

SEMANTISCHES PUBLIZIEREN  
THEORETISCHE GRUNDLAGEN & ANFORDERUNGEN  
IM INTERDISZIPLINÄREN WISSENSCHAFTSNETZWERK

NORA SCHMIDT



nora.schmidt@univie.ac.at  
<http://orcid.org/0000-0002-7105-9515>  
7. Dezember 2014  
Version 1.1



Diese Arbeit ist auch in HTML verfügbar, sowie als für doppelseitigen Druck optimierte PDF-Version: urn:nbn:de:kobv:11-100221877, erschienen in den Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft, hrsg. v. Konrad Umlauf, Band 375, ISSN: 1438-7662; auch als gedrucktes Buch zu bestellen.

## Abstract – Zusammenfassung

### *Semantic Publishing in an Interdisciplinary Scholarly Network. Theoretical Foundations and Requirements*

*The study examines preconditions to adopt semantic web technologies for a novel specialized medium of scholarly communication that – also interdisciplinary – enables the synchronicity of publication and knowledge representation on the one hand and the dynamic bundling of assertions on the other hand. Therefore it is first of all necessary to determine a concept of “(scholarly) publication” and of neighbouring concepts. These considerations are fertilized by theories that can be related to the radical constructivism. Therefrom derives a critique of the mainstream of knowledge representation that resigns to being not able to represent the dynamics of knowledge. Finally the study evinces a conceptual outline of a technical system that is built upon the known concept of nanopublications and is called “scholarly network”. The increased effort while publishing in the scholarly network is outweighed by the benefits of this publication medium: It may help to render research outputs more precisely as well as to raise their connectivity through reducing the complexity of assertions. Beyond that it would generate an openly accessible and finely structured discourse archive – a wide participation provided.*

Die Studie untersucht Voraussetzungen der Nutzung von Semantic-Web-Technologien für ein neuartiges Spezialmedium der wissenschaftlichen Kommunikation, das – auch interdisziplinär – einerseits die Gleichzeitigkeit von Publikation und Wissensrepräsentation und andererseits die dynamische Bündelung von Aussagen ermöglicht. Dafür ist zunächst die Bestimmung eines Begriffs der »(wissenschaftlichen) Publikation« und von benachbarten Begriffen erforderlich. Befruchtet werden diese Vorüberlegungen durch Theorien, die dem radikalen Konstruktivismus zuzuordnen sind. Daraus leitet sich dann eine Kritik am Mainstream der Wissensrepräsentation ab, der sich damit abfindet, die Dynamik des Wissens nicht repräsentieren zu können. Am Ende der Studie steht eine konzeptionelle Skizze eines technischen Systems, das auf dem bekannten Konzept der Nanopublikation aufbaut und »Wissenschaftsnetzwerk« genannt wird. Trotz des wahrscheinlich erhöhten Aufwandes beim Publizieren im Wissenschaftsnetzwerk überwiegen die Vorteile dieses Publikationsmediums: Es kann helfen, Forschungsergebnisse zu präzisieren sowie durch eine Reduzierung der Komplexität der Aussagen deren Anschlussfähigkeit erhöhen. Darüber hinaus würde es bei breiter Beteiligung ein frei zugängliches und fein strukturiertes Diskurs-Archiv hervorbringen.

## Inhaltsverzeichnis

<b>Vorwort</b>	<b>5</b>
<b>1 Einleitung</b>	<b>6</b>
<b>2 Funktion und Form einer wissenschaftlichen Publikation</b>	<b>12</b>
2.1 Die wissenschaftliche Publikation . . . . .	12
2.2 Die Form des wissenschaftlichen Dokuments . . . . .	19
<b>3 Bedarf für ein neues Medium wissenschaftlicher Kommunikation</b>	<b>25</b>
3.1 Die Notwendigkeit neuer selektionsunterstützender Verfahren . . . . .	25
3.2 Von der Wissens- zur Wissenschaftsrepräsentation: Eine Skizze . . . . .	28
3.3 Wissenschaftskommunikation und Semantik . . . . .	31
3.4 Mit Ontologien das ontologische Weltbild überwinden? . . . . .	36
<b>4 Die Evolution eines möglichen Wissenschaftsnetzwerks</b>	<b>45</b>
4.1 <i>Memex, Xanadu</i> und das <i>WorldWideWeb</i> . . . . .	45
4.2 Wissensdatenbanken . . . . .	48
4.3 Semantisches Mark-up . . . . .	50
4.4 Nanopublikationen . . . . .	56
<b>5 Fazit: Ad-hoc-Wissenschaftsnetzwerk statt Ex-ante-Ontologien</b>	<b>61</b>
<b>Literatur</b>	<b>72</b>

## Abbildungsverzeichnis

1 Modellierung für das Mark-up der Argumentationsstruktur von wissenschaftlichen Texten aus Ribaupierre und Falquet 2013 . . . . .	55
--	----

## Tabellenverzeichnis

1 Neue Paradigmen zum Aufbau eines Wissenschaftsnetzwerks . . . . .	63
---	----

## Vorwort

Bei der vorliegenden Arbeit handelt es sich um eine überarbeitete Version meiner Masterarbeit, die ich im Mai 2014 unter dem Titel »Wissenschaftliches Publizieren im Semantic Web – auch in den Kulturwissenschaften« am Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin einreichte. Die Änderung des Titels ist dem Umstand geschuldet, dass ich mich ursprünglich darauf beschränken wollte, ein Ideen für ein semantisches Publikationsmedium zu entwickeln, dass den Anforderungen der Kulturwissenschaften genügt, weshalb ich ein entsprechend formuliertes Thema beim Prüfungsausschuss einreichte. Erst gegen Ende der Bearbeitungszeit stellte sich heraus, dass diese disziplinäre Einschränkung weder nötig noch sinnvoll ist.

Ich möchte meinen Betreuern Peter Schirnbacher und Martin Gasteiner für alle Anregungen danken. Ihre Gutachten bewerteten die ursprüngliche Arbeit durchschnittlich mit der Note 1,3 (sehr gut). Die Konstellation, einerseits von einem Informationswissenschaftler und -praktiker und andererseits von einem informatikaffinen Kulturhistoriker betreut zu werden, war mein Wunsch, um bereits in einem frühen Stadium meiner Überlegungen sowohl aus der Gruppe der potentiellen DienstleisterInnen als auch aus der anvisierten Zielgruppe Hinweise zu erhalten. Dieser Wunsch findet sich auch bei der Zusammensetzung meiner Lektoren wieder, die sich aus Informationspraxis und Literaturwissenschaft (Wolfram Seidler), Alter Geschichte (Stefan Paul Trzeciok), Informatik (Gerhard Gonter) und Soziologie (Kaspar Molzberger) rekrutierten. Allen bin ich unendlich dankbar für ihre Anmerkungen und Korrekturen!

## 1 Einleitung

Die Dynamik des Internets kann für den Erfolg wissenschaftlicher Kommunikation nur dann voll ausgenutzt werden, wenn den Forschenden Spezialmedien zur Verfügung stehen, welche auch die Dynamik der Kommunikation selbst unterstützen. Bedeutungen stehen im Rahmen von Kommunikation niemals fest. Die Komplexität von wissenschaftlicher Kommunikation sollte nicht auf Kosten der Bedeutungspluralität reduziert werden, sondern durch die Destillation von Bedeutungen und Aussagen, ihre dynamische Bündelung und Redundanzvermeidung. Das Semantic Web sollte bis auf Weiteres als experimentelles Kommunikationsmedium gesehen werden denn als Wahrheiten bereithaltender Wissensspeicher, der exakte Antworten auf alle Fragen liefern kann. Denn: Die gibt es nicht, auch wenn Ontologien das behaupten.

Offenbar wird von Teilen der ExpertInnen<sup>1</sup> der Wissensorganisation die Beschäftigung mit Ontologien für das Semantic Web abgelehnt: Spricht man von Ontologie, solle man sie als philosophische Grundlage verstehen, »not to be confused with homonym schemes for machine treatment of semantic information« (Gnoli, McIlwaine et al. 2008)<sup>2</sup>, womit gleichzeitig offenbart ist, dass hier 2008 noch eine eher vage Vorstellung davon herrschte, was das Semantic Web kann und soll. Auch die Publikationspraxis gibt kaum Hinweise darauf, dass man die Chance, Wissenserzeugung erheblich zu erleichtern und zu beschleunigen, indem computergestützt über Publikationen transportierte Aussagen automatisch verknüpft werden, bereits erkannt hat (siehe z. B. Neylon 2012 sowie Bourne, Fink et al. 2008).

Der Stand der wissenschaftlichen Reflexion zum semantischen Publizieren beschränkt sich nahezu vollständig auf Vorschläge konkreter Anwendungen. Die vorliegende Studie untersucht mittels der Theorie sozialer Systeme Niklas Luhmanns die Grundlagen für die Adaption semantischer Publikationstechnologien in der Wissenschaft. Diese Theorieentscheidung wurde getroffen, da die Systemtheorie nicht nur ausgiebige Analysen des Wissenschaftssystems und einen ausgefeilten Kommunikations- und damit Gesellschaftsbegriff bereitstellt, sondern auch durch ihre Integration einer Form- und damit Medientheorie. Bei der Untersuchung der Grundlagen von informationswissenschaftlicher Wissensrepräsentation ermöglicht mir dieses Instrumentarium faszinierende Einsichten, die insbesondere in Kapitel 3 nachvollziehbar werden sollen.

<sup>1</sup> Im Plural verwende ich das Binnen-I, um zu markieren, dass alle Geschlechter gemeint sind, im Singular wechsele ich unsystematisch zwischen weiblicher und männlicher Form, insbesondere weil der/die LeserIn nicht *doppelt* stolpern soll: über Artikelverdopplungen *und* Binnen-I.

<sup>2</sup> In den Geistes- und Sozialwissenschaften ist es nach wie vor verbreitet, Zitate mit einer Seitenzahl zu referenzieren. Ich breche mit dieser Tradition bei Quellen, die primär digital und volltextdurchsuchbar verfügbar sind. Immer mehr *publisher* verzichten zudem generell auf eine Paginierung, z. B. *PLOS* und *Frontiers*.

Alle mir bekannten, konkreteren Vorschläge zum semantischen Publizieren – bis auf einen, siehe Kapitel 4 – entstammen den Lebenswissenschaften.<sup>3</sup> Obwohl das Semantic Web aufgrund der zahlreichen Editionsprojekte der *Digital Humanities* auch in den Kulturwissenschaften bekannt sein sollte, wurde es als Anwendung für zeitgenössische Publikationen meines Wissens nach bislang kaum in Betracht gezogen. Die Kulturwissenschaften sind für meine Fragestellung besonders interessant, weil der Fall gleichzeitig fern- und naheliegend erscheint. Einerseits hat sich das Semantic Web bislang nur in der Verarbeitung klar strukturierter Daten durchgesetzt: Taxonomien in der Biologie, Genforschung, Geographie, aber auch im kulturellen Bereich: z. B. für Informationen über Musik und Filme – Fakten: ja, Analysen: nein. Andererseits sind die Kulturwissenschaften für die Verwendung eines Mediums, das Bedeutungspluralitäten zulässt und Diskurse abbilden kann, prädestiniert. Dass das Semantic Web dazu bislang kaum eingesetzt wird, ist kein Zufall, denn das ontologische Denken, auf dessen Grundlage man begonnen hat, das Semantic Web aufzubauen, lässt auch nichts anderes zu als Faktensammlungen. Wie gelangen abstrakte Begriffe mit dem Ringen um ihre Bestimmung und die Verknüpfung dieser Begriffe zu Aussagen, also Theorie, in bestreitbarer Form, als wissenschaftliche Publikation, ins Semantic Web? Dabei wird sich jedoch schnell zeigen, dass diese Anforderung nicht disziplinspezifisch ist.

Weite Teile der Geistes- und Sozialwissenschaften haben aufgrund ihrer Methoden und der ihnen innewohnenden Bedeutung von Theorien grundsätzlich Bedarf an Werkzeugen, die das schnelle Erfassen von Begriffsbestimmungen unterschiedlicher Provenienz und ihrer theoretischen Kontexte erleichtern. Die Verwendung der Bezeichnung »Kulturwissenschaften« soll im Folgenden genau diese Bereiche der Geistes- und Sozialwissenschaften umfassen, deren Methoden nicht primär empirisch ausgerichtet sind, da dort erweiterte, hier nicht zu behandelnde Werkzeuge den spezifischen Bedürfnissen gerecht werden mögen.

Die Kulturwissenschaften sind in ihren Themen und Methoden unbestimmt und bestimmen sich genau dadurch: Alles kann zu ihrem Thema werden; einen gesetzten Methodenkanon gibt es nicht. Es herrscht, mit Paul Feyerabend gesprochen, wissenschaftstheoretischer Anarchismus. Was die kulturwissenschaftliche Forschung als solche erkennbar macht, ist ihre besondere Form der zirkulären Selbstreferenz: Die Güte eines kulturwissenschaftlichen Forschungsergebnisses ist nur anhand der Ausschöpfung des gewählten methodischen und theoretischen Programms für die Analyse eines gesellschaftlichen Problems zu beurteilen. Durch diese Ausschöpfung führt solche Forschung dann zwangsläufig zu einer Anpassung von Theorie und Methode, die gleichsam als Forschungsergebnis zu betrachten ist. »Man kann also durch die Verwendung einer

<sup>3</sup> Ein vollständiger Überblick über den Entwicklungsstand des semantischen Publizierens in den Naturwissenschaften kann hier weder gegeben werden, noch wurde er bislang andernorts zusammengetragen.

bestimmten Methode, eines bestimmten Vorgehens, eines bestimmten Vokabulars Ergebnisse nicht auf ein und derselben Bezugsskala besser machen – man kann sie nur anders machen« (Daniel 2006, S. 15). Das führt zu enormen Freiheitsgraden insbesondere in der Strukturierung von wissenschaftlichen Dokumenten und der Gestaltung von Publikationen, was wiederum ihre Aufbereitung für die Verarbeitung durch Maschinen erschwert, da sich weder im Aufbau der Texte noch der Argumentation klare, erwartbare Muster erkennen lassen. Solche Muster dienen in den Naturwissenschaften als Einstieg ins Semantic Web für Publikationen.

Die Studie setzt voraus, dass die Funktionsweise von semantischen Webtechnologien, wie sie gern als Stufenmodell<sup>4</sup> dargestellt werden, in ihren wesentlichen Aspekten bekannt ist. Die Vorteile dieser Technologien sollen mit den Anforderungen der wissenschaftlichen Kommunikation an ein Kommunikationsmedium in ein vielversprechendes Verhältnis gesetzt und eine Vorstellung von einem geeigneten System entwickelt werden. Dazu müssen im Rahmen dieser Studie kommunikationstheoretische Vorschläge genügen. Diese Beobachtungen können dann für empirische Forschungen operationalisiert werden, um die vorausgesetzten Anforderungen zu verifizieren. Darüber hinaus können sie als Grundlage für technische Entwicklungen dienen.

Der gesellschaftliche Diskurs ermöglicht mittlerweile, für transitive Überlegungen zur Wissenschaftskommunikation eine Open-Access-Prämisse zu setzen,<sup>5</sup> nicht zuletzt aufgrund von politischen Entscheidungen vom Zweitveröffentlichungsrecht in Deutschland über das 8. EU-Forschungsförderungsrahmenprogramm *Horizon 2020* hin zu ähnlichen Entwicklungen in vielen Ländern der Welt. Darüber hinaus arbeiten bedeutende Organisationen wie die *Open Knowledge Foundation* an der Etablierung der Open-Science-Idee, mit der nicht nur die freie Zugänglichkeit zu Forschungsergebnissen verknüpft ist, sondern unter anderem auch zu den Mitteln und Werkzeugen der Forschung. Aus der Perspektive dieser Studie ergibt sich dann die Anforderung, dass auch die Software, mit der Publikationen hergestellt werden, Open Source sein muss.<sup>6</sup> Außerdem ist für Forschungsergebnisse eine permanente und persistente Verfügbarkeit geboten.<sup>7</sup> Elektronische Publikationen, die hier als Normalfall vorausgesetzt werden (siehe auch Kaden 2013), haben gegenüber gedruckten den Nachteil, leichter

<sup>4</sup> Siehe W3C, Semantic Web Levels, <http://www.w3.org/2001/12/semweb-fin/swlevels.png>, zuletzt besucht am 23. April 2014. Als Einführung eignet sich Shadbolt, Hall et al. 2006, praxisorientierter ist Domingue 2011.

<sup>5</sup> Zur Verbreitung von Open Access siehe nur Archambault 2013 und die Hochrechnung von Lewis 2012, der Open-Access-Journals als disruptive Technologie beschreibt.

<sup>6</sup> Für eine Diskussion der Vor- und Nachteile sowie Adaptionsbarrieren von Open-Source-Software siehe Gwebu und Wang 2011.

<sup>7</sup> Für Einzelanforderungen an Formate siehe Funk 2010 und die dort angegebene Literatur.

imitiert und verfälscht werden zu können, wenn ihre Integrität nicht durch besondere Maßnahmen geschützt wird, die jedoch hier nicht diskutiert werden.

Davon einmal abgesehen, scheinen gewisse Akzeptanzprobleme des elektronischen Publizierens in den Geisteswissenschaften dadurch zu entstehen, dass es mittlerweile recht leicht ist, ohne die Inanspruchnahme von Verlagsservices eine Publikation herzustellen und zu präsentieren, die keinen geringeren professionellen Anschein macht als klassische Verlagsprodukte. Wie auch Kaden (2013) anmerkt, wird dadurch die *inhaltliche* Professionalität von elektronischen Publikationen mitunter in Frage gestellt. Alle Vorschläge für elektronische Publikationsformate werden bis auf Weiteres diese Skepsis hervorrufen. Dies scheint ein typisches Übergangsphänomen zu sein, derer immer welche auftreten, wenn der Mediengebrauch sich ändert. Eine Chance wäre jedoch, ein Format zu kreieren, das nicht mehr plausibel als Imitation des analogen Formats beschrieben werden kann. Daran ist auch eine Entwicklung von neuen Qualitätssicherungs- und damit Reputationserzeugungsmechanismen geknüpft, deren Untersuchung jedoch anderen Studien vorbehalten bleiben soll.<sup>8</sup> Das Publikationswesen ist fraglos in einem Wandel begriffen,<sup>9</sup> aber eine Änderung der sozialen Funktion und inneren Formen der Publikation scheint derzeit nicht in Sicht zu sein.

Das Adaptionsproblem bei der Einführung eines neuartigen Publikationsmediums hat mehrere Dimensionen:

1. In der *sachlichen Dimension* scheint die ungebrochene Nachfrage nach gedruckten Büchern in den Kulturwissenschaften und nach PDFs in allen Fachbereichen die auf diesem Wege publizierenden Forschenden zu bestätigen (vgl. Börsenverein des Deutschen Buchhandels 2013). Nach dem Semantic Web hingegen fragen insbesondere in den Sozial- und Geisteswissenschaften, und hier fast ausschließlich in den *Digital Humanities*, nur sehr wenige Forschende. Diese Studie begründet, inwiefern ein erneuter Medienbruch die Wissenschaft von überlieferten, an die neuen digitalen Bedingungen nicht angepassten Selektionsmechanismen, von Doppelforschung und mühsamen Recherchen befreien kann. Gleichzeitig wird nach einer Vorstellung von diesem neuen Medium gesucht, die Forschenden leicht vermittelt werden kann, ohne dabei bereits ins technische Detail zu gehen.
2. In der *zeitlichen Dimension* muss man einen enormen Aufwand betreiben, um Aussagen so zu formulieren, dass sie maschinenlesbar werden. Auch die Urform der semantischen Auszeichnung, die Registererstellung, konnte sich trotz der großen Beliebtheit bei den Rezipierenden niemals zu einem allgemeinen Standard entwickeln. Jedoch: Wenn man annimmt, dass KulturwissenschaftlerInnen ihren

<sup>8</sup> Siehe z. B. Taubert 2010; Birukou, Wakeling et al. 2011 sowie Brembs, Button et al. 2013.

<sup>9</sup> Siehe dazu auch Schmidt 2014 und die dort angegebene Literatur.

mitunter auch als künstlerische Leistung verstandenen Text weiterhin als Ausgangspunkt begreifen, so scheint es wahrscheinlich, dass die Auseinandersetzung mit dem eigenen Text, die ein Destillat der enthaltenen Aussagen erzeugen soll, die Klarheit der Argumentation auch im Ausgangstext fördert.

3. Die *soziale Dimension* des Problems lässt sich nicht so leicht ausräumen: Da im Wissenschaftssystem derzeit keinerlei formale Anerkennung für den angedeuteten Aufwand vorgesehen ist, bestehen kaum Anreize, ihn zu leisten. Zwischen den etablierten Forschenden herrscht ein Nash-Gleichgewicht: Semantisch zu publizieren ist nur dann attraktiv, wenn auch die anderen dies tun. Ohne eine kritische Masse an verknüpfbaren Aussagen ist die Nutzung des zugrundeliegenden Systems für Recherchezwecke unbrauchbar und die eigene Investition wäre verschenkt. Niemand wird als erster die eigene Publikationspraxis ändern, da die dafür benötigte Zeit der Konkurrenz Vorteile verschafft. Dem Nash-Gleichgewicht ist nur durch Überzeugung zu entkommen.

Schließlich können sich die Ergebnisse dieser Studie nicht in erster Linie an die Publizierenden selbst richten, sondern an jene Einrichtungen, die mit der Forschung gleichzeitig das Wissenschaftssystem fördern, zu dem das Publizieren fraglos gehört. Die AutorInnen sollten »von den konkreten technischen Problemen entlastet, aber zur Mitarbeit bei der Entwicklung fachgerechter Arbeitsumgebungen bereit sein« (Mittler 2011), um eine sinnvolle Arbeitsteilung zu erreichen. Die *publisher*<sup>10</sup> sind nicht diejenigen, die das Wissenschaftssystem fördern, aber sie können als Serviceeinrichtung im Auftrag jener Fördernden verstanden werden. Immer häufiger befinden sich vor allem jene *publisher* in der Kritik, die riesige Marktanteile innehaben.<sup>11</sup> Sowohl Forschende, als auch Wissenschaftsmanagement und Bibliotheken sehen sich mit einem schlechten, aber kaum verhandelbaren Preis-Leistungs-Verhältnis konfrontiert. Die Innovationsfreudigkeit der Verlage ist nicht immer mit dem technischen Entwicklungsstand in Einklang zu bringen. Daher sollten die Serviceeinrichtungen von Forschungsinstitutionen den Verlagen das Feld auch nicht gänzlich überlassen und mindestens den Entwicklungsstand kennen, um auf die – vermutlich noch seltenen – Anfragen von AutorInnen reagieren zu können. Letztendlich entscheiden die Forschenden, wessen Service sie

<sup>10</sup> Aufgrund der besonderen historischen Entwicklung des Verlagswesens im angelsächsischen Raum kann man diese Bezeichnung nicht eindeutig mit »Verlag« ins Deutsche übertragen. Sie umfasst nicht nur Verlagshäuser, sondern auch Herausgeber, die direkt mit einer Druckerei zusammenarbeiten und Verlagstätigkeiten meist semi-professionell selbst leisten, z. B. einige Fachgesellschaften.

<sup>11</sup> Exemplarisch sei nur die Aktion »The Cost of Knowledge« genannt, auch bekannt als »Elsevier-Boycott«, der Anfang 2012 vom Mathematiker William Timothy Gowers gestartet wurde. Mittlerweile wurde das Positionspapier von 14 600 ForscherInnen unterzeichnet, die sich dazu bekennen, ihre Arbeit Elsevier nicht mehr kostenfrei zur Verfügung zu stellen. Die Aktion zog eine große Medienaufmerksamkeit nach sich. Siehe <http://thecostofknowledge.com>, zuletzt besucht am 12. April 2014.

bevorzugen und wer sich dadurch etablieren und weiterentwickeln kann. Insbesondere für öffentlich geförderte Einrichtungen sollten sich jedoch genügend Anreize finden lassen, im Verbund Alternativen zu kommerziellen Lösungen anzubieten, darunter vor allem: langfristig beherrschbare Kosten und die Schwächung der Marktposition der Kommerziellen. Noch werden immer wieder hochinnovative Kleinstprojekte gefördert, die von vornherein eher als Experiment denn als tragfähige Breitentechnologie konzipiert werden. Die vorliegende Studie soll einen Beitrag zur Konzeption von universalen wissenschaftlichen Publikationsinfrastrukturen leisten.

Im Mittelpunkt des 2. Kapitels dieser Studie stehen Begriffsbestimmungen, das Eröffnen des Möglichkeitshorizonts. Dazu gehören system- und formtheoretisch inspirierte Definitionen der wissenschaftlichen Publikation und des Dokuments sowie eine Erläuterung ihrer Bedeutung für die wissenschaftliche Kommunikation. Im 3. Kapitel soll einerseits begründet werden, warum nach einem neuen wissenschaftlichen Publikationsmedium gesucht werden muss und inwiefern sich das Semantic Web dafür anbietet. Andererseits werfen die gegenwärtig in den Informationswissenschaften dominierenden Konzeptionen von Wissensrepräsentation, Semantik und Ontologien Schwierigkeiten auf: Sie sind schlecht vereinbar mit der Funktionsweise wissenschaftlichen Kommunizierens, wie sie im vorhergegangenen Kapitel erläutert wurde. Gegen Ende dieses Kapitels werden die Anforderungen an das neue Medium formuliert. Das 4. Kapitel sucht nach Spuren für bereits erfolgte Umsetzungen dieser Anforderungen und bietet einen gestrafften Überblick über die aktuellen Beiträge zum Thema semantisches Publizieren, um Anschlussmöglichkeiten aufzuzeigen. Schließlich wird im 5. Kapitel aus dem Raum des Möglichen heraus eine Vorstellung von einem neuen Medium entwickelt und mit dem Titel »semantisches Wissenschaftsnetzwerk« benannt. Seine Umsetzung erfordert jedoch mehrere Paradigmenwechsel, wie sie im 3. Kapitel bereits angedeutet wurden. Die Anstrengung verspricht nicht nur ein besseres Erfüllen der gegenwärtigen Anforderungen an das Publikationswesen, sondern auch das automatische Erstellen von neuen Hypothesen und das leichte Aufspüren von Inkonsistenzen in der wissenschaftlichen Kommunikation.

## 2 Funktion und Form einer wissenschaftlichen Publikation

Eine Bezeichnung sollte nicht nach der jeweils passend erscheinenden Bedeutung als Begriff definiert werden, sondern umgekehrt: Die Definition wird abgeleitet aus dem alltäglichen Sprachgebrauch des Wortes einerseits, und seinen professionell-wissenschaftlichen Verwendungen andererseits, also durch die Analyse von unterschiedlichen Bedeutungen derselben Bezeichnung und bestenfalls auch von damit ähnlichen Bedeutungen anderer Bezeichnungen. Durch diese Rücksichtnahme auf Semantiken kann Missverständnissen vorgebeugt werden.

### 2.1 Die wissenschaftliche Publikation

Als Grundlage der Bestimmung eines Publikationsbegriffs soll die *Berner Übereinkunft zum Schutz von Werken der Literatur und Kunst*<sup>12</sup> von 1979 dienen, da sie als Ergebnis eines internationalen Diskussionsprozesses den Charakter einer rechtlich-politisch bindenden Entscheidung hat. Man rang um gemeinsame Begriffe nicht nur im Hinblick auf die Harmonisierung mit den unterschiedlichen nationalen Urheberrechtsbestimmungen: Die Vereinbarung musste sich in der Praxis bewähren und daher den damaligen Schutzbedürfnissen der UrheberInnen genügen. Man stellte also einen Konsens her, der das Rechtssystem mit den nötigen Voraussetzungen ausstattete, die sowohl für das Wissenschafts-, als auch für das Kunstsystem notwendigen Leistungen erbringen zu können.

In der Berner Übereinkunft sind vor allem drei Merkmale einer Publikation hervorgehoben, die eine Veröffentlichung ausmachen: Die Verbreitung des Werks unter *Einwilligung der AutorInnen* in einem Maße, das den *Bedürfnissen der Öffentlichkeit* angemessen erscheint. Durch diesen Publikationsbegriff ausgeschlossen wird die Präsentation eines Originalwerks, das gegenüber seiner *Reproduktion* ein wesentlich höheres gesellschaftliches Ansehen transportiert, so z. B. bei performativen Werken oder in der Architektur.

Wenn Riehm et al. (2004) einen Kommunikationsprozess als Publikation definieren, der als indirekt, vermittelt und asynchron charakterisiert wird und »für die Öffentlichkeit, für ein mehr oder weniger anonymes Publikum bestimmt« ist, fehlen nahezu alle drei Aspekte. Eine solche Definition gerät, selbst wenn man sie nur auf wissenschaftliches Publizieren beziehen würde – eine Einschränkung, die sie selbst nicht explizit beinhaltet – in einen erheblichen Konflikt mit der internationalen Urheberrechtsgesetzgebung: Ein Werkbegriff fehlt völlig und der Grad der Verbreitung erscheint zu unbestimmt.

<sup>12</sup> World Intellectual Property Organization, Berne Convention for the Protection of Literary and Artistic Works, Article 3, [http://www.wipo.int/treaties/en/text.jsp?file\\_id=283698#P98\\_14701](http://www.wipo.int/treaties/en/text.jsp?file_id=283698#P98_14701), zuletzt besucht am 8. April 2014.

Auch erscheint es fragwürdig, die Publikation als Kommunikationsprozess zu bezeichnen, da der allgemeine Sprachgebrauch den Begriff zwar einerseits als Objekt-, andererseits als Prozessbezeichnung kennt, dann aber den Vorgang meint, ein Werk aus seiner singulären Existenz heraus in den Bereich der Wahrnehmbarkeit durch Viele zu befördern. Hierin findet sich auch der Aspekt der Reproduktion aus der Berner Übereinkunft wieder, auch wenn es sich beim elektronischen Publizieren im technischen Sinne um keine Reproduktion handelt (siehe Kapitel 2.2), wohl aber in einem sozialen Sinne: Die wiederholte und doch je individuelle Wahrnehmung gekoppelt an die Reproduktion von wissenschaftstypischem Verhalten. Beobachtbar ist die Publikation selbst dabei jedoch zunächst einmal nur als Ereignis. Inhaltliche Aspekte, hier: Forschungsergebnisse prägen sich dem Publikationsmedium ein, erfordern aber die Rezeption, um soziale Prozesse in Gang zu bringen.

Riehm et al. optieren allerdings auch gar nicht auf eine Kommunikation über Inhalte, sondern setzen die Annahme oder die Ablehnung eines Manuskripts an den Anfang des Publikationsprozesses, der wohl im zweiten Fall gleichzeitig sein Ende wäre, das andernfalls vor Beginn der Rezeption läge. Was aber, wenn diese ausbleibt? Kann man schon von Kommunikation sprechen, wenn lediglich ihre Voraussetzungen in Anspruch genommen werden, die realisierte soziale Komponente jedoch gegen Null tendiert?<sup>13</sup> Der Fall der Nicht-Wahrnehmung der Publikation wird offenbar durch die zwingende Verwendung »anerkannter Kanäle« ausgeschlossen. Damit würden Preprints nicht als Publikation gelten und selbstverständlich auch keine Werke, die ohne die Beihilfe eines *publishers* im Internet zugänglich gemacht wurden. Einzuwenden wäre, dass die Funktion jeder Form von Begutachtung fraglich erscheint, solange sie ohne die Bestimmung inhaltlicher Kriterien auskommen soll und allein formale Bedingung bleibt. Folgt man der Definition von Riehm et al., muss man daraus schließen, dass Graue Literatur damit nicht als Publikation gilt. Das Urheberrecht gilt jedoch. Und: Was wäre sie dann? Riehm et al. meinen: verbreitete Dokumente.

Die Deutsche Nationalbibliographie enthält in »Reihe B - Monografien und Periodika außerhalb des Verlagsbuchhandels. Bücher, [...] und elektronische Publikationen«.<sup>14</sup> Nach einem klassischen Lehrbuch der Bibliothekswissenschaft beinhaltet »publizierte Information [...] Dokumente [...] in analoger oder digitaler Form, die von Verlagen, politischen, gesellschaftlichen oder privaten Vereinigungen, Organisationen bzw. Institutionen hergestellt, vervielfältigt und für die Öffentlichkeit bzw. eine Teilöffentlichkeit

<sup>13</sup> Für die Kunst ist dieser Gedanke charmant: »Der Harlekin mag im Dunkeln tanzen – aber selbst das wäre noch Kommunikation, die ihre eigene Vollendung sabotiert, um sich zu bestätigen, daß sie sich nur sich selber verdankt und nicht den Blicken eines Beobachters« (Luhmann 1995b, S. 131). Aber gilt das auch für die Wissenschaft?

<sup>14</sup> DNB, Deutsche Nationalbibliographie, [http://www.dnb.de/DE/Service/DigitaleDienste/DNBBibliografie/dnbbibliografie\\_node.html](http://www.dnb.de/DE/Service/DigitaleDienste/DNBBibliografie/dnbbibliografie_node.html), zuletzt besucht am 12. April 2014.

bestimmt, herausgegeben werden« (Umstätter 2011, 10f.). Darüber hinaus wandelt sich allmählich auch die Bewertung von Grauer Literatur: Studien zeigen, dass ihre Qualität in den meisten Fällen gesichert ist und eine Nachfrage besteht. Es gibt daher nicht länger Gründe dafür, dass sie von Bibliotheken weniger intensiv gesammelt wird als durch *publisher* veröffentlichte Literatur (siehe Gelfand und Lin 2013).

Die Asynchronizität des Schreibens und Lesens und auch die Vermitteltheit sind keine Besonderheiten des Kommunikationsmediums Publikation, da diese Merkmale für jede Kommunikation zutreffen, die auf den Gebrauch des Mediums Schrift zurückgreift oder die Form audiovisueller Sendung annimmt. Es handelt sich um eine »Notwendigkeit der Massenkommunikation« (Luhmann 1997, S.308). Die Interaktion unter Anwesenden<sup>15</sup> ist geprägt durch die Möglichkeit, sich wechselseitig am Verhalten des anderen zu orientieren, Erwartungen zu unterstellen, reflexiv wahrzunehmen, also darüber zu spekulieren, wie das eigene Verhalten vom anderen wahrgenommen wird. Handlungen müssen, um soziale Relevanz zu erhalten, von einem Beobachter als solche zugerechnet werden. Luhmann nennt diesen Beobachter »Ego« und den Handelnden »Alter«, weil erst dadurch, dass Ego sich angesprochen fühlt, das soziale System in Gang kommt. Unter normalen Bedingungen fällt es schwer, Mitteilungen innerhalb eines bestehenden Interaktionssystems zu ignorieren, ohne das System zu (zer)stören. Ego unterscheidet dann eine Information von der Mitteilung und versteht damit etwas, wohlgermerkt völlig unabhängig davon, was der Andere, Alter, gemeint haben mag – das bleibt gezwungenermaßen intransparent. Dieses Selektionsgeschehen verändert sich durch den Gebrauch von Medien, die reflexive Wahrnehmung unmöglich machen. Am Telefon beispielsweise ist sie durchaus noch möglich, begrenzter auch im Chat oder im persönlichen Brief- oder E-Mailwechsel, die sich durch ihren Mediengebrauch als Sonderformen der Interaktion durch eine anwesende, also in der Interaktion ständig präsente, Abwesenheit auszeichnen.

Sobald eine Wissenschaftlerin einen Aufsatz für ein nicht näher spezifiziertes Publikum zur Lektüre verbreitet, variiert das beschriebene kommunikative Selektionsgeschehen<sup>16</sup> folgendermaßen: Das Publikum muss sich erstens angesprochen fühlen, zweitens eine Information aus der Form der Mitteilung selektieren und drittens Möglichkeiten eines kommunikativen Anschlusses erkennen, also: verstehen. Die Form der Mitteilung ergibt sich durch das spezifische Kommunikationsmedium – darauf wird zurückzukommen sein. Anschlüsse in Form von Zitationen werden hier nicht die ersten kommunikativen Anschlüsse sein, sondern eher Aktivitäten im Rahmen des Zeitschriftenmarketings oder die Verbreitung einer Erscheinungsmeldung über soziale Netzwerke. Die Differenz

<sup>15</sup> Zu diesem Begriff Luhmanns siehe insbesondere einen Aufsatz von 1975a.

<sup>16</sup> Zu Luhmanns synthetischem Kommunikationsbegriff und den drei Selektionen siehe am besten Luhmann 1984, Kapitel 4, Abschnitt II.

zum Selektionsgeschehen in Interaktionen, nämlich die Unmöglichkeit der reflexiven Wahrnehmung, hat massive Konsequenzen: Ignorieren bleibt für das Publikum in den meisten Fällen folgenlos, aber negative Anschlüsse auf derselben Ebene von Kommunikation<sup>17</sup>, nämlich jener der theoretisch weltweiten Verbreitung, können hingegen enorm folgenreich sein. Ein weiterer bedeutsamer Unterschied sind die Chancen auf Korrektur des einmal Kommunizierten. Während es in Interaktionen zum Normalen gehört, sich selbst zu widersprechen, ist dies auf der Ebene der gesellschaftsweiten Kommunikation schwer möglich, insbesondere, solange diese Kommunikation auf den Druck des Artikels angewiesen war. Und hier kehren wir zurück zur Asynchronizität: Diese war solange höchst bedeutsam, wie enorm zeit- und kostenintensive Verfahren in Anspruch genommen werden mussten, um gesellschaftsweit kommunizieren zu können. Noch dazu war es extrem unwahrscheinlich, mit der Mitteilung einer Korrektur Aufmerksamkeit zu erlangen. Das gilt in viel stärkerem Maße für wissenschaftliche wie für massenmediale Verbreitungsmedien, da sie sich durch eine längere Frequenz auszeichnen und Monate später wahrscheinlich niemand die Korrektur eines kaum mehr in der Erinnerung präsenten Artikels wahrnehmen wird. Mit dem Internet gibt es jetzt allerdings die Chance der zeitnahen Korrektur, die eng mit der Erstveröffentlichung verknüpfbar ist. Wenn die verwendete Publikationsplattform Annotationstools zur Verfügung stellt, mit denen während der Rezeption öffentlich kommentiert werden kann, nähert sich die Asynchronizität bis zur Bedeutungslosigkeit an die Synchronizität an, wie sie in Interaktionssystemen herrscht.

An dieser Stelle soll eine allerdings weiter zu erläuternde Definition formuliert werden, die sich dadurch auszeichnen hat, auf alle Publikationen in der modernen Wissenschaft<sup>18</sup> vom 17. Jahrhundert bis heute und möglichst auch zukünftig anwendbar zu sein, also unabhängig von den verwendeten Verbreitungsmedien und den jeweils anerkannten Publikationsformaten, sowohl inhaltlich-strukturell als auch technisch. Sie soll wie folgt lauten: *Eine wissenschaftliche Publikation ist die durch ihre AutorInnen veranlasste und gezeichnete Verbreitung eines auf originale und frühere Forschungsergebnisse anderer*

<sup>17</sup> Luhmann unterscheidet die drei Ebenen der Kommunikation: Interaktion, die sich über Anwesenheit begrenzt, Organisation, deren Kommunikation nur unter Mitgliedern abläuft, und Gesellschaft, die sich schließlich durch weltweite Erreichbarkeit auszeichnet (Luhmann 1975b).

<sup>18</sup> Gemeinhin wird dieser Zeitpunkt mit der ersten Ausgabe der ersten wissenschaftlichen Zeitschrift, also mit den *Philosophical Transactions* von 1665 markiert. Der Grund dafür lässt sich schon am Titelblatt ablesen: »Giving some ACCOMPT of the present Untertakings, Studies, and Labours of the INGENIOUS in many considerable parts of the WORLD« (Hervorhebung nach Original). Die Intention der Reproduktion und Verfestigung eines dem Prinzip nach weltweiten Kommunikationszusammenhangs ist deutlich kommuniziert und damit auch eine flexible Definition eines Publikums geschaffen. Die Epochenbezeichnung »Moderne« wird für diese Studie weniger dem kontinentaleuropäischen als dem angelsächsischen Diskurs entlehnt, wo man mit »modern history« den Zeitraum zwischen Renaissance und Gegenwart benennt. Auf die Adaption des Begriffs der Postmoderne wird hier verzichtet, da er eine Unterscheidung einführt, die hier zu keinem Erkenntnisgewinn beiträgt.

referierenden wissenschaftlichen Dokuments, um es einem weltweiten Publikum zugänglich zu machen.<sup>19</sup> Der Bedeutung von Reputation im gegenwärtigen Wissenschaftssystem würden anonyme Werke zuwider laufen, weshalb die Zeichnung durch AutorInnen zwingend zu einer wissenschaftlichen Publikation gehört. Außerdem dient ein Personenname in der Wissenschaft als primärer Referenzpunkt. Nun ist diese Definition weiter zu erläutern: Was zeichnet ein originales Forschungsergebnis aus? Warum muss die Definition den Verweis auf frühere Forschungsergebnisse enthalten? Was ist ein wissenschaftliches Dokument? Dies lässt sich nur ergründen, wenn man die Funktion der Publikation für die moderne Wissenschaft kennt.

Publikationen in der Wissenschaft stehen analog zu Zahlungen in der Wirtschaft: Sie verweisen auf andere Publikationen, also auf Elemente gleichen Typs. Man kann nicht zahlen, ohne zuvor eine Zahlung erhalten zu haben. Auch wird die neuerliche Zahlung ebenfalls wieder Zahlungen auslösen: Der vergrabene Schatz ist vorerst der Wirtschaft entzogen. Mit Publikationen verhält es sich genauso: Der Geniestreich in der Schublade ist eine kommunikative Sackgasse. Ebenso kann es keine wissenschaftliche Publikation geben, die keine Selbstreferenzen auf das Wissenschaftssystem durch Zitationen erzeugt. Solche Selbstreferenzen sind die derzeit für die Wissenschaft bedeutendste und unverzichtbare Methode, Komplexität zu reduzieren, um gezielt an die frühere Kommunikation des Wissenschaftssystems anzuschließen. Damit man zu einer produktiven wissenschaftlichen Diskussion kommt, muss die Kenntnis bestimmter Forschungsergebnisse, insbesondere Publikationen, vorausgesetzt werden. Würde dieser Kanon nicht mit jeder Publikation aufs Neue reproduziert und erweitert werden, bliebe die Diskussion stecken. Publikationen sorgen also für die operationale Schließung des Systems, für seine Einheit. Zitationen erzeugen ein Netzwerk von Interaktionen, die auf anwesende Abwesenheit angewiesen sind. Es entspinnt sich ein Dialog, auf dessen vorhergehenden Vollzug die Publikation selbst beruht, und gleichzeitig spinnt sie selbst am Netzwerk mit, während sie es beschreibt. Publikationen sind also gleichzeitig für die Strukturbildung und die Strukturbeschreibung in der Wissenschaft verantwortlich (Stichweh 1994).<sup>20</sup> Das klingt ein wenig so, als würde sich die Kommunikation nach einem akkuraten Plan ausrichten. Dem ist nicht so: »Manche Texte werden gelesen, einige im rechten Moment. Mit einem hohen Anteil an Zufälligkeit ergeben sich daraus neue Texte, für die dasselbe gilt« (Luhmann 1990, S. 59).

<sup>19</sup> Eine Definition einer nicht näher spezifizierten Publikation lässt sich leicht durch den Austausch von »auf originale und frühere Forschungsergebnisse anderer referierenden wissenschaftlichen Dokuments« durch »originalen Werks« erzeugen.

<sup>20</sup> In Luhmanns Worten klingt das hier Gesagte so: »Erst in der Form von Publikationen erreicht die moderne Wissenschaft autopoietische Anschlußfähigkeit. Publikationen sind gleichsam das Zahlungsmittel der Wissenschaft, das operative Medium ihrer Autopoiesis« (Luhmann 1990, S. 432).

Zitationen gehören nach Gérard Genette (2001) zum »Paratext« eines Textes. Er nannte sein bekanntes Werk, das ausschließlich Paratexte der künstlerischen Literatur zum Gegenstand hat, im französischen Original *Seuils: Schwellen*. Analog dazu lässt sich auch die Vorsilbe »Para-« verstehen, beschreibt sie doch etwas Antithetisches, das Oszillieren zwischen zwei Seiten einer Unterscheidung, die jedoch eine präferierte Seite hat, nämlich den Text. Der Paratext verweist einerseits vom Text nach außen, andererseits von außen auf den Text. Allerdings kann der Text dabei nicht schlicht gedacht werden als Abfolge von Sätzen in den Paragraphen des Hauptteils eines Dokuments: Paratext schleicht sich auch zwischen die Zeilen. Ein Satz kann gleichzeitig Text und Paratext enthalten, und zwar, wenn er als Peritext auftritt, sich also direkt in den Text einschreibt, wie z. B. Zitationen. Es gibt aber auch Peritext, der sich im Umfeld des Textes positioniert, wie Schlagworte. Die Steigerung davon wären Epitexte, die selbst eine geschlossene Textform annehmen und eigene Paratexte haben, wie Rezensionen. Paratext kann als »faktischer Paratext« auch auf ein explizites Mitteilungsformat verzichten und unterschwellig auf der Rezeption des Textes lasten, z. B. wenn der Autor bekanntermaßen einer bestimmten Theorieschule angehört, die jedoch von Beobachtern der Wissenschaftskommunikation konstruiert wird und keine Selbstpositionierung des Autors im Text ist.

Der Paratext von wissenschaftlichen Publikationen ist genau jenes, das nicht zum bekannt zu machenden Forschungsergebnis gehört, die Publikation aber als wissenschaftliche erst erkennbar macht, z. B. der Fußnotenapparat und die Referenzen auf die Publikation in einer anderen Publikation. Genette geht noch weiter: »Der Paratext ist also jenes Beiwerk, durch das ein Text zum Buch wird und als solches vor die Leser, und allgemeiner, vor die Öffentlichkeit tritt« (ebd., S. 10). Eine wichtige Einschränkung gibt es jedoch: Genette schränkt den Begriff ein auf jenes Begleitwerk, das die Intention des Autors, hier nämlich: als wissenschaftliche Publikation erkannt und rezipiert zu werden, stützt. Auch wenn bei Genette vom Autor und seinen Verbündeten die Rede ist, so soll hier nicht gemeint sein, dass die Autorin jeden Paratext autorisieren muss oder dass er keine negative Wertung des Textes enthalten darf. Jeder Paratext, der den Text mit der gesellschaftlichen Kommunikation verknüpft, also als – positiv oder negativ – anschlussfähig erkennbar macht, erhöht grundsätzlich erst einmal die Wahrscheinlichkeit von Kommunikation.<sup>21</sup> Genette selbst lässt diese Folgerung zu, wenn er die Funktion des Paratexts, hier ja auf die Kunstkommunikation zugeschnitten, zusammen-

<sup>21</sup> Die Unwahrscheinlichkeit von Kommunikation nach Luhmann 2005 ergibt sich erstens durch die Unmöglichkeit, Informationen direkt zu übertragen: Was verstanden wird, hängt von Egos Gedächtnis und Sozialisation ab. Zweitens ist jede Erreichbarkeit oder Aufmerksamkeit begrenzt: Hier greifen Verbreitungsmedien. Drittens ist eine erfolglose Kommunikation wahrscheinlicher, weil Personen normalerweise nicht ohne Weiteres ihre Handlungsprämissen ändern und Selektionsvorgaben akzeptieren, um an Kommunikation anzuschließen – es sei denn Erfolgsmedien wie die wissenschaftliche Wahrheit werden verwendet. Die Medientypologie wird im Folgenden klarer werden.

fassend als Schleuse beschreibt, die »zwischen der idealen und relativ unwandelbaren Identität des Textes und der empirischen (soziohistorischen) Realität seines Publikums« eingerichtet wird, um »durch sie ›auf gleicher Höhe‹ bleiben zu können« (Genette 2001, 388f.). Bei einer Publikation in einer wissenschaftlichen Zeitschrift werden Titel und URL der Zeitschrift gleichsam zum Paratext, ebenso bei der Publikation auf einem Dokumentenserver. Für die Rezeption des Textes, für die Wahrscheinlichkeit von kommunikativen Anschlüssen macht es jedoch einen enormen Unterschied, wo der Paratext sich einschreibt.

Zitationen haben die besondere, wenn auch nicht exklusive, Eigenschaft, Paratext zweier Publikationen zu sein und diese dadurch zu verbinden. Je nach Beobachtungstandpunkt ist ein Text dann Text oder Paratext. Eine Zitation reduziert die vorausgegangene Publikation auf mindestens einen Teil ihres Erkenntnisgewinns und erzeugt damit selbst einen Erkenntnisgewinn, also ein Forschungsergebnis. Außer Frage scheint zu stehen, dass darüber hinaus Neuheit eine besondere Rolle in der Wissenschaft spielt, und zwar in zweifacher Hinsicht: Einerseits gibt es insbesondere in den Naturwissenschaften eine Präferenz für neuere, also jüngere Forschungsergebnisse, an die es anzuschließen gilt.<sup>22</sup> Andererseits stellt sich neues Wissen in Differenz zum bereits bekannten. Neuheit ist hier also je nachdem ein Merkmal in einer zeitlichen oder in einer sachlichen Dimension. Für wissenschaftliche Neuheit in der sachlichen Dimension muss nicht zwingend ein neues Paradigma aufgestellt werden (vgl. Kuhn und Krüger 1978), es genügt auch ein Abweichen vom Erwartbaren, das zu neuen Erklärungen führt, die mehr bewirken, als nur den vorhandenen Erklärungen etwas hinzuzufügen: Sie verschieben jene Perspektiven, mit Hilfe derer wiederum nach Wissenserneuerungen gesucht wird (vgl. Bachmann-Medick 2009 sowie Luhmann 1990, 216f.).

Solche originalen Erkenntnisse müssen nicht nur ein Verbreitungsmedium verwenden, um die Chance zu erhalten, Teil eines Zitationsnetzwerkes zu werden, sondern es kommen unabhängig von dieser äußeren Form nur bestimmte inhaltliche Formen in Frage, in denen solche Wahrheitskommunikation überhaupt in Erscheinung treten kann. Durch die vorangegangene Erkenntnisproduktion der Wissenschaft ist es ausgeschlossen, bestimmte Ergebnisse für wahr zu erklären: Die Anwendung gewisser Methoden ist unseriös und jegliche Interessenskonflikte sind mitzukommunizieren. Solche und andere Formen prägen ein Erfolgsmedium, ein symbolisch generalisiertes

<sup>22</sup> Siehe zuletzt Bouabid und Larivière 2013. Allerdings hängt die sogenannte Halbwertszeit von wissenschaftlicher Literatur nicht nur vom Fachbereich ab. Außerdem lässt sich in allen Feldern zunehmend eine Verlängerung der »Lebensdauer« von Publikationen beobachten, die möglicherweise nicht zuletzt der Verwendung von Datenbanken und Suchmaschinen zuzuschreiben ist, die eine Recherche nach konkreten Forschungsfragen erleichtern: Man ist weniger auf das Lesen aktueller Zeitschriftenausgaben angewiesen.

Kommunikationsmedium<sup>23</sup>, das dann vorgeformt für Wahrheitskommunikation zur Verfügung steht und selbst Wahrheit genannt werden kann. Aber, so Stichweh (1994), was nicht publiziert ist, ist keine Wissenschaft, auch wenn es wahr ist. Es kann also das Medium der Wissenschaft, die Wahrheit, in Anspruch genommen werden, z. B. beim intellektuellen Disput in gemütlicher Runde, indem man sich an gewisse Regeln hält, z. B. die Ideen anderer nicht als die eigenen ausgibt. Wissenschaftliche Kommunikation auf der Ebene der Gesellschaft aber erfordert zusätzlich ein Verbreitungsmedium, das ihr zu weltweiter Wahrnehmbarkeit verhilft.

Verbreitungsmedien können ihren Zweck für die Wissenschaft, nämlich wahrscheinlicher zu machen, dass an Kommunikationsangebote angeschlossen wird (siehe Fn. 21), je besser erfüllen, desto mehr potentiell interessierte, ebenfalls Publizierende sie erreichen. Das kann einerseits durch Spezialisierung dieser Medien auf Wissenschaft, aber auch auf bestimmte Disziplinen oder sogar Themen erreicht werden. Auch eine komplementäre mediale Strategie kann Kommunikation wahrscheinlicher machen: Ein universales *one stop system*, ausgestattet mit individuellen Benachrichtigungsdiensten und ausgefeilter Suchmaschine sowie für das Auffinden durch externe Suchmaschinen optimierten Inhalten. Den Anforderungen an das Publikationssystem und die Publikation, die sich durch diese oder jene selektionsfördernde Methode ergeben, ist nicht leicht zu entsprechen. Allerdings kann selbst ihre unterdurchschnittliche Erfüllung durch die Reputation der Autorin ausgeglichen werden und ein Hinterfragen des Mediums hinfällig machen (siehe Ellison 2011). Reputation wird, neben vielen anderen Aspekten, die sich horizonthaft um das Thema des Forschungsergebnisses aufbauen, mit jeder Publikation mitkommuniziert oder vielmehr: kommunikativ reproduziert.

Es ist nicht möglich, allgemeine Merkmale festzulegen, die ein Publikationsmedium im Hinblick auf Ausrichtung und Reputation aufweisen muss. Hier gilt es, im je spezifischen Fall Voraussetzungen und Möglichkeitshorizont abzugleichen. Stattdessen soll hier aber auf die Potentiale verwiesen werden, die sich durch inhaltliche und technische Formate ergeben, zunächst unabhängig von einem konkreten Publikationsmedium, derer sich zwar viele im Internet finden, jedoch, wie wir am Ende dieser Studie sehen werden, wenige, welche die gegebenen Potentiale auszuschöpfen in der Lage sind.

## 2.2 Die Form des wissenschaftlichen Dokuments

Wenn eine Publikation die Verbreitung von Forschungsergebnissen ist, müssen diese in einer verbreitbaren Form vorliegen: als Dokument. Das können Messergebnisse sein,

<sup>23</sup> Zum Begriff des symbolisch generalisierten Kommunikationsmediums siehe Luhmann 1997, Kapitel 2, IX-XII. Medien generell können die Unwahrscheinlichkeit von Kommunikation bearbeiten, erfolgreiche Kommunikation wahrscheinlicher machen (siehe Fn. 21).

jedoch sind diese durch Menschen schwer konsumierbar – und: Ihnen würden jene Selbstreferenzen fehlen, die sie als *wissenschaftliche* Publikation leicht erkennbar machen. Es ergibt sich also in Anlehnung an die Logik George Spencer-Browns (1969) folgende Form:<sup>24</sup>



Selbstverständlich ist das Ansetzen zu einer Unterscheidung wie dieser kontingent, sie greift in einen Möglichkeitsraum ein, der nur durch seine Unbestimmtheit näher zu bestimmen ist.<sup>25</sup> Die Innenseite der Unterscheidung, hier z. B. das wissenschaftliche Dokument, muss hingegen vor dem Hintergrund des Unbestimmten bestimmt, bezeichnet werden, womit die Außenseite, hier das Forschungsergebnis, als davon Ausgeschlossenes nicht mehr alle Möglichkeiten enthält. Die Unterscheidung bezieht sich also, auch wenn nur die Innenseite bezeichnet wird, immer auch auf ihre Außenseite. Die Form ist untrennbar von den zwei Seiten, die sie unterscheidet. Damit bringt sie etwas Drittes in die Form: den Beobachter mit seinem je individuellen Möglichkeitsraum. Neue, nun nicht mehr mit allen Freiheitsgraden ausgestattete Möglichkeitsräume tun sich mit jeder Bezeichnung auf, die Form schafft ein neues Medium, in das sich weitere Formen einprägen, hier die wissenschaftliche Publikation, für die jeweils dasselbe gilt. Ein formloser Raum lässt sich nicht beobachten. Im Hinblick auf den Grad der Bestimmtheit bzw. die Freiheitsgrade jeder weiteren Unterscheidung ergibt sich durch Formkaskaden eine Hierarchie, die hier und im Folgenden helfen soll, besser zu verstehen, was eine wissenschaftliche Publikation ausmacht.

Bei der vorgeschlagenen Form einer wissenschaftlichen Publikation beginnt das Unterscheiden durch das Abgrenzen von wissenschaftlichen Dokumenten von der umfassenden Kategorie des Forschungsergebnisses. Als Forschungsergebnisse sollen hier als durch die Forschenden definierte Produkte des eigenen Forschungshandelns bestimmt werden. Rudolf Stichweh gilt das Forschungshandeln neben der Publikation

<sup>24</sup> Die Unterscheidung verhält sich analog zu jener in den Functional Requirements for Bibliographic Records (FRBR) allgemein für Vorlagen jeder Art formulierten (IFLA 2009), wobei die Exemplebene hier naheliegenderweise ausgespart wird:



<sup>25</sup> Für eine ausführlichere Begründung der sozialtheoretischen Anwendung der Formtheorie, einschließlich der naheliegenden Referenzen auf Gotthard Günther und Heinz von Foerster siehe Luhmann 1993.

als zweites autopoietisches Element der Wissenschaft (1994).<sup>26</sup> Natürlich erfolgt die Bestimmung von Forschungsergebnissen nach gewissen Normen und Werten, die eine Wahrnehmung von anderen und Anschlüsse wahrscheinlicher machen. Während sich die Publikation auf die Erzeugung von Selbstreferenzen und Selbstbeschreibungen der Wissenschaftskommunikation auf der Ebene der Gesellschaft spezialisiert hat, scheint das Forschungshandeln vielmehr für Rückbindungen an die Ebene der Interaktion zu sorgen, die notwendig sind, um den Alltag im Labor und am Schreibtisch zu meistern.

Es kann vielfältige wissenschaftliche Dokumente geben, mit deren Hilfe Forschungsergebnisse publiziert werden: Vor allem ist hier an Forschungsdaten zu denken, die aufgrund der gegenwärtigen Bemühungen um Standards für ihre Zitierbarkeit zunehmend den Charakter einer wissenschaftlichen Publikation erlangen. Die Unterscheidung zwischen wissenschaftlicher Publikation und wissenschaftlichem Dokument muss also heute als prekär angesehen werden, wobei jedoch derzeit nur sehr wenige Forschungsdaten Referenzen auf frühere Forschung anderer enthalten, die jedoch notwendig sind, um ihnen diesen Status verleihen. Solange Forschungsdaten primär als Paratext von Publikationen fungieren, wird sich daran nichts ändern.

Die Form der wissenschaftlichen Publikation lässt sich auch vom Dokument ausgehend konstruieren:



Wie einleitend festgelegt, ist von einer elektronischen Publikation als Normalfall auszugehen, entsprechend interessiert hier vor allem eine Konzeption des digitalen Dokuments (vgl. Buckland 1998), dessen Bestimmung dennoch von der Außenseite der Form her versucht werden soll: mit der Etymologie des Dokumentbegriffs. Das römisch-lateinische *documentum* bedeutet soviel wie »Beispiel, Beweis, Lektion«, vor allem in einem didaktischen Sinne, und verweist damit nicht zwingend auf etwas Materielles. Mit der Herausbildung der Nationalstaaten seit dem 17. Jahrhundert erhält der Dokumentbegriff einerseits seine Funktion zur Bezeichnung von etwas Geschriebenem, und andererseits wird die primär didaktische Konnotation durch eine legale und bürokratische abgelöst (Lund 2009). Ein Dokument musste eine Information beinhalten, die über den persönlichen Nahbereich des Besitzenden oder Verfassenden hinaus Relevanzpotential besitzt. Dem Zeitgeist des 18. Jahrhunderts entspricht dann auch das neue Erfordernis von Authentizität. Dokumente können »echt« oder »gefälscht«

<sup>26</sup> Das kleinste Element von Systemen ist nach Luhmann das Ereignis (1984, S. 76). Sowohl Publikationen als auch Forschungshandeln sind als einzelne Ereignisse dekomponierbar.

sein. Wer entscheidet über die Originalität? Je nachdem, um welche Art von Dokument es sich handelt, braucht es eine Person, der entsprechende Kompetenz und Macht zugeschrieben werden, z. B. einen Staatsbediensteten. Über die Originalität und Dokumentenhaftigkeit kann z. B. auch vor Gericht entschieden werden, wenn ein einfacher Brief zum Beweisstück wird. Bis dann beispielsweise die ersten archäologischen Fundstücke als Zeitdokumente galten, vergingen etwa weitere 100 Jahre. Dazu musste zunächst die Archäologie als Wissenschaft anerkannt sein.

Die Auflösung dieser Bedingungen – verbürgte Originalität und Beweisdienlichkeit – beginnt womöglich mit der Verbreitung von Textverarbeitungssoftware auf Heim-PCs, denn seitdem gilt bereits als Dokument, was die Dateierweiterung `.doc` hat; und wenn schon jeder Heim-PC-Inhabende solcherlei Dokumente erstellen kann, ist es nicht mehr weit, jedwede auf einem Speichermedium – für wen oder was auch immer – lesbar vorgehaltene Zeichenkette als digitales Dokument zu begreifen. Die Bezeichnung hat viel von ihrem Potential zur inhaltlichen Beschreibung verloren und wird vermutlich daher im englischsprachigen Raum zunehmend durch *resources* abgelöst, die allerdings noch inklusiver, aber dafür mit weniger störenden historischen Konnotationen versehen ist. Wenn also ein Begriff benötigt wird, der die vorgenannten historischen Aspekte des Dokuments exklusiv beinhaltet oder andere, zusätzliche, sollte nach einem Begriff Ausschau gehalten werden, der nicht bereits so sehr wie der Dokumentbegriff Teil der Alltagssprache geworden ist. Um beispielsweise die festgestellte Archivierungswürdigkeit eines digitalen Dokuments zu markieren, könnte von einer digitalen Archivalie gesprochen werden. Die Digitalität zeitigt selbstverständlich Folgen: Dadurch, dass ein digitales Original von seiner Kopie nicht unterschieden werden kann, wird auch der Begriff der Reproduktion hinfällig: Entweder die Kopie ist exakt und damit identisch oder sie es nicht und damit keine Kopie. Der Aspekt der Reproduzierbarkeit, der am Anfang dieser Studie der Berner Konvention für die Definition einer Publikation entlehnt wurde, ist bereits in der Voraussetzung der Digitalität einer gegenwärtigen wissenschaftlichen Publikation enthalten und muss daher nicht eigens in der Definition erwähnt werden. Nun mag man einwenden, dass nicht jede elektronische Publikation reproduzierbar, da kopiergeschützt ist. Solche Publikationen widersprechen der Definition an einer anderen Stelle, wenn man diese streng auslegt: Sie sind verbreitungsgehemmt. In einer weniger strengen Auslegung gelten sie als Publikation, sobald sie überhaupt an ein interessiertes Publikum verbreitet werden.

Wie steht das digitale Dokument zur Datei? Ein gutes Beispiel ist ein LaTeX-Dokument: Ist das der »Output«, z. B. das PDF, das erzeugt wird mit Hilfe von etwa einem Dutzend Dateien, die selbst ebenfalls in Frage kämen, oder ist es der »Input«, eine Datei nicht näher spezifizierten Inhalts mit der Endung `.tex`? Wie steht dann eine BibTeX-Literaturdatenbank dazu, deren Inhalte in die `tex`-Datei eingebunden und zu einem

PDF kompiliert werden können? Jede einzelne dieser Dateien bezeugt etwas und weist eine nach außen abgeschlossene Struktur auf, die an ihr Format gebunden ist. Im Gegensatz zum Datei- lässt der Dokumentbegriff jedoch eine technische Vielteiligkeit zu: Die äußeren Grenzen des Dokuments sind Sinn Grenzen: Ein LaTeX-Dokument könnte also sehr gut alle genannten Dateien umfassen, wenn sie in einem festen Sinnzusammenhang miteinander stehen, was bei aus einer bib-Datei erzeugten Zitationen ebenso festzustellen ist wie bei einer log-Datei, die eine Dokumentation über die Erstellung des PDFs enthält. Um den Gedanken an anderen Beispielen zu demonstrieren: Viele Digitalisate bestehen aus so vielen Bilddateien wie die Vorlage Seiten hatte und sowohl das gesamte Digitalisat als auch die einzelnen Seiten werden als Dokumente bezeichnet. Analog können die gesammelten Akten der Nürnberger Prozesse ein Dokument sein, ebenso jede einzelne Akte, jedes einzelne Schriftstück . . . Entscheidend ist der Bezug zu einem Sinnzusammenhang. Die Eingrenzung von einem nicht näher bestimmten Dokument auf ein digitales Dokument beschränkt den Begriff in nur einer Hinsicht: Es muss maschinenlesbar sein. Damit ist hier lediglich gemeint, das irgendein Computer in der Lage sein muss, das Dokument in einer Weise zu öffnen, so dass sich *mindestens entweder* für die Maschine weiterführende Verarbeitungsoptionen ergeben oder sich für den Menschen ein Sinnzusammenhang erschließen könnte. Andernfalls hätten wir es womöglich mit einer korrupten Datei zu tun und nicht mit einem Dokument. Für die Zwecke dieser Studie, die von einem soziologischen Zugang ausgeht und zu erkennen versucht, wie die Wissenschaft kommuniziert, ist dieser vergleichsweise offene, wenige Anforderungen stellende, nah am Sprachgebrauch gebaute Dokumentbegriff hilfreich.<sup>27</sup>

Für wissenschaftliche Publikationen hat dieser Dokumentbegriff die Folge, dass auch mehrere Dokumente in einer wissenschaftlichen Publikation zusammengebunden sein können, solange wenigstens eines der Dokumente oder die gesamte Aggregation die oben beschriebenen Merkmale aufweist, insbesondere das Merkmal der Referenzen auf andere wissenschaftliche Publikationen. Die Möglichkeit, durch Webtechnologien temporär »Wirkungszusammenhänge« zu visualisieren (Kaden 2013) ist zwar neu, aber lediglich eine andere Darstellungsform für Verweisungszusammenhänge, die Publikationen immer schon aufwiesen. Die Systematisierung der Sammlung von Kontextinformationen über AutorInnen wie sie z. B. mittels ORCID forciert wird, verleiht diesen Zusammenhängen natürlich eine neue Quantität, aber ist für das Wissenschaftssystem nichts dem Prinzip nach Neues. Noch ist das globale AutorInnenidentifikationssystem nicht durchgesetzt; einige Details von ORCID bieten neuartige Optionen, z. B. durch die Möglichkeit der Selbstverwaltung durch die AutorInnen. Es gibt derzeit keine Beispiele dafür, dass die Visualisierung von solchen Verweisungszusammenhängen mit einer zunehmend hohen Engmaschigkeit die Publikation als abgeschlossenen Sinnzusammen-

<sup>27</sup> Für ein alternatives Konzept vgl. Pédaque 2003.

hang sprengen würde, wie Kaden vermutet, denn die sie konstituierenden Anschlüsse in beide Richtungen bleiben kommunikationsstrukturell unverändert, wenn sie auch in ihrer technischen und inhaltlichen Gestaltung variieren und somit die kommunikativen Formen der Anschlüsse in der sachlichen und zeitlichen, jedoch kaum in der sozialen Sinndimension<sup>28</sup> unvertraut sind.

Letztendlich handelt es sich bei solchen Anwendungen um neuartige Paratexte, denn sie verorten Publikationen. Metadaten haben eine ähnliche Funktion, unterscheiden sich von Paratexten jedoch dadurch, dass der Begriff nicht mehr besagt als ihre Funktion, Daten zu beschreiben. Der Begriff lässt beliebige perspektivische Verschiebungen zu: Gemeint sind jene beschreibenden Daten, die bei der Beobachtung eines bestimmten Datenclusters unmittelbar relevant werden. Eine wissenschaftliche Publikation enthält Metadaten, wenn ihre Struktur durch Textauszeichnung maschinenlesbar gemacht wurde. Eine Maschine kann dann also z. B. einen Titel auslesen, bestimmen, in wie viele Kapitel der Text unterteilt ist und zitierte Publikationen erkennen und ggf. auflisten. All dies lässt sich auch anhand der Untersuchung des Paratexts zeigen. Der entscheidende Unterschied ist, dass Metadaten zwingend auf ein (heute nahezu durchweg technisches) System angewiesen sind, das sie in Beziehung zu anderen Metadaten setzen kann, weil sie sonst ihre Funktion und damit ihr Definiens verlieren. Früher dienten in Bibliotheken z. B. die Zettelkataloge als ein solches System, heute sind es Datenbanken. Paratexte können sowohl intellektuell als auch technisch kontextualisiert werden. Es ergibt sich also folgende Unterscheidung:



Die gegenwärtig vielgenutzten Publikationsformate ermöglichen, lediglich einen Teil des Paratextes in Form von Metadaten maschinell als solchen erkennen und verarbeiten zu können. Diese Publikationsformate verhindern, dass Maschinen Kernaussagen aus einer Publikation extrahieren und mit Kernaussagen anderer Publikationen verknüpfen und vergleichen können, um beispielsweise einen Überblick über den Forschungsstand zu einem bestimmten Thema vorzuschlagen: Eine Aufgabe, die häufig Servicepersonal oder studentische Hilfskräfte übernehmen, wo sie zur Verfügung stehen. Der zeitliche Aufwand dafür ist enorm und die Fehleranfälligkeit groß.

<sup>28</sup> Zum Sinnbegriff bei Luhmann siehe Luhmann 1984, Kapitel 2.

### 3 Bedarf für ein neues Medium wissenschaftlicher Kommunikation

#### 3.1 Die Notwendigkeit neuer selektionsunterstützender Verfahren

Die wissenschaftliche Publikation ist nach wie vor für die Wissenschaftskommunikation elementar: vorläufig unverzichtbar und unersetzbar. Das Wissenschaftssystem kennt derzeit kein anderes Instrument, seine Einheit herzustellen. Es ist auf Nachvollziehbarkeit angewiesen, um auf das bekannte Wissen aufsetzen zu können und seine Funktion zu erfüllen, nämlich die Gesellschaft mit neuem Wissen zu versorgen. Alle anderen Strukturen des Wissenschaftssystems sind Hilfsmechanismen, die selbst davon abhängig gemacht werden, dass publiziert wird: Allen voran Reputation, Stellen in Wissenschaftsorganisationen, Förderungen etc.

Eine Druckvorlage für die höchsten Ansprüche an Typographie herzustellen, kann heute leicht gelingen. Voraussetzungen sind lediglich ein durchschnittlich leistungsstarker Heimcomputer und die freie Software LaTeX, sowie Computerliterarität und Geduld. Das Ergebnis, zumeist eine PDF-Datei, wird den Anforderungen der meisten Buchverlage, die sich der Zuständigkeit für die Erstellung einer Druckvorlage entledigt haben, gerecht. Es braucht keine empirische Erhebung, um festzustellen, dass PDF das vorherrschende und auch meistnachgefragte Format für gegenwärtige wissenschaftliche Publikationen ist. Das mag verwundern, wenn man einmal versucht hat, es mit einem mobilen Lesegerät zu verwenden. Selbst ein PDF/A ist für die Langzeitarchivierung weit weniger geeignet als auf XML basierende Formate, da es nicht ausschließlich standardkodierte Zeichen enthält und eine eventuell vorhandene Struktur nicht ohne spezielle Software zugänglich ist. Jedes *text mining* von PDFs ist mit mühsamer Handarbeit verbunden, so dass man sich trotz bemerkenswerter Fortschritte in entsprechenden Technologien unweigerlich die Frage stellen muss: Wäre es nicht an der Zeit, jetzt, da digitale Dokumente immer häufiger hergestellt und rezipiert, aber immer seltener ausgedruckt werden, über Alternativen nachzudenken?

Für das Wissenschaftssystem kündigen sich Veränderungen an, die es von einem vergleichsweise exklusiven System zu einem hyperinklusiven machen. Das lässt sich an Statistiken ablesen: Während die Weltbevölkerung seit den 1970er Jahren nicht mehr hyperexponentiell wächst, da die Wachstumsrate seitdem sinkt, kann man das für den Anteil der Forschenden an dieser Bevölkerung nicht behaupten: Das globale Bildungsniveau steigt; im Jahr 2009 gab es auf der Welt etwa 8 Mio. Forschende. Sowohl die Anzahl beteiligter Individuen als auch der Publikationen wächst seit zwei Jahrzehnten jährlich um etwa 3,3% (Ware und Mabe 2009). Zu erwarten ist ein noch stärkeres Wachstum. Das Bevölkerungswachstum liegt dagegen 2010 bei

1,1%.<sup>29</sup> Hinzu kommen Entwicklungen wie jene der *Citizen Science*, die Laien z. B. bei der Sammlung von Beobachtungsdaten einbezieht. Insbesondere in Astronomie und Ornithologie sind Bürgerbeteiligungen an wissenschaftlichen Entdeckungen sehr traditionsreich. Dennoch ist das Phänomen in seiner gegenwärtigen Disposition neu: 2012 fand eine erste Konferenz zum Thema statt, die *Conference on Public Participation in Scientific Research* in Portland, Oregon. Das EU-geförderte *Socientize Project* hat ein Forum eingerichtet, in dem man sich über Citizen-Science-Projekte informieren kann. Die Beteiligung soll gefördert werden, weswegen man auch für die EU-Mitgliedsstaaten entsprechende Empfehlungen formuliert hat. Ausdrückliches Ziel ist, »durch das Öffnen der Labor-Türen und die Einbindung von Laien-WissenschaftlerInnen Spitzenforschung zu betreiben«<sup>30</sup>, und zwar über elektronische Infrastrukturen, wie sie erst seit der Etablierung des Internets entwickelt werden.<sup>31</sup>

Überhaupt hat sich durch technische Entwicklungen die Erzeugung und Analyse von Forschungsdaten und damit die Produktion von Forschungsergebnissen enorm beschleunigt. Da überrascht es fast, dass relativ betrachtet gar nicht mehr publiziert wird als früher, aber die Anzahl jener Publikationen, die gelesen werden müssen, steigt trotzdem permanent. Zwar reagiert das Wissenschaftssystem mit einer immer feineren Binnendifferenzierung in Subdisziplinen und nur mehr einzelne Themen, auf die sich kleine, Interaktionen ermöglichende *communities* beschränken, aber das entlastet Forschende nicht gänzlich vom erhöhten Lektürepensum, denn die eigene Forschung muss kontextualisiert werden, um auch außerhalb eines sehr beschränkten Kreises wahrgenommen zu werden und damit die Funktion der wissenschaftlichen Publikation zu erfüllen. Die zunehmende Differenzierung zeitigt also auch unerwünschte Folgen, die dann wiederum eine Aktualisierung von Konzepten der Inter- und Transdisziplinärität auslösen. Die gegenwärtige Entwicklung lässt jedoch nicht vermuten, dass sich die Komplexität damit in den Griff bekommen lässt.

In der wissenschaftlichen Kommunikation ist zunächst einmal alles, was der angewendeten Theorie und Methode nach Wahrheit für sich beansprucht und sich dabei an den geltenden Normen orientiert, potentiell relevant. Selektionen relevanten neuen Wissens

<sup>29</sup> UNO, World Population Prospects: The 2012 Revision, <http://esa.un.org/unpd/wpp/>, zuletzt besucht am 13. April 2014.

<sup>30</sup> Societize, FP7-Projektnummer RI-312902, Homepage: <http://societyize.eu/>, zuletzt besucht am 13. April 2014.

<sup>31</sup> Aufgrund dieser Entwicklungen könnte das Wissenschaftssystem nachträglich eine asymmetrische Rollenstruktur nach dem klassischen Muster z. B. des Rechtssystems ausbilden: Leistungsträger treffen auf ein Publikum, das eine wesentliche Rolle für die Strukturierung der Kommunikation spielt. Die Aussage »Das Publikum der Wissenschaftler sind die Wissenschaftler.« (Bühl 1974, S. 242; siehe auch Luhmann 1990, S. 624) würde dann abgelöst.

finden vor dem Hintergrund von Variationen des bereits bekannten Wissens statt.<sup>32</sup> Für diesen zentralen Selektionsmechanismus braucht es unterstützende Mechanismen, sozusagen Vorfilter, um die enorme Komplexität bewältigen zu können. Dazu gehört z. B. die Einteilung in Disziplinen. Luhmann unterscheidet außerdem eine implizite von einer expliziten Selektion (1990, S. 577). Die explizite Selektion setzt voraus, dass der Beitrag diskutiert und damit bestätigt oder widerlegt wird. Die implizite Selektion setzt davor an: Sie schließt bereits einen Großteil der Kommunikationsofferten aus, und zwar schlicht dadurch, dass sie aus den unterschiedlichsten Gründen nicht rezipiert werden. Verschiedene Mechanismen unterstützen die implizite Selektion, z. B. die Zuordnung zu bestimmten Publikationsmedien per klassischem *peer review* oder der *impact factor*. Sie setzen zumeist bereits vor der Publikation an und verhindern somit, dass es zu einer expliziten Selektion kommen kann.

Solche Mechanismen verlieren jedoch zunehmend an Legitimation, da sie z. B. dem politisch geförderten Anspruch nach Demokratisierung und Transparenz sowie Kosteneffizienz zuwider laufen.<sup>33</sup> Sie begünstigen die unbemerkte Wiederholung ähnlicher Forschung, die nur dann verhindert werden kann, wenn handwerklich gute Forschungsergebnisse nicht bereits der impliziten Selektion zum Opfer fallen. Die Kapazitätsgrenzen von Printmedien haben bislang<sup>34</sup> die relevanzgeleitete Selektion wissenschaftlicher Dokumente vor der Publikation erfordert – dieses Erfordernis fällt ersatzlos weg. Die Selektion von Kommunikationsangeboten ist höchst heikel. Sie muss sich heute verwehren können gegen jeden Nepotismus- oder Korruptionsverdacht. Es scheint derzeit nur eine Instanz zu geben, die in diesen Hinsichten unbedenklich ist, weil man ihr keine Intention unterstellen kann: die Maschine, die kein einzelner Mensch programmieren, aber vollständig auf der Grundlage offener Standards arbeitend von einer offenen *community* kontrolliert werden kann. Maschinelle Verfahren bieten nur dann einen Vorteil gegenüber den etablierten Selektionsmechanismen, wenn alle Publikationen für Mensch und Maschine frei zugänglich und nach gewissen Standards aufbereitet, also vergleichbar sind. Der Vorteil liegt dann in der potenzierten Verarbeitungskapazität und einer Eindämmung von zufälligen Selektionen. Die enorme Herausforderung, leistungsfähige Verfahren und Standards zu entwickeln, zu etablieren und für Menschen nachvollziehbar und überprüfbar zu machen, lässt diese Idee auf den ersten Blick jedoch wenig aussichtsreich und seine Durchsetzung unwahrscheinlich erscheinen.

<sup>32</sup> Nach Luhmann 1990, S. 576, erzeugen die beiden Unterscheidungen einen Zirkel: Variationen können nur dann von der Wissenschaftskommunikation selektiert werden, wenn sie als potentiell wahr gelten. Die Selektion selbst verändert jedoch das, was als wahres Wissen gilt und zum Ausgangspunkt für neue Variationen und Selektionen wird.

<sup>33</sup> Der gleiche Anspruch hat höchstwahrscheinlich zur aktuellen Förderung von *Citizen Science* geführt.

<sup>34</sup> In den Naturwissenschaften gibt es mittlerweile eine Reihe von *publishern*, die ihre Selektionskriterien formal und methodologieorientiert, nicht nach Relevanz, ausgerichtet haben. So muss PLOS ONE trotz teils scharfer Kritik faktisch als Erfolgsgeschichte gelten.

Diese Studie unternimmt den Versuch, das Unwahrscheinliche wahrscheinlicher zu machen.

### 3.2 Von der Wissens- zur Wissenschaftsrepräsentation: Eine Skizze

Der Erfolg des Internets legt nahe, ein Publikationsservice aufzubauen, das sich an eine globale Öffentlichkeit richtet. Die Chancen, die sich aus interdisziplinärem Austausch ergeben, sind kaum zu unterschätzen. Es spricht also viel für ein wissenschaftsweites Service. Viele der wertvollen Innovationen, die das Publikationswesen durch die Verbreitung des Internets und Open Access erfahren hat, waren jedoch meist zunächst nur auf einzelne Zeitschriften oder Portale bezogen, haben sich bei Erfolg innerhalb derselben Disziplin verbreitet und sind so für ferne Disziplinen aufgrund der Behandlung von disziplinären Spezialbedürfnissen wenig attraktiv gewesen. Mittlerweile sind eben jene Bedürfnisse viel bekannter geworden und können bei der Entwicklung von Publikationssystemen besser berücksichtigt werden. Nur: Wie soll man Bedürfnisse erheben, die sich immer nur an den Mängeln des Gegenwärtigen orientieren können? Es scheint zunächst angezeigt, ein grobes Konzept zu entwickeln, das dann diskutiert werden kann. In dieser frühen Phase einer Entwicklung ist viel wichtiger, die zu lösenden Probleme zu erkennen, als einen detaillierten Entwicklungsplan vorzulegen. Dieser Modus soll nun dennoch kurz verlassen werden, um die Grundanforderung zu formulieren: Benötigt wird ein System, das folgendes zweistufiges Verfahren unterstützt:

1. die inhaltliche Aufbereitung der zu publizierenden bzw. publizierten Forschungsergebnisse, die zunächst Maschinen und dadurch den Menschen in die Lage versetzen, bei möglichst geringem Aufwand die Informationen zu kontextualisieren, und
2. die fachliche Beurteilung von Redlichkeit, Qualität und Relevanz, die unterschiedlichste Formen annehmen kann.

Die beiden Stufen sollten sich idealerweise wechselseitig aneinander steigern, indem die fachliche Beurteilung als Kontextinformation in die inhaltliche Aufbereitung einfließt. Umgekehrt muss eine fachliche Beurteilung in einem offenen Publikationssystem zunächst einmal akquiriert werden, was am ehesten gelingt, wenn die Publikation vielfach mit bereits als relevant beurteilter Information verknüpft ist. Vorschläge für Verfahren der Post-Publikationsqualitätssicherung liegen bereits einige vor (siehe z. B. Kriegeskorte 2012). Solche Verfahren wirken sich dann zwangsläufig auf die typischen Geschäftsprozesse von *publishern* aus, weshalb ein neues Modell vorgeschlagen wurde: Das Overlay-Modell (Ginsparg 1997; Smith 1999 sowie Priem und Hemminger 2012), das darauf beruht, ein Manuskript zunächst auf einem Open-Access-Dokumentenserver

sicher und persistent zu archivieren, so dass Dienstleister oder andere Einrichtungen Beiträge auswählen und eine oder mehrere Funktionen, die sonst üblicherweise *publisher* übernehmen, nicht-exklusiv ausführen können, z. B. die Qualitätssicherung. In dieser Studie soll es nur um die erste Stufe gehen: Wie sollte dieses Manuskript aufbereitet werden, damit es als wissenschaftliche Publikation möglichst ohne Selektionsvorbehalte selektionsunterstützenden Verfahren zugeführt werden kann?

Gerade bei wissenschaftlichen Publikationen lassen sich Kontextinformationen, die weiter oben Paratexte genannt wurden, vergleichsweise einfach identifizieren, weil sie entweder durch die Publikation mitgeliefert werden, wie bei den basalen Titeldaten und Zitationen, oder leicht in Erfahrung zu bringen sind, z. B. die Organisationszugehörigkeit über AutorInnenidentifikationssysteme oder Institutionswebseiten. Für die inhaltliche Erschließung werden immer präzisere Verfahren entwickelt, die sich durch einen zunehmenden Grad der Automatisierung auszeichnen. Zur Zeit fokussiert man dabei auf die Verwendung eines prädefinierten Vokabulars, um automatisiert große Mengen von Dokumenten zu klassifizieren und zu verschlagworten. Solche Verfahren haben den Nachteil, das Entscheidende an der Publikation nur eingeschränkt abbilden zu können: die Variation des bekannten Wissens. Nach Stock und Stock (2008, 30f.) können Wissensordnungen nur für die Repräsentation von Normalwissenschaft (Kuhn und Krüger 1978) Anwendung finden. Paradigmenwechseln in der Wissenschaft folgen gemeinhin Paradigmenwechsel in der Wissensorganisation. Wäre nicht auch ein Vorgehen denkbar, durch das die Publikation, autorisiert durch denjenigen, der seine Forschungsergebnisse veröffentlicht, angeschlossen wird an ein frei zugängliches, dynamisches Wissenschaftsnetzwerk, das die Wissenschaft in Echtzeit repräsentieren kann? Es geht hier um die Gleichzeitigkeit der Erzeugung neuen Wissens – das, selbst wenn es nach Kuhn als Normalwissenschaft charakterisiert werden muss, durch bestehende Wissensorganisationssysteme nicht abgebildet werden kann – und der Wissensrepräsentation, die zunächst einmal seiner Erschließung dient, dann aber auch emergentes Potential hat.

Selbstverständlich sind auch dafür prädefinierte Vokabulare hilfreich, aber sie könnten, statt *top-down* aus einer die Wissenschaft beobachtenden Perspektive, auch *bottom-up* durch die Wissenschaft selbst entwickelt werden, gleichsam als neue universale Wissenschaftssprache. Neu deswegen, weil sie als Medium nicht lediglich die lose gekoppelten Elemente einer natürlichen Sprache verwenden oder sich auf eine Disziplin beschränken dürfte. Das Medium sollte vielmehr ein fachsprachlicher Wortschatz sein, der sich aus den Wissenschaftssprachen aller Disziplinen zusammensetzt und durch individuelle Beiträge erweiterbar ist.

Der Informationswissenschaft und -wirtschaft sowie den Bibliotheken und verwandten Serviceeinrichtungen fiel dann die Aufgabe zu, dafür geeignete Infrastrukturen bereit-

zustellen und zu pflegen. Vergleichsweise unproblematisch kann man sich eine enorm umfangreiche Enzyklopädie vorstellen, die jede Bezeichnung, seine Definition und damit seinen Begriff, wie er in der resümierenden Formulierung eines Forschungsergebnisses vorkommt, in den unterschiedlichsten Sprachen festhält. Über die Zeit würde sich so unvermeidlich ein Überblick über die jeweiligen Begriffsgeschichten ergeben, da die Bedeutung abstrakter Konzepte insbesondere in den Geisteswissenschaften ständig zur Disposition steht. Schwerer zu denken ist dann schon an ein universales Vokabular anderen Typs, ohne das die Formulierung von solchen Ergebnissen unmöglich wäre: Konnektoren müssen die Begriffe einerseits je nach intendierter Aussage treffend verbinden und andererseits auf eine Weise, dass Aussagen, die einen der Bezeichner dieser Begriffe verknüpft mit einer anderen Bedeutung verwenden, damit verglichen werden können, um z. B. widerstreitende Aussagen auch für Maschinen erkennbar zu machen. Erleichtert wird die intellektuelle Bearbeitung des so entstehenden Wissenschaftsnetzwerks durch vergleichsweise stabile Konnektoren bzw. Relationen. Außerdem muss dieses Konnektor-Vokabular erst recht fachunspezifisch, universal und vor allem überschaubar sein, um auch interdisziplinär intellektuell unmittelbar erfassbare Verknüpfungen zu ermöglichen. Dieses Vorgehen schließt eine hierarchische Ordnung der Begriffe keinesfalls aus. Im Gegenteil: Solche Ordnungen könnten die Suche nach Analogien der einerseits in der Publikation und andererseits im Wissenschaftsnetzwerk verwendeten Begriffe erleichtern und der unbemerkten Schaffung von Synonymen entgegen wirken.

Hierarchien sind jedoch so lange ungeeignet für diesen Verwendungskontext, wie sie auf exklusive Positionen einzelner Begriffe innerhalb der Hierarchie beharren. Solche Hierarchien mögen für stabile Taxonomien in der Biologie geeignet sein, nicht jedoch für Wissen, das umstritten ist. Ein gutes Beispiel sind die Formen einer wissenschaftlichen Publikation aus dem vorherigen Kapitel: Die »wissenschaftliche Publikation« kann gleichzeitig hierarchisch unterhalb von »Dokument« und »Forschungsergebnis« eingeordnet sein – beides und weiteres ist ebenso gut möglich. Welche Unterscheidung zum Zuge kommt, ist allein abhängig vom Beobachter. Ein die aktuelle wissenschaftliche Kommunikation gut repräsentierendes Wissenschaftsnetzwerk muss alle dokumentierten Beobachtungen in sich aufnehmen können und wird dadurch zwar nicht beobachterunabhängig, aber überaus flexibel. Quantitativ stärkere Unterscheidungen und Relationen können von einer Maschine bei der Verarbeitung der Informationen entsprechend gewichtet werden. Es kann nicht mehr darum gehen, als einzelner Beobachter den Überblick über eine Wissensordnung behalten zu wollen, wie das für Thesauri noch gut möglich und erwünscht ist – wenigstens innerhalb eines Fachgebiets. Dort hat man mit Assoziationsrelationen versucht, kontingente Hierarchisierungen aufzufangen, musste sich jedoch immer für eine dominante Hierarchisierung entscheiden. Dieses Konzept muss aufgegeben werden, wenn das Ziel sein soll, die Wissenschaft

möglichst umfassend zu repräsentieren und, vielmehr noch: medial zu stützen. Die Voraussetzungen für den Einsatz von Technologien, die es ermöglichen, nicht mehr nachträglich Wissensordnungen erzeugen zu müssen, sondern ad hoc, während der Erzeugung des Wissens, sind mit semantischen Webtechnologien heute gegeben, wie unten gezeigt werden wird.

Wenn AutorInnen mit der Hilfe eines möglichst leicht zu bedienenden Systems in die Lage versetzt werden könnten, ihre Beiträge selbst in eine netzwerkartige Struktur einzubauen oder einen Dienstleister dazu zu autorisieren, wäre dies eine Chance zur Bearbeitung eines der gravierendsten Probleme der informationswissenschaftlichen Wissensrepräsentation: Jenes der Unmöglichkeit der verlustfreien Übertragbarkeit von Informationen zwischen psychischen Systemen. Kommunikation ist nicht die Lösung, sondern das Problem. Im Kontext der Erläuterung der drei Selektionen der Kommunikationen wurde es bereits deutlich: Verstehen kann nicht viel mehr bedeuten, als sich angesprochen zu fühlen. In der Informationswissenschaft ist das durchaus bekannt,<sup>35</sup> aber bislang schien man sich mit diesem in der Tat nicht zu eliminierenden Problem tendenziell resignierend zu arrangieren. Selbstverständlich besteht dasselbe Problem auch zwischen Autorin und Rezipient, aber immerhin wird keine zusätzliche Ebene eingezogen, die manchmal eben eher verstellt als vermittelt.

Wenn das Verstehen schon bei zwischenmenschlicher Kommunikation unwahrscheinlich ist, wieso sollten wir der Illusion anhängen, Computer könnten die besseren Vermittler sein? Es geht hier nicht um Kommunikation zwischen Computern oder zwischen Computern und Menschen: Nur die Gesellschaft kommuniziert, und zwar beruhend auf den vielfältigsten Erwartungsstrukturen. Sozialstrukturen *sind* Erwartungen. Computer bauen keine Erwartungen auf, die enttäuscht werden könnten und tragen daher nicht unmittelbar zur Bildung von Gesellschaftsstrukturen bei. Menschen erwarten etwas von Computern, z. B., die Wahrscheinlichkeit von Kommunikation, von Anschlussfähigkeit und damit von Verstehen – im Rahmen des Möglichen – zu erhöhen. Wie diese Erwartungen aufgebaut werden, wäre dann die eigentlich interessante Frage. Nun soll jedoch genauer gezeigt werden, wie das angedeutete Wissenschaftsnetzwerk auf Grundlage wissenschaftlicher Publikationen entstehen könnte und welche Ansätze dazu bereits bestehen. Das treffende Stichwort scheint hier das semantische Publizieren zu sein.

### 3.3 Wissenschaftskommunikation und Semantik

Semantiken dienen der Selbstbeobachtung der Gesellschaft; es sind »bewahrenswerte Sinnvorgaben« (Luhmann 1997, S. 887). Die Autorität von Semantiken wird kommuni-

<sup>35</sup> Siehe z. B. Gradmann 2009 und die dort angegebene, teilweise ältere Literatur.

kativ hergestellt, ihre Zweckdienlichkeit im jeweiligen Kontext permanent überprüft. In der Systemtheorie steht der Begriff der Semantik in engem Zusammenhang mit jenem der Gesellschaftsstruktur (siehe Stäheli 1998). Damit beispielsweise innerhalb einer wissenschaftlichen Disziplin Forschungsergebnisse konzentriert diskutiert werden können, reproduziert die Kommunikation dort immer wieder eine Spezialsemantik, eine Fachsprache. Insbesondere in Interaktionssystemen, im Labor, erlaubt großer Zeitdruck keine explizite Definition der jeweils verwendeten Bezeichnungen. Man geht von einem geteilten Verständnis aus: Semantiken reduzieren die Komplexität von andernfalls zu kommunizierendem Sinn. Die universitäre Ausbildung ist daher nicht zuletzt darauf ausgelegt, die Verwendung von Fachsprachen anzutrainieren.

Eine Technik, die von der Gesellschaft also bereits vor der Erfindung der Schrift zur Reduktion von Komplexität erfolgreich eingesetzt wurde – Semantik ist lediglich auf den Einsatz von Symbolen angewiesen –, könnte nun als Vorlage für ein computergestütztes Wissenschaftsnetzwerk verwendet werden, an dem sich die wissenschaftliche Kommunikation orientieren kann. Dieses System heißt nicht Informationsnetzwerk, weil es über die Strukturierung von Daten hinausgehen muss:<sup>36</sup> Entscheidend ist die Kontextualisierung der Informationen durch die Verknüpfung mit wissenschaftlicher Kommunikation. Ein allgemeines Wissensnetzwerk, das sich an der gesellschaftlichen Kommunikation ohne die Einschränkung auf die Wissenschaft orientiert, wird für seinen nutzenbringenden Einsatz auf die Besonderheiten eines Wissenschaftsnetzwerks weitestgehend verzichten müssen, allem voran das ständige Mitführen von Kontingenz. Anwendungsbereiche für semantische Technologien entwickeln sich derzeit insbesondere auch in Wirtschaft und Politik, wo sich jeweils auch ein spezifischer Umgang mit Wissen ausgeprägt hat. So ist man in beiden Bereichen darauf angewiesen, Bestimmtes als unbestreitbar anzunehmen, z. B. die Eigenschaften eines angebotenen Produkts oder die Haushaltsdaten einer Volkswirtschaft. Die Kontingenzen kommen hier z. B. bei Preisen vor, die sich von Anbieter zu Anbieter unterscheiden, oder bei parteipolitischen Positionen zu bestimmten Haushaltsposten, die sich von Fraktion zu Fraktion unterscheiden. Hierfür bietet sich also tatsächlich eine Ontologie nach mittlerweile klassischem Model an. Wohl aber birgt ihre Anwendung Risiken, wenn nicht hinreichend Kontingenzen zugelassen werden, denen hier aber nicht weiter nachgegangen werden kann.

Eine Vernetzung aller genannten Anwendungsbereiche untereinander liegt nahe. So kann es Forschung über Unternehmen oder politische Entscheidungen geben oder poli-

<sup>36</sup> Hier ist die in der Informationswissenschaft klassische Trias von Daten, Information und Wissen angesprochen, die jedoch mit den systemtheoretischen Verwendungen dieser Bezeichnungen nicht harmoniert: Information ist in der Systemtheorie ereignishaft und hat nur so lange Bestand, wie sie überrascht, während Wissen sich davon durch Aufbewahrungs- und Übertragungsmöglichkeiten unterscheidet.

tische Förderung von Unternehmen, die computergestützt repräsentiert werden. Um diese Fremdreferenzen abzubilden, ist man nicht auf eine verbindende *upper ontology* (siehe auch Kapitel 3.4) angewiesen, sondern lediglich auf URIs und passende Relationen. Semantiken hängen von der Gesellschaftsstruktur ab und diese ist funktional differenziert. Das macht es möglich, auf die Wissenschaft spezialisierte Technologien einzusetzen, ohne auf Beobachtungsmöglichkeiten in der Umwelt der Wissenschaft verzichten zu müssen. Wenn es eine gemeinsame Ontologie im Wissenschaftsnetzwerk braucht, dann, um z. B. disziplinübergreifend zwischen Forschungsgegenstand und Methode unterscheiden zu können.

Begriffe der Fachsemantiken werden auch ohne Computerunterstützung immer wieder verwendet – das erst macht sie zu Semantiken. Die Idee ist demnach, diesen bereits aus der vollständig analogen Kommunikation bekannten Vorgang durch die Verwendung des Internets zu erleichtern, und zwar nicht als Selbstzweck, sondern zur Unterstützung der – analogen – Forschung. Ein weiterer Vorteil ist, dass so argumentiert leicht vermittelt werden kann, wozu das Netzwerk dient. Es muss keine neue Logik angenommen oder Verständnis für das Funktionieren von Technologien aufgebracht werden. Es genügt die Reflexion der alltäglichen Erfahrung mit der wissenschaftlichen Kommunikation, darunter der Bedeutung der Autorschaft, und ihre Übertragung auf ein neues Medium:

bringing humanity fully into the information loop requires data structures and computational techniques that enable us to treat social expectations and legal rules as first-class objects in the new Web architecture (Hendler und Berners-Lee 2010).

Die Entwicklung von Semantiken wird dann automatisch mitgeschrieben und die Autorschaft von Variationen nachgewiesen. Da trotz immer weiter verbreiteter mobiler Technologien zunächst nicht davon ausgegangen werden kann, dass Menschen in Interaktionen Semantiken computergestützt in Anspruch nehmen werden, ist eine vorläufige Beschränkung der Anwendung des Wissenschaftsnetzwerks für schriftliche Kommunikation, also für Recherche und Publikationen, sinnvoll. Wenn im Folgenden also von einer semantischen Publikation die Rede ist, sei eine wissenschaftliche Publikation gemeint, *in der Aussagen über Forschungsergebnisse mittels Semantiken formuliert werden, die in einem frei zugänglichen und wiederum durch semantische Publikationen erweiterbaren Publikations- und Recherchesystem nachgewiesen sind*. Das System, das Wissenschaftsnetzwerk genannt werden soll, arbeitet dabei *auf der Grundlage standardisierter semantischer Webtechnologien und reduziert erheblich die Komplexität und Unwahrscheinlichkeit wissenschaftlicher Kommunikation*.

Der systemtheoretische Begriff von Semantik harmoniert besser mit jenem durch das Semantic Web in Anspruch genommenen als die in den Geisteswissenschaften verbreitetere Lesart, wo man »Semantik« mit »Bedeutung« übersetzt. David Shotton (2009 und

2012) entwickelte im Anschluss an den »Erfinder« des Semantic Web, Tim Berners-Lee (2001) eine extensionale Definition des semantischen Publizierens von wissenschaftlichen Artikeln unter Verwendung einer solchen Bestimmung von »Semantik«: Beim semantischen Publizieren geht es demnach um »anything that enhances the meaning« (Shotton 2009), im Einzelnen könne dies durch Folgendes erreicht werden

- die Anreicherung der Bedeutung der Publikation, z. B. durch die visuelle Verstärkung von Aussagen oder die Einbettung in einen dynamisch abzurufenden, selbst nicht in der Publikation enthaltenen Kontext,
- ihre Verlinkung mit in irgendeiner Weise ähnlichen Publikationen,
- das Verfügbarmachen von zugrundeliegenden Daten, auch Bildern, in einer Form, die einen größtmöglichen Wiederverwendungswert hat,
- das Versehen mit maschinell verarbeitbaren, umfangreichen Metadaten und
- die Erleichterung ihrer Auffindbarkeit.

Nicht jedes dieser Merkmale leuchtet als Bedeutungsanreicherung unmittelbar ein, insbesondere die letzten beiden Aspekte. Umfangreiche Metadaten erleichtern die Auffindbarkeit, deshalb erweitert sich aber noch nicht die Bedeutung der Publikation, es sei denn, man begreift Bedeutung nicht in seiner sachlichen, sondern in seiner sozialen Dimension. Überhaupt scheint eine zur Erweiterung offene Sammlung von Maßnahmen für ein sehr allgemeines, auch durch klassische Techniken des wissenschaftlichen Schreibens erreichbares Ziel als Definition ungeeignet, da sie dadurch extrem inklusiv ist und wenig instruktiv wirkt. Von den zahlreichen Anreicherungstechniken, die Shotton 2009 anhand eines bereits publizierten Artikels demonstriert hat, können nur zwei hier Relevanz beanspruchen: Das semantische Mark-up von im Text verwendeten Termen, das sie mit Informationsquellen im Internet verknüpft, z. B. Fachdatenbanken, und eine semantische Repräsentation aller verwendeten Referenzen, die nicht nur die Titeldaten enthält, sondern auch Informationen darüber, *wie* diese Literatur referenziert wurde. Auf diese Ontologie wird zurückzukommen sein (siehe Kapitel 3.4).

Ganz ähnlich wie Shotttons Definition ist jene des nun naheliegenden *enhanced publishing* gebaut, die aus dem DRIVER-II-Projekt (Verhaar 2008) hervorgegangen ist: Eine angereicherte Publikation ist demnach ein um Quellen, Daten, Visualisierungen, Kommentare etc. ergänztes Dokument (»ePrint«). Die Anreicherung bezieht sich nicht auf eine erweiterte Funktionalität durch die maschinenlesbare Repräsentation des Textes selbst, sondern auf eine Erweiterung der Information:

the fact that digital media make even more information available will only increase the problem. Digital texts, if we merely conceive of them as delimited containers that carry a

certain amount of information, will not help us to solve this problem either (Gradmann und Meister 2008).

Das Problem, das Gradmann und Meister hier für die Sozial- und Geisteswissenschaften bestimmen, trifft letztendlich auch für Wissenschaft im Allgemeinen zu: Semantiken sind dynamisch und instabil. Solange wissenschaftliche Publikationen als Container konzipiert und nicht mit anderen Publikationen maschinenlesbar verknüpft werden, handelt es sich bei den genannten Anreicherungen um Techniken, welche die Komplexität der wissenschaftlichen Kommunikation ausschließlich erhöhen. Den Publikationen wird zusätzliches Material hinzugefügt, statt die Komplexität in irgendeiner Weise zu reduzieren. Erst semantische Methoden ermöglichen durch klarere Anschlüsse an Begriffe (und damit Bezeichnungen und Bedeutungen), AutorInnen und Diskurse mittels URIs die leichtere Nachvollziehbarkeit von Dynamiken und Kontexten. Es handelt sich bei den Vorschlägen aus dem DRIVER-II-Projekt und in weiten Teilen auch bei jenen von Shotton also um Projekte, die andere Ziele verfolgen als semantisches Publizieren, wie es oben definiert wurde.

In diese Reihe gehört auch das im Januar 2014 auf zwei Jahre bewilligte DFG-Projekt »Future Publications in den Humanities (Fu-PusH)«, das Handlungsempfehlungen zur Produktion und Gestaltung digitaler Publikationen entwickeln soll (Fu-PusH 2014). Es ist hier vor allem deswegen erwähnenswert, weil dadurch den Geisteswissenschaften ein Bedarf attestiert wird, die gewohnten Publikationsformate auf ihre Funktionalität in einer primär digitalen Rezeptionsumgebung zu prüfen. Das Semantic Web ist jedoch bislang offenbar kein zwingend einzubeziehender Bestandteil dieser Umgebung, insbesondere nicht auf einer inhaltlichen Ebene im Hinblick auf die Vernetzung von Wissen, sondern höchstens im Hinblick auf die mittlerweile in Etablierung begriffene Notwendigkeit, zumindest die Metadaten von Publikationen als *Linked (Open) Data* im Internet verfügbar zu machen: Eine gute Gelegenheit für Bibliotheken, sich mit den Technologien des Semantic Web vertraut zu machen, was derzeit noch nicht flächendeckend als Aufgabe verstanden wird. Es ist also nicht zu erwarten, dass die hier vorgestellten Ideen in näherer Zukunft breit diskutiert werden. Jedenfalls sollte bis hierhin deutlich geworden sein, dass es sich auch für die Geistes- und Sozialwissenschaften lohnt, die Entwicklungen in den Lebenswissenschaften zu beobachten und über den Einsatz ähnlicher Technologien nachzudenken, die das wissenschaftliche Wissen wenigstens zu einem größeren Themenkomplex, über die annotierte Edition des Œuvres einer Autorin hinausgehend, integrieren, so dass es dynamisch an der wissenschaftlichen Kommunikation wächst.

### 3.4 Mit Ontologien das ontologische Weltbild überwinden?

Wenn man das Semantic Web als ein mit »Bedeutung« angereichertes Internet versteht, liegt Kritik nahe, weil es den durch seinen Titel vermittelten Anspruch nicht einlösen kann, Bedeutung zu transportieren. So hält Kim Veltman (2004) die Bezeichnung für verfehlt und schlägt »transactions web« oder »logic web« vor, da formale Logik die einzig objektive Dimension von Bedeutung sei und daher auch Eingang in die Entwicklung des Semantic Web gefunden habe. Mit der Etablierung von OWL um 2002 habe man darauf verzichtet, auch andere Dimensionen darstellbar zu machen. Diese Beschränkung wird allerdings nicht erläutert. Selbst wenn durch Logik verbundene Elemente als in sich geschlossenes System es erlauben, innerhalb des Systems gültige Schlüsse zu ziehen,<sup>37</sup> kann doch von Objektivität nicht die Rede sein: Sowohl die Struktur des Systems, sein Einsatz selbst und die Wahl und Anordnung der Elemente sind kontingent und beobachterabhängig.

Die Vorstellung von festen denotativen Bedeutungsbeziehungen, das Ignorieren der Kontextabhängigkeit von Sprache im Sinne Wittgensteinscher Sprachspiele, der Unübersetzbarkeit kultureller Konzepte [...] statischer »Ontologien« – kurz: das Ignorieren der gesamten Traditionslinie von Nietzsches Sprachkritik über die Frankfurter Schule bis zum Poststrukturalismus – lassen diesen Ansatz verkürzt und problematisch erscheinen (Dudek 2012).

Dadurch, dass man sich bei der Entwicklung des Semantic Web auf die »semantic primitives«<sup>38</sup> (Sowa 2000) beschränkt habe, würde man Entscheidendes übersehen, so Veltman weiter, dessen im Folgenden referierte Position durch Gradmann (2009) unterstützt wird:

1. Statt die mittlerweile durchgesetzte funktionale Weltsicht mit Hilfe des Semantic Web zu repräsentieren, würde man weiter der substantialistischen Weltsicht von Aristoteles anhängen: »Everything is presented as if this is the way ›it is‹ ontologically, rather than providing frameworks whereby what a thing ›is‹, what it means, and how it relates to other things, change as the framework changes.«
2. Man würde nun den Begriff der Existenz feiner differenzieren können und müssen: Scheinbar, nur dem Namen nach oder dinglich bestehend.

<sup>37</sup> Um es in den Worten von Niels-Oliver Walkowski 2011 auszudrücken: »Mittels einer Inferenz wird aus einer Klassenzugehörigkeit (durch ein Prädikat) eines konkreten Subjekts eine andere Eigenschaft dieses Subjekts – in diesem Fall transitiv — deduziert.«

<sup>38</sup> Damit sind Ausdrücke für Dispositionen von Dingen oder Konzepten gemeint, deren Verwendung auch intersubjektiv im Normalfall nicht als Erzeuger von Missverständnisse auffallen: Existenz, Koreferenz, Verknüpfung, Zusammenfügung und Verneinung.

3. Worte sind nicht eindeutig, sondern ihre Bedeutung ist veränderbar je nach Anwendungskontext (z. B. fachsprachlich). Man sollte daher unterscheiden zwischen Bedeutungen, Bezeichnungen und ihrer definierten Verbindung: Begriffen.
4. Beziehungen können auch unterordnend, determinierend oder ordinal sein: »Needed is an approach to semantics that places it in a larger context of semiotics, lexicology, lexicography, semasiology and onomasiology.«
5. Begriffe werden je nach Ort, Zeit, politischen Umständen etc. unterschiedliche definiert. »We need databases to reflect that meaning changes both temporally (whence etymology) and spatially, even within a culture (e. g. national, regional and local differences) and especially between cultures.« Aber auch die Bezeichnungen von Bedeutungen können sich wandeln.

Alle Forderungen Veltmans können hier mitgetragen werden, sind allerdings teilweise mit OWL, RDFS und SKOS bereits umsetzbar und werden mitunter auch umgesetzt; von einer weiten Verbreitung dieser Grundsätze kann man jedoch noch nicht sprechen. Um die Veränderung von Relationen zwischen Dingen oder Konzepten in spezifischen Kontexten deutlich zu machen, braucht es die Formulierung von Aussagen, z. B. in semantischen Publikationen, die keine Allgemeingültigkeit beanspruchen, sondern bestreitbar sind. Mit Ontologien und URIs allein kann das nicht gelingen. Für ein Wissenschaftsnetzwerk muss es auf begrifflicher Ebene frei durch AutorInnen (mehrfach) besetzbare Leerstellen geben. Ein großes Problem ergibt sich weiter dadurch, dass zur Zeit die Entwicklung von Ontologien nicht systematisch versioniert wird. Überarbeitete *labels* von Klassen und Relationen werden selten dokumentiert.

Die zitierten KritikerInnen des Semantic-Web-Mainstreams vernachlässigen allerdings, genauer zu differenzieren zwischen dem Begriff der Ontologie in der Philosophie, der eine Weltsicht meint, die alle anderen Unterscheidungen der Unterscheidung Sein/Nichtsein unterordnet, und der Ontologie als »formal, explicit specification of a shared conceptualisation« (Gruber 1993) als Hilfsmittel für Systeme zum automatischen Schließen. Letzterer Begriff beinhaltet nämlich nicht notwendig, wenn auch häufig so ausgeführt, die Konsistenz mit ersterem. Rein theoretisch könnte ein Computer in die Lage versetzt werden, auf Paradoxien zu schließen, nur wurde dazu bislang meines Wissens nach noch kein Experiment unternommen, vielleicht, weil alles aktuelle Denken noch immer primär von mehreren Tausend Jahren ontologischer Philosophie sozialisiert ist, was zur Folge hat, dass Paradoxien als zu vermeiden gelten. Allerdings reflektiert die Philosophie selbst die Reflektionsarmut des ontologischen Beobachtens, indem sie die Neigung der Ontologie erkannt hat, folgenreiche Fehler zu produzieren: »Die Ontologie garantiert mithin die Einheit der Welt als Einheit des Seins. Nur das Nichts wird ausgeschlossen, aber damit geht ›nichts‹ verloren« (Luhmann 1997, S. 896).

Was aber tun, wenn sich das als etwas Beobachtete als Nichts herausstellt? Nur wenn man jede Unterscheidung als grundsätzlich kontingent markiert, kann man sich vor Folgefehlern bewahren. Alle primär hierarchischen Systeme der Wissensrepräsentation tragen sich mit diesem Risiko.

Um die Fehleranfälligkeit zu reduzieren, führte man bereits in der Antike die Unterscheidung *episteme/doxa* ein, um das nicht hinreichend Erwiesene abzuspalten, jedoch wiederum mit der vorgelagerten ontologischen Unterscheidung, denn sowohl das Unzweifelhafte als auch das Vermeintliche *sind*. Die Einführung der Logik und später – mit erweiterten Methoden – das Wissenschaftssystem retten die Kommunikation aus der ontologischen Falle, indem sie (seiende) Unwahrheiten bezeichnen können. Das Problem des Entweder-Oder, das alles Dritte mit dem Ziel der Vereinfachung ausschließt, ist also lediglich invisibilisiert – bis man feststellt, dass das Unwahre doch nicht sein kann! Dabei sind Unwahrheiten höchst aufschlussreich und bewahrenswert, aber das ist in der Wissenschaft nicht zu vermitteln: Negativergebnisse fallen der impliziten Selektion zum Opfer (Brembs, Button et al. 2013), obwohl sie sogar als Wahrheit formuliert sein müssen, um überhaupt als Wissenschaft anerkannt zu werden. Möglicherweise lassen sich jede Menge kommunikativer Probleme vermeiden, wenn man das Dritte, das bereits in jeder Unterscheidung weitestgehend unbemerkt anwesend ist, zulässt: nämlich den Beobachter.

Die Idee der mehrwertigen Logik ist nicht neu, seit etwa 100 Jahren diskutiert man darüber. Gerade die Entwicklung des Computers hat dazu beigetragen, Werte zwischen wahr und falsch zuzulassen, oder den bereits von Aristoteles vorgeschlagenen Wert des (noch nicht) Bekannten einzuführen. Aber all diese Vorschläge sind nichts als Varianten der zweiwertigen Logik. Der entscheidende Punkt, auf den Gotthard Günther z. B. 1980 hingewiesen hat, ist die mit der zweiwertigen Logik unreflektiert verbundene Konstruktion des Subjekts, das fehlbar darin ist, die unzweifelhaft seiende Welt zu erkennen. Diese Sicht blendet die Möglichkeit aus, dass die Konstruktion des Subjekts überhaupt erst das Seiende erzeugt, auf das man sich dann nachträglich kommunikativ einigen kann. Daraus kann man nur schließen, dass das Seiende unabhängig von einem Beobachter nicht ist. Was nicht ist, kann werden, aber alles, was sein könnte, ist als solches nicht greifbar, weil viel zu komplex. Aber eine Beobachterin kann nichtsdestotrotz in diesen Raum, bei Luhmann: diese Welt, hinein- und *etwas* herausgreifen und es bezeichnen. Wenn dies im Rahmen von Forschung passiert, so dreht sich die wissenschaftliche Kommunikation um nichts anderes, als Unterscheidungsgebräuche wie diese zu etablieren, so dass sie der Kommunikation zur Verfügung stehen. Für wie lange, ist dabei ungewiss – und die Wissenschaft reflektiert diese Prekarität viel stärker als andere Bereiche. Eben diesen Vorgang sollte ein System, das der wissenschaftlichen Kommunikation ein tüchtiges Medium sein soll, auf jeden Fall abbilden können, um

nicht hinter die Leistungsfähigkeit von analogen Medien zurückzufallen. Dann nämlich wird es sich – zu Recht – kaum etablieren können.

Ontologien halten üblicherweise zwei Typen von Entitäten vor, was gut zeigt, worin die ontologische Vereinfachung besteht, die es in der Wissenschaft eigentlich zu vermeiden gilt: Einerseits die *terminology box* (TBox), für die man davon ausgeht, dass sie auch auf Dauer stabil bleibt und andererseits die *assertion box* (ABox), die in erster Linie temporäre Instanzen von Entitäten der TBox aufnimmt. Ein klassisches Beispiel entstammt der Politik, wo Ämter die TBox, konkrete AmtsinhaberInnen der ABox zuzuschreiben sind. Für die Wissenschaft wird es schon schwieriger, ein Beispiel zu finden. Hier werden niemals dauerhafte Fakten festgeschrieben, denn darüber bräuchte man ja nicht wissenschaftlich diskutieren. Ein faktischer Standard kann in der Wissenschaft nur über längere Zeit bzw. über sehr häufige Reproduktion geschaffen werden. Ein dynamisches Wissenschaftsnetzwerk müsste geeignet sein, solche Fakten dann erkennbar zu machen. In verwandter Form findet sich dieser Gedanke bei Stefan Heßbrüggen-Walter (2013), der zwar grundsätzliche Skepsis äußert, dass geisteswissenschaftliche »Forschende jenseits der digitalen Geisteswissenschaften davon zu überzeugen wären, dass es sinnvoll sein kann, die Formulierung von Tatsachen an den Anforderungen einer Ontologie auszurichten«. Auch scheint Heßbrüggen-Walter selbst einem ontologischen Weltbild anzuhängen, so dass für meine Zwecke »Tatsachen« mit »Beobachtungen« ausgetauscht gehört. Dennoch scheint sein Vorschlag die auch hier eingebrachte Möglichkeit zu enthalten, induktiv vorzugehen, um wissenschaftliche Diskurse mittels semantischer Technologien sich nach und nach manifestieren zu lassen, so dass sie »sich im Laufe der Zeit zu einem vollständigen begrifflichen Modell der jeweiligen Disziplin entwickeln mögen« (ebd.). Unklar bleibt, warum er eine disziplinäre Parzellierung für erforderlich hält, die doch ebenso eine (temporäre) Folge der Emergenz sein kann – oder auch nicht (mehr).

Eine Wissenschaftsrepräsentation kann allerdings nicht den Anspruch verfolgen, eine exakte Kopie der wissenschaftlichen Kommunikation zu sein. Trotz einer erheblichen Flexibilität und großer Freiheitsgrade wird ein Autor immer versuchen, an bestehende Begriffe anzuschließen, um die eigenen Aussagen möglichst stark zu vernetzen. Dabei mag manche bei genauem Betrachten zu erkennende größere Inkongruenz übersehen werden. Wie bei allem beobachtungsgeleiteten Handeln kann es nur darum gehen, Unterscheidungen zu treffen, wie Veltman schließlich selbst feststellt. Damit kann man beliebig beginnen. Beobachten ist immer mit Risiken behaftet. Die Frage ist nur, welchen Unterscheidungsgebrauch man anderen unterstellt: Wird die Inkongruenz auffallen? Nur darum geht es. Wenn also beispielsweise festgestellt wird, dass die Verwendung eines bestimmten Wortes missverständlich sein könnte, weil andere dieses Wort verwendende Aussagen ihm einen im vorliegenden Kontext nicht treffenden Sinn

verleihen, wird man die Notwendigkeit erkennen, nach einer alternativen Bezeichnung zu suchen und damit einen neuen Begriff zu formulieren, vorausgesetzt es steht eine Infrastruktur bereit, die das Treffen dieser Unterscheidung durch eine Autorin erlaubt, es zu maschinenlesbaren Daten verarbeiten und diese ins Wissenschaftsnetzwerk einspeisen kann. Es lassen sich einige leicht greifbare Anforderungen an semantische Publikationen aus den Forderungen Veltmans und den dargestellten Überlegungen ableiten:

1. Das Wissenschaftsnetzwerk besteht aus folgenden Elementen: Bezeichnungen, Bedeutungen, Begriffen, Ressourcen und Aussagen. Aus Bezeichnungen und Bedeutungen werden Begriffe gebildet; aus Begriffen Aussagen, die durch Ressourcen angereichert sein können.
2. Bedeutungen können entweder unter der Verwendung anderer Begriffe formuliert werden – dann nehmen sie die Form von Aussagen an – oder durch URIs auf Ressourcen in Umweltsystemen verweisen, z. B. in ein AutorInnenidentifikationssystem oder in ein Repositorium. Die Bedeutung abstrakter Begriffe kann nicht durch den Verweis auf Ressourcen formuliert werden.
3. Konnektoren, die für die Formulierung von Aussagen benötigt werden, sind ebenfalls Begriffe. Meist eindeutiger als Konkreta und Abstrakta, fallen sie seltener als missverständnisproduzierend auf. Ihre Definitionen werden häufig als Selbstverständlichkeit behandelt. Trotz ihrer vergleichsweise hohen Stabilität sollte in der technischen Umsetzung dieser Relationen ermöglicht werden, die Verbindung zwischen Bezeichnung und Bedeutung der jeweiligen Relation neu zu bestimmen.
4. Alle Elemente des Wissenschaftsnetzwerks können evolvieren und müssen daher in standardisierter Form versioniert werden. Die Verwendung einer bestimmten Version eines Elements zu einem bestimmten Zeitpunkt muss rekonstruierbar sein. Daher erhält jede Version eines Elements – außer externen Ressourcen – einen URI, der dynamisch aus *base URL*, Elementbezeichner, ID und Zeitstempel generiert wird, so dass er intellektuell interpretierbar ist.
5. Auch die Persistenz von externen Ressourcen muss sichergestellt werden. Das dafür verwendete System muss die Verknüpfung des jeweiligen Elements mit einer neuen Version bzw. ihre Stabilität registrieren und dem Wissenschaftsnetzwerk bei der Verwendung der Ressource melden.
6. Bei der Aktualisierung eines Elements entstehen automatisch neue Aussagen. Es ist notwendig, klar zwischen einer ersten, autorisierten Version eines wissenschaftlichen Dokuments und seiner automatischen Aktualisierung zu differenzieren,

die dann nicht mehr allein dem Autor zugeschrieben werden darf, obgleich die entsprechenden Verknüpfungen über die Versionierung erhalten bleiben.

Wie kann ein semantisches Wissenschaftsnetzwerk von den bisherigen Erfahrungen mit Ontologien zehren? Die Vielzahl der bereits erzeugten Spezialontologien kann nur unter größten Anstrengungen in einen anderen als den intendierten Anwendungskontext übertragen werden, z. B. aufgrund von Restriktionen in Form von Axiomen, vorgeschriebener Hierarchiepositionen bestimmter Klassen (*range* und *domain*). Interessanterweise kritisiert der »Erfinder« des Semantic Web die Entwicklung solcher hochspezialisierter und -komplexer Ontologien, wie sie auch für ein Wissenschaftsnetzwerk wenig hilfreich wären:

Rather than focusing on the challenges of creating large and expressive ontologies by specialized knowledge experts, the large scale social mechanisms we envision require that we must instead figure out how we can maximally break down the task of turning messy human knowledge into a shared information space that is useful to everyone. The smaller we can make the individual steps of this transformation, the easier it will be to find humans who can be incentivized to perform those steps (Hendler und Berners-Lee 2010).

Über diese grundsätzliche Problematik hinaus kann die Anwendung bestehender Ontologien für ein Wissenschaftsnetzwerk dadurch vereitelt werden, dass Definitionen von Entitäten über unumstrittenes Lehrbuchwissen – das übrigens nichtsdestotrotz ebenso versioniert werden muss – hinausgehen. Diese und die oben genannten Schwierigkeiten erfordern dann für die Wiedernutzung der Ontologie eine Anpassung, für die es nach Dröge et al. 2013 drei Möglichkeiten gibt, hier am Beispiel von Klassen erläutert:

1. Man übernimmt die Klasse in den eigenen Namensraum, verweist jedoch nicht auf das ursprüngliche Element. Hier wird der Vernetzung des Semantic Web entgegengewirkt.
2. Man übernimmt die URI aus dem externen Namensraum und erstellt Unterklassen mit `rdfs:subClassOf`, um Freiheiten in der Definition zu erlangen. Dieses Vorgehen kann jedoch nicht in jedem Fall helfen, z. B. keine problematischen Restriktionen auflösen.
3. Man übernimmt die Klasse in den eigenen Namensraum, verbindet sie aber über eine Relation wie beispielsweise `owl:equivalentClass` mit dem Ursprungselement.

Ähnlichkeit und Identität lassen sich schwer bestimmen. Gerade bei der letzten Option ist die Frage, wie genau man Äquivalenz deutet: Wenn sie vorliegt, warum war es dann nötig, eine neue URI zu vergeben? Die Dokumentation von OWL sieht durchaus andere Anwendungen, nämlich innerhalb einer Ontologie, für diese *property* vor und lässt dabei immer noch gehörigen Interpretationsspielraum. Für ein Wissenschaftsnetzwerk

scheint es angemessener, die wissenschaftliche Kommunikation über die Ähnlichkeit von Begriffen und Aussagen entscheiden zu lassen und als Grundgerüst verwendete Ontologien, die sich ja gerade dadurch auszeichnen müssen, wenigstens zum gegebenen Zeitpunkt als unumstritten zu gelten, möglichst grob zu stricken. Das wirft allerdings das Problem auf, nur eingeschränkt auf bestehende Ontologien zurückgreifen zu können. Hier ist eine genaue Bestandsaufnahme nötig, die jedoch im Rahmen dieser Studie nicht zu leisten ist. Es soll nur dargestellt werden, dass sich im Grunde drei Typen von Ontologien eignen könnten:

1. thematisch sehr umfassende Ontologien, auch *upper ontologies* genannt, die grundlegende Unterscheidungen wie jene zwischen Lebewesen und unbelebtem Gegenstand einführen,
2. Ontologien, die nur singular definierte Begriffe aus einem Themenbereich umfassen und aktuell von in der Wissenschaft anerkannten Autoritäten wie den UN als geltend festgelegt sind, z. B. GeoNames, und
3. Ontologien, die vor allem Relationen beschreiben.

Zu letzterem Typus gehört die *Citation Typing Ontology*<sup>39</sup>, die hier nur als Beispiel für Ontologien dieses Typs stehen soll. Eine umfassende Bestandsaufnahme muss anderen Studien vorbehalten werden. CiTO enthält über 40 Relationen, *properties*, um die Art einer Zitation zu charakterisieren – das ist möglicherweise mehr, als man braucht und tendiert bereits zu Unübersichtlichkeit. Die Ontologie gehört zur leitend von David Shotton entwickelten SPAR-Suite.<sup>40</sup> Es kann sich erst im Gebrauch und im Zuge einer interdisziplinären Diskussion erweisen, ob CiTO für eine breite Anwendung geeignet ist. Zwar binden mehrere Projekte CiTO bereits ein, aber laut LODStats<sup>41</sup> verwenden erst vier Datensätze und darin 16 *RDF-statements* die Ontologie. Zum Vergleich: Die bekannte Ontologie zur Beschreibung von Personen FOAF ist in 156 Datensätzen mit 29 841 822 Einzelanwendungen zu finden. Die geringe hier dokumentierte Verwendung von CiTO überrascht angesichts der breiten Wahrnehmung, die sich beispielsweise in Blogbeiträgen zeigt, welche die zeitnah zur Veröffentlichung der Ontologie eingeführte Möglichkeit begrüßen, sie auch innerhalb des *social bookmarking service CiteUlike* benutzen zu können.<sup>42</sup> Auch umfassendere Projekte wie das an anderer Stelle noch näher

<sup>39</sup> CiTO, <http://purl.org/net/cito/>, zuletzt besucht am 23. April 2014.

<sup>40</sup> Siehe Linked Open Vocabularies, [http://lov.okfn.org/dataset/lov/details/vocabularySpace\\_SPAR.html](http://lov.okfn.org/dataset/lov/details/vocabularySpace_SPAR.html), zuletzt besucht am 4. Mai 2014. Alle anderen dort enthaltenen Ontologien haben entweder die Bibliographie oder generell die Struktur eines Artikels zum Gegenstand, jedoch nicht einen Katalog von Relationen, die helfen würden, wissenschaftliche Aussagen zu formulieren.

<sup>41</sup> Siehe <http://stats.lod2.eu/vocabularies?search=cito>, zuletzt ausgeführt am 4. Mai 2014.

<sup>42</sup> Siehe z. B. Jodi Schneider, CiTO in the wild, 18.10.2010, <http://jodischneider.com/blog/2010/10/18/cito-in-the-wild/>, zuletzt besucht am 4. Mai 2014.

zu untersuchende *user centric model* für semantische Annotation in der Wissenschaft (Ribaupierre und Falquet 2013) setzt CiTO ein (siehe Kapitel 4.3).

*ScholOnto* (siehe Groza, Handschuh et al. 2009) hat einen breiteren Einsatzbereich: Mittels seiner Relationen können z. B. Kausalitäten, Ähnlichkeiten, taxonomische Verknüpfungen und auch viele Relationen, wie sie auch CiTO bereitstellt, erzeugt werden. Das Besondere ist die Polarisierung aller Relationen, d. h. man kann ausdrücken, ob die Relation positiv oder negativ und auch, wie stark sie ist.

Der zweite Typus, Ontologien für spezielle Einsatzgebiete, ist naturgemäß sehr vielfältig. Da diese Untersuchung auf die Kulturwissenschaft fokussiert, sei deshalb als Beispiel eine Ontologie ausgewählt, die sich mit dem Quellenstudium Wittgensteins beschäftigt und möglicherweise auf andere philosophische Quellen, auf die die wissenschaftliche Diskussion Bezug nimmt, und sogar für die Formulierung von gegenwärtiger Philosophie wenigstens in Teilen übertragbar ist. Es soll hier nur kurz die Modellierung der oberen Klassen beschrieben werden (vgl. Pichler und Zöllner-Weber 2013): *source* (Primär- und Sekundärquellen), *person* und *subject*. Die Primärquellen werden in immer kleinere Einheiten zerlegt, deren kleinste die *bemerkung* ist. Besonders interessant scheint hier, wie die thematische Klasse, die natürlich immer an mindestens eine *source* gebunden ist, weiter zerlegt wird (Auswahl):

- *issue* markiert strittige Punkte, die insbesondere durch die mit ihnen verbundene Sekundärliteratur aufscheinen.
- *point* für Schlussfolgerungen und Festlegungen,
- *perspective* zeigt Widersprüche auf, die sich zwischen Instanzen von *issue* und/oder *point* ergeben.
- *fields* dienen der Klassifikation von Instanzen von *issue* und/oder *point*.

Diese Ontologie erlaubt also die komplexe Verknüpfung von Primär- und Sekundärquellen sowie die Formulierung eigener Aussagen, die sich auf diese Quellen beziehen. Es ließe sich damit also leicht ein Überblick darüber gewinnen, wie eine bestimmte Aussage Wittgensteins in der Forschung diskutiert wird und welche anderen Aussagen dabei herangezogen werden etc.

Was die *top-level ontologies* angeht, reicht ihre Geschichte weit zurück bis vor die Erfindung des Internets. Während man in der Frühphase solche Ontologien für die Anwendung in Wissensdatenbanken entwickelte (siehe folgendes Kapitel), dienen sie heute vielmehr der Verknüpfung von spezialisierten Fachontologien. Hier können unmöglich alle entsprechenden Projekte untersucht oder gar verglichen werden, aber

schon eine oberflächliche Bestandsaufnahme lässt folgende Schlüsse zu: Viele Ontologien dieser Art wurden für die kommerzielle Nutzung entwickelt und sind daher auch nicht, jedenfalls nicht in vollem Umfang, frei zugänglich. Wenn man die gemessen am enormen Entwicklungsaufwand große Zahl von vergleichbaren Projekten sieht, stellt sich leicht die Frage, warum man die Energien nicht bündelte. Mögliche Gründe können in wirtschaftlicher Konkurrenz, also konkurrierenden Auftraggebern, gelegen haben oder auch in einem eher politischen Kampf um Definitionsmacht. Ein kurzer Blick auf frei nutzbare und aktuell aktive Projekte ist ernüchternd: Ansätze wie *Common Semantic MOdel* (COSMO)<sup>43</sup>, das sich ein Standard-Wörterbuch der englischen Sprache zur Grundlage gemacht hat, scheint kaum geeignet, wissenschaftlichen Zwecken zu dienen. Elaboriertere Ansätze, die auf manche Strömungen der gegenwärtigen Philosophie Rücksicht nehmen – darunter fällt allerdings nicht der Konstruktivismus – wie die *General Formal Ontology* (GFO)<sup>44</sup> erscheinen da zunächst interessanter, weil sie z. B. auch zeitliche Aspekte differenziert behandelt. Gleichzeitig macht genau das sie wieder fragwürdig, denn gerade Kategorien wie die Zeit sind philosophisch nach wie vor umstritten. Ohne die Betrachtung hier vertiefen zu können: Vor dem Hintergrund der oben formulierten Kritik am ontologischen Weltbild und an Hierarchisierungen fällt es schwer, die Festlegungen von bestehenden Ontologien anzunehmen und die dafür erforderlichen Kompromisse gegen den Nutzen abzuwägen. Ob und wenn ja, welche Ontologien für ein Wissenschaftsnetzwerk genutzt werden können und sollten, muss an anderer Stelle untersucht werden.

<sup>43</sup> Siehe <http://micra.com/COSMO/>, zuletzt besucht am 4. Mai 2014.

<sup>44</sup> Siehe <http://www.onto-med.de/ontologies/gfo/>, zuletzt besucht am 4. Mai 2014.

## 4 Die Evolution eines möglichen Wissenschaftsnetzwerks

### 4.1 *Memex*, *Xanadu* und das *WorldWideWeb*

Das Konzept des semantischen Publizierens – mit dieser oder jener Definition – wurde noch vor der Entwicklung des Internets durch Ideen vorbereitet, von denen die eine gar nicht, die andere teilweise realisiert wurde: *Memex* und *Xanadu*.

A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory (Bush 1945).

Heute würde man hierin die Beschreibung eines *smart phones* erkennen, aber 1945 dachte Vannevar Bush eher an einen Schreibtisch mit einer bildschirmartigen Oberfläche, auf die Inhalte projiziert werden können, die auf einem neuartigen Mikrofilm gespeichert sind. Das bahnbrechende Feature ist jedoch »associative indexing, the basic idea of which is a provision whereby any item may be caused at will to select immediately and automatically another« (ebd.). Die Verknüpfungen sollten in erster Linie über einen gemeinsamen Zeitstempel für jene Texte und Bilder hergestellt werden, an denen gleichzeitig oder nacheinander gearbeitet wurde. Auch die Verbreitung des so verknüpften Wissens, des *trails*, ist vorgesehen: »So he sets a reproducer in action, photographs the whole trail out, and passes it to his friend for insertion in his own memex, there to be linked into the more general trail.« Bush bedachte mit, dass nicht jede individuelle Arbeitsweise zu sinnvollen *trails* führt, weshalb eine neue Berufsgruppe mit der Aufgabe der Verknüpfung betraut werden sollte. Wie die Bände einer Enzyklopädie sollte man auch *trails* auf Mikrofilm kaufen können. Selbst die Erweiterung des menschlichen Körpers um elektronische Geräte, die damit weit über *wearables* hinaus geht, wird eingeführt: Warum elektrische Impulse mechanisch so umsetzen, dass sie mit den menschlichen Sinnen wahrnehmbar werden, wenn man die Impulse auch direkt dem Gehirn übermitteln kann?

Bushs Grundidee war jedoch, der zunehmenden Komplexität der menschlichen Zivilisation mit einer Erweiterung des je individuellen Gedächtnisses entgegenzutreten. Die einzige Alternative wäre gewesen: Komplexitätsreduktion durch Vergessen. Zwar war durch die Erwerbung von Mikrofilmen ein kommerzieller Aspekt enthalten, Rechtemanagement und Versionierung scheinen für Bush jedoch zunächst einmal vernachlässigbar gewesen zu sein.

Gerade diese zwei Aspekte, die über die Vernetzung von Wissen durch Querverweise, Hyperlinks, hinausgehen – dem Prinzip nach bekanntlich eine frühe Erfindung, die spätestens seit dem Talmud angewendet und hier folgenreich technisch speziell formatiert wird –, gehören zu den Hauptinnovationen von Ted Nelsons *Xanadu*. Jahrzehntelang –

seit 1960 – entwickelt, aber nie über den Prototyp-Status hinausgelangt, war es als kommerzielles System angelegt, in dem einzelne, versionierte Sinneinheiten, Dokumente, eine feste Adresse haben. Ähnlich wie heute in Wikis realisiert, sollten Änderungen sichtbar gemacht und eingehende Verlinkungen automatisch auf die aktuelle Version aktualisiert werden. Die folgenden aus Nelsons Projektbericht (1975) herausgegriffenen und zur besseren Verständlichkeit neu geordneten Grundsätze sollen das Konzept skizzieren:

- Everything in the system is a document and has an owner. ... The owner is the only one who can remove or change the document.
- The unified structure containing all versions (history, alternatives) may be called *docuplex*. The storage system for documents with parts, versions and links among them may be called *hyperfile*.
- A user should have the possibility to set any types of link. Links are point-to-point, point-to-span or span-to-span. A link can have multiple endpoints. Links can be directional or not, depending on the type of link.
- A link between or within documents resides in one document. ... Other cases are possible, such as links between documents residing in yet others.
- A document can quote as much as desired from another document.
- Publishing is irreversible and allows every reader to link to the document. No withdrawal is possible for a published document. But the owner can allow unrestricted distribution of a private document (privashing) and remove it later on.
- Categories are not system level but user directories (documents, too).

Das System konnte wohl aufgrund seiner Komplexität nie vollständig entwickelt werden. Insbesondere die Hauptanforderungen an die Software, also das diffizile Versions- und Rechtemanagement waren sehr ambitioniert. Obwohl es sowohl beim *Memex* als auch bei *Xanadu* darum ging, das Wissen der Menschheit vernetzt verfügbar zu machen – ein Ziel, das auch dem semantischen Publizieren für den Teilbereich der Wissenschaft zugeschrieben werden soll –, fehlte beiden der Aspekt der *openness*, der Bedingung dafür ist, dass einerseits die Barrieren für die Nutzung der angebotenen Inhalte und damit auch für die Beteiligung mit eigenen Inhalten am Wissenschaftsnetzwerk niedrig gehalten werden und andererseits die Verantwortung für die Langzeitarchivierung nicht in den Händen eines Unternehmens liegt, dessen Bestand von seinem wirtschaftlichen Erfolg abhängt. Gegenüber der in dieser Studie geltenden Definition des semantischen Publizierens fehlt den beiden Systemen außerdem die Analogie zur Rolle von Semantik für die gesellschaftliche Kommunikation. Die dort vorgesehenen Knotenpunkte im Netzwerk zeichnen sich jeweils zwar durch hohe Flexibilität aus – sie können die unterschiedlichsten, auch nicht-textuellen Formen annehmen –, erfordern aber Erschließungstechnologien, die nicht bereits durch die *trails* bzw. die *Xanadu*-Dokumente mit

den Verlinkungen angelegt, sondern zusätzlich entwickelt werden müssen, um überhaupt das bereits in den Systemen verfügbare Wissen aufspüren zu können, das man mit den eigenen Beiträgen vernetzen will.

Bereits enthalten ist – und darauf kommt es hier an – in beiden mittlerweile historischen Ideen die Auflösung des linear-zirkulären Prozesses der Wissenserzeugung, hin zu einer netzartigen Struktur. Stefan Gradmann (2011) schlägt dafür die Auflösung des monolithischen Dokuments vor, die bereits dann gegeben sei, wenn die Qualitätssicherung nicht mehr vor der Publikation, sondern danach stattfindet und neue Versionen zum Bestandteil der Publikation werden. *Xanadu* enthält die Möglichkeit, neue Dokumente durch die Neukombination von bereits bestehenden zu schaffen, durch Transklusionen, wobei der Originalkontext leicht nachvollzogen werden kann. Der Zirkel vom Verfassen über das Publizieren zum Rezipieren kann durchbrochen werden, wenn einzelne Aussagen, ja sogar einzelne Begriffsdefinitionen auch ohne das eingehende Studium von jenen im selben Dokument getroffen rezipiert und zitiert werden können. Das erfordert jedoch auch andere Techniken des Verfassens. Abgeschlossene Sinneinheiten müssen dazu durch eine Autorin definiert werden.

Tim Berners-Lees Konzept (1990), nach dem das heutige Internet, genauer: *WorldWide-Web*, entwickelt wurde, war *Memex* und *Xanadu* in mehreren Hinsichten überlegen (vgl. McCracken 2014): Berners-Lees Ziele waren nicht sehr hoch gesteckt, denn zunächst ging es um ein internes, textbasiertes Wissensmanagementsystem für das CERN. Er verwendete Konzepte, die seinen Kollegen bekannt waren: z. B. Hypertext, vorbereitet natürlich durch *Xanadu*. Von Anfang an setzte er auf Plattformunabhängigkeit. Jeder sollte mit dem System umgehen und, vor allem: es für seine individuellen Zwecke anpassen können. Er setzte also auf Universalität und Flexibilität. Auf Grundlage einer universalen Client-Server-Struktur, die es rein theoretisch jedem ans Internet angeschlossenen Computer ermöglicht, Inhalte über ein gemeinsames Protokoll auszutauschen, lässt sich ein Wissensnetzwerk Schritt für Schritt modellieren. Die zugrundeliegende Technologie musste nicht verändert werden, um Semantic-Web-Technologien, Software für das nutzerseitige Editieren von Inhalten oder Micropayment-Systeme darauf aufzusetzen.

Das Internet wurde so schnell adaptiert, weil es leicht war, die Neuerung gegenüber etablierten Verbreitungsmedien zu beschreiben: Es ist wie eine Zeitung, nur mit Links, so dass man von einem Artikel zu einem anderen springen kann. Es braucht ein breit geteiltes Verständnis von etwas Neuem, damit es sich etablieren kann. MC Schraefel (2007) meint, dass auch das Semantic Web eine analoge Entsprechung brauchen würde, die es erweitern könne. Das trifft für außerwissenschaftliche Anwendungen gewissermaßen zu, da die Voraussetzungen für Kommunikation, eben die Semantik, nur in der wissenschaftlichen Auseinandersetzung reflektiert wird. Wenn man das Semantic

Web als Ergänzung des Internets beschreibt, die selbiges maschinenlesbar machen soll, leuchtet für ein nicht-wissenschaftliches Publikum der Mehrwert für Mensch und Gesellschaft nicht ein. Stattdessen sollte man vielmehr die Nähe zur Kommunikation anrufen: Was wäre, wenn wir über das Internet alle öffentlich wahrgenommen Äußerungen in dieser Welt leicht wiederfinden würden, nur weil wir wissen, worum es dabei ungefähr ging? Was wäre, wenn es nur einen Link braucht, um alle öffentlichen Äußerungen, die sich auf dieses Ereignis beziehen, aufzurufen? Es geht nicht zuletzt darum, jeden Menschen, der keine Schulung in Recherchetechniken genossen hat, in die Lage zu versetzen, ohne Umwege zu einer gesuchten Information zu kommen und sich mit ihren verschiedenen Versionen auseinandersetzen zu können. Die analoge Entsprechung des Semantic Web ist die gesellschaftliche Kommunikation, nur, dass das Kommunikationsmedium Formen bewahrt, man so jederzeit auf sie zugreifen und auch an historische Kommunikation anschließen sowie Sozialstrukturen in Echtzeit maschinell analysieren kann.

## 4.2 Wissensdatenbanken

Neben der Entwicklungslinie vom Hypertext her sollte auch der enzyklopädische Ansatz einbezogen werden: *Cyc* ist eine universale Wissensdatenbank mit logischen Verknüpfungen, die seit 1984 entwickelt wird, seit 1994 im Rahmen des Unternehmens *Cycorp, Inc.* Im Jahr 2002 wurde ein Teil der Datenbank erstmals als *OpenCyc* zugänglich gemacht, der gegenwärtig etwa 239 000 Bezeichnungen und 2 093 000 *triples* in OWL und Verlinkungen zu anderen Wissensdatenbanken enthält,<sup>45</sup> nicht jedoch die Regeln zur logischen Deduktion. Zwar kann man zur Darstellung, Bearbeitung und Befragung der Datenbank eine Software nutzen, aber deren Quelltext ist nicht zugänglich. *OpenCyc* wurde zur Grundlage vieler Ontologien. 2006 wurde außerdem *ResearchCyc* veröffentlicht: Ein proprietäres Programm, das *OpenCyc* um viele Fachbegriffe und Fakten erweitert und zu sogenannten »Mikrotheorien« zusammenfasst: einer Klassifikation in Disziplinen und Subdisziplinen. Fehlende Dokumentationen und Qualitätsstandards erschweren eine Beurteilung dieses Projekts. Sein Anspruch, das Wissen der Welt abzubilden, macht es zu einem der Vorläufer für ein neues Medium der wissenschaftlichen Kommunikation.

In diese Linie gehört dann auch die *Wikipedia*, die selbst zwar nicht auf Semantic-Web-Technologien basiert, deren strukturierte Fakten aber seit 2007 von der *DBpedia*<sup>46</sup> mittels einer Ontologie verknüpft werden (für Details siehe Lehmann, Isele et al. 2014). Im vorliegenden Zusammenhang ist vor allem interessant, dass die *Wikipedia*

<sup>45</sup> Siehe <http://www.cyc.com/platform/opencyc>, zuletzt besucht am 5. Mai 2014.

<sup>46</sup> Siehe <http://dbpedia.org/>, zuletzt besucht am 5. Mai 2014.

von jedem editiert werden kann – allerdings insbesondere in der deutschen Version unter recht strengen Auflagen. So ist Originalforschung in der Enzyklopädie generell nicht erwünscht; darüber, was relevant ist, wird häufig debattiert. »Einträge in einem anerkannten Lexikon oder einer anerkannten Enzyklopädie beziehungsweise in einem fachspezifischen Nachschlagewerk«<sup>47</sup> gelten als Absicherung für das Erreichen der enzyklopädischen Höhe eines Gegenstands und verhindern so gleichzeitig die Einlösung der Chancen, die sich durch die Überwindung des analogen Mediums ergeben, allen voran der unbeschränkte Umfang und die Möglichkeit von dynamischer Rekombination durch Transklusionen möglichst atomarer Artikel.

Die *DBpedia* enthält nicht nur strukturierte Fakten, sondern auch Definitionen und Kurzbeschreibungen der Artikel: Abstracts, die leider sehr unterschiedlicher Qualität sind. Ein Vorteil ist die Mehrsprachigkeit und die vergleichsweise regelmäßige Aktualisierung der zugrundeliegenden *Wikipedia*-Daten, deren Geschichte jedoch in der *DBpedia* nicht per *timestamp* nachgewiesen wird. So hilft es auch nicht, ältere Versionen des *dumps* herunterladen. Die *DBpedia* scheint mit allen bekannten Wissensdatenbanken das Problem zu teilen, dass sie ihre eigene Geschichte nicht in geeigneter Form mitschreibt: Verwendet man ihre URIs, kann man nicht sicher sein, ob der zu diesem Zeitpunkt gemeinte Begriff sich nicht in einer Weise verändern wird, dass die Aussage später nicht mehr zutrifft. Auch die Artikelstruktur wirft erhebliche Probleme auf, wenn z. B. Begriffe unterschiedlicher Provenienz als ein Artikel unter einer gemeinsamen Bezeichnung zusammengebunden werden, so dass man nur auf das Bündel verweisen kann und damit die Verwendung der Bezeichnung als URI in einem *triple* sehr unscharf wird. Auch wird dadurch mitunter die Kategorie, in der der Artikel eingeordnet ist, gesprengt. Die *DBpedia* muss also als Grundlage für ein Wissenschaftsnetzwerk ausscheiden.

*Freebase*<sup>48</sup> scheint auf den ersten Blick der *DBpedia* zu ähneln, was daran liegt, dass diese 2010 von *Google* gekaufte semantische Wissensdatenbank ihre Inhalte unter anderem ebenfalls von *Wikipedia* bezieht. Die *Freebase*-Inhalte stehen vollständig unter einer freien Lizenz, aber die Anwendung, die für das *harvesting* verantwortlich ist, steht nicht zur Verfügung. Der Mehrsprachensupport ist begrenzt. Ein angemeldeter User kann Instanzen anlegen, aber weder Klassen noch Relationen kreieren; in die allgemeine Datenbank übernommen werden Beiträge nur nach der Begutachtung durch eine *Freebase*-Mitarbeiterin. Das verspricht einheitliche Qualitätsstandards, die jedoch natürlich keine wissenschaftlichen sind. *Freebase* dient *Google* in erster Linie als Basis für die semantische Suchmaschine *Knowledge Graph*, die nun in die Standard-Suche integriert

<sup>47</sup> Wikipedia:Relevanzkriterien, zuletzt am 4. Mai 2014 um 21:51 Uhr geändert, <https://de.wikipedia.org/wiki/Wikipedia:Relevanzkriterien>, besucht am 5. Mai 2014.

<sup>48</sup> Siehe [freebase.com](http://freebase.com), für eine Beschreibung des Systems siehe [https://developers.google.com/freebase/guide/basic\\_concepts](https://developers.google.com/freebase/guide/basic_concepts), zuletzt besucht am 5. Mai 2014.

wurde. Die Ontologie hat zwei Grundstrukturen: Einerseits werden die Instanzen – *topics* – hierarchischen *domains*, also einer groben Klassifikation zugeordnet. Andererseits dient ein zweites Ordnungsprinzip, die *types*, die selbst *domains* zugeordnet sind, einer Facettierung. Ein *topic* kann unbegrenzt vielen *types* angehören. Auch wenn *Freebase* gegenüber *DBpedia* Vorteile aufweist – die einfache Ontologie-Struktur und die Qualitätssicherung – ist es für ein Wissenschaftsnetzwerk solange nicht geeignet, wie *Google* die erforderliche Nachhaltigkeit nicht garantiert. Allerdings verspricht die mögliche Entwicklung von Applikationen auf der Grundlage eines solchen Wissenschaftsnetzwerks auch wirtschaftlichen Gewinn. Entsprechende vertragliche Vereinbarungen mit einer Allianz internationaler öffentlicher Wissenschaftsorganisationen zu schließen, wäre also gar keine so abwegige Idee.

### 4.3 Semantisches Mark-up

Auch von den *science publishers* selbst wurde die Idee eines Netzwerks für wissenschaftliches Wissen gefördert. Im selben Jahr, in dem auch der einmalige *Workshop on Semantic Web Applications in Scientific Discourse* (SWASD) bei der *International Semantic Web Conference* (ISWC) stattfand, 2009, beherbergten die *Elsevier Labs* erstmalig den *HypER workshop* mit dem Titel »Hypotheses, Evidence and Relationships« in Amsterdam, von dem jedoch kein Programm aufzufinden ist. Es gibt auch inhaltlich einen Zusammenhang, denn beim SWASD wurde der *HypER Approach* (Waard, Buckingham Shum et al. 2009) sozusagen konkurrierend zum Nanopublikationskonzept (Mons und Velterop 2009, siehe auch Kapitel 4.4) vorgestellt,<sup>49</sup> das selbst mit der Person Jan Velterop im kommerziellen Publikationswesen zu verorten ist.

Eine genauere Analyse der Organisationszugehörigkeiten des relativ überschaubaren Kreises an Personen, die an den Initiativen zum semantischen Publizieren beteiligt waren, würde sehr wahrscheinlich aufzeigen, dass die *community* gut vernetzt ist, dadurch aber auch aller Wahrscheinlichkeit nach recht wenig Irritationen von außen erfährt. Der eben bereits angedeutete Eindruck von Konkurrenz zwischen zwei unterschiedlichen Ansätzen verstärkt sich dadurch, dass der auf semantisches Mark-up von Fließtexten setzende Ansatz, der mit *ClaiMaker* (siehe Mancini und Shum 2006) seinen Ausgang nahm und als *HypER Approach* fortgesetzt wurde, fast durchgehend von Entwicklern des *Knowledge Media Institute* (KMI)<sup>50</sup> der *Open University* verantwortet wird. *ClaiMaker* firmiert auf der KMI-Webseite unter den »Classics« und seine Projekt-Homepage ist nicht mehr erreichbar. Die Einordnung des Projekts in die Kategorie »Narrative Hyper-

<sup>49</sup> Die Diskussion während des SWASD ist leider nicht überliefert.

<sup>50</sup> Siehe <http://kmi.open.ac.uk/>, zuletzt besucht am 6. Mai 2014.

media« stellt eine Verbindung zu *COHERE*<sup>51</sup> her, das Ideen miteinander verbinden soll, diese aber nicht in ihre Bestandteile zerlegt, um Variationen maschinenerkennbar zu machen. Man geht hier über Hypertext nur wenig hinaus.

Im Kontext dieser KMI-Projekte wurde nicht nur die bereits in Kapitel 3.4 vorgestellte *ScholOnto* entwickelt, sondern auch diverse andere Ontologie-Frameworks (für einen Vergleich siehe Groza, Handschuh et al. 2009 sowie Shum, Clark et al. 2010). Mit dem *HypER Approach* bekommt das KMI-Konzept implizit einen projektübergreifenden Namen:

The shift to author intent means shifting our conceptualization of the text towards discourse: that is, a move from viewing the text as a collection of verbs and nouns, to a view of the contextualized pragmatic language used for science (Waard, Buckingham Shum et al. 2009).

Meines Erachtens kranken jedoch alle Vorschläge dieser Entwicklungslinie, die mit weiteren Beiträgen ergänzt werden könnte, an einer starken Konzentration auf das Erkennen von Argumentationsstrukturen in Artikeln, die man als hinter der Semantik verborgen vermutet. Der Artikel selbst wird als Medium nicht in Frage gestellt, wenn das Ziel heißt:

improve access to collections of scientific papers represented as networks of collection of claims that have a defined epistemic value, with links to experimental evidence and argumentative relationships to other statements and evidence (ebd.).

*COHERE* scheint als bestehende Anwendung des *HypER Approach* die Möglichkeiten, die durch die entwickelten Ontologien bereitgestellt werden, nicht voll auszunutzen und den Anspruch nicht einzulösen. Insbesondere erscheint *COHERE* vielmehr als unverbindliche Diskussionsplattform, denn als wissenschaftliches Kommunikationsmedium.<sup>52</sup>

David Shotton (2012) sieht den HyPER-Folgeworkshop 2010 als Vorläufer der *Beyond the PDF Conference*<sup>53</sup>, die im Januar 2011 von Philip Bourne an der University of California, San Diego, organisiert wurde. Man kann behaupten, dass sie noch heute den Stand der gegenwärtigen Diskussion spiegelt. Im selben Jahr folgte ein Treffen desselben Personenkreises, zu dem neben den bereits genannten Philip Bourne und Anita de Waard auch Tim Clark (Harvard University), Robert Dale (Macquarie University), Ivan Herman (W3C), Eduard Hovy (University of Southern California) und David Shotton

<sup>51</sup> Siehe <http://cohere.open.ac.uk/>, zuletzt besucht am 6. Mai 2014.

<sup>52</sup> Für ein besseres Verständnis der Gründe dafür, dass diese Entwicklungslinie heute verwaist wirkt, scheint das Führen von Interviews mit den ProtagonistInnen erforderlich.

<sup>53</sup> Die Veranstaltung ist u. a. mit Aufnahmen der Vorträge dokumentiert, siehe *Beyond the PDF, Workshop: January 19-21, 2011 University of California San Diego*, <https://sites.google.com/site/beyondthepdf/>, zuletzt besucht am 25. April 2014.

(University of Oxford) gehörten, unter dem Titel »The Future of Research Communication«, diesmal in Deutschland als *Dagstuhl Perspectives Workshop*.<sup>54</sup> Hier gründete sich die *Force11 Community*, die sich zum Ziel setzte, die (digitale) wissenschaftliche Kommunikation zu verbessern.<sup>55</sup> Es folgte die Publikation eines *White Papers* (Bourne, Clark et al. 2011). Darin wird die Vorstellung erläutert, dass wissenschaftliche Kommunikation darin bestehen könnte, wörtlich am Netz wissenschaftlichen Wissens mitzuweben: »Adding new elements of scholarly knowledge is achieved by adding nodes and relationships to this network« (ebd.). So würden »reusable scholarly artifacts« (ebd.) entstehen. Die bestehenden sozialen Praktiken seien um die technischen Möglichkeiten, die einst nicht mehr als ein monolithisches Dokument zuließen, herum entstanden. Es reiche nicht, die Technologien auszutauschen; das Problem ist eines des »social ecosystem of communication« (ebd.). Dieser knappen Beschreibung des Problems und dem programmatischen *statement* wurde auch mit *Beyond the PDF 2* im Jahre 2013<sup>56</sup> keine tiefgehendere Analyse hinzugefügt, die semantisches Publizieren als Alternative im Blick hat.

Ebenfalls 2011 findet während der *Extended Semantic Web Conference* (ESWC) erstmals der Satellitenworkshop *SePublica* statt, an dessen Organisation neben Anita de Waard auch andere bekannte zwischen kommerziellem *publishing* und Softwareentwicklung im akademischen Umfeld stehende Personen maßgeblich beteiligt sind. Inhaltlich ist der starke Fokus auf Anwendungen in den Lebenswissenschaften auffällig. Auch hier wird das Nanopublikationskonzept wieder vorgestellt. Für 2012 verrät die Angabe in den Proceedings, dass acht von neun Einreichungen angenommen wurden, keine übermäßige Dynamik des Themas. Statt Grundsatzdebatten zu führen, geht es hauptsächlich um sehr spezielle Workflows. 2013 hinterlässt der *SePublica*-Workshop kaum Spuren, insbesondere keine eigenen Proceedings, die nur mehr anhand der Themen vague aus den ESWC-Satellite-Proceedings zu filtern sind.<sup>57</sup>

Ein naheliegender Gedanke ist, statt ein völlig neues Publikationsformat zu entwickeln, das altbewährte zu semantischem Publizieren zu befähigen: Solange Bedarf besteht, Argumentationslinien natürlichsprachlich darzustellen und zu rezipieren, sollte dieser auch bedient werden. Das PDF allein ist zweifelsohne ungeeignet, ein Wissensnetzwerk aufzubauen, aber über in das PDF einbettbare XML-Metadaten (XMP) können PDFs an

<sup>54</sup> Siehe <https://sites.google.com/site/futureofresearchcommunications/> sowie <http://drops.dagstuhl.de/opus/volltexte/2011/3315/>, beide zuletzt besucht am 25. April 2013..

<sup>55</sup> Siehe Force11, <http://www.force11.org/>, zuletzt besucht am 25. April 2013.

<sup>56</sup> Beyond the PDF2: Outcomes, Force11, <https://www.force11.org/outcomes>, zuletzt besucht am 6. Mai 2013.

<sup>57</sup> Für die Proceedings siehe 2011 <http://ceur-ws.org/Vol-721/>, 2012 <http://sepublica.mywikipaper.org/sepublica2012.pdf> und 2013 <http://link.springer.com/book/10.1007/978-3-642-41242-4>, alle zuletzt besucht am 6. Mai 2014.

das Semantic Web angeschlossen werden. Auch umgekehrt kann auf PDFs, solange sie im Internet über eine URI identifiziert werden, im Wissenschaftsnetzwerk verwiesen werden. PDF und XML/RDF ergänzen sich, wenn man die Anforderungen von Menschen und Maschinen an Publikationsformate gleichermaßen berücksichtigen möchte:

publishers have for many years used XML to store the underlying article, and HTML and PDF as vehicles for its dissemination. All that has been missing is an explicit recognition of the relationships between these, and a technology to link them all together (Pettifer, McDermott et al. 2011).

Pettifer et al. (ebd.) schlagen mit *Utopia Documents* einen neuartigen PDF-Reader für die Darstellung und die nutzerInnenseitige Erweiterung von Verknüpfungen, die über Annotationen auf Datenbanken verweisen, vor. Unter anderen bewirbt die *Royal Society of Chemistry* die Nutzung von *Utopia Documents* zur Anzeige und zur eigenen Erstellung der mit den Artikel-PDFs verknüpften Verweise auf ihre eigene frei zugängliche Datenbank *ChemSpider*.<sup>58</sup>

Obwohl *Utopia Documents* die selbstgestellten Anforderungen gut bewältigt und in der Literatur keine Stimmen zu vernehmen sind, die diese Software oder ihre Anforderungen kritisieren, ist auch nur wenig positives Feedback zu vernehmen und keine größere Verbreitung, insbesondere über die Lebenswissenschaften hinaus, nachzuweisen.<sup>59</sup> Zur Zeit beschränkt sich die Software auch ausdrücklich auf diesen Fachbereich, wenn auch die Entwickler Datenbankbetreiber anderer Fachbereiche dazu aufrufen, mit ihnen zu kollaborieren.<sup>60</sup> Durch das Generieren von Kontextinformationen anhand der Nutzung der Daten von Crossref, Altmetrics und sozialen Netzwerken ist *Utopia Documents* jedoch fachunabhängig den sonst verbreiteten PDF-Readern überlegen. Die Software ist nicht Open Source, wenn auch frei nutzbar.<sup>61</sup> Ebenso ist die Technologie nicht öffentlich dokumentiert. Es lässt sich vermuten, dass es durch *Utopia Documents* gelingt, die PDF-Versionen von Artikeln der kooperierenden *publisher* mit XML-Versionen derselben Artikel zu verknüpfen, die wiederum durch die NutzerInnen über eine API annotiert werden können. Deshalb können für eigene Annotationen auch nur Datenbanken ver-

<sup>58</sup> Richard Kidd, *Utopia Documents highlights RSC Publishing's semantic chemistry*, 21. 2011, <http://blogs.rsc.org/technical/2011/06/21/utopia-documents-highlights-rsc-publishings-semantic-chemistry/>, zuletzt besucht am 25. April 2014.

<sup>59</sup> Als einzigen zugänglichen Hinweis auf die Verbreitung muss hier die Anzahl der Twitter-Follower hinreichen: 413, siehe <https://twitter.com/utopiadocs>, zuletzt besucht am 25. April. Auch sind aktuellere Berichte oder Empfehlungen nicht zu finden.

<sup>60</sup> Help us extend Utopia, <http://utopiadocs.com/contact.php>, zuletzt besucht am 25. April.

<sup>61</sup> In den Twitter-Diskussionen, siehe Fn. 59, wird dieser Punkt mehrfach kritisiert und von den Entwicklern damit begründet, dass die Entwicklung wirtschaftlich sein muss. In einem Twitter-Beitrag wird auch angesprochen, dass z. B. Pharma-Unternehmen eine spezielle Version der Software lizenzieren, die es ermöglicht, Annotationen nur intern sichtbar zu machen.

wendet werden, die bereits in *Utopia Documents* integriert wurden. Es handelt sich also um ein geschlossenes System, das zwar für LebenswissenschaftlerInnen das monolithische Dokument ein Stück weit mit Gewinn aufzulösen in der Lage ist, aber für ein Wissenschaftsnetzwerk, wie es in dieser Studie konzipiert wurde, wenig Perspektiven bietet. AutorInnen, die ihre semantischen Publikationen mit jenen der teilnehmenden Verlage auf gleichem Niveau verknüpfen möchten, sind darauf angewiesen, in den Zeitschriften dieser Verlage zu publizieren, was die Anschlussmöglichkeiten – statt aus wissenschaftlichen – aus wirtschaftlichen Gründen reduziert und damit der Entwicklung einer Wissenschaftskommunikation entgegenläuft, die zum Ziel haben sollte, eigene Kriterien für Komplexitätsreduktion anzuwenden.

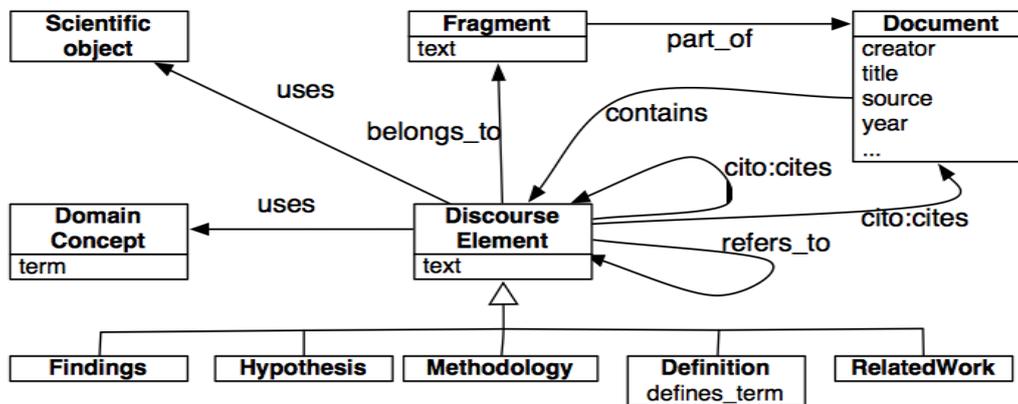
Schlagwort- und volltextindexierungsbasierte Information-retrieval-Systeme können nicht alle Informationsbedarfe abdecken (Ribaupierre und Falquet 2013), nämlich

- wenn die gesuchte Information im Kontext eines umfassenderen Dokuments steht, für das diese Information nicht zentral ist. Information-retrieval-Systeme »represent each document as a whole and do not take their subparts into account«.
- wenn der Informationsbedarf nur mit Hilfe einer logischen Verknüpfung mehrerer Begriffe ausgedrückt werden kann. Information-retrieval-Systeme »fail to represent knowledge about the semantic or rhetorical role of document elements«.

In den Naturwissenschaften seien diese Probleme bereits erkannt worden, allerdings haben die daraufhin entwickelten Lösungen eine von den Geistes- und Sozialwissenschaften zu unterscheidende Ausgangslage, denn dort herrsche »relatively little variation in describing the results, hypothesis, conclusions, etc.« (ebd.). Meist gliedern sich Aufsätze entsprechend, was in den Geistes- und Sozialwissenschaften nur selten der Fall ist. Darüber hinaus würden die entsprechenden Projekte immer nur Teillösungen vorschlagen, z. B., indem sie die Art einer Referenz genauer bestimmen oder eine (automatische) Annotation von gewissen Argumentationsstrukturen erlauben.

Die AutorInnen stützen ihren Vorschlag auf empirische NutzerInnenstudien in den *Gender Studies*. Das daraufhin entwickelte Modell, siehe Abbildung 1, basiert auf Diskurselementen, die theoretisch an jeder Stelle im Text auftauchen können: Ergebnisse, Definitionen, Methoden, Hypothesen und die Beschreibung des Forschungsstands. Das Mark-up eines Textfragments mit einem entsprechenden Diskurselement ist nicht exklusiv und einzelne Diskurselemente sind mittels CiTO verlinkbar. Außerdem werden Fachbegriffe in der Publikation identifiziert (*domain concepts*), die es erlauben, die Publikation inhaltlich einzuordnen.

Für die *Gender Studies* existierte allerdings noch keine umfassende Ontologie, weshalb die AutorInnen eine mit mehr als 600 Klassen entwickelten. Dadurch sind komplexe



**Abbildung 1.** Modellierung für das Mark-up der Argumentationsstruktur von wissenschaftlichen Texten aus Ribaupierre und Falquet 2013

Abfragen möglich, wofür einige überzeugende Beispiele genannt werden. Da schließlich nicht davon ausgegangen werden kann, dass Forschende generell in der Lage sind, SPARQL-Abfragen zu formulieren, fehlte ein »adaptive-faceted search interface«, über das nicht nur das Mark-up, sondern auch Recherchen möglich sind. Das fragmentarische Mock-up überzeugt zwar intuitiv, wird aber nicht näher beschrieben. Die Darstellung ist jedenfalls die einzige in der Literatur auffindbare, die sich einerseits den Geistes- und Sozialwissenschaften zuwendet und andererseits von einer tieferen Analyse des Problems her eine Lösung vorschlägt, die nicht nur eine Ontologie enthält, sondern auch einen konkreten Vorschlag, wie die Forschenden diese nutzen können.

Schließlich gibt es Initiativen für automatisiertes Mark-up, deren bekannteste *OpenCalais*<sup>62</sup> des *publishers Thomson Reuters* ist. Als Webservice ist es kostenlos, erfordert jedoch die manuelle oder programmgesteuerte Abfrage einer API. Ohne entsprechende EntwicklerInnen-Kenntnisse ist man auf ein ebenfalls kostenfrei angebotenes Tool angewiesen, das es nur für *Windows* gibt. Alternativ kann man auch *Plug-ins* für *Drupal* oder *WordPress* nutzen, die jedoch nicht an die aktuellen Programmversionen angepasst sind. Eine weitere Hürde ist insbesondere für KulturwissenschaftlerInnen, dass derzeit nur Englisch als Eingabesprache verarbeitet werden kann. Auch gibt es keine Dokumentation darüber, welche Datenquellen abgefragt werden. Laut Shotton, Portwin et al. (2009) erweist sich *OpenCalais* als hilfreich insbesondere für das Mark-up von geographischen Namen, Institutionen und Personen, was einige kurze Tests nicht bestätigen konnten.

<sup>62</sup> OpenCalais, <http://www.opencalais.com>, zuletzt besucht am 6. Mai 2014.

Diese Bestandsaufnahme verweist auf ein weitestgehend offenes Feld, auf dem bislang kaum mehr als Versuche stattfanden:

there is no overall coordination of the various initiatives from diverse publishers and institutions, many of which strike out in different directions in their attempts to transform scholarly publishing (Rinaldi 2010).

#### 4.4 Nanopublikationen

Wie der *HyPER Approach* wurde das Nanopublikationskonzept erstmals 2009 beim SWASD vorgestellt. Es steht in erster Linie in der Tradition von Mikroattributionen für Datenbankeinträge (zuerst Nature Genetics Editors 2007). Mons und Velterop (2009) beschreiben das Konzept als

approach to data interoperability across language barriers, jargon, database formats, and eventually, ambiguity and redundancy [als Perspektive; derzeit noch nicht in vollem Umfang möglich]. The basic principle is: natural guidance of human authors to structure their data in such a way that computers understand them.

Kern ist jeweils der Begriff als kleinste Einheit des Denkens. Mons und Velterop sind allerdings entgegen der vorstehenden Analysen überzeugt davon, dass Begriffe vollständig disambiguiert werden können und intersubjektiv übertragbar sind. Wie man mit identischen Bezeichnern bzw. Bedeutungsformulierungen in so disambiguierten Begriffen umgeht, ist unklar. In den Naturwissenschaften fällt es eventuell auch nicht leicht, ein Beispiel zu finden, das solche Fälle aufzeigt, aber um diesen Zustand zu erreichen, wurden in den Naturwissenschaften enorme Anstrengungen unternommen, um eben nicht die Kommunikation mit ständig mitlaufenden Begriffsdefinitionen belasten zu müssen.

Mittlerweile ist ein Entwurf für Nanopublikationsrichtlinien erschienen (Concept Web Alliance 2013), der beschreibt, wie das Format beschaffen sein soll (siehe dazu auch Groth, Gibson et al. 2010):<sup>63</sup> Minimal muss es eine Behauptung (*assertion*), zusammengesetzt aus Begriffen, enthalten, einen Herkunftsnachweis (*provenance*) dieser Behauptung, also z. B. den Link zur ausgewerteten Quelle oder zum Datensatz, und einen Herkunftsnachweis der Nanopublikation selbst: Autorschaft, URI der Nanopublikation und *timestamp*. Diese drei Komponenten werden jeweils als *triple*-Bündel unterschieden und sind verlinkt. Empfohlen wird, die Turtle-Erweiterung *TriG* zum Verfassen von

<sup>63</sup> Für diese Elemente wurde eine Ontologie entwickelt: *nschema*, zuletzt überarbeitet am 23. Oktober 2013, <http://www.nanopub.org/nschema>. Angesichts der regen Diskussion auf der Mailingliste, <https://trac.nbic.nl/mailman/listinfo/nanopub-users>, beide zuletzt besucht am 7. Mai 2014, ist davon auszugehen, dass diese Ontologie nicht stabil ist. Auch sind die Vorschläge aus Mons und Velterop 2009 darin nur teilweise umgesetzt.

Nanopublikationen zu verwenden.<sup>64</sup> Durch die Verlinkung in sich selbst entstehen aus den *triples* dann *quads*, also *named graphs*, einer Erweiterung des *triple* um eine Stelle.

Die *provenance* begründet gewissermaßen die Hauptfunktion der Nanopublikation, denn ihre Erfinder begreifen sie in erster Linie als Metadaten zu Forschungsdatensätzen, die jenen dann die für eine wissenschaftliche Publikation erforderlichen Referenzen hinzufügt. Für die Kulturwissenschaften mag dies zunächst befremdlich sein. Aber zur *provenance* gehören auch Statusinformationen (Mons und Velterop 2009): »community, authority, peer-reviewed, curated, disputed, retracted, hypothetical, observational, repetitive, et cetera«. Noch haben diese Status keinen Eingang in die Richtlinien gefunden, könnten aber ein Weg sein, den grundsätzlich anderen Charakter von kultur gegenüber lebenswissenschaftlichen Aussagen zu markieren. Im übrigen dürften auch einige Naturwissenschaften sich hier nicht angesprochen fühlen, z. B. die Mathematik. Unter den zitierten Vorschlägen für Status ist vor allem der hypothetische interessant: Hypothesen können Aussagen sein, die nicht durch Experimente oder dergleichen erwiesen wurden, sondern sich aus erwiesenen Aussagen (aus Nanopublikationen) mittels Algorithmen erzeugen lassen. Es lassen sich vielfältige Anwendungen ersinnen, die durch die Markierung des Status erreichbar wären:

Nanopublication of new triples in all three categories [clustered, observational, hypothetical] should lead to real time alerts to scientists who have indicated that they are interested in one of the concepts in the statement or in closely related areas of this »concept web«. With appropriate recognition and traceability of the statements this could enable an entirely different way of scholarly communication, much more adapted to the current rate of data production (ebd.).

Auch solle in semantischen Editoren automatisiert im Hintergrund eine Suche nach Hypothesen laufen, die mit den verwendeten Begriffen gebildet werden können. Dadurch, so soll hier ergänzt werden, dass die Bezugnahme auf bestimmte Grundlagen in Nanopublikationen deutlich als solche markiert sind, kann die Maschine einerseits das neue Wissen leicht filtern und andererseits bei der Widerlegung der als Grundlage verwendeten Aussagen leicht eruieren, welche der daraus abgeleiteten Aussagen nun in Frage zu stellen sind.

Die Umsetzung wollen die Autoren mittels eines *Concept Wiki*<sup>65</sup> erreichen, in dem jeder Begriff eine eindeutig identifizierbare Seite mit Verweisen auf Folgendes erhält: Ontologie(n), denen das Konzept entstammt, Synonyme, auch in anderen Sprachen, die selbst wieder Konzepte sind, der Fachbereich sowie Informationen zur Entwicklung

<sup>64</sup> TriG, <http://www.w3.org/TR/trig/>, zuletzt besucht am 7. Mai 2014.

<sup>65</sup> Concept Wiki, <http://www.conceptwiki.org/>, zuletzt besucht am 7. Mai 2014. Es handelt sich um eine Erweiterung der Software *MediaWiki*, entwickelt im Rahmen des *OpenPHACTS project (European Innovative Medicines Initiative)*. Policy: »Anything you can write a Wikipedia article about is a concept«. Wie fragwürdig das ist, wurde in Kapitel 4.2 erläutert.

des Konzepts, insbesondere terminologische und strukturelle, die das Konzept in eine Typologie einordnen. Mit einem solchen Wiki ist ein Problem gelöst, mit dem sich Forschende unmittelbar konfrontiert sehen, wenn sie Nanopublikationen einfach mittels eines Texteditors erstellen wollen: Woher die URIs für die Fachbegriffe nehmen? Rein theoretisch ist dafür jede Webseite verwendbar, aber als Einzelperson ist die Garantie der Persistenz wohl kaum zu leisten. Darüber hinaus bieten sich (fachspezifische) Datenrepositorien an, wo man theoretisch nicht nur RDF-Statements zur Begriffsdefinition, sondern auch die Nanopublikation selbst ablegen könnte. Das wird jedoch normalerweise daran scheitern, dass sie ihren eigenen Identifikator enthalten muss, was wiederum voraussetzt, diesen vor der Publikation zu kennen. Repositorien bieten diese Möglichkeit zunächst einmal nicht. Denkbar ist eine automatische Ergänzung des entsprechenden Statements. Die andere Möglichkeit erfordert eine, naheliegenderweise über einen *publisher* vermittelte Zuweisung z. B. eines DOI. Allerdings können sich nur Organisationen, also *publisher* um einen DOI-Namensraum bewerben.

Unter den gegebenen Voraussetzungen ist es also nicht möglich, in den Kulturwissenschaften ohne Weiteres mit dem Nanopublizieren zu beginnen. Solange es keine bestenfalls wissenschaftsweiten oder zumindest alle Disziplinen abdeckenden Anwendungen ähnlich dem *Concept Wiki* gibt, die gleichzeitig als Recherche- und Publikationsplattform dienen, scheint das Verfassen von Nanopublikationen wenig attraktiv, da nicht sichergestellt ist, dass sie als wissenschaftliche Publikation wahrgenommen werden, obwohl sie grundsätzlich der in Kapitel 2.1 formulierten Definition entsprechen. Wie jedoch dort festgestellt wurde, ist freie Zugänglichkeit im Internet – in einem technischen Sinne – nicht dasselbe wie die weltweite Zugänglichkeit für die wissenschaftliche Kommunikation, wie sie in der Definition als Kriterium gefordert wird: Ob die Publikation wahrgenommen wird, erweist sich immer erst nachträglich, aber gewisse Bedingungen, die ein Kommunikationsmedium bereitstellt, machen dies wahrscheinlicher.

Dafür, aber auch, um im großem Stil das in Publikationen manifestierte Wissen Maschinen zur Analyse zugänglich machen zu können, braucht es Standards. Mons und Velterop (2009) sehen das Problem darin, dass, wenn eine *Community* diesen Bedarf erkennt, viele Gruppen beginnen, konkurrierende Standards zu entwickeln. Selbst wenn sich dann einer durchgesetzt hat, wird er dennoch von vielen ignoriert.

We strongly believe in the process of »bottom up standard emergence«, a process by which useful and intuitive standards emerge from joint community action. Therefore we have proposed [...] to develop systems that [...] aim] to accommodate all standards developed so far. This means that we [...] can subsequently accept all non-ambiguous identifiers denoting these concepts as long as they are properly mapped to a universal reference in a public environment, which is »owned« and governed by the user community.

Grundsätzlich ist dieser Ansatz vielversprechend, aber für die Autoren führt er dazu, bereits in der Etablierungsphase eine geeignete Ontologie für Begriffe auswählen zu wollen, was ja dann doch wieder eine Top-down-Aktion wäre. Andererseits scheint dies aber erforderlich, wenn man ein GUI anbieten möchte, um *triples* einfach erstellen zu können. Als *Concept Web Alliance* entwickeln die Autoren mit ihren MitstreiterInnen für die Lebenswissenschaften eine Ontologie auf der Basis von Bestehendem, die voraussichtlich 200 Mio. Terme enthalten wird, die selbst nicht in Nanopublikationen verwendet werden, sondern mit den rund 50 Mio. Begriffen verknüpft sind. Diese Zahl potenziert sich natürlich durch die Einbeziehung anderer Sprachen.

Übertragen auf die Geistes- und Sozialwissenschaften würde das aber aufgrund der vielen teilweise nur winzige Teilbereiche betreffenden bestehenden Ontologien eine enorme Starthürde bedeuten. Theoretisch müsste man alle Ontologien harmonisieren, integrieren und gleichzeitig für Übersichtlichkeit sorgen, denn immerhin sind z. B. Luhmanns Begriffe für weite Teile der Sozial- und Geisteswissenschaften relevant; Aristoteles' Begriffe im Grunde für die gesamte Wissenschaft. Daran anknüpfende Aussagen sollten immer zusammengebracht werden, sonst funktioniert das Netzwerk nicht. Außerdem – bedeutender – läuft dieser Ansatz der wichtigen Voraussetzung des ständig evoluiierenden Wissens entgegen: Selbst wenn man eine Ontologie vorlegen könnte, die in diesem Moment allumfassend erscheint, ist sie es im nächsten Moment nicht mehr. In der Nanopublikation verwendete Begriffe können im Laufe der Zeit eine veränderte Bedeutung annehmen, dadurch, dass sich der Inhalt, der hinter den URIs steht, sich verändert hat. Natürlich muss durch einen *timestamp* immer nachvollziehbar bleiben, welcher Inhalt sich zum Zeitpunkt der Nanopublikation hinter den URIs verborgen hat. Es kann jedoch sein, dass durch die Veränderung dieser Inhalte sich auch die Bedeutung des *triple* in erkenntnisreicher Weise verändert.

Stellt man nun die Nanopublikation und den zuvor untersuchten Ansatz des semantischen Mark-ups gegenüber, kommt man zu dem Schluss, dass beide Ansätze sich zunächst einmal nicht ausschließen. In der Literatur gibt es auch Beispiele für die Verknüpfung von beiden (Clare, Croset et al. 2011). Die Vorteile der Nanopublikation liegen vor allem in der Flexibilität für die Anwendungsentwicklung und in der leichten Archivierbarkeit. Nanopublikationen enthalten nichts als klar und einheitlich strukturierte RDF-Statements. Die zentralen, in der Nanopublikation getroffenen Aussagen werden jedoch wahrscheinlich insbesondere in den Kulturwissenschaften bessere Aussichten auf Rezeption haben, wenn sie sich im Kontext einer Gesamtargumentation verorten lassen. Dadurch aber, dass Texte hier keine so klaren, erwartbaren Strukturen aufweisen wie naturwissenschaftliche Artikel, müsste man dafür ein manuelles Mark-up des zugrundeliegenden Textes vornehmen, bestenfalls nach dem Muster der Nanopublikation: Sicher kein leichtes Unterfangen. Da Nanopublikationen der Maschine klar vermitteln

können, worum es im Forschungsergebnis geht, ist ein zusätzliches Mark-up des narrativen Textes zunächst einmal nicht erforderlich. Das wissenschaftliche Dokument – und damit die Publikation – könnte aber auch beides enthalten.

Möglicherweise sperrt sich das Semantic Web gegen eine analoge Analogie und damit gegen eine breite Adaption, weil es momentan noch enormen intellektuellen Aufwandes bedarf, ihm etwas hinzuzufügen. Doch müssen alle NutznießerInnen des Semantic Webs selbst zu seiner Erzeugung beitragen? Nein, es ist noch nicht einmal erforderlich, dass sie verstehen, wie sie das Internet mit Informationen versorgt. Wer weiß schon, wie die Massenmedien funktionieren? Für die Lebenswelt genügt es, zu wissen, dass es so etwas gibt. Aber wenigstens die Forschenden sollten beitragen, weil

1. sie es (lernen) können,
2. die Aussicht besteht, dadurch ihre Forschungsergebnisse präzisieren zu können, und
3. sie zeitnah zur Veröffentlichung ihrer *triples* und *quads* automatisch erzeugte Verknüpfungen untersuchen und zur Fortsetzung ihrer Forschung verwenden können.

## 5 Fazit: Ad-hoc-Wissenschaftsnetzwerk statt Ex-ante-Ontologien

Das wahrscheinlich konsensfähige Ziel der Wissensrepräsentation ist, per intuitiv formulierter Anfrage an ein System, das ggf. Rückfragen stellt, eine möglichst große Menge relevanter Informationen aus allen im Internet nachgewiesenen Quellen zu erhalten. Kontrollierte Vokabulare können – vorausgesetzt sie werden *einheitlich* angewendet – dieses Auffinden erleichtern. Dabei verdecken sie jedoch das Neue, das für die Wissenschaft von größter Bedeutung ist und setzen die Duldung der Festsetzung des eigentlich kontingenten Unterscheidungsgebrauchs voraus, noch dazu unautorisiert stellvertretend gegenüber dem Verfasser des so beschriebenen Dokuments. Solange man mangels entsprechender Technologien nicht direkt im Zuge des Publizierens in maschinenlesbarer Weise an frühere Publikationen anschließen konnte, waren kontrollierte Vokabulare unverzichtbar. Man musste die fehlende Dynamik in Kauf nehmen und durch Aktualisierungen der Vokabulare und der Erzeugung von Konkordanzen kompensieren, die jedoch das Problem aufwarfen, mit früheren Versionen konsistent sein zu müssen. Darüber hinaus weisen z. B. alle historisch gewachsenen Klassifikationssysteme einen lokalen *bias* auf, was beim Versuch auffällt, im deutschsprachigen Raum rechts- oder literaturwissenschaftliche Publikationen mit der *Dewey Decimal Classification* zu klassifizieren. Möchte man tatsächlich zu einem »Weltkatalog« kommen, steht man vor dem nahezu unbewältigbaren Problem, sich zunächst auf eine global verwendbare Klassifikation einigen zu müssen und dann Mappings anzufertigen. In der Praxis ist dann allerdings nicht die globale Gültigkeit, sondern die Verbreitung einer Klassifikation das entscheidende Argument dafür, ein bestimmtes Mapping einzusetzen.

Nun kann man dank der Semantic-Web-Technologien ein dynamisches Wissenschaftsnetzwerk aufbauen, das wissenschaftliche Aussagen in vielfältige Beziehungen zueinander setzt. Dafür ist die wiederholte Verwendung von Entitäten und ein erweiterbares Vokabular für die Bezeichnung von Beziehungen dieser Entitäten erforderlich, nicht aber ein kontrolliertes Vokabular für Entitäten, denen dann lediglich Instanzen hinzugefügt werden dürften. Für Anwendungen in Bildung, Wirtschaft und Politik mag ein solches Vokabular zur Komplexitätsreduktion nützlich sein, nicht jedoch in der Wissenschaft, die auf Autorschaft zur Erzeugung von Reputation und auf Innovation geradezu beruht. Die Autorin webt am Netz mit, das die Kommunikation zusammenhält und muss daher frei sein in ihrem Unterscheidungsgebrauch und Autoritäten selbst auswählen. Im Folgenden wird daher auch nicht mehr von Entitäten, sondern von Elementen gemäß den auf S. 40 definierten Anforderungen die Rede sein.

Nachdem Wissensrepräsentation bis zur jungen Erfindung von Folksonomien immer *top-down* durch Autoritäten konzipiert werden musste, um eine kritische Menge an zu repräsentierendem Wissen beschreiben zu können, und dieses damit erschloss, haben

wir mit dem Semantic Web die Chance, dafür nicht mehr auf das Streben nach einer großen Übersicht über alles Wissen angewiesen zu sein. Vielmehr: Nun wird offenbar, dass Wissen nicht mehr im eigentlichen Sinne repräsentiert werden muss, um sich mit anderem Wissen in einen Kontext zu stellen. Gewohnte Unterscheidungsmuster für Daten und Metadaten verschieben sich (siehe Kapitel 2.2). In Nanopublikationen sind alle Daten als Metadaten verfügbar, also maschinell lesbar, da die Beschreibung eines jeden dort verwendeten textuellen Elements des Dokuments für den algorithmischen Gebrauch vorliegt. Die Unterscheidung von Daten und Metadaten wird hinfällig. Paratexte jedoch bleiben erhalten, in erster Linie in Form von Referenzen. Woran die eigenen Forschungsergebnisse anschließen, kann nach wie vor allein intellektuell bestimmt werden. Aus



wird



Der Schwerpunkt verschiebt sich damit so, dass eine Autorin in die Lage versetzt wird, größere Anteile des für die Rezeption so entscheidenden Paratextes selbst verfassen zu können, statt auf die heute fast immer durch Dienstleister für Fachinformation, selten durch FachwissenschaftlerInnen erstellten Metadaten angewiesen zu sein, deren nachträgliche Korrektur häufig allein schon deswegen vereitelt wird, weil diese Dienstleister ihre Daten vor dem Zugriff durch fremde Systeme schützen.

Eben weil die Metadatenvergabe durch Nicht-Fachwissenschaftler so fehleranfällig ist, braucht man heute kontrollierte Vokabulare. Daran wird sich auch bei der Umstellung auf semantisches Publizieren nichts ändern, wenn diese nicht auf ein Ad-hoc-Wissenschaftsnetzwerk setzen, das von seiner zugrundeliegenden Ontologie wiederum kaum zu trennen sein wird, da die wissenschaftliche Kommunikation bestimmt, was »ist« – und das ist auch geprägt von (produktiven) Missverständnissen, unseriösen Methoden und handfesten Streits, die sich jedoch erst in der nachträglichen Reflexion offenbaren und für Dienstleister schon gar nicht zu erkennen sind – und selbst wenn: nüchterne Neutralität und damit Ignorieren fordern. Doch ist diese Neutralität überhaupt möglich?

Für die Konstruktion von Begriffen und anderen Elementen zur Erstellung einer umfassenden Ontologie müssen unumgänglich Unterscheidungen getroffen werden. Dieser Unterscheidungsgebrauch bringt zusätzliche Komplexität in das System wissenschaftlicher Kommunikation hinein, gleichwohl eine Ontologie ja das unabhängig von ihr bestehende Wissen zusammenführen und damit die beobachtete Komplexität reduzieren soll. Das geschieht auch, aber jede neue Unterscheidung bedeutet gleichzeitig neue Komplexität, der also niemals wirklich beizukommen ist: Es handelt sich um eine endlose Schleife, verursacht durch die Beobachterabhängigkeit des Unterscheidungsgebrauchs (siehe auch Kapitel 2.2). Statt diese durch Kompromisse zu verdecken, sollte der Beobachter einbezogen werden. Statt das durch Autoritäten Ausgeschlossene nur zu reflektieren, erlaubt das Semantic Web die Einbeziehung aller Beobachter, die einem Wissenschaftsnetzwerk Unterscheidungen hinzufügen und damit selbst Verantwortung dafür tragen, ihren Unterscheidungsgebrauch zu reflektieren. Wissen ist immer nur das Wissen eines Beobachters, der freilich auch ein soziales System sein kann, auf das dann aber durch psychische Systeme nicht zugegriffen kann. Stattdessen beobachten sie das soziale System selbst wieder mit den eigenen Unterscheidungen – mit denselben Konsequenzen.

Um die sich daraus ergebende Chance, auf das vor- oder nachgelagerte Einziehen einer Meta-Ebene mit allen problematischen Konsequenzen verzichten zu können, einzulösen, müssen eine Reihe von immer wieder angezweifelte (siehe Kapitel 3), aber dennoch weiter bestehenden Selbstverständlichkeiten aufgegeben und durch andere Grundsätze ersetzt werden, die in Tabelle 1 gegenübergestellt sind.<sup>66</sup>

Neue Grundsätze	Bestehende Selbstverständlichkeiten
Wissen ist dynamisch, flüchtig und inkonsistent.	Der Wissenkorpus kann vollständig erfasst und erschlossen werden. (Nur wurde diese Aufgabe noch nicht bewältigt.)
Das Ganze ist mehr als seine Teile. Jede Unterscheidung ist kontextabhängig.	Wenn man alle Worte aller Sprachen plausibel und erschöpfend mit Bedeutungen und die so gebildeten Begriffe miteinander verknüpft, erreicht man die Darstellung allen Wissens.
Wissensrepräsentation ist gleichzeitig die Erzeugung von Wissen.	Wissensrepräsentation dient dem <i>information retrieval</i> , ist niemals Selbstzweck und bildet vorhandenes Wissen lediglich ab (siehe Stock und Stock 2008, S. XI).

**Tabelle 1.** Neue Paradigmen zum Aufbau eines Wissenschaftsnetzwerks

<sup>66</sup> Vergleiche die Anforderungen für ein Wissenschaftsnetzwerk S. 40.

Soweit ich sehen kann, wurde diese Befruchtung der Informationswissenschaft mit der Form- und Systemtheorie bislang noch nicht vorgeschlagen.<sup>67</sup> Zum Abschluss sollen zwei Bilder helfen, die Vorstellung von einem zu konstruierenden Wissenschaftsnetzwerk auf der Basis von Technologien des Semantic Web, darunter Nanopublikationen, zu evozieren: Um zentrale Aussagen als Forschungsergebnis zu formulieren, muss man Unterscheidungen treffen: entweder in einem *unmarked space* (= unreferenziert, ein Neologismus, wie z. B. Genettes »Paratext« 1987) oder indem man eine beobachtete Unterscheidung reproduziert – der Normalfall. Solche Kopien sind jedoch einzigartig und erhalten jeweils einen Stempel der Autorschaft inklusive der aktuellen Zeit, wie übrigens alles, was man in der Nanopublikation verwendet. Gleichzeitig sind alle anderen Kopien der Unterscheidung und natürlich das Original nun damit verbunden.

Man denke sich nun einen Stapel transparenter Folien, die man übereinanderlegen kann. Jede zeigt einen Ausschnitt des Wissenschaftsnetzwerks, dargestellt als Graphen. Auf jeder Folie wird auf eine andere Nanopublikation fokussiert. Gemeinsam haben alle Folien im Stapel, dass sie eine bestimmte Element des Wissenschaftsnetzwerks entweder erstellt oder reproduziert haben, z. B. eine Bezeichnung, mit einer Bedeutungsformulierung kombiniert zu einem Begriff. Nehmen wir an, dieser Begriff wird durch die anderen AutorInnen wiederverwendet, so dass dieses Begriffsbündel, das den Begriff in immer anderen Kontexten zeigt, im Zentrum der gestapelten Folien liegt, um das sich alle anderen, damit verbundenen Aussagen anordnen. Solange alle Folien übereinanderliegen, zeigt sich ein komplexes Bild von Verknüpfungen zu vielen anderen Begriffen und Begriffsbündeln. Sobald man einen *trail* – und nichts anderes als die *Memex*-Idee von 1945 sind diese Verknüpfungen (siehe Kapitel 4.1) – genauer untersuchen und herausgreifen will, zeigt sich, auf welchen Folien Teile dieses *trails* zu finden sind und welche Verknüpfungen durch die Überlagerung semi-automatisch erstellt wurden. Semi-automatisch deswegen, weil die Reproduktion von Elementen des Wissenschaftsnetzwerks nur intellektuell geschehen kann, wenn auch gestützt durch eine Suchmaschine, die ihr Auffinden erst ermöglicht. Automatisch erfolgt die Einordnung in dieses dreidimensionale Netzwerk jedoch nur dadurch, dass auch andere WissenschaftlerInnen Elemente reproduzieren. Wenn man die Folien aller Kopien einer Unterscheidung – nichts anderes sind Elemente vom Typ Bezeichnung – übereinanderlegt, hat man nicht nur einen Überblick über den Gebrauch der jeweiligen Unterscheidung, sondern kann das entstandene Netzwerk, in dessen Zentrum diese Unterscheidung vorübergehend steht, auf interessante Assoziationen hin untersuchen. Es entstehen neue *trails*, die kein einzelner entwickelt hat, dadurch, dass sich die durch die Aussagen erzeugten *trails* in ihrem Unterscheidungsgebrauch überschneiden.

<sup>67</sup> Ein Vergleich der vorliegenden Ergebnisse mit jenem durch Søren Brier (2008) vertretenen, *Cybersemiotics* genannten Ansatz erfordert eine genauere Analyse.

Dieses Modell wird den oben ausführlich erläuterten Bedürfnissen in den Sozial- und Geisteswissenschaften gerecht, wo viel stärker auffällt, dass es so etwas wie allgemeingültige Begriffe nicht gibt. Wenn man sich in den Naturwissenschaften so erfolgreich auf Begriffe verständigt hat, dass man sich gemeinhin darauf verlassen kann, muss man dort auch nicht die »Tiefenstruktur« eines Begriffstapels berücksichtigen und kann unüberlegt an den »Urbegriff« anschließen. Es handelt sich um ein Modell, das für alle Disziplinen geeignet ist und damit den bedeutenden Vorteil bietet, interdisziplinäre Verknüpfungen zu ermöglichen.

Natürlich kann man Ontologien als Ausgangspunkt für die Konstruktion des Wissenschaftsnetzwerks verwenden, doch sollte für den wissenschaftlichen Gebrauch jeder Begriff, einmal durch die Kombination von Bezeichnung und Bedeutung formuliert, als »Original« markiert sein, also verknüpft mit einem Stempel der Autorschaft. Selbstverständlich kann auch Autorschaft umstritten sein: Die Lösung dafür sieht – wie im folgenden Abschnitt deutlich werden wird – dem Umgang mit demselben Problem bei der Darstellung eines Forschungsstandes in herkömmlichen Publikationen ähnlich. Das ist naheliegend, denn das Wissenschaftsnetzwerk ist nichts anderes als ein neues Medium, das den Anforderungen genügen muss, welche die wissenschaftliche Kommunikation derzeit an ihre Medien stellt. Typische Techniken des Unterscheidungsgebrauchs in wissenschaftlichen Publikationen werden recht genau in das neue Medium kopiert und stehen ebenso zur Verfügung wie im alten Medium.

Die Versorgung der wissenschaftlichen Kommunikation mit Aussagen ist selbst Gegenstand der wissenschaftlichen Kommunikation. Wenn der Erfinder einer Aussage sich selbst aus biologischen Gründen nicht mehr per Nanopublikation als solcher im Wissenschaftsnetzwerk verewigen kann, müssen diese Zuschreibungen posthum erzeugt werden. Ein durch einen zeitgenössischen Autor X erzeugter Begriff einer historischen Aussage muss also gleichzeitig Elemente für vermeintliche historische Erfinder und ihre Werke erzeugen. Wenn der zeitgenössische Autor Y meint, eine andere Herkunft belegen zu können, reproduziert er zwar das fragliche Element, erzeugt aber zugleich eine ablehnende Verknüpfung zu den Aussagen von Autor X. Autor Z kann nun Autor Y beipflichten und dessen Kopie des Elements reproduzieren – und verzichtet somit auf die direkte Verknüpfung mit dem durch Autor X erzeugten Element, dem »Original«.

Das Beispiel zeigt: Obwohl nicht jede Beschäftigung mit einem bestimmten Konzept sofort integriert werden kann, wird die wissenschaftliche Diskussion ihre eigene Geschichte nach und nach integrieren müssen – sonst arbeitet sie nicht seriös. Es stellt sich das gleiche Problem wie beim Übergang von analoger zu digitaler Publikation: der Medienbruch. Auf einmal muss an zwei Stellen gesucht werden: In den analogen Beständen von Bibliotheken und Archiven und in den digitalen, weil eine Digitalisierung aller relevanten, also zu referenzierenden analogen Bestände lange nicht abgeschlossen

sein wird.<sup>68</sup> Beide Medientypen erfordern unterschiedliche Kompetenzen und Infrastrukturen. Sollte sich die Wissenschaft auf die Umstellung von der digitalen auf die semantische Publikation orientieren, wird man in einigen Fachbereichen mit einem doppelten Medienbruch konfrontiert sein. Das trifft vor allem die Sozial- und Geisteswissenschaften, hier insbesondere die Geschichts- und Kulturwissenschaften. Man darf annehmen, dass ein doppelter Medienbruch eine enorme Adaptionsbarriere aufbaut, wenn die Vorteile des neuen Mediums nicht hinreichend plausibilisiert werden und nicht gleichzeitig auch die Erwartung der Nutzung eines solchen Publikationsmediums im Wissenschaftssystem verankert wird.

Das zweite Bild soll weniger das Wissenschaftsnetzwerk selbst plastischer machen, sondern die Vorteile des beschriebenen Vorgehens zu seiner Erzeugung erläutern, und zwar gegenüber der Option, semantisches Publizieren ausschließlich durch die Verknüpfung von Nanopublikationen mit prädefinierten, also kontrollierten Ontologien zu ermöglichen. Es handelt sich um ein klassisches Henne-Ei-Problem, dem man beikommen kann, indem man eine scheinbar perfekte Henne konstruiert. Das ist der derzeit begangene Weg, bei dem man weder Kosten noch Mühe scheut, umfassende Ontologien zu konstruieren. Nur: die Henne ist äußerst kompliziert aufgebaut, was man nicht erkennt, weil man sie nach einem Modell baut – es ist ja die erste Henne –, sondern weil man weiß, was sie können soll. Man beginnt, was den »Bauplan« angeht, mit einem leeren Blatt, mit einem *unmarked space*, in den man Unterscheidungen zeichnen muss. Macht man einen Fehler bei der Konstruktion, wird man niemals zu einem Ei kommen. Alternativ kann man ein Ei konstruieren, das bekanntlich sehr einfach strukturiert ist. Das gleiche gilt übrigens für eine Nanopublikation. Ein Ei allein macht möglicherweise keine Henne (sondern z. B. einen Hahn). Man braucht also eine kritische Menge von Eiern, die sich dann – als Hühner ausgewachsen – gegenseitig befruchten können. Natürlich kann man auch Fehler bei der Konstruktion des Eis machen, aber diese werden aufgrund der vergleichsweise wenig komplexen Struktur schwerer zu machen und leichter zu finden sein. Ist der Durchbruch einmal geschehen, wird in kürzester Zeit eine Hühnerschar entstehen . . . und dann Probleme, mit denen man einen Umgang finden muss.

Um das Bild nun zu verlassen: Es werden unbemerkte Redundanzen entstehen, unverbundene Synonyme etc. Das sind einerseits Inkonsistenzen, wie sie bei der Auseinan-

<sup>68</sup> Am effektivsten dürfte diese Herausforderung nach einem Modell ähnlich dem Projekt *E-Books on Demand* (EoD, siehe <http://books2ebooks.eu/>, zuletzt besucht 23. Mai 2014) zu meistern sein, wenn es nicht vor allem um die Erhaltung des Werks durch Formatkonversion geht: Normalerweise finden Digitalisierungsprojekte auf der Ebene einzelner Einrichtungen oder kleinerer Kooperationen statt. Es werden dann so viele Titel digitalisiert, wie es die erteilten Ressourcen zulassen. EoD schont Ressourcen dadurch, dass ein Bibliotheksnutzer die Digitalisierung erst anstoßen muss. Finanziert wird das Projekt in einem größeren institutionellen Verbund und durch die Nutzenden selbst.

dersetzung mit dem eigenen Wissen ebenfalls aufscheinen mögen, andererseits aber auch Inkonsistenzen, wie sie wahrscheinlich nur von sozialen Systemen erzeugt werden können, z. B. indem Beiträge durch den Gebrauch eines eigentlich wissenschaftsfernen symbolisch generalisierten Kommunikationsmediums – wie Macht – die Kommunikation irritieren. Ein Wissenschaftsnetzwerk darf nicht den Anspruch der Konsistenz verfolgen, denn das Wissen der Welt ist nicht konsistent. Es kann helfen, redundante Forschung zu vermeiden. Für erste Schlussfolgerungen im Wissenschaftsnetzwerk benötigt man auch kein Deduktionsregelwerk, da durch die Verknüpfung von etwas mit etwas anderem immer etwas Zusätzliches, Drittes entsteht. Auch in der Mathematik ist die eine Seite einer Gleichung nicht identisch mit der anderen, auch wenn ihnen der gleiche Wert zugeschrieben wird. Sonst bräuchte man mit Gleichungen nicht rechnen! Auch kann man das Gleichheitszeichen selbst nicht kürzen. Das wäre die Vernichtung der Gleichung und die Erschaffung von etwas wiederum Neuem.

Mit einem Wissenschaftsnetzwerk ist man beispielsweise nicht mehr darauf angewiesen, von einem leeren Suchschlitz oder den Referenzen einer Publikation ausgehend Quellen zu finden, zu vergleichen, Widersprüche zu entdecken und diese Analyse per Publikation zugänglich zu machen: als Beitrag, der abgelehnt werden kann. Solche prekären Operationen kann ein semantisches und dabei noch immer triviales System weitestgehend vorbereiten, wenn man ihm unter den notwendigen technischen und informatorischen Voraussetzungen entsprechende Selbstbeobachtungsaufträge erteilt. Es beteiligt sich dann gewissermaßen an der Kommunikation,<sup>69</sup> da es selbst, wie Mons und Velterop (2009) beschreiben, Hypothesen vorschlagen kann.

Dazu sind die bestehenden Ontologien enorm hilfreich, da sie logische Beziehungen zwischen Begriffen, die in Publikationen als »Lehrbuchwissen« nicht referenziert werden müssen, bereits enthalten. Alle Beziehungen müssen jedoch möglichst flexibel sein, keine anfechtbaren Restriktionen enthalten, um auch Widerspruch ausdrücken zu können. Gegenwärtige Ontologien für die Wissenschaft teilen dieselbe in Fachbereiche ein und gehen von diesen ersten Unterscheidungen immer weiter in die Tiefe. Die so begründeten Hierarchien lassen zwar Querverweise zu, aber behaupten die Gültigkeit von einem oder mehreren Pfaden. Auf diese Weise ist es unmöglich, ein »echtes« Semantic Web zu bauen.<sup>70</sup> Die Überwindung des ontologischen Weltbilds ist nicht ohne die Überwindung auch der dieses Weltbild verfestigenden gegenwärtigen Ontologien der Computerlinguistik zu erreichen.

<sup>69</sup> Hier liegt der Verweis auf Luhmanns berühmte Kommunikation *mit* Zettelkästen nahe. Siehe Luhmann 1992.

<sup>70</sup> Damit ist auch eine Deutungsmöglichkeit für Kim Veltmans oben (siehe Kapitel 3.4) referierte Position gefunden, dass OWL die Chancen für ein im Luhmannschen Sinn echtes Semantic Web verbaut hätte.

Um die Chancen, die ein solches Wissenschaftsnetzwerk bietet, voll ausnutzen zu können, werden einerseits Methoden der Qualitätssicherung für alle neuen Beiträge und andererseits ein *Graphical User Interface* (GUI) für die Recherche und Eingabe von Nanopublikation sowie Algorithmen benötigt, die Inkonsistenzen sichtbar machen. Diese Inkonsistenzpunkte markieren Forschungsbedarf: Automatisch erzeugte Hypothesen über Widersprüche sind von unschätzbarem Wert für die Begründung von Forschungsprojekten.

Wie verändert sich das Verhältnis von Technik und Gesellschaft, wenn eine Maschine in die Lage versetzt wird, solche sozial folgenreichen Ausgaben zu generieren? Das eingeübte sozio-technische Verhältnis ist durch strukturelle Kopplungen zwischen den beiden unterschiedlichen Systemtypen begründet: Das soziale System bietet die Entwicklungs- und Einsatzvoraussetzungen für das technische System, das wiederum der Kommunikation Komplexität zur Verfügung stellt: Das soziale System kann »mehr Überraschendes und noch Unbekanntes typisieren und verarbeiten und damit den Bereich von Kommunikation vergrößern, der verstanden werden kann« (Luhmann 2000, S. 46). Dadurch, dass technische Systeme in Kausalstrukturen operieren, steigern sie die Erwartbarkeit bestimmter Ereignisse und reduzieren damit die Unwahrscheinlichkeit von Kommunikation, allerdings nur, wenn Umwelteinflüsse weitestgehend ausgeschaltet werden. Selbst wenn man den Zufall einbaut, ist dann eben der Zufall erwartbar. Dadurch, dass auch technische Systeme eine Umwelt haben, die sie nicht vollständig kontrollieren können, erzeugen sie gleichermaßen Unbestimmtheit.<sup>71</sup> Das liegt schlicht daran, dass sie neue Unterscheidungen einführen, was z. B. mit dem von Latour (1988) entlehnten Beispiel vom Türenöffnen treffend beschrieben ist: Um das Problem zu lösen, kann man einen Portier anstellen oder zu technischen Lösungen greifen, z. B. eine Klinke oder einen automatischen Türschließmechanismus einbauen. Zwar erwartet man vom Portier, dass er unaufgefordert die Tür öffnet, sobald man auf die Tür zugehend eine kritische Nähe erreicht hat, aber aufgrund der Erfahrungen, die wir mit Menschen haben, ist immer auch im Horizont die Möglichkeit präsent, wieder einmal nicht wahrgenommen zu werden. Eine verschlossene Tür mit Klinke als Ausgang zu erkennen, aktualisiert die Erwartung, dass die Tür verschlossen sein könnte, aber eher nicht, dass die Klinke abbrechen wird und wir deshalb festsitzen werden. Viel voraussetzungsreicher noch ist der automatische Öffner, der uns nötigt, in einem gewissen Winkel an die Tür heranzutreten, was wiederum voraussetzt, dass wir erkannt haben, dass wir es mit einem Automatismus zu tun haben. Nur der Portier signalisiert bereits, bevor er uns wahrgenommen hat, dass wir ohne größere Probleme hinausgelangen werden. Die Konsequenz ist: Die Technik spielt mit, erzeugt, bestätigt und enttäuscht Erwartungen, dirigiert das menschliche Verhalten, manchmal in Folge

<sup>71</sup> Siehe hierzu insbesondere Luhmanns Soziologie des Risikos 1991 und 1995a.

eines völlig unerwarteten Einwirkens von Umwelteinflüssen auf die Technik sogar mit katastrophalen Folgen.

Der Einsatz von Technik ist also immer mit Risiken verbunden, die auch im vorliegenden Zusammenhang zu reflektieren sind. So ist ein offenes Wissenschaftsnetzwerk auch leicht manipulierbar und angreifbar, wenn nicht entsprechende Vorkehrungen getroffen werden. Das Risiko kann niemals ganz beseitigt, lediglich minimiert werden, da sich die Grundkonstellation zwischen Technik und seiner Umwelt, zu der auch die Gesellschaft und die Menschen gehören, niemals ändert. Die Kopplung des vorgeschlagenen technischen Systems gerade mit dem Wissenschaftssystem bietet allerdings gute Voraussetzungen, da man in der Wissenschaft immer schon Manipulationsmöglichkeiten mitdenkt. Gerade weil es hier um Leben und Tod<sup>72</sup> geht, ist die Wissenschaft gleichzeitig auf Beschleunigung und die Erzeugung von Kausalität angewiesen, wie sie Kommunikation allein nur begrenzt leisten kann.

Was aber ist, wenn technische Systeme nicht mehr nur Sinn in Form von kausal verketteten Ereignissen produzieren, sondern als sogenannte selbstlernende Systeme gerade darauf programmiert sind, neue Unterscheidungen zu errechnen, die niemand vorhergesehen hat? Die Differenz zwischen trivialer Maschine und Bewusstsein bzw. Sozialsystem verschwimmt: Beide prozessieren Sinn dann als kontingente Selektionen. Rammert und Schulz-Schäffer (2002) meinen, selbstlernenden Systemen könne dann Intentionalität unterstellt werden. Ich halte das für *Science Fiction*. Ob es tatsächlich Menschen gibt, die Maschinen Intentionalität unterstellen, wäre empirisch zu prüfen. Wenn selbstlernende oder intelligente, »smarte« Maschinen plötzlich tatsächlich so beobachtet werden, ist es höchste Zeit, den BenutzerInnen zu vermitteln, inwiefern es möglich ist, dass Maschinen intransparent für das Bewusstsein arbeiten und Entscheidungen treffen können. Wenn heute Computer unerwartete Operationen ausführen, rechnet man dies eher einem Virus oder einem Programmfehler zu. Dass bereits derart komplexe Operationen von Computern ausgeführt werden können, wissen nur wenige, selbst wenn sie deren Resultate nutzen. Es ist nicht erforderlich, ihr Zustandekommen zu verstehen oder zu hinterfragen, solange die Resultate als solche erwartbar sind. Wenn innerhalb des Wissenschaftsnetzwerks aufgrund logischer Verknüpfungen in Ontologien automatisch auf Aussagen geschlossen wird, fällt das allein deshalb auf, weil eine sonst obligatorische Autorschaft, die aufgrund von Identifikationssystemen zunehmend schwieriger zu fälschen sein wird, nicht nachgewiesen ist. Die bahnbrechende Funktionalität, neue Aussagen generieren zu können, die der unverzichtbaren wissenschaftlichen Prüfung standhalten, wird großes Aufsehen erregen und daher

<sup>72</sup> Natürlich ist zunächst an die medizinische Forschung zu denken, aber auch die Kulturwissenschaften sind bedeutsam für die Krisen- und Konfliktbewältigung in der gegenwärtigen Gesellschaft.

nicht unterschiedslos zu den intellektuell erzeugten Aussagen zur Kenntnis genommen werden.

Gerade das Semantic Web zeichnet sich durch die Besonderheit aus, Daten in unbegrenzt mögliche Formen der Mitteilung zu bringen (vgl. Luhmann 1997, 309f.). Wer Daten in ein Wissenschaftsnetzwerk eingibt, kann nicht wissen, wie diese einmal dargestellt werden. Zwar werden ebenso wie beim analogen Medium Buch die Form der Mitteilung von der Information unterschieden, aber die Autorin verliert jedweden Einfluss auf die Gestaltung dieser Form. Das kann skeptisch machen, weshalb es enorm wichtig ist, sich für den Betrieb eines solchen Netzwerks genau zu überlegen, was mit diesen sensiblen, weil personenbezogenen Daten getan werden darf, ob z. B. einzelne, in einer Nanopublikation zusammengefasste Aussagen oder einzelne Mark-ups eines abgeschlossenen Textes nur unter bestimmten Umständen voneinander getrennt verarbeitet werden dürfen.

Es bleibt zu fragen, ob dieser Vorschlag von einem Wissenschaftsnetzwerk nicht viel zu weit greift und überzogene Erwartungen an das Semantic Web weckt. Stefan Gradmann – sonst selbst mit der Forderung nach der Abkehr vom monolithischen Dokument nicht eben pessimistisch – warnt:

Fakt ist jedenfalls, dass die auf komplexen Logikansätzen basierenden Versprechungen der Frühzeit des Semantic Web mit ihren intelligenten Agenten und einer großen Nähe zu den weit überzogenen Versprechen der KI das Semantic Web große und kaum einlösbare Erwartungen hinsichtlich der Repräsentierbarkeit von Bedeutung und Interpretation im menschlichen Sinne erzeugt haben, die der erfolgreichen Entwicklung des Projektes nicht eben förderlich waren (Gradmann 2013).

Ich würde dagegen ebenso ungeschützt vermuten, dass der Durchbruch des Semantic Web viel eher deshalb auf sich warten lässt, weil den potentiellen UserInnen bislang weder vermittelt wurde, wie, weshalb, worüber und wohin sie *triples* oder *quads* schreiben sollen, noch Systeme bereitgestellt wurden, die ihnen erlauben, mit den gewohnten Mitteln der Eingabe in Webformulare selbst relativ voraussetzungslos mit ihren Beiträgen am Semantic Web teilzunehmen. Auch haben sie vielleicht bemerkt, dass in den letzten Jahren die Suchmaschinen erfolgreicher und ihre Ergebnisse besser geworden sind, aber es wurde ihnen in nennenswertem Umfang bislang weder von der Presse noch von anderen Medien zugemutet, die Gründe dafür mitgeteilt zu bekommen, vermutlich, weil es jenen, die das Semantic Web verstanden haben, schwer fällt, eine Botschaft zu formulieren, die ohne rätselhafte Akronyme auskommt und klarmachen kann, worum es geht, was gerade passiert und was möglich wäre – zum Nutzen der Kommunikation und also des Users.

Das hier vorgestellte Konzept ist, um das noch einmal zu betonen, keine völlig neue Form von wissenschaftlicher Kommunikation. Es wird lediglich ein zusätzliches Me-

dium vorgeschlagen, das das alte nicht völlig ersetzen, sondern seine Schwächen ausgleichen soll. Selbst eine Orientierung der wissenschaftlichen Kommunikation auf die Nanopublikation als primäres Publikationsformat wird das Verfassen von Fließtexten, insbesondere in den Geistes- und Sozialwissenschaften, bis auf Weiteres nicht gänzlich verdrängen.<sup>73</sup> Vor allem die Kulturwissenschaften sind auf Narrative und kreatives Schreiben angewiesen, da diese enorm zur Reputation eines Forschenden beitragen. Auch werden bereits Technologien unter dem Titel *ontology learning* entwickelt, die das automatische Erstellen von Ontologien aus größeren Textkorpora ermöglichen (siehe Wong, Liu et al. 2012). Da diese noch nicht ausgereift sind und auch zu untersuchen wäre, ob sie für ein Wissenschaftsnetzwerk mit den hier formulierten Anforderungen überhaupt geeignet sind, kann man sich fragen, welchen Anreiz KulturwissenschaftlerInnen haben, am Netz mitzuweben und zusätzlich zum Verfassen eines Textes auch noch Nanopublikationen zu erstellen: Einerseits kann dadurch die Anschlussfähigkeit der eigenen Forschungsergebnisse erhöht werden, andererseits, und das wird in einer Konsolidierungsphase wichtiger sein, kann die Formulierung von *triples* die Klarheit der Argumente und Anschlüsse fördern und auf die parallele konventionelle Publikation zurückwirken, die übrigens im beschriebenen Wissenschaftsnetzwerk leicht einzubinden ist, sobald sie durch eine URI identifiziert werden kann.

Diese Beschreibungen, formuliert als Ausblick, stehen am Ende dieser Studie, die als früher Beitrag zu einem nahezu unbestellten Feld für Forschung und Entwicklung verstanden werden soll. Es fehlt eine Anbindung an und nötigenfalls die Erweiterung von bestehenden Analysen des wissenschaftlichen Publikationsverhaltens. Außerdem wurde ein Bedarf für ein neues Publikationsmedium unterstellt, jedoch nicht empirisch erhoben. Schließlich muss auch ein konkretes Umsetzungsbeispiel zukünftigen Studien vorbehalten sein.

Auch wenn die hier entwickelte Utopie fern einer Umsetzung ist, vermag sie vielleicht auch für die gegenwärtige Bewältigung des Medienbruchs von analog auf digital hilfreich sein, da sie gleichzeitig zeigt, wie beschränkt die derzeitige Nutzung der technischen Möglichkeiten ist. Sie macht deutlich, wie dringend die Wissenschaft auf eindeutige Identifikation und Persistenz ihrer Publikationen und damit auf die Digitalisierung oder wenigstens eindeutige digitale Nachweise von analogen Quellen angewiesen ist, und wie notwendig die Überwindung überkommener, aus dem analogen Medium übernommener Techniken der Qualitätssicherung ist. Es liegt auch eine Chance darin, Recherche- und Publikationssystem zu vereinen, um Adaptionsbarrieren

<sup>73</sup> Wie solche Texte technisch und textstrukturell formatiert sein sollten, um möglichst funktional für die wissenschaftliche Kommunikation zu sein, ist eine andere Frage, der an dieser Stelle nicht nachgegangen werden kann. Denkbar ist jedenfalls, z. B. mittels RDFa-Mark-up Nanopublikation und Fließtext punktgenau miteinander zu verknüpfen.

abzubauen. Es ist an der Zeit, Sprachbarrieren zu überwinden, Teilnahmevoraussetzungen zu senken und die wissenschaftliche Kommunikation mit neuen Freiheitsgraden auszustatten.

## Literatur

- Archambault, Éric (2013). »The Tipping Point - Open Access Comes of Age«. In: *ISSI 2013 Proceedings Volume II*, S. 1665–1680. URL: <http://www.issi2013.org/proceedings.html> (besucht am 13.05.2014).
- Bachmann-Medick, D. (2009). *Cultural turns. Neuorientierungen in den Kulturwissenschaften*. Reinbek bei Hamburg: Rowohlt.
- Berners-Lee, Tim (1990). *Information Management: a Proposal*. URL: <http://www.w3.org/History/1989/proposal.html> (besucht am 23.04.2014).
- Berners-Lee, Tim und James Hendler (2001). »Publishing on the Semantic Web«. In: *Nature* 410.6832, S. 1023–1024. DOI: 10.1038/35074206.
- Birukou, Aliaksandr, Joseph Rushton Wakeling, Claudio Bartolini, Fabio Casati, Maurizio Marchese, Katsiaryna Mirylenka, Nardine Osman, Azzurra Ragone, Carles Sierra und Aalam Wassef (2011). »Alternatives to Peer Review: Novel Approaches for Research Evaluation«. In: *Frontiers in Computational Neuroscience* 5.56. DOI: 10.3389/fncom.2011.00056.
- Bouabid, Hamid und Vincent Larivière (2013). »The Lengthening of Papers' Life Expectancy: a Diachronous Analysis«. In: *Scientometrics* 97.3, S. 695–717. DOI: 10.1007/s11192-013-0995-7.
- Bourne, Philip, Tim Clark, Robert Dale, Anita de Waard, Ivan Herman, Eduard Hovy und David Shotton, Hrsg. (2011). *Force11 White Paper: Improving The Future of Research Communications and e-Scholarship*. URL: <http://force11.org/> (besucht am 10.03.2013).
- Bourne, Philip E., J. Lynn Fink und Mark Gerstein (2008). »Open Access: Taking Full Advantage of the Content«. In: *PLoS Computational Biology* 4.3, e1000037. DOI: 10.1371/journal.pcbi.1000037.
- Brembs, Björn, Katherine Button und Marcus Munafò (2013). »Deep Impact: Unintended consequences of journal rank«. In: *Frontiers in Human Neuroscience*. DOI: 10.3389/fnhum.2013.00291.
- Brier, Søren (2008). *Cybersemiotics. Why Information is not enough!* Toronto: University of Toronto Press.
- Buckland, Michael (1998). *What is a »Digital Document«?* Shortened Version of »What is a »Document?«, *Journal of the American Society for Information Science* 48/9, 1997, S. 804-809. URL: <http://people.ischool.berkeley.edu/~buckland/digdoc.html> (besucht am 10.04.2014).
- Bush, Vannevar (1945). »As We May Think«. In: *The Atlantic*. URL: <http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/> (besucht am 04.04.2013).

- Börsenverein des Deutschen Buchhandels, Abt. Marktforsch. u. Statistik, Hrsg. (2013). *Buch und Buchhandel in Zahlen 2013*. Frankfurt am Main: MVB.
- Bühl, Walter (1974). *Einführung in die Wissenschaftssoziologie*. München: C. H. Beck.
- Clare, Amanda, Samuel Croset, Christoph Grabmueller, Senay Kafkas, Maria Liakata, Anika Oellrich und Dietrich Rebolz-Schuhmann (2011). »Exploring the Generation and Integration of Publishable Scientific Facts Using the Concept of Nano-publications«. In: *Proceedings of the First Workshop on Semantic Publication (SePublica2011), 8th Extended Semantic Web Conference Hersonissos, Crete, Greece, May 31, 2011*. Hrsg. von Anita De Waard, Alexander García Castro, Christoph Lange und Evan Sandhaus. CEUR Workshop Proceedings. Vol. 721, S. 13–17. URL: <http://ceur-ws.org/Vol-721/> (besucht am 28.04.2014).
- Concept Web Alliance (2013). *Nanopublication Guidelines*. Working Draft. URL: <http://www.nanopub.org/2013/WD-guidelines-20131215/> (besucht am 22.03.2014).
- Daniel, Ute (2006). *Kompendium Kulturgeschichte. Theorien, Praxis, Schlüsselwörter. Fünfte durchgesehene und aktualisierte Auflage*. Frankfurt am Main: Suhrkamp.
- Domingue, John (2011). *Handbook of Semantic Web Technologies*. Berlin, Heidelberg: Springer.
- Dröge, Evelyn, Julia Iwanowa, Violeta Trkulja, Steffen Hennicke und Stefan Gradmann (2013). »Wege zur Integration von Ontologien am Beispiel einer Spezifizierung des Europeana Data Model«. In: *Informationswissenschaft zwischen virtueller Infrastruktur und materiellen Lebenswelten. Tagungsband des 13. Internationalen Symposiums für Informationswissenschaft (ISI 2013), Potsdam, 19.-22. März 2013*. Hrsg. von Hans-Christoph Hobohm. Glückstadt: Verlag Werner Hülsbusch, S. 273–284. URL: <http://nbn-resolving.de/urn:nbn:de:kobv:525-4161>.
- Dudek, Sarah (2012). »Die Zukunft der Buchstaben in der alphanumerischen Gesellschaft. Text und Dokument unter digitalen Bedingungen«. In: *Bibliothek Forschung und Praxis* 36, S. 189–199. DOI: 10.1515/bfp-2012-0023.
- Ellison, Glenn (2011). »Is Peer Review in Decline?« In: *Economic Inquiry* 49.3, S. 635–657. DOI: 10.1111/j.1465-7295.2010.00261.x.
- Fu-PusH (2014). *Future Publications in den Humanities (Fu-PusH) – Beschreibung des DFG-Projekts*. URL: <http://www.ub.hu-berlin.de/shared/dokumente/dfg-projekt-future-publications-in-den-humanities-fu-push> (besucht am 26.04.2014).
- Funk, Stefan E. (2010). »Digitale Objekte und Formate«. In: *nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Hrsg. von Heike Neuroth, Achim Oßwald, Regine Scheffel, Stefan Strathmann und Karsten Huth. Version 2.3. URL: <http://nbn-resolving.de/urn:nbn:de:0008-20100617136>.
- Gelfand, Julia und Anthony (Tony) Lin (2013). »Grey Literature. Format Agnostic yet Gaining Recognition in Library Collections«. In: *Library Management* 34.6/7, S. 538–550. DOI: 10.1108/LM-03-2013-0022.
- Genette, Gérard (2001). *Paratexte. Das Buch vom Beiwerk des Buches*. Zuerst 1987. Frankfurt am Main: Suhrkamp.

- Ginsparg, Paul (1997). »Winners and Losers in the Global Research Village«. In: *The Serials Librarian* 30.3-4, S. 83–95. doi: 10.1300/J123v30n03\_13.
- Gnoli, Claudio, I. C. McIlwaine und Joan S. Mitchell (2008). »Ten Long-Term Research Questions in Knowledge Organization«. In: *Knowledge Organization* 35.2-3, S. 137–149.
- Gradmann, Stefan (2009). *Signal. Information. Zeichen. Zu den Bedingungen des Verstehens in semantischen Netzen. Antrittsvorlesung*. URL: <http://nbn-resolving.de/urn:nbn:de:kobv:11-10096931>.
- (2011). »Die Zukunft der Bibliotheken. Traditionelle versus neue Formen der Dokument- und Wissensorganisation«. In: *Community of Knowledge. Wissensmanagement in Theorie und Praxis*. URL: <http://www.community-of-knowledge.de/beitrag/die-zukunft-der-bibliotheken> (besucht am 18.07.2013).
- (2013). »Semantic Web«. In: *Grundlagen der praktischen Information und Dokumentation. Band 1: Handbuch zur Einführung in die Informationswissenschaft und -praxis*. Hrsg. von Rainer Kuhlen, Thomas Seeger und Dietmar Strauch. 6. völlig neu gefaßte Ausgabe, S. 509–519.
- Gradmann, Stefan und Jan Christoph Meister (2008). »Digital Document and Interpretation: Re-thinking ›text‹ and Scholarship in Electronic Settings«. In: *Poiesis & Praxis* 5, S. 139–153. doi: 10.1007/s10202-007-0042-y.
- Groth, Paul, Andrew Gibson und Johannes Velterop (2010). *The Anatomy of a Nano-publication*. URL: [http://www.w3.org/wiki/images/c/c0/HCLSIG\\$\\$\\$SWANSIOC\\$\\$\\$Actions\\$\\$\\$RhetoricalStructure\\$\\$\\$meetings\\$\\$\\$20100215\\$cwa-anatomy-nanopub-v3.pdf](http://www.w3.org/wiki/images/c/c0/HCLSIG$$$SWANSIOC$$$Actions$$$RhetoricalStructure$$$meetings$$$20100215$cwa-anatomy-nanopub-v3.pdf) (besucht am 18.07.2013).
- Groza, T., S. Handschuh, T. Clark, Buckingham Shum, S. und A. de Waard (2009). »A short survey of discourse representation models«. In: *Proceedings. 8th International Semantic Web Conference, Workshop on Semantic Web Applications in Scientific Discourse. Lecture Notes in Computer Science*. CEUR Workshop Proceedings. Vol. 523. URL: <http://ceur-ws.org/Vol-523/Groza.pdf> (besucht am 13.05.2014).
- Gruber, TR (1993). »A Translation Approach to Portable Ontologies«. In: *Knowledge Acquisition* 5, S. 199–220.
- Gwebu, Kholekile L. und Jing Wang (2011). »Adoption of Open Source Software: The Role of Social Identification«. In: *Decision Support Systems* 51, S. 220–229. doi: 10.1016/j.dss.2010.12.010.
- Günther, Gotthard (1980). »Das Problem einer transklassischen Logik«. In: *Beiträge zu einer operationsfähigen Dialektik. Band 3*. Meiner: Hamburg, S. 73–94.
- Hendler, Jim und Tim Berners-Lee (2010). »From the Semantic Web to Social Machines: a Research Challenge for AI on the World Wide Web«. In: *Artificial Intelligence* 174, S. 156–161. doi: 10.1016/j.artint.2009.11.010.
- Heßbrüggen-Walter, Stefan (2013). »Tatsachen im semantischen Web: Nanopublikationen in den digitalen Geisteswissenschaften?«. In: *Historyblogosphere – Bloggen in den Geschichtswissenschaften*. Hrsg. von Peter Haber und Eva Pfanzelter, S. 149–160. doi: 10.1524/9783486755732.149.

- IFLA Study Group on the Functional Requirements for Bibliographic Records (2009). »3. Entities«. In: *Functional Requirements for Bibliographic Records. Final Report – Current Text*. HTML version. IFLA. URL: [http://archive.ifla.org/VII/s13/frbr/frbr\\_current3.htm](http://archive.ifla.org/VII/s13/frbr/frbr_current3.htm) (besucht am 15.04.2014).
- Kaden, Ben (2013). »Elektronisches Publizieren«. In: *Grundlagen der praktischen Information und Dokumentation. Band 1: Handbuch zur Einführung in die Informationswissenschaft und -praxis*. Hrsg. von Rainer Kuhlen, Thomas Seeger und Dietmar Strauch. 6. völlig neu gefaßte Ausgabe, S. 509–519. DOI: 10.1515/9783110258264.509.
- Kriegeskorte, Nikolaus (2012). »Open Evaluation (OE): A Vision for Entirely Transparent Post-Publication Peer Review and Rating for Science«. In: *Frontiers in Computational Neuroscience* 6.79. DOI: 10.3389/fncom.2012.00079.
- Kuhn, Thomas Samuel und Lorenz Krüger (1978). *Die Entstehung des Neuen: Studien zur Struktur der Wissenschaftsgeschichte*. Frankfurt am Main: Suhrkamp.
- Latour, Bruno (1988). »Mixing Humans and Nonhumans Together. The Sociology of a Door-Closer«. In: *Social Problems* 35.3, S. 298–310.
- Lehmann, Jens, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer und Christian Bizer (2014). »DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia«. In: *Semantic Web Journal*. URL: <http://www.semantic-web-journal.net/content/dbpedia-large-scale-multilingual-knowledge-base-extracted-wikipedia-0> (besucht am 05.05.2014).
- Lewis, David W. (2012). »The Inevitability of Open Access«. In: *College & Research Libraries* 73.5, S. 493–506. URL: <http://crl.acrl.org/content/73/5/493.abstract> (besucht am 12.04.2014).
- Luhmann, Niklas (1975a). »Einfache Sozialsysteme«. In: *Soziologische Aufklärung* 2. Opladen: Westdeutscher Verlag, S. 21–39.
- (1975b). »Interaktion, Organisation, Gesellschaft. Anwendungen der Systemtheorie«. In: *Soziologische Aufklärung* 2. Opladen: Westdeutscher Verlag, S. 5–20.
- (1984). *Soziale Systeme. Grundriß einer allgemeinen Theorie*. Frankfurt am Main: Suhrkamp.
- (1990). *Die Wissenschaft der Gesellschaft*. Frankfurt am Main: Suhrkamp.
- (1991). *Soziologie des Risikos*. Berlin/New York: Walter de Gruyter.
- (1992). »Kommunikation mit Zettelkästen. Ein Erfahrungsbericht«. In: *Universität als Milieu. Kleine Schriften*. Hrsg. von André Kieserling. Bielefeld: Haux, S. 53–61.
- (1993). »»Was ist der Fall?« und »Was steckt dahinter?« Die zwei Soziologien und die Gesellschaftstheorie«. In: *Zeitschrift für Soziologie* 22.4, S. 245–260.
- (1995a). »Das Risiko der Kausalität«. In: *Zeitschrift für Wissenschaftsforschung* 9/10, S. 107–119.
- (1995b). *Die Kunst der Gesellschaft*. Frankfurt am Main: Suhrkamp.
- (1997). *Die Gesellschaft der Gesellschaft*. Frankfurt am Main: Suhrkamp.
- (2000). *Die Religion der Gesellschaft*. Frankfurt am Main: Suhrkamp.

- Luhmann, Niklas (2005). »Die Unwahrscheinlichkeit der Kommunikation«. In: *Soziologische Aufklärung* 3. Zuerst 1981. Wiesbaden: VS, S. 29–40.
- Lund, Niels Windfeld (2009). »Document Theory«. In: *Annual Review of Information Science and Technology* 43.1, S. 1–55. DOI: 10.1002/aris.2009.1440430116.
- Mancini, Clara und Simon J. Buckingham Shum (2006). »Modelling Discourse in Contested Domains: A Semiotic and Cognitive Framework«. In: *Int. J. Human-Computer Studies* 64, S. 1154–1171. DOI: 10.1016/j.ijhcs.2006.07.002.
- McCracken, Harry (2014). »The Web at 25: Revisiting Tim Berners-Lee’s Amazing Proposal«. In: *Time*. URL: <http://time.com/21039/tim-berners-lee-web-proposal-at-25/> (besucht am 16. 04. 2014).
- Mittler, Elmar (2011). »Wissenschaftliche Forschung und Publikation im Netz«. In: *Medienkonvergenz - Transdisziplinär*. Hrsg. von Stephan Füssel. DeGruyter, S. 31–80.
- Mons, Barend und Jan Velterop (2009). »Nano-Publication in the E-Science Era«. In: *Proceedings of the Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009), collocated with the 8th International Semantic Web Conference (ISWC-2009), Washington DC, USA, October 26, 2009*. Hrsg. von Tim Clark, Joanne S. Luciano, M. Scott Marshall, Eric Prud’hommeaux und Susie Stephens. CEUR Workshop Proceedings. Vol. 523. URL: <http://ceur-ws.org/Vol-523/Mons.pdf> (besucht am 29. 04. 2014).
- Nature Genetics Editors (2007). »Editorial. What is the Human Variome Project?«. In: *Nature Genetics* 39.4, S. 423. DOI: 10.1038/ng0407-423.
- Nelson, Theodor Holm (1975). »Chapter Two: Proposal for a Universal Electronic Publishing System and Archive«. In: *Literary Machines*. Zitiert nach Hervada-Sala, Francesc: Computing Pages. Mindful Press. URL: <http://u-tx.net/ccritics/lm2.html> (besucht am 15. 04. 2014).
- Neylon, Cameron (2012). »More than just Access: Delivering on a Network-Enabled Literature«. In: *PLOS Biology* 10.10. DOI: 10.1371/journal.pbio.1001417.
- Pédauque, Roger T. (2003). *Document: Form, Sign, and Medium, as reformulated for Electronic Documents*. URL: [http://archivesic.ccsd.cnrs.fr/sic\\_00000594/fr/](http://archivesic.ccsd.cnrs.fr/sic_00000594/fr/) (besucht am 10. 04. 2014).
- Pettifer, S., P. McDermott, J. Marsh, D. Thorne, A. Villegier und T. K. Attwood (2011). »Ceci n’est pas un Hamburger: Modelling and Representing the Scholarly Article«. In: *Learned Publishing* 24.3, S. 207–220. DOI: doi:10.1087/20110309.
- Pichler, Alois und Amélie Zöllner-Weber (2013). »Sharing and Debating Wittgenstein by using an Ontology«. In: *Literary and Linguistic Computing* 28.2, S. 700–707. DOI: 10.1093/lc/fqt049.
- Priem, Jason und Bradley M. Hemminger (2012). »Decoupling the Scholarly Journal«. In: *Frontiers in Computational Neuroscience* 6.19. DOI: 10.3389/fncom.2012.00019.
- Rammert, Werner und Ingo Schulz-Schäffer (2002). »Technik und Handeln. Wenn soziales Handeln sich auf menschliches Verhalten und technische Abläufe verteilt«. In: *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*. Hrsg. von Werner Rammert und Ingo Schulz-Schäffer. Frankfurt am Main: Campus, S. 11–64.

- Ribaupierre, H el ene de und Gilles Falquet (2013). »A User-Centric Model to Semantically Annotate and Retrieve Scientific Documents«. In: *ACM Sixth International Workshop on Exploiting Semantic Annotations in Information Retrieval. ESAIR'13, October 28-27 2013, San Francisco, CA, USA*. DOI: 10.1145/2513204.2513213.
- Riehm, Ulrich, Knud Bohle und Bernd Wingert (2004). »Elektronisches Publizieren«. In: *Grundlagen der praktischen Information und Dokumentation. Band 1: Handbuch zur Einf uhrung in die Informationswissenschaft und -praxis*. Hrsg. von Rainer Kuhlen, Thomas Seeger und Dietmar Strauch. M unchen: K. G. Saur, S. 549–560. DOI: 10.1515/9783110964110.549.
- Rinaldi, Andrea (2010). »For I Dipped into the Future«. In: *EMBO reports* 11.5, S. 345–349. DOI: 10.1038/embor.2010.57.
- Schmidt, Nora (2014). *Der Goldene Weg des Open Access zum funktionalen Publikationswesen. Handlungsoptionen f ur die Universit at Wien*. URL: <http://hdl.handle.net/10760/22655>.
- Schraefel, MC (2007). »What is an Analogue for the Semantic Web and Why is Having One Important?«. In: *ACM Hypertext, September 10–12, 2007, Manchester, United Kingdom*, S. 123–132. URL: <http://eprints.ecs.soton.ac.uk/14274/> (besucht am 03.04.2014).
- Shadbolt, Nigel, Wendy Hall und Tim Berners-Lee (2006). »The Semantic Web Revisited«. In: *IEEE Intelligent Systems*. DOI: 10.1109/MIS.2006.62.
- Shotton, David (2009). »Semantic Publishing: the Coming Revolution in Scientific Journal Publishing«. In: *Learned Publishing* 22.2, S. 85–94. DOI: 10.1087/2009202.
- (2012). »The Five Stars of Online Journal Articles – a Framework for Article Evaluation«. In: *D-Lib Magazine* 18.1/2. DOI: 10.1045/january2012-shotton.
- Shotton, David, Katie Portwin, Graham Klyne und Alistair Miles (2009). »Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article«. In: *PLOS Computational Biology* 5.4, e1000361. DOI: 10.1371/journal.pcbi.1000361.
- Shum, Simon Buckingham, Tim Clark, Anita de Waard, Tudor Groza, Siegfried Handschuh und Agnes Sandor (2010). »Scientific Discourse on the Semantic Web: A Survey of Models and Enabling Technologies«. In: *Semantic Web Journal*. URL: <http://www.semantic-web-journal.net/content/scientific-discourse-semantic-web-survey-models-and-enabling-technologies> (besucht am 28.04.2014).
- Smith, John W. T. (1999). »The Deconstructed Journal: a New Model for Academic Publishing«. In: *Learned Publishing* 12.2. URL: <http://hdl.handle.net/10760/5841>.
- Sowa, John F. (2000). *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, California: Brooks/Cole.
- Spencer-Brown, George (1969). *Laws of Form – Gesetze der Form*. Hier in der Ausgabe von 2008. Leipzig: Bohmeier.
- St aheli, Urs (1998). »Die Nachtr aglichkeit der Semantik. Zum Verh altnis von Sozialstruktur und Semantik«. In: *Soziale Systeme* 4.2, S. 315–340. URL: <http://www.uni-bielefeld.de/sozsys/deutsch/leseproben/staheli.htm> (besucht am 13.05.2014).

- Stichweh, Rudolf (1994). *Wissenschaft, Universität und Professionen. Soziologische Analysen*. Frankfurt am Main: Suhrkamp.
- Stock, Wolfgang und Mechtild Stock (2008). *Wissensrepräsentation*. München: Oldenbourg.
- Taubert, Niels C. und Peter Weingart (2010). »Open Access« – Wandel des wissenschaftlichen Publikationssystems«. In: *Medienwandel als Wandel von Interaktionsformen*. Hrsg. von Tilmann Sutter und Alexander Mehler. Wiesbaden: VS, S. 159–181. URL: <http://pub.uni-bielefeld.de/download/1902672/2329375> (besucht am 13. 05. 2013).
- Umstätter, Walther: (2011). *Lehrbuch des Bibliotheksmanagements*. Stuttgart: Hiersemann.
- Veltman, Kim H. (2004). »Towards a Semantic Web for Culture«. In: *Journal of Digital Information* 4. URL: <http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Veltman/> (besucht am 15. 02. 2014).
- Verhaar, Peter (2008). *DRIVER, Digital Repository Infrastructure Vision for European Research II. D4.2: Report on Object Models and Functionalities*. URL: [http://www.driver-repository.eu/component?option=com\\_jdownloads/Itemid,83/task,view.download/cid,54/](http://www.driver-repository.eu/component?option=com_jdownloads/Itemid,83/task,view.download/cid,54/) (besucht am 28. 04. 2014).
- Waard, Anita de, Simon Buckingham Shum, Annamaria Carusi, Jack Park, Matthias Samwald und Ágnes Sándor (2009). »Hypotheses, Evidence And Relationships: the Hyper Approach for Representing Scientific Knowledge Claims«. In: *Proceedings of the Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009), collocated with the 8th International Semantic Web Conference (ISWC-2009), Washington DC, USA, October 26, 2009*. Hrsg. von Tim Clark, Joanne S. Luciano, M. Scott Marshall, Eric Prud'hommeaux und Susie Stephens. CEUR Workshop Proceedings. Vol. 523. URL: <http://ceur-ws.org/Vol-523/deWaard.pdf> (besucht am 25. 04. 2014).
- Walkowski, Niels-Oliver (2011). »Semantic Web Techniken im explorativ geisteswissenschaftlichen Forschungskontext«. In: *Digitale Wissenschaft. Stand und Entwicklung digital vernetzter Forschung in Deutschland, 20./21. September 2010, Köln*. Beiträge der Tagung, 2., ergänzte Fassung. Silke Schomburg u. a., S. 27–32. URL: [http://www.hbz-nrw.de/dokumentencenter/veroeffentlichungen/Tagung\\_Digitale\\_Wissenschaft.pdf](http://www.hbz-nrw.de/dokumentencenter/veroeffentlichungen/Tagung_Digitale_Wissenschaft.pdf) (besucht am 13. 05. 2014).
- Ware, Mark und Michael Mabe (2009). *An Overview of Scientific and Scholarly Journal Publishing*. Hrsg. von Technical International Association of Scientific und Medical Publishers. URL: [http://www.stm-assoc.org/2009\\_10\\_13\\_MWC\\_STM\\_Report.pdf](http://www.stm-assoc.org/2009_10_13_MWC_STM_Report.pdf) (besucht am 12. 11. 2012).
- Wong, W., W. Liu und M. Bennamoun (2012). »Ontology Learning from Text: A Look back and into the Future«. In: *ACM Computing Surveys* 44.4. DOI: 10.1145/2333112.2333115.