

2016-04-28

Curation module

<https://clarin.oeaw.ac.at/curate/#!/Collections>

<https://clarin.oeaw.ac.at/curate/#!/Public%20Profiles>

Menzo: Huygens repository - 60.000 are not valid - check why the report does not show invalid files in collection

Davor: performance problem with too big files form CMD-records (> 20 MB) from Leipzig
Process killed after 30 minutes

Move code-base to clarin-github

(contact Menzo after transfer so the right developers get write rights)

Example of profile with 0 concept links but ~65% facet coverage:

https://catalog.clarin.eu/ds/ComponentRegistry#!/?itemId=clarin.eu%3Acr1%3Ap_1337778924932®istrySpace=public

(due to xpath in facet mapping -> potential match assumed)

Chromosome Example: clarin.eu:cr1:p_1337778924932


<https://lux17.mpi.nl/isocat/clarin/vlo/mapping/index.html>

Menzo: I'll extend it with an analysis of the fallback XPath

Update SMC-data

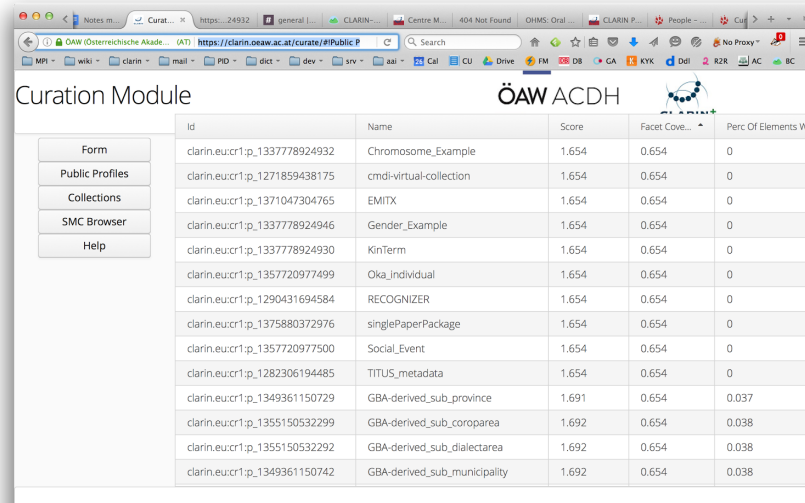
Feature requests:


- Allow to search in table (substring match, filter)
- Assesment -> assessment
- Asses -> assess
- Layout issues in Chrome on Mac (Logo + menu/title whitespace)

Curation Module ÖAW ACDH 

	Name	Avg Score	Num Of
Form	HZSK_Repository	12.093	2,547
Public Profiles	BAS_Repository	12.088	24,002
Collections	Unnamed_provider_at_dspace_library_uu_nl	11.898	64,645
SMC Browser	CLARIN_DK_UCPH_Repository	11.746	87,551
Help	CLARIN_PL_digital_repository	11.538	140
	IDS_Repository	11.447	22,831
	Clarin_D_Repository_Tubingen	11.252	58
	LINDAT_CLARIN_digital_library_at_Institute_of_Formal_and_Applied_Linguistics_Charles_University_in_Prague	11.211	1,163
	IMS_Repository	11.176	66
	The_Language_Archive_s_IMDI_portal	11.023	116,30
	Common_Language_Resources_and_Technology_Infrastructure_Norway_CLARINO_Bergen_Repository	10.903	36
	Slovenian_language_resource_repository_CLARIN_SI	10.889	35
	TalkBank	10.886	16,352
	Speech_and_Language_Data_Repository_SLDL_ORTOLANG_	10.855	271

- Layout issues in Firefox on Mac



Curation Module ÖAW ACDH 

	id	Name	Score	Facet Cove...	Perf: Of Elements With
Form	clarin.eu:cr1:p_1337778924932	Chromosome_Example	1.654	0.654	0
Public Profiles	clarin.eu:cr1:p_1271859438175	cmdi-virtual-collection	1.654	0.654	0
Collections	clarin.eu:cr1:p_1371047304765	EMITX	1.654	0.654	0
SMC Browser	clarin.eu:cr1:p_1337778924946	Gender_Example	1.654	0.654	0
Help	clarin.eu:cr1:p_1337778924930	KinTerm	1.654	0.654	0
	clarin.eu:cr1:p_1357720977499	Oka_individual	1.654	0.654	0
	clarin.eu:cr1:p_1290431694584	RECOGNIZER	1.654	0.654	0
	clarin.eu:cr1:p_1375880372976	singlePaperPackage	1.654	0.654	0
	clarin.eu:cr1:p_1357720977500	Social_Event	1.654	0.654	0
	clarin.eu:cr1:p_1282306194485	TITUS_metadata	1.654	0.654	0
	clarin.eu:cr1:p_1349361150729	GBA-derived_sub_province	1.691	0.654	0.037
	clarin.eu:cr1:p_1355150532299	GBA-derived_sub_coroparea	1.692	0.654	0.038
	clarin.eu:cr1:p_1355150532292	GBA-derived_sub_dialectarea	1.692	0.654	0.038
	clarin.eu:cr1:p_1349361150742	GBA-derived_sub_municipality	1.692	0.654	0.038

Basic integration with Component registry possible (as soon as a 'production' version is available - to be deployed on a CLARIN host)

Set up a development version for further development

HTML-version via the REST-API

Expose the XSL-stylesheets

=> make a printable simple html page (via content negotiation or xslt processing instruction)

Integration with harvesting and vlo - most efficient to have it on one server.

So the production version would run on the vlo/catalog server.

VLO & curation/mapping integration

Update availability mapping - see

<https://vlo.clarin.eu/search?fq=collection:Leipzig+Corpora+Collection>

License: CC BY-NC

Values already mapped:

CC BY-NC 3.0 DE
CC BY-NC 3.0
CC BY-NC 4.0
CC BY-NC-SA 3.0
CC-BY-NC
CC - BY - NC

For now just update the mapping in git-repo and notify Twan

Consider use of regular expressions?

VLO

Application of CLARIN styling https://github.com/clarin-eric/interface_guidelines

Work in progress (based on current work of Amir): <http://alpha-vlo.clarin.eu/vlo/search>

Schedule: ???

Schedule and implementation will be discussed in a meeting with Amir later today (28 April)

VLO 3.4.1 has been deployed to <http://vlo.clarin.eu> - see

<https://github.com/clarin-eric/VLO/releases/tag/3.4.1>

Davor: Google indexing?

Twan: Sitemap submitted - It is being processed, but slowly and not completely yet.

Next meeting:

- overview of the post-processing and to which degree the can and should be shared between vlo-importer and curation module
- CMDI 1.2 support in curation module

Thomas:

<https://trac.clarin.eu/ticket/819> - metadata helper class - related to current discussion

- move it away from VLO 4.0 milestone

And try to align with the discussion on shared library for facet mapping and normalisation