

Google michelangelo

Alle Bilder Videos News Bücher Mehr Einstellungen Tools

Ungefähr 35.100.000 Ergebnisse (0,49 Sekunden)

Michelangelo – Wikipedia
<https://de.wikipedia.org/wiki/Michelangelo>
 oft nur Michelangelo (vollständiger Name Michelangelo Buonarroti [mikeˈlandʒelo buoˈnarˈroti]); geboren als Michelangelo di Lodovico Buonarroti Simoni; * 6. März 1475 ...
 David · Kategorie:Michelangelo · Sixtinische Kapelle · Fresko

Michelangelo - Wikipedia
<https://de.wikipedia.org/wiki/Michelangelo> · Diese Seite übersetzen
 Michelangelo di Lodovico Buonarroti Simoni or more commonly known by his first name Michelangelo was an Italian sculptor, painter, architect and poet of the ...
 Notable work: David, Pieta, The Last Judgment; ... Movement: High Renaissance
 Known for: Sculpture, painting, architecture, and ...

Michelangelo Buonarroti - Maler, Bildhauer, Architekt und Dichter der ...
<https://www.opplaworld.de/zeit/michelangelo/>
 Michelangelo wird am 6. März 1475 als zweiter Sohn des Florentiner Beamten Ludovico di Lionardo di Buonarroti Simoni (1444-1534) und der Francesca di ...

Michelangelo (Michelangelo Buonarroti) , Maler(Sixtinische Kapelle ...
www.art-michelangelo.de/
 Daten und Fakten: Michelangelo Maler, Bildhauer, Dichter, Architekt. Michelangelo Maler von Fresken, die berühmtesten Fresken befinden sich in der ...

Michelangelo – Klexikon - Das Freie Kinderlexikon
<https://klexikon.zum.de/wiki/Michelangelo>
 03.03.2018 - Michelangelo war ein italienischer Maler, Bildhauer, Architekt und Dichter. Er lebte von 1475 bis 1564. Er war der berühmteste italienische ...

Portfolio: Michelangelo Buonarroti - [GEO]
<https://www.geo.de> · Magazine · GEO Epoche Edition ·
 "Il divino" nennen ihn seine Bewunderer, den Göttlichen. Sein ganzes Leben widmet Michelangelo Buonarroti (1475-1564) der Kunst - und was für einer. Kein ...

Videos

LIDO : Geheimnis Michelangelo
BR - 23.10.2017

Michelangelo Biography: Who Was This Guy, Really? | Art History ...
Artrageous with Nate YouTube - 28.04.2016

Pieta (marble sculpture) (video)
Smarthistory, art... Khan Academy - 04.03.2015

Michelangelo
Maler

Michelangelo Buonarroti, oft nur Michelangelo, war ein italienischer Maler, Bildhauer, Baumeister und Dichter. Er gilt als einer der bedeutendsten Künstler der italienischen Hochrenaissance und weit darüber hinaus. Wikipedia

Geboren: 6. März 1475, Caprese Michelangelo, Italien
Gestorben: 18. Februar 1564, Rom, Italien
Epochen: Hochrenaissance, Italienische Renaissance, Renaissance, Manierismus
Bestattet: Santa Croce, Florenz, Italien
Bauwerke: Petersdom, Basilica di San Lorenzo, New Sacristy, MEHR

Kunstwerke Über 10 weitere ansehen

David 1504
Römische Pieta 1499
Decke der Sixtinischen Kapelle 1512
Das Jüngste Gericht 1541
Die Erschaffung Adams

Wird auch oft gesucht Über 15 weitere ansehen

Leonardo da Vinci
Raffael
Donatello
Sandro Botticelli
Tizian

Mehr zu Michelangelo

DEUTSCHE DIGITALE BIBLIOTHEK
 KUNST UND KULTUR

STARTSEITE ÜBER UNS JOURNAL HILFE

Link auf diese Seite

Michelangelo, Buonarroti
 Künstler, Bildhauer, Maler, Architekt, Schriftsteller, Lyriker
 Geboren: 6. März 1475, Caprese, Michelangelo
 Gestorben: 27. Februar 1564, Rom

Beteiligt an:

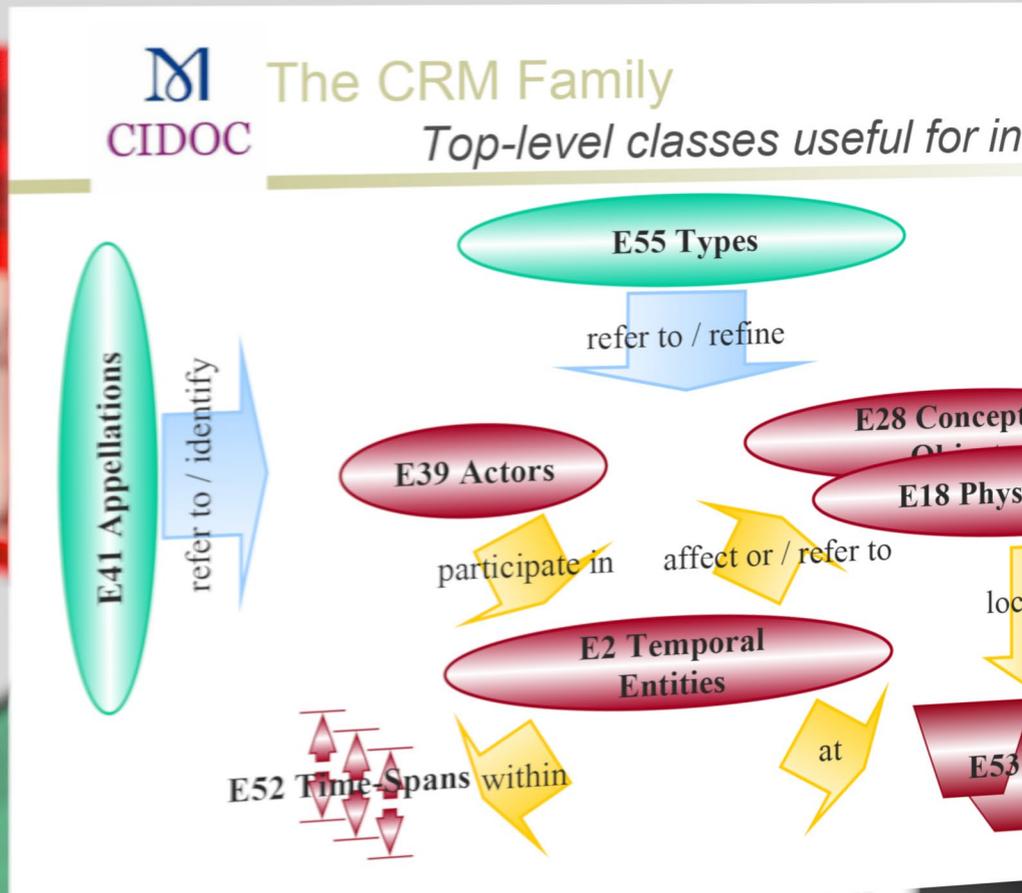
- Regel der Safford Order, Von der Architecture
- Vignola - Amortium, 1554
- Regola dell'Ungueto Ordini D'Architettura
- Vignola - Anheim, 1620
- Städtische Decke — Die Vorfahren Christi — Jakob und Joseph
- Büchel
- Städtische Decke — Propheten und Söhnen — Die Perische Skylla
- Büchel

Alle Objekte (228)

Thema in:

- Starbender Säule in Glas (umlegierte Zeichnung der von Michelangelo Buonarroti gefertigten Skulptur; Louvre, Paris)
- 22. Dezember 1905, Landesarchiv Sigmaringen, FAS SA A 7 T 1 Nachlass Xaver Henkel
- Rom: St. Peter (Pieta von Michelangelo)
- Landesarchiv Baden-Württemberg, Abt. Staatsarchiv Sigmaringen, FAS H 1/2 T 1 Nachlass Albert Waldenspul-Glasplatten, Diapositive und Alben
- Rom: Sixtinische Kapelle (Das Jüngste Gericht von Michelangelo, Detail)
- Landesarchiv Baden-Württemberg, Abt. Staatsarchiv Sigmaringen, FAS H 1/2 T 1 Nachlass Albert Waldenspul-Glasplatten, Diapositive und Alben
- Rom: Städtische Kapelle (Das Jüngste Gericht von Michelangelo)
- Landesarchiv Baden-Württemberg, Abt. Staatsarchiv Sigmaringen, FAS H 1/2 T 1 Nachlass Albert Waldenspul-Glasplatten, Diapositive und Alben

Alle Objekte (187)





Angela Kailus, Regine Stein

F. Besser vernetzt: Über den Mehrwert von Standards und Normdaten zur Bilderschließung

→ Informationszugang, semantische Interoperabilität, Linked Open Data, Normdaten, Koreferenzierung

Das dokumentarisch genutzte Bild hat einen Doppelcharakter: Die Kunstgeschichte verwendet es seit jeher als Stellvertreter ihrer Untersuchungsgegenstände, der Kunst- und Bauwerke, gleichzeitig aber ist das Bild ein eigenständiges Artefakt, dessen historische Kontextualisierung und Materialität immer stärker selbst zum Gegenstand wissenschaftlichen Interesses werden. Daraus erwächst die Anforderung, bei der Erschließung solcher Bildbestände klar differenzierend zu dokumentieren.

Die Zugänglichkeit zu solchen Forschungsmaterialien definiert sich heute über ihre Zugänglichkeit im Web. Die Erschließung und Aufbereitung von Objekten und ihren Abbildungen erfolgt bislang oft maßgeschneidert für spezifische Projektkontexte und/oder institutionelle Anforderungen. Die Publikation der gewonnenen Daten und Texte im Web ist daher häufig noch durch heterogene semantische Strukturen gekennzeichnet – wertvolle Informationen bleiben damit trotz Nutzung moderner Publikationsformate in unverbundenen Inselangeboten letztlich unzugänglich. Semantic Web-Technologien versprechen eine qualifizierte Vernetzung und damit eine verbesserte Auffindbarkeit und Nachnutzbarkeit von Ressourcen.

Der Beitrag stellt die hinter den Schlagworten Semantic Web und Linked Open Data stehenden Strategien zu einer nachhaltigen und spartenübergreifenden Vernetzung vor und erklärt die Bedeutung von Standards und Normdaten als Knotenpunkte für einen zunehmend offeneren Zugang zu kunsthistorischen Forschungsdaten im Web. Dabei wird die Brücke von der Graph-basierten Repräsentation von Wissen (**Knowledge Graph**) aktueller Suchmaschinentechnologie bis zu den im Kulturerbe-Bereich wichtigsten Normdateien und Koreferenzierungsprojekten geschlagen.

Integration

Actual

ical Thing

ation

Places

F.1 Der kunsthistorische Forschungskontext

Über digitale und analoge Abbildungen als Gegenstand einer digitalen Kunstgeschichte zu sprechen erscheint selbstverständlich, ist aber bei genauerer Betrachtung erklärungsbedürftig. Die Anforderung, digitale Repräsentationen eines Objekts selbst zu thematisieren und sie in die semantischen und stilistischen Kontexte des Objekts einzubetten, verweist auf den Doppelcharakter, der realistischen Abbildungen innewohnt: Sie versprechen, das natürlich gesehene Bild eines Gegenstands zu rekonstruieren, sind aber gleichzeitig selbst Produkte, die durch die bei ihrer Herstellung wirksamen Intentionen und technischen Bedingungen geprägt sind. ⁰¹ Dies gilt für die Erzeugnisse der analogen Fotografie und sogar in höherem Maße für digitale Bilder, die mit einer Vielzahl von Möglichkeiten digital weiter bearbeitet werden und damit das ihnen zugrunde liegende Objekt verfremden können, ohne dass dies bei Betrachtung des Resultats auf Anhieb erkennbar ist.

Die Kunstgeschichte war eine der ersten Disziplinen, die den dokumentarischen Nutzen der Fotografie für ihr Fach erkannte und in relativ kurzer Zeit die neue Technik umfassend und gewinnbringend in Forschung und Lehre zum Einsatz brachte. Das Foto, zumal als Kunstreproduktionsfotografie, fungiert hier als Stellvertreter, als Repräsentant des abwesenden, häufig unikalten Original-Objekts, das es abbildet und dem das Forschungsinteresse gilt. Es veranschaulicht seine Beschaffenheit und belegt die Argumentationsführung der Forschenden. Auch wenn die zuerst euphorisch begrüßte Objektivität der Darstellung ⁰² durch das neue bildgebende Verfahren schon bald relativiert wurde ⁰³, hat sich in der vorherrschenden kunsthistorischen Praxis der Gebrauch von Fotografien nicht wesentlich verändert. Das Foto entkoppelt das Objekt von den räumlichen und zeitlichen Bedingungen seiner Verfügbarkeit, es verhilft zur Kenntnis einer breiten Auswahl von Vergleichsobjekten, die aufgrund ihrer verteilten Standorte nicht alle zu besuchen wären oder in den dokumentierten Zuständen nicht mehr existieren. Es ist gleichzeitig ein beliebig verfügbares Vehikel der Erinnerung. Die auf seinem angenommenen verlässlichen Zeugnischarakter basierende Zweckbestimmung des dokumentarischen Fotos ist mit dem Wandel von der analogen zur digitalen Fotografie weitreichender denn je, denn erst das digitale Foto und die mit ihm verbundene textliche Information verschaffen dem analogen Objekt eine Präsenz im neuen virtuellen Raum, in dem mehr Repräsentationen von Objekten als jemals zuvor verfügbar sind.

Das Foto selbst, ob analog oder digital, wird dabei in seinem fachlich-kunsthistorischen Gebrauch durchsichtig, es wird nicht als eigenständiges Artefakt wahrgenommen, sondern verschwindet hinter seinem Gegenstand, dessen Darstellung sein eigentlicher Zweck zu sein scheint. ⁰⁴ Folgerichtig konzentriert sich die kunsthistorische Erschließungspraxis auf den dargestellten Gegenstand und nicht auf das Foto selbst. In musealen Sammlungen hingegen findet sich das analoge Foto schon bald als Sammlungsobjekt und wird primär in seiner Materialität beschrieben, die differenzierte Erschließung des Dargestellten ist demgegenüber nachrangig.

Mit dem Siegeszug der Digitalfotografie ändert sich die Perspektive auch in der kunsthistorischen Forschung: Wie das künstlerische wird auch das dokumentarische Foto vom einfachen Gebrauchsmedium zum historischen

■ 01

Joel Snyder, *Das Bild des Sehens* (1980), in: Herta Wolf (Hg.): *Paradigma Fotografie. Fotokritik am Ende des fotografischen Zeitalters*, Frankfurt (Main) 2002, S. 23–59.

■ 02

William Henry Fox Talbot, *Der Zeichenstift der Natur (1844–1846)*, in: Wilfried Wiegand (Hg.), *Die Wahrheit der Photographie. Klassische Bekenntnisse zu einer neuen Kunst*, Frankfurt (Main) 1981, S. 79; vgl. auch die Textausgabe, URL <https://www.gutenberg.org/files/33447/33447-pdf.pdf>, S. 1, 19, 20, 33, 48.

■ 03

Vgl. die Positionen von Berthold Haendcke 1893, Paul Kristeller 1908 in: Angela Matyssek, *Kunstgeschichte als fotografische Praxis (Humboldt-Schriften zur Kunst und Bildgeschichte Bd. VII)*, Berlin 2009, S. 130 ff., und Richard Hamann 1911, in: Matyssek 2009, S. 125 ff.

■ 04

Ingeborg Reichle, *Fotografie und Lichtbild: Die unsichtbaren Medien der Kunstgeschichte*, in: Anja Zimmermann (Hg.), *Sichtbarkeit und Medium. Austausch, Verknüpfung und Differenz naturwissenschaftlicher und ästhetischer Bildstrategien*, Hamburg 2005, S. 169–181, URL http://hup.sub.uni-hamburg.de/volltexte/2008/85/chapter/HamburgUP_Sichtbarkeit_Reichle.pdf.

Objekt, das es zu konservieren und zu erforschen gilt. Die fotohistorische Forschung interessiert sich für das Werk bestimmter Fotografinnen oder Fotografen, für die jeweiligen materiellen Befunde zusammengehöriger Fotos ebenso wie für den Authentizitätsgrad und die jeweils angewandten technischen Verfahren. Dies gilt für analog erzeugte genauso wie für digital erzeugte Fotobestände, deren historische Bedingtheiten trotz des noch relativ geringen Alters aufgrund der rasanten technischen Entwicklung ebenfalls als aufzeichnungswürdig erscheinen. Das dokumentarische Foto ist sowohl Repräsentant seines (künstlerisch gestalteten) Gegenstands als auch selbst ein gestaltetes Objekt. Dieser Doppelcharakter ist prinzipiell allen bildhaften Objekten eigen, deren Zweck die Dokumentation, Reproduktion oder Rekonstruktion eines anderen Gegenstands der realen Welt ist. Als Repräsentanten genutzte Bilder, ob sie als Druckgrafik, als analoge oder digitale Fotografie oder zunehmend auch als 3D-Rekonstruktion realisiert sind, sind damit auch selbst potenziell Forschungsgegenstand.

Das Bild ist ein Mittel visueller, nichtsprachlicher Kommunikation. Die Nutzungszugänge zu dokumentarischen Bildbeständen und den durch sie repräsentierten Objekten, Personen oder Ereignissen sind aber nach wie vor überwiegend textbasiert. Eine grundlegende Voraussetzung für die Auffindbarkeit des Bildes ist daher die Benennung der maßgeblichen beteiligten Entitäten **05** und ihre Kontextualisierung. In der textbasierten, digitalen Erfassung von Informationen werden dabei in der Praxis häufig mehrere Ebenen miteinander vermischt: Ein Objekt der bildenden Kunst, sei es ein Gemälde oder eine Grafik, kann eine Entität der realen Welt zum Thema haben. Dieser Bildinhalt wird ikonografisch erschlossen, beispielsweise dass es sich bei dem Gemälde um ein Portrait von Martin Luther handelt. Zugleich hat das Objekt selbst bestimmte Eigenschaften, es ist ein Gemälde von Lucas Cranach aus dem Jahre 1528. Und schließlich liegt eine digitale Abbildung des Objekts vor, deren spezifische Eigenschaften ebenfalls benannt werden können. Der Verweischarakter ist also ein doppelter: Vom Objekt der bildenden Kunst auf die reale Welt und von der digitalen Abbildung auf das Objekt. **06**

Eine breitere, systematischere und gleichzeitig differenziertere Nutzung von Bildern ist möglich, wenn die Dokumentation die Unterscheidung dieser Ebenen ernst nimmt. **07** Ein Gegenstand kann auf beliebig vielen Objekten der bildenden Kunst dargestellt werden, es liegt daher nahe, diesen mithilfe strukturierter Informationen zu identifizieren. Dies geschieht mithilfe der Kategorien **Titel** und **Thema**. Als separate Einheit wird das Objekt greifbar, wenn seine Entstehungsumstände, seine Materialität, sein aktueller Standort und seine Besitzverhältnisse ebenso systematisch dokumentiert werden. Zu den Merkmalen, die eine analoge oder digitale Abbildung charakterisieren, gehören der Name des Fotografen oder der Fotografin, das Aufnahmedatum, technische Verfahren der Herstellung und Bearbeitung sowie die Art des innerhalb einer Sammlung vorliegenden Bildträgers (Negativ, Abzug, Bilddatei).

Die Bilderschließung, das Sprechen über Bilder und in Bildern repräsentierte Objekte findet wiederum in verschiedenen textlichen Formen statt: Zum einen werden Informationen in der klassischen und i. d. R. datenbankbasierter Sammlungsdocumentation semantisch strukturiert in Datenfeldern erfasst. Die in einer Datenbank aufgenommenen Informationen sind teilweise mit

05

Unter Entität wird hier eine eindeutig bestimmbare Einheit der realen oder abstrakten Welt verstanden, über die Informationen (Daten) erhoben werden oder erhoben werden können: eine konkrete Person wie Michelangelo Buonarroti, ein Objekt der bildenden Kunst wie Michelangelos David oder ein Epochenbegriff wie Renaissance. Als Objekte werden in diesem Beitrag Werke der Kunst und Architektur bezeichnet, die typischerweise Gegenstand kunsthistorischer Forschung sind.

Zum Begriff der Entität in der Informatik siehe [https://de.wikipedia.org/wiki/Entität_\(Informatik\)](https://de.wikipedia.org/wiki/Entität_(Informatik)) sowie <https://de.wikipedia.org/wiki/Entity-Relationship-Modell>.

06

Die Komplexität der Zuordnung der Informationen zu den einzelnen Ebenen wird noch erhöht, wenn ein Objekt der bildenden Kunst einen realen Gegenstand darstellt, dieses Objekt (oftmals schon vor mehreren Jahrzehnten) in einer analogen Fotografie reproduziert wurde und das analoge Foto schließlich digitalisiert worden ist.

07

Visual Resources Association, VRA Core. A Data Standard for the Description of Images and Works of Art and Culture. Introduction, URL http://www.loc.gov/standards/vracore/VRA_Core4_Intro.pdf, Version: 28/10/2014.

■ 08

Eingabekontrollen basieren wiederum auf kontrolliertem Vokabular oder Normdaten, vgl. den später folgenden Abschnitt Normdaten: Knotenpunkte im Netz.

■ 09

https://en.wikipedia.org/wiki/Colorless_green_ideas_sleep_furiously.

■ 10

Zum Begriff der Semantik siehe die Abschnitte *The Human's View of Semantics* und *The Computer's View of Semantics* in: David Poole, Alan Mackworth, *Artificial Intelligence: Foundations of Computational Agents*, 2nd Edition, Cambridge (GB) 2017, URL <http://artint.info/2e/html/ArtInt2e.html>, Kapitel 5.1.2 *Semantics of the Propositional Calculus*, URL <http://artint.info/2e/html/ArtInt2e.Ch5.S1.SS2.html>.

Eingabekontrollen **08** versehen, teilweise aber auch freitextlich abgelegt. Zum anderen liegt die Information in aus informationstechnischer Sicht semantisch unstrukturierten Fließtexten vor, die manchmal innerhalb einer virtuellen Forschungsumgebung mit strukturierten Informationen verknüpft sein können, häufig aber einfach in klassischen Publikationsformaten vorliegen. Alle diese Informationen sind potenziell für die Forschung relevant bzw. selbst Ergebnis von Forschung. Im Kontext dieses Beitrags werden sämtliche in textlicher Form vorliegenden Informationen als Daten bezeichnet, die für die Wissenschaft Informationsquellen darstellen und verschiedene Organisationsgrade, von semantisch strukturierten – kontrollierten oder unkontrollierten – Daten bis zu semantisch unstrukturierten Daten (Fließtexte), abdecken.

Zur Klärung des Begriffs der Semantik ist der Rückgriff auf den Satz des Linguisten Noam Chomsky erhellend: »Colorless green ideas sleep furiously.« **09** Dieser Satz wurde 1957 von Chomsky konstruiert, um den Unterschied zwischen Syntax und Semantik in natürlichen Sprachen zu demonstrieren. Der Satz ist grammatisch korrekt, aber semantisch sinnlos. Übertragen auf die informationstechnische Aufbereitung und Verarbeitung von Daten lässt sich damit verdeutlichen, dass die Festlegung technischer Formate nicht ausreicht, um Inhalte sinnvoll zu präsentieren. Für semantisch strukturierte Daten ist mithilfe eines Datenmodells die Bedeutung der Aussagekomponenten definiert und ihre Kombinierbarkeit beschränkt. **10**

F.2 Zugang schaffen: Abbildungen – Objekte – Kontexte

Textlich erfasste Informationen – Daten – beziehen sich also auf die dargestellten Inhalte, auf die Objekte selbst und deren Entstehungskontexte, auf Personen, Orte, Ideen ebenso wie auf weitere bildhafte und textliche Quellen. Gleichzeitig sind die Forschenden daran interessiert, für ihre Fragestellungen weitergehende Informationen und Kontextmaterialien zu finden. Mit welchen anderen Objekten, Personen oder Ideen stehen die Gegenstände des Forschungsinteresses vielleicht im Zusammenhang? Der Ausgangspunkt liegt für Martin Doerr bereits in den grundlegenden Fragestellungen der historischen Wissenschaften an die Objekte:

»A museum object is more like an illustration, or witness of the past, than information in its own right. Cultural historical research means understanding **possible pasts**, the facts, events, material, social and psychological influences and motivations. It lives from understanding contexts, by pulling together bits and pieces of related facts from disparate resources, which can typically not be classified under subjects

■ 11

Martin Doerr, *Technological Choices of the Research Space Project* (August 2010), URL <http://www.researchspace.org/researchspace-concepts/technological-choices-of-the-researchspace-project>.

in an obvious way. It lives from taking into account all known facts.« ¹¹

Im Bereich der Erschließung geht es daher, besonders im Museums- und Archivbereich, um die Kontextualisierung des Objekts. Zur Identifikation und Bewahrung seines Zeugnischarakters ist die Aufklärung und Erläuterung des Kontexts erforderlich. Dies erleichtert die Authentifizierung dieser Objekte und unterstützt wiederum die Forschung bei ihrer Analyse und Interpretation. Um ein bestimmtes künstlerisches Objekt zu identifizieren, nutzt man Informationen seiner Geschichte und Materialität: Es geht um die Benennung der Akteure, die das Objekt zu einer bestimmten Zeit an einem bestimmten Ort unter Benutzung bestimmter Materialien und Techniken hergestellt oder verwaltet haben. Für archäologische Objekte kommt die Dokumentation des Fundereignisses hinzu, das sich ebenfalls über die Benennung von Akteuren, Ort, Zeit und Verfahren konstituiert. Zu den weiteren typischerweise erfassten Informationen gehören die klassifizierende Einordnung sowie die Erschließung der Darstellungsinhalte, die Ikonografie. Die Gewinnung dieser kontextualisierenden Informationen ist nicht selten Resultat einer aufwändigen Forschungsleistung, da das bildhafte Objekt – im Gegensatz zu den typischerweise in Bibliotheken erschlossenen Medien – diese oftmals nicht selbst mitteilt.

Zu den Bedingungen, wie für kunst- und kulturhistorische Fragestellungen relevante Bilder und Objekte überhaupt aufgefunden werden, wie Forschende sich weitergehende Informationen beschaffen können, gibt das obige Zitat in mehrfacher Hinsicht Aufschluss und verdeutlicht die Komplexität der Aufgabe: Die Daten sind naturgemäß unvollständig, unsicher, komplex, über verschiedene Quellen verteilt und weisen unterschiedliche semantische Strukturierungsgrade auf. Darüber hinaus konnte sich in der Dokumentation bildhaften Kulturguts in den Museen, bei der Denkmalpflege, in Universitäten oder Bildarchiven des deutschsprachigen Raums trotz verschiedener Initiativen kein Regelwerk oder Standardisierungskonzept für die strukturierte Beschreibung oder Klassifizierung von bildhaften Objekten durchsetzen. Die Landschaft ist durch eine Vielzahl von Insellösungen geprägt, viele Sammlungen arbeiten nach wie vor mehr oder weniger systematisch nach Hausregeln. Daraus resultieren große Herausforderungen für Informationssysteme, die einen möglichst einfachen und zielgenauen Zugang zu den Daten bieten sollen.

F.3 Suchen und finden: Closed-World-versus Open-World-Paradigma

Um die trotz oder gerade wegen der rasant wachsenden Menge an Bildern und Daten empfundene Diskrepanz zwischen Verfügbarkeit und Auffindbarkeit zu verstehen, muss man sich ein wenig mit den Voraussetzungen für ihre gezielte Auffindbarkeit beschäftigen: Wie funktionieren Suchzugänge, wo können Forschende nach für sie relevanter Information suchen und was finden sie? Für

das Information Retrieval sind zunächst zwei Paradigmen grundsätzlich zu unterscheiden.

Herkömmliche Datenbanksysteme, wie sie etwa im Museum als Sammlungsmanagementsysteme eingesetzt werden, basieren auf der Annahme einer Closed World, in der sämtliche bekannte und (im Sinne des jeweiligen verwendeten Datenmodells) vollständige Informationen zu allen relevanten Entitäten vorliegen und ausgewertet werden können. ¹² Die Informationen sind zudem einheitlich strukturiert und der in sich geschlossene Datenbestand ist nach definierten, auch kombinierbaren Kriterien abfragbar, etwa nach »allen Objekten, die nach 1945 durch Ankauf von Auktionshaus X in die Sammlung kamen« oder nach »allen von Autorin Y seit 1990 publizierten Monografien«. Bibliothekarische OPACs, aber auch Online-Datenbanken von Museen, funktionieren nach diesem Prinzip, das eine hohe Recherchequalität im Sinne von Recall und Precision ¹³ sichert, wenn die Dateninhalte zudem auf Basis kontrollierter Vokabulare und Normdaten erfasst sind. Solche Informationssysteme können für spezialisierte Anwendungskontexte passgenau entwickelt und konfiguriert werden, und auf Basis dieses Paradigmas erfolgt auch typischerweise die strukturierte Erschließung von Abbildungen und Objekten, maßgeschneidert für spezifische Projektanforderungen und/oder institutionelle Bedürfnisse. Daraus resultierende Webangebote sind demzufolge in sich homogen und schlüssig, sie stehen aber zunächst unverbunden neben weiteren, inhaltlich vielleicht höchst interessanten verwandten Angeboten. Zudem sind diese Webangebote aufgrund der Abgeschlossenheit häufig schwierig, jedenfalls nicht gemeinsam auffindbar.

Demgegenüber steht die Informationsbeschaffung unter der Annahme einer Open World, in der die Recherche über einen offenen und verteilten Datenbestand mit heterogenen Strukturen stattfindet. Die zu durchsuchende Datenbank ist potenziell das Web selbst. Der prominenteste Anwendungsfall sind Web-Suchmaschinen wie Google, die die für alle erreichbaren Angebote im Web (**Surface Web**) zugänglich machen. Hierfür wird Suchmaschinentechnologie eingesetzt, die auf Basis von Schlüsselwort-Suche und mit inzwischen sehr fortgeschrittenen Indexierungs- und Rankingalgorithmen eine Ergebnismenge mit einem tendenziell hohen Recall, d. h. vielen Treffern, aber geringer Precision, d. h. auch vielen nicht relevanten Treffern zurückliefert. Je besser das Ranking funktioniert, desto weniger nachteilig wirkt sich die geringe Precision aus, weil die als am relevantesten eingestufteten Treffer zuerst angezeigt werden. Die zugrunde gelegten Relevanzkriterien werden dabei vom Service-Anbieter konfiguriert und sind i. d. R. nicht transparent für die Nutzerin und den Nutzer. Suchmaschinentechnologie setzt von der historischen Entwicklung her zunächst auf textbasierter, semantisch wenig oder gar nicht strukturierter Information – typischerweise aus Webseiten – auf und wertet syntaktische Auszeichnungen wie Seitentitel oder Gliederungen aus.

Obwohl die Nutzung von Web-Suchmaschinen aus keinem Forschungsalltag mehr wegzudenken ist, muss doch festgestellt werden, dass sie mit Blick auf den kunsthistorischen Forschungskontext häufig unbefriedigende Treffermengen und Treffer liefert: Die Ergebnisse sind unspezifisch und stehen in Kontexten, die für die aktuelle Fragestellung nicht relevant sind. Eine weitergehende Filterung von großen Treffermengen auf Basis der gefundenen Treffer ist nicht möglich, so dass hochrelevante Treffer aufgrund ihrer niedrig gerankten

■ 12

Zur Closed World versus Open World Assumption siehe Poole/Mackworth 2017, dort 5.6 Complete Knowledge Assumption, siehe auch https://en.wikipedia.org/wiki/Open-world_assumption.

■ 13

Recall ist ein Maß für die Vollständigkeit einer Treffermenge: »Wurden alle relevanten Treffer gefunden?« bzw. »Wie hoch ist der Verlust?«, Precision ist ein Maß für die Genauigkeit einer Treffermenge: »Wurden nur relevante Treffer gefunden?« bzw. »Wie hoch ist der Ballast?«.

■ 14

Deep Web, im Gegensatz zum Surface Web, siehe https://de.wikipedia.org/wiki/Deep_Web.

Position de facto unzugänglich bleiben, weil man nur eine überschaubare Anzahl von Treffern auf den ersten Ergebnisseiten überhaupt zur Kenntnis nimmt. Andererseits werden viele relevante Ressourcen nicht gefunden, nicht zuletzt, weil die Inhalte in für Suchmaschinen nicht auswertbaren Formaten oder in unzugänglichen Datenbanken (**Deep Web** ¹⁴) vorliegen. Zudem stellen die sprachliche und kulturelle Heterogenität textbasierter Sucheinstiege wie auch die mangelnde automatisiert auswertbare Kontextualisierung und damit Verlinkung der bereitgestellten Information eine große Hürde dar.

F.4 Lösungsansätze für Daten des kulturellen Erbes

Der Zugang zu kunst- und kulturhistorischen Daten ist in der Nutzung also alles andere als einfach: Selbst wenn spezialisierte Webdatenbanken gefunden werden, muss man sich mit den jeweils unterschiedlichen Recherchemöglichkeiten, Funktionalitäten und Präsentationsformen vertraut machen. Wünschenswert sind hingegen Einstiegsmöglichkeiten, die keine Kenntnisse über Organisationsformen und Erschließungsregeln des spezifischen Angebots voraussetzen. Wie bereits festgestellt, ist es charakteristisch, dass die Daten ohnehin unvollständig und über viele Quellen verteilt sind, sie bilden also per se eine Open World.

Um den Begrenzungen der genannten Ansätze zur Informationsbeschaffung zu begegnen, wurden für den Kulturerbe-Bereich in den vergangenen Jahren Portalanwendungen und Plattformen mit dem Ziel aufgebaut, einen vereinfachten, dennoch qualifizierten sowie institutionenübergreifenden Zugang zum kulturellen Erbe im Web zu schaffen. Es gibt unterschiedlichste, thematisch, regional oder spartenspezifisch fokussierte Angebote. Zentrale Initiativen mit spartenübergreifendem, d. h. auf die Bestände in Museen, Bibliotheken, Archive und Forschungseinrichtungen zielendem Anspruch sowie großer geographischer Reichweite sind Europeana ¹⁵ und die Deutsche Digitale Bibliothek ¹⁶. Die Angebote verfolgen den Ansatz, hinsichtlich ihrer Herkunft wie auch ihrer Strukturierung außerordentlich heterogene Daten auf Basis eines einheitlichen Datenformats zusammenzuführen und auf diese Weise vereinheitlichte Zugangswege zu schaffen, über die die Nutzerin oder der Nutzer wiederum zu den Spezialangeboten navigieren kann. In der Tendenz gilt jedoch auch für diese Portale und Plattformen, dass die Recherchequalität, d. h. Recall und Precision, mit steigender Heterogenität der Quelldaten und wachsender Datenbasis sinkt. Dafür lassen sich im Wesentlichen zwei Gründe anführen: Um mit den etablierten Technologien einen nutzerfreundlichen und schnellen Zugang zu schaffen, kommen zum einen vereinfachende, flache Datenformate zum Einsatz, die vielfach mit einem erheblichen Verlust an Informationstiefe und Genauigkeit im Vergleich zu den Quelldaten verbunden sind. Zum anderen sind selbst gut strukturierte Quelldaten hinsichtlich der Dateninhalte häufig wenig kontrolliert, so dass durch die textliche Form die Hürde der sprachlichen und kulturellen

■ 15

Europeana, <https://www.europeana.eu>.

■ 16

Deutsche Digitale Bibliothek, <https://www.deutsche-digitale-bibliothek.de>.

Heterogenität fortbesteht. Die integrierte Suchbarkeit von Materialien, die z. B. eine gemeinsame Herkunft oder ein gemeinsames Thema aufweisen, aber auf Institutionen verschiedener Sparten verteilt sind, bleibt im Resultat bislang nur unzureichend umgesetzt.

F.5 Semantic Web und Linked Open Data

Die Schlagworte zur Überwindung dieser Unzulänglichkeiten lauten Semantic Web und Linked Open Data (LOD). Hierbei handelt es sich um informationstechnische Konzepte, deren Einsatz auch im Kulturerbe-Bereich seit einigen Jahren als zukunftsweisend propagiert und vielfach – allen voran von Europeana und Deutscher Digitaler Bibliothek – verfolgt wird, deren wirksame Umsetzung jedoch ein Umdenken im gesamten Prozess von der Bestandserschließung bis hin zur Publikation gewonnener Daten im Web erfordert.

Was verbirgt sich dahinter? Die Grundidee von Linked Open Data, 2006 von Tim Berners-Lee skizziert ¹⁷, lässt sich im Kern in der folgenden Erwartung paraphrasieren: Es soll für alle, also auch für die nicht-digitalen Entitäten – etwa für die historisch nachgewiesene Person Michelangelo Buonarroti, für die von ihm geschaffene Skulptur David, für den Epochenbegriff Renaissance –, eindeutige Namen im Web, nämlich HTTP URIs ¹⁸ geben. Diese URIs sollen von möglichst vielen Institutionen, die Daten produzieren und nutzen, wiederverwendet werden. Die Bezüge zwischen diesen Entitäten, etwa für die Aussage: **David** – wurde geschaffen von – **Michelangelo Buonarroti** sollen ebenfalls mit URIs als eindeutigen Namen und auf Basis von Standards, genauer RDF-Standards ¹⁹, ausgedrückt werden. Damit entstünde ein Netzwerk von semantisch verlinkten und über Systeme, Standorte und Anwendungen hinweg zu teilenden und zu nutzenden Daten. Tony Gill beschreibt das Potenzial einer solchen Vernetzung wie folgt:

»As of this writing, LOD offers great promise for semantically rich, easier, and more widespread use, reuse, and sharing of both metadata records and the controlled vocabularies that are used to populate those records and provide meaningful connections among resources. LOD has the potential to revolutionize the way data can be disseminated and integrated in ways that will significantly enhance the process of information- and resource-seeking and utilization.« ²⁰

Diesem Ansatz folgend müsste jede Entität, die im Kontext eines Projekts oder in einer Institution erschlossen wird, einen URI erhalten, und die zur Entität gewonnene identifizierende und die wesentlichen Merkmale beschreibende Information müsste mit dazu vorliegenden URIs der beteiligten Entitäten verknüpft werden. Auch für semantisch unstrukturierte Fließtexte bestünde die

■ 17

Tim Berners-Lee, W3C Design Issues – A Roadmap to the Semantic Web – Linked Data, URL <https://www.w3.org/DesignIssues/LinkedData.html>, Version: 18/06/2009.

■ 18

HTTP URIs (Uniform Resource Identifier) sind von einem URL zu unterscheiden: Ein URL ist die Adresse einer notwendig digitalen Ressource, ein HTTP URI repräsentiert eine – nicht notwendig digitale – Entität und liefert im Idealfall (aber selbst das ist nicht zwingend) identifizierende Information zu dieser Entität zurück.

■ 19

Resource Description Framework, <https://www.w3.org/RDF>, siehe auch https://de.wikipedia.org/wiki/Resource_Description_Framework.

■ 20

Tony Gill, Metadata and the Web, in: Murtha Baca (Hg.), Introduction to Metadata, Los Angeles, 2016 (3. Aufl.), URL <http://www.getty.edu/publications/intrometadata/metadata-and-the-web/>.

Anforderung, dass die im Text referenzierten relevanten Entitäten und deren Bezüge für Semantic Web-Techniken zugänglich gemacht werden.

Semantic Web-Techniken bilden die Brücke aus einer Closed World zur Open World, sie erlauben die Vernetzung heterogener und verteilter Daten auf Basis von Standards. Zunehmend werden semantische Strukturen auch von Web-Suchmaschinen berücksichtigt, zur besseren Interpretation einer Suchanfrage genutzt und in die Indexierungs- und Rankingalgorithmen integriert.

Ein wichtiger Schritt auf diesem Weg ist die Initiative Schema.org ^[21], die zum Ziel hat, qualifizierte Daten auf Basis einheitlicher Beschreibungselemente in Webseiten einzubetten und diese für die Verbesserung der Suchergebnisse zu nutzen. Im Rahmen von Suchmaschinenoptimierung (SEO) haben Webseiteninhaber die Möglichkeit, ihre Angebote mithilfe dieser Technik besser zu positionieren. Auch die Webangebote der kunst- und kulturhistorischen Dokumentation und Forschung mit ihren spezifischen hochwertigen Inhalten können versuchen, auf diese Weise ihre Sichtbarkeit und Zugänglichkeit zu verbessern. ^[22]

Inzwischen setzen die Suchmaschinenanbieter darüber hinausgehend auf die Graph-basierte Repräsentation von Wissen – Knowledge Graphs – die große strukturierte Datensammlungen zur Identifizierung von Entitäten (Personen, Orte, Werke, Publikationen, Körperschaften/Firmen, Marken) in Kombination mit Text Mining und Linked Data-Technologien nutzen, um Zusammenhänge besser erschließen zu können. ^[23] Bekanntestes Beispiel dafür ist der Google Knowledge Graph ^[24] ^[01]. Ausgangspunkt ist eine Datenbasis, die u. a. andere große, mit freien Lizenzen versehene Datenbestände wie die von Wikipedia,

■ 21

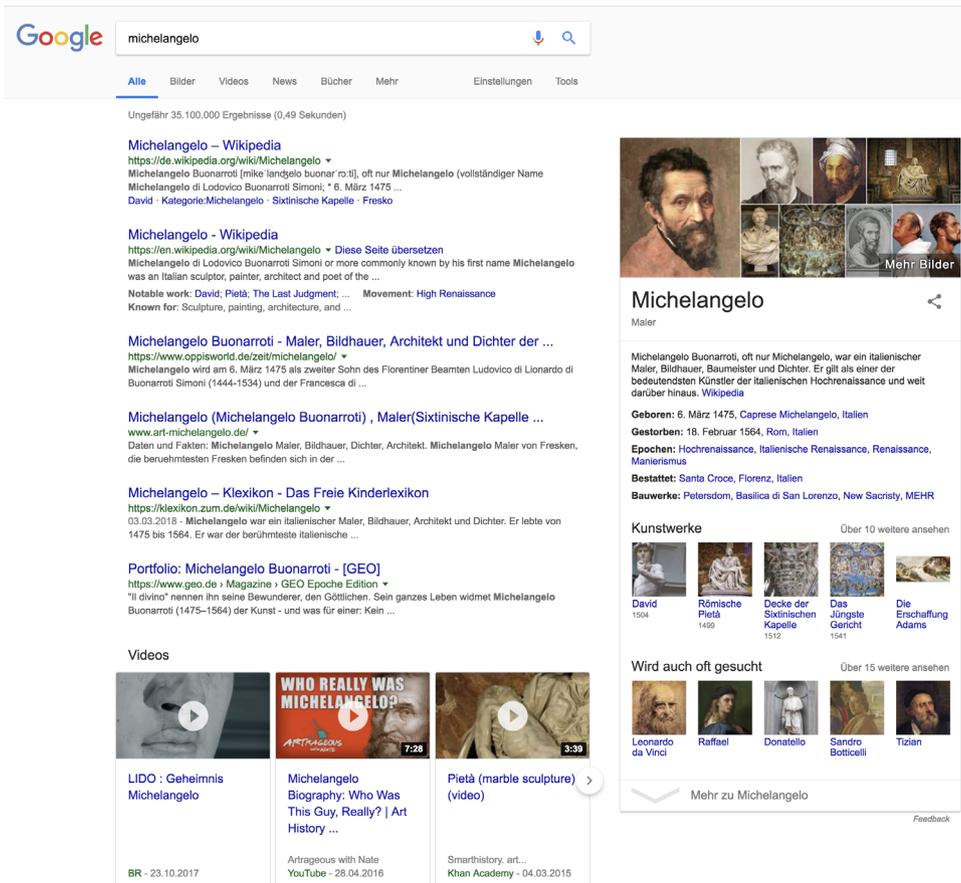
Schema.org, <http://schema.org>. Die Initiative wurde 2011 von den Suchmaschinenanbietern Google, Bing und Yahoo! gemeinsam lanciert.

■ 22

Google Search Documentation zur SEO-Optimierung von Webangeboten, <https://developers.google.com/search/docs/guides>.

■ 23

Bing Snapshot, <http://blogs.bing.com/search/2013/03/21/understand-your-world-with-bing>; Yahoo, <http://searchengineland.com/yahoo-has-their-own-knowledge-graph-not-without-their-own-embarrassing-issues-218538>.



□ 01

Google-Ergebnisseite zur Suchanfrage Michelangelo (https://www.google.de/?gws_rd=ssl#q=michelangelo).

■ 24

Google Knowledge Graph, <https://www.google.com/intl/bn/insidesearch/features/search/knowledge.html>, <https://googleblog.blogspot.de/2012/05/introducing-knowledge-graph-things-not.html>, https://en.wikipedia.org/wiki/Knowledge_Graph.

■ 25

Zum Beispiel bei Google-Suche nach Akropolis, Weizsäcker, Verne. Bei anderen Begriffen (z. B. Jaguar, Jupiter) wird dem Nutzer erst am Ende der ersten Trefferseite eine Verfeinerung der Suche vorgeschlagen.

■ 26

Laut Duden bedeutet Interoperabilität die »Fähigkeit unterschiedlicher Systeme, möglichst nahtlos zusammenzuarbeiten«, siehe <http://www.duden.de/rechtschreibung/Interoperabilitaet>.

■ 27

Zum Begriff der Ontologie siehe Poole/Mackworth 2017, Kapitel 14.3 Ontologies and Knowledge Sharing, <http://artint.info/2e/html/ArtInt2e.Ch14.S3.html>.

Wikidata oder des CIA World Factbook aggregiert, um Suchanfragen zu kontextualisieren und damit die Genauigkeit (Precision) der Treffermenge zu verbessern. So werden zum Suchbegriff passende Entitäten vorgeschlagen, die mit identifizierenden Kurzinformationen präsentiert werden, aber auch Vorschläge zur Disambiguierung ²⁵ gemacht und auf ähnliche Entitäten zur weiteren Navigation verwiesen. Seit der öffentlichen Einführung 2012 hat Google diese Technologie beständig fortentwickelt. In allen Kontexten, in denen große standardisiert erfasste Mengen an Daten vorgehalten oder erworben werden können, wird Google seinen Wirkungsgrad erweitern.

Die Herausforderung besteht nun darin, auf der Basis von zwangsläufig unvollständigen Kenntnissen – sei es, weil etwas tatsächlich unbekannt ist oder weil im Web vorliegende Ressourcen unbekannt oder unzugänglich sind – bestmöglich zum Aufbau eines – per se offenen – semantischen Netzes für den Kulturerbe-Bereich beizutragen.

F.6 Semantische Interoperabilität

Die von Tim Berners-Lee formulierten Prinzipien erlauben es, dass auch in einer LOD-Publikation die semantische Heterogenität herkömmlicher Erschließung fortgesetzt werden kann, die Nutzung von Standards bezieht sich lediglich auf die technische, nicht aber semantische Interoperabilität ²⁶ bei der Verlinkung – oder einfacher ausgedrückt: Die Standards legen die technische Struktur der Information fest, nicht aber ihren Inhalt, vergleichbar mit grammatikalischen Regeln. Insofern bringt nicht allein die Anwendung von LOD-Techniken wirklich qualitativ verbesserte Zugangsmöglichkeiten mit sich, wenn auch die Öffnung der Datenbestände und die Publikation mit Standard-techniken eine wesentliche Voraussetzung ist. Entscheidend für einen verbesserten Suchzugang bleibt darüber hinaus die übergreifende Verwendung von gemeinsamem Vokabular, sowohl für die semantischen Strukturen in Form von Datenformaten oder von Ontologien ²⁷ als auch für die Dateninhalte. Notwendig ist also die Verständigung auf Standards für die semantische Interoperabilität der Daten, denn die Tragfähigkeit und Angemessenheit der semantischen Strukturen bestimmt die Tragfähigkeit und Qualität der Suchergebnisse.

Welches sind die Fragestellungen, unter denen ein Datenbestand durchsucht wird? Zweifelsohne bieten Fragen nach wer – was – wann – wo typische Sucheinstiege, also nach Personen bzw. allgemeiner nach Akteuren, nach Objekten, nach Zeiträumen und Orten. Wenn diese nicht unterscheidbar sind, bleibt das Suchergebnis unbefriedigend, denn bei undifferenzierter Semantik kann die Suche in strukturierten Daten kaum bessere Ergebnisse erzielen als die Volltextsuche auf Basis aktueller fortgeschrittener Suchmaschinenteknologie. Es muss also eine der Komplexität kunst- und kulturhistorischer Daten angemessene Ontologie für die semantische Auszeichnung verwendet werden, um die Vorteile von Semantic Web-Ansätzen tatsächlich fruchtbar zu machen. Dies bedeutet nicht, dass sämtliche Daten in derselben Art und Weise und derselben Tiefe strukturiert werden müssen, vielmehr stellt sich die Frage nach der An-

gemessenheit der semantischen Strukturierung für den jeweiligen Zweck. Um für Informationssysteme Zugangspunkte zu einem Fließtext zu schaffen, kann die Verknüpfung im Sinne einer thematischen Erschließung z. B. mit den im Text herausgehoben behandelten Objekten und Personen bereits ausreichend sein. Auch um etwa einen großen Sammlungsbestand in kurzer Zeit mit eng begrenzten Personalmitteln überhaupt in eine Datenbasis zu überführen, können zunächst nur Grunddaten erhoben werden. Wichtig ist dabei, dass die dadurch entstehenden Daten derart in ein semantisches Netz überführt bzw. integriert werden können, dass dieses Netz wiederum konsistent ist und interoperabel ausgebaut werden kann. Um in einem weiteren Schritt Bezüge ebenfalls für eine Suche auswerten zu können, z. B. dass ein Objekt eine Person abbildet oder von einer Person gekauft wurde, dass ein druckgrafisches Porträt Teil einer Serie, Reproduktion eines Gemäldes, Abzug eines bestimmten Zustands ist, müssen diese explizit maschinenlesbar formuliert oder aufbereitet werden.

Eine wesentliche Dokumentationsanforderung in Forschungskontexten ist es auch, dass Daten dahingehend qualifiziert werden können, ob es sich um unumstrittene historische Fakten handelt, ob die Information ungesichert oder in der Forschung strittig ist. Die Zuschreibung eines Gemäldes an eine bestimmte Künstlerin oder einen Künstler muss etwa mit der Information verknüpft werden können, durch wen und wann die Zuschreibung erfolgt ist und ob sie weiterhin gültig ist. Je besser die Daten semantisch strukturiert sind, desto mehr Auswertungen mit semantischen Techniken sind möglich, ohne dass die Information explizit in den Daten enthalten sein muss. Das Druckdatum der druckgrafischen Reproduktion eines Gemäldes kann z. B. automatisch zur zeitlichen Eingrenzung einer wann-Suche auch in Bezug auf das Gemälde ausgenutzt werden, selbst wenn zum Gemälde selbst keinerlei Datierung in den Daten vorhanden ist.

Die Datenstrukturierung sollte auf der Basis der für den Kulturerbe-Bereich entwickelten Referenzontologie ISO 21127, des CIDOC Conceptual Reference Model (CIDOC-CRM) ²⁸ erfolgen. Hierbei handelt es sich um ein formalisiertes Begriffsmodell, das die Integration, Zugriffsvermittlung und den Austausch verschiedenartig strukturierter Informationen aus dem Bereich des kulturellen Erbes unterstützt. Es wurde seit Mitte der 1990er Jahre in Arbeitsgruppen der CIDOC, dem Fachkomitee für Dokumentation im Internationalen Museumsbund (ICOM), entwickelt. Anstelle eines festgeschriebenen, monolithischen Datenmodells bildet es ein erweiterbares gemeinsames Verständnismodell, in dem in formalisierter Weise die Dinge beschrieben sind, von denen im Kulturerbe-Bereich die Rede ist, und wie diese miteinander in Beziehung stehen. Dies beinhaltet insbesondere die saubere Differenzierung zwischen einem Bild und einer darauf dargestellten Entität. Ein CIDOC-CRM-konformes Datenmodell ermöglicht die Überführung jedweder Daten in eine CIDOC-CRM-konforme LOD-Publikation. Die Granularität der Datenmodellierung im konkreten Anwendungskontext ist dabei skalierbar: Entscheidend für eine qualifizierte Datenintegration im Sinne der wer-was-wann-wo-Fragen ist die explizite Ereignisorientierung des Modells. Im Kern sind dabei die Entitätstypen Objekte (E18 Physical Thing), Ideen (E28 Conceptual Object), Personen bzw. allgemeiner Akteure (E39 Actor), Orte (E53 Place), Datierungen (E52 Time-span) und deren Bezüge über zeitliche Entitäten (E2 Temporal Entity), insbesondere als Ereignisse, relevant. ^[2]

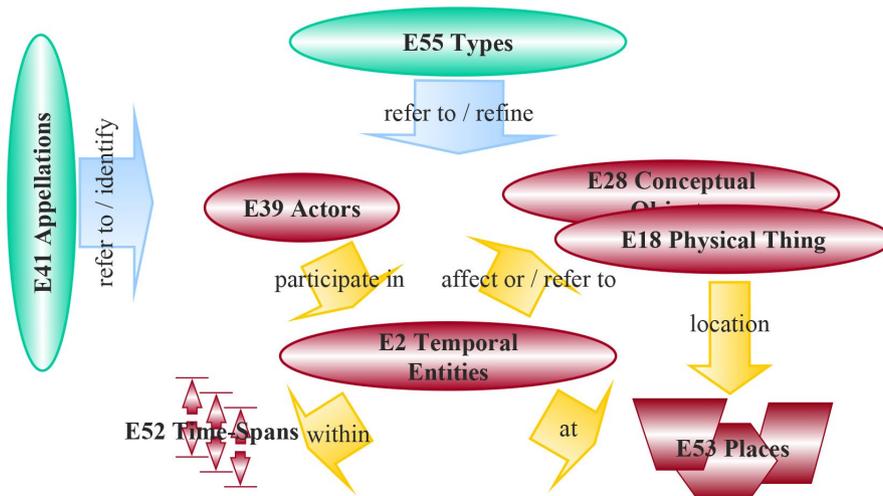
■ 28

CIDOC Conceptual Reference Model, <http://cidoc-crm.org>. Der offizielle ISO-Standard ist die Version 5.0.4, zugänglich als Online-Publikation unter <http://cidoc-crm.org/html/5.0.4/cidoc-crm.html>. Vgl. hier auch in der Terminologie des CIDOC-CRM die Definitionen der Begriffe *interoperability* und *semantic interoperability*, http://cidoc-crm.org/html/5.0.4/cidoc-crm.html#_Toc310250696.



The CRM Family

Top-level classes useful for integration



□ 02

CIDOC-CRM Top-level classes useful for integration (Martin Doerr, CIDOC CRM Family. Harmonized Models for the Digital World: CIDOC CRM and Extensions, Nürnberg, 19.5.2015. http://www.cidoc-crm.org/sites/default/files/1_CRM_Family%28MD%29.ppt).

■ 29
Lightweight Information Describing Objects, <http://www.lido-schema.org>.

■ 30
ICOM International Committee for Documentation, <http://network.icom.museum/cidoc>.

■ 31
LIDO ist auch als Standardformat zur Bereitstellung von Erschließungsdaten zu bildhaften und dreidimensionalen Objekten aus DFG-geförderten Digitalisierungsprojekten vorgeschrieben, siehe DFG-Praxisregeln Digitalisierung, http://www.dfg.de/formulare/12_151.

■ 32
Europeana Data Model Documentation, <http://pro.europeana.eu/page/edm-documentation>.

Für die CIDOC-CRM-konforme Publikation und Weitergabe von Daten auf XML-Basis hat sich das Harvestingformat LIDO (Lightweight Information Describing Objects) ²⁹ etabliert, welches ebenfalls unter dem Dach der CIDOC ³⁰, entwickelt wurde. LIDO dient als Standardformat für strukturierte Daten zu kunst- und kulturhistorischen Objekten, auf dessen Basis die Publikation solcher Bestände z. B. in der Europeana und in der Deutschen Digitalen Bibliothek erfolgt. ³¹ CIDOC-CRM- bzw. LIDO-basierte Daten lassen sich in spartenübergreifend konzipierte Anwendungsentologien wie das Europeana Data Model EDM ³² übertragen. Die Nutzung solcher Standards, die die Daten der kunst- und kulturhistorischen Domäne ausreichend spezifisch repräsentieren, erlauben qualifizierte Bezüge zwischen Entitäten. Um insbesondere für große Datenmengen spezifischere Zugangswege zu realisieren, müssen jedoch auch für die Bezugspunkte selbst, also die Entitäten, eindeutige Namen verwendet werden.

F.7 Normdaten: Knotenpunkte im Netz

Schon vor der Umstellung der Bibliothekskataloge von Karteikarten-Katalogen auf Datenbanken legten die bibliothekarischen Katalogisierungsregeln fest, dass die Benennung von Personen, Körperschaften, Orten oder thematischen Schlagworten mit einer Vorzugsbezeichnung zu erfolgen habe. Die eingesetzten kontrollierten Vokabulare bzw. Begriffssysteme umfassen nicht nur Schlagwortlisten und -systeme, sondern auch nicht natürlichsprachlich basierte

Systeme wie Klassifikationen. Die konsistente Verwendung der Vorzugsformen bei der Angabe der Personennamen oder der thematischen Verschlagwortung der katalogisierten Medien sichert die Homogenität des Kataloges und damit die Wiederauffindbarkeit des Eintrags.

In den verwendeten kontrollierten Vokabularen werden die Bezeichnungen (Schlagworte, Deskriptoren, Notationen) jeweils eindeutig auf Begriffe bezogen und Ambiguitäten der natürlichen Sprache wie Polysemie, Homonymie und Synonymie durch Definitionen oder Verweisungen aufgelöst. Die präferierte Form kann aus einer alphanumerischen Zeichenfolge (wie bei Personennamen) oder einem numerischen Code (wie bei der zur thematischen Eingruppierung von Literatur verwendeten Dewey-Dezimalklassifikation) bestehen.

Im Rahmen der Umstellung auf datenbankgestützte Katalogisierung konnte man auf dieses Organisationsprinzip aufbauen, sicherte es doch ebenfalls die Konsistenz der elektronischen Kataloge. Die Bestimmung der Vorzugsbezeichnungen zur Ansetzung erfolgte nun mithilfe ständig wachsender Normdateien, die zuerst vor allem Namensformen enthielten, zunehmend aber zu Systemen ausgebaut werden, die die Identifizierung und damit Referenzierung individueller Entitäten erlauben. Wegen ihrer zentralen Rolle für die Auffindbarkeit der Materialien ist die Pflege und Qualitätssicherung bei Normdaten eine ständige Herausforderung: Pro nachgewiesener Entität soll nur ein Datensatz existieren, Konsistenz und Erschließungstiefe sollen inhaltlich und strukturell möglichst einheitlich sein. Bislang war der Ausbau von Normdaten eine Aufgabe einschlägig qualifizierter Fachkräfte in der Verantwortlichkeit von wissenschaftsaffinen Institutionen wie Bibliotheken und Forschungsinstitutionen, die Verlässlichkeit und stabile Verfügbarkeit der Informationen garantieren. Auch bei Übernahme von Inhalten aus anderen Quellen (z. B. Crowdsourcing) gibt es nachgeschaltete Redaktionsvorgänge oder Kennzeichnungen der Vertrauenswürdigkeit der Quelle. Eine zentralisierte Pflege und Vorhaltung der Normdateien ermöglicht eine vereinfachte ressourcensparende Bestandserschließung und unterstützt vereinheitlichte Sucheinstiege für die Nutzerinnen und Nutzer.

F.8 Identifizierung beteiligter Entitäten

Ein Normdatensatz ermöglicht die Identifikation einer Entität durch Zusammenführung von Vorzugsbenennung und Namensvarianten mit identifizierenden Attributen. Der Datensatz der Gemeinsamen Normdatei (GND) der Deutschen Nationalbibliothek zur Person des Künstlers Michelangelo Buonarroti (<http://d-nb.info/gnd/118582143>) führt neben der Vorzugsbenennung **Michelangelo, Buonarroti** über 90 Namensformen auf, dazu die Lebensdaten, Geburts- und Sterbeort, Beruf und verschiedene Quellen. Diese Angaben reichen aus, um die individuelle Person zu identifizieren und eine Zuordnung vorzunehmen. Für Linked-Open-Data-Nutzungen ist hier die Referenzierung der Entität über den URI entscheidend, nicht die Übernahme der Vorzugsbenennung. Der URI liefert den kontrollierten Zugangspunkt zu den mit Michelangelo verbundenen Informationen in der jeweiligen Anwendung und gewährleistet zudem

eine zuverlässige Differenzierung zum gleichnamigen italienischen Schriftsteller und Dramatiker, der von 1568–1646 lebte und den URI <http://d-nb.info/gnd/115428364> hat.

Normdaten garantieren also die Identifizierung der Entitäten, sie liefern ihrerseits strukturierte semantische Information zu der Entität und bilden so den Ausgangspunkt für weitere qualifizierte semantische Verknüpfungen. Der URI fungiert als **Name** für die Entität, er ermöglicht damit die Auflösbarkeit von mit dem URI verbundener Information. Die automatischen Agenten des Web können nun also auswerten, dass es sich um dieselbe Entität handelt, die in verschiedenen Ressourcen mit demselben URI zitiert wird. Dies ist ein wichtiger Schritt von der Maschinenlesbarkeit zur Maschinenverständlichkeit von Daten.

F.9 Häufig genutzt: Gemeinsame Normdatei, Getty-Vokabulare, Iconclass, Geonames

Gegenwärtig gibt es keine universelle Normdatei, die für alle Dokumentationsbelange im Museum, Archiv oder in der Bibliothek geeignet wäre. Jede Sparte tendiert dazu, ihre eigenen Systematiken und Vokabulare zur Unterstützung der Informationsbelange ihrer Zielgruppen zu entwickeln.

Für die Referenzierung der in kunsthistorischen Daten relevanten Entitäten wird zunehmend die strategische Bedeutung der bibliothekarischen Normdaten wie der GND, aber auch der speziell für die Dokumentation der Kunstgeschichte und ihrer Nachbarwissenschaften konzipierten Vokabulare des Getty Research Institute – Art and Architecture Thesaurus (AAT) ³³, Thesaurus of Geographic Names (TGN) ³⁴, Union List of Artist Names (ULAN) ³⁵ – und des vom Rijksbureau voor Kunsthistorische Documentatie Den Haag (RKD) betreuten Iconclass erkannt. Mit der Nutzung dieser Vokabulare haben Sammlungen begonnen, ihre lokalen Datenbestände an weit verbreitete, damit übergeordnet anschließbare, nachnutzbare und von vertrauenswürdigen Institutionen langfristig gepflegte Normvokabulare anzuschließen. Sie profitieren damit auch von der konzeptionellen und technischen Weiterentwicklung der Informationsvernetzung, die für die Gedächtnisinstitutionen maßgeblich vom Bibliothekssektor vorangetrieben wird. Alle genannten Vokabulare stehen unter freien Lizenzen zur Verfügung, eine zentrale Voraussetzung für die unbeschränkte Nutzung und Wiederverwendbarkeit der Normdaten in Linked Open Data-Kontexten.

Die Gemeinsame Normdatei der deutschsprachigen Bibliotheksverbände als wichtigste und umfassendste deutschsprachige Normdatei deckt eine Reihe von Entitäten ab, die für die Dokumentation von Kunst- und Bauwerken oder Bildbeständen relevant sind. Sie umfasst aktuell Datensätze zu über 3,6 Millionen individualisierten Personen, 1,2 Millionen Körperschaften, 620.000 Veranstaltungen, 290.000 Geografika (mit Bauwerken und Denkmälern), 205.000 Sachschlagwörtern und 244.000 Werken verschiedener Kunstgattungen. Seit Einführung der gemeinfreien Nutzungslizenz (CC0) 2012 und der damit einhergehenden kostenfreien Bereitstellung der Daten in mehreren Formaten

■ 33

Art and Architecture Thesaurus, <http://www.getty.edu/research/tools/vocabularies/aat/index.html>.

■ 34

Thesaurus of Geographic Names, <http://www.getty.edu/research/tools/vocabularies/tgn/index.html>.

■ 35

Union List of Artist Names, <http://www.getty.edu/research/tools/vocabularies/ulan/index.html>.

■ 36

Deutsche Nationalbibliothek Digitale Dienste, http://www.dnb.de/DE/Service/DigitaleDienste/digitaledienste_node.html.

über verschiedene Schnittstellen und andere Bezugswege ³⁶ haben eine Reihe von Sammlungen begonnen, die GND zu nutzen und teilweise auch ihre bestehenden Datenbestände nachgängig zu referenzieren. Vor allem die Personennormdaten finden bislang Verwendung. In deutlich geringerem Ausmaß werden die Sachschlagworte für die Objektklassifikation oder die Thematik und die Bildinhalte angewendet. Bei den Sachschlagworten der GND handelt es sich um ein Allgemeinvocabular für alle Disziplinen, das bislang nicht im Sinne einer konsistenten Ontologie durchstrukturiert ist. In Bezug auf die gewünschte Abdeckung und die Spezifik der Bedeutungen und der Oberbegriffe erfüllt diese GND-Facette außerhalb des Bibliothekskontexts daher häufig nicht die fachspezifischen Anforderungen.

2013 hat die Deutsche Nationalbibliothek begonnen, die GND gezielt für die Nutzung und den Datenzuwachs aus dem Spektrum der nichtbibliothekarischen Gedächtnisinstitutionen zu öffnen. Damit gibt es neue Möglichkeiten, die Abdeckung der im Museum, im Archiv oder verwandten Einrichtungen benötigten Entitäten und damit die Reichweite GND-basierter Dienste zu verbessern. Die Verwendung der GND-URLs ist die Voraussetzung für eine zukünftig verstärkt auszubauende Verlinkung von Ressourcen über die Spartengrenzen hinaus, so dass es nicht mehr entscheidend sein wird, ob die Materialien zur gesuchten Entität in einem Museum, einem Archiv oder einer Bibliothek vorgehalten werden.

Mit den bereits angelegten Facetten zur Beschreibung von Werken, auch solchen der bildenden Kunst, und von ortsgebundenen Bauwerken und Denkmälern besitzt die GND ein großes Potenzial, auch für kunsthistorisch relevante Objekte selbst URIs an autoritativer Stelle anbieten zu können. Viele bekannte Kunstwerke oder Bauwerke haben bereits einen Normdatensatz, die wachsende Abdeckung in der Fläche wird aber von den noch zu entwickelnden Verfahren für den Massenimport von an anderer Stelle erfassten Daten abhängen.

Über den Datendienst Entity Facts ³⁷ der DNB bietet die Deutsche Digitale Bibliothek seit einiger Zeit ein LOD-basiertes Zusatzangebot zu GND-referenzierten Personen an. Ihre **Personenseiten** zeigen biografische Informationen zur Person mit qualifizierten Links auf GND-referenzierte Daten und Bilder innerhalb und außerhalb der DDB. ^[03]

Das Getty Research Institute bietet mit dem Art and Architecture Thesaurus (AAT) ein etwa 40.000 Begriffe umfassendes, hierarchisch strukturiertes und mehrsprachiges Vokabular für die Beschreibung von Kunstwerken, Architektur und Objekten der materiellen Kultur. Aufgrund seiner konsistent polyhierarchischen Strukturierung gemäß ISO 25964 ³⁸ und seiner fachlichen Fokussierung auf die Kunstgeschichte stellt er ein umfassendes Fachvokabular vor allem für die Bereiche der Sachklassifikation, der Stil- und Epochenbegriffe und der Material- und Technikangaben dar. International weit verbreitet, wird sein Einsatz im deutschen Sprachraum sicherlich zunehmen, wenn die begonnene Übersetzung ins Deutsche ³⁹ vollständig vorliegt.

Der Thesaurus of Geographic Names (TGN) weist über eine Million Geografika weltweit nach, darunter auch historische Orte. Die Union List of Artist Names (ULAN) enthält Datensätze zu ca. 250.000 mit der Kunstproduktion verbundenen Personen und Körperschaften. Alle Getty-Vokabulare stehen unter einer freien Lizenz als RDF-Daten über einen SPARQL-Endpoint oder als Exportdaten zur Verfügung.

■ 37

Wolfgang Boiger, Entity Facts – ein neuer Dienst der Deutschen Nationalbibliothek, in: B.I.T Online 18 2015, S. 329–338, URL <http://www.b-i-t-online.de/heft/2015-04/fachbeitrag-boiger.pdf> und Entity Facts, <http://www.dnb.de/DE/Wir/Projekte/Abgeschlossen/entityFacts.html>.

■ 38

ISO 25964 – the international standard for thesauri and interoperability with other vocabularies, <http://www.niso.org/schemas/iso25964>.

■ 39

Art and Architecture Thesaurus Deutsch, <http://www.aat-deutsch.de>.

DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

STARTSEITE ÜBER UNS JOURNAL HILFE

Link auf diese Seite

Michelangelo, Buonarroti
Künstler, Bildhauer, Maler, Architekt, Schriftsteller, Lyriker
Geboren: 6. März 1475, Caprese Michelangelo
Gestorben: 27. Februar 1564, Rom

Beteiligt an:

- Regel der fünf Orden, Von der Architectur Vignola, - Amsterdam, 1664
- Regola dell cinque Ordini D'Architettura Vignola, - Arnheim, 1620
- Sixtinische Decke — Die Vorfahren Christi — Jakob und Joseph Bildfeld
- Sixtinische Decke — Propheten und Sibyllen — Die Persische Sibylle Bildfeld

Alle Objekte (219)

Thema in:

- Sterbender Sklave in Gips [unsignierte Zeichnung der von Michelangelo Buonarroti gefertigten Skulptur; Louvre, Paris] 21. Dezember 1905, Landesarchiv Baden-Württemberg, Abt. Staatsarchiv Sigmaringen, FAS Sa A 7 T 1 Nachlass Xaver Henselmann
- Rom: St. Peter (Pietà von Michelangelo) Landesarchiv Baden-Württemberg, Abt. Staatsarchiv Sigmaringen, FAS H 1/1 T 1 Nachlass Albert Waldenspul: Glasplatten, Diapositive und Alben
- Rom: Sixtinische Kapelle (Das jüngste Gericht von Michelangelo, Detail) Landesarchiv Baden-Württemberg, Abt. Staatsarchiv Sigmaringen, FAS H 1/1 T 1 Nachlass Albert Waldenspul: Glasplatten, Diapositive und Alben
- Rom: Sixtinische Kapelle (Das jüngste Gericht von Michelangelo) Landesarchiv Baden-Württemberg, Abt. Staatsarchiv Sigmaringen, FAS H 1/1 T 1 Nachlass Albert Waldenspul: Glasplatten, Diapositive und Alben

Alle Objekte (92)

Anmelden Deutsch

michelangelo buonarroti
Nur Objekte mit Digitalisat Erweiterte Suche

03
Personenseite der Deutschen Digitalen Bibliothek zum Künstler Michelangelo Buonarroti (<https://www.deutsche-digitale-bibliothek.de/entity/118582143>).

Michelangelo, Buonarroti
Quelle: Wikimedia Commons

Informationen zum Lizenzstatus eingebundener Mediendateien (etwa Bilder oder Videos) können im Regelfall durch Anklicken dieser abgerufen werden.

Suche:
„Michelangelo, Buonarroti“ in der DDB

Externe Links:

- Gemeinsame Normdatei (GND) im Katalog der Deutschen Nationalbibliothek
- Bibliothèque nationale de France
- Kalliope Verbundkatalog
- Digitaler Portraitindex der druckgraphischen Bildnisse der Frühen Neuzeit
- Wikipedia (Deutsch)
- Wikisource
- Wikipedia (Englisch)
- NACO Authority File
- Virtual International Authority File (VIAF)
- International Standard Name Identifier (ISNI)
- Wikidata

■ 40
Iconclass, <http://www.iconclass.nl>.

■ 41
GeoNames, <http://www.geonames.org>. ■ 42
Allgemeines Künstlerlexikon – Internationale Künstlerdatenbank – Online, <https://www.degruyter.com/view/db/akl> (zugangsbeschränkt).

Iconclass **40** ist ein international verbreitetes alphanumerisches Klassifizierungssystem, das sich besonders zur Verschlagwortung von Bildinhalten und Thematiken der abendländischen Kunst eignet. Es deckt ca. 28.000 Themen ab und liegt in mehreren Sprachen vor, darunter auch auf Deutsch. Durch die hierarchische Organisationsweise sind spezifisch benennbare Sachverhalte zugleich systematisch und konsistent in generellere Kontexte eingeordnet, bleiben aber immer auf die Fachthematik bezogen. Iconclass ist im Web frei zugänglich und liegt in für LOD-Applikationen geeigneten Formaten vor.

Weitere relevante Normdateien in der Nutzung deutschsprachiger Kulturerbe-Institutionen sind Geonames **41**, das Daten zu über 9 Millionen geographischen Entitäten aus freien und öffentlichen Datenquellen aggregiert und bereits einen wichtigen Knotenpunkt im Semantic Web bildet, und das Allgemeine Künstlerlexikon Online **42**, das allerdings als Verlagsprodukt subscriptionspflichtig ist und wegen enger Nachnutzungskonditionen nicht für LOD-Anwendungen geeignet ist.

F.10 Strategien zum Umgang mit Koreferenz: VIAF und Wikidata

Die Effektivität des Semantic Web steigt mit der Offenheit und der semantischen Interoperabilität verschiedener Datenbestände, die maschinenlesbar miteinander verknüpft werden können. Diese wird durch das Problem der Koreferenz, des Auftretens mehrerer oder inkonsistenter Benennungen oder Identifier für eine bestimmte Entität, eingeschränkt. Zwei große Projekte haben es sich zur

Aufgabe gemacht, vielfach referenzierte Webressourcen gegeneinander abzugleichen, um eindeutige, instanzbezogene Einsprungs- und Knotenpunkte zu schaffen. Sie können als übergeordnete Drehscheiben für die großen Normdateien fungieren.

Der Virtual International Authority File (VIAF) ⁴³ hat das Ziel, eine weltweit einheitliche virtuelle Normdatei bereitzustellen. Das Projekt wurde 2003 von der Library of Congress, der deutschen und der französischen Nationalbibliothek und dem Bibliotheksdienstleister OCLC initiiert und leistet den qualifizierten Abgleich der Normdateien von über 40 Nationalbibliotheken und Bibliotheksverbänden weltweit. VIAF stellt zu einer Entität (z. B. einer Person) Konkordanzen über die jeweils zu ihr vorhandenen Normdatensätze her und bietet einen URI zu diesem Cluster an. Auf diese Weise kann die Entität unter allen Sprach- oder Namensvarianten oder URIs der einzelnen Normdateien adressiert werden. Der Schwerpunkt liegt bislang auf Daten zu Personen und Körperschaften, aber zunehmend werden auch Geografika und Werke nachgewiesen.

Wikidata ⁴⁴ ist eine von der Wikimedia Foundation 2012 initiierte, schnell wachsende, frei nachnutzbare domänenübergreifende Datenbank, zu der jeder durch persönliches Editieren, aber auch durch automatisierte Anreicherung beitragen kann. Die Anschubfinanzierung erfolgte u. a. durch das Allen Institute for Artificial Intelligence des Microsoft-Mitbegründers Paul Allen und Google. Wikidata wurde zunächst aufgebaut, um die Wikimedia-Schwesterprojekte sprachlich zu synchronisieren und strukturell zu verbessern, es ist durch wechselseitige Referenzierung eng auf diese bezogen. Anders als DBpedia korreliert Wikidata im vorgesehenen Abdeckungsbereich und der strukturellen Abgrenzung seiner Datenobjekte nicht zwingend mit Wikipedia. Daten werden zunehmend und im großen Umfang aus weiteren Quellen aggregiert und gegen den vorhandenen Datenbestand abgeglichen. Im März 2017 enthielt Wikidata über 25,4 Millionen Datenobjekte (items), von denen 27 % Aussagen (Eigenschaft-Wert-Paare) aus anderen Quellen als Wikipedia enthalten. ⁴⁵ Wikidata integriert VIAF, die GND, den AAT und Iconclass, wenn auch noch nicht mit vollständiger Abdeckung. ⁴⁶ Kulturerbe-Institutionen sind eingeladen, die Daten zu ihren Objekten in Wikidata einzubringen oder zu verlinken. ⁴⁷ Im Gegenzug können diese Institutionen nicht nur von der strukturellen Weiterentwicklung der für sie relevanten Datenschnittmenge durch Wikidata profitieren, z. B. durch Nutzung der Übersetzung vieler Aussagen in verschiedene Sprachen. Über eine der wichtigsten LOD-Drehscheiben wird ihr Bestand granular und qualifiziert zugänglich, der Traffic auf dem eigenen Webangebot nimmt dadurch zu. Wikidata wird, wie DBpedia, zu einem der wichtigsten Knotenpunkte der LOD Cloud werden. Manche trauen der Datenbank eine zukünftige Rolle als **Super-Authority** zu. ⁴⁸

Doch während bei Pflege und Ausbau der oben erwähnten Normdateien großer Wert auf die redaktionelle Absicherung der publizierten Informationen und Dublettenkontrolle der Datensätze ⁴⁹ gelegt wird und bislang kein Material aus unautorisierten Quellen integriert wird, ist dieses bei Wikidata anders. Etwa 10 % der Datenobjekte und 50 % der Aussagen sind bislang unreferenziert, weitere bedeutende Anteile wurden automatisiert aus den Crowdsourcing-Projekten Wikipedia oder Freebase ⁵⁰ gewonnen. ⁵¹ Das redaktionelle Qualitätsmanagement der Wikipedia und die Transparenz der Artikelentwicklung über die Versionsgeschichte sind wesentliche Momente für die Beurteilung ihrer Glaub-

■ 43

Virtual International Authority File, <https://viaf.org> und <http://www.oclc.org/viaf.en.html>.

■ 44

Wikidata, https://www.wikidata.org/wiki/Wikidata:Main_Page.

■ 45

Wikidata Stats, <https://tools.wmflabs.org/wikidata-todo/stats.php>, Anzeige-modus Table.

■ 46

Mix'n'Match, <https://tools.wmflabs.org/mix-n-match>.

■ 47

Wikidata:Flemish art collections, Wikidata and Linked Open Data, https://www.wikidata.org/wiki/Wikidata:Flemish_art_collections,_Wikidata_and_Linked_Open_Data.

■ 48

Maximilian Klein et. Al., Authority Addicts: The New Frontier of Authority Control on Wikidata, in: Wikimania, Hong Kong, 7.-11.08.2013, URL https://wikimania2013.wikimedia.org/wiki/Submissions/Authority_Addicts:_The_New_Frontier_of_Authority_Control_on_Wikidata und <https://www.youtube.com/watch?v=4oFeo38Ydz8>; Luca Martinelli, Wikidata: A New Way to Disseminate Structured Data, in: Faster, Smarter and Richer. Reshaping the library catalogue, Rom, 27.-28.02.2014, URL <http://eprints.rclis.org/22754/>.

■ 49

Eines der grundlegenden Konstruktionsprinzipien einer Normdatei ist, dass es pro abgrenzbarem Begriff einen Datensatz gibt, der diesen Begriff vertritt.

■ 50

Freebase, <https://en.wikipedia.org/wiki/Freebase>. Freebase wurde 2016 von Google, das den Betreiber Metaweb 2010 erworben hatte, zugunsten von Wikidata eingestellt.

■ 51

Wikidata Stats, Tabellen Overview und Statements per item.

■ 52

Andreas Kolbe, Andreas: OP-ED: Whither Wikidata? in: *The Signpost*, 2.12.2015, URL https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2015-12-02/Op-ed.

■ 53

Wikipedia: Belege, <https://de.wikipedia.org/wiki/Wikipedia:Belege>.

■ 54

Vgl. das Wikidata-Item zu Michelangelo Buonarroti in der Visualisierungsapplikation Reasonator, <https://tools.wmflabs.org/reasonator/?&q=5592>.

■ 55

PETRUS – Prozessunterstützende Software für die digitale Deutsche Nationalbibliothek, <http://www.dnb.de/DE/Wir/Projekte/Archiv/petrus.html>; Imma Hinrichs, Gérard Milmeister et al., *Computergestützte Sacherschließung mit dem Digitalen Assistenten (DA-2)*, in: *o-bib*, 3 (4) 2016, S. 156–185, URL <http://dx.doi.org/10.5282/o-bib/2016H4S156-185>.

würdigkeit und Qualität. Bei der Extraktion der strukturierten Information für Wikidata geht die Verbindung zu dieser Autorisierungsschicht aber verloren. Es wird lediglich pauschal die Datenquelle, im Falle der Wikipedia die jeweilige Sprachversion, genannt, nicht aber der oder die Quellartikel bzw. das Datum der Übernahme. ⁵²

Während im Falle der nach den gesetzten Maßstäben von Wikipedia ⁵³ verfassten, von vielen Autoren überarbeiteten und redaktionell sorgfältig geprüften Artikel auch Wikidata korrekte Informationen übernimmt, ist dieses dort nicht zuverlässig gewährleistet, wo die Qualitätssicherung von Wikipedia bislang zu kurz greift, so bei manchen Spezialthemen, gezielt manipulativ veränderten Informationen oder qualitativ unzureichenden Artikeln (die häufig als solche gekennzeichnet sind). Auch wenn der Wikipedia-Artikel später korrigiert wird, ist unklar, ob und wann Wikidata dies übernimmt.

Bei den Qualitäts- und Konsistenzprüfungen sowie bei der Zusammenführung der Daten setzt Wikidata gegenwärtig vor allem auf technische, softwarebasierte Lösungen. Besonders zu bekannteren Entitäten ist das Niveau der Informationsverknüpfung aus unterschiedlichen Quellen oft bereits beeindruckend ⁵⁴ und auch für weniger bekannte Entitäten kann sich dieses umso mehr verbessern, je mehr neue Datenquellen erfolgreich integriert werden. Es bleibt allerdings abzuwarten, ob die transparente Autorisierung der eingehenden Informationen eine größere Rolle als bisher spielen wird. Die CC0-Lizenz ist die Basis einer völlig offenen Partizipation beim Aufbau wie bei der Nachnutzung. Für maßgebliche, auf Wikidata aufbauende Applikationen wie den Google Knowledge Graph ist das Thema der Autorisierung wahrscheinlich von niedrigerer Bedeutung als für die wissenschaftsaffinen Kulturerbe-Institutionen.

F.11 Normdaten – Erschließungsstrategien für wachsende Datenmengen

Die intellektuell-manuelle Referenzierung von Datenbeständen mit Normdaten ist ressourcenintensiv. Für Suchangebote, die stark wachsende Mengen zunehmend heterogener Ressourcen nachweisen sollen, kann sie auf diese Weise, bei Einhaltung eines bestimmten definierten Niveaus, nicht mehr umfassend geleistet werden. Im Bibliotheksbereich reagiert man darauf einerseits mit der Gruppierung der zu katalogisierenden Ressourcen nach verschiedenen Erschließungsstufen. Außerdem geht man zunehmend zu (halb-) automatischen und verteilten Verfahren der Klassierung und Normdatenreferenzierung über, bei denen linguistische und Textmining-Methoden zum Zuge kommen oder bereits vorhandene Erschließungen nachgenutzt werden. ⁵⁵ Von der Zukunftsfähigkeit dieser Verschlagwortung ist Ulrike Junger (DNB) überzeugt:

»Die Anpassung der Erschließungsverfahren soll also gewährleisten, dass der strategische Mehrwert der Normdaten auch in Zukunft genutzt werden kann: Klassifikationen

■ 56

Ulrike Junger, Quo vadis Inhalterschließung der Deutschen Nationalbibliothek? Herausforderungen und Perspektiven, in: o-bib, 2 (1) 2015, S. 20, URL <http://dx.doi.org/10.5282/o-bib/2015H1S15-26>.

■ 57

Joachim Brand, Frank Dührkohp, Semantic Indexing of Figure Captions, in: EVA Berlin 2016 – Elektronische Medien & Kunst, Kultur und Historie (Konferenzband), S. 62–66, URL <http://books.ub.uni-heidelberg.de/arthistoricum/catalog/book/256>.

■ 58

John Resig, Using Computer Vision to Increase the Research Potential of Photo Archives, in: Journal of Digital Humanities, 3 (2) 2014, URL <http://journalofdigitalhumanities.org/3-2/using-computer-vision-to-increase-the-research-potential-of-photo-archives-by-john-resig/>; Peter Bell, Björn Ommer, Visuelle Erschließung. Computer Vision als Arbeits- und Vermittlungstool, in: EVA Berlin 2016 – Elektronische Medien & Kunst, Kultur und Historie (Konferenzband), S. 66–73, URL <http://books.ub.uni-heidelberg.de/arthistoricum/catalog/book/256>.

■ 59

Dies gilt auch für den Google Knowledge Graph, vgl. den Post von Tom Morris, führender Entwickler des Open Source-Tools OpenRefine: »Some of the data is just typed in by hand by oDeskers in the Philippines or other places. Google is a large customer of oDesk [...] and they use these contractors both to provide gold standard data for their machine learning algorithms, to QA sampled output of those algorithms and to hand pick (multi-vote) difficult cases.« URL <https://www.quora.com/Where-does-Google-get-their-Knowledge-Graph-data-from>.

und kontrollierte Vokabulare haben sich über viele Jahrzehnte als Mittel zur Wissensorganisation und Informationsverdichtung bewährt. Sie ermöglichen es auch im Kontext von Suchmaschinen und Discovery Systemen, die thematische Recherche zu unterstützen und sachlich Zusammengehöriges auffindbar zu machen. Indem beide Instrumente ins Zentrum gestellt werden, kann eine Einheitlichkeit der Sucheinstiege erreicht werden, unabhängig vom Verfahren, mit dem verbale oder klassifikatorische Daten erzeugt wurden. Darüber hinaus weisen DDC und GND Eigenschaften auf, die sie für die Anwendung im Web nützlich machen (z. B. Identifier, Strukturierung als Entitäten).« 56

Diese Strategie lässt sich verallgemeinern: Ein relativ hoher Vorstrukturierungsaufwand wird in die semantischen Hintergrunddaten gelegt – dies gilt unabhängig von den dabei eingesetzten Methoden gleichermaßen für die GND, Wikidata oder die Datenaggregationen der Knowledge Graphs. Bei der Verknüpfung mit den zu erschließenden Daten gilt es, wirtschaftliche Verfahren anzuwenden, die den Herausforderungen von großen Datenmengen und unterschiedlichen Strukturierungsgraden genügen und gleichzeitig ein bestimmtes Qualitätsniveau in den Ergebnissen gewährleisten. In diesem Zusammenhang werden einerseits Verfahren entwickelt, mit denen freitextliche Daten durch Markups nachträglich semantisiert und strukturiert werden können (z. B. durch Named-entity recognition) 57.

Zur Erschließung digitaler Bildressourcen bietet sich eine weitere Strategie an, die beim Bild selbst ansetzt: Bereits textlich oder gar semantisch erschlossene Bilder sind der Ausgangspunkt einer automatisierten Analyse, die Bilder als eine geordnete Menge von Bildpunkten mit bestimmten Mustern, Farb- und Tonwerten vergleicht. Bei entsprechender Übereinstimmung werden die mit dem Ausgangsbild verbundenen Metadaten auf das zu erschließende Bild übertragen. 58 Ihre Einsetzbarkeit für große Datenmengen hängt hier wesentlich von leistungsfähigen Verfahren zur Validitätsprüfung der automatisierten Zuordnungen ab. In Sammlungen oder Projekten kann ihr unterstützender Einsatz aber einen wichtigen Rationalisierungsgewinn bei der Bilderschließung bedeuten.

Letztlich sind es unabhängig vom konkreten Anwendungskontext dieselben Fragestellungen, die sich aus dem Umgang mit großen Beständen und Datenmengen ergeben: Wie können Zusammenhänge effizient erschlossen werden? Diese Herausforderung stellt sich für die eigentliche Bestandserschließung gleichermaßen wie für die Verbesserung des Zugangs zu Ressourcen mithilfe von Suchmaschinentechnologie. Wohl nicht zufällig sind die Lösungsansätze ebenfalls dieselben, nur mit unterschiedlichem Fokus: Das Rückgrat bildet eine qualifizierte Datenbasis und es kommt eine Kombination aus manuell-intellektuellem Input und automatischen Verfahren zum Einsatz. 59 Dabei gilt es mit einer Bottom-up-Strategie für zunehmend bessere Qualität in der Datenbasis zu sorgen. Konkret bedeutet dies für die verschiedenen Fachdisziplinen,

ihre spezifischen Vokabulare und Normdateien mit den größeren, übergeordneten Normdaten qualifiziert zu verknüpfen oder in die Datenbasen einzubringen, um einen zunehmend größeren Wirkungsgrad damit zu erzielen.

F.12 Besser vernetzt – aber wie?

Für die Geisteswissenschaften stellt die Bereitstellung von Informationen über Linked Open Data und Semantic Web-Technologien aktuell in erster Linie ein großes Versprechen und eine Zukunftsoption dar. Die Technologien eröffnen Möglichkeiten, die Zugänge zu den Informationsinseln des Web und den analogen Wissensspeichern, die heute noch in nicht geringem Ausmaß durch aktive aufwändige Recherche, Kennerschaft und glücklichen Zufall bestimmt sind, effektiv und qualitativ zu verbessern. Der Forschung verschafft dies neue Perspektiven: Bislang schlecht auffindbare, verteilte Materialien werden für neue Fragestellungen zugänglich, neue Formen der Zusammenarbeit werden möglich. Die große Herausforderung liegt dabei nicht so sehr in der Entwicklung der technischen Möglichkeiten, mit denen man auf anspruchsvolle Fragestellungen zu reagieren imstande wäre, als vielmehr im noch unvollständigen Verständnis ihres Anwendungspotenzials für die geisteswissenschaftliche Forschung. Dieses Verständnis verbreitet sich erst langsam, so wie auch die konkreten Strategien und Kenntnisse, mit denen die Geisteswissenschaften und die Kulturerbe-Institutionen ihre Daten für diese Art des globalen Publizierens ausrichten können, nur zögerlich Anwendung finden. Hier bedarf es einer sehr viel engeren und verstärkten Zusammenarbeit zwischen den Akteurinnen und Akteuren der Geisteswissenschaften, des Informationsmanagements und der Informatik, um gemeinsam die notwendigen Entwicklungen anzustoßen. Kunstwissenschaftlerinnen und Kunstwissenschaftler sollten dabei die Herausforderung annehmen, die Spezifik, Authentizität und den Strukturierungsgrad ihrer mit Semantic Web-Technologien zu öffnenden Daten selbst kompetent und gemäß den fachlichen Anforderungen festzulegen. Dies gilt für alle webbasierten Publikationsformen, ob es sich nun um natürlichsprachliche Texte, Medien-sammlungen, Datenbanken oder mehrere Präsentationsformen vereinende Forschungsumgebungen handelt. Die immer schon übliche assoziative Kontextualisierung in der Arbeit mit Fließtexten wird dabei in eine semantisch explizite Form überführt, die mithilfe neuer Technologien die Zugänglichkeit und Nutzbarkeit der Daten verbessert. Die Forschenden sollten dabei selbst mit ihrer Kompetenz die korrekte Identifizierung und die Qualität der Aussagen sowie ihrer expliziten (und nicht maschinell erstellten) Verknüpfungen absichern und diese Aufgabe bereits in der Planungsphase eines Projekts berücksichtigen. Sie haben ihr Material schließlich immer schon für die Publikation strukturiert, etwa durch die Aufbereitung für Register und Glossare.

Mehr denn je sind unterschiedliche Nutzungsszenarien für Daten umsetzbar oder überhaupt erst vorstellbar. Unter dem Aspekt der Nachhaltigkeit erfordert dies, die wertvollen, meist intellektuell erhobenen Daten für zukünftige Nutzungen zugänglich zu machen. Daten im Nachhinein für die Integration in

ein semantisches Netz aufzubereiten ist unvergleichlich aufwändiger und nicht selten mit Qualitätseinbußen behaftet, wie Erfahrungen mit der nachgängigen Aufbereitung von älteren Datenbeständen zeigen. ⁶⁰

■ 60

Zum Beispiel im DFG-Projekt Entwicklung eines zentralen historisch-biographischen Informationssystems für den deutschsprachigen Raum (2012–2016), <https://www.deutsche-biographie.de/ueber>.

Dies bedeutet nicht, dass sämtliche Daten in derselben Art und Weise und derselben Tiefe aufbereitet werden müssen, vielmehr stellt sich die Frage nach der Angemessenheit der semantischen Strukturierung für den jeweiligen Zweck. Entscheidend ist, dass auf Basis von vereinbarten Standards und Normdaten Zugangspunkte in jeglichen Daten so verankert werden, dass sie in ein semantisches Netz überführt bzw. integriert werden können, und dass dieses Netz in sich konsistent sowie auch konsistent erweiterbar ist. Wenn dieser Übergang von einer Closed World zur Open World gelingt und, technisch unterstützt, von vornherein in die geisteswissenschaftliche Forschungspraxis einbezogen wird, wird die angestrebte Rekontextualisierung von kunsthistorisch relevanten Objekten und Bildern in ihrem historischen Umfeld tatsächlich wirksam und eröffnet den Zugang zu weitergehenden Materialien.

Der anfangs geforderte differenzierte Bezug der Daten auf die in der Abbildung angesprochenen Ebenen, die des abgebildeten Objekts, seinen Themen oder der Abbildung selbst, und ihre entitätenbasierte Referenzierung wäre die Voraussetzung für die Gewinnung von qualifiziertem Material für eine größere Bandbreite an Fragestellungen: Welche anderen Objekte stellen die biblische Figur des David dar? Welche stellen die Statue Michelangelos dar? Welche Arbeiten anderer Künstler rezipieren dieses Werk? Wie unterstützt die sich historisch entwickelnde fotografische Praxis Kanonisierungsprozesse? Welche Rolle spielte dabei die Publikationen der Fratelli Alinari? Wo gibt es fotografische Dokumentationen zu den Restaurierungen des Originals? In welchen Kontexten werden Abbildungen des ikonischen Werks heute eingesetzt? Der konsequente Einsatz der hier vorgestellten Strategien verspricht einen qualitativen Sprung für die Identifizierbarkeit und einfache Erreichbarkeit der relevanten verteilten (Daten-)Bestände jenseits herkömmlicher fachwissenschaftlicher Kennerschaft. Durch die übergreifende Verwendung gemeinsamer Anker können die Versprechen eines Mehrwerts für die Forschung auf Basis von Semantic Web-Technologien zur Entfaltung kommen, wissenschaftliche Fachangebote können ihre qualifizierte Auffindbarkeit und ihre Reichweite im Netz verbessern. Um es mit den Worten von Tim Berners-Lee zu sagen: »It is the unexpected re-use of information which is the value added by the web.« ⁶¹

■ 61

Berners-Lee 2009.