

# *AGRISAP and DSpace*

ARD Prasad

Documentation Research & Training Centre  
Indian Statistical Institute, Bangalore

Johannes Keizer

FAO, Rome



# *AGRIS*

## *International Information System for Agricultural Science and Technology*

- Metadata from about 268 centers across the world
- Presently holds about 2.5 million metadata records in ISO-2709 and AGRISAP
- Data available from 1975 to present



# *Issues in adopting AgrisAP*

- Multi-Lingual data:
  - Presently about 22 languages and may extend to a few more
- Migration (import):
  - Legacy data in AGRISAP XML and in many languages/scripts, using varied encoding Schema
- OAI-PMH:
  - metadataPrefix=agrisap



# *AgrisAP*

- Based on Qualified DC and extends it, providing more granularity
  - corporate author, conference
  - publisher name and place as separate fields
  - keywords: various thesauri as qualifiers
  - all relations: have qualifiers like URI, ISBN, DOI etc.
- Ref: <http://www.fao.org/docrep/008/ae909e/ae909e00.htm>



# *Assumption*

- ⇒ Though, many records have metadata in English, even when the resource is in a national language
- ⇒ We can assume, in future
  - Resource and metadata are in the same language, only the translated title will be in some other language
  - In other words, we do not assume resource in one language and metadata in some other language



*An IR of a university*  
*or*  
*International organization like FAO*

- authors/submitters have publications in different languages
- They may enter Other title (translated) in English or some other language/s
- They enter metadata in the language of the resource/publication



# *Pros and Cons of DSpace*

- Pro: DSpace allows metadata to be entered in any language/script
- Con: DSpace provides interface in only one language configured using Messages.properties
- Pro: metadatafieldvalue table has a field for text\_lang. It is different from <dc:language>
- Con: We can not store the language of other title in text\_lang using input-forms.xml



## *Pro & Cons (cont...)*

- ⇒ Con: Even for element `<dc:language>`, it stores the default value in `text_lang`
- ⇒ DSpace simply stores the language mentioned in `dspace.cfg` in `text_lang` in `metadatabalue` table
- ⇒ **Ideally it should pick up from the input-forms**





# *Metadatafieldvalue Table*

- Columns include:
  - metadata\_value\_id,
  - item\_id | metadata\_field\_id,
  - text\_value,
  - text\_lang,
  - place
- What is stored in place?



dspace@localhost... (Logout)

**Search DSpace**

[Advanced Search](#)

[Home](#)

**Browse**

- [Communities & Collections](#)
- [Titles](#)
- [Authors](#)
- [Subjects](#)
- [By Date](#)

**Sign on to:**

- [Receive email updates](#)
- [My DSpace](#)  
authorized users
- [Edit Profile](#)
- [Administer](#)

[DSpace at My University](#) >  
[Test](#) >  
[Normal DSpace](#) >

Please use this identifier to cite or link to this item: <http://hdl.handle.net/123456789/7>

Full metadata record

DC Field	Value	Language
<b>dc.date.accessioned</b>	2007-10-11T12:44:06Z	-
<b>dc.date.available</b>	2007-10-11T12:44:06Z	-
<b>dc.date.issued</b>	2007-10-11T12:44:06Z	-
<b>dc.identifier.uri</b>	<a href="http://hdl.handle.net/123456789/7">http://hdl.handle.net/123456789/7</a>	-
<b>dc.language.iso</b>	it	en_US
<b>dc.title</b>	The title in in Italian	en_US
<b>dc.title.alternative</b>	Other title is in English	en_US

Appears in Collections: [Normal DSpace](#)

**Files in This Item:**

File	Description	Size	Format
<a href="#">application.html</a>		66.41 kB	HTML <a href="#">View/Open</a>

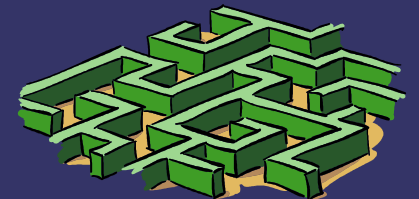
# *Our Approach*

- Ideally, avoid modifying the existing Dspace code
- May add extra programs to suite AGRIS AP
- More elements are added to accommodate AGRIS AP to metadatafieldregistry table
- Modification of input-forms.xml



# *Our Approach*

- Migration modules from the AGRIS AP xml records to DSpace import format using Java/SAX2
- Changing encoding to UTF-8
- metadataPrefix takes agrisap for OAI-PMH



```
- <ags:resources>
- <ags:resource ags:ARN="QJ2003000001">
- <dc:title xml:lang="en_US">
  title
  <dcterms:alternative xml:lang="es">Spanish Other Title</dcterms:alternative>
  <dcterms:alternative xml:lang="de">German Other Title</dcterms:alternative>
  <ags:titleSupplement xml:lang="fr">French Supplement Title</ags:titleSupplement>
  <ags:titleSupplement xml:lang="it">Italian Supplement Title</ags:titleSupplement>
</dc:title>
- <dc:creator>
  <ags:creatorPersonal>Prasad, ARD</ags:creatorPersonal>
  <ags:creatorPersonal>DUtta, Biswanath</ags:creatorPersonal>
  <ags:creatorCorporate>Ministry of Agriculture</ags:creatorCorporate>
  <ags:creatorCorporate>Ministry of Education</ags:creatorCorporate>
  <ags:creatorConference>All India Conference</ags:creatorConference>
  <ags:creatorConference>Indian Conference</ags:creatorConference>
</dc:creator>
- <dc:publisher>
  <ags:publisherName>DRTC</ags:publisherName>
  <ags:publisherPlace>Bangalore</ags:publisherPlace>
</dc:publisher>
- <dc:date>
  <dcterms:dateIssued>2007-08-10T13:07:09Z</dcterms:dateIssued>
</dc:date>
- <dc:subject>
  <ags:subjectClassification scheme="ags:ASC">4.44</ags:subjectClassification>
  <ags:subjectClassification scheme="ags:CABC">321</ags:subjectClassification>
  <ags:subjectThesaurus scheme="ags:AGROVOC">Fish</ags:subjectThesaurus>
  <ags:subjectThesaurus scheme="dcterms:MeSH">disease</ags:subjectThesaurus>
</dc:subject>
- <dc:description>
  <ags:descriptionNotes xml:lang="en_US">This is description Notes</ags:descriptionNotes>
  <ags:descriptionEdition>2nd Editon</ags:descriptionEdition>
  <dcterms:abstract xml:lang="en_US">This is abstract</dcterms:abstract>
```

# *Suggestion for Import/export*

- Should convert any encoding scheme of xml records into UFT-8 (similar to that of iconv of Unix)
- Export/import may follow DC or QDC format produced by oai\_dc or qdc





# *Conclusion*

- End user should be able to change the language of the DSpace interface
- The value of `text_lang` in `metadata` field value should be taken from input forms rather than `dspace.cfg`
- Export/import should convert different encodings into UTF-8
- `dublincore.xml` may follow `oai_dc` and `qdc` format rather than the present one



*Thank You*

