

#### **SMC Browser**

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

#### Idea



- growing size and complexity of the CMD<sub>s</sub> 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)

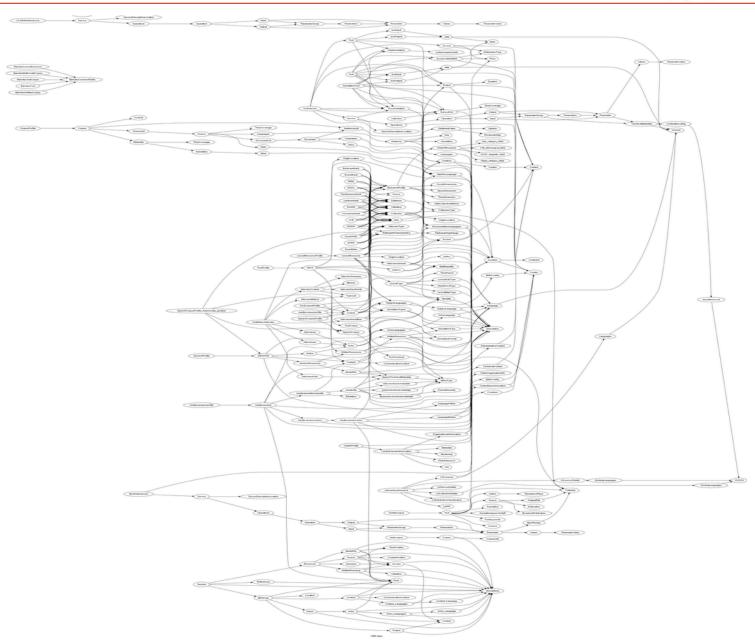
#### **Features**



- web application
  - old <a href="http://clarin.aac.ac.at/smc-browser/">http://clarin.aac.ac.at/smc-browser/</a>
    new <a href="http://clarin.arz.oeaw.ac.at/smc-browser">http://clarin.arz.oeaw.ac.at/smc-browser</a>
    (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 <u>d3</u> data driven documents wonderful javascript/SVG visualization library
- automatic graph layout multiple alternative layouts based on <u>force layout</u>
- reuse and size statistics
- send link, download svg
- usable for any graph data

### Initial static dot-based approach

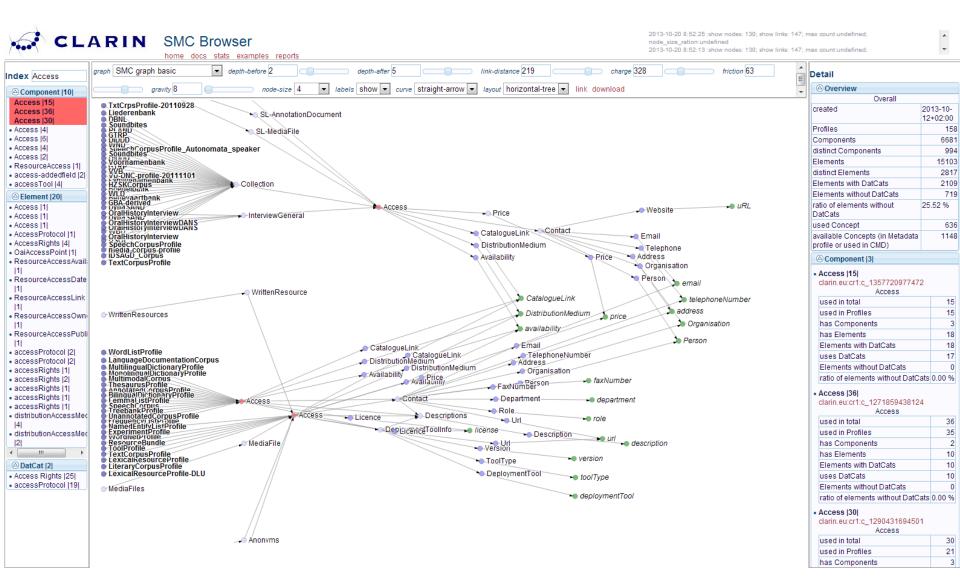




### SMC Browser - User interface layout



#### link

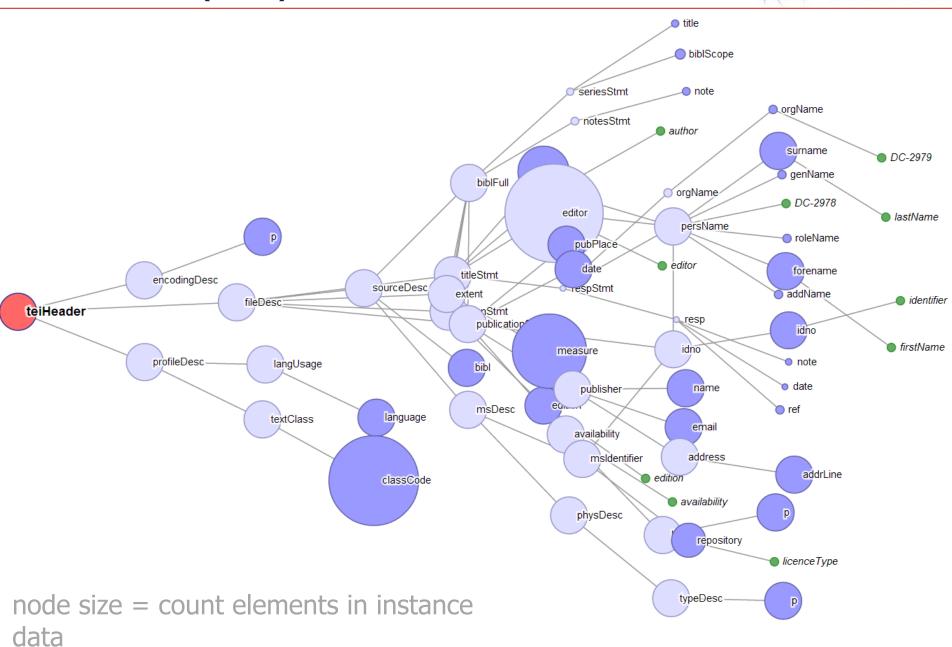


# Examples of output



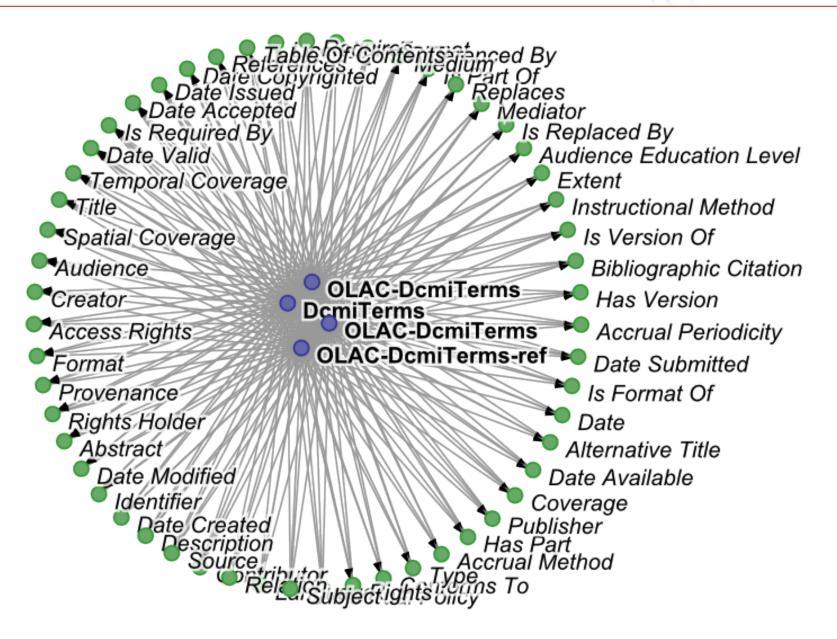
### teiHeader (DTA)





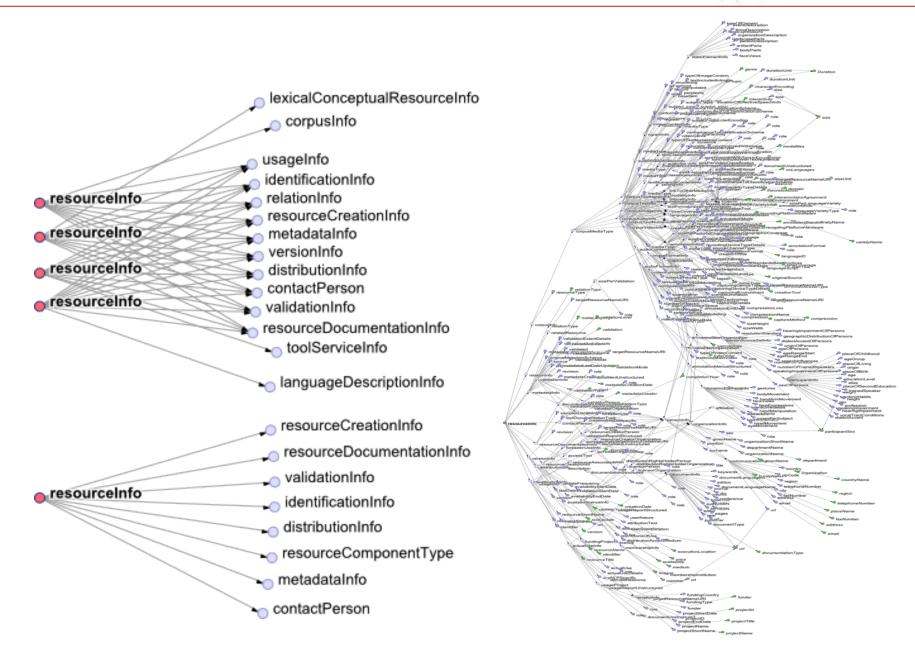
### dublincore - overlap





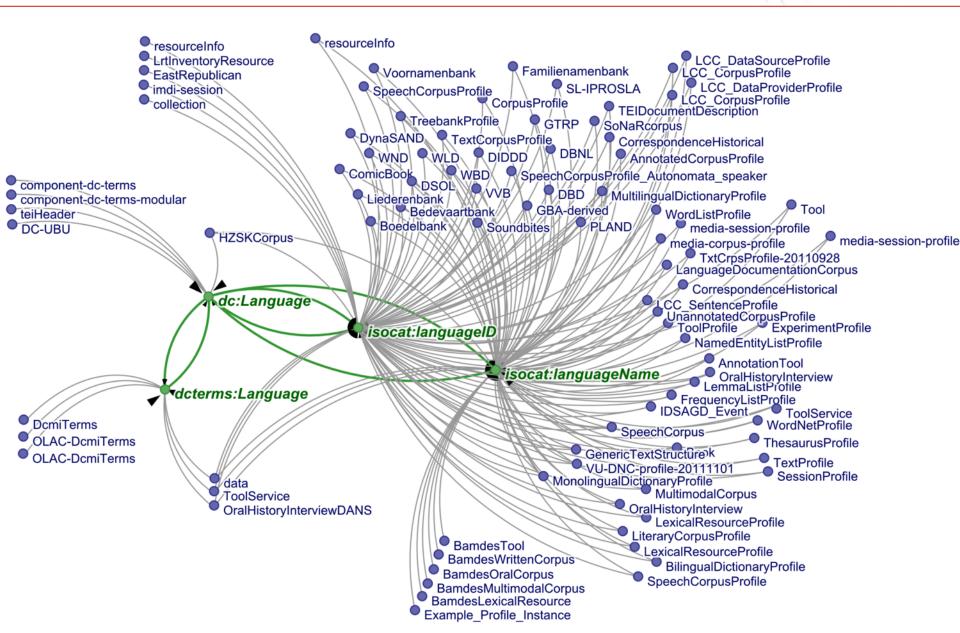
## resourceInfo (META-SHARE)

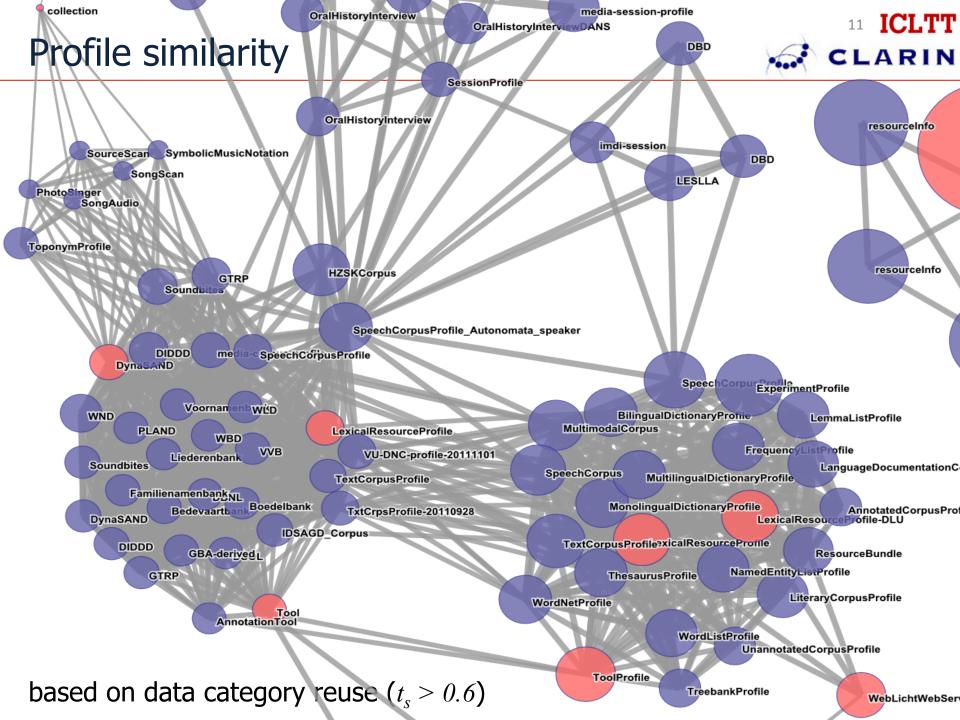




### data categories: Language







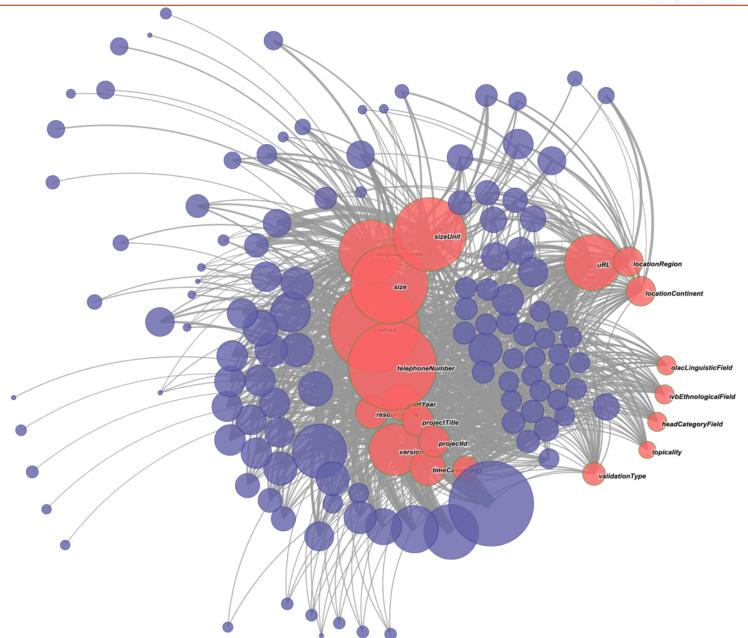
# Profile similarity - matrix



																								1	1					
 Amazini Engan Pr	allia Accordation Plani	le the lee Police	Sanda Jako Para	na Karaja Malimaia Ka	upa Karak Dulka	pa. Ramin Sai Ra	eroka IV: Dar Carpas R	OF STATE OF STREET	Salama Sank Kilopadikhina y	ralla Rambilanta Rasi	Code Profit Circumscame, Par	ngia CI ARRONARia dan		da milaka Cari	-	- delen-	Carpa Profits	Correspondentellida	land Car on paradiomat Natural	rind dids DRS DRS DRS Davillano.	نسزيشر 0000 0000	in deserve Malakida Corp	as DECK Dynamics	المسترشر المستردة	a Papidon PV	Ti Fampio Fralla Jalos	n Papar install Profile	Parillement of P	Napangi kiPrat	to COSA de Invest
												-			-	- maile														
P-MPC			:	:		: :			:	: :	: :	1 1	- :	: :	- 1						: :	:	: :	: :	- :	:	:	: :		
 2-325L 4-325L 4-325L	2-2000 2-2000 2-2000		1-00%		_:	1 1		:		1 1	: :		- :	: :	- 1				:		: :	:	: :	: :	- 1	1	:	: :		: :
	1-626		1-007L 1-007L	0 - 00.00 0 - 00.00	E-03%	- :			:	: :	: :	1 1	- 1	1 1	- 1							:	1 1	: :		:	:	: :		: :
4-318L 3-62L	2-3184 2-4284		1-200 (-200	G-GER.	(-13%	E-03% (-1	are .	:	:	: :	: :	: :		: :		:		:	:	: : : : :	: :	:	: :	: :	- :	:	:	: :		: :
21-43% 41-723%	3-050 3-050		2-20-9L 6-24-7L	1-20-00 1-20-00	1-2000 1-2000	4-3000 3-3 4-3000 3-3	mark 1	-0.2% -0.7%	21-4290.	: :	: :	: :		: :		:			:	: : : : :	: :	:	: :	: :	- :	:	:	: :		: :
21-02% 1-02%	11-12-70 4-12-70		2-20-90 (-7-20-	2-20-00 (-7-20-	2+25/KL (+7/2%	0-3000 0-0 0-000 0-0	Est 1	-042K	0-000 2-000 (-0.0. 2-000	419.0	: :	: :	- 1	: :		:			:	: : : : :		:	: :	: :		:	:	: :		: :
t-max	0.20		1-3676	1-312%	1-3626	1-3176 1-1			1-00/FC 1-3E/FC	2-00% CFG																				
3-72%	2-72%		:	:		: :		0-00.7%	-100 1-700	1-200	(-EFL -	: :		: :		:			:	: : : : :	: :	:	: :	: :		:	:	: :		:
							-											-			: :				-					
7-EDL	1-2020		3-3350	1+31.0%	2-2224	2+310% 2+3			HERE FHERE	1-22%		(-EDL -																		
01-262%	0.00%		2-36/6	2-36/K	2-2000	1-3600 1-3			0-EFL H-EFL	0-EE 60	HERE .	(-20L																		
4-012%	4-026		(-EPL	(+EBL	(-EDC	(+7.2% (+6	ESC. (-	-191.	-an san	4-03%	2-842K -	(1476	4-83	N. 2-0.N.																
										63%				. (-22																
44-38.PL	g-9476		1.45	020	0.00.00	2-9-75 4-2	_		e-HIPL G-SEPL	- 634				- (-22												•				•
		1-18L	1-050		4-32/6	2-16/6L 4-2	3L/K 3			4-44E 44			4-20	_																
1-82%	1-026	-246	(-12)L	(-E2C	(-12)				1-EE 1-EE	1-03K GR				(-E2C 2-(2)			1-0.2%													
1-82L 0-82L	1-026	I-BAL	1-22C 7-22C	(-12L	0-13%	0-12% 0-1 1-10% 1-1			1-020 1-070 11-020 0-0270	1-03t GR		(-14%		0-620, 3-63 20, 3-630, 3-72			1-02L 2-02L		1-02%											
23-21.0L	0.25%		7-8880 2-8770	1-27%	1-mm	2-20% 2-2			0-090 0-090 0-090 0-090	21-626 2- 21-626 2- 21-626 2-		1-686 -		00 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0			0-30%		1-05t	g- gage										
22-2424	0.25%		a-mar.	1-226	2-2-50	1-max 1-1			***** ******	31-244 CA		3-676		0- 0- 0- 0- 0-			0-30-50 0-31-50		1-020	G										
21-14PL	2-22 2-22		rene.	I-max	1-24	4-300 2-3		- CARL	8-200 2-200 8-200 2-200	8-025 (c)		(-126		ME SERVICE CO.			61-21.0% (8-61.0%		1-026	4. 2. 2. 4. 2. 2.										
1-104									1-12%	42.0				mpt.						EN EN EN								. :		
2-42c	3-00		r-mov.	1-200	1-20%	4-313% I-I		- N.DL	1-11-11-11-11-11-11-11-11-11-11-11-11-1	G-MPL GP		(-12L	21.75	DE G-			g-gpt	1-424	1-02%	## ## ## ## ## ## FF ##				. :				. :		
														dir.						430 HALLES										
										. 2-					ar .					· · · <u>· ·</u> ·										
21-42K	31-0070		2-20-90	1-20-00	2-20%	4-3686 2-3		-04EG	8-00F 20-00F	4- and 4-	E-00/00	(-10L -	8-20	mm			G-4186		1-12%	H- H- H- H- GR GR EA GR H- H- H- H- GR HR FR HR	er.				-					
n-an	31-00.0%		2-200%	1-2000	2-2000	4-3636 E-3	200 P	- 042% ·	2-4%	e-mr. cr	t-day.	(-18t. ·	8-00	EE G-10	× .		4-420	1-02%	1-02%	41.20 21.00 21.20 20.00 41. 21. 21. 21. 41.	en.		METE.							
1-20-%	2-30/6		1-36/K	2-20/6	2-36/6	1-300 1-3	3L/E (-	-72%	- CARL S-MAN	4-42FL (2.0	1-7.04	(-20L	3-4	marana area	n. 7-42	1-430	1-686	(-EFL	(-E2L	MEN NOW NOW HERE THERE	er.		er cer							
1-36/K	2-3686		(-an	(- C/L	(- G/L	(-0.76 (-)			1-12FL 1-12FL	3-626 6-		2-1526		ALCORDO 1- G				2-2000	2-2070	2 - 1 - 2 - 2 - 2 - 2 - 2 - 2 - 2 - 2 -	62K		2- 2-43C		-036 -	-				
m-max	31-42-70		7-E/K	1-480	1-026	4-3026 1-4			8-300L B-703L	21-20 PL (EP)		3-646 -		Sec. 8-40		•		4-0/6	4-45/6	G- 21- 21- 21- GR ET- 31- 21- 3-27-	rac.		ne nere		-2006	0-20-FL				
H-43%	31-0276		2-20-00	1-22-01	2-22/60	4-3000 2-3		-04BE -	21-42FL	4-mm		(-10L -		- E-C			G-612C		1-03%	CALCELER CO.	M.P.	•	GENERAL GENERAL		- carc	3+4EPL	3-25			
\$1-\$1.0K	31-2274		8-3474	2-3120	1-3124	4-3200. 2-3			2-42C F-22C	22-01PL (2.7)		2-60°C -		ar ar			O-MAK		2-E/K	CO. H. H. H. 1-130	aer .		SEPT. MINISTRAL		-max.	1-115%		22-00.0%		
n-an	2-070		1-EAL	1-20-00	1-2000	4-3686 2-3	E-60 1	- HERE	8-00E 2-0E	d-mar (-	1-00FC	(-10L -	1-11	ME 4-01	DE -		G-HIPL	1-020	1-03%	er er er er	MPL.		COLUMN CO		- CERC	1-1EPL	31-20.PL	4-mx 3	20-0020	t :
	:		:	:	-	: :			:	: :	: :	: :	- :	: :	- :				:	: : : : :	: :	:	: :	: :	- :	:	:	: :		:
	:		:	:	-	: :			:	: :	: :	: :	- :	: :	- :				:	: : : : :	(+7EBL	•	: :	0.7250	- :	:	:	: :		
H-EAL	0-036		1-22%	1-1120	1-22%	1-1100 1-1	11 PL 1	- 974	H-02E H-2EE	01-02E 2-	1-200	3-00%	3-4	or are	PL .		01-07E	1-026	2-026	E- H- H- H- HPL HPL HTL GPL	01- 41.0%	(- GAL	CLEE CHARLE	i i	ne.	1-31FL	41-22.PL	H-GPL I	H-227L	H-GPL
										. (-				. (-12						ERREN GR	+LBC	(+215%	ar.							
n-an	3-02		2-2000	1-2000	1-20%	4-3170 2-3	mou s	-0.2%	8-MPL 21-42-%	4-MPL (-		(-190.	1-2	DE G. 6-10			G-G26	1-020	1-420	61- 22- 24- 48- 42- 28- 27- 28-	<b>E</b> .		er-		-020	2-162K	N-TAL	4-82 S	2-an	4-MPL -
																					· PER			PER						
n-arc	21-670	1-28L	7-282%	1-626	1-620	E-0020 E-0			B-25.00 P-00.00	2-24 CA		1-626 -		Br 20-00		ne re-environ			2-626	CO. EX. 60. E4. 2-10.	20-7K		200 21-20 PC		1-43%	e-marc			r-arc	B-86.
H-10/K	3-0%		1-EAL	0-0020	E-3626	4-3EPL 6-3		- HERE	B-819L 31-819L	21-22-PL (2-7)		(-10L -		or state 4- co			G-HIPL		1-03%	G- 2- 2- 2- 20 GR GR GR	N.St.		HATEL HATEL		- CERC -	2-1EDL	21-45-FL		H-2270.	m-met.
1-7/60	4-020		1-656	(-E2C	1-626	2-06/66 0-6			1-02E 2-02E	7-026 03		(+426 -		N. 3-836 8-30		•		7-2200	7-2201	ESC MAN GEST CON	Car.		Care 7-Care		-686	1-4.%			-cas	7-026
1-0%	4-620		(-E2L	(-E2L	(-13).	(-200. (-2			-mm 1-mm	7-2000 GE		(-E/K		PL 2-03PL2-00		•		4-43L	0-020	COUNTY MOUNTS	EPL		non remove to		-626.	3-315L	2-626		1-0/6	7-2000
1-00%	1-00%		(+3C%	(+3C%)	0-2076	(+213E (+2			1-mm 1-mm		2-012%	(+1E4E		PC 2+20-9C (+40-			(+ (13)L				mou.		ne i-me a-		-cac	(-21FL	3+013%		1-22.00	1-2250
C-MAL	04-27.0%	+38L	2-30%	1-313%	2-303% -	2-2020 2-3	20.	- H2E	8-3476 SH2746	G-3456 G-0	7-ziac	3-73%	3-30	201-2070 1-10			E-06/9L	1-12%	0.000	2-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0	no.	:	SERE G-MAN	ļ .	-cas	2-20.0%	20-22.0%	G-3484 3	H-2676	0.00%
						- 1								M+	-					9- 3- F- W-			20.					. :		
e-mac	3-22		I-may	1-22%	1-100	0-313% I-3		-an	n-arc n-arc	22 - 62.00 GL/G		3-7.0%		DU NO. 0-40			G-TARL		1-120	G. H. P. H. GRINGER GR	an.		an near		-22%	1-31%	ED-TERM.	20-426	N-1174	20-4EK
7-3070	7-3686		3-03%	1-215%	1-2250	(-13% 1-3			0-P3% 1-3E%	0-2000 GEN		(-E/K		e-			O-FLEE		1-036	5- 7- 5- 6- 6-22-6-22-6-22-6-	an.		rate doubt		-0.0%	1-1EBL	C-EER		1-000%	0.236
C-26/C	4-44	-13L	2-313%	2-2136	2-2126	4-2724 2-2			0-PAL 0-BEL	G-27-01 GR		2-E/K		ar e-a			2-505		1-03K	7. C. C. C. C. D. D. D. P.C.	CSL.		e- e-en e-en		-an	1-313%	0-23%		0-2270	6-27/6
H-22/C	44-22.0%	1-42%	2-3000 2-3000	2-34.04	2-30.00	4-300C 2-3			0-80% 0-80%	G-35/K 274		1-E4L -		31.1-4A.4-4			4-22%		1-020	7- G- H- G- G- H- H- G- 7- G- H- G-	GPL		G. HGER		-0.00	2-27.5%	4-236		0-22%	0.36%
G-HZL	01-227C	1-62% 1-62%	1-mm	2-200s	1-220	OHE 10		-30%		G-20.00 G.N		1-00%		and other son			01-212E		1-626	7- G- H- G- GR GR BY BY: 7- G- H- H-	20.00L		a. a. a.m.		-1125	1-00% 1-00%	01-012FL 01-02/FL		0-47L 0-47L	G-28/K
G-MPL	1-115t	- LEC	1-MPL 1-MPL	1-200	1-200	0-20.00 E-2			0-45t	21-4274 41A		3-72%		ne meret e-en			G-SERL		2-826	7- Q- H- H- MPL BPL BPL QPL H- B- B- B-	euru M-		are wear		-1176	1-12FL	0-00 0-00		0-0/L	20-42FL
D-20%	0-22A		2-363C	1-220	1-200	1-may 1-1		-cas	3-45 F-85			1-686 ·					0-31FL		1-E/C	6. B. B. B. 2. 2.20	St.				- 1120	1-110L 1-42PL	21-2224		2-FSL	31-45K
21-21.0% 21-21.0%	0-EK		1-M2C	1-317L	1-100	0-2020, 0-2		-0.00	9-08 B-08	21-025 OA		3-586		SE SE SES			4-67L		1-076	G. E. E. E. GR. HN. ER GR. G. E. E. E. DR. GR. GR. TR.	mo. m		STATE STATE		-0191	1-20/6	31-225L 31-625L	_	D-20%	21-72-24
21-3030. 21-7030.	3-25		7-200.	1-4.%	1-4%	4-365C E-3			9-876 B-876	2-45 CF		1-686		20 May 0-42			4-2% 		2-126	State of the State	N.D.		No. No. of Str.		-max	1-3000	51-455C		r-max	2000
21-7626	3-45		7-2020	1-60%	1-62%	0-2076 E-0			2-0A E-EA	21-00 CB		1-686		N 80 1-0			G-TAL		2-72%	0 - 2 - 2 - 2 - 2 - 2 - 2 - 2 - 2 - 2 -	n.	(-02L	S. S.CK		na.	1-22%	10-10-10		B-850	2-05
21-42K	2-0X		1-E-6	1-20-0L	1-200	4-318L 2-3		- HERE		G-CEN GA		(-200		Mark con			G-0125		1-026	4. 2. 3. 4. 4. 2. 3. 4.	er.		GAL G-REA		-626	3-1626	21-22 PL		2-626	4-mx
41-9186	2-45		0.000	1-0076	1-626	1.30% 1.1			N-676 G-6676	N-676 H-		2-62%		No. o.z		(-10)	0.076		2:105	0. B. B. S.	M.P.	1-04PL	N. N. MAR.		-26.00	1-1176	F-674	_	a-car.	31-676
6-200L	1-225		3-42%	2-020	1-026	0-82% 2-0			-22 1-22	7-32% (43)		0-276		012FL 310FL 310FL			1-80%		3-026	to to to to the rectangle	•.		1- 200 1-200		-313%	1-0.00	1-22		1-274	7-33%
21-2124	2-24		0-34FL	7-68	7-68	2-2120 7-4			B-7474 N-2274	22-7434 (c)		3-726		Mark con			e-ma		1-04	COLUMN TERM	TAL EAL		21.00 21.00 PM		-076	1-mg.	B-dr.	_	N-2474	21-7174
0-026	1-020	+12L	4-2000	1-27.0%	1-726	1-026 2-2			0-45t C-25t	G-626 GR		3-62%		00. 4 - 01.00. 7 - 02.			2-112K		2-626	7- 40- 60- 40- 7- 40- 60- 40- 7- 40- 60- 40-	G-		G. G.Car.		· care	1-42K			c-an	0-68
0-026	1-086	-12L	4-312L	2-2720	1-224	1-424 1-2	T31 0	-036	0-48 C-88	0-02t (-		1-424 -	2-22	W-4-020-7-02			E-112%		2-424	7- G- H- G-	g		G. G.CNF		-076	1-42FL	2-25	0-026 0	C-02L	0.424

## Reuse of data categories





#### **Future**



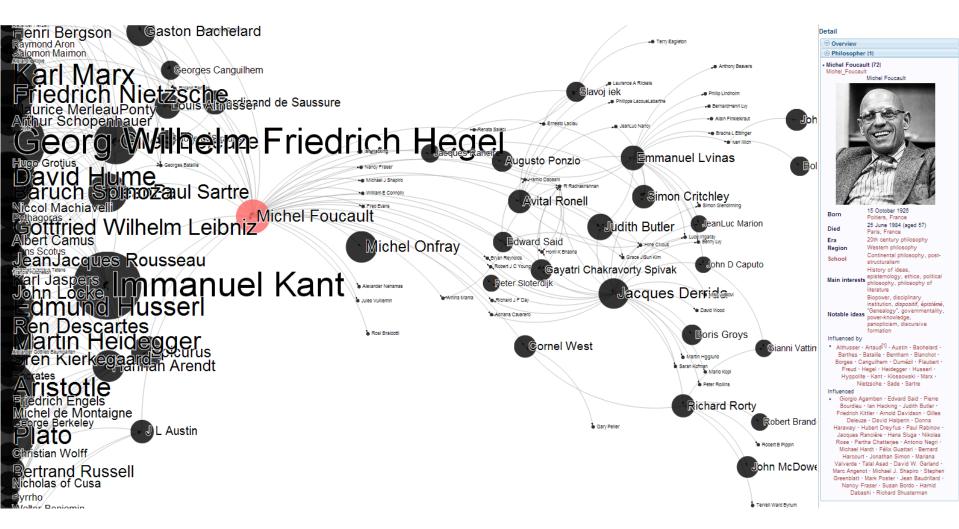
- set operations on subgraphs
  - => differential views (e.g. show just shared components)
- integrate instance data (CMD records)
- integrate with Component Registry
  - on the fly visualization of developed profiles
  - reuse/size statistics as input for modeller's reuse decision
- open for external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

### Philosopher influence



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





# Thank you



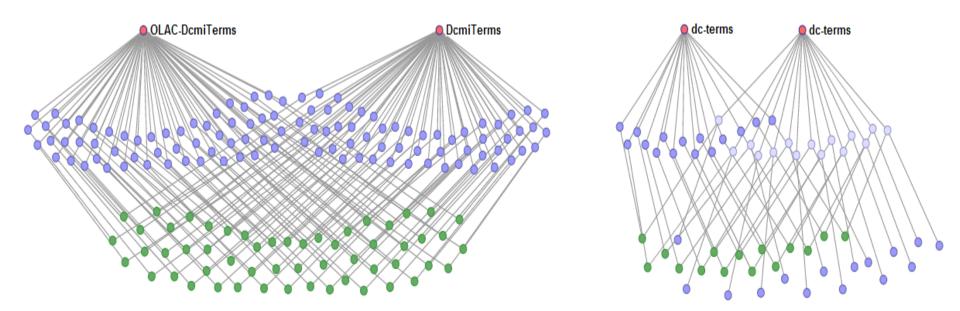
# More graphs



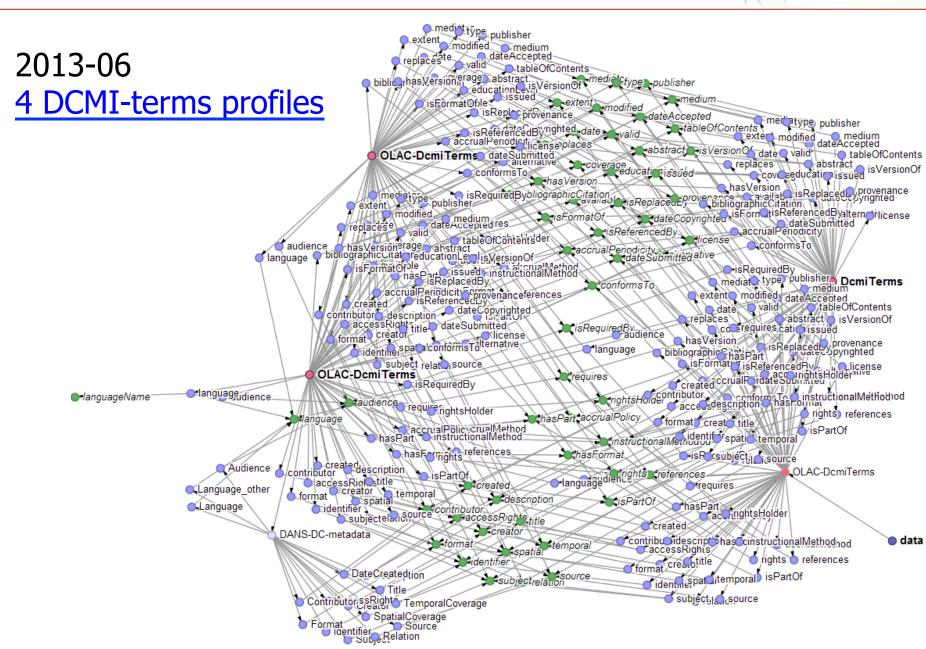
#### dublincore I



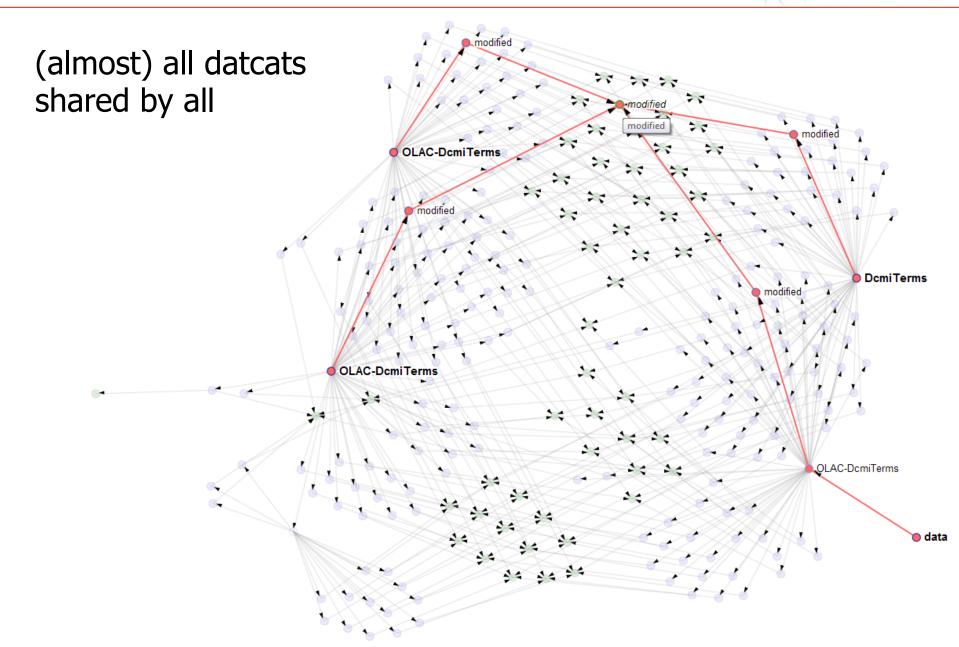
- 2 profiles with dc-terms (55 data categories)
- 2 profiles with dc-elements (called "dc-terms") as of 2013-01



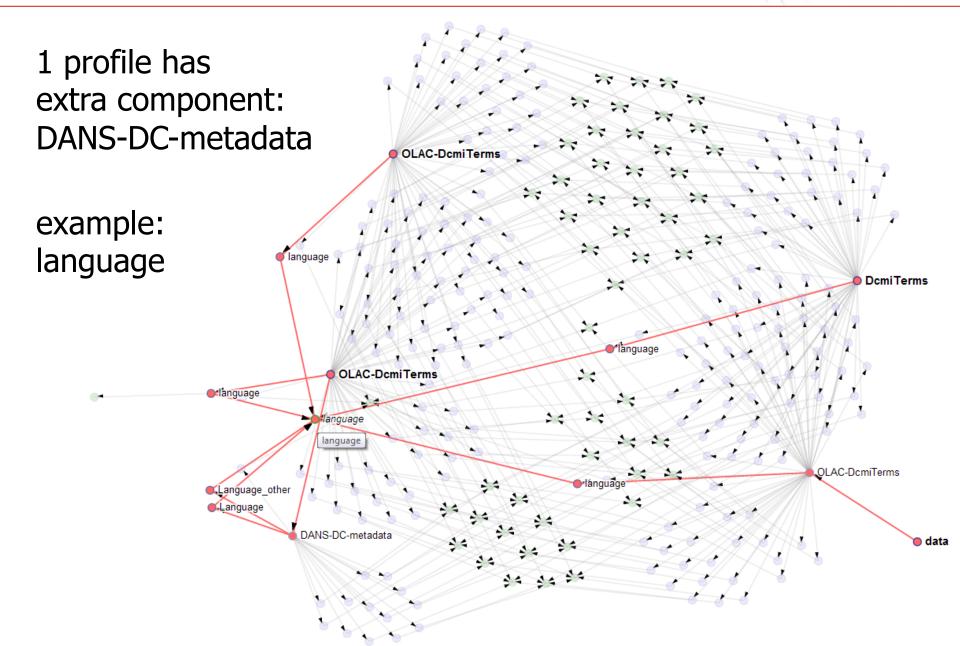






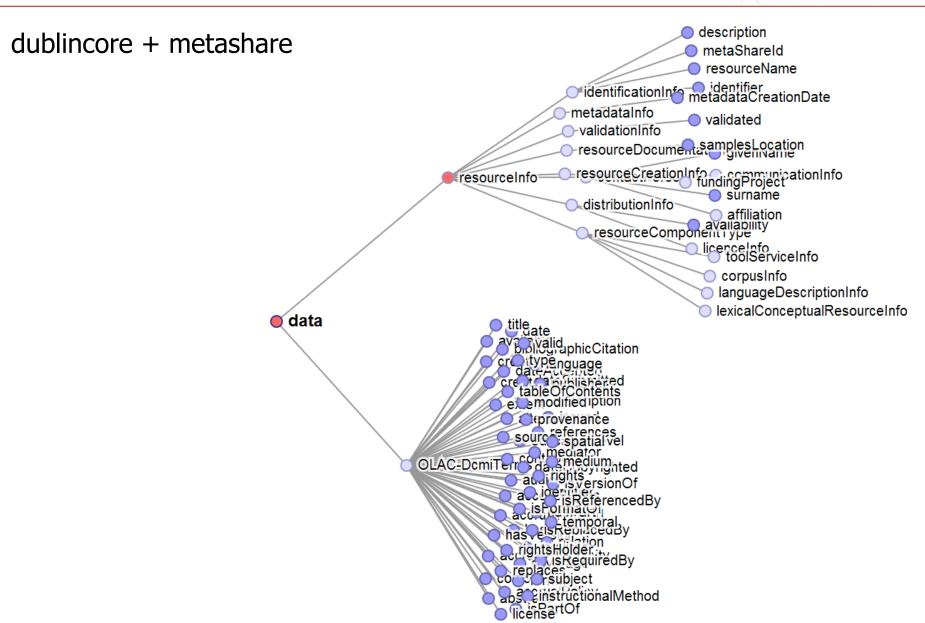




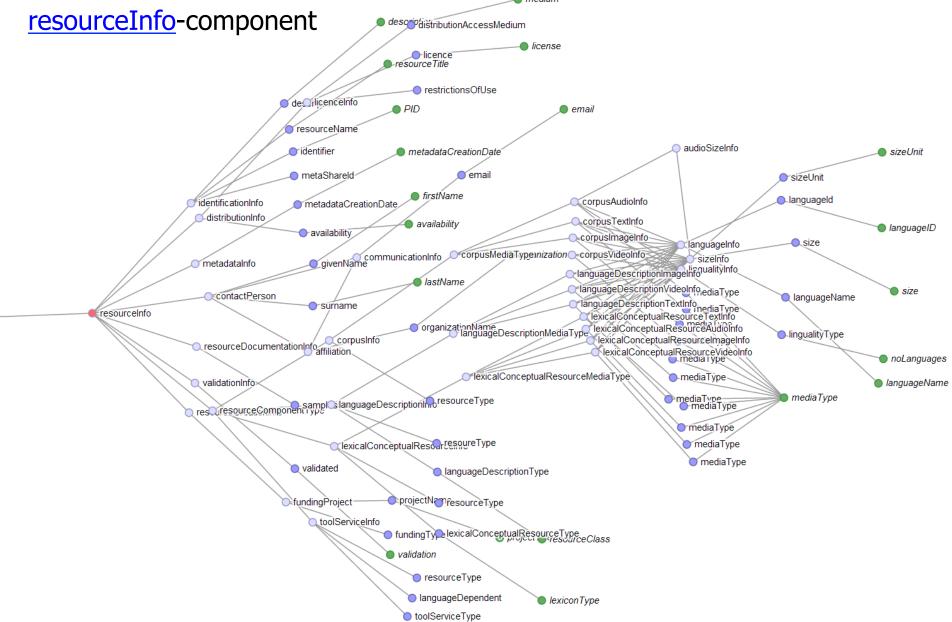


### META-SHARE: data profile (LINDAT)









### teiHeader (ICLTT)



