Wednesday June 16 - Breakout Sessions
Flexible Metadata/Controlled Vocabularies

Please add your questions and comments below. We also ask you to put your name and affiliation next to the question.

Which controlled vocabularies will be supported? (Philipp Conzett, UiT/DataverseNO)

- The design is open/configurable. At present (as in the demo as well) we have support for vocabularies coming from skosmos, ORCID PIDs, and some ARDC Field of Research vocabularies. Adding new vocabularies from skosmos involves creating appropriate json configuration information. Supporting other PIDs/vocabulary services involves developing a Javascript as well. Once we have the basic PR in place, we hope the community will quickly add more options.

For which fields will external controlled vocabulary support be possible? For Topic Classification and Keywords? Or also for self defined fields? (Dorothea Iglezakis, Uo Stuttgart/DaRUS)

- The design as demoed supports using a single field (which stores the term URI) or a combination field where one of the child fields is the term URI. Unfortunately, the existing Keywords and Topic fields only include vocabulary name/uri and term name fields. Part of the proposal to support them will be to add a fourth child field that covers the term URI. With that change in the citation metadata block (which should be backward compatible in that existing entries with an empty term URI wouldn't break things), one or more vocabularies from a single service could be connected to these fields.

Can the metadata to be used change in the future? (Rafael Castillo, UChile/Dataverse yes)

- The association of vocabularies with given metadata fields is configurable, so in that sense, one can change which vocabularies are allowed over time. Further, since Dataverse is only storing the term URI as a string, one can turn vocabulary support on/off for a given field without breaking anything.

How feasible is it to incorporate machine learning to improve description, versus controlled vocabulary that is more static? (Rafael Castillo, UChile/Dataverse yes)

- How selections are made is completely up to the Javascript involved, so how sophisticated that is is up to whoever creates it. For example, as I hope to show in the demo, since ORCID indexes many things from your profile, a search for 'Slava CoronaWhy' finds the ORCID for the Slava who's involved in the CoronaWhy project. Javascripts could use machine learning, use skos relationships to guide you to narrower/broader terms, etc.

For "Subject" Field and associated metadata in DV, would it be possible to work on a multilingual list: the current one implemented in DV is in English, a French one has been translated by Scholars Portal Dataverse, what about other languages? (Marie-Claude Deboin, CIRAD/Dataverse yes)
Can we propose some alignments with other national or/and international Subjet Classifications?

- When used with a single Dataverse metadata field, Javascripts can take input and display fields in any language that the underlying service supports. For skosmos, and any service that provides multi-language labels in a format Dataverse can parse, the current code also indexes the multiple languages for search and includes them in metadata exports. (When used with a compound field, the current code only displays/indexes/exports the language in which the information was entered. It probably makes more sense to migrate such fields to just using a single field over time, but it is possible that we could extend the code to support multiple language in compound fields instead.)

For “Topic Classification” and “Keyword” Fields, are some controlled vocabularies recommended and already implemented in some Dataverses? (Marie-Claude Deboin, CIRAD/Dataverse yes)

- As far as I know, there are no recommendations yet although I think it would be easy to collect current practices. (For example, QDR uses ICPSR’s keyword terms). This would be a great community activity.

For “Geographical Metadata”, should some controlled vocabularies be recommended such as ISO 3166-1 for “Countries” and “Nations” ; Geonames for « Province », « City », « Other »? (Marie-Claude Deboin, CIRAD/Dataverse yes)

- These would be great targets - gazetteers. The compound field support might be very useful here (i.e. selecting a place would fill out the country/province/city fields)

Would a webservice look-up be possible/considered for the roadmap. Instead of managing vocabs (and authorities) in Dataverse, it’s more efficient to call upon either LOD vocabs via API or own managed vocabs in specialised vocabulary and authority management tools. (Roxanne Wyns, KU Leuven/Dataverse set-up in progress/planned 1st production release end of 2021)

- ~yes. There are probably some complexities in doing this, but it would be a great next step. Until then, sharing the json config objects required to set up a vocabulary for a given field should help make setup a matter of cut/paste.

SKOS allows for improved find-ability by using multi lingual terms or concept labels and also supports hierarchy, with broader and narrower concepts.
Are there plans to somehow leverage this SKOS information?

- I’m (Jim) not aware of what plans may exist, but this type of thing is definitely possible and should be able to be done just by changing the Javascript (no changes to Dataverse).