# The Semantically-enriched Translation Interoperability Protocol

**Sven Christian Andrä**
ONTRAM Inc.
10 South Third Street
San José, CA 95113, USA
`sca@ontram.com`

**Jörg Schütz**
bioloom group
Bahnhofstraße 12
66424 Homburg, Germany
`joerg@bioloom.de`

## Abstract

In this paper, we present the semantically enriched (SE) version of the Translation Interoperability Protocol (TIP). TIP is designed to foster and enable the seamless sharing of data and information between different TMS based on open standards for data representations by means of the TIP Package (TIPP) as transport container. SE-TIP is a research sideline that employs Semantic Web technologies to support modeling identification and interaction of sharing tasks, and uses the Web architecture to ensure extensibility and scalability of the SE approach.

## 1 Introduction

### 1.1 User Story

Imagine a small young bio-pharmaceutical organization, say nanopharm, that recently has developed a new nano scale therapy for drug targeting. Their technology has been approved by national and European health administration authorities, and now the company wants to expand their market to several European countries as well as the entire world market. To accomplish this challenge, the company is faced with a huge amount of administrative business tasks, processes and workflows which have to be fulfilled in several languages and adapted to multiple cultures according to local rules and regulations.

Initial tests with freely available machine translation services on the Web have shown promising and even partly acceptable results. However, these results come with a serious lack of an appropriate vocabulary coverage of the bio-pharmaceutical field, and particularly the company's own terminology and style. In addition, general globalization, internationalization, localization and translation (GILT) project management capabilities are missing with these services as well as severe security and trust problems including appropriate and convenient configuration and customization facilities.

Buying and maintaining their own translation automation solution is currently too expensive for the small company, and so they are looking for a Web based solution for their translingual communications needs. The envisioned solution should also support the effective communication with language service providers (LSPs), individual translators, and possible social network services for crowdsourced translation.

In summary, the actual needs and requirements of the nanopharm example enterprise are those of a fully fledged secure Web-scale translation service management framework (TSMF) that requires no software installation, can be personalized easily to suit individual habits and preferences, is secure and extensible, and works fast, reliable and effective in solving multilingual global business challenges, extending the company's value, and helping to decide what translation process size fits their needs in changing environments.

### 1.2 Ultimate Solution

The envisioned Web-scale TSMF can be seen as an innovative cloud computing application within the broad field of GILT process modeling, automation and intelligence. This framework combines, controls and manages a number of services that are accessible through the Web by intuitive webbrowser interfaces. The services are dedicated particularly to quality, competence and performance in terms of their result delivery, and they enable users to optimize and maximize their trans-

lingual communications processes and workflows, and to gain new revenues in existing and new customer relationship operations.

This service scenario might also include the checking and automated streamlining of information content for machine translation readiness, various terminology related operations such as term mining, the deployment of fully automated translation workflows and post-editing tasks to gain optimized quality, as well as the static and dynamic configuration and the management of complete internationalization and localization project life-cycles. Since multiple services are involved within such a framework of services, interoperability between the service connectors and components is key, particularly the sharing of data and metadata, is a necessary and challenging requirement.

## 1.3 Good Practice Solution

In this paper, we will solely focus on the interoperability of translation related data and metadata between translation management systems (TMS) because this is one of the essential needs and requirements within the broader application scenario of the nanopharm example company. Firstly, the already existing solution of the Translation Interoperability Protocol[1] (TIP) ensures the freedom of tool choice for GILT service buyers, vendors, and individuals within GILT workflows. Secondly, TIP is based on existing standards and best practices for data exchange formats employed in the TIP Package (TIPP) as transport container. Thirdly, the semantically-enriched TIP extension SE-TIP employs Semantic Web technologies for different modeling purposes, and is grounded in the Web architecture to allow for a thorough extensibility and scalability.

## 2 TIP Package

### 2.1 Basics

Exchanging and interchanging various types of data between different TMS gains more and more attention in the field of product and media localization and translation. It comprises multiple workflows with various activities and tasks of humans and machines on different data types and formats

in tandem with several actors, technologies and tools.

The need interoperability issue arises because a translation buyer, remember the example company, might use other systems and tools for handling and managing language data than the translation vendor LSP, or uses even multiple systems within their enterprise infrastructure. Additionally, there might be different freelance individuals involved in these processes who again employ yet another computational infrastructure, say, mainly based on free or open source software. In either case, lossless data and information sharing is considered a valuable asset in many natural language related processes that deal with terminological data, translation memory content, machine translation systems, etc.

Today, many proprietary solutions of LSPs exist, however, with the ultimate danger of entering into a vendor lock-in. Therefore, over the last two decades several initiatives – public and private – have been working on standardized data representation formats, frameworks and best practices to support the interchange of natural language vocabulary material and translation memory content. But even if we rely on these open standards, they mostly deal with the content part only, and not with associated processing information and general metadata.

### 2.2 Open Standards and Best Practices

Over the last two decades, a set of open standards related to localization and translation has been developed to support the various data and processing needs in technical communications and documentation of the software and manufacturing industry in close collaboration with internal and external translation services. Today, the most widely accepted open standards in the GILT industry are:

**ITS:** The Internationalization Tag Set (ITS) is a markup language for the identification of internationalization related aspects in XML documents including terminological and glossary information. The work on ITS is with W3C.

**TBX:** The Term Base Exchange (TBX) is a means to describe terminological data either as concept-oriented data or flat glossary data in an XML style. After the demise of LISA in February 2011, the continuation of TBX and other localization related standards maintained by LISA is still an open issue. Recently, the GALA localization

---

[1] TIP is the result of the Interoperability Now! (IN!) initiative of several independent enterprises, which started in the second quarter of 2010.

organization has started a standards initiative in this context.

**TMX:** The Translation Memory Exchange (TMX) is a collection of translation memory data in possibly multiple languages. The formal means are based on XML; LISA was also responsible for this exchange format.

**OLIF:** The Open Lexicon Interchange Format (OLIF) is a highly complex description format for lexical material. It has been created to support the needs of NLP tools that operate with linguistic rules for morphological, syntactic and semantic processing, including machine translation (mainly RBMT). OLIF has been pursued by industrial and research organizations and partners such as SAP, SDL/Trados, Systran, DFKI, IAI, etc.

**XLIFF:** The XML Localization Interchange File Format (XLIFF) is a transport container that stores and carries extracted text through the various steps of a localization process. As such it is the only format that was designed with a process oriented view on the represented data. An XLIFF file is bilingual, i.e. only one source language and one target language are permitted. The work on XLIFF is under the supervision of the OASIS group, and several tools are available for handling the different aspects of XLIFF including editing.

**Related Other Standards:** Other standards comprise, for example, formats for describing segmentation rules (SRX – Segmentation Rules Exchange, LISA) of natural language expressions, quantitative measures of documents (GMX – Global Information Management Metrics Exchange, LISA), authoring memories (xml:tm – XML-based text memory, OASIS), and the GNU gettext for Portable Objects (PO) in software engineering. Complete frameworks for metamodel markup languages for lexical data and terminology data are LMF (Lexical Markup Framework) and TMF (Terminological Markup Framework) that have been designed and developed within ISO/TC37 (ISO 16642) contexts and the EU project SALT which also initiated the work on TBX.

## 2.3    Existing Gaps and TIP

The introduced standards for GILT data have all in common that they are markup languages for content data with only a limited support of metadata, mainly for administrative purposes. XLIFF is an exception because it also allows for the specification of process related data and metadata through its support of XML namespaces

for non-XLIFF elements and attributes. This approach opens a multitude of possibilities and thus interpretations across applications which also discourages interoperability.

What is needed is a framework that combines content, resource information and workflow information in a coherent and agreed upon or even standardized way with one single interpretation across applications. For each of these types of data we need to provide specifications for identification, representation and interaction to ensure effective interoperability. The aim of TIP is therefore to integrate the description models of the various disruptive GILT technologies and their associated data, and to allow for optimizing their deployment in even disruptive GILT workflows. The main challenging areas in GILT workflows across different industries are:

- coordination and distribution of data and information within and across organization department boundaries in multiple languages

- harmonization and monitoring of translation business processes

- language and cultural specific, i.e. locale specific, challenges with time-to-market delivery issue

TIP and especially SE-TIP combine these technologies through a dynamic object view that links data, resources and possible functions and processes with metadata models. In addition, TIP consuming applications may modify the TIP Package content in an automated way.

## 2.4    TIP Package Layout

The TIP Package (TIPP) is a container that consists of a TIP *Manifest File* (MF) encoded in XML which includes references to and administrative information about the different TIP objects, and a series of either object files or object folders with object files. The latter structure applies if more than one object file of a given object type is part of the TIPP distribution. As of this writing, we distinguish the following object types with their possible representation formats including the extensions of SE-TIP[2]:

---

[2] Currently, the TIP, TIPP and XLIFF:doc specifications are under beta review, and they will be presented to the general public at TM Europe 2011 with implementations that also demonstrate the round-trip capabilities of TIP.

- *Translation Object* contains in one file the source language input, and after translation the target language output, both represented in XLIFF:doc[3] as described in (Bly, 2011).

- *Translation Memory Object* is a partial or a complete database extract of already translated and aligned natural language segments in source and target language, and encoded either in TMX or XLIFF. SE-TIP uses an RDF serialization format for TM content.

- *Terminology Object* is either a partial, i.e. relevant for the translation task, or a complete extract of a term database represented in either OLIF or TBX. SE-TIP is experimenting with SKOS and OWL .

- *Reference Object* contains general reference material. TIPP does not specify any data format yet, and in SE-TIP references are modeled in RDF.

- *Workflow Object* encodes process information in RDF (SE-TIP only).

- *Metrics Object* delivers accompanying administrative information such as word counts, pricing, quality, etc. in RDF (SE-TIP only).

- *Style Object* describes translation, editing and general governance rules in RDF (SE-TIP only).

All these object ingredients constitute the entire TIP Package in both flavors. In SE-TIP, the MF information is also encoded in RDF, and maintains direct links to the TIP objects.

## 2.5   TIP Supported Formats

When building distributed applications the employed formats of the resources matter mostly. The meaning, or semantics, behind the data and information in a resource must be understood by all parties involved in an interaction in order to successfully achieve a business goal. In this section, we explore and discuss in detail the formats that are supported in TIP with a particular emphasis on SE-TIP. In the following sections, the term *data* means the raw and uninterpreted

streams of bits, *information* refers to the interpretation of the data within the context of a particular application domain or a specific task, and *knowledge* represents the understanding of a domain after collecting, analyzing, and reasoning over the available data and information.

### The XML Case and XLIFF:doc

Now consider, for example, a text translation task of our example company nanopharm with a particular set of vocabulary and a certain style of localization, which has to be executed in a specific sequence of steps and in compliance with already existing translations stored in TMX format to ensure natural language consistency on different levels.

The description of each step of this translation task and the sequential ordering of the steps can be encoded in several ways. Nowadays an XML based representation is favored because it explicitly expresses hierarchical structures, and is often self-describing due to its textual nature. This allows us to separate the structured data and the represented information in terms of the data's interpretation.

This idea has been the general guidance for the design of XLIFF which in real-life applications, however, turned out as being too broadly specified in some cases, and too narrowly in others. On the one hand, because some XLIFF definitions are unclear and provide no orthogonality, or different mechanisms apply for the same concept, it is often impossible to support the specifications adequately across XLIFF tool implementations and interchanging applications. On the other hand, flexibility in storing, for example, translation project information, terminological data, or particular software contexts of user interfaces is missing or is too narrowly specified in order to being effectively applied in real-life translation projects.

Therefore, in the context of TIP the streamlining of XLIFF was a major task because it appeared easier to fully specify a usable and workable core subset of XLIFF than to invent the wheel anew. The XLIFF:doc (Bly, 2011) of the TIP approach takes care of the mentioned shortcomings, and directly supports interoperability between TMS.

Within SE-TIP we aim at an even  tighter integration of the TIPP objects through link relations in order to provide a semantic context for

---

[3] XLIFF:doc is a robust, fully documented subset of XLIFF 1.2 with the namespace extension "dx:" designed within the IN! initiative to ensure interoperability between TMS.

better controlling and monitoring workflows and resources. The use of links and their relation to objects is similar to (software) contracts, which also ensure the fulfillment of the interoperability requirement. In TIP, we still have the unsatisfactorily need to employ some level of human involvement. In order to accomplish full machine automation, we have to enrich such contracts for machines particularly on the level of choice of information representation to ensure the ability to share that information in an interoperable manner.

**Semantic Web Case of TIP Package Objects**

The main challenge in our interoperability scenario is to interpret information consistently across TMS applications. In this context, we use the term *semantics* to refer to the shared understanding that is defined by the TIPP objects in a contract-like way, and by which the meaning of, for example, a sequence of request-response exchanges, or the way in which a resource representation should be interpreted and used is modeled unambiguously.

In the following, we distinguish between the general approach of computing based on semantic technologies, such as machine learning, ontologies, inferencing, etc., and the Semantic Web (SW), which is the term used to refer to a specific ecosystem of technologies, such as RDF, RDFS, RDFa, OWL, etc. maintained by W3C. We only provide some brief insights on how we utilize RDF and OWL as well as SPARQL for SE-TIPP object representations and access because a fully fledged introduction to the SW technologies is beyond the scope of this paper – see (Schütz, 2010) for their employment within the business process and business performance field.

One could ask why should we use SW technologies because they are apparently very similar to a pure XML representation? The strength of RDF with its model of representing data as a directed, labeled graph lies in its processing model and the use of Uniform Resource Identifiers (URIs) to build statements, i.e. all aspects related to any TMS application and the associated processes can be dynamically described by using RDF statements about resources and their interrelationships.

Statements in RDF are of the form [subject, predicate, object], also known as triples, and they are quite near to a natural language expression which makes them evenly consumable by humans and machines. Subject and predicate of an RDF statement are always URIs, and an object can be either a URI or a literal. RDF also permits the specification of complex expressions based on the simple s-p-o schema. Within the SE-TIPP object scenario, additional statements can be either embedded directly in an already existing TIPP object representation or delivered to consumers through yet another object incarnation. In addition, RDF makes it easy to combine information from different graphs, as long as matching URIs are used to ensure the identity relationship. This allows software libraries to bring together the known statements about a resource in a variety of levels and complexity.

In Figure 1 the following simple natural language statements, which describe two qualities of a fictional task of our example company, are represented in the graph notation of RDF:

- *task 1* has *taskname term-harvest* (s-p-o statement with URIs only)

- *task 1* has *costbase 2.0* (s-p-o statement with a literal in object position)
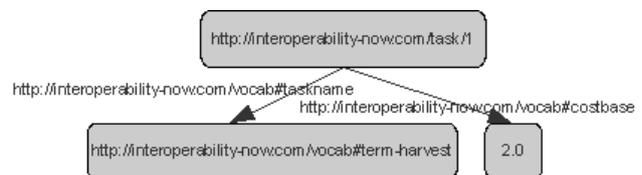


*Figure 1: RDF Graph - Part of Task Description*

The complete task description of task 1 with the additional information slot "costitem" that accounts for "wordcount," "maxsize," and "reference" in RDF/XML notation is depicted in Figure 2.

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [<!ENTITY xsd "http://www.w3.org/2001/XMLSchema#">]>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:interop="http://interoperability-now.org/vocab#">

    <rdf:Comment>This graph represents a simple Interoperability-Now task</rdf:Comment>
    <rdf:Description rdf:about="http://interoperability-now.org/task/1">
        <interop:taskname rdf:resource="http://interoperability-now.org/vocab#term-harvest"/>
        <interop:costbase rdf:datatype="&xsd;decimal">2.0</interop:costbase>
        <interop:costitem rdf:resource="_:item1" />
    </rdf:Description>

    <rdf:Description rdf:about="_:item1">
        <interop:wordcount rdf:resource="http://interoperability-now.org/vocab#all" />
        <interop:maxsize rdf:resource="http://interoperability-now.org/vocab#full" />
        <interop:reference rdf:resource="http://interoperability-now.org/vocab#harvesting" />
    </rdf:Description>
</rdf:RDF>
```

*Figure 2: RDF Statements for Task Description*

Additional information can be integrated easily in such an RDF representation. For example, the

representation of the task might also state that the URI representing the domain choice "biopharm" is associated with the corresponding label "bio-pharmaceutical" in English and the appropriate label "biopharmazeutisch" in German by using a link to a vocabulary specification; that the company's origin is a small town in Germany by using a geographical name service; and that its application domain is "drug targeting" by using a proprietary and shareable biotechnology vocabulary.

The processing model of RDF defines a set of basic rules and constructs that software applications can use as the building blocks for constructing the objects they might exchange. Because these constructs can also be used as the basis for developing vocabularies of concepts, such as "order," "cost," "metric," "wordcount," etc., which we employ to describe particular task qualities within our TMS application, they might evenly describe the meaning of certain XLIFF constructs which are beyond the XLIFF:doc specifications.

As such, the RDF approach allows us to define task-specific information by means of employing vocabularies for different purposes and specified in the Web Ontology Language (OWL) of W3C. For example, similar to the case that due to the absence of a widely used bio-pharmaceutical industry terminology, nanopharm can define a vocabulary that only applies within its own specific localization tasks. Such a vocabulary can be extended to provide a shared knowledge base that ensures effective interoperability and assures a common understanding of the employed SE-TIPP objects[4]. In both cases, an application-specific ontology[5] is defined.

## 2.6    SE-TIP Information Processing

In this section, we introduce the processing of SE-TIPP objects, and how applications can access the information encoded in these object data elements, i.e. s-p-o triples. We distinguish two main TMS application scenarios with each having its own SE-TIPP processing style:

- An application that becomes aware of SE-TIPP and starts to consume, understand

and interpret the package content in the intended way.

- An application that accepts SE-TIPP and just routes it through a particular workflow.

The former application scenario represents an active and dynamic processing style that accomplishes changes the originally delivered SE-TIPP objects in a controlled manner, whereas the latter scenario is a passive processing style with only a delivery and routing functionality.

A particular SE-TIPP within a given workflow always contains an information record of the applications in the form of additional s-p-o statements in the workflow objects, which are obligatory to ensure full traceability, control and monitoring, and possibly in the other objects, which extend or amend the represented data and information with, for example, revised and new translation memory and terminological content.

As we have seen, RDF and OWL can be combined into a single information graph of s-p-o triples. To access and to query these statements by matching a graph or subgraph, the W3C language SPARQL was designed to support the RDF data model with a query language for graphs. The result of a SPARQL query may consist of a set of resources and the interrelationships that satisfy the given conditions, answers to true and false questions based on the encoded knowledge, or entirely new graphs that are generated by inferring new triples from the existing set of statements – inference is the only mechanism at work in the SW context. Figure 3 shows an example SPARQL query which makes use of the publicly available vocabulary "FoaF" (Friend-of-a-Friend) to describe attributes of persons such as "person" and "age."

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX nanopharm: <http://nanopharm.com/vocab#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?name
FROM <http://internal.nanopharm.com/translators.rdf>
WHERE { ?x foaf:name ?name;
        foaf:age ?age.
    FILTER (xsd:positiveInteger(?age) > 30) }
```

*Figure 3: Example SPARQL Query*

## 2.7    SE-TIP Security and Trust

In this section, we discuss the aspects of SE-TIP that are related to:

---

[4] The Semantic Web community refers to such vocabularies as ontologies.

[5] A less complex formal means for terminology data is SKOS (Simple Knowledge Organizing System) of W3C. The difference between SKOS and OWL is their intention for different purposes: OWL allows the explicit modeling of a domain, whereas SKOS provides vocabulary and navigational structure.

29

- Confidentiality which keeps SE-TIP information private while in transit or storage.

- Integrity which prevents SE-TIP information from being changed undetectably.

- Identity which authenticates the parties involved in interactions.

- Trust which authorizes a party to interact with a package in a prescribed manner.

For these areas, the Web community has developed a number of higher-order protocols that address the issues of identity and trust which sit atop of HTTP, and allow systems to interoperate securely. HTTP natively supports authentication to establish identity, and authorization to basically help to establish trust. In a SE-TIP application, we can secure access to the SE-TIP resources with these capabilities. For instance, we may allow only authorized "consumers" to access a terminology resources. Privileged resources are accessed by providing certain credentials in an authorization header.

The integrity of SE-TIP Package objects is maintained through the built-in control and monitoring capabilities which allow for a dynamic "supervision" of the involved processes without influencing the actual processing of the shared data and information. As such, even the transitivity of application or process sequences is guaranteed as long as the information records maintained by consuming applications are not exposed to being attacked or harmed (vulnerability).

Additionally, package objects might be encrypted for privacy reasons; for example, to fully secure a company's terminology and translation memory content, and to grant access to these resources only to trusted "consumers".

## 2.8    Related Work

Currently, we are not aware of any directly related work to SE-TIP. Most approaches in GILT environments are still dealing with the syntactic level. There are also other container based approach emerging but none of these envisiones to employ explicit semantic descriptions. In the field of cloud computing, the community discusses similar aspects for modeling and representation purposes including aspects of security and trust.

Because SE-TIP maintains workflow information as one essential resource, there is also an in-direct relationship to business process management (BPM) and business process intelligence (BPI) as well as to SOA, and particularly to the area of governance which is reflected in SE-TIP through the objects that deal with references and style rules.

## 2.9    SE-TIP Next Steps and Future

One of the advantages of SW technologies is that we can build graphs of information facts without having to decide on a predefined and fixed data schema as it is the case when designing information structure schemes. Sometimes we might not even have a schema for our information model at all, see, for example, the ongoing discussions on how to effectively organize terminologies and translation memories in a sharable manner. Unlike relational database technologies, RDF allows us to combine information in arbitrary ways, without having to adhere to a data layout that is defined and fixed in advance of an application's deployment.

To fully employ the power of RDF, OWL, etc. in interoperability scenarios, RDF in attributes (RDFa) might fill an initially existing technology gap by bringing RDF to pure XML based approaches. While RDFa is targeted primarily at the human use of the Web, we believe it is also useful as a first step for understanding and building distributed Web-scale applications in combination with our SE-TIP approach.

The premise of RDFa is that Web documents can convey both presentation and semantic information. Through the use of XML attributes, presentation constructs are annotated with semantic information. This allows software applications other than webbrowsers to process and reason over the embedded information. As an example, Figure 4 exemplifies how an XHTML nanopharm translation ticket – here an offer for a translation task – could be presented in a way that allows both the person Joanna Da Rui and a software application to process the ticket appropriately. In the example, the relevant data elements are highlighted with a bold font.

A webbrowser can render this information for a human to read, while a software application that is part of a machine-to-machine interaction can extract the necessary information for making forward progress in a business process involving a translation offer for an individual.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
          "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
<html xmlns="http://www.w3.org/1999/xhtml"
    xmlns:nanopharm="http://nanopharm.com/vocab#"
    xmlns:foaf="http://xmlns.com/foaf/0.1/"
    version="XHTML+RDFa 1.0" xml:lang="en">
  <head>
    <title>Glocalization Task Request</title>
    <!-- Digitally signed thumbprint of a ticket number -->
    <meta property="nanopharm:ticketnumber" content="1234-56" />
  </head>
  <body>
    <h1>Receipt for Translation Delivery 1234</h1>
    <p about="nanopharm:ticket-recipient">Dear
      <span property="foaf:name" typeof="foaf:Person">Joanne Da Rui</span>,
    </p>
    <p about="nanopharm:translation-product">Thank you for your recent translation
      service. Since you have been always an excellent foreign language expert,
      we would like to offer you an additional <span property="nanopharm:taskname">
      Glocaltask</span>.
    </p>
  </body>
</html>
```

*Figure 4: RDFa Translation Ticket Example*

For example, we might leverage RDFa statements in nanopharm's XML documents in order to avoid the initially expensive transition – in terms of costs and time – to fully fledged RDF, OWL, etc. for translation related interactions. Such a step-by-step move to SW technologies might be appropriate to introduce initial TIP based applications. In such a scenario the TIPP objects would be represented in RDFa instead of a RDF, OWL, etc.

Last but not least, the introduced approaches to interoperability between TMS obviously allow for a seamless integration into the Web architecture (Fielding, 2000, and Richardson, 2008).

## 3 Conclusions and Perspective

In this paper, we have presented a semantically enriched version of TIP which further extends this solution to overcome the interoperability shortcomings of todays GILT industry. Based on the needs and requirements of the example company nanopharm, we have outlined the capabilities and potentials of the SE-TIP solution, and also shown that it is very important that a shared understanding of exchanged data and information does not get translated into a shared way of processing that data and information. Participants in loosely coupled distributed applications, as it is the use case with different TMS, shall remain free to deal with the data and information they receive in any way and by any tool they wish, but with the ability of a shared understanding.

Natural language specifications provide a mechanism for designers and developers to agree on the meaning of the data they exchange and share. However, as the volume, complexity and scale of distributed data and applications grow exponentially, it is important to consider a repres-

entation of information that employs machine-processable formats. Today, SW technologies are ready and mature to support the definition of data formats, protocols, and contracts.

SE-TIPP contains the data, information and knowledge that is necessary to fulfill the GILT tasks of nanopharm in an effective and efficient way encoded in SW formalisms and processable by machines. This encoding model provides the representation basis to ensure full interoperability based on a shared understanding of the resource descriptions. In addition, SE-TIP can also be seen as an enabler of forthcoming cloud-based services and sustainable language resources ecosystems (see Andrä and Schütz, 2009; and Andrä and Schütz, 2010).

## References

Sven C. Andrä and Jörg Schütz. 2009. MT Bazaar: Translation Ecosystems in the Cloud. In *Proceedings of the 12th Machine Translation Summit*, pp. 395-402, Ottawa, Ontario, Canada, August 26-30.

Sven C. Andrä and Jörg Schütz. 2010. Effectual MT within a Translation Workflow Panopticon. In *Proceedings of the 9th Conference of the Association for Machine Translation in the Americas (AMTA),* Denver, Colorado, USA, October 31-November 5.

Micah Bly. 2011. XLIFF Representation Guide for Documents. Internal Working Paper of the Interoperability-Now! Initiative. Draft v0.9.0, August.

Roy Thomas Fielding. 2000. Architectural Styles and the Design of Network-based Software Architectures. PhD Dissertation, University of California, Irvine, USA.

Leonard Richardson. 2008. Introducing Real-World REST. Presentation at Qcon (particularly: Act 3), San Francisco, California, USA, November 19-21.

Jörg Schütz. 2010. Semantic Technologies in Multilingual Business Intelligence. Invited Talk at the 1st Multilingual Web W3C Workshop, Madrid, Spain, October 26-27.