

SKOS Authority Control for DSpace:

This development was triggered by the asserted needs of two multilingual institutions:

- 1) The internal scientific documentation database of the Belgium Poison Centre (Dutch, French, English); Background document:
<http://convegni.cilea.it/conf/viewpaper.php?id=197&print=1&cf=11>
- 2) CDMC Haute Alsace (Guebwiller, France) : <http://www.windmusic.org/dspace>
(French, German, English, Spanish, Dutch)

This development is based on the terminology and the recommendations of the SKOS (Simple Knowledge Organisation System) standard established by the W3C:
<http://www.w3.org/TR/skos-primer/>

Reading the SKOS primer is a prerequisite to understand fully the technical aspects of this document.

Some more features have been added foreseeing the size and the complexity of vocabularies like the MeSH (Medical Subject Headings) with its Supplementary Concept Records (<http://www.nlm.nih.gov/pubs/factsheets/mesh.html>).

This development has touched many parts of DSpace (controlled vocabularies, update form, input forms, search forms and their results, item display, configuration, metadata registry, etc.). We are now in the process of identifying and motivating all the changes made to publish this contribution to the DSpace community.

We welcome any support to make this contribution merged for the satisfaction of the DSpace community.

Comments and constructive critics on this document are also very welcome.

Table of Contents

Main Motivations.....	3
Following a user:	3
The Subject Search:.....	3
The Subject Search Result:.....	4
Thesaurus management directly in DSpace:	5
Internationalization (I18n):.....	6
Search on a Chinese name:.....	7
Result for this search:.....	7
Displaying a retrieved record:	8
Storage of this record in DSpace:.....	10
Record for a Person being an Author:.....	11
Updating a Record:	12
Configuration	14
DSpace.cfg	14
Existing Parameters re-used:	14
New parameters:	16
SKOS directory	19
SKOS XML.....	20
SKOS Tab delimited Text :.....	21
SKOS Pre-loaded SQL	21
SKOS Dynamic SQL.....	22
A Test using Agrovoc from FAO:.....	23
Displaying a Record:.....	24
Controlled Vocabulary Search using Agrovoc:.....	24
“Auto-Complete”:.....	26
Searches:.....	27
About DESTIN inc. SSEB:.....	28

Main Motivations

- Overview of the database content by classification but also by hierarchies of subjects;
- Precise retrieval of the documents linked to a concept selected by the user (selection using the concepts' names but also concepts' classification in hierarchies);
- Exhaustive searches using concepts hierarchies to encompass all specific topics and using translations + synonyms to encompass all terminological variants;
- Easier searches for the users: auto-complete giving clues of the actual content of the database;
- Multilingual operation;
- Faceted browsing of any result set (still to be implemented);
- Data quality control;
- Easier Cataloguing (auto-complete or menus for most fields)

Following a user:

Please find below, for different DSpace screens, the new features proposed to the users.

The Subject Search:

Search DSpace WindMusic > A A A

Subject Search Help...

Check the boxes next to the categories that you wish to search under, then hit "Search...". Categories can be expanded to refine the search terms, and as many categories can be selected as required.

Filtering the list of categories will remove from the list below any categories that do not match the filter term. Expanding each category will show you which terms did match the filter.

Find a subject in the controlled vocabulary:
Filter:

- ARTS (874)
- ARCHITECTURE (4)
- ARTS DU SPECTACLE (328)
- ARTS PLASTIQUES (34)
- PEINTURE (30)
- SCULPTURE (4)
- ESTHETIQUE (27)
- LITTERATURE (514)
- DANSE (3966)
- BOSSA NOVA (84)

1. The Subjects are automatically obtained from any SKOS sources (an XML file, Tab delimited Text files, SQL database; RDF to be implemented). In this example, they are obtained from a DSpace collection containing one record for each subject (see the "Subject Record" further below).
2. Result counts are provided for each subject.

- Subjects names are displayed in the language of the user when they are available (it is not still the case for WindMusic: translation is in progress. Please look at the Agrovoc test further below).
- The search can (or not) include all subdivisions for a given subject. Here, there is one record indexed to “Arts Plastiques”, 30 records indexed by “Peinture” and 4 indexed by “Sculpture”. The “OR” of those three sets contains 34 records.
- The search is done on the specific code of each subject selected: it is a precise search that will have no noise like a word search.

The Subject Search Result:

WindMusic >



ARTS PLASTIQUES

Results 1-20 of 34.

		Call Number	Composer	Date issued	Newly Available	Inventory	Loris	Identifier	Similarity
n°	Newly Available	Titre			Author		Type		
1	2009-02-11		Dutch Masters Suite	Composer:MEIJ (DE), Johan		Full score			
2	2008-11-14		Un jour au Louvre	Composer:FIENGA, Robert		Full score			
3	2008-09-25		Art in the park	Composer:SHELDON, Robert E.		Full score			
4	2008-09-25		Art in the park	Composer:SHELDON, Robert E.		Works on CD			
5	2008-09-10		Three paintings by Lautrec	Composer:JOHNSON, Laurie		Full score			
6	2008-04-24		Vor der sonne : Impressionen nach einem Bild von Joan Miro : für flöte, oboe und fagott	Composer:RUDIN, Rolf		Ensemble music			

This is the result if one chooses “Arts Plastiques” in the Subject Search Index. It is sorted and contains exactly the records using the selected subjects and their specifics (if so configured).

- Circled in red:
 - the author names coming from the Persons (or Orchestras) authority list;
 - their roles (composer, arranger, etc.) are also coded and translated using an SKOS Concept Scheme of the « Authors' roles »;
 - the type of the record coded and translated in the user language, using a SKOS Concept Scheme of the records' types.
- Circled in green, the print basket function allows choosing records to be printed together. An export function could be based on this mechanism.
- Circled in blue, the possible sort keys (configured for this application). The second column displays the sort key for each records (here the availability date of each document). The user can choose anytime a different sort key.

Thesaurus management directly in DSpace:

Subjects can be catalogued in a DSpace collection like any other bibliographic record: <http://www.windmusic.org/dspace/handle/68502/22360>

WindMusic >
6. Autour des œuvres musicales >
6.4. Sujets (index) >



Title:	ARTS PLASTIQUES <small>Simil.</small> 1 → 34
in:	ARTS → 8 → 874
Loris:	lorthes_565
URI (PermaLink):	http://dspace/handle/68502/22360
Appears in Collections:	6.4. Sujets (index)
Submitted by:	loris

- A title may be specified for each language (translation in progress for WindMusic)
- Next to the title, the two numbers are for “horizontal searches”:
 1. the number of records specifically linked to the concept (here: 1)
 2. the number of records linked to the concept or any of its specifics (here: 34)
- dc.title.alternative can be used for synonyms.
- “in:” (dc.relation.ispartof) indicates that this concept is a subdivision of ARTS. The number of linked records is also indicated for ARTS, the parent concept (the more generic subject).
- “lorthes_565” is a SKOS notation, equivalent to keyword “sujet_22360” but in another application.

Advanced Search:

<http://www.windmusic.org/dspace/advanced-search>

Search: All of DSpace

Composer/Performer MOZART, Wolfgang Amadeus (386)

AND Solo instruments Flute (217)

AND Every fields

AND

Recent Sub

Orchestre c

Instrument

Oeuvre originale

Arrangement

Batterie-fanfare

Orchestre de jeunes

Search Clear

n 1.4.1 - Feedback

The authority list varies depending on the search type chosen. Some search types do not use authority lists. The size of the result is indicated. The possible search types are defined in a SKOS source “dSPACESearch”.

Internationalization (I18n):

Names are indexed with all variants and translations of their names. Lucene search does work with asian characters.

Search on a Chinese name:

Search:

Contributors

AND

Result for this search:

Search:

for:

Results 1-1 of 1.

Call Number	Composer	Date issued	Date available	Inventory	Loris	Identifier	Similarity
n ^o	Titre	Contributor				Score type	
1	 Three moon	Scriptwriter: Price, John Wesley				Full score	

(The Chinese characters in the search form are the translation (or transliteration) of the person name “Price, John Wesley”)

Displaying a retrieved record:

<http://www.windmusic.org/dspace/handle/68502/59731>

WindMusic >

1. Partitions pour Orchestres à vent >

1.1. Partitions pour Orchestres d'harmonie >



<< 14 / 20 >>

Page 1 / 2 >



Title:	Art in the park <small>Simil.</small>
Additional Title:	Art dans le parc <small>Simil.</small>
Contents list:	I. Le Cycliste (an example of cubism by Gino Severini) II. The Shadowed Stream (an example of aquarelle by John Singer Sargent) III. Excavation (an example of sgraffito by Willem Kooning) IV. Transfiguration (an example of chiaroscuro by Raphael)
Music type:	Wind orchestra
Type:	Full score ▶-12818
Language:	English ▶-18753
Composer:	SHELDON, Robert E. ▶-159 ▶-159
Abstract:	Cette pièce en quatre mouvements dresse musicalement le portrait de différents styles et techniques utilisés dans les arts graphiques. Quatre tableaux ont inspiré cette composition. Dans chaque cas le compositeur a essayé non seulement de capturer le caractère et l'énergie de la peinture, mais aussi de recréer musicalement le style du coup de pinceau.
Keywords:	21E SIECLE ▶-10313 ▶-10313 OEUVRE ORIGINALE ▶-25703 ▶-25703 ORCHESTRE D'HARMONIE ▶-15506 ▶-15506 PEINTURE ▶-30 ▶-30 SUITE DE PIECES ▶-1040 ▶-1040
Level:	level 5 ▶-1401
Orchestras:	Concert band 1 ▶-32538
Additional instruments:	Glockenspiel 1 ▶-30 ▶-13 sol. Xylophone 1 ▶-166 ▶-68 sol.
Publisher:	Alfred ▶-716 ▶-+1094
Place of publication:	Van Nuys, Etats-Unis
Issue date:	2008
Collection:	Alfred's Concert Band Series ▶-60
Publisher Reference:	29479
Description:	47 p.
Duration:	00:08:30
Linked Recordings:	Art in the park CDHAR001823-308
CIM Call Number:	210.1 SHE
Loris:	74884
Inventory #:	033365
Record type:	Notice locale brève
Loris Cataloging date:	2008-Sept-25
URI (PermaLink):	http://dspace/handle/68502/59731
Appears in Collections:	1.1. Partitions pour Orchestres d'harmonie
Submitted by:	Gerard

- **Title:** Fields presentation is prepared using little HTML templates (JSP UI). Here an icon is added and also a search for similar titles (**Simil.**)
- **Type, Language, Level, Orchestras, Additional instruments:** These fields are controlled by small and stable authority lists. An “horizontal” search is proposed with figures indicating the number of records potentially retrieved. Two numbers are indicated for instruments, one for instruments used by the orchestra, one for instruments used as soloist.

- **Author, Keywords, Collection:** The role (Composer) of the Author is indicated in the left margin. These fields are controlled by Dynamic SQL SKOS Authority lists. In this example, the permitted entries (and their translations) are stored in DSpace collections.
The number of related records is shown on the right. If there are multiple results possible (for instance, search on a precise concept + search when using it with all its specific concepts), multiple numbers are shown. *It remains to suppress the repeated numbers when two types of search return the same result.*
- **Publisher:** There is a DSpace collection for allowed publishers. The two different figures are respectively:
 - the number of CDs or Music Scores published by this publisher
 - the number of Recordings (tracks) on those CDs
- **Linked Recordings:** There are multiple DSpace collections containing items about Music Recordings. A link is generated toward the recording associated with the music score displayed.

Storage of this record in DSpace:

DC Field	Value	Language
dc.contributor.author	Compositeur_ loraut_621	-
dc.date.accessioned	2009-04-20T15:00:16Z	-
dc.date.available	2009-01-05T04:56:59Z	-
dc.date.issued	2008	-
dc.date.submitted	2008-09-25	-
dc.identifier	033365	-
dc.identifier.other	210.1 SHE	-
dc.identifier.uri	http://dspace/handle/68502/59731	-
dc.identifier.loris	74884	-
dc.identifier.publisher	29479	-
dc.description.abstract	Cette pièce en quatre mouvements dresse musicalement le portrait de différents styles et techniques utilisés dans les arts graphiques. Quatre tableaux ont inspiré cette composition. Dans chaque cas le compositeur a essayé non seulement de capturer le caractère et l'énergie de la peinture, mais aussi de recréer musicalement le style du coup de pinceau.	-
dc.description.tableofcontents	I. Le Cycliste (an example of cubism by Gino Severini)	-
dc.description.tableofcontents	II. The Shadowed Stream (an example of aquarelle by John Singer Sargent)	-
dc.description.tableofcontents	III. Excavation (an example of sgraffito by Willem Kooning)	-
dc.description.tableofcontents	IV. Transfiguration (an example of chiaroscuro by Raphael)	-
dc.audience	niveau_Orchestre_Difficulte_5	-
dc.format.extent	000830	-
dc.format.medium	47 p.	-
dc.language.iso	en	-
dc.publisher	lorpub_127	-
dc.publisher.country	Van Nuys, Etats-Unis	-
dc.relation.ispartofcollection	lorcoll_1440	-
dc.subject	lorthes_337	-
dc.subject	lorthes_457	-
dc.subject	lorthes_1114	-
dc.subject	lorthes_566	-
dc.subject	lorthes_797	-
dc.subject.classification	musique_1	-
dc.subject.orchestra	typeOrchestre_orchestre_d_harmonie_1	-
dc.subject.additionalinstrument	glockenspiel_1	-
dc.subject.additionalinstrument	xylophone_1	-
dc.title	Art in the park	-
dc.title.alternative	Art dans le parc	-
dc.type.score	typePartition_Orchestre_a_Vent_Orchestration_Complete	-
dc.relation.hasformat	lorocd_74896	-
dc.type.record	notice_breve	-
Appears in Collections:	1.1. Partitions pour Orchestres d'harmonie	
Submitted by:	Gerard	

- **Contributor.Author:** the role of the author is coded “Compositeur_”. The author is specified using a “notation” (alternate coding system) loraut_621 (the main code is personne_8144). The record describing this author is in DSpace (please look further below).
- **Description.Abstract:** this is a “free text” field. It must be manually translated as it is not coded. WindMusic being in 5 languages, we want to minimize this kind of fields.
- **Audience, Subject.Classification, Subject.Orchestra, Type.Score, Type.Record:** these fields are controlled by small and stable authority lists. Those lists are defined in simple Tab delimited Text files. For example,

Subject.Classification is controlled by the SKOS Scheme (Authority List) “musique”:

about	label-fr	label-en
0	Divers	misc.
1	Orchestre à vent	Wind orchestra
2	Pédagogie	Educational methods
3	Ensemble	Ensemble
4	Eveil et formation musicale	Musical awakening and training
5	Orchestre symphonique	Symphonic orchestra
6	Musique pour enfants	Music for children
V	Orchestre à vent(sujet)	Wind orchestra(subject)

- **Publisher, Relation.IsPartOfCollection, Subject, Relation.HasFormat** are all controlled by one or more DSpace Collections.
- **Subject.AdditionalInstrument** is controlled by the small and stable SKOS Scheme (Authority list) “instrument”. A suffix is used to indicate the number of instruments (“_1”, “_2”, ...). A prefix could be used to indicate a role (“soloist_” for instance). This allows making a precise search on the code of an instrument. This also allows combining a quantity and a role in this search.

Record for a Person being an Author:

WindMusic >
 6. Autour des œuvres musicales >
 6.1. Auteurs (Biographies) >



Title:	† SHELDON, Robert E. <small>Smil.</small> † 159
Additional Title:	Scheldon Robert <small>Smil.</small>
Abstract:	Diplômé de la Eastman School of Music. Spécialiste de muséographie pour les instruments de musique. Pendant vingt ans, ses fonctions inclueront restauration et conservation des instruments de musique pour le compte de la Smithsonian Institution division of Music Instruments. Occupera ensuite des fonction similaires dans la section instruments de musique à la Library of Congress. Pratique différents instruments avec plusieurs ensembles et enregistre des disques pour Mercury, Erato, Decca.....
Place of publication:	United States
Loris:	loraut_621
URI (PermaLink):	http://dspace/handle/68502/8144
Appears in Collections:	6.1. Auteurs (Biographies)
Submitted by:	loris

- If authors are within the DSpace application, you use DC fields (or any other schema like vCard 3 or FOAF) to store all the desired information about each author. All usual DSpace functions are available to manage the authors' pages. The current cataloguing expertise of an institution is directly applicable to author information.
- In this example, a collection is dedicated to receive the records about the authors. This is a “Dynamic SQL SKOS source” for which accesses are cached: entries are not pre-loaded to save application memory.
- Permanent “categories” of authors can be added if necessary (country of birth for instance) and searches are possible by those categories (for instance: “find all records authored by a belgian”).
- Circled in grey, the number of records authored by Robert E. Sheldon. This is a precise count, not the result of a word search.

- Circled in green, a synonym for the author's name that can be used in word searches. Multiple synonyms are allowed for different languages and / or alphabets.
- Circled in red, an alternate notation for this author (please look at SKOS standard).
- Circled in blue, the key for this author. It is therefore referred by "personne_8144"

Updating a Record:

The update form is completely overhauled:

- a. The form data is displayed in the same order than for the record display; Empty fields are present; Supplementary fields are allowed at the end. The proposed fields may then be different from one collection to another.
- b. A table of content for groups of fields (direct access)

Below is the image of the beginning of the update form for a musical score:

WindMusic >

Edit Item 68502/29992 (internal id.29976), Last modified: 2009-Marc-7 16:41:07

In Collections: 1.1. Partitions pour Orchestres d'harmonie

<ul style="list-style-type: none"> ▪ Title ▪ Music type ▪ Author ▪ Abstract ▪ Level ▪ Orchestras ▪ Ensembles ▪ Publisher ▪ Collection ▪ Linked Recordings ▪ CIM Call Number ▪ ... ▪ Add Bitstream 	<div style="border: 1px solid gray; padding: 2px; margin-bottom: 5px;">occ.1 ; occ.2 ; ...</div> <p>occ.1 occ.2 ...</p>	<p>Update: =</p> <p>Remove: X</p> <p>Add: +</p>
ff	Title	
Title :	Three moon	<input type="button" value="="/> <input type="button" value="X"/> <input type="button" value="+"/>
Additional Title :	Les trois lunes	<input type="button" value="="/> <input type="button" value="X"/> <input type="button" value="+"/>
Contents list :		<input type="button" value="="/> <input type="button" value="X"/> <input type="button" value="+"/>

- c. A reminder of button meanings (= : update; X : remove; + : add). Empty fields are mentioned and receive only a "+" button (add).

- d. Numbers are validated
- e. Short lists are shown as menus (Honorific Title for instance)

Example of a menu + Example of Authors having a role and a code (data entry using auto-complete based on the name of author):

ff		Music type		
Music type :	Wind orchestra			= X +
Score type :	misc. Wind orchestra Educational methods Ensemble Musical awakening and training Symphonic orchestra Music for children Wind orchestra(subject)			= X +
Language :	language_en			= X +
ff		Author		
Author :	HULTGREN, Ralph			= X +
Compositeur	personne_8329			
Arranger :				+
Contributor :	Price, John Wesley			= X +
Scenariste	personne_82607			

- f. A menu of allowed languages can be associated to any field and constrained differently for each. The languages can be an auto-complete field (with any existing ISO code allowed) or a menu (when only some languages are allowed).

Our customers have measured that, with this form, they achieve the same efficiency to update a record with DSpace than with specialized library management software.

Configuration

DSpace.cfg

We added, in this file, the main parameters to use SKOS Schemes within DSpace.

Existing Parameters re-used:

- `webui.controlledvocabulary.enable = true`
`webui.controlledvocabulary.index = broadkeyword`
`webui.controlledvocabulary.schemes = sujet`

Those parameters enable the “Subject Search”, specify the index to use for search but also (this is an addition) the SKOS Concept Scheme (hierarchy) to display.

- `webui.itemdisplay.enregistrement.collections=68502/7,68502/8,68502/9,68502/10,68502/13`

This set of parameters controls the list of collections that follows a given display profile (here: *enregistrement*). When SKOS Schemes are stored in DSpace collections, it is necessary to have profiles for them. Those display profiles wear the same name than the concept scheme. Here, the Scheme containing all the recordings (*enregistrement*) is displayed using the profile with the same name.

In case a SKOS Scheme is stored in multiple collections having different display rules, it is allowed to add an underline and specify a more specific display profile for a subset of a Scheme. For instance, with music scores (*partition*), we have:


`webui.itemdisplay.partition.collections=68502/12,68502/15`

`webui.itemdisplay.partition_ensemble.collections=68502/2,68502/3,68502/4,68502/5`

`webui.itemdisplay.partition_repertoire.collections=68502/15,68502/16,68502/17`

- `webui.itemlist.columns = dc.title=title, dc.contributor.author!author, dc.type.score!type`

This parameter defines the columns to display in a search result. It is using “styles” (title, author, type) to parameterize the display of each column.

n°	Titre	Author	Type
141	 The Tennessee Dance	Composer:JABOVSKI, Tony	Works on CD

Styles are explained further below.

- `webui.itemdisplay.enregistrement = *dc.title=title, \`

This set of parameters controls the fields displayed for given set of collections. We added a few new features to the field list:

- A star “*” indicates a section boundary within the display.
To get the item display below, `dc.publisher` and

dc.relation.ispartofcollection are both prefixed with a star:

	xylophone 1	166	68 sol.
Publisher:	Alfred	716	+1094
Place of publication:	Van Nuys, Etats-Unis		
Issue date:	2008		
Collection:	Alfred's Concert Band Series	60	
Publisher Reference:	29479		

- After the field designation, you can have:
 - =style to use the named “style” to display the field in one “full width” column
 - [style] to. use the named “style” to display the field in two columns (one for the content, one for the links)
 - {style} to. use the named “style” to display the field in three columns (the field name column for the roles, one column for the content, one column for the links)

Styles are explained further below.

A complete example:

```
webui.itemdisplay.enregistrement =*dc.title=title, \
  dc.title.alternative=othertitle, \
  dc.description.tableofcontents=text, \
*dc.subject.classification[code], \
  dc.type.score[code], \
  dc.language.iso[code], \
*dc.contributor.author{contributor}, \
  dc.contributor.arranger{contributor}, \
  dc.contributor.performer{contributor}, \
  dc.contributor{contributor}, \
*dc.description.abstract=abstract, dc.description=abstract, \
  dc.subject.SORTED[subject], \
*dc.audience[code], \
*dc.subject.orchestra[code], \
  dc.subject.additionalinstrument[code], \
  dc.subject.soloist[subject], \
  dc.subject.voice[code], \
  dc.subject.choir[code], \
  dc.subject.instrument[subject], \
  dc.subject.ensemble[code], \
  dc.description.instrumentation, \
*dc.identifier.publisher, \
  dc.format.medium, \
  dc.format.extent=duration, \
*dc.relation.isformatof[link], \
  dc.relation.haspart[link], \
  dc.relation.ispartof[link], \
  dc.relation.listening=listening, \
  dc.relation.seealso[link], \
  dc.relation.isversionof[link], \
  dc.source.uri=url, \
*dc.identifier.other=cim, \
  dc.identifier.loris, dc.identifier, dc.type.record[code], \
  dc.identifier.isbn, \
  dc.identifier.ismn, \
  dc.date.submitted=date, \
  dc.identifier.uri=inbound
```

New parameters:

- `skos.identification = windmusic`
This name is used by the SKOS management layer as the “application name” when keeping counts of the usage of each concepts by the different indexes of the different applications. The goal is to ensure that, in any application (DSpace or not), it is easy to display the number of times a concept is used by the different applications (and to link there).
- `linkout.html.style=field presentation template`

An example:

```
linkout.html.field = [value]
```

This is the default style which simply displays the field content.

```
linkout.html.date = [datevalue]
```

Dates `yyyymmdd` can be displayed nicely (like regular DSpace)

Publisher e-mail: plavender@halleonard.com

Internet Web Site: <http://www.halleonard.com/>

```
linkout.html.url = <a target="_blank" href="[PLAINvalue]">[value]</a>
```

For a field containing an URL, the `PLAINvalue` ensures no HTML entities are generated.

```
linkout.html.email = [valueEMAILS]
```

Emails can be detected in the middle of a text and displayed nicely using `linkout.html.EMAILS` below:

```
linkout.html.EMAILS = <a href="mailto:$1" style="text-decoration:none;color:MediumSeaGreen;font-weight:bold">$1</a>
```

Keywords:	 DANSE	▶-1364 ▶▶-3966
	 ARRANGEMENT	▶-12462 ▶▶-12471
	 ORCHESTRE D'HARMONIE	▶-15506 ▶▶-15506
	 GRECE	▶-48 ▶▶-48
	 20E SIECLE-1E MOITIE	▶-1300 ▶▶-1314
	 MELODIE POPULAIRE	▶-12 ▶▶-12

```
linkout.html.subject = [linkval THES]
```



```
linkout.html.ispartof = [linkval THES]
```

Each entry will be translated using the associated SKOS Concept.

Links are generated using the `linkout.html.THES` parameter:

```
linkout.html.THES=[schemeIcon][prefix] [label] [suffix]</td><td class="FieldValueCell"><small>[localSearch]</small>
```

`localSearch` uses the templates `LATERAL` and `FILTER` to generate searches based on the displayed concept.

Title:  Art in the park Smil. 

```
linkout.html.title = <font size="+1">[value]</font> <a href="[root]/simple-search?query=[URLvalue]" style="text-decoration:none;font-size:66%;color:MediumSeaGreen">Smil.</a> [linkage INBOUND]
```

The title is displayed prefixed with the `.GIF` image associated with the collection containing current item.

If the collection is used to store a SKOS Scheme, a search is proposed to

look at items using this Scheme: the linkout.html.INBOUND template is used for this.



```
linkout.html.contributor = [linkval CONTRIBUTOR]
```

Each contributor will be retrieved in its SKOS Scheme and processed by linkout.html.CONTRIBUTOR

```
linkout.html.CONTRIBUTOR=[prefixField]:&nbsp;</td><td
class="FieldValueCell">[schemeIcon]<a
href="[root]/handle/68502/[about]?searchname=[context]&backlink=[hand
le]">[label]</a></td><td
class="FieldValueCell"><small>[localSearch]</small>
```

This template generates three columns (the first one being within the field title column) for each contributor to the item.

```
linkout.html.INBOUND=[schemeIcon][prefix][localSearch]
```

This template processes the SKOS Concept represented by the current item. [localSearch] uses the templates below to generate searches starting from this concept:

```
linkout.html.LATERAL=&nbsp;<a href="[root]/simple-
search?query=%22[URLvalue]%22 NOT
identifier:%22[URLvalue]%22&searchname=[context]&backlink=[handle]"><
small>[label][quantity]</small></a>[valueFILTER]
```

This template is used to search the items using a given SKOS Concept.

```
linkout.html.FILTER=<a href="[root]/simple-
search?query=[URLquery]%20AND%20%22[URLvalue]%22&searchname=[URLquery
]%20[context]&backlink=[handle]">[label]</a>
```

This template is used to drill down the current search using the current SKOS Concept.

```
webui.itemlist.html.title= <a
href="[root]/handle/[handle]"><b>[value]</b></a>
webui.itemlist.html.type=[linkval TYPE]
webui.itemlist.html.TYPE=<small>[label]</small>
webui.itemlist.html.author=[linkval AUTHOR]
webui.itemlist.html.AUTHOR=[prefix]:[label]
```

These styles (templates) are formatting fields within search results:

n°	Titre	Author	Type
141	The Tennessee Dance	Composer:JABOVSKI, Tony	Works on CD

- `search.sort = available`
Default sort for search results
- The search indexes and sort indexes explained in <http://wiki.dspace.org/index.php/SortingSearchResults> have been implemented but slightly differently.
There is four types of indexes:
 - Search indexes (tokenized, not stored)
 - Sort indexes (not tokenized, stored)
 - Date indexes (not tokenized, stored)
 - Number indexes (not tokenized but right aligned with zeroes, stored)

Example:

```

search.index.title=dc.title,dc.title.alternative
search.index.author=dc.contributor.author,dc.contributor.arranger,dc.
contributor.performer
search.index.keyword=dc.subject
broadsearch.index.broadkeyword=dc.subject
...
# Sort keys:
# 1. Call number + First author + Title
sort.index.cote=dc.identifier.other,dc.contributor.author,dc.title
# 2. Composer + Title
sort.index.composer=dc.contributor.author,dc.title
# 3. Decreasing publication date
date.index.issued=dc.date.issued
# 4. Decreasing End of processing Date
date.index.available=dc.date.submitted
# Simple management sort :
sort.index.cim=dc.identifier.other
number.index.loris=dc.identifier.loris
...

```

- The potential result of a search (using a given index) may be displayed. The following templates provide the required formatting:

```

index.html.author={0} &nbsp;
index.html.keyword={0} &nbsp;
index.html.broadkeyword={0}
&nbsp;

```

SKOS directory

A parameter (web.xml ? dspace.cfg ?) should allow to place this directory anywhere. For the moment, it is a sub-directory of the "config" directory of the DSpace application.

For instance, SKOS directory may contain the following files:

communitiesCollections.cfg	Configuration of the SKOS Scheme "communitiesCollections" (XML file)
communitiesCollections.xml	Multilingual names of all DSpace Communities and Collections
dc.cfg	Configuration of the SKOS Scheme "dc" (XML file)
dc.xml	Properties and multilingual names of all metadata fields in the DC schema
documentType.cfg	Configuration of the SKOS Scheme "documentType" (Tab delimited Text file)
documentType-labels.txt	Labels for the SKOS Concepts in the Scheme "documentType"
dspaceSearch.cfg	Configuration of the SKOS Scheme "dspaceSearch"
dspaceSearch-labels.txt	Labels for all searches available in the "Advanced Search" box. Each search can use different indexes and different SKOS Schemes for auto-complete. Advanced Search .JSP does not have to be modified anymore.
dspacesort.cfg	Configuration of the SKOS Scheme "dspaceSort"
dspacesort-labels.txt	Labels for all search results sort
language.cfg	Configuration of the SKOS Scheme "language" (XML)
language.xml	Languages... in different languages! SKOS Collection of languages available for the user's interface of WindMusic
userLanguage.cfg	Configuration of the SKOS Notation Scheme "loris" which provides an alternate access by record number of previous OPAC
loris.cfg	Configuration of the SKOS Scheme "musique"
musique.cfg	Labels for the SKOS concepts in the "musique" Scheme
musique-labels.txt	Configuration of the SKOS Scheme "partition" (Dynamic SQL source)
partition.cfg	Configuration of the SKOS Scheme "personne" (Dynamic SQL source)
personne.cfg	Configuration of the SKOS Scheme "prefixes" (XML)
prefixes.cfg	Labels of all allowed prefixes for a concept reference
prefixes.xml	XSLT to display the Controlled Vocabulary Index
SKOS2html.xsl	Configuration of the SKOS Scheme "suffixes" (XML)
suffixes.cfg	Labels of all allowed suffixes for a concept reference
suffixes.xml	Configuration of the SKOS Scheme "sujet" (Pre-loaded SQL source)
sujet.cfg	Configuration of the SQL connection pool for the "WindMusic" database
wind-pool.cfg	

All these files are accessible in the directory <http://www.askosi.org/dspaceSKOS>

SKOS XML

When using a XML SKOS source, the .CFG file contains only one line:

```
type=XML
```

SKOS is standardized on RDF. We did not implement up to now the OpenRDF RIO parser because we wanted first to have a very fast XML parsing for big thesauri like MeSH and Agrovoc.

The SKOS XML format is conforming to the following XSD:

<http://www.askosi.org/ConceptScheme.xsd>

The XML tags follow the terminology and the structure of the SKOS standard.

Example with <http://www.askosi.org/dspaceSKOS/communitiesCollections.xml> :

```
<conceptScheme about="communitiesCollections">
<title lang="de">Categorien</title>
<title lang="es">Categorias</title>
<title lang="en">Categories</title>
<title lang="fr">Catégories</title>

<concept about="1">
<prefLabel lang="fr">1. Partitions pour Orchestres à vent</prefLabel>
<prefLabel lang="en">1. Wind Music Scores</prefLabel>

<scopeNote lang="fr">1. TEXTE D'ACCUEIL pour une catégorie...</scopeNote>
</concept>

<concept about="2">
<prefLabel lang="fr">1.1. Partitions pour Orchestres d'harmonie</prefLabel>
</concept>

<concept about="3">
<prefLabel lang="fr">1.2. Partitions pour Batteries-Fanfaires</prefLabel>
</concept>

<concept about="4">
<prefLabel lang="fr">1.3. Partitions pour Big Band</prefLabel>
</concept>

...
<concept about="23">
<prefLabel lang="fr">6. Autour des œuvres musicales</prefLabel>
</concept>

<concept about="24">
<prefLabel lang="fr">6.1. Auteurs (Biographies)</prefLabel>
<prefLabel lang="en">6.1. Authors (Biographies)</prefLabel>
<prefLabel lang="es">6.1. Autor</prefLabel>
<prefLabel lang="de">6.1. Autor</prefLabel>
<prefLabel lang="nl">6.1. Schepper</prefLabel>
</concept>

<concept about="25">
<prefLabel lang="fr">6.2. Éditeurs</prefLabel>
<prefLabel lang="de">6.2. Verlagsangabe</prefLabel>
<prefLabel lang="en">6.2. Publishers</prefLabel>
<prefLabel lang="es">6.2. Casa editorial</prefLabel>
<prefLabel lang="nl">6.2. Uitgever</prefLabel>
</concept>
```

```

<concept about="26">
<prefLabel lang="en">6.3. Orchestra</prefLabel>
<prefLabel lang="fr">6.3. Orchestres</prefLabel>
<prefLabel lang="de">6.3. Orchester</prefLabel>
<prefLabel lang="es">6.3. Orquestas</prefLabel>
<prefLabel lang="nl">6.3. Orkest</prefLabel>
</concept>

<concept about="27">
<prefLabel lang="en">6.4. Keywords (index)</prefLabel>
<prefLabel lang="fr">6.4. Mots-clés (index)</prefLabel>
<prefLabel lang="de">6.4. Stichwörter</prefLabel>
<prefLabel lang="es">6.4. Palabras claves</prefLabel>
<prefLabel lang="nl">6.4. Trefwoord</prefLabel>
</concept>

<concept about="28">
<prefLabel lang="fr">6.5. Collections</prefLabel>
<prefLabel lang="en">6.5. Collections</prefLabel>
</concept>
</conceptScheme>

```

A very important XML Scheme is dc.xml which contains not only terminology but also presentation + validation parameters for all metadata fields:

<http://www.askosi.org/dspaceSKOS/dc.xml>

SKOS Tab delimited Text :

When using a XML SKOS source, the .CFG file contains:

```
type=TAB
```

and also the labels for the Scheme itself (not for the concepts : they are in the Tab delimited Text files) :

```

title-en=Music Type
title-es=Tipo de música
title-fr=Type de musique
title-nl=Soort muziek
title-de=Art von muzik

```

This example was already shown for SKOS Scheme “musique” in section “Storage of this record in DSpace” above. <http://www.askosi.org/dspaceSKOS/musique-labels.txt>

Tab Delimited Text files can also be used to define notations, aliases, scope notes, related concepts, broader concepts, etc.

SKOS Pre-loaded SQL

We implemented first an SQL source (and not a DSpace specific one) because:

- It is the most efficient
- It may access other databases than the one managed by DSpace
- SQL is known by most programmers

An example of connection pool definition can be found at <http://www.askosi.org/dspaceSKOS/wind-pool.cfg>

Example of a CFG file to access a DSpace collection and preload it as an SKOS Scheme:

```
type=SQL
pool=wind
title-en=Keywords
title-fr=Mots-clés
title-de=Stichwörtern
title-es=Palabras claves
title-nl=Trefwoord
## about, label-fr, label-en...,
labels=SELECT h.handle AS about, t.text_value AS label from item as m,
handle as h, metadatavalue as t where h.resource_type_id=2 and
h.resource_id=m.item_id and t.item_id=m.item_id and
t.metadata_field_id={IDdc@title} and m.owning_collection={IDhandle@27}
## about, broader
broaders=SELECT h.handle AS about, t.text_value AS broaderLoris, k.handle
AS broader from item as m, handle as h, handle as k, metadatavalue as i,
metadatavalue as t where i.metadata_field_id={IDdc@identifier.loris} and
k.resource_type_id=2 and k.resource_id=i.item_id and t.item_id=m.item_id
and t.metadata_field_id={IDdc@relation.ispartof} and
m.owning_collection={IDhandle@27} and h.resource_type_id=2 and
h.resource_id=m.item_id and substring (lower(t.text_value) from 1 for 50) =
substring (lower(i.text_value) from 1 for 50)
```

SKOS Dynamic SQL

Dynamic SQL SKOS Source is a very efficient way to control a metadata field with dynamic content coming from any database.

Example of configuration for the “personne” (authors):

```
type=DynSQL
pool=wind
title-en=People
title-fr=Personnes
title-de=Person
title-es=Personas
title-nl=Persoon
# Find a concept by its key (about)
about=SELECT resource_id AS id, handle AS about FROM handle,item WHERE
handle='68502/'||? AND resource_type_id=2 AND resource_id = item_id AND
item.owning_collection={IDhandle@24}
size=SELECT count(*) AS size FROM item WHERE
item.owning_collection={IDhandle@24}
# find labels for a concept
labels=SELECT '' AS language, t.text_value AS label from item as m,
metadatavalue as t where m.item_id=cast(? as integer) and
t.item_id=m.item_id and t.metadata_field_id={IDdc@title}
# find synonyms for a concept
alternates=SELECT '' AS language, t.text_value AS alternate from item as m,
metadatavalue as t where m.item_id=cast(? as integer) and
t.item_id=m.item_id and (t.metadata_field_id={IDdc@title.alternative})
# find concepts matching a part of a label or synonym
match=SELECT h.handle AS about, m.text_lang AS language, m.text_value AS
label, m.text_value AS match FROM handle AS h, metadatavalue AS m, item AS
i WHERE m.metadata_field_id in ({IDdc@title},{IDdc@title.alternative}) AND
m.text_value ILIKE ? AND h.resource_type_id=2 AND h.resource_id = m.item_id
AND i.item_id = m.item_id AND i.owning_collection={IDhandle@24}
```

A Test using Agrovoc from FAO:

The following has been tested:

- 1) Download of Agrovoc in SKOS format
- 2) Conversion to XML
- 3) Installation within an existing DSpace application. From DSpace log file:
02:44:34,718 DEBUG SchemeFactory:101 - Load C:\DSpace\config\SKOS\c.cfg
02:44:34,718 DEBUG SchemeFactory:214 - Source type=XML
02:44:41,343 DEBUG SkosManager:378 - Reload of scheme c ?
02:44:41,515 DEBUG SkosManager:358 - 28953 concepts loaded for scheme c (FAO Multilingual Thesaurus.)
 - The whole thesaurus (all languages, all synonyms) is loaded in about 7 seconds elapsed time (3 GHz, 3 years old portable PC with MS-Windows XP)
 - Tomcat is using 300 megs of RAM but there is also many other thesaurus loaded and 80 thousand records in the repository.
- 4) Normalization of the repository has taken a day for more than 80 thousand records (15 keywords per record) to replace terms by their codes and index them: this certainly can be improved.

The result is interesting (aligning and indexing MeSH data based on Agrovoc):

A matching was done between the terms in dc.subject.mesh and the Agrovoc thesaurus (only a part of the keywords matched). This is done with an «automatic normalization» and modifies the database by replacing the terms by their concept codes (c_9999).

For example (“detailed” item display in DSpace):

dc.language	eng	-
dc.relation.ispartofseries	Current allergy and asthma reports	-
dc.subject.mesh	Allergens[+]	en_US
dc.subject.mesh	Allergens[administration_and_dosage]	en_US
dc.subject.mesh	Allergens[immunology]	en_US
dc.subject.mesh	c_444	en_US
dc.subject.mesh	c_1391	en_US
dc.subject.mesh	c_2502	en_US
dc.subject.mesh	Food Hypersensitivity[+]	en_US
dc.subject.mesh	c_4586	en_US
dc.subject.mesh	Inhalation	en_US
dc.subject.mesh	c_4826	en_US
dc.subject.mesh	c_6946	en_US
dc.title	Allergic reactions to foods by inhalation.	en_US
dc.type	Journal Article	en_US
dc.type	Review	en_US

The original record can be viewed in <http://www.pubmed.org> , searching for PMID **17448326**

Displaying a Record:

Concept codes are dynamically translated in item displays. The example item above is displayed (in English):

Title:	Allergic reactions to foods by inhalation.
Authors:	James, John M Crespo, Jesús Fernández
E-Mail:	jm.james@coloradoallergy.com
Affiliation:	Colorado Allergy and Asthma Centers, 1136 East Stuart Street, Suite 3200, Fort Collins, CO 80524, USA.
Abstract:	Although allergic reactions to foods occur most commonly after ingestion, inhalation of foods can also be an underlying cause of these reactions. For example, published reports have highlighted the inhalation of allergens from fish, shellfish, seeds, soybeans, cereal grains, hen's egg, cow's milk, and many other foods in allergic reactions. Symptoms have typically included respiratory manifestations such as rhinoconjunctivitis, coughing, wheezing, dyspnea, and asthma. In some cases, anaphylaxis has been observed. In addition, there have been many investigations of occupational asthma following the inhalation of relevant food allergens. This report reviews the current literature focusing on allergic reactions to foods by inhalation.
MESH:	Allergens[+] Allergens[administration_and_dosage] Allergens[immunology] Animals Cattle Eggs Food Hypersensitivity[+] Humans Inhalation Milk Seeds
Substances:	Allergens

Controlled Vocabulary Search using Agrovoc:

The complete view of Agrovoc (Controlled Vocabulary Search) is generated in 4 seconds

- The JavaScript for closing/opening hierarchies is slower than the server to generate the whole thesaurus! (an optimization of this JavaScript will have to be made for this)
- [Image of the Agrovoc Hierarchy when displayed as a DSpace Subject hierarchy:](#)

Subject Search

[Help...](#)

Check the boxes next to the categories that you wish to search under, then hit "Search...". Categories can be expanded to refine the search terms, and as many categories can be selected as required.

Filtering the list of categories will remove from the list below any categories that do not match the filter term. Expanding each category will show you which terms did match the filter.

Find a subject in the controlled vocabulary:

Filter:

- Acceptable daily intake
- Acquired characters
- Acupuncture
- Additives
 - Acidulants
 - Anticaking agents
 - Bulking agents
 - Emulsifiers
 - Enzyme preparations
 - Feed additives
 - Food additives
 - Humectants
 - Stabilizers
- Administration
 - Devolution
 - Public administration
- Africa
 - Africa South of Sahara
 - Anglophone Africa
 - Francophone Africa
 - North Africa
 - Portuguese speaking Africa
- Aggregate data
 - Value added
- Agrarian countries
- Agricultural sector
- Agricultural structure

The size of the potential result has been added since this test (see above).

“Auto-Complete”:

- An “auto-complete” helps choosing a concept to search:

The screenshot shows a search interface titled "Recherche avancée" with a dropdown menu set to "Tout le dépôt". Below this, there are three rows of search criteria, each with a dropdown menu set to "Partout" and a search term field. The first search term field contains "hu". A dropdown menu is open below the first search term field, displaying a list of concepts with their occurrence counts in parentheses:

- Alcool (fr)Huile de fusel (17)
- Antigène (sk)hubové antigény (2)
- Appétit (de)HUNGER (PHYSIOLOGIE)
- Chien (de)HUND (412)
- Comportement humain (en)Human behaviour (13)
- Densité de population (sk)hustota populácie (2)

Please notice the number of occurrences of a concept in the database, indicated between parentheses. This works also with other alphabets than latin.

The “auto-complete” list changes based on the search index selected by the user (drop-down list of available search indexes).

- Another “auto-complete” helps choosing a concept when updating a record to add to metadata:

The screenshot shows a metadata update interface. A table with a "MESH" column is visible. The first row has the value "lacta" in the MESH field. A dropdown menu is open below the "lacta" field, displaying a list of concepts with their occurrence counts in parentheses:

- Albumine (de)LACTALBUMIN (2)
- Alimentation au sein (es)Lactancia materna (172)
- Allaitement (es)Lactancia
- Beta galactosidase (de)LACTASE (BETA GALAKTOSIDA)
- Durée de la lactation (en)Lactation duration
- Induction de la lactation (es)Lactación inducida

To the right of the table, there are buttons for each row: an equals sign (=) and a red X.

(For testing purposes, we overloaded an existing MESH field with Agrovoc links)

Searches:

- All synonyms and translations are indexed in the DSpace “full text” index: this means that typing any Agrovoc term (or word of a term) in any language in the “simple search” box, will retrieve all records using the corresponding concept(s).

Example with “mleko” (Slavonic word for milk):

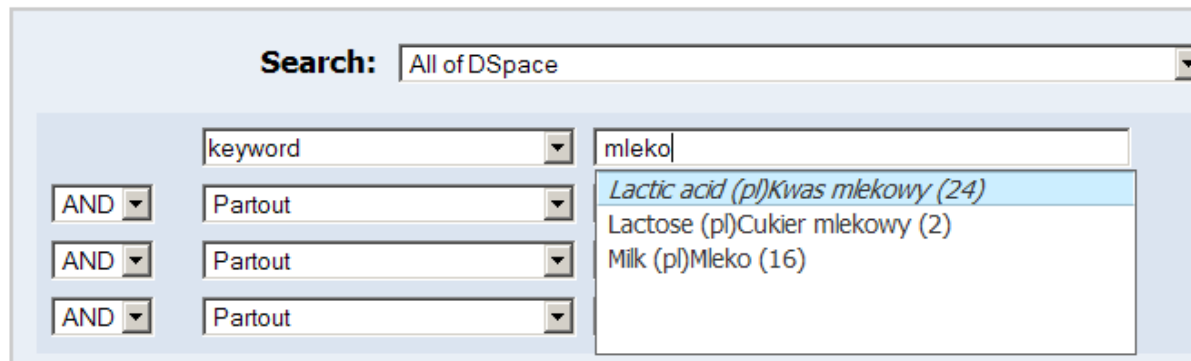
- Searched as a truncated word, the result is:



The screenshot shows a search interface with a dropdown menu set to "All of DSpace". Below it, a search box contains "mleko*" and a "Go" button. The text "mleko*" is also displayed below the search box.

Results 1-20 of 42.

- Searched as a word within a keyword:



The screenshot shows a search interface with a dropdown menu set to "All of DSpace". Below it, there are three rows of search criteria, each with a dropdown menu set to "keyword" and a "Partout" dropdown. The first row has the search term "mleko" in the input field. The results list shows three items: "Lactic acid (pl)Kwas mlekowy (24)", "Lactose (pl)Cukier mlekowy (2)", and "Milk (pl)Mleko (16)".

One may notice here that $24 + 2 + 16 = 42$.

- If the third concept is chosen (Milk), the result is:



The screenshot shows a search interface with a dropdown menu set to "All of DSpace". Below it, a search box contains "keyword:c_4826 ((keyword:Milk))" and a "Go" button.

Results 1-16 of 16.

Last but not least: The authority lists can be pre-loaded (menus, short lists) but they can also be dynamically loaded from any SQL source (strong caching already implemented). This means that very big authority lists can be managed without problems.

For instance the DSpace repository itself (or specific collections) can be an authority list which constrains the relation between DSpace records.

About DESTIN inc. SSEB:



Founded in Montreal (Canada) by Christophe Dupriez in 1986, established in Belgium since 1990, DESTIN is small company providing document management solutions to Quebec, European and Belgium Federal institutions (RIZIV/INAMI, Amazone, Poison Centre) and some private organizations.

We are deeply involved in the DSpace project with two big repositories running:

- 3) The internal scientific documentation database of the Belgium Poison Centre;
- 4) <http://www.windmusic.org/dspace> for the CDMC Haute Alsace (Guebwiller, France)

We would be happy to collaborate to other projects providing services to make things work perfectly for local needs. The required source code is provided freely to DSpace community as long as authorship is mentioned and improvements are also made available to the community.

This publication of the source code requires a momentum built over multiple projects:

- 1) Documentation (users, configuration, adaptation);
- 2) Source code organization in different independent patches;
- 3) The implication of a DSpace Committer to merge essential parts in DSpace SVN and possibly trunk.

We hope to see this momentum arise from the ongoing and foreseen projects.

Christophe Dupriez

*DESTIN inc. SSEB, c.o. m-Village, rue des Palais 44 boîte 82,
B-1030 Schaarbeek (Brussels, Belgium)*

<http://www.destin.be> Tel. +32 475 77 62 11, dupriez@destin.be