry, probably on a separate spreadsheet, with its own ancestry.

**Rule:** Track ancestry locally only as far as where the ancestor term is defined in its OWN variant of the same indented CSV structure.

**Rule:** Don’t replicate other indented sets of terms.

### Next Steps

Collate the 2 XLS resources we have

Design the indented version

Combines: Generalization hierarchy and the column requirements for T&D.

Can go into RDF or SKOS.

For SKOS: we should define the sub properties of broaderThan and narrowerThan (broaderTransitive) and define how to treat of the Generalization relation

In the CSV: Always reflect only Generalization as the relation represented by the indentation

* Use another kind of feature for object properties including those OPs that are parthood, containment, etc. etc. that CAN (in some accounts) be broader / narrower concepts
* Add some narrative / explanation on how to recognize that a relation is a Generalization relation.

## Presenting at December QM

Ontology PSIG – ask Evan Wallace for a short time slot.

**Action:** MB to mail EW on this

FDTF? Agenda awaiting 2 possible talk from Lars, that would close out the time. If not heard from could present at FDTF also.

BC-PSIG – present in the context of the current ARIAS in the RFC and RFP documents.

**Action:** Add to Blockchain PSIG

GovDTF: ask Lars

**Action:** Email LT about this.

## Next Meeting

We will meet on 30 Nov (last call before OMG Week).

**Agenda:** present the draft spreadsheet formats and proposed presentations to the SGs.

## AoB?

Nope

# Abbreviations, Terms and Definitions (!)

**ARIAS**: Abbreviations, References, Initialisms, Acronyms, and Symbols

**MIF:** multi-indentation format (spreadsheet format)

**MVF:** Metadata Vocabulary Facility (OMG proposed specification)

**SG:** Sub-0group = Task Force or SIG

**SKOS**: Simple Knowledge Organization System (vocabulary standard)