Implementation of Rich Metadata Formats and Semantic Tools using DSpace

Imma Subirats, ARD Prasad, Andrew Bagdanov and Johannes Keizer

Food and Agriculture Organization of the United Nations, FAO (Rome, Italy) and Indian Statistical Institute (Bangalore, India)

AGROVOC is a multilingual, structured and controlled vocabulary designed to cover the terminology of all subject fields in agriculture, forestry, fisheries, food and related domains (e.g. environment), and currently contains over 29 000 terms. **Available in 17 languages:** *Arabic, Chinese, Czech, English, Farsi, French, German, Hindi, Hungarian, Italian, Japanese, Lao, Polish, Portuguese, Slovak, Spanish, Thai.*

Dspace was created as digital repository with the objective of storing, indexing, preserving and disseminating an organization's research material in digital formats. MIT and Hewlett Packard Company designed the system between 2000 and 2002. **DSpace** *supports the Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH) v2.0 as a data provider.*

http://www.dspace.org/
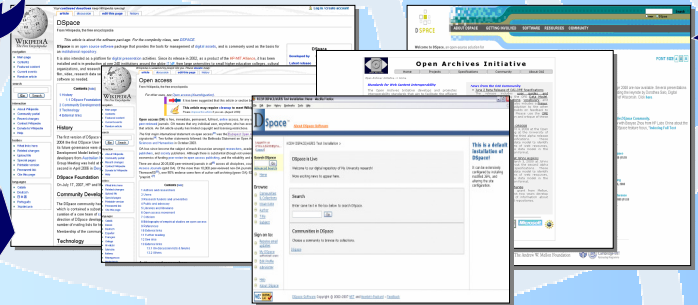
The **AGRIS AP** *was created specifically to enhance the description, exchange and subsequent retrieval of agricultural Document-like Information Objects (DLIOs). It is a metadata schema which draws elements from Metadata standards such as Dublin Core (DC), Australian Government Locator Service Metadata (AGLS) and Agricultural Metadata Element Set (AgMES) namespaces. The* **AGRIS AP** *provide mechanisms for sharing information in a standardized manner by recommending the use of common semantics and interoperable syntaxes.*

Currently there is no control on the subject keywords. This leaves lots of freedom for the choice of subject keywords and leads to inconsistent use, which makes interoperability on the level of subject difficult. **DSpace** *controlled vocabulary mechanism is incapable of handling the large multilingual nature of the AGROVOC thesaurus.*

DSpace *uses a qualified version of the Dublin Core schema based on the Dublin Core Libraries Working Group Application Profile (LAP).*

The OAI standards mandates the use of unqualified Dublin Core metadata schema for exposing metadata through the OAI-Protocol for Metadata Harvesting, however this is not sufficient for the requirements of the Agricultural Scientific Community.

The **AGRIS AP** *and* **AGROVOC** *offer advantages to the service providers:*
- *defining richer interoperability layer to aid additional resource discovery through the use of a qualified element set;*
- *giving flexibility to provide more relevant query results through targeted searching;*
- *providing rich, qualified metadata*
- *and giving adequate information about the content of the resource through the use of* **agriculture specific qualifiers**.

Content description through quality metadata creation and use of standard terminologies is the basis of efficient content management as well as the development of value added services.
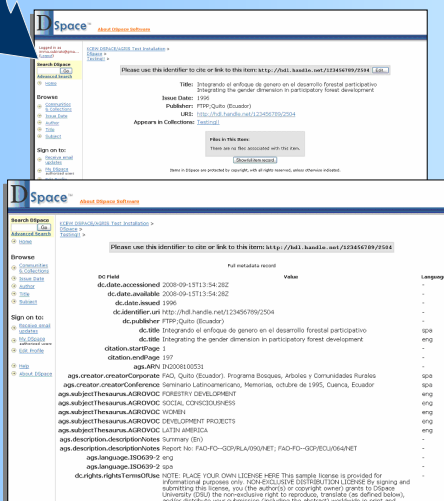
The objective of this project is to customize **DSpace** *to make use of rich metadata formats and controlled, multilingual and semantically-enriched vocabularies.*

Still to do annotation using multilingual thesauri, local or remote versions
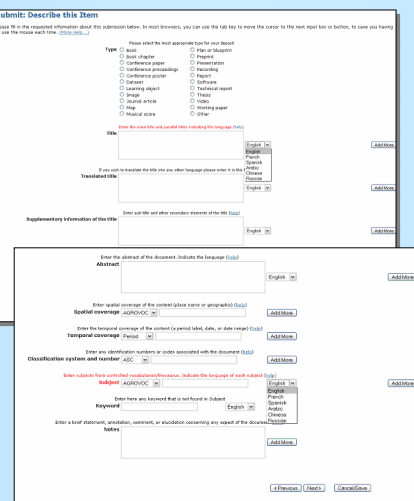
Features already customized

Custom metadata schemas
A schema for representation of AGRIS AP metadata elements

Enriched metadata input
Custom input types for language qualification of AGRIS AP metadata elements

Enriched metadata dissemination
- A Crosswalk plug-in for OAI-PMH dissemination of metadata in AGRIS AP.
- Mapping of AGRIS AP elements to DC for vanilla DC dissemination of AGRIS AP