deutsches literatur archiv marbach

Data processing of ILS data to facilitate a new discovery layer for the German Literature Archive (DLA)

Thomas Meyer, <u>DLA Marbach</u> Felix Lohmeier, <u>Open Culture Consulting</u>

ELAG2019, Berlin, 2019-05-09

a new catalogue for DLA



- current catalogue about 20 years old ...
- new catalogue prototype project started 2016
- current project started 2017
- currently working on release 2 (of 8)



what to expect from a new catalogue



what are users of DLA working on?



<u>aStec</u> aDIS/BMS - Kallías data foundation: authority records serving as glue



Тур	Person (piz)				
Name	Benn, Gottfried				
Kennzeichen	individualisiert				
Geschlecht	männlich				
Personenstatus	Literarische/r Autor/in im Sinne der Bibliothek Person der Zeitungsausschnittsammlung Bestandsbildner/in im Sinne der Handschriftensammlung Bestandsbildner/in im Sinne der Sammlung Bilder und Objekte				
Lebensdaten	02.05.1886 - 07.07.1956				
Land	Deutschland				
Notation	L				
Erstveröff Jahr	1912				
Syst. Dok.stelle	xy Autoren, Kritiker und Übersetzer				
Geographischer Bezug	Mansfeld [Geburtsort: Geografikum] Berlin [Sterbeort: Geografikum]				
Sachbezug	Arzt [Beruf/Funktion: Schlagwort] Dr. med. [Akad. Grad: Schlagwort] Schriftsteller [Charakt. Beruf: Schlagwort]				



Тур	Person (piz)
Name	Benn, Gottfried
Kennzeichen	individualisiert
Geschlecht	männlich
Personenstatus	Literarische/r Autor/in im Sinne der Bibliothek Person der Zeitungsausschnittsammlung Bestandsbildner/in im Sinne der Handschriftensammlung Bestandsbildner/in im Sinne der Sammlung Bilder und Objekte
Lebensdaten	02.05.1886 - 07.07.1956
Land	Deutschland
Notation	L
Erstveröff Jahr	1912
Syst. Dok.stelle	xy Autoren, Kritiker und Übersetzer
Geographischer Bezug	Mansfeld [Geburtsort: Geografikum] Berlin [Sterbeort: Geografikum]
Sachbezug	Arzt [Beruf/Funktion: Schlagwort] Dr. med. [Akad. Grad: Schlagwort] Schriftsteller [Charakt. Beruf: Schlagwort]

Til is the se- cin - new

Gottfried Benns Werk in der Musik





-

Тур	Person (piz)
Name	Benn, Gottfried
Kennzeichen	individualisiert
Geschlecht	männlich
Personenstatus	Literarische/r Autor/in im Sinne der Bibliothek Person der Zeitungsausschnittsammlung Bestandsbildner/in im Sinne der Handschriftensammlung Bestandsbildner/in im Sinne der Sammlung Bilder und Objekte
Lebensdaten	02.05.1886 - 07.07.1956
Land	Deutschland
Notation	L
Erstveröff Jahr	1912
Syst. Dok.stelle	xy Autoren, Kritiker und Übersetzer
Geographischer Bezug	Mansfeld [Geburtsort: Geografikum] Berlin [Sterbeort: Geografikum]
Sachbezug	Arzt [Beruf/Funktion: Schlagwort] Dr. med. [Akad. Grad: Schlagwort] Schriftsteller [Charakt. Beruf: Schlagwort]

4 . The firsten ser vin - ven





-





en. atus Jahr elle Geograpmscher Bezug

Sachbezug

Person (piz)
Benn, Gottfried
individualisiert
männlich
Literarische/r Autor/in im Sinne der Bibliothek Person der Zeitungsausschnittsammlung Bestandsbildner/in im Sinne der Handschriftensammlung Bestandsbildner/in im Sinne der Sammlung Bilder und Objekte
02.05.1886 - 07.07.1956
Deutschland
L
1912
xy Autoren, Kritiker und Übersetzer
Mansfeld [Geburtsort: Geografikum] Berlin [Sterbeort: Geografikum]
Arzt [Beruf/Funktion: Schlagwort] Dr. med. [Akad. Grad: Schlagwort] Schriftsteller [Charakt. Beruf: Schlagwort]

one author - various ca	atalogue items	deutsches literatur archiy marbach
Autor and andres feel with the second	enns Werk enns Werk der Musik enns Werk der Musik	Person (piz) Benn, Gottfried individualisiert männlich Literarische/r Autor/in im Sinne der Bibliothek Person der Zeitungsausschnittsammlung Bestandsbildner/in im Sinne der Handschriftensammlung Bestandsbildner/in im Sinne der Sammlung Bilder und Objekte 02.05.1886 - 07.07.1956 Deutschland L 1912 xv Autoren, Kritiker und Übersetzer
Ge	ograpms cher Bezug	Mansfeld [Geburtsort: Geografikum] Berlin [Sterbeort: Geografikum]
Sad	chbezug	Arzt [Beruf/Funktion: Schlagwort] Dr. med. [Akad. Grad: Schlagwort] Schriftsteller [Charakt. Beruf: Schlagwort]















<u>aStec</u> aDIS/BMS - Kallías data foundation: an ILS adressing all DLA holdings including authority files





deutsches literatur archiv marbach





- > Totenmasken
- > Buchumschläge Verlag Hanser
- > Noten
- > Netzliteratur
- > Hörspiele

Digitalisate
Friedrich Schiller Manuskripte



Geburtstage in diesem Monat

- > Fritz Rudolf Fries
- > Harry Kessler
- Monika Plessner
- > Peter Szondi
- > Winfried Georg Sebald

Neu im Katalog

- > Bibliothek Ernst Jünger
- > Bilder und Objekte im Nachlass Dieter Kühn
- Bilder und Objekte im Vorlass Hans Magnus Enzensberger
- > Handschriftlicher Nachlass Karlheinz Barck



- Literature starting from the 19th century up to the present, including scholarly literature at maximum completeness
- Literary holdings, autographs, manuscripts, correspondence, printed works, AV material, portraits, memorabilia
- German-speaking authors, critics, translators, philosophers, philologists and other scholars
- Documents and their impact from 1750 to the present day
- Archives of publishers and editors



- currently no open format available addressing all DLA holdings in their detail level
 - source data: 1200 fields
 - relevant for the project: 470 fields
- implement a custom (non-public) format, following aDIS/BMS internal format (available as "Token" export)

design principles

- Open Source throughout
- ✓ modular system architecture
- enable DLA to customize and enhance
- ✓ ability to reuse data in other environments



for detailed (German) information visit https://felixlohmeier.de/dla/systemarchitektur.html

lookup tables (Pandas)

- preprocessing with
 <u>Python Data Analysis Library</u>
- records contain codes (controlled vocabulary) that need to be resolved based on a csv file
- additional global replacements are done based on a second csv file

deutsches literatur archiv marbach

read csv file with global replacements
mappings = pd.read_csv('mappings.csv', sep=',')

select only mappings from subset "AK"
mappings =
 mappings[mappings['Bestand'].str.contains('AK',
 na=True, case=False)]

join mappings with existing dataframe
df = df.join(mappings, on=['key', 'value'])

```
# update values and drop new columns
df['value'].update(df['neu'])
df.drop('neu', axis=1, inplace=True)
```

lookup tables (Pandas)

deutsches literatur archiv <mark>marbach</mark>

Branch: master - opac-ng_etl / conf / mappings.csv	Find file Copy path
dia-meyer #1529 FUNKTI in weiteren Feldern ersetzen	439e3a1 on Feb 11
2 contributors 🗝 🛗	
457 lines (456 sloc) 16 KB	Raw Blame History

Q AI710

1	Bestand	Adis Feld	alt	neu
219	AK	AI710	anla_Ts	Anlass
220	AK	AI710	anla_Tu	Anlass
221	AK	AI710	feie_Tb	Gefeierte Körperschaft
222	AK	AI710	feie_Tp	Gefeierte Person
223	AK	AI710	feie_Ts	Gefeiertes Ereignis
224	AK	AI710	geoa_Tg	Geografikum, allgemein
225	AK	AI710	geoa_Ts	Geografikum, allgemein

grouping (Pandas)

- records are transposed from line-based (n lines per record) to column-based (one record per line) format
 - multiple values are stored in one cell with separator "s (<u>U+241F</u>)

```
# pivot
```

```
df = df.groupby(['risn', 'key'])['value'].
    agg('us'.join).unstack()
```

 output: csv files (will be imported into OpenRefine in the next step)

export
df.to csv('ak.csv')

grouping (Pandas)

deutsches literatur archiv marbach

- records are transposed from line-based (n lines per record) to column-based (one record per line) format
 - multiple values are stored in one cell with separator ^u_s (<u>U+241F</u>)
- output: csv files (will be imported into OpenRefine in the next step)

pandas statistics for subset AK

print(df.info(verbose=True, null_counts=True))

<class 'p<="" th=""><th>oandas.core.frame.DataFrame'></th></class>	oandas.core.frame.DataFrame'>
Index: 16	538796 entries, 10 to 999999
Data col	umns (total 114 columns):
id 10	638796 non-null object
A0001	1638796 non-null object
A0038	31161 non-null object
A00578	6 non-null object
A01030	1493909 non-null object
A03300	1368812 non-null object
A0331	1398861 non-null object
A0335	723560 non-null object
A0359	1279365 non-null object
A0360	1700 non-null object
()	
dtypes: o	object(114)
memory	usage: 1.4+ GB

grouping (Pandas)

•	records are tra	466 r	natch	ing rows	(1625213	total)					Extensions:	VIB-Bits -	
	line-based (n l	Show a	as: row	s records	Show: 5	10 25 50 rows				« first « p	revious 1-5	next > last >	»
				💌 id	A0001	A0038	A00578	A01030	A03300	A0331	A 0335	A 0359	•
	(one record pe		6890.	AK01006426	01006426	Deutsch ". Tschechisch		Abgeschlossen	txt	Autorenportraits	München - Praha - Dresden ; Euro-Regio- Tour 1994	[hrsg. von Peter Becher im Auftr. des Adalbert- Stifter- Vereins]	
	stored in c separator		9346.	AK01008707	01008707	Portugiesisch % Englisch		Abgeschlossen	txt	Genie und Wahnsinn	Schriften zu einer intellektuellen Biographie	Fernando Pessoa. Aus dem Portug. und Engl. übers., mit Anm. und einem Nachw. vers. von	
•	output: csv file (will be import OpenRefine in											Steffen Dix. Basierend auf der von Jerónimo Pizarro hrsg. Kritischen Ausg.	
			9707.	AK01009033	01009033	Schwedisch ^{1,} Französisch		Abgeschlossen	txt	Natur-Trilogie		August Strindberg. Aus dem Schwed. und Franz. übertr. von Emil Schering	

scripting OpenRefine

- OpenRefine can be managed via command line with <u>OpenRefine-client</u> (one of many clients)
- <u>Short video demo</u> of transforming DOAJ articles
 - create project from csv file
 - \circ apply rules from json file
 - export project to stdout

deutsches literatur archiv marbach

video demo

joins/lookups (OpenRefine)

- records contain fields with multiple foreign keys to internal authority data (e.g. persons)
- 223 new columns are generated with an <u>fork</u> of OpenRefine that enhances cross function to allow splitting multivalued fields when joining projects)

002E6642 004240EE

deutsches literatur archiv <mark>marbach</mark>

Snow	as: rows	s records	Show: 5 10 25	5 50 rows				
	-				5081	6		
001357:	Add c	olumn bas	sed on colum	n PE0100	26	E		
	New column name		XX_PE_P0800_PE0100					
000817			• set to blank	s \bigcirc store error \bigcirc copy value from original column	26			
	Expres	sion	Language	General Refine Expression Language (GREL) \checkmark		C		
	forEa c v f	ch(ells["PE010 , orNonBlank(0"] <u>.value</u> .spli	t(""),				
002603	Pre	eview Hi	story Starred	i Help	09	EEC		
	row	/ value		forEach(cells["PE0100"].v		B		
	1.	0000003		Finck, Adrien				
	2.	00000079		Beyer, Uwe				
	3.	00000718		Krug, Hans-Jürgen				
		00003564		Heidenreich, Gert	17	E		
0117491	4.					S		
D117491	4. 5.	00084094		Schneider, Ute		E		
)11749:	4. 5. 6.	00084094 00021615%	00033588	Schneider, Ute Jacobson, Eduard [®] Salingré, Hermann	-	B		

00021616 00022699

20001212 20100122

00006496

joins/lookups (OpenRefine)

- records contain fields with multiple foreign keys to internal authority data (e.g. persons)
- 223 new columns are generated with an <u>fork</u> of OpenRefine that enhances cross function to allow splitting multivalued fields when joining projects)

```
# source: ak, PE0100 (contains keys)
# target: pe, foreign key in P0001, P0800
forEach(
   cells["PE0100"].value.split(" "s"),
  V,
  forNonBlank(
     v.cross(
         "pe",
         "P0001",
         пυ_п
      )[0].cells["P0800"].value,
     Х,
      Х,
     II <sub>R</sub> II
).join("<sup>"</sup>")
```

result list (OpenRefine)

- additional fields are generated with custom rules for each subset
 - listview_title
 - listview_associate
 - listview_additional1
 - listview_additional2

deutsches literatur archiv <mark>marbach</mark>

Treffer 1-25 von 139

<	1 2 3 4 5 6 >	Sortierung	Standard	•
Ēl	Relevante Namen & Werke Bobrowski, Johannes (1917-1965) Schriftsteller, Lektor < Verlag>			
ĒI	Relevante Namen & Werke Bobrowski, Johannes <1917-1965>. Levins Mühl Bobrowski, Johannes 1964	e (Romai	ı : 1964)	
Ţ	Levins Mühle Seemann, Horst; Bobrowski, Johannes; Geschonneck, Erwin; Paryla, Katja; Düren, Fred [Berlin] : Icestorm Entertainment, 2006 - 1 DVD Video (113 Min.)	listvie	listview_title ew_associate w_additional1	*
	Bobrowskis Roman. [Levins Mühle.]. [Rez. zu:] Levins Mühle / Johannes Bobrowski : [Rezension]	iistvi	listview_title	*
ĺ	Erschienen in: Sinn und Form : Beiträge zur Literatur 17 (1965), Seite 631-636	listvie	w_additional1	

Johannes Bobrowski, Levins Mühle. 1964. [Rez.

result list (OpenRefine)

- IDs of linked authority records are stored in several fields in the source data
- these IDs are copied to a common field "entity_ids"
- top 2 authority records for the result set are determined by a facet on field "entity_ids"

deutsches literatur archiv <mark>marbach</mark>

Treffer 1-25 von 139



Johannes Bobrowski, Levins Mühle. 1964. [Rez.

facets (OpenRefine)

 custom rules for each subset Inhaltsverzeichnis Zusatzfelder für Facetten und Suchfilter AK facet medium facet_form_content facet names facet roles facet language facet time facet time stat facet source facet location facet subject filter digital filter new filter collection filter benutzungscode filter hidden AU filter benutzungscode filter hidden BF facet form content facet names facet time facet time stat facet source

facets (OpenRefine)

- custom rules for each subset
- example: Solr's DateRangeField expects formats like
 - [2000-11-01 TO
 2014-12-01]

 [2014 TO 2014-12-01]



Column tmp / Customized facets / facet by blank / false
 Column facet_time / Edit cells / Transform...

with(cells["tmp"].value,x,forNonBlank(value,v,v+"s"+x,x))

• Column tmp / Edit column / Remove this column

facet_time_stat

• Column facet_time / Edit colum / Add column based on this column... / facet_time_stat

```
with(value.replace("[","").replace("]","").replace(" ","").replace("TO"," ","),v1,
forEach(filter(v1.split(" ",").uniques().sort(),v2,isNumeric(v2)),v3,
if(v3.length()==4,v3+"-01-01T00:00:00Z","") +
if(v3.length()==6,v3+"-01T00:00:00Z","") +
if(v3.length()==8,v3+"T00:00:00Z","")
).join(" ","))
```

facets (OpenRefine)

- custom rules for each subset
- example: Solr's DateRangeField expects formats like
 - [2000-11-01 TO
 2014-12-01]
 - [2014 TO 2014-12-01]



detail page (OpenRefine)

custom rules for each subset

deutsches literatur archiv <mark>marbach</mark>

Inhaltsverzeichnis

Zusatzfelder für Detailseiten

AK

Wert X in Verfasserangabe (A0359) entfernen enthaltene_materialien (aus BI und HS, ggf. über Exemplarsätze) detail_urheber detail_urheber_ids detail_beteiligt detail_beteiligt_ids detail_bestandsangaben detail_relation_ids und detail_relation_type detail_relation_title HS helper_benutzungscode (effektiver Benutzungscode) helper alive (Lebende Personen?)

helper_ausleihstatus_bf (Ausleihstatus BF)

detail_restrictions (Benutzungsbeschränkungen)

detail page (OpenRefine)

- custom rules for each subset
- example persons
 - differentiated into authors and contributors
 - with birth/death year and role from authority record
 - with links to authority records

deutsches literatur archiv <mark>marbach</mark>

OPAC-NG-Hauptprojekt

detail_urheber

siehe Spezifikation in Ticket #149

• Column SUCHEF / Edit column / Add column based on this column... / detail_urheber

```
with(
    forEachIndex(
        cells["SUCHEF"].value.split("","),
        i1,
        v1,
        v1 + if(
            cells["GFUPER"].value.split(""")[i1].contains("_GS"),
            "_GS",
    ),
    х,
    forNonBlank(
        forEachIndex(
            filter(
                х,
                 z2,
                z2.contains("PE")
            i2,
            v2,
            if(
                v2.contains("_GS"),
                cells["XX_PE_listview_title_PE0100"].value.split(""")[i2]
                  + cells["GFUPE"].value.split("";")[i2]
                  + "]",
                 null
        ).join("""),
        y2,
        " "s"+y2,
    ) + forNonBlank(
```

detail page (Ope

custom rule: each subset

examp

Ο

Ο

Ο

Ы Levins Mühle - Gekürzte Fassung vom Komponisten

Zimmermann, Udo; Zimmermann, Ingo; Bobrowski, Johannes; Kurz, Siegfried; Rundfunk-Sinfonieorchester <Leipzig>; Mitteldeutscher Rundfunk <Leipzig> / Chor Hamburg : Edel Classics, 2001 - 1 CD (58 Min.)

mple pe	Details	Zugang (?)	Platzhalter 1	Platzhalter 2			
different	Medium			Audio-CD			
into autl	Urheber			Zimmermann,	Udo (1943-) [Komponist]		
contribu				Zimmermann,	Zimmermann, Ingo (1940-) [Librettist]		
with birt	Titel			Levins Mühle / Udo Zimmermann. Text: Ingo Zimmermann nach Johannes Bobrowski. Rundfunk-Sinfonie-Orchester Leipzig. Dirigent: Siegfried Kurz.			
year and				Rundfunkchor	Leipzig		
from au ⁻	Beteiligt			<u>Bobrowski, Jo</u>	hannes (1917-1965) [Verfasser der literarischen Vorlage]		
record				Kurz, Siegfried	d [Dirigent]		
				Rundfunk-Sinf	fonieorchester <leipzig> [Sonstige Person, Familie und</leipzig>		
with link				Korperschatt	ar Bundfunk <lainzig> / Char [Sanatiga Parsan, Familia und</lainzig>		
authority				Körperschaft]	er Kundrunk <leipzig familie="" ferson,="" joonstige="" onor="" th="" und<=""></leipzig>		
	Erschiene	en		Hamburg : Ed	lel Classics, 2001		

×



Kallías – der Online-Katalog des Deutschen Literaturarchivs Marbach	iteraturarchivs Marbach		
😥 Gesamtbestand 📧 Gedrucktes 💵 Handschriften	ĒI	JGCotta'sche Buchhandlung <stuttgart> (1810-1889) Körperschaft</stuttgart>	*
≊ Expertensuche i≡ Bestandsübe	ĒÌ	Goethe, Johann Wolfgang von (1749-1832) Schriftsteller, Politiker, Naturwissenschaftler	*
	Ξł	Unseld, Siegfried (1924-2002) Verleger	*
	Ξl	Jünger, Ernst (1895-1998) Schriftsteller, Entomologe, Offizier	*
	=1	Hesse, Hermann (1877-1962) Schriftsteller, Buchhändler, Maler, Pazifist, Nobelpreisträger	*

Kallías – der Online-Katalog des Deutschen Literaturarchivs Marbach	ĒÌ	Suhrkamp-Verlag <frankfurt, main=""> (1950- Frankfurt am Main Körperschaft</frankfurt,>	(1950-)		*
			entity_score: 1	120046	~
Gesamtbestand 🖬 Gedrucktes 📲 Handschriften	ĒÌ	JGCotta'sche Buchhandlung <stu< th=""><th>ttgart></th><th></th><th>*</th></stu<>	ttgart>		*
		(1810-1889) Körperschaft	entity_score:	64203	~
₽ Evpertensuche := Bestandsübe	ĒÌ	EI Goethe, Johann Wolfgang von (1749- Schriftsteller, Politiker, Naturwissenschaftler	-1832)		*
			entity_score:	29401	~
	Ξ1	Unseld, Siegfried (1924-2002) Verleger			*
			entity_score:	28874	~
	ĒÌ	Jünger, Ernst (1895-1998)			*
		Schriftsteller, Entomologe, Offizier	entity_score:	28701	~
	ĒÌ	Hesse, Hermann (1877-1962)			*
		Schriftsteller, Buchhändler, Maler, Pazifist, Nobelpreisträge	entity_score:	24828	~

- all values of field "entity_ids" (contains IDs of linked authority records, cf. result list) are extracted into a list and loaded into a separate project "entities"
 - o { { forNonBlank(cells["entity_ids"].value,v, forEach(v.split(" "s"),x,x+"\n"),"") } }
- in this new project "entities" a new column "entity_count" is added that contains the "score"

 value.facetCount("value","entity_ids")
- finally all authority records (persons, corporate bodies, work records) are enriched with that score

 forNonBlank(cell.cross("entities","entity_ids")

[0].cells["entity_count"].value,v,v,0)



entities				
entity_ids	entity_count			
PE0000001	2			
PE0000003	1			
PE0000002	2			
PE0000001	2			
PE0000002	2			

persons				
id	entity_score			
PE0000001	2			
PE0000002	2			
PE0000003	1			

enrichment (in development with OpenRefine)

- source data contains IDs from Gemeinsame Normdatei (GND)
- HBZ/lobid provides a <u>GND reconciliation service</u> for OpenRefine including data from EntityFacts.
- Wikidata itself provides a <u>Wikidata reconciliation service</u> for OpenRefine. Matches can be refined by property <u>P227</u> (GND-ID).
- Reconciliation will be implemented as a separate (monthly?) process that provides a "cache" project for enrichments in daily routine
- external data of interest to DLA:
 - Link to German Wikipedia
 - Link to official website (Wikidata: <u>P856</u>)
 - Link to image at Wikimedia Commons (Wikidata: P18)
 - Link to signature at Wikimedia Commons (Wikidata: P109)
 - Link to Filmportal (Wikidata: <u>P2639</u>)



output

- CSV files ready to be loaded into the search index (Solr)
 - customized: relevance ranking, synonym graph, suggester, ... (not covered in this presentation)
- **OpenRefine** for data analytics and extracting data (e.g. filtered lists for digitization projects)
- EAD-XML provided via <u>file-based</u>
 <u>OAI-PMH Data Provider</u>



deutsches

automated workflow (Bash scripts)

deutsches literatur archiv marbach

felix@tux:~ × ext lohm@neo:~/opac-labor/logs> tail -n 19 20190430 022032.log Ende Aktualisierung OpenRefine-Demo: Di 30. Apr 07:50:57 CEST 2019 === Statistik === Beginn und Laufzeiten der einzelnen Prozessierungsschritte: Beginn Gesamtprozessierung Di 30. Apr 02:20:32 CEST 2019 (00:00:00) Arbeitsverzeichnis leeren Di 30. Apr 02:20:32 CEST 2019 (00:00:00) Vorprozessierung aDIS Token Di 30. Apr 02:20:32 CEST 2019 (00:00:00) Transformation aDIS Token Di 30. Apr 02:20:32 CEST 2019 (01:10:39) Kopplungen und Bestandspezifisches Di 30. Apr 03:31:11 CEST 2019 (01:17:48) Zusammenfassung und Ergebnisprüfung Di 30. Apr 04:48:59 CEST 2019 (00:00:37) Neuindexierung Solr 1 Di 30. Apr 04:49:36 CEST 2019 (02:01:51) csv2ead 1/2 OpenRefine Di 30. Apr 06:51:27 CEST 2019 (00:39:06)Server: csv2ead 2/2 XSLT Di 30. Apr 07:30:33 CEST 2019 (00:03:53) 2 vCPUs OpenRefine Demo aktualisieren Di 30. Apr 07:34:26 CEST 2019 (00:16:31) Ende Gesamtprozessierung Di 30. Apr 07:50:57 CEST 2019 (00:00:00)

Gesamtlaufzeit: 05:30:25 (hh:mm:ss)

ext lohm@neo:~/opac-labor/logs> echo "hello #elag2019"

Source data: 4.3 million records with up to 116 fields (6.9 GB txt)

Search index: 4.2 million records with 561 fields (6.87 GB Solr)

64 GB RAM

fully automated with bash scripts (planned: Rundeck)

frontend (TYPO3-find)

- the frontend is built with TYPO3 extension "find"
 - developed initially by State and University Library Göttingen
 - available at GitHub: <u>https://github.com/subugoe/typo3-find</u>
- customized (not covered in this presentation):
 - custom templates for each subset with search suggestions
 - custom facets: time range, dynamic hierarchical facet
 - custom result list with top 2 authority record hits
 - custom search box with tokenized autocompletion

interaction of open source components – example autocompletion



interaction of open source components – example autocompletion



interaction of open source components – example autocompletion



deutsches literatur archiv marbach

live demonstration

- project required and brought deep insights into DLA ILS data structures
- high effort spent in transforming highly specialized ILS data without losing detail information
- increased in-house librarian/archivists data literacy
- OpenRefine as a valuable self-service information tool
- data transformation enabled ILS data reuse
- Redmine as a proven tool to communicate and to document

testing environment

deutsches literatur archiv marbach

During ELAG the beta test env is freely available online. We are looking forward to your feedback!



https://www-test.dla-marbach.de (available until May 12th 2019)

deutsches literatur archiv marbach

Questions are gladly answered by: <u>Thomas.Meyer@dla-marbach.de</u> <u>Felix.Lohmeier@opencultureconsulting.com</u>

Participating service providers:

<u>Open Culture Consulting</u> (OpenRefine) <u>effective WEBWORK GmbH</u> (TYPO3)

Markus Mandalka (Solr)

Kai Mertens (Design)

Kallias ILS aDIS/BMS by aStec GmbH