



`<CMD_Component />`



Metadata and DCR

Dieter Van Uytvanck

Max Planck Institute for Psycholinguistics

Dieter.VanUytvanck@mpi.nl

ISOcat workshop

Utrecht

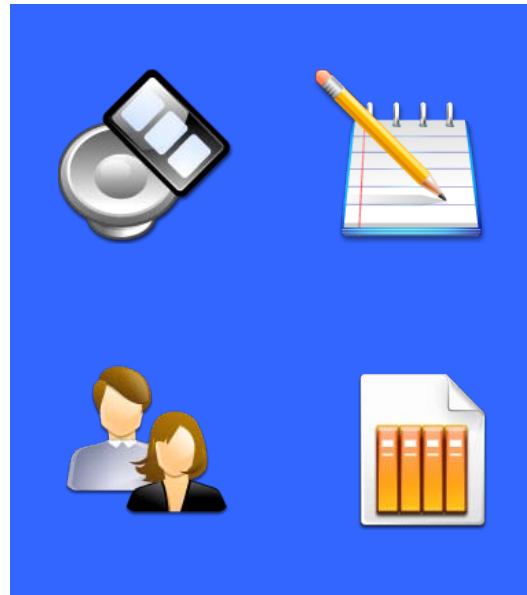
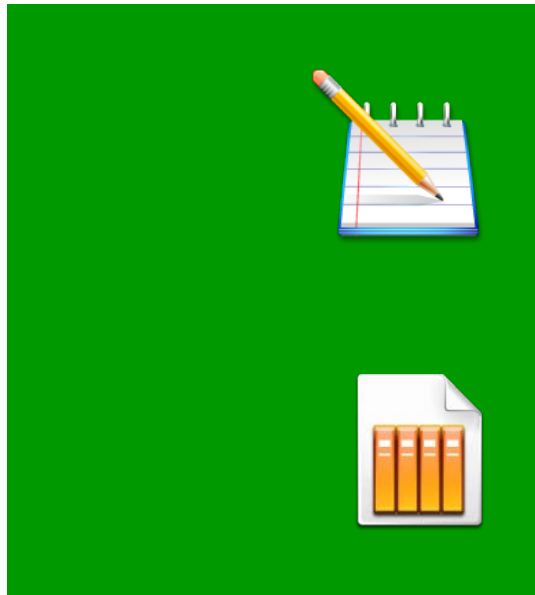
2010-03-25

Overview



- Traditional metadata
- Component metadata
- Data categories
- The big picture
- In practice:
 - Building components
 - Using components

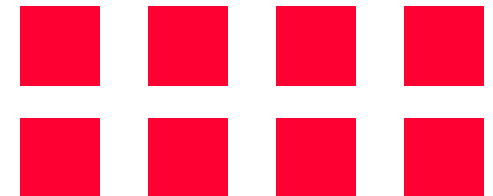
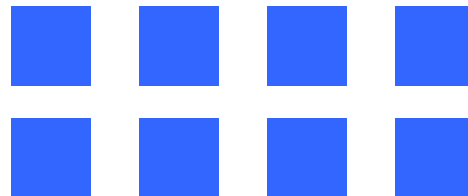
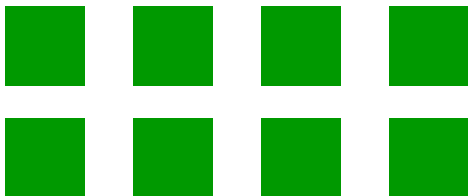
Traditional Metadata



project 1

project 2

project 3



Traditional Metadata: problems

- Lack of flexibility
 - Too many fields...
 - ... but not the ones I am looking for!
- Lack of interoperability
 - My metadata does not work with your infrastructure!
 - Nederland? Netherlands? The Netherlands? Holland? NL?



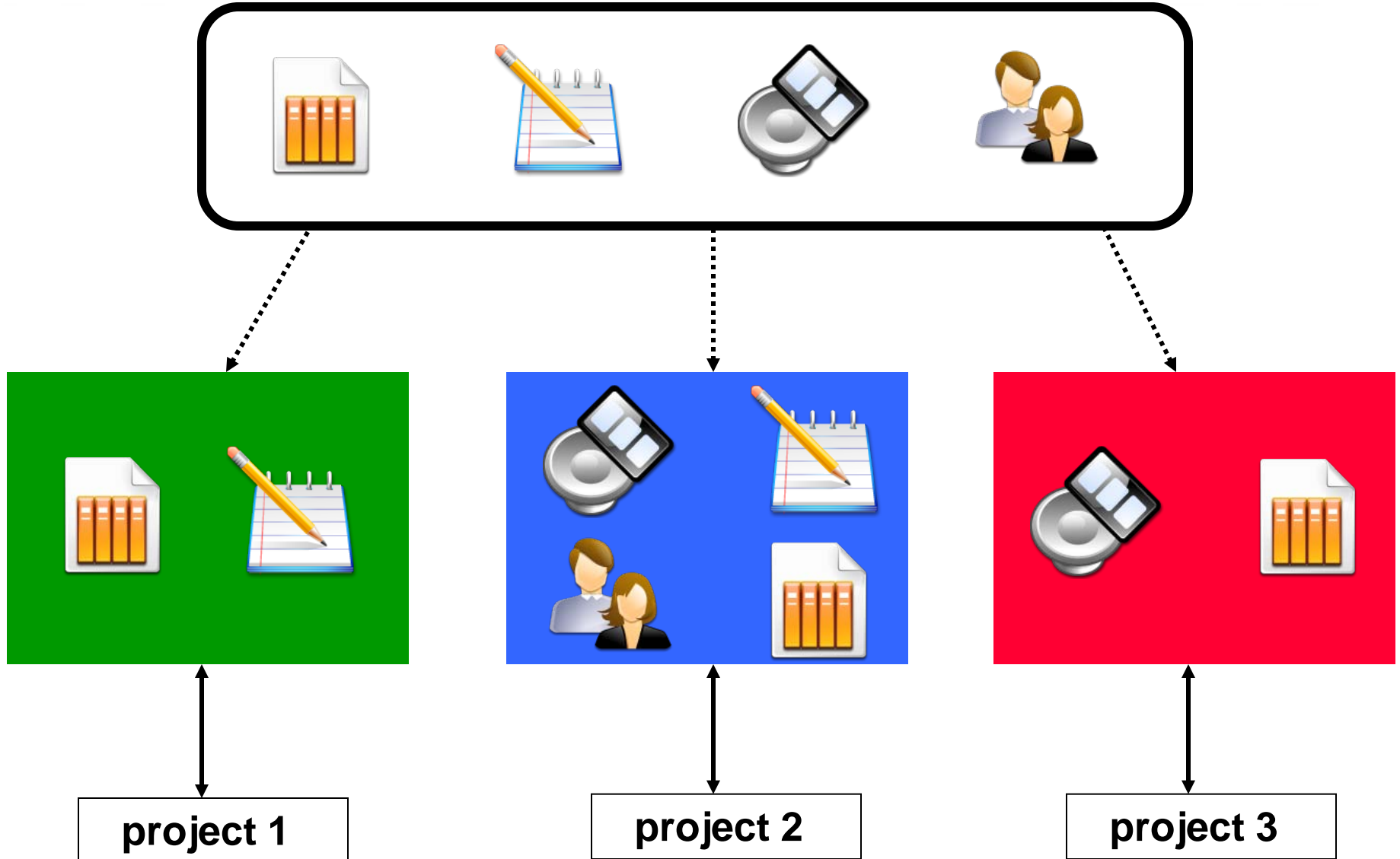
- Other Metadata Infrastructures in our domain:
 - IMDI, OLAC/DC, TEI
- Problems:
 - Inflexible: too many (IMDI) or too few (OLAC) fields
 - Limited interoperability
 - Problematic (unfamiliar) terminology for some sub-communities.
 - etc.

CLARIN Project - CMDI



- Metadata infrastructure based on a “Component Metadata Model”
- Aims
 - Flexibility
 - Researcher should themselves decide what metadata fits their needs
 - Offer ready made metadata components
 - Allow creation of new metadata components needed
 - Interoperability built-in
 - Complete Infrastructure: software for editing, harvesting, exploitation
 - Compatibility with existing frameworks: OLAC, IMDI

Component Metadata



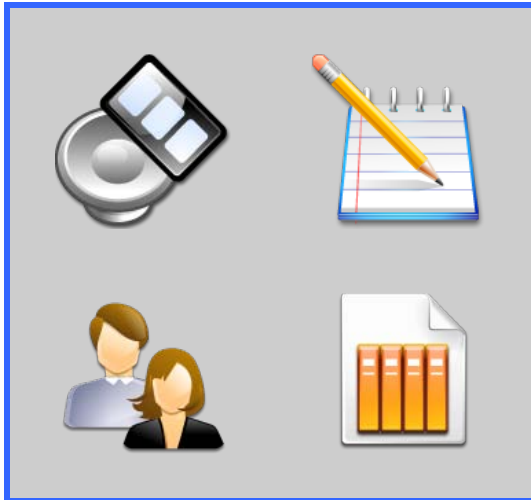
Some terminology



- Element = atomic unit (a “field”) – e.g. recording date
- Component = set of elements – e.g. Actor
- Profile = set of components – e.g. OLAC profile
- Schema = technical (formal) grammar describing a profile – e.g. olac.xsd
- Instance = one metadata description – e.g. myresource.xml

Metadata components?

Metadata Profile
("components à la carte")



XSLT



XML schema
("grammar")

```
<xs:schema>  
  ...  
</xs:schema>
```

XML validator

XML editor



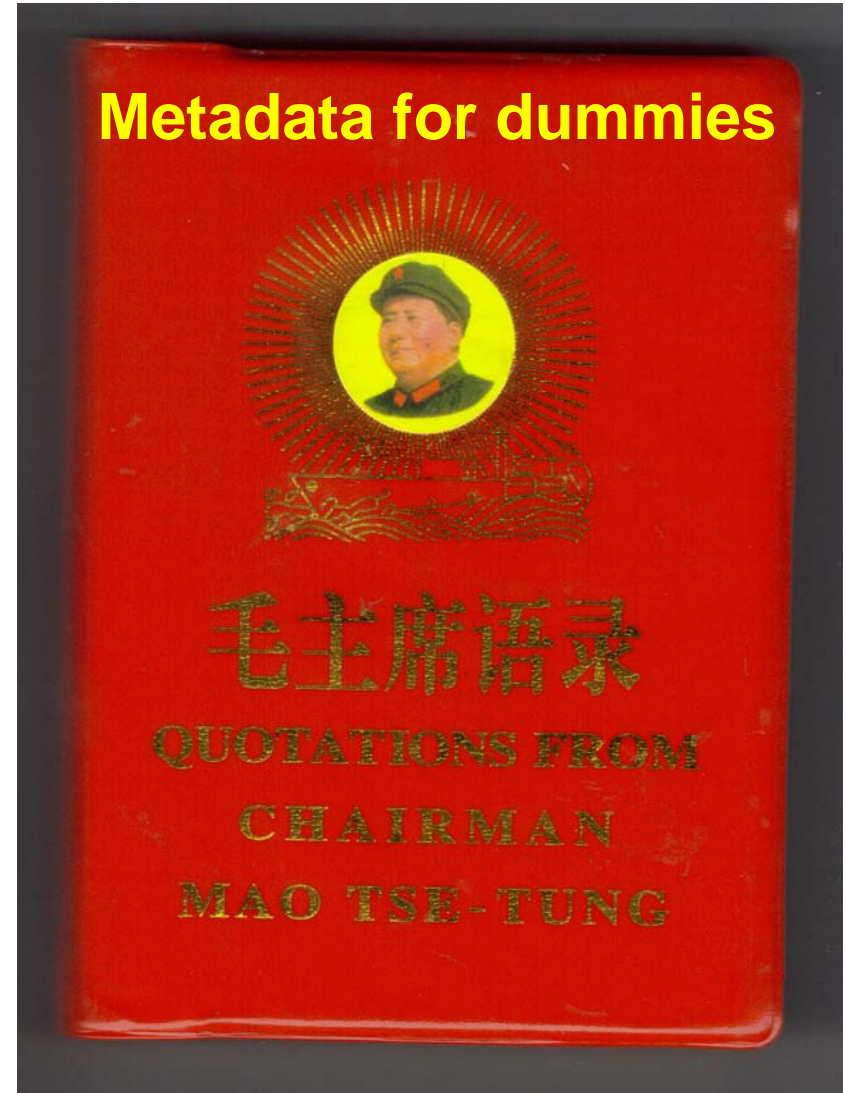
```
<CMD>  
  ...  
</CMD>
```

Metadata instance
(the real resource description)

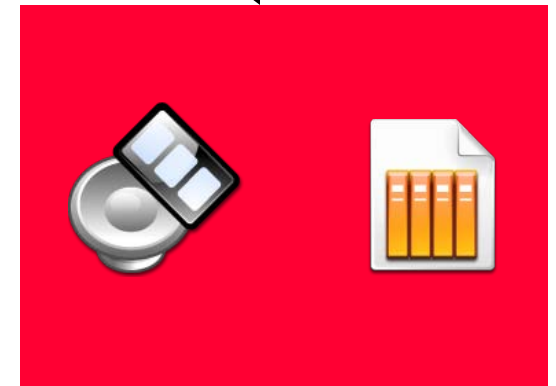
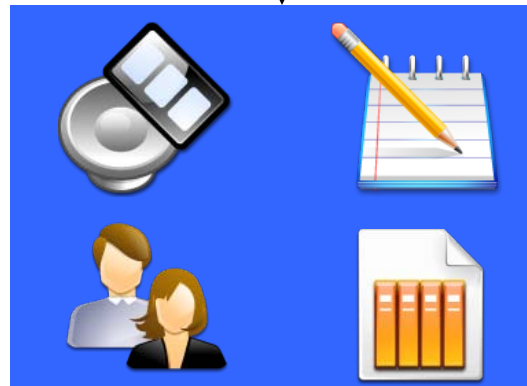
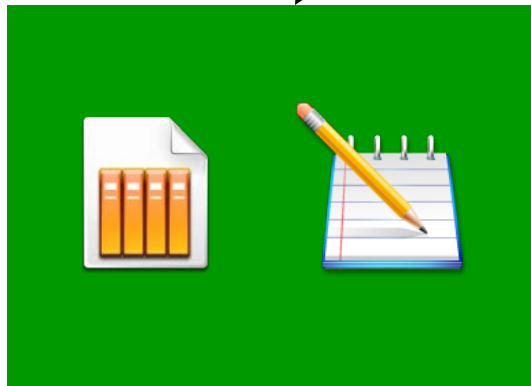
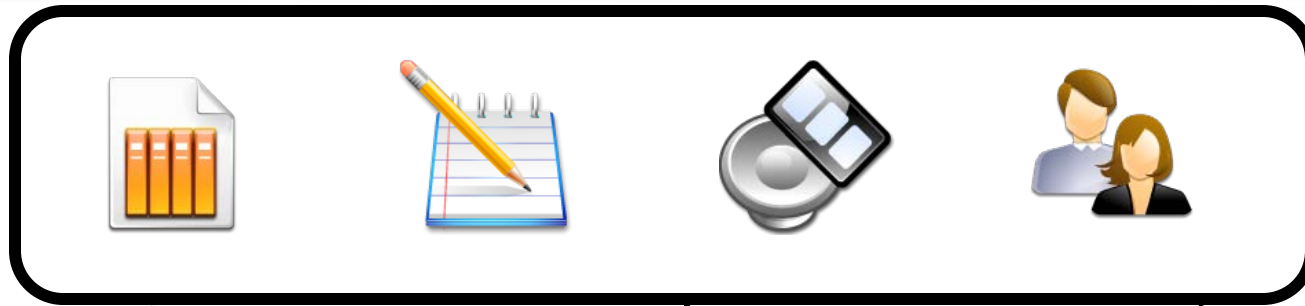
Communist Metadata Infrastructure?



- Are we all forced to use the same components?
 - No!
 - (although re-use is generally a good idea)
- But how to guarantee interoperability while using different components?



Data Categories

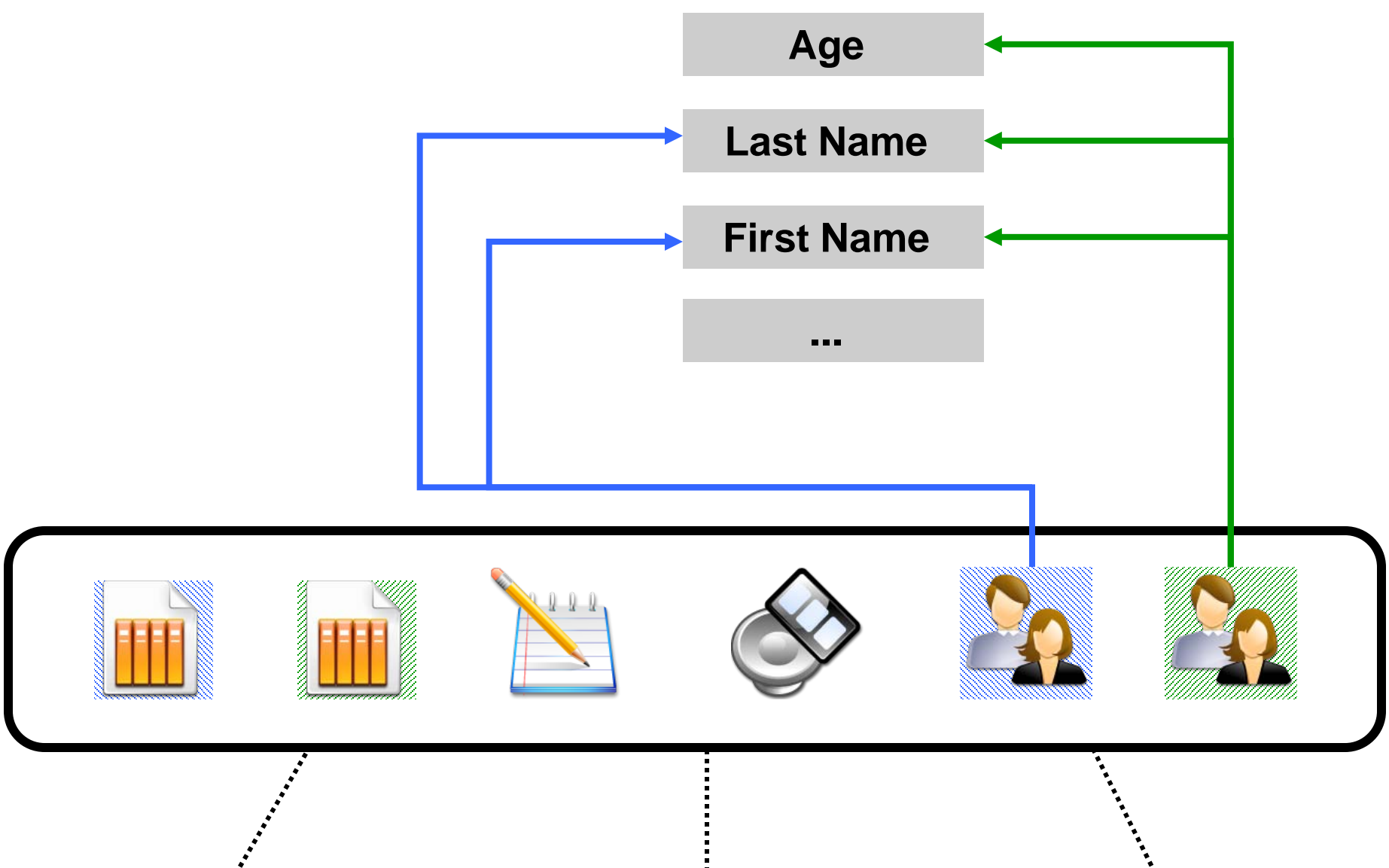


project 1

project 2

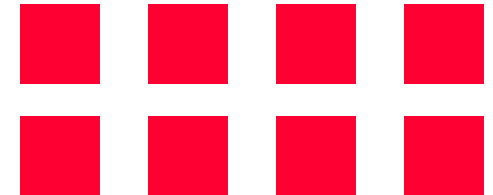
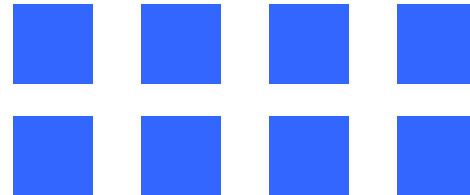
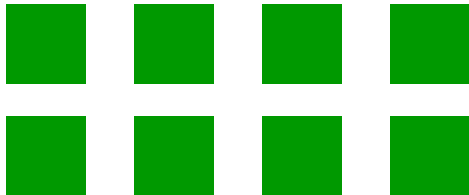
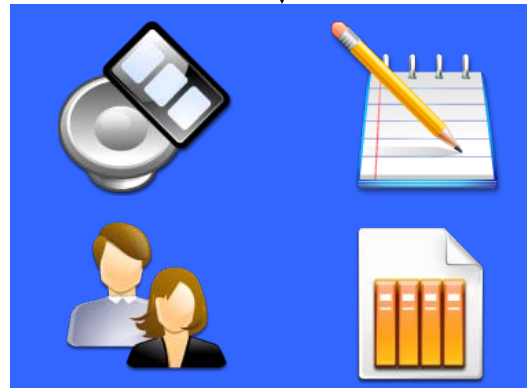
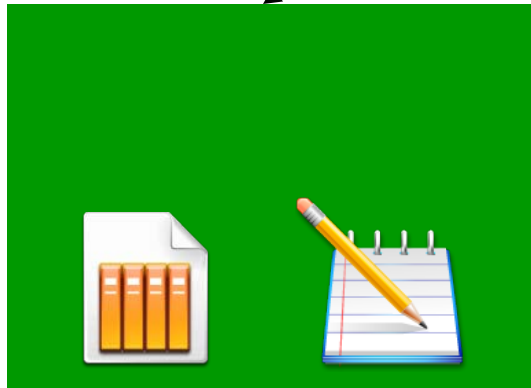
project 3

Data Categories

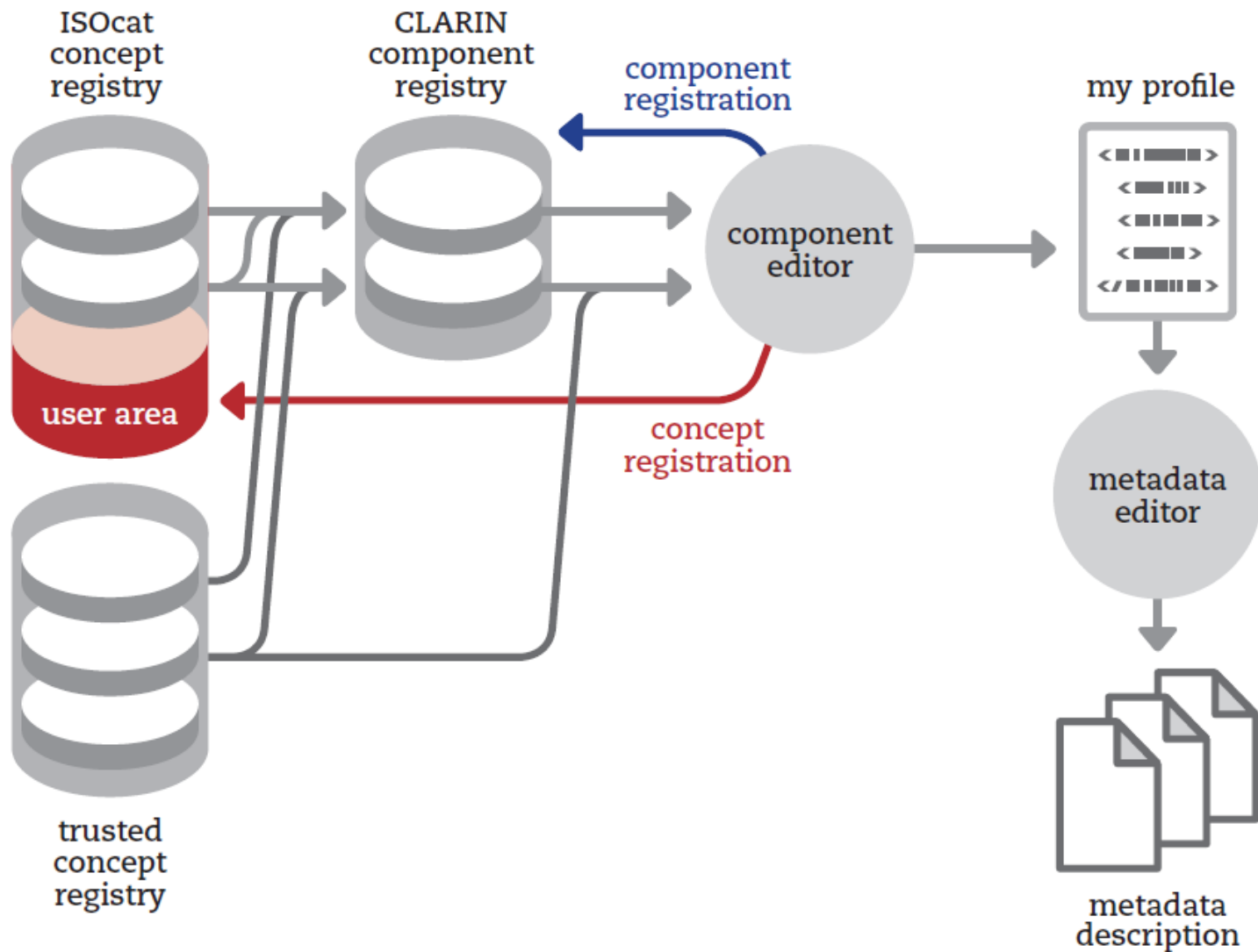


The big picture

Data Category



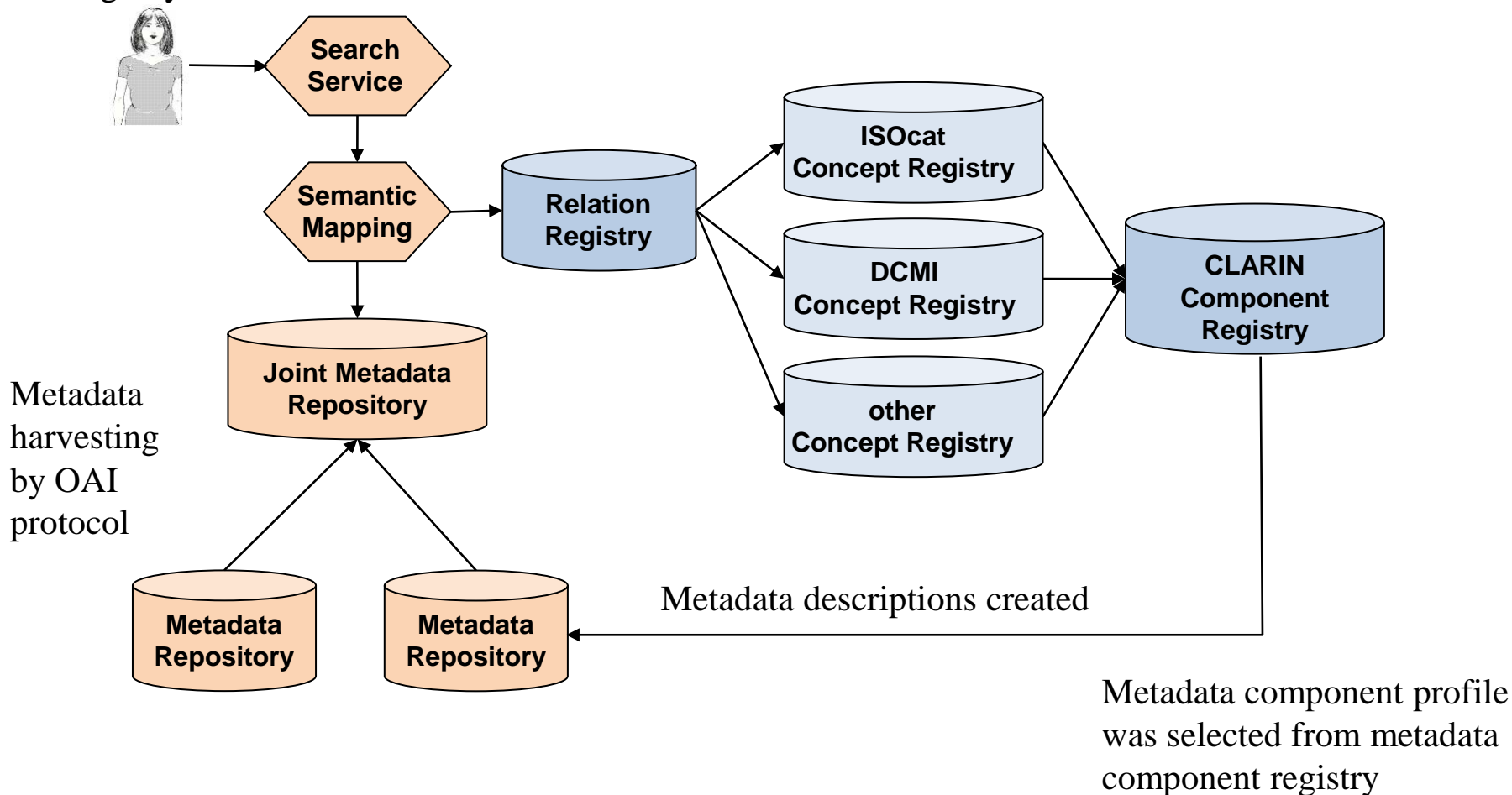
Metadata creation flow



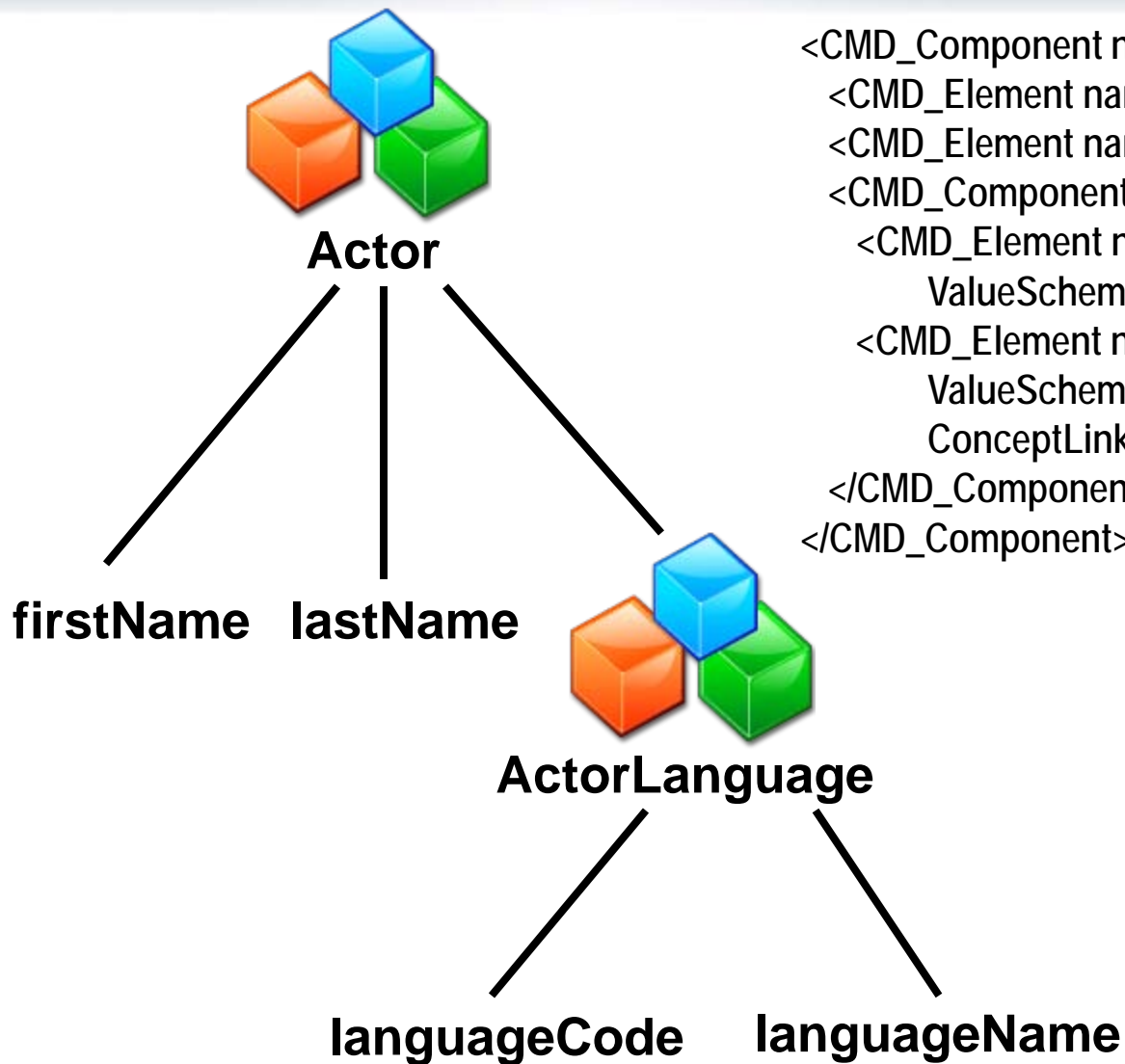
CLARIN MD Live-cycle

Perform search/browsing on the metadata catalog using the ISO DCR and other concept registries and CLARIN relation registry

Create metadata schema from selection of existing components. Allow creation of new components if they have references to ISOcat

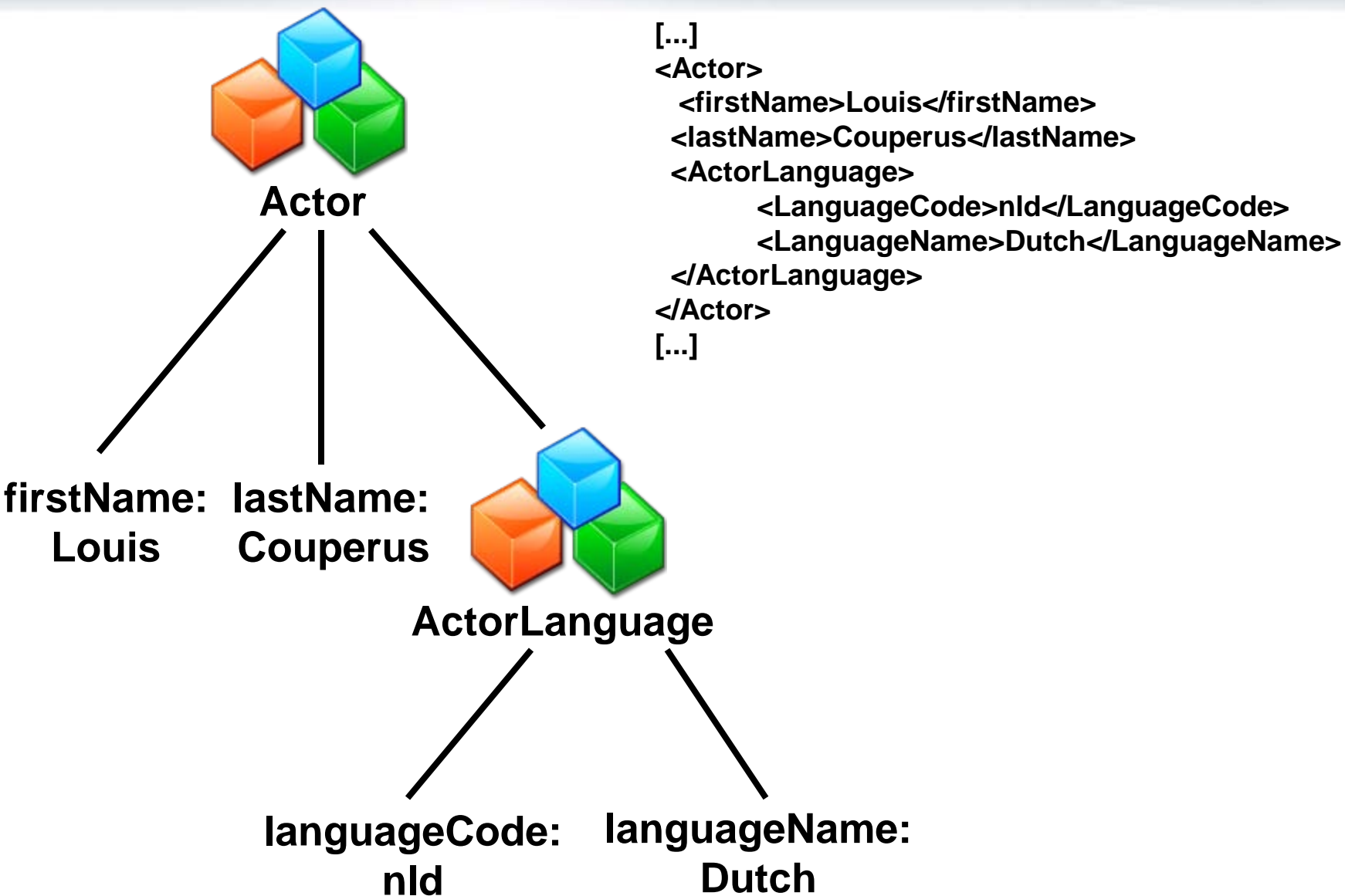


Building a component



```
<CMD_Component name="Actor">
  <CMD_Element name="firstName" ValueScheme="string"/>
  <CMD_Element name="lastName" ValueScheme="string"/>
  <CMD_Component name="ActorLanguage">
    <CMD_Element name="LanguageCode"
      ValueScheme="string" />
    <CMD_Element name="LanguageName"
      ValueScheme="string"
      ConceptLink="http://www.isocat.org/datcat/DC-1766"/>
  </CMD_Component>
</CMD_Component>
```


Using a component



New metadata datcats? (1)



- Register at isocat.org
- Send a mail with your name to dieter.vanuytvanck@mpi.nl
- You will be added to the CLARIN-NL group
- My Workspace > button “create new data category”
- My Workspace > Private > CLARIN-NL > MD proposal > button “edit this data category selection”
- My Workspace > Private > + (add this data category to selection)
- Click on the icon for “save the selected data categories”

New metadata datcats? (2)



- After inspection the new datacats can be moved to the Metadata thematic view
- Then the datacats in the Metadata thematic view can be submitted to the Thematic Domain Group for official approval

Important points



- Re-use components + profiles
- See eg CLARIN-NL components:
 - <http://www.clarin.eu/cmd/components/clarin-nl/>

Component Registry + ISOcat



- Lookup function, see demo
 - <http://lux16.mpi.nl:8080/ds/ComponentRegistry/#>

Metadata workshop



ISOcat workshop

Utrecht

2010-03-25

www.clarin.eu

- May 27, more information soon ...

Conclusions



- Building your own components and profiles is already possible
- Creating CLARIN metadata descriptions too
- Both things require some technical (XML) skills
- This is **not** the final infrastructure:
 - Format will be supported in the future
 - To be expected: user friendly
 - editor
 - Browser
 - Component registry
 - search engine

Where to get the toolkit?

- <http://www.clarin.eu/toolkit>



CLARIN

Common Language Resources and Technology Infrastructure



Thank you for your attention

CLARIN has received funding from
the European Community's Seventh Framework Programme
under grant agreement n° 212230

Backup slides



ISOcat workshop
Utrecht
2010-03-25

www.clarin.eu