

Repositorien: vom Datencontainer zum Diskursraum

Vortrag gehalten am 17.11.2017 anlässlich der Vergabe eines Projektes zur Einrichtung eines medienwissenschaftlichen Repositoriums an die Medienwissenschaft in Marburg

Abstract: Digitale Repositorien für wissenschaftliche Texte aus einem bestimmten Fach entstehen immer mehr. Sie fördern die Sichtbarkeit und lassen sich auf dem "grünen Weg" auch für schon einmal gedruckte Arbeiten verwenden. Die These hier: Ein solches Repositorium ist nur Ausgangspunkt. Es könnte und müsste weiterentwickelt werden zu einem wissenschaftlichen Habitat, in dem die Identifikation von interessanten Texten, deren Kommentierung und Bewertung möglich wird.

Ich darf der Marburger Medienwissenschaft zunächst einmal zum Erhalt des Projektes gratulieren, mit dem die Deutsche Forschungsgemeinschaft sie damit beauftragt, für das Fach ein Publikationsrepositorium aufzubauen. Ein solches Unternehmen scheint mir – und das ist jetzt keine *captatio benevolentiae* – eine der bedeutendsten wissenschaftspolitischen und wissenschaftspraktischen Aufgaben zu sein, die in unseren Fächern am Beginn des digitalen Zeitalters anstehen. Falls Sie sich hier über den Begriff des Beginns wundern: Ich verstehe ihn so wie Jeff Jarvis, der den Stand der digitalen Technik mit dem des Buchdrucks um 1480 verglich. Mir scheint, das erlaubt einen eindrücklichen, wenn auch einigermaßen diffusen Blick auf das, was da noch alles kommen wird. Ich werde den Rest meines Beitrages dem Versuch widmen, die These von der Bedeutung der Aufgabe plausibel zu machen. Und zwar werde ich das im wesentlichen aus der Perspektive meines eigenen Faches machen, der Kunstgeschichte. Die Probleme aber dürften *cum grano salis* überall die gleichen sein. Sie werden keine Schwierigkeiten haben, das Gesagte bei allen Unterschieden auf die Medienwissenschaften zu übertragen bzw. zu adaptieren.

Mein eigenes Fach also vertritt gewöhnlich nicht die Speerspitze des technologischen Fortschrittes bei der Ausstaffierung seiner Methoden. Es ist schon von der Materialität seines Gegenstandes her ein durch und durch analoges Fach. Im Feld der Publikationsrepositorien aber stimmt dies nicht, ganz im Gegenteil, obwohl viele meiner Fachgenossen, wie manche andere Geisteswissenschaftler auch, weiterhin gerne darüber lamentieren, dass die heutigen Studierenden keine gedruckten Bücher mehr lesen. Die Kunstgeschichte dürfte zu den ersten Disziplinen gehören, die ein solches Repositorium aufgebaut haben und die damit naturgemäß weiterhin beschäftigt sind. Nun ja, wenn man genauer hinsieht, hat sie dies eher über sich ergehen lassen, als dass sie es betrieben hätte. Aber inzwischen sitzt sie doch einigermaßen aktiv mit im Boot. Betrieben hat es vielmehr eine Archäologin und ein Historiker, letzterer in seiner Funktion als Direktor einer Universitätsbibliothek nämlich hat erstere, seine Fachreferentin für den Bereich Archäologie und Kunstgeschichte ermutigt, ihrem Enthusiasmus in dem Feld freien Lauf zu lassen. Die Rede ist von Veit Probst von der Heidelberger Universitätsbibliothek, die gleichzeitig ihren Sammelschwerpunkt in der Kunstgeschichte hat. Zudem von Maria Effinger, die sich ebendort seit 2012, aber einschließlich des Vorlaufes schon viel länger mit einer geradezu sagenhaften Intensität dem Aufbau eines elektronischen Fachinformationsdienstes für die Kunstgeschichte widmet. Wenn ich richtig sehe, gibt es so etwas ja bei Ihnen auch, und zwar an der Universitätsbibliothek Leipzig, die mit Ulrich Johannes Schneider wiederum einen Protagonisten der deutschen Open Science-Bewegung zum Direktor hat. Hier haben wie einen Unterschied: In der Medienwissenschaft ist FID und Repositorium auf zwei Orte verteilt. In der Kunstgeschichte nicht, obwohl auch hier mit der SLUB Dresden ein zweiter Protagonist hinzukommt, der für

den Bereich der Kunst nach 1945 zuständig ist. Nach nunmehr 4 Jahren ist es Frau Effinger gelungen, im Rahmen von „arthistoricum“ das Art Dok genannte Repositorium so weit zu etablieren, dass es insgesamt über 5300 Schriften von schätzungsweise deutlich über tausend meist deutschen Kunsthistorikern und Kunsthistorikerinnen beinhaltet.

ART-Dok
Publikationsplattform Kunstgeschichte

UNIVERSITÄTS-BIBLIOTHEK HEIDELBERG

HOME Suche Volltextsuche Browsen Neuzugänge Publizieren

Willkommen bei ART-Dok

ART-Dok, der Volltextserver von arthistoricum.net, wird von der Universitätsbibliothek Heidelberg im Rahmen ihres von der DFG geförderten Fachinformationsdienstes Kunst, Fotografie, Design bereitgestellt. Die Publikationsplattform bietet Wissenschaftlerinnen und Wissenschaftlern weltweit die Möglichkeit, ihre Veröffentlichungen aus dem Fachbereich Kunstgeschichte kostenlos im Sinne des Open Access zu verbreiten.

Aktuell enthält ART-Dok 5.290 Publikationen mit freiem Zugang zum

open-access.net – News

- BMBF richtet Post-Grant-Fund e 04.09.2017
- SAVE THE DATE – 7. November 2017 26.07.2017
- BMBF Förderrichtlinie zu Open Access 02.06.2017
- 11. Open-Access-Tage: Call for Proposals veröffentlicht 22.02.2017

Das sind im wesentlichen Publikationen, die auf dem grünen Weg ins Open Access gefunden haben, also als Zweitveröffentlichung nach einer konventionellen gedruckten. Auch wenn das schlecht zu beweisen ist, führt diese Konversion häufig zu einem Boom in der Rezeption. Speziell bei einer Generation von Studierenden, die die Existenz einer Schrift häufig mit ihrer Präsenz im Internet gleichsetzt. Ich erinnere mich an die ebenso überraschte wie erfreute Aussage meines Doktorvaters Werner Busch, der eine entschieden verbreitetere Kenntnis seiner Schriften seit dem Zeitpunkt feststellte, an dem er sie auf Art Dok hochgeladen hatte.

Zugriffszahlen für *Das Einfigurenhistorienbild und der Sensibilitätskult des 18. Jahrhunderts* (urn:nbn:de:bsz:16-artdok-17841)

letzte Wochen		aktuelles Jahr		Vorjahr		alle Jahre	
KW	Downloads	2017	Downloads	2016	Downloads		Downloads
41	6	Dez	-	Dez	2	2017	54
40	0	Nov	-	Nov	10	2016	119
39	2	Okt	6	Okt	1	2015	85
38	1	Sep	3	Sep	7	2014	75
35	2	Aug	3	Aug	8	2013	25
33	1	Jul	6	Jul	7	2012	21
32	0	Jun	1	Jun	6		
30	3	Mai	7	Mai	6		
29	1	Apr	7	Apr	20		
		März	5	März	12		
		Feb	6	Feb	19		

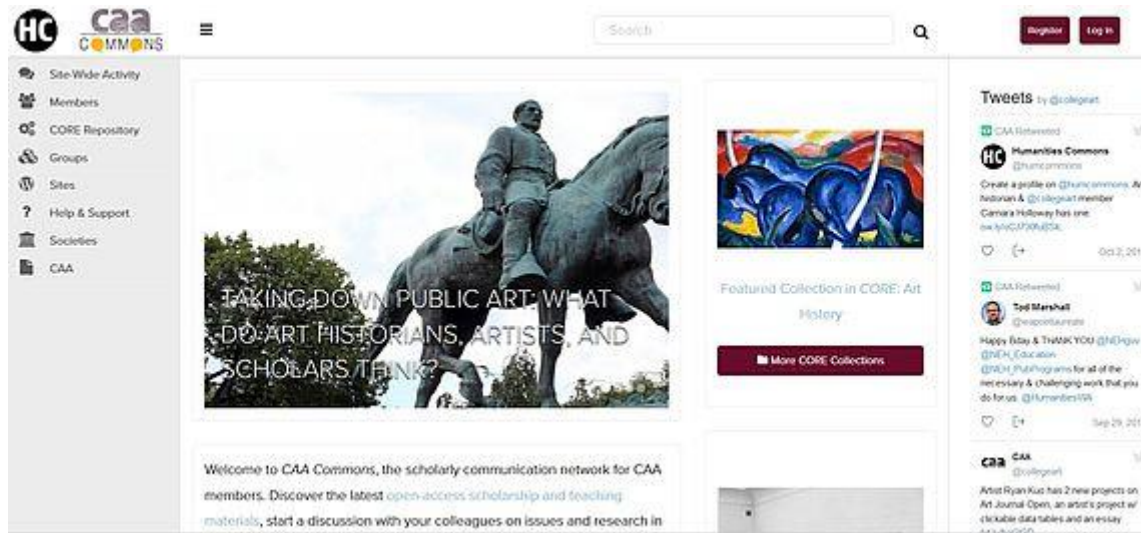
Anderen geht es ähnlich. Ein objektiveres Urteil wird man zu Schriftennutzungen aber erst fällen können, wenn man Nutzungs- und Zitationsstatistiken über einen längeren Zeitraum verfolgt und dann eventuell bemerkt, dass sie einige Zeit nach online-Veröffentlichung ansteigen. Für die hier gezeigte Statistik darf ich vergleichsweise immerhin darauf verweisen,

dass laut Wolfgang Kemp der durchschnittliche wissenschaftliche Aufsatz genau 1.4 Mal gelesen wird. Und wenn Sie jetzt einwenden, dass ein download ja noch kein Lesen ist, dann antworte ich Ihnen, dass Jürgen Osterhammels 100.000 Mal verkaufter 2000-Seiten Wälzer zur Globalgeschichte des 19. Jahrhunderts auch nicht 100.000 Mal von vorne bis hinten gelesen wurde. Oder alternativ auch, dass ein bei der UB ausgeliehener Zeitschriftenaufsatz häufig ebenfalls nicht gelesen, sondern kopiert wird. Denn mit Hans Ulrich Gumbrecht müssen wir davon ausgehen, dass man dann immerhin dafür gewappnet ist, den Aufsatz irgendwann einmal lesen zu können. 5300 Schriften ist übrigens natürlich weniger, als das auf den ersten Blick den Anschein hat, zumal fast gar keine Monographien dabei sind. Meine Vermutung wäre, dass damit erst ca. ein 100stel der gesamten Fachpublizistik abgedeckt ist, immerhin tendenziell trotz der rechtlichen Schwierigkeiten ein rezenterer Teil. Aber manches weitere ist ja auch über Nationallizenzen online greifbar bzw. aufgrund des ausgelaufenen Urheberrechtsschutzes über große Retrodigitalisierungskampagnen vorhanden. Da dies andererseits nicht das Gleiche ist und, wie zu zeigen sein wird, das Ganze auch von seiner Integration lebt, wäre zu überlegen, ob man nicht irgendwie diese Schriften auch in Art Dok importiert oder wenigstens verlinkt. In jedem einzelnen Fall ist aber von Heidelberg die Erlaubnis zur Speicherung auf art dok nachgefragt worden, um die sich der Autor oder die Autorin zu kümmern hatte. Sie erinnern sich sicherlich an die Diskussionen um die Erneuerung des Urheberrechtsgesetzes im letzten Jahr. Diese besagt, dass in Zukunft wenigstens in bestimmten Fällen nicht mehr gefragt werden muss, sondern dass das Recht auf Zweitveröffentlichung in einem Repositorium ungefragt ausgeübt werden kann.

Mit Art Dok war die Heidelberger Universitätsbibliothek nicht nur national früh dran, sondern auch international. Zumindestens glaube ich nicht, dass etwas Vergleichbares irgendwo sonst zu diesem Zeitpunkt in der Kunstgeschichte vorhanden war. Und genau genommen hat sich daran bis heute nicht viel geändert. Einer der Gründe dafür liegt in der Tatsache, dass es übergeordnete Repositorien gibt. In ihnen nimmt die Kunstgeschichte zwar nur eine Randrolle ein, da sie allen Fächern gewidmet sind, dies aber scheint den Fachvertretern und Fachvertreterinnen im Ausland bislang offenbar zu genügen. Ich nenne hier die Ihnen sicherlich bekannten researchgate und academia.edu, die den Prozess der Einbindung der Texte in die Hände des Nutzers delegieren. Beide verbinden das einfache Repositorium mit einem sozialen Netzwerk, das einen erheblichen Mehrwert gegenüber dem schlichten Datencontainer aufweist. Denn hier wird alles dafür getan, Forscher und Forscherinnen aus einem bestimmten Bereich mit solchen ähnlichen Interessess zusammenzuschalten. Researchgate macht zur Zeit von sich reden, weil der Umgang mit Verwertungsrechten dort offenbar einigermaßen lax ist. Und um academia.edu ist zuletzt ein Konflikt entbrannt, der für uns alle lehrreich sein dürfte. Mit Einführung eines

neuen Services nämlich, der den bequemen Nachweis von Zitationen eigener Artikel betrifft, führte academia.edu eine Bezahlschranke ein. Mit ihr wurde dieser Service nur Leuten geöffnet, die für ihn eine monatliche Gebühr zu leisten gewillt waren. Auf einen Schlag war die Zwiespältigkeit der Kostenlos-Kultur des Internet auch in der Wissenschaft angekommen. Mit gratis wird gelockt, und wenn man dann süchtig ist, fällt das Bezahlbeil. Geschickt war der genannte Serviceteil gewählt, denn wir interessieren uns natürlich immer sehr für das, was andere von uns selber denken. Wer ehrlich war, musste sich schon länger gefragt haben, wie denn z.B. die kalifornischen Anbieter von academia.edu ihren aufwändigen und technologisch anspruchsvollen Dienst finanzierten. Der Ärger war groß, und er zeigt an, mit welchen Problemen eine Wissenschaft zu rechnen hat, die sich privatisiert. Wir sehen das ja auch zur Zeit bei den Streitigkeiten mit Firmen wie Elsevier, die – es sollte immer wieder in Erinnerung gerufen werden - mit 35% Kapitalrendite den 4-fachen Gewinn von BMW schafft. Die paar Euro im Monat für den Zusatzservice wird man zwar verkraften, aber wenn erst einmal der Weg vorgezeichnet ist, dann werden ihn die Anbieter auch weitergehen – so die Vermutung. Er hat – wieder in der Kunstgeschichte – die schon virulenten Tendenzen befördert, ein eigenes Repository aufzubauen (CAACCommons, Abb.), das vor kurzem unter den Fittichen der College Art Association, des amerikanischen Fachverbandes, seine Tätigkeit aufgenommen hat. Dieser Verband ist zwar ebenfalls privatrechtlicher Natur, und auch bei ihm kostet es eine Menge Geld, ein solches Repository aufzubauen, aber hier ist die

Leistung eben über die Mitgliedsbeiträge gedeckt. Das wiederum heißt allerdings, dass nur Mitglieder in den vollen Genuss der Leistungen kommen. Dazu gehören übrigens auch Lehrmaterialien und die Möglichkeit, Gruppen von wissenschaftlich Gleichgesinnten zu bilden. Das ist eine Funktionalität, die man wohl den Häuptlingen der weltweiten Datenkommunikation, also vor allem google bzw. alphabet, abzuluchsen bestrebt ist. Zumindestens auf den ersten Blick ist bei CAACommmons aber noch nicht unbedingt der Teufel los. Viel mehr als 100 Mitglieder haben sich noch nicht eingetragen, Gruppen bestehen aus durchschnittlich 5 Teilnehmern.



Es dürfte kein Zufall sein, dass in Deutschland eine Bibliothek die Aufgabe übernommen hat, ein solches Repository aufzubauen, und nicht etwa der Fachverband, der bei uns VDK: Verband deutscher Kunsthistoriker heißt. Bibliotheken nämlich haben die Bedeutung der Digitalisierung in Deutschland viel früher entdeckt und sie dann auch in ihre Taten einfließen lassen, als die (Geistes)Wissenschaft selber. Es wäre allerdings dringend zu wünschen, dass dieser Verband eine Aufgabe übernehme, die auch in der Ankündigung dieser kleinen Tagung Erwähnung findet, die nämlich, die mit dem Begriff der Akzeptanzschaffung benannt ist. Bei Art Dok lief das zunächst über Mund zu Mund Propaganda, einige Enthusiasten haben kräftig die Werbetrommel gerührt und Kollegen beknet, doch auch ihre Schriften zur Verfügung zu stellen, frei nach dem Motto: die meisten müssen zu ihrem Glück gezwungen werden. Inzwischen kommen viele von sich aus auf die Idee, hier aktiv zu werden. Aber viel besser noch wäre es natürlich, wenn der genannte VDK hier einmal einen Aufruf starten würde. Das wäre also mal gleich mein Tip für Sie: Machen Sie Ihrem Fachverband den Mund wässrig und überzeugen Sie ihn davon, dass der grüne Weg inzwischen unter bestimmten Bedingungen, also der Einhaltung von Karenzzeiten, durchaus legal ist. Oder dass eine freundliche Nachfrage beim Verlag auch in den Fällen Erfolg versprechend ist, in denen eigentlich kein Rechtsanspruch besteht. Und noch besser: Dass auch eine native open access Publikation nicht ehrenrührig ist. Reputation ist auch in diesem Feld alles, und diese dürfte in erster Linie von einem solchen Fachverband ausgehen. By the way: Ich sehe meine Aufgabe hier nicht darin, Ihnen die Notwendigkeit des open access bzw. der online-Veröffentlichung zu erklären, denn ohne Internet ist ein Repository ja ganz sinnlos. Es ist zwar in der Tat so, dass man selbst bei der Behandlung komplexer Probleme der digital humanities zumindestens in meinem Fach schlussendlich niemals danach gefragt wird. Stattdessen geht immer alles wieder von vorne los, und es wird nach der grundsätzlichen Sinnhaftigkeit des Digitalen oder gerne alternativ auch nach den fiesen copyright-Restriktionen gefragt. Diese sind natürlich in

der Kunstgeschichte besonders virulent und werden gerne einmal vorgeschoben, um das Desinteresse an online als von außen erzwungenes vorzugaukeln. Aber ich gehe mal davon aus, dass das in den Medienwissenschaften vielleicht doch inzwischen abgehakt ist. Ausdrücklich ausschließen will ich dabei allerdings nicht die Frage, wie sehr sich unsere Fächer unter dem Einfluss des Digitalen grundsätzlich verändern werden. Allerdings schiene mir eine solche Diskussion an dieser Stelle wenig sinnvoll.

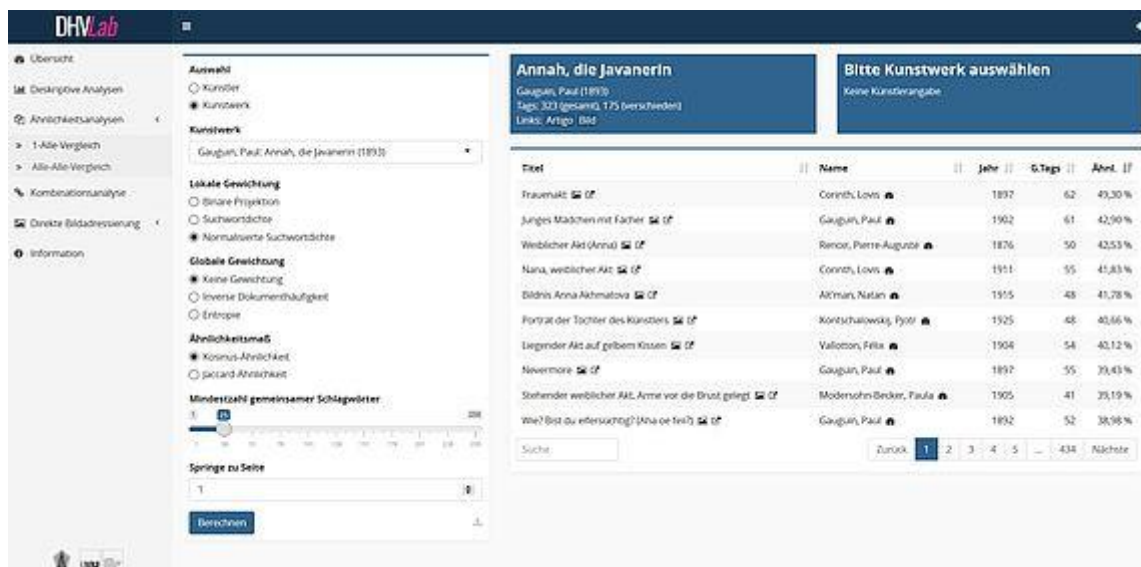
So, nachdem ich Ihnen jetzt mal eben vorgeschrieben habe, was nachher in der Diskussion nicht passieren soll, komme ich zu den Repositorien zurück. Bibliotheken – das wäre die wohl kaum zu wiederlegende These - haben letztlich das bessere know how und sind sicherlich am besten in der Lage, die komplexen technischen und fachlichen Voraussetzungen für ein solches Repositoryum zu garantieren. Ich nehme daher mal an, dass Sie das hier in Marburg auch irgendwie in Zusammenarbeit mit einer Bibliothek planen. Wenn man dann noch das strategische Geschick einer Bibliothek wie der Heidelberger besitzt, dann betrachtet man dort die Alternativen aus dem privatwirtschaftlichen Bereich auch nicht so sehr als Konkurrenz, sondern als Mitstreiter. Frau Effinger in jedem Fall empfiehlt bei jeder sich bietenden Gelegenheit, die überragenden Qualitäten von academia.edu oder researchgate in Richtung Netzworkebildung in Anspruch zu nehmen. Gleichzeitig hält sie es für sinnvoll, dann auf die Volltexte in arXiv zu verweisen, um sich nicht mit Haut und Haaren in den Einflussbereich eines amerikanischen oder auch deutschen kommerziellen Anbieters zu begeben. Denn von dem weiß man weder, ob er in 10 Jahren noch existiert, noch, wie teuer sein Service in Zukunft werden wird. Gerade für die Langzeitarchivierung dürften die Kompetenzen der Bibliothek unübertroffen sein und bleiben. Das wird ja im übrigen auch bei der Deutschen Forschungsgemeinschaft so gesehen.

So sehr ich diese Strategie für zielführend halte, so sehr würde ich doch auch dafür plädieren, dass wir uns im Rahmen einer öffentlichen Institution unsere Services verstärkt selber aufbauen sollten, gerade auch im Hinblick auf die Dimensionen, die über den reinen Containercharakter von Literatur in Richtung auf ein social medium für Wissenschaftler hinausgehen. Das meine ich in meinem Vortragstitel mit den Begriffen Datencontainer und Diskursraum. Wir sollten zusehen, mit dem Repositoryum eine Sphäre zu entwickeln, die langfristig zum Produktions- wie zum Diskussionsraum, das heißt z.B. auch zum Bewertungsraum für die wissenschaftliche Produktion wird. Besonders dann, wenn die Texte als maschinenlesbare Volltexte vorliegen, ergeben sich hier weitreichende Möglichkeiten, von denen gleich zu reden sein wird. Voraussetzung dafür ist allerdings, dass wir mehr und mehr auch den goldenen und nicht mehr nur den grünen Weg der Veröffentlichung beschreiten, also gleichsam born digital publizieren. Das ist natürlich ein weiter Weg, aber er scheint mir auch für die Wissenschaft essenziell, wenn nicht lebenswichtig. An dieser Stelle sorgt man sich immer um das Wohl der Verlage, aber diese scheinen mir in dem angestrebten System durchaus eine Rolle spielen zu können, wenn auch eine andere als bisher. Aber das ist ein anderes Thema, so dass ich das hier nicht weiter verfolgen will. Mal abgesehen davon, dass es an dieser Stelle immer und grundsätzlich Ärger gibt.

Wenn man den Datencontainer in einen Diskursraum weiterentwickeln will, so beginnt das mit den Vernetzungsperspektiven. Eine Mitarbeiterin von mir namens Stefanie Schneider, Statistikerin und Informatikerin, hat im Rahmen unseres dhvlab ein Analysezentrum für die Kunstgeschichte programmiert, in dem sie auf der Basis des open source Statistik-Programms R unterschiedliche Formen von Ähnlichkeiten zwischen Bildkunstwerken messen kann.



Das dhvlab, hier nur in Parenthese, ist eine Umgebung, mit der es ermöglicht wird, digital gestützte Seminare in der cloud abzuhalten, also ohne Sorge um vergleichbare digitale Voraussetzungen bei den Teilnehmern. Bei dieser Ähnlichkeitsanalyse werden die in einem crowdsourcing-Spiel namens artigo für Kunstwerke vergebenen tags in gerichtete Vektoren umgerechnet und in ihrer Summe mit denen von anderen Kunstwerken verglichen, so dass sich deren Nähe oder Ferne ergibt.



Der Witz für unseren Zusammenhang: Ähnlichkeiten lassen sich für Kunstwerke genauso wie für Texte bestimmen. Und zwar ohne, wie bislang meist praktiziert, dass eigene Metadaten vergeben werden, sondern im direkten Zugriff auf die Originaltexte. Unter der Voraussetzung natürlich, dass diese maschinenlesbar im Volltext vorliegen. Ein Testfall war für mich ein wenig peinlich. Es ergab sich nämlich eine Ähnlichkeit von 85% zwischen zweien meiner Aufsätze in artdok. Dieser Wert ist extrem hoch und bedeutet bei den Kunstwerken im Normalfall, dass es sich um identische Arbeiten handelt, die sich nur im Blick auf ihre unterschiedliche Reproduktion unterscheiden. Identisch waren meine beiden Aufsätze nicht, aber in dem einen hatte ich doch einen ganz wesentlichen Teil vom anderen abgeschrieben, weil mir in meiner Not und unter Zeitdruck zu dem Thema nichts Neues mehr eingefallen war. Aber egal, die Eignung des Verfahrens war auf diese Weise bewiesen. Und wir kennen

es ja auch aus der sogenannten Plagiatsoftware, die ähnlich funktioniert wie das, was meine Mitarbeiterin gemacht hat.

Wozu das Verfahren? Ich denke, wir können es auch für unseren Zusammenhang nutzbar machen. Wenn jemand seinen oder ihren Text in einem Repository veröffentlicht, dann verfolgt er damit verschiedene Zwecke. Einmal will er oder sie natürlich gefunden werden. Ausweislich meines eben zitierten Doktorvaters funktioniert das auch ziemlich gut. Aber er oder sie möchte auch mit Gleichgesinnten kommunizieren, also mit Personen, die ähnliche Forschungsgegenstände traktieren. Wir alle möchten Teil einer „république des lettres“ werden, und sei es nur einer Republik, die sich mit unseren speziellen Forschungsgegenständen beschäftigt. Unter anderem darin begründet sich ja auch der Erfolg von Systemen wie academia.edu und researchgate. Mit der beschriebenen Ähnlichkeitsanalyse funktioniert das perfekt, wenn auch rein rechnerisch-quantifizierend durchgeführt. Sobald ich einen Aufsatz hochgeladen habe, sagt mir das System, welches die in ihm enthaltenen ähnlichsten Aufsätze sind. Und die werde ich dann tunlichst lesen. In einem zahlenmäßig begrenzten Umfeld mag das trivial sein, denn ich kenne ja die Leute, die über ähnliche Dinge arbeiten wie ich, oder ich kenne doch eben die Texte. Außerdem ist so etwas ja auch leicht über normale bibliographische Angaben herauszufinden. Aber das Potential könnte sich vor allem in einem mittleren Bereich entfalten, also nicht etwas bei höchster Ähnlichkeit, sondern bei vielleicht 50% (das müsste alles noch getestet werden). Denn hier stoße ich auf Dinge, die über bibliographische keywords etc. nicht mehr zu erschließen sind, die aber trotzdem hinreichend nahe dran sind, um nützlich zu sein. Und: das System entwickelt seine Macht umso mehr, je umfangreicher die verwalteten Daten werden. Daher wäre es wichtig, dass möglichst bald eine Vernetzung mit internationalen Angeboten erfolgte. Denn auch für nicht-deutsche Texte lässt sich das Verfahren verwenden, wenn man bedenkt, wie gut inzwischen die automatischen Übersetzungen geworden sind. **Das** haben wir getestet, und es funktioniert umso besser, als die bei der automatischen Übersetzung immer noch knifflige Dimension der Satzbildung ja bei einem reinen word-count wegfällt. Also noch einmal konkret: In Zukunft könnte das Repository unserer 5.300 Texte so gestaltet werden, dass zu jedem Einzeltext eine Rangliste der ihm ähnlichsten angezeigt wird. Das ist dann wie bei amazon und co: Menschen, die sich für das von ihnen gekaufte Buch interessiert haben, haben auch folgende Bücher gekauft. Hier allerdings sind es wieder Metadaten, die ausgewertet werden.

Das, was mit einer solchen Umgebung befördert wird, ist aus meiner Sicht typisch für die Wirkungsweisen der Digital Humanities. Der Computer ist häufig kein direktes Erkenntnisinstrument, sondern ein Beförderer dessen, was mit dem schönen englischen Begriff der serendipity beschrieben wird. Er bietet mir Material an, das eine Ähnlichkeit zu meiner Fragestellung aufweist und diese damit produktiv erweitert. So gesehen könnte man den Rechner als Serendipitomaten sehr adäquat beschreiben und ihm damit berechtigt eine Rolle zuweisen, die seinem üblichen Image als „stahlhartem Gehäuse der Moderne“ so ganz und gar nicht entspricht.

Sobald solche Formen der Relationierung implementiert sind, ist schon einmal ein wesentlicher Schritt auf dem Weg zu einem wissenschaftlichen Diskursraum getan, und dieser wird umso effektiver, je mehr er sich in einem interdisziplinären Umfeld bewegt. Denn die Ähnlichkeitsanalyse kennt keine irgendwie gearteten überkommenen Zugehörigkeiten, sondern führt mich plötzlich zu einem Text, der in einem völlig anderen Fach verfasst wurde, aber zu meiner Fragestellung passt. Gerade bei unterschiedlichen Fachzugehörigkeiten zeigen klassische bibliographische Methoden doch ihre Schwächen. Eine Vision, werden Sie sagen, weil wir dann nicht nur in Richtung auf eine internationale Fachvernetzung gehen müssten,

sondern auch auf eine interdisziplinäre. Aber wenn wir die Einzelrepositorien erst einmal aufgebaut haben, und zwar so, dass sie interoperabel sind, dann dürfte die Verbindung nicht mehr schwer sein. Es ist wie immer bei diesen elektronischen Aufgaben. Das Wachstum verläuft zunächst mühsam und kaum merkbar, um dann an einem bestimmten, aber schlecht bestimmbar kritischen Punkt exponentiell anzusteigen. Im übrigen ist der Zusammenschluss auch überlebenswichtig, um die erwähnte kritische Masse zu erreichen, denn es ist hier wie an vielen Stellen: fehlendes Wachstum könnte auch ein Todesurteil sein. Erinnern Sie sich bitte an das eben erwähnte CAACommons: Wenn da nicht innerhalb des nächsten Jahres ordentlich was passiert, ist das Angebot mausetot.

So richtig zum Diskursraum aber wird das Repositorium erst dann, wenn ich nicht nur Texte entdecke, die meinem Interesse entsprechen, sondern wenn ich mich mit deren Autoren oder Autorinnen in Auseinandersetzung begeben. Ein erster Schritt bei dieser Auseinandersetzung ist die Bewertung. Damit sind wir bei dem wichtigen Feld der Evaluierung, das in Teilaspekten gerne auch unter dem Oberbegriff des peer reviewing läuft. Unter analogen Bedingungen war dieses peer reviewing, und zwar eines, das über Veröffentlichung oder nicht Veröffentlichung entschied, notwendige Praxis, weil das endliche Papier der überkommenen Zeitschrift die Beschränkung des Publikationswürdigen erzwang. Das ist Ihnen als Medienwissenschaftlern ein bekanntes Argumentationsschema, wenn es auch klassischen Geisteswissenschaftlern allzu materialistisch vorkommen mag. Wenn peer review inzwischen zu einem Fetisch geworden ist, den insbesondere auch die Herausgeber von open access Zeitschriften permanent vor sich hertragen, dann ist das ein Versuch, das Reputationsmanko von online gegenüber gedruckt zu kompensieren. Meine These – und meine Hoffnung – wäre, dass das neue Medium Internet irgendwann die Merkmale des alten Mediums abstreifen wird, also auch die des peer reviewing. Denn elektronischer Speicherplatz ist im Überfluss vorhanden, Begrenzungen sind zumindestens in dieser Hinsicht überflüssig. Wichtig wäre eventuell nur die Hierarchisierung der Qualität, die dann aber post festum vorgenommen werden kann, und zwar unter anderem von eben denjenigen, die sich in einem solchen Repositorium organisieren.

Das alte peer review Verfahren ist ja auch unabhängig von der Medienfrage inzwischen deutlich unter Beschuss geraten. Untersuchungen haben gezeigt, dass es von Willkürlichkeiten geradezu strotzt, dass es aus naheliegenden Gründen eine Tendenz zeigt, das Alte gegenüber dem Neuen zu privilegieren, und dass es im Verdacht steht, Hierarchien und Machtstrukturen zu erhalten. Der bekannte Bonner Philosoph Markus Gabriel – um hier nur ein rezentes Beispiel zu zitieren - hat in einem Interview der Süddeutschen Zeitung vor ein paar Wochen zu Protokoll gegeben: „Peer-Review in der Philosophie ist die Vortäuschung von Qualitätskontrolle.“ Ich würde Gabriel hier nur in der Einschränkung auf die Philosophie widersprechen und zuspitzen: „Peer-Review in der Wissenschaft ist die Vortäuschung von Qualitätskontrolle.“ Außerdem kostet sie einen Haufen Zeit. Ein nachgeschaltetes peer reviewing erlaubte hier eine erheblich größere Transparenz und würde im übrigen die Arbeit auf viel mehr Schultern verteilen. Dabei können die Kriterien der Evaluierung vielfältig definiert werden. Downloadstatistiken, Gewichtungen der Bewerter, Kommentare in diskursiver Form oder mittels Sternchenvergabe usw. usf. Der Phantasie sind da keine Grenzen gesetzt und technisch ist es auch kein Problem, diese Kriterien in unterschiedlichster, eventuell auch noch frei wählbarer Kombination und Gewichtung einzusetzen. Wie stark hier anonymisiert zu verfahren ist oder nicht, wäre auszuhandeln. Das Ergebnis aber könnte einen Ausgleich herstellen zwischen dem berechtigten Interesse einzelner Autorinnen und Autoren, seine oder ihre Gedankengänge der Öffentlichkeit zu präsentieren und der Notwendigkeit, die naturgegeben begrenzte Aufmerksamkeit jedes einzelnen Wissenschaftlergehirns nicht überzubeanspruchen. Schlecht bewertete oder selten

heruntergeladene Beiträge würden im long tail verschwinden und die Breite der Wissenschaftler nicht weiter belästigen. Das ist das Ergebnis eines wichtigen, zunächst einmal wirtschaftswissenschaftlich orientierten Buches des amerikanischen Internet-Theoretikers Chris Anderson. Gleichzeitig aber blieben sie über die exakte Adressierbarkeit bei spezifischen Frageinteressen wiederauffindbar, frei nach dem Motto: auch in einem weniger bemerkenswerten Beitrag kann doch eine wichtige Information stecken, die bei Nicht-Veröffentlichung eben auch nicht das Licht der Welt erblicken würde. Sobald aber letzteres z.B. bekannt würde, könnte es auf der Leiter des long tail wieder nach oben klettern und stärker in die Aufmerksamkeitsspanne der Wissenschaft geraten. Das Ganze würde damit zu einem atmenden System aufwachsen, das hoffentlich meine anfängliche Behauptung rechtfertigt, wir hätten es bei diesen Repositorien, wenn man diese nur richtig anpackt, mit „einer der bedeutendsten wissenschaftspolitischen und wissenschaftspraktischen Aufgaben“ der Gegenwart zu tun.

Publizieren, verorten, bewerten: mit diesen wichtigen drei Phasen des wissenschaftlichen Prozesses ist aber das Ende der Fahnenstange noch immer nicht erreicht. Ein wenig fehlt noch die Diskussion, obwohl die in der Bewertung, je nachdem, wie diese konkret vorgenommen wird, natürlich schon mit enthalten ist. Gut geeignet für ein solches internet-gestütztes Repitorium ist auch der Vorgang der Annotation.

Der wissenschaftliche Prozess ist im Kern ein Ablauf von Anmerkungen, Kommentaren, Ergänzungen, Kritik. Jeder Aufsatz referenziert zwar in erster Linie auf seinen Gegenstand, aber er tut dies explizit oder implizit, indem er andere Referenten kommentiert. Auch dort nämlich, wo er sich scheinbar direkt auf die historische Außenwelt bezieht, vollzieht er dies immer über deren Wahrnehmung durch andere. Die Chance (zugegeben, vielleicht auch die unabsehbare Gefahr), die sich in einem Repitorium wie dem von mir hier visionierten ergibt, wäre es, das Gewebe der Assoziationen enger zu knüpfen. Das würde auch für eine sehr grundlegende Form der wissenschaftlichen Produktion gelten, die das beschriebene Verhältnis wörtlich nimmt, das eben der Annotation. Bei ihm wird der Kommentar explizit, ganz unmittelbar auf das Kommentierte bezogen, und dafür eignet sich das Internet natürlich besonders gut. Es gibt inzwischen eine Vielfalt von software-Produkten, die diesen Annotationsprozess in allen seinen Filiationen perfekt unterstützt. Und zwar so, dass er, wenn gewollt, auch öffentlich sichtbar ist bzw. in frei bestimmter Tiefe öffentlich sichtbar sein kann. Wenn den Leser nur die Kommentare der Ordinarien interessieren, dann stellt er seinen Browser eben so ein, wiederum unter der Voraussetzung, dass diese ihren Status im System entsprechend markiert haben. Im Annotieren wird das wissenschaftliche Arbeiten gleichsam zur Kenntlichkeit entstellt und als ein kollektiver Prozess geoutet. Die Voraussetzungen für ein Verfahren wie das eben beschriebene sind bei der augenblicklichen Verfasstheit der Wissenschaft allerdings nicht gut. Der leider viel zu früh verstorbene Andreas Thielemann hat schon vor über 10 Jahren (also in Internet-Perspektive vor einer Ewigkeit) an der Bibliotheca Hertziana, einer Vorzeiginstitution der deutschen Kunstgeschichte, ein solches System eingeführt, konnte aber nur wenige Erfolge erzielen. Der Grund dafür war einfach. Er lag in der Tatsache begründet, dass die bei ihm produzierten Annotationen nicht ernst genommen wurden, nicht ernst genommen werden konnten, weil sie gar nicht zitierbar waren bzw. in Bibliographien nicht aufgenommen werden konnten. Es ist auch sehr die Frage, wie das überhaupt funktionieren soll. Ich würde daraus aber nicht die Konsequenz ziehen, neue Verfahrensweisen einfach abzubügeln und beim Alten zu bleiben. Vielmehr wären die Evaluierungsverfahren der neuen Wirklichkeit anzupassen, so dass auch jemand ernst zu nehmen ist, der den Schwerpunkt seiner Tätigkeit auf die Annotation legt. Durchführbar wäre dies natürlich wiederum im Rahmen eines zum Diskursraum erweiterten Repositoriums, in dem auch die Annotation in der beschriebenen Mischung von automatisierten und manuellen

Prozeduren ihren verdienten Platz erhalte. Konkret hieße dies, dass etwa bei Berufungsverfahren nicht mehr irgendwelche ausgedruckten Lebensläufe zugrunde gelegt würden, sondern (wer hätte es gedacht) – unter anderem eben auch unser Repositorium.

Sie werden vielleicht denken, dass mit dem zum Diskursraum entwickelten Repositorium ein geschlossenes Monstrum wissenschaftlicher Produktion entstünde, ein totalitäres Instrument der Überwachung, ein Gefängnis, dem niemand seiner Gestalter mehr entkommen kann. Um hier ein wenig Entwarnung zu geben, möchte ich etwas anmerken, was mir bei allem zuzutreffen scheint, was in den Digital Humanities produziert wird. Der Charakter von deren Ergebnissen ist immer durch den Begriff der „Empfehlung“ zu beschreiben, die von einem menschlichen Verstand zu gewichten ist. Damit ist aber schon eine Menge geleistet. Auszugehen ist dabei von dem klugen Diktum des italienischen Internet-Philosophen Luciano Floridi, der in seinem Buch über die „4. Revolution. Wie die Infosphäre unser Leben verändert“ schrieb: Der Computer ist und bleibt eine syntaktische Maschine, so wie der Mensch eine semantische Maschine ist. Regen Sie sich bitte nicht darüber auf, dass Floridi den Menschen als Maschine beschreibt. Viel wichtiger ist, dass er den Unterschied zum Computer als einen von Syntax und Semantik beschreibt. Damit ist auch ein klares Votum zur Zukunft der künstlichen Intelligenz gefasst: Der Computer wird niemals in seiner Funktionsweise an den Menschen heranrücken. Das heißt nicht, dass er nicht viele Dinge auf seine Art und sehr viel besser simulieren kann, die sehr stark nach menschlicher Intelligenz aussehen. Der Unterschied wird bleiben. Für unsere Repositorien heißt das noch einmal, dass alles, was in ihnen ausgeführt wird, immer nur den Charakter von Angeboten hat, die letztlich von einer menschlichen Intelligenz, an dieser Stelle schlicht dem Leser, zu überprüfen sind. Die große Gefahr, die wir bei der Quantifizierung der Wissenschaften eingehen – und das, was ich hier beschrieben habe, ist natürlich eine Form von Quantifizierung – ist, dass sie kritiklos als Grundlage angenommen wird. Dass also ein Berufungsverfahren z.B. nur noch oder doch zu einem erklecklichen Teil über den berühmt-berüchtigten Hirschfaktor eines Kandidaten entschieden wird, also seine Zitationsstatistik. So weit ich höre, ist das bei den Naturwissenschaften durchaus schon gängig. Entscheidend sollte in meinen Augen weiterhin das Urteil des Lesers und die Abwägung in der Diskussion bleiben. Dabei sind die quantitativen Merkmale ein wichtiger, aber eben unhintergebar fragmentarischer Orientierungsmarker, mehr aber auch nicht.

Meine Damen und Herren, mir ist völlig klar, dass Sie mit Ihrem Repositorium erst einmal sehr viel bodenständigere Probleme zu lösen haben, als die, die ich Ihnen hier vor allem in den letzten 10, 15 Minuten vorgetragen habe. Und ich bin auch nicht so sicher, ob meine Vorstellungen, die mehr mit dem Diskursraum als dem Datencontainer zu tun haben, die Bereitschaft Ihre Fachkollegen eher fördern oder eher behindern, ihre Texte dem Repositorium zur Verfügung zu stellen. Aber man muss sich doch schon klar machen, dass im Internet eine Information nicht einzuhegen ist, wie das vielleicht noch im Analogen stärker der Fall war. Wer A sagt, muss auch B sagen. Wer meint, dieses Repositorium einfach nur als eine Art Sicherheitskopie verwenden zu können, wird sich noch wundern, was mit seinen Texten auf die Dauer dort alles so passiert, ohne dass er oder sie es intendiert oder sich überhaupt nur vorgestellt hätte. In meinen Augen überwiegen hierbei die Vorteile die Nachteile. Aber im Grunde sind wir hier mitten drin in der im wahrsten Sinne des Wortes weltbewegenden Diskussion um die positiven oder negativen Folgen von Big Data.