SMC Browser

CLARIN Annual Meeting

2013-10-21 - Prague
Matej Ďurčo, ICLTT, Vienna

Idea

growing size and complexity of the CMD_s data domain
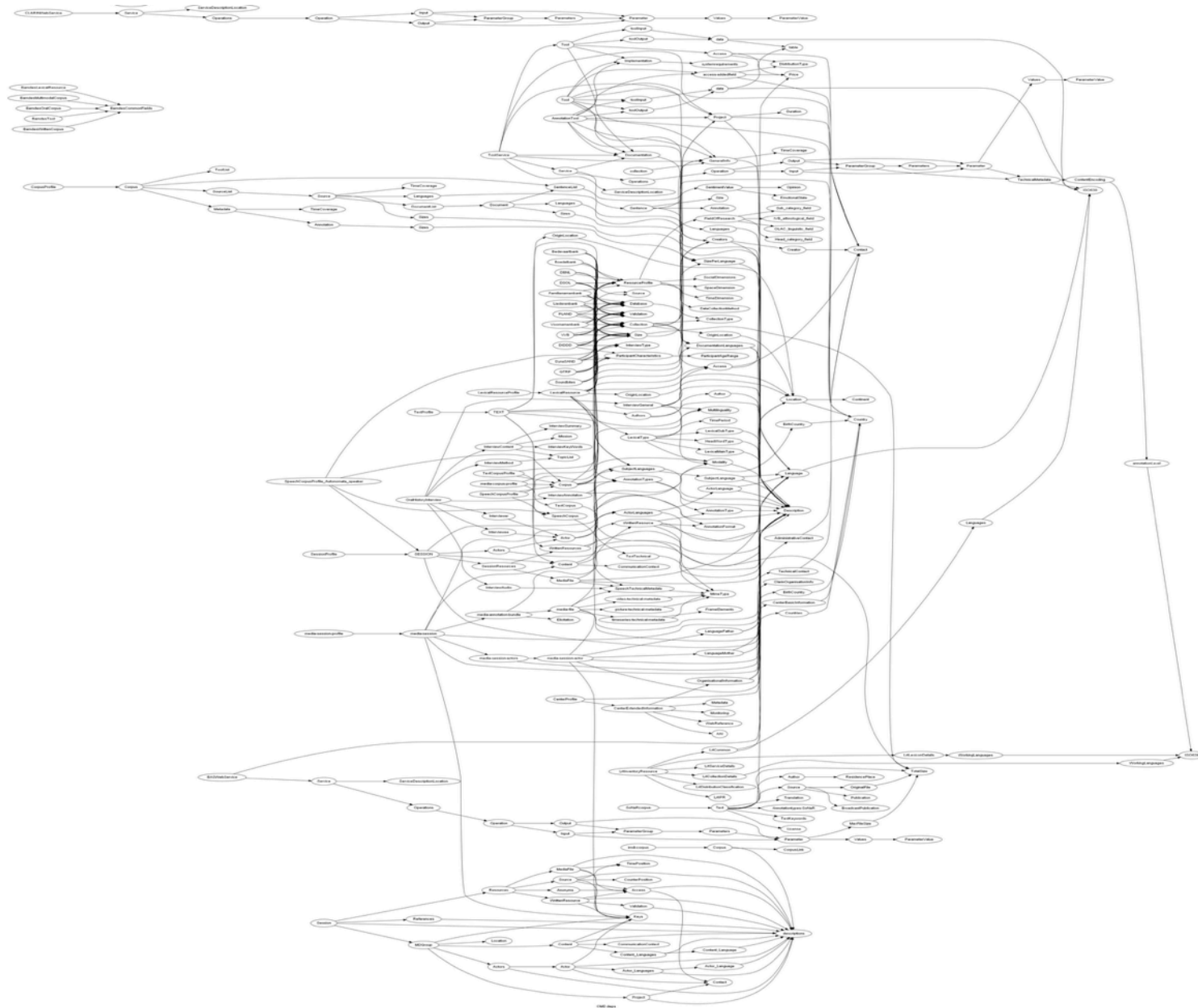> i.e. the profile and component definitions

=> 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

Initial static dot-based approach

SMC Browser - User interface layout

Examples of output

teiHeader (DTA)

node size = count elements in instance data

dublincore - overlap
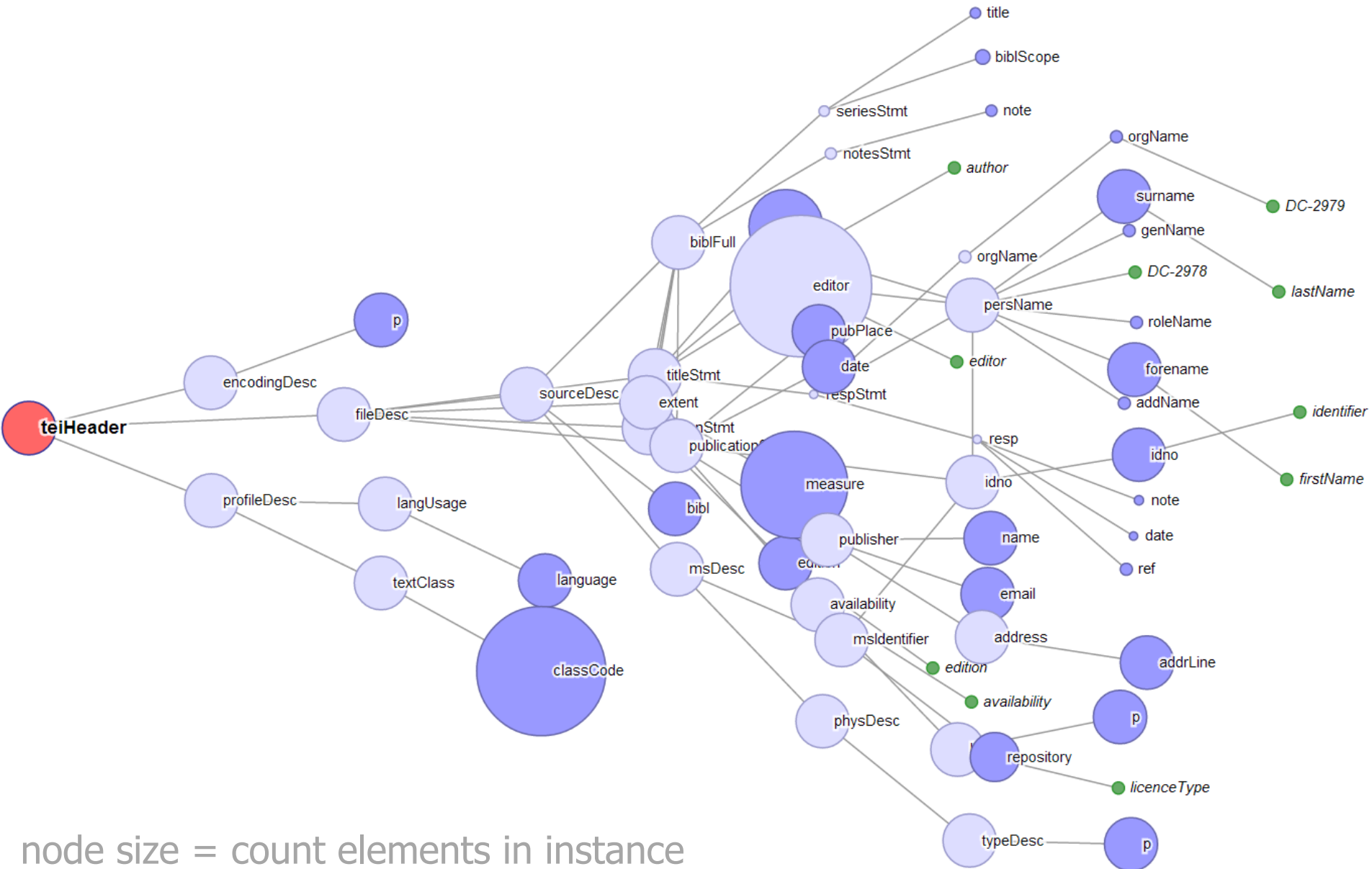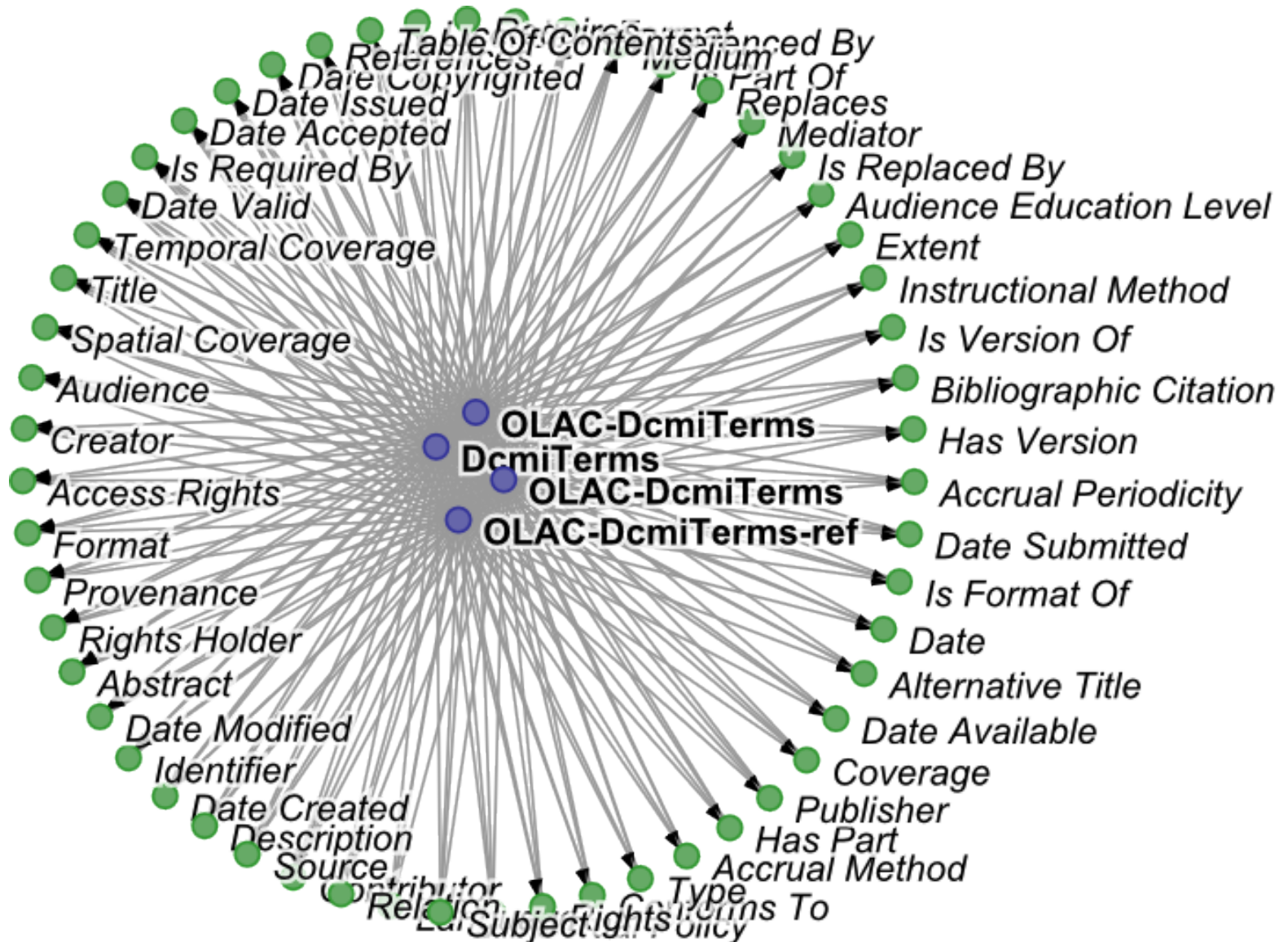
resourceInfo (META-SHARE)

data categories: Language

based on data category reuse ($t_s > 0.6$)
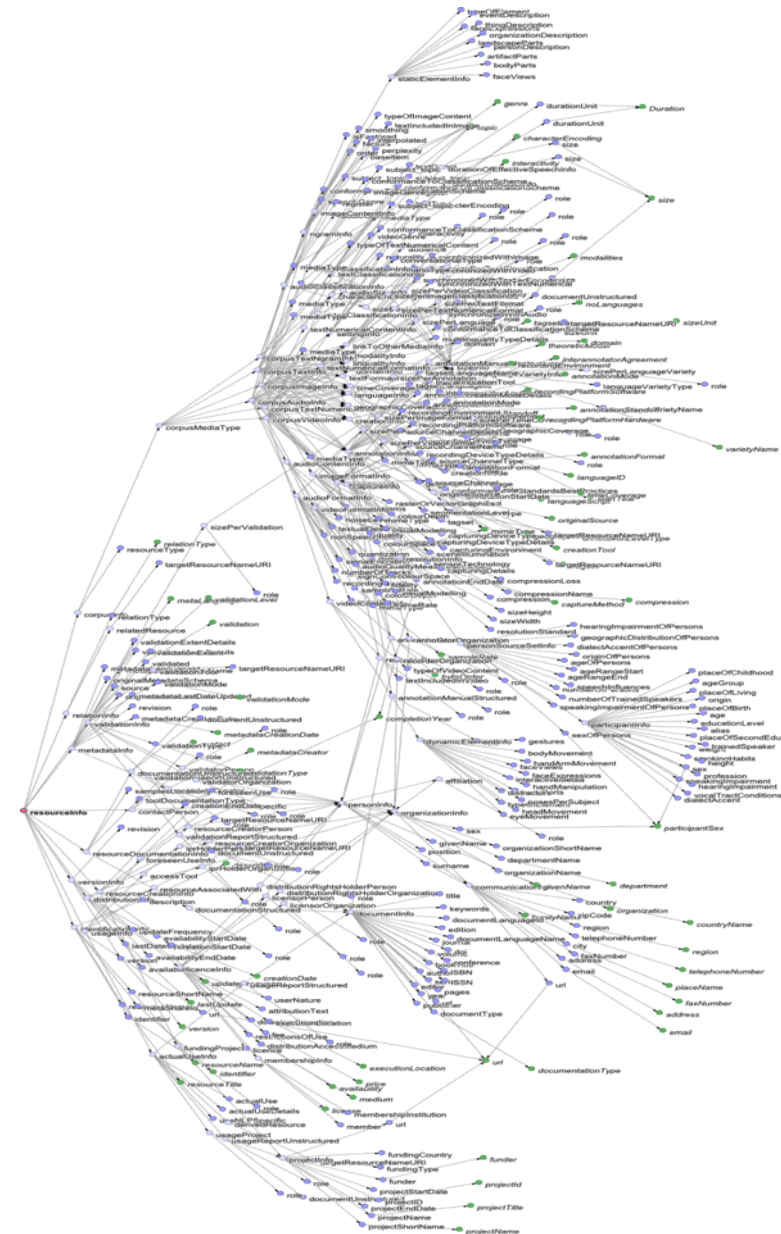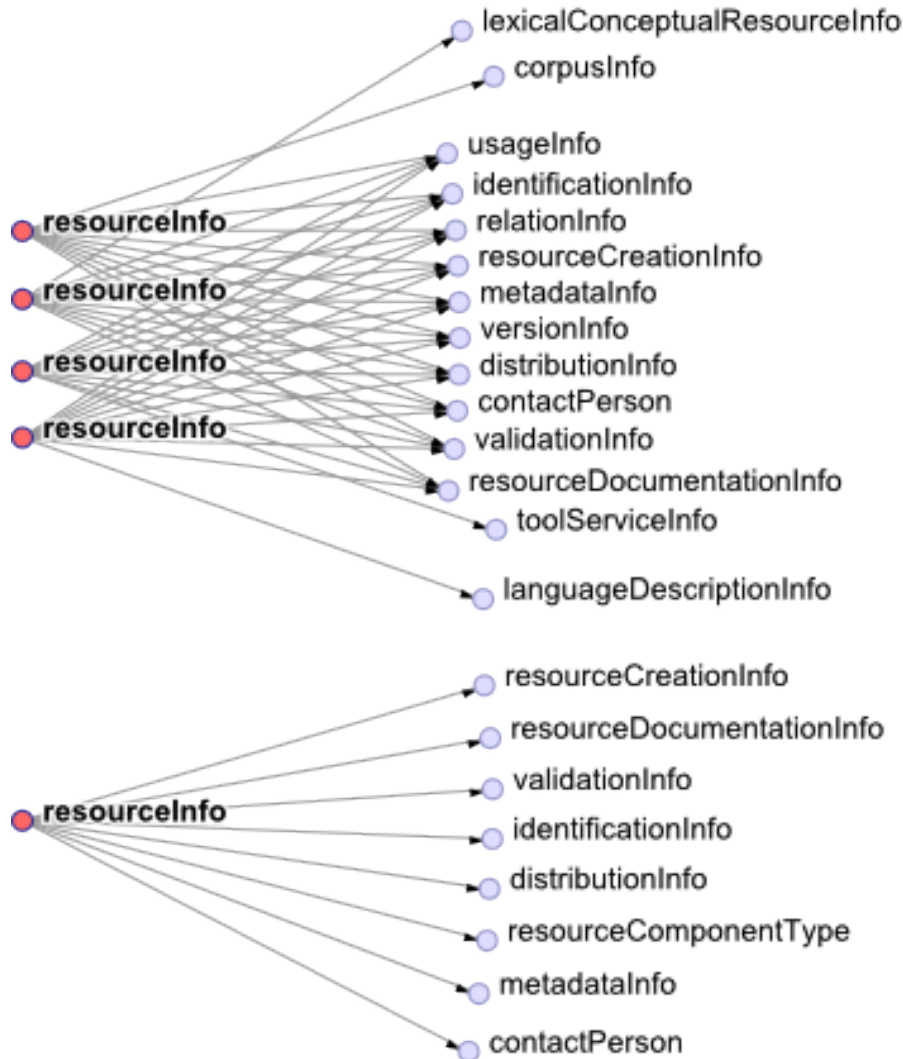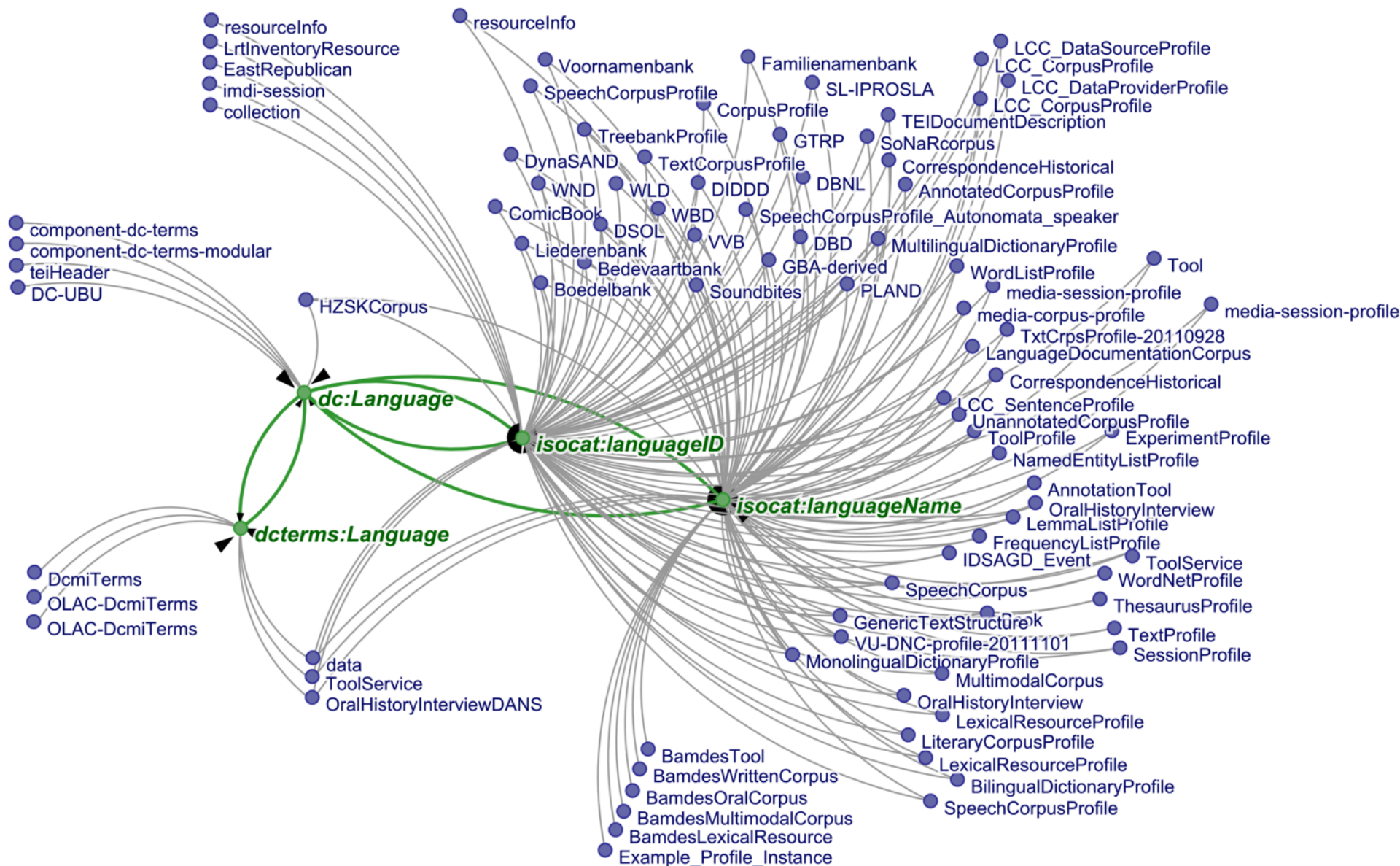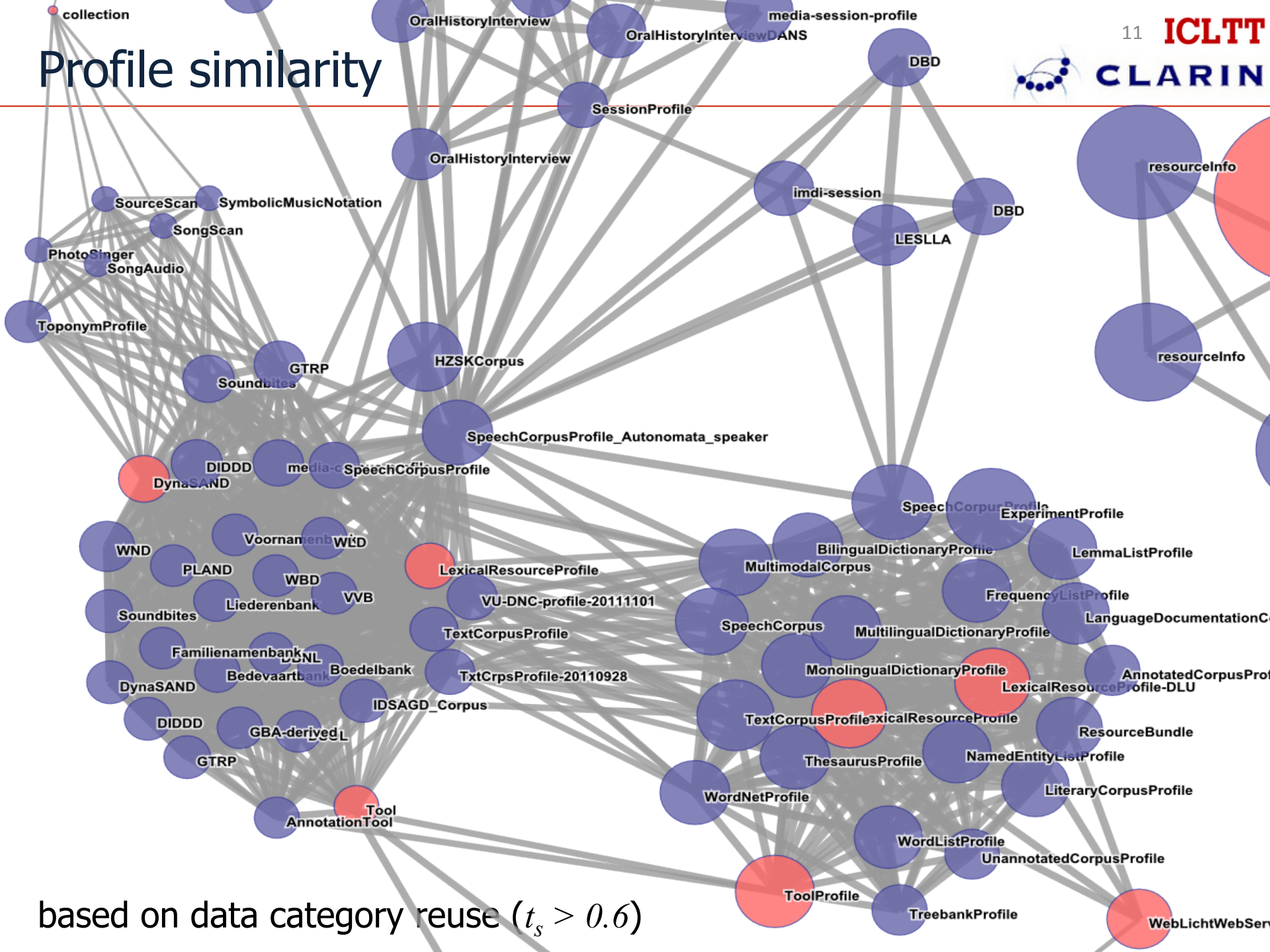
Profile similarity - matrix

Reuse of data categories
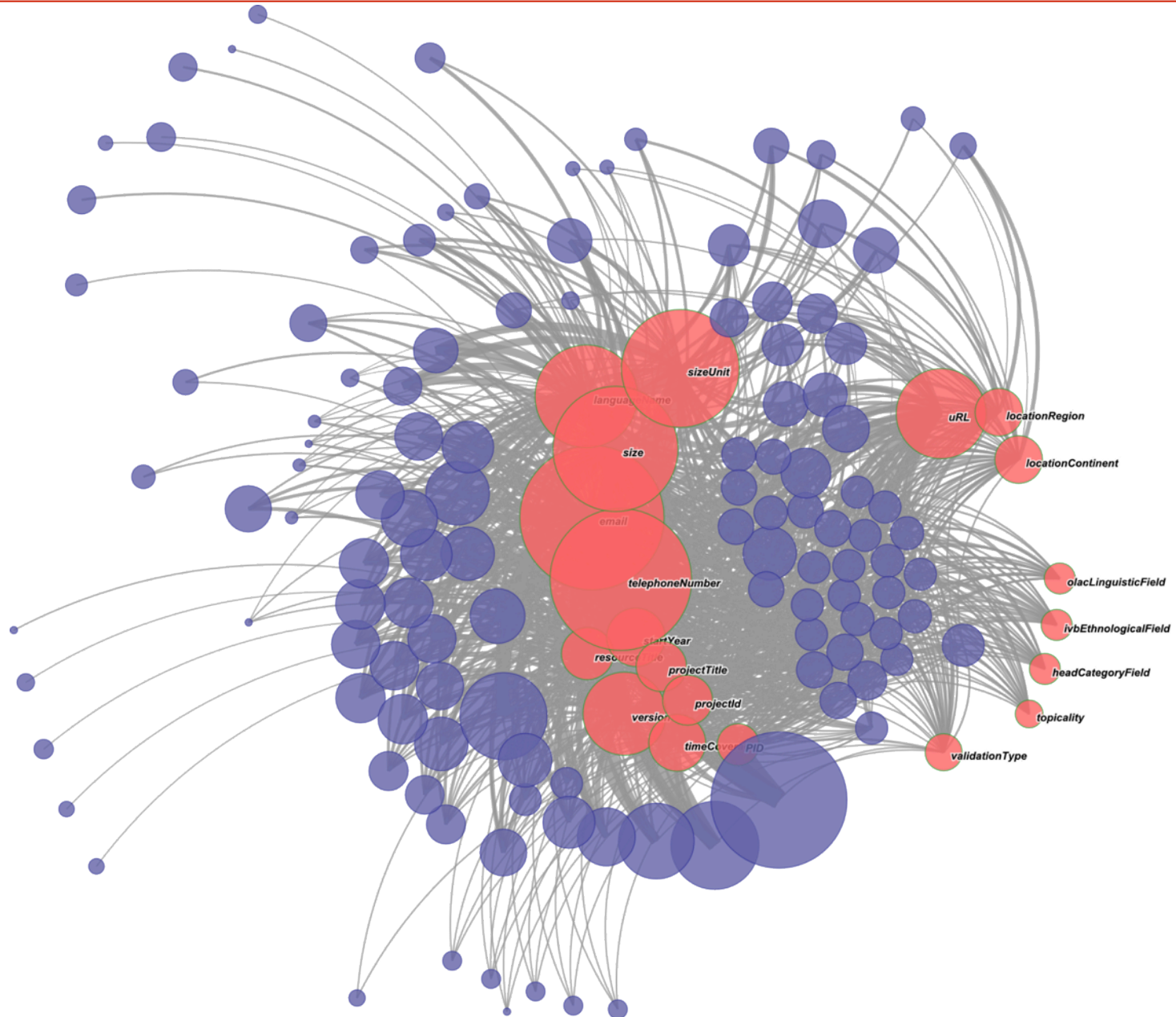
Future

- 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

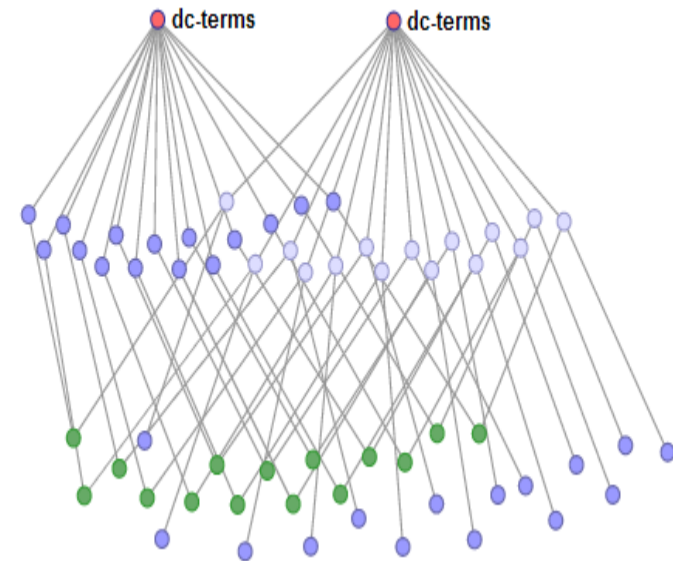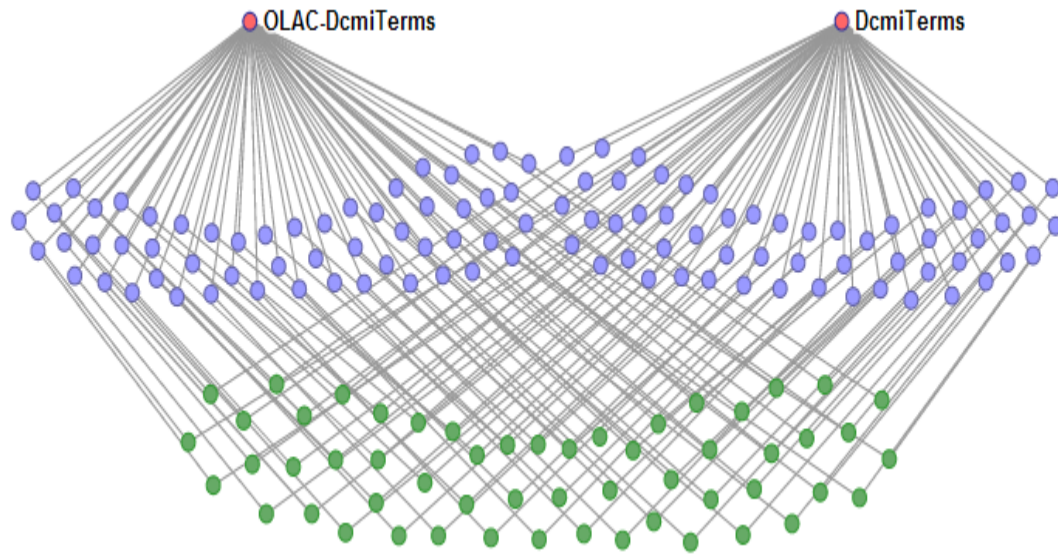extract influence relations from dbpedia, convert to graph, feed to smc-browser
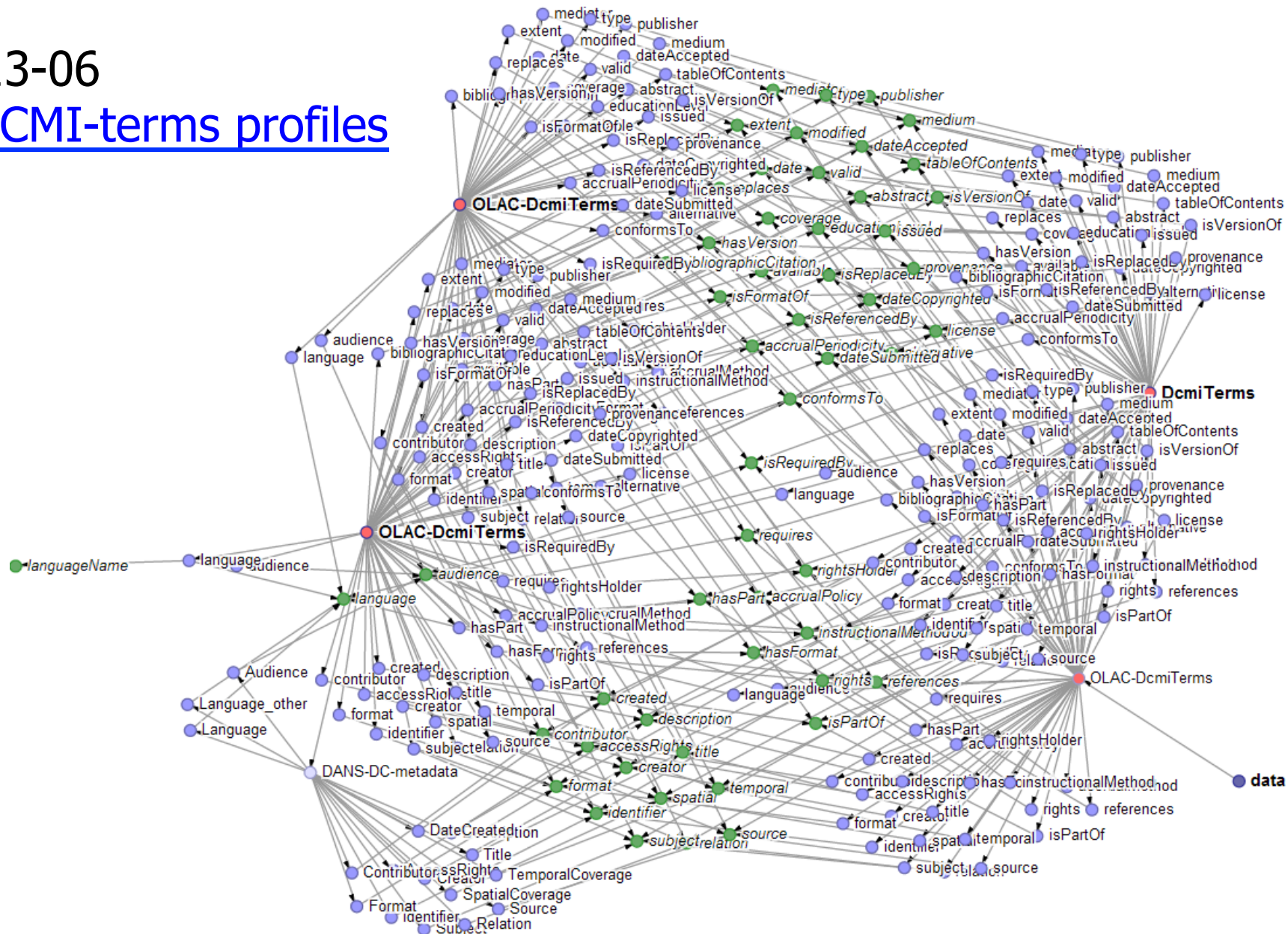
link
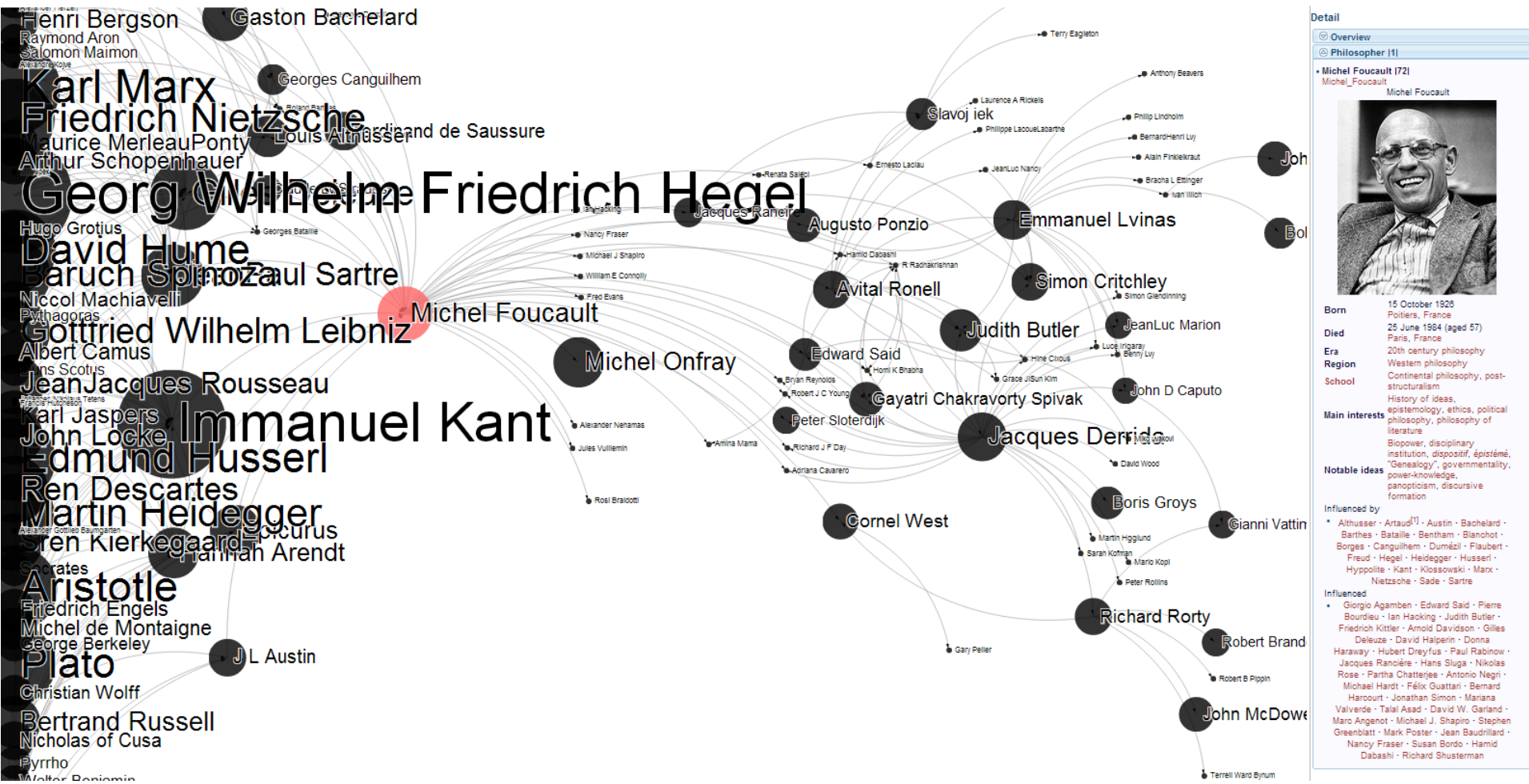
Thank you

More graphs

dublincore I

- 2 profiles with dc-terms (55 data categories)
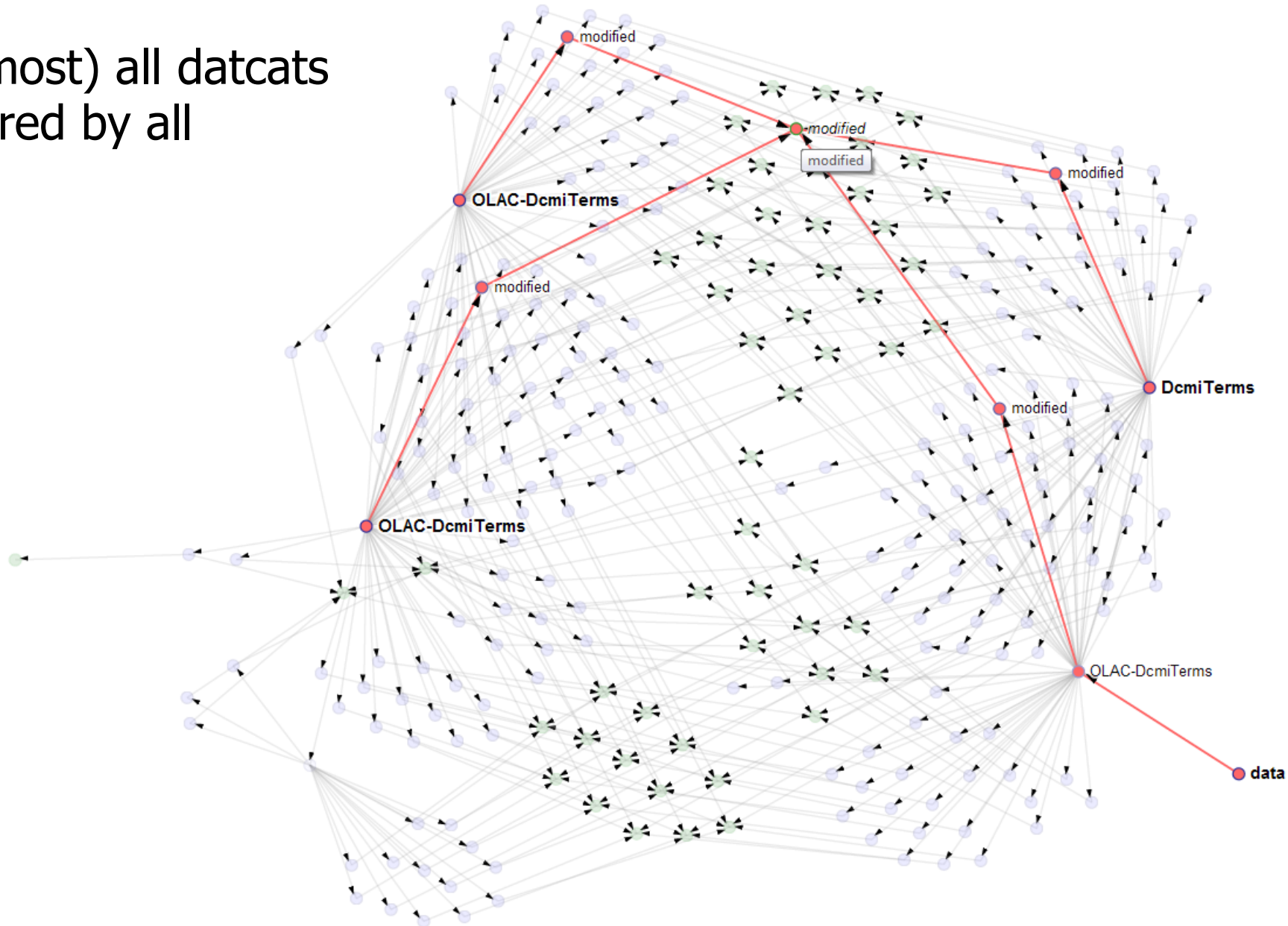- 2 profiles with dc-elements (called „dc-terms")

as of 2013-01

dublincore II

2013-06
4 DCMI-terms profiles

dublincore III

(almost) all datcats shared by all

dublincore IV
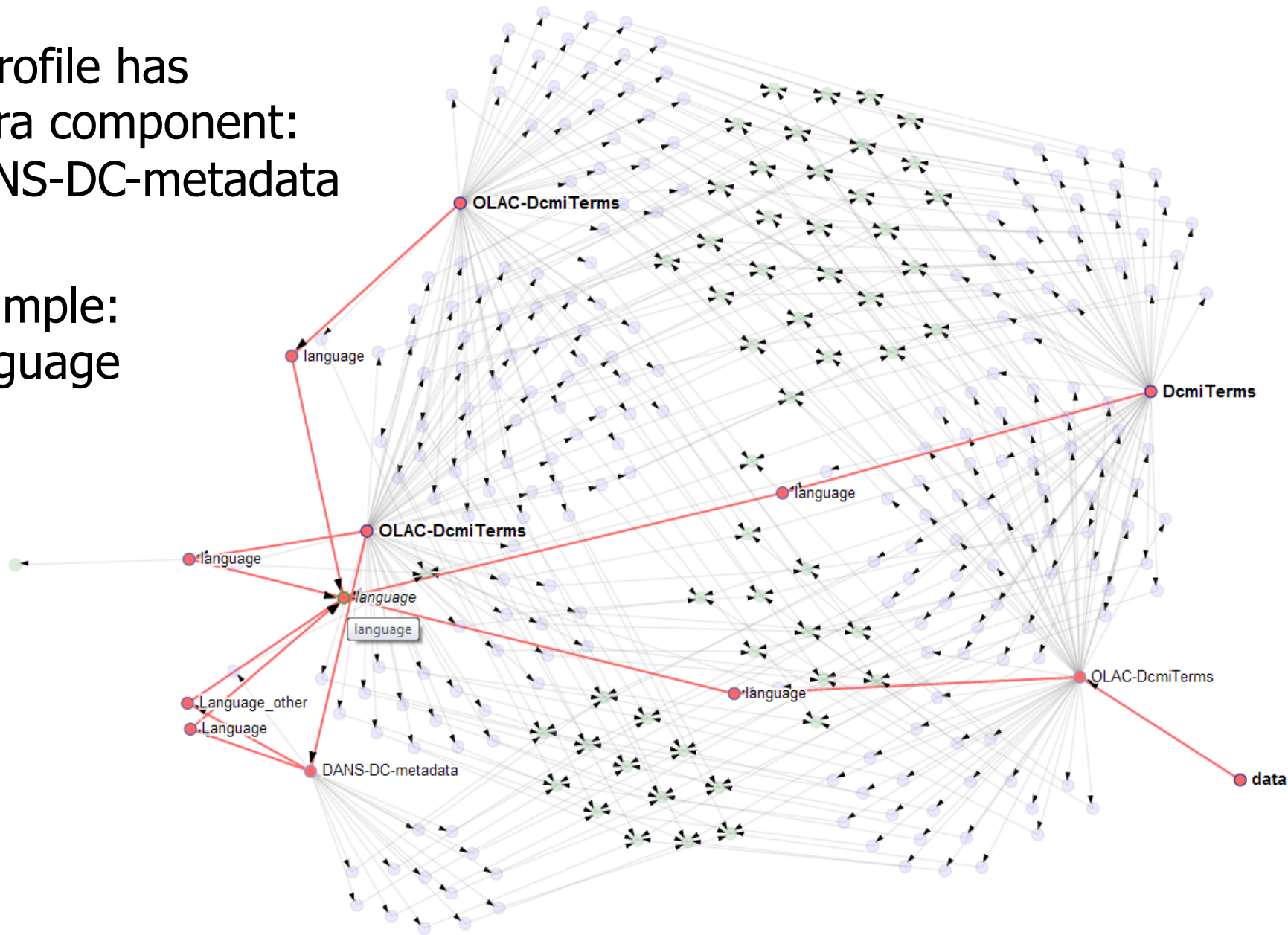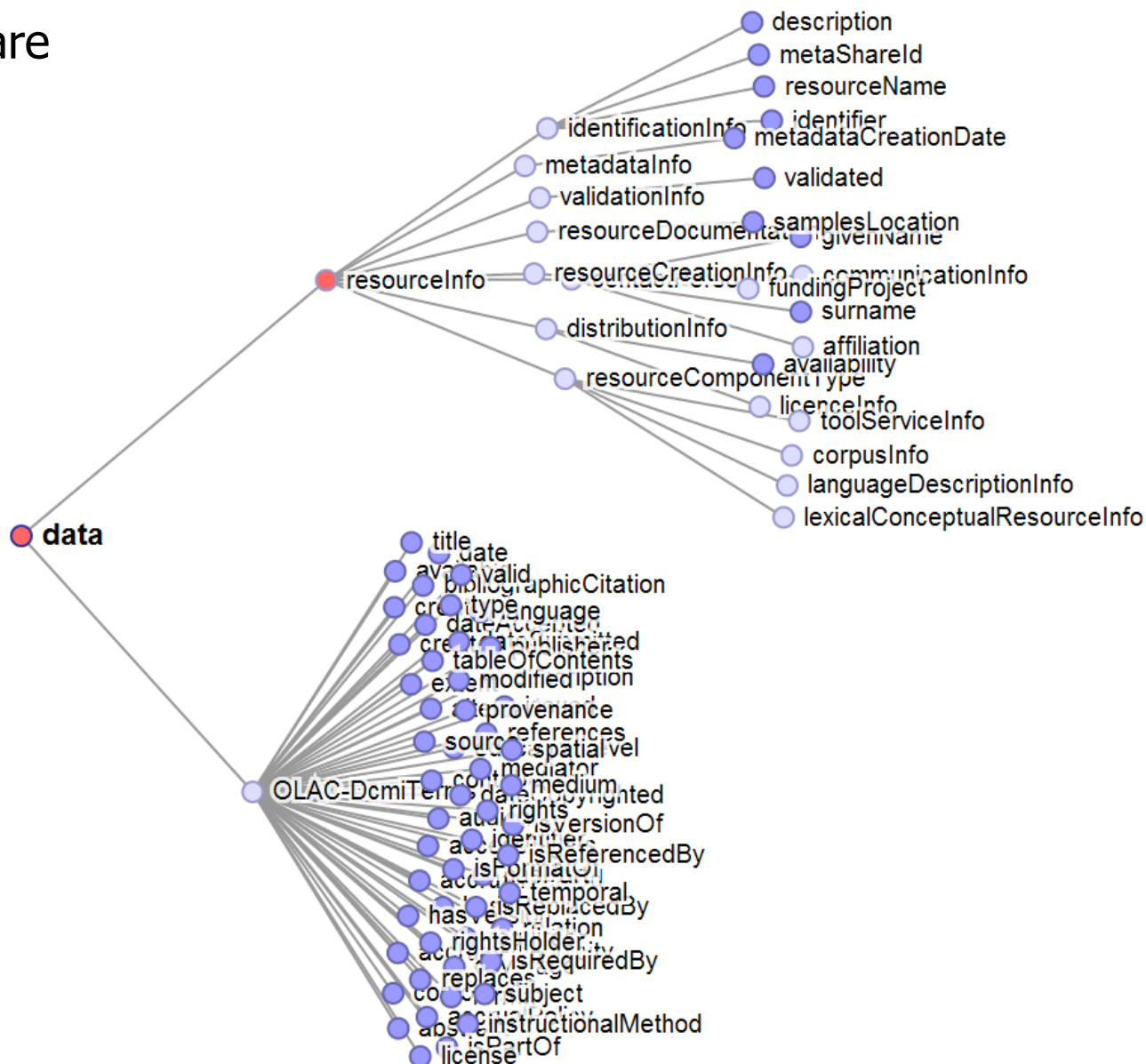
1 profile has
extra component:
DANS-DC-metadata

example:
language

META-SHARE: data profile (LINDAT)

dublincore + metashare
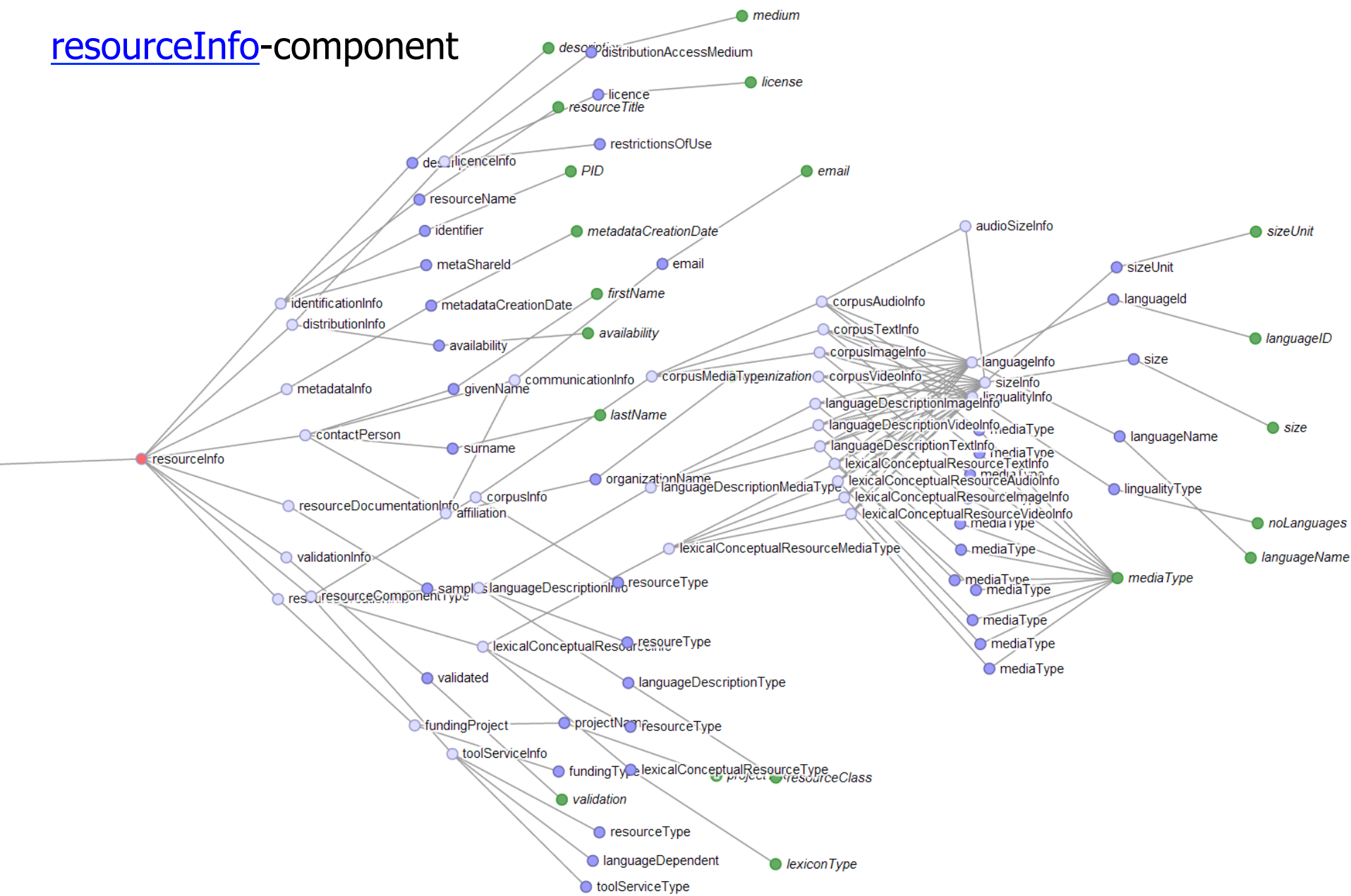
META-SHARE: data profile (LINDAT)

resourceInfo-component

teiHeader (ICLTT)

based on the recommended header

node size = reuse in other profi