

# Semantic Network Services (SNS)

Michael Angrick, Richard Bös, Maria Rüter<sup>1</sup>; Thomas Bandholtz<sup>2</sup>

## Abstract

Talking about “*Agents, Communication and Ontologies*” within “*European Environmental Networks*” we announce the *Semantic Network Services* that just have been implemented in Germany. SNS realizes an *Agent* to provide the application of a common *Ontology* by means of *Web Services Communication* for the environmental authorities that have gathered in the German Environmental Information Network (gein®).

The ontology is bi-lingual so far, but as it already includes the German and English part of GEMET, it wouldn't be much to extend SNS to support any GEMET language.

## 1 Introduction

In 1998, the European Union released the Aarhus convention on "access to information, public participation in decision-making and access to justice in environmental matters". This turned out to be an unheard-of challenge to the information policies and technologies of all the governmental authorities concerned.

Today, environmental information is published in masses in the Internet. Everybody can access this information with common Internet browsers from his office, home, public internet cafe, or using mobile media in the near future.

Still the information is distributed among many locations, and each information provider follows his own rules and preferences. There is no harmonization but being "in the internet", which is a rather technical consensus.

In Germany, since 2000 gein® has implemented a common content classification system as a first step of any further content-related integration of the different Internet information sources. Following this encouraging experience, we have started an R&D project (SNS)<sup>3</sup> in 2001 to improve these early efforts. SNS will be finalized end of 2002, “to leverage human capability and better exploitation of human

---

<sup>1</sup> Umweltbundesamt, Berlin, Germany. [michael.angrick@uba.de](mailto:michael.angrick@uba.de); [richard.boes@uba.de](mailto:richard.boes@uba.de); [maria.ruether@uba.de](mailto:maria.ruether@uba.de).

<sup>2</sup> SchlumbergerSema, Cologne, Germany. [TBandholtz@slb.com](mailto:TBandholtz@slb.com).

<sup>3</sup>Research project UFOPLAN-Ref. No. 20111612, promoted by BMU/UBA, Germany

knowledge and content in large scale distributed systems” (IST) by “Adding explicit semantics to content, services and processes” and “Acting upon semantic descriptions” (FP6KT).

## **2The gein® Taxonomy**

The gein® Taxonomy has been introduced and revisited in several presentations, so we only will summarize the key points here. GEIN uses three semantic structures:

1. a thesaurus of currently 39,143 environmental terms (UmThes®),
2. a gazetteer including the intersections between 48,213 geographical objects of all kinds,
3. a “calendar” – the synopsis of historical and contemporary events that affected the environment.

UmThes® is a full-fledged Thesaurus supporting all the associations required by ISO 2788/5964 (Broader/Narrower; Synonym; Related; Component), and it contains the full word morphology. It is also used by several German-speaking authorities such as the German and Austrian Environmental Data Catalogue, and it is the German source of the GEneral Multilingual Environmental Thesaurus (GEMET, Batschi).

The gein® Gazetteer is based on the GN250 (by Federal Agency for Cartography and Geodesy), but it adds several layers relevant for the environment, and it contains all the spatial intersections as explicit relations in the data, ready-to-use in a rapid query.

The gein® Calendar has been started by GEIN from scratch, and it intends to become some kind of “temporal Thesaurus” – not to be misunderstood as a pin board for conference announcements.

All the three threads of ontology have been integrated into one semantic model by a Topic Map (see Chapter 4). So the term “Topic” refers to any of the ontology nodes regardless whether they are descended from the Thesaurus, the Gazetteer, or the Calendar. From the metadata point of view, Topic may be understood as equivalent for “keyword”.

Currently there is an ongoing discussion about adding “Actors” (persons and organisations) to enhance the classification facilities.

## **3Semantic Web Services**

Since 1999, the gein® Broker has been hosting all the taxonomy. It has been used for the classification of currently 140,000 static Web pages published by 70 information providers. There have been several requests by the information providers to be enabled to use the same taxonomy and auto-classification methods

for their own purposes as well. Thinking about the effort to prepare a compact module to be distributed for implementation in 70 different technical environments, we preferred to think about a centralized service that can be accessed online by all of them. This service is designed to support the basic needs for content classification, such as:

1. Topic matching – searching for any kind of occurrence of a given character string in the Topic characteristics,
2. Topic navigation – moving from topic to topic using associations as stepping stones,
3. Auto-Classification – using a new topic-map-based linguistic analysis to extract the most characteristic Topics of a given document.

gein® looks back to very positive experiences with distributed queries using XML packed into HTTP requests, so we considered to consult the W3C Web Services Activity (<http://www.w3.org/2002/ws/>). According to the Web Service Description Usage Scenarios (W3C, 2002), SNS represents a “UC0015 Request-Response” scenario defined as: “Ability to describe an operation of a web service that responds with an output message or a fault based on at least one or more input messages received.”

Examples of messages and responses may be:

*SNS-client*: “My user is searching for information using the string “...”. Is there any Topic matching with a string like that?”

*SNS*: “I have X Topics, here I give you their definitions in the XML structure that we have agreed on.”

*SNS-client*: “I see. Please give me all the associations of Topic XYZ.”

*SNS*: “Topic XYZ is *narrower term* of Topic ABC, it is *related to the Calendar* Topic DEF, etc, etc.”

*SNS-client*: “Please give me the auto-classification results for the document at <http://www.abc.xyz/any.foo>. Please find the 10 most significant topics!”

*SNS*: “The 10 most significant topics for this document are ....”

Additionally, there may be messages according to the “UC0025 Event notification” scenario described as: “An application subscribes to notifications of certain named events from an event source. When such events occur, notifications are sent back to the originating application (first party notification) or to another application (third party notification).” (W3C, 2002)

In this scenario, SNS is the “event source”, sending “named events” (such as a Topic Map update has happened) to the SNS-clients that have subscribed to this services before.

The XML structure and syntax of this communication are well defined using XML Schema and Web Service Description (W3C, 2001), being registered in an Universal Description, Discovery, and Integration (UDDI) registry. Details on these documents are beyond the scope of this paper.

Conforming with these standardization processes provides a general frame for interoperability of SNS. Not only the gein® information providers, but any potentially interested system all over the world will be able to understand immediately the technical means of integration (given an awareness of global standardization in this field).

#### **4Topic Maps**

The three semantic structures of the gein® Taxonomy introduced in Chapter 2 have been integrated into one single Topic Map by setting up associations between thesaurus descriptors, events and locations.

Topic Maps (ISO 13250) have proved to be the structure of choice whenever different thesauri, gazetteers, classifications, or even simple keyword lists have to be networked. Topic Maps are a rather new standard in the world of taxonomy and ontology, and they have emerged to become one of the most discussed contributions. At XML Europe 2002 (XMLe), there have been 10 contributions dedicated to Topic Maps, and the Knowledge Technology track has been dominated by this movement. Topic Maps have their roots in SGML, and have been standardized in parallel to and in permanent discussion with the Semantic Web and RDF process (Bandholtz, 1999). Today it looks like these two threads are going to find their deserved integration.

Basically, Topic Maps consist of three components: Topics, Associations, Occurrences.

*Topics* represent any kind of *subject*:

"In the most generic sense, a subject is anything whatsoever, regardless of whether it exists or has any other specific characteristics, about which anything whatsoever may be asserted by any means whatsoever." (ISO 13250)

So, a Topic may be a Thesaurus descriptor or synonym, a geographic object of a Gazetteer, an Event, a Person, an Organization, whatsoever. Distinct kinds of Topics are defined as Topic Types in a Topic Map application.

*Associations* may interconnect Topics in some kind of semantic relation. Descriptors and Synonyms, Events and Locations, Events and Persons, Persons and Organizations, whatsoever. Distinct kinds of Associations, bound to certain Topic Types as their members, are defined as Association Templates in a Topic Map application.

*Occurrences* are pieces of information that contribute to the definition of the Topic. Generally, an Occurrence may be seen as any kind of existing information about a Topic, but, as Occurrences are “groupings of addressable information objects around topics” (ISO 13250), this should not be misunderstood to be the general information index of a “corpora” like *gein*®. In SNS, the document index is separated from the Topic Map. Topics are used as classification properties in the document metadata, which rather means: “groupings of topics around addressable information objects” (Bandholtz, 2002).

The XML model of the SNS Topic Map is based on ISO/IEC 13250, but it does not use the (XTM) 1.0 DTD. XTM has been designed to be an *interchange format*, not a processing format (Bandholtz, 2002). We preferred to develop an XML Schema for Topic Maps after having reviewed the basic ISO SGML DTD, the XTM XML DTD, and Martin Bryan’s early XML Schema proposal (Bryan). Again, XML details are beyond the scope of this paper.

## **5Future Integration**

While this paper gets written, a first Beta release of SNS has been implemented so far, basically the integration of the *gein*® Taxonomy, technically based on an XML Topic Map Engine provided by SchlumbergerSema. We expect the first version ready for integration tests in September. The final release of Version 1 is scheduled for Q4, 2002.

There are several integration options, targeted to different users in different application areas:

1. *gein*® Broker: The first client to make experimental use of SNS will be the *gein*® Broker. SNS is intended to replace the existing implementation of taxonomy features later as a “second generation”.
2. *gein*® Information Providers: the (currently 70) contributing organizations are invited to integrate SNS by Web Services for any kind of information activities, starting with Q4 of 2002. Some of them intend to implement a local version of SNS themselves. Finally, there may be a network of cascading Topic Maps parts of which may be consulted depending on whatsoever localized or thematic focus of an application.
3. UDK (German Catalog of Data Sources): The (VV UDK/GEIN), an administrative agreement of the Federal and Länder authorities in Germany about an integrated future maintenance and development of UDK and *gein*®, will become effective on January, 1st, 2003. One of the key subjects currently in discussion is, how to integrate the taxonomy features of both systems. Currently, both use *UmThes*®, but in different implementations. SNS is intended to become

the common basis of both systems in the future, in a revised version to come in 2003.

4. GeoMIS.Bund: the “Metainformation-System for geodata of the Federation” of (IMAGI) within their “national Geo data infrastructure”, a prototype of which has been developed by SchlumbergerSema in early 2002, has considered to make use of SNS by integrating an early prototype interface of the SNS Gazetteer. Currently there are negotiations with the Interministerial Committee for Geoinformation ( IMAGI ) about the future support and extended application.
5. Europe: The eEIONET workconference within EnviroInfo (that this paper is part of) discusses “environmental web services e.g. Reportnet, country networks, and metadata, as well as terminology/ontology issues” on a European level (EIONET). As the relation between GEMET and UmThes® is very close, and as SNS already is working bi-lingual (German/English), it is a candidate to be extended to a European Scope (gazetteer) and to the full multi-lingual context of currently 19 GEMET languages. This has been proposed in (Bandholtz, 2001) on the EIONET work conference last year, which led to an Expression of Interest (EoI) within the 6<sup>th</sup> Framework Programme of the European Commission.

## **6Bibliography**

- Bandholtz, T. (1999): GEIN 2000 and beyond: Information about the environment in the „semantic web“. In: Environmental Markup Language (EML). Proceedings of Workshop 1, Berlin 1999. Metropolis. Marburg, Germany 2000
- Bandholtz, T. / Bös, R. / Rüter, M. (2000): The German Environmental Information Network (GEIN). In Cremers, A.; Greve, K. (eds.): Computer Science for Environmental Protection '00. 2000.
- Bandholtz, T. (2001): Semantic Network Services (SNS) - a webservice for environmental taxonomy. CDS and e-EIONET Work Conference 2001: Building environmental web services. <http://www.eionet.eu.int/events/cdsittag2001/abstracts/band.htm>
- Bandholtz, T. (2002): A Taxi in Knowledge Land. XMLEurope 2002, Barcelona. <http://62.231.133.220/idea-eks-nav/papers/03-05-03/03-05-03.html>
- Batschi, W.D. Felluga, B., Legat, R., Plini, P., Stallbaumer, H., Zirm, K.: „SuperThes“: A New Software for Construction, Maintenance and Visualisation of Multilingual Thesauri. In: EnviroInfo 2002, Vienna 2002.
- Bryan, Martin : XML Schema for ISO 13250 Topic Maps. Proposed Syntax 29th January 2001. <http://www.diffuse.org/TopicMaps/schema.html>
- EIONET: European Environment Information and Observation Network. eEIONET Work Conference from 26-28 September 2002 in Vienna. Released: 2002/07/03. [http://eea.eionet.eu.int/Best\\_Practice/eEIONET2002](http://eea.eionet.eu.int/Best_Practice/eEIONET2002)

- EoI: European Environmental Topic Map Engine with Multilingual Auto-Classification (EETM). Expression of Interest to the 6<sup>th</sup> Framework Programme of the European Commission. June 2002. [http://www.jiscmail.ac.uk/files/DC-ENVIRONMENT/EoI\\_Bandholtz.doc](http://www.jiscmail.ac.uk/files/DC-ENVIRONMENT/EoI_Bandholtz.doc)
- FP6KT: Knowledge Technologies Workshop. Luxembourg, 15 May 2002. Summary report. (ISTweb | KA3 | IAF) [http://www.cordis.lu/ist/ka3/iaf/iaf\\_workshop.htm](http://www.cordis.lu/ist/ka3/iaf/iaf_workshop.htm).
- IMAGI: Interministerial Committee for Geoinformation ( IMAGI ).  
[http://www.imagi.de/welcome\\_eng.htm](http://www.imagi.de/welcome_eng.htm)
- ISO 13250: Topic Maps (ISO/IEC FCD 13250:2000). Prepared by: ISO/IEC JTC1/SC34 - Document Description and Processing Languages.  
<http://www.ornl.gov/sgml/sc34/document/0058.htm>
- IST: Knowledge Technologies within the IST 2002-2006 programme. Some reflections. Internal report on Knowledge Technologies (aka IRG Report). Luxembourg 2002.
- SNS: Federal Environmental Agency of Germany, SchlumbergerSema: Semantic Network Services. Research Project UFOPLAN-Ref. No. 20111612, promoted by BMU/UBA, Germany
- UDDI: UDDI Version 3.0 Specification. Open Draft, Dated 03 July 2002.  
<http://www.uddi.org/pubs/UDDI-V3.00-Open-Draft-20020703.htm>
- VV UDK/GEIN: Verwaltungsvereinbarung zwischen Bund und Ländern über den gemeinsamen Betrieb und die gemeinsame Entwicklung und Pflege des Metainformations-systems Umwelt -Datenkatalog UDK und des Umweltinformationsnetzes Deutschland GEIN.
- W3C (2001): Web Services Activity: Web Services Description Language (WSDL) 1.1. W3C Note 15 March 2001. <http://www.w3.org/TR/wsdl>
- W3C (2002): Web Service Description Usage Scenarios. W3C Working Draft 4 June 2002.  
<http://www.w3.org/TR/2002/WD-ws-desc-usecases-20020604/>
- XMLe: XML Europe 2002. Down to business: Getting serious about XML. Barcelona, 20-23 May 2002. Proceedings. <http://www.idealliance.org/papers/xml02/>
- XTM: Members of the TopicMaps.Org Authoring Group: XML Topic Maps (XTM) 1.0. TopicMaps.Org Specification. <http://www.topicmaps.org/xtm/index.html>