www.ASKOSI.org

Application Services
for Knowledge Organisation
and System Integration

A Short Presentation

May 2010

Christophe Dupriez
dupriez@askosi.org

Thesauri:
Take a walk on the « Why? » slide !

❖ **Search** : Linking users words to
the concepts used in the Information System […]

❖ **Presentation** : […] and vice-versa.

❖ **Translation** : Multilingual terms equivalence for a concept.

❖ **Integration** : Semantical equivalence between separately
designed systems.

❖ **Classification** : Hierarchies to organize data along different
facets.

❖ **Validation** : Authority lists for controlled data entry.

❖ **Training** : Newcomers can learn the Domain of Discourse of
the Organisation.

❖ **CoP Motivation** : Solving terminology problems is one of
the best way to involve specialists in discussions,
negotiations, consensus building around the concepts of a
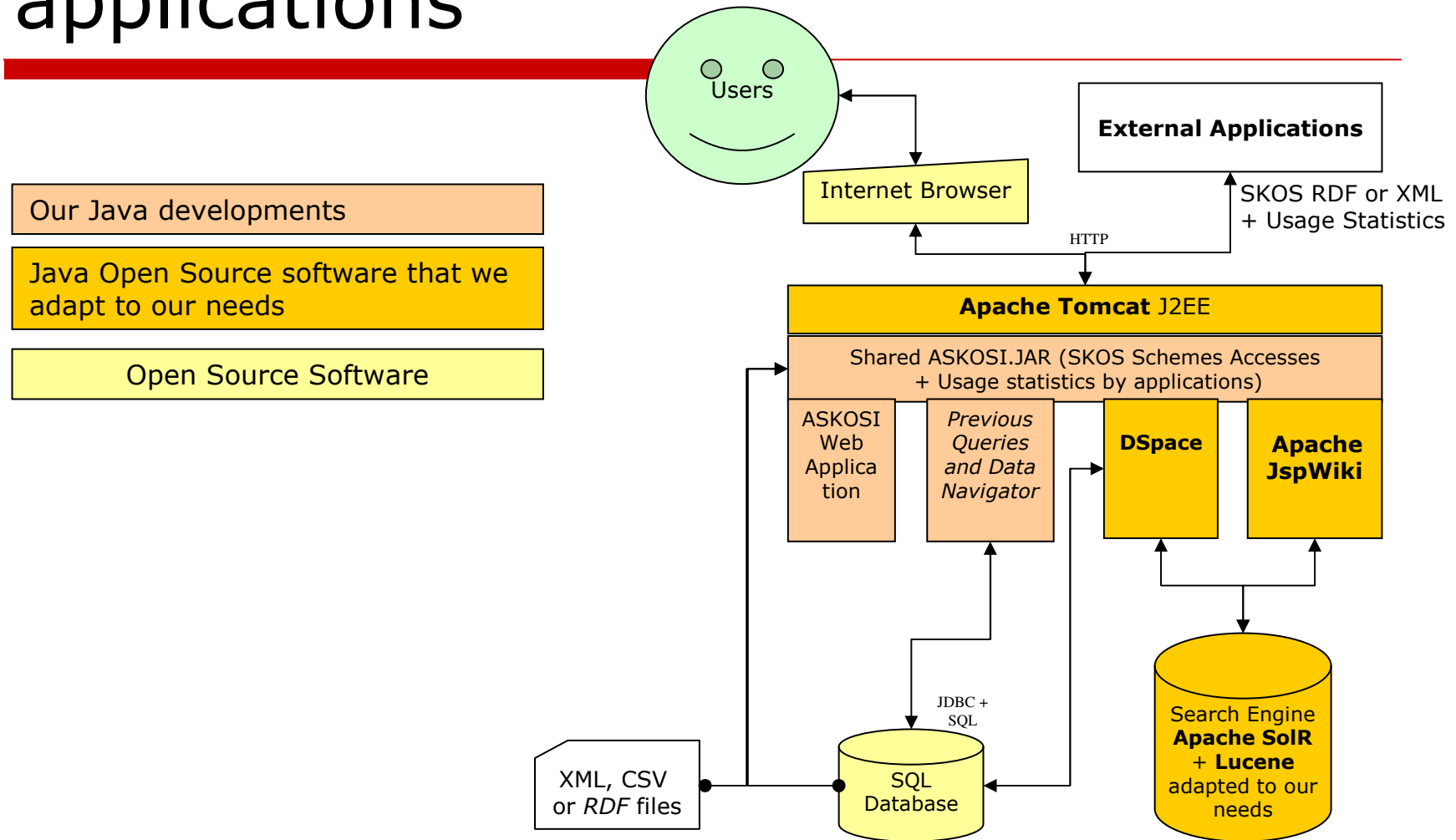scientific or technical domain.

What is ASKOSI?

A project to create Java tools to integrate (to benefits of) terminology management within applications…

1. A Java Archive (JAR)
2. A Web Application
☐ To be developed: A Web Service and/or a RMI Server

Integrating ASKOSI with other applications

Users

Internet Browser

External Applications

SKOS RDF or XML + Usage Statistics

Our Java developments

Java Open Source software that we adapt to our needs

Open Source Software

HTTP

Apache Tomcat J2EE

Shared ASKOSI.JAR (SKOS Schemes Accesses + Usage statistics by applications)

ASKOSI Web Application

Previous Queries and Data Navigator

DSpace

Apache JspWiki

JDBC + SQL

XML, CSV or *RDF* files

SQL Database

Search Engine **Apache SolR** + **Lucene** adapted to our needs

The ASKOSI JAR

- API aligned on W3C SKOS data structure
    - ISO 25964 is considered.
- SQL, CSV and XML data sources
- Dynamic or Static (reload) access to the data sources with SKOS goggles
- Usage statistics: which applications are using which SKOS concepts, how (roles) and how many times?
- Designed for data sharing: all applications in the same Web Application Container (J2EE) access a single copy of the data.

The ASKOSI Web Application

- ☐ Authority lists browsing:
  - ■ Thesauri trees
  - ■ Alphabetical
  - ■ Usage Frequency
- ☐ Powerful word and string search tool
- ☐ SKOS Concepts display in different formats / extents
- ☐ Generation of SKOS RDF
- ☐ Validations:
  - ■ Data errors
  - ■ Terminology validations (ambiguity, *missing translation*)
  - ■ Hierarchy validations (loops, *siblings*)
- ☐ Links to applications using the SKOS concepts
- ☐ In development: search history manager

SKOS Schemes
used by WindMusic.org

alphabetical [frequency] sort

27 concept schemes registered:

33224 Scores partition
WindMusic: linkedscore26

28070 Recordings ocd
WindMusic:

19143 People personne
WindMusic: Composer/Performer:5941   Performer:2338   Nationality:5941

3218 CDs cd
WindMusic: Contains:178

1930 Publisher editeur
WindMusic: Publisher:12565   Contains:18434

1617 Collection collection
WindMusic: Collection:506

1588 Orchestra orchestre
WindMusic: Composer/Performer:2590   Performer:2590   Nationality:2590

1024 Keywords sujet , max.depth ↓6
ASKOSI: 1117   1117
WindMusic: Keyword:233848   Keyword+NTT:148297   Contains:1072

255 Country country , 4 errors
WindMusic: Publisher:6547   Nationality:50848

186 Languages language
WindMusic: Language:35321

160 Instruments instrument , 4 errors
WindMusic: Instrument:45514   Soloist:10214

61 metadataPropertys for DC prefLabelspace dc

- Scores, Recordings, Persons, CDs, Publishers, Collections, Orchestras: database objects can be the source of an Authority List
- The Keywords Thesaurus is 6 levels deep
- Countries, Languages, Instruments: these closed lists control different database fields.
- Even the DC (Dublin Core) database fields are described in a SKOS Scheme.

http://www.windmusic.org/askosi/toc.jsp

Keywords, A Thesaurus Scheme

Scheme list: alphabetical or [hierarchy]

⊙1024 concepts ∈ sujet , max.depth ↓6

Keywords

* ☑ [   ] * by pieces: ☐ 🔍

☛ nl Trefwoord
☛ fr Mots-clés
☛ de Stichwörtern
☛ es Palabras claves

Last loaded: Mon May 24 12:46:21 CEST 2010[Reload]

ASKOSI:
　1117
　1117
WindMusic:
　Keyword:233848
　Keyword+NTT:148297
　Contains:1072

[Check duplicates]

⊙**1024 concepts:**

↕ Keywords	Keyword:	Keyword+NTT:	Contains:	👤	🔺
1　Artistic profession	26	104	31		7
2　ENFANT ARTISTE				1	
2　INTERMITTENT DU SPECTACLE				1	
2　METIER DE LA DANSE		7	3	1	3
3　CHOREGRAPHE	5	5		1	
3　DANSEUR	2	2		1	
3　THEORICIEN DE LA DANSE				1	
2　METIER DE LA SCENE			4	1	4

http://www.windmusic.org/askosi/view.jsp?scheme=sujet&display=concepts

"Singing", A Thesaurus Concept

Concept ⊟22594 ∈ sujet:

Singing

- fr CHANT
- en Singing
- de Gesang
- nl Zang
- es Canto

Synonyms:

Cántico

	Keywords	Keyword:	Keyword+NTT:	Contains:	👤	👥
1	Education	14	5795	80		5
2	Education in music	16	2279	18	1	8
3	Aural training	107	107		1	
3	Initiation into music	23	174	3	2	2
4	Listening	57	67	1	2	1
3	Listening ↑	57	67	1	2	1
3	Musical game	7	7		1	
3	Notes reading	142	142		1	
3	Rhythm	269	308	1	1	1
3	Rhythm education	1	1		1	
3	Singing	630	1699	7	2	6
4	Choir	567	567		1	
4	Folk songs	507	523	1	2	1
5	Calypso	16	16		2	
4	Lay				2	
4	Seguedilla	2	2		2	
4	Shanty	26	26		2	
4	Singing exercise	6	6		1	
1	Music	156	57786	369		9
2	Musical writing	30	50633	60	1	15
3	Repertoire	1984	50092	33	1	11
4	Adaptation	60	60		1	
4	Arrangement	15676	15686	2	1	1
4	Character piece	8	8		1	
4	Composer	180	578	11	1	11
4	Condensed score	80	240	1	2	1
4	Excerpt	111	111		1	
4	Orchestral excerpt	20	20		2	
4	Original work	31615	31615		1	
4	Singing ↑	630	1699	7	2	6
4	Transcription	235	235		1	
4	World music	35	35	1	1	1

BT	👤	More general concept, something "above" in the tree
BTT	👥	All the more general concepts, all of the "above" in the tree
NT	👥	More specific concept, something "lower" in the tree
NTT	👥	All the more specific concepts, all of the "lower" in the tree

Listening ↑, Singing ↑
indicate previous apparitions of the concept in the hierarchy

http://www.windmusic.org/askosi/concept.jsp?concept=sujet_22594

Possible Contributions of ASKOSI to your projects

- ☐ Concept Surrounding Display
- ☐ Concepts' Usage analysis
- ☐ Dynamic SKOSification of existing applications: an SQL Data Source
  - ■ Generated SKOS file
- ☐ Example of improvements to existing applications:
  - ■ DSpace

Parameters to "SKOSify" the SQL Data Source for the WindMusic Thesaurus

type=SQL
pool=wind
title-en=Keywords
title-fr=Mots-clés
title-de=Stichwörtern
title-es=Palabras claves
title-nl=Trefwoord
title.lorthes-en=Keywords
…

url=jdbc:postgresql://dbserver:5432/dspace
driver=org.postgresql.Driver
username = dspace
password = xxxxxxxxx
validation=SELECT 1 #Oracle: SELECT 1 FROM DUAL
IDdc=select … as key, metadata_field_id as value
        from metadatafieldregistry;
IDhandle=select … as key, resource_id as value from handle;

display-en=http:/dspace/handle/68502/**[about]**
icon-en=/dspace/image/68502/27.gif
create-en=http:/dspace/submit?post=yes&collection=**{IDhandle@27}**&step=0
…
notation.lorthes=SELECT h.handle AS **about**, i.text_value AS **notation**
        from item as m, handle as h, metadatavalue as I
        where i.metadata_field_id=**{IDdc@identifier.loris}**
            … and m.owning_collection=**{IDhandle@27}**
labels=SELECT h.handle AS **about**, t.text_value AS **label**, t.text_lang AS **lang**
        from item as m, handle as h, metadatavalue as t
        where h.resource_type_id=2
            … and t.metadata_field_id=**{IDdc@title}**
        and m.owning_collection=**{IDhandle@27}**
…alternates…broaders…broadmatches…notes…

Generated SKOS: Preamble

```xml
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xmlns="http://www.windmusic.org/askosi/skos.jsp?about=sujet#"
xml:base="http://www.windmusic.org/askosi/skos.jsp?about=sujet"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:skos="http://www.w3.org/2004/02/skos/core#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:zbwext="http://zbw.eu/namespaces/skos-extensions/">
  <skos:ConceptScheme rdf:about="http://www.windmusic.org/askosi/skos.jsp?about=sujet">
    <dc:title xml:lang="nl">Trefwoord</dc:title>
    <dc:title xml:lang="fr">Mots-cl&#195;&#169;s</dc:title>
    <dc:title xml:lang="de">Stichw&#195;&#182;rtern</dc:title>
    <dc:title xml:lang="es">Palabras claves</dc:title>
    <dc:title xml:lang="en">Keywords</dc:title>
    <dcterms:available>Wed May 05 10:20:07 CEST 2010</dcterms:available>
  </skos:ConceptScheme>
…
```

Validates in Schemarama, **SKOS Thesaurus Compatibility**

Generated SKOS: a Concept...

```xml
<skos:Concept rdf:about="#21942">
    <skos:inScheme rdf:resource="http://www.windmusic.org/askosi/skos.jsp?about=sujet" />
    <skos:prefLabel xml:lang="en">Music</skos:prefLabel>
    <skos:prefLabel xml:lang="fr">MUSIQUE</skos:prefLabel>
    <skos:prefLabel xml:lang="de">Musik</skos:prefLabel>
    <skos:prefLabel xml:lang="nl">Music</skos:prefLabel>
    <skos:prefLabel xml:lang="es">M&#195;&#186;sica</skos:prefLabel>
    <skos:narrower rdf:resource="#22080" />
    <skos:narrower rdf:resource="#22215" />
    <skos:narrower rdf:resource="#22239" />
    <skos:narrower rdf:resource="#22245" />
    <skos:narrower rdf:resource="#22264" />
    <skos:narrower rdf:resource="#22318" />
    <skos:narrower rdf:resource="#22608" />
    <skos:narrower rdf:resource="#22737" />
    <skos:narrower rdf:resource="#89944" />
    … … …
```

Generated SKOS:
Usage of the Concept

```xml
<zbwext:indexedItem>
  <rdf:Description>
    <dcterms:requires>askosi</dcterms:requires>
    <dcterms:isPartOf>narrower</dcterms:isPartOf>
    <dcterms:extent>9</dcterms:extent>
  </rdf:Description>
</zbwext:indexedItem>
<zbwext:indexedItem>
  <rdf:Description rdf:about="http:/dspace/simple-search?query=keyword:sujet_21942">
    <dcterms:requires>WindMusic</dcterms:requires>
    <dcterms:isPartOf>keyword</dcterms:isPartOf>
    <dcterms:extent>156</dcterms:extent>
  </rdf:Description>
</zbwext:indexedItem>
<zbwext:indexedItem>
  <rdf:Description rdf:about="http:/dspace/simple-search?query=broadkeyword:sujet_21942">
    <dcterms:requires>WindMusic</dcterms:requires>
    <dcterms:isPartOf>broadkeyword</dcterms:isPartOf>
    <dcterms:extent>57997</dcterms:extent>
  </rdf:Description>
</zbwext:indexedItem>
<zbwext:indexedItem>
  <rdf:Description rdf:about="http:/dspace/simple-search?query=publishedby:sujet_21942">
    <dcterms:requires>WindMusic</dcterms:requires>
    <dcterms:isPartOf>publishedby</dcterms:isPartOf>
    <dcterms:extent>369</dcterms:extent>
  </rdf:Description>
</zbwext:indexedItem>
</skos:Concept>
```

Query for a SKOS Concept

```
<skos:Concept rdf:about="#21942">
    <skos:inScheme rdf:resource="http://www.windmusic.org/askosi/skos.jsp?about=sujet" />
    <skos:prefLabel xml:lang="en">Music</skos:prefLabel>
… … …
        <rdf:Description
            rdf:about="http:/dspace/simple-search?query=keyword:sujet_21942">
… … …
        <dcterms:extent>156</dcterms:extent>
```

ASKOSI can improve an existing application…

- 156 records for "MUSIC"
- 57786 when including Narrower Terms (all kinds of music)
- 369 Narrower Terms

ASKOSI integration in DSpace

- ☐ Open Source software developed by the MIT, Hewlett-Packard and Google to manage Digital Repositories (Open Archives)
- ☐ Used by more than thousand universities around the world. *In France: INIST*
- ☐ Based on SQL storage + Apache Lucene Search engine (SolR in development)
- ☐ Presentation in Götegorg (SW): http://gupea.ub.gu.se/dspace/handle/2077/21341
- ☐ Our presentation to FAO about how DSpace is used in a Poison Centre: http://www.aepic.it/conf/DSUG2007/viewabstractd263.html?id=197&cf=11

Benefits of this integration

- ☐ Multilingual data
- ☐ Exhaustivity: Searches retrieves specifics, synonyms, translations
- ☐ Precision: precise result for a given concept
- ☐ Displays with usage statistics (and search link) near each value
- ☐ Strongly validated updates; Data entry helped by Autocomplete
- ☐ Better metadata model and easier to maintain
- ☐ Example: http://www.windmusic.org

Current developments

- ☐ Meetings with Users to improve presentation and interactions;
- ☐ Concept display with different formats / extents
- ☐ Supplementary Global validations (e.g. siblings vs. specifics)
- ☐ Cascading Authorities: local, national, international…
- ☐ Queries History…
- ☐ Better Search Engine

Queries History

- ☐ Remembers the different searches done by users
- ☐ Users can browse, sort, examine and save results of searches
- ☐ Boolean combinations (AND, OR, NOT) of previous searches
- ☐ Sharing of searches between users
- ☐ Tagging individual records or a whole search result
- ☐ Translation of a result set into a target type

  (for instance, translation of a set of recordings into a set of musical scores, using relation « isFormatof »)

This within a supportive and adapted User Interface

Apache SolR 1.4

- ☐ Based on Apache Lucene
- ☐ May federate searches
  for multiple local applications
- ☐ N-gram or phonetic searches:
  tolerance to small errors in words
  typed.
- ☐ Facetted browsing: list of most
  frequent concepts within a search
 result and help drill down.

DSpace uses Lucene and not yet SolR…

Contributions welcome !

- ☐ Sharing of efforts for Specifications, Development and Implementation
- ☐ Sharing of current and future results
- ☐ Maintenance of a common and coherent set of tools for Thesauri maintenance and use.
- ☐ What is the most useful for your project NOW?

Symbols Standardisation

☐ Who could be interested?

☐ Your opinion about ConceptScheme symbol?

1. ⚭ countries ≥ 250⊙
2. ✖ countries ≥ 250⊙
3. ⚘ countries ≥ 250⊙
4. ∴ countries ≥ 250⊙
5. ∀ countries ≥ 250⊙

⊙ Europe ∈ ∴ countries, ∪ politicalOrganisations
⊡ eu
⊟ ec
⊞ eec
☞ en Europe
☞ es Europa
☞ en European Union
☞ en european
⇈ ⊙ Earth
⊂ ⊙ Continents
⊃ ⊙ Belgium
⊃ ⊙ England
⊃ ⊙ France
⊃ ⊙ Germany
...
⋈ ⊙ European Parliament
= ∴ Eurovoc ⊙ Europe
≈ ∴ Géologie ⊙ Europa
≳ ∴ Agrovoc ⊙ Benelux
≲ ∴ Géologie ⊙ Pangée
≍ ∴ Eurovoc ⊙ European Union
≋ Continent north of Africa and west of Asia
≙ Use this concept for the geographical region, not the political organization
≚ *An example of this concept use...*
≙ *(history of the evolution of the concept and of its terminology)*
≙ *(changes in the database about european continent)*
☛ *Internal notes to terminologists...*

Proposal available at:
http://www.destin.be/ASKOSI/Wiki.jsp?page=Icons%20for%20SKOS