



Collection Registry (M 3.4.2)

Arbeitspaket 3.4

Verantwortlicher Partner BBAW, GWDG

DARIAH-DE Aufbau von Forschungsinfrastrukturen für die e-Humanities

Dieses Forschungs- und Entwicklungsprojekt wird / wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF), Förderkennzeichen 01UG1110A bis N, gefördert und vom Projektträger im Deutschen Zentrum für Luft- und Raumfahrt (PT-DLR) betreut.

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Projekt: DARIAH-DE: Aufbau von Forschungsinfrastrukturen für die e-Humanities

BMBF Förderkennzeichen: 01UG1110A bis N

Laufzeit: März 2011 bis Februar 2016

Dokumentstatus: final

Verfügbarkeit: öffentlich

Autoren:

Niels-Oliver Walkowski (BBAW)

Daniel Kurzawe (GWDG)

Revisionsverlauf:

Datum	Autor	Kommentare
14.07.2014	Niels-Oliver Walkowski	Zusammenfassung der Beiträge der AG Daten und Sammlungen für den Meilenstein 3.4.2
August 2014	Daniel Kurzawe	Sichtung und Ergänzung
02.07.2016	Niels-Oliver Walkowski	Modifikation Status und Überprüfung

Inhaltsverzeichnis:

1. Einleitung	4
2. Arbeitsschritte und Ablauf	4
3. Beschreibungsschema und kontrollierte Vokabulare	5
4. Definition einer Content Provider Strategie	11
5. Rechtesituation	12
5.1. Fragestellungen, Erläuterungen und Beispiele.....	12
6. Moderation	14
6.1. Workflows.....	14
6.2. Aufgaben und Lösungsansätze im moderierten Ansatz.....	15
6.3. Aufgaben, die eine andauernde Moderation erfordern würden.....	16
7. Anhänge	18
7.1. Beispiel einer Sammlungsbeschreibung.....	18
7.2. Eingetragene Sammlungen.....	20

1. Einleitung

Die BBAW hat zusammen mit der GWDG und der AG Daten und Sammlungen im Rahmen des Meilensteins 3.4.2 132 geisteswissenschaftliche Datensammlungen identifiziert, evaluiert, mit Metadaten erfasst und in die im Meilenstein 1.2.2 von DARIAH entwickelte Collection Registry eingetragen. Der Abschluss des Meilensteins erfolgte durch Eingabe der Metadaten in die Collection Registry durch das dafür vorgesehene Webinterface. Das vorliegende Dokument dokumentiert die geleisteten Arbeiten und beschreibt den Prozess seiner Vollendung.

Kapitel 2 bietet einen kurzen Überblick über die einzelnen Arbeitsschritte, die auf dem Weg zur Vollendung des Meilensteins vollzogen wurden. In Kapitel 3 sind die Ergebnisse der Arbeiten dokumentiert, die zur Erstellung eines geeigneten Beschreibungsschemas für die Sammlungsbeschreibungen führten. In Kapitel 4 wird die Content Provider Strategie präsentiert, die das inhaltliche Profil der Collection Registry und damit der eingetragenen Sammlungen vorgegeben hat. Kapitel 5 gibt Überblick über die Diskussion von möglichen rechtlichen Problemen, die bei der Erfassung von externen Sammlungen innerhalb einer Infrastruktur wie der Collection Registry entstehen könnten. Das Kapitel 6 gibt die Ergebnisse der Evaluation von Moderationsaufgaben und -strategien für die nachhaltige Pflege von Sammlungsbeschreibungen in der Collection Registry wieder. Der Anhang bietet das Beispiel einer Sammlungsbeschreibung sowie eine Liste der Sammlungen, die im Rahmen des Meilensteins in die Collection Registry eingepflegt wurden.

2. Arbeitsschritte und Ablauf

Zu Projektbeginn wurden zwei Kernaufgaben für die Bearbeitung des Meilensteins 3.4.2 identifiziert. Diese bestanden:

1. in der Identifizierung eines geeigneten Schemas zur Beschreibung von Sammlungen, das sowohl den Sammlungen innerhalb des Profils der Collection Registry gerecht werden musste, als auch die Funktion der Collection Registry innerhalb der DARIAH Infrastruktur gerecht werden sollte.
2. die Definition eines inhaltlichen Sammlungsprofils, welches Kriterien dafür anbietet wonach entschieden werden kann, welche Sammlungen in die Collection Registry aufgenommen werden sollen und nach welchen Sammlungen im Rahmen des Meilensteins gesucht werden sollte.

Der erste Schritt für die Bearbeitung der ersten Aufgabe bestand in der Evaluation unterschiedlicher Disziplinen im Hinblick auf benötigte Sammlungsmetadaten. Arbeitspaket 3.4 und die AG Daten und Sammlungen schlossen sich hier einer in Arbeitspaket 3.3 angestoßenen Initiative an.

In der Folge wurden unterschiedliche Schemata zur Beschreibung von Sammlungen analysiert, sowie der Collection Registry vergleichbare Applikationen wie zum Beispiel *EASY.nl*, *ISIDORE*, *DHO:Discovery* untersucht. Das Ergebnis dieser Aktivitäten war die Entscheidung für das Dublin Core Collection Application Profile, einer Dublin Core Variante mit einigen Erweiterungen speziell für die Beschreibung von Sammlungen. Die Entscheidung basierte im wesentlichen auf der weiten Verbreitung von Dublin Core sowie auf dessen niederschwelli-

gen, neutralen Semantik. Das Schema wurde in einem nächsten Schritt für die spezifischen Bedürfnisse in DARIAH leicht angepasst. Weite Teile der von Dublin Core vorgeschlagenen kontrollierten Vokabulare für die Eingabe von Metadaten wurden übernommen und einige hinzugefügt.

Die Vorgaben für die Identifikation von Sammlungen, die im Rahmen des Meilensteins eingetragen werden sollten wurden in einer sogenannten Content Provider Strategie festgehalten. Nach der Definition dieser Strategie wurde zusammen mit allen DARIAH Partnern eine Liste von ca. 155 Sammlungen erstellt, die für den Meilenstein 3.4.2 in Frage kamen. Diese Liste wurde in Zusammenhang mit der Content Provider Strategie evaluiert und priorisiert. Das Ergebnis war ein weitere Liste mit ca. 130 Sammlungen, die an Partner in Arbeitspaket 3.4 und der AG Daten und Sammlungen zur Eintragung in die Collection Registry unter Verwendung des zuvor verabschiedeten Modells verteilt wurden.

Während der Evaluation von Sammlungen sowie der Eintragung der Sammlungen in die Collection Registry wurden zwei Themenbereiche identifiziert, die dicht an die genannten Aufgaben angrenzen und eine hohe Relevanz für die Collection Registry in DARIAH besitzen.

1. die Frage nach der rechtlichen Situation der Collection Registry, sowie
2. Perspektiven für die Moderation von Sammlungen in der Collection Registry, bzw. der Hinzufügung von Sammlungen nach Beendigung des Meilensteins.

Im Kontext des ersten Themas wurden mögliche rechtliche Schwierigkeiten gesammelt und an Hand einiger in die Collection Registry aufgenommenen Sammlungen evaluiert. Das Ergebnis wurde dem Steuerungsgremium vorgelegt.

Ein gleicher Ansatz wurde für die Bearbeitung des zweiten Themas gewählt. Es wurden unterschiedlich betreuungsintensive Szenarien für die Moderation der Collection Registry definiert und angrenzende Aufgaben beschrieben. Außerdem wurde versucht den Ressourcenaufwand zu identifizieren, der für die Bearbeitung der einzelnen Aufgaben benötigt wird. Die Ergebnisse dieser Arbeiten gingen ebenfalls in das Steuerungsgremium.

Schlussendlich wurde das Eintragen der Sammlungen in die Collection Registry für ein *User Testing* der Collection Registry als Applikation genutzt. Die Ergebnisse wurden von den teilnehmenden Partner in das DARIAH Jira eingetragen.

3. Beschreibungsschema und kontrollierte Vokabulare

Dieses Kapitel enthält eine Übersicht über das für die Beschreibung der Sammlungen definierte Schema. Das Schema folgt im wesentlichen dem zuvor erwähnten Dublin Core Collection Application Profile (DCCAP). DCCAP besteht aus 4 verschiedenen Objekten: Sammlung, Agent, Location und Service. Ein Übersicht über die inhaltliche Spezifizierung dieser vier Objekte bieten die folgenden Tabellen in der Reihenfolge ihrer eben erfolgten Nennung.

Die Abkürzungen in der Spalte Status lesen sich wie folgt:

- M=Pflichtfeld
- O=Optional,
- u. a. = und andere Vokabulare, die durch den Benutzer eingetragen werden können,
- 1=maximal einmal zu verwenden
- 1+=ein- oder mehrmals zu verwenden.

Feld	Prädikat	Beschreibung	Werte	Status
Titel	dc:title	Titel der Sammlung	Freitext	M1
Akronym	dclap:acronym	Akronym	Freitext	O1
ID	dc:identifier	Eindeutige Identifikatoren	ColReg (intern); u. a.	M1+
Landingpage	dc:identifier @type landingpage	URL auf die Webpräsenz der Sammlung im Netz	URL	M1
Typ	dc:type	Beschreibungstyp	“Collection”	M1
Beschreibung	dcterms:abstract	Eine Beschreibung der Sammlung	Freitext	O1
Größe	dcterms:extent	Die Größe der Sammlung (Objekte, MB ...)	Freitext	O1
Sprache	dc:language	Sprache der Objekte in der Sammlung	ISO639-3	O1+
Objekttyp	cld:itemType	Typ der Objekte in der Sammlung nach MODS	<u>MODS typeOfResource</u>	O1+
Objektformat	cld:itemFormat	physisches und/oder digitales Format der Objekte in der Sammlung	<u>IANA MimeTypes; Obergruppen</u>	O1+
Rechte	dc:rights	Rechte	vordefinierte Vorschlagsliste	O1
Zugangsbeschränkungen	dcterms:accessRights	Rechte bzgl. des Zugangs zur Sammlung	free free but registration needed paid collection, but free sample/free preview paid content	O1

			paywall	
			not specified	
Erwerbsmethode	dcterms:accrualMethod	Methoden mittels derer Objekte zur Sammlung hinzugefügt werden	Freitext	O1
Aktualisierungshäufigkeit	dcterms:accrualPeriodicity	Frequenz innerhalb derer die Sammlung aktualisiert wird	vorgegebene Liste	O1
Richtlinien	dcterms:accrualPolicy	Richtlinien für der Erstellung der Sammlung	Freitext	O1
Sammlungsgeschichte	dcterms:provenance	Besitzumswechsel und andere relevanten Herleitungsinformationen	Freitext	O1
Zielgruppe	dcterms:audience	Personenkreise für die die Sammlung von Interesse	Freitext	O1+
Schlagwort	dc:subject	Schlagwörter für die Sammlung	DDC; LCSH; GND +	O1+
geographische Relevanz	dcterms:spatial	Bezeichnungen oder Koordinaten für die geographische Relevanz der Sammlung	GeoNames; Getty Thesaurus	O1+
zeitliche Relevanz	dcterms:temporal	Bezeichnungen oder Zeitangaben für die zeitliche Relevanz der Sammlung	ISO8601	O1+
Sammelzeitraum	dcterms:created	Zeitraum innerhalb dessen die Objekte der Sammlung zusammengetragen wurden	ISO8601	O1+
Datierungszeitraum	cld:dateItemsCreated	Zeitraum innerhalb dessen die Objekte der Sammlung erstellt wurden	ISO8601	O1+
Sammler	dc:creator	Person oder Organisation, die als Sammler der Zusammenstellung fungiert	AGENT	O1+
Besitzer	marcrel:own	Person oder Einrichtung, die als Rechteinhaber der Sammlung fungiert	AGENT	M1+
Ort der Sammlung	cld:isLocatedAt	Physischer oder virtueller Ort der die Sammlung hosted	LOCATION	M1+
Zugang zur Sammlung	cld:isAccessedVia	Schnittstelle über die auf eine Sammlung zugegriffen werden kann	SERVICE	O1+
Sub-Sammlung	dcterms:hasPart	Teilsammlung der vorliegenden Sammlung	COLLECTION	O1+
Super-Sammlung	dcterms:isPartOf	Sammlung deren Teil die	COLLECTION	O1

Metadatenschemata	dclap:metadataEncoding-Scheme	vorliegende Sammlung ist Das Schema, welches für die Metadaten der Objekte der Sammlung verwendet wurde	Freitext	O1+
Objektschemata	dclap:itemEncodingScheme	Schema welches in den Objekten der Sammlung verwendet wird	Freitext	O1+

Tabelle 1: Anwendung des Dublin Core Collection Application Profiles (Collection Class) für die Beschreibung von Sammlungen in DARIAH

Feld	Prädikat	Beschreibung	Werte	Status
Titel	dc:title	Name des Agenten	Freitext	M1
Typ	dc:type	Klassentyp	“Person Organisation”	M1
ID	dc:identifizier	Persistente Identifikator	ColReg (intern)	M1+
Acronym	vcard:nickname	“		O1
Adresszusatz	vcard:extended-address	Adresszusatz		O1
Straße	vcard:street-address	“		O1
Post Box	vcard:post-office-box	“		O1
Lokalität	vcard:locality	“		O1
Postleitzahl	vcard:postal-code	“		O1
Region	vcard:region	“		O1
Land	vcard:country-name	“		O1
Telefon	vcard:Tel	“		O1
Handy	vcard:Cell	“		O1
Fax	vcard:Fax	“		O1
Email	vcard:Email	“		O1
Homepage	vcard:url	“		O1

Tabelle 2: Anwendung des Dublin Core Collection Application Profiles (Agent Class) für die Beschreibung assoziierter Akteure von Sammlungen in DARIAH

Feld	Prädikat	Beschreibung	Werte	Status
Titel	dc:title	Titel der Sammlung	Freitext	M1

Feld	Prädikat	Beschreibung	Werte	Status
Typ	dc:type	Klassentyp	“Location”	M1
ID	dc:identifizier	Persistenter Identifikator		M1+
Acronym	vcard:nickname	“		O1
Adresszusatz	vcard:extended-address	Adresszusatz		O1
Straße	vcard:street-address	“		O1
Post Box	vcard:post-office-box	“		O1
Lokalität	vcard:locality	“		O1
Postleitzahl	vcard:postal-code	“		O1
Region	vcard:region	“		O1
Land	vcard:country-name	“		O1
Telefon	vcard:Tel	“		O1
Handy	vcard:Cell	“		O1
Fax	vcard:Fax	“		O1
Email	vcard:Email	“		O1
Homepage	vcard:url	“		O1
Administrator	marcrel:consultant	Ansprechpartner für den beschriebenen Ort	AGENT	O1+

Tabelle 3: Anwendung des Dublin Core Collection Application Profiles (Location Class) für die Beschreibung geographischen Orten an denen Sammlungen administriert oder gehosted werden.

Feld	Prädikat	Beschreibung	Werte	Status
Titel	dc:title	Name der Schnittstelle	Freitext	M1
ID	dc:identifizier	Persistenter Identifikator	ColReg (intern); u. a.	M1+
Typ	dc:type	Klassentyp	“Service” (intern)	M1
Acronym	dclap:acronym	Abkürzung für die Schnittstelle	Freitext	O1
Service Access Method	dclap:serviceAccess-Method	Der Name des Schnittstellentyps über den der Zugriff auf die Daten der Sammlung erfolgt (z. B. OAI-PMH - bisher alleinig von der DARIAH Infrastruktur unterstützt)	Eine Liste mit standardisierten Schnittstellenbezeichnungen wird zur zeit von Christoph erstellt. Die Verwendung von OAI-PMH für die entsprechende Schnittstelle ist bis dahin abgesprochen.	M1
Service Function	dclap:serviceFunction	Die Methoden (Lesen, Schreiben, Aktualisieren u.s.w) die die Schnittstelle zur Verfügung stellt. In den meisten Fällen ein lesender Zugriff	Eine Liste mit standardisierten Bezeichnungen von Servicefunktionen wird zur zeit von Christoph erstellt (voraussichtlich CRUD. Unterstützt wird von der DARIAH Infrastruktur nur READ.	M1+
Service URL	dclap:serviceURL	Basis URL der Schnittstelle, also	URL	M1

Feld	Prädikat	Beschreibung	Werte	Status
Service Subset	dclap:subset	z.B. die OAI-PMH Base URL Die Angabe von Subsets ist nur bei OAI-PMH möglich und werden ansonsten von der Collection Registry automatisiert ermittelt	Name des Subsets	O1+
Schema-ID	dclap:schemaID	ID des vom Service angebotenen Schemas aus der Schema Registry	ID, Interaktion mit Schema Registry	O1
Service Interface	dclap:serviceInterface	Adresse einer wsdl oder anderen Datei, die den Service maschinenlesbar beschreibt	URL	O1
Service Help URL	dclap:serviceHelpURL	URL der menschenlesbaren Dokumentation für die Schnittstelle	URL	O1
Access Control	dclap:accessControl	maschinenlesbare Information bzgl. des Zugangs zur Sammlung	MARC21 - 506 (siehe dcterms:accessRights)	M1
Administrator	marcrel:consultant	Person die die Schnittstelle administriert	Instanz der AGENT Klasse	O1+

Tabelle 4: Anwendung des Dublin Core Collection Application Profiles (Service Class) für die Beschreibung Sammlungsschnittstellen in DARIAH

Vokabular	Namensraum	Prefix
The Dublin Core Metadata Element Set, v1.1	http://purl.org/dc/elements/1.1/	dc
Dublin Core Terms	http://purl.org/dc/terms/	dcterms
Dublin Core Type Vocabulary	http://purl.org/dc/dcmitype/	dcmitype
MARC Relator Code Properties	http://www.loc.gov/loc/terms/relators/	marcrel
Collection Description Terms	http://purl.org/cld/terms/	cld
Collection Description Type Vocabulary Terms	http://purl.org/cld/cdtype/	cdtype
DCLAP	-	dclap
VCard	http://www.w3.org/2006/vcard/ns#	vcard
ISO639-3	http://www.sil.org/iso639-3/	-
ISO8601	http://purl.org/dc/terms/ISO8601	-
IANA Mime Types	http://purl.org/NET/mediatypes (http://mediatypes.appspot.com/)	imt
Dewey Dezimal Klassifikation	http://purl.org/NET/decimalised#d	ddc
Library of Congress Subject Headings	http://id.loc.gov/authorities/subjects	lcsch
Geonames	http://www.geonames.org	gn
Getty Thesaurus of Geographical Names		dc:TGN
MODS	http://www.loc.gov/mods/v3	mods

Tabelle 5: Kontrollierte Vokabulare in der Collection Registry

4. Definition einer Content Provider Strategie

Die Content Provider Strategie umfasst insgesamt 10 Punkte. In dieser spricht sich das Arbeitspaket sowie die AG Daten und Sammlungen dafür aus:

1. dass DARIAH ein aktives Community Engagement mit dem Ziel betreiben sollte, das Akteure Sammlungen in die Collection Registry eintragen und das zu Grunde liegende Infrastrukturen im Sinne der DARIAH Qualitätskriterien weiter entwickelt werden. Für diesen Zweck plädiert das Arbeitspaket und die AG Daten und Sammlungen für die Einkalkulierung entsprechender Ressourcen in Folgeanträgen.
2. dass die AG Prozess und Dienste zusammen mit dem Exekutivkommittee einen nachhaltigen Business Plan für die Collection Registry entwickelt der den Moderationsaufgaben gerecht wird.
3. dass die Collection Registry keine geographischen Restriktionen dafür vorgibt, was die Sammlung inhaltlich repräsentiert oder an welchem Ort sie gehosted wird.
4. dass die Collection Registry keine technischen Restriktionen dafür vorgibt, welche Sammlungen in sie eingetragen werden. Die Vorhandenheit einer maschinenlesbaren API wird als wünschenswert erachtet, ihre Abwesenheit ist jedoch kein Ausschlusskriterium.
5. dass die Collection Registry keine medialen Restriktionen macht. Sammlungen jeglichen Medientyps werden in der Collection Registry akzeptiert.

6. dass die Collection Registry keine Restriktionen im Sinne der institutionellen Herkunft der Sammlung macht. Dennoch werden Sammlungen aus dem Kontext von Forschungsprojekten bevorzugt behandelt, da für Sammlungen in anderen institutionellen Kontexten wie zum Beispiel Museen bereits ein Infrastrukturen zur Verfügung stehen.
7. dass die Collection Registry keine geisteswissenschaftliche Disziplin besonders bevorzugt.
8. dass die Collection Registry sich einer inhaltlichen Qualitätskontrolle der Sammlungen enthält, jedoch die Metadaten über die Sammlung in der Collection Registry evaluiert.
9. dass die Collection Registry Sammlungen, die unter einer offenen Lizenz stehen bevorzugt. Eine offene Lizenz ist jedoch nicht zwingend für die Aufnahme einer Sammlung in die Collection Registry.
10. dass die Collection Registry andere Collection Registries aufnimmt und für diesen Zweck das Prinzip von Collections und Sub-collections einführt.

5. Rechtesituation

In Zusammenhang mit der Bearbeitung des M1.2.2 und M3.4.2 haben sich eine Reihe von Diskussionen über rechtliche Problemstellungen in Zusammenhang mit der Collection Registry entwickelt. Diese Problemstellungen sind nicht abhängig von der Frage ob Sammlungen lediglich mit Metadaten beschrieben oder auch gehosted werden.

Da für die laufende Projektphase keine juristische Hilfe vorhanden ist, wird laut Beschluss der AG Daten versucht, eine rechtliche Absicherung auf Basis der Terms of Use der Collection Registry zu ermöglichen, die ein Content Provider der Collection Registry akzeptieren muss, bevor dessen Sammlungsbeschreibung überhaupt in die Collection Registry eingetragen wird. In dieser Hinsicht soll also vorerst unter Zuhilfenahme eines Disclaimers verhindert werden, dass z.B. urheberrechtlich geschützte Metadaten veröffentlicht werden.

5.1. Fragestellungen, Erläuterungen und Beispiele

Ziel der Collection Registry ist es, Informationen und Metadaten über Forschungsdatensammlungen und Quellensammlungen über vielfältige Suchmöglichkeiten geisteswissenschaftlichen Forschern zur Verfügung und zur Nachnutzung zu stellen. Es muss von Seiten DARIAHs sichergestellt werden, ob die von DARIAH den Nutzern zur Verfügung gestellten Informationen dem geltenden Urheber- und Leistungsschutzrecht entsprechen. Eine Vielzahl rechtlicher Rahmenbedingungen ist hierbei zu klären. Prinzipiell muss davon ausgegangen werden, dass selbst Metadaten von Sammlungen - sofern der Sammlung direkt entnommen - nicht Open-Access sind und damit frei verwaltet, verändert und angeboten werden können, sondern dass vor Aufnahme von Sammlungsbeschreibungen zu prüfen ist, ob dies aufgrund von Urheber- und Leistungsschutzrechten und auch z.B. aufgrund von der Sicherstellung von Persönlichkeitsrechten überhaupt rechtlich möglich ist. Aus diesem Grund können nur Sammlungsbeschreibungen übernommen werden, auf denen keine Zugangs- und Nutzungsbeschränkungen liegen und bei denen eindeutig die Urheberschaft zu identifizieren ist. Inwieweit diese

Urheber- und Copyrightprobleme vorhanden sind, wenn die Sammlungsbeschreibung während der Registrierung der Collection in Eigenarbeit erstellt wird, ist ungeklärt. Persönlichkeitsrechte können auch in diesem Fall verletzt werden. Folgende Aspekte sollten deshalb beachtet werden.

5.1.1. Bedarf es zur Eintragung einer Sammlung einer Anfrage beim Urheber?

Findbücher und Informationen aus Sammlung werden beispielsweise an Wissenschaftler nur zum eigenen Gebrauch weitergegeben. Eine weitere Verwendung der Informationen wird oftmals untersagt. Insofern ist der einzelne Wissenschaftler in der Regel nicht Rechteinhaber, sondern lediglich Nutzer von Sammlungsbeschreibungen. Aus diesem Grund muss vor Eintragung die Provenienz der Sammlungen geklärt werden und zweitens ob der Rechteinhaber mit einer Veröffentlichung einverstanden ist. Solche terms of use bzw. Nutzungsvereinbarungen werden standardmäßig vom Bundesarchiv und fast allen öffentlichen Archiven (Kirchen, Verbände, Parteien, Stiftungen usw.) der Bundesrepublik benutzt.

Vorschlag: pragmatischer Umgang für M3.4.2: Wenn eine Sammlung im Web frei zugänglich ist, steht einer Eintragung nichts entgegen, da da von ausgegangen wird, dass dies einer öffentlichen Publikation gleichkommt.

5.1.2. Werden im Zweifelsfall Persönlichkeitsrechte durch die aufgenommenen Metadaten verletzt?

Neben der Klärung der Frage des Urheber- und Leistungsschutzrechtes ist bei jedem Eintrag zu klären, ob z.B. Persönlichkeitsrechte durch die Metadatenbeschreibungen verletzt werden. So können beispielsweise in Sammlungsbeschreibungen Namen von Personen genannt werden, die entweder noch leben, oder die erst vor kurzem verstorben sind. Hier stellt sich die Frage, ob von diesen die Genehmigung und Zustimmung vor einer Veröffentlichung eingeholt werden? Gerade Quellenbestände aus dem 20. Jahrhundert - so z.B. aus der Zeit des Nationalsozialismus - könnten hier zu rechtlichen Schwierigkeiten führen. Beispielsweise könnten in Bestandsbeschreibung (Collections) Namen von Personen angegeben sein, die einer Veröffentlichung per se nicht zugestimmt haben und dementsprechend die Aufnahme zu einer Verletzung der Persönlichkeitsrechte führen würde.

Vorschlag: Wenn die Metadaten auf Basis der öffentlich zugänglichen Daten der Sammlung oder in Kommunikation mit Provider erhoben werden, können keine Rechte verletzt werden.

5.1.3. Bedarf es für die Metadaten zu den Sammlungen eines License Agreements mit dem Metadatenersteller?

Diese Frage wurde in Zusammenhang mit Frage 5.1.3 evaluiert.

5.1.4. Wie müssen die Terms of Use gestaltet werden, um uns rechtlich abzusichern?

DARIAH ist europäisch ausgerichtet, Urheber- und Leistungsschutzrechte sind hingegen nationalstaatlich ausgerichtet und von einer Vereinheitlichung weit entfernt. Es muss geklärt werden, welche Bedeutung dies für die Collection Registry hat und inwiefern man dies operationalisieren kann.

Archive - ob staatliche, kommunale, private oder auch z.B. kirchliche - haben unterschiedliche Nutzungsbedingungen und Beschränkungen. Wie ist sichergestellt, dass diese in der Collection Registry abgebildet werden? Dürfen Veränderungen/Ergänzungen vorgenommen werden, ohne dass die originäre Information verändert wird?

Die ausgearbeitete Problematik wurde dem Steuerungsgremium zur Diskussion von Lösungsvorschlägen vorgelegt.

6. Moderation

6.1. Workflows

Workflow 1: DARIAH identifiziert eine Sammlung, die in der CR registriert werden soll.

1. Identifikation der Sammlung durch DARIAH
2. Identifikation eines Verantwortlichen für die Sammlung und Bitte um Eintragung der Sammlung
3. *Sofern Verantwortlicher nicht in AAI registriert:* Benutzerregistrierung
4. Registrierung der Sammlung
5. *Sofern License Agreement nötig:* License Agreement muss abgeschlossen werden
6. Validierung der Metadaten (automatisch oder manuell)
7. Sammlung wird registriert

Workflow 2: Registrierter Benutzer möchte eine bestehende Sammlung in die CR einbringen

1. Benutzer möchte Sammlung registrieren
2. AAI prüft die Benutzerrechte
3. Benutzer registriert Sammlung
4. *Bei aktiver Review:* Administrator/Maintainer der CR überprüft Vorschlag anhand der Vorgaben aus der Content Strategy
5. Prüfung, ob Duplikation einer existierenden Sammlung
6. *Bei aktiver Review:* Verfügungsgewalt des Benutzers über die Sammlung muss geklärt werden.
7. *Sofern License Agreement nötig:* License Agreement muss abgeschlossen werden
8. Validierung der Metadaten (automatisch oder manuell)
9. Sammlung wird registriert

Workflow 3: Nicht-registrierter Benutzer möchte eine Sammlung vorschlagen

- Fall 2: Nicht-registriertem Benutzer sind keine Vorschlagsrechte erlaubt
 1. Benutzer muss sich registrieren
 2. *Weiter bei Workflow 2*

Workflow 4: Registrierter Benutzer möchte eine Sammlung editieren

- Fall 1: Benutzer hat Sammlung registriert
 1. Metadaten werden überarbeitet

- 2. Validierung der Metadaten (automatisch oder manuell)
- Fall 2: Benutzer hat Sammlung nicht registriert
 - 1. Benutzer der Sammlung registriert hat wird benachrichtigt
 - 2. weiter Fall 1

Workflow 5: Registrierter Benutzer möchte eine Sammlung löschen

- Fall 1: Benutzer hat Sammlung registriert
 - 1. Benutzer meldet Sammlung zur Löschung an (Mit Begründung)
 - 2. *Bei aktiver Review*: Begründung wird durch Administrator geprüft
 - 3. Sammlung wird gelöscht
- Fall 2: Benutzer hat Sammlung nicht registriert
 - 1. Benutzer meldet Sammlung zur Löschung an (Mit Begründung)
 - 2. Benutzer der die Sammlung registriert hat wird benachrichtigt und prüft Wunsch zur Löschung
 - 3. *Bei aktiver Review*: Begründung wird durch Administrator geprüft
 - 4. Sammlung wird gelöscht

Workflow 6: Externer Projektantrag sieht Kooperation mit DARIAH CR vor

1. Kooperationsvorlage muss geprüft werden
2. DARIAH Recommendations und Policies sollten Eingang in den Projektantrag finden
3. Guidelines zur Registrierung der Sammlung in die Collection Registry sollten im Projektarbeitsplan verankert sein
4. Projektpartner muss mit entsprechenden Rechten in der AAI registriert werden
5. *Weiter bei Workflow 2*

6.2. Aufgaben und Lösungsansätze im moderierten Ansatz

6.2.1. Aufgaben, die von DARIAH in der Projektzeit erfüllt werden sollten

- Es müssen Anleitungen und Empfehlungen für den Umgang mit der Collection Registry geschrieben werden
 - Wie wird eine Sammlung eingetragen, aktualisiert, gelöscht (AP1.2, AP3.4)
 - Was für ein Standard wird verwendet, was für Authority Files, best practices (AP3.3)
- License Policy und Terms of Use müssen formuliert werden (AP3, AP1, AG Daten, benötigt einen Juristen)
 - Terms of Use Kriterien aus diesem Dokument für die Content Provider Rolle
 - Beschreibungen dürfen keine sensiblen Daten enthalten
 - Verletzt mit dem Eintragen der Sammlung keine Rechte oder Persönlichkeitsrechte Dritter.
 - Eintragung erfordert kein Einverständnis des Rechthebesitzers oder das Recht ist eingeholt worden

- Absicherung: Sammlung ist keine Closed Collection, Existenz der Sammlung ist öffentlich
- Content Provider muss aus einem Forschungszusammenhang kommen
- Sammlung muss aus einem Forschungszusammenhang kommen
- Metadaten, die bei der Beschreibung erstellt werden, werden unter eine offene Lizenz gestellt
- Beschreibung muss nach bestem Wissen und Gewissen korrekt sein
- Business Model für die langfristige Aufrechterhaltung der Collection Registry muss entwickelt werden
- Klärung der rechtlichen Situation der Collection Registry (Eine ausführlichere Beschreibung der juristischen Problematiken der Collection Registry siehe vorheriges Kapitel)
 - Bedarf es zur Eintragung einer Sammlung einer Anfrage beim Urheber
 - Werden im Zweifelsfall Persönlichkeitsrechte durch die aufgenommenen Metadaten verletzt
 - Bedarf es für die Metadaten zu den Sammlungen eines License Agreements mit dem Metadatenersteller
 - Wie müssen die Terms of Use gestaltet werden, um uns rechtlich abzusichern
 - Bei der Eintragung von Collections Prüfung von Urheber- und ggf. Persönlichkeitsrechte

6.3. Aufgaben, die eine andauernde Moderation erfordern würden

- Sofern der zukünftige Eintragungsprozess von DARIAH kontrolliert werden soll bedarf es einer aktiven Entscheidung und Kontrolle für die Erstellung einer Sammlung. Um den Moderationsaufwand zu kontrollieren, könnte eine benutzergesteuerte Filterung eingerichtet werden. Dieser Prozess könnte ebenfalls den Arbeitsaufwand erleichtern, sofern 3.1.3 zugunsten nicht-registrierter Benutzer entschieden wird.
 1. Benutzer schlägt eine Sammlung vor
 2. Vorschlag wird in einer Liste veröffentlicht
 3. Benutzer können für Vorschläge voten
 4. Die Vorschläge werden priorisiert und die mit den meisten Stimmen führen zur Erstellung JIRA-Tickets und zur Bearbeitung
- Qualitätsmanagement und Qualitätskontrolle zur Aufrechterhaltung der Qualität der Metadaten in der Collection Registry. Für diesen Prozess braucht es einen aktiven Moderator, der durch ein benutzergesteuertes System und ein Alerting Systems entlastet werden kann.
 1. Plan - Einarbeitung von Qualitätsverbesserungsmaßnahmen in die Content Strategy
 2. Do - Datenbereinigung

- Aussortieren überholter Sammlungen
- Aussortieren nicht mehr erreichbarer Sammlungen
- Überarbeiten überholter Beschreibungen von Sammlungen
- 3. Check - Qualitätskontrolle
 - Bewertung der Metadaten nach Mindestanforderungen aus den Empfehlungen/Richtlinien für CR und Schemakonsistenz
- 4. Act - Unterstützung durch Benutzersysteme
 - zum Qualitätsmanagement: Benutzer oder URL-Resolver prüft Metadaten und meldet eine Unstimmigkeit; JIRA Ticket wird erzeugt; Moderator bearbeitet Ticket
 - zur Qualitätskontrolle: Benutzer gibt Bewertung für die Beschreibung einer Sammlung ab
- Langfristiger Aufwand für das Hosting der Collection Registry (bemessen auf 4-5 Stunden/Woche)
 1. Aktualisierung
 - Einspielen von Updates (normal und sicherheitsrelevant)
 - Sicherheitskonzepte pflegen
 - Anpassung der Konfiguration bezgl. Veränderungen im Hosting oder zum Zugriffsschutz
 - Systemtests
 - Monitoring
 2. Überwachung der Dienstfunktionalität
 - Fehlerbehebung
 - Systemneustarts
 - Fehleranalyse
 - Behebung von Konfigurationsfehlern
 - Pflege der Hardware
 3. Konfiguration
 - Anpassungen für neue Nutzungsszenarien oder Forschungsprojekte
 4. Backup
 - Datensicherung
 - Konfigurationssicherung
- Ein aktives Community Engagement und die Implementierung der Content Provider Strategy erfordern den größten Zeitaufwand, würden aber die Tragweite und Attraktivität von DARIAH erheblich vergrößern.
 1. Nutzen
 1. Ein Reagieren auf Nutzerbedürfnisse ist möglich.
 2. Das Profil der Collection Strategy durch gezielte Suche kann gestärkt werden.
 3. Die Informationspolitik zu Sammlungen kann verbessert werden (z. B. gibt es eine nicht dokumentierte API).
 4. Es kann auf Entwicklungen aktiv Einfluss im Sinne der DARIAH Empfehlungen genommen werden.

5. Kooperationen können zu einer Bindung an DARIAH und zu einer Verbesserung des DARIAH Angebotes führen.
 6. Weiterentwicklungen (API) oder Lizenzveränderungen (hin zu offenen Lizenzen) können angeregt werden.
2. Aufwand
 1. Aktive Recherchearbeit
 2. Interaktion mit Benutzern
 3. Unterstützung von Entwicklungsprozessen
 4. Schulungen
 5. Juristische Arbeit
 3. Unterstützung
 1. Einrichtung von Alertingalgorithmen zur Meldung von neuen interessanten Sammlungen innerhalb der DARIAH Content Provider Strategy auf der Basis definierter Quellen
 2. Berücksichtigung von API Wrappern für bestehende Systeme in den Empfehlungen und durch How-To's
 3. Archive-in-a-Box
 4. Erstellung von Kooperationsvorlagen

7. Anhänge

7.1. Beispiel einer Sammlungsbeschreibung

Im folgenden soll das DCLDAP am Beispiel von Arachne des Deutschen Archäologischen Instituts demonstriert werden:

Feld	Wert
Type	Collection
Collection Identifier (1 internal + opt.)	6546547584765
Title	Arachne
Description	Arachne is the central Object database of the German Archaeological Institute (DAI) and the Archaeological Institute of the University of Cologne (for further info see Arachne's homepage). It contains DAI's collection of images of archaeological objects and books (iDAI.BookBrowser) related to Classical Archaeology.
Size	836.142 digital Objects of 250.000 physical Objects
Language	lat, grc
Item Type	Image, Physical Object, Text
Item Format	Mosaik, Malerei, Vasen, Plastik, Idealplastik, Porträtplastik, Sarkophage, Reliefs, Figürliche Bauplastik, Inschriften, Stuck, Terrakotta, Bronze, Kleinkunst, Denkmäler, Architektur
Item Format @scheme=IANA	image/jpeg, pdf
Rights	Some contents are licensed under Creative Commons, others are not.
Access Rights	The Access to the content via webinterface is not restricted

Feld	Wert
Accrual Method	-
Accrual Periodicity	-
Accrual Policy	-
Custodial History	-
Audience	Classicists, Archaeologists
Subject @scheme=LCSH	Archeology, Photography in archaeology, ...
Subject @scheme=DDC	930
Spatial Coverage @scheme=Getty	Hellas, ...
Temporal Coverage	-21000/+1400
Dates Collection Accumulated	-
Dates Items Created	-
Relationships between the Collection and Agents	
Collector	AGENT ID, http://d-nb.info/gnd/160680751
Owner	AGENT ID, http://d-nb.info/gnd/160680751
Relationships between the Collection and Location, Collection and Service	
Is Located At	LOCATION ID
Is Accessed Via	SERVICE ID
Relationships between Col- lections (and between Col- lections and Catalogues or Indices)	
Sub-Collection	-
Super-Collection	-
Additional Fields of the DARIAH Collection Regis- try	
Metadata Scheme	CIDOC-CRM, Dublin Core, METS/MODS, Arachne XML
Item Scheme	jpeg, pdf, TEI
Agent	
Title	Reinhard Foertsch
Type	Person
Identifiers @scheme=PND	http://d-nb.info/gnd/160680751
Acronym	
Street	
Post BOX	
Locality	
Post Code	
Region	
Country	
Phone	
Mobile Phone	

Feld	Wert
Fax	
Email	
Homepage	http://www.arachne.uni-koeln.de/
Service	
Title	Arachne OAI-PMH Schnittstelle
Identifizier	
Type	Service
Acronym	OAI-PMH
Service Access Method	OAI-PMH
Service Function	READ
Service URL	http://arachne.uni-koeln.de:8080/OAI-PMH/oai-pmh.xml
Service Interface	
Service Help URL	http://www.arachne.uni-koeln.de/drupal/?q=de/node/234
Acces Control	Open
Administrator	AGENT

Tabelle 6: Beispiel einer Sammlungsbeschreibung

7.2. Eingetragene Sammlungen

- Arachne
- Berliner Klassik Datenbanken
- Carl-Maria-von-Weber-Complete-Edition
- CEEC – Codices Electronici Ecclesiae Coloniensis
- Controversia et Confessio
- Deutsches Textarchiv
- Digital Averroes Research Environment
- Digitales Wörterbuch der deutschen Sprache
- Digitale Bibliothek in TEI
- Edoc Server
- Epigraphische Datenbank
- European History Online
- Idai.Field

- Jahresberichte für deutsche Geschichte online
- MEGA Digital
- Monasterium/VdU
- Monumenta Germaniae Historica
- OPAC des IEG
- Personendaten-Repository
- Quarterly Kalonymos
- Ritterkatalog
- Thesaurus Linguae Aegyptiae
- Universal-Kirchenzeitung
- Verteilte Digitale Inkunabelbibliothek
- Wissensspeicher
- ZENON
- Ägyptisches Totenbuch
- Bone Commons
- CoNE
- e-codices
- Global Performing Arts Database
- Homer MultiText project
- Open Context
- Virtuelles Skriptorium St. Matthias
- Digitale Keilschriften Bibliothek – cdli
- Perseus
- Deutsch-Jüdische-Publizistik
- Leopold-Zunz-Archiv

- Regesta Imperii
- Computer Assisted Tools for Septuagint
- digitale Monumenta Germaniae Historica
- Düsseldorfer Theaterzettel
- Gesamtkatalog der Wiegendrucke
- Gottlob Freges Grundgesetze der Arithmetik
- International Women's Periodicals
- Kant Corpora
- Kinematic Models for Design
- Kritikon
- LIMAS-Korpus
- Nietzsche Source
- Regionaldatenbank Archäologie Museum der Westlausitz
- Yad Vashem Photo Archive
- The Archimedes Project
- NAVIS I
- NAVIS II
- NAVIS III
- TOMBA
- Briefe und Texte aus dem intellektuellen Berlin um 1800
- UbuWeb
- Kant Opus Postumum
- Corpus Coranicum
- Registres de l'Academie
- Monumenta Germaniae Historica

- Alexander von Humboldt Abhandlungen
- Alexander von Humboldt – Christian Gottfried Ehrenberg Briefwechsel
- Die unselbständigen Schriften Alexander von Humboldts
- Corpus Inscriptionum Latinarum
- Corpus Medicorum Graecorum/ Latinorum
- Deutsche Texte des Mittelalters
- Inscriptiones Graecae
- Magazin zur Erfahrungsseelenkunde
- Turfanforschung
- Digitales Bildarchiv des Bundesarchivs
- EMTO
- Hethitische Forschungen
- Historical Monographs Collection
- JSTOR Data for Research
- Thesaurus Linguae Graecae
- Wittgenstein Source
- Edition der politischen Schriften Leibniz
- Katalog von Heidelberger Papyri, Pergamenten, Haderpapieren, Griechisch
- Kataloge von Heidelberger Papyri, Pergamenten, Haderpapieren, Demotisch
- Kataloge von Heidelberger Papyri, Pergamenten, Haderpapieren, Koptisch
- International Dunhuang Project
- Christian Classics Ethereal Library
- Judaica-Sammlung Frankfurt
- Digital Image Archive of Medieval Music (DIAMM)
- Chopin's First Editions Online (CFEO)

- Bildkulturen ökologischer Forschung
- Bonner Frühneuhochdeutschkorpus
- Codice diplomatico della Lombardia Medievale
- Corpus Vitrearum Medii Aevi
- Deutsche Inschriften Online
- Deutsche Geschichte in Dokumenten und Bildern (DGDB)
- Forschungsstelle für Personalschriften Online
- KernScores
- LBA
- Mittelalterliche Handschriften in österreichischen Bibliotheken
- Online Catalogue of Musical Sources
- Lippisches Landesmuseum Archäologische Sammlung
- Kreismuseum Jerichower Land, Sammlung Archäologie / Paläontologie
- Sachsen-Anhalt und das 18. Jahrhundert
- Archäologie der anhaltischen Fürsten
- Regionalgeschichte.net
- Reichserzkanzlerarchiv Online
- textlog.de
- VD 16
- VD 17
- Numismatische Bilddatenbank Eichstätt
- Online-Filmdatenbank
- Cinovid – database for experimental film and video art
- Sekundärliteraturdatenbank zu Alexander von Humboldt
- Virtuelles Berlin um 1800

- Census of Antique Works of Art and Architecture Known in the Renaissance
- Wilhelm von Humboldt – Briefdatenbank
- Online Chopin Variorum Edition (OCVE)
- Arnold Schönberg Center
- Beethoven- Haus Bonn Digitales Archiv
- Thesaurus Lingua Graece
- NT-VMR (New Testament Virtual Manuscript Room)
- Bach Digital
- Strauss Online
- Medieval Aschkenas Online
- Race, Ethnicity and Religion
- SPK-digital
- Arsenal
- Schubert-Autographe
- Brahms Digital
- DigiBaeck | Leo Baeck Institute
- Das wissenschaftliche Bibelportal der Deutschen Bibelgesellschaft
- Prosopographie der mittelbyzantinischen Zeit
- Dt. Kinemathek
- Internet Movie Database
- Luxonline – Educational Resource about British Film and Video Artists
- Video Data Bank
- Electronic Arts Intermix
- imai
- NIMk

- Moralische Wochenschriften
- Théâtre classique
- Les Bibliothèques virtuelles humanistes
- Old Bailey Online
- David Rumsey Historical Map Collection
- Epigraphic Database Heidelberg
- Europeana
- Wörterbuchportal
- Porta Hebraica
- Corpus Vitrearum Medii Aevi Deutschland/Potsdam
- Demokratiegeschichte
- Digitales Flurnamenlexikon
- Arbeitskreis Filmbibliotheken
- Video History Project
- Digitale Mozart- Edition
- Pseudoepigrapha, Apocrypha, and Sacred Writings
- Perseus Digital Library
- Gesamtkatalog deutschsprachiger Leichenpredigten (GESA)
- Titelblattkatalog der Leichenpredigten und sonstiger Trauerschriften (TBK)
- Bibliographie zur Leichenpredigten-Literatur
- Thesaurus Locorum (THELO)
- Thesaurus Professionum (THEPRO)
- Sicherungsverfilmung ausgewählter deutschsprachiger Altbestände (SIBRES)
- Wittgenstein Source: Bergen Text Edition (BTE)
- Wittgenstein Source: Bergen Facsimile Edition (BFE)

- Constitutiones et acta publica imperatorum et regum
- Digitale Sammlung Universität Frankfurt
- Lichtbildarchiv älterer Originalurkunden bis 1250 (LBA)