

CLARIN

Common Language Resources and Technology Infrastructure



CMDI Component Registry and Editor

Twan Goosen

CLARIN ERIC

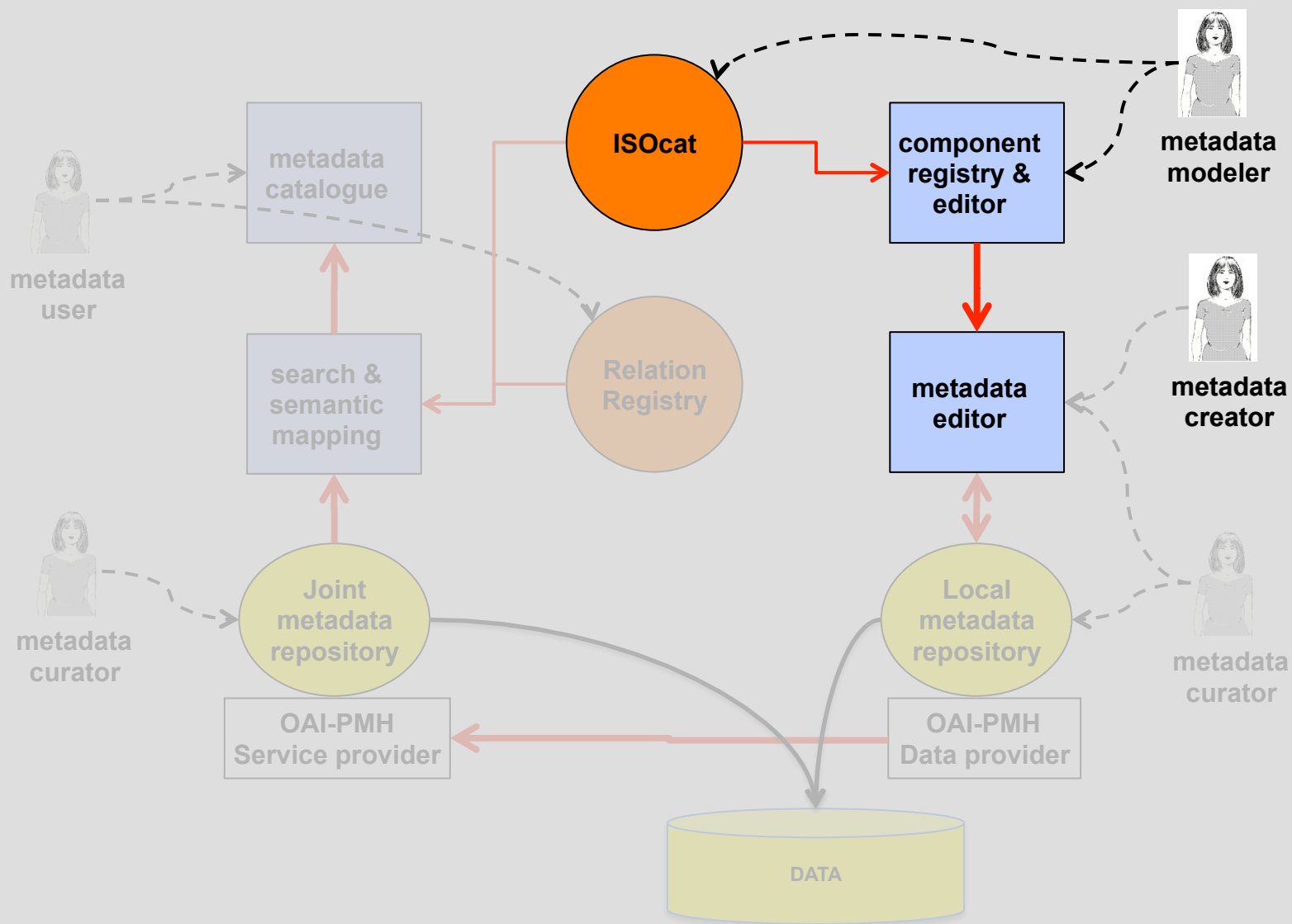
twan@clarin.eu

CLARIN-NL CMDI Tutorial

Utrecht, 4 March 2014



CMDI Metadata life-cycle



Purpose



Registration, Browsing and Editing

Registration: provide long term storage of Component/Profiles

Browsing: provide easy access to Components/Profiles

Editing: create new and edit existing Components/Profiles

Recap: Component/Profiles



CLARIN Metadata **Component**: An aggregation of metadata elements and other components aimed at describing a specific aspect of a resource.

Reusable building block



CLARIN Metadata **Profile**: An aggregation of metadata components and elements that can be used to create metadata descriptions. The profile is used to describe all relevant aspects of a resource or collection.

**Blueprint for
metadata description of a resource**



Example



Component named “Contact”. Has metadata elements to describe a persons contact information: Name, Address, Email and Organisation.

```
<CMD_ComponentSpec isProfile="false">
  <Header>
    <ID>clarin.eu:cr1:c_1264757339280</ID>
    <Name>component-contact</Name>
    <Description>IMDI components</Description>
  </Header>
  <CMD_Component CardinalityMax="1" CardinalityMin="1" name="Contact">
    <CMD_Element CardinalityMax="1" CardinalityMin="0" ValueScheme="string"
ConceptLink="http://www.isocat.org/datcat/DC-2454" name="Name"/>
    <CMD_Element CardinalityMax="1" CardinalityMin="0" ValueScheme="string" name="Address"/>
    <CMD_Element CardinalityMax="1" CardinalityMin="0" ValueScheme="string"
ConceptLink="http://www.isocat.org/datcat/DC-2521" name="Email"/>
    <CMD_Element CardinalityMax="1" CardinalityMin="0" ValueScheme="string"
ConceptLink="http://www.isocat.org/datcat/DC-2459" name="Organisation"/>
  </CMD_Component>
</CMD_ComponentSpec>
```

Example Profile



- Goal: Create metadata for a comic book resource



Source: Archive.org

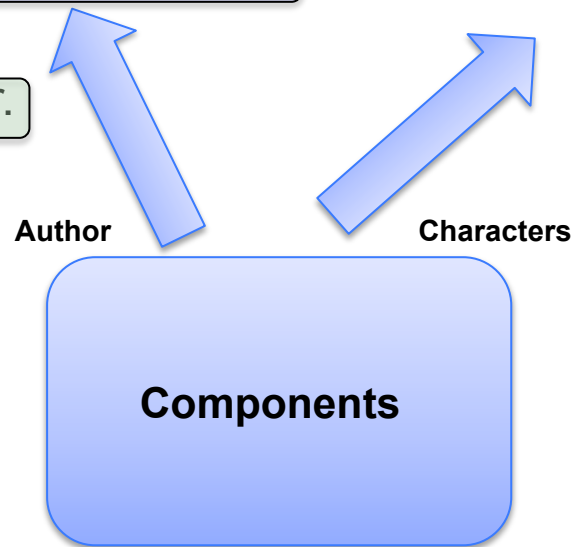
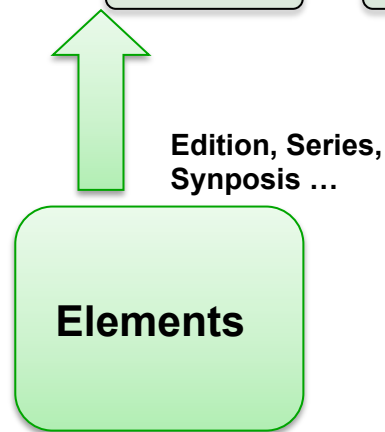
Comic book: Resource



First edition of "The Green Hornet" in which The Green Hornet and Kato combat a ruthless gang whose victims are forced to obey criminal commands!

Written by Fran Striker (U.S., 1903), published by Gold Key in 1966.

Has 36 pages in color.



Comic book: Profile



ComicBook

Series

Title

Synopsis

Author(s)

Name

Year of birth

Country

Character(s)

Name

Role



Using the Component Registry



- Browse to <http://www.clarin.eu/cmdr>
- Click the link to the Component Registry

Clarín Component Registry

Component Browser

User: anonymous [login](#)

Profiles Components **Public space**

Create new Edit as new Import filter... Showing 63 of 63

Name	Group Name	Domain Name	Creator	Description	Registration Date	Comments
AnnotationTool			Eric Sanders	Description of a tooladapt...	09 February 2011 14:13:59	0
ArthurianFiction		Other	Rik Hoekstra	Profile for Arthurian Fictio...	04 September 2012 14:50:05	0
BamdesLexicalResource		computational_linguistics	Dieter Van Uytvanck	Lexical Resource as used b...	27 October 2010 15:47:42	0
BamdesMultimodalCorpus		computational_linguistics	Dieter Van Uytvanck	Oral Corpus as used by BA...	27 October 2010 16:00:41	0
BamdesOralCorpus		computational_linguistics	Dieter Van Uytvanck	Oral Corpus as used by BA...	27 October 2010 16:00:05	0
BamdesTool		computational_linguistics	Dieter Van Uytvanck	Tool as used by BAMDES (f...	27 October 2010 15:48:54	0
BamdesWrittenCorpus		computational_linguistics	Dieter Van Uytvanck	Written Corpus as used by...	27 October 2010 15:59:28	0
Bedevaartbank			Folkert de Vriend	Profile for Bedevaartbank	30 July 2010 9:33:53	0
Boedelbank			Folkert de Vriend	Profile for Boedelbank	30 July 2010 9:36:25	0
cmdr-virtual-collection			CLARIN (cmdr@clarin.eu)	A component for describin...	21 April 2010 16:18:30	0
collection			Matel Durco	minimal collection profile f...	15 October 2010 20:58:37	0

ArthurianFiction view xml Comments (0)

Name: **ArthurianFiction**
Description: Profile for Arthurian Fiction database
ConceptLink: <http://www.isocat.org/rest/dc/4347>

Component: manuscript
ConceptLink: <http://www.isocat.org/datcat/DC-4371>
Number of occurrences: 0 - 1

Element: id string
ConceptLink: <http://www.isocat.org/datcat/DC-3894>

Component Browser



Clarin Component Registry

Component Browser

Profiles Components **Public space**

Create new Edit as new Import

Name	Group Name	Domain Name	Creator
AnnotationTool			Eric Sanders
ArthurianFiction		Other	Rik Hoekstra
BamdesLexicalResource		computational_linguistics	Dieter Van Uytendaele
BamdesMultimodalCorpus		computational_linguistics	Dieter Van Uytendaele
BamdesOralCorpus			
BamdesTool		computational_linguistics	Dieter Van Uytendaele
BamdesWrittenCorpus		computational_linguistics	Dieter Van Uytendaele
Bedevaartbank			Folkert de Vries
Boedelbank			Folkert de Vries
cmdi-virtual-collection			CLARIN (cmdi)
collection			Matej Duro

ArthurianFiction view

Name: **ArthurianFiction**
Description: Profile for Arthurian Fiction database
ConceptLink: <http://www.isocat.org/rest/dc/4347>

Component: manuscript
ConceptLink: <http://www.isocat.org/datcat/DC-4371>
Number of occurrences: 0 - 1

Element: id string
ConceptLink: <http://www.isocat.org/datcat/DC-3894>

Available Templates & Profiles

IMDI Templates

- Catalogue (internal)
- Corpus (internal)
- Session (internal)
- Sign-Language Session (internal)

Clarin Profiles

Add URL Add File

- AnnotationTool
- ArthurianFiction
- BamdesWrittenCorpus
- BASWebService
- Bedevaartbank
- Boedelbank
- CenterProfile
- Chromosome_Example
- cmdi-virtual-collection
- collection
- collection
- ...icBook
- ...component-dc-terms
- Component-dc-terms-modular
- CorpusProfile
- DBNL
- DcmiTerms
- DIDDD
- DIDDD_sub_location
- DSOL
- DynaSAND
- DynaSAND_sub_location
- EastRepublican
- Example_Profile_Instance
- Familienamenbank
- Gender_Example
- GenericTextStructure

Reload Clarin Profiles

Only load profiles selected for manual editing

New Template Close

Logging in



Clarin Component Registry

Component Browser

User: anonymous [login](#)

Profiles Components Public space

Create new Edit as new Import

Sign in to Clarin EU Service Provider

Select your Identity Provider

If you cannot find your institution in the list above please select the "Clarin.eu website account" and use your credentials of the CLARIN website. For questions please contact webmaster@clarin.eu.

- Clarin.eu website account
European Union
- Max Planck Institute for Psycholinguistics
Netherlands
- Academisch Ziekenhuis Maastricht
Netherlands
- AMOLF
Netherlands
- Amsterdamse Hogeschool voor de Kunsten
Netherlands
- ArtEZ Hogeschool voor de Kunsten
Netherlands
- ASTRON
Netherlands
- Avans Hogeschool
Netherlands

		Comments
Profile for Arnhemman Pict...	07 September 2012 14:30:03	0
Lexical Resource as used b...	27 October 2010 15:47:42	0
Oral Corpus as used by BA...		
Oral Corpus as used by BA...		
Tool as used by BAMDES (f...		
Written Corpus as used by...		
Profile for Bedevaartbank	30 July 2010 9:33:53	
Profile for Boedelbank	30 July 2010 9:36:25	

User: twagoo@mpl.nl [settings](#)

Max Planck Institute for Psycholinguistics

Shibboleth Identity Provider Login

Username:

Password:

Element: id string
Concept link: <http://www.isocat.org/datcat/DC-3894>

Edit: Add a new Component



Clarin Component Registry

Component Browser

User: twagoo@mpl.nl [settings](#)

Profiles Components **Work space** ▼

Create new Edit

Component Editor

User: twagoo@mpl.nl [settings](#)

Save Save as new Publish in public space Cancel

Profile Component

Name:

Group Name:

Description:

Domain Name: **Select a domain...** ▼

ConceptLink:

AttributeList: No Attributes
[+attribute](#)

[Collapse all](#) [Expand all](#)

Element: X ▼ ▲

Drag existing components onto edit canvas

Work space ▼ Showing 0 of 0

Name	Group Name	Domain Name	Creator	Description	Registration Date	Comments



Edit: Add a new Component

Create Author Component

Save Save as new Publish in public space Cancel

Profile Component

Name: Author

Group Name: Demo

Description: The author of a comic book

Domain Name: Select a domain...

ConceptLink: http://www.isocat.org/datcat/DC-9123

Search in isocat...

AttributeList: No Attributes

+attr Search in Isocat.org

Q type keyword and press 'enter' to search Search 0 results

Name	Definition	Persistent Identifier	Type	Version

OK Cancel

Collapse +element

Edit: Add a new Component



Save Save as new Publish in public space Cancel

Create Author Component

Profile Component

Name: Author

Group Name: Demo

Description: The author of a comic book

Domain Name: Select a domain...

ConceptLink: <http://www.isocat.org/datcat/DC-9123> Search in isocat...

AttributeList:
+attribute

Collapse all Expand all

Element

Max Occurrences: 1

Type: string

Multilingual:

Edit and choose a type

Type

Value

- primary
- secondary
- other
- Click to add

Edit and choose a type

Type

Controlled vocabulary

Pattern

Enter pattern:

`[[A-Z][A-Z0-9]*)\b{^>}*>(.*)` Use Pattern

Cancel

isocat...



Edit: Create a new profile

Reuse existing components!

Profile Component

Name:

Group Name:

Description:

Domain Name:

ConceptLink:

AttributeList: No Attributes
[+attribute](#)

[Collapse all](#) [Expand all](#)

Element:

Name:	Type:	Cardinality:	Display priority:
Title	string	1 - 1	0

[+element](#)

Component:

Min Occurrences:

Max Occurrences:

ComponentId:

[+component](#)

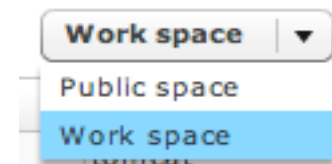
[Collapse all](#) [Expand all](#)

Save/publish



- **save** = save in workspace and overwrite original (only available if it is an edit of an existing profile/component)
- **save as new** = save in workspace as new, so original will still be there if editing an existing profile/component
- **publish** = move to public space, remove from workspace

- Public space is open for everyone
- Profile or Component can be made public when:
 - It is valid (e.g. all fields are filled in correctly).
 - It relies on already public components.



Browse: Result



ComicBook	Demo
Name: ComicBook	
Group Name: Demo	
Description: A comic book resource	
Element: Title string	
DisplayPriority: 1	
Number of occurrences: 1 - 1	
Multilingual: false	
Component:	Author
Name: Author	
Group Name: Demo	
Description: The author of a comic book	
ConceptLink: http://www.isocat.org/datcat/DC-9123	
Element: Name string	
DisplayPriority: 1	
Number of occurrences: 1 - 1	

Ready to be used in
Metadata Editor (Arbil).



Profile to Metadata

- Show info
- Download as XML...
- Download as XSD...
- Edit Item as New...
- Delete Item...

Metadata Profile

Series

Title

Synopsis

Author(s)

Name

Year of birth

Country

Character(s)

Name

Role

Available Templates & Profiles

IMDI Templates <input checked="" type="checkbox"/> Catalogue (internal) <input checked="" type="checkbox"/> Corpus (internal) <input checked="" type="checkbox"/> Session (internal) <input type="checkbox"/> Sign-Language Session (internal)	Clarin Profiles <input type="button" value="Add URL"/> <input type="button" value="Add File"/> <input type="checkbox"/> AnnotationTool <input type="checkbox"/> ArthurianFiction <input type="checkbox"/> BamdesWrittenCorpus <input type="checkbox"/> BASWebService <input type="checkbox"/> Bedevaartbank <input type="checkbox"/> Boedelbank <input type="checkbox"/> CenterProfile <input type="checkbox"/> Chromosome_Example <input type="checkbox"/> cmdi-virtual-collection <input type="checkbox"/> collection <input type="checkbox"/> collection <input checked="" type="checkbox"/> ComicBook <input type="checkbox"/> component-dc-terms <input type="checkbox"/> component-dc-terms-modular <input type="checkbox"/> CorpusProfile <input type="checkbox"/> DBNL <input type="checkbox"/> DcmiTerms <input type="checkbox"/> DIDDD <input type="checkbox"/> DIDDD_sub_location <input type="checkbox"/> DSOL <input type="checkbox"/> DynaSAND <input type="checkbox"/> DynaSAND_sub_location <input type="checkbox"/> EastRepublican <input type="checkbox"/> Example_Profile_Instance <input type="checkbox"/> Familienamenbank <input type="checkbox"/> Gender_Example <input type="checkbox"/> GenericTextStructure <input type="button" value="Reload Clarin Profiles"/> <input checked="" type="checkbox"/> Only load profiles selected for manual editing
---	---

XML schema ("grammar")

```

: schema>
. .
s : schema>
  
```



CLARIN

Common Language Resources and Technology Infrastructure



Thank you for your attention

Component Registry and documentation can be found at:

<http://www.clarin.eu/cmdr>

Questions/Remarks sent to:

cmdr@clarin.eu