

Metadaten

Jörn Sieglerschmidt, Mannheim

Metadaten meint auf den ersten Blick Daten über Daten,¹ entspricht also nicht ganz dem ursprünglichen griechischen Gebrauch, der eher einem Dazwischen oder Herüber entsprach. An diese Bedeutung knüpfen noch heute einige Wortbildungen an, z. B. Metapher. Vor allem im philosophischen Sprachgebrauch hat sich die Bedeutung der Letztbegründung einer Sache etabliert: Metalogik, Metatheorie, Metaphysik.² Gerade am Begriff der Metaphysik läßt sich zeigen, wie derjenige der Metadaten methodisch zu verstehen sein könnte. Bereits im Abschwung der philosophischen Beschäftigung mit der Metaphysik, ja ihrer Diskreditierung als reine, d.h. hier: müßige Spekulation hat KANT den Versuch einer Neubegründung der Metaphysik gemacht.³ Das Ergebnis ist die in den *Prolegomena* zu findende Kurzfassung des Argumentes, das er in der ersten *Kritik* entwickelt hat. Dort stellt er nochmals die Frage nach den Bedingungen der Möglichkeit von Erkenntnis bzw. im engeren Sinne: wissenschaftlicher Erkenntnis. Das Ergebnis seiner Überlegungen ist bekannt: Erkenntnis ist nur möglich, wenn vorausgesetzt wird, daß es kategoriale Begriffe gibt, die jeder Erfahrung vorausliegen, apriori gegeben sind und das begründen, was als transzendentaler Idealismus bezeichnet wird. Kategoriale Begriffe dieser Art sind z. B. Raum und Zeit, ohne die empirische Erkenntnis nicht möglich wäre. Methodisch wäre also in Hinsicht auf Daten zu fragen, was ihnen kategorial vorausliegt, um Daten überhaupt zu ermöglichen. Das wäre ein philosophisch begründeter Begriff von Metadaten, der hier aber –das wird sofort deutlich– viel zu weit greift. Es geht beim Begriff der Metadaten mehr um die Normierung des Gebrauchs, also etwa das, was WAIDACHER als Funktion der Metamuseologie im Verhältnis zur Museologie (theoretische wie praktische) definiert hat.⁴ Es geht also um Standards, die Kommunikation ermöglichen sollen. Damit sind wir von einer philosophischen Begründung weit entfernt. Standards sind ausschließlich eine Frage der Konvention.

Bei Metadaten handelt es sich also eher um eine Sprache, die versucht, den aktuellen sprachlichen Gebrauch zu normieren. Auf dieser metaphorischen Ebene kann deutlicher gemacht werden, was unter Metadaten zu verstehen sein könnte. COULMAS hat in Hinsicht auf die visuelle Fixierung des Wortes davon gesprochen, daß diese die Standardisierung seiner Form notwendig und die Beobachtung der Abweichung von einem vorherigen Zustand erst möglich mache.⁵ Genau das ist die Funktion von Metadaten. Die folgende kleine Übersicht versucht, das in Analogie zur Sprache schematisch zu fassen:

Daten	Semantik	Inhalt
Datenstruktur	Syntax/ Grammatik	Gebrauchsregeln
Metadaten	Orthographie	Aufschreiberegeln

¹ vgl. die gute Einführung unter: <http://www.onlineinc.com/onlinemag/metadata>

² Vgl. die unterschiedlichen Artikel in: MITTELSTRAß, JÜRGEN/ WOLTERS, GEREON (Hgg.): Enzyklopädie Philosophie und Wissenschaftstheorie. Bd. 2. Mannheim/ Wien/ Zürich: Bibliographisches Institut 1984, 864-876.

³ Kant, Immanuel: Prolegomena zu einer jeden künftigen Metaphysik die als Wissenschaft wird auftreten können. In: ders.: Werke in sechs Bänden. Band 3. Darmstadt: Wiss. Buchgesellschaft 1963, 111-264 [Erstveröff.: Riga: Johann Friedrich Hartknoch 1783], hier 114, weist auf die Probleme der Beschäftigung mit der Metaphysik hin: "Auch haben sich ihre Anhänger gar sehr verloren, und man siehet nicht, daß diejenigen, die sich stark genug fühlen, in andern Wissenschaften zu glänzen, ihren Ruhm in dieser wagen wollen, wo jedermann, der sonst in allen Dingen unwissend ist, sich ein entscheidendes Urteil anmaßt, weil in diesem Lande in der Tat noch kein sicheres Maß und Gewicht vorhanden ist, um Gründlichkeit von seichem Geschwätz zu unterscheiden."

⁴ Waidacher, Friedrich: Handbuch der Allgemeinen Museologie. Wien/ Köln/ Weimar: Böhlau 1993, 29.

⁵ Coulmas, Florian: Über Schrift: (suhrkamp taschenbuch wissenschaft 378) Frankfurt am Main: Suhrkamp 1981, 26.

Daten umfassen den Inhalt der Mitteilung, sind also nach semantischen Regeln zu analysieren. Die Datenstruktur ist nicht vollkommen vom Inhalt zu trennen, da die Gebrauchsregeln auf den Inhalt einwirken, wie auch der Inhalt auf Gebrauchsregeln wirkt. Grammatik und im engeren Sinne die Syntax präformieren die Semantik, determinieren sie aber nicht. Das Argument von COULMAS geht nun da hin, daß auch Aufschreiberegeln die Grammatik (sowieso) und die Semantik präformieren. In Hinsicht auf die Sprache wird diese These hier nicht weiter verfolgt. Doch in Hinsicht auf digitale Daten und ihre Analogie zur Sprache, stellt sich die Frage, ob Metadaten die Datenstruktur und Daten präformieren bzw. gar determinieren. Wenn dem hier vorgeschlagenen metaphorischen Gebrauch des Metadatenbegriffes gefolgt wird, so stellen Metadaten die Oberfläche, das Superfizielle (facies – Gesicht) der Daten dar.

Metadaten transportieren Inhalt, sind aber nicht der Inhalt selbst. Wenn Metadaten inhaltlich verstanden würden, käme das Problem der Letztbegründung wieder ins Spiel, denn dann müßte begründet werden, warum bestimmte Daten, eben Metadaten, anderen vorgeordnet sind, weil ihr Verstehen die Bedingung der Möglichkeit von Daten ist, d. h. das Verständnis von Daten erst ermöglicht. Es ist daher sinnvoller, pragmatisch von der normierenden und damit Datenaustausch ermöglichenden Funktion von Metadaten auszugehen. Ein Beispiel für den pragmatischen Einsatz von Metadaten ist die Definition von maschinenlesbaren Austauschformaten für Bibliotheken (MARC, MAB). Es handelt sich dabei um die Definition einer Datenstruktur, die immer nur eine Teilmenge differenzierter Datenerfassung sein kann, einschließlich der für die – in die Datenstruktur einzutragenden – Inhalte geltenden Aufschreiberegeln. Es muß z. B. definiert werden, wie ein Datum (9.9.99) zu erfassen ist, damit es von einem datenverarbeitenden System richtig gelesen werden kann. Aber nicht nur die komplex strukturierte Datenerfassung kann durch solche Metadaten nicht vollständig abgebildet werden, auch die gesamten Verarbeitungsfunktionen, das was in Bibliotheken zuweilen als Lokalsystem bezeichnet wird, sind auf Datenbankfunktionen angewiesen, die über die Möglichkeiten vernetzter Systeme bisher hinausgehen. Metadaten dienen also der Normierung von Grammatik/Syntax und Orthographie, um Daten bzw. Inhalte ohne Verständigungsschwierigkeiten austauschen oder in vernetzten Umgebungen finden zu können.

Daten sind in Wort (graphisch), Bild (visuell) und/ oder Ton (akustisch) niedergelegte Informationen zu Objekten jeglicher, d. h. auch geistiger Art. Metadaten müßten dann etwas aussagen über die Anordnung und Aufteilung der aufzubewahrenden Daten, also über die Schreibweisen, mit denen ein Lesesystem die Informationen als solche identifizieren kann. Die Datenstruktur umfaßt die Definition von Datenfeldern. Sie stellen gewissermaßen die Grammatik und Syntax eines Informationssystems dar, mit denen Informationen bereits vorstrukturiert werden, indem ggf. auch Relationen zwischen Elementen der Datenstruktur definiert sind. Außerdem muß eine gewisse Orthographie beachtet werden. Aufschreiberegeln dieser Art gehören zu den Datenfelddefinitionen, die – für die hier vorgeschlagene Verwendung mißverständlich – üblicherweise als Syntax bezeichnet werden. Syntaxkontrollen prüfen daher die Richtigkeit der Schreibweise von Informationen. Definitionen von Datenstruktur und Aufschreiberegeln müssen sich auf Metadaten beziehen, um wiedererkennbar zu sein.

Solche Metadaten sind in vielerlei Anwendungsgebieten verbreitet, meist so, daß die Nutzer/innen eines Systems dieses garnicht erkennen.⁶ Denn Datenverarbeitungssysteme sind darauf angelegt, Informationen sequentiell abzuarbeiten. Physikalisch gesehen gibt es zwar Möglichkeiten, diese sequentielle Suche nach Textmarken zu beschleunigen, indem z. B. mit festen Satz- und Blocklängen gearbeitet wird, doch ist ein solches Verfahren außerordentlich speicherplatzaufwendig und wird daher vermieden. Überall also, wo Daten verarbeitet werden, muß mit Metadaten gearbeitet werden, die dem System sagen, welche Operationen mit den folgenden Informationsstrings auszuführen sind und welche nicht.

⁶ Krüger, Manfred: SGML – Praxis eines langen Weges. In: Möhr, Wiebke/ Schmidt, Ingrid (Hgg.): SGML und XML. Anwendungen und Perspektiven. Berlin/ Heidelberg: Springer 1999, 54f.

Es ist daher bereits häufiger darauf verwiesen worden, daß Metadaten u. a. in allen Textverarbeitungssystemen verbreitet sind. Wenn bestimmte Funktionen wie Schriftarten, Textauszeichnungen, Textgliederungen usw. aufgerufen werden, müssen die entsprechenden Textpassagen mit Marken, sog. *tags*, versehen werden, die vom jeweiligen System angemessen interpretiert werden. Sichtbar ist das heute noch bei Systemen wie TUSTEP, die komplexe und flexible Textverarbeitung ermöglichen, dafür aber der/ dem Nutzer/in die Definition der notwendigen Textmarken aufbürden. Ein anderes, seit einigen Jahren wesentlich bekannteres Beispiel sind Sprachen wie SGML,⁷ deren Derivat HTML und deren Erweiterung XML, die nicht Texte, sondern auch Ton- und Bildinformation für die Verarbeitung strukturieren können.

Metadaten strukturieren inhaltliche Information so, daß sie im Netz unabhängig von *hard-* und *software-*Voraussetzungen genutzt werden kann. Personen und Institutionen, die Informationen im Netz anbieten, können auf diese Weise die Recherchen der Internet-Suchmaschinen unterstützen. Es zeigt sich nämlich, daß die Ergebnisse von Recherchen im Internet sehr unterschiedlich, häufig nicht zufriedenstellend, da zu unspezifisch ausfallen. Das liegt nicht nur an den unterschiedlichen Konzepten der genutzten Suchmaschinen,⁸ sondern auch an der angebotenen Information. Was die Suchmaschinen betrifft, so ist auf die Definitionsmacht von Suchmaschinen hingewiesen worden, die nach Klassifikationsschemata vorgehen, z. B. *Yahoo*, und Inhalte nicht nur indexieren, wie z. B. *Altavista*.⁹ Metadaten bringen eine Struktur in die Information, die von Suchmaschinen entsprechend genutzt werden kann so, daß der Erzeuger der Information in erheblichem Maße darüber entscheiden kann, wer seine Informationen wie im Netz findet.

David Bearman hat bereits 1997 darauf hingewiesen, daß Museen die Möglichkeit haben, ihre Internetpräsenz mit Dublin Core zu strukturieren und damit ihre Sammlungen einem größeren Publikum bekannt zu machen. Der Vorteil ergibt sich für ihn –trotz mancher Probleme der Beschreibung von Museumsdaten mit Dublin Core– gerade durch die Integration von Suchvorgängen nach Objekten anderer Institutionen mit den Museumsobjekten.¹⁰ Genau das ist die gerade jetzt von der EU unterstützte Initiative gemeinsamer Projekte von Bibliotheken als den fortgeschrittensten Netzanwendern, mit Archiven und Museen (daher BAM-Projekte genannt). Der Internetnutzer sollte bei der Suche nach bestimmten Wörtern oder noch neutraler: Zeichenfolgen aus den drei genannten Institutionen gezielt Informationen erhalten können über das hinaus, was evtl. bereits jetzt verfügbar ist.

Von den unterschiedlichen Beschreibungen der Erfolgsgeschichte der DC-Entwicklung soll keine an dieser Stelle nochmals erzählt werden.¹¹ Das Ergebnis der mehrjährigen internationalen Diskussion ist die Definition des DC Element Set sowie seiner Erweiterungen um qualifizierende Angaben (*qualifiers*), d.h. Subelemente (SUBELEMENTS), Angaben zur jeweils verwendeten Sprache (LANGUage) und zu Hilfsmitteln der Sacherschließung (SCHEME). Eine Übersicht gibt die vorliegende Tabelle, in der die Definitionen der autoritativen Quellen ebenfalls zitiert sind. Die einzelnen Elemente sollen nicht erläutert werden, da das in

⁷ ISO 8879; über die näheren Hintergründe der Entwicklung von SGML besonders in Europa als Werkzeug für die strukturierte Textverarbeitung informiert Krüger 1999 [wie Anm.4], 51-76.

⁸ Dazu Thiele, Harold: The Dublin Core and Warwick framework. A review of the literature, march 1995-September 1997. In: D-Lib Magazine January 1998.

⁹ Grötschel, Martin/ Lügger, Joachim: Scientific Information Systems und metadata. ZIB report SC-98-26 October 1998.

¹⁰ Bearman, David: Standards – Museum metadata integrated with the metadata of other distributed information resources. In: Qualität und Dokumentation. CIDOC Jahrestagung 1997. Nürnberg: Germanisches Nationalmuseum 1998, 18-7ff.

¹¹ Informativ die Beschreibung von Rusch-Feja, Diann: Auf dem Weg zur Entwicklung eines Internet-Standards – 4. Dublin Core Metadata Workshop in Canberra. In: Bibliotheksdienst 1997, Heft 4. Zugänglich auch über: <<http://www.mpib-berlin.mpg.de/dok/gmr/>>, wo sich weiteres Material zu Dublin Core befindet, u. a. der Bericht über DC5, den fünften Dublin Core Workshop in Helsinki 1997. Deutsche Projekte sind zu finden unter: <<http://www.bsz-bw.de/diglib/medserv/>> (Bibliotheksservicezentrum Konstanz); <http://www2.sub.uni-goettingen.de/> (Staats- und Universitätsbibliothek Göttingen).

zahlreichen Publikationen bereits geschehen ist und nicht nochmals wiederholt werden muß.¹² Der *Guide to best practice* weist außerdem darauf hin, daß die Abfolge der Elemente vollkommen unerheblich ist.¹³ Wichtig ist, daß in den letzten Jahren der notwendige Weg der Einbettung der DC-Elemente in die entsprechende HTML (XML)-Umgebung erfolgt ist. Das unten folgende Beispiel bezieht sich auf diese syntaktische Umgebung.

Im ersten und fünften Dublin Core Workshop sind Eigenschaften definiert worden, die wesentlich für die Entwicklung waren und auch der weiteren Entwicklung zugrundeliegen sollten:

- Intrinsicity (immanente Verstehbarkeit),
- extensibility (Erweiterbarkeit),
- syntax independence (Syntaxunabhängigkeit),
- optionality (Optionalität),
- repeatability (Wiederholbarkeit),
- modifiability (Veränderbarkeit),
- simplicity (Einfachheit),
- semantic interoperability (Austauschbarkeit mit anderen Metadatendefinitionen),
- international consensus (internationale Übereinstimmung),
- flexibility (Flexibilität).

Bei den unterschiedlichen Projekten haben sich jedoch Anwendungsprobleme gezeigt, die zu einer Überarbeitung von Struktur und Auslegung des Dublin Core geführt haben. Es ging um die eben bereits erwähnten qualifizierenden Angaben. Grundsätzlich sind solche Änderungen diskutiert worden vor dem Hintergrund einer Spaltung in Minimalisten und Strukturalisten, wobei erstere ausschließlich die 15 DC-Elemente in einfacher Form beibehalten wollten, letztere die Erweiterungen als notwendig erachteten, um Dublin Core überhaupt zu retten. BAKER hat von der Alternative zwischen *pidginization* und *creolization* gesprochen.¹⁴ Es gilt inzwischen als akzeptiert, daß beide Wege richtig sind, jeweils auf unterschiedliche Funktionen bezogen.

Die *qualifier*, durch die einzelne Elemente ergänzt werden können sind: LANGUage, SUBELEMENT und SCHEME. Für die Werte von Dublin-Core-Elementen muß anzugeben sein, in welcher Sprache sie eingetragen sind, um richtig interpretiert werden zu können. Der *qualifier* LANG dient also der Internationalisierung von Dublin Core in besonderer Weise. SUBELEMENT dagegen erlaubt es, ein vorhandenes Element –im Augenblick vor allem die Angaben zum Verfasser oder Urheber (*creator*), zum Datum (*date*), zur Objektbeziehung (*relation*) und zu weiteren Bezügen (*coverage*)– um präzisierende Angaben zu ergänzen. Es soll in keinem Fall gestattet sein, darüber hinauszugehen in der Absicht, auf diese Weise Dublin Core um neue Elemente zu erweitern. Schließlich ist es wichtig, bei einzelnen Elementen sagen zu können, daß sich der dort eingetragene Wert auf ein Standardwerk für die Sacherschließung bezieht. Dafür ist der *qualifier* SCHEME gedacht.¹⁵

In der vom März diesen Jahres stammenden *draft version* des *Guide to best practice: Dublin Core* von CIMI werden zu Anfang die Probleme aufgezählt, die Museumsanwender/innen mit der einfachen Version des Dublin Core haben wegen:

- der einmaligen Betonung der Eigenschaften physischer Objekte;

¹² Eine Einführung in das Verständnis der einfachen, d. h. nicht um *qualifier*, *sub-elements* und *language* erweiterten Dublin Core Elemente gibt der *User Guide*, der über die Dublin-Core-Haupseite und die Seite der entsprechenden *Working Group* zugänglich ist: <http://purl.org/dc/core/documents/working_drafts/wd-guide-current.htm>.

¹³ Consortium for the Computer Interchange of Museum Information: *Guide to best practice: Dublin Core* (DC1.0=RFC2413). Version 0.3.1: 30 March 1999.

¹⁴ Baker, Thomas: *Languages for Dublin Core*. In: D-Lib Magazine December 1998.

¹⁵ In Deutschland wären das: PND (Personennamendatei der DDB), GKD (Gemeinsame Körperschaftsdatei), SWD (Schlagwortnormdatei). In den USA: LCSH (Library of Congress Subject Headings), DDC (Dewey Decimal Classification), AAT (Arts and Architecture Thesaurus), LCNAF (LoC Name Authority File) usw.

- dem Wunsch, diese physischen Objekte mit Personen, Orten und Ereignissen zu verknüpfen;
- der notwendigen Verantwortung für Sammlungen;
- der notwendigen Verantwortung für Objektsurrogate wie Photos;
- des historisch erklärbaren Fehlens von Standards zur Sacherschließung.

Wer die Elemente genau durchmustert, wird sicherlich zahlreiche Punkte finden, an denen zu fragen ist, ob denn Dublin Core für den Museumsgebrauch geeignet ist¹⁶. Mit Sicherheit kann Dublin Core kein Ersatz für eine gut strukturierte Datenbank zur wissenschaftlichen Erschließung und Verwaltung von Objekten sein, zumindest ist das weder bei weiterer Entwicklung absehbar noch intendiert. Wesentliche Aufgabe einer DC-kompatiblen Beschreibung musealer Objekte und ihrer Derivate ist ihre Präsenz im Netz und damit die Präsenz bildungs- und forschungsrelevanter Information aus den Museen für ein großes Publikum. Wer also seine Objekte in einer aufwendig strukturierten Datenbank erschließt, kann eine Teilmenge der dort gesammelten Information für den Netzbetrieb aufbereiten. Die gründliche Erschließung selbst wird er nicht im Rahmen der Möglichkeiten von DC und einer ML leisten können.

Es soll abschließend versucht werden, ein museales Objekt mithilfe des DC zu beschreiben, um deutlich machen, wie einfach die Einbettung musealer Objektinformation in diese Umgebung ist und wie weit die Verwaltung von Objektinformation mit Mitteln des DC und einer ML gehen kann.

¹⁶ Vgl. die im Anhang zu findenden Vergleiche zwischen dem Datenfeldkatalog des Deutschen Museumsbundes bzw. dem General International Standard of Archival Description (ISAD(G)) und dem Dublin Core Element Set. Die deutsche Übersetzung zur Version 1.1 des DCES ist zugänglich unter: <http://www.mpib-berlin.mpg.de/metadaten/gmr/metatagd.htm>.

```
<?xml version="1.0"?>
<dc-record>
<type>physisches Objekt</type>
<type>Original</type>
<type>Sammlung</type>
<type>Kultur</type>
<title>Bleistift TK9500</title>
<description>Sechseckig geformter Zeichenstift mit Bleimine</description>
<description>Abmessungen Höhe: 15 cm, Länge: 0,6 cm, Breite 0,6 cm</description>
<subject>Zeichenstift; Bleistift; Zeichnen</subject>
<subject>Kunststoff; Eisen; Blei</subject>
<creator.corporate name>Faber-Castell</creator.corporate name>
<date.created>1987</date.created>
<coverage>Nürnberg </coverage>
<coverage>Franken</coverage>
<coverage>Deutschland</coverage>
<rights>Darf nicht ausgestellt werden.</rights>
</dc-record>
```

Englische und deutsche (Kurz)Fassung der Dublin Core Elemente (Version 1.1)

(Jörn Sieglerschmidt)

Name/ Identifier	Definition	Name/Bezeichnung	Definition
Title/ Title	A name given to the resource.	Titel	Der an die Ressource vergebene Name.
Creator/ Creator	An entity primarily responsible for making the content of the resource.	Verfasser oder Urheber	Das- oder derjenige, der in erster Linie für die Erstellung des Inhalts der Quelle verantwortlich ist.
Subject and Keywords/ Subject	The topic of the content of the resource.	Thema und Stichwörter	Das Thema des Inhalts einer Quelle/Ressource.
Description/ Description	An account of the content of the resource.	Inhaltliche Beschreibung	Eine Darstellung des Inhalts einer Ressource.
Publisher/ Publisher	An entity responsible for making the resource available.	Verleger oder Herausgeber	Die Instanz, die für die Veröffentlichung bzw. Verfügbarkeit der Ressource verantwortlich ist.
Contributor/ Contributor	An entity responsible for making contributions to the content of the resource.	Weitere beteiligte Personen u.Körperschaften	Eine Instanz, die einen Beitrag zum Inhalt der Ressource macht.
Date/ Date	A date associated with an event in the life cycle of the resource.	Datum	Ein Datum, das ein Ereignis im Lebenszyklus der Ressource darstellt.
Resource Type/ Type	The nature or genre of the content of the resource.	Ressourcenart	Die Art bzw. das Genre des Inhalts der Ressource.
Format/ Format	The physical or digital manifestation of the resource.	Format	Die physische oder digitale Festschreibung der Ressource.
Resource Identifier/ Identifier	An unambiguous reference to the resource within a given context.	Identifikation der Ressource	Eine eindeutige Bezeichnung der Ressource in einem bestimmten Kontext.
Source/ Source	A Reference to a resource from which the present resource is derived.	Quelle	Ein Hinweis auf Quellen, von der die vorliegende Ressource abgeleitet werden kann.
Language/ Language	A language of the intellectual content of the resource.	Sprache	Die Sprache(n) des intellektuellen Inhalts der Ressource.
Relation/ Relation	A reference to a related resource.	Beziehung zu anderen Ressourcen	Eine Beziehung zu einer verwandten Ressource.
Coverage/ Coverage	The extent or scope of the content of the resource.	abgedeckter räumlicher u. zeitlicher Bereich	Die Reichweite oder der Anwendungsbereich des Inhalts der Ressource.
Rights Management/ Rights	Information about rights held in and over the resource.	Rechteverwaltung	Informationen über die Rechte, die in der Ressource und diese betreffend vorhanden sind.

Quellen: <http://purl.org/dc/documents/rec-dces-19990702.htm>
<http://www.mpib-berlin.mpg.de/dok/metatagd.htm>