

# La Redocumentarisation ... ein informationswissenschaftliches Editorial

Liebe Mitglieder und Freunde der DGI, gerade komme ich zurück von der Tagung WWW 2012 in Lyon. Ich sprach dort in einem Panel über die potentielle Rolle des ‚Semantic Web‘ in den ‚Digital Humanities‘ und hatte Gelegenheit, diverse Säulenheilige des WWW – einschließlich Tim Berners-Lee höchstpersönlich – kennen zu lernen.

Vor allem aber ist mir aufgefallen, wie aufgeschlossen das in Lyon natürlich stärker als in anderen Jahren frankophone Publikum für Fragestellungen war, die normalerweise im stark technisch dominierten Umfeld des W3C keine Heimat haben: Fragen nach ‚Sinn‘ und ‚Bedeutung‘ im WWW, nach der semiologischen Verfasstheit des Web (inwieweit ist es als ‚Text‘ zu begreifen) und damit eben auch immer wieder Fragen, die unsere informationswissenschaftliche ‚community‘ unmittelbar angehen: was geschieht im Übergang vom Web der ‚Dokumente‘ zum (semantischen?) Web der ‚Dinge‘, der ‚Linked Data‘? Wieweit haben die Informationsaggregationen des Web noch Dokumentcharakter? Der gesamte Abschlussstag der Konferenz war von Themen dieses Kalibers bestimmt, beginnend mit Bernard Stiegler's faszinierender Keynote zum Thema der ‚Aufklärung in the Age of Philosophical Engineering‘.

All dies erinnerte mich stark an einen Moment während der Podiumsdiskussion auf unserer letzten (wirklich gelungenen!) Konferenz in Düsseldorf: Marlies Ockenfeld wies dort zu Recht darauf hin, dass es wahrscheinlich ein Fehler gewesen ist, den für unsere Zunft so wesentlich konstitutiven Begriff des Dokuments so weitgehend aufzugeben, wie wir dies in den vergangenen 15 Jahren in der Tat getan haben. Und ich habe in diesem Zusammenhang darauf hingewiesen, dass wir es uns in der deutschen Informationswissenschaft erlauben haben, die neben den grundlegenden Arbeiten von Michael Buckland wahrscheinlich gehaltvollsten Beiträge zur Entwicklung des Dokumentbegriffs im Digitalen einfach zu ignorieren – und dies wahrscheinlich nur (oder zumindest vor allem), weil sie fast sämtlich ‚nur‘ in französischer Sprache erschienen sind.

Die Rede ist von den unter dem Kollektivpseudonym Roger T. Pédaque erschienenen Beiträgen, von denen ein erster immerhin auch ins Englische übertragen wurde<sup>1</sup>.

Die weiteren Beiträge sind jedoch im Netz und in Buchform nur noch auf Französisch erschienen<sup>2</sup> und dem entsprechend im nicht-frankophonen Umfeld praktisch nicht rezipiert worden.

Dies möchte ich gerne ändern – und vielleicht hilft ja dieses Editorial dabei.

Vorweg zur Verfasserschaft: es handelt sich bei den mit R. T. Pédaque gezeichneten Dokumenten durchweg um besonders gelungene Fälle von „écriture collective“, um das gemeinsame Werk eines Autorenkollektivs des „Réseau Thématique Pluridisciplinaire“ des CNRS mit dem Titel „Documents et contenu: création, indexation, navigation“ (abgekürzt als „RTP Doc“ – von daher das sprechende Pseudonym!), an dem zahlreiche Philosophen, Linguisten, Semiologen, Informatiker, Mathematiker und Informationswissenschaftler beteiligt waren.

Der 2006 erschienene Band *Le document à la lumière du numérique* fasst die drei zentralen, in den Jahren 2003 bis 2005 entstandenen Texte des Kollektivs zusammen. Thema ist dabei die Entwicklung des Dokumentbegriffs im Übergang vom gedruckten zum digitalen Dokument entlang dreier nicht-exklusiver Vektoren:

- Form (le vu, das Gesehene, in einer Sprachanalogie die Ebene der Morphosyntax): das Dokument als materielles oder nicht-materielles strukturiertes Objekt;
- Zeichen (le lu, das Gelesene, das Gegenstück zur Semantik im Sprachmodell): das Dokument als sinntragende Instanz und damit zugleich intentional determiniert und Teil eines Zeichensystems;
- Medium (le su, das Gewusste, analog zur linguistischen Pragmatik die Ebene des Verstehens, des Begreifens): das Dokument als Kommunikationsvektor, Teil einer sozialen Realität, in der zeitlich und räumlich situierte Vermittlungsprozesse konstitutiv werden.

Diese drei Vektoren werden dann in den drei Texten wiederum jeweils in Dreischritten (vom traditionellen Dokument über das elektronische zum XML-/Web-Dokument) durchdekliniert, die jeweils in für die Perspektive des Vektors spezifischen Definitionen resultieren.

Im ersten Betrachtungswinkel des Dokuments als Form ist dessen Objektcharakter konstitutiv: das traditio-

<sup>1</sup> Roger T. Pédaque: Document: Form, Sign and Medium, As Reformulated for Electronic Documents (= [http://archivisic.ccsd.cnrs.fr/docs/00/06/22/28/PDF/sic\\_00000594.pdf](http://archivisic.ccsd.cnrs.fr/docs/00/06/22/28/PDF/sic_00000594.pdf))

<sup>2</sup> Roger T. Pédaque: *Le document à la lumière du numérique*. Caen 2006.

nelle Dokument wird dabei als die Einheit von Träger und Beschriftung, das elektronische Dokument als die Kombination von Strukturen und Daten und schließlich das XML-Dokument als die Einheit von strukturierten Daten und Stylesheet aufgefasst. Daraus ergibt sich eine erste Definition: *Ein elektronisches Dokument ist eine Menge von Daten, die in einer stabilen Struktur organisiert sind, welche mit Formatierungsregeln verbunden ist und eine Lesbarkeit für den Autor und die Rezipienten des Dokuments herstellt.*

Aus dieser Betrachtungsweise ist eine Reihe von zum Teil auch heute noch aktuellen Forschungsfragen ableitbar: Wie unabhängig voneinander sind Strukturen und Daten? Wie stark muss der Zusammenhang zwischen ihnen sein, um elektronische „Lesbarkeit“ herzustellen? Wie geräteunabhängig ist diese „Lesbarkeit“? Mit Blick auf Langzeitverfügbarkeit: wie viel Struktur und/oder Daten und/oder Styles müssen erhalten bleiben, um Lesbarkeit zu sichern? Und: wie weit geht der mit dem digitalen Dokument als Form verbundene Übergang von der Präzisierung zur Postkoordination? Wie konstitutiv für die menschliche Perzeption ist die Analogie mit etablierten kulturellen Objekten (wie sie z. B. in PDF-Dokumenten noch gegeben ist)?

Der zweite mit Pédaque gezeichnete Text nimmt eine semiologische Perspektive ein: es geht um das Dokument als (komplexes) Zeichen, womit der bedeutungstragende Aspekt konstitutiv wird: das traditionelle Dokument ist in diesem Blickwinkel die Einheit von Schrift und Sinn, das digitale Dokument die Kombination aus digitalem Text und Wissen und das Dokument im Semantic-Web die Kombination von digitalem Text und Ontologien. Auch aus diesem Dreischritt ergibt sich eine Definition, diesmal lautet sie: *Ein digitales Dokument ist ein ‚Text‘, dessen Elemente potentiell durch ein Wissensverarbeitungssystem mit dem Ziel einer Auswertung durch einen kompetenten ‚Leser‘ analysierbar sind.*

Auch aus dieser Definition ist eine – unverändert hochaktuelle! – Forschungsagenda ableitbar: das ‚Dokument‘ wird dabei sekundär: primär ist nun der ‚Text‘ und sein ‚Kontext‘, welche letzterer Interpretierbarkeit überhaupt erst herstellt. Es stellen sich in diesem Zusammenhang Fragen nach dem Verhältnis von Text und Markup, von Text und Metadaten. Kategorien wie ‚Vertrauen‘ und ‚Authentizität‘ erhalten hier große Bedeutung – aber vor allem geht es um ‚Bedeutung‘ und ‚Referenz‘: Kernstück dieses aus meiner Sicht wichtigsten der drei Texte ist eine fundierte Kritik des Semantic Web in seiner semiologischen Naivität, gepaart mit dem Versuch, eine genuine Textsemiologie zu begründen.

Der dritte Vektor schließlich betrachtet das Dokument als Medium und damit als soziales Phänomen. In dieser

Sichtweise stellt sich das traditionelle Dokument als die Einheit von Schrift und Legitimation dar. Das elektronische Dokument wäre demgegenüber die Kombination aus Text und Verfahren und das Web-Dokument die Kombination von Publikation und gemessener Nutzung. Die aus diesem dritten Dreischritt resultierende Definition lautet dann: *Ein elektronisches Dokument ist die Spur sozialer Beziehungen, die mit informationstechnischen Methoden rekonstruiert wurde.*

Auch hieraus ergeben sich wieder Forschungsfragen, und zwar allererst an die Adresse der Archive: ist es ihre primäre Rolle, Vergangenes zu dokumentieren oder sollen sie in erster Linie Geschehen aufzeichnen? Und wieweit verschwimmen in diesem neuen Kontext die Grenzen zwischen Archiv und Publikation? Aber auch allgemeiner: wie wird ‚Aufmerksamkeit‘ generiert? Und wer zahlt? Und wofür, in welcher Währung?

Wenn man diesen ersten drei Texten einen Vorwurf machen kann, so diesen: das Potential von RDF und dem darauf bauenden Web der ‚Linked Data‘ wurde zum damaligen Zeitpunkt noch nicht antizipiert, was allerdings auch eine prophetische Leistung gewesen wäre, befand sich doch das Semantic Web zum Entstehungszeitpunkt der ersten Texte von RTP Doc noch in seiner ersten, stark von Ansätzen der Künstlichen Intelligenz befrachteten Phase: Linked Data entdeckte auch Berners-Lee erst um 2007.

Dies Versäumnis (wenn es denn eines ist) holt Roger T. Pédaque derzeit allerdings nach: im Netz befinden sich derzeit zwei Versionen einer Fortschreibung der Textserie. In der ersten, ausführlichen Version<sup>3</sup> aus dem Jahr 2011 geht es vor allem um die mit der zunehmenden Dominanz des Informations- und Wissensraums WWW einhergehende Umwälzung und die Konstitution einer veritablen „Web Science“ – mit erstaunlich ausführlichen Rückgriffen auf die Dokumentationswissenschaft und insbesondere Otlet. Von der zweiten Version<sup>4</sup> existiert momentan nur eine manifestartige Kurzfassung, sie ist stärker auf den im Web in der Tat problematischen Zeitaspekt fokussiert.

Ergänzt werden diese Veröffentlichungen um einen weiteren Sammelband des Pédaque-Kollektivs (nunmehr aber mit Individualverfassern) mit dem Titel *La redocumentarisation du monde*<sup>5</sup>, drei Tagungsbände zu der von dieser Gruppe initiierten Tagungsreihe „Document nu-

<sup>3</sup> Le web sous tension (v1), <https://espacestemp.com/text/vsMAqHUTfli/view/>

<sup>4</sup> Le web sous tension (v2), <https://espacestemp.com/text/IFg22TP6glM/view/>

<sup>5</sup> Roger T. Pédaque: la redocumentarisation du monde. Paris 2007

mérique et société“ und diverse Einzelveröffentlichungen der Protagonisten der Gruppe – dabei besonders hervorzuheben die zuletzt erschienene Monographie des spiritus rector der Gruppe, Jean-Michel Salaün<sup>6</sup>, der darin eine Geschichte des WWW als Teil einer Geschichte des Dokumentbegriffs zu schreiben versucht.

Die Pédauque-Gruppe ist wahrlich nicht die einzige Gruppe, die weltweit über den Begriff des elektronischen Dokuments arbeitet: zu nennen sind in diesem Kontext mindestens noch die Gruppe um Michael Buckland in Berkeley sowie die Gruppe um Roswitha Skare und Niels Windfeld Lund in Tromsø. Doch ist die französische Gruppe wahrscheinlich die produktivste und am ehesten interdisziplinär ausgerichtete, vielleicht auch die epistemologisch anspruchsvollste.

Und zugleich verwendet sie in der Schlussbetrachtung ihrer ersten Veröffentlichung eine Formulierung, von der her eine sehr überzeugende Brücke zur Informationswissenschaft (der früheren Dokumentation) zu schlagen wäre. Es heißt dort:

Un document ne serait finalement qu'un contrat entre des hommes dont les qualités anthropologiques (lisibilité-perception), intellectuelle (compréhension-assimilation) et sociales (sociabilité-intégration) fonderaient une part de leur humanité, de leur capacité à vivre ensemble (Pédauque 2006, S. 50–51).

In deutscher Übersetzung heißt dies: „Letztlich ist ein Dokument nichts anderes als eine Übereinkunft zwischen

Menschen, deren anthropologische (Lesbarkeit, Wahrnehmung), intellektuelle (Verstehen, Aneignung) und soziale (gesellschaftliche Integrierbarkeit) Dimensionen ein Teil ihres Menschseins und ihrer Fähigkeit zum Zusammenleben selbst begründen.“ Und daran schließe ich die folgende Doppelthese:

Die Bedingungen dieser Übereinkunft unter vernetz-digitalen Bedingungen neu zu verstehen ist vital für die Zukunft der Informationswissenschaft, die dabei primär als eine Wissenschaft vom Verstehen, als „Geisteswissenschaft“ im besten Sinne gedacht sein muss.

Und ohne eine so verstandene Informationswissenschaft werden wir umgekehrt die Bedingungen dieser Vereinbarung nicht einmal ansatzweise verstehen können.

So besehen ist für mich in der Tat eine stärkere Befassung mit dem Kernbegriff „Dokument“ unter den Auspizien des WWW klar überfällig, und ich danke Frau Ockenfeld für ihren dahin gehenden Hinweis in unserer Podiumsdiskussion.

Und vielleicht findet sich für Roger T. Pédauque ja doch noch ein Übersetzer – vielleicht sogar unter den Mitgliedern der DGI?

In diesem Sinne wünsche ich Ihnen eine gute Lektüre des vorliegenden Heftes und uns ein baldiges Wiedersehen – spätestens bei den DGI-Praxistagen in Frankfurt am Main. DGI-Mitglieder bitte ich, sich den Termin unserer außerordentlichen Mitgliederversammlung am 8. November bereits vorzumerken.

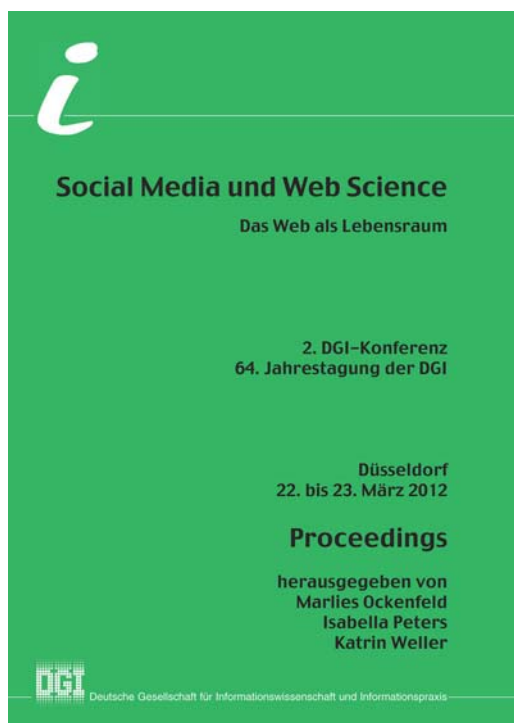
Herzliche Grüße – Ihr Stefan Gradmann

<sup>6</sup> Jean-Michel Salaün: Vu, lu, su. Les architectes de l'information face à l'oligopole du Web. Paris 2012.

# Aktuelle Veröffentlichungen der DGI

Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e.V.

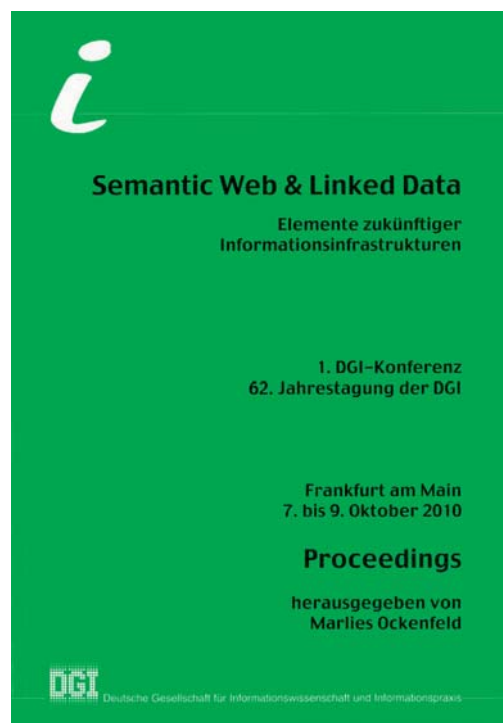
## Tagungsbände der DGI-Konferenzen



2. DGI-Konferenz / 64. Jahrestagung der DGI

### **Social Media und Web Science Das Web als Lebensraum**

Düsseldorf, 22. bis 23. März 2012  
Proceedings, hrsg. von Marlies Ockenfeld,  
Isabella Peters und Katrin Weller  
DGI: Frankfurt am Main 2012, 472 Seiten  
50,- Euro Schutzgebühr  
(für DGI-Mitglieder 40,- Euro)  
ISBN 978-3-925474-72-9



1. DGI-Konferenz / 62. Jahrestagung der DGI

### **Semantic Web & Linked Data Elemente zukünftiger Informations- infrastrukturen**

Frankfurt am Main, 7. bis 9. Oktober 2010  
Proceedings, hrsg. von Marlies Ockenfeld  
DGI: Frankfurt am Main 2010, 272 Seiten  
50,- Euro Schutzgebühr  
(für DGI-Mitglieder 40,- Euro)  
ISBN 978-3-925474-70-5

**DGI-Geschäftsstelle**  
Windmühlstraße 3  
60329 Frankfurt am Main



Telefon 069 43 03 13  
Fax 069 4909096  
E-Mail [mail@dgi-info.de](mailto:mail@dgi-info.de)  
Website [www.dgi-info.de](http://www.dgi-info.de)