

Adaptation of DSpace to the Specific Needs of the Agricultural Sciences and Technology Community

Imma Subirats¹, Andrew Bagdanov¹, ARD Prasad² and Johannes Keizer¹

{Imma.Subirats, Andrew.Bagdanov, Johannes.Keizer}@fao.org

ard@drtc.isibang.ac.in

The aim of this project is to explore the customization of DSpace according to the use of controlled vocabularies. The plug-in shall be easy – to – install modules available free for download on the DSpace site. The objective is the adaptation of DSpace to the specific needs of the Agricultural Sciences and Technology community in order to assure quality in metadata creation. Knowledge Organization Systems such as the AGROVOC thesaurus provide mechanisms for sharing information in a standardized manner by recommending the use of common semantics.

AGROVOC¹ is a multilingual structured thesaurus of all subject fields in Agriculture, Forestry, Fisheries, Food security and related domains (e.g. Sustainable Development, Nutrition, etc). Its main role is to standardize the indexing process in order to make searching simpler and more efficient, and to provide users with the most relevant resources. AGROVOC is developed by FAO².

DSpace is OAI compliant and thus bibliographic records can be easily harvested by service providers for inclusion into their systems. Bibliographic resources are described with qualified Dublin Core (DC) metadata. By default only 3 fields (title, language and date) are mandatory. However, DSpace can be customized to use other application profiles for metadata like the AGRIS AP used for the description of documents in the agricultural domain. This allows the use of more sophisticated and specialized metadata. Amongst other fields there is the subject field, where it is possible to describe the resource with one or more subject keywords and choose a vocabulary, where this keyword has been taken from, for example the AGROVOC thesaurus. However, in DSpace, currently there is no control on the subject keywords, i.e. a user can type in any keyword freely and no check against a controlled vocabulary is executed. This leaves too much freedom in the choice of subject keywords and thus inconsistent use, which will eventually make interoperability on the subject level difficult.

The use of AGROVOC in DSpace will have the following direct advantages:

- use of advanced knowledge organization (thesauri)
- to provide an openSource tool for data and metadata management for the agricultural community and especially the AGRIS network.

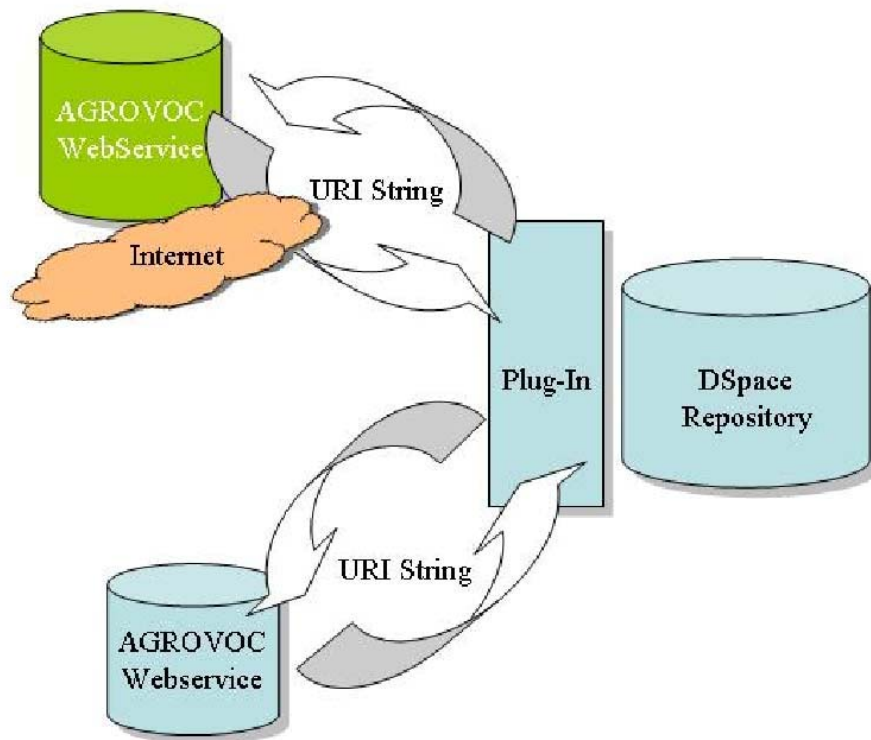
The first aim is to be able to use the AGROVOC thesaurus in DSpace in order to be able to pick and assign keywords from AGROVOC to describe agricultural resources. However, since DSpace is not domain specific, and since there are numerous agricultural terminology systems, the plug-in shall be able to import and use any vocabulary and terminology system provided in standard SKOS or OWL formats. The plug-in is needed on two levels: to support the data entry process and to support browsing by subject.

1. Language-neutral metadata representation

A key feature of the plugin will be the representation of terms from multilingual controlled vocabularies, e.g. AGROVOC, as URIs instead of only their textual representation. The URI can be localized in any supported language. Further, the URI representation of subject keywords can be localized when indexing to automatically provide multilingual search capabilities “for free.” Our initial belief is that the best way to implement these features will be to store *both URIs and selected localized textual representations in the metadata records*. Administrators should be able to select which languages are used for localizing textual representation for

¹ Agricultural Information Management (AIMS) Web site. Retrieved, March 30, 2008, from <http://www.fao.org/aims/>

² Lauser, Boris, Margherita Sini, Gauri Salokhe, Johannes Keizer, Stephen Katz (2006) Agrovoc Web Services: Improved, real-time access to an agricultural thesaurus. *IAALD Quarterly Bulletin*, 2. Retrieved, March 30, 2008, from <ftp://ftp.fao.org/docrep/fao/009/ah767e/ah767e00.pdf>



metadata storage and indexing. The URI will be used for localization and the selected textual localizations for indexing. This strategy should limit the amount of customization required for the DSpace indexing module.

Figure 1. Architecture

The plugin shall facilitate the use of language-neutral URIs representing subject keywords. The user shall not be aware of the fact that URIs instead of terms are being used for indexing, and the internal representation of keyword metadata must support on-the-fly localization of URIs in any configured language. Localization must be supported during metadata entry, metadata display and indexing.

When metadata records containing entries from controlled vocabularies are displayed, these entries shall be localized in the user's language whenever possible. There shall be a configuration option allowing the user to set his or her preferred languages for metadata display. Terms can be localized using the internal language-neutral URI representation and the local/remote webservices of the multilingual controlled vocabulary.

Similar to metadata display, on export all metadata fields coming from multilingual controlled vocabularies shall be localized in the languages selected by the user. The system shall provide a configuration option allowing the user to choose which languages are used for localizing controlled metadata entries. Localization of controlled metadata terms in one or more languages shall be supported also for OAI-PMH services as well.

2. Interfacing

The metadata entry module of DSpace shall be interfaced with the multilingual thesaurus browser so as to allow users to enter one or more terms from controlled vocabularies using the multilingual thesaurus browsing functionality described above.

The plugin shall provide users with the ability to browse metadata collection using the multilingual controlled vocabulary browsing functionality described above. Like the existing “Browse by Subject” functionality in the core DSpace application, records annotated with selected terms in the browser will be displayed.