

Die Metadateninfrastruktur des CLARIN-Projektes

Thomas Eckart

Abteilung Automatische Sprachverarbeitung

Universität Leipzig, Deutschland

teckart@informatik.uni-leipzig.de

UNIVERSITÄT LEIPZIG

1. Forschungsinfrastruktur CLARIN
2. Metadaten in CLARIN (CMDI)
3. Erstellungsprozess von Metadaten
4. Bereitstellung und Aufbereitung von Metadaten in föderierten Infrastrukturen
5. Problemfälle und Lösungsansätze

Metadaten?

Metadaten [...] sind Daten, die Informationen über Merkmale anderer Daten enthalten, aber nicht diese Daten selbst.

Interoperable Metadaten sind [...] Metadaten aus potenziell unterschiedlichen Quellen, zwischen denen eine Beziehung in der Weise besteht, dass mit ihnen gemeinsam gearbeitet werden kann.

(<http://de.wikipedia.org/wiki/Metadaten>)

Wozu?

FRBR User Tasks:

- *Find:* „meeting a user's search criteria“
- *Identify/Select:* „confirm they have found what they looked for“, „meeting a user's requirements with respect to content, physical format, etc.“
- *Obtain:* „enables to acquire an entity“
- *(Navigate)*

(<https://www.loc.gov/cds/downloads/FRBR.PDF>)

Schema vs. Instanz

Bauplan vs. Konkrete Ausprägung

Person vs. „Thomas Eckart“, ...

1. CLARIN

Eine Forschungsinfrastruktur für Sprachressourcen



CLARIN-D

CLARIN-D – Eine web- und zentrenbasierte Forschungsinfrastruktur für die Geistes- und Sozialwissenschaften

Linguistische Daten, Werkzeuge und Dienste werden bereitgestellt ...

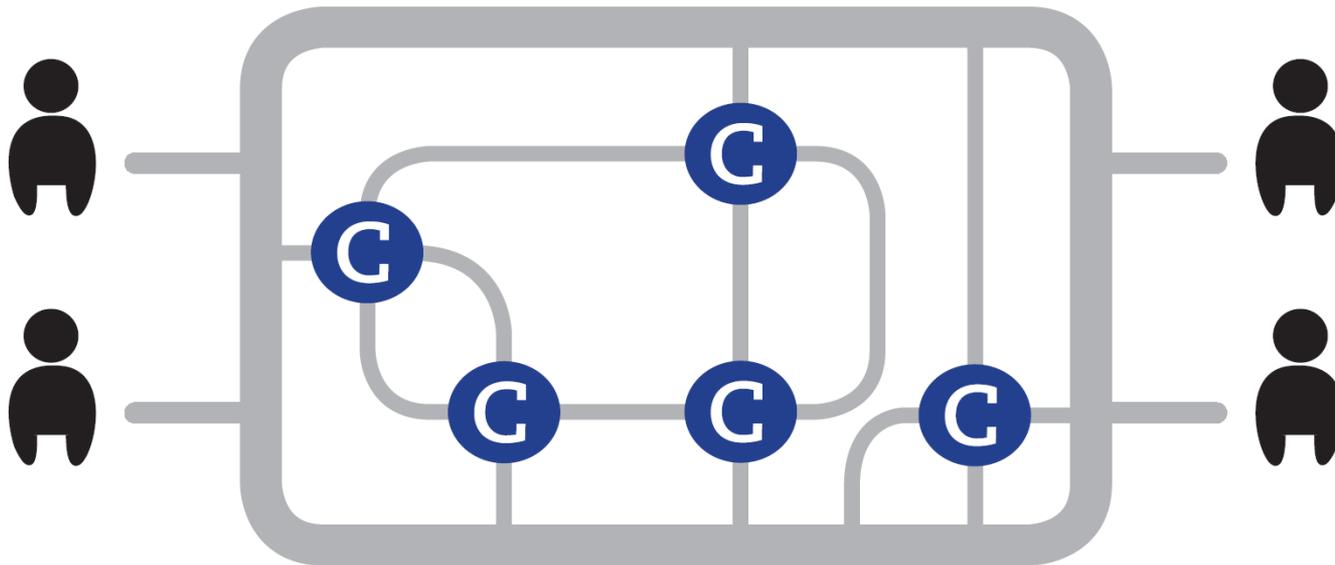
- In einer integrierten, interoperabel und skalierbaren Infrastruktur
- Entwickelt für und in enger Zusammenarbeit mit den Geistes- und Sozialwissenschaften
- Gefördert vom Bundesministerium für Bildung und Forschung (BMBF)
- CLARIN ist eine europäische Infrastruktur-Initiative mit zur Zeit 15+ Mitgliedsländern
- Aktuelle Förderung bis September 2020
- Website: <http://de.clarin.eu>



- Textkorpora / Treebanks / Wörterbücher
- Audio (+Transkriptionen)
- Multimodale Daten (z.B. Video + Audio + Gesten)
- Werkzeuge: Annotation, Visualisierungen
- Services: SOAP/RESTful Webservices
- Beschreibung interner Projektstrukturen

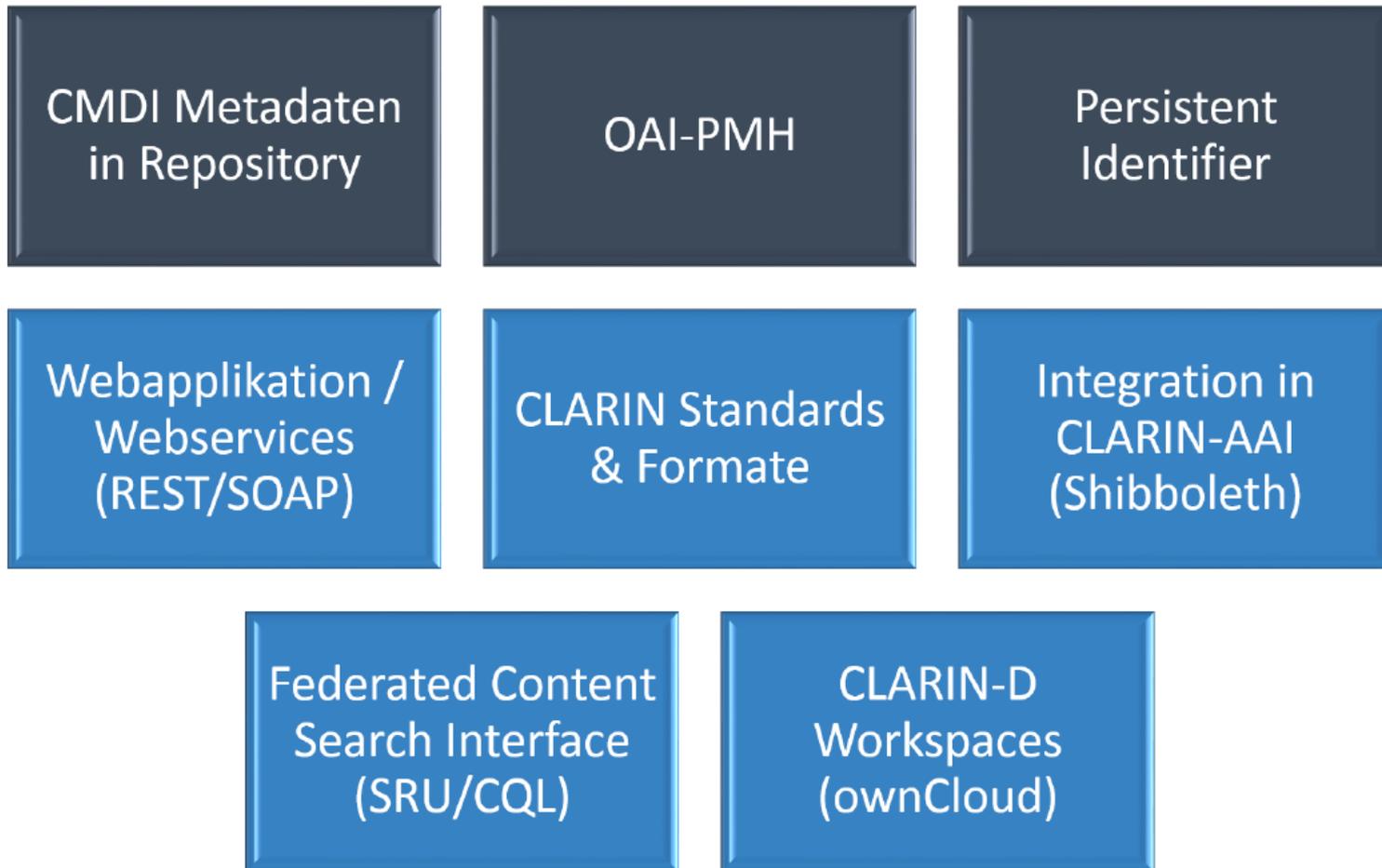
- Keine monolithische Architektur
- Stattdessen: verteilte Architektur / föderiertes System

services to researcher



VLO -> FCS -> WebLicht

- Zertifizierungsprozess für Zentren
- u.a.
 - Kontaktpersonen (administrativ/technisch)
 - Data Seal of Approval Assessment
 - Verwendung persistenter Identifier (PIDs)
 - ...
 - **Beschreibung vorliegender Ressourcen über Metadaten**



2. Component **MetaData** Infrastructure

Metadaten in CLARIN

- Beispiel: Beschreibung eines „texttragenden Objektes“

Beispiel: Aspekte der Beschreibung eines „texttragenden Objektes“

- Autor
- Titel/Name
- Beschreibung
- Sprache
- Identifier (ISBN, Handle, DOI, ...)
- Zugriffsort (Regalnummer, URI)
- Zeitangaben: Publikation, Änderungen
- Schlüsselwörter/Keywords
- Bereitstellende Institution/Person
- Kontaktdaten
- Angaben zur Größe (Wörter, Sätze, Fläche, Gewicht, ...)
- Angaben zum Zustand
- usw. usf.

Beispiel: Aspekte der Beschreibung eines „texttragenden **Objektes**“

- ~~Author~~ → Ersteller
- Titel/Name
- Beschreibung
- Sprache
- Identifier (ISBN, Handle, DOI, ...)
- Zugriffsort (Regalnummer, URI)
- Zeitangaben: Publikation, Änderungen
- Schlüsselwörter/Keywords
- Bereitstellende Institution/Person
- Kontaktdaten
- Angaben zur Größe (Wörter, Sätze, Fläche, Gewicht, ...)
- Angaben zum Zustand
- usw. usf.

- CMDI: Allgemeines Framework zur Beschreibung von Metadatenformaten
- Metaformat zur Spezifikation von MD-Schemata (ISO 24622-1:2015)
- Ressourcentypen-übergreifend

- Grundideen (1):
 - Alle Ressourcen sollten beliebig (fein-)granular beschreiben werden können
 - Alle Ressourcen(-typen) können durch verschiedene Aspekte beschrieben werden
 - Aspekte können Ressourcentypen-spezifisch sein (sind es aber häufig nicht)
 - Wiederverwendbarkeit ist gut
 - Spart Zeit
 - Erlaubt Berücksichtigung von Expertenwissen
 - Verringert Fehler

Beispiel: Aspekt *Kontaktdaten*

Personenname
Postadresse
Email-Adresse
Organisation
Telefonnummer
Webseite



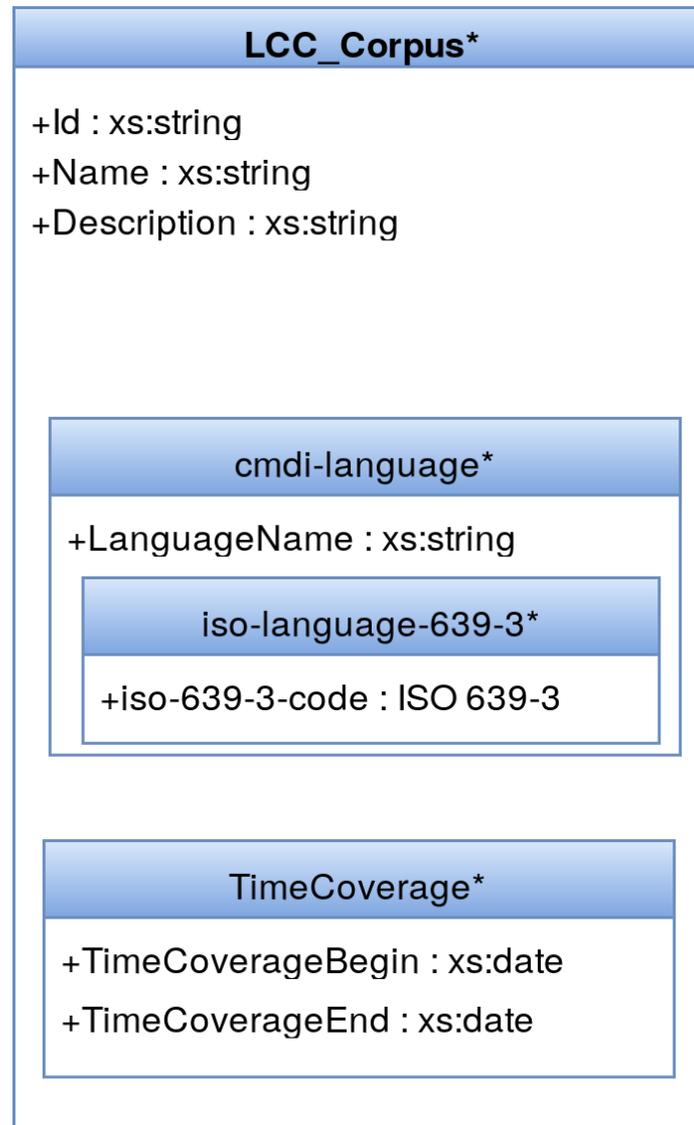
In CMDI:

- 1 Aspekt ↔ 1 Komponente
- Komponente = „Bauplan“ / Template / Schema

- Grundideen (2):
 - Jeder (im Prinzip) kann solche Komponenten erstellen
 - Gespeichert an zentralem Ort
 - Ziel: beliebig detaillierte Beschreibung eigener Ressourcentypen

- Ziele:
 - Formale Interpretierbarkeit
 - Wiederverwendbarkeit
 - Verständlichkeit
 - Austauschbarkeit
 - Ausnutzung von Expertenwissen
 - Effizientere Arbeitsprozesse
 - Offenheit

- Bündelung aller Aspekte eines Ressourcentypen in einem **Profil**
- Profil = vollständiger Bauplan zur Beschreibung einer Klasse von Ressourcen

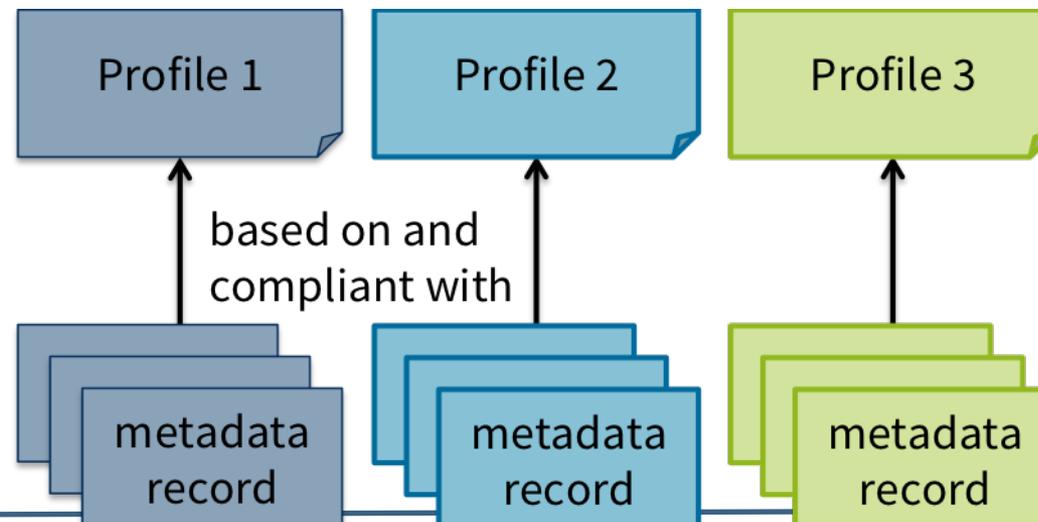


- Zusammenfassung:

Jeder Ressourcentyp = 1 Profil

Profil besteht aus beliebig vielen Komponenten

Profil dient zur Beschreibung beliebig vieler Instanzen (Ressourcen)



3. Erstellungsprozess von Metadaten in der CMDI

- Erstellung von Metadaten
 1. Erstellen bzw. Finden eines passenden Schemas
 - zentrale Registratur (Freie Lesezugriff, Login für Schreibzugriff)
 2. Erstellen konkreter Instanzen als Beschreibung konkreter Objekte
 - dezentral (zentrenabhängig)

- Component Registry (Testversion)
 - ComponentRegistry

CMDI Component Registry  

       Showing 100 of 100 

Name	Group Name	Domain Name	Creator	Description	Registrati...	Com...
AnnotatedCorpusProfile	CLARIN		nalida	A CMDI profile for annotated...	2013-01-31	0
AnnotatedCorpusProfile-DLU	DLU		GrietDepoorter	A CMDI profile for annotated...	2013-10-24	0
AnnotationTool			Eric Sanders	Description of a tooladapted...	2011-02-09	0
ArthurianFiction		Other	Rik Hoekstra	Profile for Arthurian Fiction ...	2012-09-04	0
Bamdes_LexicalResource		Computational Li...	Dieter Van Uytva...	Lexical Resource as used by ...	2010-10-27	0

Sofern Ihre Institution nicht an diesen Dienst angeschlossen ist, können Sie ein neues Benutzerkonto bei einem der Zugangsanbieter für Gäste erstellen. Sie können auch Ihre Institution kontaktieren um einen Anschluss an diesen Dienst vorzuschlagen.

 clarin.eu website account (New)
 European Union 

Email:

Password:

[Forgotten password?](#)

Authenticate

Cancel

- Instanzerstellung
 - Manuell
 - XML-Editoren (Oxygen, XMLSpy,...)
 - Web-Editoren (z.B. [COMEDI](#))
 - „Automatisch“
 - Transformationen aus bestehenden Metadaten (z.B. XSLT)
 - XML-Bibliotheken (z.B. XStream, JAXB)

- Beispiel für Instanzdatei eines Textkorpus:
 - Name: *deu_news_2010_1M*
 - Identifier: *hdl:11022/0000-0000-20E0-E*
 - 1.000.000 Sätze deutscher Zeitungstext von 2010

<http://hdl.handle.net/11022/0000-0000-20E0-E>

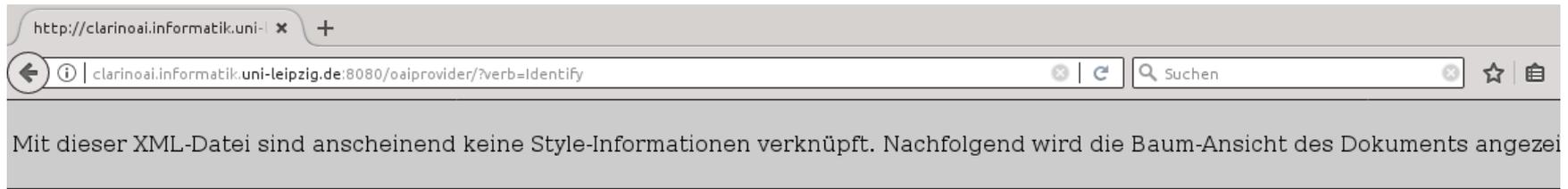
4. Bereitstellung und Aufbereitung von Metadaten in förderierten Infrastrukturen

- Metadateninstanzen liegen lokal vor – was nun?
- Rechtliche Unterscheidung von Ressourcen und ihrer Metadaten
- Ziel: freie Verteilung über Standardprotokolle
- Populärer Standard: OAI-PMH

- OAI: **O**pen **A**rchives **I**nitiative
- PMH: **P**rotocol for **M**etadata **H**arvesting
- Einfaches maschinenlesbares Protokoll
- Abfrage von XML-Dokumenten die lokalen Ressourcen-Bestand beschreiben

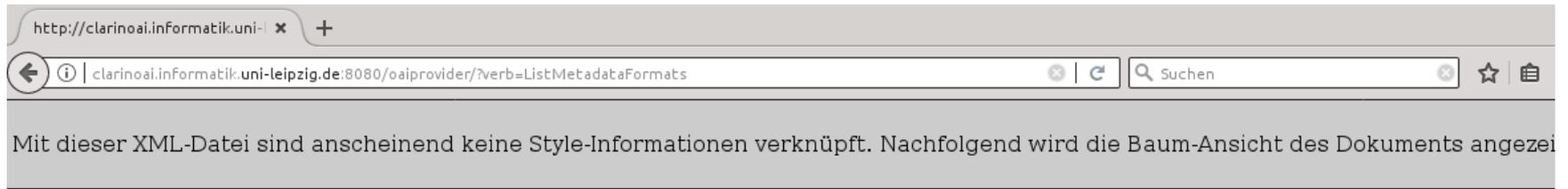
- Verschiedene Anfragen („verbs“) an einen OAI-PMH-Endpunkt:
 - Identify
 - ListMetadataFormats
 - ListIdentifiers
 - GetRecord
- Aufruf: *http://ENDPUNKT_URL?verb=**VERB***
- *http://clarinoai.informatik.uni-leipzig.de:8080/oaiprovider/*

verb=Identify



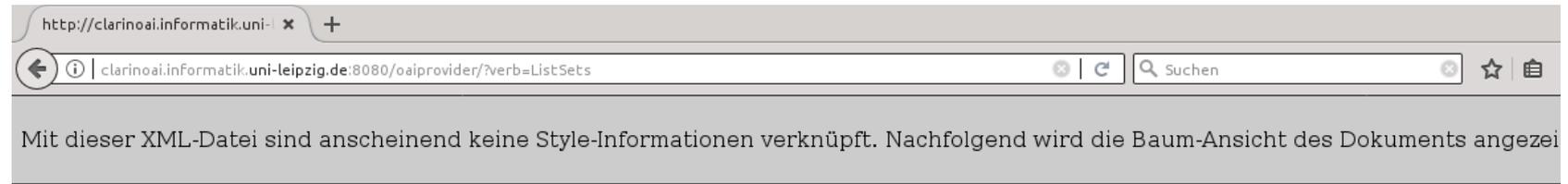
```
-<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2017-05-24T11:51:55Z</responseDate>
  -<request verb="Identify">
    http://clarinoai.informatik.uni-leipzig.de:8080/oaiprovider/
  </request>
  -<Identify>
    -<repositoryName>
      CLARIN OAIProvider - Institute of Computer Science - NLP Group - University of Leipzig
    </repositoryName>
    -<baseURL>
      http://clarinoai.informatik.uni-leipzig.de:8080/oaiprovider
    </baseURL>
    <protocolVersion>2.0</protocolVersion>
    <adminEmail>clarin@informatik.uni-leipzig.de</adminEmail>
    <earliestDatestamp>2012-05-31T00:00:00Z</earliestDatestamp>
    <deletedRecord>no</deletedRecord>
    <granularity>YYYY-MM-DD</granularity>
  -<description>
    -<friends xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/friends/ http://www.openarchives.org/OAI/2.0/friends.xsd">
      <baseURL>http://arXiv.org/oa2</baseURL>
      <baseURL>http://memory.loc.gov/cgi-bin/oa2_0</baseURL>
    </friends>
  </description>
```

verb=**ListMetadataFormats**



```
-<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2017-05-24T11:52:39Z</responseDate>
  -<request verb="ListMetadataFormats">
    http://clarinoai.informatik.uni-leipzig.de:8080/oaiprovider/
  </request>
  -<ListMetadataFormats>
    -<metadataFormat>
      <metadataPrefix>oai_dc</metadataPrefix>
      <schema>http://www.openarchives.org/OAI/2.0/oai_dc.xsd</schema>
      <metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc/</metadataNamespace>
    </metadataFormat>
    -<metadataFormat>
      <metadataPrefix>cmdi</metadataPrefix>
      -<schema>
        https://infra.clarin.eu/CMDI/1.x/xsd/cmd-envelop.xsd
      </schema>
      <metadataNamespace>http://www.clarin.eu/cmd/1 </metadataNamespace>
    </metadataFormat>
  </ListMetadataFormats>
</OAI-PMH>
```

verb=ListSets



```
-<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2017-06-06T11:19:42Z</responseDate>
  -<request verb="ListSets">
    http://clarinoai.informatik.uni-leipzig.de:8080/oaiprotider/
  </request>
  -<ListSets>
    -<set>
      <setSpec>oai:wortschatz</setSpec>
      <setName>oai:wortschatz</setName>
    </set>
    -<set>
      <setSpec>oai:webservices</setSpec>
      <setName>oai:webservices</setName>
    </set>
    -<set>
      <setSpec>oai:cts.pbc</setSpec>
      <setName>oai:cts.pbc</setName>
    </set>
    -<set>
      <setSpec>oai:cts.muqtabas</setSpec>
      <setName>oai:cts.muqtabas</setName>
    </set>
    -<set>
      <setSpec>oai</setSpec>
      <setName>oai</setName>
    </set>
  </ListSets>
</request>
</OAI-PMH>
```

verb=ListIdentifiers



```
-<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2017-05-24T11:53:01Z</responseDate>
  -<request verb="ListIdentifiers" metadataPrefix="cmdi">
    http://clarinoai.informatik.uni-leipzig.de:8080/oaiprovider/
  </request>
  -<ListIdentifiers>
    -<header>
      <identifier>oai:asv.corpus:110220000000020E0E</identifier>
      <datestamp>2017-05-08T13:12:01Z</datestamp>
      <setSpec>oai:wortschatz</setSpec>
    </header>
    -<header>
      <identifier>oai:asv.corpus:11022000000008F185</identifier>
      <datestamp>2017-05-08T13:12:02Z</datestamp>
      <setSpec>oai:wortschatz</setSpec>
    </header>
    -<header>
      <identifier>oai:corpus:110220000000020DE2</identifier>
      <datestamp>2017-05-08T13:12:03Z</datestamp>
      <setSpec>oai:wortschatz</setSpec>
    </header>
    -<header>
      <identifier>oai:corpus:110220000000020DF1</identifier>
```

verb=**GetRecord** (1/2)

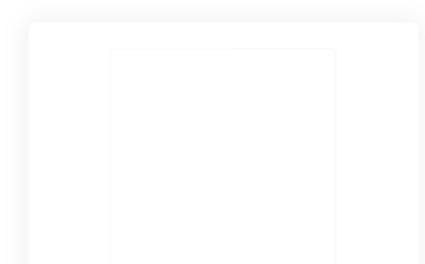


```
-<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2017-05-24T11:53:01Z</responseDate>
  -<request verb="ListIdentifiers" metadataPrefix="cmdi">
    http://clarinoai.informatik.uni-leipzig.de:8080/oaiprovider/
  </request>
  -<ListIdentifiers>
    -<header>
      <identifier>oai:asv.corpus:110220000000020E0E</identifier>
      <datestamp>2017-05-08T13:12:01Z</datestamp>
      <setSpec>oai:wortschatz</setSpec>
    </header>
    -<header>
      <identifier>oai:asv.corpus:11022000000008F185</identifier>
      <datestamp>2017-05-08T13:12:02Z</datestamp>
      <setSpec>oai:wortschatz</setSpec>
    </header>
    -<header>
      <identifier>oai:corpus:110220000000020DE2</identifier>
      <datestamp>2017-05-08T13:12:03Z</datestamp>
      <setSpec>oai:wortschatz</setSpec>
    </header>
    -<header>
      <identifier>oai:corpus:110220000000020DF1</identifier>
```

verb=**GetRecord** (2/2)



```
<cmd:Resources></cmd:Resources>
- <cmd:Components>
- <cmdp:LCC_CorpusProfile ProfileId="clarin.eu:cr1:p_1381926654508">
- <cmdp:LCC_Corpus ComponentId="clarin.eu:cr1:c_1381926654507">
  <cmdp:Id>11022/0000-0000-20E0-E</cmdp:Id>
  <cmdp:Name>deu_news_2010_1M</cmdp:Name>
- <cmdp:Description xml:lang="eng">
  1.000.000 sentences of a German newspaper corpus based on material from 2010
  </cmdp:Description>
  <cmdp:Version>1.0</cmdp:Version>
- <cmdp:ResponsibleOrganisation>
  CLARIN-D center, Natural Language Processing Group, University of Leipzig
  </cmdp:ResponsibleOrganisation>
  <cmdp:ResourceType>Written Corpus</cmdp:ResourceType>
- <cmdp:Contacts>
- <cmdp>Contact>
  <cmdp:Person>CLARIN Center Leipzig Staff</cmdp:Person>
  <cmdp:Role>Technical contact</cmdp:Role>
  <cmdp:Email>clarin@informatik.uni-leipzig.de</cmdp:Email>
- <cmdp:Organisation>
  CLARIN-D center, Natural Language Processing Group, University of Leipzig
  </cmdp:Organisation>
  <cmdp:Url>http://clarin.informatik.uni-leipzig.de/repo</cmdp:Url>
  </cmdp>Contact>
</cmdp:Contacts>
- <cmdp:Types>
  <cmdp:Type>text/plain</cmdp:Type>
  <cmdp:Type>application/gzip</cmdp:Type>
```



Weitere Abfragemöglichkeiten:

- Liste alle Einträge im CMDI-Format auf, welche zum Set „mySet“ gehören:

PROVIDER_URL?

verb=ListRecords&metadataPrefix=cmdi&set=mySet

- Liste alle Einträge im CMDI-Format auf, welche nach dem 15.5.2017 um 8:30 hinzugefügt/geändert wurden:

PROVIDER_URL?

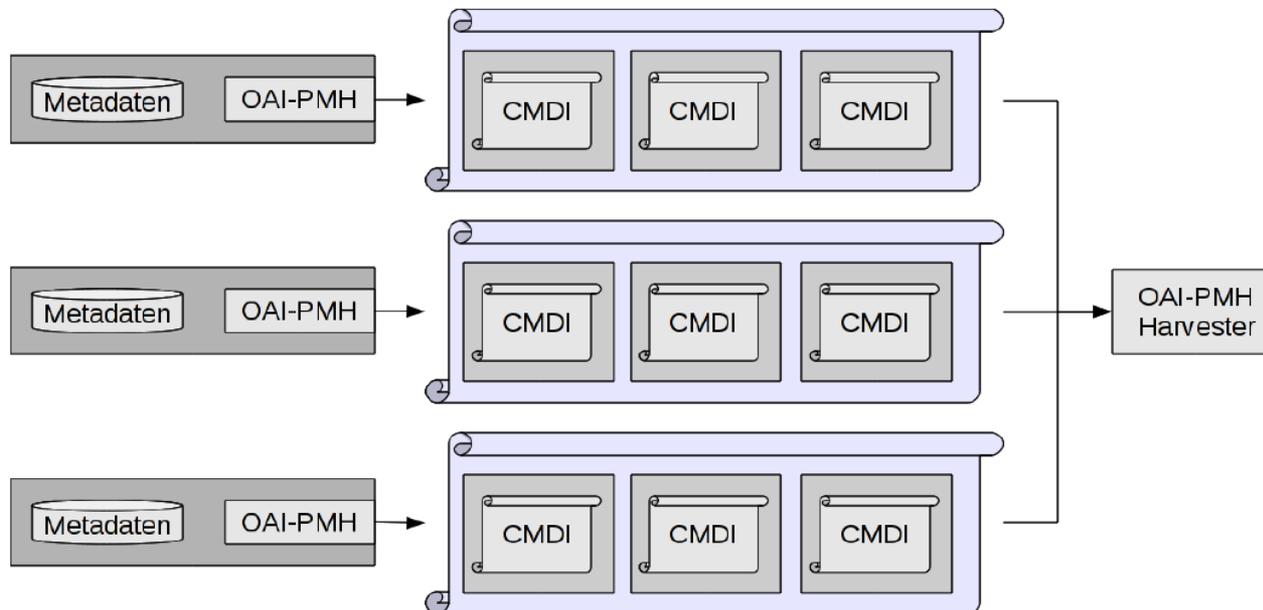
verb=ListRecords&metadataPrefix=cmdi&from=2017-05-05T08:30:00Z

Weitere Beispiele für OAI-PMH-Endpunkte:

- TextGrid: *textgridlab.org/1.0/tgoaipmh/oai*
- Europeana: *oai.europeana.eu/oaicat/*
(restricted)
- DAI: *opac.dainst.org/OAI*
- DNB: *services.dnb.de/oai/repository*
(restricted)
- *OAI-Liste (>3000 Einträge):*
www.openarchives.org/Register/BrowseSites

OAI-PMH-Harvester:

- Sammelt die dezentral in den Zentren hinterlegten Metadaten ein
- Anfrage: Welche Datensätze im CMDI-Format wurden seit der letzten Harvesting-Runde hinzugefügt?



Momentan zwei Harvester in CLARIN:

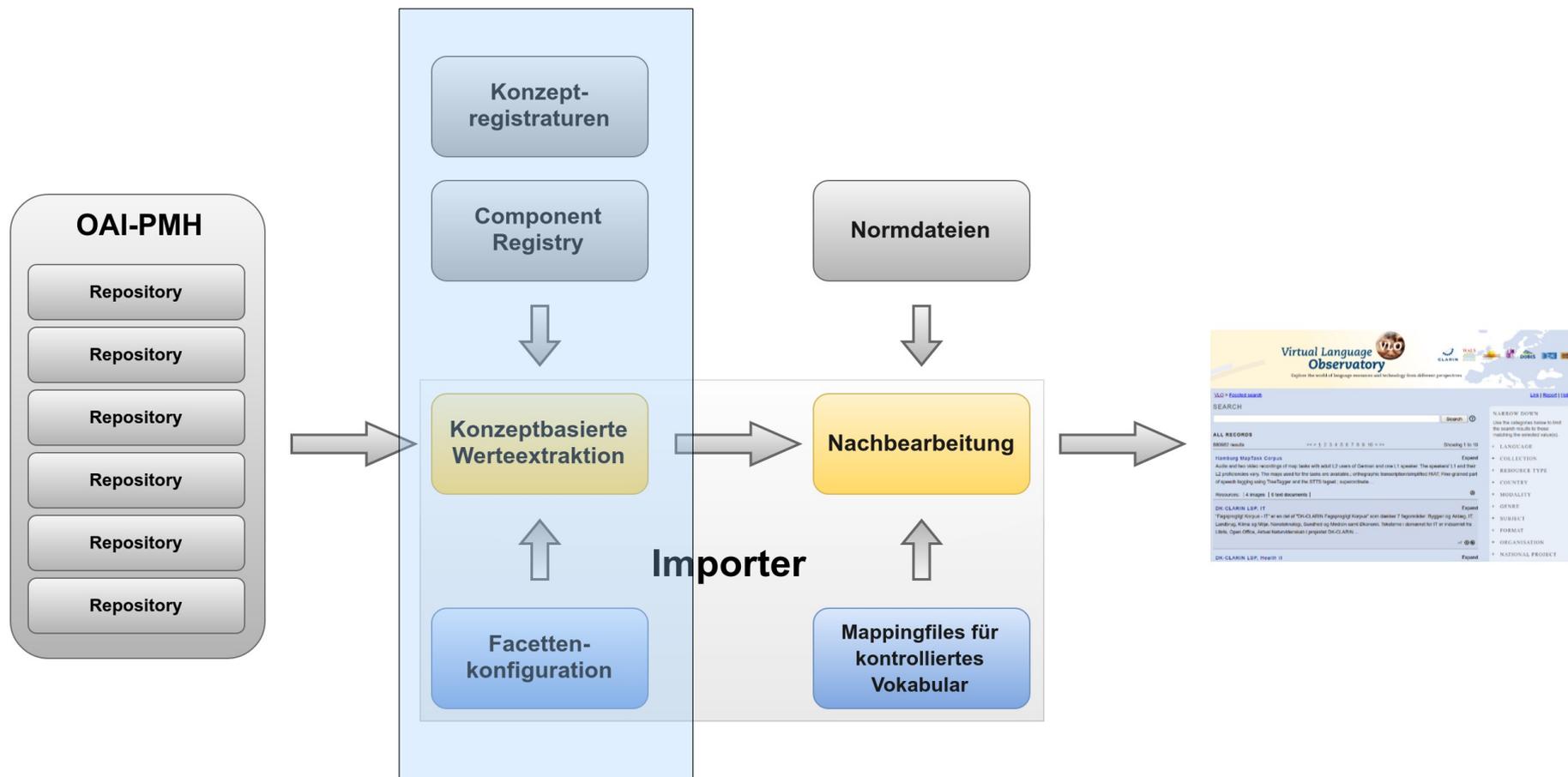
1. **Alle** Metadaten für zentrale Suchmaschine

→ automatische Integration in das VLO

2. Alle Metadaten die **Webservices** beschreiben

→ automatische Integration in Workflow-Engine WebLicht

Das Finale: Import in das VLO



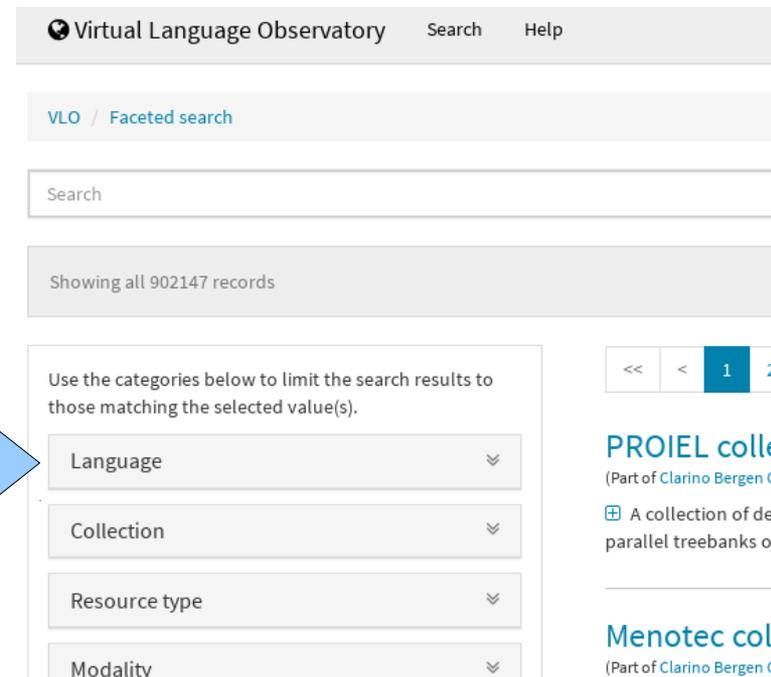
Facettendefinition „language“

Language ID

(hdl:11459/CCR_C-2482_08eded24-4086-7e3f-88e5-e0807fb01e17)

Language Name

(hdl:11459/CCR_C-2484_669684e7-cb9e-ea96-59cb-a25fe89b9b9d)



Virtual Language Observatory Search Help

VLO / Faceted search

Search

Showing all 902147 records

Use the categories below to limit the search results to those matching the selected value(s).

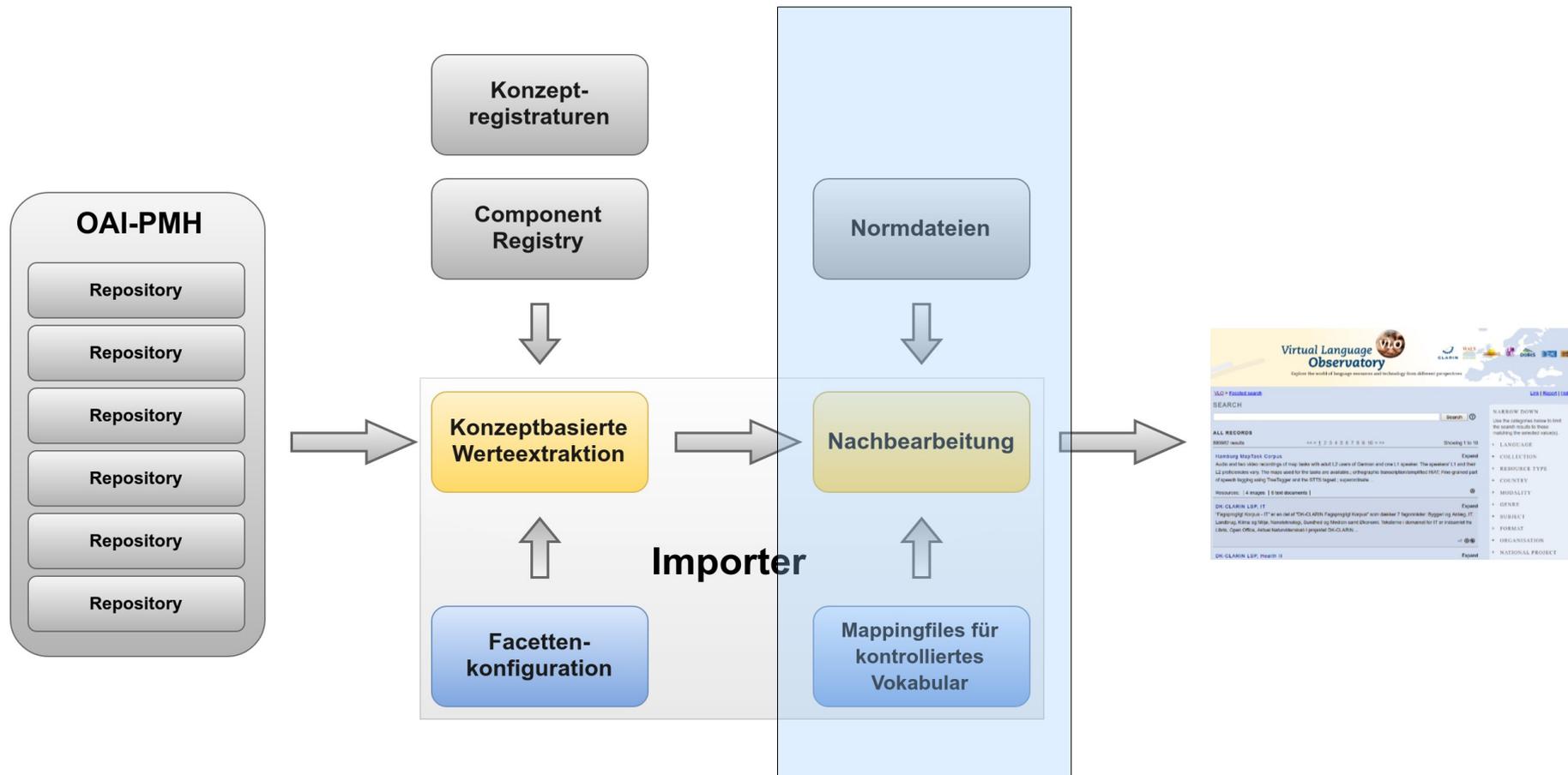
- Language
- Collection
- Resource type
- Modality

<< < 1 2

PROIEL colle
(Part of Clarino Bergen C)
A collection of de parallel treebanks of

Menotec coll
(Part of Clarino Bergen C)

Das Finale: Import in das VLO



Search



Showing all 902147 records

Results per page: 10

Use the categories below to limit the search results to those matching the selected value(s).

Language



Collection



Resource type



Modality



Format



Keyword



<< < 1 2 3 4 5 6 7 8 9 10 > >>

PROIEL collection

(Part of [Clarino Bergen Centre - INESS](#))

A collection of dependency treebanks for early attestations of Indo-European, including a set of parallel treebanks of the New Testament.



Menotec collection

(Part of [Clarino Bergen Centre - INESS](#))

The Menotec collection in INESS contains four central manuscripts in Old Norwegian (although the ISO language code is termed "Old Norse"): The Old Norwegian Homily Book in AM 619 4to (ca. 1200–1225), the legendary saga of St Olaf in Upps DG 8 II (ca. 1225–1250), Strengleikar in Upps DG 4–7 4to (ca. 1270) and the Law of...



5. Problemfälle und Lösungsansätze

- Robin Wendler: „**On a horse**“-Problem
- „Theodore Roosevelt Collection“

Robin Wendler, 2004. “The eye of the beholder: Challenges of image description and access at Harvard,” In: Diane I. Hillmann and Elaine Westbrooks (editors). *Metadata in Practice*. Chicago: ALA Editions, pp. 51–69.

- (ein) Grundproblem:
 - „*Relevance under change of context*“
 - Abschätzung zukünftiger Interessenten/Kontexte
 - Beispiel: transkribiertes Werk

Avé-Lalleman:

Das Deutsche Gaunerthum (BBAW) im VLO

Woher wissen was den Nutzer interessiert?

- Nutzerfeedback!

Kritik am VLO (u.a. Nutzerbefragung 2010):

- Nutzer wollen **direkten** Zugriff auf Ressourcen
 - Rechtefrage
- Nutzer hassen falsche/veraltete Informationen
 - u.a. inhärentes Problem föderierter Systeme
- Nutzer möchten möglichst **schnell** Ergebnisse
- Ungeeignete Suchfacetten (ongoing struggle)
- Inkompatible Vokabulare in Instanzen

Search



Showing all 902147 records

Results per page: 10

Use the categories below to limit the search results to those matching the selected value(s).

Language



Collection



Resource type



Modality



Format



Keyword



<< < 1 2 3 4 5 6 7 8 9 10 > >>

PROIEL collection

(Part of [Clarino Bergen Centre - INESS](#))

A collection of dependency treebanks for early attestations of Indo-European, including a set of parallel treebanks of the New Testament.

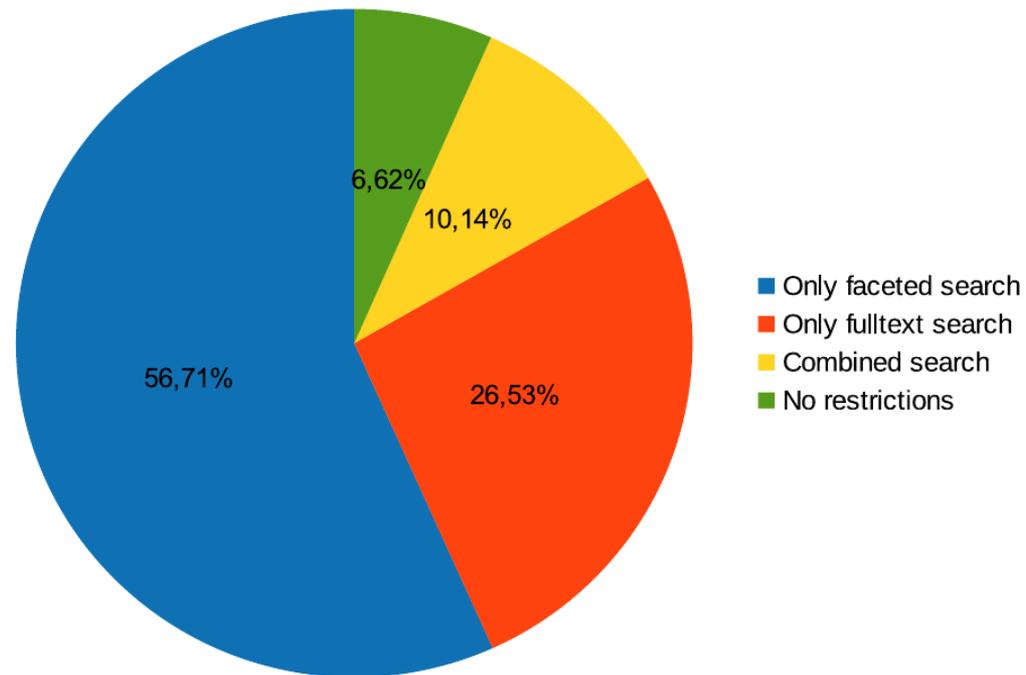


Menotec collection

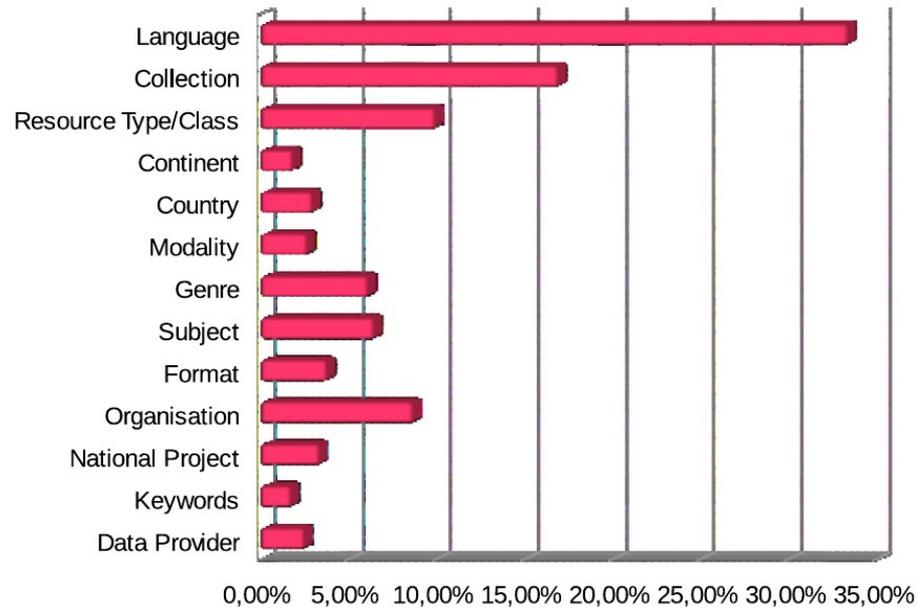
(Part of [Clarino Bergen Centre - INESS](#))

The Menotec collection in INESS contains four central manuscripts in Old Norwegian (although the ISO language code is termed "Old Norse"): The Old Norwegian Homily Book in AM 619 4to (ca. 1200–1225), the legendary saga of St Olaf in Upps DG 8 II (ca. 1225–1250), Strengleikar in Upps DG 4–7 4to (ca. 1270) and the Law of...





VLO 3



VLO 3

Auswahl relevanter Metadaten:

- Sowohl Domänen-abhängig als auch
- Allgemeingültig (siehe Dublin Core)

Problem 2: Vokabulare (oder „Was schreibe ich in die MD-Instanzen?“)

- Natürlicher Prozess
- Nutzung einheitlichen Vokabulares ist **schwer**
- Wird noch schwerer, falls
 - „Lange“ Zeiträume
 - Verteilte Arbeitsprozesse
 - Viele Teilprojekte
 - Multilinguale / „Multikulturelle“ Umgebung
 - Dezentrale Strukturen

- Weitere Probleme (kleiner Auszug):
 - Datums/Zeitangaben („02/05/2017“ vs. „2017-02-05“)
 - Lizenzangaben („CC-BY 2.0“ vs. „Creative Commons 2.0“)
 - Granularität der Beschreibung: „Text“, „Document“, „Manuscript“, „Poem“
 - Volltext vs. Identifier („German“ vs. „deu“)
 - Groß/Kleinschreibung (Problemfälle *TüBa-D/Z*)
 - ...

- 1. mögliche Lösung: Manuelles Nachbearbeiten
- VLO: 16 „Aufbereitungsdienste“
- Beispiel: Sprachangabe

Showing items 1 - 50 of 8041 matching values:

Danish (120758)
nld (116563)
ISO639-3:eng (75732)
Unspecified (54735)
nl (46170)
deu (44132)
da (36763)
English (35101)
ISO639-3:nld (23774)
eng (22141)
lat (16962)
ISO639-3:deu (15788)
ISO639-3:spa (15072)
ger (14110)
Deutsch (13919)
ISO639-3:fra (12264)
fra (8178)
ISO639-3:jpn (7732)



Showing items 1 - 50 of 6460 matching values:

Dutch (187857)
Danish (159013)
English (98317)
German (83169)
Unspecified (54739)
French (21339)
Latin (18338)
Spanish; Castilian (17630)
Deutsch (13829)
Japanese (9160)
Arabic (8420)
Chinese (6987)
Turkish (6895)
undetermined (6287)
Farsi; Persian (5844)
Indonesian (4443)
Français (4144)
Italian (3705)

- 2. mögliche Lösung: Zentrale Normdateien
- *„ein Verzeichnis von normierten Begriffen zur Verwendung als Deskriptor in der Dokumentation“*
(de.wikipedia.org/Normdatei)
- z.B. Gemeinsame Normdatei (GND) der DNB:
„Normdatei für Personen, Körperschaften, Kongresse, Geografika, Sachschlagwörter und Werktitel“
- Idee: Statt Beschreibung lieber Verweis auf autoritative Instanz
- Weitere Normdateien auch in Ihrem Fachbereich!

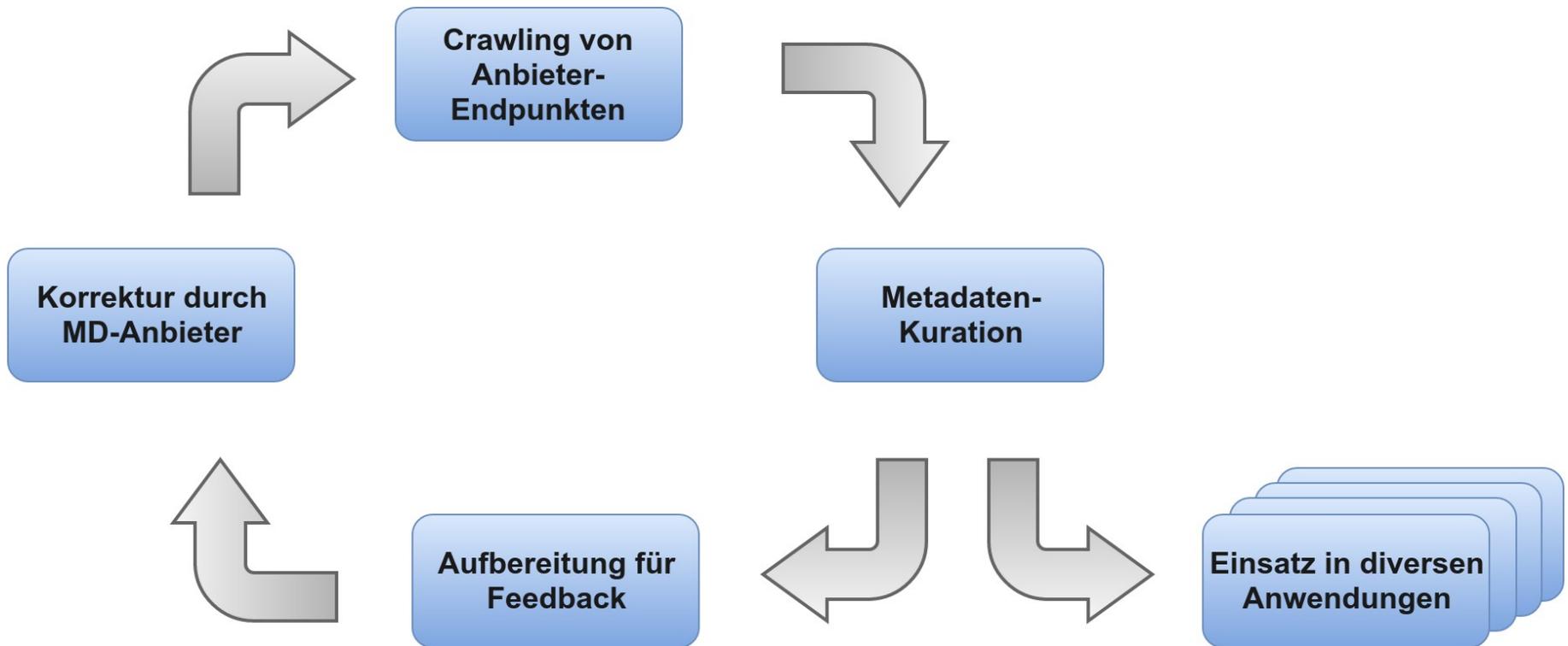
- Beispiel:

Organisation	Max Planck Instituut voor Psycholinguïstiek (Nimwegen)
PPN:	191120391 Zitier
GND-Nummer:	511293-X Link zu diesem Datensatz in der GND
Frühere Ansetzung:	<i>in gkd: a Max Planck Instituut voor Psycholinguïstiek <Nijmegen></i> <i>in swd: c Nimwegen / Max-Planck-Institut für Psycholinguistik</i>
Typ:	Organisation (kiz)
Quelle:	GKD
Beziehung zu Organisationen:	Max-Planck-Gesellschaft zur Förderung der Wissenschaften [administrative Überordnung]
Geografischer Bezug:	Nimwegen [Ort, Sitz (allgemein)] Niederlande [geographischer Wirkungsbereich]
Thematischer Bezug:	Psycholinguistik [Thema]
Ländercode:	XA-NL [Niederlande]
Zeitangaben:	1.1.1980 - [Zeit, Bestehen]
GND-Systematik:	6.5 [Wissenschaft] ; 11.2a [Allgemeine Sprachtheorie] ; 5.3 [Sozial-, Kultur- und Völkerpsychologie]
Weitere Namen	MPI voor Psycholinguïstiek (Nimwegen) Max-Planck-Instituut voor Psycholinguïstiek (Nimwegen) Max-Planck-Institut für Psycholinguistik (Nimwegen) Max Planck Institute for Psycholinguistics (Nimwegen) Planck Instituut voor Psycholinguïstiek (Nimwegen) Planck-Institut für Psycholinguistik (Nimwegen) Planck Institute for Psycholinguistics (Nimwegen) MPI for Psycholinguistics (Nimwegen) Max-Planck-Gesellschaft zur Förderung der Wissenschaften. Max Planck Instituut voor Psycholinguïstiek

- Weitere Normdateien/kontrolliertes Vokabular auch in Ihrem Fachbereich!
 - Inter-Active Terminology for Europe (IATE)
 - Schlagwortnormdatei (mittlerweile in GND)
 - Library of Congress Control Number
 - ...

- Qualität von Metadaten als zentrale Aufgabe
- Entsprechende Taskforces in CLARIN
 - CMDI-Taskforce
 - VLO working group
 - **Metadata curation**

- Idee: Feedbackschleife



- Curation Module des ACDH Wien
- Automatische Auswertung aller bereitgestellten Metadaten
- Bewertet sowohl Schema als auch Instanz
- Überprüft:
 - Anteil leerer Elemente
 - Anteil gefüllter Suchfacetten im VLO
 - ...
- Zusammenfassung für einzelne Zentren

- Curation Module (ACDH Wien)
- Beispiel:

<https://clarin.oeaw.ac.at/curate>

Metadaten:

1. sind ein spannendes Thema
2. stellen häufig ersten Einstieg in Ressourcen her
3. erschließen große (riesige) Ressourcenbestände
4. sind Grundlage für verschiedenste Anwendungen, weit über einfache Kataloge hinaus
5. sind nur bei Einhaltung von Qualitätsstandards hilfreich
6. sind nicht statisch, sondern ~~sollten~~ müssen an veränderte Kontexte angepasst werden

Fragen? Anmerkungen?