

# Vom Weblog lernen... Community, Peer-to-Peer und Eigenständigkeit als ein Modell für zukünftige Wissenssammlungen

Jörg Kantel  
Der Schockwellenreiter  
[www.schockwellenreiter.de](http://www.schockwellenreiter.de)

27. April 2003

## **Zusammenfassung**

Das Internet erscheint vielen als unübersichtlicher, anarchistischer Raum, in dem zwar alles zu finden ist, aber nicht das, wonach man sucht (*Lost in Cyberspace*). Die bisherigen Lösungsansätze bestanden in der Regel darin, daß man versuchte, die Informationen zentral zu sammeln, zu bündeln und sie „geordnet“ wieder zur Verfügung zu stellen. Demgegenüber sind in den letzten Jahre mit den *Weblogs* und um die Weblogs herum Strategien entstanden, wie verteilte Informationen behandelt und dennoch zugänglich gemacht werden können. Dieser Artikel zeigt auf, was für verteilte Informationssammlungen spricht, wie Weblogs mit über das gesamte Netz verstreuten Informationen umgehen, um dann zu untersuchen, wie die dabei entstandenen Techniken auch auf andere Wissenssammlungen im Internet angewandt werden können. Beispielhaft wird das an der Implementierung einer verteilten Musiknoten-Sammlung aufgezeigt.

## **1 Motivation**

### **1.1 Weblogs und persönliche Wissenssammlungen**

#### **1.1.1 Was sind Weblogs**

Auch wenn manchmal etwas anderes behauptet wird, werden Weblogs als Medienereignis frühestens seit 1999 (in Deutschland nicht eher als 2000) von der

Öffentlichkeit wahrgenommen<sup>1</sup>. Im Gegensatz zu der schon vor 1999 existierenden Tagebuchszene (die ihre Webseiten noch liebevoll mit „handgestricktem“ HTML pflegte), sind Weblogs ohne die dazugehörige Software, wie z.B. Blogger ([www.blogger.com](http://www.blogger.com)), Radio UserLand ([radio.uslerland.com](http://radio.uslerland.com)) oder Movable Type ([www.movabletype.org](http://www.movabletype.org)) nicht zu denken. Erst diese Software ermöglichte es, ständig aktualisierte Webseiten zu erstellen und zu pflegen, ohne allzutief in die Abgründe der HTML-Codierung eintauchen zu müssen. Der Weblogger (kurz: Blogger) konnte sich so ganz dem Inhalt (neudeutsch auch *content* genannt) seiner Seiten widmen.

Und genau das sind Weblogs<sup>2</sup>: Ständig aktualisierte, persönliche Webseiten mit kurzen, oder auch längeren Beiträgen, die umgekehrt zeitlich geordnet sind, d.h. der jüngste Beitrag steht immer ganz oben an erster Stelle. Ältere Beiträge rutschen immer weiter nach unten, bis sie irgendwann endgültig im Archiv verschwunden sind.

Der Fama nach sind Weblogs aus dem Brauch entstanden, sich Hinweise auf interessante Webseiten per Email zuzuschicken<sup>3</sup> und was lag dann näher, statt der Email die persönliche Homepage für diese **häufig, bis zu mehrmals täglich aktualisierten, kommentierten Linklisten** (so ein weiterer Definitionsversuch) zu benutzen.

Über die Bedeutung von **Links** in einem Weblog habe ich schon an anderer Stelle geschrieben<sup>4</sup>, daher hier nur noch einmal so viel:

- Weblogbeiträge verweisen in der Tat häufig auf eine oder mehrere externe Quellen (*verlinken* auf diese Quelle), mehr noch: Der Hinweis auf diese Seite(n) ist oft der (einzige) Grund für diesen Weblogbeitrag.
- Die Kommentare zu diesem Links sind persönlich und oft sehr sarkastisch. Ausgewogenheit und *political correctness* darf man von einem Blogger nicht erwarten<sup>5</sup>.

Die Weblogsoftware – in der Regel sind das kleine Content-Management-Systeme (CMS) – erlaubt ein schnelles und einfaches Publizieren der Web-

---

<sup>1</sup>Der Streit darüber, wer das erste Weblog führte und ob nicht etwa schon das Tage- oder Logbuch von XX ein Weblog gewesen sei, erinnert mich immer ein wenig an den in den 70er Jahren erbittert durchgeführten Versuch, Science-Fiction-Literatur für die Hochkultur zu legitimieren, indem man nachträglich die Utopien von Platon, Morus, Campanella und Swift zur Science-Fiction-Literatur erklärte. Und so bringt es auch nichts, Slashdot ([www.slashdot.org](http://www.slashdot.org)) nachträglich zu einem Weblog zu ernennen.

<sup>2</sup>Auch wenn natürlich niemand die Definitionshoheit über den Begriff *Weblog* besitzt.

<sup>3</sup>Allerdings habe ich bis heute noch keinen Beleg für diese vielkolportierte Behauptung gefunden.

<sup>4</sup>Vgl. [Kantel 2002]

<sup>5</sup>Dies gilt natürlich in besonderem Maße für mein eigenes Weblog, den Schockwellenreiter.

logbeiträge. Dabei kann man drei verschiedene Typen von Weblog-Software unterscheiden<sup>6</sup>:

- Weblogsoftware, die auf einem zentralen Server läuft und dessen Server die Weblogs hostet. (Blogger, Manila, Antville.)
- Weblogsoftware, die als CGI-Script auf dem – meist bei einem Internet Service Provider (ISP) gemieteten – Webspaces des Bloggers läuft<sup>7</sup>. (Movable Type, pMachine, Sunlog, Bloxom und viele andere, meist PHP-basierte Weblog-Tools.)
- Weblogsoftware, die auf dem Rechner des Benutzers installiert ist und statische Seiten erzeugt, die via FTP auf dem – ebenfalls meist angemieteten – Webspaces bei einem ISP heruntergeladen werden. Diese Variante hat den Vorteil, daß beim ISP keine weitere Software installiert sein muß. (Radio UserLand)

Auf Basis dieser Grundfunktionalität (Beitrag schreiben und Publizieren) entstanden im Laufe der letzten zwei bis drei Jahre zahlreiche zusätzliche Funktionalitäten und Techniken, die das Publizieren erleichterten und erweiterten, ohne die grundsätzliche Einfachheit des Publizierens eines Weblogs aufzugeben.

Und genau darum geht es hier in diesem Beitrag: Ich möchte untersuchen, ob und wieweit diese Grund- und Zusatzfunktionen auch genutzt werden können, um Wissenssammlungen, insbesondere verteilte Wissenssammlungen, ebenso einfach wie Weblogs zu erstellen und zu pflegen.

### 1.1.2 Persönliche Wissenssammlungen

Weblog-Software wurde sehr schnell zu mehr genutzt, als nur kommentierte Linksammlungen anzulegen. Da die meisten der Weblog-Tools auch die Möglichkeit boten, neben den täglichen Nachrichten auch sogenannte „Stories“, das sind längere Beiträge, die unter einer eigenen URL abgelegt werden können, anzulegen, entstanden sehr schnell auch „Features“, längere Beiträge zu einem Thema.

Desweiteren gibt es in den meisten Weblog-Tools die Möglichkeit, Beiträge „Kategorien“ (Schlag- und/oder Stichworte) zuzuordnen und sie unter diesen Kategorien abzulegen.

---

<sup>6</sup>Vgl. [Doctorow et al. 2002]

<sup>7</sup>Dies setzt Zugriff auf das `cgi-bin`-Verzeichnis des Servers und die Installation mindestens einer Scriptsprache (Perl, PHP, Python) dort voraus, ein Service, den nicht jeder ISP bietet und der in der Regel auch mehr Geld kostet.

Außerdem bot zumindest zeitweise mindestens ein Weblog-Tool (Radio-UserLand zusammen mit Manila) die Möglichkeit, sich Directories (das sind Linksammlungen im Yahoo-Stil) anzulegen.

So wuchsen viele der ursprünglich einfachen Weblogs sehr schnell zu kleinen, oft liebevoll gepflegten persönlichen Wissenssammlungen zusammen.

### 1.1.3 Weblogs und Wikis

Wikis<sup>8</sup> liegen mir ihrem Verlinkungsmechanismus noch am nächsten zur ursprünglichen Hypertext-Idee. Obgleich Wikis eigentlich für alle offene Webseiten sind, die jeder editieren und so Texte ändern und/oder hinzufügen kann, sind in der Realität viele Wikis eher persönliche Informations- und Wissenssammlungen, die von einer Person oder wenigen Personen gepflegt werden – *Personal Information Management Software (PIM)*. (Was ich für eine durchaus legitime Anwendung von Wikis halte.)

Wikis stehen nicht in Konkurrenz zu Weblogs, können aber in vielen Fällen eine sinnvolle Ergänzung zu einem Weblog als „Persönliche Wissenssammlung“ sein. Wie oben schon angedeutet, gehören auch längere Beiträge in eine Wissenssammlung. Frontier ([frontier.userland.com](http://frontier.userland.com)), die Software, auf die Manila basiert, bot schon sehr früh mit dem automatischen Anlegen von *Shortcuts* für Beiträge Wiki-ähnliche Möglichkeiten der Verlinkung, noch weiter in die Richtung „längere Beiträge mit einem Weblog kombinieren“ ging dann Radio Userland<sup>9</sup> mit dem „eingebauten“ CMS. Vielfach wird aber auch einfach ein Wiki parallel zu einem Weblog betrieben.

## 1.2 (Wissenschaftliche) Sammlungen und Publikationen im Internet

Es folgen einige Beispiele von dem, was ich unter „Wissenssammlungen“ verstehe. Die Beispiele wurden auch unter dem Gesichtspunkt ausgewählt, wie und ob sie sich für eine verteilte Wissenssammlung eignen.

### 1.2.1 Das Projekt Gutenberg

Das Projekt Gutenberg-DE ([gutenberg.spiegel.de](http://gutenberg.spiegel.de)) ist eines der ambitioniertesten Digitalisierungsprojekte im deutschsprachigen Internet. Es enthält Texte in deutscher Sprache von mehr als 400 Klassikern und über achtzig heutigen Autoren. Das Archiv enthält zur Zeit ungefähr 50.000 HTML-Dateien

---

<sup>8</sup>Den ultimativen Text zu Wikis liefert [Mattison 2003]

<sup>9</sup>Zu Radio UserLand und die Möglichkeiten, die diese Software über das Führen eines Weblogs hinaus bietet, vgl. [Kantel 2002a].

(250.000 Textseiten), 850 MB an Text- und Bildmaterial, 10.800 Gedichte, 1.300 Märchen, 1.000 Fabeln, 3.000 Sagen, 1.200 vollständige Romane, Erzählungen, Novellen und 7.300 Bilder. Es wurde 1994 als Freizeitprojekt begonnen, als es nur wenige deutschsprachige Texte im Internet gab. Bis heute sind mehrere zehntausend Arbeitsstunden für das Projekt vom Gutenberg-Team aufgebracht worden. Außerdem haben viele weitere Freiwillige Text- und Bildmaterial für dieses Archiv beigesteuert. Es ist heute einer der ersten Adressen, wenn es um die Quellen deutschsprachiger Literatur im Internet geht. Das belegen die ca. 1,5 Millionen Pageviews und ca. 4,5 Millionen abgerufenen Dokumente im Monat. Laut Eigenaussage betrug das abgerufene Datenvolumen seit 1977 mehr als ein Terrabyte.

Und hier liegt eines der Probleme: Im Jahre 2001 wurde das Bereitstellen des Servers für das Projekt dem bisherigen Sponsor, der AOL, zu teuer und es wurde kurzerhand abgeschaltet. Zwar fand sich nach kurzer Zeit der Spiegel-Verlag bereit, daß Projekt weiterzuführen, doch unter der URL des Spiegel-Verlags. Das kam einer mittleren Katastrophe gleich. Zehntausende von Links, die auf die Seiten des Projekts Gutenberg verwiesen, zeigten auf einmal ins Leere.

### 1.2.2 Perseus-Projekt

Während das Gutenberg-Projekt im Prinzip nur mit einer Art von Daten arbeitet, nämlich mit eingescannten und via Optical Character Recognition (OCR) maschinenlesbar gemachten Daten, ist das Perseus-Projekt (URL?) bedeutend umfangreicher. Es hat sich nicht mehr und nicht weniger vorgenommen, als die gesamte Kultur des antiken Griechenlands online zur Verfügung zu stellen. Das Perseus-Projekt bietet Texte der griechischen Klassiker, Abbildungen von antiken Kunstwerken, Landkarten etc. Dafür hat es in seiner langen Geschichte schon einmal das Medium wechseln müssen. Damals als gigantischer HyperCard-Stack angefangen, wechselte es mit dem Aufkommen des WWW dorthin und wurde unter einem gewaltigen Aufwand von Perl-Skripten ins Internet gestellt.

Neben den eigentlichen Materialien stellt das Perseus-Projekt auch Hilfsmittel zur Erschließung dieses Materials zur Verfügung: Es gibt eine große Anzahl von Wörterbüchern, Übersetzungshilfen etc., die mit Hilfe eines morphologischen Analyzers auch online genutzt werden können.

**Exkurs:** Es wäre interessant, einmal zu erforschen, wie viele der heute noch existierenden Web-Projekte ihren Ursprung in HyperCard-Stacks hatten. Ich glaube, daß es ziemlich viele sind.

### 1.2.3 The Virtual Laboratory of Physiology (VLP)

Das 1997 von Sven Dierig und mir konzipierte „Virtuelle Laboratorium für Physiologie“ ([vlp.mpiwg-berlin.mpg.de](http://vlp.mpiwg-berlin.mpg.de)) ist ein Projekt, das einen begrenzteren Radius besitzt, als die vorherigen Projekte. Es erforscht die „Wurzeln“ der Physiologie von etwa 1870 bis 1930. Ähnlich wie das Projekt Gutenberg begann es als kleines, eher in den Bereich „private Wissenssammlung“ gehörendes Projekt, wuchs aber – auch bedingt durch die von außen kommende, positive Resonanz – sehr schnell darüber hinaus. Im Gegensatz zum Projekt Gutenberg war das VLP aber von vorneherein darauf ausgelegt, mit heterogenen Medien zu arbeiten. Neben der „klassischen“ Bibliothek, die aus gescannten Images der Literatur besteht, waren von Anfang an Zeichnungen von Experimenten, Lebensläufe der Wissenschaftler, Beschreibungen der Labore etc. Bestandteil der Sammlung. Zur Zeit wird das VLP im Rahmen des von der VW-Stiftung geförderten Projektes „Experimentalisierung des Lebens“ am Berliner Max-Planck-Institut für Wissenschaftsgeschichte weitergeführt und -entwickelt.

### 1.2.4 Das eDoc-Projekt des ZIM

Der eDoc-Server<sup>10</sup> des *Heinz Nixdorf Zentrums für Informationsmanagement (ZIM) in der Max-Planck-Gesellschaft* ist als *institutional repository*, das die gesamten Forschungsergebnisse der Max-Planck-Gesellschaft (MPG) und ihrer Institute sammelt, archiviert und präsentiert. Es soll als Schaufenster für die wissenschaftlichen Leistungen der MPG dienen. Darüber ist eDoc als technologische Plattform konzipiert, um das Internet und seine ganzen Möglichkeiten für die wissenschaftliche Kommunikation und Information den Max-Planck-Instituten nutzbar zu machen. Es sollen Dokumente und Informationen sowohl institutsintern als auch in Forschungs Kooperationen gemeinsam verwaltet werden können. (Dokumenten-Server)

Die grundsätzliche Organisation von Dokumenten und Material erfolgt auf eDoc in sogenannten *Collections* (Sammlungen), die von Mitarbeitern von Max-Planck-Instituten angelegt, gepflegt und verwaltet werden.

Ich halte das eDoc-Projekt für ein sehr spannendes Projekt, weil es in seiner Grundidee meiner Vision von einer verteilten Wissenssammlung am nächsten kommt. Zwar wird das eDoc-Material (noch) auf einem zentralen Server gespeichert, für das Einspeisen und Pflegen sind jedoch die Wissenschaftler an den beteiligten Instituten zuständig.

Außerdem haben die Mitarbeiter des ZIM ausführlich zu den Problemen der Metadaten von Sammlungen Überlegungen angestellt. Ich komme darauf

---

<sup>10</sup>Dieser Abschnitt folgt weitgehend der Einleitung aus [ZIM 2003].

in dem Abschnitt über Metadaten noch einmal zurück.

## 1.3 Das Client-Server-Dilemma

### 1.3.1 Das Manila-Syndrom

Manila ([manila.userland.com](http://manila.userland.com)) ist eine Weblog-Server-Software, die auf den Application-Server Frontier ([frontier.userland.com](http://frontier.userland.com)) aufsetzt. Beides, Manila wie auch Frontier stammen aus dem Hause Userland ([www.userland.com](http://www.userland.com)). Um seine Software zu promoten, unterhielt Userland einen kostenlosen Hosting-Service für Manila-Weblogs auf einem (oder sogar mehreren) seiner Server. Als der Traffic so überhand nahm, daß ein vernünftiges Bloggen nicht mehr möglich war, veranstaltete Dave Winer (CEO von UserLand) mal wieder einen seiner berühmten *Cornerturns* und erklärte das zentrale Hosten von Weblogs für erledigt. Sein neues Produkt hieß Radio UserLand ([radio.userland.com](http://radio.userland.com)), war „Weblogging on your desktop“ und erlaubte es, statische Seiten auf einem Server der Wahl abzulegen. Nur die notwendigsten Funktionen wurden noch zentral verwaltet, die da wären

- der „Ping“ an [Weblogs.com](http://Weblogs.com)
- der „Radio Community Server“ (RCS)
- Syndication mittels RSS

### 1.3.2 Antville

Zur Zeit leidet ein weiterer kostenloser Hoster, das österreichische, auf den *Helma Application Server* basierende Antville ([www.antville.org](http://www.antville.org)) unter dem *Manila-Syndrom*. Dies wurde kurzfristig mittels einer vielbeachteten Spendenaktion behoben, bei der die User knappe Euro 3.500,- aufbrachten, um einen neuen, leistungsfähigeren Server zu finanzieren. Langfristig ist dies m.E. jedoch keine Lösung – jede Client-Server-Architektur wird über kurz oder lang unter dem ständig wachsenden Traffic zusammenbrechen.

Bei Antville kommt m.E. verschärfend hinzu, daß – soweit ich informiert bin – Java-Servlets die Grundlage der Applikation bilden. Und Servlet-basierte Architekturen litten schon immer unter einem massiven Ressourcen-hunger.

Beide Beispiele zeigen, daß die klassische Client-Server-Architektur aus der Welt der begrenzten, lokalen und überschaubaren Netzwerke sehr schnell ihre Grenzen erreicht, wenn sie diese lokale Umgebung verläßt und auf die Gegebenheiten des World-Wide-Web übertragen wird.

### 1.3.3 One Point of Failure

Ein von henso ([www.henso.com](http://www.henso.com)) geäußertes Argument (URL?), daß ja auch große Hoster alles auf einen Server setzen und dies dann doch funktioniert, zieht meines Erachtens nicht. Das Beispiel der beiden großen deutschen Hoster zeigt dies: Während Strato ([www.strato.de](http://www.strato.de)) alles auf eine Karte, sprich einen Server setzt, setzt Puretec (1&1, [www.1und1.de](http://www.1und1.de)) bewußt eine Serverfarm von mindestens 1.200 Linux-Servern ein. Und auch Strato verfügt meines Wissens über mehrere Sun-Enterprise-Maschinen, die so redundant ausgelegt sind, daß so ziemlich jedes Teil bei laufendem Betrieb ausgewechselt werden kann, ohne daß irgendein Verlust auftritt. Trotzdem fallen die Ausfälle bei Strato häufiger auf, da bei einem tatsächlichen Serverstillstand alle Kunden betroffen sind. Fällt dagegen bei 1&1 ein Server aus, sind nur etwa 1/1200 Teil der Kunden betroffen.

## 1.4 Konformität vs. Individualität

Es gibt etliche Literatursammlungen im Internet, viele werden von großen Institutionen getragen. Andere entstehen aber auch aus privater Initiative, wie z.B. die großartige Quellensammlung zur Biologiegeschichte von Kurt Stüber ([www.mpiz-koeln.mpg.de/~stueber/stueber\\_library.html](http://www.mpiz-koeln.mpg.de/~stueber/stueber_library.html)).

Viele dieser Sammlungen haben ein liebevolles und individuell gestaltetes Design, daß von den Betreibern sicher nur ungern zu Gunsten einer zentralen, konformen Sammlung aufgegeben wird. Paradebeispiel hierfür sind für mich die überall im Internet gepflegten Sammlungen mit den Texten der Songs der beliebtesten Pop-Artisten (übrigens auch ein gutes Beispiel für verteilte Sammlungen – ich komme später noch einmal darauf zurück).

Daher scheitern viele, auch ambitionierte Web-Projekte oft daran, daß die Macher ihrer bestehenden Projekte aufgeben und sie zentral auf einem Server neu installieren müßten. Das bedeutet, alle bisherigen Projekte werden vereinheitlicht, bekommen ein einheitliches Layout und die Individualität der verschiedenen Macher verschwindet darin. Neben Fragen des persönlichen Geschmacks spielen hier z.B. auch Fragen der *Corporate Identity* der Organisation (auch Wissenschaftsorganisationen haben eine oder sollten eine haben) oder des Zuwendungsgebers eine Rolle und können im Zweifelsfalle solch ein Gemeinschaftsprojekt scheitern lassen.

Auch hier wäre es sinnvoller, Bedingungen zu entwerfen, unter denen die Projekte wie gewohnt weiterlaufen könnten und trotzdem ein gemeinsames Projekt dabei herauskommt. Oder, wie Jürgen Renn schrieb:

**Information will become a patchwork-like structure.** The forms of scientific representation will radically change as a con-



sequenece of the information revolution. There is no reason why future scientific representations should take the forms of journals or books, forms that are largely determined by the print medium and the respective agents. Instead they may take any form suitable as a contribution to a global scientific information network and its structure.<sup>11</sup>

Und konsequent weitergedacht, besteht auch kein Grund mehr, daß Informationen und (digitalisierte) Materialien zentral verwaltet und gelagert werden müssen.

## 2 Die Werkzeuge der Weblogger

Weblogger neigen dazu, eine Community, einen sozialen Organismus<sup>12</sup> zu bilden. Warum das so ist, ist bisher noch nicht untersucht worden. Ohne Zweifel tragen aber die oben erwähnten Werkzeuge dazu bei. Über Weblogs.com und den Community-Server werden die einzelnen Blogger informiert, welche Blogs neuen Content ins Netz gestellt haben und welche Blogs zu *ihrer Community* gehören, mittels *Syndication* werden die Updates der abonnierten Blogs frei Haus geliefert.

### 2.1 CMS und Template-Mechanismen

Template-Mechanismen stammen aus den Vorbildern der Weblogtools, den „großen“ *Content Management Systemen (CMS)*. Sie ermöglichen im Wesentlichen die logische Trennung von Bearbeitungs- und Präsentationsschicht, von Inhalt (Content) und Design.

### 2.2 XML

Weblog-Tools sind *Content Management Systeme* und wie fast alle CMS basieren sie auf XML. Das klingt trivial, daraus folgt aber eine wichtige Konsequenz: Zumindest potentiell liegen alle Daten als reine Nutzdaten vor, befreit vom Ballast der Layout-Informationen<sup>13</sup>.

---

<sup>11</sup>[Renn 2000], Seite 3

<sup>12</sup>Vgl. [Kantel 2002]

<sup>13</sup>Einige Weblog-Tools benutzen ein anderes – aber ebenfalls strukturiertes Format – zur Speicherung der einzelnen Artikel. Jedoch ist eine Wandlung nach XML aus diesen Formaten in der Regel trivial.

Nur dadurch ist es möglich, daß die Inhalte von anderen Clienten plattform- und sprachenunabhängig „abgegriffen“ und in anderen Applikationen weiterverarbeitet werden können.

Das bekannteste Beispiel der Wiederverwendung dieser XML-Daten ist RSS (siehe dort).

XML wurde ursprünglich als „abgespeckte“ Version für das von vielen als zu unhandlich und zu kompliziert empfundene SGML für das Web entwickelt. Leider droht XML an seinem Erfolg zu ersticken. Mit wachsender Popularität entdeckten immer mehr das Format und begannen, ihre eigenen Erweiterungen zu entwickeln. Mit dem Effekt, daß XML heute von vielen als noch komplizierter empfunden wird als SGML. Daher ist bei XML-basierten Entwicklungen darauf zu achten, daß der Einfachheit den Vorzug gegeben wird. Die meisten Weblogtools halten sich daran. Soweit ich weiß, gibt es für kaum eine RSS-Version eine DTD, sondern nur verbale Beschreibungen. Weblog-Tools nutzen XML-RPC (URL?) (siehe nächsten Abschnitt) statt dem bedeutend komplizierteren SOAP (URL?).

## 2.3 XML-RPC

Die meisten der Bloggertools kommunizieren über XML-RPC miteinander. XML-RPC steht für *Remote Procedure Call auf Basis XML* und ist ein einfaches, aber trotzdem effektives Protokoll, um Prozeduren über Betriebssystem- und Sprachgrenzen hinweg aufzurufen.

[=> Weiter ausführen!]

### 2.3.1 Weblog API

### 2.3.2 Meta Weblog API

### 2.3.3 Und was ist mit SOAP?

## 2.4 Der „Ping“

Der *Ping* an Weblogs.com ist der zentrale Punkt des oben angesprochenen *cornerturns* von Dave Winer. Schon vorher hatten sich die Weblogs, die das wollten (und das waren fast alle) bei Weblogs.com eingetragen und Weblogs.com selber fragte im Stundenrhythmus die eingetragenen Weblogs ab, ob es ein *Update* gibt. Schon nach wenigen Monaten hatten sich soviele Weblogs eingetragen, daß der Server mit der stündlichen Nachfrage nicht mehr mitkam, die Abfrage war noch nicht beendet, aber die Stunde schon vorbei und der Server mußte mit einer neuen Abfrage beginnen.

Die Lösung war, daß nicht mehr der Server die Clients – also die Weblogs – abfragt, sondern die Clients dem Server mitteilen, wann es bei ihnen etwas Neues gibt. Der *Ping* war geboren. Er ist eine einfache XML-RPC-Prozedur, die dem Server nur mitteilt, daß es bei diesem Weblog etwas Neues gibt.

Die Umsetzung innerhalb der Community erfolgte erstaunlich schnell. Obwohl UserLand selber nur seine eigene, proprietäre Scriptsprache *UserTalk* unterstützte, waren schon nach wenigen Tagen XML-RPC-Pings für alle gängigen und weniger gängigen Programmier- und Scriptsprachen vorhanden. Es gibt heute meines Wissens kein Weblog-Tool, das den Ping an Weblogs.com *nicht* unterstützt.

#### 2.4.1 Andere „Pings“

Auch von anderen wurde das Prinzip des „Anpingens“ übernommen. Während blo.gs (URL?) eine schlichte Kopie des Weblogs.com-Pings ist, haben andere tools das Ping-Prinzip verfeinert:

#### 2.4.2 Trackback

#### 2.4.3 Blogrolling

#### 2.4.4 Erweiterungen der bestehenden Werkzeuge.

Der Ping an Weblogs.com war eine geniale Idee. Immer mehr Werkzeuge entstehen, die auf diesem Ping aufbauen. So z.B. Blogdex.com mit dem Versuch, ein erweitertes *Ranking* einzuführen und Technorati.com, die versuchen, so etwas wie einen *Blogger-Cosmos* aufzuzeigen.

Eine dritte Entwicklung, die aber direkt nicht von Weblogs.com beeinflusst wurde, trotzdem für meine Argumentation wichtig ist, nämlich die *Community- oder Kollektiv-Blogs* werden in einem späteren Abschnitt behandelt.

### 2.5 Der „Community Server“

Der „Community-Server“ hat eine ähnliche Funktion wie Weblogs.com. Nur daß er nicht die Updates *aller* Weblogs anzeigt, sondern nur die Updates aller Weblogs, die zu einer *Community* gehören. Grob läßt sich der Begriff *Community* erst einmal damit übersetzen, daß zu dieser Community alle Weblogs gehören, die vor dem *Cornerturn* auf einem Server zentral untergebracht waren. Das dies die Community-Bildung fördert, sieht man heute noch an den Antville-Blogs auf [www.antville.org](http://www.antville.org) und [www.twoday.net](http://www.twoday.net). [=> Erläutern!]

[Exkurs: Community/Communities auf den Antville-Server(n)]

Erst einmal hat dieser Server tatsächlich nur die Funktionen, die eine klassische serverbasierte Weblog-Software anbietet: Rankings, Last Update, Most Read Sites etc. Doch zwei weitere Entwicklungen beginnen sich abzuzeichnen.

Hier ist/sind die derzeitigen Implementierungen noch inkonsequent. Statt daß der *Ping* an den Community-Server geht, der ihn dann an Weblogs.com durchreicht, geht ein Ping an Weblogs.com und über einen zweiten XML-RPC-Mechanismus wird der Community-Server über ein Update informiert. Hier wäre folgende Hierarchie anzustreben, die auch wieder Weblogs.com entlasten würde:

[Zeichnung]

### 2.5.1 Anforderungen an einen Community-Server

#### XML-RPC

#### Open Source

### 2.5.2 Communities, die durch gemeinsame Interessen gebildet wurden.

Pionier im deutschsprachigen Raum war in diesem Falle sicher das *Bloghaus* ([www.blogworld.de](http://www.blogworld.de)), das sich als *Gemeinschaft der deutschsprachigen Weblogs* verstand. Während das einzige gemeinsame Interesse bisheriger Communities in der Nutzung der gleichen Software bestand, entstand hier etwas völlig Neues, eine Community, die sich – emanzipiert von der Software – durch andere, gemeinsame Interessen definiert.

Interessant dabei ist, daß diese Communities im Prinzip keine neuen Tools benötigen. Der Weblog-Monitor des Bloghauses bedient sich bei Weblogs.com, nur daß er eben nur die Updates der im Bloghaus angemeldeten Weblogs zeigt.

## 2.6 Syndication mittels RSS

### 2.6.1 Push vs. Pull

Bei dem oben erwähnte Ping an Weblogs.com erfährt der Benutzer – wenn er sich auf die Seite von Weblogs.com begibt – zwar, daß es auf einer Seite resp. in einem Weblog etwas Neues zu lesen gibt, aber nicht, was es Neues zu lesen gibt. Immer noch muß der Benutzer, um etwas über den Inhalt der Seite zu erfahren, diese aktiv ansurfen und sich den Inhalt auf seinen Rechner laden (*to pull*).

Ein anderer Weg, der – natürlich mit unterschiedlichen Techniken und Formaten, wie sollte es anders sein - schon vor einigen Jahren von Netscape und Microsoft beschritten wurde, ist der, daß der Anbieter selber informiert, wenn sich seine Seite geändert hat. Er stellt eine Datei ins Netz, die der Webbenutzer via http (dem Standardprotokoll der Browser) abrufen kann und die ihn informiert, ob sich die Seiten geändert haben. Als Format dafür bot sich XML an. Spezielle Webserver (Aggregatoren) sammeln diese Informationen und stellen sie dem Benutzer (konfigurierbar, d.h. aus den gesammelten Informationen stellt sich der Surfer seine Auswahl zusammen) zur Verfügung. Aus historischen und marketing-technischen Gründen fangen diese Server meist mit *my* an, wie z.B. myNetscape.com oder myUserLand.com. Der Benutzer bekommt die Daten gewissermaßen geliefert (*to push*) und muß nicht mehr selber nach den aktuellen Seiten suchen. Die einzelnen Seiten, die auf den Aggregatoren zusammengefaßt werden, nennt man *channel*.

Syndication-Formate sollten ursprünglich nur die Headlines und ggf. einen *Abstract* der Nachrichten zeigen. (Der User soll schließlich - nachdem er sich überzeugt hat, daß die Information ihn interessieren könnte, auf die eigenen Seiten gelockt werden.) Die meisten heutigen RSS-Feeds leisten mehr. So liefert z.B. Radio UserLand von Anfang an den kompletten Artikel incl. aller Bilder und Links im RSS-Feed mit<sup>14</sup>. Das erlaubte einmal den Bezieher des Feeds, diesen Channel vollständig in seiner Website zu integrieren.

## 2.6.2 RSS

Damit nun die Aggregatoren die Dateien überhaupt lesen konnten, mußte man sich auf ein gemeinsames Format einigen - und dieses Format ist RSS (Rich Site Summary)<sup>15</sup>.

Die spannendere Geschichte ist aber, daß dieses Format es dann ermöglicht, den Inhalt pur auch auf der Website des Client für weitere Informationen, Spider und ähnliches abzulegen.

[=> Erläutern!]

---

<sup>14</sup>Diese Art des RSS-Feeds ist allerdings umstritten [=> erläutern]

<sup>15</sup>Zu Aufbau und Funktion einer RSS-Datei vgl. auch [Kantel 2001]

## 2.7 (re)Structured Text

## 3 Peer-to-Peer-Netzwerke

## 4 Metadaten

Eine Diskussion über Metadaten hat in der Weblog-Welt bisher kaum stattgefunden, obwohl sie an einigen Stellen – wenn auch zaghaft – angestoßen wurde. Dabei existieren durchaus schon rudimentäre Metadaten in den Headern der RSS-Feeds. Mit der weiteren Verbreitung und auch Spezialisierung von Weblogs ist damit zu rechnen, daß diese Diskussion bald verstärkt in Gang kommen wird.

### 4.1 Wozu Metadaten?

### 4.2 Dublin Core

### 4.3 RDF

## 5 Community-Blogs

Gibt es überhaupt ein Potential unter den Netzbenutzern, die bereit sind, an gemeinsamen Projekten wie einer gemeinsamen, verteilten Wissenssammlung mitzuwirken? Auch hier gibt die Weblogwelt eine Antwort:

[Slashdot, Metafilter, Rollberg, Kollektivblogs]

## 6 Szenario einer virtuellen, verteilten Sammlung

Betrachten wir nun einmal eine Sammlung wie das bereits oben erwähnte *Projekt Gutenberg-DE*, nur daß wir uns vorstellen, daß diese Sammlung verteilt angelegt ist. Jeder Sammler legt die Dateien auf einem Server seiner Wahl ab, zusammen mit einem Satz Metadaten. Diese Metadaten bestehen aus einem Kern, der möglichst klein und unkompliziert zu halten ist, damit die freiwilligen (sic!) Mitarbeiter nicht die Lust verlieren, diesen Metadaten-satz anzulegen. Neben diesem Kern steht es jedem Sammler und jeder Gruppe natürlich frei, weitere Metadaten (in einem anderen *Namespace*) anzulegen, die sie für ihre spezielle Sammlung oder für ihr spezielles Gruppenprojekt benötigen.

Nach Fertigstellung erfolgt ein Ping an einen oder mehrere „Community-Server“. Wenn mehrere Community-Server vorhanden sind – was aus Gründen der Ausfallsicherheit zu empfehlen ist – sollten sich diese Server untereinander synchronisieren. Dies kann ebenfalls mittels eines XML-RPC-Pings geschehen. Der oder die angepingten Server holen sich vom Client den Metadatenatz und fügen ihn in ihren Datenbestand ein.

Anfragen an das Projekt gehen immer an einen der Community-Server, die mit Hilfe dieses Datenbestandes ein *Redirect* vornehmen und den Anfragenden auf den Server und die Seiten verweisen, auf denen die Informationen tatsächlich liegen.

[=> Load Balancing]

Gleichzeitig bieten die einzelnen Clients einen aus den Metadaten und (XML-) Datensätzen automatisch generierten RSS-Feed an, den einmal die Server zum Generieren von Inhaltsverzeichnissen nutzen können, als auch von den Clients innerhalb und außerhalb der „Community“ als Update-Information bezogen werden kann.

## 6.1 Beispiel-Implementierung einer (verteilten) Notensammlung

## 6.2 Was ist für eine Implementierung notwendig?

[Daten aus XML => Ping => (Metadaten (RDF) => XML-RPC => Community-Server) => führt zu Wissensmanagement. (Community-Blog?)]

## 6.3 Der Client: Statische Seiten vs. dynamische Seiten?

# 7 Verteiltes „Knowledge Management“

Voraussetzung ist, dass der Content als XML/(X)HTML auf der Website des Client vorliegt und nicht in irgendwelchen Datenbanken versenkt wird.

Das von mir aufgezeigte Szenario ist ein Projekt, das im Sinne einer *Informationslogistik* einen universellen Zugriff auf heterogene Wissensressourcen ermöglicht. Für den Content sind nicht die Betreiber der *Community-Servers*, sondern nach wie vor die einzelnen, teilnehmenden (Unter-) Projekte selber verantwortlich. Eine redaktionelle Vor- oder Nachbearbeitung liegt ebenfalls in der Hand der einzelnen Projekte. Technische Bedingung zur Teilnahme an dieser Community ist einzig und allein die Bereitschaft, den Community-Server anzupingen, einen einheitlichen Metadatenatz und einen RSS-Feed

anzubieten. Auf der anderen Seite ist es mehr als eine *Metasuchmaschine* mit einem einheitlichem Interface für Suche und Präsentation, wie z.B. das vom kunstgeschichtlichen Seminar der Berliner Humboldt-Universität initiierte Projekt *Prometheus* [Bredekamp 2002].

=> Wir entwickeln ein Weblogs.com für Wissenssammlungen.

## Literatur

- [Bredekamp 2002] Horst Bredekamp, Ingeborg Reichle: *Prometheus*. Das verteilte digitale Bildarchiv für Forschung und Lehre, in: Humboldt-Spektrum, 9. Jahrgang, Heft 4/2002, S. 48-53
- [chromatic, Aker und Krieger 2002] chromatic, Brian Aker & Dave Krieger: *Running Weblogs with Slash*, Sebastopol (O'Reilly) 2002
- [Dierig, Kantel und Schmidgen 2000] Sven Dierig, Jörg Kantel und Henning Schmidgen: *The Virtual Laboratory for Physiology*, A Project in Digitalising the History of Experimentalisation of Nineteenth-Century Life Sciences *with an enclosed CD-ROM*, Berlin (Max Planck Institute for the History of Science, Preprint No. 140) 2000, ([www.mpiwg-berlin.mpg.de/Preprints/P140.PDF](http://www.mpiwg-berlin.mpg.de/Preprints/P140.PDF)), PDF, 6 MB
- [Doctorow et al. 2002] Cory Doctorow, Rael Dornfest, J. Scott Johnson, Shelley Powers, Benjamin Trott & Mena G. Trott: *Essential Blogging*. Selecting and Using Weblog Tools, Sebastopol (O'Reilly) 2002
- [Grötker 2003] Ralf Grötker: *Dosenware*. DRM aus Leipzig: „Copy Hall“ setzt Bibliotheken unter Druck, Telepolis 2003 ([www.heise.de/tp/deutsch/special/copy/14044/1.html](http://www.heise.de/tp/deutsch/special/copy/14044/1.html)), zuletzt besucht am 19. Februar 2003
- [Hammersley 2003] Ben Hammersley: *Content Syndication with RSS*, Sebastopol (O'Reilly) 2003
- [Ianella 1998] Renato Ianella: *An Idiot's Guide to the resource Description Framework*, 1999 [archive.dstc.edu.au/RDU/reports/RDF-Idiot/](http://archive.dstc.edu.au/RDU/reports/RDF-Idiot/), zuletzt besucht am 31. März 2003
- [Kantel und Dierig 1998] Jörg Kantel, Sven Dierig: „*Hypermediale Wissensanordnung und collagierender Umgang mit dem Material?*“ - *Digitales „Bildersammeln“ am Beispiel des VIPP (Virtual Institute of Physiology Project)*, Vortrag, gehalten auf



- der HyperKult VII am 18. Juli 1998, Universität Lüneburg, ([derschockwellenreiter.editthispage.com/vlp/hypermedia.html](http://derschockwellenreiter.editthispage.com/vlp/hypermedia.html)), zuletzt besucht am 27. April 2003
- [Kantel und Dierig 1999] Jörg Kantel, Sven Dierig: *Zur Verwaltung großer Datenmengen im WWW am Beispiel des VIPP (Virtual Institute for Physiology Project)*, in: Friedbert Kaspar, Hans-Ulrich Zimmermann (Hrsg.): *Internet- und Intranet-Technologien in der wissenschaftlichen Datenverarbeitung, 15. DV-Treffen der Max-Planck-Institut 18. - 20. November 1998 in Göttingen*, Göttingen (GWDG-Bericht Nr. 53) 1999, auch online zu lesen unter [derschockwellenreiter.editthispage.com/vlp/datenmengen.html](http://derschockwellenreiter.editthispage.com/vlp/datenmengen.html), zuletzt besucht am 27. April 2003
- [Kantel 2001] Jörg Kantel: *Was ist RSS?*. Eine Einführung in Content Syndication, ([www.schockwellenreiter.de/webdesign/rss.html](http://www.schockwellenreiter.de/webdesign/rss.html)), zuletzt besucht am 13. Februar 2003
- [Kantel 2002] Jörg Kantel: *Archäologie des Bloggens*. Ein erster und unvollständiger Versuch, dem „Phänomen Weblog“ mittels einer Geschichte seiner einzelnen Teile näher zu kommen ([www.tzw.biz/www/home/article.php?p\\_id=2028](http://www.tzw.biz/www/home/article.php?p_id=2028)), in Christian Eigner (Hg.): *Blogging und die neue Medienkultur des Netzes*, 2002, ([www.tzw.biz/www/home/article.php?p\\_id=2034](http://www.tzw.biz/www/home/article.php?p_id=2034)), zuletzt besucht am 13. Februar 2003.
- [Kantel 2002a] Jörg Kantel: *Radio UserLand*, Bloghaus, März 2002 ([www.blogworld.de/tm.article.php?article\\_id=5](http://www.blogworld.de/tm.article.php?article_id=5)), zuletzt besucht am 27. April 2003
- [Langham 2003] Matthew Langham: *Kleines Format ganz groß*. RSS – Eine Einführung in Content Syndication mit XML, in: *Linux Enterprise 3.03*, S. 81 - 84
- [Mattison 2003] David Mattison: *Quickwiki, Swiki, Twiki, Zwiki and the Plone Wars – Wiki as a PIM and Collaborative Content Tool*, April 2003 ([www.infoday.com/searcher/apr03/mattison.shtml](http://www.infoday.com/searcher/apr03/mattison.shtml)), zuletzt besucht am 8. April 2003
- [Neuburg 1998] Matt Neuburg: *Frontier. The Definitive Guide*, Sebastopol (O'Reilly) 1998

- [Renn 2000] Jürgen Renn: *Challenges of the Information Revolution for the Max Planck Society*, Berlin (Max-Planck-Institut für Wissenschaftsgeschichte, Preprint 151) 2000 ([www.mpiwg-berlin.mpg.de/Preprints/P151.PDF](http://www.mpiwg-berlin.mpg.de/Preprints/P151.PDF)), PDF, 512 KB
- [Renn 2001] Jürgen Renn: *Erwirb es um es zu besitzen: Kulturelles Erbe im Zeitalter der Informationsrevolution*, Berlin (Max-Planck-Institut für Wissenschaftsgeschichte, Preprint 176) 2001 ([www.mpiwg-berlin.mpg.de/Preprints/P176.PDF](http://www.mpiwg-berlin.mpg.de/Preprints/P176.PDF)), PDF, 64 KB
- [Renn 2002] Jürgen Renn [ed.]: *ECHO - an infrastructure to bring European Cultural Heritage Online*, Berlin (Max-Planck-Institut für Wissenschaftsgeschichte, Preprint 191) 2002 ([www.mpiwg-berlin.mpg.de/Preprints/P191.PDF](http://www.mpiwg-berlin.mpg.de/Preprints/P191.PDF)), PDF, 1,4 MB
- [Schatten 2003] Alexander Schatten: *Die offene Wissensgesellschaft und ihre Feinde* ([www.schatten.info/](http://www.schatten.info/)), Version vom 27. Februar 2003
- [Udell 1999] Jon Udell: *Practical Internet Groupware*, Sebastopol (O'Reilly) 1999
- [ZIM 2003] Heinz Nixdorf Zentrum für Informationsmanagement (ZIM) in der Max-Planck-Gesellschaft: *eDoc Handbuch* ([edoc.mpg.de/3564](http://edoc.mpg.de/3564)), Fassung vom 10. März 2003