



Eintragung wissenschaftlicher Sammlungen in die Collection Registry (M 4.2.5)

Version 21.2.2016
Cluster 4
Verantwortlicher Partner HAB

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: Dario Kampkaspar, HAB

Revisionsverlauf:

Datum	Autor	Kommentare
21.02.2016	Dario Kampkaspar	Erstfassung
21.02.2016	Dario Kampkaspar	Finalisierung

Inhaltsverzeichnis

1. Ausgangssituation.....	4
2. Bewertungskriterien für die Auswahl von Sammlungen	4
3. Weitere Übernommene bzw. in Übernahme befindliche Sammlungsbeschreibungen.....	7

1. Ausgangssituation

Mit der Collection Registry entwickelt DARIAH-DE einen Service, der verschiedene Sammlungen von Forschungsdaten¹ unabhängig von ihrem Aufbewahrungsort, ihres Datenformates und der Form ihrer Beschreibung am Sammlungsort vergleichbar recherchierbar macht. Sie bildet damit einen Baustein einer Gruppe von Werkzeugen zur übergreifenden Beschreibung und Durchsuchbarkeit von Forschungsdaten.

Eine grundlegende Herausforderung ist dabei die Modellierung eines Beschreibungsmodells, das eine größtmögliche Flexibilität bei gleichzeitiger hoher Vergleichbarkeit bietet. Um diesem Umstand gerecht zu werden, wurde das zugrunde liegende Datenmodell im Laufe von DARIAH-DE II angepasst², was notwendige Nacharbeiten an der Collection Registry nach sich zog.

Ziel der Arbeit von Cluster 4, *Wissenschaftliche Sammlungen*, in der II. Förderphase war unter anderem die gezielte Auswahl bestehender Sammlungen aus Datenbanken und ihr Eintrag in die Collection Registry. In diesem Dokument sollen die erarbeiteten Auswahlkriterien, die gewonnenen Erkenntnisse und weitere Sammlungen, die zum Eintrag kommen werden, zusammengefasst werden. Für detailliertere Beschreibungen wird dabei auf weitere im DARIAH-DE-Kontext veröffentlichte Texte verwiesen.

2. Bewertungskriterien für die Auswahl von Sammlungen

Für die initiale Übernahme von bestehenden Sammlungen in die Collection Registry wurde ein Kriterienkatalog erarbeitet, anhand dessen eine Liste von Aufnahmekandidaten erstellt wurde. Diese Kriterien gelten aber nicht nur für die Übernahme von

¹ Die Erarbeitung einer Definition von »Forschungsdaten« im Kontext von DARIAH-DE erwies sich als außerordentlich komplex. Die endlich genutzte Definition findet sich unter <https://de.dariah.eu/forschungsdaten>.

² Vgl. hierzu auch R 4.2.4, *Aufbau und Nutzung wissenschaftlicher Sammlungen* (<https://wiki.de.dariah.eu/download/attachments/26150061/R4.2.4-final.pdf?version=1&modificationDate=1458823511080&api=v2>) und R 4.2.1, *Modellierung und Dokumentation von Use-Cases für wissenschaftliche Sammlungen* (<https://wiki.de.dariah.eu/download/attachments/26150061/R%204.2.1%20Use%20Cases%20zur%20Sammlungsmodellierung.pdf?version=1&modificationDate=1446624763793&api=v2>). Eine ausführliche Beschreibung des aktuellen Datenmodells wird zu einem späteren Zeitpunkt separat veröffentlicht werden.

Daten aus anderen Datenbanken, sondern bieten auch eine grundsätzliche Checkliste, anhand derer sich bewerten lässt, inwiefern eine Eintragung der Sammlung in die Collection Registry sinnvoll ist.

Ein Hauptaugenmerk liegt dabei darauf, dass die produzierten Daten für die Wissenschaft frei verfügbar sind und durch bereitgestellte Schnittstellen in offenen und üblichen Formaten abgerufen werden können. Keiner dieser Punkte stellt jedoch ein Ausschlusskriterium dar, das eine Aufnahme in die Collection Registry verbietet. Vielmehr sollen im Endzustand so viele Sammlungen wie möglich verzeichnet sein, da der Nutzen der DARIAH-DE-Dienste zur übergreifenden Suche mit der Zahl der verzeichneten Sammlungen steigt. Dies gilt insbesondere für die hohe Zahl an Datensammlungen, die im Rahmen befristeter Projekte entstanden sind und die häufig kaum bis gar nicht außerhalb ihres ursprünglichen institutionellen Horizontes sichtbar sind.

Es wurde beschlossen, dass bei der Auswahl der Sammlungen

- keine geographischen Restriktionen angewendet werden;
- alle Objekttypen Berücksichtigung finden;
- keine Einschränkung des institutionellen Hintergrundes stattfindet;
- keine disziplinäre Ausrichtung vorgegeben und
- keine inhaltliche Qualitätskontrolle der Sammlungen erfolgen wird.

Unter geographischen Restriktionen wurden dabei sowohl der Ort des Hostings eines Angebotes wie auch seine regionale Zuordnung verstanden. Beide sollen hier nicht als Ausschlusskriterium dienen, da allein hieraus keine Aussage über die Relevanz für andere Forschungsfragen getroffen werden kann. Aus dem gleichen Grund wurde für nicht sinnvoll erachtet, sich ausschließlich auf textzentrierte Sammlungen zu beschränken, auch wenn diese bislang die Mehrzahl der Online-Angebote ausmachen. Eine Einschränkung des institutionellen Hintergrundes erschien ebenfalls als nicht zielführend, da Forschungsprojekte häufig verschiedene Institutionen umfassen und auch Projekte von Einzelforschern Berücksichtigung finden sollten. Zentral ist vielmehr, dass die Sammlung als Ergebnis einer konkreten Forschungstätigkeit entstanden ist und somit wissenschaftlichen Grundansprüchen genügt.

Die Festlegung, alle Disziplinen zu berücksichtigen ergibt sich allein schon aus der stetig wachsenden Zahl interdisziplinärer Forschungsvorhaben, zumal die Interdisziplinarität eine der grundlegenden Eigenschaften der Digital Humanities darstellt. Eine

– willkürlich zu treffende – Einschränkung würde die Nutzbarkeit der Collection Registry somit unmittelbar reduzieren.

Beispielhaft wurde die Datenbank GEPRIS der DFG³ ausgewertet und die dort vorhandenen Einträge in vier Kategorien eingeteilt (Übernahme; Übernahme wünschenswert nach Prüfung; Nach Prüfung Gründe gegen Aufnahme gefunden; Eindeutige Gründe gegen Aufnahme). Hierbei wurden mehrere Punkte berücksichtigt:

- Vorhandensein einer offenen Schnittstelle;
- Vorhandensein einer freien bzw. offenen Lizenz;
- geringe Sichtbarkeit des Projektes;
- wissenschaftliches Forschungsprojekt.

Da es sich bei den in GEPRIS verzeichneten Projekten um DFG-geförderte Forschungsprojekte handelt, die somit einer institutionellen Prüfung unterliegen, wurde der letzte Punkt, der eine zumindest grundlegende Qualitätsprüfung darstellt, grundsätzlich bejaht. Die Ergebnisse der Prüfung sind zu finden unter https://wiki.de.dariah.eu/download/attachments/38087341/Tabelle_%20Collection%20Registry_Gepris.doc?version=1&modificationDate=1432800204301&api=v2.

Einer weiteren Prüfung unterzogen wurden außerdem diejenigen Sammlungen, die bereits in der I. Förderphase (s. unter <https://wiki.de.dariah.eu/display/DARIAHDE/Daten-Sammlungen+und+-Portale>) identifiziert wurden.

Ebenfalls ausgewertet wurden Sammlungen, die in den Repositorien LAUDATIO⁴ sowie »Wissenschaftliche Sammlungen digital«⁵ verzeichnet sind. Eine ausführliche Aufstellung der geprüften Sammlungen und des Ergebnisses findet sich unter <https://wiki.de.dariah.eu/display/DARIAH2/Inhaltlicher+Ausbau+der+CR>.

³ <http://gepris.dfg.de>.

⁴ Long-term Access and Usage of Deeply Annotated Information; <http://www.laudatio-repository.org/repository/documentation>.

⁵ <http://wissenschaftliche-sammlungen.de/de/portal/>.

3. Weitere Übernommene bzw. in Übernahme befindliche Sammlungsbeschreibungen

Aufgrund der unter anderem durch die Überarbeitung des Datenmodells notwendig gewordenen Anpassungsarbeiten an der Collection Registry sowie der Neugestaltung der webbasierten Eingabemaske kam es zu Verzögerungen bei der Eintragung von Sammlungen in den Bestand. Insbesondere von Einzelforschern sind daher bislang kaum Beiträge zu verzeichnen. Es konnten jedoch einige institutionelle Partner gewonnen werden, die Daten für die Eintragung zur Verfügung stellten bzw. die Dienste von DARIAH-DE in ihrem eigenen Kontext nachnutzen wollen.

Der BMBF-geförderte Forschungsverbund Marbach-Weimar-Wolfenbüttel (MWW)⁶ nutzt das Datenmodell der Collection Registry zur einheitlichen Beschreibung verschiedener Sammlungen der Teilinstitutionen. Die Erfassung erfolgt dabei zunächst durch einen Projektmitarbeiter des MWW an der Herzog August Bibliothek lokal im XML-Format, das später in die Collection Registry importiert werden kann. Die hier gemachten Erfahrungen konnten dabei direkt in die Weiterentwicklung des Datenmodells der Collection Registry einfließen.

Innerhalb der III. Förderphase von DARIAH-DE soll diese Zusammenarbeit intensiviert werden. Es werden nicht nur die bereits vorbereiteten Sammlungsbeschreibungen in die Collection Registry übernommen werden, sondern in Kooperation eine weitere Verbesserung des Dienstes erfolgen, sodass eine umfangreiche Eintragung von Sammlungsbeschreibungen seitens MWW erfolgen und diese Daten als Rechercheinstrument für den Verbund dienen können.

Als weitere relevante und für die Weiterentwicklung der Dienste relevante Sammlung wurde *Wikidata* identifiziert. Diese umfangreiche Sammlung ist allgemein bekannt und wird häufig als Quelle für Daten verschiedenster Art (z.B. für biographische Daten aber auch für Verweise auf verschiedene Normdatensätze) genutzt – so auch innerhalb von DARIAH-DE, wo R 5.3.1, *Konzept für die Umsetzung von Lösungen für die Use Cases*, ein Beispiel für die Nutzung und eine Bewertung gibt.

⁶ <http://www.mww-forschung.de>.

Zum Testen der Datenübernahme wurden zunächst 100 Beschreibungen an DARIAH-DE übermittelt. Anhand dieser Daten wird zunächst die Funktionsweise der Schnittstellen der Collection Registry zum Import von Daten geprüft.⁷ Die Eintragung weiterer Daten erfolgt sukzessive ab März 2016.

Der dritte große Beiträger wird die vom BMBF geförderte *Koordinierungsstelle für wissenschaftliche Universitäts-sammlungen in Deutschland*.⁸ Im Laufe des Frühjahres 2016 werden die hier vorhandenen Beschreibungen in den Datenbestand der Collection Registry aufgenommen werden.

Diese und weitere potentielle Quellen von Sammlungen ergeben sich aus Rückmeldungen des von Cluster 4 koordinierten Stakeholdergremiums Wissenschaftliche Sammlungen, dessen erfolgreiche Arbeit in der III. Förderphase von DARIAH-DE fortgeführt werden wird.

Eine Herausforderung stellt die Nutzung der Collection Registry durch Einzelforscher dar. Um hier eine hohe Zahl an Eintragungen zu erhalten, ist es nötig, die Einstiegs-hürden möglichst gering zu halten. Eine Zulieferung von Daten wird in der Regel nicht über eine Import-Schnittstelle erfolgen, sondern über das angebotene Web-Interface.

Häufig wird hierbei jedoch die Erfassung in Kooperation des Forschers mit einer Einrichtung erfolgen. Als Beispiel hierfür wird die Herzog August Bibliothek im Frühjahr 2016 die Datensammlung eines Forschers übernehmen und ihre Beschreibung eintragen können. Ob Einzelforscher ihre Sammlungen unabhängig von einem institutionellen Kontext eintragen werden, ist unklar. Es steht jedoch zu vermuten, dass dies am ehesten bei jüngeren WissenschaftlerInnen der Falls sein wird.

Deshalb und auch aufgrund mehrerer Gespräche ergibt sich, dass die Nutzung der von DARIAH-DE bereitgestellten Dienste (nicht nur der Collection Registry) neben einer aktiven Bewerbung vor allem auf die potentiellen Nutzer zugeschnittenen Schulungsangebote erfordert. Diese werden im Rahmen der III. Förderphase von DARIAH-DE überarbeitet und teils neu erstellt werden, sodass sich mit der Verfügbarkeit

⁷ Vgl. hierzu den (zum Verfassungszeitraum noch ausstehenden) M. 4.2.2, *Weiterentwicklung der Collection-Registry-Schnittstellen*.

⁸ <http://wissenschaftliche-sammlungen.de/>

einer neuen Oberfläche der Collection Registry eine deutliche Erhöhung der verzeichneten Sammlungen erwarten lässt.