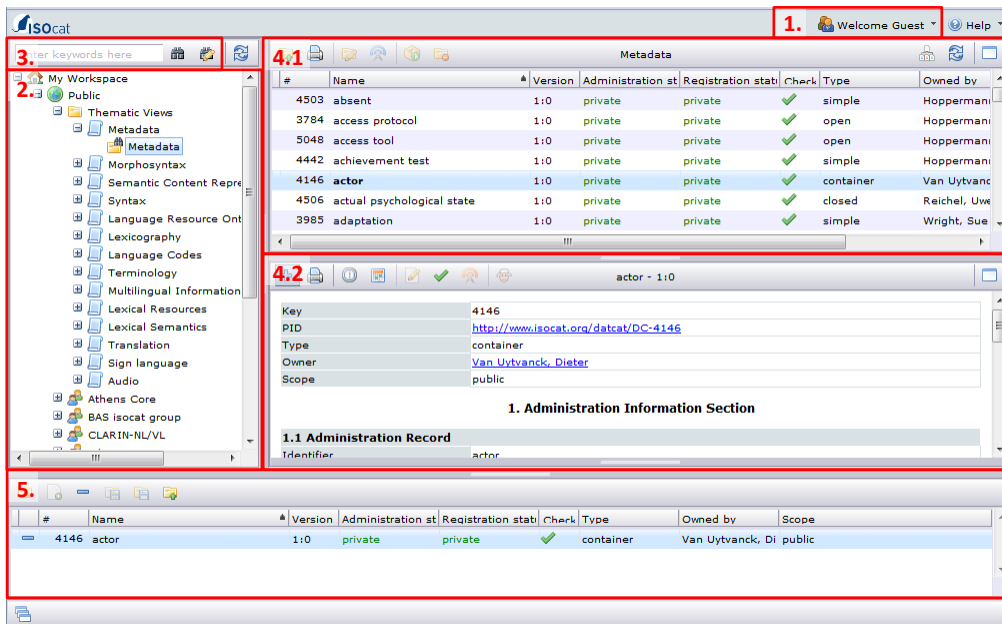


# LREC 2012 ISOcat tutorial – exercises: ‘answers’

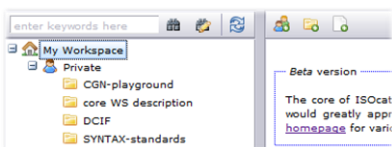


1. Welcome menu
2. Workspace explorer
3. Search box and options
4. Main area
  - 4.1. DCS viewer
  - 4.2. DC viewer
5. Basket area

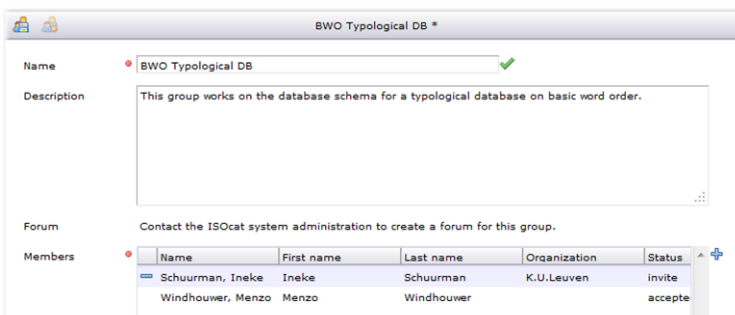


- If needed, register (👤) with the ISOcat (testbed).
- Log in (👤) to ISOcat (testbed) using your own credentials.
- Notice that after logging in you can change the automatic generated password into something easier to remember.
  - You find the Password (🔑) menu item in the Welcome menu (👤).

## Hands-on: ISOcat basics

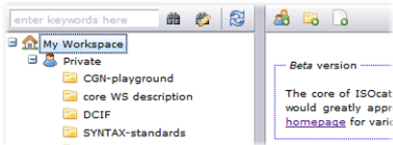


- Select ‘My Workspace’.
- In the right menu bar there are three buttons to create new groups (👤), new selections (📁) and new data categories (📄). Create a new group (👤).

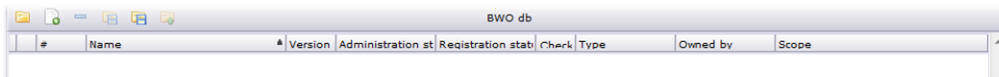


- Give your new group a unique name. This name has to be globally unique as when the group makes selections public it will appear in the workspace of every ISOcat user.
- Describe the purpose of the group.
- Add users to the group (+). These users will receive an invitation email and will only have access to the selections and categories shared with the group once they have accepted the invitation.
- Save the new or changed group (👤).

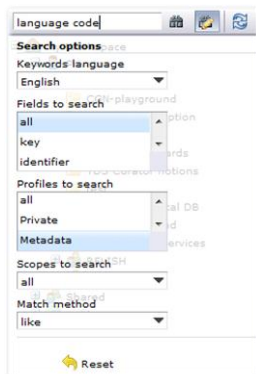
### Hands-on: creating Data Category selections



- Select 'My Workspace'.
- Create a new Data Category Selection (📁).
- Give the selection a name.



- The selection is directly loaded into the basket area.
- Use the search (🔍) and/or navigation facilities to find matching data categories for the database columns.
  - For the generic columns, e.g., language name and code, one can use the options (👤) to restrict the search to the metadata domain:



- You can also restrict the search to name and identifier to find data categories named similar to the database columns
- If you found a proper category it can be added (+) to the basket.
- The following is a possible selection of Data Categories:

#	Name	Version	Administration st	Registration st	Check	Type	Owned by	Scope
2482	language ID	1:0	private	private	✓	constrained	Athens Core	public
2484	language name	1:0	private	private	✓	open	Athens Core	public
2534	original source	1:0	private	private	✓	open	Athens Core	public
3759	source	1:0	private	private	✓	open	Nevskaya, Irina	public

- Why might you prefer these categories?

Column	Data Category	Discussion
language name	/language name/ ( <a href="#">DC-2484</a> )	The owner of this data category is Athens Core, which is actually a group of people working in the metadata thematic domain. This means the category is reviewed by a number of people and likely to be standardized somewhere in the future. However, it might still undergo changes.
language code	/language ID/ ( <a href="#">DC-2482</a> )	Same as for /language name/. Notice that /language ID/ doesn't list all valid ISO 639-3 codes (this conceptual domain is too big to be conveniently managed in ISOcat), but it does define a constraint to limit the value domain to 3 lower case letters.
Source	/source/ ( <a href="#">DC-3759</a> )	The MDF /source/ is targeted towards people as the source of information. We do have a separate column for informant. If this column is only used for bibliographical references the /original source/ owned by Athens Core.
	/original source/ ( <a href="#">DC-2534</a> )	
informant	/source/ ( <a href="#">DC-3759</a> )	See the source row above.

- Save (💾) the changes to your Data Category Selection.

#	Name	Version	Administration st	Registrati
2482	language ID	1:0	private	private
2484	language name	1:0	private	private
2534	original source	1:0	private	private
3759	source	1:0	private	private

- Load the DCS in the DCS viewer.
- Change the scope (👤) of the DCS.

**Change the scope of this Data Category Selection**

Groups with read and write access to your/their Data Categories in this Data Category Selection

- BWO Typological DB
- CGN-playground
- CLARIN Web Services
- CLARIN-NL/VL
- GilAndSusan
- RELISH
- STTS
- TDS Curator
- VU-DNC

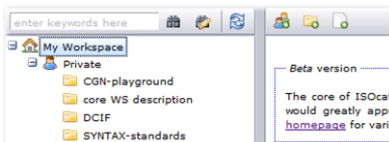
Set scope to public (everyone has read access)

Make all your DCs in this DCS also public

Cancel Change

- Mark the checkbox for the group you've created before.
- Save the changes (✔). Now all group members, who have accepted the invitation, will have access to this DCS.

## Hands-on: creating Data Category specifications



- Select 'My Workspace'.
- Create a new Data Category (📄).

**subjectObjectVerb - 1:0 \***

Administrative Information Section

**Identification**

Identifier: subjectObjectVerb ✓

Version: 1:0

Type: simple

Justification: Needed for a typological database on basic word order.

Origin:

**Status**

Administration status: private

Registration status: private

**Dates**

Effective date: yyyy-MM-dd

Until date: yyyy-MM-dd

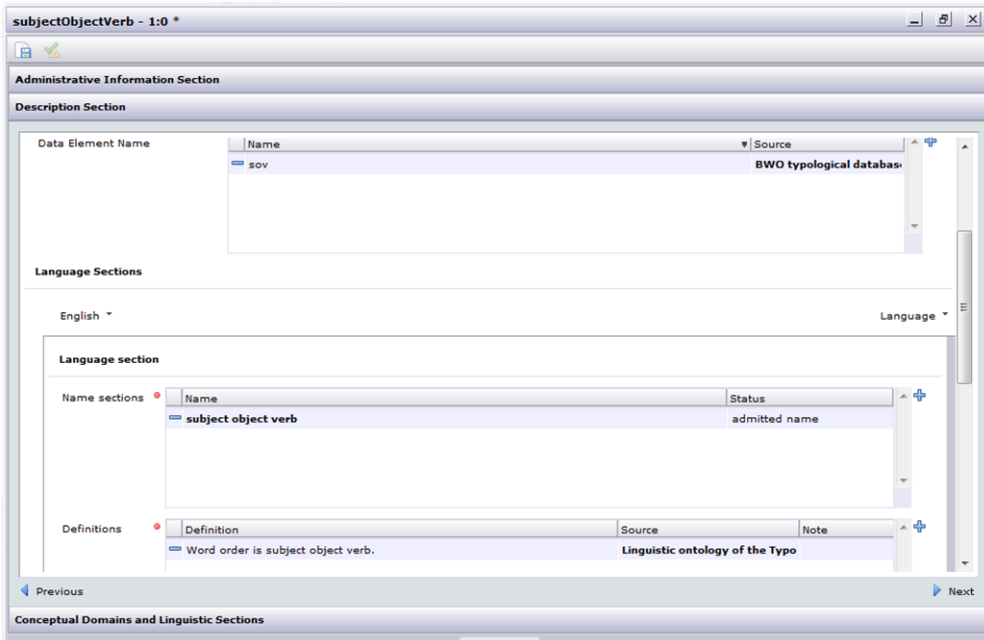
**Additional information**

Cancel Next

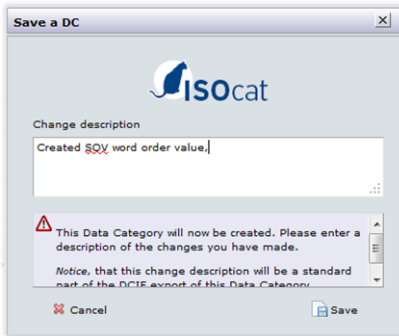
**Description Section**

**Conceptual Domains and Linguistic Sections**

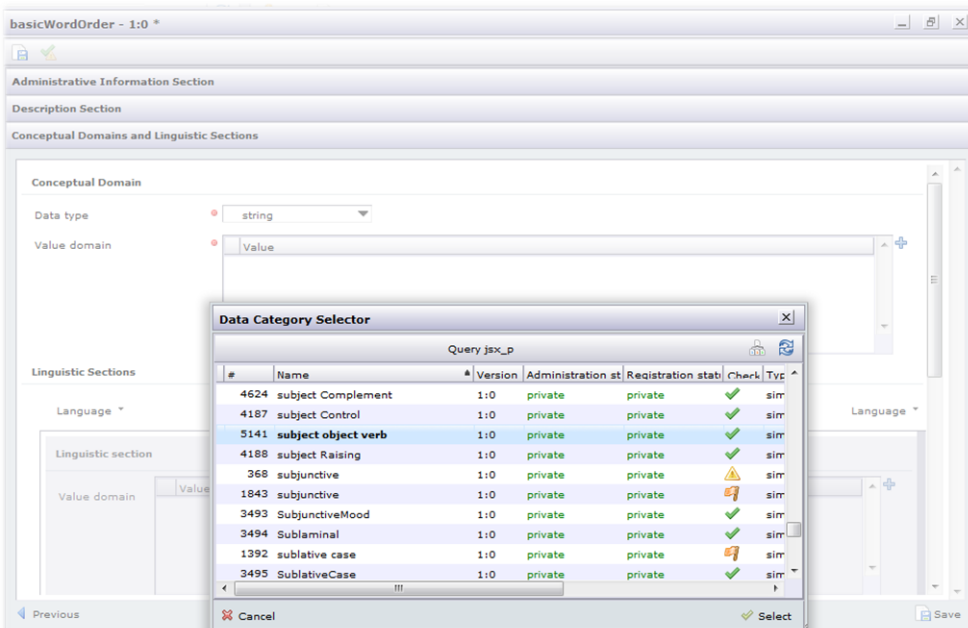
- Enter a camel case identifier.
- Choose the proper type.
- Enter a justification, just in case you want to standardize the Data Category later.



- The abbreviation we use in our database we can enter as a Data Element Name.
- The English language section is mandatory and should contain at least one name and exactly one definition.
- Save (F4) the new Data Category.



- Enter a meaningful change description and save (F4).
- Agree to add the new Data Category to the basket which contains the DCS we're filling.
  - Don't forget to save (F4) the changes.
- Same can be done for the other word order values and for the basic word order closed data category.



- Add (✚ and ✓) the simple Data Categories, like */subject object verb/*, to the value domain of the closed complex Data Category */basic word order/*.

## Take home: beyond ISOcat – data category references

A database schema is defined using the Data Definition Language (DDL) part of SQL. Using @dcr:datcat annotations inside comments the DDL statements can be annotated:

```
-- @dcr:datcat 'sov' http://www.isocat.org/datcat/DC-...
-- @dcr:datcat 'svo' http://www.isocat.org/datcat/DC-...
-- @dcr:datcat 'osv' http://www.isocat.org/datcat/DC-...
-- @dcr:datcat 'ovs' http://www.isocat.org/datcat/DC-...
-- @dcr:datcat 'vos' http://www.isocat.org/datcat/DC-...
-- @dcr:datcat 'vso' http://www.isocat.org/datcat/DC-...

CREATE TYPE bwo AS ENUM (
    'sov',
    'svo',
    'osv',
    'ovs',
    'vos',
    'vso'
);

-- @dcr:datcat code http://www.isocat.org/datcat/DC-2482
-- @dcr:datcat name http://www.isocat.org/datcat/DC-2484
-- @dcr:datcat bwo http://www.isocat.org/datcat/DC-...
-- @dcr:datcat source http://www.isocat.org/datcat/DC-2534
-- @dcr:datcat informant http://www.isocat.org/datcat/DC-3759

CREATE TABLE "typ-word-order" (
    code character varying(3) NOT NULL,
    name character varying NOT NULL,
    bwo bwo,
    source character varying,
    informant character varying
);
```

## Take home: beyond ISOcat – ontological relationships

verb-object-subject (vos) and verb-subject-object (vso) can be seen as subclasses of predicate-initial:

```
/verb-object-subject/ rel:subClassOf /predicate-initial/
/verb-subject-object/ rel:subClassOf /predicate-initial/
```

In the ontology of the Typological Database System (TDS) project these kind of theoretical perspectives were both included but kept separate in their own parts of the subsumption hierarchy. In the TDS different theories describing the same linguistic phenomena are linked using the special tds:equatesWith predicate<sup>i</sup>. Depending on your preference the same can be stored in the Relation Registry RELcat:

```
/verb-object-subject/ tds:equatesWith /predicate-initial/
/verb-subject-object/ tds:equatesWith /predicate-initial/
```

However to make it possible for generic semantic search algorithms to make use of this relationship tds:equatesWith has to be embedded in the taxonomy of relation types:

1. rel:related
  - 1.1. rel:almostSameAs
    - 1.1.1. tds:equatesWith

<sup>i</sup> A. Dimitriadis, A. Saulwick, M. Windhouwer. Semantic relations in ontology mediated linguistic data integration. In Proceedings of the E-MELD Workshop on Morphosyntactic Annotation and Terminology: Linguistic Ontologies and Data Categories for Linguistic Resources (E-MELD 2005), Cambridge, Massachusetts, July 1-3, 2005. <http://www.emeld.org/workshop/2005/papers/saulwick-paper.html>