Sustainability of Environmental Information

Thomas Bandholtz[1]

1 Introduction

This article talks about the sustainability of information. It compares patterns of non-sustainability, resulting in *noise* for information. I discuss reasons for and impacts of sustainable information. Coming from the practical area as the developing project manager of the German Environmental Information Network (GEIN, http://www.gein.de), webservices are examined to be a solution. A basic semantic webservice, covering taxonomy and auto indexing, is announced to be available early next year.

2 General Patterns of Non-Sustainability

When we analyse examples of failed sustainability we usually find similar patterns. Differences appear mainly between non-sustainability of material production, and non-sustainability of information and communication.

2.1 Material Production

The pattern may be described as shown in Figure 1:

natural resources
human idea
success
growth
progress
limitation of resources
migration

Figure 1: a pattern for non-sustainability of material production

[1] XML Competence Center, SchlumbergerSema. Sema GmbH, Kaltenbornweg 3, D-50679 Köln (Cologne) thomas.bandholtz@de.sema.com

Material production like slash-and-burn farming, or fuel-based transports today, generally starts with the discovery of a resource that can be exploited by some innovative technique. In this innocent state the resource seams to be endless. In the historical beginning of slash-and-burn farming nobody could imagine that the forest might be a limited resource one day - today we know more.
The new way of production is successful - otherwise it will not be continued. As it is successful, population grows, and production and productivity increase. First problems with resource consumption are overcome by progress of techniques, but the regeneration of the forest (or fuel) is far too slow. In the end, the population has to migrate - either geographically (if there is another forest, or another world), or by migrating technology and the kind of resources used. Following the American Petroleum Institute (1996), oil "... would sustain the current rate of consumption between 63 to 95 years."

2.2 Information

The problem of information sustainability is not that much a question of material production and resource limitations. The physical media production consumes resources, sometimes even in a destructive way, but there has never been any "desert" of information that would have been the result of resource limitations. Information did prosper continuously since the early days of oral tradition. Information has undertaken many migrations of techniques, starting with cave-painting and carving stones, proceeding to papyrus, and so on, till the days of printed paper, telecommunication and digital media.

Germany	3775	Italy	1126	Belgium	350	Netherlands	250
Great Britain	2500	Russia	500	Sweden	296	Turkey	200
France	2000	Switzerland	450	Denmark	250	Greece	200
Austria-Hungary	1200	Spain	400	Portugal	250	Norway	128

Table 1: Newspapers in Europe (1876/7)

All of these techniques have proved that they can be sustainable, and the step from carving stones to writing on papyrus was not undertaken because of restrictions in the availability of stones. Papyrus just provided portability of information, which obviously had been more important than the risk of inflammability.
Many collections of information have been destroyed by fire or human violence - but others have survived - without consuming resources in a hazardous way.

Resource consumption increased meaningfully with the invention of printed paper, especially newspapers.
Today, we establish recycling techniques in paper and digital media production, and we are using digital media and telecommunication more and more. Though, "an analysis of global paper consumption shows that use of paper products has tripled over the past three decades and is expected to grow by half again before 2010". World Resource Institute (2001). Sounds like we need an accelerated technical migration. These aspects are discussed more deeply in other contributions to this symposion, e.g. Reichart (2001).

oral tradition
symbols & letters
mobility
distribution
reproduction
multimedia
noise

Figure 2: a pattern for non-sustainability of information

But there is another aspect that is not so obvious, though most of us have experienced it somehow: the growth of information quantity results in a loss of meaning because of noise and interferences. If you are looking for a general evolutionary pattern of information, you will find something like this:
This is not is not a generic pattern that has been repeated many times in the history of men - it is one pattern for the history of information in all. Today we are somewhere between multimedia and noise. The problem of noise is that it starts to interfere with the sustainability of information.

3 Why and how should information be sustainable?

When we discuss climate change today, we usually talk about increasing average temperature like shown in Figure 3 (left).
"Measurement records indicate a warming of 0.3o-0.6oC in global average temperature since 1860. .... But observations are sparse before 1900 and much of the warming occurred between 1910 and 1940, before the largest rise in greenhouse gases. There is clearly more going on than a simple, direct response to emissions" UNFCC (2001).
What else is going on? When you oversee a longer period of time, like 10,000 years in Figure 3 (right), you find a different pattern then simply *increasing* temperature. I

don't want to discuss the possible meaning of this long-time-observations" in this article. I am just arguing for long-time sustainability of data content.

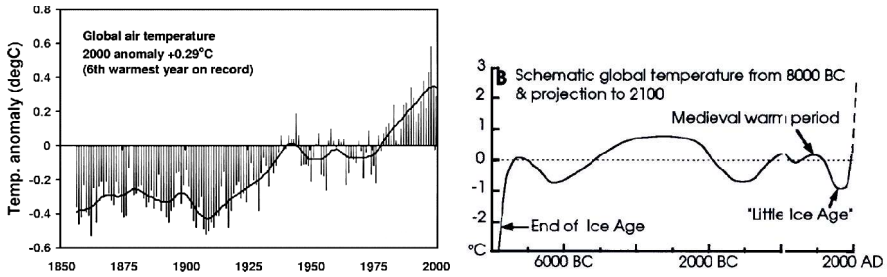

Figure 3: Global air temperature since 1850 (left) and since 8000 BC

What were those "sparse observations before 1900", and how have they been documented? Or - does anybody know the results of the population census in the year 0, in Bethlehem?
Today we have the chance to keep our current observation data readable without any time limitations - and we shouldn't miss this by producing noise instead.
Does this mean we shall save all the information we are generating every day till the end of time?
Of course, not! Each piece of information has a dedicated ***life cycle***. If you take most of the daily news (or the daily "soap"), the life cycle will last one day at most. You can see the life cycle *happen*, when you watch the daily waste paper collection.

4 Impacts of sustainable information

Besides a solution of sustainable resource consumption, information has to provide several *informational* features to become sustainable all over its individual life cycle.

- **Localisation**

First of all, information must be localised by the client. Not much information is addressed directly to the recipient. Most of it is searched for using libraries, catalogues (like the *Umweltdatenkatalog*), search engines, or information brokers.

- **Physical Access**

After information has been localised, it needs to be accessed physically.
Imagine you have found the library, but the book is rotten and it turns to dust when you touch it. Imagine an alien finding some of today's CD-ROMs in the far future - will he know which device is needed to access it? Will the media be readable?

And what about 5 1/4 inch floppy disks or tape material from the early computing days?

- **Encoding**

When you have physical access to information, you have to decode it. Can you read cuneiform characters? Latin letters? EBCDC? ASCII? Unicode?

- **Structure & Semantics**

After you have decoded the information, you have to understand its structure and semantics. Is it plain text with headlines and chapters? Database rows and columns? What is the meaning of the columns, which units for numbers, which methods?

- **Aggregation & garbage collection**

At the end of the life cycle, you may aggregate or erase the information.

5 A technical answer: Webservices

All of these impacts are addressed somehow by an upcoming de-facto-standard, based on Internet and XML Following the Webservices Definition Diagram (Webservices.org 2001), webservices are: "... encapsulated, loosely coupled contracted functions offered via standard protocols".

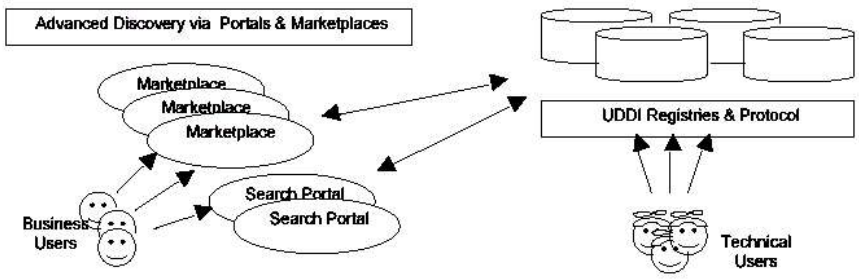

Figure 4 Universal Description, Discovery and Integration (UDDI) of webservices

From the end user's point of view, one can scan a registry of webservices that provide different kinds of information, and he can immediately access each service, as it works with a common standard protocol. Those who have to offer one of these services "just" have to implement it conforming to the same protocol, and describe it in a registry.

As Figure 5 shows, most of the proposed architecture and protocols are ready to use. It looks like webservices are becoming the practical answer to Tim Berners-Lee's vision of the *Semantic Web* ("The Semantic Web is a web of data, in some ways like a global database"), which already has been an important influence to the *German*

Environmental Information Network (Bandholtz, 1999, 2000). The "global database" is not made from static HTML-pages, but from dynamic webservices, and the description language is not the Resource Description Framework (RDF), but the Webservice Description Language (WSDL, see http://www.w3.org/TR/wsdl). But both approaches are members of the XML standards family, and they may co-exist without general interferences.

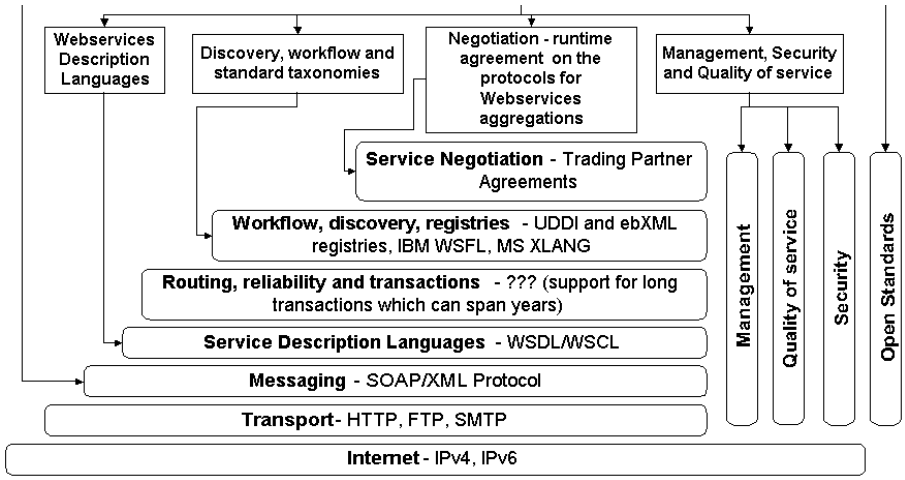

Figure 5 Webservice Layers. From: Webservices.org (2001)

6 Webservices in environmental information

In the 90s most of the environmental authorities started to use the Internet for public information. Now we can say: most of the commonly requested information is accessible via the Web. But still the information is distributed among many places, and each information provider follows his own rules and preferences. There is no harmonisation but being "in the Internet", and there is no localisation support by common search engines. Thus the end user, though amidst of a wide variety of environmental information, just *happens* to find some of it without being able to compare it to what could be found. Early dedicated portals, using semantically meaningful metadata, can be found on the national (e.g. German Environmental Information Network, GEIN) or European (EEA) level. They serve as centralised access points, not only as passive catalogues like before.

While developing the GEIN broker, we found that a growing amount of information providers connected their databases to the Internet dynamically. These contributions

cannot be indexed using search engine crawler methods. So we established a distributed query mechanism using XML in HTTP requests. This might be called an early webservice, offering result sets of described hyperlinks as an answer to a search condition. Currently we have 8 database servers participating in Germany, including the *Umweltdatenkatalog*, but there are many others waiting to be connected.
This started before webservices have been standardised in the way described before, but the basic techniques (XML via HTTP) are the same, and the protocols can be adapted to the Simple Object Access Protocol (SOAP) and WSDL easily after this experience. We don't need something completely new - we just have to adjust (or add) generic interfaces and make them conform to a new global standard. This is also what the European Environmental Agency means when they write:
"... a set of emerging standards and practices from the e-business world will have to be adopted by the environment sector. These include but are not limited to XML/edi, SOAP, ebXML, UDDI, ISO/IEC 11179, and more." CDS (2001).

7 A sustainable webservice for environmental taxonomy

One of the first true environmental webservices in Germany is currently developed in an R&D project by the Federal Environmental Agency and SchlumbergerSema - based on the experiences with GEIN. The project is named "Semantic Network Services (SNS)"[2], and it will result in:
"... a set of semantic webservices that make thesaurus-based content analysis accessible by any environmental web server via HTTP. GEIN should be on the way to become an interactive environmental topic map for public use." Bandholtz (2001).
This means a standard taxonomy service in the Workflow-Discovery-Registry-Layer of the webservices definition diagram (Figure 5), and it will be implemented using the proposed protocols. GEIN will become a *Service Broker* as described in Newport (2001).
This supported services benefit from the inheritance of the early catalogues: thesauri and metadata standards. SNS uses GEMET (or the extended German thesaurus) to build indices in a well-structured data model defined by an XML Schema.
SNS converts thesauri and additional taxonomically sources into an *Environmental Topic Map*. Basically, a Topic Map [ISO/IEC 13250:2000] is a network of *topics* (~ terms), *associations* between topics, and of *occurrences* (information) of these topics. Associations are used to semantically interconnect terms, geographic names, and environmental events to provide understanding of questions like: *What happened since Chernobyl*?

[2] Research Project UFOPLAN-Ref. No. 20111612, promoted by BMU/UBA, Germany.

SNS provides simple query-services that find topics matching a given search condition, and there will be text analysis facilities which result in auto-classification and topic-based indexing of documents (passed by URL or as complete text).
The "beta"-phase is scheduled for this winter, and every information provider that is interested to participate should contact gein@gein.de soon.

8 Literature

Arndt, H./ Günther, O. (eds.) (1999): Environmental Markup Language (EML). Proceedings of Workshop 1, Berlin 1999. Metropolis. Marburg, Germany 2000.

Bandholtz, T. (1999): GEIN 2000 and beyond: Information about the environment in the „semantic web“. In: Arndt (1999)

Bandholtz, T. / Bös, R. / Rüther, M. (2000): The German Environmental Information Network (GEIN). In: Cremers (2000).

Bandholtz, T. (2001): Inhaltsanalyse und Navigation im Web. In: Tochtermann (2001).

CDS and e-EIONET Work Conference 2001: Building environmental web services http://www.eionet.eu.int/events/cdsittag2001

Newport, B. (2001): Requirements for Building Industrial Strength Web Services: The Service Broker. http://www.theserverside.com/resources/article.jsp?l=Service-Broker

Reichart, I.; Hischier R (2001): Environmental impact of electronic and print media: In: [this compilation].

Stolzenburg, W. (2001): Fire in the Rain Forest, The Nature Conservancy, May/June 2001, http://nature.org/magazine/2001/mayjun/issues/art3166.html

The American Petroleum Institute (1996): Oil Supplies -- Are We Really Running Out of Oil? http://www.api.org/oilsup.htm

Tochtermann, K. / Arndt, H. (eds.): Proceedings of 4th Workshop "Hypermedia in Environmental Protection", Metropolis. Marburg, Germany 2001

Topic Maps (ISO/IEC FCD 13250:1999). Prepared by: ISO/IEC JTC1/SC34. http://www.ornl.gov/sgml/sc34/document/0058.htm

UDDI (2000): Universal Description, Discovery and Integration. UDDI Technical White Paper. September 6, 2000. http://www.uddi.org

Cremers, A.; Greve, K. (eds.): Computer Science for Environmental Protection '00. 2000.

UNFCC (2001): Climate Change Information Kit, http://www.unfccc.int/resource/iuckit/infokit.pdf

Webservices.org (2001): Webservices Definition Diagram. http://www.webservices.org/whatarewebservices.php

World Resource Institute (2001): Global Trends. Production and consumption, http://www.wri.org/trends/consume.html