

# SMC Browser

CLARIN Annual Meeting  
2013-10-21 - Prague  
Matej Ďurčo, ICLTT, Vienna

growing size and complexity of the  $CMD_s$  data domain

i.e. the profile and component definitions

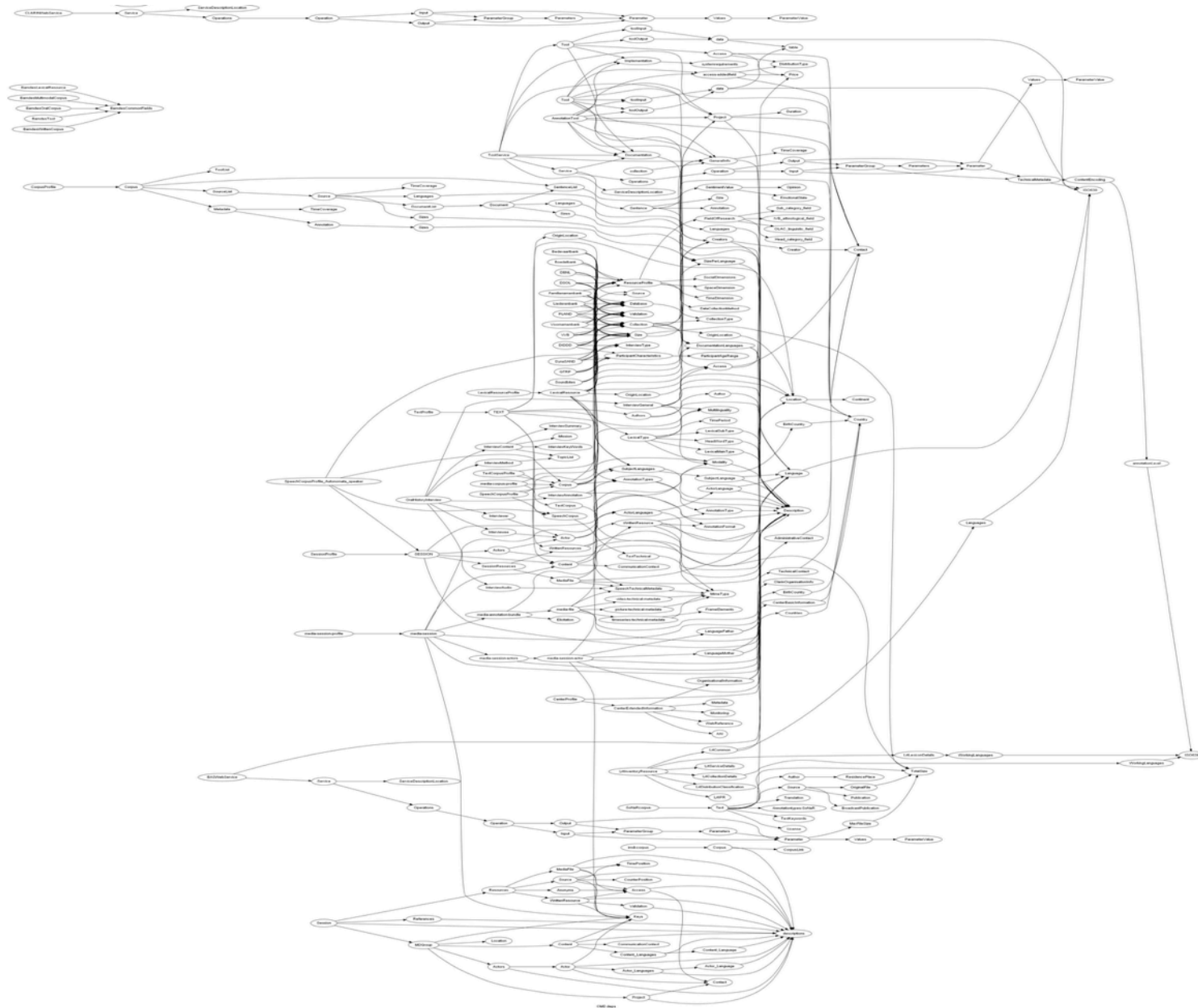
=> need for enhanced ways of exploration

- component structure → tree
  - component and data category reuse → graph
- => interactive (web-based) graph visualization
- primary user: metadata modeller
  - some requirements:
    - select multiple nodes,
    - traverse “up” and “down” the graph
    - display auxiliary information (link back to definition)



# Features

- web application
  - old <http://clarin.aac.ac.at/smc-browser/>
  - new <http://clarin.arz.oeaw.ac.at/smc-browser>
  - (best performance in Chrome)
- data served static – generated in preprocessing step  
(fetch all data from Component Registry, Data Category Registry,...)  
=> UI initialization can take some time
- based on [d3](#) – data driven documents  
wonderful javascript/SVG visualization library
- automatic graph layout  
multiple alternative layouts based on [force layout](#)
- reuse and size statistics
- send link, download svg
- usable for any graph data



link



CLARIN SMC Browser

[home](#) [docs](#) [stats](#) [examples](#) [reports](#)

```
2013-10-20 8:52:25 :show nodes: 130; show links: 147; max count:undefined;
node_size_ration:undefined
2013-10-20 8:52:13 :show nodes: 130; show links: 147; max count:undefined;
```

home docs stats examples reports

graph SMC graph basic depth-before 2 depth-after 5 link-distance 219 charge 328 friction 63

gravity 8 node-size 4 labels show curve straight-arrow layout horizontal-tree link download

**Index** Access

**Component [10]**

- Access [15]
- Access [36]
- Access [30]
- Access [4]
- Access [6]
- Access [4]
- Access [2]
- ResourceAccess [1]
- access-addedfield [2]
- accessTool [4]

**Element [20]**

- Access [1]
- Access [1]
- Access [1]
- Access [1]
- AccessProtocol [1]
- AccessRights [4]
- AccessPoint [1]
- ResourceAccessAvail [1]
- ResourceAccessDate [1]
- ResourceAccessLink [1]
- ResourceAccessOwn [1]
- ResourceAccessPubl [1]
- accessProtocol [2]
- accessProtocol [2]
- accessRights [1]
- accessRights [2]
- accessRights [1]
- accessRights [1]
- distributionAccessMed [4]
- distributionAccessMed [2]

**DatCat [2]**

- Access Rights [25]
- accessProtocol [19]

**Detail**

**Overview**

| Overall   |                  |
|---|------------------|
| created   | 2013-10-12+02:00 |
| Profiles  | 158              |
| Components  | 6681             |
| distinct Components                                     | 994              |
| Elements  | 15103            |
| distinct Elements                                       | 2817             |
| Elements with DatCats                                   | 2109             |
| Elements without DatCats                                | 719              |
| ratio of elements without DatCats                       | 25.52 %          |
| used Concept  | 636              |
| available Concepts (in Metadata profile or used in CMD) | 1148             |

**Component [3]**

**Access [15]**

clarin.eu:cr1:c\_1357720977472

| Access                            |        |
|-----------------------------------|--------|
| used in total                     | 15     |
| used in Profiles                  | 15     |
| has Components                    | 3      |
| has Elements                      | 18     |
| Elements with DatCats             | 18     |
| uses DatCats                      | 17     |
| Elements without DatCats          | 0      |
| ratio of elements without DatCats | 0.00 % |

**Access [36]**

clarin.eu:cr1:c\_1271859438124

| Access                            |        |
|-----------------------------------|--------|
| used in total                     | 36     |
| used in Profiles                  | 35     |
| has Components                    | 2      |
| has Elements                      | 10     |
| Elements with DatCats             | 10     |
| uses DatCats                      | 10     |
| Elements without DatCats          | 0      |
| ratio of elements without DatCats | 0.00 % |

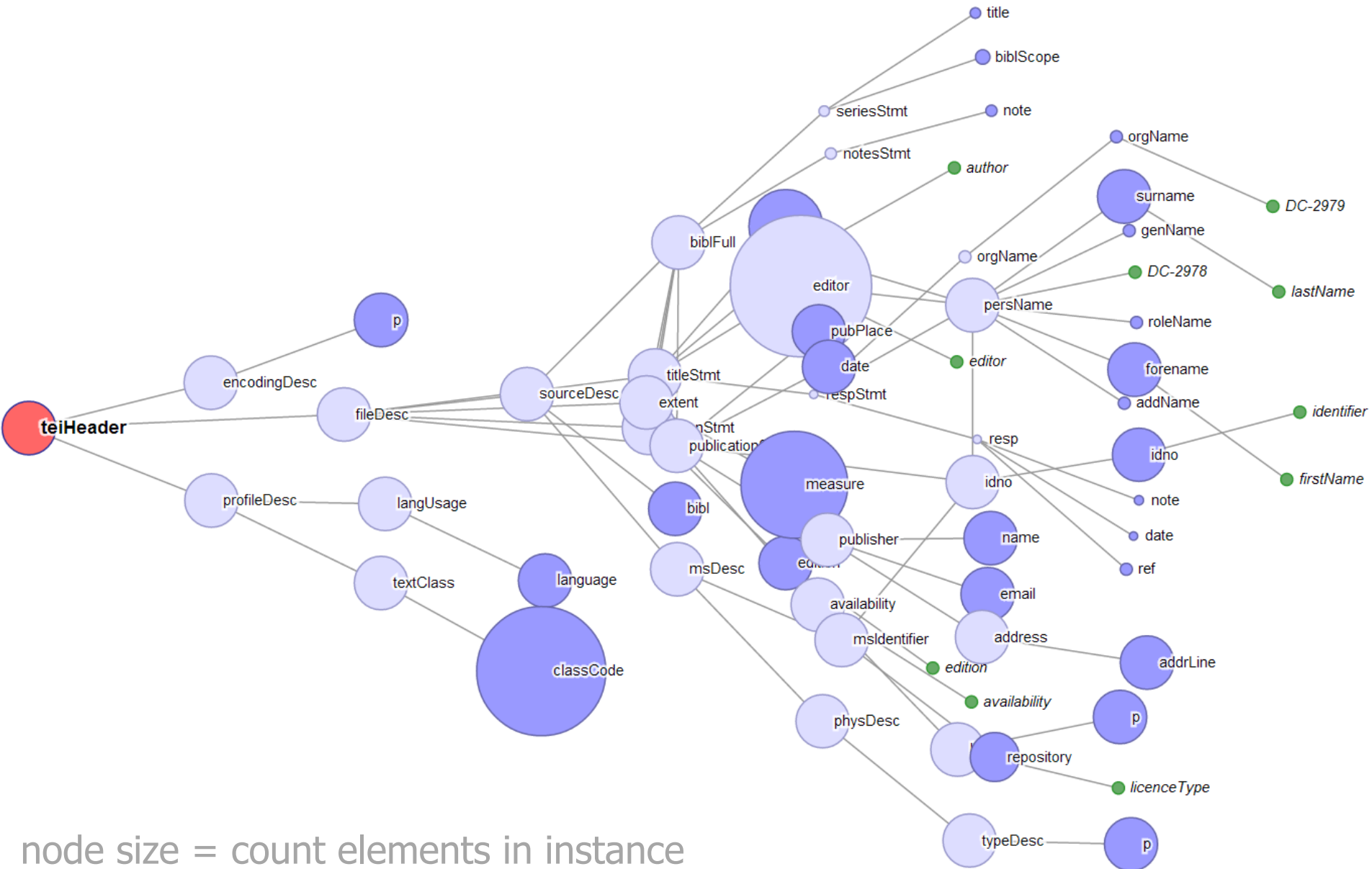
**Access [30]**

clarin.eu:cr1:c\_1290431694501

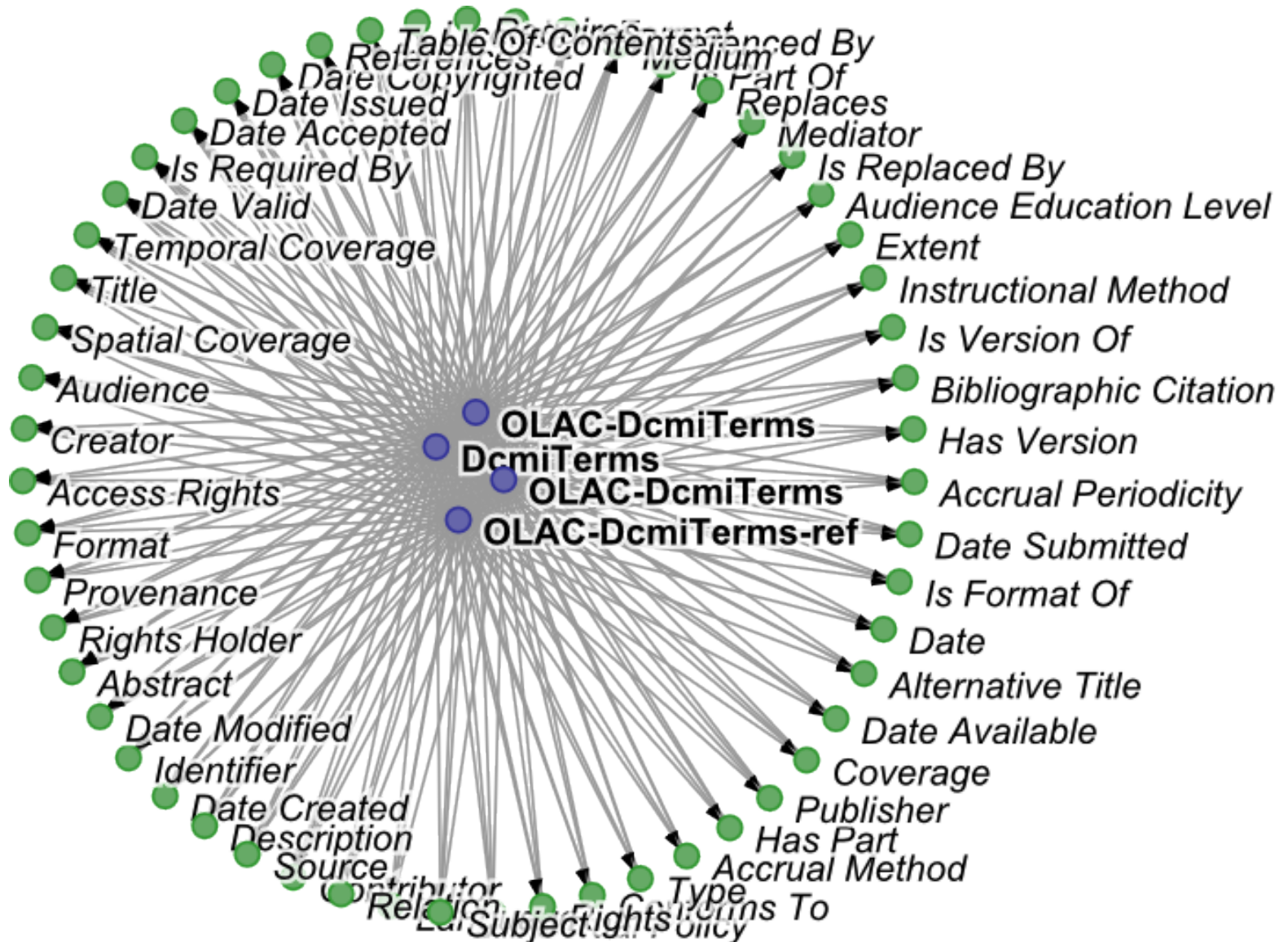
| Access           |    |
|------------------|----|
| used in total    | 30 |
| used in Profiles | 21 |
| has Components   |    |

# Examples of output

# teiHeader (DTA)

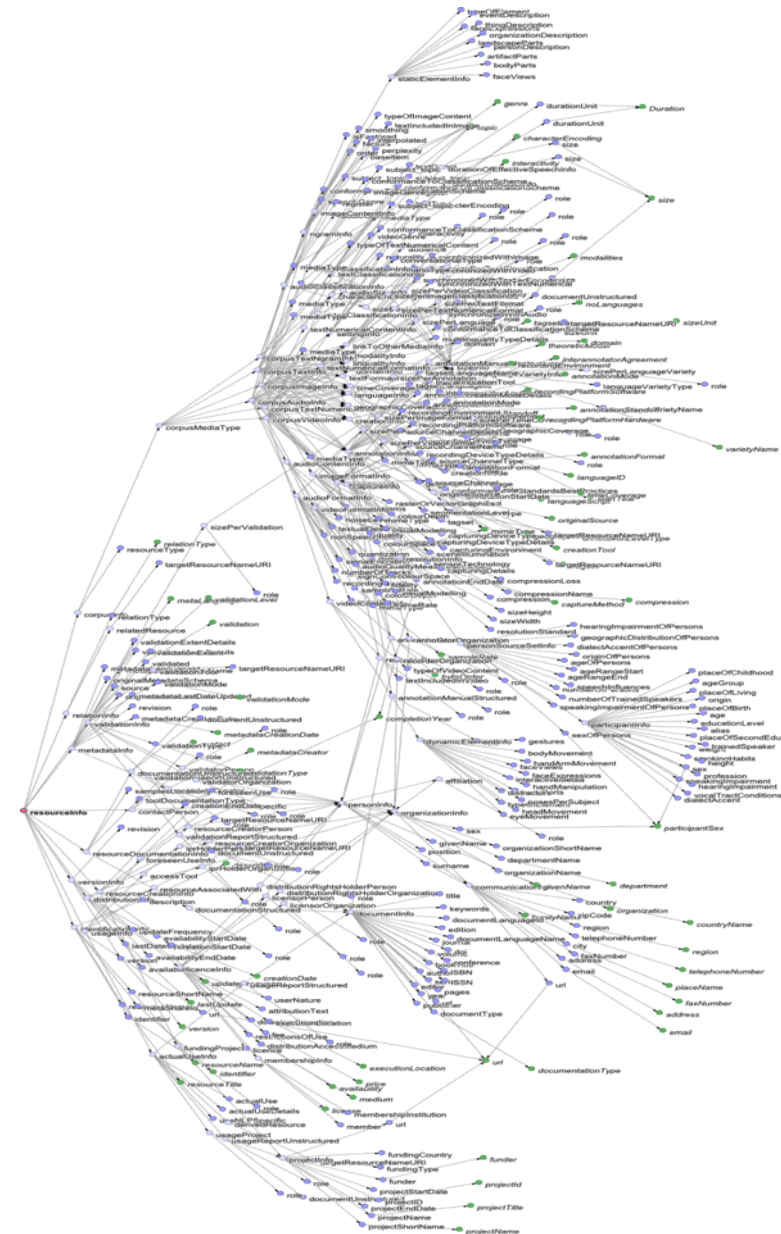
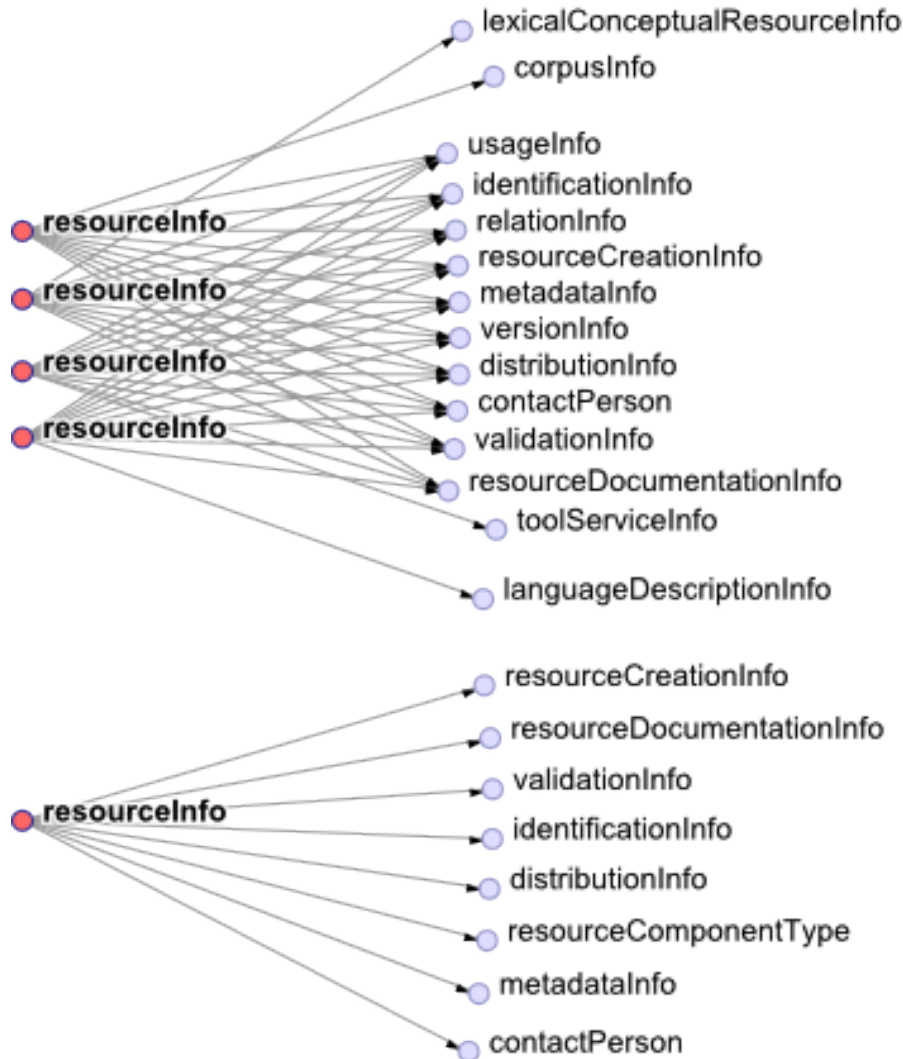




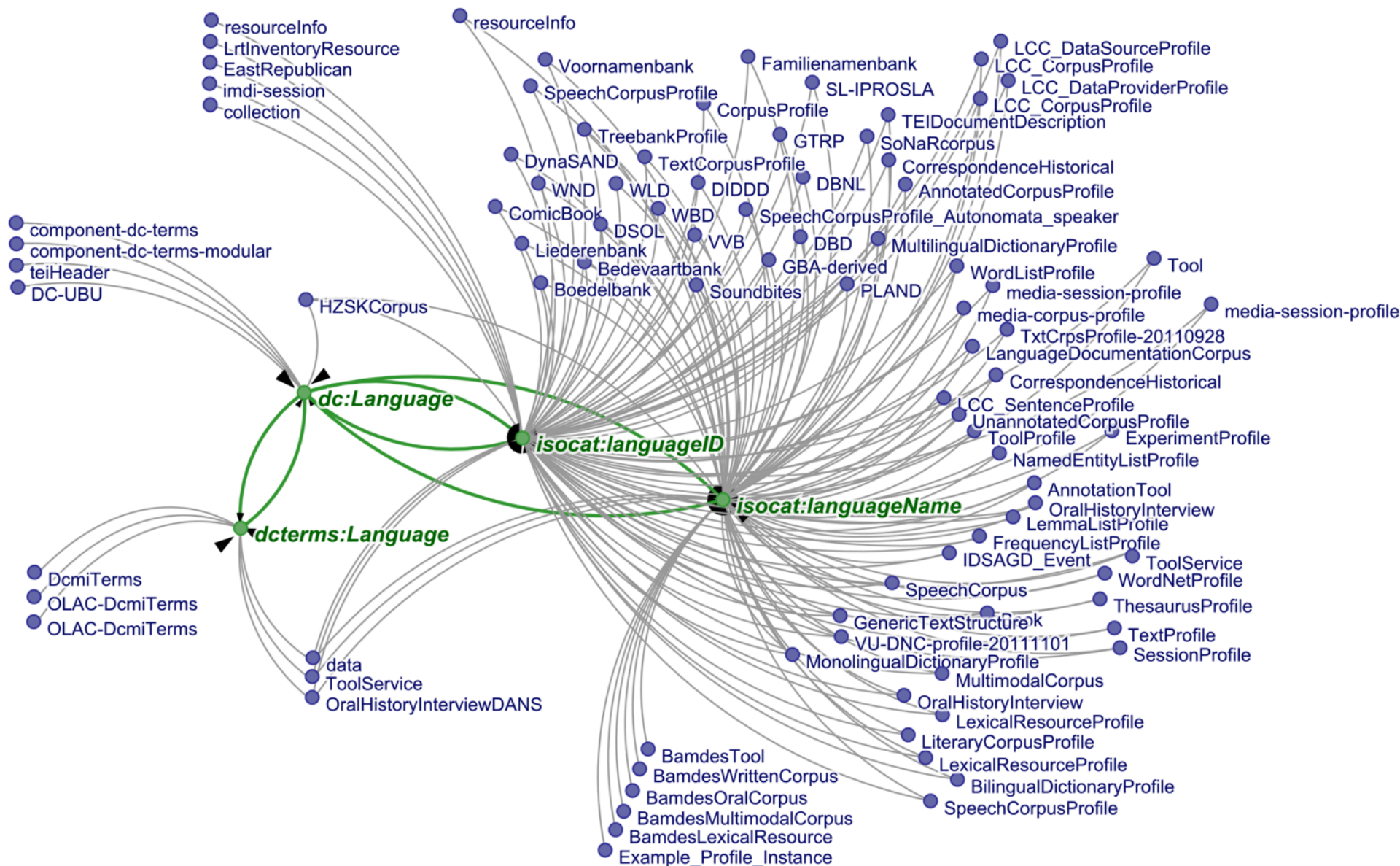




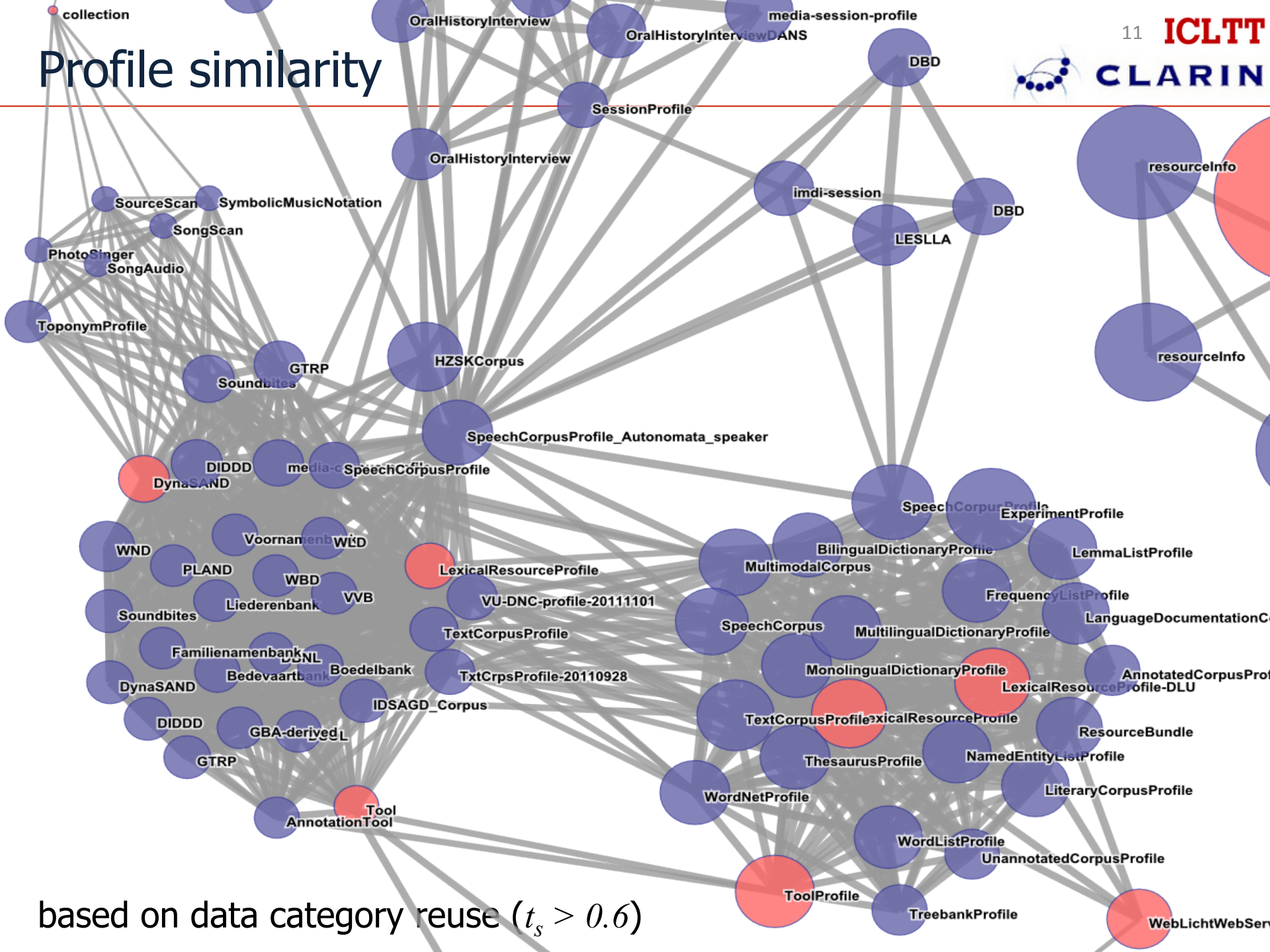
## resourceInfo (META-SHARE)



# data categories: Language



based on data category reuse ( $t_s > 0.6$ )





## Profile similarity - matrix

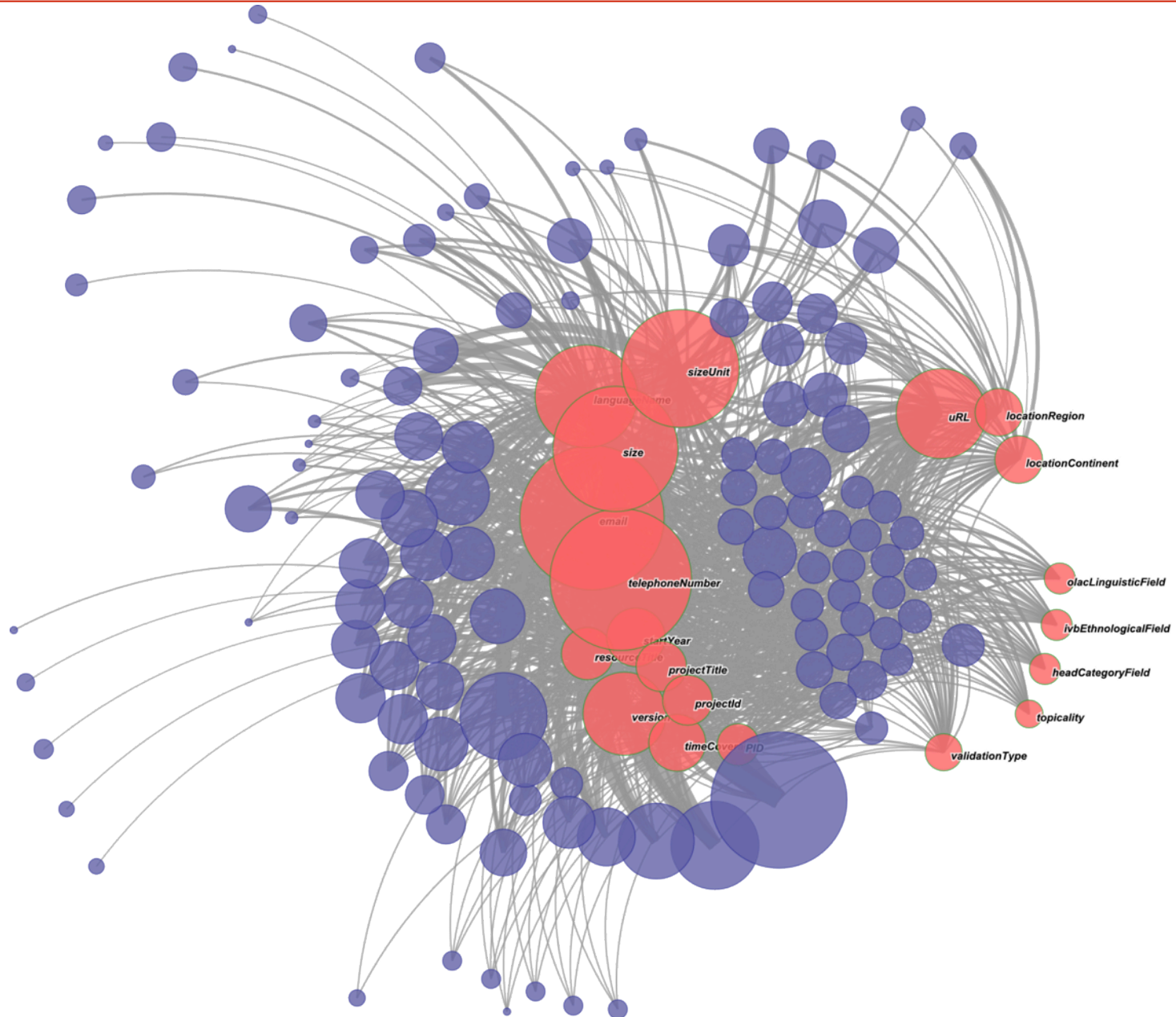
[illegible]

# Reuse of data categories

13



ICLTT  
CLARIN

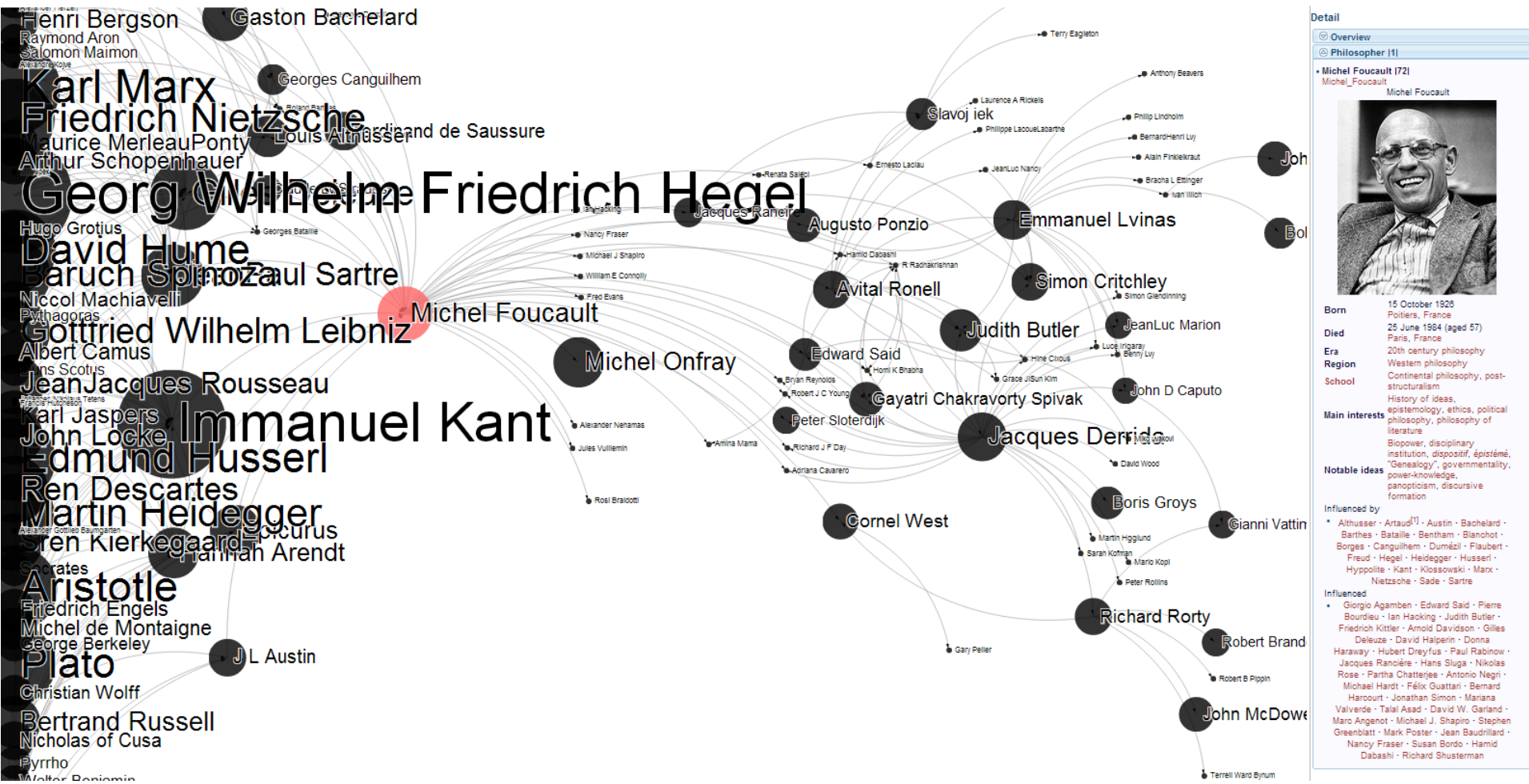


- set operations on subgraphs
  - => differential views (e.g. show just shared components)
- integrate instance data (CMD records)
- integrate with Component Registry
  - on the fly visualization of developed profiles
  - reuse/size statistics as input for modeller's reuse decision
- open for external (graph) data
  - = generic stand-alone graph viewer

# Philosopher influence

extract influence relations from dbpedia, convert to graph, feed to smc-browser

[link](#)

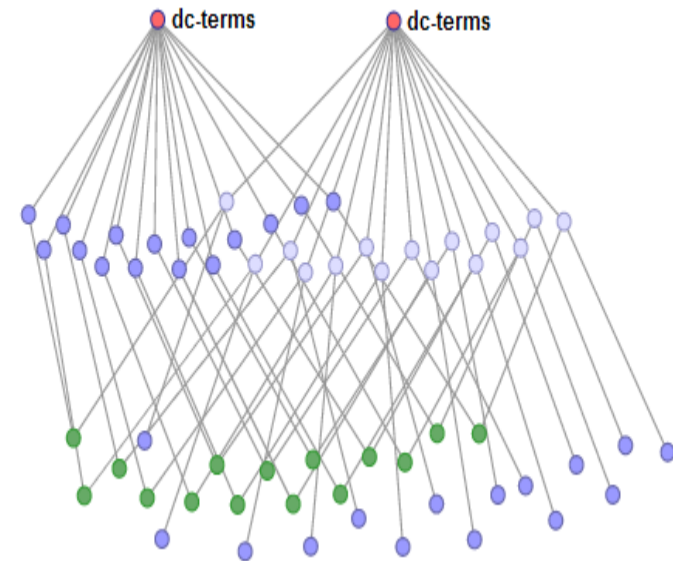
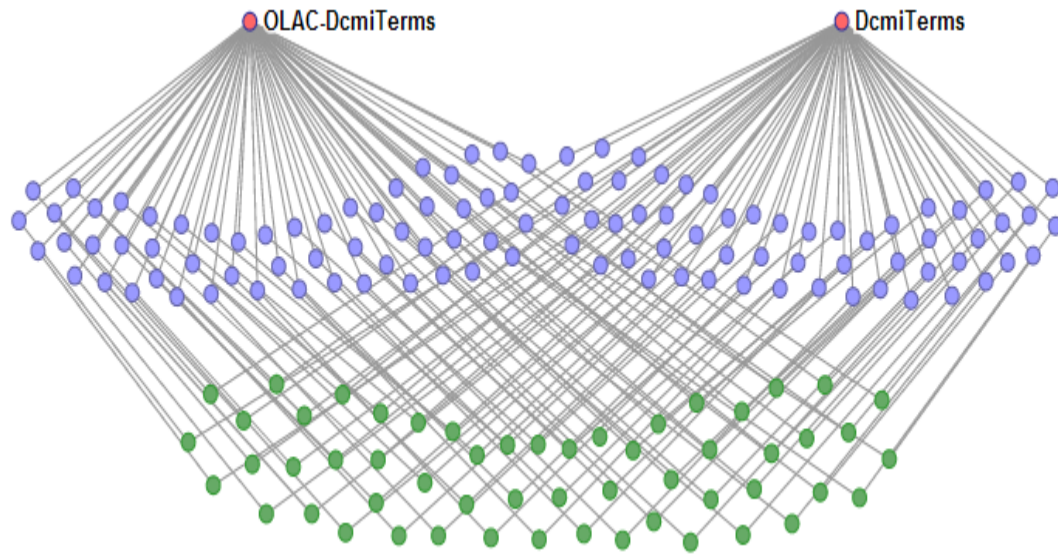




# Thank you

# More graphs

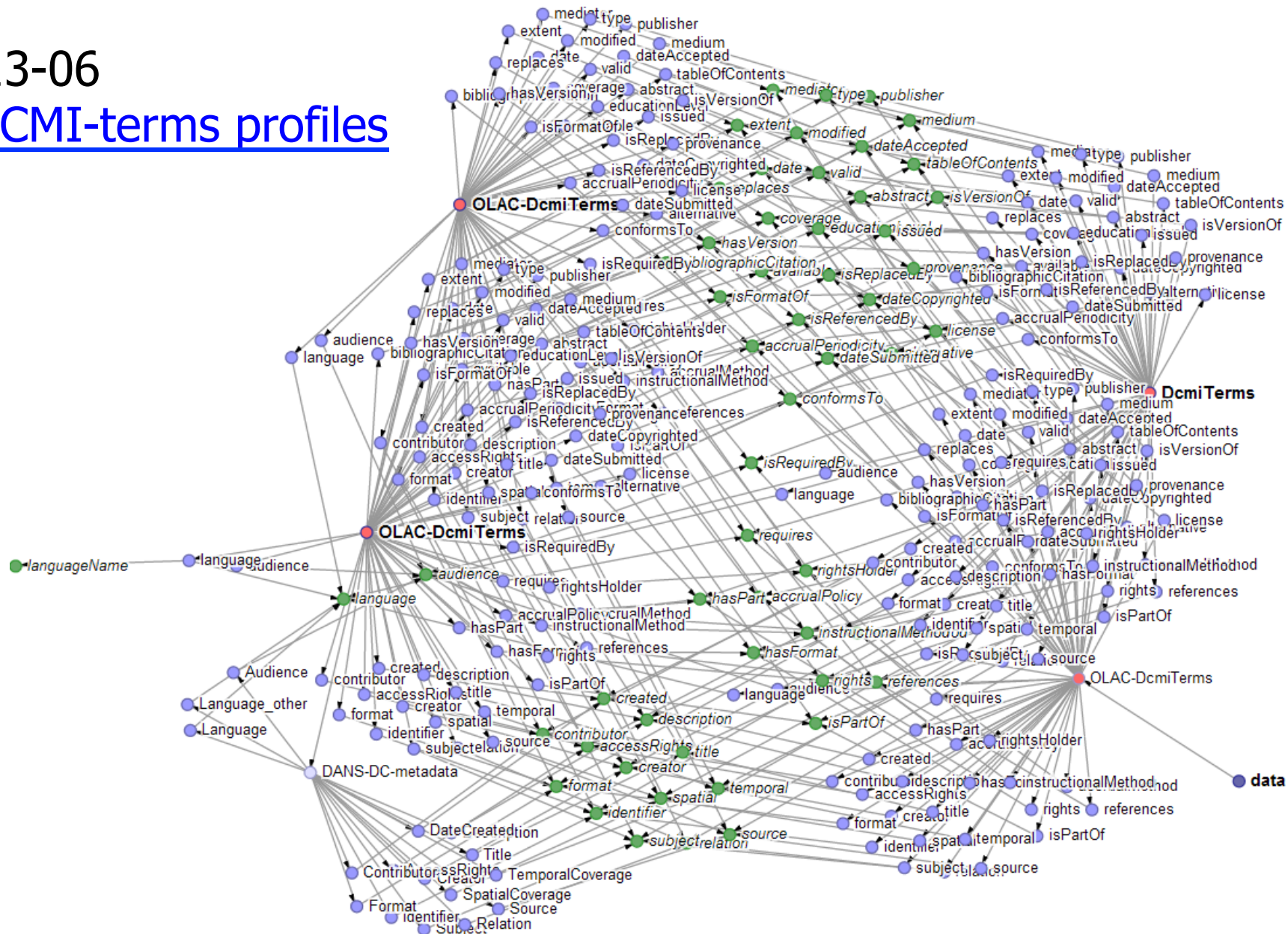
- 2 profiles with dc-terms (55 data categories)
- 2 profiles with dc-elements (called „dc-terms“)  
as of 2013-01



# dublincore II

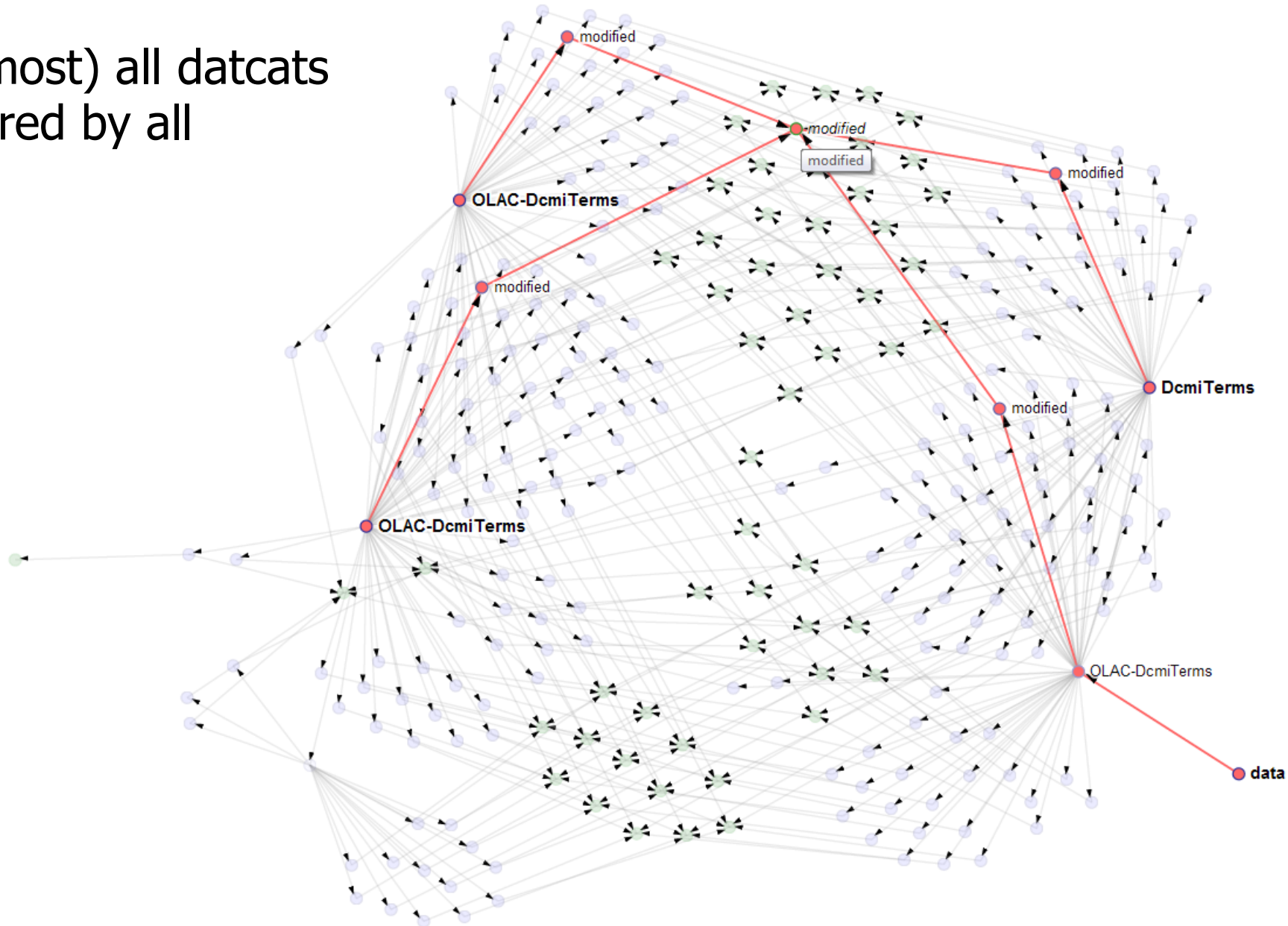
2013-06

## 4 DCMI-terms profiles





(almost) all datcats  
shared by all

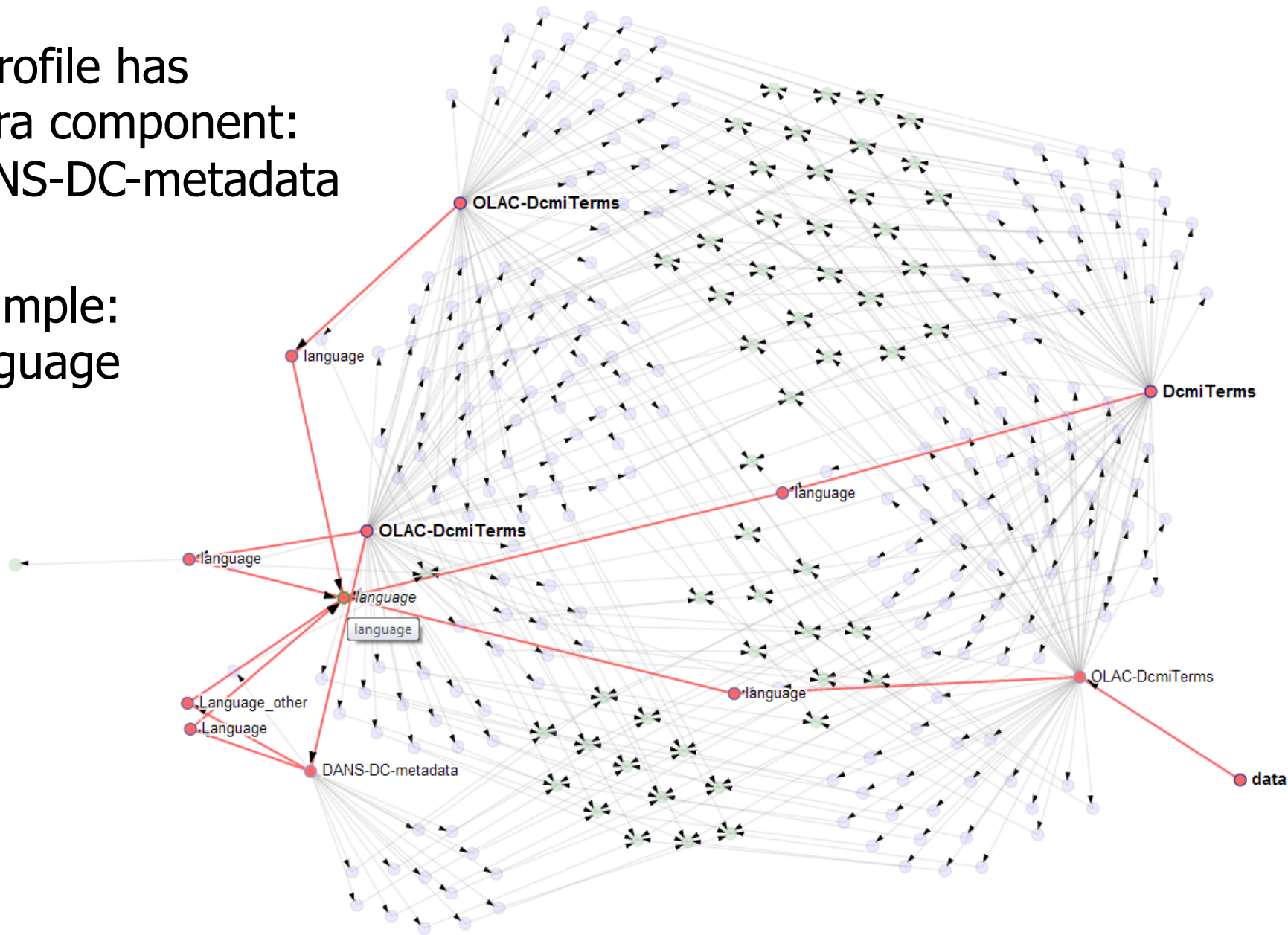






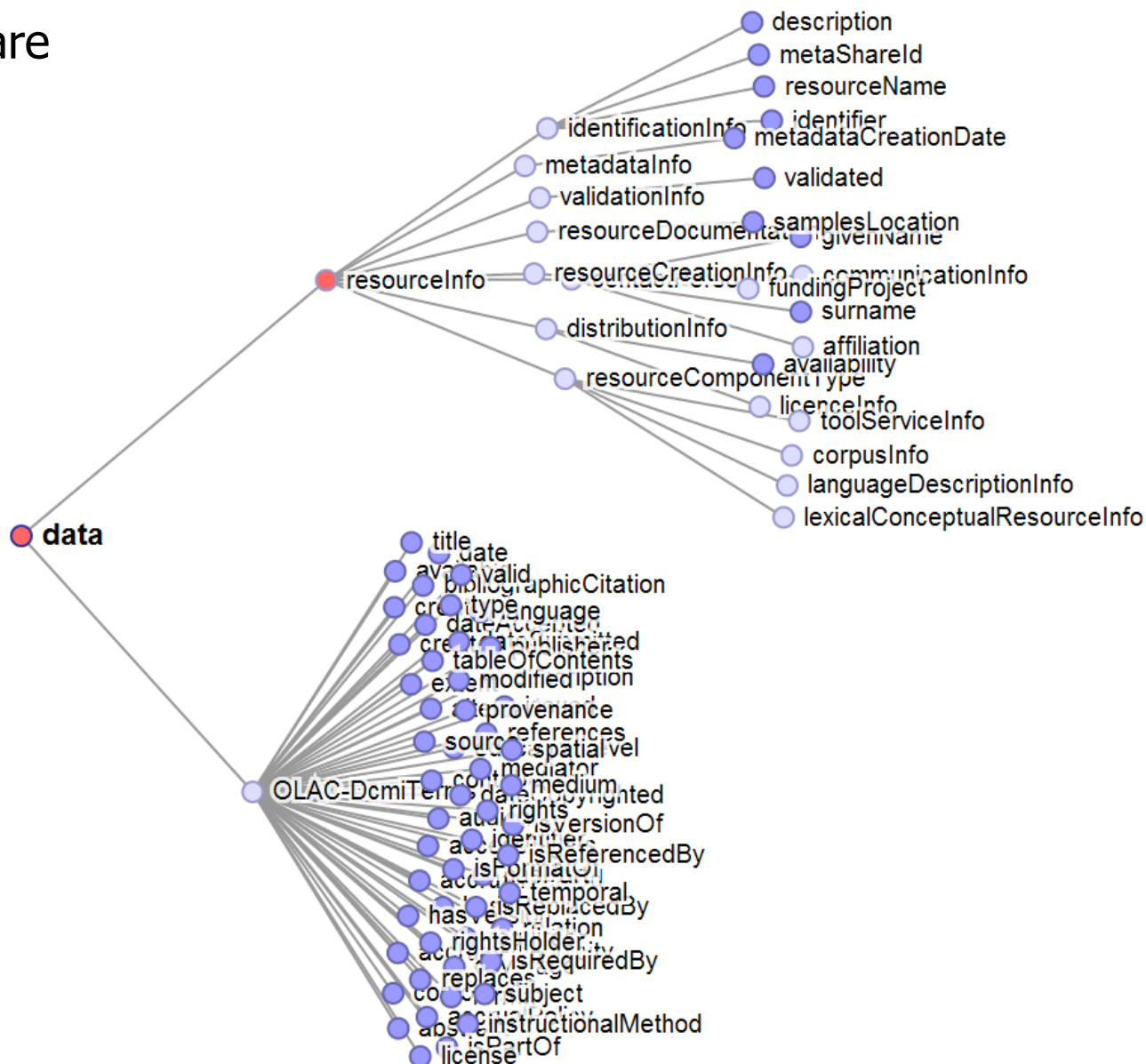
1 profile has  
extra component:  
DANS-DC-metadata

example:  
language



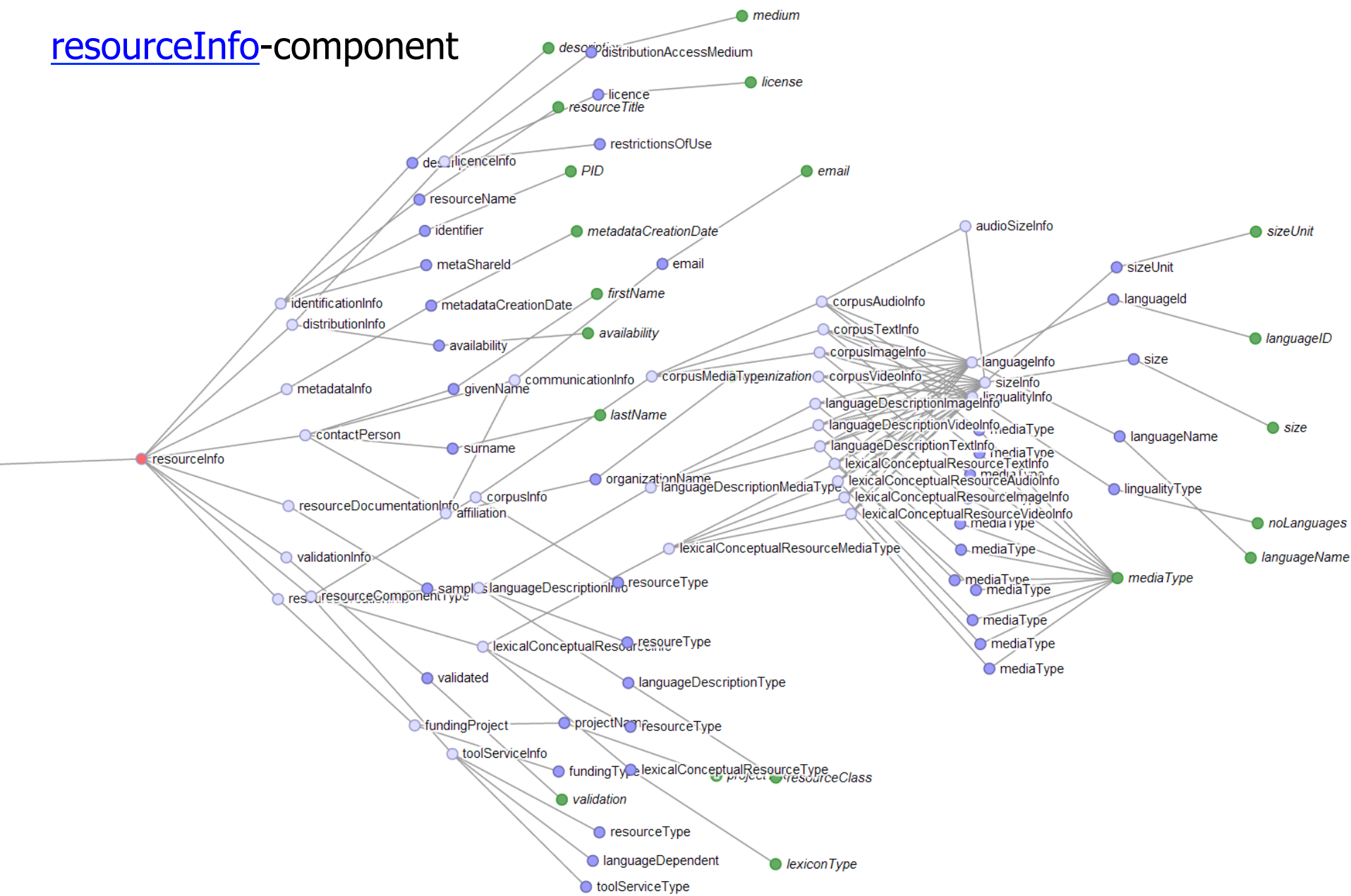
# META-SHARE: data profile (LINDAT)

dublincore + metashare





## resourceInfo-component



# teiHeader (ICLTT)

based on the [recommended header](#)

