



Dokumentation theorie- und verfahrensgeleiteter Sammlungskonzepte (R 4.2.3)

Version 2015-12-15
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:

Andorfer, Peter

Kampkaspar, Dario

Wolf, David

Inhaltsverzeichnis:

1. Einleitung	5
1.1. Forschungsstand	6
2. Sammlungskonzepte – Eine Bestandsaufnahme	11
2.1. Konkret verwendete Suchportale und Kataloge	14
2.1.1. Sammlungen HAB-StipendiatInnen	14
3. Eine Ontologie von Sammlungskonzepten	18
3.1. Die Sammlungsklassen	19
3.1.1. Bibliographie	19
3.1.2. Katalog	19
3.1.3. Sammlung	20
3.1.4. Wörterbuch, Lexikon/Enzyklopädie und Normdatensammlung	20
3.1.5. Portal	21
4. Eine Ontologie von Sammlungskonzepten als OWL-Modell	22
4.1. Klassen und Eigenschaften	22
4.1.1. Die Klassen der Sammlungskonzepte	23
4.1.2. Die Klassen der Ressourcen	25
4.1.3. Die Eigenschaften	27
5. (Weitere) Eigenschaften von Sammlungskonzepten	29
5.1. Vollständigkeit	29
5.1.1. Vollständige Sammlung	30
5.1.2. Unvollständige Sammlung und Auswahlammlung	30
5.2. Wissenschaftlichkeit	31
5.3. Prozessierbarkeit/Interoperabilität	32
5.4. Schnittstellen	33
5.5. Auffindbarkeit	34
6. Implementierung des Gesagten in das DCLAP-Schema	35
7. Ausblick	37
8. Literatur	38
9. Benutzte Webseiten	40
10. Anhang	41

1. Einleitung

*Was sammeln ist, weiß jeder. Und jeder kann es. Was also gibt es darüber noch zu sagen?*¹

Die Semantik des Begriffes ‚Sammlung‘ ist groß. Eine Suche im online Synonymlexikon „Woxikon“² liefert etwa 26 Treffer. Diese Bedeutungsbandbreite reicht demnach von ‚Konzentration‘, ‚Aufmerksamkeit‘ und ‚Gottesdienst‘ über ‚Depot‘, ‚Museum, und ‚Garnitur‘ bis hin zu ‚Menge, ‚Anhäufung‘ und ‚Zusammenstellung‘. Der Brockhaus beschreibt Sammlung als „Zusammenbringen von Geld oder anderen Dingen“ sowie als „zusammengetragenen Besitz (z. B. Kunst-, Münz-S.)“, geht im restlichen Teil des Artikels aber nur noch auf das „S[ammlungs]-Recht“ in Deutschland ein. Außerdem finden sich im Brockhaus noch ein Artikel zu „Sammler“ mit Verweisen auf die Artikel „Kanalisation“ (Abwassertechnik), „Flotationsmittel“ (Aufbereitung), „Akkumulator“ (Elektrotechnik), „Dränung“ (landwirtschaftlicher Wasserbau) und „Jäger und Sammler“ (Völkerkunde). Dazu gibt es noch einen eigenen Artikel zur „Sammlung Deutscher Drucke“.³ Der Duden schließlich kennt insgesamt vier Bedeutungsfelder und fünf Synonymgruppen von Sammlungen.⁴

Für die vorliegende Arbeit soll unter Sammlung – dem Duden folgend – eine „Gesamtheit gesammelter Gegenstände“ verstanden werden, wobei anstelle von ‚Gegenstand‘ die Bezeichnung ‚Ressource‘ bevorzugt wird. Die sich aufdrängende Frage, was wiederum unter Ressource zu verstehen sei,⁵ soll hier nur soweit verfolgt werden, dass unter Ressource prinzipiell alles verstanden werden kann, was in irgendeiner Form eindeutig identifizierbar ist.

Dass der Begriff der Sammlung trotz einer solchen ersten semantischen Einengung aber immer noch ein breites Spektrum verschiedener Sammlungskonzepte umfasst, ist das Thema dieser Arbeit. Ziel ist, die verschiedenen, für geistes- und kulturwissenschaftliches Forschen relevanten Sammlungskonzepte zu *identifizieren*, zu *benennen* und zu *systematisieren*.⁶

Als Ausgangspunkt für die Identifizierung solch relevanter Sammlungskonzepte dient der im Anschluss befindliche Überblick zum aktuellen Forschungsstand. Dieser eher theoretische Zugang wird dazu noch ergänzt und konkretisiert durch die in Kapitel II aufgelisteten Bibliographien, Bibliotheks-

¹ Manfred Sommer, *Sammeln. Ein philosophischer Versuch*, Frankfurt am Main 1999, S. 7.

² <http://synonyme.woxikon.de/>.

³ Vgl. Brockhaus s.v. „Sammlung“, „Sammler“ und „Sammlung Deutscher Drucke“.

⁴ Vgl. dazu: <http://www.duden.de/rechtschreibung/Sammlung>.

⁵ Vgl. dazu etwa Michael Buckland, *What is a "digital document"*, in: *Journal of the American Society for Information Science* 48/9 1997, online und leicht gekürzt:

<http://people.ischool.berkeley.edu/~buckland/digdoc.html>; siehe auch Roger T. Pédaque, *Document: Form, Sign and Medium, As Reformulated for Electronic Documents*, 2003 <sic_00000594> http://archivesic.ccsd.cnrs.fr/sic_00000594.

⁶ Vgl. zum Begriff Konzept u.a. Stock, *Handbook of Information Science*, 2013, S. 531-546.

kataloge, Datenbanken, Suchmaschinen, Webportale, etc., die in geistes- und kulturwissenschaftlichen Forschungsprojekten verwendet werden. Kapitel III entwickelt eine Terminologie zur Benennung der verschiedenen, in Kapitel II gesammelten, Sammlungskonzepte und beschreibt die Systematik der Beziehungen der einzelnen Konzepte zueinander, basierend auf der Sammlungseigenschaft ‚Funktionalität‘. Die Umsetzung einer solchen Ontologie von Sammlungskonzepten in ein maschinenlesbares und prozessierbares Datenmodell, erfolgt in Kapitel IV. Kapitel V evaluiert anschließend noch weitere Kriterien und Eigenschaften, die für KuratorInnen v.a. aber NutzerInnen von Sammlungen relevant sein könnten.

In Kapitel VI wird schließlich untersucht, inwieweit DCLDAP (DARIAH-DE Collection Level Description Application Profile),⁷ das derzeitige in DARIAH-DE verwendete XML-Schema zur Sammlungsbeschreibung, die semantischen Relationen der in Kapitel IV beschriebenen Klassen, Typen und Eigenschaften bereits abbildet, bzw. welche Adaptionen dafür ggf. notwendig wären.⁸

1.1. Forschungsstand

Die Frage, was eine Sammlung ist, wird in der für diesen Artikel eingesehen Forschungsliteratur⁹ gleich mehrmals aufgeworfen, gelegentlich auch gleich im Titel. „What is a collection?“ fragten sich etwa James Curall, Micheal Moss und Susan Stuart im Jahr 2004. Damit trägt ihr in der 58. Nummer der Zeitschrift „Archivaria“ erschienener Aufsatz¹⁰ exakt die gleiche Überschrift, wie ein bereits vier Jahre früher im Journal of the American Society for Information Science erschienenes Paper von Hurl-Li Lee.¹¹ Nur am Rande erwähnt sei hier, dass Curall et al zwar John Locke, Immanuel Kant und Jacques Derrida zitieren,¹² die Arbeit Lees aber unerwähnt bleibt.

Ein weiteres, das Thema Sammlungskonzepte tangierendes Problem stellt jenes der Definition von Sammlung dar. Einschlägige Arbeiten dazu stammen von Carol L. Palmer (u.a.): „Collection Definition in Federated Digital Resource Development“¹³ oder von Carl Logoze und David Fielding: „Defining

⁷ DARIAH-DE Collection Level Description Application Profile (<https://dev2.dariah.eu/wiki/x/VoFn>)

⁸ Zum Zeitpunkt der Publikation dieses Berichtes befindet sich das beschriebene Datenmodell bereits in einer Überarbeitung. Die Ergebnisse dieser Überarbeitung, die in Abstimmung mit externen Partnern erfolgt und auf den hier vorgetragenen Überlegungen basieren, werden in einem eigenen Working Paper dokumentiert werden.

⁹ Siehe dazu Kapitel VI Literatur.

¹⁰ Curall, J., Moss, M. and Stuart, S. (2004). What is a collection?, in: Archivaria, 58, 131-146.

¹¹ Lee, H. (2000). What is a collection?, in Journal of the American Society for Information Science, 51/12, 1106-1113.

¹² Curall et al, What is a collection, S. 133f.

¹³ Carol L. Palmer u.a., Collection Definition in Federated Digital Resource Development, in: 69th Annual Meeting of the American Society for Information Science and Technology (ASIST), Austin (US), 3-8 November 2006. [Conference paper], (<http://eprints.rclis.org/8638/>).

Collections in Distributed Digital Libraries“, veröffentlicht 1998 in der Novemberausgabe des D-Lib Magazin.¹⁴

Abgesehen von dem behandelten Thema teilen sich VerfasserInnen der genannten Artikel auch einen gemeinsamen fachlichen oder beruflichen Hintergrund. Dieser ist im Bereich der Informationswissenschaften und/oder von Gedächtnisorganisationen wie Museen, Archive, vor allem aber Bibliotheken zu verorten, selbst wenn in einzelnen Arbeiten durchaus die Perspektive der SammlungsbenutzerInnen in den Vordergrund gerückt wird.¹⁵ Dies trifft – wenig überraschend – insbesondere auf die AutorInnen jener Arbeiten zu, die sich mit der Konzeption, Realisierung und Evaluation von Sammlungsbeschreibungen bzw. von Schemata und Modellen zur Sammlungsbeschreibung auseinandersetzen. Als Beispiele für zwei Arbeiten jüngeren Datums, die sich vornehmlich der Sammlungsmodellierung und -beschreibung widmen, können Stephan J. Millers 2014 bereits in dritter Auflage erschienenes Handbuch „Metadata for Digital Collections“¹⁶ und der 2013 veröffentlichte Aufsatz „Modeling Cultural Collections for Digital Aggregations and Exchange Environments“,¹⁷ verfasst u.a. von Karen M. Wickett und Carole L. Palmer, ins Feld geführt werden. Diese beiden Titel dienen auch als Ausgangspunkt für die eigene Literaturrecherche, womit zumindest teilweise die bereits erwähnte informationswissenschaftliche Dominanz erklärbar wird.

Liest man diese und weitere Arbeiten mit Blick auf das Thema Sammlungskonzepte durch, so können mehrere Aspekte vermerkt werden, wobei die für die Existenz dieses Aufsatzes wohl wichtigste Beobachtung gleich vorangestellt sei: Eine möglichst umfassende Benennung, Beschreibung und Systematisierung von Sammlungskonzepten spielte in der eingesehenen Literatur wenn überhaupt nur eine untergeordnete Rolle. Dies ist insofern verwunderlich, da in vielen Texten nicht nur die Frage gestellt wird, was denn überhaupt eine Sammlung sei, sondern häufig auch eine ganze Reihe von Beispielen mehr oder weniger konkreter Sammlungen und Sammlungskonzepte zum Zwecke der Beantwortung besagter Frage genannt werden. Damit einher geht dann meist die Feststellung, dass eine allgemein verbindliche Definition von Sammlung nicht gegeben oder nur schwer zu erreichen ist. Um letztere Behauptung zu belegen, hier ein kleiner Auszug entsprechender Zitate:

*The word “collection” has been common currency in what we accept as the real world of objects and events, and has been imported with seemingly little effort into our discourse about the digital world, **yet there is no clear definition in either domain of what is meant by the***

¹⁴ Carl Lagoze/David Fielding, Defining collections in distributed digital libraries, in D-Lib magazine 54/11. (<http://www.dlib.org/dlib/november98/lagoze/11lagoze.html>)

¹⁵ Lee, H. (2005). The concept of collection from the users's perspective. The Library Quarterly, 75/1, 67-85.

¹⁶ Stephan J. Miller, Metadata for Digital Collections, London 2014.

¹⁷ Karen M. Wickett u.a., Modeling cultural collections for digital aggregation and exchange environments, o. O. 2013. (<http://hdl.handle.net/2142/45860>).

term.; The collection, which must continue to be described, **is the bunch of objects bound together by a defined activity or institutional framework.**¹⁸

*It probably is impossible to arrive at a definition of the concept of collection in the digital library environment that everyone will agree to as an operational definition.*¹⁹

*Our findings reveal high variability and ambiguity in the collection construct; ... there is little agreement on what constitutes a collection ...; We have a problem with that word collection. We fought about that word, so when you use it what do you mean?" ...; an overloading of the meaning of the word "collection".; In undertaking this project, we have been surprised at the variety of ways that the term "collection" has been used, and indeed that the term is noted as fuzzy or problematic by some of our respondents.*²⁰

*... a consistent definition of collection has not emerged.*²¹

*The three international organizations ICA, IFLA and ICOM have never engaged in any common definition of what a collection is.*²²

Diese durchwegs ernüchternden Aussagen sind meist das Ergebnis von Reflexionen zu den Eigenschaften, Kriterien und Konzepten von Sammlungen, wie sie gerne in der Einleitung der einzelnen Arbeiten anzutreffend sind und wofür, wie erwähnt, gerne verschiedene Sammlungen oder Sammlungskonzepte aufgezählt werden.

Eine Zusammenschau dieser Beispiele erweist sich als äußerst bunt, was wenig verwundert, wenn man bedenkt, dass – je nach Intention der VerfasserInnen – ja gerade die Vielfältigkeit oder Unschärfe des Sammlungsbegriffes veranschaulicht werden soll. Die Bandbreite reicht von Internetsuchmaschinen Yahoo! („Consider Yahoo! as an example. [...] It certainly fits the dictionary definition of collection – a group of objects.“²³) und Alta Vista,²⁴ wodurch heutigen LeserInnen drastisch die Historizität dieser Texte vor Augen geführt wird, über Lesezeichen („favorites“) im Browser²⁵ und generell Webpages,²⁶ bis hin zu konventionelleren Konzepten wie Datenbanken²⁷ und Katalogen²⁸ oder gene-

¹⁸ Curall, What is a collection, S. 131 und S. 145.

¹⁹ Hill, Collection metadata solutions, S. 1171.

²⁰ Palmer, Collection Definition.

²¹ Wickett, Modeling cultural collections, S. 7.

²² Wickett, Modeling cultural collections, S. 7.

²³ Lee, What is a collection, S. 1109.

²⁴ Heaney, M. (2000). An analytic model of collections and their catalogues, in UK Office for Library and Information Science.

²⁵ Vgl. Lee, The concept of collection, S. 80.

²⁶ Lee, What is a collection, S. 110: „When a World Wide Web page has hyperlinks to other WWW pages, is it a collection or a directory“

²⁷ Beagrie, Plenty of Room; Lee, The concept of collection, S. 69.

²⁸ Heaney, Analytical Model, S. 4f.

rell zu zu Sammlungen synonym verwendeter Bezeichnungen wie Archiv, Museum, Ausstellung und Bibliothek.²⁹

Eine nähere Beschäftigung mit den genannten Sammlungen und den damit stellvertretend genannten Sammlungskonzepten, etwa eine Analyse von Gemeinsamkeiten und Unterschieden fehlt jedoch in den allermeisten Fällen, was häufig damit begründet wird, sich vor allem auf bibliothekarische Sammlungen konzentrieren zu wollen.³⁰ Definitionen von Sammlungen, sofern solche überhaupt vorgenommen werden, variieren in den einzelnen Texten und reichen von eher minimalistischen Aussagen wie: „collections themselves are the entities that meet the information needs of researchers“³¹ oder „[a] collection is logically defined as a set of criteria for selecting resources from the broader information space“,³² bis hin zu komplexeren Angaben wie:

*A collection is a group of objects gathered together for some intellectual, artistic, or curatorial purpose. In addition, we limit our attention to collections that satisfy the following constraints: (1.) The collection has members that have been gathered together in the past or will be gathered together in the future. (2.) Membership in a collection is determined by some criteria that fit the purpose or intentions of the collector. (3.) The collection may be treated as an individual object for purpose of description, access, and curation.*³³

Wenn tatsächlich Sammlungskonzepte an sich, etwa Möglichkeiten und Kriterien zur Systematisierung und Klassifizierung konkreter Sammlungen diskutiert werden, dann handelt es sich dabei vornehmlich um Sammlungskonzepte, die vor dem digitalen Zeitalter entwickelt und verwendet wurden. Hill nennt etwa „seven criteria [...] to conceptualize collections“, darunter unter anderem „Catalog representation“, „Proprietary of ownership“ oder „Material permanency“,³⁴ in einem früheren Paper referiert sie außerdem noch „four major presumptions associated with collections“ nämlich „tangibility [Greifbarkeit, Verständlichkeit], ownership, a user community, and an integrated retrieval mechanism“,³⁵ um anschließend darzulegen, dass angesichts zunehmender Digitalisierungsbestrebungen diese bisher relevanten Kriterien, allen voran natürlich „tangibility“, an Bedeutung eingebüßt haben.

Ein anderes Beispiel zur Systematisierung von Sammlungskonzepten, wird in Curall et al referiert.³⁶ Konkret handelt es sich dabei um eine von dem US-amerikanischen Archivwissenschaftler Theodore R. Schellenberg vorgeschlagene Kategorisierung von Sammlung in drei Typen, nämlich in eine

²⁹ Palmer, Collection Definition; Wickett, Modeling cultural collections, S. 7.

³⁰ Vgl. beispielsweise Lee, What is a Collection, S. 1106: „In other words, “collection” should be thought of as in the term “a library collection”. (S. 1006).

³¹ Wickett, Modeling, S. 7.

³² Lagoze, Defining Collections, S. 6.

³³ Wickett, Modeling, S. 13.

³⁴ Lee, The concept of collection, S. 75.

³⁵ Lee, What is a collection, S. 1106.

³⁶ Vgl. Curall, What is a Collection?, S. 138.

„organische“ und eine „künstliche“ Sammlung sowie in eine „collection of miscellany“ – also eine Mischsammlung. Wie Schellenberg selbst festhält, konstituiert sich seine Typologie von Sammlungskonzepten aufgrund der Art und Weise, wie sie angelegt wurden: „Collections of private papers fall into three types that may be distinguished from one another *by the way they came into being*“.³⁷

Eine ressourcenbasierte duale Typologie von Sammlungskonzepten beschreibt hingegen Heaney, indem er zwischen „collections of entities [...] or derived representations of entities“, wie sie etwa Bücher oder „photographs of pieces of sculpture“ darstellen und „collections of information about such entities“ unterscheidet.³⁸ Ein ähnliches Konzept verfolgen auch Wickett et al, die eine „Holdings Collection“ und eine „Referential Collection“ beschreiben. Unter „Holdings Collection“ verstehen sie „a collection of items in the custody or control of an organization or curator“ und unter „Referential Collection“ eine „collection referring to rather than directly holding its items.“³⁹ Mit dieser Unterscheidung geht auch eine Differenzierung zwischen institutionell und privaten Sammlungen einher, wobei Wickett et al eine solche Zweiteilung aber selbst hinterfragen: „It is worth considering whether to make a parallel distinction between institutionally developed collections and collections developed by private individuals, i.e. whether ‘amateur’ collections follow the same principles as ‘professional’ ones.“⁴⁰ Auch wenn die AutorInnen schließlich zu dem Schluss kommen, dass die „differences between physical holdings collections and referential collections“ nicht signifikant genug sind um diesen in ihre Datenmodelle von Sammlungen (IMLS und europeana) einfließen zu lassen, so kann daran für die in dieser Arbeit angestrebte Klassifizierung von Sammlungskonzepten durchaus angeknüpft werden.

³⁷ Theodore R. Schellenberg, T.R. Schellenberg, *The Management of Archives*, New York 1965, Reprint Washington DC 1988, S. 172 (<http://hdl.handle.net/2027/mdp.39015015526695?urlappend=%3Bseq=213>). Kursivsetzung nicht aus dem Original übernommen.

³⁸ Heaney, *Analytical Model of Collections*, S. 4.

³⁹ Wickett, *Modeling*, S. 11.

⁴⁰ Ebd. S. 12.

2. Sammlungskonzepte – Eine Bestandsaufnahme

Die Deutsche Forschungsgemeinschaft DFG⁴¹ unterstützt zahlreiche Forschungs- und Digitalisierungsprojekte. Solche Projekte sind in der Onlinedatenbank „GEPRIS – Geförderte Projekte der DFG“⁴² verzeichnet, die Einträge über alle von der Deutschen Forschungsgemeinschaft geförderten Projekte seit Beginn des Jahrtausends enthält. Im November 2014 waren dort 16514 Projekte mit geisteswissenschaftlicher Ausrichtung verzeichnet.

Die von der DFG verwendete Fachsystematik erleichtert das Auffinden bestimmter Projekte nur bedingt, geförderte Projekte aus dem Bereich Geisteswissenschaften sind relativ leicht auffindbar, hierunter fallen aber neben der Erstellung von Sammlungen auch reine Digitalisierungsprojekte ohne Anreicherung mit Metadaten oder Plänen zur Veröffentlichung, der Ankauf von Nachlässen und der Aufbau von Infrastrukturen ohne präzise definierte Inhalte.

Trennt man diese bei der DFG unter den Punkt „Infrastrukturförderung“ gelisteten Projekte nach Fachrichtung zeigt sich dessen ungeachtet ein deutliches Bild:

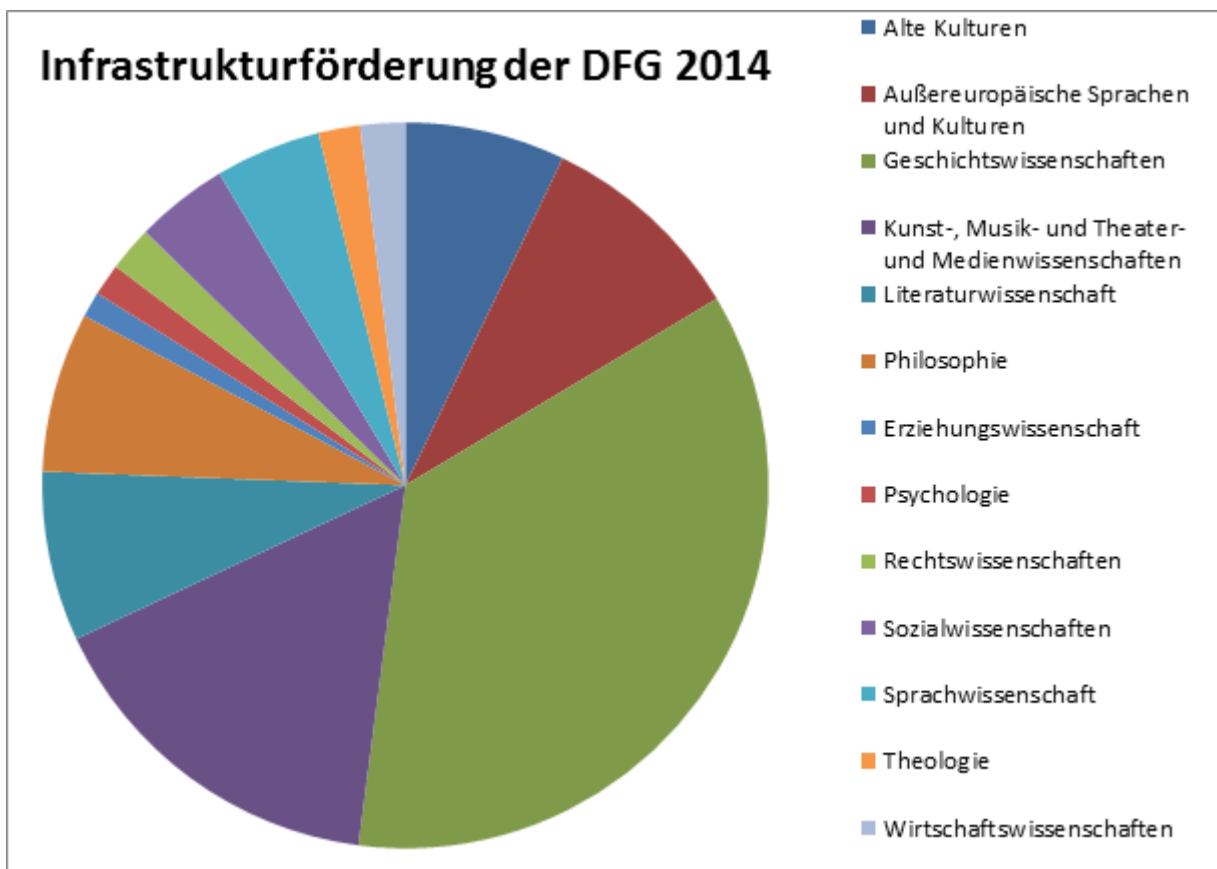


Abbildung 1: Verteilung der geförderten DFG-Projekte nach Fachrichtungen.

⁴¹ <http://www.dfg.de>

⁴² <http://gepris.dfg.de/gepris/OCTOPUS>

Die Geschichts- und Altertumswissenschaften sowie die Sprach- und Literaturwissenschaften machen zusammen weit mehr als die Hälfte der DFG-Förderungen im Bereich Infrastrukturen aus. Der vermutliche Hauptgrund ist der Vorsprung, der durch die vergleichsweise leicht umsetzbare Digitalisierung, Anreicherung und Bereitstellung von Text und Bild in Online-Datenbanken gewonnen wurde.

In der GEPRIIS-Datenbank verzeichnete Projekte, deren Ergebnisse Sammlungen sind, die als (meistens freie) Ressource online zugänglich gemacht wurden, finden ihren Weg in die DARIAH-DE *Collection Registry*, in der im November 2014 ca. 100 Einträge verzeichnet waren.

Eine genauere Untersuchung der Einträge in der Collection Registry verschärft das oben angedeutete Bild: als Portal angebotene, mit DFG-Fördermitteln entstandene Sammlungen sind zu 90% den Fachbereichen Geschichte und Sprache/Literatur zuzuordnen und fast ausschließlich Sammlungen von Text- oder Bildmaterial. Die restlichen Angebote sind zu etwa gleichen Teilen aus Philosophie oder Musik oder fachlich nicht klar zuzuordnende Nachlässe und Privatsammlungen oder Wörterbücher und Thesauri. Der Zweck von GEPRIIS ist reiner Nachweis, der Auflistung liegt kein klar definierter Begriff zugrunde. Zweck der Collection Registry ist ebenfalls Nachweis, ein klarer Sammlungs-begriff ist aber erwünscht.

Um sich dem Begriff der „Sammlung“ besser nähern zu können, wird im Zuge einer aktuell laufenden Befragung der StipendiatInnen, GastwissenschaftlerInnen und wiss. MitarbeiterInnen der Herzog August Bibliothek zum Thema „Forschen und Forschungsdaten in den Geisteswissenschaften“ unter anderem folgende Frage gestellt:

„Welche Suchportale und Kataloge verwenden Sie um (digitale) Ressourcen für Ihr Forschungsprojekt zu suchen? Hier einige Beispiele: *google, KVK, Verbundkataloge (GBV), OPAC, europeana, einschlägige Datenbanken (virtuelles-kupferstichkabinett.de, arachne, ANNO – AustriaN Newspapers Online)*“.⁴³

Eine explizite Frage nach den verwendeten wissenschaftlichen Sammlungen wurde bewusst vermieden, da, wie in den Vorbemerkungen dargelegt, der Begriff der wissenschaftlichen Sammlungen, ebenso wie der Begriff Forschungsdaten, nicht ausreichend etabliert ist.⁴⁴ Die bisherigen Antworten sind in folgender Wortwolke zusammengefasst.

⁴³ Der Fragebogen in seiner aktuellsten Fassung kann unter diesem Link eingesehen werden: digital-archiv.at:8081/exist/rest/db/spielen/xforms/umfrage/Interview_2.xml. Zur zitierten Frage siehe Punkt III. 1.

⁴⁴ Vgl. dazu Carol L. Palmer u.a., *Collection Definition in Federated Digital Resource Development*, Austin 2006, (<http://eprints.rclis.org/8638/>), S. 4: „We have a problem with that word collection. We fought about that word, so when you use it what do you mean.“

mes – und darüber hinaus – erfasst sind und wodurch die wesentliche Sekundärliteratur sämtlicher Fachdisziplinen zumindest auf der Eben der selbstständigen Publikationen abgedeckt ist, bietet sich als erster Einstiegspunkt zur Literaturrecherche für VertreterInnen sämtlicher Disziplinen an. Das „Gesamtverzeichnis der Wiegendrucke“⁴⁷ hingegen bedient ein ungleich schmaleres Segment geistes- und kulturwissenschaftlicher Bedürfnisse. Entsprechend unterschiedlich häufig werden die beiden Kataloge/Portale auch bei den Interviews genannt.

2.1. Konkret verwendete Suchportale und Kataloge

Bis dato wurden insgesamt rund 65 unterschiedliche Kataloge und Portale genannt. Eine jeweils aktuelle und vollständige Übersicht über die bei der Befragung der HAB-StipendiatInnen und GastwissenschaftlerInnen genannten Kataloge und Suchportale, kann unter folgendem Link abgerufen werden:

<http://digital-archiv.at:8081/exist/apps/DARIAH-Collection/xquery/distinctCollections.xq>

Dabei gilt es jedoch zu beachten, dass hier die Antworten in einer bereits normalisierten Form angezeigt werden. Betroffen davon sind die Groß- und Kleinschreibung sowie Leerzeichen zwischen einzelnen Wörtern. Außerdem gilt es zu bedenken, dass die Interviewten gebeten wurden, die von ihnen verwendeten Portale und Kataloge zu *nennen* und nicht die Möglichkeit hatten, aus einer bereits existierenden Liste auszuwählen. Dies erklärt vielleicht auch die eine oder andere idiosynkratische Formulierung, wie auch die Verwendung unterschiedlicher Namen für ein und dasselbe Portal. Eine überarbeitete Version dieser Liste zeigt untenstehende Tabelle:⁴⁸

2.1.1. Sammlungen HAB-StipendiatInnen

Antwort	standardisierter Name / link	Typ
adb	http://www.deutsche-biographie.de/index.html	biographisches Lexikon
amazon	http://www.amazon.de/	Katalog käuflich erwerbbarer Produkte
archive.org	https://archive.org/index.php	Sammlung digitalisierter Materialien
bbkl	Biographisch-Bibliographisches Kirchenlexikon http://www.bbkl.de/	biobibliographisches Lexikon
bds1-online.de	Bibliographie der deutschen Sprach- und Literaturwissenschaft http://www.bds1-online.de/	fachspezifische Bibliographie

⁴⁷ <http://www.gesamtkatalogderwiegendrucke.de/>.

⁴⁸ Eine um die Spalte „Klasse“ erweiterte Version dieser Tabelle kann unter diesem Link eingesehen werden: https://docs.google.com/spreadsheets/d/1_EPLFs68nisgGcyrtuVZshk6nyANgkA5M8rMmA5xoeA/edit?usp=sharing.

Antwort	standardisierter Name / link	Typ
beolingus	BEOLINGUS - Ihr Online-Wörterbuch http://dict.tu-chemnitz.de/	Fremdsprachenwörterbuch
bibliothekskatalogeunifreiburg	https://katalog.ub.uni-freiburg.de/opac/	Bibliothekskatalog
bnf	Bibliothèque nationale de France http://www.bnf.fr	Nationalbibliothek
britishlibrary	http://www.bl.uk/	Nationalbibliothek
britishmuseumcollectionsearch	http://www.britishmuseum.org/research/collection_online/search.aspx	Museumskatalog
ceecöln	Codices Electronici Ecclesiae Coloniensis http://www.ceec.uni-koeln.de/	Sammlung digitalisierter Handschriften
census	Census of Antique Works of Art and Architecture Known in the Renaissance http://www.census.de/	Normdatensammlung
digitalerportraitindex	http://www.portraitindex.de/	Sammlung digitalisierter Portraits
dnb	Deutsche Nationalbibliothek http://www.dnb.de/	Nationalbibliothek
e-codices	Virtuelle Handschriftenbibliothek der Schweiz http://www.e-codices.unifr.ch/	Sammlung digitalisierter Handschriften
ebSCO	http://www.ebsco.com	
europaena	www.europeana.eu	Sammlung digitalisierter Materialien
fotomarburg	http://www.fotomarburg.de/ [siehe auch bildindex.de]	Sammlung digitalisierter (?) Bilder
gallica	http://gallica.bnf.fr/	Sammlung digitalisierter Materialien
gbv	https://www.gbv.de/	Verbundkatalog
gesamtverzeichnisderwiegenrucke	http://www.gesamtkatalogderwiegendrucke.de/	Katalog von Wiegedrucken
gnd	http://www.dnb.de/gnd	Normdatensammlung
google	https://www.google.de/	
googlebooks	http://books.google.de/	Volltext(?)datenbank
gvk	http://gso.gbv.de/	Verbundkatalog
handschriftencensus	http://www.handschriftencensus.de/	Handschriftenkatalog
handschriftendatenbankhab	http://dbs.hab.de/mss/	Handschriftenkatalog
heidi	http://www.ub.uni-heidelberg.de/helios/kataloge/heidi.html	Bibliothekskatalog
hszkult	http://www.hsozkult.de/	Portal

Antwort	standardisierter Name / link	Typ
jstor	http://www.jstor.org/	Volltextsammlung
kubikat	www.kubikat.org/	Verbundkatalog
kvk	http://www.ubka.uni-karlsruhe.de/kvk.html	allgemeine bibliographische Datenbank
leichenpredigt-datenbank	http://dbs.hab.de/leichenpredigten/	Katalog von Leichenpredigten
leo	http://dict.leo.org/	Fremdsprachenwörterbuch
libraryofcongress	http://www.loc.gov/	Bibliothekskatalog
manuscriptamediaevalia	http://www.manuscripta-mediaevalia.de/	Handschriftenkatalog
mittelhochdeutschebegriffsdatenbank	http://mhdbdb.sbg.ac.at/	Volltextsammlung
mla	modern language association http://www.mla.org/	Portal
nkp.cz	National Library of the Czech Republic http://www.nkp.cz/	Nationalbibliothek
opachab	http://opac.lbs-braunschweig.gbv.de/	Bibliothekskatalog
oxfordmusiconline	http://www.oxfordmusiconline.com/	Volltextsammlung
rilm	Répertoire International de Littérature Musicale http://www.rilm.org/	fachspezifische Bibliographie
rism	Répertoire International des Sources Musicales https://opac.rism.info/	Katalog musikalischer Quellen
rsa.org	The Renaissance Society of America http://www.rsa.org/	Portal
theaterforschung.de	http://theaterforschung.de/	Portal
treccani	L' Enciclopedia Italiana http://www.treccani.it/	allg. Lexikon
vd16	www.gateway-bayern.de/index_vd16.html	fachspezifische Bibliographie
vd17	http://www.vd17.de/	fachspezifische Bibliographie
vd18	http://vd18.de/	fachspezifische Bibliographie
vkk	http://www.virtuelles-kupferstichkabinett.de/	Sammlung digitalisierter Kupferstiche
warburgiconographicdatabase	http://warburg.sas.ac.uk/photographic-collection/iconographic-database/	Sammlung digitalisierter Bilder
wikipedia	wikipedia	allg. Lexikon
worldbiographicalinformatics	http://db.saur.de/WBIS	biographisches Lexikon

Antwort	standardisierter Name / link	Typ
nssystem	World Biographical Information System	
worldcat	https://www.worldcat.org/	Verbundkatalog
wörterbuchnetz	http://woerterbuchnetz.de/	historisches Wörterbuch
yale.deu/dissertations	http://guides.library.yale.edu/dissertations	Katalog
zeno.org	http://www.zeno.org/	Volltextsammlung

3. Eine Ontologie von Sammlungskonzepten⁴⁹

Ontologie wird hier im Sinne von Wolfgang und Mechtild Stock verstanden als:

[A] specific knowledge organization system, which is available in a machine-readable, formal language and also disposes of mechanisms of automatic reasoning. When a nomenclature, a classification system or a thesaurus is realized in a machine-readable language, we call this KOS a "simple KOS" (or SKOS) – in contrast to a "rich" KOS, i.e. an ontology.⁵⁰

Die Tabelle zu den ‚Sammlungen‘, die von den StipendiatInnen der Herzog August Bibliothek im Kontext ihrer Forschungsprojekte verwendet werden, wurde um eine Spalte mit der Überschrift ‚Typ‘ erweitert. Besagter Spalte zufolge verteilen sich 61 Sammlungen dieser Tabelle auf knapp 30 Typen, wobei diese Typen zu sieben Gruppen oder Klassen zusammengefasst werden können. Die Verteilung der Sammlung über die sieben Klassen ist in nachfolgender Tabelle zusammengefasst:

Klasse	Anzahl der Sammlungen
Normdatensammlung	1
Bibliographie	2
Wörterbuch	3
Portal	4
Lexikon	5
Sammlung	14
Katalog	25

Die Zuordnung der einzelnen Typen zu ihren jeweiligen Klassen zeigt diese Tabelle:

Klasse	Typen
Normdatensammlung	<ul style="list-style-type: none"> ● Normdatensammlung
Bibliographie	<ul style="list-style-type: none"> ● Fachbibliographie ● Nationalbibliographie ● Sachbibliographie
Wörterbuch	<ul style="list-style-type: none"> ● Fremdsprachenwörterbuch ● historisches Wörterbuch
Portal	<ul style="list-style-type: none"> ● Portal
Lexikon/Enzyklopädie	<ul style="list-style-type: none"> ● allg. Lexikon ● biobibliographisches Lexikon ● biographisches Lexikon
Sammlung	<ul style="list-style-type: none"> ● Sammlung digitalisierter Bilder, ● - Kupferstiche, ● - Portraits, ● - Materialien, ● - Handschriften; ● Volltextsammlung
Katalog	<ul style="list-style-type: none"> ● Bibliothekskatalog ● Verbundkatalog

⁴⁹ Auf Basis der Sammlungsobjekte.

⁵⁰ Stock, Handbook of Information Science, 2013, S. 526.

	<ul style="list-style-type: none"> ● Handschriftenkatalog ● Katalog deutscher Drucke, ● - musikalischer Quellen, ● - Leichenpredigten, ● - Wiegendrucke ● Museumskatalog
--	--

In Anbetracht der großen Heterogenität der in geistes- und kulturwissenschaftlichen Forschungsprojekten verwendeten Sammlungen muss obige Tabelle nur als ein erster und früher Entwurf einer Ontologie von Sammlungsbegriffen verstanden werden. Dies gilt sowohl für die Beziehungen zwischen den Typen und Klassen wie auch für die verwendete Terminologie.

Während einige der genannten Begriffe wohl weitgehend selbsterklärend sein dürften (Lexikon/Enzyklopädie, Wörterbuch), bedarf es wohl vor allem hinsichtlich der Differenzierung von Bibliographie, Katalog und Sammlung einer knappen Erläuterung der jeweiligen Klassifizierungs- und Typisierungsmerkmale.

3.1. Die Sammlungsklassen

3.1.1. Bibliographie

Unter Bibliographie wird hier eine Sammlung *ausschließlich* bibliographischer Angaben von wenigstens zwei Objekten (Publikationen) verstanden.⁵¹ Angaben zum Standort des erfassten Objektes, darunter wäre auch ein Link zu dem jeweiligen Objekt zu verstehen, zählen explizit nicht dazu. Eine bibliographische Sammlung kann zu einem Thema bzw. einer *Sache* angelegt werden, sich speziell an ein bestimmtes *Fach* richten oder versuchen, die Gesamtzahl der Objekte eines klar erkennbaren Raumes, einer *Nation* zu erfassen, wobei Kombinationen der verschiedenen Typen möglich sind. Ein Beispiel einer nationalen fachbibliographischen Sammlung wäre etwa die Österreichische Historische Bibliographie (ÖHB).

3.1.2. Katalog

In engem Zusammenhang zur Bibliographie steht der Katalog. Bei einer solchen handelt es sich um eine Sammlung bibliographischer Daten und *Informationen zum Standort der jeweiligen Objekte*. Somit ist jeder Katalog immer auch eine Bibliographie, im Unterschied zur Bibliographie *verweist* die bibliographische Datenbank jedoch auf die Standorte ihrer Objekte.

⁵¹ Siehe einführend zu Thema und Begriff Bibliographie: Eberhard Bartsch, Die Bibliographie. Einführung in Benutzung, Herstellung, Geschichte, München u.a. 1979 und Georg Schneider/Friedrich Nesterl, Handbuch der Bibliographie. Völlig neu bearbeitet von Friedrich Nestler, Stuttgart 1999.

Wie eine Bibliographie kann auch ein Katalog noch näher typisiert werden. So werden von einem *spezifischen Katalog* nur bestimmte Typen (z. B. Dissertationen) von Ressourcen (Publikationen) erfasst. Zu unterscheiden ist außerdem zwischen einem auf den Bestand einer einzigen Institution beschränkten Katalog, der eben nur die Ressourcen dieser einen Institution beschreibt (z. B. der opac einer Bibliothek) und dem *Verbundkatalog*, welcher die Kataloge mehrerer Institutionen zusammenführt.

Bibliographie und Katalog entsprechen dem, was Wickett als „referential collection“ bezeichnet: „Referential Collection: A collection referring to rather than directly holding its items.“⁵²

3.1.3. Sammlung

Wie der Katalog beschreibt auch die Sammlung wenigstens zwei *einmalige (konkrete)* Objekte. Im Unterschied zum Katalog beschreibt bzw. verweist die Sammlung aber nicht nur auf ihre Objekte – jede Sammlung ist somit gleichzeitig auch ein Katalog – sondern ermöglicht einen *direkten Zugriff auf das Objekt*.⁵³ Analog zu den übrigen Klassen kann auch die Klasse der Sammlung typisiert werden, vornehmlich über das Thema- und/oder über die gesammelten Objekttypen.

3.1.4. Wörterbuch, Lexikon/Enzyklopädie und Normdatensammlung

Bei einem Wörterbuch und einem Lexikon/einer Enzyklopädie⁵⁴ kann es sich Sammlungskonzepte bzw. um Typen der Klasse Sammlung handeln, vorausgesetzt man begreift die Einträge und Artikel eines Wörterbuches, einer Enzyklopädie, als Ressourcen, die auch unabhängig von ihrer Zuordnung zu einem Wörterbuch oder einer Enzyklopädie eindeutig unterscheid- und identifizierbar sind. Letztlich wird die Frage, ob es sich bei einem Wörterbuch, einer Enzyklopädie, einem Lexikon nun um eine Sammlung diskreter Ressourcen handelt von Fall zu Fall entschieden werden müssen.

Bei einer Normdatensammlung wie zum Beispiel der Gemeinsamen Normdatei der Deutschen Nationalbibliothek (GND) ist die Distinktion der Ressourcen, der Normdatensätze, alleine schon aus dem Grund gegeben, da ein Normdatensatz für gewöhnlich ja mittels eigener ID eindeutig identifiziert ist. Da die Hauptfunktion einer Normdatensammlung, wie auch eines Wörterbuches, eines Lexikons oder einer Enzyklopädie vor allem jene des Nachschlagens, Verweisens oder Vergewisserns ist, scheint die Bildung einer gemeinsam Klasse gerechtfertigt zu sein.

⁵² Wickett, Cultural Collections, S. 11.

⁵³ Vgl. dazu die Definition von „digital collection“ in Miller, Metadata for Digital Collection 2014, S. 6: „A digital collection,[...] is a collection of digital resources, along with metadata about those resources, made available online through an interface that allows users to search and/or browse the contents of that collection.“

⁵⁴ Im deutschsprachigen Raum werden Lexikon und Enzyklopädie durchaus synonym verwendet (vgl. etwa das Biographisch-Bibliographische Kirchenlexikon oder Johann Heinrich Zedlers Grosses vollständiges Universal Lexicon), während das englische „Lexicon“ eher im Sinne von Wörterbuch verwendet wird.

Sammlung, Wörterbuch, Lexikon/Enzyklopädie und Normdatensammlung entsprechen also dem, was Wickett als „Holdings Collection“ bezeichnet: „Holdings Collection: A collection of items in the custody or control of an organization or curator.“⁵⁵

3.1.5. Portal

Die zuvor genannten Klassen von Sammlungskonzepten, bzw. die jeweiligen Instanzen dieser Klassen, konnten aufgrund ihrer doch weitestgehend eindeutigen Kriterien relativ präzise beschrieben und schlüssig miteinander in Beziehung gesetzt werden. Jene Onlinere Ressourcen hingegen, die in der Tabelle „Sammlungen der HAB-StipendiatInnen“ als Portale typisiert bzw. klassifiziert wurden⁵⁶ erweisen sich hinsichtlich ihrer Funktionalität und/oder in Bezug auf die darin versammelten Objekte als zu vielfältig und heterogen um einer der oben genannten Klassen zugeordnet werden zu können. Das Portal H-Soz-Kult beispielsweise, ließe sich angesichts der darin anzutreffenden Inhalte zwar der Klasse Sammlung zuordnen, handelt es sich bei den darin erschlossenen Inhalten doch beinahe ausschließlich um Sammlungen von Objekten wie Rezensionen, Tagungsberichten, Stellenausschreibungen oder einschlägigen Artikeln. Was jedoch die Funktion dieses Portales betrifft, so dürfte diese eher im Bereich Verbreitung und Distribution als im Bereich von Sammlung und Archivierung von Informationen liegen. Eindeutiger lässt sich der Unterschied zwischen den Klassen Sammlung und Portal am Webaufttritt der „Renaissance Society of America“ (<http://www.rsa.org/>) exemplifizieren. Gegen eine Klassifizierung als Sammlung spricht hier, abgesehen von der deutlich erkennbaren kommunikativen Funktion, vor allem auch die Heterogenität der gesammelten Objekte. Als Portale sollen also all jene Sammlungskonzepte klassifiziert werden, die aufgrund der Heterogenität ihrer gesammelten Objekte, vor allem aber aufgrund ihrer über die reine Sammlungstätigkeit hinausgehende Funktionalität nicht zweifelsfrei einem der oben genannten Sammlungskonzepten zugeordnet werden können.

⁵⁵ Wickett, Cultural Collections, S. 11.

⁵⁶ Konkret handelt es sich dabei um <http://www.mla.org/>, <http://www.hsozkult.de/>, <http://www.rsa.org/> und <http://theaterforschung.de/>

4. Eine Ontologie von Sammlungskonzepten als OWL-Modell

Die in Kapitel III formulierten Überlegungen zu einer Ontologie von Sammlungskonzepten werden in diesem Kapitel in ein maschinenlesbares Modell überführt. Der sich daraus ergebende Mehrwert besteht erstens in der genannten maschinellen Prozessierbarkeit des Modells, zweitens in der darin verwendeten eindeutigen Terminologie und drittens in den darin beschriebenen ebenfalls eindeutigen semantischen Beziehungen zwischen den modellierten Klassen bzw. deren Instanzen. Die Beschreibung der für eine Ontologie von Sammlungskonzepten notwendigen Klassen und Eigenschaften sowie die der Beziehungen dieser Klassen und Eigenschaften erfolgt mittels OWL. OWL steht für Web Ontology Language, eine vom World Wide Web Consortium (W3C) spezifizierte Beschreibungssprache für Ontologien. Ein Standard, der seinerseits wiederum stark auf dem Resource Description Framework (RDF) und dem Resource Description Framework Schema (RDFS) aufbaut.⁵⁷ Realisiert wird das Modell mit Hilfe des Ontologie-Editors Protégé, entwickelt von der Stanford University.⁵⁸

4.1. Klassen und Eigenschaften

Das für die hier entwickelte Ontologie von Sammlungskonzepten zentrale Element ist die Beziehung zwischen Sammlung und Ressource: *Eine Sammlung sammelt bzw. verfügt über oder hat Ressourcen*. Diese Eigenschaft, Charakteristik oder Fertigkeit (Property) einer Sammlung wird im Modell als owl:ObjectProperty beschrieben und **colcon:hasResource** benannt.⁵⁹ Die in RDFS spezifizierten Eigenschaften rdfs:range und rdfs:domain stellen dann die oben genannte Beziehung her. Mithilfe von rdfs:domain wird die Klasse **colcon:CollectionConcept** als Subjekt, mithilfe von rdfs:range die Klasse **colcon:Resource** als Objekt der Eigenschaft (Property) colcon:hasResource definiert.

```
...
<owl:ObjectProperty rdf:about="colcon:hasResource">
  <rdfs:subPropertyOf rdf:resource="owl:topObjectProperty"/>
  <rdfs:domain rdf:resource="colcon:CollectionConcept"/>
  <rdfs:range rdf:resource="colcon:Resource"/>
</owl:ObjectProperty>
...
```

Nachfolgend eine schematische Darstellung zu obigem Codeausschnitt:

⁵⁷ Vgl. einführend zu RDF, RDFS und OWL: Dean Allemang, Jim Hendler, Semantic Web for the Working Ontologist. Effective Modeling in RDFS and OWL, 2. Auflage, Amsterdam u.a. 2011 sowie Shelly Powers, Practical RDF, Peking u.a. 2003.

⁵⁸ <http://protege.stanford.edu/>.

⁵⁹ Der hier verwendete Namespace „colcon:“ (CollectionConcepts) ist ein Provisorium, da noch keine endgültigen URIs vergeben wurden.

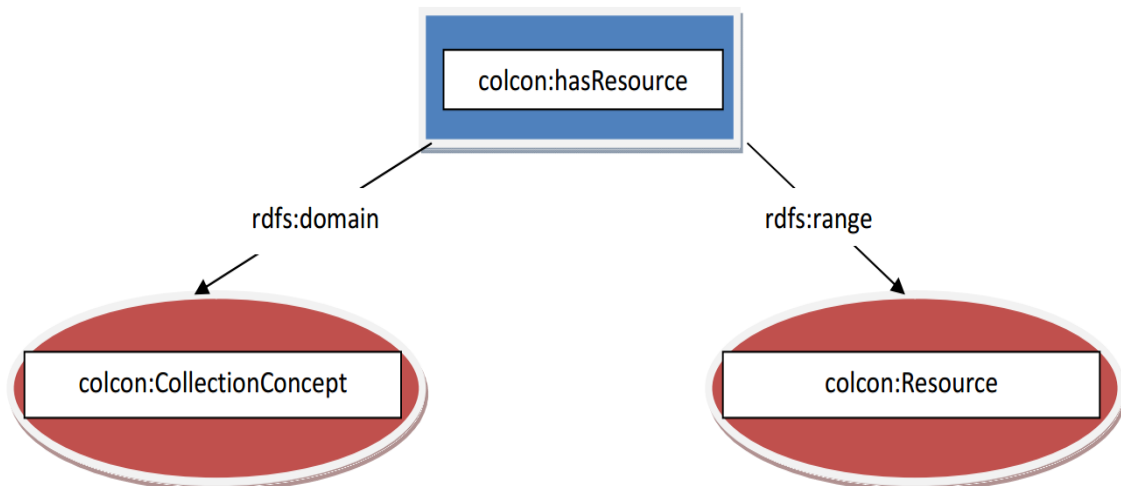


Abbildung 3: Schematische Darstellung der owl:ObjectProperty `colcon:hasResource`.

4.1.1. Die Klassen der Sammlungskonzepte

Wie aus dem verwendeten Namensraum der Klassen `CollectionConcept` und `Resource` hervorgeht, sind diese definiert als owl:Class. Hier die entsprechenden Auszüge in RDF/XML, ergänzt um die Eigenschaften `rdfs:label` und `rdfs:comment` samt `language`-Attributen (`xml:lang`):⁶⁰

```

...
<owl:Class rdf:about="colcon:CollectionConcept">
  <rdfs:label xml:lang="de">any collection</rdfs:label>
  <rdfs:comment xml:lang="en">This is the class of all kind of possible con
    cepts of collections.</rdfs:comment>
</owl:Class>
...
...
<owl:Class rdf:about="colcon:Resource">
  <rdfs:label xml:lang="de">resource</rdfs:label>
  <rdfs:comment xml:lang="en">Any possible kind of resource which can be
    collected</rdfs:comment>
</owl:Class>
...

```

Der OWL-Logik entsprechend, handelt es sich bei beiden Klassen um Unterklassen der Klasse `owl:Thing`, auch wenn dies nicht extra codiert werden muss. Sehr wohl codiert werden müssen aber sämtliche Unterklassen von `colcon:CollectionConcept` und `colcon:Resource`. Auch dafür wird wieder auf ein RDFS-Konstrukt zurückgegriffen. Mittels `rdfs:subClassOf` wird definiert, dass es sich bei den Klassen `colcon:ReferentialCollection` und `colcon:HoldingCollection` um Unterklassen von `colcon:CollectionConcept` handelt. Wie aus der Namenswahl dieser beiden Klassen deutlich wird,

⁶⁰ Aus Platzgründen wird in den übrigen Codebeispielen auf die Wiedergabe der Elemente `rdf:label` und `rdf:comment` verzichtet.

beziehen sich diese auf die von Wickett et al vorgestellten Sammlungskonzepte „Holdings Collection“ und „referential collection“.⁶¹ Gleichzeitig wird auch bestimmt, dass diese beiden Klassen komplementär zueinander sind (owl:complementOf). Dies bedeutet, dass alle Klassen, die KEINE Unterklassen der einen Klasse sind, zwangsläufig Unterklassen der anderen Klasse bilden.

```
...
<owl:Class rdf:about="colcon:ReferentialCollection">
  <rdfs:subClassOf rdf:resource="colcon:CollectionConcept"/>
  <rdfs:subClassOf>
    <owl:Class>
      <owl:complementOf rdf:resource="colcon: HoldingCollection"/>
    </owl:Class>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="colcon: HoldingCollection">
  <rdfs:subClassOf rdf:resource="colcon:CollectionConcept"/>
  <rdfs:subClassOf>
    <owl:Class>
      <owl:complementOf rdf:resource="colcon:ReferentialCollection"/>
    </owl:Class>
  </rdfs:subClassOf>
</owl:Class>
...
```

Analog zu den Unterklassen von colcon:CollectionConcept, colcon:ReferentialCollection und colcon: HoldingCollection, werden auch die weiteren Unterklassen in dieser OWL-Ontologie von Sammlungskonzepten beschrieben.⁶² Die konkreten hierarchischen Beziehungen unter den einzelnen Klassen sowie deren Namen zeigen untenstehende Grafik.

⁶¹ Wickett, Cultural Collections, S. 11. Vgl. dazu auch die von Heaney vorgeschlagene Differenzierung von Sammlungen in „collection of entities“ und „collections of information about such entitites.“, Heaney, Analytical model of collections, S. 4.

⁶² Vgl. zu colcon:Catalogue auch cdtype:CatalogueOrIndex (<http://dublincore.org/groups/collections/collection-application-profile/>)

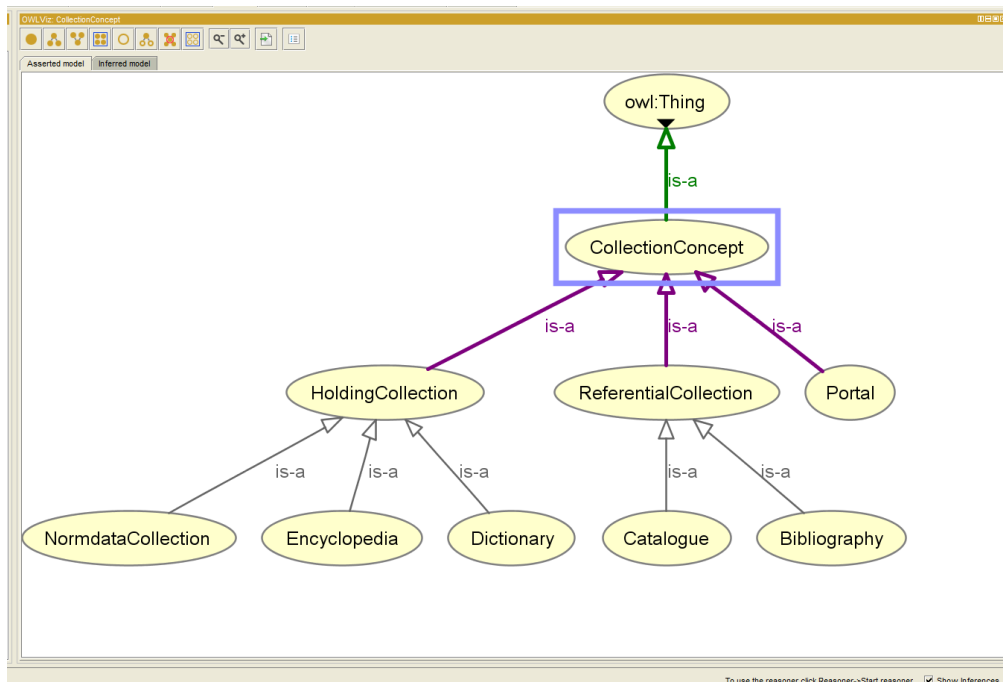


Abbildung 4: Schematische Darstellung der Klassenhierarchie der einzelnen Sammlungskonzepte. Die Darstellung wurde mit dem Protégé-Tool OWLViz generiert.

Was aus dieser Grafik nicht hervorgeht, abgesehen vielleicht von den gewählten Klassennamen, ist, wodurch diese Klassifizierung von Sammlungskonzepten begründet ist. Es wurde zwar bereits das zentrale Element dieser Ontologie, nämlich *die Beziehung zwischen Sammlung und Ressource* genannt und auch eine entsprechende formale Definition dieser Beziehung vorgestellt (`colcon:hasResource`), allerdings bezog sich diese auf die oberste Klasse der hier entworfenen Klassenhierarchie. Da alle Unterklassen die Eigenschaften ihrer Oberklassen übernehmen bzw. erben (Inference), kann eine solche Beziehung aber bei allen Klassen von Sammlungskonzepten und der Klasse `colcon:Resource` beobachtet werden.

4.1.2. Die Klassen der Ressourcen

Zu einer semantisch fundierten Ontologie von Sammlungskonzepten wird die oben skizzierte Hierarchie von Sammlungskonzepten aber erst mit der Ausbildung von Unterklassen der Klasse `colcon:Resource` und der Definition klassenspezifischer `colcon:hasResource`-Beziehungen. Eine erste Unterklasse der Klasse `colcon:Resource` bilden die **`colcon:ComplexResource`** und **`colcon:SimpleResource`**.⁶³ Auch hierbei handelt es sich um komplementäre Klassen und auch hier folgt die Namensgebung der Literatur. So schreiben Meghini et al:

⁶³ Vgl. dazu auch die im Rahmen von EDM (Europeana Data Model) diskutierten Begriffe Item, EDM Aggregation und EDM Proxy. (<http://labs.europeana.eu/api/record/>).

Today, digital information comes in the form of digital objects such as JPEG images or PDF documents, and these objects can be assembled into more complex objects. For an example, a photograph in digital form, accompanied by a caption (also in digital form) is a *complex object* consisting of two other objects, the picture and the caption. In general, a complex object is created from *simpler objects* that collectively form a meaningful unit.⁶⁴

Die vollständige Hierarchie der Klasse colcon:Resource zeigt untenstehende Grafik:

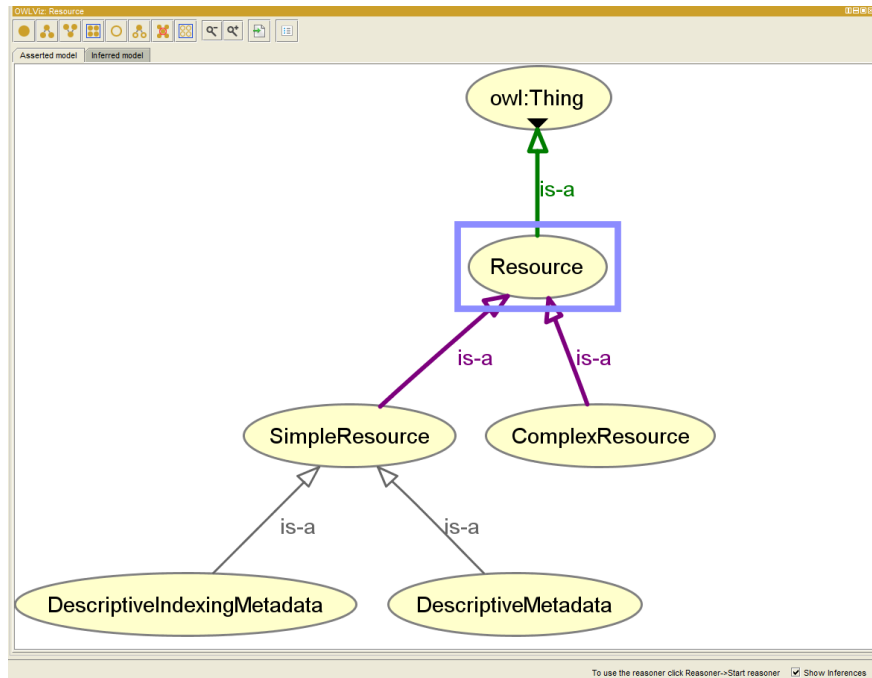


Abbildung 5: Schematische Darstellung der Klassenhierarchie verschiedenen Unterklassen von colcon:Resource.

Die Überlegungen, die zur Unterscheidung der Ressourcen colcon:ComplexResource und colcon:SimpleResource geführt haben, dürften mit dem Hinweis auf Meghini et al erschöpfend behandelt worden sein. Mittels obigem Zitat kann jedoch auch die Unterscheidung in die Klassen **colcon:Content** und **colcon:Metadata** argumentiert werden, wonach etwa das von Meghini als Beispiel angeführte digitale (aber auch analoge) Foto als eine Instanz der Klasse colcon:Content, eine Beschreibung zu („caption“) diesem Foto aber als Instanz der Klasse colcon:Metadata verstanden werden kann. Wie aus Abbildung 5 aber weiter hervorgeht, ist der Klassifizierungsprozess für die Instanz der Klasse colcon:Metadata noch nicht abgeschlossen. Es besteht nämlich noch die Möglichkeit einer weiteren Differenzierung von Ressourcen der Klasse colcon:Metadata. Die Namen dieser Unterklassen von colcon:Metadata, **colcon:DescriptiveMetadata** und **colcon:DescriptiveIndexingMetadata** geben dafür die Richtung vor. Während Instanzen erstgenannter Klasse eine Ressource ausschließlich (!) beschreiben, so wie etwa die Beschriftung zu einem Foto oder ein

⁶⁴ Meghini et al (2010), Modelling digital libraries based on logic, in: Research and Advanced Technology for Digital Libraries. Proceedings of the 14th European conference on Research and advance technology for digital libraries, pp. 2-13; eigene Kursivsetzung

Eintrag in einem Literaturverzeichnis der den Titel, den Autor, Erscheinungsort- und Erscheinungsjahr nennt, kommt bei Instanzen der Klasse `colcon:DescriptiveIndexingMetadata` noch ein Hinweis, ein Anzeiger oder eben Index auf den Standort der beschriebenen Ressource hinzu. Welche Form dieser Hinweis hat, ob es sich dabei um einen Hyperlink zur beschriebenen Ressource handelt oder um eine Standortsignatur ist unerheblich.

4.1.3. Die Eigenschaften

Um die Klassen `colcon:CollectionConcept` und `colcon:Resource` mit einander in Beziehung zu setzen, wurde die Property `colcon:hasResource` definiert (siehe oben). Will man nun die jeweiligen Unterklassen von `colcon:Collection Concept` und `colcon:Resource` ebenfalls in ein solches semantisches Verhältnis bringen, bedarf es konsequenterweise ähnlicher Konstrukte des Typs `owl:objectProperty`. Konkret handelt es sich dabei um **`colcon:hasComplexResource`** und **`colcon:hasSimpleResource`**, die beide als ‚Untereigenschaften‘ (`rdfs:subPropertyOf`) von `colcon:hasResource` definiert und mittels `rdfs:domain` und `rdfs:range` mit ihren ‚Subjekten‘ und ‚Objekten‘ `colcon: HoldingCollection` und `colcon:Complex` bzw. `colcon:ReferentialCollection` und `colcon:SimpleResource` in Beziehung ‚verknüpft‘ wurden.

Die Definition dieser beiden Eigenschaften als Untereigenschaften von `colcon:hasResource` führt dazu, dass jede mit diesen Eigenschaften beschriebene Subjekt-Objekt-Beziehung immer auch eine Subjekt-Objekt-Beziehung vom Typ `colcon:hasResource` ist.

Mit der Definition der Eigenschaften **`colcon:hasDescriptivIndexingMetadata`** und **`colcon:hasDescriptiveMetadata`**, jeweils als Untereigenschaften von `colcon:hasSimpleResource` werden abschließend noch die Ressourcen `colcon:DescriptiveIndexingMetadata` und `colcon:DescriptivMetadata` als Objekte an die Subjekte `colcon:Catalogue` und `colcon:BibliographicDatabase` bzw. an `colcon:Bibliography` gebunden.

Nachfolgende Grafik zeigt die in der OWL-Ontologie von Sammlungskonzepten beschriebenen Klassen und Eigenschaften und deren Beziehungen zueinander. Die durchgezogenen Pfeile zeigen dabei die Beziehung `rdfs:subClassOf` und die unterbrochenen Pfeile zeigen die Beziehung `colcon:hasResource`.

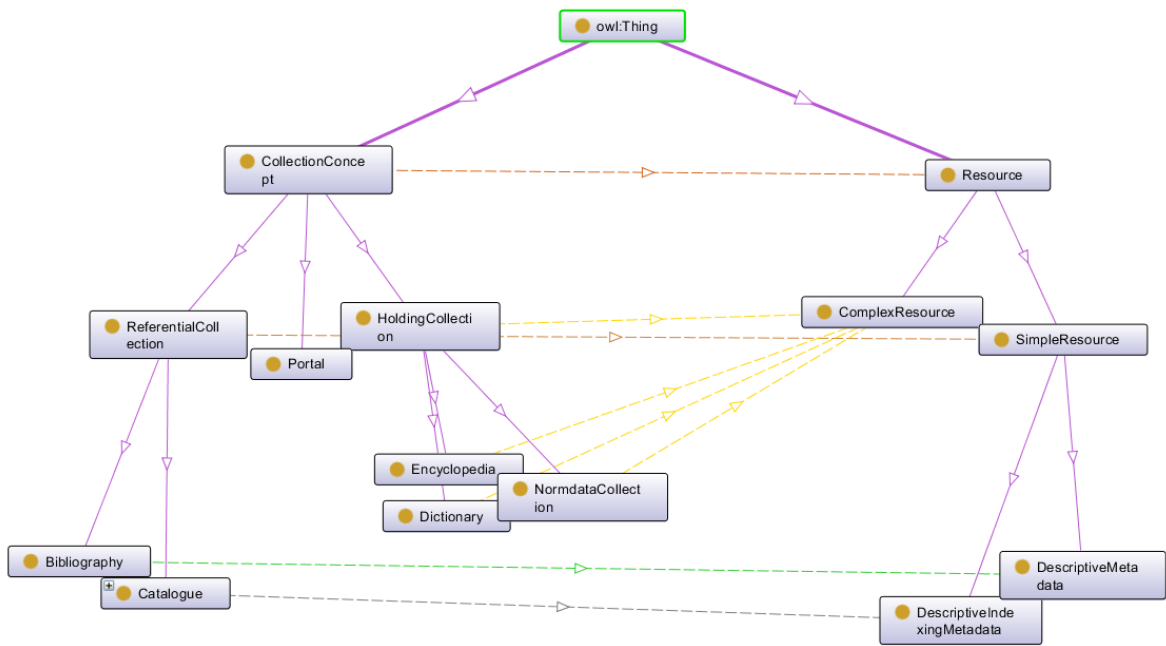


Abbildung 6: Diese mit OntoGraf generierte Darstellung der OWL-Ontologie von Sammlungskonzepten zeigt alle darin beschriebenen Klassen sowie die mittels `rdfs:subClassOf` und `colcon:hasProperty` definierten Beziehungen dieser Klassen zueinander.

5. (Weitere) Eigenschaften von Sammlungskonzepten

Die in Kapitel III verbalisierte und in Kapitel IV formalisierte Ontologie von Sammlungskonzepten fußt auf der Funktionalität der in den verschiedenen Klassen von Sammlungskonzepten gesammelten Ressourcen. Die zentrale Eigenschaft dieser Ontologie ist somit jene der Funktionalität und zwar der Funktionalität der Objekte und der sich daraus zwangsläufig ableitenden Funktionalität der Sammlung.

Es steht natürlich völlig außer Frage, dass Funktionalität weder die einzige Eigenschaft einer Sammlung ist, noch die einzige Eigenschaft, mittels derer sich eine Ontologie von Sammlungskonzepten gestalten lässt. Wie die vorigen beiden Kapitel aber gezeigt haben, eignet sich genau diese Eigenschaft sehr gut dafür, eine möglichst generische, gleichzeitig aber auch eine den realen Bedürfnissen entsprechende Ontologie von Sammlungskonzepten zu entwerfen. Nichtsdestotrotz sollen in diesem Kapitel nun weitere Eigenschaften oder Kriterien beschrieben werden, die aus Sicht der SammlungskuratorInnen, vor allem aber aus Sicht der SammlungsbenutzerInnen und je nach Anwendungsfall von Bedeutung sein können.

5.1. Vollständigkeit

Ein vor allem aus der NutzerInnenperspektive sehr relevantes Typisierungsmerkmal einer Sammlung ist das Kriterium der Vollständigkeit. Konkret geht es dabei darum, größtmögliche Gewissheit zu haben, dass ein Katalog, ein Portal, eine Datenbank auch tatsächlich die Gesamtheit aller in Frage kommenden Objekte, gleich ob als Digitalisat oder als Metadatensatz, umfasst und in diesem Sinne als eine *vollständige Sammlung* bezeichnet werden kann. Das logische Gegenteil ist dann natürlich die *unvollständige Sammlung*, die nur einen Teil aller in Frage kommenden Objekte beinhaltet. Um eine solche Typisierung von Sammlungen vorzunehmen, bedarf es konsequenterweise eines klaren Sammlungsziels, einer Bildregel, worin klar formuliert ist, welche Objekte (in welcher Form) Teil der Sammlung sind bzw. einmal sein sollen, und welche nicht.

Die Frage, ob oder ab wann eine Sammlung aber als vollständig gilt, stellt sich jedoch nicht nur aus NutzerInnenperspektive. Dies zeigt u. a. eine Studie aus dem Umfeld der ILMS Digital Collection Registry,⁶⁵ der zufolge für die Entwicklung digitaler Sammlungen verantwortliche Personen anführten, nicht zu wissen "where the collections begin and end."⁶⁶

⁶⁵ <http://imlsdcc.grainger.uiuc.edu/>

⁶⁶ Palmer, Collection, S. 4.

5.1.1. Vollständige Sammlung

Ein Beispiel einer vollständigen Sammlung wäre eine nationale Fachbibliographie. Wer wissen will, was in Österreich geschichtswissenschaftliches zu Hall in Tirol geschrieben publiziert wurde, könnte dazu in der Datenbank der „Österreichische[n] Historische[n] Bibliographie“ (ÖHB) recherchieren. Ihrem Sammlungsziel, ihrer Bildregel entsprechend sind in dieser Sammlung „die Publikationen der österreichischen Geschichtsforschung, soweit sie in Österreich erschienen sind“ erfasst. Als „Publikationen“ werden „Monographien, Reihenwerke, Artikel aus Festschriften, Sammelwerken, Zeitschriften und Periodika, Kongress- und Tagungsbände, Aufsatzsammlungen und Katalogen sowie Diplomarbeiten, Dissertationen und Habilitationsschriften, aber auch Sach- und Personalbibliographien zu historischen Themen“ verstanden. Beginn der Sammlung ist das „Berichtsjahr 1945“. Will man der Bildregel der ÖHB glauben schenken und erachtet man die für die Bibliographie zuständige Institution, die Alpen-Adria-Universität als hinreichend zuverlässig, dann darf man wohl davon ausgehen, dass in der ÖHB auch tatsächlich alle geschichtswissenschaftlichen Publikationen zur Stadt Hall in Tirol gefunden werden können, die seit 1945 in Österreich publiziert worden sind.

Die Vorteile beim praktischen Arbeiten mit einer vollständigen Sammlung dürften auf der Hand liegen. BenutzerInnen einer solchen vollständigen Sammlung haben erstens die Gewissheit, dass diese eben alle (relevanten) Objekte der Sammlungsthematik enthält, wodurch zuverlässige Aussagen und Rückschlüsse zum Gesamtbestand, wie auch zu Teilaspekten der Sammlung getroffen werden können. Anhand der anzunehmenden Vollständigkeit der ‚Sammlung ÖHB‘ kann beispielsweise festgestellt werden, dass seit 1945 150 geschichtswissenschaftliche Publikationen veröffentlicht wurden, in deren Titel die Zeichenfolge „Hall in Tirol“ anzutreffen ist. Zweitens erweist sich eine Suche in einer vollständigen Sammlung als umfassend. Wer in der ÖHB also nach geschichtswissenschaftlichen Arbeiten jüngerem Datums (seit 1945) aus ÖHB zur Stadtgeschichte Hall in Tirol gesucht hat, muss nach dieser Art von Titeln nicht mehr in anderen Sammlungen, wie etwa der Tiroler Landes- und Universitätsbibliothek suchen.

5.1.2. Unvollständige Sammlung und Auswahlammlung

Wie bereits erwähnt, steht der vollständigen Sammlung die unvollständige Sammlung gegenüber. Da der Zustand der Vollständigkeit gemeinhin als Soll-, jener der Unvollständigkeit jedoch als Mangelzustand verstanden wird, die beiden Sammlungstypen durch ihre jeweilige Benennung also in ein eindeutiges System von Wertigkeit gezwängt werden, sei als zusätzliche alternative Bezeichnung zur unvollständigen Sammlung auch der positiver konnotierte Begriff *Auswahlammlung* vorgeschlagen. Letztere unterscheidet sich von der unvollständigen Sammlung aber dahingehend, dass hier bewusst darauf verzichtet wird, eine vollständige Sammlung zu sein. Eine Beispiel einer solchen Auswahl-

sammlung ist das Deutsche Textarchiv⁶⁷. Die Auswahl der Objekte bzw. der Texte, die in diese Sammlung aufgenommen wurden, „erfolgte auf der Grundlage einer von Akademiemitgliedern erstellten und ausführlich kommentierten, umfangreichen Bibliographie.“ Bei der praktischen Arbeit mit einer solchen Auswahlammlung muss daher entschieden werden, ob die darin anzutreffende Auswahl an Objekten für die Bearbeitung einer konkreten Forschungsfrage ausreicht, oder ob andere Sammlungen hinzugezogen werden müssen. Dies ist natürlich ebenso in umgekehrter Richtung möglich, wenn ein Wissenschaftler beim Verfolgen einer spezifischen, vorher definierten Forschungsfrage eine Auswahlammlung aus anderen Sammlungen erstellt.

Bei einer unvollständigen Sammlung ist die Anzahl der darin enthaltenen Objekte – im Gegensatz zur Auswahlammlung – ungleich stärker fremdbestimmt. Anders formuliert: Eine unvollständige Sammlung ist nicht das Ergebnis einer bewussten Auswahl, sondern die Gesamtheit all jener Objekte, die unter Aufwendung der vorhandenen Ressourcen unter den gegebenen Rahmenbedingungen gefunden und zusammengetragen bzw. gesammelt werden konnten. Ein Beispiel einer solchen unvollständigen Sammlung wäre etwa der „Gesamtkatalog deutschsprachiger Leichenpredigten“. So enthält dieser Katalog bzw. diese Metadatenammlung nämlich nicht die Gesamtheit aller deutschsprachigen Leichenpredigten sondern ‚nur‘ jene 210.432 Datensätze, die der für den Gesamtkatalog zuständigen „Forschungsstelle für Personalschriften“ bekannt sind.

Den Unterschied zwischen einer unvollständigen und einer vollständigen Sammlung kann manchmal nur ein einziges Wort im Sammlungstitel oder in der Sammlungsbeschreibung ausmachen. Im Falle des eben genannten Beispiels würde es sich dann um eine vollständige Sammlung handeln, wenn diese hieße: „Gesamtkatalog bekannter deutschsprachiger Leichenpredigten“. Wie sinnvoll, aussagekräftig oder hilfreich ein solcher Sammlungstitel aber ist, muss letztendlich im Gesamtkontext der Sammlung und vor dem Hintergrund der von NutzerInnen der Sammlung konkret verfolgten Forschungsfrage/n bewertet werden.

5.2. Wissenschaftlichkeit

Das DARIAH-DE Stakeholdergremium „Wissenschaftliche Sammlungen“⁶⁸ setzt sich aus VertreterInnen von Fachwissenschaften, Archiven, Bibliotheken und Museen zusammen. Dieses Gremium hat gemeinsam eine allgemeine Definition der Begriffe „Wissenschaftliche Sammlung“ und „Forschungsdaten“ für den wissenschaftlichen Gebrauch vorgeschlagen.

⁶⁷ deutschestextarchiv.de

⁶⁸ Vgl. dazu Oltersdorf et al. 2014

Nach dieser Definition besteht eine Sammlung wissenschaftlich relevanter digitaler Forschungsdaten aus diskreten, voneinander verschiedenen, logisch unabhängigen Einheiten und wurde intentional angelegt, d.h. es liegt eine *Bildregel* vor.

Die Sammlung muss nicht aus einem Forschungsanliegen heraus entstanden sein, ist aber Gegenstand wissenschaftlicher Fragestellungen bzw. hat das Potential zu einer wissenschaftlichen Analyse und dient der Validierung von Aussagen, Methoden, Thesen, Hypothesen oder Theorien in Forschung und Lehre, um die *Reproduzierbarkeit* der Ergebnisse zu gewährleisten.

Nach dem Konzept des *research data life cycle* kann eine Sammlung sowohl Ursprung als auch Ergebnis wissenschaftlicher Arbeit sein und ist zusätzlich in einer regelhaften Form maschinenlesbar, dokumentiert und idealerweise nach internationalen Standards erfasst und mit Normdaten ausgezeichnet für automatisierte *Prozessier- und Interpretierbarkeit*.

Des Weiteren gibt eine Sammlung Auskunft über ihren Rechtsstatus, vor allem über zu beachtende *Nutzungsbedingungen und Lizenzen*.

Sie dient der Ordnung der Sammlungsgegenstände und der archivischen Sicherung (*Archivierung*).

Daten, unabhängig ihrer Provenienz, werden im Kontext einer geistes- und kulturwissenschaftlichen Forschungsfrage zu *digitalen Forschungsdaten*, sobald sie gesammelt, beschrieben, ausgewertet und/oder erzeugt und in maschinenlesbarer Form zum Zwecke der Nachvollziehbarkeit von Forschungsergebnissen sowie zur Archivierung, Zitierbarkeit und Weiterverarbeitung aufbewahrt werden.

Sobald mehrere solcher Daten nach den o.g. Kriterien aggregiert werden, kann demzufolge von einer *Wissenschaftlichen Sammlung* gesprochen werden.

5.3. Prozessierbarkeit/Interoperabilität

Auch durch fortschreitende Digitalisierung von Beständen werden mehr und mehr Sammlungen im Internet öffentlich zugänglich gemacht. Gerade die „Gedächtnisorganisationen“, also Museen, Archive und Bibliotheken präsentieren so ihre Daten. In gleichem Maße steigt auch der Bedarf von NutzerInnen an Möglichkeiten, diese Datenbanken zu vergleichen, global zu durchsuchen und Daten rechnerunterstützt automatisiert auszutauschen (Vgl. Doerr 2003, S. 75).

Ein grundlegendes Problem der letzten vierzig Jahre, das auch bisher die Durchsetzung eines Standards verhinderte, ist die Verschiedenartigkeit der Nutzungsanforderungen an die Daten und die daraus resultierenden unterschiedlichen Metadatenschemata, die bei der Erstellung der digitalen Sammlungen genutzt werden (Haslhofer und Neuhold 2012, S. 2; Vierkant et al. 2012).

Weinberger (2012) postuliert:

„*We are not going to agree on a single metadata standard.*“

Um diesem Umstand Rechnung zu tragen, wurde in den letzten zehn Jahren Linked Data (Bsp. Semantic MediaWiki⁶⁹) wieder verstärkt in den Fokus genommen, als „Nebenprodukt“ der Entwicklungen des www und im Speziellen der Durchsetzung von Ontologien⁷⁰ für spezifische Bereiche (wie z.B. MPEG-7 für Multimedia). Auch argumentieren Haslhofer und Neuhold, dass keine allgemeine Lösung für das Problem der Interoperabilität absehbar ist. Das „Mapping“ eines Datenmodells auf ein anderes wird weiterhin dem Bedarf von Einzelnen entsprechend unternommen und Variationen werden diesem Bedarf folgend entstehen und nebeneinander existieren (Haslhofer und Neuhold 2012, S.19).⁷¹

Baker et. al (2002) beschreiben drei Stufen der Interoperabilität, die adressiert werden müssen: Protokolle für Transport und Austausch, die maschinenlesbare Datenrepräsentation sowie Werte und Attribute in Kontrollierten Vokabularien.

Daraus lässt sich folgern, dass ein Mapping eines Datenmodells auf ein anderes Datenmodell derzeit nur mit Kenntnis und Verständnis der Eigenheiten und Details der beiden Schemata sinnvoll angegangen werden kann und auf diesen einen Use Case beschränkt bleiben sollte.

Probleme beim Prozessieren der Metadaten können sowohl auf Ebene der Sammlungsbeschreibung als auch einzelner Objekte entstehen. Allen et al. (2008, S.80) merken an, dass gerade die spezifischen Designanforderungen von Modellen, die zur Unterstützung einer speziellen Forschungsfrage erstellt wurden in der Regel nicht austauschbar sind und demzufolge in der Realität sammlungsbeschreibende Metadaten häufig völlig ignoriert werden.

5.4. Schnittstellen

Um diese Bedeutungsverluste beim automatisierten Austausch möglichst gering zu halten, ist es sinnvoll, einen etablierten Standard wie das Protokoll OAI-PMH⁷² zu nutzen, dessen derzeitige Iteration bereits seit 2002 besteht und dementsprechend stabil und in der Community verankert ist. Beer et al. (2014, S.11ff) zeigen verschiedene Wege auf, wie z.B. in XML vorliegende Metadaten für das OAI-PMH lesbar und prozessierbar gemacht werden können.

Die Entscheidung für oder gegen ein Metadatenschema hängt von verschiedenen Faktoren ab. Zentral ist hier der wissenschaftliche Hintergrund bei Erstellung der Sammlung, die Einordnung des Projektes – die „Forschungsfrage“. Welche Schnittstellen dann angeboten werden folgt oft aus dieser Entscheidung. Versuche, reines HTTP zu nutzen oder TEI in RDF/XML abzubilden sind eher ein Anzei-

⁶⁹ semantic-mediawiki.org

⁷⁰ Nach Gruber (1993) „Spezifikationen von Konzeptualisierungen“.

⁷¹ Zu den vielfältigen Problemen der technischen Umsetzung von Interoperabilität vgl. Haslhofer, und Klas 2010.

⁷² <http://www.openarchives.org/pmh/>.

chen, dass umfangreicher automatisierter Datenaustausch auf der Prioritätenliste der meisten Projekte nicht besonders weit oben steht (Beer et al. 2014, S.14).

XML als das in den Geisteswissenschaften am weitesten verbreitete Metadatenschema hat den Vorteil, dass es auch für Menschen leicht lesbar ist. Ausgehend von RDF/XML empfehlen Beer et al. (2014, S.19f), derzeitigen Entwicklungen des WWW folgend, auch auf Linked Open Data zu setzen. Linked Open Data ist eine Bezeichnung für Beschreibungen in RDF/XML, die HTTP nutzt. Beschreibungen in XML erhalten durch das RDF-Modell eine Semantik, die es ermöglicht, automatisiert Bedeutungsverbindungen zu prozessieren (Beer et al. 2014, S.21).

5.5. Auffindbarkeit

Welche Öffentlichkeit eine Sammlung erreicht, ist ohne Zugriff auf den Server der haltenden Institution kaum messbar. Es liegt im Interesse und ist Auftrag vieler Gedächtnisinstitutionen, allen interessierten NutzerInnen kostenlosen Zugriff auf vorhandene Sammlungen zu geben, gerade wenn es sich um mit DFG-Fördermitteln digitalisierte Sammlungen handelt. Diese Sammlungen haben in der Regel eine eigene designierte Einstiegsseite mit integrierter Suchmaschine, um Objekte im Internetbrowser anzeigen zu lassen.

Der Zugriff auf solche Sammlungen ist in erster Linie potenziell dadurch beschränkt, ob und wie sie von gängigen Internetsuchmaschinen angezeigt werden. Um die hier möglichen Probleme zu umgehen, entstehen vermehrt Sammlungen über Sammlungen wie die DARIAH-DE Collection Registry⁷³ oder Archivportal D⁷⁴, über deren integrierte Suchen einzelne Sammlungen und von dort aus Objekte gefunden werden können.

⁷³ <http://colreg.de/dariah.eu/colreg/colreg/main> - Prototyp, der von DARIAH-DE entwickelten Generischen Suche mit Zugriff auf die Collection Registry für Januar 2015 geplant

⁷⁴ <https://www.archivportal-d.de/>

6. Implementierung des Gesagten in das DCLAP-Schema

In der ersten Projektphase von DARIAH-DE wurde mit dem DARIAH-DE Collection Level Description Application Profile, in weiterer Folge DCLDAP genannt, ein technisch wie inhaltlich auf der Höhe der Zeit befindliches Schema zur Beschreibung von Sammlungen entwickelt.⁷⁵ Abgesehen davon, dass damit Sammlungen in detaillierter Form und unter Einbeziehung etablierter wie einschlägiger Standardvokabularien aus dem Umfeld von Dublin Core, MARC, PURL oder vCard bis auf die Ebene der Sammlungsobjekte erfasst werden können, ist das Schema auch auf die technische Infrastruktur DARIAH-DEs abgestimmt. Allen voran zeichnet sich DCLDAP für die Validierung sämtlicher von der DARIAH-DE Collection Registry verwalteten Datensätze von Sammlungsbeschreibung verantwortlich. Was im DCLDAP-Schema jedoch fehlt, ist die Möglichkeit der expliziten Benennung des Sammlungskonzepts, dem eine im DCLDAP-Schema beschriebene Sammlung zuzurechnen wäre. Die Formulierung explizite Benennung muss hier insofern betont werden, als das jeweilige Sammlungskonzept sicherlich zumindest in Ansätzen aus der Zusammenschau der in der Sammlungsbeschreibung angeführten Informationen erschlossen werden kann, etwa aus den Angaben zu den gesammelten Objekttypen (cld:itemType) und -formaten (cld:itemFormat). Insofern wäre es eine lohnenswert zu untersuchen, ob und wie sich zwischen den hier beschriebenen Klassen von Sammlungskonzepten und Ressourcen und den verschiedenen Objekttypen und -formaten semantische Beziehungen definieren ließen, sodass etwa anhand der gesammelten Objekttypen und -formaten auf das Sammlungskonzept geschlossen werden kann.

Weniger elegant, dafür aber einfacher in der Realisierung ist sicherlich eine explizite Benennung des Sammlungskonzeptes. Bisher werden die Ressourcen, die nach DCLDAP beschrieben werden, über die Eigenschaft dc:type der Collection-Klasse beschrieben, die als einzigen Wert das Literal „Collection“ kennt.⁷⁶ Ersetzt man diese Vorgabe nun aber mit den in diesem Paper bzw. in der OWL-Ontologie von Sammlungskonzepten definierten colcon:CollectionConcept-Klassen, so kann nicht nur das Sammlungskonzept der Sammlung benannt, sondern auch die Sammlung in die Ontologie von Sammlungskonzepten eingebunden werden.

Ein Beispiel: Im Falle der DCLDAP-Beschreibung von „epidat – epigraphische Datenbank zur jüdischen Grabsteinepigraphik“,⁷⁷ wird in der dclap:Collection-Klasse als dc:type dieser Ressource angeführt:

⁷⁵ Ausführlich zum DCLDAP-Schema siehe DARIAH-DE Collection Level Description Application Profile (<https://dev2.dariah.eu/wiki/x/VoFn>).

⁷⁶ Vgl. dazu DARIAH-DE Collection Level Description Application Profile (<https://dev2.dariah.eu/wiki/x/VoFn>).

⁷⁷ <http://colreg.de.dariah.eu/colreg/colreg/collectionDetails?id=101>.

```

<dclap:dclapDescription ... rdf:about="101">
  <dclap:Collection>
    ...
    <dc:type>Collection</dc:type>
    ...
  </dclap:Collection>
  ...
</dclap:dclapDescription>

```

Die Beschreibung entspricht somit dem DCLDAP-Schema, ist in diesem Punkt aber völlig identisch zu den 111 restlichen, in der Collection Registry verzeichneten Einträge.⁷⁸ Eine Implementierung des colcon:CollectionConcepts-Ontologie könnte wie folgt aussehen:

```

<dclap:dclapDescription ... rdf:about="101">
  <dclap:Collection>
    ...
    <dc:type rdf:resource="colcon: HoldingCollection"/>
    ...
  </dclap:Collection>
  ...
</dclap:dclapDescription>

```

Obwohl nur ein minimalinvasiver Eingriff, wären wenigstens auf drei Ebenen Verbesserungen gegenüber dem aktuellen Stand zu erwarten.

- (1) Zuallererst erzwingt die verpflichtende Bestimmung des dc:type Elements eine minimale Auseinandersetzung mit dem Wesen der beschriebenen Sammlung und eine Reflektion über das darin eingeschriebene Sammlungskonzept.
- (2) Eine systematische Bestimmung und Benennung des konkreten Sammlungskonzeptes würde eine maßgeschneiderte Entwicklung von Abfrage- und Darstellungs-Interfaces ermöglichen.
- (3) Daran anschließend würde eine systematische Bestimmung und Benennung des konkreten Sammlungskonzeptes auch bessere Suchmöglichkeiten und Ergebnisse mit sich bringen. BenutzerInnen können von vornherein nach den gewünschten Sammlungskonzepten respektive den entsprechenden Ressourcen suchen bzw. die Qualität des Suchergebnisses besser einschätzen.

Dass aber selbst ein kleiner Eingriff, wenn er 112-Mal ausgeführt wird, durchaus einen beträchtlichen Arbeitsaufwand mit sich zieht, soll an dieser Stelle nicht verschwiegen werden.

⁷⁸ Stand vom 08.12.2014.

7. Ausblick

Die vorherigen Überlegungen werden z.Zt. – auch aufgrund von Rückmeldungen aus der Entwicklung der Collection Registry und externen Partner – überarbeitet. Das DCLDAP wird dabei zum DARIAH Collection Description Data Model (DCDDM) weiterentwickelt.⁷⁹ Nach Abschluss dieses Prozesses und der Implementierung des Datenmodells in die von DARIAH-DE angebotenen Dienste, werden die Hintergründe dieser Weiterentwicklung sowie das Ergebnis in einem eigenständigen Working Paper dokumentiert werden.

⁷⁹ Der aktuelle Entwurf des Datenmodells ist dokumentiert unter <https://github.com/DARIAH-DE/DCDDM/blob/master/DCDDM.odt>

8. Literatur

DARIAH Project Bibliography⁸⁰ wird berücksichtigt und (ggf.) erweitert.

Allemang, Dean & Hendler, Jim (2011): *Semantic Web for the Working Ontologist. Effective Modeling in RDFS and OWL*. Amsterdam u.a. (2. Aufl.).

Renear, Allen H. et al (2008): "Collection/Item Metadata Relationships". <http://edoc.hu-berlin.de/conferences/dc-2008/renear-allen-80/PDF/renear.pdf>.

Baker, T. et al (2002): "Principles of metadata registries". In: *DELOS Network of Excellence on Digital Libraries*. <http://dl.acm.org/citation.cfm?doi=1667062.1667064>.

Bartsch, Eberhard (1979): *Die Bibliographie. Einführung in Benutzung, Herstellung, Geschichte*. München.

Beagrie, Neil (2005): "Plenty of Room at the Bottom? Personal Digital Libraries and Collections". In: *D-Lib Magazine* 11/6. <http://www.dlib.org/dlib/june05/beagrie/06beagrie.html>.

Bekiari, C., Doerr, M. & Le Boeuf, P. (2008): "FRBRoo, a Conceptual Model for Performing Arts". In: *2008 Annual Conference of CIDOC*, Athens. <http://www.ics.forth.gr/publications/drfile.2008-06-42.pdf>.

Buckland, Michael K., (1989): "The Roles of Collections and the Scope of Collection Development". In: *Journal of Documentation* 45 (1989), 213-226.

Currall, J., Moss, M. & Stuart, S. (2004): "What is a collection?". In: *Archivaria* 58 (2004), 131-146. <http://journals.sfu.ca/archivar/index.php/archivaria/article/viewArticle/12480>.

Doerr, Martin (2003): "The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata". In: *AI Magazine* 24/3 (2003). doi:10.1609/aimag.v24i3.1720.

Geisler, G. et al (2002): "Creating virtual collection sin digital libraries: benefits and implementation issues", in: *Joint Conference on Digital Libraries*. Portlan 210-218.

Gruber, Tom (1993): "A translation approach to portable ontology specifications". In: *Knowledge Acquisitions* 5, 199–220. <http://tomgruber.org/writing/ontolingua-kaj-1993.pdf>.

Haslhofer, Bernhard & Klas, W. (2010): "A survey of techniques for achieving metadata interoperability". In: *ACM Comput. Surv.* 42/2 Article 7 (2010). <http://doi.acm.org/10.1145/1667062.1667064>.

Haslhofer, Bernhard & Neuhold, Erich J. (2011): "A Retrospective on Semantics and Interoperability Research". In: *Foundations for the Web of Information and Services*, hg. Dieter Fensel. Berlin:Springer, 3–27. <http://cs.univie.ac.at/research/research-groups/multimedia-information-systems/publikation/infpub/2921/>.

Heaney, M. (2000): "An analytic model of collections and their catalogues, in UK Office for Library and Information Science". <http://www.ukoln.ac.uk/metadata/rsip/model/amcc-v31.pdf>.

Hill, L. at al (1999): "Collection metadata solutions for digital library applications". In: *Journal of the American Society for Information Science* 50/13 (1999), 1169-1181.

⁸⁰ <https://www.zotero.org/groups/45002>

- Jalal, Samir Kumar (2009): "Collection development in digital environment: A case study". In: *Indian Journal of Library and Information Service* 3 (2009), vol.3, n. 3. S. 151-160.
- Marie R. Kennedy (2008): "Nine questions to guide you in choosing a metadata schema". In: *Journal of Digital Information* 9/1 (2008).
<https://journals.tdl.org/jodi/index.php/jodi/rt/prINTERfriendly/226/205>.
- Lagoze, C. & Fielding, D. (1998): "Defining collections in distributed digital libraries". In: *D-Lib magazine* 54/11 (1998). <http://www.dlib.org/dlib/november98/lagoze/11lagoze.html>.
- Lee, H. (2000): "What is a collection?". In: *Journal of the American Society for Information Science* 51/12 (2000), 1106-1113.
- Lee, Hur-Li (2003): "Collection Development as a Social Process". In: *Journal of Academic Librarianship* 29/1 (2003), 23–31.
- Lee, H. (2005): "The concept of collection from the users` perspective". In: *The Library Quarterly* 75/1, 67-85.
- Meghini, C. & Spyratos, N. (2007): "Viewing collections as abstractions". In: *Digital Libraries: Research and Development*, 207-217.
- Meghini, C. & Spyratos, N. (2010): "Unifying the concept of collection in digital libraries". In: *Advances in Intelligent Information Systems*, 197-224.
- Meghini et al (2010): "Modelling digital libraries base on logic". In: *Proceedings of the 14th European conference on Research and advance technology for digital libraries*. hg. von Mounia Lalmas et al., 2-13.
- Miller, Stephan J. (2014): *Metadata for Digital Collections*. London 2014.
- Oltersdorf, Jenny, Schmunk, Stefan & Stäcker, Thomas (2014): "Stakeholdergremium Wissenschaftliche Sammlungen", Report.
<https://dev2.dariah.eu/wiki/download/attachments/14651583/R%204.1.1%20Stakeholder%20Wissenschaftliche%20Sammlungen.pdf?version=1&modificationDate=1404226725584&api=v2>.
- Palmer, Carol L. et al (2006): "Collection Definition in Federated Digital Resource Development". Austin. <http://eprints.rclis.org/8638/>.
- Schneider, Georg & Nestler, Friedrich (1999): *Handbuch der Bibliographie. Völlig neu bearbeitet von Friedrich Nestler*. Stuttgart.
- Stock, Wolfgang G. & Stock Mechtild (2013): *Handbook of Information Science*. Berlin:De Gruyter.
- Vierkant et al. (2012): "Vocabulary for the Registration and Description of Research Data Repositories". http://gfzpublic.gfz-potsdam.de/pubman/item/escidoc:76875/component/escidoc:76874/re3data_vocabulary_v2-0.pdf.
- Weinberger (2012): "[The inevitable messiness of digital metadata](http://www.hastac.org/blogs/ruby-sinreich/2012/07/13/accepting-challenges-representing-scholarly-knowledge)".
<http://www.hastac.org/blogs/ruby-sinreich/2012/07/13/accepting-challenges-representing-scholarly-knowledge>.
- Wickett, Karen M. et al (2013): "Modeling cultural collections for digital aggregation and exchange environments". <http://hdl.handle.net/2142/45860>.

9. Benutzte Webseiten

http://archivesic.ccsd.cnrs.fr/sic_00000594

<http://www.deutschestextarchiv.de/>

<https://dev2.dariah.eu/wiki/x/VoFn>

<http://www.dfg.de>

<http://digital-archiv.at:8081/exist/apps/DARIAH-Collection/xquery/distinctCollections.xql>

https://docs.google.com/spreadsheets/d/1_EPLFs68nisdGcyrtuVZshk6nyANgkA5M8rMmA5xoeA/edit?usp=sharing

<http://dublincore.org/groups/collections/collection-application-profile/>

<http://www.duden.de/rechtschreibung/Sammlung>

<http://www.europeana.eu/>

<http://gepris.dfg.de/gepris/OCTOPUS>

<http://www.gesamtkatalogderwiegendrucke.de/>

<http://www.gutenberg.org>

<http://www.hsozkult.de/>

<http://imlsdcc.granger.uiuc.edu/>

<http://www.mla.org/>

<http://www.personalschriften.de/datenbanken/tbk.html>

<http://protege.stanford.edu/>

<http://www.rsa.org/>

semantic-mediawiki.org

<http://synonyme.woxikon.de/>

<http://theaterforschung.de>

http://www.ubka.uni-karlsruhe.de/kvk_en.html

<http://www.virtuelles-kupferstichkabinett.de/>

<https://www.zotero.org/groups/45002>

<http://colreg.de.dariah.eu/colreg/colreg/main>

<https://www.archivportal-d.de/>

10. Anhang

[CollectionConcepts RDF/XML](#)