Use and integration of vocabularies in digital repositories

Imma.Subirats@fao.org
http://aims.fao.org

TaCCIRe Project
Towerbuilding (Norway)
November 15th, 2013

Free Webinars @AIMS

What is aims.fao.org?

- Web portal managed by FAO of the UN that disseminates standards and good practices in information and knowledge management for the support of the right to food, sustainable agriculture and rural development
- Supports the implementation of structured and linked information by fostering a community of practice centered on the themes of interoperability, reusability and cooperation

What do we do?

- To help to make agricultural information increasingly accessible
- To promote good practices widely applicable and easy to implement
- To adopt, create and publicize standards, tools and services enabling stakeholders to build open and interoperable information systems

Aid to Semantic Navigation and Visibility

Use of Controlled Vocabularies

What is a controlled vocabulary?

- List of terms (e.g. words, phrases) that is used to tag (label) information in a consistent way
- There are different types of vocabularies like
  - authority files, classification systems, controlled lists, ontologies, taxonomies, glossaries, subject headings etc.
- Objective is to facilitate the retrieval of content

Uses: indexing

Subject vocabularies (words or phrases taken from standardized, organised knowledge structures) should be employed to resolve indexing problems such as **plurals**, **spelling variants**, **synonyms**, **homographs** (words with same spelling but different meaning), and **polysemes** (words with multiple meanings) by ensuring that each concept is described using only one authorized term and each authorized term in the controlled vocabulary describes only one concept.

Free Webinars @AIMS

Uses: enhancing access

The use of subject vocabularies guarantees meaningful metadata while also enhancing the quality of the interoperability and effectiveness of information exchange among data providers, thus facilitating the re-usage of data by other repositories/services and in the process adding value to the local researcher.

Free Webinars @AIMS

AGROVOC

A controlled vocabulary in the agricultural field

AGROVOC

- Controlled vocabulary covering all areas of interest to FAO, including food, nutrition, agriculture, fisheries, forestry, environment etc.
- Contains over 32,000 concepts organized in a hierarchy, each concept may have labels in up to 22 languages

Uses

- Standardizes the indexing process in order to make searching simpler and more efficient and to guide the user to the most relevant sources
- Used worldwide (researchers, information management specialists) for indexing, retrieving and organizing data in agricultural information systems

Concept Scheme

- Developed in the early 80s, as a thesaurus to support uniform indexing of the AGRIS bibliographic database, and then of the whole FAO bibliographic catalogue.
- Thesaurus expressed as a concept scheme using SKOS
- This conversion from a relational database has provided added semantics value to term relationships

Concepts

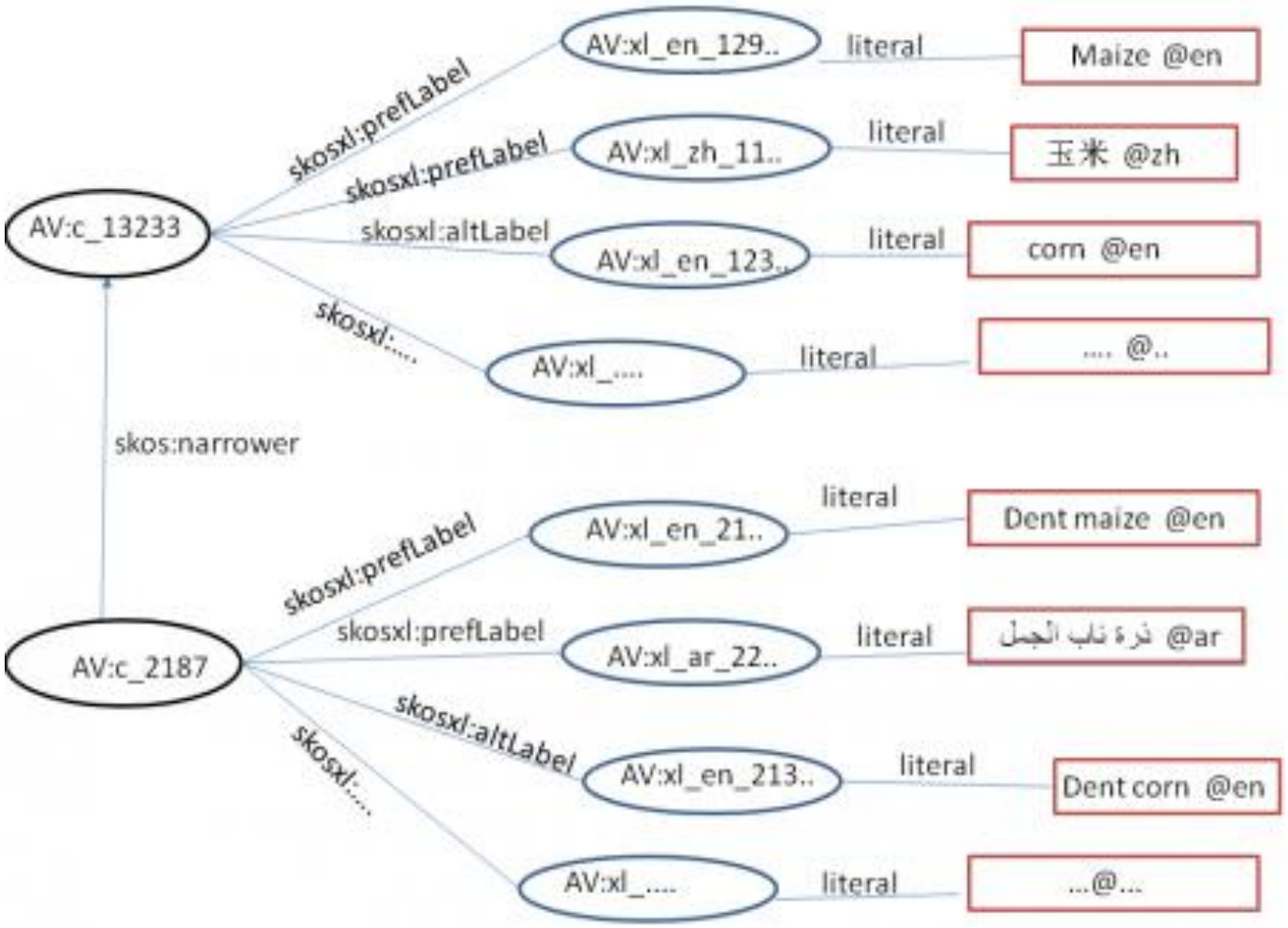
- Meant to represent the *meaning* of terms
- Set of all terms considered to be translation of one another in various languages
- Concepts are given dereferenceable URIs (= URL), such as http://aims.fao.org/aos/agrovoc/c_12332 for maize.

Terms

- Or labels, are the actual terms used to name things or abstract concepts
- For example maize, maïs, 玉米, ข้าวโพด, are all labels for the same concept in English, French, Hindi respectively.

Relations, between concepts or terms

- **Concepts**: Hierarchical relations between concepts correspond to the classical thesaurus relations broader/narrower (BT/NT)
- **Terms**: range of forms that can occur for each term such as spelling variants, singular or plural, for example organization or organisation, cow or cows


AV:c_13233
skosxl:prefLabel
AV:xl_en_129..
literal
Maize @en
skosxl:prefLabel
AV:xl_zh_11..
literal
玉米 @zh
skosxl:altLabel
AV:xl_en_123..
literal
corn @en
skosxl:....
AV:xl_....
literal
.... @..
skos:narrower
AV:c_2187
skosxl:prefLabel
AV:xl_en_21..
literal
Dent maize @en
skosxl:prefLabel
AV:xl_ar_22..
literal
@ar ذرة ناب الجمل
skosxl:altLabel
AV:xl_en_213..
literal
Dent corn @en
skosxl:....
AV:xl_....
literal
...@...

Maintenance

Collaborative effort

AGROVOC is kept up to date by the AGROVOC team in FAO, by a number of involved institutions serving as focal points for specific languages, and by individual domain experts

AGROVOC and Linked Open Data

Towards the Semantic Web

The Semantic Web

- The main difference between the web of hypertext and the Semantic Web is that while the first links in html pages or documents, the second calls for going beyond the concept document and link structured data
- In this context, Linked Data is the set of best practices for publishing and connecting structured data on the Web
- Its main objective is to liberate data from silos that are framed by proprietary database schemas

Four rules

- Defined by Tim Berners-Lee in 2006:
  - to use URIs (uniform resource identifiers) to identify resources uniquely;
  - to use http URIs so people can access the information about the resource;
  - to provide information about the resources using standard formats like RDF/XML; and
  - to include links to other resources, URIs, enhancing the linking between different resources distributed on the web.

What is Linked Open Data?

- Linked Open Data (LOD) is Linked Data distributed under an open license that allows its reuse for free.
- In 2010, Tim Berners-Lee defined a 5-star rating scheme to encourage data providers to provide linked data under open licenses.
  - The scheme uses gold stars to evaluate the availability of linked data as linked open data.

How to facilitate the linking between resources?

- The easiest way is the use of standard vocabularies, including standard vocabularies for describing data/metadata elements and standard vocabularies for indicating values.

AGROVOC

Resource	Topics	Linked concepts	Languages	Linked Data	Type of link
FAO Biotechnology Glossary	Biotechnologies	810	EN, ES, FR, +3 more	Yes	skos:closeMatch
EUROVOC	General EU	1 297	EN, ES, FR + 21 more	Yes	skos:exactMatch
GEMET	Environment	1 191	EN, ES, FR + 30 more	Yes	skos:exactMatch
Library of Congress Subject Headings (LCSH)	General	1 093	EN	Yes	skos:exactMatch
NAL Thesaurus	Agriculture	13 390	EN, ES	Yes	skos:exactMatch
RAMEAU Répertoire d'autorité-matière encyclopedique et alphabetique unifie	General	686	FR	Yes	skos:exactMatch
STW - Thesaurus for Economics	Economy	1 136	EN, DE	Yes	skos:exactMatch
TheSoz - Thesaurus for the Social Sciences	Social sciences	846	EN, DE	Yes	skos:exactMatch
Geopolical Ontology	Geopolitical entities	253	AR, CH, EN, ES, FR, RU	Yes	skos:exactMatch
Dewey Decimal Classification (DDC)	General	409	EN, ES, FR + 8 more	Yes	skos:exactMatch
DBpedia	General	10 989	EN, ES, FR + 8 more	Yes	skos:exactMatch skos:closeMatch
SWD (Schlagwortnormdatei)	General	6 245	DE	Yes	skos:exactMatch skos:closeMatch skos:broadMatch skos:narrowMatch
GeoNames	Geographical entities	212	EN, ES, FR + 63 more	Yes	skos:exactMatch

AGROVOC as Linked Open Data

- The additional value that linking AGROVOC to other vocabularies provides is that data repositories attached to those vocabularies become discoverable
- This is a very simple classic case of exposing repository contents automatically across datasets through AGROVOC indexing.

AGROVOC LOD- links

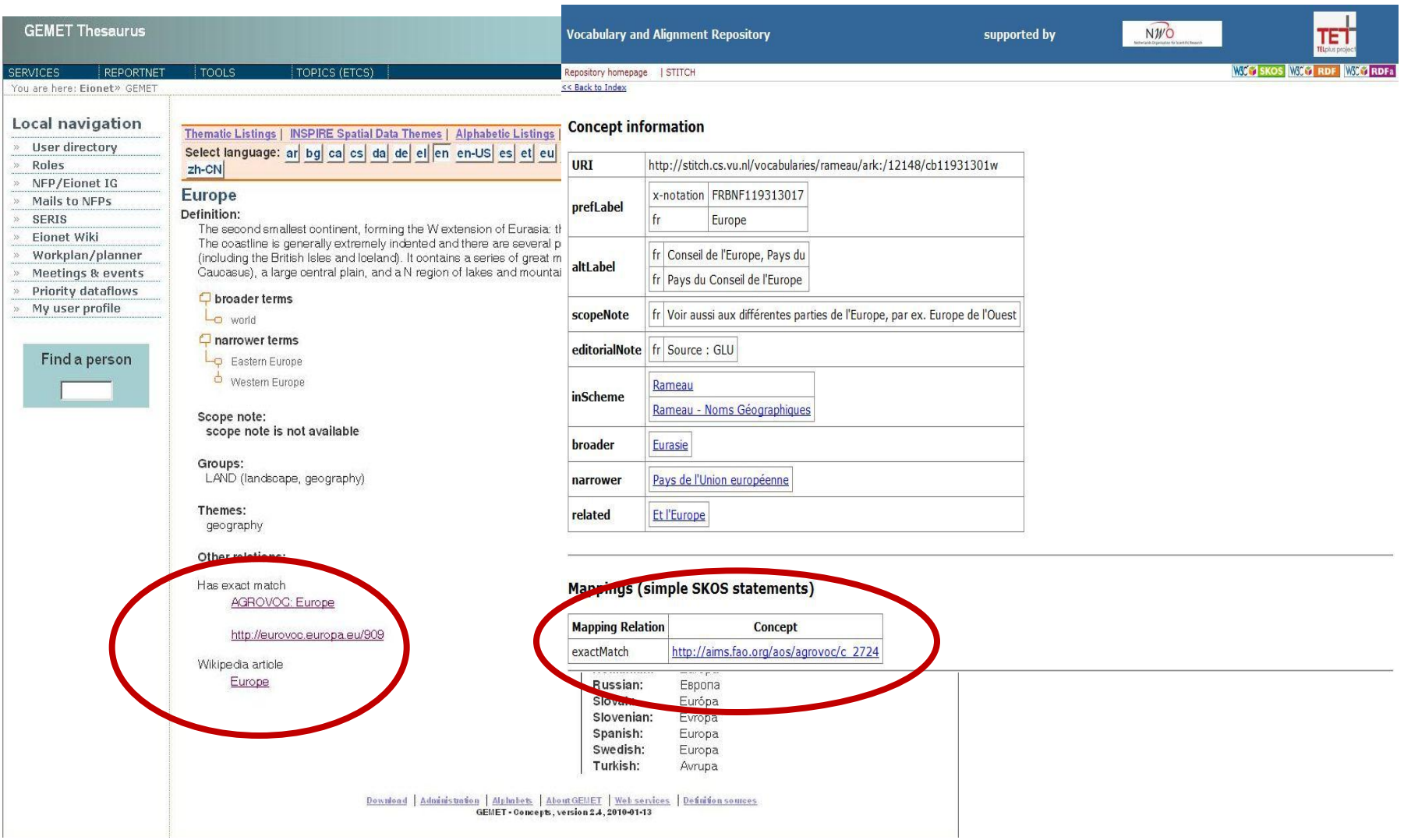
Europe at AGROVOC Thesaurus
http://aims.fao.org/aos/agrovoc/c_2724

Property	Value
broader	■ <http://aims.fao.org/aos/agrovoc/c_2723>
exactMatch	■ <http://agclass.nal.usda.gov/nalt/2011.xml#5987> ■ <http://eurovoc.europa.eu/909> ■ <http://id.loc.gov/authorities/sh85045631#concept> ■ <http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb11931301w> ■ <http://www.eionet.europa.eu/gemet/concept/2992> ■ <http://zbw.eu/stw/descriptor/16815-3>
date of creation	■ 1981-01-09
date of last update	■ 2010-05-12
isPartOfSubvocabulary	■ Geographical above country level (string)
isSpatiallyIncludedIn	■ <http://aims.fao.org/aos/agrovoc/c_24920>
prefLabel	■ EUROPA (de) ■ Europa (it) ■ Europa (es) ■ Europa (pl) ■ Europa (pt) ■ Europe (fr) ■ Europe (en) ■ Európa (hu) ■ Európa (sk) ■ Evropa (cs) ■ европа (ru) ■ أوروبا (ar) ■ اروپا (fa) ■ यूरोप (hi) ■ ยุโรป (th) ■ ເອີຣົບ (lo) ■ ヨーロッパ (ja) ■ 欧洲 (zh) ■ 유럽 (ko)
spatiallyIncludes	■ <http://aims.fao.org/aos/agrovoc/c_2726>

The international agricultural thesaurus AGROVOC is maintained by FAO and part of AIMS. The AGROVOC LOD services are a collaboration project with the MIMOS Bhd.
As Turtle | As RDF/XML

Free Webinars @AIMS

AGROVOC LOD- links



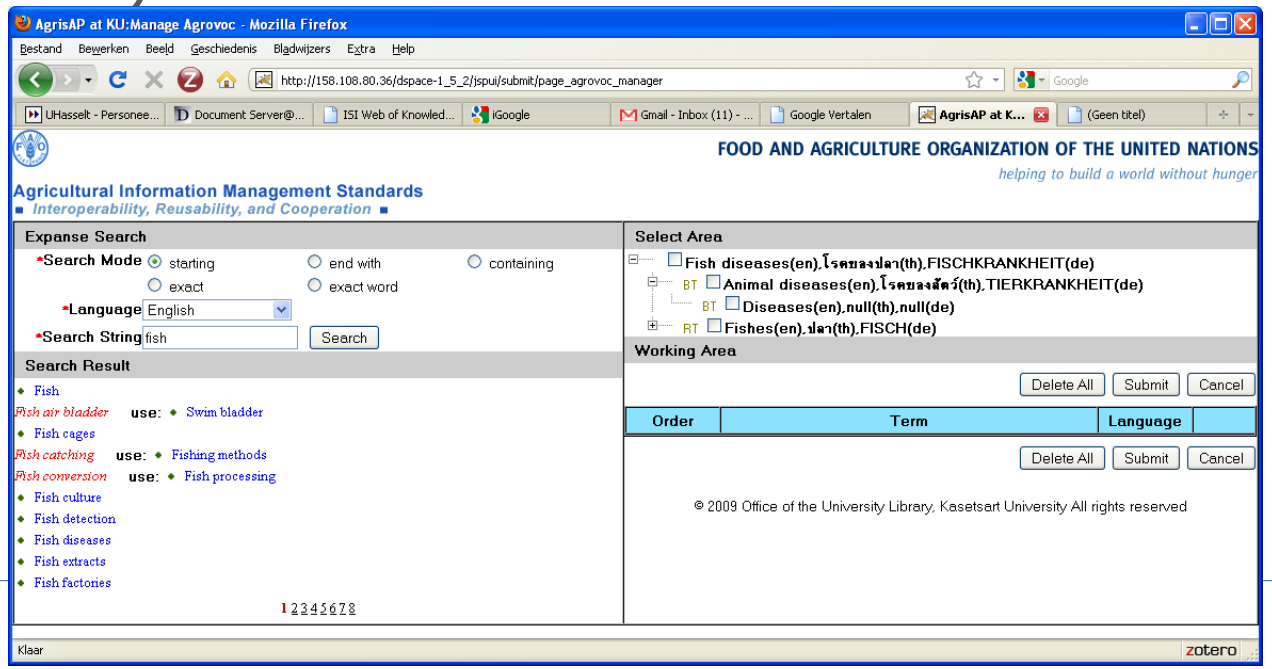
Free Webinars @AIMS

DSpace and AGROVOC

AgriOcean DSpace and the Ontology Plugin

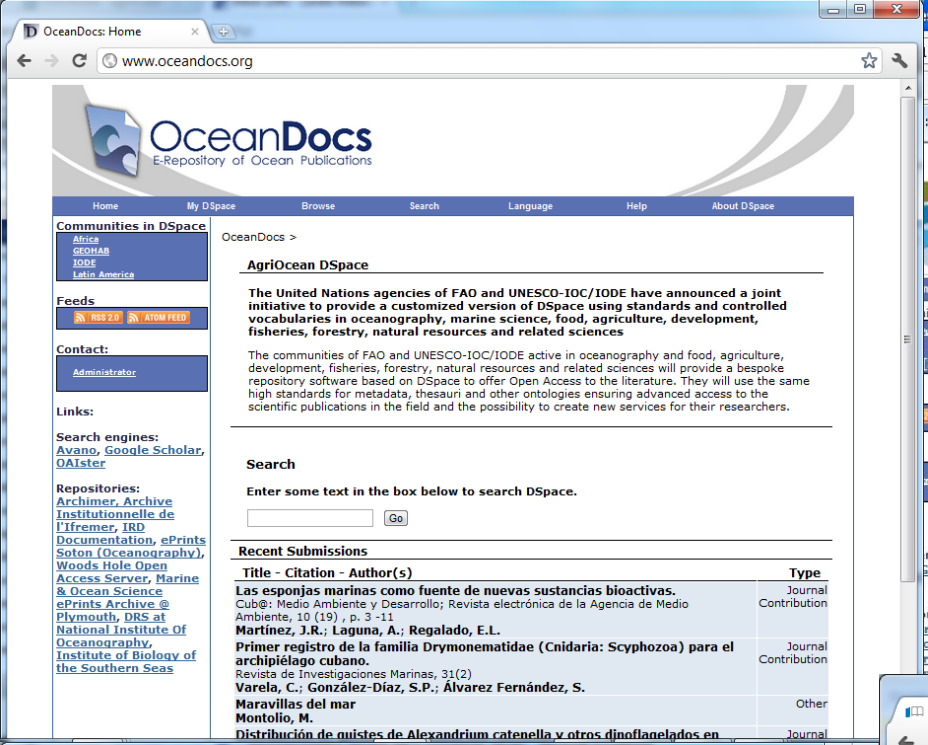
Thesaurus plug-in for AGROVOC

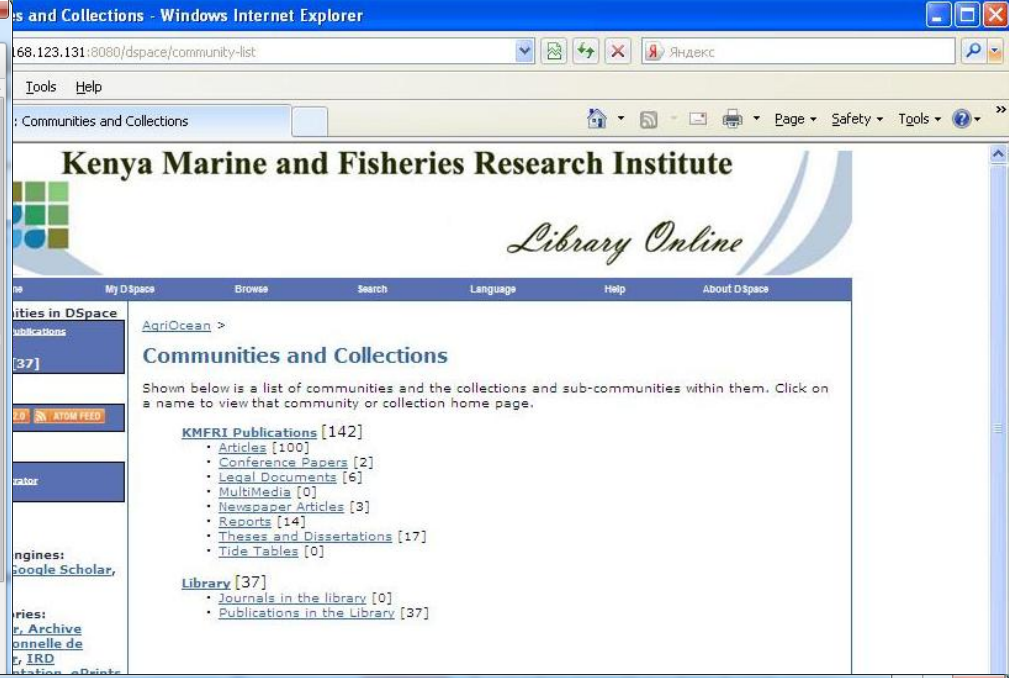
- Plug-in for Dspace defined by FAO (Feb. 2009)
  - Authority control of AGROVOC terms during submission in Dspace - Implementation of Semantic Tools
- Developed for Dspace 1.4 by Kasetsart University (Bangkok, Thailand)



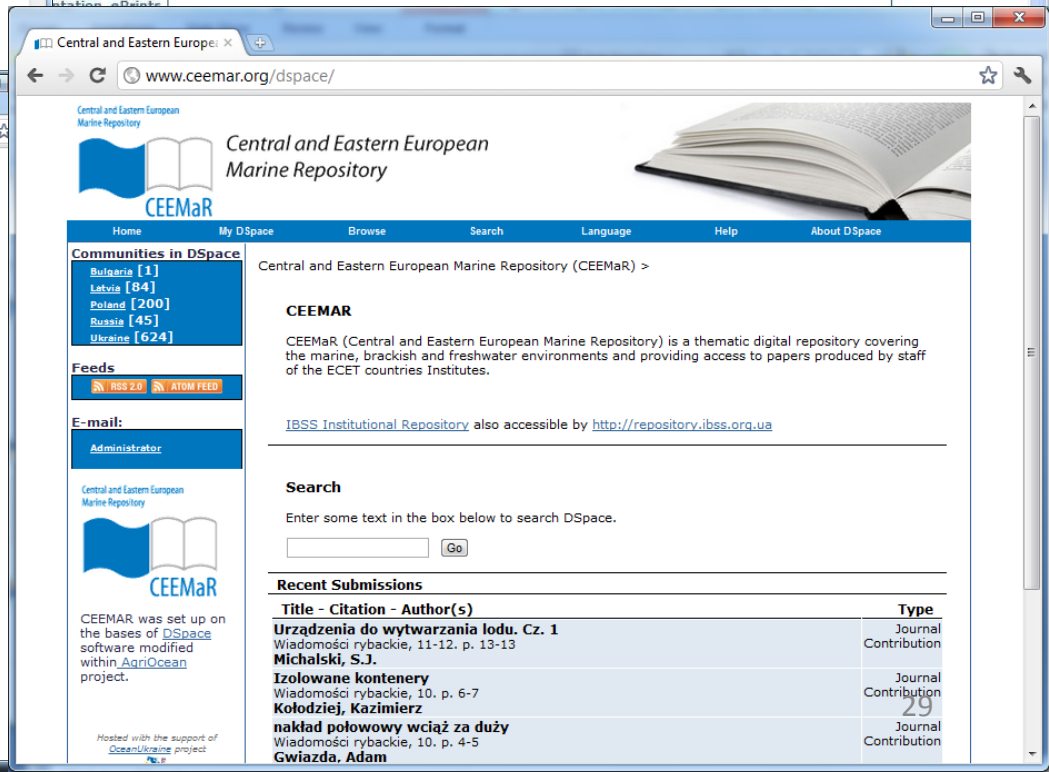
AgriOcean DSpace

- Joint initiative between the United Nations agencies of FAO and UNESCO-IOC to provide a customized version of DSpace using standards and controlled vocabularies in oceanography, marine science, food, agriculture, development, fisheries, forestry, natural resources and related sciences.
- The communities supported by FAO and UNESCO-IOC/IODE are synergistic and the standards on metadata and controlled vocabularies are similar for both.
  - Communities: AGRIS – ASFA - ODINS
  - Standards: Agris AP – MODS
  - Thesauri: AGROVOC – ASFA


OceanDocs: Home
www.oceandocs.org
OceanDocs
E-Repository of Ocean Publications
Home
My DSpace
Browse
Search
Language
Help
About DSpace
Communities in DSpace
Africa
GEOHAB
IODE
Latin America
Feeds
RSS 2.0
ATOM FEED
Contact:
Administrator
Links:
Search engines:
Avano, Google Scholar, OAIster
Repositories:
Archimer, Archive Institutionnelle de l'Ifremer, IRD Documentation, ePrints Soton (Oceanography), Woods Hole Open Access Server, Marine & Ocean Science ePrints Archive @ Plymouth, DRS at National Institute Of Oceanography, Institute of Biology of the Southern Seas
OceanDocs >
AgriOcean DSpace
The United Nations agencies of FAO and UNESCO-IOC/IODE have announced a joint initiative to provide a customized version of DSpace using standards and controlled vocabularies in oceanography, marine science, food, agriculture, development, fisheries, forestry, natural resources and related sciences
The communities of FAO and UNESCO-IOC/IODE active in oceanography and food, agriculture, development, fisheries, forestry, natural resources and related sciences will provide a bespoke repository software based on DSpace to offer Open Access to the literature. They will use the same high standards for metadata, thesauri and other ontologies ensuring advanced access to the scientific publications in the field and the possibility to create new services for their researchers.
Search
Enter some text in the box below to search DSpace.
Go
Recent Submissions
Title - Citation - Author(s)
Type
Las esponjas marinas como fuente de nuevas sustancias bioactivas.
Cub@: Medio Ambiente y Desarrollo; Revista electrónica de la Agencia de Medio Ambiente, 10 (19) , p. 3 -11
Martínez, J.R.; Laguna, A.; Regalado, E.L.
Journal Contribution
Primer registro de la familia Drymonematidae (Cnidaria: Scyphozoa) para el archipiélago cubano.
Revista de Investigaciones Marinas, 31(2)
Varela, C.; González-Díaz, S.P.; Álvarez Fernández, S.
Journal Contribution
Maravillas del mar
Montolio, M.
Other
Distribución de quistes de Alexandrium catenella y otros dinoflagelados en
Journal

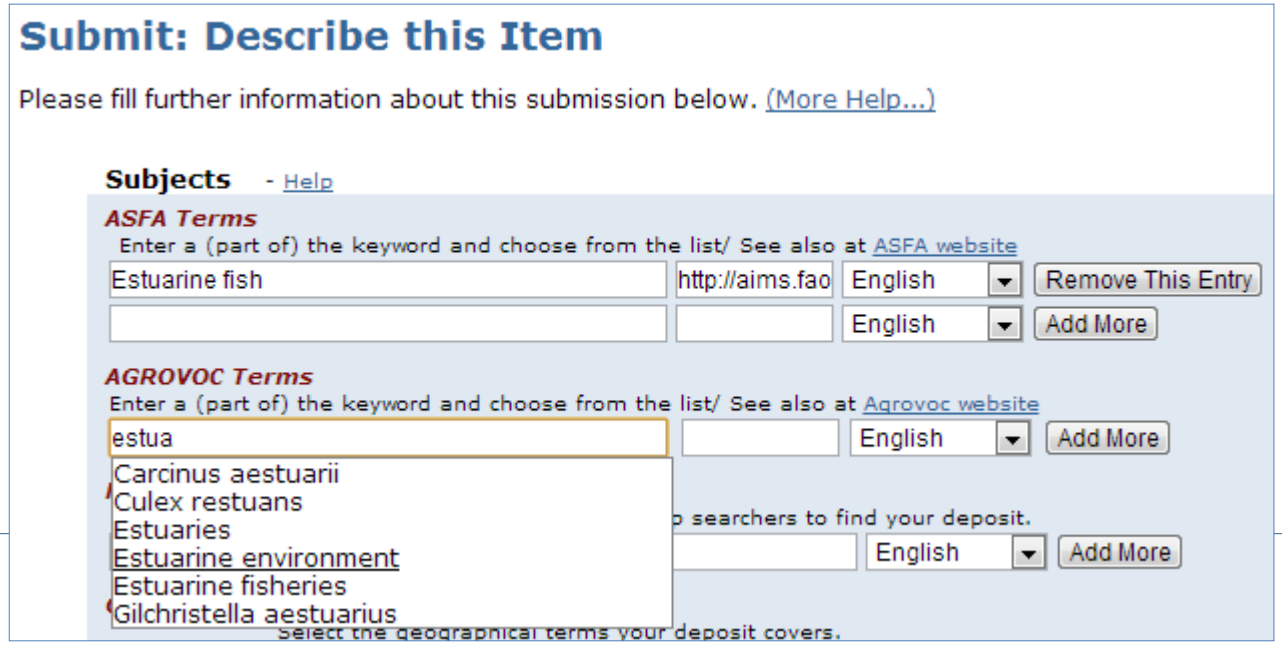

Communities and Collections - Windows Internet Explorer
168.123.131:8080/dspace/community-list
Tools
Help
Communities and Collections
Page
Safety
Tools
Kenya Marine and Fisheries Research Institute
Library Online
My DSpace
Browse
Search
Language
Help
About DSpace
AgriOcean >
Communities and Collections
Shown below is a list of communities and the collections and sub-communities within them. Click on a name to view that community or collection home page.
KMFRI Publications [142]
Articles [100]
Conference Papers [2]
Legal Documents [6]
MultiMedia [0]
Newspaper Articles [3]
Reports [14]
Theses and Dissertations [17]
Tide Tables [0]
Library [37]
Journals in the library [0]
Publications in the Library [37]

IBSS Repository: Home
repository.ibss.org.ua/dspace/
Institute of Biology of the Southern Seas
E-Repository
Home
My DSpace
Browse
Search
Language
Help
About DSpace
Communities in DSpace
Books [43]
Conferences [83]
Dissertation's abstracts [1]
IBSS serials [1565]
Overprints [76]
Preprints [8]
Feeds
RSS 2.0
ATOM FEED
Contact:
Administrator
1871
IBSS
Supported by IOC/IODE.
www.iode.org
National node of OceanDocs
OceanDocs
IBSS Institutional Repository provides a stable, secure and accessible repository for the long-term
IBSS Repository >
Welcome to the IBSS Institutional Repository!
Repository preserves and distributes digital collections including:IBSS serials, theses, research reports, preprints, papers, image collections etc.
If you are a member of the IBSS research community and you are interested in contributing digital content, please contact the repository administrator repository@ibss.org.ua
We are moving to the new DSpace version, modified in the frame of AgriOcean project. Some bugs may be present at this moment.
Search
Enter some text in the box below to search DSpace.
Go
Recent Submissions
Title - Citation - Author(s)
Type
Shell Disease in the Giant Oyster OSTREA EDULIS LINNE, 1758 (BIVALVIA) in the Lake Donuzlav (Black Sea)
Pirkova, A. V.
Preprint
On the Article "A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus" by F. Wolfe-Simon, J.S. Blum, T.R. Kulp, G.W. Gordon, S.E. Hoeft, J. Pett-Ridge, J.F. Stolz, S.M. Webb, P.K. Weber, P.C. W. Davies, A.D. Anbar, and R.S. Oremland
Inland water biology, 4(3), p. 400-401
Shadrin, N. V.
Journal Contribution
Александр фон Гумбольдт - первый эколог
Россия и Германия,(2), p. 2-11
Шадрин, Н. В.
Journal Contribution
Alternative stable states of ecosystems, tipping points and management of the coastal seas
Shadrin, N. V.
Other


Central and Eastern Europe
www.ceemar.org/dspace/
Central and Eastern European Marine Repository
CEEMaR
Central and Eastern European Marine Repository
Home
My DSpace
Browse
Search
Language
Help
About DSpace
Communities in DSpace
Bulgaria [1]
Latvia [84]
Poland [200]
Russia [45]
Ukraine [624]
Feeds
RSS 2.0
ATOM FEED
E-mail:
Administrator
CEEMaR
CEEMAR was set up on the bases of DSpace software modified within AgriOcean project.
Hosted with the support of OceanUkraine project
Central and Eastern European Marine Repository (CEEMaR) >
CEEMAR
CEEMaR (Central and Eastern European Marine Repository) is a thematic digital repository covering the marine, brackish and freshwater environments and providing access to papers produced by staff of the ECET countries Institutes.
IBSS Institutional Repository also accessible by http://repository.ibss.org.ua
Search
Enter some text in the box below to search DSpace.
Go
Recent Submissions
Title - Citation - Author(s)
Type
Urządzenia do wytwarzania lodu. Cz. 1
Wiadomości rybackie, 11-12. p. 13-13
Michalski, S.J.
Journal Contribution
Izolowane kontenery
Wiadomości rybackie, 10. p. 6-7
Kołodziej, Kazimierz
Journal Contribution
nakład połowowy wciąż za duży
Wiadomości rybackie, 10. p. 4-5
Gwiazda, Adam
Journal Contribution

Authority control in AgriOcean DSpace

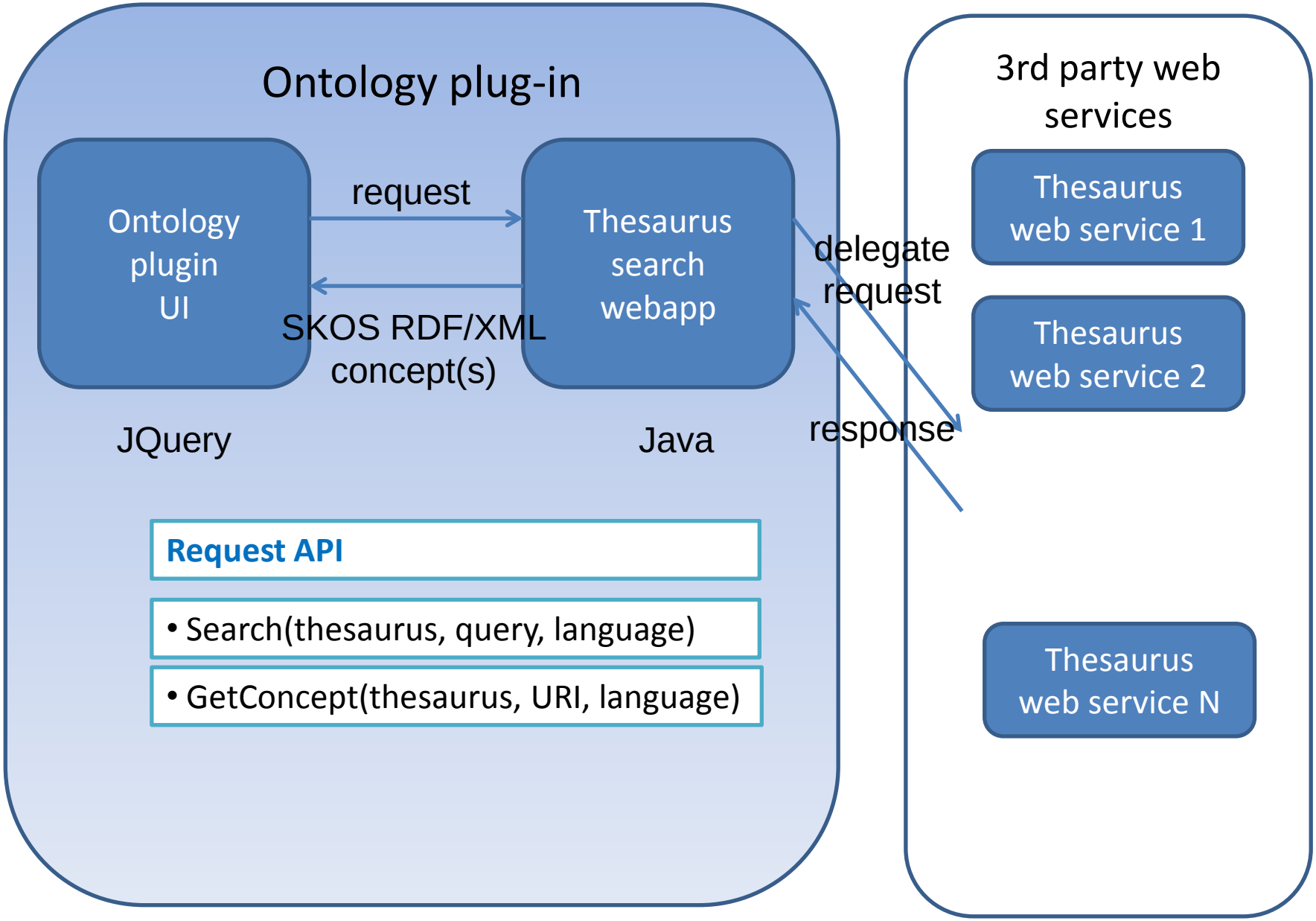
- Journal titles, Subject terms (AGROVOC, ASFA)
- Search option in alphabetic lists (as tables) with autosuggest possibility
- For every vocabulary additional input field



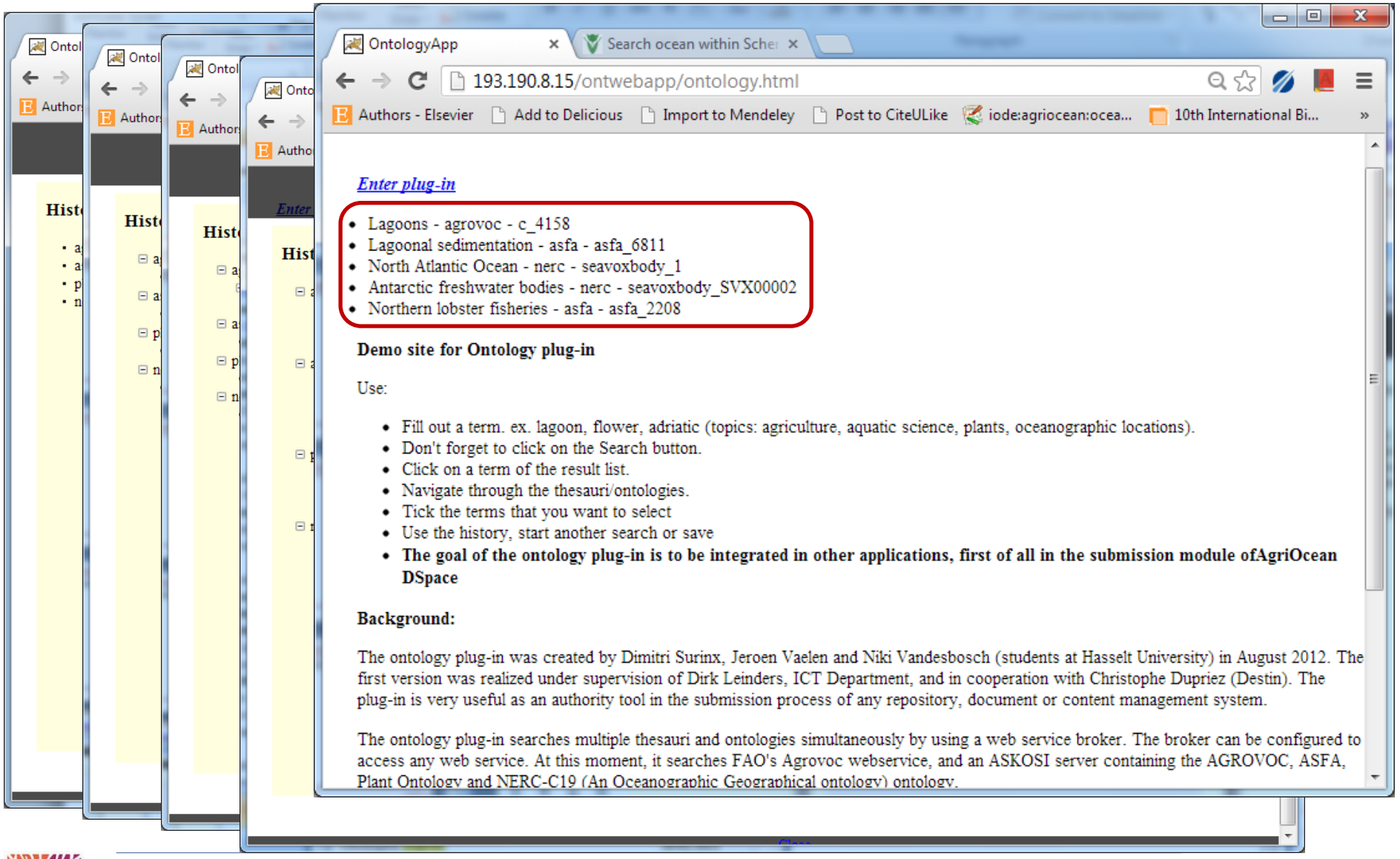
Free Webinars @AIMS

Ontology plug-in: concept & implementation

- Tool for submission of metadata
- Search over different vocabularies:
  - Grouping of vocabularies is possible
  - submitter does not have to think about which vocabulary to use
- Independent tool that can be integrated in different systems
- It searches AGROVOC webservices, and an ASKOSI server containing the AGROVOC, ASFA, Plant Ontology and NERC-C19 (An Oceanographic Geographical ontology) ontology.
  - The broker can be extended to access any web service


Ontology plug-in
Ontology plugin UI
request
SKOS RDF/XML concept(s)
Thesaurus search webapp
JQuery
Java
delegate request
response
3rd party web services
Thesaurus web service 1
Thesaurus web service 2
Thesaurus web service N
Request API
• Search(thesaurus, query, language)
• GetConcept(thesaurus, URI, language)

Demo at http://193.190.8.15/ontwebapp/ontology.html



Free Webinars @AIMS

Roadmap

- Inclusion in AgriOcean Dspace 2.0 (March 2014)
- Availability for other projects:
  - On Google Code
- Options for further developments:
  - Extended view of relations:now only direct relations
  - Graphical interface
  - Other implementation

Ontology plug-in

- The ontology plug-in was created by Dimitri Surinx, Jeroen Vaelen and Niki Vandesbosch (students at Hasselt University) in August 2012.
- The first version was realized under supervision of Dirk Leinders, ICT Department, and in cooperation with Christophe Dupriez (Destin – ASKOSI - http://www.destin-informatique.com/ASKOSI/).

More information at

- AIMS
  - http://aims.fao.org/
- AGROVOC
  - http://aims.fao.org/standards/agrovoc
- AgriOcean Dspace
  - http://aims.fao.org/agriocean-dspace
- Ontology plug-in
  - Source code is as open source Apache License 2.0 at https://code.google.com/p/ontology-plugin/
  - Access to the Ontology Plug-in demo at http://193.190.8.15/ontwebapp/ontology.html

Thanks!

Turn of questions