EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Seminar für Sprachwissenschaft

The CLARIN Language Resource Switchboard

Claus Zinn (dev. 2015-2019) and Emanuel Dima (dev. 2019-)

Oberseminar, Winter Term 2021, 19.01.2022

CLARIN

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Motivation

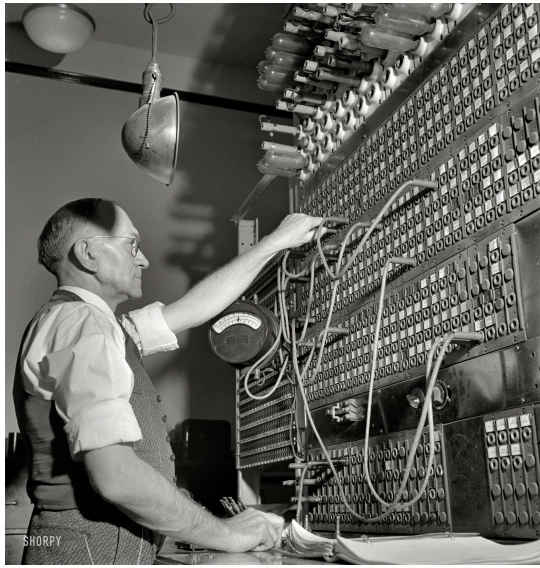
The CLARIN Language Resource Switchboard (LRS) aims at bridging the gap between resources (as identified in the VLO, FCS, and VCR) and tools that can process these resources in one way or another.

-- CLARIN-PLUS deliverable D2.5
https://office.clarin.eu/v/CE-2015-0684-LR_switchboard_spec.pdf

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Motivation

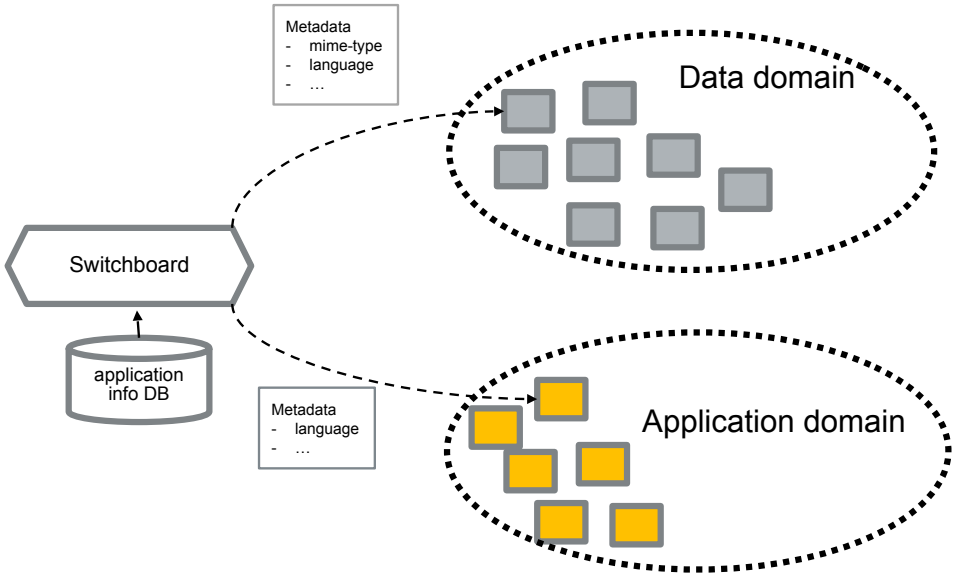
- improve user experience for users (from the VLO, FCS, VCR *etc.*)
- users experience a gap between the resources they find, and the processing workflows they would like to call
- for each resource found, the Switchboard suggests applicable web-based tools that can process the resource
- users can invoke applicable tool by **simple** click
- relevant information forwarded to tool in question
- switchboard has achieved its task, it does no more

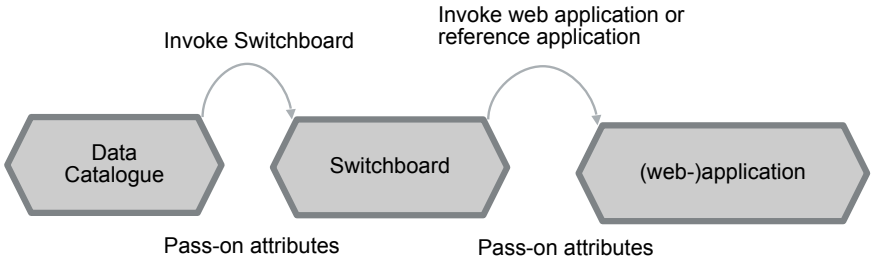


https://switchboard.clarin.eu

The Switchboard

- brokers between data and (web-)applications
- can be integrated into data-catalogues
- contains a catalogue of (web-) applications suitable for specific input data
- presents a list of applicable (web-)applications for resources based on the data formats





Figures: Adapted from Daan Broeder

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

The Switchboard

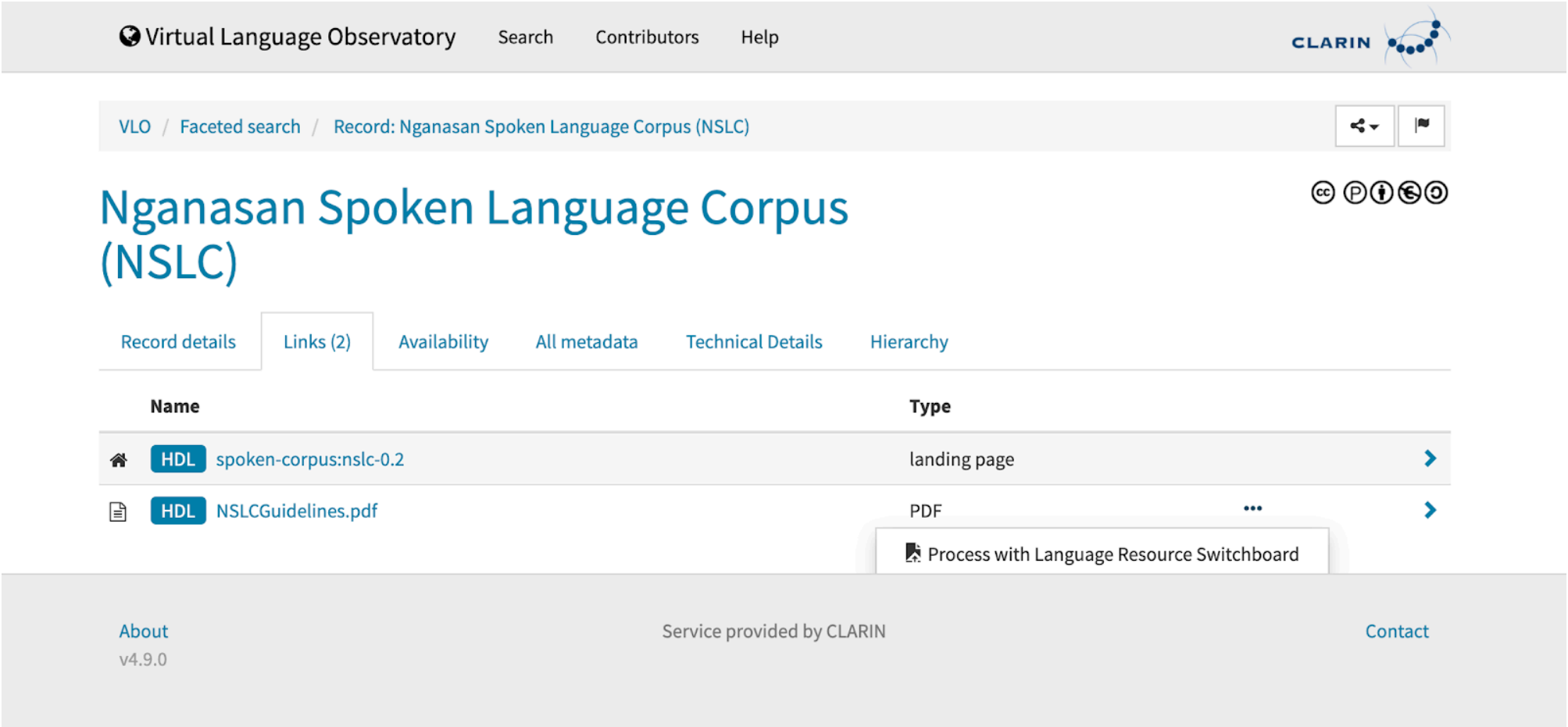
- From an infrastructure building perspective:
  - good scalable way to connect many data-catalogues type to relevant applications
  - connect different infrastructures, potential of cross-discipline fertilisation
  - light-weight; currently only a REST call from the data catalogue or a link to the application
- From a user perspective
  - easy access to different (web-) applications
  - (offer could be personalised for user profile and discipline, and according to interest/usage)
- From a funder perspective
  - offered applications become explicit and verifiable
  - applications become FAIR

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Examples

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Switchboard integration with the VLO



https://vlo.clarin.eu/record/http_58__47__47_hdl.handle.net_47_11022_47_0000-0007-C9C5-8

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Switchboard integration with the FCS (mock-up)

Tagesspiegel – Homepage

Berlin-Brandenburg Academy of Sciences and Humanities

Tagesspiegel - Newspaper archive at the Berlin-Brandenburg Academy of Sciences.

German

Display as Key Word In Context | Download | Switchboard

Kiri wird ausgestopft Der in der Nacht zu Donnerstag verstorbene **Elefant** bleibt Berlin erhalten .

Nachrichten 2. 02. 2001 Justizkrimi Dreiundzwanzig Jahre Einsamkeit Frankreich hat mit der Wiederaufnahme des Verfahrens Seznec einen neuen Fall Dreyfus Jörg von Uthmann Die Wiederaufnahme eines rechtskräftig abgeschlossenen Verfahrens ist der weiße **Elefant** , die blaue Mauritius , die ganz große Ausnahme im Strafprozessrecht - in Frankreich nicht anders als in Deutschland .

01. 02. 2003 Sonntag Wie es ist , ein Soldat zu sein mit Dieter Wellershoff Pocken – ein Virus kehrt zurück Fürs Leben lernen Rezzo Schlauch im Fragebogen Was ich mag , und was ich nicht mag : Die Waffen der Frauen Frauen und Männer Donald H. Rumsfeld Zahlen , bitte Das Geheimnis der Wachteleier Essen und Trinken Wolfgang Schmidbauer im Jemen Menschen und **Elefanten** Hellmuth Karaseks Vergleiche

Für Köllner jedenfalls ist ein Verkauf lukrativ , weil in Europa der Import von Wildtieren verboten ist und **Elefanten** derzeit gut 60000 Euro wert sind .

Und allein darf Köllner Kenia laut Gesetz nicht halten : **Elefanten** sind hochsoziale Tiere .

In einer Nacht- und Nebelaktion verließ Karl- Heinz Köllner , Direktor des Zirkus „ Harlekin " , sein Winterquartier in Göritz bei Dessau – den **Elefant** im Gepäck .

Für Thomas Höller vom Tierschutzverein „ Animal Public ein klarer Beweis für Köllners Unverantwortlichkeit : „ Er darf nicht länger **Elefanten** halten " , fordert er .

„ Wir würden aber sofort geeignete Partner suchen , denn mit unseren asiatischen **Elefanten** können wir die afrikanische Kenia nicht zusammenlegen " , sagt Török .

Derzeit befindet sich der **Elefant** „ unter amtstierärztlicher Kontrolle " , wie es bei der Dessauer Polizei heißt .

Eigentlich gehörten **Elefanten** in Zirkussen verboten , fordert der Vorsitzende Alexander Haufellner .

••• More Results

Close

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Henrik Ibsen: works and analyses

General

Name: Henrik Ibsen: works and analyses

Type: extensional

Creation Date: 2014-10-20

Description: A collection of works by and secondary sources about Henrik Ibsen. Inspired by the Hathi Trust collection (http://babel.hathitrust.org/cgi/mb?a=listis;c=1024421342) and extended with resources available via CLARIN.

Purpose: reference

Reproducibility: fluctuating

Reproducibility Notice: Additional references might be added over time.

Persistent identifier: hdl:11372/VC-1002

Keywords:
- Henrik Ibsen
- plays

Switchbord integration with the VCR

https://collections.clarin.eu

Creators

Person: Dieter Van Uytvanck

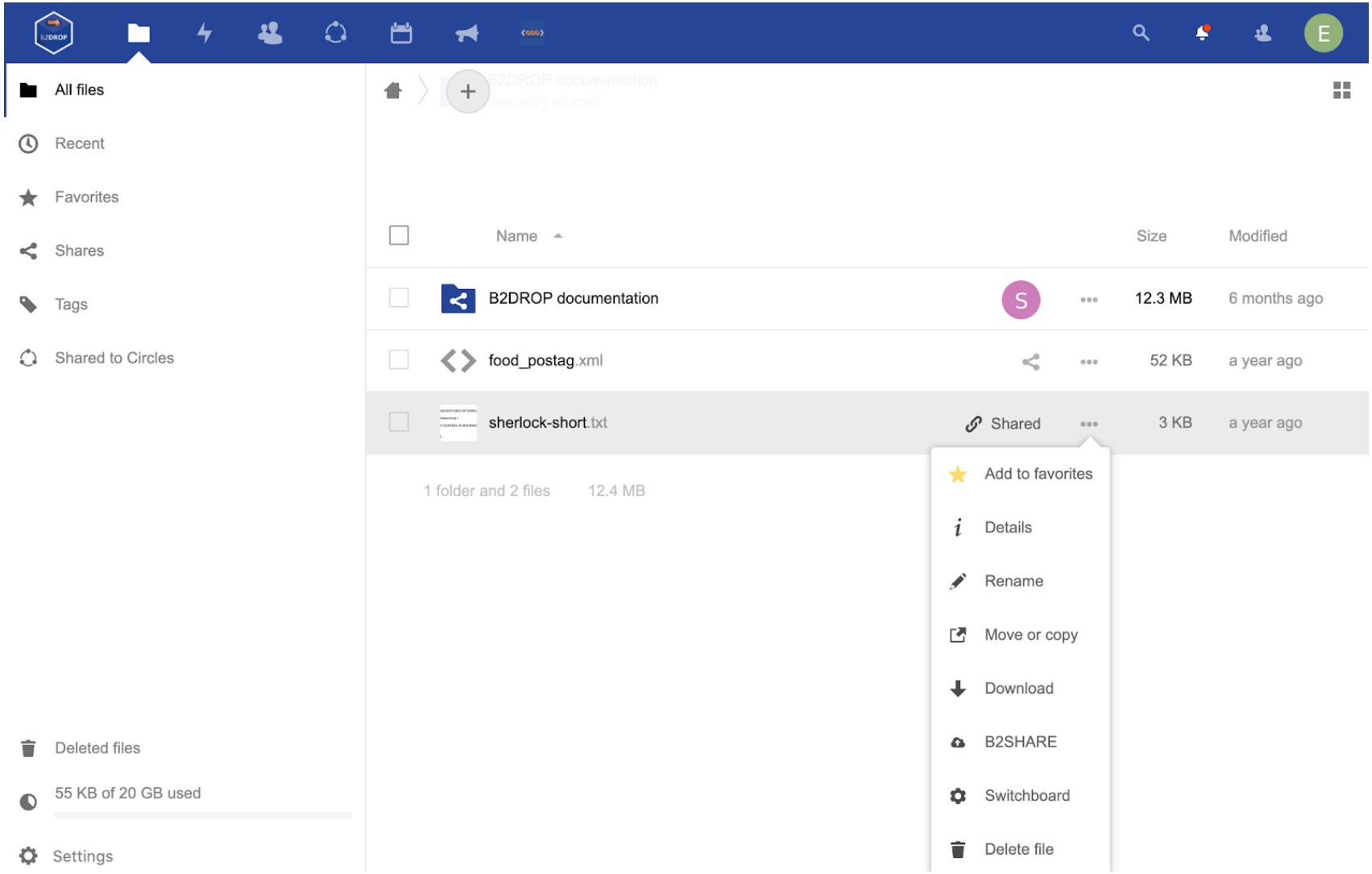
Email: dieter@clarin.eu

Organisation: CLARIN

Resources

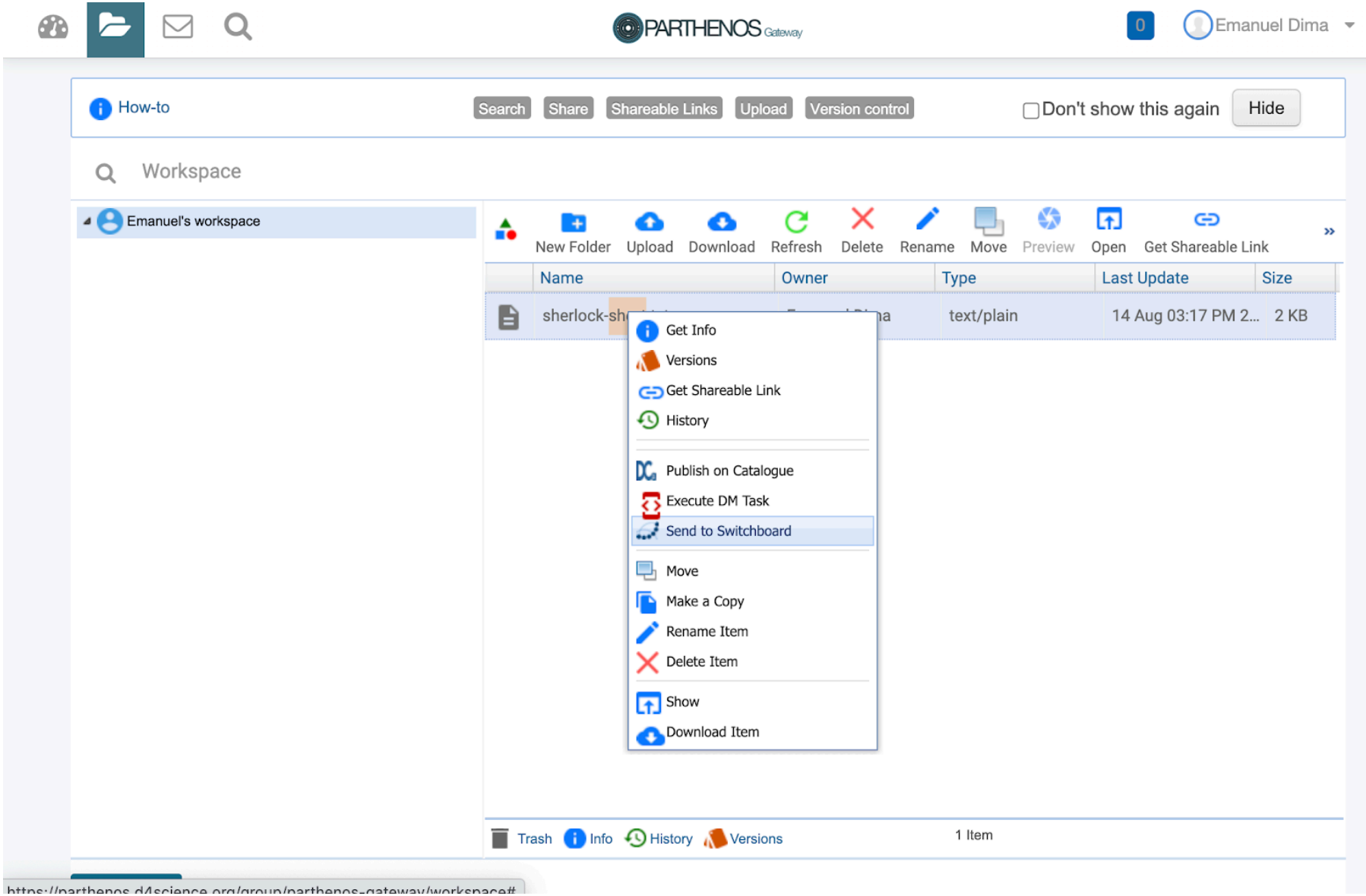
Reference		Type
Peer Gynt (in English) An untraced edition, apparently from 1875.	•••	Resource
Henrik Ibsen, poet, mystic and moralist / by Henry Rose	•••	Resource
Henrik Ibsens Einfluss auf Hermann Sudermann / von Hans Jürgensen	•••	Resource
Ibsen in England / by Miriam Alice Franc	•••	Resource
Letters of Henrik Ibsen / translated by John Nilsen Laurvik and Mary Morison	Process with Language Resource Switchboard	Resource
Ibsen op de planken. Een ensceneringsgeschiedenis van het werk van Henrik Ibsen in Nederland 1880-1995 Dutch PhD thesis (with English abstract) by R.G.C. van der Zalm	•••	Resource
Typisch Ibsen. Transformatie Nederlandse type Ibsen-ensceneringen in de periode 1995-2011 Dutch BA thesis by L.G. de Jong	•••	Resource

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



https://b2drop.eudat.eu

EBERHARD KARLS UNIVERSITÄT TÜBINGEN



https://parthenos.d4science.org/group/parthenos-gateway/workspace

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Metadata

File Type
text/xml

PID
hdl:11858/00-1734-0000-0005-1421-2
Citation Suggestion

Download

Object (TEI)
Metadata (XML)
Tech. Metadata (XML)
Plain Text (txt)
E-Book (epub)
HTML
ZIP

Tools

Switchboard (TEI)
Switchboard (txt)
Voyant
Annotate

Table of contents

Johanna Spyri

Johanna Spyri
Heidi kann brauchen, was es gelernt hat

[11] Reisezurüstungen

Der freundliche Herr Doktor, der den Entscheid gegeben hatte, daß das Kind Heidi wieder in seine Heimat zurückgebracht werden sollte, ging eben durch die breite Straße dem Hause Sesemann zu. Es war ein sonniger Septembermorgen, so licht und lieblich, daß man hätte denken können, alle Menschen müßten sich darüber freuen. Aber der Herr Doktor schaute auf die weißen Steine zu seinen Füßen, so daß er den blauen Himmel über sich nicht einmal bemerken konnte. Es lag eine Traurigkeit auf seinem Gesichte, die man vorher nie da gesehen hatte, und seine Haare waren viel grauer geworden seit dem Frühjahr. Der Doktor hatte eine einzige Tochter gehabt, mit der er seit dem Tode seiner Frau sehr nahe zusammen gelebt hatte und die seine ganze Freude gewesen war. Vor einigen Monaten war ihm das blühende Mädchen durch den Tod entrissen worden. Seither sah man den Herrn Doktor nie mehr so recht fröhlich, wie er vorher fast immer gewesen war.

Auf den Zug an der Hausglocke öffnete Sebastian mit großer Zuvorkommenheit die Eingangstür und machte gleich alle Bewegungen eines ergebenen Dieners; denn der Herr Doktor war nicht nur der erste[11] Freund des Hausherrn und dessen Töchterchens, durch seine Freundlichkeit hatte er sich, wie überall, die sämtlichen Hausbewohner zu guten Freunden gemacht.

»Alles beim alten, Sebastian?« fragte der Herr Doktor wie gewohnt mit freundlicher Stimme und ging die Treppe hinauf, gefolgt von Sebastian, der nicht aufhörte, allerlei Zeichen der

https://textgridrep.org/browse/vqmz_0

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Citation Hermes (2018). Der Hermes Koffer. DARIAH-DE. https://doi.org/10.5072/0000-0011-b8c6-e

HTML References

- Repository landing page
- DataCite record

Metadata References

- Descriptive [TTL * RDF/XML * JSON * JSONLD * NTRIPLES]
- Administrative [TTL * RDF/XML * JSON * JSONLD * NTRIPLES]
- Technical [XML]
- Handle
- DOI
- OAI-PMH [Set * Record]

Data References

- Download file only
- Download ZIP (BagIt)
- Download RDA BagPack ZIP (BagIt)

IIIF References

- Collection's IIIF manifest
- View collection in Mirador

Language Resource Switchboard

- Call Switchboard with this resource (text/vnd.dariah.dhrep.collection+turtle)

https://trep.de.dariah.eu/1.0/dhcrud/21.T11991/0000-0011-B8C6-E/index

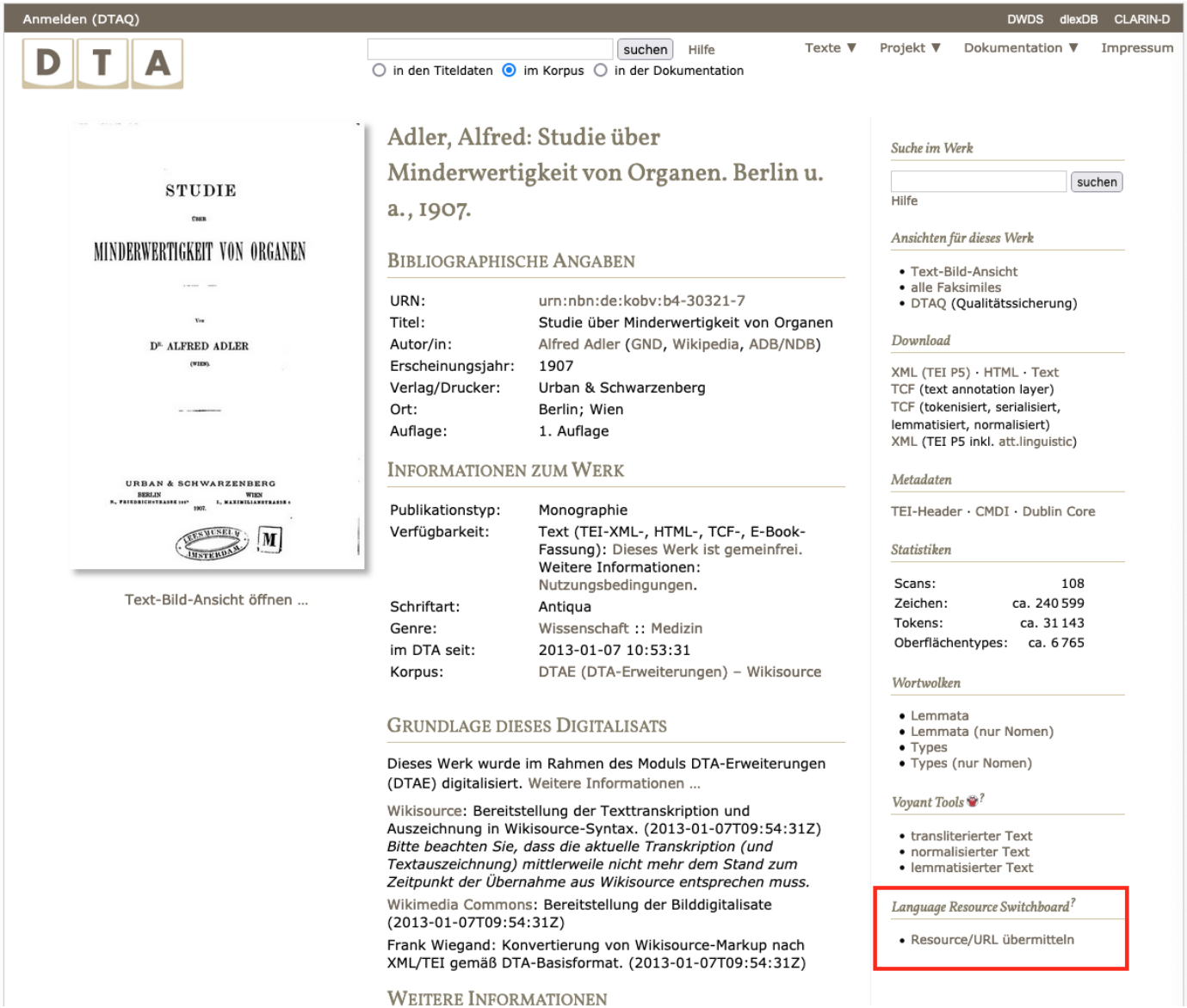
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Resource Information / Actions:

file name:	11858_00-1778-0000-0005-896C-F_tuebadz-stylebook-0607.pdf
file size:	1.29 MB
mime type:	application/pdf
file accessibility:	no
clipBoard:	copy
download/show:	https://hdl.handle.net/11858/00-1778-0000-0005-896C-F@tuebadz-stylebook-0607.pdf
switchboard:	send

https://weblicht.sfs.uni-tuebingen.de/CMDIExplorer/

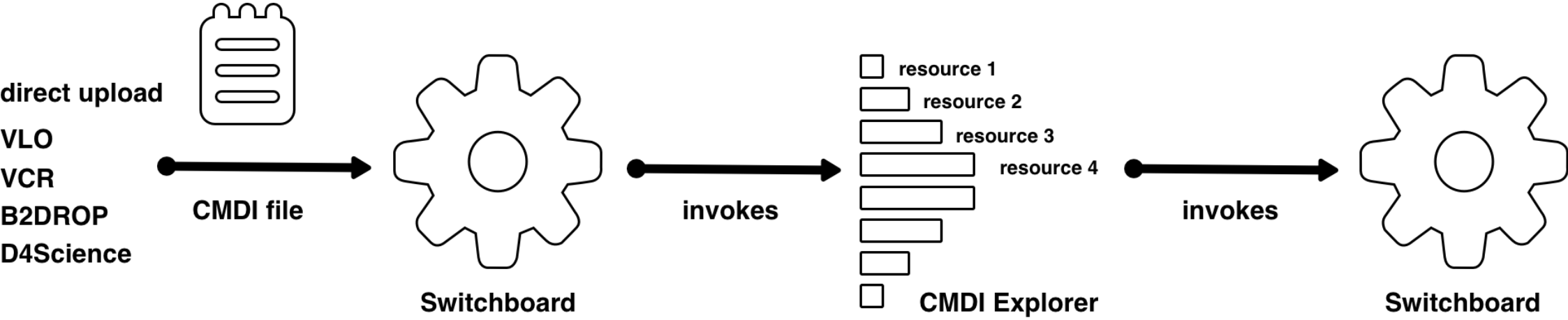
EBERHARD KARLS UNIVERSITÄT TÜBINGEN



https://www.deutschestextarchiv.de/book/show/adler_studie_1907

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Networking



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Tool Space

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

	Task	Deploymer	Software T	Name	Logo	Homepage	Location	Creators	Contact: Pe	Contact: E-	Version	Authen
1	Constituency Parsing	production	qualitative	WebLicht-Const-Parsing-DE	weblicht.jpg	https://weblicht.sfs.uni-tuebingen.de/weblicht	Tuebingen,	CLARIN-D Centre	CLARIN W	wlsupport@s	v1.0	Requir
2	Constituency Parsing	production	qualitative	WebLicht-Const-Parsing-EN	weblicht.jpg	https://weblicht.sfs.uni-tuebingen.de/weblicht	Tuebingen,	CLARIN-D Centre	CLARIN W	wlsupport@s	v1.0	Requir
3	Coreference Resolution	production	qualitative	Concraft -> Bartek	zil.png	http://zil.ipipan.waw.pl/Bartek	Warsaw, Po	Institute of Comp	MultiServic	rjawor@amu	v1.0	no
4	Dependency Parsing	developme	qualitative	Alpino	alpino.jpg	http://www.let.rug.nl/vannoord/alp/Alpino/	Nijmegen,	Maarten van Gom	Maarten va	proycon@an	null	Yes. Be
5	Dependency Parsing	production	qualitative	MaltParser	clarin-pl.png	https://ws.clarin-pl.eu	Wrocław, P	Clarin-PL	Tomasz Wa	tomasz.walk	1.0	null
6	Dependency Parsing	production	qualitative	Spacy (hosted by D4Science) - DE", "lifecycle:	d4science.png	https://www.d4science.org	Pisa, Italy	D4Science staff	D4Science	switchboard	v1.0	no
7	Dependency Parsing	production	qualitative	Spacy (hosted by D4Science) - EN", "lifecycle:	d4science.png	https://www.d4science.org	Pisa, Italy	D4Science staff	D4Science	switchboard	v1.0	no
8	Dependency Parsing	production	qualitative	UDPipe	lindat.jpg	http://ufal.mff.cuni.cz/udpipe	Charles Un	Milan Straka, Jana	Milan Strak	straka@ufal.	v1.2	no
9	Dependency Parsing	production	qualitative	WebLicht-Dep-Parsing-DE	weblicht.jpg	https://weblicht.sfs.uni-tuebingen.de/weblicht	Tuebingen,	CLARIN-D Centre	CLARIN W	wlsupport@s	v1.0	Requir
10	Dependency Parsing	production	qualitative	WebLicht-Dep-Parsing-EN	weblicht.jpg	https://weblicht.sfs.uni-tuebingen.de/weblicht	Tuebingen,	CLARIN-D Centre	CLARIN W	wlsupport@s	v1.0	Requir
11	Dependency Parsing	production	qualitative	WebLicht-Dep-Parsing-HR (RELDI)	weblicht.jpg	https://weblicht.sfs.uni-tuebingen.de/weblicht	Tuebingen,	CLARIN-D Centre	CLARIN W	wlsupport@s	v1.0	Requir
12	Dependency Parsing	production	qualitative	WebLicht-Dep-Parsing-NL-ALPINO	weblicht.jpg	https://weblicht.sfs.uni-tuebingen.de/weblicht	Tuebingen,	CLARIN-D Centre	CLARIN W	wlsupport@s	v1.0	Requir
13	Dependency Parsing	production	qualitative	WebLicht-Dep-Parsing-SL (RELDI)	weblicht.jpg	https://weblicht.sfs.uni-tuebingen.de/weblicht	Tuebingen,	CLARIN-D Centre	CLARIN W	wlsupport@s	v1.0	Requir
14	Dependency Parsing	production	qualitative	WebLicht-Dep-Parsing-SR (RELDI)	weblicht.jpg	https://weblicht.sfs.uni-tuebingen.de/weblicht	Tuebingen,	CLARIN-D Centre	CLARIN W	wlsupport@s	v1.0	Requir
15	Dependency Parsing	production	qualitative	Concraft -> DependencyParser	zil.png	http://zil.ipipan.waw.pl/DependencyParser	Warsaw, Po	Institute of Comp	MultiServic	rjawor@amu	v1.0	no
16	Distant Reading	production	quantitative	Voyant Tools	voyant-tools.jpg	http://voyant-tools.org	Canada (Qu	Stéfan Sinclair (M	Unknown F	switchboard	v2.1	no
17	Extraction of Polish term	production	quantitative	TermoPL	clarin-pl.png	https://ws.clarin-pl.eu	Wrocław, P	Clarin-PL	Tomasz Wa	tomasz.walk	1.0	null
18	Inclusion detection	production	quantitative	Inkluz	clarin-pl.png	https://ws.clarin-pl.eu	Wrocław, P	Clarin-PL	Tomasz Wa	tomasz.walk	1.0	null
19	Keyword Extractor	production	quantitative	ReSpa	clarin-pl.png	https://ws.clarin-pl.eu	Wrocław, P	Clarin-PL	Tomasz Wa	tomasz.walk	1.0	null
20	Lemmatization	production	qualitative	CSTLemma (hosted by D4Science)	d4science.png	https://www.d4science.org	Pisa, Italy	Bart Jongejan (toc	D4Science	switchboard	v1.0	no
21	Lemmatization	production	qualitative	WebLicht-Lemmas-DE	weblicht.jpg	https://weblicht.sfs.uni-tuebingen.de/weblicht	Tuebingen,	CLARIN-D Centre	CLARIN W	wlsupport@s	v1.0	Requir
22	Lemmatization	production	qualitative	WebLicht-Lemmas-EN	weblicht.jpg	https://weblicht.sfs.uni-tuebingen.de/weblicht	Tuebingen,	CLARIN-D Centre	CLARIN W	wlsupport@s	v1.0	Requir
23	Machine Translation	developme	qualitative	Oersetter (FRY-NLD)	oersetter.jpg	http://oersetter.nl/	Nijmegen,	Maarten van Gom	Maarten va	proycon@an	null	Yes. Be
24	Machine Translation	developme	qualitative	Oersetter (NLD-FRY)	oersetter.jpg	http://oersetter.nl/	Nijmegen,	Maarten van Gom	Maarten va	proycon@an	null	Yes. Be
25	Morpho-syntactic tagger	production	qualitative	MorphoDiTa	clarin-pl.png	https://ws.clarin-pl.eu	Wrocław, P	"Clarin-PL"],	Tomasz Wa	tomasz.walk	1.0	no
26	Morpho-syntactic tagger	production	qualitative	Tagger NLTK	clarin-pl.png	https://ws.clarin-pl.eu	Wrocław, P	Clarin-PL	Tomasz Wa	tomasz.walk	1.0	null
27	Morpho-syntactic tagger	production	qualitative	WCRFT2	clarin-pl.png	http://nlp.pwr.wroc.pl/redmine/projects/wcrft	Wrocław, P	Clarin-PL	Tomasz Wa	tomasz.walk	1.0	null
28	Morphological Analysis	production	qualitative	Morfeusz 2	clarin-pl.png	https://ws.clarin-pl.eu	Wrocław, P	Clarin-PL	Tomasz Wa	tomasz.walk	1.0	null
29	Morphological Analysis	production	qualitative	WebLicht-Morphology-DE	weblicht.jpg	https://weblicht.sfs.uni-tuebingen.de/weblicht	Tuebingen,	CLARIN-D Centre	CLARIN W	wlsupport@s	v1.0	Requir
30	Morphological Analysis	production	qualitative	WebLicht-Morphology-EN	weblicht.jpg	https://weblicht.sfs.uni-tuebingen.de/weblicht	Tuebingen,	CLARIN-D Centre	CLARIN W	wlsupport@s	v1.0	Requir
31	N-Gramming	developme	quantitative	Colibri Core (folia+xml)	colibriCore.jpg	https://github.com/proycon/colibri-core	Nijmegen,	Maarten van Gom	Maarten va	proycon@an	null	Yes. Be

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Tool Name	Task	Language(s)	Mime-type(s)
Ucto	Tokenization	nld, eng, deu, fra, ita, fry	text/plain
CST Tokenizer	Tokenization	bul, ces, dan, deu, ell, eng, spa, est, fas, fra, hun, isl, ita, lat, mkd, nld, pol, por, ron, rus, slk, slv, srp, tur, ukr	many formats
CST Lemmatizer	Lemmatization	see above	many formats
CST Name Recognizer	Named Entity Recognition	dan	many formats
espeak	Voice Synthesis	many languages	many formats
Tesseract — CuneiForm	OCR Engine	dan, eng, generic	image/gif, image/jpeg, image/png, image/tiff
WebLicht-Lemmas-EN	Lemmatization	eng	text/plain, text/rdf, application/pdf, application/msword, others
WebLicht-Tokenization-TK	Tokenization	tur	see above
WebLicht-Lemmas-DE	Lemmatization	deu	see above
WebLicht-POSTags-Lemmas-DE	PoS Tagging	deu	see above
WebLicht-POSTags-Lemmas-IT	PoS Tagging	ita	see above
WebLicht-POSTags-Lemmas-EN	PoS Tagging	eng	see above
WebLicht-Dep-Parsing-NL	Dependency Parsing	nld	see above
WebLicht-Dep-Parsing-EN	Dependency Parsing	eng	see above
WebLicht-Morphology-DE	Morphology Analysis	deu	see above
WebLicht-Morphology-EN	Morphology Analysis	eng	see above
WebLicht-NamedEntities-DE	Named Entity Recognition	deu	see above
WebLicht-NamedEntities-EN	Named Entity Recognition	eng	text/plain
Voyant Tools	Text Analytics	eng, deu, spa, nld, fra, generic	text/plain, test/rdf, application/pdf, application/doc, others
T-scan	Text Analytics	nld	text/plain
Oersetter (NLD-FRY)	Machine Translation	nld	text/plain
Oersetter (FRY-NLD)	Machine Translation	fry	text/plain
Fowlt	Spelling correction	eng	text/plain, text/folia+xml
Frog	NLP suite for Dutch	nld	text/plain, text/folia+xml
FoLiA-stat	N-Gramming	nld, generic	text/folia+xml
Valkuil	Spelling correction	nld	text/plain, text/folia+xml
Colibri Core	N-Gramming	nld, eng, deu, fre, esp, por, fry, generic	text/plain, text/folia+xml
Alpino	Dependency Parsing	nld	text/plain

Tool Space

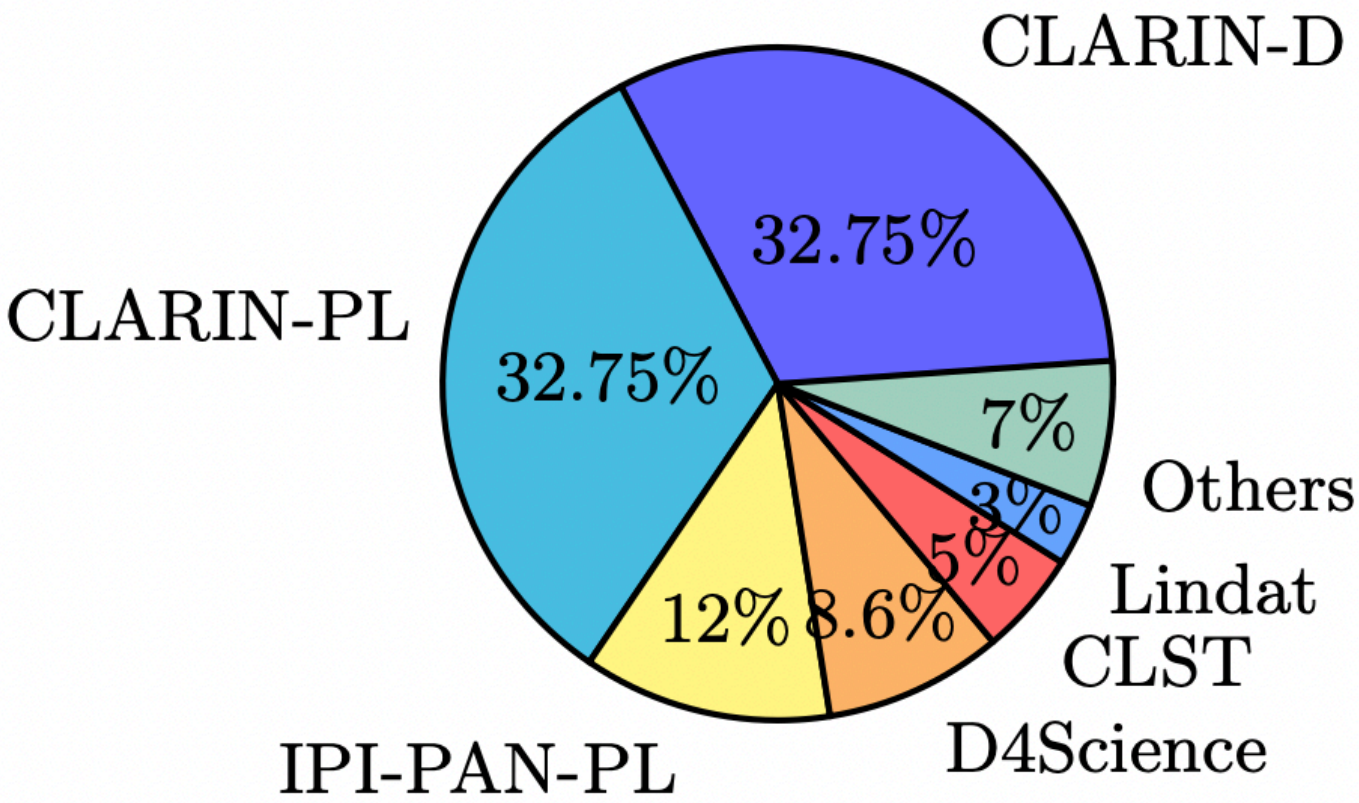


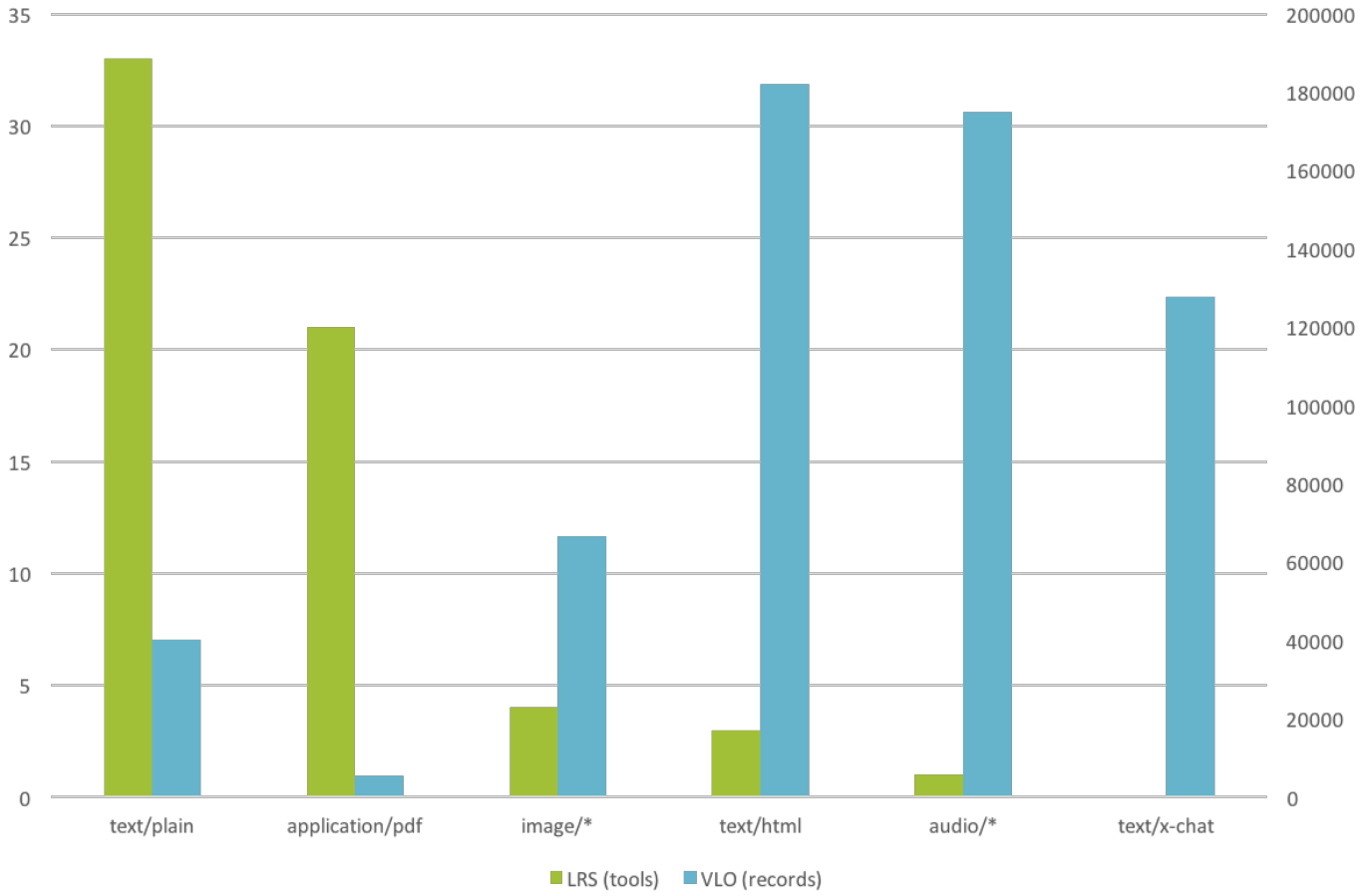
Fig. 5: Switchboard's toolset, sorted by consortiums.

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Most common resource languages (VLO) vs tool support (LRS)

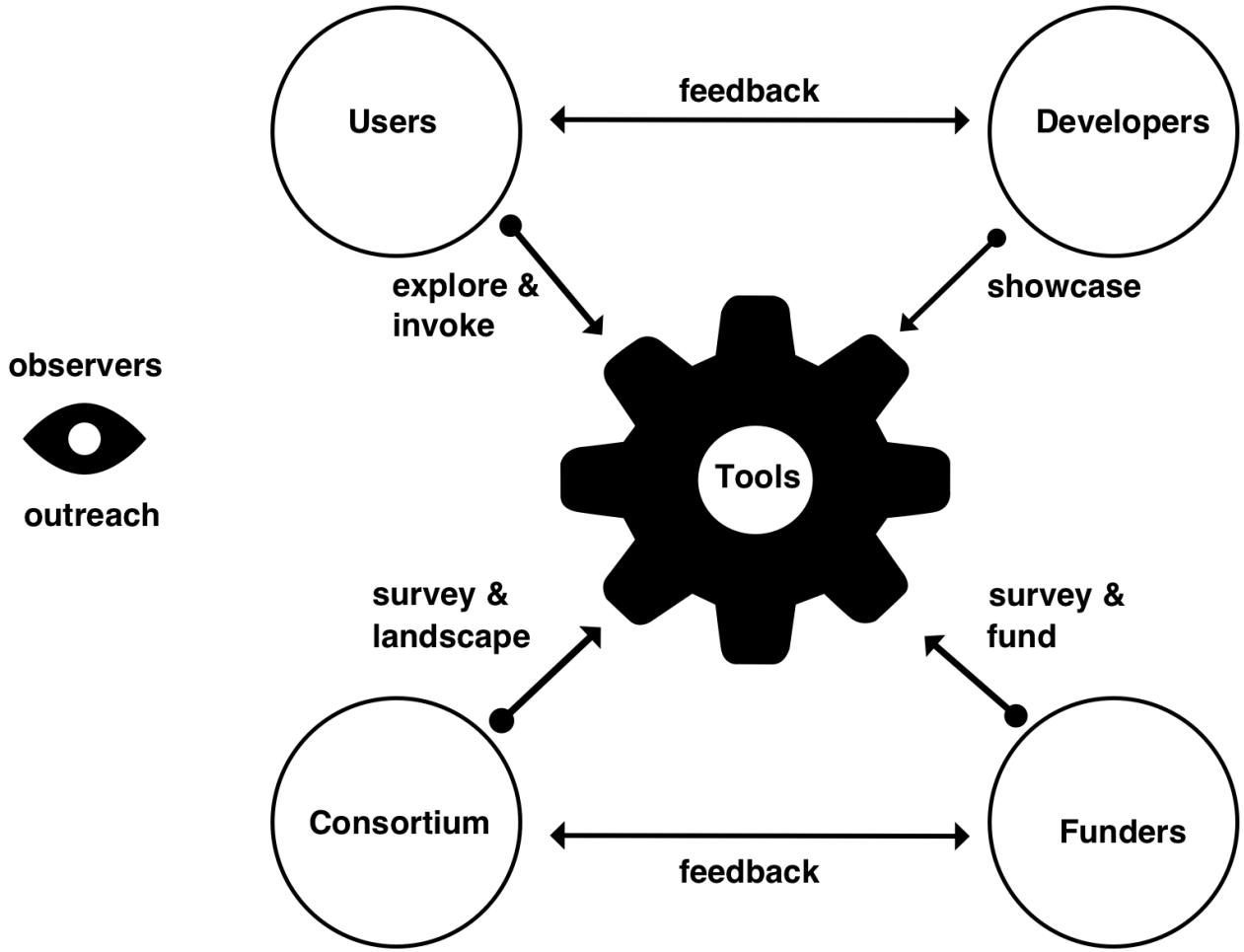
	Tokenisation	Lemmatisation	POS tagging	Constituency Parsing	Dependency Parsing	Named Entity	OCR	Speech recognition
Dutch								
Danish								
English								
German								
French								
Latin								
Spanish; Castilian								
Japanese								
Arabic								
Chinese								
Turkish								
Farsi; Persian								
Indonesian								
Italian								
Portuguese								
Polish								
Russian								
Javanese								
Catalan; Valencian								
Czech								
Estonian								

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Impact



EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Publications

- Zinn, C. 2016. The CLARIN Language Resource Switchboard. In *Proceedings of the CLARIN Annual Conference*, Aix-en-Provence, France.
- Zinn, C. 2018. A Bridge from EUDAT's B2DROP cloud service to CLARIN's Language Resource Switchboard. *Linköping Electronic Press*, (147):36–45.
- Zinn, C. 2018. The Language Resource Switchboard. Computational Linguistics, 44(4), pages 631-639, December 2018.
- Zinn, C. and Dima, E. The CLARIN Language Resource Switchboard. CLARIN Book 2022 (Ed. by Darja Fišer and Andreas Witt), DeGruyter, 2022 (to appear).

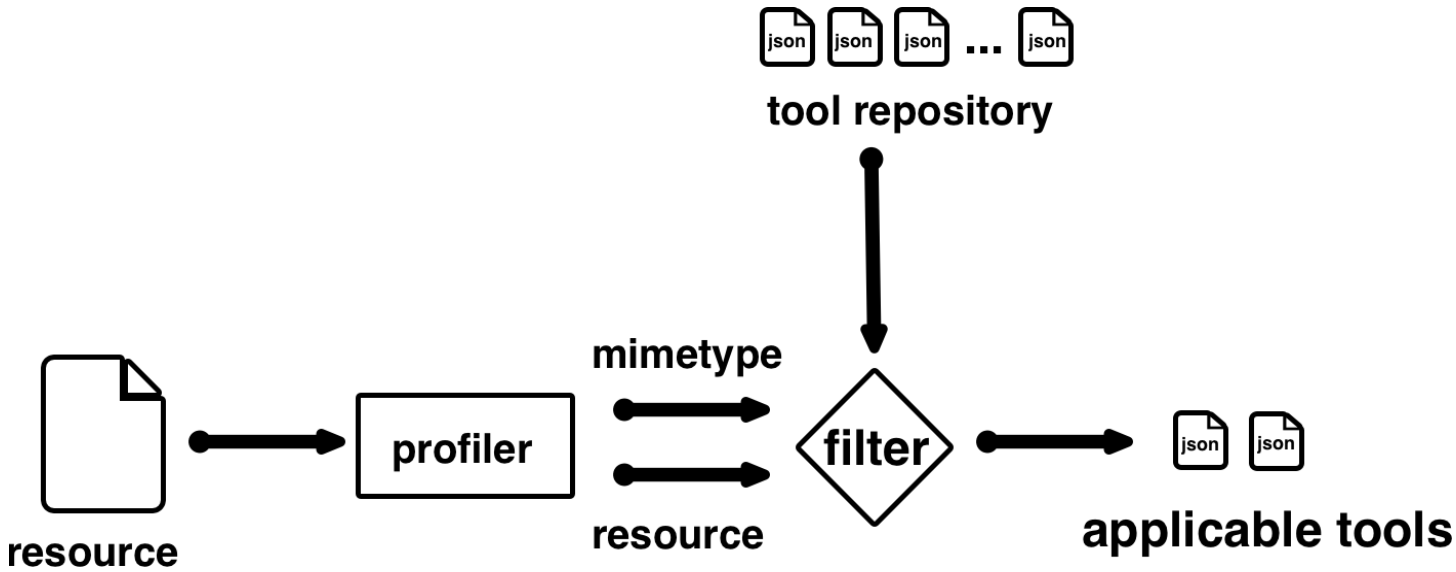
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Discussion

- Difference between Switchboard and WebLicht
  - Tool has GUI
  - Tool understands TCF
  - …
- Gap analysis in tool space
- Switchboard in other domains?
- What does the future hold?

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Technical Aspects



Same technical concept & backbone (Java + ReactJS) from the very start, but substantial improvements in both code quality and deployment workflows.