



Schnittstellenkonzept (M 4.3.1)

Version 2017-03-03

Cluster 4

Verantwortlicher Partner MInf-BA, SUB

DARIAH-DE Überführung der digitalen Forschungsinfrastrukturen für die e-Humanities in die Operational Phase (Betriebsphase)

Dieses Forschungs- und Entwicklungsprojekt wird / wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF), Förderkennzeichen 01UG1610A bis J, gefördert und vom Projektträger im Deutschen Zentrum für Luft- und Raumfahrt (PT-DLR) betreut.

GEFÖRDERT VOM



**Bundesministerium
für Bildung
und Forschung**

Projekt: DARIAH-DE: Überführung der digitalen Forschungsinfrastrukturen für die e-Humanities in die Operational Phase (Betriebsphase)

BMBF Förderkennzeichen: 01UG1610A bis J

Laufzeit: März 2016 bis Februar 2019

Dokumentstatus: Final

Verfügbarkeit: öffentlich

Autoren:

Tobias Gradl, MInf-BA

Stefan E. Funk, SUB Göttingen

Revisionsverlauf:

Datum	Autor	Kommentare
20.01.2017	Tobias Gradl	Entwurf ohne Repository
03.02.2017	Stefan E. Funk	Repository-Schnittstellen hinzugefügt sowie Beteiligung zur RDA
06.02.2017	Stefan E. Funk	Repository API-Dokumentation angelegt und referenziert
16.02.2017	Tobias Gradl	Formatierung
03.03.2017	Tobias Gradl	Final / Öffentlich



Dieses Werk ist unter einer Creative Commons Lizenz vom Typ Namensnennung 3.0 Deutschland zugänglich. Um eine Kopie dieser Lizenz einzusehen, konsultieren Sie <http://creativecommons.org/licenses/by/3.0/de/> oder wenden Sie sich brieflich an Creative Commons, Postfach 1866, Mountain View, California, 94042, USA.

Inhaltsverzeichnis:

1. Einleitung	4
2. Technische Ebenen und Schnittstellen	4
2.1. Generische Suche	6
2.2. Föderationsschicht	10
2.2.1. Collection Registry	10
2.2.2. Schema Registry.....	11
2.2.3. Mapping Registry	13
2.3. DARIAH-DE Repository.....	14
2.3.1. Schnittstellen-Dokumentation des Repositorys.....	16
2.3.2. Öffentliche Schnittstellen der einzelnen Repository-Dienste.....	17
3. Schnittstellen zu anderen Projekten / Infrastrukturen	19
3.1. MWW.....	20
3.2. WissKI	20
3.3. Cosmopolitanities	21
3.4. CLARIN-D	21
3.5. RDA – Research Data Alliance.....	22
4. Ausblick	22
Literaturverzeichnis	23

1. Einleitung

Die Digital Humanities umfassen eine komplexe und vielseitige Landschaft wissenschaftlicher Disziplinen der Kultur- und Geisteswissenschaften und der Informatik. Gerade wegen der Vielseitigkeit werden die Digital Humanities derzeit mit Hilfe verschiedener Infrastrukturen und Forschungsprojekte erschlossen, die bislang weitgehend isoliert voneinander durchgeführt werden.

Für DARIAH-DE ist die Interoperabilität von Daten und Diensten seit der ersten Projektphase ständiger Begleiter. In der derzeit durchgeführten Phase mit dem Ziel der *Überführung der digitalen Infrastruktur für die e-Humanities in die Operationale Phase* und damit der *Verstetigung und dem dauerhaften Nutzbarmachen von Projektergebnissen* erhält die Kommunikations- und Kooperationsfähigkeit von Diensten eine noch zentralere Bedeutung.

Der vorliegende Meilensteinbericht fasst für den inhaltlichen Bereich von *Cluster 4: Wissenschaftliche Sammlungen* zunächst die umgesetzten und geplanten Schnittstellen der Dienste der DARIAH-DE Föderationsarchitektur zusammen und bietet einige Beispiele der technischen Schnittstellen und ihrer Anwendbarkeit. Im darauffolgenden Abschnitt wird eine Auswahl von Projekten vorgestellt zu denen entsprechende Schnittstellen eingerichtet wurden oder werden. Diese Liste reflektiert dabei lediglich diejenigen Projekte, die sich thematisch und technisch für eine Interaktion mit der DARIAH-DE Föderationsarchitekturen qualifizieren und repräsentiert keinesfalls eine vollständige Aufzählung aller Partnerprojekte von Cluster 4 oder gar des DARIAH-DE Verbundprojekts.

2. Technische Ebenen und Schnittstellen

Abbildung 1 zeigt einen Überblick über die Komponenten (Kästen) und Interaktionen (Pfeile) innerhalb der DARIAH-DE Infrastruktur. Sämtliche Schnittstellen an den dargestellten DARIAH-DE Komponenten wurden mit einem primären Fokus auf eine Nutzung für interne Anwendungsfälle definiert und implementiert: die Generische Suche von DARIAH-DE greift für den Zugriff auf Daten, sowie deren Verarbeitung und Indexierung auf zwei Schnittstellen zu:

- Durch Interaktion mit der **Collection Registry** werden insbesondere die Fragen geklärt,

- welche Kollektionen zugegriffen werden können,
 - unter welcher Adresse ein solcher Zugriff erfolgen kann (im Beispiel des **Repository** anhand der standardisierten OAI-PMH¹ Schnittstelle) und
 - wie Daten in den Kollektionen strukturiert sind (Verweis auf ein Schema in der DARIAH-DE Schema Registry).
- Jeder Verweis auf ein Datenmodell kann durch die Nutzung der Schnittstellen der **Schema Registry** aufgelöst werden und weitere Fragen zu lösen:
- Welchen strukturellen Rahmenbedingungen folgen die Daten einer Kollektion?
 - Mit Hilfe welcher Regeln können Aspekte des Erstellungskontexts so in Daten aufgenommen werden, dass eine Interpretation auch ohne kontextuelles Sammlungswissen erleichtert werden kann?

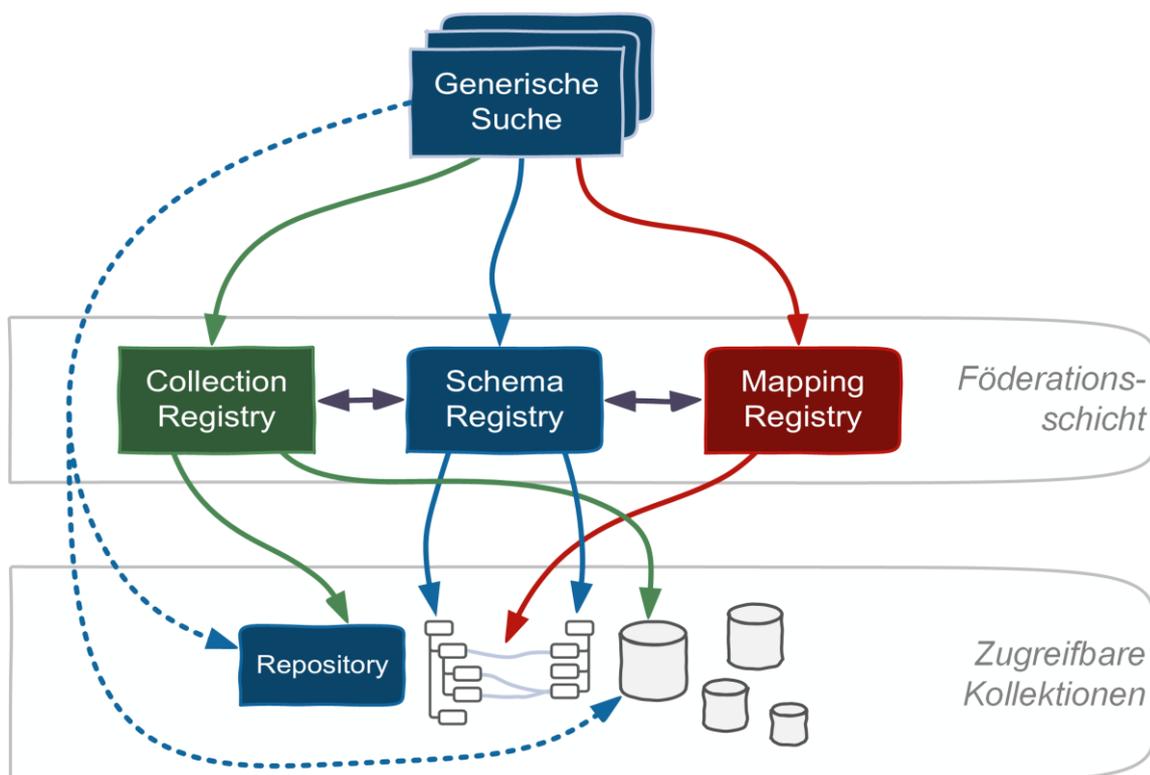


Abbildung 1: DARIAH-DE Föderationsarchitektur²

Für eine Auflösung übergreifender Suchanfragen und die Transformation von Suchergebnissen bezieht die Generische Suche zusätzliche Informationen aus der **Mapping Registry**:

¹ Open Archives Initiative Protocol for Metadata Harvesting; vgl. Lagoze et al. (2002)

² Gradl et al. (2015), Abschnitt 3; Gradl/Henrich (2016), S. 227

- Welche Schemata weisen Assoziationen miteinander auf und wie genau sehen diese Assoziationen aus?
- Welche Transformationsregeln sind anzuwenden, um Daten aus ihrem originalen Kontext in einen neuen Verwendungskontext zu überführen.

Während diese Schnittstellen (mit Ausnahme der standardisierten OAI-PMH Schnittstelle des DARIAH-DE Repository) initial vornehmlich für genau diesen Anwendungsfall der Generischen Suche konzipiert und implementiert wurden, wurde bei deren Konzeption durchaus auf eine spätere Nachnutzung durch weitere interne und externe Dienste (angedeutet in Abbildung 1 durch das Multi-Objekt Symbol der Generischen Suche) geachtet.

Im folgenden Abschnitt werden die im Rahmen der DARIAH-DE Föderationsarchitektur implementierten und geplanten Schnittstellen kurz vorgestellt, insbesondere um einen Ausblick auf Möglichkeiten zur Nachnutzung anzubieten. Für weiterführende Informationen zu den Komponenten der DARIAH-DE Föderationsarchitektur sei vor allem auf Gradl et al. (2015), Gradl/Henrich (2016) sowie Schmunk/Funk (2016) verwiesen.

2.1. Generische Suche

Die von der Generischen Suche – und auf ihr basierenden Dienste – angebotenen Schnittstellen ermöglichen einen Zugriff auf die im Rahmen der DARIAH-DE Föderation aggregierten Metadaten und Daten. Die Schnittstellen der Generischen Suche sind explizit für eine Verwendung außerhalb der Föderationsarchitektur selbst konzipiert und daher nicht der Übersichtsgrafik in Abbildung 1 enthalten.

Mit Hilfe der Schnittstellen der Generischen Suche können externe Anwendungen sehr einfach von den Vorzügen der DARIAH-DE Föderationsarchitektur profitieren ohne die Probleme der Datenverarbeitung, -anreicherung und -transformation selbst – durch Interaktion mit den Registries – behandeln zu müssen.

Abbildung 2 skizziert die Interaktionsmöglichkeiten die internen und externen Infrastrukturen und Diensten durch die Schnittstellen der Generischen Suche angeboten werden können.

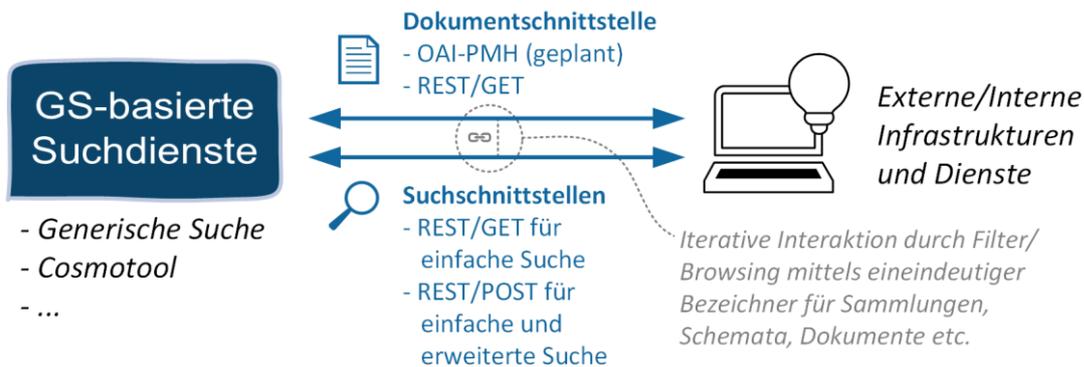


Abbildung 2: Schnittstellen der Generischen Suche von DARIAH-DE

Suchschnittstellen

Über die REST-basierte Schnittstelle der Generischen Suche kann stets mindestens alle Suchfunktionalität erreicht werden, die auch über das Browserinterface der Suche angeboten wird. Dies wird insbesondere dadurch erreicht, dass die für die Umsetzung des Browserinterfaces implementierte Logik selbst auf die REST-basierte Schnittstelle zugreift.

Einfache Suche

Eine einfache Suche kann durch die Übergabe einer Anfrage in Form eines GET-Parameters (z. B. <http://search.de.dariah.eu/search/query?q=Bamberg>) resultiert in der Rückgabe einer Menge von Suchergebnissen.

Einfache und erweiterte Anfragen können durch Verwendung von POST-Parametern genauer spezifiziert werden.

Unter Verwendung der Such-URI <http://search.de.dariah.eu/search/query> kann die einfache Suche nach „Bamberg“ durch Übergabe von `{"queryString":"bamberg"}` erreicht werden.

Erweiterte Suche

Eine beispielhafte, erweiterte Suche unter Verwendung von strukturbasierten Suchfacetten könnte durch die Übergabe der folgenden Anfrageparameter ausgeführt werden.

```
{
  "facets": [
    {
      "schemaElementId": "564ca7a8e4b088abc7b7fc68",
      "expression": "goethe"},
  ]
}
```

```

    {
      "schemaElementId": "564ca7a8e4b088abc7b7fc7b",
      "expression": "1820"
    }
  ],
  "schemaId": "564ca78fe4b088abc7b7fc62"
}

```

Abbildung 3: Anfrageparameter für Generische Suche

Die in Abbildung 4 dargestellte, gekürzte Antwort auf obenstehende Anfrage liefert sowohl Informationen zu den einzelnen Suchergebnissen, als auch den beinhaltenden Kollektionen. Zudem wird gegen Ende der Antwort die konkret ausgeführte Suchanfrage, sowie ggf. in andere Schemata übersetzte Anfragen mitgeliefert.

```

<QueryResult xmlns="">
  <resultElements>
    <resultElement>
<id>564ca78fe4b088abc7b7fc62.5814773ce4b091b5ddfbf496.oai:bdr.oai.bsb-
muenchen.de:europaana:BDR-BV007878360-62885</id>
      <score>13.062655</score>
      <explanation/>
      <sourceName>Bayerisches Digitales Repositorium</sourceName>
      <sourceId>5814773ce4b091b5ddfbf496</sourceId>
      <schemaId>564ca78fe4b088abc7b7fc62</schemaId>
      <snippet.../>
      <rank>0</rank>
      <info>
        <ResultElement>
          <Title>
            <~>Erlkönig : Ballade von Goethe ; 1tes Werk</~>
          </Title>
          <Link>
            <~>http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:12-bsb00041494-4</~>
          </Link>
          <Identifier>
            <~>http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:12-bsb00041494-4</~>
          </Identifier>
          <Description>
            <~>in Musik gesetzt für eine Singstimme mit Begleitung des Piano-Forte ... von Franz Schubert</~>
          </Description>
          <Description>
            <~>Musikdruck</~>
          </Description>
        </ResultElement>
      </info>
    </resultElement>
  </resultElements>
  <resultDatasources>
    <resultDatasource>
      <providerId>5814773ce4b091b5ddfbf496</providerId>
      <providerName>Bayerisches Digitales Repositorium</providerName>
      <hitCount>4</hitCount>
      <loadTags>false</loadTags>
    </resultDatasource>

```

```

</resultDatasources>
<maxScore>13.062655</maxScore>
<totalHits>4</totalHits>
<took>20</took>
<status>OK</status>
<providerQuery>false</providerQuery>
<schemaUuid/>
<startIndex>0</startIndex>
<hasMore>false</hasMore>
<executedQueries>
  <executedQuery>{
"bool" : {
  "must" : [
    {
      "match" : {
        "_content.Dc.Title.~" : {
          "query" : "goethe",
          "operator" : "OR",
          "prefix_length" : 0,
          "max_expansions" : 50,
          "fuzzy_transpositions" : true,
          "lenient" : false,
          "zero_terms_query" : "NONE",
          "boost" : 1.0
        }
      }
    },
    {
      "match" : {
        "_content.Dc.Date.~" : {
          "query" : "1820",
          "operator" : "OR",
          "prefix_length" : 0,
          "max_expansions" : 50,
          "fuzzy_transpositions" : true,
          "lenient" : false,
          "zero_terms_query" : "NONE",
          "boost" : 1.0
        }
      }
    }
  ],
  "disable_coord" : false,
  "adjust_pure_negative" : true,
  "boost" : 1.0
}
}</executedQuery>
</executedQueries>
</QueryResult>

```

Abbildung 4: Gekürzte Beispielantwort auf Anfrage aus Abbildung 3

Interaktionsformate

Während die Spezifikation der Suchparameter im Falle der POST-basierten Übergabe auf einer Definition in JSON basiert, können Suchergebnisse beliebig in JSON oder XML zurückgegeben werden.

Die oben verwendete URL <http://search.de.dariah.eu/search/query> versucht hierbei eine automatisierte Auswahl des korrekten Formats auf Basis übergebener Header-Informationen.

Durch die Verwendung von <http://search.de.dariah.eu/search/query.xml> bzw. <http://search.de.dariah.eu/search/query.json> kann das gewünschte Rückgabeformat durch die konsumierende Anwendung selbst vorgegeben werden.

2.2. Föderationsschicht

Aufgrund existierender Publikationen bekannter sind die Schnittstellen der Föderationsschicht, die – wie schon im einleitenden Text dieses Abschnitts erwähnt – zwar vornehmlich für die Verwendung durch die Generische Suche implementiert wurden, aber auch für eine Nachnutzung durch weitere interne und externe Dienste zur Verfügung stehen.

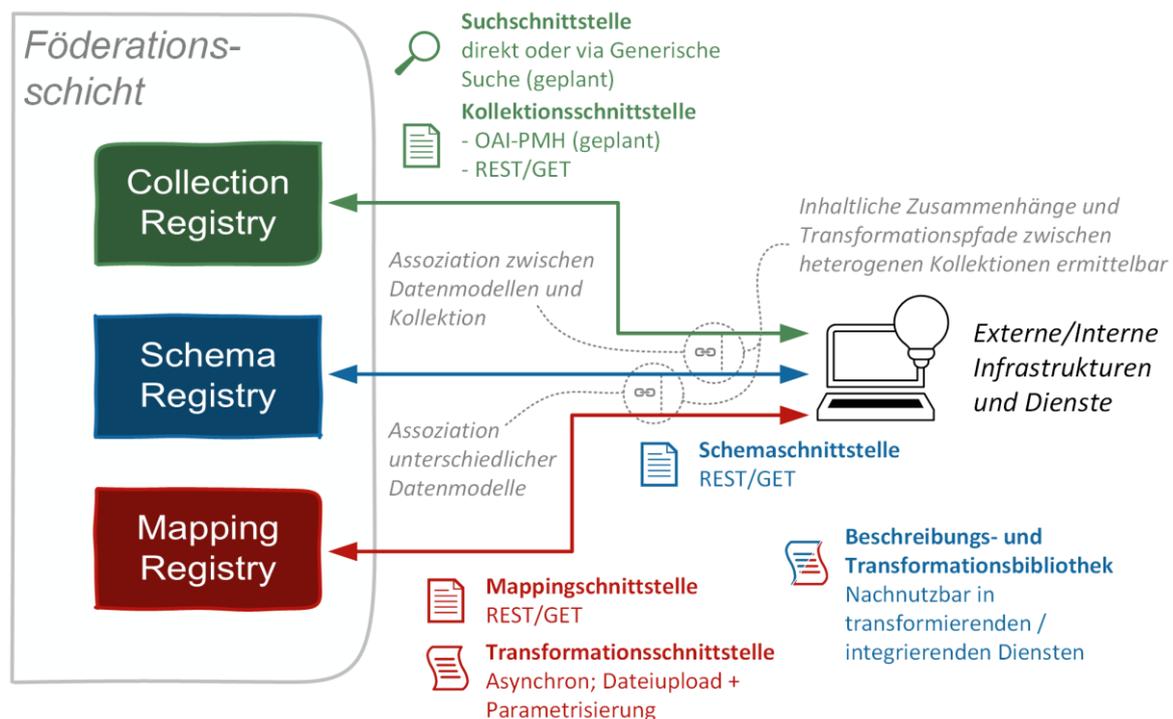


Abbildung 5: Schnittstellen der Föderationsschicht

2.2.1. Collection Registry

Die Collection Registry implementiert derzeit zwei direkt für die Generische Suche und das Repository konzipierte Schnittstellen, um einen erhöhten Overhead durch

ständige Zugriffe zu vermeiden. Insbesondere die Generische Suche greift zyklisch im Abstand von etwa fünf (konfigurierten) Minuten auf die Collection Registry zu und überwacht so, ob neue Sammlungen oder Entwürfe z. B. aus dem Repository aufgenommen wurden. Dieses zyklische Polling wird noch im Rahmen der Laufzeit von DARIAH-DE III durch eine Event-basierte Schnittstelle ersetzt: hiernach wird die Collection Registry bei einer Veränderung ihres Datenbestandes ein gezieltes Update in der Generischen Suche bzw. auch weiteren angebundenen Diensten auslösen. Ein zyklisches Abrufen aller Daten wird zwar vermutlich auch dann noch ausgeführt werden, allerdings in weitaus größeren Abständen.

Die Umsetzung einer OAI-PMH Schnittstelle ist ebenso geplant, wie eine Synchronisationsschnittstelle zwischen mehreren Instanzen der Collection Registry. Letztere ist z. B. für die Umsetzung der Kooperation mit dem Forschungsverbund Marbach Weimar Wolfenbüttel (MWW)³ erforderlich und zielt auf die Nachnutzbarkeit von Kollektionsbeschreibungen.

Derzeit noch unklar ist die Art der Umsetzung einer Suchschnittstelle für die Collection Registry. Hierbei wird noch evaluiert, inwiefern eine Abkopplung und Nachnutzung von Funktionalität der Generischen Suche oder die direkte Abfrage in der Generischen Suche unter Verwendung eines Kollektionsfilters sinnvoller ist.

2.2.2. Schema Registry

In der Schema Registry hinterlegte Datenmodelle können einschließlich aller Interpretations- und Transformationsregeln über eine REST/GET Schnittstelle bezogen werden.

Eine auflistende Abfrage <http://schereg.de.dariah.eu/schereg/api/schemas> liefert dabei basale Informationen zu den Schemata (siehe Abbildung 6).

```
[
  {
    "schema":
    {
      "class": "eu.dariah.de.minfba.core.metamodel.xml.XmlSchema",
      "id": "564ca78fe4b088abc7b7fc62",
      "label": "oai_dc",
      "description": "",
      "versionId": "5824ae7ce4b003fa9f1bbfeb",
      "entityId": "564ca78fe4b088abc7b7fc62",
      "type": "XmlSchema",
    }
  }
]
```

³ Vgl. Abschnitt 3.1

```

        "simpleType": "XmlSchema"
    },
    ...
]

```

Abbildung 6: Listenzugriff auf Schemata

Auf Basis der eindeutigen ID eines Schemas führt eine Einzelabfrage zu der Rückgabe umfassender Informationen zu dem ausgewählten Datenmodell.

Die Anfrage nach der ID 564ca78fe4b088abc7b7fc62 – einem Dublin Core Modell auf Basis des oai_dc Schemas (<http://schereg.de.dariah.eu/schereg/api/schemas/564ca78fe4b088abc7b7fc62>) führt so beispielsweise zu der in Abbildung 7 dargestellten Antwort.

```

{
  "schema": {
    "class": "eu.dariah.de.minfba.core.metamodel.xml.XmlSchema",
    "id": "564ca78fe4b088abc7b7fc62",
    "label": "oai_dc",
    "description": "",
    "terminals": [
      {
        "class": "eu.dariah.de.minfba.core.metamodel.xml.XmlTerminal",
        "id": "564ca7a8e4b088abc7b7fc6b",
        "name": "creator",
        "namespace": "http://purl.org/dc/elements/1.1/",
        "attribute": false,
        "simpleType": "XmlTerminal",
        "type": "eu.dariah.de.minfba.core.metamodel.xml.XmlTerminal"
      },
      ...
    ],
    "versionId": "5824ae7ce4b003fa9f1bbfeb",
    "namespaces": [
      {
        "id": "564ca7a8e4b088abc7b7fc95",
        "prefix": "",
        "url": "http://www.openarchives.org/OAI/2.0/oai_dc/",
        "simpleType": "XmlNamespace",
        "type": "eu.dariah.de.minfba.core.metamodel.xml.XmlNamespace"
      },
      ...
    ],
    "rootElementNamespace":
    "http://www.openarchives.org/OAI/2.0/oai_dc/",
    "rootElementName": "dc",
    "entityId": "564ca78fe4b088abc7b7fc62",
    "type": "XmlSchema",
    "simpleType": "XmlSchema"
  },
  "root": {
    "class": "eu.dariah.de.minfba.core.metamodel.Nonterminal",
    "id": "564ca7a8e4b088abc7b7fc66",
    "entityId": "564ca78fe4b088abc7b7fc62",
    "name": "Dc",
    "terminalId": "564ca7a8e4b088abc7b7fc65",

```

```

"childNonterminals": [
  {
    "class": "eu.dariah.de.minfba.core.metamodel.Nonterminal",
    "id": "564ca7a8e4b088abc7b7fc68",
    "entityId": "564ca78fe4b088abc7b7fc62",
    "name": "Title",
    "terminalId": "564ca7a8e4b088abc7b7fc67",
    "childNonterminals": [
      {
        "class": "eu.dariah.de.minfba.core.metamodel.Nonterminal",
        "id": "564ca7a8e4b088abc7b7fc6a",
        "entityId": "564ca78fe4b088abc7b7fc62",
        "name": "Lang",
        "terminalId": "564ca7a8e4b088abc7b7fc69",
        "transient": false,
        "simpleType": "Nonterminal",
        "type": "eu.dariah.de.minfba.core.metamodel.Nonterminal"
      }
    ],
    "transient": false,
    "simpleType": "Nonterminal",
    "type": "eu.dariah.de.minfba.core.metamodel.Nonterminal"
  },
  ...
],
"transient": false,
"simpleType": "Nonterminal",
"type": "eu.dariah.de.minfba.core.metamodel.Nonterminal"
},
"grammars": {}
}

```

Abbildung 7: Antwort auf Einzelanfrage an Schema Registry (Auszug)

Wie auch die Generische Suche unterstützt die Schema Registry verschiedene Rückgabeformate, die durch Spezifikation üblicher HTTP Header Informationen oder aber die explizite Angabe des gewünschten Formats ausgewählt werden können, so also:

<http://schereg.de.dariah.eu/schereg/api/schemas.json> bzw.

<http://schereg.de.dariah.eu/schereg/api/schemas.xml>.

2.2.3. Mapping Registry

Technisch sind die Komponenten der Schema und Mapping Registry derzeit als eine Instanz zusammengefasst. Sobald es aufgrund der Mengen- bzw. Zugriffsverhältnisse notwendig erscheint können beide Komponenten autonom betrieben werden, weshalb diese auch an dieser Stelle separat voneinander beschrieben werden.

Die Schnittstelle der Mapping Registry unterscheidet sich dabei nur in der Art des aufgerufenen Konzepts von der Schema Registry: mit Hilfe des REST/GET Befehls

<http://schereg.de.dariah.eu/schereg/api/mappings> kann eine Liste verkürzter Informationen zu allen Mappings bezogen werden.

Durch Auswahl einer ID können auch für Mappings sämtliche relevanten Informationen

(<http://schereg.de.dariah.eu/schereg/api/mappings/58147981e4b0bcaaaeb9d43f>)

bezogen werden.

2.3. DARIAH-DE Repository

DARIAH-DE entwickelt ein Repository als digitales Langzeitarchiv für geistes- und kulturwissenschaftliche Forschungsdaten, das nun in einer fortgeschrittenen Beta-version vorliegt. Das DARIAH-DE Repository⁴ ist eine zentrale Komponente der DARIAH-DE Forschungsdaten-Föderationsinfrastruktur, die verschiedene Dienste und Anwendungen aggregiert und so komfortabel nutzbar macht. Das DARIAH-DE Repository erlaubt es, Forschungsdaten nachhaltig und sicher zu speichern, diese mit Metadaten zu versehen und die Forschungsdaten durch die Generische Suche aufzufinden. Jedes Objekt bekommt einen eindeutigen und dauerhaft gültigen persistenten Identifikator (PID), mit dem es nachhaltig referenziert, zitiert und langfristig verfügbar gehalten wird. Darüber hinaus ermöglicht das DARIAH-DE Repository die nachhaltige und sichere Archivierung von Datensammlungen bzw. Kollektionen. Dies ist komfortabel über den DARIAH-DE Publikator⁵, das Web-Interface des DARIAH-DE Portals, im Browser möglich.

⁴ <https://de.dariah.eu/repository>

⁵ <https://de.dariah.eu/publish>

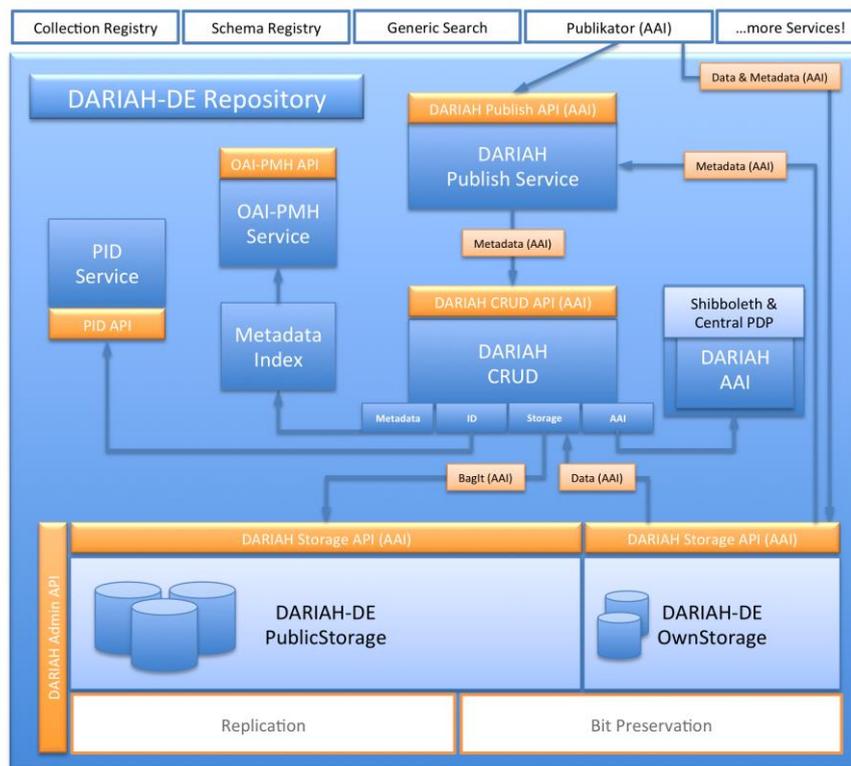


Abbildung 8: Das DARIAH-DE Repository und angeschlossene Dienste der Forschungsdaten-Föderationsinfrastruktur

Daten im DARIAH-DE Repository sind in Kollektionen oder Sammlungen organisiert. Eine solche Kollektion wird zunächst vom Nutzer über den Publikator angelegt und mit Metadaten ausgezeichnet. Einer Kollektion können dann beliebig viele Dateien zugeordnet werden, die ebenfalls über die GUI hochgeladen und mit Metadaten ausgezeichnet werden. Die Daten werden zunächst in einem nur den jeweiligen Forschenden zugänglichen Speicherbereich des DARIAH-DE Storage abgelegt, wo sie – auch über einen längeren Zeitraum hinweg – bearbeitet werden können, bis die Kollektion von den Forschenden zur Veröffentlichung freigegeben wird. Eine publizierte Kollektion sowie alle darin enthaltene Objekte können unmittelbar nach dem Publizieren per Persistent Identifier (PID) referenziert werden und sind damit öffentlich zugänglich. Im nächsten Schritt kann die Kollektion – dieser Schritt wird beim Publikationsprozess vorbereitet – in der DARIAH-DE Collection Registry nachgewiesen und auch dort veröffentlicht werden. Sobald eine Kollektionsbeschreibung der Kollektion selbst auch in der Collection Registry publiziert wurde, und erst dann, sind die Daten auch mit der Generischen Suche von DARIAH-DE recherchierbar.

Das DARIAH-DE Repository stützt sich auf die Codebasis des TextGrid Repository und wurde mit verschiedenen Service-Instanzen und unterschiedlichen – an das DARIAH-DE Repository angepassten – Modulen für Funktionen wie Speicher- und AAI-Zugriff implementiert.

Die aktuelle Beta-Version des DARIAH-DE Repositorys ist vor allem auf den manuellen Ingest (Datenaufnahme) von Forschungsdaten via Publikator und der Collection Registry ausgerichtet. Durch die weiteren Entwicklungstätigkeiten soll in der zukünftigen Produktiv-Version des Repositorys die Möglichkeit gegeben werden, einen „Masseningest“ von Forschungsdaten mithilfe der bereits entwickelten APIs zu gestatten. Eine ausführliche Beschreibung und Anleitung zur Benutzung des Publikators und des DARIAH-DE Repositorys ist im öffentlichen DARIAH-DE Wiki⁶ veröffentlicht.

Das DARIAH-DE Repository wird zukünftig nicht nur DARIAH-DE assoziierten Forschungsprojekten zur Verfügung stehen, sondern auch EinzelforscherInnen sowie Forschungsprojekten, die ihre Forschungsdaten persistent, referenzierbar und langzeitarchiviert speichern und Dritten zur Verfügung stellen wollen. Hierbei steht vor allem der einfache und nutzerorientierte Zugang von FachwissenschaftlerInnen zu einer Langzeitspeicherung von Forschungsdaten im Vordergrund.

2.3.1. Schnittstellen-Dokumentation des Repositorys

Die internen Schnittstellen, die von den Diensten genutzt werden, um untereinander zu kommunizieren, sind alle HTTP und REST-basiert, und nutzen internationale Standards wie RDF und XML. Da einige Teile des Repositorys als eine Einheit organisiert sind, sind diese Schnittstellen nicht von extern zu erreichen und auch nicht öffentlich.

Die Dokumentation der externen Schnittstellen des DARIAH-DE Repositorys⁷ ist noch in Arbeit, allerdings kann an dieser Stelle auf die Schnittstellendokumentation der TextGrid-Dienste⁸ hingewiesen werden, die im Großen und Ganzen auch für das DARIAH-DE Repository zur Verfügung stehen.

⁶ Vgl. Das DARIAH-DE Repository. <https://wiki.de.dariah.eu/display/publicde/Das+DARIAH-DE+Repository>

⁷ Vgl. The DARIAH-DE Repository Documentation. <http://repository.de.dariah.eu/doc/services>

⁸ Vgl. The TextGrid Repository Documentation. <http://textgridlab.org/doc/services>

2.3.2. Öffentliche Schnittstellen der einzelnen Repository-Dienste

Wie in Abbildung 8 zu sehen ist, existieren viele interne Schnittstellen, die Daten zwischen den einzelnen Diensten des Repositories austauschen. Die folgenden externen Schnittstellen, die öffentlich verfügbar sind, sind entweder über die DARIAH AAI abgesichert oder frei nutzbar.

DH-crud public (öffentlich)

- *Version:*

<http://repository.de.dariah.eu/dhcrud/version>

- *Index-Seite:*

<http://repository.de.dariah.eu/dhcrud/11022/0000-0001-328E-7/index>

- *Datei:*

<http://repository.de.dariah.eu/dhcrud/11022/0000-0001-328E-7/data>

- *Deskriptive Metadaten:*

<http://repository.de.dariah.eu/dhcrud/11022/0000-0001-328E-7/metadata>

- *Administrative Metadaten:*

<http://repository.de.dariah.eu/dhcrud/11022/0000-0001-328E-7/adm>

- *Technische Metadaten:*

<http://repository.de.dariah.eu/dhcrud/11022/0000-0001-328E-7/tech>

- *Herkunftsmetadaten:*

<http://repository.de.dariah.eu/dhcrud/11022/0000-0001-328E-7/prov> (noch nicht implementiert)

- *Den gepackten Bagit-Bag* (Datei samt aller Metadaten-dateien in einer standardisierten Bagit ZIP-Datei):

<http://repository.de.dariah.eu/dhcrud/11022/0000-0001-328E-7/bag>

DH-publish

- *Version:*

<http://repository.de.dariah.eu/dhpublish/version>

- *Status und Prozessinformation* (per AAI gesichert):
<http://repository.de.dariah.eu/dhpublish/EAEA0-4540-9327-9156-0/status>
<http://repository.de.dariah.eu/dhpublish/EAEA0-4540-9327-9156-0/info>
<http://repository.de.dariah.eu/dhpublish/EAEA0-4540-9327-9156-0/ministatus>
- *Publizieren* (per AAI gesichert):
<http://repository.de.dariah.eu/dhpublish/EAEA0-4540-9327-9156-0/publish>

OAI-PMH (öffentlich)

Der DARIAH-DE OAI-PMH-Service⁹ bedient alle standardisierten OAI-PMH-Abfragen¹⁰, zusätzlich sind Resumption-Tokens implementiert für eine bequeme Abfrage von Teilinhalten des gesamten Repositorys.

Handle-Service (PIDs / DOIs)

Über den Handle-Service der Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen¹¹ (GWDG) werden die Persistenten Identifikatoren für alle Objekte des DARIAH-DE Repositorys erzeugt. Alle diese PIDs sind technisch gesehen EPIC2-Handles¹² und können über jeden Handle-Service aufgelöst werden:

- <http://hdl.handle.net/11022/0000-0001-328E-7>

Alle mit einem Handle gespeicherten Metadaten lassen sich wie folgt aufrufen:

- <http://hdl.handle.net/11022/0000-0001-328E-7?noredirect>

Für das produktive DARIAH-DE Repository werden sehr wahrscheinlich DOIs¹³ als persistente Identifikatoren eingesetzt, da sowohl die SUB Göttingen als auch die GWDG Mitglied von DataCite¹⁴ sind. DOIs sind in der wissenschaftlichen Community sehr viel weiter verbreitet und weiterhin ist an jeden DOI auch ein verpflichtender Satz von deskriptiven Metadaten gebunden. Technisch gesehen bauen DOIs auf dem System von Handles auf und sind so auch technisch kompatibel.

⁹ <http://repository.de.dariah.eu/oaipmh/oaipmh/>

¹⁰ Vgl. Open Archives Initiative. Protocol for Metadata Harvesting. <http://www.openarchives.org/pmh/>

¹¹ <http://www.gwdg.de>

¹² Vgl. Nachhaltige Referenzierung von Digitalen Objekten mit Hilfe von persistenten Identifikatoren (PID). <https://de.dariah.eu/pid-service>

¹³ Vgl. Digital Object Identifier System. <http://www.doi.org/>

¹⁴ <https://www.datacite.org/>

3. Schnittstellen zu anderen Projekten / Infrastrukturen

Technische Funktionalität und Schnittstellen der DARIAH-DE Föderationsarchitektur haben zu einer Vielzahl von Kooperationen und Anbahnungen geführt. Aktuell aktiv verfolgt und ausgebaut werden hierbei die in Abbildung 9 dargestellten Kooperationen.

Die Kooperationen haben insbesondere gemein, die Ideen und Konzepte der DARIAH-DE Föderationsarchitektur implizit auf ihre Eignung, Übertragbarkeit und Nutzbarkeit im wissenschaftlichen Kontext zu prüfen. Aufgrund der sowohl für die Informatik als auch Geisteswissenschaftlichen Relevanz der durch die jeweiligen Kooperationen adressierten Fragen wird die Platzierung von Publikationen und ggf. eigenständige Weiterführung angestrebt.

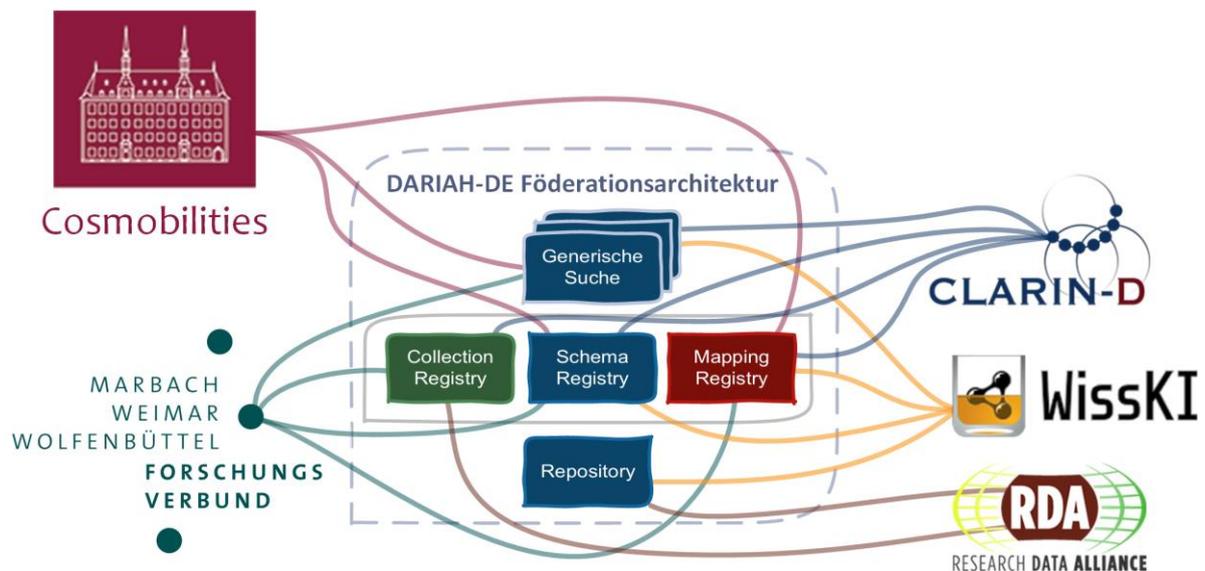


Abbildung 9: Aktuelles Forschungsumfeld zur Föderationsarchitektur

Im Folgenden werden die in der Grafik dargestellten Assoziationen kurz beschrieben und eingeordnet. Hierbei ist insbesondere zu beachten, dass an dieser Stelle weder die projektpolitische Zusammenarbeit, noch die ggf. weitere Zusammenarbeit an anderen Stellen von DARIAH-DE in die Betrachtung einfließen. Da der Fokus des vorliegenden Dokuments auf den Arbeiten von Cluster 4 bzw. insbesondere der DARIAH-DE Föderationsarchitektur liegt, erhebt dieser Abschnitt keinerlei Anspruch auf Vollständigkeit in Hinblick auf die Gesamtmenge der Kooperationen von DARIAH-DE.

3.1. MWW

Der Forschungsverbund Marbach Weimar Wolfenbüttel (MWW) ist ein vom Bundesministerium für Bildung und Forschung (BMBF) seit 2013 gefördertes Projekt. Durch die gemeinsamen Aktivitäten des Verbunds soll insbesondere eine erhöhte Sichtbarkeit der Sammlungen der beteiligten Partner auf nationaler und internationaler Ebene erreicht werden.¹⁵

Die Gesamtinitiative ist in drei Teilprojekte¹⁶ untergliedert – die Föderationsarchitektur von DARIAH-DE unterstützt bei der Umsetzung von Projekt A: Datenmodellierung und Metadaten, sowie Projekt B: Aufbau eines virtuellen Forschungsraum.

Gemeinsam mit dem MWW Verbund entwickelt DARIAH-DE eine auf die Anwendungsdomäne und das Zielpublikum angepasste Variante der Generischen Suche und betreibt eine eigenständige Instanz von Collection, Schema und Mapping Registry. Sämtliche Komponenten – obwohl vollständig autonom – sollen diejenigen Sammlungen, Datenmodelle und Assoziationen mit den Hauptinstallationen der Föderationsarchitektur teilen ohne dass eine redundante Pflege notwendig ist.

3.2. WissKI

WissKI ist eine wissenschaftliche Kommunikationsinfrastruktur für die Dokumentation, Erforschung und Archivierung von Kulturerbe. Von dem Germanischen Nationalmuseum in Nürnberg wurde vorgeschlagen, die auf dem ISO Standard 21127 (CIDOC CRM) basierten öffentlich zugänglichen Forschungsdatensammlungen des GNM zur Nachnutzung an die DARIAH-DE-Infrastruktur anzubinden. Angestrebt ist eine Registrierung der Sammlungen über die DARIAH-DE Collection Registry und die Modellierung und bidirektionale Transformation von Sammlungsdaten mit Hilfe der DARIAH-DE Schema Registry. Langfristig soll es auch möglich sein, Forschungsdaten aus WissKI direkt im DARIAH-DE Repository zu speichern. Die Kooperation von WissKI und DARIAH-DE adressiert eine wachsende Zielgruppe in den Geistes- und Kulturwissenschaften und fördert die Vernetzung von verschiedenen Fachkulturen.

¹⁵ Vgl. <http://www.mww-forschung.de/forschungsverbund-mww/>

¹⁶ Vgl. <http://www.mww-forschung.de/digitale-forschungsinfrastruktur/>

Für die konzeptionelle Schnittstelle der DARIAH-DE Föderationsarchitektur und WissKI wurden vier initiale Use-Cases definiert, deren Umsetzung derzeit prototypisch implementiert und evaluiert wird:

1. Use-Case 1: Import von Daten in WissKI
2. Use-Case 2: Nachnutzung auf WissKI basierender Sammlungen
3. Use-Case 3: Persistente Datenhaltung für WissKI auf DARIAH-DE-Basis
4. Use-Case 4: Ergänzung des DARIAH-DE Dienste-Portfolios

3.3. Cosmopolitanities

Cosmopolitanities ist eine am Institut für Europäische Geschichte (IEG) durchgeführte Machbarkeitsstudie¹⁷, die untersucht inwiefern qualitative Forschung durch quantitative Methoden unterstützt werden kann.

DARIAH-DE implementiert in Cluster 5 einen Prototypen, der auf Basis paralleler Installationen der Collection, Schema und Mapping Registry bestehende Funktionalität von Cluster 4 Komponenten erweitert und Schnittstellen verfeinert. Wie auch für die Kooperation zum MWW Verbund ist für die Übernahme von Sammlungen, Datenmodellen und Mappings eine Schnittstelle zur Synchronisation zwischen parallel installierten DARIAH-DE Föderationsschichten erforderlich.

3.4. CLARIN-D

CLARIN-D ist eine „digitale Forschungsinfrastruktur für Sprachressourcen in den Geistes- und Sozialwissenschaften. Diese Infrastruktur umfasst Sprachdatenbanken, hochgradig interoperable sprachtechnologische Werkzeuge sowie web-basierte sprachverarbeitende Dienste.“¹⁸

DARIAH-DE und CLARIN-D können gut als Geschwisterinfrastrukturen verstanden werden, die inhaltlich ähnliche Zielsetzungen aufweisen, aufgrund unterschiedlicher Anwendungsdomänen jedoch eine an vielen Stellen unterschiedliche Umsetzung dieser Zielsetzungen verfolgen.

¹⁷ Vgl. http://www.ieg-mainz.de/forschungsprojekte/cosmopolitanities_grenzueberschreitende_lebenslaeufe_in_europaeischen_nationalbiographien

¹⁸ Vgl. <https://www.clarin-d.de/de/ueber/kurzuebersicht>

So implementiert CLARIN-D mit Komponenten wie der Suchmaschine des Virtual Language Observatory (VLO)¹⁹ und der CLARIN Centre Registry²⁰ Komponenten, die gegenüber der DARIAH-DE Föderationsarchitektur vergleichbare Ziele mit jedoch verschiedenen, jedoch auch zu DARIAH-DE komplementären Methoden implementieren.

Die Kollaboration mit CLARIN-D auf der Ebene der Datenföderation hat bislang zu einem ersten Austausch von Datenquellen geführt. Mit Hilfe der Schema Registry konnten heterogene CMDI-basierte Schemata assoziiert und in der Generischen Suche integriert werden. Umgekehrt ist der Zugriff auf die in der Collection Registry verzeichneten Kollektionen auf Basis von Dublin Core ein Zugriff auf via das VLO eingerichtet. Erste einfache Schnittstellen zwischen den Infrastrukturen wurden demnach bereits prototypisch umgesetzt und werden im Zuge der Überführung in den nachhaltigen Betrieb beider Projekte nun verfestigt.

3.5. RDA – Research Data Alliance

DARIAH-DE hat unter anderem einige Mitglieder in der *Research Data Repository Interoperability Working Group*²¹ der RDA, die sich damit befasst, Standards für Interoperabilität zwischen verschiedenen Forschungsdatenrepositorien und -Plattformen zu etablieren. Diese Standards werden wahrscheinlich eine generische API sowie Import- und Exportformate beinhalten, weitere Entwicklungen sind denkbar und möglich. DARIAH-DE bringt hier die Schnittstellen des DARIAH-DE Repositories sowie die DARIAH Storage API ein und plant, in der Arbeitsgruppe entwickelte Schnittstellen auch für DARIAH-DE zur Verfügung zu stellen. Auch weitere Dienste von DARIAH-DE wie beispielsweise die Collection Registry werden berücksichtigt.

4. Ausblick

Die technischen Schnittstellen der DARIAH-DE Föderationsarchitektur wurden primär für eine DARIAH-interne Verwendung konzipiert, um durch die Orchestrierung einzelner Dienste komplexere Probleme lösen zu können. Am Beispiel der generischen

¹⁹ <https://vlo.clarin.eu/>

²⁰ <https://centres.clarin.eu/>

²¹ Vgl. <https://www.rd-alliance.org/groups/research-data-repository-interoperability-wg.html>

Suche wird deutlich, dass Lösungen zur Interpretation und Transformation von Daten so von der letztendlichen Anwendung entkoppelt wurden, dass eine Nachnutzung durch weitere Dienste möglich ist und in aktuell durchgeführten Vorstudien zu Kooperationsprojekten bereits eingesetzt werden. Dafür stehen allen DARIAH-Nutzerinnen und -Nutzern die hier dokumentierten externen Schnittstellen der DARIAH-DE Föderationsarchitektur zur Verfügung.

Die DARIAH-DE Föderationsarchitektur ist derzeit in Hinblick auf eine externe und nachhaltige Nachnutzung bereits ein gut durchdachtes und verbreitetes Konzept. Durch die anhaltende Nachfrage von außen können weitere Kooperationen und schrittweise Verfeinerungen der Schnittstellen nutzbar und nachhaltig weiterentwickelt werden.

Literaturverzeichnis

Gradl, Tobias; Henrich, Andreas; Plutte, Christoph (2015): Heterogene Daten in den Digital Humanities: Eine Architektur zur forschungsorientierten Förderung von Kollektionen. In: Constanze Baum und Thomas Stäcker (Hg.): Grenzen und Möglichkeiten der Digital Humanities. Zeitschrift für digitale Geisteswissenschaften 2015 (1). Wolfenbüttel: HAB - Herzog August Bibliothek.

Gradl, Tobias; Henrich, Andreas (2016): Die DARIAH-DE-Föderationsarchitektur – Datenintegration im Spannungsfeld forschungsspezifischer und domänenübergreifender Anforderungen. In: *Bibliothek Forschung und Praxis* 40 (2). DOI: 10.1515/bfp-2016-0027.

Lagoze, Carl; Van de Sompel, Herbert; Nelson, Michael; Warner, Simeon (2002): The Open Archives Initiative Protocol for Metadata Harvesting. Open Archives Initiative. Online verfügbar unter <http://www.openarchives.org/OAI/openarchivesprotocol.html>, zuletzt aktualisiert am 07.12.2008, zuletzt geprüft am 13.01.2014.

Schmunk, Stefan; Funk, Stefan E. (2016): Das DARIAH-DE- und das TextGrid-Repositorium. Geistes- und kulturwissenschaftliche Forschungsdaten persistent und referenzierbar langzeitspeichern. In: *Bibliothek Forschung und Praxis* 40 (2). DOI: 10.1515/bfp-2016-0020.