

Schwager, Thomas

Entwicklung eines Metadatensets für AV-Daten im Rahmen des Forschungsdatenzentrums "Forschungsdaten Bildung" am DIPF

2012, 45 S. - (Potsdam, Fachhochschule, 2012)

urn:nbn:de:0111-dipfdocs-73833



Nutzungsbedingungen / conditions of use

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

Deutsches Institut für
Internationale Pädagogische Forschung
Mitglied der Leibniz-Gemeinschaft
Frankfurter Forschungsbibliothek
Schloßstraße 29
D-60486 Frankfurt am Main
publikationen@dipf.de
www.dipf.de/de/bildungsinformation/ffb

Fachhochschule Potsdam

Institut für Information und Dokumentation

Kurs 2012

Entwicklung eines Metadatensets für AV-Daten im Rahmen des Forschungsdatenzentrums „Forschungsdaten Bildung“ am DIPF

Abschlussarbeit im Rahmen der Fortbildung zum / zur Wissenschaftlichen Dokumentar/in / Information Specialist

vorgelegt von

Thomas Schwager

Thematischer Schwerpunkt:

Wirtschaftsinformatik-Grundlagen für Informationsspezialisten

Dozent: Sven Hirsch

Frankfurt/Main, August 2012

Abstract

Die Arbeit fokussiert die Bedingungen bei der Implementierung des Projekts „Forschungsdaten Bildung“ am Deutschen Institut für Internationale Pädagogische Forschung (DIPF) für den Metadatenbereich. Dabei wird zunächst der Begriff Forschungsdatenzentrum exemplarisch samt Entwicklungsprozessen skizziert. Nach einer Vorstellung der konkreten Ausgangslage für „Forschungsdaten Bildung“ inklusive des Datenbestandes widmet sich die Arbeit dem Thema Metadaten und Metadatenstandards. Hierbei wird das Augenmerk auf wesentliche Aspekte bei der Schaffung eines Anwendungsprofils, „Application Profile“, wie die Nutzung von XML und die Einbindung relevanter Metadatenstandards gelegt. Im Mittelpunkt stehen Dublin Core und die Data Documentation Initiative (DDI) in ihren Funktionsweisen sowie übliche audiovisuelle Metadatenstandards. Schließlich werden Anforderungskriterien für die Entwicklung des Metadatensets für „Forschungsdaten Bildung“ angeführt und eine erste Präzisierung der Metadatenelemente in Anlehnung an Dublin Core und DDI vorgenommen.

Inhaltsverzeichnis

1 Einleitung	1
2 Forschungsdatenzentren in Deutschland	2
2.1 Allgemeine Ausgangslage.....	2
2.2 Technische Voraussetzungen für Forschungsdatenzentren.....	3
2.3 Forschungsdatenzentren einzelner Wissenschaftsdisziplinen	4
2.3.1 Forschungsdatenzentren in den Geowissenschaften.....	5
2.3.2 Forschungsdatenzentren in den Sozialwissenschaften.....	5
2.3.3 Das Beispiel des Datenarchivs für Sozialwissenschaften GESIS.....	6
3 Das Portal „Forschungsdaten Bildung“ des DIPF	7
3.1 Zielsetzung und Selbstverständnis als Datenservice- und Forschungsdatenzentrum.....	8
3.2 Technische Infrastruktur	9
3.3 Datenbestand.....	10
3.4 Ausstehende Anforderungen.....	12
4 Metadaten und Metadatenstandards	12
4.1 Ein Kategorisierungsansatz für Metadaten	13
4.2 Grundprinzip Interoperabilität	14
4.3 XML, XML-Schema und Application Profile	15
4.4 Metadateninitiativen und Standards zur Datenstrukturierung.....	17
4.4.1 Dublin Core	18
4.4.2 Data Documentation Initiative (DDI).....	21
4.4.3 Die Situation im audiovisuellen Bereich	23
5 Der Entwicklungsprozess eines Metadatensets für AV-Daten am DIPF	25
5.1 Bezugspunkte für die Auswahl der Metadatenelemente	25
5.2 Metadatenanforderungen für AV-Daten der Bildungsforschung.....	27
5.3 Modulare Formen und exemplarische Umsetzung.....	30
6 Abschließende Bemerkung.....	33
7 Literatur- und Quellenverzeichnis.....	34
8 Anhang	42

1 Einleitung

Jahrzehntlang oft stiefmütterlich behandelt, mehren sich gerade in diesbezüglich schlechter ausgestatteten Wissenschaftsdomänen Stimmen für einen Ausbau und die Weiterentwicklung von Forschungsinfrastrukturen zwecks Sicherung von Forschungsdaten. Trotz mancherorts fortgeschrittener Entwicklungen besteht Nachholbedarf hinsichtlich geeigneter Möglichkeiten, Forschungsdaten für Sekundäranalysen aufzubereiten. Dies trifft in unterschiedlichem Maße auf die einzelnen Forschungsbereiche und Datentypen zu und obliegt deutschlandweit einer relativ starken Heterogenität. Vor allem in interdisziplinären Forschungsfeldern mangelt es an Qualitätskriterien – ein Zustand, der mittlerweile thematisiert wird.¹

Auch für den in diesem Rahmen behandelten Bereich der audiovisuellen Forschungsdaten der Bildungsforschung trifft dies zu. Hier mangelt es bisher noch an klaren Strukturen und Konventionen für die Metadatenvergabe, Standardauswahl sowie weiterführende Datenaufbereitungsprozesse. Aus dieser Motivlage initiiert das Deutsche Institut für Internationale Pädagogische Forschung (DIPF) für den Bereich Bildungsforschung das Forschungsdatenprojekt „Forschungsdaten Bildung“. Im Zuge dieses Vorhabens sollen Datenstrukturen für die Dokumentation von qualitativen Forschungsprimärdaten geschaffen werden.

Die hier vorliegende Arbeit thematisiert einen wesentlichen Teilbereich dieses Entwicklungsprozesses: Sie unternimmt einen ersten Versuch, einen Kern an Metadatenelementen für den speziellen Bereich der audiovisuellen Forschungsdaten zu bestimmen und den ersten Schritt hin zu einem Application Profile zu bewältigen. Für eine genaue Bestimmung ist es dabei von elementarem Interesse, einerseits die aus dem Datenbestand und dem Nutzerinteresse erwachsenen Anforderungen samt Zielsetzung von Forschungsdatenzentren und des Projekts selbst im Speziellen zu untersuchen und andererseits standardisierte Methoden für die Vergabe von Metadaten auf ihre Tauglichkeit zu prüfen. Deswegen sollen hier der Entstehungshintergrund des Projektes ebenso beleuchtet werden wie die speziellen Datenbestände und relevante Metadatenstandards in ihrer Funktionsweise und Umsetzung im Rahmen von „Forschungsdaten Bildung“. Dies soll abschließend in einer entsprechenden Auswahl von notwendigen Metadatenelementen in Form eines Sets münden. Dabei muss betont werden, dass es sich auf Grund des Projektstatus nur um eine vorläufige Momentaufnahme handeln kann, die aber eine wichtige Ausgangsbasis für künftige Arbeitsschritte bietet.

¹ Vgl. Wissenschaftsrat (2012), S.56.

2 Forschungsdatenzentren in Deutschland

Die von der Europäischen Kommission ins Leben gerufene High Level Expert Group on Scientific Data konstatierte 2010 in ihrer „Vision 2030“:

“A fundamental characteristic of our age is the rising tide of data – global, diverse, valuable and complex. (...) With a proper scientific e-infrastructure, researchers in different domains can collaborate on the same data set, finding new insights. They can share a data set easily across the globe, but also protect its integrity and ownership. They can use, reuse and combine data, increasing productivity.”²

Dieses Statement zeigt die zentrale Motivlage beim Ausbau der Strukturen zum Forschungsdatenerhalt. Im Zuge des enormen Anstiegs an erhobenen Datenmengen sollen Infrastruktureinrichtungen geschaffen werden, die die interoperable Nachnutzung von Forschungsdaten ermöglichen. Der Begriff Forschungsdatenzentrum (FDZ) impliziert genau dies. Es geht per definitionem um einen Ort, an dem alle erdenklichen Datentypen aus verschiedenen Wissenschaftsdisziplinen und -projekten archiviert und aufbereitet werden mit dem Ziel, Data Sharing für die Nachnutzung, Data Enrichment (Datenanreicherung) sowie eine Verifikation³ der Forschungsergebnisse zu ermöglichen. Im Folgenden wird der Entwicklungsstand in Deutschland skizziert, der dabei in etwa als Spiegelbild der internationalen Situation gelten kann.⁴

2.1 Allgemeine Ausgangslage

Forschungsprimärdaten entstehen täglich in verschiedenem Ausmaß. Der Umgang mit diesem Datenmaterial divergiert aber je nach Wissenschaftsdisziplin. Manche Fachgebiete haben einen wesentlich höheren Professionalisierungsgrad als andere, was zum Beispiel die Schaffung von Standards für eine Erfassung und Beschreibung von Forschungsdaten betrifft, und haben auch bereits fortgeschrittene Datenarchive etabliert.⁵ In anderen Disziplinen hingegen mangelt es immer noch an Konventionen, einer Dokumentationsstrategie und dadurch am Zugang zu

² High Level Expert Group on Scientific Data (2010), S. 4.

³ Hinsichtlich des Bedürfnisses nach Verifikation herrscht nicht immer Einigkeit, vgl. Medjedović (2010), S. 61f.

⁴ Vgl. Strathmann (2012), S. 41.

⁵ Vgl. Kleiner (2012), S. 10. Vor allem in der Medizin, vgl. Dickmann (2012), und z.B. in der Psycholinguistik, vgl. Wittenburg (2012), haben sich handfeste Strukturen entwickelt. Dies führte in jenen betroffenen Fachbereichen früh zu festen Absprachen, Projekten und Standards mit Vorteilen für die Datenmanagementpraxis.

Forschungsdaten und Nachnutzungsmöglichkeiten.⁶ Außerdem war es bei einer oft dezentralen Ausrichtung der Forschungsstrukturen durchaus Usus, diese verstreuten Datenmengen nur für das eigene Projekt zu verwenden. Da eine nachträgliche Dokumentation oft schwer möglich ist, resultierte daraus oft ein schleichender teilweiser Datenverlust essentieller Forschungsdaten⁷ – mit seinen forschungspolitischen und -ökonomischen Konsequenzen.

Im Jahr 1998 erkannte die Deutsche Forschungsgemeinschaft (DFG) diese Problematik und formulierte die Empfehlung, „Primärdaten (...) auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für zehn Jahre“⁸ aufzubewahren. Von einer effizienten Nachnutzung, Validierung, interdisziplinärer Einbindung oder interoperablem Datenaustausch konnte jedoch meist noch keine Rede sein. Deswegen unternahm die DFG in ihrem Positionspapier aus dem Jahr 2006 einen weiteren Schritt zur Datensicherung. Sie legte als Förderungsschwerpunkt auf Grund der „notwendigen Vernetzung der heterogenen Informationsangebote (...) den Aufbau interoperabler, überregional und international kompatibler Strukturen in technischer wie auch in organisatorischer Hinsicht“⁹ fest. Durch diese klare Stellungnahme dieser einflussreichen Förderinstanz wurde ein Umdenkprozess der deutschen Wissenschaftslandschaft angestoßen und die Bereitschaft für die Errichtung von FDZ erhöht.

Bei der Durchsetzung stoßen die jeweiligen Initiatoren jedoch oft auf zahlreiche Hindernisse. Grund hierfür ist die Heterogenität der Wissenschaftslandschaft und der dadurch benötigte lange Koordinations- und Entwicklungsaufwand für Konventionen. Zudem können FDZ mit fortgeschrittener Infrastruktur wegen unterschiedlicher Ausrichtung nur grob Orientierungs- und Anknüpfungspunkte bieten, so dass vielerorts Experimentierbedarf vorherrscht.¹⁰

2.2 Technische Voraussetzungen für Forschungsdatenzentren

Für die Etablierung von FDZ bedarf es der Erfüllung gewisser technischer Ansprüche: Datensicherung erfordert, dass sich Daten speichern lassen. Die produktive Nachnutzung von Forschungsdaten bedingt, „dass digital kodierte Information über Jahrzehnte hinweg authentisch

⁶ Vgl. Wissenschaftsrat (2012), S. 55. Hier spielen auch finanzielle, technische und organisatorische Faktoren eine wichtige Rolle.

⁷ Oßwald (2012), S. 13.

⁸ DFG (1998), S. 12.

⁹ DFG (2006), S. 6.

¹⁰ Vgl. Neuroth (2012).

verfügbar bleibt¹¹. Dieser Kernaspekt findet seine Berücksichtigung in den Bestrebungen, geeignete Maßnahmen in Form von Digitalisierungs- und Langzeitarchivierungsprojekten von Forschungseinrichtungen in die Wege zu leiten. Darin wird, bestenfalls standardisiert¹², die Erhaltung von Datenmengen in digitaler Form auf Servern, in Datenbankstrukturen und Repositorien vorangetrieben. Diese langfristige Verfügbarmachung via Langzeitarchivierungsstrategien wird auch von der DFG immer wieder thematisiert und findet ihren Niederschlag in den „Praxisregeln zur Digitalisierung“¹³. Besondere Erwähnung obliegt einem hierbei unverzichtbaren Element, den persistent Identifiers: Sie bieten die Möglichkeit, Forschungsdaten eindeutig zu identifizieren, zu zitieren und dauerhaft verfügbar zu machen.¹⁴

Persistente Identifizierung ist ebenso grundlegend wie die Frage des Hostings. Je nach Art der Daten handelt es sich teilweise um voluminöse zu speichernde Mengen, die finanzielle und logistische Anstrengungen für FDZ requirieren: Videos erfordern mehr Speicherplatz als ein digitalisierter Fragebogen. Das relativ umfangreiche Datenarchiv der GESIS mit Daten und Dokumenten zu etwa 6000 Studien und ca. 600000 Objekten beläuft sich z.B. nach eigenen Angaben auf einen Speicherbedarf von unter einem Terabyte.¹⁵ Gerade bei aufwendigen medizinischen Screeningverfahren oder AV-Daten ist dieser Umfang schon schnell erreicht.¹⁶

Bei AV-Daten bietet sich optional auch das Verfahren der Bitstream Preservation¹⁷ an, das aber keine dauerhafte Lösung bei der Datensicherung darstellt, weil sich technische Nutzungsumgebungen schnell verändern und damit auch schlecht zu prognostizieren sind.¹⁸

2.3 Forschungsdatenzentren einzelner Wissenschaftsdisziplinen

In einigen Wissenschaftsbereichen, vor allem in der außeruniversitären Forschung, haben sich in Deutschland bereits FDZ etabliert. Zwei Bereiche werden hier kurz vorgestellt.

¹¹ Vgl. Kleiner (2012), S. 9.

¹² Ein wichtiger Standard im Bereich der Langzeitarchivierung ist PREMIS mit der Formulierung des PREMIS Data Dictionary, vgl. PREMIS (2012) und Caplan (2009).

¹³ DFG (2009). 2009 wurde deshalb in Deutschland das Projekt „nestor“ ins Leben gerufen, vgl. Nestor (2012).

¹⁴ Möglich wird dies durch Uniform Resource Identifiers (URI), vgl. W3C (2012), wie beispielsweise Uniform Resource Names (URN), vgl. Persistent Identifier (2012), und Digital Object Identifiers, vgl. DOI (2012). Der Wissenschaftsrat verlangt ebenfalls explizit deren Nutzung, vgl. Wissenschaftsrat (2012), S. 57.

¹⁵ Vgl. Quandt (2012), S. 71.

¹⁶ Allein das Datenvolumen der Pythagoras-Studie beläuft sich auf geschätzte 60 GB, vgl. Bambey (2012a), S. 125. Unkomprimiert erfordert ein Film von 90 Minuten Dauer ca. 1,5 TB, vgl. Hercher (2011), S. 7.

¹⁷ Eine ausführliche Schilderung dieser Praxis unter Neuroth (2009).

¹⁸ Vgl. Oßwald (2012), S. 17.

2.3.1 Forschungsdatenzentren in den Geowissenschaften

Ein anschauliches Beispiel für diese unterschiedliche Ausprägung innerhalb der Wissenschaftslandschaft stellt der Bereich der Geowissenschaften dar.¹⁹ Dort haben sich an größeren Forschungsinstituten Infrastrukturen integrieren lassen, die Austauschformate²⁰ oder anerkannte Normen wie z.B. von der International Organization for Standardization (ISO) übergreifend nutzen.²¹ Beispielhaft soll hier das Helmholtz-Zentrum Potsdam Deutsches GeoForschungszentrum (GFZ)²² genannt werden. Das GFZ stellt für einzelne Forschungsschwerpunkte wie u.a. Seismologie, Geodäsie oder Gravimetrie eigenes und fremdes Datenmaterial in Datenarchiven unter Berücksichtigung genannter Kriterien wie der Nutzung internationaler Standards und der Vergabe von DOI unter Einbeziehung der TIB Hannover zur Verfügung. Dies erfolgt in guter Vernetzung und im internationalen Rahmen über das Information System and Data Center for geoscientific data (ISDC).²³

2.3.2 Forschungsdatenzentren in den Sozialwissenschaften

In den Sozialwissenschaften gab es - im Gegensatz z.B. zu den Geisteswissenschaften²⁴ - stärkere Koordinierungsbemühungen und ein stärkeres Interesse der Forschungscommunity an FDZ.²⁵ Seit 1999 reichte im Zuge des erwähnten Bewusstseinswandels die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) und 2001 der Rat für Sozial- und Wirtschaftsdaten (RatSWD) im Auftrag des Bundesministeriums für Bildung und Forschung (BMBF) Vorschläge zur Verbesserung der Dateninfrastruktur ein mit der Empfehlung des Ausbaus weiterer FDZ und Datenservicezentren (DSZ). Daraufhin entstanden vier vom RatSWD akkreditierte und vom BMBF geförderte

¹⁹ Bei Klump (2012) wird der Bereich der Geowissenschaften überblicksartig dargestellt.

²⁰ Zu nennen ist hier der Standard Directory Interchange Format (DIF) als Maßstab für Datenportale der Geowissenschaften, siehe DIF (2012).

²¹ So beispielsweise ISO 19115:2003 Geographic Information – Metadata, vgl. ISO 19115:2003 (2012).

²² GeoForschungszentrum Potsdam (2012).

²³ ISDC des GFZ Potsdam (2012).

²⁴ Dort werden wegen der breiten Streuung erst seit 2006 tragfähige Initiativen im Forschungsverbund in Form einer virtuellen vernetzten Forschungsumgebung unter dem Namen TextGrid aufzubauen versucht, vgl. TextGrid (2012); eine nähere Beschreibung findet sich auch bei Pempe (2012), S. 144ff.

²⁵ Bereits 1960 wurde das Zentralarchiv für Empirische Sozialforschung für die quantitative empirische Sozialforschung sowie Forschungsdatenzentren für Daten der amtlichen Statistik ins Leben gerufen, vgl. Medjedović (2010), S. 63, sowie Statistische Ämter des Bundes und der Länder (2012).

FDZ sowie zwei DSZ.²⁶ Mittlerweile führt der RatSWD 17 akkreditierte FDZ und drei DSZ auf seiner Webseite auf.²⁷ „Bei aller Heterogenität hinsichtlich der betreuten Daten und Themen ist diesen Institutionen [dabei] gemein, dass sie einen starken Fokus auf Produktion und Anreicherung qualitativ hochwertiger Daten und insbesondere auf eine Verbesserung der Zugänglichkeit solcher Daten setzen.“²⁸ Begünstigt wird die Langzeitverfügbarkeit durch die Schaffung eines zentralen Registrierungsservices durch die GESIS inklusive der Vergabe von DOI für den gesamten sozialwissenschaftlichen Bereich in Deutschland: die da|ra²⁹.

Die Sozialwissenschaften sind somit eine derjenigen Disziplinen, die die gestellten Aufgaben des Datenmanagements und Data Sharings „am stärksten institutionell gebündelt haben. Mit GESIS (...) existier[t] eine eigenständige Institution, zu deren Kernaufgaben es gehört, diese Dienstleistung anzubieten.“³⁰

2.3.3 Das Beispiel des Datenarchivs für Sozialwissenschaften GESIS

Das „Datenarchiv für Sozialwissenschaften“ des Leibniz-Instituts GESIS repräsentiert eine zentrale Bezugsstelle für sozialwissenschaftliche empirische Forschungsprimärdaten.³¹ Verschiedene Schwerpunkte sind dabei zu konstatieren. Zum Einen wird Wert gelegt auf die methodisch-technisch klar definierte Datenakquise, den Datenservice und Data Enrichment nach international anerkannten Standards für quantitative Forschungsdaten³² und die Zugänglichmachung dieser Daten. Auf Grund dieser Spezialisierung können jedoch bisher „jene Subdisziplinen der Sozialwissenschaften weniger gut bedient werden, die sehr stark auch andere Datentypen wie etwa Quellen aus der amtlichen Statistik oder qualitative Daten nutzen.“³³ Viele Daten sind dadurch nur unsystematisch zugänglich, Gesamtbestände schwer abschätzbar und Aussagen über die Langzeitsicherheit solcher Bestände nicht präzise. Dies verhält

²⁶ Vgl. RatSWD (2010).

²⁷ RatSWD (2012).

²⁸ Quandt (2012), S. 67f.

²⁹ GESIS (2012a).

³⁰ Ludwig (2012), S. 297.

³¹ GESIS(2012b).

³² Eine Abgrenzung zwischen quantitativen und qualitativen Daten unter Kapitel 3.3.

³³ Quandt (2012), S. 63.

sich anders bei Umfragedaten, historischen Zeitreihen bzw. Aggregatdaten und biografischen Daten. Dort bestehen Spezialisierungen in Aufbereitung und Bereitstellung.³⁴

Zum Anderen liegt das Augenmerk auf der Einhaltung und Weiterentwicklung entscheidender technischer und organisatorischer Kriterien im Datenaustauschprozess. Die GESIS und damit das Datenarchiv für Sozialwissenschaften richten nicht nur den Fokus auf die Nutzung international anerkannter Standards, sondern wirken aktiv in internationalen Kooperationen an deren Entwicklung mit, wie das Beispiel der Data Documentation Initiative (DDI) auch gegenwärtig zeigt.³⁵ Das bereits erwähnte Registrierungssystem da|ra und die damit verbundene Nutzung persistenter Identifikatoren sind weitere wichtige Bausteine für eine zielgerichtete Langzeitverfügbarkeit von Forschungsdaten samt Sekundäranalysen. Das Engagement durch die Mitwirkung an DDI Work Groups und die Entwicklung von geeigneten Metadatenstandards für die Sozialwissenschaften, sowohl für quantitative als auch bedingt für qualitative Forschungsdaten, zeigt ebenfalls, welchen Weg das Datenarchiv GESIS einschlägt. „So liegen zu jeder Studie detaillierte und standardisierte deskriptive und strukturelle Metadaten vor.“³⁶

Der Entwicklungsprozess ist allerdings teilweise noch im vollen Gange. Trotzdem zeigt die Vorgehensweise innerhalb des Workflows des Archivs für Sozialwissenschaften GESIS, wie die effiziente Nachnutzung, Interoperabilität und Data Exchange unter der Miteinbeziehung von Datenanreicherung und internationaler Standards in die Wege geleitet werden können.

3 Das Portal „Forschungsdaten Bildung“ des DIPF

Die Digitalisierung der Wissenschaft und die Etablierung digitaler Informationsumgebungen sorgen für ein gesteigertes Potential bei der Publikation von Forschungsdaten.³⁷ Diese Chance für die Wissenschaft, verknüpft mit den bereits erwähnten Anleitungen im Forschungsbereich, führte am Deutschen Institut für Internationale Forschung (DIPF)³⁸, einem außeruniversitären Forschungsinstitut und Mitglied der Leibniz-Gemeinschaft (WGL) mit einem Schwerpunkt auf empirischer Bildungsforschung, zur Initiierung und Konzipierung eines

³⁴ Quandt (2012), S. 63ff.

³⁵ DDI (2012a); weitere Kooperationen bestehen im Rahmen des Councils of European Social Science Data Archives (CESSDA) und der International Federation of Data Organisations for the Social Sciences (IFDO).

³⁶ Quandt (2012), S. 68.

³⁷ Vgl. Dallmeier-Tiessen (2012), S. 76.

³⁸ DIPF (2012).

zentralen Angebots für Data Sharing und Datenaufbereitung von Forschungsprimärdaten. Das DIPF besitzt einerseits durch die Unterhaltung einer eigenen dokumentarischen Infrastruktur, dem Informationszentrum Bildung (IZB), als auch auf Grund seiner erziehungswissenschaftlichen Expertise mit umfangreichem Forschungsoutput die notwendigen Basiskapazitäten.

3.1 Zielsetzung und Selbstverständnis als Datenservice- und Forschungsdaten-zentrum

Wie bei anderen Forschungsdatenzentren auch, ist die Ausgangslage im Bereich der Bildungsforschung ähnlich. Die stark interdisziplinär verwurzelten Forschungsansätze inklusive Methodenvielfalt, dezentraler Datenbestände und verteilter Strukturen erschweren die Datenakquise, das Datenmanagement und die institutsübergreifende Zusammenarbeit. Durch Anlaufstellen wie dem Zentrum „Forschungsdaten Bildung“ soll deswegen ein Vereinigungsprozess mit einer Bündelung relevanter erziehungswissenschaftlicher Forschungsdaten gefördert werden.³⁹

Gemäß den Kategorien des Rat SWD⁴⁰ entwickelt das DIPF beim Aufbau der Forschungsdateninfrastruktur eine zweigleisige Struktur. Dies liegt in der heterogenen Datenfülle und gleichzeitig dem beschriebenen Wunsch nach Datenaggregation begründet. Deswegen soll sowohl ein Meta-Informationsportal und die entsprechenden Services eines DSZ als auch ein Datenrepositorium inklusive der datenbezogenen Prozessierungen wie etwa Datenanonymisierung in Form eines FDZ aufgebaut werden.⁴¹ Das Portal, zukünftig integriert in das Fachportal Pädagogik⁴², ist dabei als zentrales Informationsangebot und Service zu Daten und Instrumenten der empirischen Bildungsforschung konzipiert, das im Bereich Forschungsdaten seinen Fokus auf die qualitative Forschung legt. Da Studien umfassend im Gesamtkontext dargestellt werden sollen, werden darüber hinaus auch quantitative Daten mit berücksichtigt.

Das Hauptanliegen bei der Einrichtung des FDZ besteht darin, einen übergreifenden und nutzungsfreundlichen Zugang zu den verteilt vorliegenden Datenbeständen der Bildungsforschung sowie dem entsprechenden Fundus an Erhebungsinstrumenten des Forschungsgebietes

³⁹ Vgl. Bambey (2012a), S. 113ff.

⁴⁰ RatSWD (2010).

⁴¹ Vgl. das interne Arbeitspapier von Bambey (2012b).

⁴² Siehe Anhang: Anlage 1: Rohentwurf der Portalseite Forschungsdaten Bildung; zum Fachportal Pädagogik, siehe Fachportal Pädagogik (2012).

zu entwickeln. Über das integrierte Meta-Informationsportal soll eine zentrale Anlaufstelle für jene Wissenschaftler aus dem Bereich der Bildungsforschung entstehen, die entweder ihre selbsterhobenen Daten im FDZ speichern lassen oder auf bestehende Studien für Sekundäranalysen zugreifen wollen.

Neben diesen Serviceleistungen ist ein wesentlicher Bestandteil zudem die Aufgabe, in ausgewählten Bereichen, vor allem bei qualitativen Forschungsdaten, relevante Datensets und Instrumente der Bildungsforschung zu erschließen, zu hosten und datenschutzkonform über ein zentrales Datenrepositorium zugänglich zu machen.⁴³ Dies impliziert die Berücksichtigung entsprechender internationaler Standards, Richtlinien und Empfehlungen inklusive der Partizipation an Ausarbeitungs- und Abstimmungsprozessen auf Grund des Risikos von „Dopplungen in der Datenhaltung“⁴⁴. Deswegen setzt das Portal auf die Vernetzung mit bereits bestehenden Angeboten aus diesen Forschungsfeldern und aus den verwandten Disziplinen.⁴⁵ Diese Vernetzung soll zu festen Partnerschaften ausgebaut werden, so dass ein gemeinsamer Zugriff auf die Datenbestände möglich wird, jedoch jeder Partner seine exklusive Expertise komplementär einbringen kann. Im Falle des DIPF sind hier die computerbasierte Kompetenzmessung sowie der audiovisuelle Bereich zu nennen.

Auf diesem Weg kann das Ziel in Angriff genommen werden, inhaltlich den Output von Studien und Forschungsprojekten bzw. studienbezogene Materialien zu dokumentieren, langfristig zu archivieren und der Forschung bereitzustellen. Den Kernbestandteil für einen aggregierten und forschungsadäquaten Zugang soll dabei jeweils das Forschungsvorhaben bilden.

3.2 Technische Infrastruktur

Das DIPF bietet durch die Institutionalisierung einer erprobten Repositorien- und Portaltechnik sowie als Technologiedienstleister im Bereich „Technologiebasiertes Assessment“⁴⁶ ein hohes technologisch-methodisches Innovationspotenzial. Deshalb stehen entsprechenden Vor-

⁴³ Vgl. Bambey (2012b).

⁴⁴ Quandt (2012), S. 65.

⁴⁵ Akteure in diesem Feld sind u.a. die GESIS, das Institut zur Qualitätssicherung im Bildungswesen (IQB), das Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID), das Bundesinstitut für Berufsbildung (BiBB) und das FDZ NEPS. Projekte wie NEPS fördern einerseits die Entwicklung von Data Exchange-Verfahren und die Feinabstimmung von Standards, sind aber in ihrer Strategie des Datenmanagements nicht vorrangig auf die Weiterentwicklung von Metadatenstandards ausgerichtet; zum Projekt, siehe NEPS (2012).

⁴⁶ TBA (2012).

leistungen für das geplante „Forschungsdatenzentrum Bildung“ zur Disposition. Diese sind allein deswegen unverzichtbar, weil „die hohen Hardware- und Softwareanforderungen an den Betrieb eines Rechenzentrums für die [Langzeitarchivierung] von Forschungsdaten (...) von kleinen bis mittleren Einrichtungen meist nicht zu leisten“⁴⁷ sind.

Ziel beim Aufbau des FDZ „Forschungsdaten Bildung“ ist die Ausrichtung in Form einer generischen und modularen operativen Datenbankstruktur. Diese soll wiederum die verschiedenen Datenbanken in ein Data Warehouse zusammenführen mit der Möglichkeit, vorhandene organisatorische FDZ-Verbünde durch technologisch höhere Interoperabilität auch funktional stärker zu fundieren.⁴⁸

Hinsichtlich des Hostings werden trotz eigener Serverkapazitäten auch Kooperationen mit entsprechenden Infrastruktureinrichtungen wie der TIB Hannover ausgelotet. Dies liegt daran, dass im Gegensatz zu quantitativen gerade bei qualitativen Daten im AV-Bereich hohe Datenaufkommen zu bewältigen sind. Durch die noch nicht abschätzbare Datenmenge im Zuge weiterer Datenakquisitionen lässt sich der mögliche Datenumfang deswegen nur grob beurteilen.⁴⁹

3.3 Datenbestand

Die Intention der komplementären Ergänzung spiegelt sich in dem beabsichtigten Datenbestand des Portals wider. In Abgrenzung zum IQB und zur GESIS⁵⁰ liegt der Fokus deswegen auf den qualitativen Forschungsdaten, also audiovisuellen Daten wie Videografien oder auf Interviews. Durch diese gewisse Spezialisierung entsteht nach außen ein deutlich umrissenes Profil.

Der qualitative Zweig der sozial- und erziehungswissenschaftlichen Forschung erlebt dabei nach Jahren der Vernachlässigung seit den 1980er Jahren einen Aufschwung von beachtenswertem Ausmaß.⁵¹ „Inzwischen sind qualitative Forschungsmethoden zu einem wichtigen (...) Segment im Spektrum der Forschungsmethoden der Erziehungswissenschaften, aber

⁴⁷ Bambey (2012a), S. 133.

⁴⁸ Vgl. Bambey (2012b).

⁴⁹ Vgl. Bambey (2012b).

⁵⁰ Das IQB hostet quantitative Daten der Schulleistungsstudien; die GESIS nimmt eine führende Rolle ein bez. quantitativer Daten der empirischen Forschung, vgl. Kapitel 2.3.3 oder Quandt (2012), S. 69ff.

⁵¹ Weitere Informationen zur „Renaissance“ qualitativer Methoden bei Krüger (2000) und Garz (2010).

auch der Nachbardisziplinen (...) geworden.⁵² Dies verdeutlicht die Statistik empirischer Projekte mit qualitativer Forschungsmethodik.⁵³ Gerade diese Form von Daten stellt aber einen Bestand dar, der nicht einfach von Forschern aufbereitet werden kann, sondern durch Infrastruktureinrichtungen mit entsprechenden Kapazitäten.

Aber auch im quantitativen Datenbereich wurde am DIPF durch die „Datenbank Schulqualität“⁵⁴ bereits eine in das Gesamtportal integrierte Struktur für den Zugriff auf Forschungsdaten geschaffen, auch weil im Zuge der qualitativen Forschung quantitative Daten wie Fragebögen anfallen. Trotzdem liegt der Fokus aber eindeutig auf qualitativen Beständen, thematisch verortet in der Schul- und Unterrichtsforschung. Denn „qualitative Videostudien haben insbesondere in der Unterrichtsforschung – auch in kleineren Bildungsforschungsprojekten – in den vergangenen Jahren an Bedeutung gewonnen.“⁵⁵

Im Bereich der audiovisuellen Forschungsdaten bestehen im Rahmen von DFG-Förderprojekten erste Infrastrukturen mit verschiedenen erschlossenen, datenbankgestützten Datenbeständen. Darunter fallen - partiell transkribierte - audiovisuelle Daten aus den abgeschlossenen Projekten „Videodatenbank Schulunterricht in der DDR“⁵⁶, Pythagoras⁵⁷ und PERLE⁵⁸. Weitere Daten (u.a. der Studien DESI⁵⁹ und VERA⁶⁰) werden aktuell akquiriert und sollen wie die anderen Datenbestände auch angereichert und in die Datenbankstruktur integriert werden. Im Zentrum steht dabei die jeweilige Studie. Hierzu werden Rahmendaten mit den relevanten Angaben und Erhebungsinstrumenten zur Studie einerseits sowie qualitative Daten wie AV-Daten, Transkripte, Kodierungen, aber auch quantitative Daten, Verlinkungen zu Publikationen, rechtlichen Rahmenbedingungen, Leitfäden und mehr aufbereitet.

Dabei ist Open Access prinzipiell zwar verpflichtend, obliegt aber gerade bei sensiblen Datenbeständen mit persönlichen Daten speziellen rechtlichen Nutzungsbedingungen.⁶¹

⁵² Krüger (2000), S. 323.

⁵³ Vgl. Schulzeck (2008), S. 46.

⁵⁴ DaQS (2012).

⁵⁵ Bambey (2012a), S. 125.

⁵⁶ Das Projekt ist genauer dokumentiert unter: Schluß (2007).

⁵⁷ Über Aspekte der Videografie im Rahmen der Studie Pythagoras informiert u.a. Hugener (2006). Aus der Studie liegen bislang 360 Unterrichtsbeobachtungen, Lehrerinterviews und Aufzeichnungen tutorieller Sitzungen samt Transkripten vor, vgl. Bambey (2012a), S. 125.

⁵⁸ PERLE (2012).

⁵⁹ DIPF (2006).

⁶⁰ VERA (2012).

⁶¹ Vgl. Bambey (2012b).

3.4 Ausstehende Anforderungen

Seit Anfang 2012 hat eine eigene Arbeitsgruppe am DIPF begonnen, die Vielzahl der Requirements für eine solche Datenbankstruktur zu bündeln und zu lösen. Erarbeitet werden im Rahmen einer Dissertation außerdem empirische Bedarfsanalysen aus Nutzer-, also Forscherperspektive, die deren Bereitschaft untersuchen, Forschungsdaten zu Sekundäranalysen zur Verfügung zu stellen, bzw. den Bedarf der Nachnutzung von AV-Daten ermitteln sollen. Resultate stehen aber für den Bereich der Bildungsforschung noch aus.⁶²

Neben den üblichen Anforderungen in den Bereichen Usability und Webdesign sind noch einige weitere Problemstellungen offenkundig. Wesentliche Elemente der Datenaufbereitung müssen erst noch vorgenommen werden wie die Entwicklung von Metadatensets, die Standardauswahl und die Sichtung und Erschließung weiterer Daten, oder abgeschlossen werden wie die Erstellung von Guidelines, Datenschutzbestimmungen und die Datenakquise.⁶³

Hinsichtlich der Entwicklung eines Metadatensets ist der Beitritt des DIPF zur DDI-Alliance am 01.07.2012 zu verkünden samt geplanter Mitwirkung in der „Qualitative Data Exchange Working Group“⁶⁴ zur Entwicklung von DDI-Standards für die qualitative sozialwissenschaftliche Forschung. Allerdings vermeldet diese Zusammenkunft noch keine konkret umsetzbaren Erkenntnisse. Gerade bei den Metadatenstandards ist es allerdings für das Institut „eine kooperative Herausforderung, den vorhandenen Standard DDI in Hinblick auf die Beschreibung qualitativer Datenarten weiterzuentwickeln“⁶⁵.

Insgesamt lässt sich also konstatieren, dass in wichtigen Teilbereichen bei der Organisation und dem Aufbau eines Forschungsdatenzentrums noch Ergebnisse ausstehen.

4 Metadaten und Metadatenstandards

Einen wesentlichen Aspekt bei jeglichem dokumentarischen Vorgehen stellen Metadaten dar.

⁶² Vgl. Bambey (2012a), S. 133.

⁶³ Bildungsforschungsdaten müssen meist erst durch aufwendige Kontaktaufnahme und vertragliche Übereinkünfte übermittelt, digitalisiert und z. T. anonymisiert werden, vgl. Bambey (2012a), S. 117f.

⁶⁴ DDI (2012b).

⁶⁵ Bambey (2012a), S. 134.

Metadaten sind strukturierte „Daten über Daten“⁶⁶. Sie werden entweder automatisch oder mit menschlicher Hilfe generiert und liefern Informationen von unterschiedlicher Art: Sie beschreiben, welche Ressourcen vorliegen, organisieren diese nach kontrollierbaren Kriterien und erlauben die Suche, Aggregation und den Datenaustausch.⁶⁷ Um intern wie extern auf Datenbestände effektiv zugreifen zu können, sind Metadaten elementar und unverzichtbar. Deswegen ist eine Systematisierung im Sinne der Interoperabilität durch die Aufbereitung der Daten in standardisierter Form von entscheidender Bedeutung.⁶⁸

4.1 Ein Kategorisierungsansatz für Metadaten

Für eine bessere Übersichtlichkeit werden immer wieder Kategorisierungsversuche unternommen, die sich im Wesentlichen ähneln und Unterteilungen der Gesamtmasse Metadaten zur Präzisierung vornehmen. Dabei muss allerdings erwähnt werden, dass eine Zuweisung keineswegs exklusiv beschränkt bleibt. Vielmehr sind oftmals die Grenzen durchlässig oder Zuordnungen grob gehalten und vom Standpunkt der dokumentierenden Einrichtung, der Zielgruppe und vor allem dem Dokumenttyp im Kontext abhängig.⁶⁹ Dennoch bieten Kategorisierungen einen gewissen Anhaltspunkt und illustrieren, welche unterschiedlichen Metadatentypen existieren und welche Aspekte berücksichtigt werden können.

Beispielhaft hier eine sinnvolle subjektive Kategorisierung in: technische Metadaten⁷⁰; deskriptive Metadaten wie der Content des Dokuments, auch mit dem wesentlichen Aspekt der Unterstützung von Suchfunktionen und der Navigation⁷¹; identifizierende Metadaten in Form von persistenten und eineindeutigen Identifiern⁷²; administrative Metadaten⁷³ als „encompas-

⁶⁶ Schütz (2004), S. 340.

⁶⁷ Vgl. Zeng (2008), S. 4.

⁶⁸ An Metadaten gibt es gewisse Qualitätsansprüche. Foulonneau und Riley formulieren hierzu „completeness, accuracy, provenance, conformance to expectations, logical consistency and coherence, timeliness, accessibility“ unter besonderer Betonung der Konsistenz, vgl. Foulonneau (2008), S. 54f. und S. 92.

⁶⁹ Schütz formuliert allgemein vier Kategorien: Semantische, strukturelle, administrative und technische Metadaten, vgl. Schütz (2004), S. 340. Hercher rückt dagegen audiovisuelle Datenbestände in den Fokus, vgl. Hercher (2011), S. 11ff. Die NISO nennt deskriptive, strukturelle und administrative Metadaten, vgl. NISO (2004), S. 1.

⁷⁰ Also z.B. Angaben wie Dateiformat, Art des Dokuments, zu nutzender Software beim Zugriff auf die Daten, Auflösung oder Kodierung vgl. Hercher (2011), S. 12.

⁷¹ Zu nennen sind vor allem Produkte der inhaltlichen Erschließung wie Abstract bzw. Annotation, Klassifikationen oder Indexterme, vgl. Schütz (2004), S. 340, aber auch Titel, Autor und Studie.

⁷² Sie ermöglichen die Lokalisierung und klare Zuordnung und sind deswegen unabdingbar für den sicheren Zugriff, vgl. Hercher (2011). S. 11.

sing metadata for technical management of resources, rights management, long-term preservation, processing history, and any other administrative purpose”⁷⁴; Metadaten über die Datenproduktion bzw. das Datenmanagement selbst⁷⁵, und strukturelle Metadaten. Letztere bezeichnen den Zusammenhang mit anderen Objekten und sind gerade unter Vernetzungsaspekten erwähnenswert.⁷⁶

Gerade bei im wissenschaftlichen Kontext entstandenen und vergebenen Metadaten verhält es sich hinsichtlich einer Kategorisierung komplizierter. Hier muss ein größeres Maß an Funktionen bedient werden wie z.B. „provenance, coverage, and data collection methods that are necessary for correct and appropriate scientific use“⁷⁷.

4.2 Grundprinzip Interoperabilität

Spricht man von Metadaten und Metadatenstandards, wird das Desiderat der Interoperabilität immer wieder erwähnt. Dieser Begriff bezeichnet die Möglichkeit, Daten und Metadaten von unterschiedlichen Plattformen und Systemen austauschen, diese Settings abgleichen und dadurch die Verarbeitung und den einheitlichen Zugriff auf die Daten mit so wenig Inhaltsverlusten wie möglich gewährleisten zu können.⁷⁸

Ziel im digitalen Zeitalter dank der dadurch gestiegenen Austausch- und Kooperationsoptionen ist es, Daten ohne großen Informationsverlust zu exportieren bzw. importieren, zu integrieren und zu nutzen. Dazu bedarf es allerdings einer entscheidenden Voraussetzung: Das andere System muss dazu in der Lage sein. Ein perfekt funktionierendes interoperables System ist dabei bisher nicht umgesetzt.⁷⁹ Dennoch lassen sich Konvergenzen, „Crosswalking“ oder „Map-

⁷³ Vor allem auf rechtliche Metadaten wie Zugriffsrechte, Urheberrecht, Nutzungsrecht und Persönlichkeitsrecht, die gerade bei audiovisuellen Daten eminent wichtig sind, vgl. Hercher (2011), S. 11f.

⁷⁴ Foulonneau (2008), S. 24.

⁷⁵ Hier lassen sich u.a. Informationen zu Provenienz, Entstehungs- und Änderungszeiträumen als Datumsangabe oder zu den Datenproduzenten aufnehmen, vgl. Hercher (2011), S. 12.

⁷⁶ Vgl. Hercher (2011), S. 12.

⁷⁷ Zeng (2008), S. 76.

⁷⁸ Bertram nennt dies „Austauschbarkeit, Kommunizierbarkeit und Anschlußfähigkeit“, Bertram (2005), S. 270. Eine andere Definition führt Interoperabilität als “the ability of different types of computers, networks, operating systems, and applications to work together effectively, without prior communication, in order to exchange information in a useful and meaningful manner. There are three aspects of interoperability: semantic, structural and syntactical.” Dublin Core (2005). Die DCMI definiert vier Stufen der Interoperabilität, vgl. Dublin Core (2009).

⁷⁹ Vgl. Foulonneau (2008), S. 123.

ping⁸⁰, mittlerweile umsetzen und die Interoperabilität steigern. Unter Hinzunahme von Auszeichnungssprachen (Markup Languages) wie XML, Protokollen wie dem TCP/IP-Protocol⁸¹ oder Mapping Tools⁸² ist es mittlerweile möglich, Datenaustausch zu betreiben.

Von entscheidender Bedeutung zeigen sich dabei auch Metadatenstandards. Je internationaler und ausgewogener sie konzipiert werden, desto kompatibler können Daten verwendet und interpretiert werden. Dies trifft auch auf Dateiformate zu. Hier muss, um Interoperabilität erhöhen und Crosswalking leisten zu können, ein essentieller „Harmonisierungsprozess“⁸³ stattfinden, um den Datenaustausch im jeweiligen Fachgebiet massiv zu vereinfachen.

4.3 XML, XML-Schema und Application Profile

Um diese Form der Interoperabilität für Webinhalte und digitale Dokumente prinzipiell zu ermöglichen, bedurfte es eines Standards für Datenaustausch in Form einer Markup Language, der übergreifend gilt, offen zugänglich ist und leicht anwendbar ist. Vorher war ein Metadaten austausch auf Grund der heterogenen Standards nahezu unmöglich.⁸⁴ Die Entwicklung von XML⁸⁵ stellt dabei einen entscheidenden Schritt zur Behebung dieses Zustandes dar.

XML enthält wesentliche Vorzüge. Es offeriert ein hohes Maß an Flexibilität, eine klare Strukturierung, vereinfacht die maschinen- und menschenlesbare Abbildung von Metadaten, ist, wie der Name bereits aussagt, erweiterbar, lässt also individuellen Gestaltungsspielraum bei der Festsetzung der gewünschten Granularität, und bietet unendliche Variationsmöglichkeiten: Es lassen sich beliebig viele Schemas, Elementtypen, Attribute und andere Standards einbinden.⁸⁶ Gleichzeitig verlangt XML die Vergabe von festen „Namespaces“ für alle Ele-

⁸⁰ Beides bezeichnet den Konvertierungsprozess eines Metadatensatzes eines bestimmten Formats in einen anderen Datensatz, als Definition: “A crosswalk is a specification for mapping one metadata standard to another. Crosswalks provide the ability to make the contents of elements defined in one metadata standard available to communities using related metadata standards.” NISO (1998).

⁸¹ Vgl. Foulonneau (2008), S. 119.

⁸² Vgl. Foulonneau (2008), S. 161f.

⁸³ NISO (1998).

⁸⁴ “Old rule: If you find a language element you don't understand, ignore it.” W3C (1998).

⁸⁵ XML, eXtensible Markup Language, entstand unter Obhut des W3C, wird permanent weiterentwickelt und bildet die Basis vieler Auszeichnungssprache; zur fünften Edition, vgl. W3C (2008).

⁸⁶ Vgl. W3C (2008).

mente, Attribute und Schemas. Diese „Namespace functionality“⁸⁷ spielt eine entscheidende Rolle für die klare Zuordnung.

Zur Wahrung der Transparenz bedarf es allerdings klarer Definitionen über die festgelegten Namespaces, die Granularität, Elementtypen, Subsets, Attribute und Werte (Values) und deren Verwendung und Einbindung. Dies erfolgt zu Beginn des XML-Dokuments über eine Document Type Definition (DTD)⁸⁸ respektive XML-Schema.

Ein XML-Schema ist ein XML-Mechanismus, der die logische Struktur innerhalb des XML-Dokuments beschreibt. Es definiert u.a. Elemente samt Attribute eines Dokuments, Datentypen, den Syntax, Hierarchien, Anordnungen und Strukturen innerhalb eines Dokuments (in Form von XML-Schema-Instanzen), sowie einzutragende Werte (Values wie z.B. Zahlen, Datumsnormen, Texttypen).⁸⁹ Es stellt praktisch „a grammar for a class of documents“⁹⁰ dar, assoziiert dabei „Attribute name-value pairs“, also die Kopplung eines Attributnamens an einen bestimmten Wert, und weist auf externe Subsets hin. Insgesamt bietet ein XML-Schema neben dem klar definierten Vokabular das Gerüst für unendliche Kombinationsmöglichkeiten.⁹¹ Dabei gilt es immer auf die Erfüllung des Kriteriums der „well-formedness constraint“ und „validity constraint“⁹² zu achten. Erleichtert und um semantische Präzisierung erweitert werden kann die Anwendung durch die menschenlesbare Darstellung über ein Resource Description Framework (RDF)⁹³.

Ein XML-Schema bietet also die Option einer Art Baukasten mit vielen Bestandteilen und beigefügten Bedienungsanleitungen. Gerade der letzte Aspekt sorgt in der Theorie für bessere Interoperabilität. Gleichzeitig bietet dies dem Anwender auch die Chance, seine eigenen Requirements bezüglich der Ressourcenbeschreibung umzusetzen, Granularität nach eigenem Ermessen zu definieren, Standards frei nach Bedarf zu verwenden und dennoch Metadaten für andere Nutzer zu konservieren und für das Data Sharing zur Verfügung zu stellen. „Je nach Anwendungsbereich muss [dabei] der Grat der notwendigen Strukturierung sorgfältig bedacht

⁸⁷ W3C (1998). Jeder Namespace hat einen Identifier (ID) in Form eines URI, vgl. Zeng (2008), S. 136.

⁸⁸ Vgl. W3C (1998).

⁸⁹ Vgl. Hercher (2011), S. 17.

⁹⁰ W3C (2008).

⁹¹ Vgl. W3C (1998).

⁹² Zur Definition der beiden Prinzipien hier nur der Verweis auf weitere Informationen unter W3C (2008).

⁹³ Eine Behandlung der Funktionsweise mitsamt Vor- und Nachteilen wäre Thema für eine eigene Arbeit und soll im Rahmen dieser Arbeit nicht erfolgen, weswegen an dieser Stelle nur auf die entsprechende Website des W3C verwiesen werden soll, siehe W3C (2004).

werden, da eine hohe Strukturierung zu höheren Aufwänden bei der Erzeugung und Verwaltung der Metadaten führt.⁹⁴

Das daraus resultierende entscheidende Spezifizierungsmerkmal wird dadurch illustriert, dass Datenanbieter mit interoperabler Zielsetzung sich eigene Anwendungsprofile, sog. Application Profiles, zulegen. Heery und Patel definieren Application Profiles als "schemas which consist of data elements drawn from one or more namespaces, combined together by implementors, and optimised for a particular local application"⁹⁵. Durch ein solches Application Profile können die individuellen Ansprüche umgesetzt und in einen größeren Rahmen transportiert werden. Hunderte heterogener Standards existieren nebeneinander, unterscheiden sich z.B. nach Dokumentationsinteresse, Fachdisziplin und archivierten Objekten. Im Application Profile und dank XML kann der Datenanbieter selbst definieren, welche Strukturierung und Standardisierung er bis zu welchem Grad wählt und welche Namespaces er verwendet.⁹⁶ Ein Application Profile definiert also den fachspezifischen Einsatz der bei Ressourcenbeschreibung verwendeten eigenen Bausteine, des Metadatenelementsets, und dokumentiert gleichzeitig die empfohlene Kombination mit anderen Metadatenstandards. Durch die DTD sowie Ausgestaltung und Nutzung der Schemas lässt sich dieses Profil übersichtlich gestalten und verständlich machen. Dabei muss sich der Implementierende allerdings genau Gedanken machen über die Anforderungen des eigenen Bedarfs.⁹⁷ Es ist hierbei empfehlenswert, sich an bestehenden Element Sets zu orientieren und Basisstrukturen zu übernehmen.⁹⁸

Unabhängig von der Vorgehensweise und Auswahl der Bestandteile eines Application Profiles ist eine ausreichende Dokumentation, bestenfalls mit Best-Practice-Beispielen, eigener Policy und Guidelines notwendig für sinnvolles Data Interchange notwendig.⁹⁹

4.4 Metadateninitiativen und Standards zur Datenstrukturierung

Werden einzelne Metadatenelemente zu einem Metadatenset zusammengefasst und soll deren Anlage in einer abgestimmten Art und Weise erfolgen unter Darstellung der hierarchischen

⁹⁴ Hercher (2011), S. 12.

⁹⁵ Heery (2000). Foulonneau und Riley sehen Application Profiles als "methods that allow metadata creators to combine elements from multiple formats as needed." Foulonneau (2008), S. 27.

⁹⁶ Vgl. Zeng (2008), S. 116.

⁹⁷ Hercher et al. formulieren einige Anhaltspunkte, vgl. Hercher (2011), S. 15f.

⁹⁸ Vgl. Zeng (2008), S. 101.

⁹⁹ Über den Prozess der Entwicklung von Metadata Usage Guidelines mehr bei Foulonneau (2008), S. 29-42.

Relationen samt Aufbau, spricht man von der Entwicklung eines Metadatenstandards für Datenstrukturen. Die Formulierung dieser Standards fördert die Angleichung verschiedener Datenmengen hinsichtlich des Datentransports, der Interoperabilität, Data Sharing und Vereinheitlichung der Metadateninhalte durch klare Definitionen und Zuordnung sowie die internationale Ausrichtung durch die Schaffung gemeinsamer Bezugspunkten. Dabei ist anzumerken, dass jeder Metadatenstandard in gewisser Weise nur ein Kompromiss sein und so gut wie nie alle heterogenen Interessen der Nutzer sowie Ansprüche der Datenquellen abdecken kann.¹⁰⁰

Metadatenstandards legen in ihrem Wesen sowohl Form als auch Vokabular der Metadaten bei der Ressourcenbeschreibung fest und definieren damit die allgemein akzeptierten Beschreibungsstrukturen für den jeweiligen Datenaustausch.¹⁰¹ Um der Vielfalt heterogener Metadatenstandards¹⁰² Herr zu werden, gründen sich immer wieder meist international ausgerichtete Initiativen mit der Intention, einheitliche Standards für die Ressourcenbeschreibung zusammenzuführen oder neu zu kreieren, wie DCMI und DDI.

4.4.1 Dublin Core

Im Zuge des Bedarfs nach der Erschließung von Webressourcen einigten sich 1995 die Teilnehmer einer Fachtagung, die spätere Dublin Core Metadata Initiative (DCMI), auf eine Grundmenge von beschreibenden Termen für die Kategorisierung von Webressourcen und nannten diese Metadatenmenge nach dem Tagungsort Dublin Core Metadata¹⁰³. Dies zog schnell internationale Resonanz nach sich und führten zu einem Übereinkommen über diese Kernmenge von Metadaten als internationaler deskriptiver Metadatenstandard. Kaum ein Standard wurde in der Folge sooft übersetzt. DC kann dabei als „resource discovery lingua franca for metadata“¹⁰⁴ bezeichnet werden mit dem primären Ziel der Erzeugung und Erhaltung von Metadaten. Der Standard zielt auf Einfachheit, Allgemeinverständlichkeit, internationale Ausrichtung, Erweiterbarkeit und Interoperabilität ab und ist übergreifend intendiert.¹⁰⁵

¹⁰⁰ Vgl. Foulonneau (2008), S. 19.

¹⁰¹ Vgl. Hercher (2011), S. 17.

¹⁰² Krause (2004) sieht als Leitlinie zur Konsistenzerhaltung und Interoperabilität, Standardisierung von seiner Heterogenität her zu denken.

¹⁰³ Dublin Core (2012a). Das DC Metadata Element Set (DCMES) wird auch als internationaler Standard der ISO anerkannt, vgl. ISO 15836:2009 (2012).

¹⁰⁴ OAI-PMH (2005).

¹⁰⁵ Vgl. den Abschnitt „Goals“ bei Hillmann (2005).

Die Grundform des Standards, der „Simple Core“, besteht in seiner Strukturierung aus einem Element Set von 15 Properties¹⁰⁶. Bei diesen 15 Elementen werden Bezeichnungen und klare Definitionen in Kategorienform festgesetzt. Bewusst wurde dieser Simple Core anfangs unter einer geringen Ausprägung an Granularität strukturiert, einerseits um dem hohen Grad an Pluralität von Webressourcen Rechnung zu tragen, andererseits, um Übersichtlichkeit, Nachvollziehbarkeit und Anwendungsfreundlichkeit für Datenproduzenten und Datennutzer zu kreieren.¹⁰⁷ Somit entstand eine Basis für die übergreifende internationale Nutzung und Darstellung von Metadaten, die vereinfachtes Mapping fördert.

Dublin Core beinhaltet wesentliche Grundprinzipien¹⁰⁸: Zum einen zeichnet sich DC durch Optionalität und Einfachheit aus. Die 15 Kernelemente sind nämlich nicht verpflichtend. Vielmehr bestand der Wunsch, lieber wenige sortierte Daten in festen Kategorien verfügbar zu machen als gar keine. Somit ist jedes Element optional verwendbar. Auch sind einzelne Elemente in beliebiger Wiederholung abbildbar. So lassen sich zum Beispiel unter dem Element `dc_title` mehrere Titel untereinander auflisten für dasselbe Dokument. Die Reihenfolge spielt dabei im Grundkonzept von Dublin Core keine zu berücksichtigende Rolle.¹⁰⁹

Ein weiteres entscheidendes Charakteristikum ist das Eins-zu-Eins-Prinzip. Es legt fest, dass mit einem einzelnen Metadatensatz nur ein Objekt oder eine Quelle beschrieben werden darf.¹¹⁰ Der Vorteil hiervon ist die klare Zuordenbarkeit.¹¹¹

Außerdem bietet DC die Möglichkeit, „Refinements“, also feinere Untergliederungen eines Elements in hierarchischer Struktur zur engeren Spezifizierung, einzubauen. Hierbei wird das sog. „Dumb-down-Prinzip“ angewandt. Dieser Anweisung entsprechend, wird ein Refinement immer auf das Kernelement zurückgeführt – eine Art Rückfallmechanismus, auch wenn dadurch mitunter Spezifizierungen verloren gehen können.

¹⁰⁶ Eine Auflistung der 15 Elemente mit weiteren Informationen erfolgt unter Dublin Core (2012b).

¹⁰⁷ Vgl. MetaGuide SUB Göttingen (2001).

¹⁰⁸ Zu den Grundprinzipien, vgl. Hillmann (2005).

¹⁰⁹ “The ordering of multiple occurrences of the same element (e.g., Creator) may have a significance intended by the provider, but ordering is not guaranteed to be preserved in every user environment.” Hillmann (2005).

¹¹⁰ “Many metadata formats provide strict guidance for separating the description of one version of a resource from another. In the Dublin Core community, this notion is referred to as the one to one principle. The one to one principle, simply put, prescribes that a single entity is described in a single metadata record.” Foulonneau (2008), S. 158.

¹¹¹ “In general Dublin Core metadata describes one manifestation or version of a resource, rather than assuming that manifestations stand in for one another.” Hillmann (2005).

Zusätzlich gilt die Maßgabe, „appropriate values“ unter dem „requirement of usefulness for discovery“¹¹² zu verwenden, also auf kontrolliertes Vokabular zurückzugreifen, das konsistent genutzt und eindeutig, unter der nötigen Vorsicht, definiert werden soll.

Die beiden letzten Grundprinzipien, kontrolliertes Vokabular und das Dumb-Down-Prinzip mit der Verwendung von Refinements, spielen beim Verständnis des zweiten Levels des Dublin Cores, dem sog. „Qualified Core“ eine wesentliche Rolle. Im Unterschied zum DCMES, dem Simple Core, wird dieser Bereich DC Metadata Terms¹¹³ genannt. Diese erweitern den Simple Core um weitere Elemente wie Audience, Provenance und Rightsholder und werden zusätzlich um eine Reihe von „Qualifiers“, festgelegten Refinements und sog. „Encoding Schemes“, ergänzt, die die Wiederverwendung erleichtern. Beide Bestandteile können aber nach dem Konzept der DCMI auch frei von Anwendern entwickelt und in eigenen Application Profiles hinzugefügt werden im Sinne des Kriteriums der individuellen Erweiterbarkeit.¹¹⁴

Neben den Refinements können durch die Nutzung von XML und illustrierend via RDF auch Schemes eingebunden werden.¹¹⁵ Gerade das genutzte Vokabular mit den eingetragenen Values und die Abbildung der Syntax müssen aber für Datenmanagement und –nutzung transparent und nachvollziehbar sein. Deswegen führt der Qualified Core die Encoding Schemes. Diese Qualifiers „identify schemes that aid in the interpretation of of an element value“ und beinhalten “controlled vocabularies and formal notations or parsing rules”¹¹⁶. Diese klare Bestimmung und Kennzeichnung der verwendeten Values und der Grundstruktur ermöglicht erst den sinnvollen Datentransfer, erleichtert die Orientierung für auf die Metadaten zugreifenden Nutzer und bietet trotzdem individuellen Gestaltungsfreiraum bei der Vergabe der Metadaten.

All diese Charakteristika und Grundprinzipien führen zu den gewünschten Zielen: Zu einer einfachen Anwendbarkeit und der Option auf eigene fachspezifische Application Profiles inklusive interoperabler, übergreifender Verwendungschancen der Metadaten. Einzelne Taskgruppen versuchen auch diesen Schritt vorzudefinieren mit der Entwicklung von spezifischen

¹¹² Hillmann (2005).

¹¹³ DCMI (2012).

¹¹⁴ Vgl. Hillmann (2005). Kapitel 5: Dublin Core Qualifiers.

¹¹⁵ Außerdem existieren Syntax Guidelines und ein anschauliches Informationsmodell, frei von Encoding Schemes, das DCMI Abstract Model, welches die Rahmenbedingungen des Dublin-Core-Standards samt Profilen definiert und Kombinationsmöglichkeiten aufzeigt, vgl. DCMI (2007) und Hillmann (2005). Abschnitt Syntax Issues. Eine deutsche Übersetzung liegt vor unter KIM (2008).

¹¹⁶ Hillmann (2005). Kapitel 5: Dublin Core Qualifiers.

Application Profiles für einzelne Bereiche. Auch für den Bildungssektor gibt es Ansätze eines eigenen Application Profiles, die allerdings nicht sehr weit reichend sind.¹¹⁷

Insgesamt ist es mit Dublin Core gelungen, einen allgemein angelegten deskriptiven Metadatenstandard mit enormer Resonanz und Rezeption zu entwickeln, oder, wie Zeng und Qin anmerken:

“The DC is essential to understanding all metadata standards because of its pivotal role in standards development historically, as well as in its role as the standard of integration for divergent metadata standards. Almost all metadata standards and application profiles have also created crosswalks in order to map their elements to DC elements.”¹¹⁸

4.4.2 Data Documentation Initiative (DDI)

Die DDI ist ein institutionsübergreifendes Bestreben mit der Zielsetzung, allgemein international gültige Metadatenstandards für den gesamten sozialwissenschaftlichen Bereich zu entwickeln. Ziel der Initiative ist es, Konkretisierungen für den sozialwissenschaftlichen Bedarf auszuarbeiten und an die Anforderungen im Bereich der Datendokumentation und Data Interchange anzupassen. Dadurch sollen Prozesse des gesamten Data Lifecycles in Form von Spezifikationen unterstützt werden.¹¹⁹ Bisher liegt der DDI-Schwerpunkt auf der quantitativen Forschung und entsprechenden Methoden. Für den Bereich der qualitativen Forschung wurde im Januar 2010 eine Expertengruppe instituiert,¹²⁰ die allerdings bisher auf noch keine umsetzbaren Ergebnisse verweisen kann.¹²¹

Die 1995 gegründete Initiative stellt also einen Versuch führender internationaler Datenarchive und –produzenten auf dem Gebiet der Sozialwissenschaften¹²² dar, für ein Data Lifecycle Management Schemas und Application Profiles unter der Definition von kontrolliertem Vo-

¹¹⁷ Vgl. Dublin Core (2012c) und Zeng (2008), S. 42-52.

¹¹⁸ Zeng (2008), S. 22.

¹¹⁹ Vgl. DDI (2012c).

¹²⁰ Vgl. DDI (2012b). Der letzte verfügbare Statusbericht erfolgte im Rahmen des europäischen DDI-Meetings in Utrecht, einzusehen in Form einer Präsentation, vgl. Corti (2010).

¹²¹ Es gibt seit 2004 außerdem auf Initiative der Bundesländer Bestrebungen, für den Bereich der Bildungsforschung ein eher quantitativ orientiertes EduDDI Application Profile als kompatible Untermenge zu gestalten, vgl. Mechtel (2009). Ziel ist dabei die Entwicklung eines Metadatenvokabulars für die speziellen Bedürfnisse der Bildungsforschung. Derzeit befindet sich EduDDI allerdings wegen der Heterogenität bei der Metadatenauszeichnung noch in der Entwicklungsphase, vgl. auch Bambey (2012a), S. 129.

¹²² Momentan werden 36 Mitglieder verzeichnet, u.a. die GESIS und das DIPF, vgl. DDI (2012d).

abular¹²³ als Orientierungshilfe für den Aufbau eines eigenen Schemas auszuarbeiten. Dabei wird auf XML wegen der oben angeführten Vorteile zurückgegriffen. Der DDI-Standard wird momentan noch als XML-Dokument Type Definition (DTD) formuliert.¹²⁴

Von 2009 stammt die aktuelle Version des DDI Data Lifecycles: DDI 3.1. Darin stellt die DDI eigene Schemas mit Codebook, momentan 2.5¹²⁵, vor und definiert deren Inhalte, Elemente, Packages, Relationen und Attribute.¹²⁶ Diese Version ist zwar noch nicht perfektioniert, bildet aber einen Grundstock. Die Spezifikation ist dabei in dieser Fassung erstmals auf den kompletten Lebenszyklus von Daten innerhalb einer Studie – von Design und Planung über Datenerhebung, -aufbereitung und -analyse bis hin zur Archivierung – ausgerichtet.¹²⁷

DDI erlaubt die komplexe Abbildung von einzelnen Relationen und Elementen mitsamt Attributen in Form von fest formulierten Modulen und Schemas. Dadurch entstehen Grundstrukturen, die zum einen der Komparabilität dienen, zum anderen die Versionierung und den Reuse erleichtern sowie ein hohes Spezifizierungspotential und Maß an Granularität beinhalten. Durch diese Modularisierung können Metadaten genauer beschrieben sowie besser gepflegt und aufbereitet werden.¹²⁸ Das Design sieht dabei vor, Basismodule für alle Anwendungsgebiete und Module für besondere Datentypen oder Speicherformate zu kreieren.¹²⁹ Es wird dabei dem Nutzer empfohlen, ein eigenes spezifiziertes DDI-Profil anzulegen und darin die benötigten Module zu verwenden. Die Modifizierung und Anpassung wird dadurch erleichtert, dass die Elemente und Attribute meist optional und wiederholbar sind und je nach Bedarf der eigenen Institution beliebig oft eingesetzt werden können. Außerdem trägt dazu die Möglichkeit bei, Subsets von DDI-Tags anstelle der gesamten DDI-Spezifikation zu bilden.¹³⁰

Wie bereits erwähnt wird ein XML-Schema angelegt. Innerhalb dieses XML-Schemas kommen weitere Schemas, überwiegend eigens definierte DDI-Schemas, zum Einsatz, die eine

¹²³ Feste Namespaces werden definiert und kontrolliertes Vokabular empfohlen. Hierzu der Praxisbericht DDI Best Practice (2009a).

¹²⁴ Vgl. Medjedović (2010), S. 132.

¹²⁵ Vgl. DDI (2012e). Das Codebook befindet sich aktuell noch in der Reviewphase.

¹²⁶ DDI 3 setzt sich zusammen aus 22 DDI-Schemas, Dublin-Core-Sets und XHTML-Schemas, 14 weiteren Schemas und einer extensiven Anzahl an Klassen. Die Schemas beinhalten wiederum eine optional einsetzbare und individuell ergänzbare Anzahl von Elementen und Attributen. Eine genaue Übersicht unter DDI (2009a). DDI 3.1 Schema and Documentation und Technical Specification. DDI 3.1. Part 1: Overview. Zur Illustration stehen auf der Homepage der DDI etliche Best-Practice-Beispiele zur Verfügung, vgl. DDI (2012f).

¹²⁷ Vgl. Quandt (2012), S. 75.

¹²⁸ Vgl. DDI (2009a). Technical Specification. DDI 3.1. Part 1: Overview.

¹²⁹ Die in DDI 3.1 verwendeten Basismodulformen samt XML-Schema sind unter DDI (2011) einsehbar.

¹³⁰ Für die Anlegung eines eigenen DDI-Profiles samt Subsets, vgl. DDI Best Practice (2009b).

weitere Präzision ermöglichen.¹³¹ Zudem steigen durch die Modularisierung und Schematisierung der Flexibilitäts- und Interoperabilitätsgrad, auch durch die Miteinbeziehung von Dublin Core: Im aktuellen Codebook sind bewusst Dublin Core-Terme und ein Dublin Core Schema eingebunden, die ein Mapping ermöglichen.¹³²

Weitere Wesensmerkmale sind der hierarchische Aufbau mit einer festen Vererbung aller Attribute und Relationen von Parent-Elementen zu Child-Elementen, die empfohlene Identifizierung durch Identifiers¹³³, die Vergabe fester Namespaces, also die Nutzung eines kontrollierten Vokabulars, ein CodingScheme mit der Option, Derivationsprozesse zu beschreiben, die Nutzung sog. „Major Reusable Classes“¹³⁴ und die Option auf Versionierung¹³⁵. In DDI besteht zudem die Möglichkeit, Beziehungen und Referenzen detailliert abzubilden, durch die „Inclusion by Reference“. Außerdem existiert eine Vielzahl bereitliegender Tools, u.a. DDI-Editoren.¹³⁶

DDI befriedigt also je nach Bedarf und Wunsch nach Datenstrukturierung durch die Komplexität und den großen Umfang an Types, Schemas, Modulen, Subsets und Elementen das Bedürfnis, quantitative Forschungsmetadaten sinnvoll interoperabel aufzubereiten und im sozialwissenschaftlichen Bereich Data Sharing zu betreiben.

4.4.3 Die Situation im audiovisuellen Bereich

Angesichts der Tatsache, dass AV-Forschungsdaten im Fokus der Betrachtung stehen, soll auch eine Beurteilung der Lage hinsichtlich passender Metadatenstandardformate erfolgen.¹³⁷

¹³¹ So beispielsweise für die Bereiche Archive, Conceptual Components, Data Collection, Logical Product und Physical Data Product, vgl. DDI (2009a). Technical Specification. DDI 3.1. Part IV: XML Schema files and field-level documentation. Unter DDI Best Practice (2009c) steht auch ein Anwendungsbeispiel zur Verfügung hinsichtlich der Einbeziehung eines vollwertigen DDI-Schemas.

¹³² Die Initiative stellt eine Anleitung für ein Mapping auf Dublin Core unter DDI (2009b) zur Verfügung.

¹³³ Vgl. DDI Best Practice (2009d). Die Betonung liegt dabei auf der „Uniqueness“ und der Nutzung von URNs sowie eindeutiger Attribute beim Entwurf eines Application Profiles.

¹³⁴ Major Reusable Classes beinhalten eine Anzahl komplexer Klassen und werden durchweg durch das DDI-Schema-Set benutzt. Vgl. DDI (2009a). Technical Specification. DDI 3.1. Part 1: Overview.

¹³⁵ DDI 3.1 vereinfacht die Darstellung von Versionen. Eine neue Version wird durch die jeweilige Maintenance Agency angelegt, sobald Änderungen an der Ursprungsversion vorgenommen werden, vgl. DDI (2009a). Technical Specification. DDI 3.1. Part 1: Overview, sowie den Praxisbericht DDI Best Practice (2009e).

¹³⁶ Vgl. DDI (2012g).

¹³⁷ Auch hier lässt sich ein hohes Maß an Heterogenität bei der Organisation und Verwaltung der Datenbanken innerhalb der Film sammelnden und archivierenden Einrichtungen in Europa und Deutschland konstatieren, vgl. Hercher (2011), S. 10. Ein Überblick über aktuelle AV-Standards ist bei Hercher zu finden, vgl. ebd., S. 18-22.

Von den entwickelten Standards sind die meisten bisher nicht für die Forschung geeignet, sondern erfüllen eher Kriterien der Filmindustrie. Sie sind inhaltlich oftmals auf wissenschaftlich irrelevante Aspekte wie z.B. Schauspieler oder Regisseur ausgerichtet. Dennoch bieten audiovisuelle Standards für den technischen Bereich Anhaltspunkte, welche beispielsweise von DC oder DDI in dem Maße noch nicht umgesetzt sind. Sie stellen deswegen Orientierungshilfen dar und geben Hinweise zu Metadaten wie Format, Länge, Codecs oder Software.

Insgesamt besteht also durch diese Ausgangslage im wissenschaftlichen Bereich noch Handlungsbedarf, auch auf Grund des zu erwartenden hohen Datenvolumens und der besonderen Anforderungen an das Datenmanagement durch die notwendige Berücksichtigung einer Vielzahl von Codecs sowie Formatvalidierungen und –konvertierungen.¹³⁸ Außerdem existieren weitere diffizil zu lösende, aus den unterschiedlichen Objekttypen sowie den bereits geschilderten individuellen Ansprüchen der einzelnen Fachwissenschaften resultierende Herausforderungen.¹³⁹

Verwiesen werden soll kurz in diesem Zusammenhang auf den Standard MPEG-7 (ISO/IEC 15938)¹⁴⁰ der Motion Pictures Expert Group (MPEG). MPEG-7 ist ein Standard zur Beschreibung und Suche von AV-Inhalten und deckt zugleich deskriptive und administrative Metadaten ab inklusive der sog. „Low-Level-Features“ wie Kameraeinstellungen.¹⁴¹ MPEG-7 enthält und definiert über 450 Metadaten, ist somit der umfangreichste Metadatenstandard für AV-Daten. Dabei etablierte die MPEG Schnittstellen und Mappingmöglichkeiten zu anderen relevanten Standards wie Dublin Core. Ein wesentlicher Bestandteil sind zudem Deskriptoren für auditive und visuelle Charakteristika, „Beschreibungsschemata für Multimediateil und Spezifikationen und Kodierungsvorschriften für den Transport von MPEG7 Metadaten“.¹⁴²

Der Standard unterteilt sich in drei Hauptelemente: Descriptors, wie z.B. Color, Audio, Video, zur Beschreibung der spezifischen Merkmale der zu beschreibenden Multimediateil, Description Schemes, also vordefinierte Descriptorstrukturen, die zueinander in Beziehung gesetzt werden, sowie die Description Definition Language (DDL).¹⁴³ Merkmal ist zudem die

¹³⁸ Vgl. Pempe (2012), S. 143.

¹³⁹ Vgl. Zeng (2008), S. 76.

¹⁴⁰ MPEG-7 (2004).

¹⁴¹ Vgl. Foulonneau (2008), S. 26.

¹⁴² Hercher (2011), S. 20.

¹⁴³ Eine genauere Beschreibung findet sich bei Zeng (2008), S. 77ff, und unter MPEG-7 (2004).

starke hierarchische Struktur mitsamt zahlreichen Abhängigkeiten zwischen den einzelnen Elementen mit Nachteilen für die Handhabung.¹⁴⁴

MPEG-7 ist also dank des gefächerten Spektrums durchaus einen Bezugspunkt für die Entwicklung von AV-Metadatensets, vor allem bei der Abbildung technischer Komponenten.

5 Der Entwicklungsprozess eines Metadatensets für AV-Daten am DIPF

Nach der theoretischen Abwicklung folgt der praxisorientierte Teil mit eindeutigem Bezug auf das FDZ „Forschungsdaten Bildung“ mit dem Ziel, einen ersten Versuch eines Metadatensets für AV-Daten zu unternehmen. Ein Metadatenelementset besteht dabei immer aus zwei Komponenten: Den „Semantics“, also der Definition der Bedeutung der Elemente und ihrer Relationships, sowie dem „Content“, also der Art und Weise, wie etwas wo dokumentiert werden soll.¹⁴⁵ Dabei ist anzumerken, dass auf Grund des Status Quo am Institut lediglich ein erster Ansatz mit der Entwicklung von Grundideen erfolgen kann. Zu bedenken ist dabei immer, dass die einzelnen Datenbestände am Institut heterogen aufbereitet werden.¹⁴⁶

5.1 Bezugspunkte für die Auswahl der Metadatenelemente

Wie bereits erwähnt, beteiligt sich das DIPF seit Kurzem im Zuge des geplanten FDZ an der Data Documentation Initiative und hat sich dadurch seinen primären Bezugspunkt gewählt. Allerdings klang bereits an, dass sich DDI primär und in der Momentaufnahme noch auf quantitative Daten bezieht und nur grob die Anforderungen qualitativer Forschungsdaten, speziell im AV-Bereich, tangiert. Wegen noch ausbleibender umsetzbarer Ergebnisse aus der Arbeitsgruppe für qualitative Daten und des in der Ausarbeitung befindlichen EduDDI-Entwicklungsprozesses kann DDI - noch - nur bedingt, aber in Anwendung der charakteristischen Merkmale, verwendet werden, inklusive des in DDI implementierten Dublin Core.¹⁴⁷

¹⁴⁴ Hercher (2011), S. 19f.

¹⁴⁵ Vgl. Zeng (2008), S. 93.

¹⁴⁶ Die Pythagoras-Studie wird z.B. als Ganzes dokumentiert. Filme haben zwar eine eigene ID, aber keine eigene inhaltliche Erschließung wie beim Projekt „Videodatenbank Schulunterricht in der DDR“. Andere Studien liefern kaum Erkenntnisse auf Grund noch ausstehender Sichtung, vgl. Bambey (2012b).

¹⁴⁷ Siehe die Einbindung der Dublin Core Terms im Schema „dcterms.xsd“, vgl. Schema DC-Terms (2003).

Der Mangel an umsetzbaren Orientierungspunkten für den konkreten Fall von AV-Forschungsdaten für den Bildungsbereich schlägt hier ins Gewicht.¹⁴⁸ Es existiert also noch kein einheitlicher Metadatenstandard zur Dokumentation und Forschungsdatenaustausch, der die „Bandbreite der heterogenen Datenbestände vollständig abbildet“¹⁴⁹.

Hinzu tritt der Aspekt, dass der Bedarf der Datennutzer und Produzenten der Bildungsfor- schungsdaten, also der Forscher und Wissenschaftler, noch nicht empirisch ermittelt ist, eine Einbindung der Wissenschaftler aber zwingend erwünscht ist.¹⁵⁰ Eine Orientierungshilfe wird hier ein Leitfaden für Interviews zu Anforderungen der Forscher an entsprechende Metadaten- repräsentationen bieten,¹⁵¹ bei dem leider noch eine komplette Ergebnisauswertung aussteht. Die Fragestellungen ergeben nämlich einen entscheidenden Aspekt bei der Entwicklung eige- ner Kriterien für die Zusammenstellung eines Metadatensets.

Zusammenfassend können deswegen nur vorläufig eigene Ideen in einer oberflächlich und grob gehaltenen separaten Spezifikation als erster Grundriss eines Application Profiles umge- setzt werden, deren Modifizierung und Aufgliederung stark von den entsprechenden Entwick- lungen der eigenen Bedarfsanalyse am Institut und den Resultaten der für den Bildungsbe- reich relevanten Applications abhängig ist. Durch seine einfache Struktur und seinen allge- meinen Charakter bietet Dublin Core hierbei einen wesentlichen Anhaltspunkt. Durch die Fokussierung des DIPF auf den Standard DDI werden auch die dortigen Module einbezogen, vor allem, um die individuellen Einbindungsoptionen beliebiger Schemas und Elemente durch die Erweiterbarkeit deutlich zu machen.

Insgesamt ist die erste Entscheidung für einen Metadatenstandard also hinsichtlich der Konzi- pierung eines Metadatensets aus institutspolitischen und strategischen Gründen¹⁵² bereits im Grundsatz entschieden. Kooperationen als ein wesentliches „Must have“ bei der Etablierung

¹⁴⁸ EduDDI ist quantitativ orientiert und noch in der Entwicklungsphase, vgl. Mechtel (2009), ebenso der Ver- such, DDI auf qualitative Daten umzumünzen, vgl. Corti (2010). Das speziell geschaffene Dublin Core Educati- on Application Profile wiederum ist auf den „use in teaching and learning“ ausgelegt und für qualitative For- schungsdaten eher ungeeignet, vgl. Zeng (2008), S. 50.

¹⁴⁹ Bambey (2012a), S.128.

¹⁵⁰ Vgl. Bambey (2012a), S. 133f.

¹⁵¹ Vgl. Reinhold (2012).

¹⁵² Dublin Core ist bereits ein Bezugspunkt des Fachportals Pädagogik. Außerdem bestehen bereits Absprachen mit anderen Instituten wie dem IQB oder GESIS, vgl. Bambey (2012b). Auch darf in diesem Kontext nicht die internationale Ausrichtung des Instituts vergessen werden.

von FDZ und beim interoperablen Datenmanagement bestimmen hier den Ansatz des DIPF durchaus mit.¹⁵³

Neben der zu erfolgenden Ausarbeitung gibt es zusätzlich bereits mit Metadaten angereicherte Daten, die in einer eigenen Datenbank zur Verfügung stehen. Es handelt sich um Daten aus dem Projekt „Videodatenbank Schulunterricht in der DDR“.¹⁵⁴ Hier wurden schon Metadatenfelder konzipiert und technisch aufbereitet sowie Guidelines für die Inhalte und Values erstellt. Teilweise können diese übernommen und einbezogen werden, teilweise müssen sie auf Grund von Redundanz und unbrauchbarer Kategorisierung überarbeitet werden.¹⁵⁵

Außerdem intendiert das DIPF durch die Integration des FDZ in den Verbund des Fachportals Pädagogik, das dort implementierte kontrollierte Vokabular, vor allem den FIS Thesaurus¹⁵⁶, maßgeblich einzubinden. Des Weiteren empfiehlt sich zum Zwecke einer gestiegenen Transparenz sowie zur Erleichterung von Interoperabilität und Mapping die Verwendung internationaler Normen und allgemein gültigen definierten Vokabulars.¹⁵⁷

5.2 Metadatenanforderungen für AV-Daten der Bildungsforschung

Immer wieder betont werden muss, dass es sich bei der Schaffung eines Metadatensets um eine individuelle Auswahl für geeignet befundener Metadaten handelt. Dennoch gibt es wesentliche Metadatenelemente, die als verpflichtend bezeichnet werden können und betreffs AV-Daten in ein Application Profile mit entsprechenden Namespaces und Refinements aufgenommen werden sollten. Diese ergeben sich aus den im Folgenden konkretisierten Requirements an zu repräsentierende Metadaten von „Forschungsdaten Bildung“. Bemerkenswert ist der Aspekt, dass die Granularitätsanforderungen von AV-Daten vom üblichen Grundmus-

¹⁵³ Genauere Angaben zu Kooperationsstrukturen bei Bambey (2012a), S. 113-118.

¹⁵⁴ Videodatenbank Schulunterricht in der DDR (2012).

¹⁵⁵ Dieser Leitfaden mitsamt Verbesserungsvorschlägen ist leider nur DIPF-intern verfügbar, vgl. Schwager (2012). Er entwickelt Richtlinien und Valuedefinitionen für jedes einzelne Metadatenfeld. Beispielhaft der Auszug für die Datumsangabe für das Jahr der Aufnahme: „Einheitlich wird hier das Datum nach ISO 8601 (JJJJ-MM-TT) eingetragen. Sollte das Datum der Unterrichtsaufzeichnung nicht bekannt sein, muss das Feld dennoch ausgefüllt werden, in folgender Form: 0000-00-00.“ Schwager (2012).

¹⁵⁶ Hier der Verweis auf das online einsehbare Schlagwortregister, FIS Schlagwortregister (2012).

¹⁵⁷ Gewünschte Bezugspunkte sind u.a. die Gemeinsame Normdatei (GND) der Deutschen Nationalbibliothek (DNB) hinsichtlich Körperschaften und Personen, vgl. GND (2012), der Getty Thesaurus of Geographic Names (TGN) für Geographica, vgl. TGN (2012), für Datumsangaben der Bezug auf die ISO-Norm 8601, vgl. W3C (1997), oder bez. der Sprache auf ISO 639 mit Angaben zu entsprechenden Ländercodes, vgl. ISO 639.2 (2010).

ter des DDI-Standards teilweise abweichen, da die verwendeten Dokumenttypen, also qualitative Forschungsdaten, andere Schwerpunkte verlangen.

Reinhold fasst im Rahmen ihres Dissertationsprojektes in ihrem internen Dokument erste Erkenntnisse im Zuge einer Nutzerbefragung von Forschern oberflächlich zusammen.¹⁵⁸ Die Vorstellungen der Nutzergruppe müssen bei Kenntnis des anzureichernden Datenbestandes in diesem Rahmen miteinbezogen werden. Es wird auf jeden Fall ersichtlich, dass beschreibende Metadaten zu den Zielen einer Studie, einzelnen Studiengruppen und über das vorliegende Dokument selbst von wesentlichem Belang sind. Auffällig ist außerdem der starke Wunsch nach Dokumentation der Begleitumstände wie Settings, verwendete Materialien und Methoden. Diese werden aber, weil Nicht-AV-Daten, separat behandelt und müssen in Form von Relationen und Verweisstrukturen eingebaut werden.

Zusätzlich ergeben sich weitere Metadaten von essentiellm Interesse, die nun unter Einbeziehung des vorliegenden Datenbestandes, vor allem der sich in Betrieb befindlichen „Video-datenbank Schulunterricht in der DDR“ kurz angeführt werden.¹⁵⁹

Zunächst gilt zu konstatieren, dass grundsätzlich jedes Dokument identifizierbar sein muss. Deswegen muss immer ein eindeutiger Persistent Identifier vergeben werden, um jegliche Verwechslungsmöglichkeit von vornherein auszuschließen.¹⁶⁰

Des Weiteren müssen die wesentlichen Beteiligten genannt werden. Dies betrifft zum einen die Projektleitung, also die Urheber und Betreuer der Studien, Datenaufbereiter und wissenschaftliche Mitarbeiter sowie alle beteiligten relevanten Personen und Institutionen. Dies muss bei einer möglichen Lokalisierung in Verbindung mit geographischen Angaben erfolgen.

¹⁵⁸ Daraus ergeben sich folgende Bereiche: Zielsetzung einer Studie, Anwendungsbereich, Konstruktion, Zielgruppe, Durchführungszeitraum, Literatur, Sprache und Ursprungsjahr. Außerdem besteht der Bedarf nach Dokumentation der Methoden wie Interviews oder Fokusgruppendifkussion und des Begleitmaterials. Unter letzterem werden vor allem Transkripte, Kameraskripte, Verlaufsprotokolle mit Codierungen und Codierungsschemata, Beschreibungen der Settings (z.B. Sitzplan einer Klasse, Skizze des Tafelbildes, Unterrichtsmaterialien), Beobachtungsprotokolle und generell Informationen zur Aufbereitung der Videodaten geführt. Bei der Untersuchung des Nutzerverhaltens wurde zudem deutlich, dass vor allem Personen, Institutionen, Keywords, Projekte, Jahr, Literatur, Unterrichtsgegenstand und geographische Angaben gewünscht wurden, vgl. Reinhold (2012).

¹⁵⁹ Viele der im Folgenden angeführten Vorschläge beruhen auf einer intensiven Prüfung der Metadatenfelder der Projektdatenbank und wurden über einen längeren Zeitraum auf ihre Konsistenz und die Erfüllung der Anforderungskriterien untersucht, vgl. Schwager (2012).

¹⁶⁰ Noch einmal der Hinweis in diesem Kontext, dass DDI und Dublin Core jedes einzelne Element, Schema, Relation und Attribut per Identifier klar kenntlich zu machen, vgl. DCMI (2012). Dies wird im Folgenden für alle Bestandteile des zu entwickelnden Metadatensets vorausgesetzt.

Bei den administrativen Metadaten liegt der Schwerpunkt auf Zugangsrechten und der Wahrung des Urheberrechts. Interne Access Rights müssen klar für die Data Maintenance definiert und Änderungen dokumentiert werden. Außerdem stellen AV-Daten einen sensiblen forschungsethisch brisanten Bereich dar. Aus Gründen des Datenschutzes und Persönlichkeitsrechts muss, wenn möglich, eine Anonymisierung erfolgen. Bei Unterrichtsfilmen sind aber Ausnahmefälle verzeichnet, wenn der Aufwand einer Anonymisierungs- oder Pseudonymisierungsverfahrens zu hoch ist oder die Forschungsintention ad Absurdum geführt wird.¹⁶¹ Dies ist beim Datenbestand des DIPF teilweise der Fall. Deswegen müssen konkrete Nutzungsbestimmungen definiert und eingebunden werden, also in realiter durch eine klare Kopplung des externen Data Access an Forschungsinteressen. Bei den rechtlichen Metadaten ist auch die Frage des Rechtsnachfolgers bzw. des Rechteinhabers von großem Belang und muss angegeben werden, bedarf also eines eigenen Metadatenelements.¹⁶²

Von Nöten für eine sinnvolle Nachnutzung sind außerdem Angaben zu den Ressourceninhalten. Diese deskriptiven Metadaten werden in Form eines Titelfeldes, eines Themas, Schlagwörtern und eines Abstracts angesetzt. Außerdem sollen die im Rahmen des Filmes verwendeten Materialien und Medien, das Unterrichtsfach, das inhaltlich zu definierende Genre des Films, Settings sowie eine Kommentarfunktion zur Verfügung stehen. Durch diese Angaben ist es dem Nutzer besser möglich, ohne aufwendige Sichtung und Download der bereitgestellten Ressourcen sich schnell einen Überblick über die abgebildeten Inhalte zu verschaffen.

Betrachtet man die technischen Requirements, werden Informationen über das Dateiformat inklusive Name, Größe und Laufzeit bzw. Dauer einerseits, aber auch Angaben zum Daten- und Medientypen, System- und Softwareanforderungen sowie Kodierungsschemata benötigt. Auch die Sprache des Dokumenteninhalts ist relevant und muss festgehalten werden.

Dies trifft auch auf Datumsangaben zu. Sowohl Untersuchungszeiträume im inhaltlichen Bereich wie die Dauer einer Studie, aber auch verschiedene Datumsangaben im Untersuchungszeitraum wie Entstehung, Aufnahmedatum, Publikationsdatum der Studie sind für Sekundäranalysen erforderliche Kriterien und müssen dokumentiert werden.

¹⁶¹ Vgl. HDSG (1999), §33, ausführlich behandelt bei Medjedović (2010), S. 64ff und am Beispiel Pythagoras bei Hugener (2006).

¹⁶² Vgl. Bambey (2012b).

Außerdem bedarf es Informationen zur jeweiligen Zielgruppe, zur Datenherkunft, zu Relationen zu anderen Dokumenten, auf die Studie bezogene verwendete Begleitmaterialien wie Transkripte oder Literatur sowie natürlich zum Projekt und der Studie selbst als Quelle.

Eine auf diesen Erkenntnissen basierende tabellarische Übersicht der wesentlichen zu dokumentierenden Elemente ist inklusive eines ersten Versuchs der vorzunehmenden Eintragungen unter Nennung der Encoding Values im Anhang beigefügt.¹⁶³ Bezugspunkt ist hierbei das Grundgerüst der Dublin Core Terms¹⁶⁴ und die Specifications von DDI 3.1¹⁶⁵.

5.3 Modulare Formen und exemplarische Umsetzung

Das Grundprinzip bei der Einbindung von DDI und Dublin Core unter der Schaffung einer eigenen Spezifikation ist die Verwendung geeigneter modularer Formen aus DDI, in welche die einzelnen Dublin Core Elemente wiederum eingebaut sind. Also müssen die Module aus DDI erst ausgewählt werden und dann über die sog. „Inclusion by Reference“¹⁶⁶ in Schemas zusammengebaut und selbst in ihrer Granularität und den vorgenommenen Refinements sowie Encoding Values definiert werden. Einzelne Elemente wie die Dublin Core-Properties können dabei mehreren Modulen zugeordnet werden.¹⁶⁷

Ein Beispiel hierfür ist das Metadatenelement Date. Es kann über Refinements nach eigenem Anspruch definiert werden und wurde so im vorliegenden Entwurf¹⁶⁸ mit den Refinements Aufnahmedatum (created), Eingangsdatum (DateAccepted), Veröffentlichungsdatum (Issued) und Bearbeitungsdatum (Modified) versehen.¹⁶⁹ Gerade im Forschungsbereich sind diese Angaben für die Verwendung von Primärdaten von wesentlichem Interesse. Hierbei ist es ratsam, zur besseren Interoperabilität und wegen der Vielzahl an unterschiedlichen Datumsangaben eine einheitliche Norm zu wählen. Dies erfolgt unter Nutzung des internationalen Standards ISO 8601 und der dort festgesetzten Datums- und Zeitformate.¹⁷⁰ Am Element Date

¹⁶³ Anhang. Anlage 2: Vorläufiger Entwurf des Metadatensets für AV-Daten für das Forschungsdatenzentrum „Forschungsdaten Bildung“ des DIPF.

¹⁶⁴ DDCMI (2012).

¹⁶⁵ DDI (2009a). Technical Specification. DDI 3.1.

¹⁶⁶ Vgl. DDI (2009a). Technical Specification. DDI 3.1. User Guide, S. 52f.

¹⁶⁷ Wie in Kapitel 4.4.3 beschrieben.

¹⁶⁸ Anhang. Anlage 2: Vorläufiger Entwurf des Metadatensets für AV-Daten für das Forschungsdatenzentrum „Forschungsdaten Bildung“ des DIPF.

¹⁶⁹ Orientierungspunkt ist hierbei Schema DC-Terms (2003).

¹⁷⁰ DDI empfiehlt dies explizit, vgl. DDI (2009a). Technical Specification. DDI 3.1. User Guide, S. 48.

lässt sich gut illustrieren, wie das System und Datenmodell funktioniert: Eine Studie hat ein Entstehungsdatum und benötigt die entsprechende Dokumentation, aber auch das Dokument selbst oder der Datensatz besitzen ein Entstehungsdatum. Theoretisch wäre auch eine jeweilige Versionierung im Aufführen der Versionen nach DDI-Kriterien umzusetzen und mit einem Versionierungsdatum zu versehen.¹⁷¹ Doch ist dies bei AV-Daten mit unabänderlichen Inhalten nicht zwingend notwendig. Bei Bedarf kann eine entsprechende Erweiterung allerdings jederzeit erfolgen. Dies ist auch im Sinne der DC-Prinzipien Optionalität, Erweiterbarkeit und Wiederholbarkeit¹⁷². Das Element Date kann also jedem Modul nach Bedarf zugeordnet werden und entsprechend in den Schemas dokumentiert werden.

Ähnlich verhält es sich mit dem Core Element Description. Mit seinen Refinements kann es optional in die nötigen DDI-Module der eigenen Specification eingebunden werden. Das Modul Study Unit, also die Studie als Ganzes, benötigt ebenso beschreibende Metadaten (z.B. Abstract, Notes) wie die einzelnen Unterrichtsdokumente selbst. Die Granularität kann dabei frei nach Bedarf entschieden werden. Ebenso verhält es sich mit Schlagwörtern und dem Thema, also meist dem Unterrichtsgegenstand. Im Falle von „Forschungsdaten Bildung“ bedarf es dieser Refinements zur verbesserten Suche und Datennutzung.

Gerade bei Filmen aus der Unterrichtsforschung besteht ein hoher Bedarf an Granularität bei deskriptiven Metadaten. Dies betrifft z.B. die Dokumentation der Gattung des Films, eine eigens für diesen Bereich unter Description geführte Einteilung nach festgelegten Values wie Unterrichtsaufzeichnung, Lehrfilm oder Lehrerfortbildung, oder auch die Einteilung nach Unterrichtsfächern. Hier wird die Installation fester Bezugspunkte in Form von Indexlisten mit der Nutzung eines in Encoding Schemes verwendeten festen Vokabulars angestrebt. Dies soll die Navigation und Suche für den Nutzer erleichtern.¹⁷³

Ein weiterer Bereich sind administrative Metadatenelemente. Auch hier muss jeweils für einzelne Studien die Rechtesituation, sowohl bezüglich der Rechteinhaber via RightsHolder als auch die Zugangsberechtigung, geklärt werden.

Wie erwähnt, entsteht im Zuge der AV-Datenproduktion Begleitmaterial, das von großem Interesse für die Forschung sein kann. Zu nennen sind hier beispielsweise Fragebögen, Tafel-

¹⁷¹ Vgl. DDI Best Practice (2009e).

¹⁷² Die Wiederholungsmöglichkeit der Elemente ist Grundvoraussetzung des Konzepts, siehe Kapitel 4.4.2.

¹⁷³ Vgl. die Indexlisten bei Videodatenbank Schulunterricht in der DDR (2012); zur Verwendung kontrollierten Vokabulars, vgl. DDI Best Practice (2009a).

bilder, Sitzpläne einerseits, aber auch Transkripte, Kameraskripte oder eingesetzte Unterrichtsmedien. All dies wird hier in seiner Untergliederung nicht weiter bestimmt (ebenso wie die angewandten Methoden unter Methods)¹⁷⁴, allerdings mit dem Hinweis, dass unter OtherMaterial entsprechende Refinements nach Bedarf vorgenommen werden können, je nach dem jeweiligen konkreten Dokumentationsinteresse.¹⁷⁵

Ein wichtiger Bestandteil ist die Einbindung und Dokumentation von Beziehungen zueinander unter der Nutzung des Elements Relation. Literatur wird zwar extra unter BibliographicCitation erwähnt, doch gerade bei OtherMaterial und noch signifikanter durch notwendige Verweise zwischen den einzelnen Dokumenten zueinander¹⁷⁶ und zu den Studien selbst muss die Struktur in dieser Hinsicht klar festgehalten werden. Dies kann in zwei Formen erfolgen: Indem man die Referenz selbst durch einen anderen Datensatz oder Dokument festhält via isReferencedBy oder selbst die Referenz zu einer Ressource dokumentiert über References.¹⁷⁷

Abschließend eine Bemerkung zur Verwendung von DDI-Modulen: „Forschungsdaten Bildung“ ist an den einzelnen wissenschaftlichen Studien ausgerichtet. Die AV-Daten sind im Rahmen von wissenschaftlichen Projekten entstanden, die dokumentiert werden müssen. Deswegen muss zu jeder Studie eine entsprechende Study Unit vergeben werden. Diese Study Unit muss als Modul wiederum eine Beschreibung enthalten sowie die einzelnen Study Groups anführen. Diese führen wiederum das Element Description. Hier lassen sich in DDI in vorgeschriebener Weise Schemas und Module komplett oder im Auszug übernehmen. Auf diese Weise werden die Projekte als Module definiert. Anschließend können die einzelnen Bestandteile der Data Collection als Module entwickelt werden. So lassen sich grundsätzlich übersichtliche Strukturen schaffen, die aufeinander abbildbar sind und individuell angepasst werden können.¹⁷⁸ Da das FDZ „Forschungsdaten Bildung“ allerdings selbst noch in der Konzipierungsphase ist, viele rechtliche und technische Aspekte¹⁷⁹ noch Klärungsbedarf aufweisen und nur ein kleiner Teil der Datenbestände konkrete Anforderungen an abzubildende Metadatenstandards erkennen lässt, bleiben weitere Präzisierungsschritte wie die Einbindung von MPEG-7 aus und Aufgabe zukünftiger Arbeitsprozesse.

¹⁷⁴ Außerdem muss angemerkt werden, dass einerseits ein starkes Ausmaß an Methodenvielfalt herrscht und andererseits diese oft von den Forschern nicht ausreichend aufgefangen werden, vgl. Bambey (2012a), S. 115.

¹⁷⁵ Hierzu bietet der DDI 3.1. Overview Anhaltspunkte, vgl. DDI (2009a).

¹⁷⁶ Fragebögen und Tests zu den Studien sind beispielsweise in DaQS abgebildet, vgl. DaQS (2012).

¹⁷⁷ Vgl. Schema DC-Terms (2003).

¹⁷⁸ Einen Überblick über die Einrichtung und Definition dieser Module findet sich unter DDI Best Practice (2010).

¹⁷⁹ So fehlt es zum Beispiel an finalen Aussagen zur Kodierung.

6 Abschließende Bemerkung

Betrachtet man resümierend den Entwicklungsprozess des Metadatensets für das geplante Forschungsdatenzentrum „Forschungsdaten Bildung“, wird deutlich, dass unter den gegebenen Ausgangsbedingungen nur eine vorläufige Lösung erzielt werden konnte. Zu wenig fortgeschritten ist einerseits die interne Bedarfsanalyse für den speziellen Bereich der AV-Forschungsdaten der Bildung und andererseits die Weiterentwicklung verwendbarer Application Profiles für diesen Bereich innerhalb der vom DIPF bevorzugten Initiativen und Standards.

In Form von einzelnen exemplarischen Umsetzungsmöglichkeiten ist es dennoch gelungen, wesentliche, auch aus der institutspolitischen Rason resultierende Anforderungen für die abzubildenden Metadaten zu bündeln, die erste Stufe eines Application Profiles zu erklimmen und kontrolliertes Vokabular im Ansatz zu definieren. Weitere Spezifizierungen müssen allerdings noch in der Folgezeit unter Berücksichtigung der Entwicklungen innerhalb der entsprechenden internationalen Arbeitsgruppen vorgenommen werden. Dennoch illustriert diese Arbeit im Grundprinzip, wie anhand der Einbeziehung von DDI und unter Orientierung an Dublin Core erste modulare Formen ausgearbeitet werden können und welche Anforderungen an Metadaten dokumentiert werden müssen. Die Vielfalt der Kombinationsmöglichkeiten und die freie Wahl der Granularität in Form von individuell vorzunehmender Refinements inklusive der Nutzung von XML und entsprechender Schemas haben dabei aufgezeigt, welche gestalterische Möglichkeiten theoretisch bei der Entwicklung eines eigenen Application Profiles bestehen. Diese müssen und werden bei der zu erwartenden Weiterentwicklung des FDZ am DIPF in einem langwierigeren Prozess festgesetzt werden. Ein Ansatz hierzu ist auf jeden Fall entstanden, der darüber hinaus einen Orientierungspunkt für die notwendige Spezifizierung des Metadatensets bietet. Selbst wenn dabei eine Ausweitung des Schwerpunkts von AV-Daten auf andere qualitative Datentypen erfolgt, enthält das Modell in seinen Grundprinzipien Einbindungsmöglichkeiten anderer Standards wie MPEG-7 und Normen, also infolge der Flexibilität und des Prinzips der Modularisierung und Schematisierung ausreichend Erweiterungs- und Anpassungspotenzial.

7 Literatur- und Quellenverzeichnis

- Bambey (2012a). Bambey, Doris / Reinhold, Anke / Rittberger, Marc: Pädagogik und Erziehungswissenschaft. In: Neuroth, Heike et. al. (Hrsg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg : Verlag Werner Hülsbusch. 2012. S. 111-135.
- Bambey (2012b). Bambey, Doris: Konzept zur Entwicklung einer Forschungsdaten- und Meta-Informationsstruktur für die Bildungsforschung. Unveröffentlichtes internes Papier (Stand: 19.04.2012).
- Bertram (2005). Bertram, Jutta: Einführung in die inhaltliche Erschließung. Grundlagen – Methoden – Instrumente (Content and Communication, Bd. 2). Würzburg : ERGON. 2005.
- Caplan (2009). Caplan, Priscilla: Understanding PREMIS. The Library of Congress (2009-02-01).
<http://www.loc.gov/standards/premis/understanding-premis.pdf>; Rev. 2012-08-22.
- Corti (2010). Corti, Louise / Gregory, Arofan: Report back from the DDI Qualitative Working Group. European DDI Meeting. Utrecht 8-9 Dec 2010.
http://www.iza.org/conference_files/eddi10/EDDI10_Presentations/EDDI10_B2_LouiseCorti_Slides.ppt; Rev. 2012-08-22.
- Dallmeier-Tiessen (2012). Dallmeier-Tiessen, Sünje: Die wissenschaftsorientierte Publikation von Forschungsdaten. In: 100. Deutscher Bibliothekartag in Berlin 2011. Bibliotheken für die Zukunft - Zukunft für die Bibliotheken (Deutscher Bibliothekartag: Kongressbände). Hildesheim [u.a.] : Olms. 2012. S. 75-86.
- DaQS (2012). Datenbank zur Qualität von Schule (DaQS).
<http://daqs.fachportal-paedagogik.de/>; Rev. 2012-08-22.
- DCMI (2007). Powell, Andy et al.: DCMI Abstract Model (2007-06-04).
<http://dublincore.org/documents/abstract-model/>; Rev. 2012-08-22.
- DCMI (2012). DCMI Usage Board: DCMI Metadata Terms. 2012 (2012-06-14).
<http://dublincore.org/documents/dcmi-terms/>; Rev. 2012-08-22.
- DDI (2009a). Data Documentation Initiative. DDI 3.1 (2009-11-12).
<http://www.ddialliance.org/Specification/DDI-Lifecycle/3.1/>; Rev. 2012-08-22.
- DDI (2009b). Data Documentation Initiative. Mapping to Dublin Core (DDI Version 2) (2009-12-02).
<http://www.ddialliance.org/resources/tools/dc>; Rev. 2012-08-22.
- DDI (2011). Data Documentation Initiative. XML Schema for DDI Lifecycle – Version 3.1 (2011-11-01).
<http://www.ddialliance.org/Specification/DDI-Lifecycle/3.1/XMLSchema/>; Rev. 2012-08-22.

- DDI (2012a). Data Documentation Initiative.
<http://www.ddialliance.org/>; Rev. 2012-08-22.
- DDI (2012b). DDI Qualitative Data Exchange Working Group (2012-07-06).
<http://www.ddialliance.org/alliance/working-groups#qdewg>; Rev. 2012-08-22.
- DDI (2012c). Data Documentation Initiative. What is DDI? (2012-08-19).
<http://www.ddialliance.org/what>; Rev. 2012-08-22.
- DDI (2012d). Data Documentation Initiative. Alliance Structure (2012-07-23).
<http://www.ddialliance.org/alliance/structure#members>; Rev. 2012-08-22.
- DDI (2012e). Data Documentation Initiative. Codebook 2.5 (2012-03-06).
<http://www.ddialliance.org/Specification/DDI-Codebook/2.5/>; Rev. 2012-08-22.
- DDI (2012f). Data Documentation Initiative. DDI at work.
<http://www.ddialliance.org/ddi-at-work>; Rev. 2012-08-22.
- DDI (2012g): Data Documentation Initiative. DDI Tools.
<http://www.ddialliance.org/resources/tools>; Rev. 2012-08-22.
- DDI Best Practice (2009a). Granda, Peter et al.: Controlled Vocabularies. DDI Working Paper Series – Best Practice, No. 5 (2009-02-22).
http://www.ddialliance.org/sites/default/files/bp/DDIBestPractices_ControlledVocabularies.doc.pdf; Rev. 2012-08-22.
- DDI Best Practice (2009b). Ionescu, Sandra: Creating a DDI Profile. DDI Working Paper Series – Best Practice, No. 6 (2009-02-15).
http://www.ddialliance.org/sites/default/files/bp/DDIBestPractices_CreatingAProfile.doc.pdf; Rev. 2012-08-22.
- DDI Best Practice (2009c). Linnerud, Jenny et al.: DDI 3.0 Schemas. DDI Working Paper Series – Best Practice, No. 7 (2009-03-22).
http://www.ddialliance.org/sites/default/files/bp/DDIBestPractices_Schemes.doc.pdf;
Rev. 2012-08-22.
- DDI Best Practice (2009d). Askitas, Nikos / Gregory, Arofan / Hoogerwerf, Maarten: Management of DDI 3.0 Unique Identifiers. DDI Working Paper Series – Best Practice, No. 10 (2009-02-15).
http://www.ddialliance.org/sites/default/files/bp/DDIBestPractices_ManagementOfDDIIdentifiers.doc.pdf; Rev. 2012-08-22.
- DDI Best Practice (2009e). Edwards, Michelle et al.: Versioning and Publication. DDI Working Paper Series – Best Practice, No. 8 (2009-03-22).
http://www.ddialliance.org/sites/default/files/bp/DDIBestPractices_VersioningAndPublication.doc.pdf; Rev. 2012-08-22.
- DDI Best Practice (2010). Data Documentation Initiative: Getting Started with DDI (2010-05-17).
<http://www.ddialliance.org/getting-started#best>; Rev. 2012-08-22.

- DFG (1998). Deutsche Forschungsgemeinschaft (Hrsg.): Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“. Weinheim : Wiley-VCD. 1998.
- DFG (2006). DFG-Positionspapier. Wissenschaftliche Literaturversorgungs- und Informationssysteme – Schwerpunkte der Förderung bis 2015.
<http://dfg.de/download/pdf/foerderung/programme/lis/positionspapier.pdf>; Rev. 2012-08-22.
- DFG (2009): DFG-Praxisregeln „Digitalisierung“ zu den Förderprogrammen der Wissenschaftlichen Literaturversorgungs- und Informationssysteme. (2009)
http://www.dfg.de/download/pdf/foerderung/programme/lis/praxisregeln_digitalisierung.pdf; Rev. 2012-08-22.
- Dickmann (2012). Dickmann, Frank / Rienhoff, Otto: Medizin. In: Neuroth, Heike et. al. (Hrsg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg : Verlag Werner Hülsbusch. 2012. S. 227-256.
- DIF (2012). Directory Interchange Format (DIF). Writer’s guide (2012).
<http://gcmd.nasa.gov/User/difguide/difman.html>; Rev. 2012-08-22.
- DIPF (2006): Unterricht und Kompetenzerwerb in Deutsch und Englisch. Zentrale Befunde der Studie Deutsch Englisch Schülerleistungen International (DESI). 2006.
<http://www.dipf.de/de/projekte/pdf/biqua/desi-zentrale-befunde>; Rev. 2012-08-22.
- DIPF (2012). Deutsches Institut für Internationale Pädagogische Forschung.
<http://www.dipf.de/>; Rev. 2012-08-22.
- DOI (2012). The DOI System.
<http://www.doi.org/>; Rev. 2012-08-22.
- Dublin Core (2005). Woodley, Mary S.: DCMI Glossary (2005-11-07).
<http://dublincore.org/documents/usageguide/glossary.shtml>; Rev. 2012-08-22.
- Dublin Core (2009). Interoperability Levels for Dublin Core Metadata (2009-05-01).
<http://www.dublincore.org/documents/interoperability-levels/#DC-TEXT>; Rev. 2012-08-22.
- Dublin Core (2012a). Dublin Core Metadata Initiative.
<http://dublincore.org/>; Rev. 2012-08-22.
- Dublin Core (2012b). Dublin Core Metadata Element Set 1.1 (2012-06-14).
<http://www.dublincore.org/documents/dces/>; Rev. 2012-08-22.
- Dublin Core (2012c). DC-Education Application Profile.
http://dublincore.org/educationwiki/DC_2dEducation_20Application_20Profile; Rev. 2012-08-22; Verweis auf https://docs.google.com/View?docid=dn8z3gs_38cgwkvy; Rev. 2012-08-22.
- Fachportal Pädagogik (2012). Fachportal Pädagogik.
<http://www.fachportal-paedagogik.de/>; Rev.2012-08-22.

- FIS Schlagwortregister (2012). Schlagwortregister der FIS Bildung Literaturdatenbank.
http://www.fachportal-paedagogik.de/fis_bildung/fis_sws.html?feldnr=2&; Rev. 2012-08-22.
- Foulonneau (2008). Foulonneau, Muriel / Riley, Jenn: Metadata for Digital Resources. Implementation, System Design and Interoperability. Oxford : Chandos. 2008.
- Garz (2010): Garz, Detlef / Blömer, Ursula: Qualitative Bildungsforschung. In: Tippelt, Rudolf / Schmidt, Bernhard (Hrsg.): Handbuch Bildungsforschung, 3. Aufl., Wiesbaden : VS-Verlag. 2010. S. 571-588.
- GESIS (2012a). GESIS da|ra Registrierungsagentur für sozialwissenschaftliche Forschungsdaten.
<http://www.gesis.org/dara>; Rev. 2012-08-22.
- GESIS (2012b). Datenarchiv für Sozialwissenschaften GESIS.
<http://www.gesis.org/das-institut/wissenschaftliche-abteilungen/datenarchiv-fuer-sozialwissenschaften/>; Rev. 2012-08-22.
- GeoForschungszentrum Potsdam (2012). Helmholtz-Zentrum Potsdam Deutsches GeoForschungszentrum Potsdam (GFZ). Forschungsdaten.
<http://www.gfz-potsdam.de/portal/gfz/Services/Forschungsdaten.jsessionid=FEB1399CE47494AE145C913D93C3F284>; Rev. 2012-08-22.
- GND (2012). Deutsche Nationalbibliothek: Gemeinsame Normdatei (GND).
http://www.dnb.de/DE/Standardisierung/Normdaten/GND/gnd_node.html; Rev. 2012-08-22.
- HDSG (1999). Hessisches Datenschutzgesetz (HDSG) in der Fassung vom 7. Januar 1999.
- Heery (2000). Heery, Rachel / Patel, Manjula: Application Profiles. Mixing and Matching Metadata Schemas. (2000-09-24).
<http://www.ariadne.ac.uk/issue25/app-profiles>; Rev. 2012-08-22.
- Hercher (2011). Hercher, Johannes / Mitzscherlich, Anita. / Sack, Harald: Bestandsanalyse, Metadaten und Systematisierung. In: Vogel, Andreas / Effenberg, Peter (Hrsg.): Digitalisierungsfibel. Leitfaden für audiovisuelle Archive. Potsdam : transfermedia. 2011. S. 82-101. Verwendete Fassung:
<http://www.hpi.uni-potsdam.de/fileadmin/hpi/FG ITS/Semantic-Technologies/paper/hercher2011.pdf>; Rev. 2012-08-22.
- High Level Expert Group on Scientific Data (2010). High Level Expert Group on Scientific Data : Riding the wave – How Europe can gain from rising tide of scientific data - Final report of the High Level Expert Group on Scientific Data. Verwendete Fassung:
http://ec.europa.eu/information_society/newsroom/cf/itemlongdetail.cfm?item_id=6204; Rev. 2012-08-22.
- Hillmann (2005). Hillmann, Diane: Using Dublin Core (2005-11-07).
<http://dublincore.org/documents/usageguide/>; Rev. 2012-08-22.

- Hugener (2006). Hugener, Isabelle / Pauli, Christine / Reusser, Kurt: Videoanalysen. In: Klieme, Eckhard / Pauli, Christine / Reusser, Kurt: Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis, Teil 3 (Materialien zur Bildungsforschung, Bd. 15). Frankfurt am Main: Gesellschaft zur Förderung Pädagogischer Forschung (GFPD). 2006.
- ISDC des GFZ Potsdam (2012). Information System and Data Center for geoscientific data. <http://isdc.gfz-potsdam.de/>; Rev. 2012-08-22.
- ISO 15836:2009 (2012). ISO 15836:2009. Information and documentation -- The Dublin Core metadata element set. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=52142; Rev. 2012-08-22.
- ISO 19115:2003 (2012). ISO 19115:2003. Geographic Information – Metadata. http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020; Rev. 2012-08-22.
- ISO 639.2 (2010). ISO 639.2. Codes for the Representation of Names of Languages (2010-10-18). http://www.loc.gov/standards/iso639-2/php/code_list.php; Rev. 2012-08-22.
- KIM (2008). Kompetenzzentrum Interoperable Metadaten: Übersetzung des Dublin Core Metadata Initiative Abstract Model (DCMA) (2008-11-11). http://www.kim-fo-rum.org/Subsites/kim/SharedDocs/Downloads/DE/Uebersetzungen/dcam.pdf?__blob=publicationFile; Rev. 2012-08-22.
- Kleiner (2012): Kleiner, Matthias: Vorwort. In: Neuroth, Heike et. al. (Hrsg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg : Verlag Werner Hülsbusch. 2012. S. 9-11.
- Klump (2012). Klump, Jens: Geowissenschaften. In: Neuroth, Heike et. al. (Hrsg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg : Verlag Werner Hülsbusch. 2012. S. 179-194.
- Krause (2004). Krause, Jürgen: Standardisierung und Heterogenität. In: Kuhlen, Rainer / Seeger, Thomas / Strauch, Dietmar: Handbuch zur Einführung in die Informationswissenschaft und -praxis. Bd. 1: Grundlagen der praktischen Information und Dokumentation. 5., völlig neu gefasste Ausg., München : Saur. 2004. S. 635-641.
- Krüger (2000). Krüger, Heinz-Hermann: Stichwort: Qualitative Forschung in der Erziehungswissenschaft. In: Zeitschrift für Erziehungswissenschaft, Heft 3, 2000 (3. Jg.). S. 323-342.
- Ludwig (2012). Ludwig, Jens: Zusammenfassung und Interpretation. In: Neuroth, Heike et. al. (Hrsg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg : Verlag Werner Hülsbusch. 2012. S. 295-310.

- Mechtel (2009). Mechtel, Martin / Knigge, Michel: EduDDI. An application for DDI 3.x large scale assessments in education. Tampere : IASSIST/IFDO. 2009.
http://www.fsd.uta.fi/iassist2009/presentations/D1_Mechtels.ppt; Rev. 2012-08-22.
- Medjedović (2010). Medjedović, Irena / Witzel, Andreas: Wiederverwendung qualitativer Daten. Archivierung und Sekundärnutzung qualitativer Interviewtranskripte. Wiesbaden: Verlag für Sozialwissenschaften. 2010.
- MetaGuide SUB Göttingen (2001). MetaGuide SUB Göttingen: Metadata Server. Einführung in Metadaten (2001-03-27).
<http://www2.sub.uni-goettingen.de/intrometa.html>; Rev. 2012-08-22.
- MPEG-7 (2004). Martinez, José M.: MPEG-7 Overview (version 10). Palma de Mallorca. 2004.
<http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm>; Rev. 2012-08-22.
- NEPS (2012). NEPS Nationales Bildungspanel.
<https://www.neps-data.de/>; Rev. 2012-08-22.
- Nestor (2012). Nestor Kompetenznetzwerk zur digitalen Langzeitarchivierung.
http://www.langzeitarchivierung.de/Subsites/nestor/DE/Home/home_node.html; Rev. 2012-08-22.
- Neuroth (2009). Neuroth, Heike et al. (Hrsg.): nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0. Kapitel 8.2. Bitstream Preservation. Boizenburg : Verlag Werner Hülsbusch. 2009.
http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_346.pdf; Rev. 2012-08-22.
- Neuroth (2012). Neuroth, Heike: Vorgehensweise. In: Neuroth, Heike et al. (Hrsg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg : Verlag Werner Hülsbusch. 2012. S. 51-59.
- NISO (1998). St. Pierre, Margaret / LaPlant, William P. Jr.: Issues in Crosswalking Content Metadata Standards (1998-10-15).
http://www.niso.org/publications/white_papers/crosswalk/; Rev. 2012-08-22.
- NISO (2004). Understanding Metadata? Bethesda : NISO Press. 2004.
<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>; Rev. 2012-08-22.
- OAI-PMH (2005). Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting (2005-01-19).
<http://www.openarchives.org/OAI/2.0/guidelines-repository.htm>; Rev. 2012-08-22.
- Oßwald (2012). Oßwald, Achim / Scheffel, Regine / Neuroth, Heike: Langzeitarchivierung von Forschungsdaten. Einführende Überlegungen. In: Neuroth, Heike et. al. (Hrsg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg : Verlag Werner Hülsbusch. 2012. S. 13-21.
- Pempe (2012). Pempe, Wolfgang: Geisteswissenschaften. In: Neuroth, Heike et. al. (Hrsg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg : Verlag Werner Hülsbusch. 2012. S. 137-159.

- PERLE (2012). Projekt PERLE Persönlichkeits- und Lernentwicklung von Grundschulkindern.
<http://www.perle-projekt.de/>; Rev. 2012-08-22.
- Persistent Identifier (2012). Persistent Identifier. Der Service der Deutschen Nationalbibliothek.
<http://www.persistent-identifier.de/>; Rev. 2012-08-22.
- PREMIS (2012). PREMIS Preservation Metadata Maintenance Activity.
<http://www.loc.gov/standards/premis/>; Rev. 2012-08-22.
- Quandt (2012). Quandt, Markus / Mauer, Reiner: Sozialwissenschaften. In: Neuroth, Heike et al. (Hrsg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg : Verlag Werner Hülsbusch. 2012. S. 61-81.
- RatSWD (2010). Kriterien des Rates für für Sozial- und Wirtschaftsdaten (RatSWD) für die Einrichtung der Forschungsdateninfrastruktur (2010).
http://www.ratswd.de/download/publikationen_rat/RatSWD_FDZKriterien.PDF; Rev. 2012-08-22.
- RatSWD (2012). Rat für Sozial- und Wirtschaftsdaten. Forschungsdatenzentren.
<http://www.ratswd.de/dat/fdz.php>; Rev. 2012-08-22.
- Reinhold (2012). Reinhold, Anke: Anforderungsanalyse – Ermittlung von Nutzeranforderungen an die Instrumentendatenbank. Unveröffentlichtes internes Papier. (Stand: 02.08.2012).
- Schema DC-Terms (2003). Schema „dcterms.xsd“ (2003-04-02).
http://www.ddialliance.org/Specification/DDI-Lifecycle/2.5/XMLSchema/FieldLevelDocumentation/schemas/dcterms_xsd/schema-overview.html#a1; Rev. 2012-08-22.
- Schluß (2007). Schluß, Henning / Crivellari, Fabio: Videodokumentation von Unterricht in der DDR als Quelle - Ergebnisse eines DFG-Projekts zur medialen Unterrichtsforschung. In: Pädagogische Rundschau, Heft 4, 2007 (61. Jg.). S. 437-452.
- Schulzeck (2008). Schulzeck, Ursula: Bildungsforschungsprojekte im Forschungsinformationssystem Sozialwissenschaften (FORIS). In: Weishaupt, Horst et al.(Hrsg.): Zur Situation der Bildungsforschung in Deutschland (Bildungsforschung, Bd. 28). Berlin : Bundesministerium für Bildung und Forschung (BMBF). 2008. S. 36-55.
- Schütz (2004). Schütz, Thomas: Dokumentenmanagement. In: Kuhlen, Rainer / Seeger, Thomas / Strauch, Dietmar: Handbuch zur Einführung in die Informationswissenschaft und -praxis. Bd. 1: Grundlagen der praktischen Information und Dokumentation. 5., völlig neu gefasste Ausg. München : Saur. 2004. S. 339-349.
- Schwager (2012). Schwager, Thomas: Richtlinien für die Dokumentation von Unterrichtsfilmen. Unveröffentlichtes internes Papier (Stand: 30.05.2012).
- Statistische Ämter des Bundes und der Länder (2012). Forschungsdatenzentrum der Statistischen Ämter des Bundes und der Länder.
<http://www.forschungsdatenzentrum.de/>; Rev. 2012-08-22.

- Strathmann (2012): Stand der Diskussion und Aktivitäten. International. In: Neuroth, Heike et al. (Hrsg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg : Verlag Werner Hülsbusch. 2012. S. 41-49.
- TBA (2012). TBA Technology Based Assessment.
<http://tba.dipf.de/de/>; Rev. 2012-08-22.
- TextGrid (2012). TextGrid – Virtuelle Forschungsumgebung für die Geisteswissenschaften.
<http://www.textgrid.de/>; Rev. 2012-08-22.
- TGN (2012). Getty Thesaurus of Geographic Names.
<http://www.getty.edu/vow/TGNSearchPage.jsp>; Rev. 2012-08-22.
- VERA (2012). Vergleichsarbeiten in der Grundschule Projektgruppe VERA am Zentrum für empirische pädagogische Forschung.
<http://www.uni-landau.de/vera/>; Rev. 2012-08-22.
- Videodatenbank Schulunterricht in der DDR (2012). Videodatenbank Schulunterricht in der DDR.
<http://www.fachportal-paedagogik.de/filme/>; Rev. 2012-08-22.
- W3C (1997). Date and Time Formats (1997-09-15).
<http://www.w3.org/TR/NOTE-datetime>; Rev. 2012-08-22.
- W3C (1998). Web Architecture: Extensible Languages. W3C Note (1998-02-10).
<http://www.w3.org/TR/NOTE-webarch-extlang>; Rev. 2012-08-22.
- W3C (2004). World Wide Web Consortium. Resource Description Framework (RDF) (2004-02-10).
<http://www.w3.org/RDF/>; Rev. 2012-08-22.
- W3C (2008). Extensible Markup Language (XML) 1.0 (Fifth Edition). W3C Recommendation (2008-11-26).
<http://www.w3.org/TR/REC-xml/>; Rev. 2012-08-22.
- W3C (2012). World Wide Web Consortium. Naming and Addressing: URIs, URLs.
<http://www.w3.org/Addressing/>; Rev. 2012-08-22.
- Wissenschaftsrat (2012). Wissenschaftsrat: Empfehlungen zur Weiterentwicklung der wissenschaftlichen Infrastrukturen in Deutschland bis 2020 (2012-07-13).
<http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf>; Rev. 2012-08-22.
- Wittenburg (2012). Wittenburg, Peter / Drude, Sebastian / Broeder, Daan: Psycholinguistik. In: Neuroth, Heike et. al. (Hrsg.): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg : Verlag Werner Hülsbusch. 2012. S. 83-109.
- Zeng (2008). Zeng, Marcia Lei / Qin, Jian: Metadata, London : Facet publishing. 2008.

8 Anhang

Anlage 1: Screenshot des aktuellen Rohentwurfs der Portalseite Forschungsdaten Bildung (Stand: 02.08.2012)

Fachportal Pädagogik - Herzlich Willko... +

http://www.fachportal-paedagogik.de/forschungsdaten_bildung/

Bildungsserver · Fachportal · FIS · pedocs · DAQS · Schulweb · Edutags · ELIXIER · IWWB · Innovationsportal · Bildung+ · ISM · ZDB · BWW · LID · Bildungsbericht · Schulmediothek · HBO · Wikis · Blog

FACHPORTALpaedagogik.DE

DIPF

► Startseite ► Forschungsdaten Bildung ► Startseite ► Herzlich Willkommen im Portal Forschungsdaten Bildung

► **Metasuche**

► **FIS Bildung**

► **pedocs**

► **Themenkatalog**

► **Forschungsdaten Bildung**

Startseite

Medienarchiv

Fragebogenarchiv

Studienliste

Service

Wir über uns

Impressum

Herzlich Willkommen im Portal Forschungsdaten Bildung

Das Portal Forschungsdaten Bildung ist ein Angebot des **Deutschen Instituts für Internationale Pädagogische Forschung** zur Dokumentation von Studien der empirischen Bildungsforschung inklusive der erhobenen Forschungsdaten und eingesetzten Erhebungsinstrumente. Neben der Dokumentation von Fragebogen im Rahmen der Datenbank zur Qualität von Schule (DaQS) liegen auch audiovisuelle Daten vor. Diese können in Form von anonymisierten Transkripten und Codierschemata - sofern vorhanden - frei zugänglich eingesehen werden. Die Freigabe der AV-Medien für eine wissenschaftliche Nachnutzung ist aus datenschutzrechtlichen Gründen an eine **Registrierung** unter Angabe eines begründeten Forschungsinteresses gebunden. Das Angebot befindet sich im Aufbau und wird kontinuierlich erweitert.

Zu den Inhalten dieses Portals bieten sich Ihnen drei mögliche Zugänge:

- Das **Medienarchiv** eröffnet Ihnen einen direkten Zugang zu Video- und Audiodateien zum Thema Unterrichtsqualität.
- Das **Fragebogenarchiv** leitet sie zur **Datenbank zur Qualität von Schule (DaQS)** weiter, die Fragebogeninstrumente aus dem Bereich Schul- und Unterrichtsqualität bereitstellt.
- Die **Studienliste** gibt Ihnen einen Überblick über die dokumentierten Studien.

Kontakt

Sie haben eine Frage zum Angebot, möchten uns ein Feedback geben oder eine Studie dokumentieren lassen? Dann nehmen Sie bitte Kontakt mit uns auf.

Doris Bambey (Koordination Forschungsdaten Bildung)
Tel. +49 (0) 69.247 08 - 332

E-Mail an: forschungsdaten-bildung@dipf.de

▲ nach oben

► Startseite ► Kontakt ► Impressum

Anlage 2: Vorläufiger Entwurf des Metadatensets für AV-Daten für das Forschungsdatenzentrum „Forschungsdaten Bildung“ des DIPF*

Bezeichnung	Element	Refinement	Encoding Scheme
Abstract	Description	Abstract	Zusammenfassung der Inhalte (je nach Zuordnung Filmdokument, Studie etc.).
Anmerkungen	Description	Notes	Anmerkungen zum Datensatzinhalt, kommentarartige Notes (z.B. zu Begleitumständen der Aufnahme).
Aufnahmedatum	Date	created	Entstehungsdatum der Ressource nach ISO 8601.
Dateiformat	Format		Angabe zum vorliegenden Format, z.B. mpeg, wma etc.
Dateigröße	Format	Extent	Angabe in numerischen Zahlen in Byte (Literal, Bsp. 434444000).
Datenherkunft	Provenance		Aussage zur Provenienz, also der forschenden Instanz, evtl. bei räumlichen Einordnungsmöglichkeiten verknüpft mit Geokoordinaten nach TGN.
Dateiname	Identifizier	Filename	Eineindeutig identifizierbarer Name der downloadbaren Dateien.
Dateityp	Type		Drei Optionen ausreichend: Audio; Video; Bilddatei.
Dauer	Format	Extent	Angabe nach ISO 8601 über Dauer, in hh:mm:ss.
Eingangsdatum	Date	DateAccepted	ISO 8601, JJJJ-MM-DD.
Gattung des Films	Subject	Genre	Index angestrebt; Orientierung am FIS-Thesaurus und bestehenden Indexlisten.
ID	Identifizier		Eineindeutig zuteilbarer Name in Form eines Persistent Identifiers, für jeden Datensatz einzeln zu vergeben.
Identisches Dokument	Relation	ConformsTo	Falls mehrere Dokumente über ein Ereignis aufgezeichnet werden, kann hier eine Relation hergestellt werden (Stichwort Dubletten oder gleiche Unterrichtsstunden aus anderer Perspektive).
Literatur zum Dokument	Identifizier	BibliographicCitation	Eher optional die Angabe von direkt zu einem Dokument zuordnenbarer Literatur; eher üblich für die gesamte Studie.
Literatur zur Studie	Identifizier	BibliographicCitation	Literaturangaben zur kompletten Studie.
Medientyp	Format	Medium	Bisher Online-Ressource.
Methodik	Method		Weiter zu beschreiben, allerdings bei AV-Daten nicht relevant, sondern im Kontext der Projektbeschreibung, also des Projektdatensatzes. Hier Orientierung an DDI,
Nutzungsbedingungen	Rights	accessRights	Informationen über Zugangsberechtigung in Form von formulierten Nutzungsbedingungen (entweder verlinkt oder formuliert).

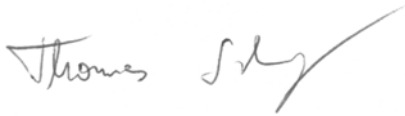
Nutzungserlaubnis	Rights	License	Sollte die Erlaubnis zu einer Verwendung außerhalb des Forschungsprojektes vorliegen, kann dies hier dokumentiert werden.
Produktionsort	Coverage	Spatial	Genaue offizielle Verzeichnung der Produktionsstätte (Institution) unter Einbindung von Geokoordinaten nach TGN.
Projekt	Relation	Source	Offizieller Titel des Projekts als Verweis auf Studie; bei der Abbildung einer ganzen Studie muss hier eine neue Zuordnung erteilt werden.
Projektleitung	Creator		Es könnte hier theoretisch auch der Aufnahmeleiter genannt werden. Da es sich allerdings um wissenschaftliche Projekte mit klarer Zuständigkeit handelt, wird hier der / die Projektleiter namentlich aufgeführt. Evtl. ist bei Creator auch nach MPEG7 in Form von Role (Author, Actor etc.) zu spezifizieren. Bisher wird dieser Bedarf allerdings noch nicht gesehen.
Publiziert von	Publisher		Publikation der Studie, also Nennung des Instituts oder des/der Publizierenden.
Rechteinhaber	RightsHolder		Person oder Institution mit den Rechten an dem Dokument, auch in Rechtsnachfolge, kombiniert mit geographischen Angaben (TGN).
Schlagwörter	Subject	Keyword	Kontrolliertes Vokabular; bisher nur aus FIS-Thesaurus. Ein Schlagwortindex ist bisher geplant.
Sprache der Aufzeichnung	Language		Eintragung nach ISO 639.2; hier kann bei reinen Unterrichtsfilmen auch eine Umbenennung in "Sprache der Unterrichtseinheit" erfolgen.
Systemanforderungen	Relation	Requires	Software (Videoplayer, Cutting-Software, Codierungssoftware), Arbeitsspeicherbedarf etc.
Thema	Subject		Klare Orientierung an Oberbegriffen, thematische Zuordnung; Unterrichtsgegenstand, aber auch Studiengegenstand, je nach Bedarf. Orientierung an FIS-Thesaurus bedingt nur möglich.
Titel	Title		Entweder Übernahme des Titels des Originaldokuments / der offiziellen Studientitulierung oder, wenn nicht vorhanden, selbst zu wählen.
Unterrichtsfach	Subject	Keyword	Kontrolliertes Vokabular aus FIS-Thesaurus (Bsp.: Deutsch).
Untersuchungszeitraum	Coverage	Temporal	Die Zeitspanne der Aufzeichnungsphasen nach ISO 8601, JJJJ-MM-DD - JJJJ-MM-DD.
Verweis	Relation	IsReferencedBy / References	Verweis oder Zitierung anderer Ressourcen.
Veröffentlichungsdatum	Date	Issued	Wie bei Publisher nur bei Bedarf, Maßstab ISO 8601.
Wissenschaftliche Mitarbeit	Contributor		Die Ebene unterhalb der Projektleitung, sowohl fachwissenschaftlich-pädagogische als auch dokumentarische Bearbeitung.
Zielgruppe	Audience		Forscher / Forschungsbereich.

Zubehör	OtherMaterial		Weiter aufzufächern u.a. durch Description, Format, Type, Relation etc. OtherMaterial ist ein komplettes Modul, das allerdings keine AV-Daten enthält und deswegen nicht genauer spezifiziert wird. Hier lassen sich aber Unterrichts- und Begleitmaterialien wie Tafelbilder, Sitzpläne, Transkripte etc. verorten.
Zuletzt bearbeitet am	Date	Modified	Hier muss das letzte Bearbeitungsdatum angeführt werden nach ISO 8601-Norm.

* Hierbei handelt es sich um eine erste Entwicklungsstufe, die in vielerlei Hinsicht noch Modifizierung bedarf. Auch kann auf Grund fehlender Erkenntnisse über Datenbestände und Anforderungen von Forschungsseite keine endgültige und vor allem vollständige Festlegung erfolgen.

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst habe und dass keine anderen als die angegebenen Hilfsmittel benutzt wurden.

A handwritten signature in black ink, appearing to read "Thomas Schy". The signature is written in a cursive style with a long, sweeping flourish at the end.

Frankfurt, den 24.08.2012