

Das DARIAH-DE Repository

Publikation und Referenzierung von geisteswissenschaftlichen Forschungsdaten

Stefan E. Funk

funk@sub.uni-goettingen.de

- Forschungsdaten und wissenschaftliche Sammlungen
- Anforderungen von Fachwissenschaftlerinnen und Fachwissenschaftlern
- Das DARIAH-DE Repository –Architektur und Implementierung
- (Nach-)Nutzungsperspektiven
- Beispiele

Unter digitalen geistes- und kulturwissenschaftlichen Forschungsdaten werden innerhalb von DARIAH-DE all jene Quellen und Ergebnisse verstanden, die im Kontext einer geistes- und kulturwissenschaftlichen Forschungsfrage gesammelt, beschrieben, ausgewertet und/oder erzeugt wurden und in maschinenlesbarer Form zum Zwecke der Archivierung, Zitierbarkeit und zur weiteren Verarbeitung aufbewahrt werden können.

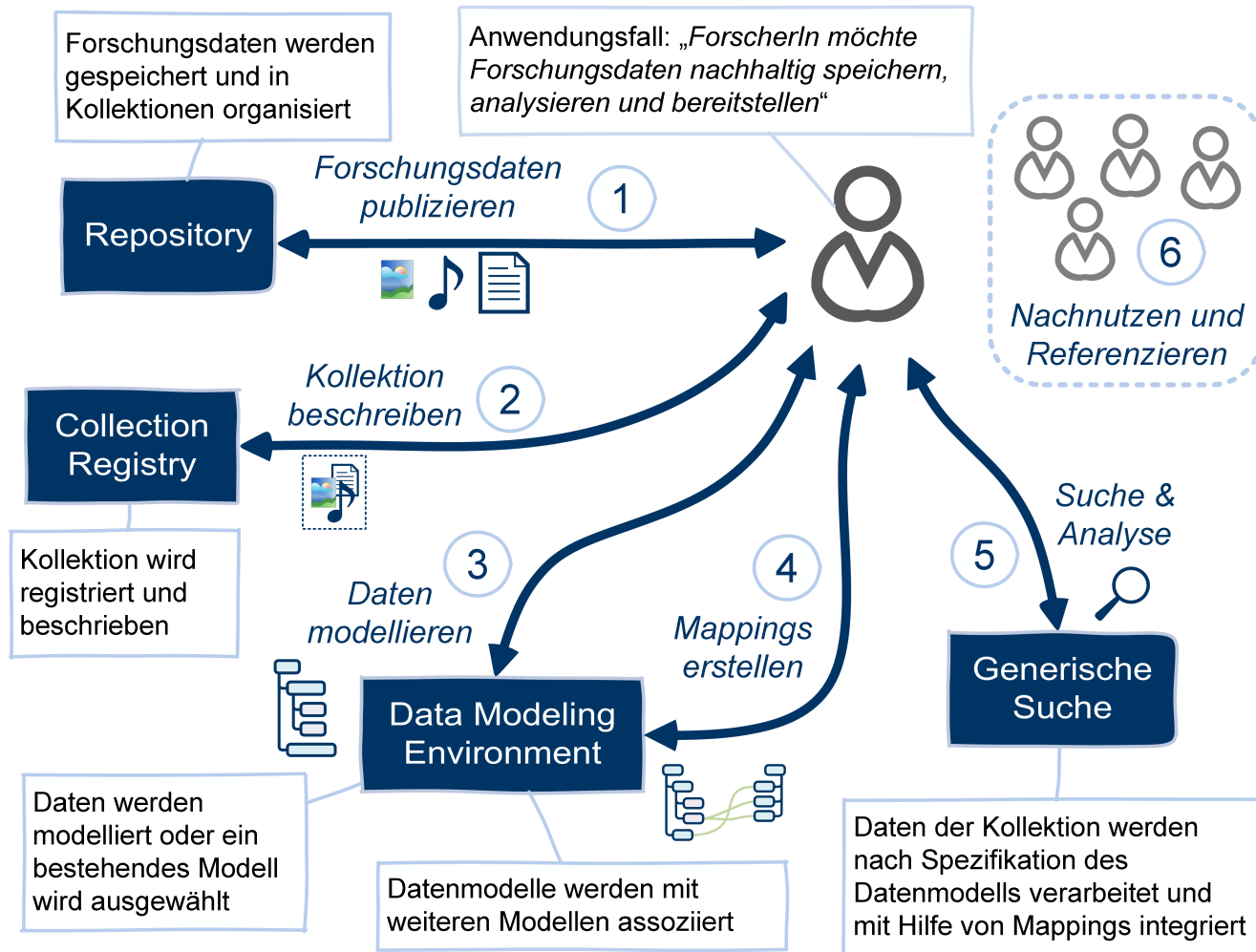
Unter digitalen geistes- und kulturwissenschaftlichen Forschungsdaten werden innerhalb von DARIAH-DE

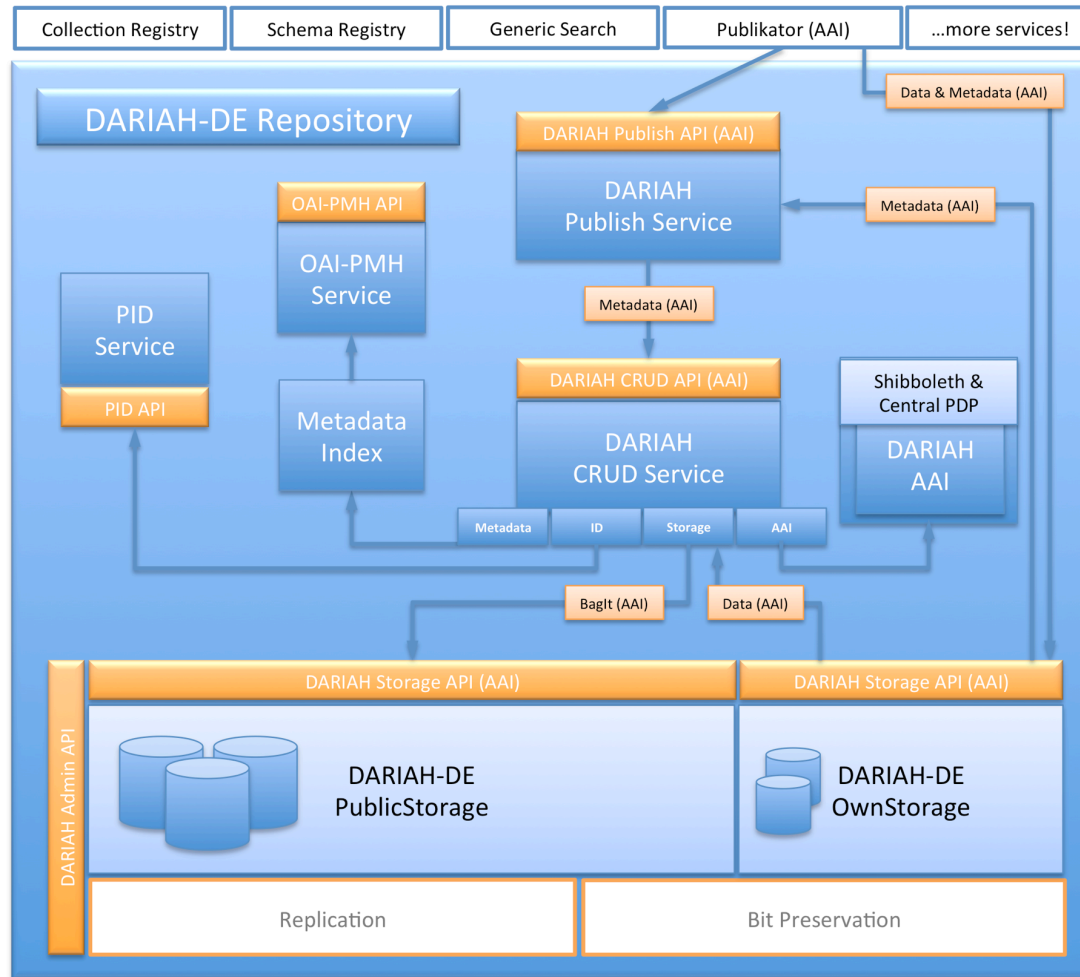
- **Digitale und / oder digitalisierte Daten**
 - **Digitale Repräsentation analoger Daten**
 - **Normdaten / Kontrollierte Vokabulare**
 - **Metainformationen**
 - **...**
- Verarbeitung aufbewahrt werden können.

- Maschinenlesbarkeit
- Interpretierbarkeit
- Prozessierbarkeit
- Referenzierbarkeit
- Speicherung / Langzeitarchivierung
- Open Access
- ...

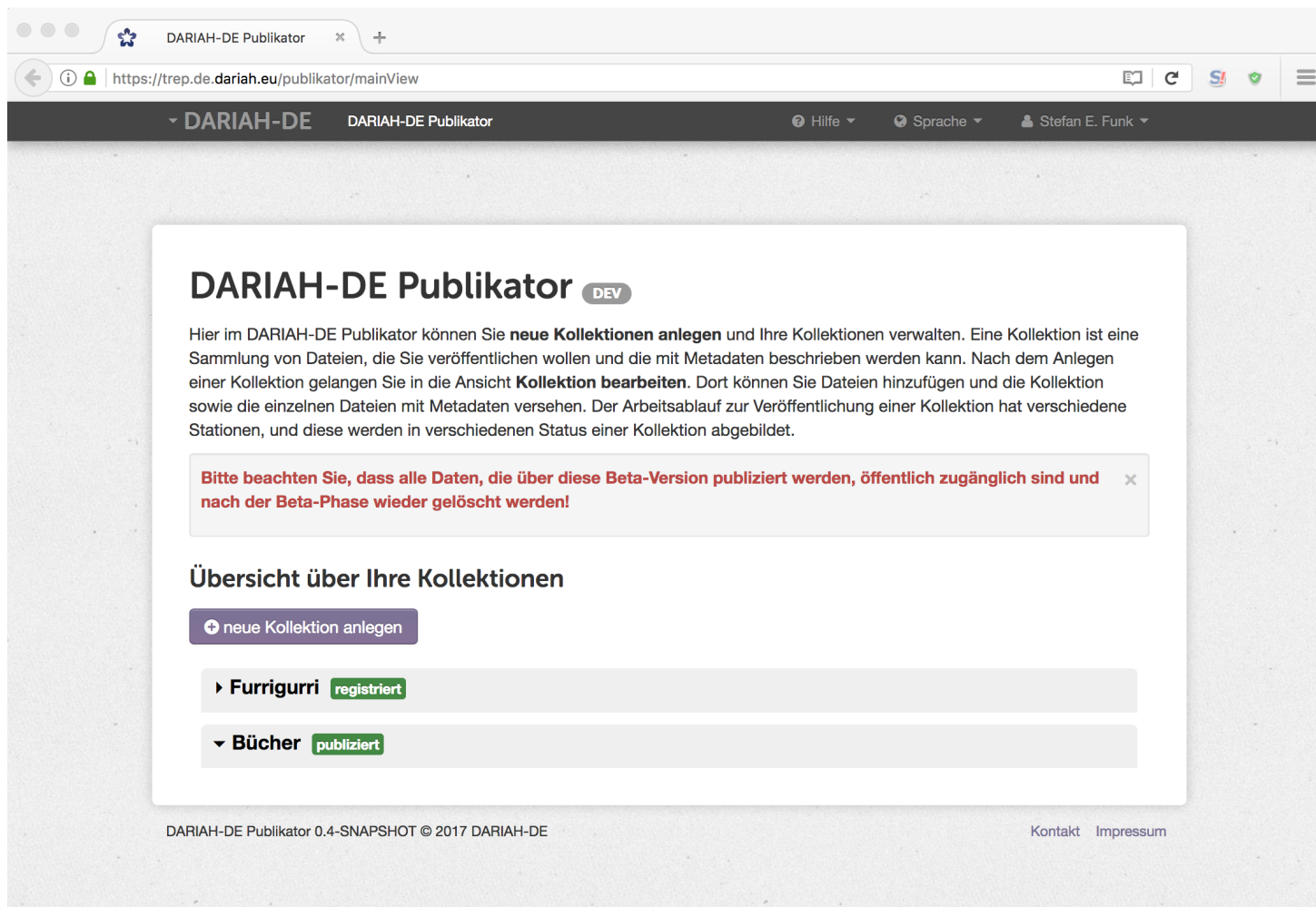
>> Nachnutzung und Veröffentlichung

- Forschungsdatenarchiv für die Geisteswissenschaften
- Eingebunden in die Forschungsdaten-Föderationsarchitektur
- Intuitive Publikation über das DARIAH-DE Portal (Publikator) und über APIs
- Nutzung über DARIAH-Account
- Gesichert durch die DARIAH-DE Authentifizierungs- und Autorisierungs-Infrastruktur (AAI)
- Nachnutzung von Komponenten der TextGrid-Kerndienste für Speicherung und Publikation
- Dauerhafte Referenzierung der Daten über DataCite DOIs und EPIC2 Handles
- Speicherung der Daten via DARIAH Storage API (OwnStorage / PublicStorage)





- **Publikator – Web Interface für das Publizieren**
Hochspielen der Daten, Anreichern mit Metadaten (DC Simple)
- **DARIAH-DE Publish Service – #publish, #status**
Vorbereiten der Daten für den Import, Metadatenvalidierung
- **DARIAH-DE CRUD Service – #create, #version, #retrieve, #retrieveMetadata**
Einspielen der Daten in das Repository, Verwalten diverser Datenbanken, Auslieferung der Daten
- **DARIAH AAI**
Authentifizierungs- und Autorisierungs-Infrastruktur
- **DARIAH-DE PID Service**
Erzeugen von PIDs für jede Datei (Zitierbarkeit) – EPIC2 Handles und DataCite DOIs
- **DARIAH-DE OwnStorage & PublicStorage | Storage API**
Speichern der Daten im Repositorium über die DARIAH Storage API
- **Metadaten-Index**
Indizieren der (Meta-)Daten mit ElasticSearch für eine performante Suche
- **OAI-PMH Service**
Ausliefern der Metadaten per standardisiertem Protokoll für weitere Dienste



The screenshot shows a web browser window with the address bar displaying <https://trep.de.dariah.eu/publikator/mainView>. The page header includes the DARIAH-DE logo, the title "DARIAH-DE Publikator", and navigation links for "Hilfe", "Sprache", and "Stefan E. Funk".

DARIAH-DE Publikator DEV

Hier im DARIAH-DE Publikator können Sie **neue Kollektionen anlegen** und Ihre Kollektionen verwalten. Eine Kollektion ist eine Sammlung von Dateien, die Sie veröffentlichen wollen und die mit Metadaten beschrieben werden kann. Nach dem Anlegen einer Kollektion gelangen Sie in die Ansicht **Kollektion bearbeiten**. Dort können Sie Dateien hinzufügen und die Kollektion sowie die einzelnen Dateien mit Metadaten versehen. Der Arbeitsablauf zur Veröffentlichung einer Kollektion hat verschiedene Stationen, und diese werden in verschiedenen Status einer Kollektion abgebildet.

Bitte beachten Sie, dass alle Daten, die über diese Beta-Version publiziert werden, öffentlich zugänglich sind und nach der Beta-Phase wieder gelöscht werden!

Übersicht über Ihre Kollektionen

[+ neue Kollektion anlegen](#)

- **Furrigurri** registriert
- ▾ **Bücher** publiziert

DARIAH-DE Publikator 0.4-SNAPSHOT © 2017 DARIAH-DE [Kontakt](#) [Impressum](#)

- Lädt die Daten in den DARIAH-DE OwnStorage
- Zugriff hat nur die/der jeweilige Hochladende
- Erzeugt aus den Eingaben der Nutzerinnen und Nutzer eine RDF/XML-Datei pro Kollektion
- Gibt die erzeugte Wurzelkollektion und ein Authentifizierungs-Token an den Publish-Dienst
- Prüft in regelmäßigen Abständen den Status des Publikationsvorgangs
- Meldet den Erfolg der Publikation bzw. auftretende Fehler

Beispiel einer RDF/XML-Datei

```
@prefix dariah: <http://de.dariah.eu/rdf/dataobjects/terms/> .  
@prefix dariahstorage: <https://de.dariah.eu/storage/> .  
@prefix hdl: <http://hdl.handle.net/> .  
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
@prefix dcterms: <http://purl.org/dc/terms/> .  
@prefix dc: <http://purl.org/dc/elements/1.1/> .
```

```
dariahstorage:EAEA0-DD79-5ADA-5D57-0
```

```
  a    dariah:DataObject ;  
  dc:creator "JB", "Johannes Biermann";  
  dc:format "text/csv";  
  dc:identifier "hdl:21.T11998/0000-0001-91F5-6";  
  dc:rights "CC-BY", "CC BY";  
  dc:title "measurement 20170814-06.csv";  
  dcterms:identifier "hdl:21.T11998/0000-0001-91F5-6" .
```

```
dariahstorage:EAEA0-D15E-CD7F-7D2C-0
```

```
  a    dariah:Collection ;  
  dc:creator "Johannes Biermann";  
  dc:identifier "hdl:21.T11998/0000-0001-91F4-7";  
  dc:rights "CC BY";  
  dc:title "ESR vs Capacitance";  
  dcterms:hasPart dariahstorage:EAEA0-DD79-5ADA-5D57-0 , dariahstorage:EAEA0-C048-7637-8C93-0 , dariahstorage:EAEA0-3B1C-C03B-  
81C7-0 ;  
  dcterms:identifier "hdl:21.T11998/0000-0001-91F4-7" .
```

- Authentifiziert / autorisiert gegen die DARIAH AAI
- Liefert dem Publikator auf Anfrage Statusinformation für z. B. Fortschrittsbalken und Status der Publikation
- Prüft, ob die genannte Kollektion bereits in der Collection Registry existiert
- Prüft auf Vollständigkeit der Metadaten
- Speichert Daten und Metadaten per CRUD-Service im OwnStorage
- Generiert Persistente Identifikatoren – EPIC2 Handles und DataCite DOIs

Statusabfrage einer Publikationsvorgangs

<http://trep.de.dariah.eu/1.0/dhpublish/EAEA0-495C-69E3-0B29-0/info>

```
<infoResponse status="PUBLISHED"  
  pid="21.T11998/0000-0001-91A0-5"  
  doi="10.20375/0000-0001-91A0-5"  
  crid="5991b3e97c8dec05ccf60b44"  
  uri="https://de.dariah.eu/storage/  
    EAEA0-495C-69E3-0B29-0"  
  progress="100"/>
```

- Authentifiziert / autorisiert gegen die DARIAH AAI
- Generiert administrative und technische Metadaten, ergänzt deskriptive Metadaten
- Kopiert Daten und Metadaten vom DARIAH-DE OwnStorage in den PublicStorage
- Indiziert die Daten / Metadaten per ElasticSearch Datenbank (Metadatenindex)

<rdf:RDF

xmlns:dcterms="http://purl.org/dc/terms/"

xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

xmlns:dc="http://purl.org/dc/elements/1.1/"

xmlns:dariah="http://de.dariah.eu/rdf/dataobjects/terms/"

xmlns:hdl="http://hdl.handle.net/" >

<rdf:Description rdf:about="hdl:21.T11998/0000-0001-91F7-4">

<dc:creator>Johannes Biermann</dc:creator>

<dc:rights>CC BY</dc:rights>

<dc:format>text/csv</dc:format>

<dc:identifier>hdl:21.T11998/0000-0001-91F7-4</dc:identifier>

<dc:relation>dariah:Collection:21.T11998/0000-0001-91F4-7</dc:relation>

<dc:title>measurement 20170814-02.csv</dc:title>

<dc:identifier>doi:10.20375/0000-0001-91F7-4</dc:identifier>

<rdf:type rdf:resource="http://de.dariah.eu/rdf/dataobjects/terms/DataObject"/>

</rdf:Description>

</rdf:RDF>


```
<rdf:RDF
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:premis="http://www.loc.gov/premis/rdf/v1#"
  xmlns:hdl="http://hdl.handle.net/"
  xmlns:doi="http://dx.doi.org/" >
  <rdf:Description rdf:about="hdl:21.T11998/0000-0001-91F7-4">
    <premis:hasMessageDigest>096a6fd300c26d63078823abc99ea4a7</premis:hasMessageDigest>
    <dcterms:relation>dariah:Collection:21.T11998/0000-0001-91F4-7</dcterms:relation>
    <dcterms:created>2017-08-14T17:16:15.344</dcterms:created>
    <dcterms:modified>2017-08-14T17:16:15.344</dcterms:modified>
    <dcterms:creator>JohannesBiermann@dariah.eu</dcterms:creator>
    <premis:hasMessageDigestOriginator>
      dhcrud-base 7.11.0-SNAPSHOT:201708141620
    </premis:hasMessageDigestOriginator>
    <dcterms:extent>6</dcterms:extent>
    <dcterms:identifier>doi:10.20375/0000-0001-91F7-4</dcterms:identifier>
    <premis:hasMessageDigestType>md5</premis:hasMessageDigestType>
    <dcterms:format>text/csv</dcterms:format>
    <dcterms:identifier>hdl:21.T11998/0000-0001-91F7-4</dcterms:identifier>
  </rdf:Description></rdf:Description>
</rdf:RDF>
```

```
<?xml version="1.0" encoding="UTF-8"?> <fits xmlns="http://hul.harvard.edu/ois/xml/ns/fits/fits_output"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://hul.harvard.edu/ois/xml/ns/fits/fits_output
http://hul.harvard.edu/ois/xml/xsd/fits/fits_output.xsd" version="1.1.0" timestamp="14.08.17 16:29"> <identification> <identity
format="Extensible Markup Language" mimetype="text/xml" toolname="FITS" toolversion="1.1.0"> <tool toolname="Droid"
toolversion="6.1.5" /> <tool toolname="Jhove" toolversion="1.16" /> <tool toolname="file utility" toolversion="5.14" /> <tool
toolname="Exiftool" toolversion="10.00" /> <tool toolname="ffident" toolversion="0.2" /> <tool toolname="Tika" toolversion="1.10" />
<version toolname="Droid" toolversion="6.1.5">1.0</version> <externalIdentifier toolname="Droid" toolversion="6.1.5"
type="puid">fmt/101</externalIdentifier> </identity> </identification> <fileinfo> <size toolname="Jhove"
toolversion="1.16">368</size> <filepath toolname="OIS File Information" toolversion="0.2" status="SINGLE_RESULT">/tmp/tomcat-
crud-tmp/data-6937802785461971100.rdf</filepath> <filename toolname="OIS File Information" toolversion="0.2"
status="SINGLE_RESULT">data-6937802785461971100.rdf</filename> <md5checksum toolname="OIS File Information"
toolversion="0.2" status="SINGLE_RESULT">cc852e7792d52c334608e06312936018</md5checksum> <fslastmodified toolname="OIS
File Information" toolversion="0.2" status="SINGLE_RESULT">1502720993000</fslastmodified> </fileinfo> <filestatus> <well-formed
toolname="Jhove" toolversion="1.16" status="SINGLE_RESULT">true</well-formed> <valid toolname="Jhove" toolversion="1.16"
status="SINGLE_RESULT">false</valid> <message toolname="Jhove" toolversion="1.16" status="SINGLE_RESULT">Not able to determine
type of end of line</message> <message toolname="Jhove" toolversion="1.16" status="SINGLE_RESULT">cvc-elt.1: Cannot find the
declaration of element 'rdf:RDF'. Line = 1, Column = 155</message> </filestatus> <metadata> <text> <charset toolname="Jhove"
toolversion="1.16" status="SINGLE_RESULT">UTF-8</charset> <markupBasis toolname="Jhove" toolversion="1.16"
status="CONFLICT">XML</markupBasis> <markupBasis toolname="Exiftool" toolversion="10.00"
status="CONFLICT">XMP</markupBasis> <markupBasisVersion toolname="Jhove" toolversion="1.16"
status="SINGLE_RESULT">1.0</markupBasisVersion> <standard> <textMD:textMD xmlns:textMD="info:lc/xmlns/textMD-v3">
<textMD:character_info> <textMD:charset>UTF-8</textMD:charset> </textMD:character_info> <textMD:markup_basis
version="1.0">XML</textMD:markup_basis> </textMD:textMD> </standard> </text> </metadata> <statistics fitsExecutionTime="341">
<tool toolname="MediaInfo" toolversion="0.7.75" status="did not run" /> <tool toolname="OIS Audio Information" toolversion="0.1"
status="did not run" /> <tool toolname="ADL Tool" toolversion="0.1" status="did not run" /> <tool toolname="VTT Tool"
toolversion="0.1" status="did not run" /> <tool toolname="Droid" toolversion="6.1.5" executionTime="53" /> <tool toolname="Jhove"
toolversion="1.16" executionTime="74" /> <tool toolname="file utility" toolversion="5.14" executionTime="119" /> <tool
toolname="Exiftool" toolversion="10.00" executionTime="324" /> <tool toolname="NLNZ Metadata Extractor" toolversion="3.6GA"
status="did not run" /> <tool toolname="OIS File Information" toolversion="0.2" executionTime="40" /> <tool toolname="OIS XML
Metadata" toolversion="0.2" status="did not run" /> <tool toolname="ffident" toolversion="0.2" executionTime="37" /> <tool
toolname="Tika" toolversion="1.10" executionTime="23" /> </statistics> </fits>
```

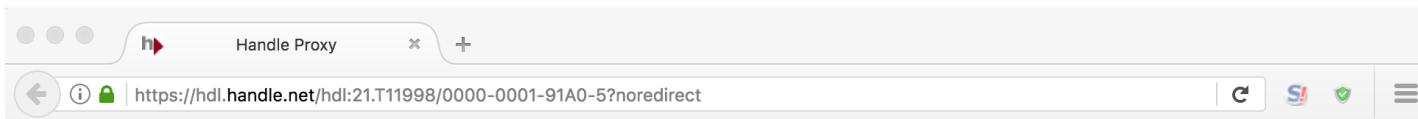
Bücher – hdl:21.T11998/0000-0001-91A0-5

https://trep.de.dariah.eu/1.0/dhcrud/21.T11998/0000-0001-91A0-5/index

DARIAH-DE REPOSITORY OBJECT INDEX

DOI	doi:10.20375/0000-0001-91A0-5
Handle	hdl:21.T11998/0000-0001-91A0-5
Title	🔊 Bücher
Creator	fu
Rights	free
Type	text/tg.collection+tg.aggregation+xml
Created	2017-08-14T16:29:53.369
Metadata	🔊 descriptive 🔊 administrative 🔊 technical 🔊 Handle 🔊 DOI
Data	🔊 download file only 🔊 download bagit bag
Embedded Objects	hdl:21.T11998/0000-0001-91A2-3 hdl:21.T11998/0000-0001-91A1-4

Delivered by dhcrud-base-7.11.0-SNAPSHOT.201708141620



Handle.Net®

Handle Values for: 21.T11998/0000-0001-91A0-5

Index	Type	Timestamp	Data
1	CREATOR	2017-08-14 14:29:48Z	PID Service pid-webapp-4.15.0-SNAPSHOT.201708141617
2	ADM_MD	2017-08-14 14:29:55Z	https://trep.de.dariah.eu/1.0/dhcrud/21.T11998/0000-0001-91A0-5/adm
3	FILESIZE	2017-08-14 14:29:55Z	8291
4	RESPONSIBLE	2017-08-14 14:29:55Z	StefanFunk@dariah.eu
5	CHECKSUM	2017-08-14 14:29:55Z	md5:cc852e7792d52c334608e06312936018
6	BAG	2017-08-14 14:29:55Z	https://cdstar.de.dariah.eu/test-nopdp/public/EAEA0-EFAA-E1B7-AC62-0
7	PUBDATE	2017-08-14 14:29:55Z	2017-08-14 16:29:53 +0200
8	PROV_MD	2017-08-14 14:29:55Z	https://trep.de.dariah.eu/1.0/dhcrud/21.T11998/0000-0001-91A0-5/prov
9	URL	2017-08-14 14:29:55Z	https://cdstar.de.dariah.eu/test-nopdp/public/EAEA0-EFAA-E1B7-AC62-0
10	DATA	2017-08-14 14:29:55Z	https://trep.de.dariah.eu/1.0/dhcrud/21.T11998/0000-0001-91A0-5/data
11	SOURCE	2017-08-14 14:29:55Z	https://de.dariah.eu/storage/EAEA0-495C-69E3-0B29-0
12	INDEX	2017-08-14 14:29:55Z	https://trep.de.dariah.eu/1.0/dhcrud/21.T11998/0000-0001-91A0-5/index
13	METADATA	2017-08-14 14:29:55Z	https://trep.de.dariah.eu/1.0/dhcrud/21.T11998/0000-0001-91A0-5/metadata
14	TECH_MD	2017-08-14 14:29:55Z	https://trep.de.dariah.eu/1.0/dhcrud/21.T11998/0000-0001-91A0-5/tech
15	DOI	2017-08-14 14:29:55Z	doi:10.20375/0000-0001-91A0-5
16	PUBLISHED	2017-08-14 14:30:01Z	true
17	COLREG_ID	2017-08-14 14:30:01Z	5991b3e97c8dec05ccf60b44
100	HS_ADMIN	2017-08-14 14:29:48Z	handle=21.T11998/TGTEST-USER01; index=1; [create hdl,delete hdl,read val,modify val,del val,add val,modify admin,del admin,add admin]

[Handle Proxy Server Documentation](#)
[Handle.net Web Site](#)

Please contact hdladmin@cnri.reston.va.us for your handle questions and comments

- Bagit: Standard zum Verpacken von Daten
 - <https://tools.ietf.org/html/draft-kunze-bagit-14>
 - <https://github.com/jkunze/bagitspec>
- Bagit-Profile zum Beschreiben, wie man seine Daten mit Bagit ablegt
 - maschinenlesbar
 - ausführlicher als die Spezifikation
 - ähnlich eines METS-Profiles
 - <https://github.com/ruebot/bagit-profiles>

```
myfirstbag/  
|-- data  
|  \-- 27613-h  
|    \-- images  
|       \-- q172.png  
|       \-- q172.txt  
|-- manifest-md5.txt  
|  49afbd86a1ca9f34b677a3f09655eae9 data/27613-h/images/q172.png  
|  408ad21d50cef31da4df6d9ed81b01a7 data/27613-h/images/q172.txt  
|-- bagit.txt
```

BagIt-Version: 0.97

Tag-File-Character-Encoding: UTF-8

- Protokoll für Metadata Harvesting
- Liefert Metadaten für alle in einem bestimmten Zeitraum publizierten Objekte einer Kollektion
- Generische Suche nutzt diese Metadaten für die Indizierung der Daten
- Regelmäßige Abfrage der Kollektionen über OAI-PMH, deren URLs in der Collection Registry angegeben sind
- Baut auf dem Datenbestand der ElasticSearch-Datenbank des CRUD-Services auf

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
xmlns:ns2="http://purl.org/dc/elements/1.1/"
xmlns:ns3="http://www.openarchives.org/OAI/2.0/oai_dc/"
<responseDate>2017-08-14T17:13:55Z</responseDate>
<request verb="Identify">repository.de.dariah.eu</request>
<Identify>
  <repositoryName>DARIAH-DE Repository</repositoryName>
  <baseURL>http://trep.de.dariah.eu</baseURL>
  <protocolVersion>2.0</protocolVersion>
  <adminEmail>funk@sub.uni-goettingen.de</adminEmail>
  <earliestDatestamp>2011-06-11T02:32:40Z</earliestDatestamp>
  <deletedRecord>no</deletedRecord>
  <granularity>YYYY-MM-DDThh:mm:ssZ</granularity>
</Identify>
</OAI-PMH>
```


- Daten sind in Repositorien publiziert
- Kollektionen sind in der Collection Registry beschrieben
- Generische Suche indiziert die Daten der Kollektionen
- Inhalte aller beschriebenen Kollektionen sind such- und findbar über die Generische Suche

- Publikator umgezogen von Liferay-Portlet zu native JavaScript / Java WebService
- Implementierung von Bagit-Bags für die Speicherung im DARIAH-DE PublicStorage
- Implementierung von Unter-Kollektionen für die einfache Strukturierung von Kollektionen
- Implementierung der DOI-Generierung bei DataCite
- Weiterentwicklung aller beteiligten Services
- Integration in die Puppet-Umgebung von DARIAH-DE
- Neuer Beta-Service bis Ende August
- Geplanter Release des produktiven DARIAH-DE Repository: Ende September
- Fertigstellung der API für den maschinellen Import: Q4 2017

<https://trep.de.dariah.eu/publikator>
funk@sub.uni-goettingen.de

<https://de.dariah.eu>
<https://textgrid.de>
info@de.dariah.eu

THANK YOU

DARIAH-DE Portal

<https://de.dariah.eu>

DARIAH-DE Repository

<https://de.dariah.eu/forschungsdaten>

DARIAH-DE Publikator (DEV)

<https://trep.de.dariah.eu/publikator>

Publikator-Dokumentation

<https://wiki.de.dariah.eu/display/publicde/Das+DARIAH-DE+Repository>

DARIAH-DE Dienste

Collection Registry <https://demo2.dariah.eu/colregs/>

Generische Suche <http://dev3.dariah.eu/search/>

EPIC2 PIDs

<https://www.pidconsortium.eu>

DataCite DOIs

<https://www.datacite.org>