

Piotr Kuroczyński

Von Big Data zu neuer Datenkultur

Die Digitalisierung unserer Gesellschaft beschleunigt in bisher nicht gekanntem Ausmaß die Bildung und den Austausch von Informationen. Gleichzeitig beobachten wir eine Auslagerung unseres kulturellen Gedächtnisses in das World Wide Web, welches seit 1990 eine rasante Entwicklung von Web 1.0 über Web 2.0 in Web 3.0 vollzieht.

Die exponentiell wachsende Informationsflut birgt Chancen, bereitet aber auch Probleme. Nach einer Schätzung aus dem Jahre 2013 waren nur ein geringer Anteil, nämlich 5% der damals geschätzt 4,4 Exabyte weltweit vorhandenen Daten sogenannte **target rich data** **01**, also Daten, deren Bedeutung von Maschinen gelesen und verstanden werden können, weil sie in strukturierter Form abgelegt und mit Informationen über sich selbst (**Metadaten**) ausgestattet waren. Im Kontext der Unmengen an unstrukturierten, heterogenen und schwer lesbaren Daten sprechen wir von **Big Data** oder **Dark Data**. Diese stellen die Wirtschaft und die Gesellschaft vor neue Herausforderungen, wenn es um Auswertung und Nachnutzung vorhandener digitaler Ressourcen geht.

Im Bereich des kulturellen Erbes gibt es sowohl von kommerzieller Seite als auch von staatlichen Einrichtungen Bestrebungen, unser analog gespeichertes Wissen der **Gutenberg-Galaxis** zu digitalisieren und jederzeit sowie überall online zugänglich zu machen. **02** Darüber hinaus werden immer mehr Objekte von kulturellem Wert mittels 3D-Retrodigitalisierung erfasst und als 3D-Punktwolken digitalisiert. Parallel entstehen digitale 3D-Rekonstruktionen von zerstörten bzw. nicht realisierten Kunst- und Baudenkmälern, die ein hypothetisches Abbild unserer Vergangenheit in Form von born-digital 3D-Modellen und Visualisierungen vermitteln.

Auch wenn im Kontext des Kulturerbes noch lange nicht von Big Data gesprochen werden kann, so stehen die kulturellen Institutionen und die sie begleitenden Wissenschaften vor ähnlich großen Herausforderungen, wenn es um die Lesbarkeit, Verwertbarkeit und Nachhaltigkeit der digitalen Daten geht. Aufgrund der schlechteren finanziellen Lage und des geringeren wirtschaftlichen Drucks sind die Herausforderungen hier eher größer als kleiner.

■ 01

Das Wachstum des Digitalen Universums wird geschätzt von 4,4 Exabyte (2013) auf 44 Exabyte (2020). Ein Wachstum von 40% je Jahr bis 2020. <https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>.

■ 02

Marshall McLuhan, *The Gutenberg Galaxy*, London 1962.

Der digitale Wandel hat bereits frühzeitig die Forschungsdaten und Forschungsmethoden erfasst und verändert die Wissenschaft grundlegend. Der **Rat für Informationsinfrastruktur (Rfii)** stellt zu Recht fest, dass

»mit der Digitalität von Forschungsdaten sich vielfältige, in ihrer Art neue Herausforderungen an deren Qualitätssicherung stellen. Neben der Qualität und Integrität von Datensätzen »an sich« spielen die Qualität des Datenmanagements und der Metadaten, die Einhaltung international gängiger Standards sowie die Qualität der Infrastruktur selbst eine wichtige Rolle.« **03**

■ 03

Rfii – Rat für Informationsinfrastrukturen, Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, Göttingen 2016, S. 52, <http://www.rfii.de/download/rfii-empfehlungen-2016/>.

Der Rfii stellt hierbei den Begriff einer **neuen Datenkultur**, einer Kultur der Offenheit und des Teilens der Daten, vor und fordert gleichzeitig eine ausreichende Datenqualität über den gesamten Lebenszyklus hinweg. Denn so lange die digitalen Forschungsdaten eine kürzere Lebensdauer haben als die materiellen Träger unseres kulturellen Gedächtnisses, wie gedruckte Bücher, Kunst- und Baudenkmäler, müssen wir uns ernsthaft die Frage nach dem Mehrwert der Digitalität in Kultur und Wissenschaft und damit der Sinnhaftigkeit des Einsatzes von Computern in den Institutionen des kulturellen Gedächtnisses und der Wissenschaft insgesamt stellen.

Arthistory's Next Topmodel?

Hinsichtlich der digitalen 3D-Rekonstruktion blicken wir auf 30 Jahre vielfältiger Anwendungen im Kontext (kunst-)historischer Rekonstruktionen zurück (**Heike Messemer → 059**). Die nativ digitalen Daten, aus denen die 3D-Modelle bestehen, stellen das Ergebnis der im Entstehungsprozess vorangegangenen wissenschaftlichen Diskurses und der hypothetischen Abbildung eines Objekts mittels computerbasierter Technologien dar. Dennoch ist es schwierig, gezielt auf die Ergebnisse der Forschungsprojekte zuzugreifen. Schuld daran ist die fehlende **digitale Erschließung**, die fehlende

»Integration und Konvergenz zwischen digitalisierten und nativ digitalen Daten in einheitlichen, integrierten Arbeitsumgebungen mit dem Ziel dynamischer Wissensintegration.« **04**

■ 04

Rfii – Rat für Informationsinfrastrukturen, Göttingen 2016, S. 13.

■ 05

André Borrmann, Markus König, Christian Koch, Jakob Beetz (Hg.), Building Information Modeling: Technologische Grundlagen und industrielle Praxis, Springer Vieweg, 2015.

Die Bauindustrie begegnete der Herausforderung fehlender digitaler Erschließung bereits in den 1990er Jahren mit der Entwicklung von **Building Information Modeling (BIM)** und des Datenaustauschmodells **Industry Foundation Classes (IFC)**. **05** Dank BIM und IFC können zeitgenössische Bauvorhaben effizienter geplant und abgewickelt werden, sodass es möglich ist, Planungs-, Bau- und Betriebskosten zu senken. Aus einfachem Grund:

Der gesamte digitale Bauprozess kann über die Schnittstelle des IFC-Datenmodells gewerkeübergreifend und ohne Informationsverluste an einem 5D-Modell simuliert, koordiniert und dokumentiert werden. Dahinter verbirgt sich die Idee eines **digitalen Zwillings** des Objekts in all seinen Lebensphasen: Von der Planung über den Bauprozess bis zum Betrieb und Abriss.

Im Zusammenhang mit geisteswissenschaftlichen objektbasierten Fragestellungen, die anhand einer digitalen 3D-Rekonstruktion bearbeitet und beantwortet werden sollen, gestaltet sich die Lösung schwieriger. Hier geht es vorrangig um die Dokumentation und Abbildung der subjektiven Quelleninterpretation und der kreativen 3D-Modellierung, der Texturierung und der Visualisierung eines Bauwerks. Anders als bei Planung zeitgenössischer Bauwerke mithilfe von BIM/IFC gilt es bei der digitalen 3D-Rekonstruktion vergangener Kulturgüter, für die mit hermeneutischen Methoden gewonnenen Rekonstruktionserkenntnisse einen Ausdruck in digitalen, hypothetischen 3D-Datensätzen (Modellen) zu finden. Die digitale (kunst-)historische Rekonstruktion wartet noch auf einen gemeinsamen Dokumentationsstandard (Datenmodell) und eine dynamische Wissensintegration innerhalb und mittels adäquater Virtueller Forschungsumgebungen (VFU). Eine Grundvoraussetzung für die digitale Erschließung leisten strukturierte Daten, welche sich an einschlägige Metadaten-Schemata, kontrollierte Vokabulare und Normdateien sowie Referenzontologien richten, wie sie in den **DFG-Praxisregeln »Digitalisierung«** festgehalten sind. ⁰⁶

■ 06

Deutsche Forschungsgemeinschaft, **DFG-Praxisregeln Digitalisierung**, Bonn 2016, S. 26, https://www.dfg.de/formulare/12_151/12_151_de.pdf.

Die Fachcommunity ist zurzeit bemüht, ein Datenmodell zu entwickeln, das die subjektiv-kreativen Interpretations- und Arbeitsprozesse bei der digitalen 3D-Rekonstruktion abbildet und das 3D-Modell zu den beteiligten Personen, zu den Orten, Quellen, Aktivitäten und Ereignissen in Beziehung setzt. Tagungssektionen und Round Tables wie **Arthistory's Next Topmodel? Der Trend zur Ontologie und Data Model beyond the digital 3D Model** eruieren das Thema der Wissensrepräsentation und die Bedeutung der Datenstrukturierung aus dem Blickwinkel der Geisteswissenschaften. ⁰⁷ Es geht dabei um nicht weniger als den Nachweis der Wissenschaftlichkeit digitaler Forschungsdaten – festzumachen an Nachvollziehbarkeit, Interoperabilität und Nachhaltigkeit – durch eine adäquate Wissensrepräsentation.

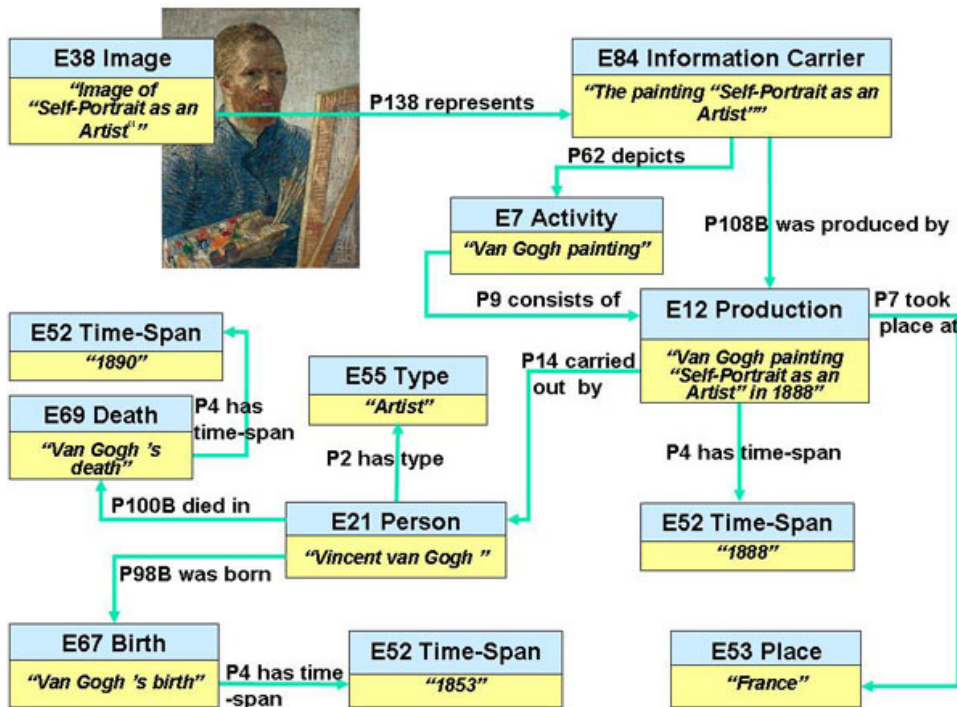
■ 07

Digital Humanities im deutschsprachigen Raum e. V., **Arthistory's Next Topmodel? Der Trend zur Ontologie**, Leipzig 2016, <http://www.dhd2016.de/abstracts/sektionen-003.html> und **22th Conference on Cultural Heritage and New Technologies, Round Table, Data Model beyond the digital 3D Model**, Wien 2017, <http://www.chnt.at/data-model-beyond-the-digital-3d-model/>.

■ 08

<http://www.cidoc-crm.org/>.

Als Referenzontologie im Kulturerbe-Bereich ist CIDOC CRM (ISO 21127:2006) weltweit anerkannt ⁰¹. **08** Mit rund 90 Entitäten und 150 Relationen ermöglicht der internationale Standard CIDOC CRM eine Formalisierung des Wissens innerhalb der kulturellen Einrichtungen (vor allem in Museen und Archiven) in menschen- und maschinenlesbarer Form.



□ 01

Formalisierung des Wissens am Beispiel des Selbstporträts von Van Gogh mithilfe der CIDOC CRM-Referenzontologie.

http://old.cidoc-crm.org/crm_core/core_examples/van_gogh.htm
(Martin Doerr).

Die Referenzontologie stellt eine übergeordnete Formalisierung des Wissens dar, auf welche sich wiederum die fachspezifischen Anwendungsentologien berufen. Die Referenzontologie CIDOC CRM bildet, analog zur IFC aus dem Bausektor, den **semantischen Kitt** zwischen den Anwendungsentologien unterschiedlicher Fachdisziplinen.

Im Bereich der Archäologie verbessern Erweiterungen wie das CIDOC CRMba die archäologische Erfassung von Gebäudebefunden und machen es damit leichter, die Dokumentation und Rekonstruktion der Gebäude zu verstehen. Wie **Paola Ronzino** (→ 354) in ihrem Beitrag schildert, können dank der Formalisierung des Wissens in CIDOC CRMba die Datensätze einer Ausgrabung beispielsweise hinsichtlich ihrer Bedeutung abgefragt werden und die entsprechenden Schlussfolgerungen daraus gezogen werden. So können automatische rechnergestützte Suchabfragen zu Raumkonfigurationen und Relationen durch die Ausgrabungsfunde (Quellen) belegt und ihre Interpretation nachvollzogen werden.

Die epistemologische Schwierigkeit, in einem Datenmodell eine subjektive Entscheidung in eine operative Objektivität zu übersetzen, kommt im scheinbaren Gegensatz von Hermeneutik einerseits und Normdaten andererseits zum Ausdruck. Für die Bearbeiter des Modells stellt sie eine große Hürde dar. Diese Übersetzungsleistung ist im Umkehrschluss jedoch Voraussetzung für die Operationalisierung des Wissens, da sie auf formalisierten Informationen gründet. Die Herausforderungen der konzeptionellen Modellie-

nung, der semantischen Segmentierung und der kognitiven Prozesse im Kontext digitaler 3D-Rekonstruktion beleuchtet **Fabrizio Ivan Apollonio** (→ 412) im letzten Beitrag von Kapitel 4. Seine analytische Ausführung und Vorstellung der Methodik erläutert der Autor im anschließenden ersten Beitrag im **Projekt-Portfolio** (→ 432) an einem Beispiel zu Architekturzeichnungen von Ledoux.

3D Digital Heritage – Exploring Virtual Research Space

Grundlegende Voraussetzung für die digitale Erschließung und Nachnutzung digitaler Forschungsdaten stellen **Virtuellen Forschungsumgebungen** (VFU) dar. Sie ermöglichen eine kollaborative, webbasierte Forschungsarbeit und fungieren im Sinne dynamischer Wissensintegration.

Die webtechnologische Entwicklung im Bereich der interaktiven Darstellung von 3D-Datensätzen, allen voran die WebGL-Technologie, eröffnet heute neue Anwendungshorizonte für den wissenschaftlichen Einsatz von 3D-Modellen in Forschung und Lehre.

Die Rolle der konzeptionellen Segmentierung eines Bauwerks bei dessen grafischer und informationstechnischer Erfassung in einer Datenbank wird von **Alexander Stenzer** (→ 372) untersucht. Dabei werden Fragen nach einer geeigneten grafischen Schnittstelle, dem Datenmodell und der Verknüpfung von Quellen und Bauwerksdaten erörtert.

Ebenfalls mit Interesse zu beobachten ist die Aktualität, die das in der zeitgenössischen Bauindustrie etablierte 3D-Datenaustauschmodell BIM/IFC in Bezug auf digitale 3D-Rekonstruktionen bekommen hat. **Paola Ronzino** und **Alexander Stenzer** stellen in ihrem Beitrag Überlegungen zur Anknüpfung des CIDOC CRM-Datenmodells an das Datenmodell von IFC an. Die Öffnung gegenüber BIM/IFC erklärt sich zum einen aus dem Wunsch, Fragestellungen aus dem Kulturerbe-Bereich in einem **digitalen Zwilling** (As-Built-Model, gängig in der Bauwerkserfassung und im Facility Management) abzubilden. Zum anderen erweitert man das Spektrum von BIM/IFC um die kulturhistorische Dimension. Diese Entwicklung sollte zusammen mit der Forschung im Bereich des **historic bzw. heritage Building Information Modelling** (hBIM), bei dem das digitale Gebäudemodell die Rolle des zentralen Informationsträgers übernimmt, weiter verfolgt werden. ⁰⁹

Für die objektorientierte Forschung bringt die VFU viele Anforderungen im Sinne der eingangs postulierten digitalen Erschließung und neuen Datenkultur mit sich. Es sind größere Forschungsprojekte, wie unter anderem **Virtuelle Rekonstruktionen in transnationalen Forschungsumgebungen – Schlösser und Parkanlagen im ehemaligen Ostpreußen** (**Kuroczyński et al** → 546), die das Mögliche auf diesem Gebiet ausloten und die Voraussetzungen für die webbasierte Zusammenarbeit verschiedener Fachkulturen bei digitalen 3D-Rekonstruktionen herausarbeiten. Aus den gewonnenen Erfahrungen kann eine ganze Reihe von Anforderungen abgeleitet werden, die gleichzeitig die Komplexität des Themas widerspiegeln. Von besonderer Bedeutung ist:

■ 09

Maurice Murphy, *Historic Building Information Modelling (HBIM) for Recording and Documenting Classical Architecture in Dublin 1700 to 1830*, Dublin, 2012; Yusuf Arayici, John Counsell, Lamine Mahdjoubi, Gehan Nagy, Soheir Hawas (Hg.), *Heritage Building Information Modelling*, Routledge, 2017.

- die Nutzung von Open-Source-Technologien für das operative System (z. B. Drupal als Content Management System)
- die Gestaltung intuitiver grafischer Schnittstellen (User Interface)
- eine Konzeptentwicklung für eine bessere Nutzererfahrung mit niedrigerer Hemmschwelle (User Experience)
- eine Strukturierung und Referenzierung der Daten in einem internationalen Standard (mensch- und maschinenlesbares Format, z. B. Resource Description Framework, Referenzontologie, z. B. CIDOC CRM)
- die Anbindung der Daten an kontrollierte Vokabulare und Normdaten (z. B. The Getty Controlled Vocabularies, Gemeinsame Normdatei)
- die semantische Segmentierung und Verknüpfung von Text, Bild und 3D-Modell mit den Datenbank-Instanzen: a) Annotation von Freitexten (HTML), b) Annotation von Bildern mit Scalable Vector Graphics (SVG), c) Annotation der 3D-Modelle im 3D-Viewer (WebGL)
- die Versionierung der 3D-Modelle (Zitierfähigkeit)
- die Open-Access-Publikation der 3D-Modelle (unter Creative Commons Lizenzen)
- die Publikation der mensch- und maschinenlesbaren Projektdaten (Export auf webbasierte Online-Dienste wie GitHub)

Den Stand der Technologie zur webbasierten 2D-/3D-Darstellung und weitere Komponenten einer VFU für die kunstgeschichtliche Forschung werden von **Peter Fornaro** (→ 392) am Beispiel einer am Baseler **Digital Humanities Lab** entwickelten Arbeitsumgebung vorgestellt. Neben der Klärung der Definition und der Ziele einer VFU beleuchtet der Beitrag das Potenzial und die Relevanz der echten Farb- und Materialwiedergabe mithilfe des **Reflexion Transformation Imaging (RTI)**-Verfahrens, welches von der WebGL-Technologie unterstützt wird.

Die größte Herausforderung hinsichtlich der Wissensrepräsentation und der digitalen Erschließung unserer rechnergestützten Forschungsdaten besteht darin, die Erkenntnisse aus den drittmittelgeförderten Forschungsprojekten zu verstetigen. Hierzu müssen zum einen die Forschungsinstitutionen die Relevanz und Bedeutung des Themas erkennen, um über die Projektlaufzeiten hinaus die personellen Ressourcen in diesem Bereich zu halten. Zum anderen muss die Fachcommunity sich endlich auf Standards hinsichtlich der Projektdokumentation und Wissensrepräsentation verständigen, die wiederum die Nachvollziehbarkeit, Lesbarkeit, Austauschbarkeit und Nachhaltigkeit der Forschungsergebnisse im Sinne einer **guten wissenschaftlichen Praxis** gewährleisten. **10**

Sollte die Wissenschaft allgemein den digitalen Wandel ernst nehmen, wird sie nicht um die Anforderungen von **Linked Open Data (Web 3.0)** herumkommen, die Tim Berners-Lee als die höchste Stufe seiner fünfstufigen Skala definiert. **11** Auf dem Weg zu **neuem Forschungsraum** und **virtuellen Forschungsumgebungen für digitale 3D-Rekonstruktionen** sowie nachhaltigen digitalen Forschungsdaten im Sinne von **Open Science** müssen wir uns ernsthaft mit einer Reihe an Fragen auseinandersetzen, die im gleichnamigen Artikel des Autors näher betrachtet wurden: **12**

■ 10

Deutsche Forschungsgemeinschaft, Sicherung guter wissenschaftlicher Praxis, Denkschrift Memorandum, Bonn 2013, S. 15, https://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf.

■ 11

Tim Berners-Lee, 5-Sterne-Modell, <http://5stardata.info/de/>, 2016.

■ 12

Piotr Kuroczyński, Neuer Forschungsraum für die Kunstgeschichte: Virtuelle Forschungsumgebungen für digitale 3D-Rekonstruktionen, in: Piotr Kuroczyński, Peter Bell, Lisa Dieckmann (Hg.), Computing Art Reader. Einführung in die digitale Kunstgeschichte, arthistoricum.net-ART-Books, Heidelberg 2018, S. 161–181, <https://books.ub.uni-heidelberg.de/arthistoricum/reader/download/413/413-17-83324-3-10-20190111.pdf>.

- Wie können Semantic-Web-Technologien in den Digital Humanities sinnvoll eingesetzt werden?
- Wie wird das Wissen im Web 3.0 strukturiert und organisiert?
- Welche Ontologien und Begriffssysteme können angewendet werden?
- Wie können Inhalte mithilfe maschineller Intelligenz semantisch angereichert werden?
- Welche Forschungsumgebungen/-infrastrukturen werden der Wissenschaft in diesem Kontext einen angemessenen Dienst leisten?
 Ein sich abzeichnender digitaler Forschungsraum basiert auf der neuen Datenkultur und einer semantischen Vernetzung mensch- und maschinenlesbaren Wissens [02].

□ 02
 Ebenenbasierte Darstellung der Komponenten einer Forschungsinfrastruktur im Sinne dynamischer Wissensintegration, Linked Open Data und Open Science. (Piotr Kuroczyński, 2017)

