



# **ISOcat content management**

Ineke Schuurman  
ISOcat content co-ordinator CLARIN-NL

Sofia 2012-10-27

# Overview



- ISOcat in general
- Usefulness for ISOcat
- Some issues in that respect
- Our approach:
  - strict rules,
  - content management
- Recent adaptations ISOcat
- Future
  - (JOIN)

# ISOcat



## ISOcat:

**Data Category Registry** defining widely accepted data categories (DCs)

<http://www.isocat.org>

All Data Categories (DCs) come with Persistent Identifiers: <http://www.isocat.org/datcat/DC-1297>

Once a data category has been created it can never be deleted; it can only be deprecated or superseded

# Panes



The screenshot shows the ISOcat web interface with several panes highlighted by red boxes and labels:

- Browser pane:** Located on the left side, showing a file tree structure under "My Workspace".
- Thematic view:** A sub-section within the browser pane, showing a list of thematic views like "Metadata", "Morphosyntax", etc.
- Folder:** A label pointing to the "Terminology" folder in the browser pane.
- Active DCS pane:** A table displaying a list of terminology entries with columns for ID, Name, Version, Administration status, Registration status, Check, Type, Owned by, and Scope.
- Add:** A button labeled "Add" located below the Active DCS pane.
- DC specification pane:** A pane showing details for a specific entry (ID 149), including a URL and administrative information.
- Detach:** A button labeled "Detach" located to the right of the DC specification pane.
- Basket pane:** A table at the bottom of the interface showing a list of items in a basket, including "abbreviated-form", "admittedTerm", and "context".

#	Name	Version	Administration st	Registration statu	Check	Type	Owned by	Scope
373	normalizedTerm	1.0	accepted	standard		open	Terminology	public
374	normativeAuthorizati	1.0	accepted	standard		closed	Terminology	public
112	normativeDocument	0.0.0	private	candidate		simple	Sue Ellen Wright	public
1487	nounAdjective	0.0.0	private	candidate		open	Isabelle KRAMER	public
383	nounClass	1.0	accepted	standard		closed	Terminology	public
1490	nounNoun	0.0.0	private	candidate		open	Isabelle KRAMER	public
1484	nounPrepositionNour	0.0.0	private	candidate		closed	Isabelle KRAMER	public

#	Name	Version	Administration st	Registration statu	Check	Type	Owned by	Scope
347	abbreviated-form	0.0.0	private	candidate		simple	Sue Ellen Wright	public
377	admittedTerm	1.0	accepted	standard		simple	Terminology	public
149	context	1.0	accepted	standard		open	Terminology	public

# DC Specification Pane



conversation - 1:0

Key	2661
PID	<a href="http://www.isocat.org/datcat/DC-2661">http://www.isocat.org/datcat/DC-2661</a>
Type	simple
Owner	<a href="#">Athens Core</a>
Scope	public

## 1. Administration Information Section

### 1.1 Administration Record

Identifier	Conversation
Version	1:0
Registration Status	private
Administration Status	private
Justification	common metadata data category
Origin	IMDI: <a href="http://www.mpi.nl/IMDI/Schema/Content-EventStructure.xml">http://www.mpi.nl/IMDI/Schema/Content-EventStructure.xml</a>
Effective Date	2009-09-11
<i>1.1.1 Creation</i>	
Creation Date	2009-09-11
Change Description	Communication event with more than two participants.
<i>1.1.2 Last Change</i>	
Last Change Date	2010-12-16
Change Description	MT

## 2. Description Section

Profile	Metadata
<b>[-] 2.1 English Language Section</b>	
Language	English (en)
<i>2.1.1 Name Section</i>	
Name	conversation
Name Status	admitted name
<i>2.1.2 Definition Section</i>	
Definition	Communication event with more than two participants.
Source	CLARIN
<b>[+] 2.2 Bulgarian Language Section</b>	
<b>[+] 2.3 Catalan Language Section</b>	

# Relevant wrt ISOcat



## Open registry:

- Everybody can register and create new entries
  - Good ones
  - Not that good ones (as such, or for CLARIN purposes)
- Everybody can make his/her DC traceable for ‘the whole world’ (making them ‘public’)
- Everybody can change his/her DCs every moment
  - even in a semantically meaningful way

# Affecting usefulness



- Quality of DCs
  - Number of DCs
  - Semantic consistency of DCs
- ▶ CLARIN-NL therefore opted for ISOcat content management

# Because ...



When nothing is done, ISOcat will not be very useful for our (CLARIN-related) purposes

In theory, it's easy:

All we need are good DCs !!

In practice:

not THAT easy ...



# So...



- Practice what is preached (i.e. follow general ISOcat guidelines),
  - **plus some more rules we came up with (do's and don'ts)**

Note, however, that we can not make others obey rules (neither general ones nor ours), only our own people

We are to accept that not all definitions in DCs will be generally accepted, as we are dealing with existing resources (legacy data)

# Expectations ISOcat



## DCs

- Unambiguous
- Clear names
- Non-tautological
- Short, neutral definitions

## ► Reusable

do NOT underestimate the ISOcat task!

- Neither for user nor for co-ordinator

Needed: tutorials, manuals, but also

## ► SUPERVISION, support, training i.e. “content management”

# General actions



- **One (1) definition per DC (otherwise: warning, ‘marker’)**

## Soon:

- **Active selection of ‘profile’ enforced**
  - Current default is ‘private’: often incorrect
- **Clickable ‘links’ for concepts used in definition**
- **Warning system that DC-*nnn* was adapted since last visit**
- **Recommendation system (level of ‘larger group’)**
  - **Available since Wednesday!!!**
- **Larger roles RELcat and SCHEMAcat**

# Aids: Group



## Group CLARIN-NL/VL

- All members can see what is (being) done, i.e. also ‘private’ DCs of fellow members
  - No ‘waste of time’ / duplication of efforts
  - members can ‘correct’ each other at an early stage

DCs remain ‘private’ as long as possible!  
(but visible for group)

# Group CLARIN-NL/VL



- CLARIN-NL/VL
  - Adelheid attributes
  - Adelheid elements
  - Adelheid tags
  - admin-accepted
  - admin-changed
  - admin-proposed
  - audio
  - CGN
  - CKCC-DANS
  - COAVA
  - coreference
  - DatacurationService
  - DUELME
  - Edisyn
  - KB\_metadata
  - MD-proposal
  - metadata
  - MetadataProject
  - MIGMAP
  - named entities
  - NEHOL
  - NLVL-recommended
  - part of speech
  - phonemes
  - pilot-submission
  - semantic roles

## CLARIN-NL/VL

Name	CLARIN-NL/VL
Description	A group for the participants in the CLARIN-NL en VL projects
Forum	<a href="#">private forum</a>

### Members

member	<a href="#">Aalstein, Anna</a> (INL)
member	<a href="#">Athens Core</a>
member	<a href="#">Audring, Jenny</a> (CLARIN)
member	<a href="#">Augustinus, Liesbeth</a> (CCL - KU Leuven)
member	<a href="#">Boelhouwer, Bob</a> (INL)
member	<a href="#">Bountouridis, Dimitrios</a> (UU)
member	<a href="#">Broeder, Daan</a> (MPI for Psycholinguistics)
member	<a href="#">Brouwer, Matthijs</a> (Meertens Institute)
member	<a href="#">Buitendijk, Bram</a> (Huygens ING)

# Aids: Group



## Scope 'private':

- Not bothering others (like CLARIN groups elsewhere) with DCs still bound to change significantly
- 'Permission' to make scope of DCs **public** *in se* to be given by content co-ordinator
  - (but CMDI, wrong button, older DCs, ...)

# Wrt recommendation



- **Recommendation** by content co-ordinator !
  - Of course: public scope
- Indicates that this DC is semantically stable, should be reused whenever possible
- ▶ When creating (or adopting) an alternative DC this should be motivated explicitly (Justification section) !!!

# Group recommendation



ISOcat interface showing a group recommendation for a DCIF resource. The interface includes a sidebar with details for the resource 'aaa - 1:0' and a main table of DCIF resources.

**Resource Details (aaa - 1:0):**

- Recommended by: TDS Curator (highlighted with a red box and '2.')
- Key: 4012
- PID: <http://www.isocat.org/datcat/DC-4012>
- Type: complex/open
- Owner: [Windhouwer, Menzo](#)
- Scope: private

**1. Administration Information Section**

**1.1 Administration Record**

Identifier	aaa
Version	1:0
Registration Status	private

**DCIF Table:**

Version	Administration status	Registration status	Check	Type	Owned by
1:0	private	private	Warning	open	Windhouwer
1:0	private	private	Flag	open	Windhouwer
1:0	private	private	Flag	open	Windhouwer
1:0	private	private	Flag	open	Windhouwer
1:0	private	private	Flag	simple	Windhouwer
1:0	private	private	Warning	open	Windhouwer
1:0	private	private	Flag	simple	Windhouwer
1:0	private	private	Flag	open	Windhouwer

**Recommendation List:**

- CLARIN Web Services
- RELISH
- TDS Curator (highlighted with a red box and '1.')

**Resource Details (aaa - 1:0):**

- PID: <http://www.isocat.org/datcat/DC-4012>
- Type: complex/open
- Owner: [Windhouwer, Menzo](#)
- Scope: private

**1. Administration Information Section**

**1.1 Administration Record**

#	Name	Version	Administration status	Registration status	Check	Type	Owned by	Scope
---	------	---------	-----------------------	---------------------	-------	------	----------	-------



# ... by CLARIN-NL/VL



search result for 'age'

#	Name	Version	Administration status	Registration status	Check	Type	Owned by	Scope
3660	age	1:0	private	private	✓	simple	de Vriend, Folkert	public
4713	education age	1:0	private	private	✓	open	Crasborn, Onno	public
2550	participant age	1:0	private	private	✓	open	Athens Core	public
3884	participant mean age	1:0	private	private	✓	open	Hoppermann, Christ	public
3653	Participant age range	1:0	private	private	⚠	closed	de Vriend, Folkert	public

age - 1:0

Recommended by	<a href="#">CLARIN-NL/VL</a>
Key	3660
PID	<a href="http://www.isocat.org/datcat/DC-3660">http://www.isocat.org/datcat/DC-3660</a>
Type	simple
Owner	<a href="#">de Vriend, Folkert</a>
Scope	public

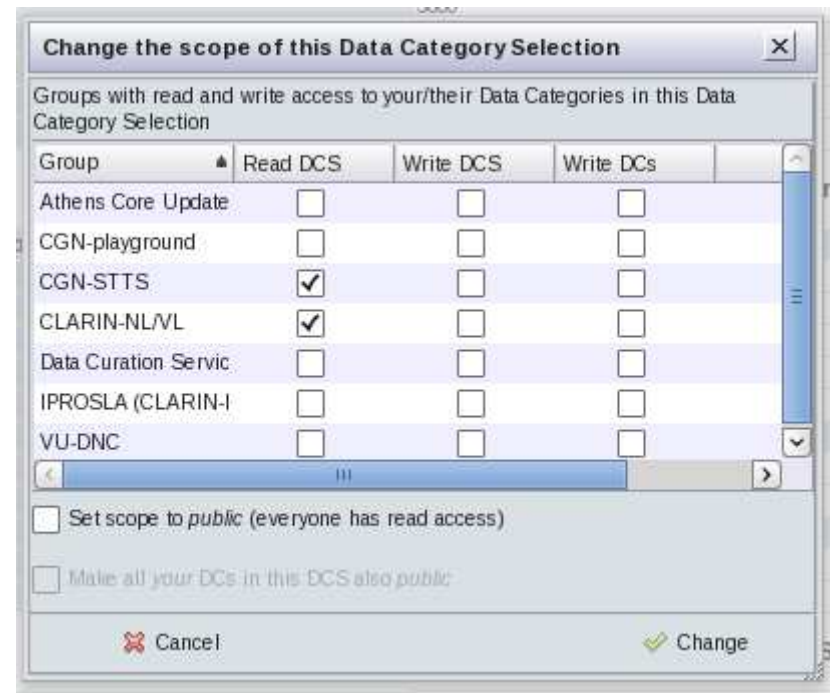
## 1. Administration Information Section

1.1 Administration Record	
Identifier	age
Version	1:0
Registration Status	private
Administration Status	private
Justification	common metadata data category
1.1.1 Creation	
Creation Date	2010-07-22
Change Description	Data category created
1.1.2 Last Change	
Last Change Date	2012-10-23
Change Description	removed capital (Name section)

# Group




- +/- 70 members
  - Some 30 active
- +/- 40 DCSs
  - Some 20 active



# Aids: Forum



 **ISOcat**  
discussion forum

Search... Search  
Advanced search

Board index · Groups · CLARIN-NL/VL

User Control Panel · View your posts

Smartfeed FAQ Members Logout [ineke]

## CLARIN-NL/VL

**NEWTOPIC\*** Search this forum... Search

Mark topics read • 6 topics • Page 1 of

ANNOUNCEMENTS	REPLIES	VIEWS	LAST POST
 <b>ISOcat CLARIN-NL/VL view</b> by mwindhouwer » Thu Oct 11, 2012 9:06 am	0	9	by mwindhouwer  Thu Oct 11, 2012 9:06 am

TOPICS	REPLIES	VIEWS	LAST POST
 <b>How to make a DCS visible for the group CLARIN NL/VL</b> by ineke » Mon Oct 15, 2012 4:09 pm	0	6	by ineke  Mon Oct 15, 2012 4:09 pm
 <b>How to put DCs in DCS 'admin-proposed' (CLARIN NL/VL)</b> by ineke » Mon Oct 15, 2012 3:55 pm	0	4	by ineke  Mon Oct 15, 2012 3:55 pm
 <b>How to decide whether an existing DC can be adopted (flowcha</b> (b) by ineke » Thu Oct 11, 2012 3:40 pm	0	7	by ineke  Thu Oct 11, 2012 3:40 pm
 <b>"do's and don'ts" (version 0.3, june 2012)</b> (b) by ineke » Thu Oct 11, 2012 2:26 pm	0	9	by ineke  Thu Oct 11, 2012 2:26 pm
 <b>Spreadsheet to collect candidate relationships for</b>	0	10	by mwindhouwer 

# Aids: View



View: access to a limited part of ISOcat, i.e. only DC(S)s from a specific (thematic domain) group (e.g., CLARIN-NL/VL)

- ▶ members of group can focus on work done within their own group

Future: switches between full and local view will be enabled

https://catalog.clarin.eu/isocat/interface/index.html?view=CLARIN-NL/VL

ISOcat

enter keywords here

- My Workspace
- Private
  - 2012022715042871 DCIF import
  - AC-update
  - admin-accepted
  - admin-changed
  - admin-proposed
  - audio
  - CGN
  - coreference
  - MD-proposal
  - metadata
  - MetadataProject
  - named entities
  - NLVL-recommended
  - noun-cgn
  - part of speech
  - phonemes
  - pilot-submission
  - semantic roles
  - spatiotemporal analysis
  - Spelling variation
  - syntax
- Athens Core Update
- CGN-STTS
- CLARIN-NL/VL
- Shared
- Public
  - CLARIN-NL/VL

Shibboleth account users

The ISOcat web user inter  
[Shibboleth forum topic](#) to le

The persistent URI for ISC  
change from time to time fo

Beta version

The core of ISOcat is stable  
that process. See [the ISOcat](#)

The ISOcat web interface allow  
registered user you can store



# view CLARIN-NL/VL

note that the 'Public'  
part is restricted to  
CLARIN NL/VL

# Content management



- Most of what is described before can not be realized without content co-ordination/management
- Introducing such a system really pays off!
- ▶ We should even think about collaboration at a higher level, cf the **JOIN**-project (cf short presentation on Friday by Menzo and Ineke)

# JOIN



- Work on ISOcat rather time-consuming for all of us
  - Collaborate at an early stage
- For example, we should avoid including certain standards (say Wordnet (the original)) x times, not knowing that X is also working on it. Or doing it half/half, but in different styles...

JOIN!

# JOIN



- Best results for CLARIN when adopting the same strategies
  - ▶ content co-ordinators/managers are essential
    - ▶ Preferably at the national level

▶ Join JOIN ◀



# JOIN



- Collaboration between national teams, via these national contact co-ordinators/managers

We are prepared to organize

- JOINt workshop
- JOINt tutorial

# Together



**... we can achieve more !!!**

Thank you for your attention.