# Integrating applications & projects
# =
# Dynamic & repeatable transformation of existing Thesauri and Authority lists into SKOS
# +
# Cross-tabulation of Concepts Linked Data

Presentation to the Linked Data Meeting

University College of London, September 14th 2010

by Christophe Dupriez, Destin SSEB, dupriez@destin.be
working for Belgium Poison Centre
rue Bruyn 1, B-1120 Brussels (Belgium)

# The main request from Users:

PARACETAMOL 893 203 2180 65

*Whenever a concept is mentioned,*

***concise** visual clues about:*

- *Where it comes from?* (e.g. : substances)
- *Where is it also mentioned?*
  (e.g. : in MDs Wiki+paper files)
- *For which role?*
  (e.g. This substance, is it a problem or a cure? )
- *About how many times is it mentioned?*
  (e.g. 893 203 2180 65 =893 bibliographic records, 203 products, 2180 calls, 65 medical reports)

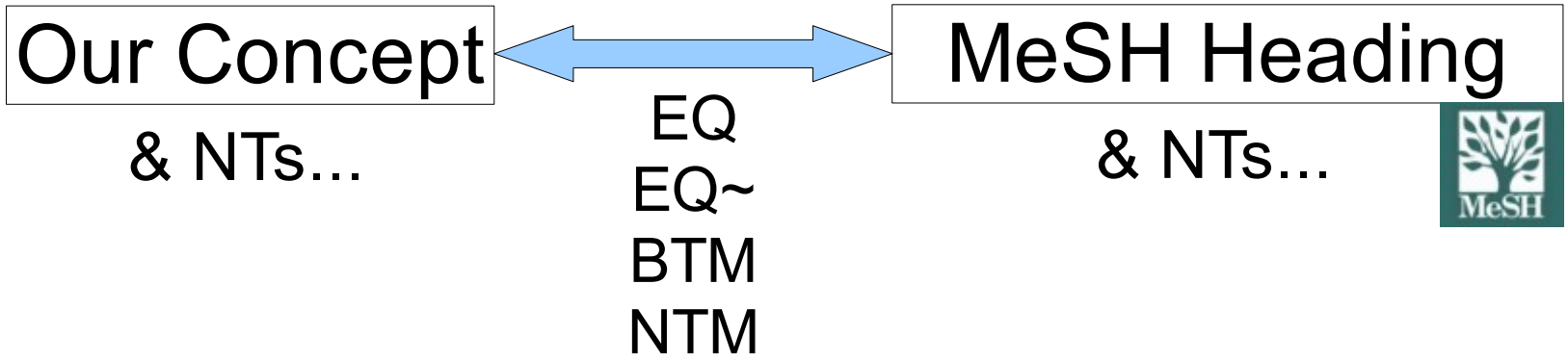
**+** *single click to access any of those when desired.*

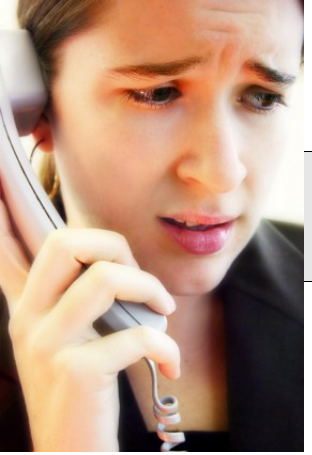
# Use Case : Current Awareness

From a list of subjects and
document types (e.g. reviews, case reports…) :

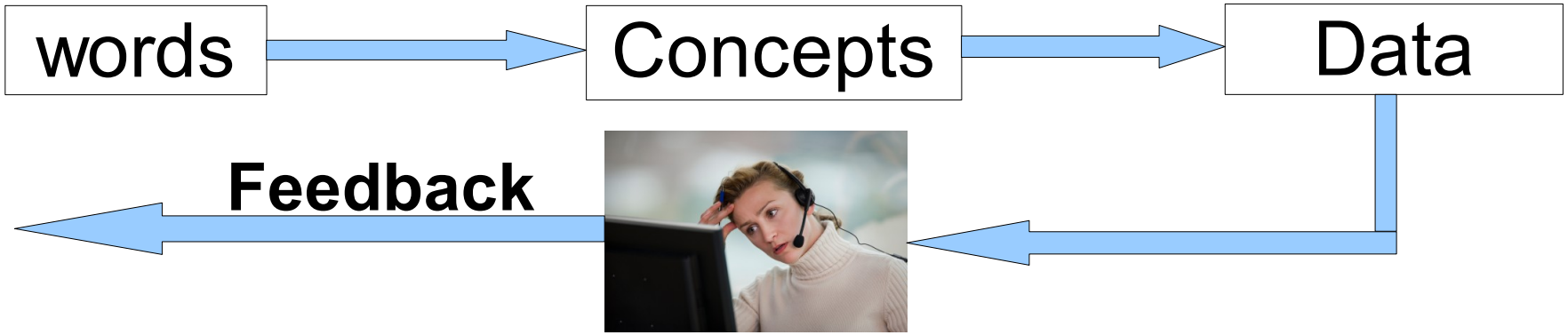
- **filter** remote sources (e.g. PubMed)
- help **index** new records with our vocabularies.

*We must **manage** equivalences*
*between our thesauri and remote ones (e.g. MeSH)*
*(browse, display, validate, update...)*





# Use Case: Emergencies




1) Powerful word search
2) Information to discriminate between Concepts
3) Identified Concepts = Clues to gather Linked Data
4) Managing sets of data for different Clues
5) Browsing Data to discriminate between Hypothesis

*Support MDs to build their recommendations;*
*Analysis of Events for Toxico-Vigilance*

# Benefits of integrating SKOS
## (terminology and concepts management)
## in all applications of users' workbench

- **Multilingual** data and user interface.
- **Exhaustivity**: Searches retrieves specifics, synonyms, translations and equivalent concepts in other “aligned” thesauri.
- **Precision**: precise result for a given concept
- Strongly **validated updates**; Data entry helped by **Auto-complete**
- **Better metadata model**, easier to maintain

# Benefits of integrating Concepts' Usages information in all applications of users' workbench

- PARACETAMOL 893 203 2180 65 : concept's references enriched with statistics and links to places where they are also used.
- Promotes direct linking from a concept to its usages within applications.
- Promotes homogenous display and functionalities to create, display, update, link, unlink concepts to applications elements.
- Usage statistics (and **search link**) near each mention of a concept (passage from one application to another)
- **Better metadata model**, easier to maintain

## 1. **BIBL** application: Articles about Human Toxicology

Internal Thesauri (Subject Vocabularies):

1) Substances
2) Living beings (plants, animals, mushrooms...)
3) Symptoms, Treatments
4) Places

External Thesauri and Vocabularies:

1) MeSH
2) NCBI Taxonomy, SP2000 Catalogue of Life
3) CAS/EINECS (REACH, ChemID+)

## 2. WIKI application ("**SAQ**")

Advices from MDs to others about how to manage situations linked to the different concepts of the internal thesauri.

## 3. **CASES** application

Data about calls received and cases reviewed.
Internal Thesauri already mentioned.

## 4. **PROD** application: Mixtures sold on BE market

Internal Thesaurus: Substances

External Thesauri and Vocabularies:

(development to be undertaken by a network of Poison Centres)

1) CAS/EINECS (REACH, ChemID+)
2) Product Usage Categories

## 5. **CONTACT** application

Topic specialists and Products' Manufacturers/Distributors

Internal Thesauri:

1) Subject thesauri already mentioned
2) Places

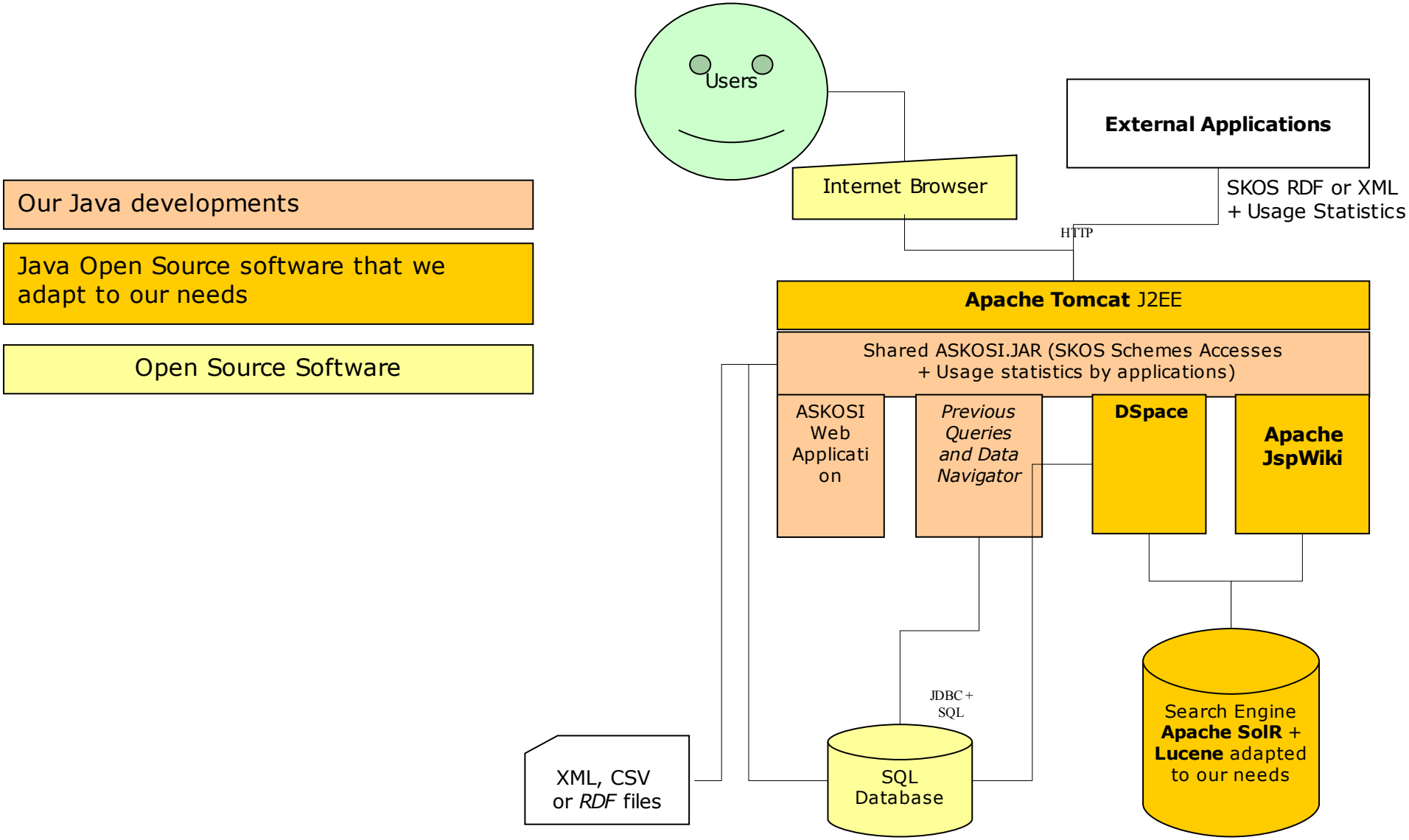
# 6. **ASKOSI**: Thesauri based Applications' Manager

*Integration under ASKOSI umbrella remains to be done for applications 3. 4. and 5. above.*

ASKOSI.org is an open project to create Java tools to integrate the benefits of terminology / concept usages management within applications. It is:

1. A Java Archive (JAR) providing an API aligned on SKOS conceptual organisation to access:
   1. Local or remote vocabularies / thesauri (being SKOS or not)
   2. Application data linked to Concepts;
2. A Web Application to browse gathered thesauri (*and to manage their interrelations*)

# Integrating ASKOSI with applications



# The ASKOSI JAR

- API aligned on W3C SKOS data structure
  = JavaBeans in-memory data structure
  = XML Structure http://www.askosi.org/ConceptScheme.xsd
  - ISO 25964 will be also considered.
- SQL, CSV, RDF and XML data sources
- Accesses can be Dynamic or Static (periodic reload) to import 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.

# Remote Sources ➔ ASKOSI

- **Big thesauri**: periodic editions of UMLS, Agrovoc, Catalogue OfLife (CoL), etc.:

1. Parameterize ASKOSI for a static SQL source
2. Load a local MySQL database with the new edition of UMLS/Agrovoc/CoL/...
3. Reload corresponding schemes

- SKOS/RDF/XML **Remote Web Services**:

1. Parameterize ASKOSI for load from a remote URL + XSLT transformation
2. (Auto)reload of corresponding schemes ("one concept at a time" must be developed)

# Internal Sources ➔ ASKOSI

- **Local Authority lists**:
  Parameterize ASKOSI for a dynamic SQL source: ASKOSI gets data up-to-date.
- **Legacy applications**:

1. Parameterize ASKOSI for XML file/URL load + XSLT transformation if necessary
2. Regularly generate the XML file with local usage data
3. (Auto)reload of corresponding schemes

- **Little lists** or small thesauri:

1. Parameterise ASKOSI for Excel CSV source.
2. (Auto)reload of corresponding schemes

# 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
…
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…
```

```
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;
```

# The ASKOSI Web Application

- Authority lists browsing:
  - Thesauri trees
  - Alphabetical lists
  - Decreasing 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*)
- <u>Links to applications using the SKOS concepts</u>

*In development:* search history manager, changes approval workflow, cross thesauri equivalence relations management.

| | | Substances | | | |
|---|---|---|---|---|---|
| 1 | | CHEMICALS AND DRUGS CATEGORIES | | | 77751 |
| 2 | | AMINO ACIDS, PEPTIDES, AND PROTEINS | | | 11614 |
| 3 | | AMINO ACIDS | | 26 P M | 2828 P M |
| 4 | | NEUTRAL AMINO ACIDS | | | 637 |
| 5 | | CYSTEINE | | 32 | 570 |
| 6 | | ACETYLCYSTEINE | | 469 | 525 |
| 7 | | THIAMPHENICOL GLYCINATE ACETYLCYSTEINATE | | | |
| 6 | | CARBOCISTEINE | | 9 | 10 |
| 6 | | CYSTEINYLDOPA | | | |
| 6 | | CYSTINE | | 1 P M | 3 P M |
| 6 | | DICHLORO(S-METHYLCYSTEINE)PLATINUM(II) | | | |
| 6 | | SELENOCYSTEINE | | | |
| 6 | | THIOCYSTEINE | | | |
| 5 | | METHIONINE | | 38 P M | 66 P M |
| 4 | | SULFUR AMINO ACIDS | | 5 P M | 952 P M |
| 5 | | CYSTEINE ↑ | | 32 | 570 |
| 5 | | D-PENICILLAMINE | | 268 P M | 303 |
| 5 | | METHIONINE ↑ | | 38 P M | 66 P M |
| 2 | | CHEMICAL ACTIONS AND USES | | | 66590 |
| 3 | | PHARMACOLOGIC ACTIONS | | | 55728 |
| 4 | | PHYSIOLOGICAL EFFECTS OF DRUGS | | | 32704 |
| 5 | | PROTECTIVE AGENTS | | 7 P M | 12382 |
| 6 | | ANTIDOTES AND SPECIFIC DRUGS | 3 | 259 | 11527 |
| 7 | | 2,3-DIMERCAPTOPROPANE SODIUM SULPHONATE | | 109 P M | 116 |
| 7 | | 4-DIMETHYLAMINOPHENOL | | 25 | 32 |
| 7 | | ACETYLCYSTEINE ↑ | | 469 | 525 |
| 7 | | ACETYLPENICILLAMINE | | 11 | 14 |
| 7 | | ACTIVATED CHARCOAL | 1 | 617 | 617 |

# Open Questions to the Community


- Users navigating the "Linked Data" Web need concise visual clues to decide what to do next, knowing what is behind each possible click:
  **How could we standardize a visual symbols system, the road signs of the Linked Data Web and its SKOS roundabouts?** *(proposals next page)*
- What is behind each concept? what is linked?
  **How could we standardize different harvesting mechanisms for Concepts Usage Data?**
  - "Push" vs "Pull" mode
  - Local vs Remote
  - Absolute vs Incremental
  - Results varying with user authorisations or preferences
- Linking with URIs allows user side or server side integration of applications.
  But between users and applications? Within an application?

# Standardising Symbols?

Your opinion about ConceptScheme symbols below?

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...*

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

# Call to Collaborations!

1. We want to integrate a “voting system” for reviewing SKOS / RDF statement contributions, including mappings between thesauri. Students welcome!
   Full proposed specs on: http://www.askosi.org/maintenance.pdf
2. Where could we discuss “Roadsigns for the Linked Data Web”?
3. Where could we discuss “Concepts Usage Data Harvesting”?
4. Where could we discuss “Concept References encoding (indexing chains) within Applications” ?

christophe.dupriez@destin.be
christophe.dupriez@poisoncentre.be