

## Glossar zu den technischen Spezifikationen der Deutschen Digitalen Bibliothek

Zum Verständnis der technischen Spezifikationen der DDB werden die folgenden Begriffe definiert. In Anlehnung an die Unterscheidung von Objekten im Kontext der Europeana<sup>1</sup>, die als Bezugspunkt für die DDB dient, nutzt die DDB eine ähnliche, aber nicht deckungsgleiche Unterscheidung der Begriffe. Die Beziehungen zwischen diesen Begriffen werden unten in der Abbildung 1 dargestellt.

- a) **Objekt** bezeichnet ein kulturelles Artefakt oder ein Mentefakt, das in analoger Form als *Analoges Objekt* oder in digitaler Form als *Digitales Objekt* in einer Kultur- bzw. Wissenschaftseinrichtung vorliegt.
- b) **Analoges Objekt** steht für ein *Objekt* in der realen Welt.
- c) **Digitales Objekt** [siehe auch: Digitaler Inhalt] bezeichnet ein *Objekt*, das in binärer Kodierung auf einem Datenträger vorliegt (s.a. *Binärcontent*). Ein digitales Objekt kann entweder ein *Digitales Primärobjekt* oder ein *Digitalisat* sein. In der DDB werden Digitale Objekte durch Medientypen klassifiziert.
- d) **Digitaler Inhalt** → siehe Digitales Objekt
- e) **Digitalisat** steht für ein *Digitales Objekt*, das durch Digitalisierung eines *Analogen Objektes* erzeugt wurde.
- f) **Digitales Primärobjekt** steht für *Digitales Objekt*, das bereits digital entstanden ist (*digital born*) und in dem Sinne nicht lediglich Repräsentation eines *Analogen Objekts* ist.
- g) Ein **Volltext** bezeichnet eine Folge von Schriftzeichen, die auch Bilder und Tabellen enthalten kann.<sup>2</sup> Volltexte zeichnen sich insbesondere dadurch aus, dass sie elektronisch durchsuchbar, verlinkbar sowie editierbar sind und weiteren linguistischen Analysen und

---

<sup>1</sup> *Logical data model: Digital Objects and Surrogates des Europeana Outline Functional Specification* <[http://version1.europeana.eu/c/document\\_library/get\\_file?uuid=a9e29cb4-a9b3-462a-a43d-0b480c677088&groupId=10602](http://version1.europeana.eu/c/document_library/get_file?uuid=a9e29cb4-a9b3-462a-a43d-0b480c677088&groupId=10602)>, S. 14 -16.

<sup>2</sup> siehe „Leitfaden zur digitalen Bestandserhaltung. Vorgehensmodell und Umsetzung ; Version 1.0“, verfasst und herausgegeben von der nestor-Arbeitsgruppe Digitale Bestandserhaltung (nestor-materialien 15) <<http://nbn-resolving.de/urn:nbn:de:0008-2011101804>>

Bearbeitungsschritten unterworfen werden können (z.B. im Rahmen von wissenschaftlichen Arbeitsumgebungen, mit sprachwissenschaftlichen Hilfsmitteln, für statistische Auswertungen etc.). Volltexte sollten in strukturierter Form vorliegen (z.B. XML). Es gibt Volltexte, die genuin digital erzeugt wurden und Volltexte, die von einem analogen Dokument in digitaler Form abgeleitet worden sind. Bei den genuin digital erzeugten als auch bei den abgeleiteten digitalen Texten lassen sich folgende Unterscheidungen anstellen:

- *selbstständige Volltexte*: Abgeleitete selbstständige Volltexte sind z.B. Transkriptionen, Transliterationen oder Übersetzungen von gedruckten oder handschriftlichen Textdokumenten (auch in Teilen). Hierbei handelt es sich um Textdokumente, die auch ohne das Original aus sich verständlich und verwendbar sind und diese als Vertreter repräsentieren können. Bei genuin digital erzeugten selbstständigen Texten handelt es sich um digital präsentierte Publikationen, z.B. in Form von Online-Artikeln, von Online-Dissertationen, von Online-Manuskripten etc. Ein gedrucktes Äquivalent dazu ist jeweils denkbar.
- *nicht-selbstständige Volltexte*: Genuin digital erzeugte nicht-selbstständige Texte sind z.B. wissenschaftliche Ergänzungen und Anmerkungen zu einer Ressource. Sie kommen in verschiedenen Darstellungsumgebungen vor und sind in der Regel mit anderen digitalen Objekten verknüpft (z.B. Fußnoten, Endnoten, editorischen Notizen, Interpretationen, Einleitungen). Dazu gehören auch Volltexte, die auf Webseiten und in virtuellen Ausstellungen als Kontextinformationen über andere Quellen und Repräsentationen erstellt worden sind.

h) *Derivat* bedeutet Abkömmling, Abgeleitetes, Ableitung und steht für ein aus einem *Digitalen Objekt* erzeugtes Extrakt. Je nach Medientyp (Text, Bild, Ton, Bewegtbild) können sehr unterschiedliche Derivate erzeugt werden. Auch ist zu beachten, dass Derivate verschiedenen Zwecken dienen können. Die DDB unterscheidet die folgenden Formen von Derivaten:

- *Miniaturbild*: unterstützendes optisches Element für die Kurzdarstellung eines Treffers (z. B. Thumbnail);
- *Vorschau*: gibt Einblick in den Inhalt des Objektes und soll dem Nutzer bei der Relevanzbewertung helfen (Anzeigebild, Inhaltsverzeichnis etc.);
- *Arbeitskopie*: am Rechner für die Arbeit und Weiterverwendung in virtuellen Arbeitsumgebungen voll nutzbares Objekt i.d.R. mit reduzierter, aber nicht druckfähiger Qualität.
- *extrahierte Daten*: Text zur Indexierung und Realisierung einer Volltextsuche; Ton zur Indexierung und Realisierung einer Volltextsuche.

i) *Binärcontent* bezeichnet entweder ein *Digitales Objekt* oder ein *Derivat*. Die DDB erlaubt grundsätzlich den Zugriff auf Binärcontent (z.B. Bilder, Videos, Audio, Text) der in

der DDB-Plattform selbst vorgehalten wird („interner Binärcontent“) als auch auf Binärcontent, der vom Kooperationspartner direkt bereitgestellt wird („externer Binärcontent“). Dabei gilt:

- *Interner Binärcontent*: wird vom Kooperationspartner an die DDB übertragen, auf den DDB-Servern gespeichert und im DDB-Webportal angezeigt.
- *Externer Binärcontent*: wird vom Kooperationspartner direkt bereit gestellt. Auf den DDB-Servern wird nur ein Link auf den Content gespeichert. Über Verweise auf Systeme des Kooperationspartners können weitere Inhalte angezeigt werden (zum Beispiel: Referenz auf einen externen Content-Viewer, Referenz auf eine Ansicht des digitalen Objekts beim Datengeber).

j) **Metadaten** bezeichnen die die digitalen bzw. analogen Objekte beschreibenden und/oder kontextualisierenden Textinformationen (einschließlich Hypertext-Links), die zur Identifizierung, Auffindung, Auslegung und/oder Verwaltung von Objekten dienen können und die gegebenenfalls für deren spätere Suche, Präsentation oder für deren Rechte-Management erforderlich sind. Metadaten im Sinne der DDB reichen von einem „Kern-Metadaten-Satz“ (s. Anhang b) Qualitätsanforderungen für Metadaten) bis zu äußerst umfassenden Erschließungsinformationen, die es ermöglichen, die betreffenden Objekte in nahezu jedem relevanten Kontext zu finden und mit anderen Objekten zu vernetzen. Sie skizzieren den Inhalt und/oder Kontext und sind nicht immer rechtfrei, sondern gegebenenfalls als Sprachwerk urheberrechtlich geschützt. Spartenübergreifend ist es schwierig, zwischen urheberrechtsfreien und urheberrechtlich geschützten Metadaten zu unterscheiden.

Spartenspezifisch werden die folgenden Arten von Metadaten unterschieden:

- a) Bibliotheken unterscheiden inhaltliche Erschließung (Abstract, Annotationen, Schlagworte, Klassifikationen), bibliographische Metadaten (Angaben wie Autor, Titel, Verlagsangaben, Beziehungen zu anderen Ressourcen), technische Metadaten (Angaben wie Umfang und Format der digitalen Ressource), administrative Metadaten (wie z. B. Angaben zum Zugriff auf die digitale Ressource, Identifier, Rechte usw.), Strukturdaten (die die inhaltlichen Strukturen von Dokumenten wie Kapitel, Abschnitt, Paginierung usw. wiedergeben) und Meta-Metadaten (die Informationen zur Entstehung der Metadaten enthalten).
- b) Für Museen umfassen Erschließungsinformation und Metadaten die zu einem analogen Objekt vorliegende wissenschaftliche Erschließung, die Benennung, den Herkunftszusammenhang (Hersteller, weitere am Herstellungsprozess Beteiligte, Besitzgeschichte), die Beschreibung (Nutzungskontexte, Ikonographie, Marken, Beschriftungen), technische Daten (Abmessungen, Material, Technik), Datierung und geographische Zuordnung.

- c) In Archiven beschreiben Erschließungsinformationen den Entstehungszweck, den Inhalt und die Struktur einer Archivalie (z. B. einer Akte). Metadaten können ergänzend hinzu kommen, um ein archivistisches Objekt zu klassifizieren. Beide sind ohne erheblichen Informationsverlust schwierig voneinander zu trennen.
  - d) In der Denkmalpflege umfassen; Metadaten die rechtfreien Angaben zum Denkmal, z. B. die Adresse. Alle übrigen Erschließungsinformationen wie z. B. Beschreibungen oder wissenschaftliche Sachdaten sind nicht rechtfrei.
  - e) In den Rundfunk-, Fernseh- und Filmarchiven liegen standardisierte filmographische Erschließungsinformationen vor sowie Metadaten zur physischen Werkmanifestation wie Länge, Medium usw.
  - f) In den Wissenschaften werden für viele Bestände bibliothekarische, museumsspezifische und archivalische Beschreibungen (s. o.) genutzt. Einen großen Stellenwert nehmen Forschungsdaten ein, die in den verschiedenen Disziplinen, in Form und im Format sehr heterogen sind. Sie umfassen z. B. empirische Daten, Messdaten, sowie Kommentare, Transkriptionen, Übersetzungen und Verknüpfungen von Objekten. Forschungsdaten sind i. d. R. Grundlage für Publikationen. Für die Erschließung der Forschungsdaten liegen zumeist keine bzw. keine einheitlichen Beschreibungskonventionen vor.
- k) DDB-Content-Einheit bezeichnet alle zu einem *Objekt* gehörenden Daten, die an die DDB geliefert, von der DDB gespeichert und ggf. weitergegeben werden. Eine DDB-Content-Einheit enthält immer zumindest *Metadaten*. Sie kann darüber hinaus *Erschließungsinformation* sowie eine beliebige Anzahl von *Derivaten* enthalten.

Abbildung 1 zeigt die Beziehungen zwischen den definierten Begriffen als Entity-Relationship-Modell (ERM). Die einzelnen Begriffe der Definitionen (a) bis (k) werden als Entitätstypen in rechteckigen Symbolen dargestellt. Die Beziehungen zwischen diesen Begriffen finden sich in Rauten-Symbolen als Relationstypen wieder. Sie sind jeweils von links nach rechts zu lesen. Schließlich stellen dreieckige Symbole mit der Beschriftung „ist ein“ eine Spezialisierung bzw. Generalisierung dar, wobei der allgemeinere Entitätstyp über, der speziellere unter dem Symbol steht. Beispielsweise ist *Digitales Objekt* ein Oberbegriff für *Digitales Primärobjekt* und *Digitalisat*. Daher gilt: Ein *Digitalisat* ist ein *Digitales Objekt*. Ein speziellerer Entitätstyp „erbt“ alle Eigenschaften und Beziehungstypen des allgemeineren Typs (Oberbegriff). Ein *Derivat* kann also beispielsweise allgemein aus einem *Digitalen Objekt* abgeleitet sein. Auf einer spezielleren Betrachtungsebene bedeutet dies, dass es entweder aus einem *Digitalen Primärobjekt* oder aus einem *Digitalisat* abgeleitet ist:

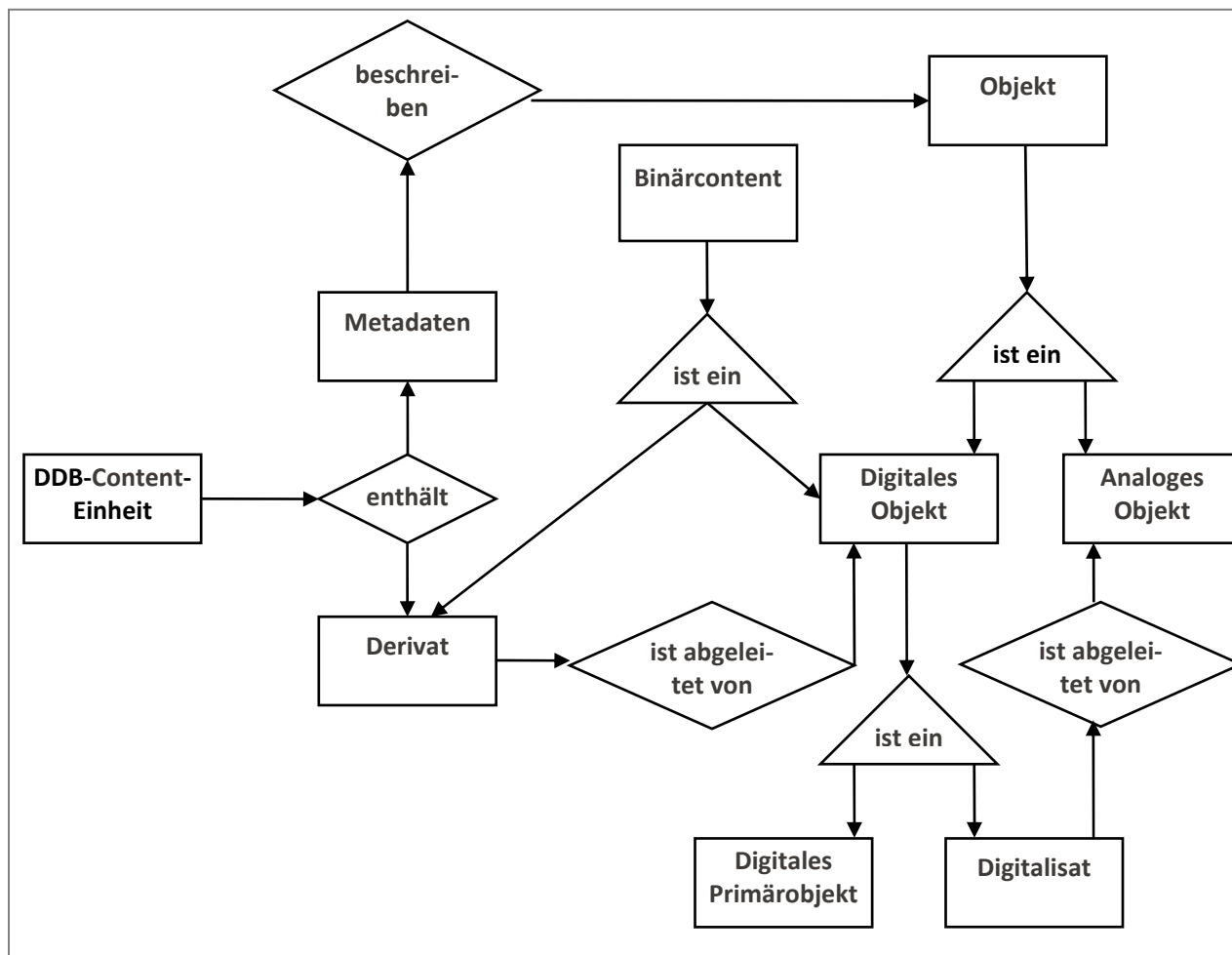


Abb. 1 Darstellung der Begriffsbeziehungen