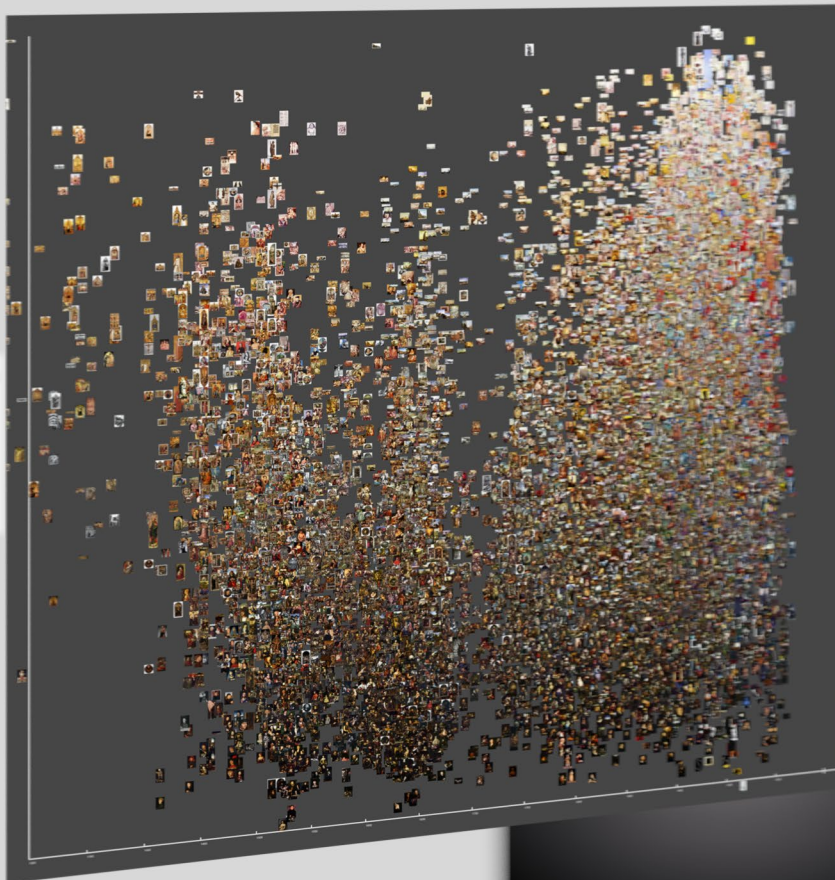


**Werk**

NAME  
Bamberger Altar (Stoß)

DATIERUNG  
1520 bis 1523

MATERIAL UND TECHNIK  
Lindenholz

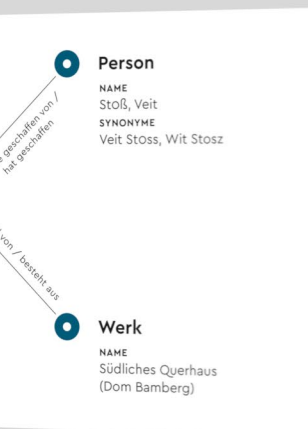


Thorsten Wübbena

## G. (Un)Ordnungen – Werkzeuge – Beziehungen: Datenbanksysteme und kunsthistorische Forschung

→ Digitale Kunstgeschichte, Datenbanken, ConedaKOR, Graphentechnologie

Der vorliegende Text geht auf einen Sektionsbeitrag mit dem Titel **Bild- und Quelldatenbanken als Instrumente der Forschung** zurück, der im Herbst 2015 auf der Summer School **Computing Art. Eine Summer School zur Digitalen Kunstgeschichte** an der Heidelberger Akademie der Wissenschaften präsentiert wurde. Die verschriftlichte Version mit dem neuen Titel **(Un)Ordnungen – Werkzeuge – Beziehungen: Datenbanksysteme und kunsthistorische Forschung** diskutiert in vier Kapiteln einzelne Gedanken zu Datenbanksystemen im Bereich der Geisteswissenschaften ((Un)Ordnungen), präsentiert Beispiele aus dem Bereich der Nutzung der erhobenen und in Datenbanken verwalteten Informationen (Werkzeuge), schneidet das Feld der Graphentechnologien an (Beziehungen) und schließt mit einer intensiveren Untersuchung des Open-Source-Datenbanksystems ConedaKOR und den hierzu gewonnenen Erkenntnissen aus der Praxis am Kunstgeschichtlichen Institut Frankfurt am Main und dem Deutschen Forum für Kunstgeschichte Paris.



»Alles kommt irgendwo her, alles hat Entsprechungen, alles steht mit allem in Beziehung, und der menschliche Verstand bezieht seinen Stolz daraus, einige dieser Beziehungen aufzudecken.« <sup>01</sup>

## ■ 01

Dieter E. Zimmer, in: V. Nabokov, *Fahles Feuer*, Reinbek 2008, S. 415.

## ■ 02

Der zweite Teil innerhalb dieser Sektion wurde seinerzeit von Thomas Hänsli (ETH Zürich) gestaltet.

## ■ 03

Hierfür siehe z. B.: Markus Burkhardt, *Digitale Datenbanken. Eine Medientheorie im Zeitalter von Big Data*, Bielefeld 2015.

Dieser Text versteht sich als erweiterte Ausführung des im Rahmen der 2015 durchgeführten Summer School **Computing Art** präsentierten Sektionsbeitrags. <sup>02</sup> Aufgrund dieses Entstehungszusammenhangs wird hier weniger ein breiter Einblick in die Welt der (Bild-)Datenbanken geboten <sup>03</sup>, als vielmehr zunächst punktuelle Überlegungen zu Daten(-banken) auf dem geisteswissenschaftlichen Feld angestellt werden. Im Anschluss daran erfolgt eine kurze – gleichsam prologartige – Beschäftigung mit Graphentechnologien, um dann mit der vertiefenden Betrachtung eines konkreten Systems (ConedaKOR) und der damit gewonnenen Erfahrung aus der praktischen Arbeit im Kunstgeschichtlichen Institut Frankfurt am Main und dem Deutschen Forum für Kunstgeschichte Paris zu schließen.

## G.1 (Un)Ordnungen

1964 schrieb Colonel Kenneth Swanson in **A Computer-Centered Data Base Serving USAF Personnel Managers**: »Each of our Managers should have a tool that will permit him to relate items relevant to his current problem.« <sup>04</sup> Damit ist im Grunde genommen bereits eine Menge über die Rolle von Datenbanksystemen in Prozessen gesagt und diese Aussage lässt sich auch sehr gut auf ihre Nutzung in den Geisteswissenschaften anwenden. Wenn man sich dazu noch den englischen Begriff **data base** genauer anschaut, der ja sowohl für **Datenbank** als auch **Datenbasis** steht, dann wird deutlich, dass hier eine Arbeitsgrundlage – wichtig, aber nicht das Ziel – benannt wird.

## ■ 04

A. Kenneth Swanson, *Development and Management of a Computer-centered Data Base: A Computer-centered Data Base Serving USAF Personnel Managers*, in: *System Development Corporation, A. Walker (Hg.), Proceedings of the Symposium on Development and Management of a Computer-Centered Data Base (June 10–11, 1963), Santa Monica 1963, S. 14.*

## ■ 05

Lev Manovich, *The Language of New Media*, Cambridge (Mass.) 2001, S. 218.

Die in derartigen Systemen abgelegten Daten »do not have a beginning or end [...] they are collections of individual items, with every item possessing the same significance as any other« <sup>05</sup>, wie Lev Manovich 2001 in seinem Buch **The Language of New Media** schrieb. Damit können auch Ansichten auf diese Daten jederzeit neu konstruiert werden.

## ■ 06

Horst Herold, in: *Der Spiegel, Kommisar Computer*, Nr. 27, 28. Juni 1971, S. 53. Mehr dazu bei David Gugerli, *Suchmaschinen. Die Welt als Datenbank*, Frankfurt/Main 2009.

Eine Betrachtung, die im Kern schon 1971 von Horst Herold, dem damaligen Präsidenten des Bundeskriminalamtes, angeführt wurde, um die Umstrukturierung des BKA zum »Großhirn der deutschen Polizei« <sup>06</sup> einzuleiten und sicherlich auch, um die Investitionen für die dazu benötigten Ressourcen zu rechtfertigen. Er argumentierte gegenüber dem *Spiegel*, dass die deutsche Polizei auf »Milliarden und Milliarden von Daten sitzt« <sup>07</sup> und es gelte, mit diesem Fundus »die einmalige Chance zu nutzen, per Computer den Täter in seiner vielfachen Verstrickung zu seiner Anlage, seiner Umwelt und seiner Gesellschaft zu erforschen, die Ursachen seines Handelns aufzuzeigen.« <sup>08</sup>

## ■ 07

Herold 2009.

## ■ 08

Herold 2009.

Aus dem ersten Teil von Herolds Aussage wird schon deutlich, was auch heute vielleicht in der Dateneuphorie leicht übersehen wird: Bereits bei der



## ■ 09

Christof Schöch, Big? Smart? Clean? Messy? Data in the Humanities, in: *Journal of Digital Humanities*, Vol. 2, No. 3, Summer 2013, URL <http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>.

Datenerhebung wird schon mit Bedeutung behaftetes Material verwendet, um daraus neue Sinnzusammenhänge zu schaffen. So wird auch im Rahmen von Forschungsprojekten bereits während der Zusammenstellung der Daten durch eine interessengeleitete Perspektive die entsprechende Modellierung beeinflusst, unter deren Gesichtspunkt die daraus entstehenden Daten zu bewerten sind. Christof Schöch wies in seinem Aufsatz »Big? Smart? Clean? Messy? Data in the Humanities«<sup>09</sup> auf den Unterschied zwischen Big Data (groß in der Zahl und variantenreich) und Smart Data (strukturiert und explizit) hin und ihm ist unbedingt zuzustimmen, wenn er dafür plädiert, beide Reservoirs kritisch zu betrachten und die jeweiligen Potenziale zu nutzen.

Darüber hinaus gilt es einen genaueren Blick auf die Auflösung der ordnenden Narrative zu werfen, welche bislang gegolten hatten und in deren Folge – z. B. bei der Wanderung von Objekten aus dem physischen Raum der Museen und Archive in den digitalen Raum der Datenbanken – Bedeutungskontexte durch einen Prozess der Egalisierung schlicht verschwinden. Das schließt an Überlegungen des Philosophen Henri-Pierre Jeudy an, der bereits vor 25 Jahren anmerkte:

»Das Anlegen von Informationen über die Objekte (und die Kunstwerke) macht glauben, daß alle Daten virtuell erfassbar sind, daß nichts vergessen wurde und daß ein Anschluß an die Netze völlig ausreicht, um jedes Verlangen nach Erkenntnis zu befriedigen.«<sup>10</sup>

## ■ 10

Henri Pierre Jeudy, Die Transparenz des Objekts, in: Florian Rötzer (Hg.), *Digitale(r) Schein. Ästhetik der elektronischen Medien*, Frankfurt/Main 1991, S. 175.

Die maschinelle Erzeugung von Daten und ihre entsprechende Weiterverarbeitung gibt vor, ein vom System unabhängiges Konvolut von Aussagen zu haben. Hier braucht es in den (digitalen) Geisteswissenschaften zukünftig verstärkt einen Diskurs über die diversen beeinflussenden Faktoren und die jeweiligen Methoden der Erhebung von Daten, die aus ihrem Entstehungskontext herausgelöst werden, denn »Raw data is both an oxymoron and a bad idea; to the contrary, data should be cooked with care.«<sup>11</sup>

## ■ 11

Geoffrey Charles Bowker, *Memory Practices in the Sciences*, Cambridge (Mass.) 2005, S. 184.

Nun steckt im zweiten Teil von Herolds mittlerweile über 45 Jahre alten Kommentar gegenüber dem *Spiegel* aber auch für die Jetztzeit und die Geisteswissenschaften so viel Potenzial, dass die digitalen Daten durchaus in einem »use, focus, re-focus«-Verfahren aus ihrer Entstehungsumgebung in diverse andere Kontexte überführt werden sollten, um neue Optionen der Interpretation zu eröffnen – immer mit entsprechender Reflexion über die Tatsache, dass den im Prozess der wissenschaftlichen Erkenntnisabsicht gewonnenen Daten jeweils auch ein (Projekt-)spezifikum eingeschrieben ist. Wenn oben von Verlust – bedingt durch die Auflösung des ordnenden Narrativs im physischen Raum – die Rede war, so eröffnen sich im digitalen Raum neue, vielgestaltige Narrationen. Diverse Pfade durch die Daten erlauben Anordnungen nach z. B. chronologischen, thematischen oder auch geografischen Aspekten, womit eine Auflösung der Bedeutungshöhe einhergeht, die mit der schlichten Tatsache zusammenhängt, dass Kontexte u. a. durch Suchanfragen und dem dahinter liegenden Algorithmus geschaffen werden.<sup>12</sup> In dieser Möglichkeit steckt damit zugleich auch eine Verantwortung, die man sich bei allen darauf aufbauenden Arbeiten vergegenwärtigen muss.

## ■ 12

In diesem Zusammenhang sei auch an die eröffnende Keynote auf der DHd-Tagung 2016 in Leipzig erinnert, in der Katharina Zweig vor einer algorithmischen Legendenbildung warnte. [Videoaufzeichnung unter [http://www.dhd2016.de/media/Eroeffnung\\_DHd\\_2016.mp4](http://www.dhd2016.de/media/Eroeffnung_DHd_2016.mp4), ab 01:10:30] (freundlicher Hinweis von Piotr Kuroczyński, Mainz).

## ■ 13

Gregory Bateson, *Ökologie des Geistes. Anthropologische, psychologische, biologische und epistemologische Perspektiven*, Frankfurt a.M. 1983, S. 488.

## ■ 14

Hayden White, *Auch Klio dichtet oder Die Fiktion des Faktischen Studien zur Tropologie des historischen Diskurses*, Stuttgart 1986, S. 119.

## ■ 15

[http://www.ddc-deutsch.de/Subsites/ddcdeutsch/DE/Home/home\\_node.html](http://www.ddc-deutsch.de/Subsites/ddcdeutsch/DE/Home/home_node.html).

## ■ 16

David Weinberger, *Everything is miscellaneous. The power of the new digital disorder*, New York 2007, S. 62f.

## ■ 17

Torsten Meyer, *Wahn(-) und Wissensmanagement. Versuch über das Prinzip Database*, in: Karl-Josef Pazzini, Marianne Schuller, Michael Wimmer (Hg.), *Wahn, Wissen, Institution. Undisziplinierbare Näherungen*, Bielefeld 2005, S. 244.

## ■ 18

Vergleiche hierzu den Beitrag von [Katrin Glinka](#) (→ 289) in diesem Band.

Die Daten an sich bilden hier eine, wie Gregory Bateson es nannte, zukünftige Information <sup>13</sup>, und sie verändern sich auch nicht in den unterschiedlichen Darstellungen: »was verschieden ist, sind die Modalitäten ihrer Beziehungen« <sup>14</sup>, wie Hayden White in seinem Buch *Auch Klio dichtet* ausführte. Diese prinzipielle Unabgeschlossenheit, die ungeordnete Multiperspektivität von Informationssammlungen ist charakteristisch für eine Datenbank. Die ihr inhärente Unordnung erlaubt etwas Neues, nämlich das Anordnen der Daten je nach Fragestellung, wobei keine Ordnung den Vorzug erhält. Gerade in dieser **digitalen Unordnung** sieht der US-amerikanische Philosoph David Weinberger einen signifikanten Unterschied zur analogen Welt. Er veranschaulicht dies daran, wie unterschiedlich sich das Auffinden eines Buches auf der Website eines Unternehmens wie Amazon im Vergleich zum Auffinden in einer Bibliothek verhält (die ihre Bücher nach der in den USA üblichen Dewey-Dezimalklassifikation <sup>15</sup> aufstellt):

»Amazon itself is about as far from a Dewey-compliant library as one can get. Dewey created a single way to cluster books; Amazon finds as many ways as it can. [...] Dewey prized neatness and order, bowing to the metric gods when he created a system based on multiples of ten; Amazon likes a friendly disorder, stuffing its pages with alternative ways of browsing and offbeat offers peculiar to each person's behavior. [...] Dewey's system prizes the stability that comes with the physical world – books on bookshelves, white ink on spines; Amazon prides itself on its ability to cluster and recluster instantly. These are differences not in the particularities of the categories and their arrangement but in the fundamental nature of organization.« <sup>16</sup>

Dieses Handling der Daten – nahezu frei von physischen Beschränkungen und darauf aufbauend vielfältige Möglichkeiten der Präsentation von Informationen an der Oberfläche bietend – macht die Dualität der Datenbank aus: Sie dient als Lagerort für vorhandene Informationen und bietet zugleich das Potenzial für die diversen Ausformungen. Torsten Meyer brachte das sehr pointiert zum Ausdruck: »Die Database ist amorph, sie hat keine Form, kann aber in alle möglichen Formen gebracht werden. Sie ist ein Potential an Formen.« <sup>17</sup> Die Variabilität der Darstellung und die Anschlussfähigkeit sind zentrale Charakteristika digitaler Datenbanken und definieren zugleich deren Leistungsvermögen. Wobei hierunter zum einen die – wie auch immer geartete – Schicht (Layer) mit der Sammlung umzugehen gemeint ist, also die Graphical User Interfaces (GUIs), welche die Schnittstelle zum Nutzer darstellen (menschenslesbar), und zum anderen die Möglichkeit, per Schnittstellen (APIs) zuzugreifen (maschinenslesbar). Was nun den zuerst genannten Aspekt der Nutzung betrifft, so ist neben dem Blick des Users auf das traditionelle Datenbankinterface ganz besonders auch an die diversen Möglichkeiten der Visualisierung von Daten zu denken <sup>18</sup>, die derzeit omnipräsent zu sein scheinen. Neben allen ästhetischen Reizen

■ 19

Womit hier unweigerlich an Franco Morettis Methodenbegriff Distant Reading zu denken wäre.

ergeben sich hier mächtige Werkzeuge, mit denen – gleichsam aus einer Art Vogelperspektive 19 – Netzwerke oder auch wechselnde, dynamische Formen (alternative, gleichzeitige Visualisierungen, mit Layern, interaktive Grafiken...) etc. erzeugt werden können.

## G.2 Werkzeuge

Auf der Grundlage strukturierter Forschungsdaten kommen heute Werkzeuge zum Einsatz, die in der Vergangenheit in dieser Form und Variabilität nicht verfügbar waren. Sehr treffend formulierte dies David Bollier 2010 in *The Promise and Peril of Big Data*:

»Perhaps one of the best tools for identifying meaningful correlations and exploring them as a way to develop new models and theories, is computer-aided visualization of data.« 20

■ 20

David Bollier, *The Promise and Peril of Big Data*, Washington 2010, S. 9.

■ 21

Maximilian Schich, Chaoming Song, Yong-Yeol Ahn, Alexander Mirsky, Mauro Martino, Albert-László Barabási, Dirk Helbing, *A network framework of cultural history*, in: *Science* 01.08.2014, Vol. 345, Issue 6196, S. 558–562.

Stellvertretend dafür sei die Arbeit des Kunsthistorikers Maximilian Schich genannt, der 2014 mit seinen Ko-Autoren in dem Aufsatz *A network framework of cultural history* 21 – bemerkenswerterweise in der amerikanischen *Science* erschienen – eine Untersuchung komplexer Netzwerke in der Kulturgeschichte vorgelegt hat. Auch zuvor konnte er schon Erkenntnisse der Graphentheorie für die Erschließung kunstwissenschaftlicher Datenbanken fruchtbar machen. 22 01

■ 22

Siehe z. B.: Maximilian Schich, *Complex Networks in Art Research – Exemplary Proofs of Concept*, Abschlussbericht: DFG Forschungsstipendium, Bonn 2012, URL [http://www.schich.info/pub/2012/DFG\\_final-report\\_1065\\_1-2.pdf](http://www.schich.info/pub/2012/DFG_final-report_1065_1-2.pdf).

Als weiteres Beispiel eines Kunsthistorikers, der sich moderner Visualisierungstechniken bedient (wobei er diese nicht mehr nur auf rein textuelle Daten anwendet), ist Lev Manovich zu nennen, der in seinen Arbeiten mit seinen

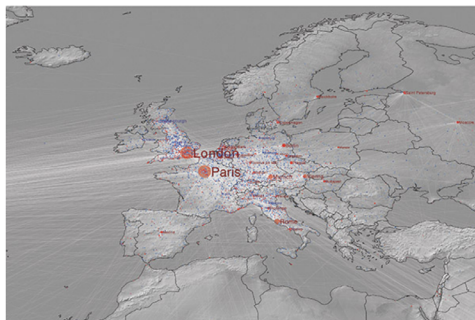
### New Report Visualizes Cultural History through “Big Data”

Scholarship | Technology

MURTHA BACA AND ANNELISA STEPHAN | JULY 31, 2014 | 1 MIN READ



Digital art history project draws on Getty Vocabulary data to map the migration of artists across 3,000 years



Birth to death migration in Europe, according to the Getty's Union List of Artist Names, cumulated over all time to CE 2012. Blue dots indicate the births of notable individuals; red dots indicate deaths. © Maximilian Schich, 2014

□ 01

Screenshot des Blogbeitrags (Anfang) *New Report Visualizes Cultural History through Big Data* von Murtha Baca und Annelisa Stephan, in: *the iris*, July 31, 2014, <http://blogs.getty.edu/iris/new-report-visualizes-cultural-history-through-big-data/>.



## ■ 23

<https://books.google.com/ngrams>.

## ■ 24

Ein von Lev Manovich – nach eigener Aussage im Jahr 2007 – eingeführter Begriff, der laut ihm wie folgt zu definieren ist: Cultural analytics is the use of computational and visualization methods for the analysis of massive cultural data sets and flows., URL <http://lab.softwarestudies.com/p/cultural-analytics.html>.

## ■ 25

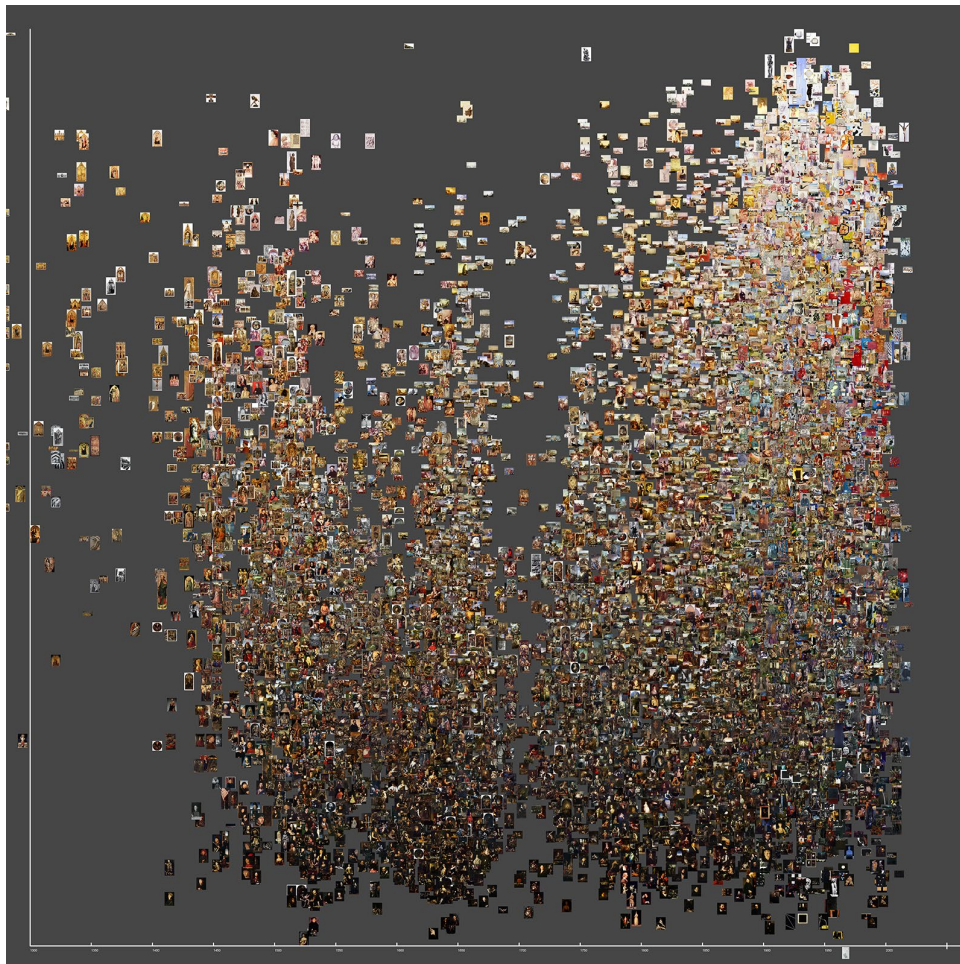
Lev Manovich, Cultural analytics: visualising cultural patterns in the era of more media, DOMUS 2009, URL [http://manovich.net/content/04-projects/063-cultural-analytics-visualizing-cultural-patterns/60\\_article\\_2009.pdf](http://manovich.net/content/04-projects/063-cultural-analytics-visualizing-cultural-patterns/60_article_2009.pdf).

## ■ 26

ImagePlot wurde von der Software Studies Initiative entwickelt und ist im Zusammenhang mit den oben erwähnten Cultural Analytics zu sehen,

Visualisierungen einen neuen Ansatz in der Erkenntnisgewinnung verfolgt. Ähnlich wie Googles Ngram Viewer <sup>23</sup> sich auf den Aspekt der Untersuchung der Frequenz von Wörtern oder Kollokationen in retrodigitalisierten Büchern fokussiert, vergleicht er große Mengen von (Bild-)Daten hinsichtlich eines Aspekts. Ein Vorgehen, welches im Kontext seines kulturwissenschaftlichen Paradigmas der Cultural Analytics <sup>24</sup> gesehen werden muss. Hiernach gilt es, sich u. a. der explodierenden Menge kultureller Daten, die im Zeitalter der New Media entstehen, mit neuen Methoden und Werkzeugen zu nähern, indem computerbasiert große Mengen kulturbezogener Daten analysiert und visualisiert werden. Denn den riesigen, allein im World Wide Web von Nutzern generierten Inhalten ist, laut Manovich, mit den althergebrachten Methoden und Werkzeugen nicht mehr Herr zu werden. <sup>25</sup> <sup>02</sup>

Mit dem von Lev Manovich mit entwickelten ImagePlot <sup>26</sup>, einem freien Visualisierungstool bzw. einem Makro für das auf Java basierende Bildverarbeitungsprogramm ImageJ <sup>27</sup>, ist es möglich, Muster in großen Bildsammlungen zu entdecken. Durch ImagePlot werden Datenpunkte in einem zweidimensionalen kartesischen Datensystem in Bezug auf zwei Variablen angeordnet. Das Besondere im Vergleich zu einem traditionellen Streudiagramm ist, dass in ImagePlot nicht nur abstrakte Datenpunkte im Diagramm dargestellt werden können, sondern die Bilder selbst. Es bietet sich daher besonders dazu an, digitalen Bildern eingeschriebene Eigenschaften wie z. B. Helligkeit, Farbton und



## □ 02

Nach Entstehungsdatum und Helligkeit angeordnete Gemälde der Bilddatenbank des Kunstgeschichtlichen Instituts Frankfurt (ConedaKOR), x-Achse: Entstehungsdatum, y-Achse: Helligkeitswertmedian, Stand November 2016, realisiert mittels ImagePlot (Sven Peter, Thorsten Wübbena).

<http://lab.softwarestudies.com/p/imageplot.html>.

■ 27

<http://imagej.nih.gov/ij/>.

■ 28

An dieser Stelle sei selbstverständlich auch auf Claus Pias, *Das digitale Bild gibt es nicht – Über das (Nicht-)Wissen der Bilder und die informatische Illusion*, in: *zeitenblicke* 2 (2003), Nr. 1 [08.05.2003], URL <http://www.zeitenblicke.historicum.net/2003/01/pias/index.html> hingewiesen.

■ 29

Laut Projektwebsite gibt es [...] no theoretical limit to the number of images that can be included in a single visualization, <http://lab.softwarestudies.com/p/imageplot.html>.

■ 30

Lev Manovich, *How to compare one million images?*, in: David M. Berry, *Understanding Digital Humanities*, Basingstoke 2012, 249–278.

■ 31

Dieses Problem ist als *Yellow Milkmaid Syndrome* in den Diskurs eingegangen. Siehe hierzu auch <https://pro.europeana.eu/post/the-yellow-milkmaid-syndrome-paintings-with-identity-problems>.

■ 32

Siehe hierzu auch den Blogbeitrag von Lena Trüper, *Das Bild als Datensatz: Ein Kommentar zu neuen Methoden der quantitativen Bildanalyse*, URL <http://blog.studiumdigitale.uni-frankfurt.de/reprometh/blog/2015/11/24/das-bild-als-datensatz-ein-kommentar-zu-neuen-methoden-der-quantitativen-bildanalyse/>.

■ 33

<https://hci.iwr.uni-heidelberg.de/compvis>.

Farbsättigung zu visualisieren. Da die Verortung der Bilder im Diagramm unmittelbar mit der menschlichen Wahrnehmung korreliert, kann die Visualisierung direkt validiert und in Bezug auf die bildimmanenten Informationen interpretiert werden. Im Gegensatz zu anderen Werkzeugen zur Visualisierung, mit denen sich hauptsächlich die Metadaten aus der jeweiligen Bilddatenbank visualisieren lassen, bietet die Kombination ImageJ/ImagePlot also die Besonderheit, in erster Linie für die Visualisierung der Bilddaten an sich <sup>28</sup> konzipiert zu sein. Darüber hinaus erlaubt es ImagePlot, auch eine sehr hohe Anzahl von Bildern zu verarbeiten, <sup>29</sup> und so hat Lev Manovich in seiner Studie zur Stilvarianz in *Mangas* <sup>30</sup> in einer einzigen Visualisierung über eine Million Bilder visualisiert.

Ein großes methodisches Problem, das sich bei der Arbeit mit den digitalen Bildern immanenten Daten in ImageJ offenbart, ist der Umgang mit der äußerst unterschiedlichen Reproduktionsqualität der Objekte in einer Bilddatenbank. <sup>31</sup> Manche Gemälde sind nur in Grautönen, manche mit, manche ohne Rahmen reproduziert, andere wiederum mit einem breiten, nicht zum eigentlichen Bild gehörenden Rand. Schon allein die Existenz von Datensätzen mit mehreren digitalen Bildern, die dasselbe Gemälde reproduzieren, jedoch in Bezug auf ihre Bildwerte völlig unterschiedlich sind, zeigt, dass eigentlich kein absoluter Bezugsrahmen existiert, der es erlauben würde, die digitalen Bilder der Datenbank miteinander zu vergleichen und daraus Rückschlüsse auf das Verhältnis der reproduzierten realen Gemälde zueinander zu ziehen. Dazu müssten die digitalen Bilder der zu untersuchenden Gemälde systematisch unter konstanten Laborbedingungen erstellt werden. <sup>32</sup>

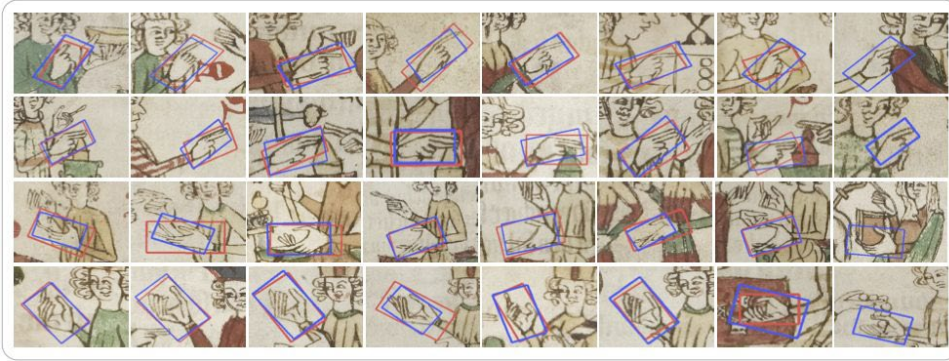
Wenngleich in diesem Vorgehen noch eine Mischnutzung aus Daten zum Bild und aus dem Digitalisat greift, wird bereits deutlich, wo das genuin Neue in der Analyse von Bilddaten zu finden ist. Anders als bei analogem Material kann eine maschinenlesbare Auswertung der Daten an sich erfolgen, wodurch sich die Interpretationsspielräume erheblich erweitern.

Wenn bei Manovichs Verfahren bereits verstärkt die textuelle (Meta-) Information durch bildimmanente Aspekte zurückgedrängt wird, so findet bei der **Automatic Image Recognition** ein rein digitales **Arbeiten im Bild** statt.

In dieser automatischen Bild- bzw. Mustererkennung steckt ein sehr großer Mehrwert für die Kunstgeschichte. Dies zeigen bereits diverse Forschungsprojekte, in denen u. a. zeitliche Entwicklungen bestimmter kunsthistorischer Motive verfolgt werden. Prominentes Beispiel dafür ist sicherlich das Vorhaben der Computer Vision Research Group <sup>33</sup> (Heidelberg Collaboratory for Image-Processing) an der Ruprecht-Karls-Universität Heidelberg, bei dem anhand illustrierter Handschriften die Kompositionalität von Bedeutung in Bild und Text aus informatischer, linguistischer und kunsthistorischer Perspektive beforscht wird. <sup>03</sup>

Aus sämtlichen oben genannten Beispielen wird deutlich, dass den Möglichkeiten der Datenvisualisierung sowie den sich daraus ergebenden Fragen, z. B. hinsichtlich der Reflexion einer Nutzung **fachfremder** Werkzeuge und der zugrunde gelegten Datenbasis, nachgegangen werden sollte. Um als Forschungsmethode eingesetzt werden zu können, muss die Visualisierung die Veränderung ihrer bestimmenden Parameter und den Zugriff auf die darunter liegende Datenbasis zulassen, denn erst dann ist eine Visualisierung ein Werkzeug, das über die reine Illustration hinausgeht und – immer komplementär zu





□ 03

Heidelberg Collaboratory for Image Processing, Computer-assisted detection and analysis of medieval legal gestures. Beispiele erkannter Gesten (blau) im Vergleich zu Ground-Truth-Begrenzungsboxen (rot). Jede Zeile zeigt dabei Erkennungen einer bestimmten Art von Geste. Die letzte Spalte jeder Zeile zeigt ein Beispiel für eine falsche Erkennung, <https://hciweb.iwr.uni-heidelberg.de/compvis/research/gestures/>.

klassisch-hermeneutischem Vorgehen gedacht – ein fruchtbares Feld eröffnet. Digitale Kunstgeschichte sollte hier als eine bestimmte Einstellung zu Transdisziplinarität, Methoden, Forschungsprozessen und zur epistemologischen Betrachtung einer qualitativen Veränderung dieser Punkte gelesen werden.

Da sich die Kunstgeschichte in den letzten hundert Jahren immer qualitativer aufgestellt hat, ist sie somit strukturell konträr zu den Ansätzen einer **quantitativen Kunstgeschichte** aufgebaut. Darüber hinaus haben wir es mit Symptomen zu tun, welche wohl typisch für Phasen technischer Umwälzung sind. Es erinnert an eine Geschichte von der Einführung des Stethoskops Ende des 19. Jahrhunderts in England, die Kathrin Passig in ihrem Artikel **Neue Technologien, alte Reflexe** erzählt: In der Londoner Times von 1895 wurde seinerzeit vonseiten der Ärzte vehement gegen den Nutzen des neuen Instruments geschrieben, und zwar, weil Diagnosegeräte an sich verpönt waren. »Ein richtiger Arzt diagnostizierte eine Krankheit mit Hilfe seiner Fachkompetenz, nicht mit Hilfe von Gerätschaften.« <sup>34</sup>

Für alle oben genannten Zugriffsformen gilt verstärkt, dass der freie Zugriff auf die jeweiligen Daten gewährleistet sein muss. Auch, weil man der oben bereits geschilderten Problematik der Verengung bei der Auswahl u. a. damit begegnen kann, dass Daten aus vielen Bereichen einbezogen werden, denn, wie Shelly Palmer es treffend formulierte: »Data is more powerful in the presence of other data.« <sup>35</sup> In der Realität gestaltet sich das häufig als aussichtsloses Unterfangen, denn obwohl z. B. bereits 2003 in der Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen die digitalen Repräsentationen von bildlichem und grafischem Material sowie wissenschaftliche Materialien in multimedialer Form mit eingeschlossen wurden <sup>36</sup>, behindert die von verschiedenen Interessengruppen <sup>37</sup> umkämpfte rechtliche Situation des Abbildungsmaterials insbesondere in der bildorientierten Forschung den produktiven Umgang mit den Digitalisaten. <sup>38</sup> Was die besitzenden, öffentlich finanzierten Kultureinrichtungen betrifft, so ist deutlich zu sagen, dass jede Ankündigung einer neu im Netz eingerichteten Sammlung von Forschungsdaten nur noch halb so erfreulich ist, wenn Wasserzeichen, fehlende API oder falsche Lizenzbezeichnungen (Stichwort: Schutzrechtsberühmung) den praktischen Nutzen für Forschung und Lehre einschränken und im Falle der hier geschilderten Methoden gar gänzlich gen Null treiben.

## ■ 34

Kathrin Passig, **Neue Technologien, alte Reflexe**, in: **Funkkorrespondenz (heute: Medienkorrespondenz)**, Ausgabe Nr. 34/14., URL <http://www.medienkorrespondenz.de/leitartikel/artikel/neue-technologien-altenbspreflexe.html>.

## ■ 35

Shelly Palmer, **Rich Data, Poor Data: What the Data Rich Do – That the Data Poor and the Data Middle Class Do Not!**, 2016, URL <https://www.linkedin.com/pulse/rich-data-poor-what-do-middle-class-shelly-palmer>.

## ■ 36

<https://openaccess.mpg.de/Berliner-Erklärung>.

## ■ 37

Das hier auch die Kunstproduzenten einzubeziehen sind, wurde u. a. in Wolfgang Ullrichs Buch **Siegerkunst – Neuer Adel, teure Lust** deutlich, wo auf zahlreichen Seiten des Buches Graufächen zu sehen sind, da die Erlaubnis zum Abdruck dort vorgesehener Abbildungen von mehreren Künstlern verweigert wurde.

## ■ 38

Siehe hierzu auch den Beitrag von **Felix Michl** (→ 255) in diesem Band.

## G.3 Beziehungen

Bei zu planenden Forschungsvorhaben mit Datenbankeinsatz gilt es, grundsätzlich die folgenden relevanten Aspekte nacheinander zu betrachten: die reale Welt, deren Abstraktion (Datenmodell), den Datenbankentwurf (Datenbankmodell), die Programmierung (Code) und das nutzbare Datenbanksystem (Abbild der realen Welt). Somit geben Datenbankinformationen zwar Auskunft über etwas außerhalb der Datenbank Befindliches (Repräsentation), andererseits legt das konzeptuelle Modell der Datenbank fest, was überhaupt als Information zählt, d. h. durch welche Informationen die Wirklichkeit beschrieben wird (Konstruktion).

Wie weiter oben bereits beschrieben wurde, gilt, dass der Blick auf die grafische Benutzeroberfläche der Datenbank nie der Blick auf die Datenbank selbst und die Strukturiertheit der inhärenten Informationen ist, sondern schlicht eine Darstellungsoption anbietet. Zu diesem Aspekt gab es in der Mitte der 1970er Jahre eine interessante Auseinandersetzung zwischen dem Mathematiker und Datenbanktheoretiker Edgar Frank Codd und dem Informatiker Charles William Bachmann, bei der u. a. heftig um die Ausrichtung der Nutzung von Datenbanksystemen durch »casual users« auf der einen Seite bzw. »technicians« auf der anderen Seite gerungen wurde. <sup>39</sup> Codd vertrat dabei die Position, dass in Datenbanksystemen die Speicherung der Daten grundsätzlich von deren Abfrage zu trennen sei, weil es den Nutzer vom Programmierer unabhängig mache. Er führt dazu aus:

»Future users of large data banks must be protected from having to know how the data is organized in the machine (the internal representation). [...] Activities of users at terminals and most application programs should remain unaffected when the internal representation of data is changed and even when some aspects of the external representation are changed. Changes in data representation will often be needed as a result of changes in query, update, and report traffic and natural growth in the types of stored information.« <sup>40</sup>

Codd entwickelte das relationale Modell, welches bis heute als Basis für die entsprechenden Datenbanken anzusehen ist, und ermöglichte damit – ganz im Sinne seines oben zitierten Diktums – Abfrageoptionen, die nicht bereits im Datenmodell angelegt sein müssten und somit auch von Personen ohne technischen Hintergrund durchgeführt werden können. Diesem relationalen Datenbankmodell, bei dem Daten in einer oder mehreren Tabelle(n) mit Spalten und Zeilen organisiert werden, wurde im Laufe der Zeit mehr und mehr der Vorzug gegenüber dem etwa zeitgleich entstandenen Netzwerkdatenbankmodell gegeben und erreichte in den vergangenen 25 Jahren eine dominante Rolle. Charles W. Bachmann gab seine skeptische Position jedoch nicht auf und interessanterweise fand sein Netzwerkdatenbankmodell, bei dem jeder Datensatz mehrere übergeordnete und untergeordnete Datensätze aufweist, mit dem Aufkommen des Semantic Web und ganz besonders in

### ■ 39

Siehe dazu auch David Gugerli, Suchmaschinen. Die Welt als Datenbank, Frankfurt/Main 2009, S. 70–88.

### ■ 40

Edgar F. Codd, A relational model of data for large shared data banks, in: Communications of the ACM, Vol. 13 No. 6, 1970, S. 377 (DOI: 10.1145/362384.362685).

jüngster Zeit mit der Entwicklung graphenorientierter Datenbanken wieder mehr Beachtung.

Was macht aber nun eine solche Graphdatenbank aus? Das ihr zugrunde liegende Prinzip ist leichter zu verstehen, als es bei vielen anderen Modellen der Fall ist. In aller hier gebotenen Kürze kann gesagt werden, dass ein Graph aus zwei Elementen besteht: einem Knoten und einer Beziehung (Kante). Jeder Knoten repräsentiert eine Entität (z. B. Person, Ort oder Sache) und jede Beziehung stellt dar, wie zwei Knoten miteinander verbunden sind. Zum besseren Verständnis hilft hier ein Vergleich mit dem Aufbau eines einfachen Satzes. Subjekt und Objekt stellen jeweils einen Knoten dar und ein Prädikat steht für die Beziehung. Beispiel: Emil (Knoten) – kennt (Beziehung) – Oskar (Knoten). Bei diesem Modell der Graphdatenbanken liegt es auf der Hand, dass, im Gegensatz zu anderen Datenbanksystemen, die Beziehungen Priorität haben. Eine Tatsache, die dem Umstand entgegen kommt, dass die in den Geisteswissenschaften <sup>41</sup> vorkommenden Daten eben nicht alleine wegen ihrer Quantität von Bedeutung sind, sondern ihren Wert aus den Verknüpfungen untereinander beziehen. Die Beziehungen liefern gerichtete, benannte, semantisch relevante Verbindungen zwischen zwei Entitäten und da eine Beziehung immer einen Start- und Endknoten hat, kann auch kein Knoten gelöscht werden, ohne nicht auch die zugehörige Beziehung zu löschen. Damit ist sichergestellt, dass eine bestehende Beziehung nie auf einen nicht existierenden Endpunkt zeigt. Nicht zuletzt durch die oben erwähnte Nähe zur Sprache lassen sich auch in einem nicht-technischen Umfeld recht schnell erste Datenmodelle erstellen, und wenn zuvor bereits mit einem Objektmodell oder einem Entity-Relationship-Modell gearbeitet wurde, erscheint das Graphmodell sehr vertraut.

Dieses bietet zwei Vorteile: zum einen können aus den Graphen komplexe Zusammenhänge abgeleitet werden und zum anderen sind die Graphwerkzeuge beim Sortieren von Daten sehr leistungsfähig, sodass sie Beziehungen und deren Relevanz in einem bestimmten Kontext zu identifizieren vermögen. Ein Beispiel: Die vom International Consortium of Investigative Journalists (ICIJ) erarbeiteten Enthüllungen zum panamaischen Offshore-Dienstleister Mossack Fonseca (**Panama Papers**) basieren auf Informationen über Verbindungen, welche in der nativen Graphdatenbank Neo4j <sup>42</sup> organisiert wurden. Für den »casual user«, um noch einmal den Begriff aus der Diskussion zwischen Codd und Bachmann aufzugreifen, wurde bei der Aufarbeitung der **Panama Papers** <sup>43</sup> auch die einfache Möglichkeit der Visualisierung des Graphen genutzt. <sup>44</sup> So bringt z. B. Neo4j ein eigenes Interface mit (Neo4j browser), welcher in einem Webbrowser läuft und mittels der leicht zu erlernenden Abfragesprache **Cypher** visuelle Antworten liefert – wenngleich die Abfragegeschwindigkeit sehr stark von der eingesetzten Hardware abhängt und natürlich auch von der Entfernung der Knoten, über die Informationen abgefragt werden sollen.

Ein weiterer Vorteil ist die einfache Erweiterbarkeit des Datenmodells. Da sich viele Forschungsfragen in geisteswissenschaftlichen Projekten erst im Verlauf des Vorhabens ergeben, muss das Datenbanksystem in der Lage sein, flexibel auf diese Anforderungen reagieren zu können.

■ 41

Einen Einblick in das breite Anwendungsspektrum der Graphentechnologie in den Geisteswissenschaften konnte jüngst auf der Tagung **Graphentechnologien. Neue Perspektiven für die Digital Humanities** (19.-20.01.2017, Akademie der Wissenschaften und der Literatur, Mainz) gewonnen werden.

■ 42

<https://neo4j.com/>.

■ 43

[https://offshoreleaks.icij.org/#\\_ga=1.95172444.1177441316.1486712642](https://offshoreleaks.icij.org/#_ga=1.95172444.1177441316.1486712642).

■ 44

Dafür wurde in diesem Fall Linkurious (<http://linkurio.us/>) eingesetzt.



## G.4 ConedaKOR

■ 45

Bruno Latour, *Tarde's idea of quantification*, London 2009, S. 155.

»Change the instruments, and you will change the entire social theory that goes with them.«<sup>45</sup>

In der praktischen Arbeit wird die Effizienz nicht nur durch die Bedingungen der Datenerhebung bestimmt, sondern wesentlich durch das Datenmodell, und eine Frage in diesem Zusammenhang wäre sicher: »Unter welchen Gegebenheiten können Datensammlungen nun als wissenschaftliche Quellen dienen und Ursprung neuer Erkenntnisse sein?«

In den Digital Humanities dominieren ja bekanntlich die Geisteswissenschaften, die mit Texten arbeiten, die bild- und objektorientierte Kunstgeschichte hingegen ist noch immer wenig präsent. Das hängt unmittelbar mit den fachspezifischen Forschungsdaten zusammen, denn wenngleich in der Kunstgeschichte Fachbibliografien und Quellenverzeichnisse erstellt werden – Priorität hat das Bild. Auch wenn das Kunstwerk der eigentliche Gegenstand der Forschung ist, hilft in den meisten Fällen ein digitaler Repräsentant, den Untersuchungsgegenstand für den Forscher im gesamten Arbeitsprozess verfügbar zu machen.

Für diese Verfügbarmachung betreibt das Kunstgeschichtliche Institut der Goethe-Universität Frankfurt für Lehre und Forschung ein digitales Bilddatenbanksystem, welches aus der früheren, analogen Diathek hervorgegangen ist. Seit 2009 wird hierfür das quelloffene Datenbanksystem ConedaKOR verwendet.<sup>46</sup> Die Software wird inzwischen in diversen Hochschulen und Forschungsinstitutionen eingesetzt<sup>47</sup> und auch als »Software as a Service«-Lösung über DARIAH-DE angeboten.<sup>48</sup>

ConedaKOR wurde für die Archivierung, Verwaltung und Recherche von Bild- und Metadaten auf einer gemeinsamen webbasierten Oberfläche entwickelt und ist konzeptuell als Graphdatenbanksystem realisiert worden. Der Datengraph stellt sich – wie oben schon beschrieben – als ein Netzwerk aus virtuellen Entitäten und deren Verknüpfungen zueinander dar. Ein Graph besteht somit im Wesentlichen aus Knoten und Beziehungen (Kanten), die jeweils zwei der Knoten miteinander verbinden. Im von ConedaKOR realisierten Graphmodell sind diese Knoten und Kanten typisiert, d. h. übergeordneten Typen zugeordnet, über welche die möglichen Verknüpfungen im Graphen eingeschränkt werden können. Durch weitere freie Attribute können die Knoten zusätzlich näher bestimmt werden.

Exemplarisch ist hier ein winziger Teilgraph aus der Datenbank des Kunstgeschichtlichen Instituts Frankfurt dargestellt. Kanten zwischen Knoten sind immer beidseitig navigierbar. Für den hier zu sehenden Graphen bedeutet dies, dass der Bamberger Altar von Veit Stoß ein Teil des südlichen Querhauses des Bamberger Domes ist und entsprechend, anders herum gelesen, das Querhaus des Doms als übergeordnetes Werk den Bamberger Altar beinhaltet.<sup>04</sup>

Aktuell enthält die Datenbank des Kunstgeschichtlichen Instituts Frankfurt 283.538 Entitäten (Knoten) und 586.212 Beziehungen (Kanten) im Datengraph.<sup>49</sup> Die wichtigsten Entitätstypen repräsentieren Medien, Werke, Personen, Literatur, Institutionen und Orte, die mit einer Vielzahl von Relationen miteinander verbunden werden können.

■ 46

Coneda UG, mit Angaben zu ConedaKOR: <https://coneda.net/> bzw. <https://github.com/coneda/kor/>.

■ 47

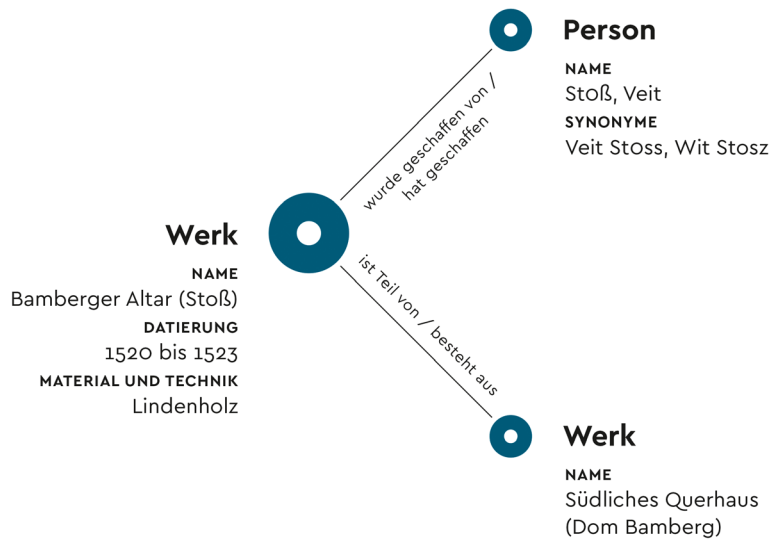
Unter anderem in Universitäten in Frankfurt am Main, Freiburg, Saarbrücken und Zürich sowie im Deutschen Forum für Kunstgeschichte Paris.

■ 48

Siehe dazu <http://dhd-blog.org/?p=7268>.

■ 49

Stand September 2017.



□ 04

Modell eines Ausschnitts aus dem ConedaKOR-Datengraph des Kunstgeschichtlichen Instituts Frankfurt.

## ■ 50

Ein großer Vorteil im universitären Kontext, da in den Hochschulrechenzentren exotische Systemkomponenten ja eher ungern gesehen werden.

## ■ 51

So wiesen dann z. B. Stefan Buddenbohm, Harry Enke, Matthias Hofmann, Jochen Klar, Heike Neuroth und Uwe Schwiigelshohn in Erfolgskriterien für den Aufbau und nachhaltigen Betrieb Virtueller Forschungsumgebungen auch zurecht auf folgenden Umstand hin: Ein werkzeugorientierter Ansatz hat gegenüber einem dem gesamten Forschungsprozess abdeckenden Ansatz Vorteile. Er konzentriert sich auf bestimmte Abschnitte des Forschungsprozesses und ist in der Lage, besser zwischen Community-Nachfrage und Infrastruktur-Angebot zu vermitteln. Er wird auch der Tatsache besser gerecht, dass generische Werkzeuge (bspw. Wiki, Clouddienste) in der Arbeitsumgebung der Forscher in der Regel bereits vorhanden sind, gegen die generische Werkzeuge einer umfassenden VRE vielleicht nur schwer bestehen können., Göttingen 2014, S. 30, URL <http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2014-7.pdf>.

## ■ 52

Suchmaschine (Java) auf Basis von Apache Lucene.

Bei einem genaueren Blick unter die Haube von ConedaKOR ist festzustellen, dass es sich hier nicht um ein natives Graphdatenbanksystem handelt. Als Storage-Backend dient MySQL und die Ablage der Medien erfolgt im Dateisystem – eine wenig revolutionäre, dafür umso robustere Lösung, was künftigen System- und DatenkuratorInnen die Nutzung dadurch deutlich erleichtert. Die Web-Applikation läuft auf der Basis von Ruby on Rails und Riot.js, sodass insgesamt nur Standard-Komponenten gewartet werden müssen.<sup>50</sup> Bei Applikationen dieser Art fallen zudem weitere Aufgaben wie die Verwaltung von Benutzern, Rechten und Sammlungen an, die mit einem relationalen Datenbankmanagementsystem (RDBMS) im Backend leichter integriert werden konnten. Das Konzept **Graph** wird in ConedaKOR in erster Linie als ontologisches Ordnungssystem verstanden und fungiert als Benutzerschnittstelle, die nicht nur Daten abfragebasiert anzeigt, sondern den Nutzer auch ermächtigt, direkt in die Modellierung einzugreifen. Die Erstellung der Graphstruktur erfolgt über ein einfach zu bedienendes Web-Interface, welches auch die Eingabe und die Anzeige umsetzt, sodass für die einzelnen Prozesse die Umgebung konstant und kohärent bleibt.

Die Erfahrungen stellen mittlerweile das monolithische Konzept einer **Virtuellen Forschungsumgebung** infrage.<sup>51</sup> Stattdessen ist eine modulare Systemarchitektur inzwischen State of the Art. Ganz in diesem Sinne ist auch der Ansatz von ConedaKOR zu sehen, welches weitere Komponenten nutzen oder Daten in diverse Umgebungen einspielen kann. So können beispielsweise Teile der Daten in Instanzen von Elasticsearch<sup>52</sup> und/oder Neo4j migriert und langfristig aktualisiert werden, sodass weitere Funktionalitäten zur Verfügung stehen. Die Rails-Applikation stellt Schnittstellen zur Verfügung, welche durch verschiedene **Verbraucher** genutzt werden können. Denn neben dem Einsatz als Datenbanksystem mit eigener Benutzeroberfläche ergeben sich auch Anwendungen, die ConedaKOR als schnittstellenstarkes Werkzeug im Backend in Erscheinung treten lassen, sei es mit einem spezifisch gestalteten Frontend, sei es als Repository, welches Daten über ein Javascript-Widget in bestehende Umgebungen einfließen lässt. Dabei erweist es sich als Vorteil, dass Inhalte u. a. in statische HTML-Seiten, WordPress-Installationen oder auch Drupal-Systeme integriert werden können, ohne die jeweilige Plattform

## ■ 53

Mittels JSON API können Inhalte in Desktop- und Mobile-Browsern angezeigt werden und über diesen Weg ist die Funktionalität der Anwendung vollständig abgebildet. Bei Vorhaben mit hohem und/oder transparentem Integrationsbedarf, bietet sich die Schnittstelle durch flexible Konfigurationsmöglichkeiten der Cross-Origin Resource Sharing-Header (CORS-Header) an. Als ein Beispiel sei an dieser Stelle nur das Projekt Wunderkammer von Günther Vogt genannt <http://wunderkammer.azdev.co/>.

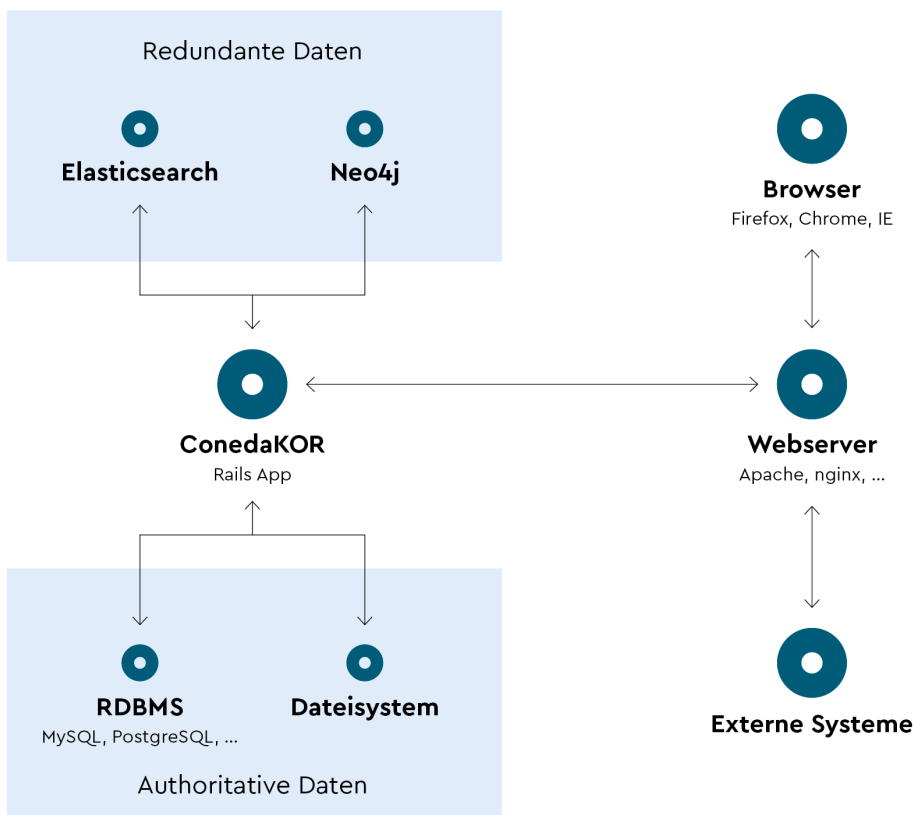
## ■ 54

<http://prometheus-bildarchiv.de/>.

anpassen zu müssen. Da hierbei an das Zielsystem keinerlei technische Anforderungen gestellt werden, können auch dann Inhalte eingebunden werden, wenn das Zielsystem nicht hinreichend veränderbar ist oder auf einer andersartigen Technologie basiert. <sup>53</sup>

Auch bietet ConedaKOR insgesamt vier Endpunkte für das Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) an, die jeweils die Inhalte von Entitätstypen, Entitäten, Verknüpfungstypen und Verknüpfungen abrufbar machen. Die Wiedergabe der Daten erfolgt hier paginiert und auf Wunsch inkrementell, sodass keine zentrierte, hohe Last entsteht. Insgesamt wird hierdurch der gesamte Nutzinhalt der Instanz geordnet abgefragt. Dieser Ansatz eignet sich vor allem dann, wenn die Reproduktion oder Zusammenführung der Daten im Vordergrund steht, also etwa für Backups, Hub-Instanzen oder z. B. auch für die Weitergabe der Daten an Harvester wie etwa prometheus – Das verteilte digitale Bildarchiv für Forschung und Lehre. <sup>54</sup>

Das Gegenstück zu diesem OAI-PMH-Repository stellt die Harvester-Komponente von ConedaKOR dar, die im Speziellen dafür Sorge trägt, dass die Abfragen idempotent sind und sich von Instanz zu Instanz transitiv verhalten. <sup>05</sup>



□ 05

Schematische Darstellung der ConedaKOR-Architektur.

## ■ 55

Siehe hierzu auch Sven Peter, *Abbildung relationaler Daten auf die Ontologie des CIDOC CRM*, 2015, URL <http://archiv.ub.uni-heidelberg.de/artdok/3454>.

Im Zuge der Anpassung und Weiterentwicklung von ConedaKOR wurde in jüngster Zeit besonderes Augenmerk auf die Möglichkeiten der Vernetzung der Instanzen unter Berücksichtigung des CIDOC Conceptual Reference Model (CIDOC CRM) gelegt. <sup>55</sup> Hier wurde eine lokale und verteilte semantische Datenintegration mittels des CIDOC CRM als Upper-Level-Ontologie entwickelt, was in der Praxis bedeutet, dass jedes Vorhaben mit einer projektspezifischen Konzeption eigener Daten samt Terminologie die Datenbank nutzen kann und sich damit nicht an ein erst einmal zu erlernendes, globales Datenschema halten muss.



Mit den Konzepten des CRM lässt sich eine (in der Regel abstraktere) Entsprechung für die spezifischen Modellierungen der jeweiligen Datenbankprojekte finden, womit die heterogenen Daten integriert werden können. Damit die konkretere Semantik der Elemente der spezifischen Datenbankmodelle nicht auf jene des abstrakten CRM reduziert werden muss, wird das CRM über den vorgesehenen Vererbungsmechanismus um die Elemente des Schemagraphen der spezifischen Datenbankmodelle erweitert. Legt man dann das CRM als standardmäßig integriertes Basisdatenmodell zugrunde, so lässt sich die semantische Datenintegration von der lokalen Datenbank auf ein verteiltes System von Datenbanken erweitern. Auf technischer Ebene wird diese Vernetzung durch die oben bereits erwähnte Implementierung des OAI-PMH realisiert. Das CRM-spezifische Verfahren ist letztlich aber losgelöst von der dahinter stehenden Softwarelösung, die ein Mapping auch nicht CRM-basiert durchführen kann.

## G.5 Schlussbemerkung

Mit der Einladung, eine Sektion auf der **Computing Art – Eine Summer School zur Digitalen Kunstgeschichte** zu gestalten, ging seinerzeit zugleich die Aufforderung der Veranstalter einher, Absolventen der Summer School sollten entweder »Rat und Inspiration für schon laufende Projekte bekommen« oder »Orientierung und Vision für die Umsetzung neuer Projekte« erhalten. Ganz in diesem Sinne kann der vorliegende Text an dieser Stelle nur Anregung für weitere Arbeiten sein. Es bleibt festzuhalten, dass als Denk- und Herangehensweise im Umgang mit Daten(-banken) eine kritische Reflektion über die jeweils vorliegenden Inhalte angebracht ist (Erhebung, Kontext etc.) und hier auch genau zwischen Daten und Informationen unterschieden werden sollte. Dieses Bewusstsein gilt es für die Repräsentation in die diversen Visualisierungsformen mitzunehmen und diese nicht als Endergebnis, sondern vielmehr als Beginn oder Hilfsmittel für eine intensivere Analyse zu begreifen. Um die Möglichkeiten einer entsprechend sensiblen und zugleich praxistauglichen Form des Datenhandlings zu gewährleisten, scheinen graphbasierte Systeme in geisteswissenschaftlichen Vorhaben von Vorteil, weil das Graphmodell den Ordnungsmodellen, die in der Forschung gebraucht werden, eher entspricht und somit eine niedrigere Einstiegsschwelle vorhanden ist. Das Streben nach der erkenntnisgeleiteten Erstellung von Referenzen und dem damit verbundenen Aufbau von Kontexten werden hier systembedingt unterstützt.