

Metadata: Boon or Bane? The Manifesto of Ignorance (Semantic Media Web Berlin 2013)

Am 26. und 27.9. fand in Berlin die "Semantic Media Web" Veranstaltung statt, unter Leitung des Innovationsforums, an dem die Humboldt-Universität zu Berlin, das Deutsche Forschungszentrum für Künstliche Intelligenz GmbH, das W3C-Büro Deutschland/Österreich, die Xinnovations e. V. und die Freie Universität Berlin beteiligt sind. An zwei Tagen ging es Schlag auf Schlag (meistens in 10minütigen Sessions) im weitesten Sinne um das Thema Semantische Technologien und Branchenbeteiligung aus Musik, Film und Verlagen. Ich selbst war (eher zufällig) Opener des Verlags-Blocks und wollte mich in meinem kurzen Vortrag ganz untechnisch der Frage widmen, warum (Buch)Verlage mit den "Daten"-Themen Probleme haben und worin wir sowohl Aufgabe des universitären Umfelds wie auch der Technologie-Anbieter sehen. Aufgrund der kurzen Zeit konnte nicht mehr als ein Kratzen an der inhaltlichen Oberfläche stattfinden, eine tiefere Analyse ist einem späteren Blogbeitrag vorbehalten. Hier aber zu Vortrag und dem "Manifesto of Ignorance" (heutzutage muss man ja auch etwas reißerischer titeln, sonst wird man nicht gelesen).

Der Redetext:

"...Mein kurzes Statement zu den Themen wie Semantic Web, Data Mining usw soll deswegen diese Herausforderung aus Sicht der deutschen Buchverlage beleuchten und bewerten.

In den folgenden Minuten möchte ich Ihnen dabei einige sehr reale Problemfelder aufzeigen, die Ihnen vielleicht einen Eindruck davon verschaffen, warum all diese wunderschönen Daten-Welten, über die gestern und heute oft euphorisch und einem mitunter unter auch sehr wissenschaftlichem Blickwinkel gesprochen wurde, sowohl Fluch als auch Segen für viele Verlage sind.

Eingangs darf dabei ein Blick darauf nicht fehlen, woher wir Verlage kommen und was wir tun. Viele von uns leben in einer physischen Welt der schön gedruckten Bücher, die wir zusammen mit Autoren mühsam zur Welt bringen und über den immer noch vorhandenen Buchhandel an Kunden verkaufen, die wir nicht kennen. Unsere Produktions-, Wertschöpfungs- und Distributionsketten sind dabei linear und zweidimensional, nicht vernetzt und dreidimensional.

Oder, um es verkürzt auszudrücken, was den Konflikt zwischen Ihrer und meiner Welt angeht: Sie sprechen über Data Mining – aber wir sind keine Minen-Arbeiter. Verlage sind Geschichtenerzähler.

Durch den Einzug der Digitalisierung, des Internets, in unsere eigenen Wertschöpfungsketten, aber auch in die täglichen Lebenswelten unserer Kunden etwa bei der Beschaffung von Gütern zeigt sich ein radikaler Wandel. Diesen Wandel will ich hier in drei Themenkreisen kurz skizzieren.

Der erste Wandel betrifft nicht einmal so sehr unsere Produkte, auch wenn E-Books, Datenbanken und vereinzelt auch Apps unser Produktportfolio erweitern. Der erste Wandel betrifft die Frage, wie Kunden zukünftig unsere Produkte finden, wie diese Produkte auch immer aussehen mögen. Unsere Kunden werden unsere Produkte zukünftig primär durch Suchmaschinen und Datenbanken finden, und wenn unsere Produkte durch sinnvolle Erschließung auf Ebene der Meta-Daten dort nicht oder nur schwer auffindbar sind werden sie auch keine Käufer finden. Hier geht es um Sichtbarkeit, um Discoverability.

Einfache Standards auf Produktebene sind etwa die ISBN oder die ISBN-A, aber auch ONIX als erste Schritte – aber es ist die Frage ob dies ausreicht in Zukunft.

Der zweite Wandel betrifft unsere Produkte selbst. Durch Digitalisierung erzeugen wir ungeheure Daten-Silos, Unmengen an Content. Um daraus aber sinnvolle Produkte zu generieren, muss dieser Inhalt erschlossen, vertaggt und vernetzt werden. Unsere Inhalte müssen ebenso granularer werden. Erst dann sind Verlage dazu in der Lage, wirtschaftlich sinnvoll, neue, nicht nur digitale, Produkte zu generieren.

Der dritte Wandel neben dem Beschaffungsweg und der Erschliessung unserer Inhalte aber betrifft den eigentlichen Mittelpunkt unserer Aktivitäten: den Leser, den Kunden. Es reicht nicht mehr, mit verlegerischer Erfahrung einem unmündigen Kunden Produkte anzubieten. Natürlich will dieser Kunde auch weiterhin überrascht werden durch nutzwertige und schöne Produkte. Aber um dies zu erreichen müssen wir mehr über ihn wissen, um zielgenauer

produzieren zu können im Zeitalter der Mikro-Interessen und der Mikro-Zielgruppen. Datenerhebung und Data Mining sind hier die Schlüsselwörter. Das Prinzip des einen Produkts für eine möglichst grosse Zielgruppe funktioniert nicht mehr. Und wir müssen auch mehr über unsere Kunden wissen, um mit ihnen sinnvoll kommunizieren zu können, denn auch die direkte Kommunikation zwischen Produzent und Käufer wird wichtiger.

Discoverability, Erschliessung unserer Inhalte und das angereicherte Wissen über unsere Kunden sind also die drei wesentlichen Problemfelder der Zukunft.

Aber das Wissen von Problemen ist nicht gleich deren Beseitigung. Die meisten deutschen Buchverlage leben in einer Produkt-Tradition, die noch auf Gutenberg zurückgeht.

Alles Wissen über die Herausforderungen der Zukunft hilft auch nichts, wenn man wie ein Frosch in einem sich langsam erhitzenen Topf voller Wasser sitzt. Es gibt keine abrupte Erhitzung, keine abrupte Disruption. Wir erleben zumindest in Deutschland in den Buchverlagen eine Zeit der Evolution, nicht der Revolution. Der Frosch wird sitzen bleiben, bis ihn die langsam steigende Wassertemperatur nach längerer Zeit schließlich umbringt, und wir Buchverlage werden in Teilen so lange die alten, bekannten Produkte in einen immer kleineren Markt hinein verkaufen, bis dies nicht mehr rentabel ist. Geringer Veränderungsdruck ist der Feind jeder Innovation.

Dazu kommen mentale wie investitionstechnische Probleme. Selbst einfachste Datenstrukturen können eine Herausforderung sein, wenn man nie in Datenstrukturen sondern gedruckten Produkten gedacht hat. Und es gibt strukturelle wie investitionstechnische Herausforderungen – oft haben Verlage nicht die richtigen Leute, um solche Themen anzugehen. Und für die mit dem nötigen KnowHow ist die Verlagsbranche oft unattraktiv.

Zudem sind solche Unternehmungen mit Investitionen verbunden, die auch eine Investition in eine unsichere Zukunft sind. Das mag banal klingen, aber viele werden die Investition in die Granulierung und semantische Erschließung ihrer Inhalte scheuen, wenn noch gar nicht klar ist, über welche zukünftigen Produkte wir reden. Und wir reden auch davon, dass der Großteil der deutschen Verlage eben kleine und mittlere Verlage sind, keine Konzerne, die ganze Abteilungen mit üppigen Budgets vorweisen können.

Was aber lässt sich für die Zukunft der deutschen Buchverlage voraussagen? Zumindest, dass es einige geben wird, die auch in vielen Jahren noch unbeirrt das schöne gedruckte Buch herstellen. Aber dieser Markt wird kleiner und viele dieser Verlage werden verschwinden. Aber einige werden sich den vorher genannten Problemfeldern zuwenden, sich diesen stellen – und dazu brauchen sie Ihre Hilfe.

Wir brauchen Hilfe aus dem universitären Umfeld – schaffen und etablieren Sie praxistaugliche Standards und erklären Sie uns diese. Und zwar in einfachem Deutsch/plain english. Noch einmal: wir sind Geschichtenerzähler, keine Informatiker.

Untersuchen Sie in der Theorie, welche digitalen Geschäftsmodelle im Kontext von Daten, von semantic Web und Meta Data sinnvoll sind – die Überprüfung auf Umsetzbarkeit in Hinblick auf den Kunden übernehmen wir.

Für die Umsetzung, für die digitalen Produkte der Zukunft brauchen wir Tools und Technik, und heute sind ja auch technologische Dienstleister anwesend. An diese ist die Bitte gerichtet, praxistaugliche Tools zu entwickeln, die auch finanzierbar sind. Denn technologische Umsetzung gehört nicht zur Kernkompetenz von Verlagen, dies muss durch Sie passieren.

Wenn wir Verlage Standards haben, Geschäftsmodelle kennen und vernünftige Tools an die Hand bekommen, dann haben wir auch in Zukunft eine solide Chance.”

english version

“My short statement on topics such as semantic web, data mining, etc. shall therefore investigate and evaluate this challenge from the perspective of German publishers.

In the next few minutes I would like to talk to you about some quite real problem areas, which might give an impression of why all these beautiful data worlds, which have been discussed yesterday and today already, in an often euphoric way and from a sometimes quite scientific point of view, why these data worlds are both boon and bane for many publishers.

First of all we need to take a look at where publishers come from, and what we do. Many of us live in a material world of beautiful, printed books, which we give birth to in cooperation with our authors with great effort. These books we sell all over the world through the still existing booktrade, to customers we do not know. Our production, value and distribution chains are linear and two-dimensional, not networked and three-dimensional.

The long and short of it: the conflict between your world and my world is, that you are talking about data mining – but we’re not mine workers. Publishers are storytellers.

A radical change appears with the appearance of the digitization and the internet in our own supply chains, but also in the daily lifestyles of our customers, for example in the procurement of goods. I will briefly outline this change here in three thematic areas.

The first change doesn’t even affect our products so much, even if e-books, databases, and also occasionally apps expand our product portfolio. The first change concerns the question of how future customers will discover our products, whatever these products may look like. Our customers will in

the future discover our products primarily through search engines and databases. If our products will [be hard to find or not even be found at all despite adequate development on the meta-data level, there will be no buyers. This is about visibility, about discoverability.

Simple standards on the product level are for example the ISBN or the ISBN-A, but also ONIX as a first step – the question is whether this will be sufficient in the future.

The second change relates to our products themselves. In the wake of Digitisation we create enormous data silos, vast amounts of content. But in order to generate useful products thereof, this content must be developed, tagged and networked. Our content, too, must be more granular. Only then will publishers be in a position to create economically sensible, new, not just digital, products.

The third change, next to the supply channel and the development of content, concerns the actual center of our activities: the reader, the customer. It is no longer sufficient to offer products to an immature customer from the experience of the publisher. Of course, the customer expects continued surprises by valuable and beautiful products. But in order to achieve this we need to know more about him, so we will be able to develop accurately targeted products in the era of micro-interests and micro-audiences. Data collection and data mining are the key words here. The principle of one product to fit the largest possible audience no longer works. We also need to know more about our customers to communicate with them in a meaningful way, because direct communication between producer and buyer will become more and more important.

Therefore discoverability, indexing of our content and enriched knowledge of our customers are the three main problem areas of the future.

But problem recognition is not equal to its elimination. Most German publishers live in a product tradition that goes back to Johannes Gutenberg.

All knowledge about the challenges of the future will not help if you sit tight like a frog in a slowly heating pot full of water. There is no abrupt heat wave, no sudden disruption. At least in Germany we are experiencing a period of evolution in book publishing, not revolution. The frog will remain seated until the slowly rising water temperature finally kills it after a long time, and we, the publishers, will in parts keep on selling the old, well-known products in a market smaller and smaller, up to the point where it will be no longer profitable. Low pressure for change is an enemy of innovation.

But there are also mental and investment-wise problems. Even simple data structures can be a challenge if you have never thought in data structures but have always thought in printed products. Furthermore we may find structural challenges and challenges concerning investment – often publishers do not have the right people to tackle issues. And for those who have the necessary know-how, the publishing industry is often not attractive.

In addition, such undertakings are connected with investments, which may be investments in an uncertain future. This may sound obvious, but many will be afraid of investing in the development of their contents granulation and semantic development, as long as it is not clear what future products we are talking about. And we also have to consider that the majority of German publishers are just small and medium-sized publishers, not corporations, who can devote entire departments with lavish budgets to the tasks at hand.

So what is the prediction of the future of German publishers? At the least, that there will be some who will continue producing the beautiful, printed book for many years to come. But this market is small and many of these publishers will disappear. However, some will turn to the previously mentioned problem areas, to address these – and for this they need your help.

We need assistance from the university environment – create and establish practical standards and teach us. And please – in simple German / plain english. Once again, we are storytellers, not computer scientists/IT specialists.

We kindly ask you to examine in theory which digital business models in the context of data, semantic web and meta data are useful and we ourselves will check for feasibility in respect to the customer.

For the implementation, for the digital products of the future, we need tools and technology, and these technological service providers are present today. So the plea is directed to you, to develop practical tools that are also affordable. Because technological implementation is not part of the core competence of publishers, we are in need of your competences.

If publishers will have standards, know different business models, and have adequate tools at our disposal, then we will participate in the future of publishing.”