Collaboration and the Knowledge Economy: Issues, Applications, Case Studies
Paul Cunningham and Miriam Cunningham (Eds)
IOS Press, 2008 Amsterdam
ISBN 978–1–58603–924-0

STASIS - Creating an Eclipse Based Semantic Mapping Platform

Sven ABELS, Stuart CAMPBELL, Hamzeh Ghazi Ahmad SHEIKHHASAN[1]
TIE Nederland BV, Antareslaan 22-24, 2132 JE Hoofddorp, The Netherlands
Tel: +31-20-658-90 00, Fax: +31-20-658-90 01, Email: sven.abels@tieGlobal.com, stuart.campbell@tieGlobal.com, hamzeh.sheikhhasan@tieGlobal.com

Abstract: Within this paper the STASIS approach for creating a comprehensive application suite is introduced which allows enterprises to simplify the mapping process between data schemas based on semantics as opposed to syntax. This paper initially introduces the current schema mapping problem and outlines the limitations of existing solutions. The STASIS approach is then presented and contrasted with other semantic projects.

Keywords: Semantic Mapping, Schema Mapping, GMF, EMF, OWL, Eclipse

1. Introduction and Problem Description

In the last decade, a significant number of different e-Business related systems have been developed such as catalogue management tools or online shopping systems. Whenever two or more companies need to collaborate, or even applications within the same company, they invariably need to exchange information electronically and integrate the results in each system. In an ideal scenario, this information exchange and integration is performed in an automatic way allowing business partners to interoperate information seamlessly. However, because of the large number of diverse information systems the data format (syntax) of each exchange (message) usually differs from company to company, or sometimes even within the same company if more than one software product is used. The situation is similar to the Tower of Babel involving many different people that want to work together on a specific task without understanding each other.

This makes it a very challenging to exchange information in an interoperable way. Interoperability in this context means “the ability of two or more systems or components to exchange information and to use the information that has been exchanged” [4]. In those cases, users either have to agree on a common standard or they have to individually map the data format of one business partner to the data format of the other one.

The STASIS project (www.stasis-project.net), funded by the European Commission’s FP6 programme [5] is addressing this problem. It provides a set of semantic methods for easing the mapping process between XML, Flatfile, EDI and Database schemas and without relying on any precedence in the ontologies used – “competitive ontologies”.

It is addressing a simple key question, which is important for almost all European companies:

“If I have information in my format, and it is integrated into my systems, I want to put minimal effort into mapping this into any format (standardised or not) to do electronic business with another party” [5].

2. Existing Solutions

Mapping processes between schemas today focus on a rather syntactical approach. Tools such as Altova XMLSpy[i] or TIE Integrator[ii] are used to successively map different

attributes of two different schemas to each other. Those attributes are in most cases either mapped in a 1:1 relationship or using various connectors such as Methlets (see [7] for explanations and details) for creating complex mappings.

Figure 1 shows an example, mapping the syntactical elements of a source XML Schema file (XSD) to a destination format.

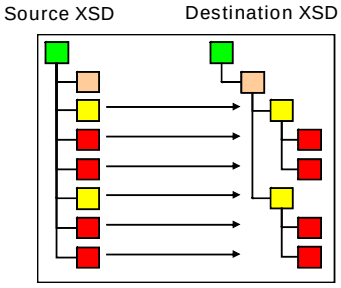

Figure 1: Syntactical Mapping

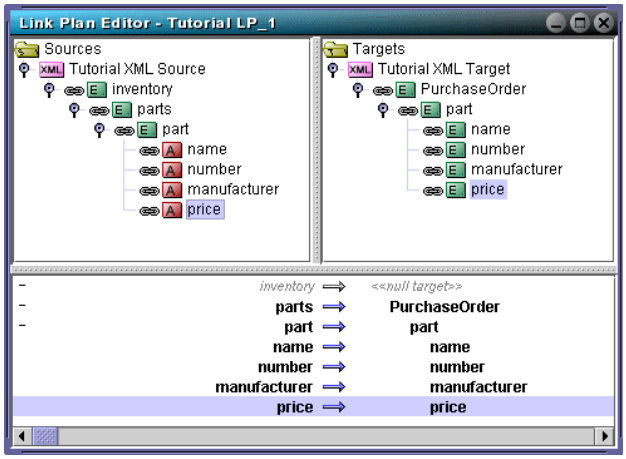

Figure 2: Link Plan creation with TIE Integrator

This approach has been pretty successful in the past and it allows fast mapping creation if the schema is not too complicated. However, restricting the mapping creation to a syntactical mapping process has some disadvantages. These include:

- The mappings must be performed by a technical person who has schema and syntax knowledge whereas what is wanted is the mapping of business information by business people. Of course comfortable GUIs such as displayed in Figure 2 may ease this task but nevertheless the process is rather technical involving skilled people that know both schemas.
- An in-depth knowledge of both the source and destination schemas is required in advance. New schemas need to be studied in detail before mapping. This is a very expensive step and cannot be avoided
- Mappings are often hard to understand and error-prone because humans normally think in terms of the actual semantics instead of the syntax – e.g. companies usually want to map two addresses instead of mapping Element [ELT4711] to [ELT0815]
- Mappings typically cannot be reused, and must be created from scratch when mapping to a new schema
- Concepts such as inheritance and logical constraints are not supported

3. STASIS Idea and Approach

The STASIS project aims to address many of the problems outlined above and also both introduce the concept of market-drive semantics – "Competitive Semantics" – and to promote a neutral standard mapping format which can be taken advantage of existing transformation tools and technologies.

3.1 Using Semantic Entities

Instead of focussing on syntactical mappings, STASIS concentrates on identifying semantic entities and mapping those semantic pieces. For example, two elements [ELT1] and [ELT2] may be grouped into one logical entity called "Address" which is mapped to a well-defined concept of an Address. After identifying those logical elements and linking the syntactical

elements to it, users may concentrate on using those semantic elements for creating their mapping. This allows users to connect elements with identical meanings instead of having to focus on the syntactical elements of two schemas.

3.2 Based on Ontologies

This process is based on ontologies used to define and link semantic entities within a schema. Those ontologies have been defined by the STASIS project team in order to describe semantic elements and their relationship. The link between a semantic element and the syntactical structure of the original schema is expressed in those ontologies. Unlike approaches such as foam [8] and GLUE [2] STASIS does not intend to create an automatic Ontology Mapping, nor does it rely on semantic mapping approaches as analysed in [1]. Instead, it is the goal of STASIS to provide an easy to use GUI, allowing users to identify semantic elements in an easy way. This means that the STASIS project uses ontologies (expressed in the Web Ontology Language OWL) in the background in order to define semantic relationships but users will not notice this nor do they need to know what ontologies are.

3.3 Using STASIS

In order to use STASIS, organisations need to import (or design) their original schema files and identify semantic entries. The overall process involves the following steps:

1. Import of original schema data.
   For example, this can be an XML schema specification (XSD) or a database schema (SQL). The first prototype implementation supports both. Other formats such as EDI are planned for the next versions of the STASIS prototype implementation. All imported data is managed in a neutral format, called SNF (STASIS Neutral Format)
2. Identifying semantic entities
   After importing all data into the neutral format, users need to mark specific elements from their schema file as being semantic entities. For example, two elements [cFirstName] and [cLastName] may be grouped into one logical entity called "Customer Name" as discussed earlier in this section. Each of those semantic entities is called an SSE (STASIS Semantic Entity).
3. Linking to generic concepts (optional)
   Once the semantic entities have been identified, users may decide to link their semantic entities to an overall concept that is provided by STASIS. For example, users may link their semantic entity that they have called "Customer Name" to the global concept "Person Name". Linking own semantic entities to those global concepts allows STASIS to derive some of the mappings automatically. However, their usage is completely optional and may be skipped by the user if preferred.

Figure 3 gives a graphical overview about the process described above.

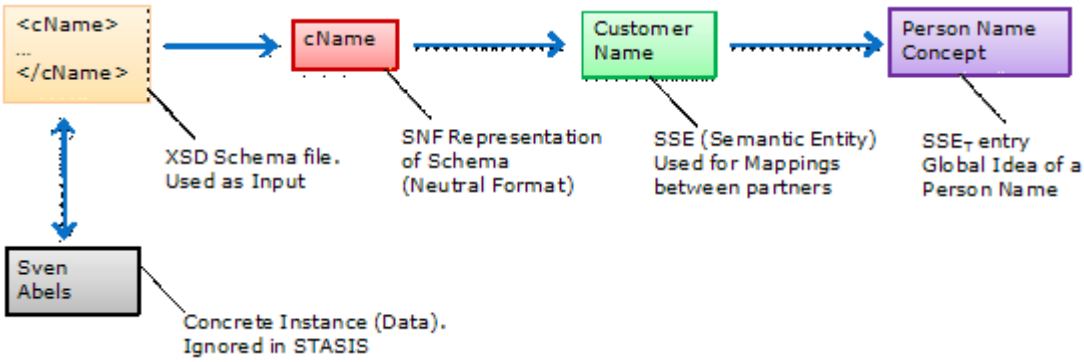

Figure 3: Importing data and preparing it for STASIS

The usage of STASIS is a straightforward process. At first, users have to import schema definitions into the SNF format. Afterwards they need to identify semantic elements and finally they may optionally link those semantic entities to generic concepts.

Once this has been performed, STASIS allows users to map their semantic entities to those of their business partners by simply connecting their Semantic Entities to the Semantic Entries of the business partner as displayed in Figure 4. This allows users to create mappings in a more natural way by considering the meaning of elements rather than their syntactical structure. Some elements might even be connected to a common third party or a chain of them. STASIS automatically detects if such a third party exists. This process will be described within the following section of this paper.

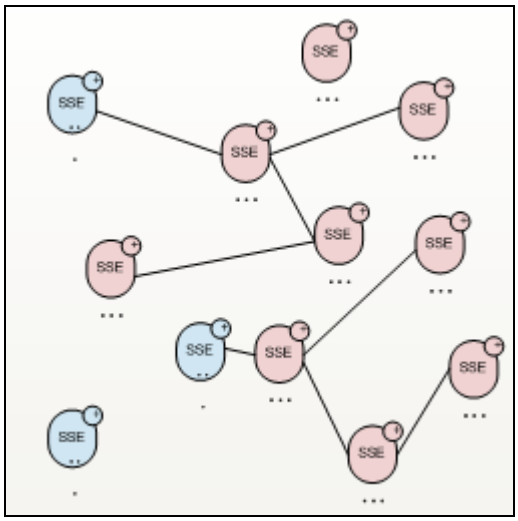

Figure 4: Connecting SSEs of two business partners.

4. Data Distribution and Sharing

All mappings that have been created by STASIS, as well as all semantic entities, are managed in a distributed registry and repository network called the SRRN (see [6]). The SRRN is a cutting-edge Peer-to-Peer network based on WebServices that allows the storage and retrieval of semantic elements in the OWL format. The SRRN has been developed within two European Research projects as described in [6]. It utilises the semantic query language SparQL, [9].

The SRRN is used within the STASIS project in order to allow sharing information among different STASIS users. This gives STASIS another significant advantage over traditional mapping creation tools as STASIS may reuse all mappings. It allows STASIS to make mapping suggestions by reusing mapping information from earlier semantic links. - For example, imagine two companies A and B that would like to map their business schemas in order to exchange information. Let's assume that both have conducted business with company C already in the past. In this situation, STASIS can map the semantic entities of A and B automatically because it knows about their mapping to a common schema from company C. This will become increasingly beneficial as more companies begin to use STASIS and turns into a significant mass of mappings. In addition to this, links of semantic entities to global concepts - as described in the last section – can be used to generate additional links automatically.

5. Technology: From OWL to EMF and Back Again

In order to realize STASIS, the consortium has selected the Eclipse Framework as a base for creating the Graphical User Interface of the mapping editor. Eclipse provides much in-

built functionality that may be reused and will therefore reduce the development time in most scenarios.

A definition of what a semantic entity is, and on how a schema is represented in STASIS, is expressed in the STASIS Common Data Model (CDM), which itself is expressed in the ontology format OWL. This logical data model is the base for the graphical editor of STASIS. In order to create this editor, STASIS is reusing two other core concepts of Eclipse: EMF and GMF.

- EMF (Eclipse Modelling Framework, see [3]) is used to create a definition of the CDM model. This is done by describing the CDM model in eCore, which is used by the EMF framework as a base for expressing models (see [3]).
- GMF [3] is a framework that allows creating a graphical editor that is based on EMF.

The overall process is illustrated in Figure 5 and is explained in detail in [10]:

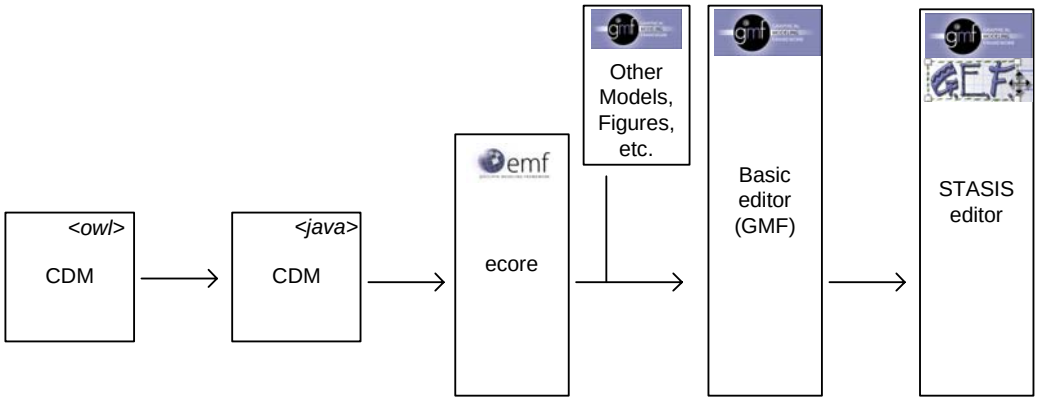

Figure 5: Editor Creation (refer to [10] for details)

As shown in the figure, the overall CDM definition is exported into annotated java classes using the Protégé[iii] EMF export methods. In the next step, the Eclipse functionality is used to derive the model file (eCore) as well as to derive a basic version of the graphical editor (GMF), which is the base for the STASIS schema editor component.

Using OWL parser libraries allows us to continue using OWL based ontologies with the STASIS editor as displayed in Figure 6.

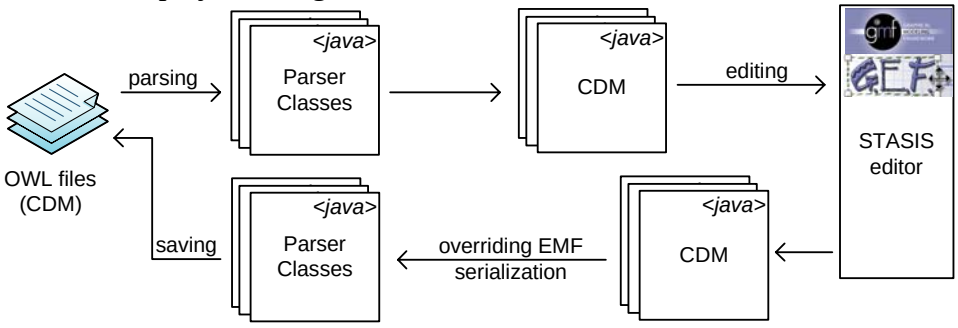

Figure 6: Roundtrip (refer to [10] for details)

6. Prototype Implementation and Results

STASIS has been realized as a prototype already and first results are available allowing the team to start with a formal evaluation process in real world scenarios. Figure 7 shows a screenshot of the first beta version.

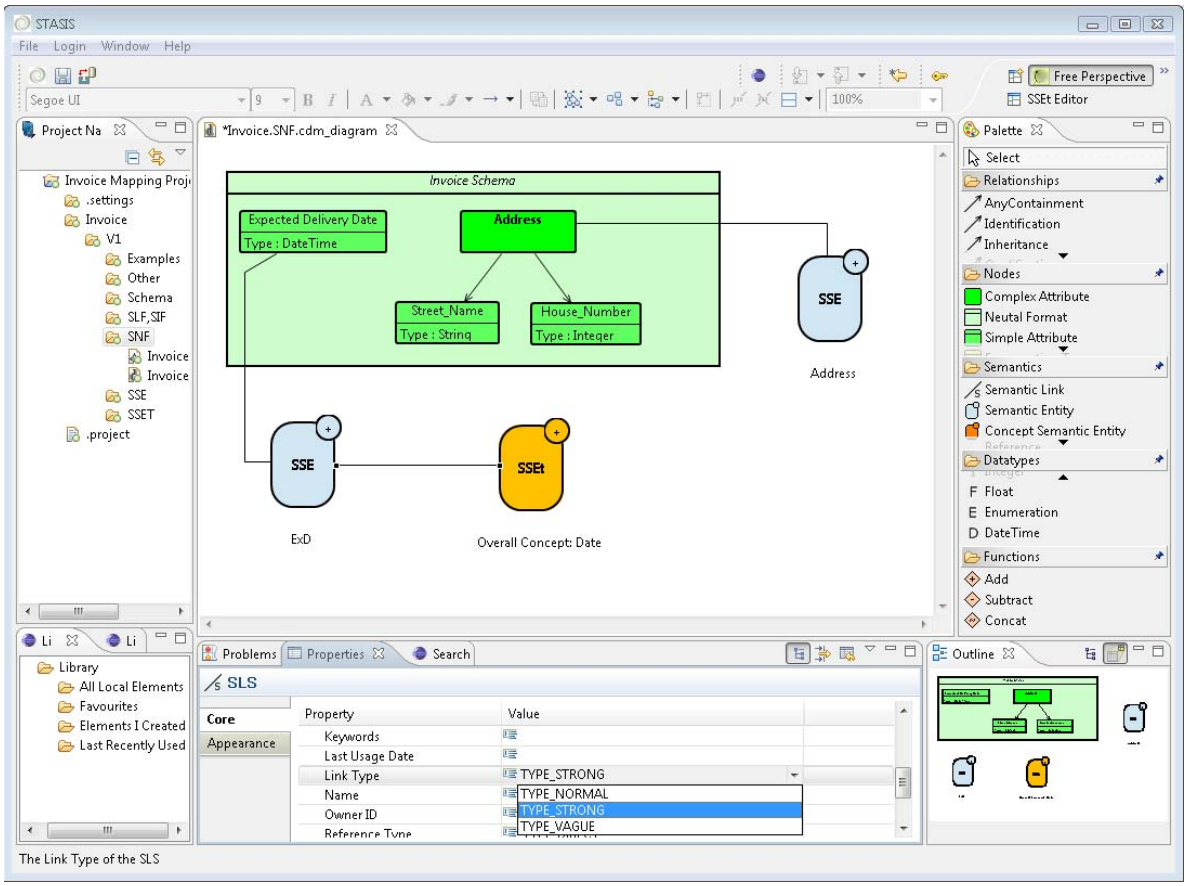

Figure 7: A screenshot of the current prototype

The screenshot shows the main modelling area of STASIS containing a very simple schema on the upper left called “Invoice Schema”. This schema contains two elements: (i) a DateTime element called “Expected Delivery Date” and an “Address” which is consisting of two sub-elements called “Street_Name” (String) and “House_Number” (integer). In addition to this, the diagram shows two semantic elements (SSEs) called “Address” and “ExD” and one global concept called “Date”.

Users may also use a set of functions to connect elements. For example, Figure 8 shows how the two schema elements “Street_Name” and “House_Number” are connected to an SSE “Address” using a concatenation function. STASIS provides additional dialogs to defile those functions and to define the order to elements.

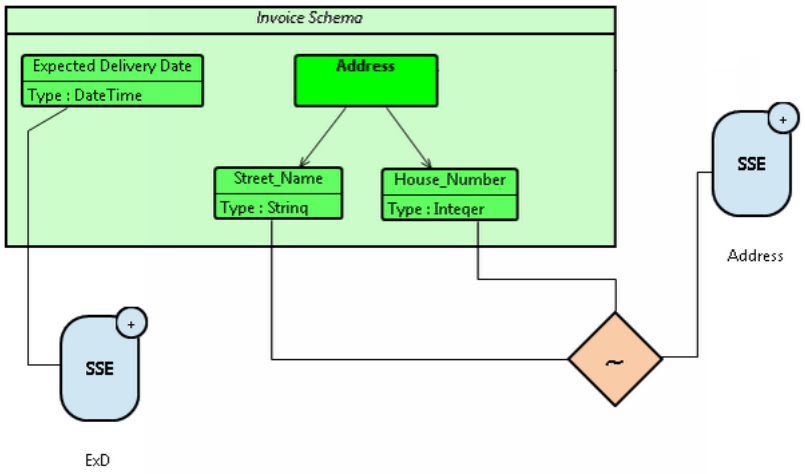

Figure 8: A complex link between schema elements to a semantic element

7. Conclusions and Summary Recommendations

The STASIS project and its implementation is a 'work in progress' and is due to be completed in 2009. In the current state, the main editors have been already implemented as well as most of the modelling functionality. However, there is a still a huge potential for extending STASIS.

Features that will be integrated in the next phase include a query environment allowing users to use simple dialogs for querying the SRRN network for existing elements. In addition to this, STASIS will be able to export mappings into XSLT in the next prototype implementations.

In parallel to the implementation, the STASIS prototype is evaluated in two real-world scenarios from the Automotive and the Furniture domain. The results will be used as an input for the third prototype, which is scheduled for the end of 2008. This will ensure that STASIS is applicable in reality and it will ensure that STASIS is usable by non-experts.

References

[1] Abels, S.; Haak, L.; Hahn, A.: Identification of Common Methods Used for Ontology Integration Tasks. Proceedings of the first international ACM workshop on Interoperability of Heterogeneous Information Systems, ACM, 2005.

[2] Doan, A. et al: Ontology matching: A machine learning approach. In Handbook on Ontologies in Information Systems, Springer-Verlag, 2004.

[3] The Eclipse Project, http://www.eclipse.org, 2008

[4] Institute of Electrical and Electronics Engineers: IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries, 1990.

[5] STASIS project, SofTware for Ambient Semantic Interoperable Services, http://www.stasis-project.net, 2008

[6] SEAMLESS project, Small Enterprises Accessing the Electronic Market of the Enlarged Europe by a Smart Service Infrastructure, http://www.seamless-eu.org, 2008

[7] TIE Integrator, An easy-to-use system that can transform and apply business rules to data in any format, passing through enterprise systems around the clock, Product Brochure, http://www.tieglobal.com, 2004

[8] Ehrig, M., Sure, Y. Ontology Mapping by Axioms. In: Proceedings of the 3rd Conference Professional Knowledge Management. Springer (Heidelberg), 2005.

[9] SPARQL Query Language for RDF, W3C Recommendation 15 January 2008, http://www.w3.org/TR/rdf-sparql-query/, 2008

[10] Abels, S.; Sheikhhasan, H.; Cranner, P.: Simplifying e-Business Collaboration by providing a Semantic Mapping Platform. In: Proceedings of the 4th International Conference "Interoperability for Enterprise Software and Applications" (I-ESA 2008), Workshop on "Semantic Interoperability: A Practical Approach", Germany, 2008

[i] http://www.altova.com
[ii] http://www.tieglobal.com
[iii] http://protege.stanford.edu