# Minimized Domain Knowledge for SOA-based Interoperability

Azin Dehmoobad and Kamran Sartipi Department of Computing and Software, McMaster University Hamilton, ON, L8S 4K1, Canada {dehmooa, sartipi}@mcmaster.ca

#### Abstract

The variety and heterogeneity of legacy systems at the application level have contributed to the complexity of interoperability provision among different application domains. In this context, most research activities are focused towards standardization and interoperability among the legacy systems within the same domain. However, an emerging challenge is to address the communication of information among heterogeneous legacy systems in different domains. The first step in achieving such a large interoperability is to follow similar development processes for collaborating domains, which provides homogeneity in their architectures. The second step would be to provide cross-domain semantic interoperability through proprietary and shared ontology systems. In this paper, we address the above challenges through description of a framework that employs healthcare standards and clinical terminology systems to achieve semantic interoperability between distributed systems in different domains. The main focus in our proposed framework is the minimal use of domain knowledge for cross-domain interoperability. Two case studies are provided, first we present how HL7 v3 is over-specified and then the proposed framework is applied to achieve semantic interoperability between two domains healthcare and insurance.

KEYWORDS: Cross-domain interoperability; Healthcare; Insurance; HL7 v3; Minimized Domain Knowledge.

# 1. Introduction

Due to increasing popularity and adoption of distributed systems, heterogeneity has become a major issue in interoperating among existing enterprise applications. Systems may be distributed in terms of data, computing and users, hence there would be several advantages to allow stand-alone systems to interoperate via well-defined services and data [3]. Most of these systems are multilingual and have been developed in different platforms so interoperability between them is a major challenge. However, due to geographical distances, integration of such systems into a monolithic system is not an option anymore. In this context, standardization is a key requirement for providing interoperability at different levels of communication hierarchy which prevents conflicts such as overlap, incompatibility and mismatching [15]. These standards may be either domain-specific or domain-neutral; in either case, there should be a proper mapping between different standards within different levels of interoperability to ensure effective operations. Developing domain-neutral standards will allow cross-domain interoperability among applications from relevant domains such as banking, insurance and healthcare to collaborate and maintain the quality of services across the domains.

Within the application domains that have embraced IT for decades, including: banking, government, reservation systems and tele-communication, most legacy systems communicate through vendor's proprietary process with no standard representation of information and messages. On the other hand, domains such as healthcare that have already experienced much difficulties in communicating medical and clinical terminologies, have developed comprehensive information and concept representations that will allow consistent interpretation of concepts among heterogeneous legacy and new healthcare systems. Within such a development framework, the domain information undergoes a sequence of refinements from a comprehensive body of knowledge representation (as class diagram) down to interoperable concepts and terminology hierarchies that are understandable by all relevant stakeholders within the same domain. Similarly, standard functionality and operations within the domain are incrementally refined from task scenarios down to a collection of standard messages that will be consequently populated by the above standard concepts and terms to interoperate. In such a generic message development framework, service oriented architecture is perfectly applicable to provide the necessary abstraction at the business rule level while maintaining the standard and vendorindependent lower-level technologies such as web services that warrant seamless interoperability at different granularity levels. In this context, task forces in different application domains (such as healthcare) have developed their own set of standards for interoperating at business rules to low-level communication protocols which hinders further interoperability across other domains such an insurance and banking.

In this paper, we address such a problem, namely "*over-specifying domain specific interoperability standards*" and propose a framework to design a cross-domain interoperability standard based on a minimal amount of domain-specific knowledge during communication between applications in two relevant domains. As a case study, we will consider HL7 v3 messaging standard in healthcare domain and will address different interoperability levels according to our framework by using web services incorporated to HL7 v3. We also present detailed arguments in adopting web services standard transmission infrastructure instead of some specifications in HL7 v3 messaging standard.

This paper has been organized as follows. Section 2 provides related work on interoperability within a domain. Section 3 briefly introduces the required technology background for the paper. Section 4 is allocated to the proposed framework and two case studies. Finally, Section 5 provides the concluding remarks.

## 2. Related work

Interoperability between heterogeneous systems has been considered in different domains, such as airport, healthcare and military. Airport interoperability standards address the following issues: information exchange model, mapping to database, spatial data standard for facilities, infrastructure and airport environment [12]. Harmonization efforts among these standards aim at filling the gap between these standards to allow them to work together. Janssen et al. [11] and Guijarro [10] leverage interoperability and address issues in electronic governance. Homann et al. [19] discuss an interoperability framework for integrating banking systems and present a case study on two European banks using web services. These approaches try to achieve standards in semantic interoperability and different domains are developing their own standards and face the same problems that healthcare domain has already attempted. In our approach, we present a framework to solve these inconsistency issues by generalizing HL7 v3 development process.

Donachy et al. [16] discuss the requirements for high quality assurance within SOA and grid infrastructures. Also there are other efforts to propose architectures and frameworks for interoperability by organizations and software vendors. CORBA (Common Object Request Broker Architecture) [3] is OMG's vendor-neutral architecture that computer applications use to collaborate over the networks. Oracle's Healthcare Transaction Base (HTB) [5] provides a means to create a comprehensive patient record that can be shared across institutions and geographic regions. Motahari et al. [14] propose a conceptual framework for analysing web services interoperability issues. In contrast to proposing different vendor-based products for interoperability, we recommend to use web services which are globally accepted and allow the users to set up low-cost networks to join.

In recent years, the proposed frameworks for interoperability between different systems have been evaluated. Lewis et al. [13] try to identify limitations of interoperability standards. They focus on two areas: semantic and organizational levels of interoperability, and provision of quality of service. Their approach concludes that standards are not enough because of capability of extension and customization and life cycles of standards, and also refer to HL7 v3 as a conflicting standard. In our proposed framework, we adopt a domain-neutral standard that to some extent resolves the above mentioned problems. Mykkanen et al. [15] propose a framework to evaluate interoperability standards; they use a case study of HL7 v3 messaging standard that is defined for scheduling sub-domain. There are some limitations that are addressed in the above two approaches that can be resolved by applying HL7 v3 process to generate standards.

# 3. Background

In this section, we briefly present the required background technologies for the proposed framework.

#### HL7 version 3 standard

Work on HL7 started in 1987 as a non-profit organization to provide standardization for data exchange between departments in the medical information system field, mainly in the USA. Standardization for version 3 (v3) supports electronic patient records using XML for document representation as well as the latest modeling, methodology, and tools. As a starting point, the HL7 v3 methodology uses: Reference Information Model (RIM), HL7-specific vocabulary domains, and data type specifications [7].

The refinement process specified in HL7 methodology is as follows. Domain Message Information Model (D-MIM) is a subset of the RIM that includes a fully expanded set of class clones, attributes and relationships that are used to create messages for any particular domain. Refined Message Information Model (R-MIM) is used to express the information content for one or more messages within a domain. Each R-MIM is a subset of the D-MIM. Hierarchical Message Description (HMD) is a tabular representation of the sequence of elements represented in an R-MIM.

Service Oriented Architecture and Web Services Service oriented architecture plays a key role to provide an architecture for integrating standalone systems and enable interactions among them. There are major benefits in using service oriented architecture as an infrastructure for integration and message passing between systems: it takes low effort for the user to join a SOA architecture, and it is independent of different implementation platforms. To implement SOA, Web Services allow the applications to expose software services using standard interoperable protocols, regardless of the platform on which they are implemented [14].

#### ACORD

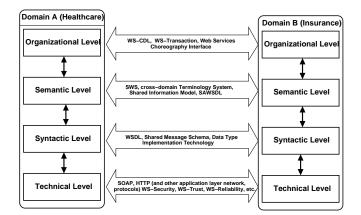
ACORD (Association for Cooperative Operations Research and Development) is a global, nonprofit insurance association. The purpose of the ACORD standards program is to provide the insurance and related financial services industry with a structured process in which industry participants may work cooperatively to create data standards for exchanging information in support of eBusiness strategy within the insurance industry [1].

## 4. Proposed framework

Figure 1 illustrates our proposed framework for crossdomain interoperability among closely related application domains based on WS-\* family of technologies. In this context, standards are needed at each level of interoperability, namely technical, syntactic, semantic and pragmatic. These standards could be either domain specific or domain neutral. The framework is intended to minimize the use of domain-specific standards that traverse different interoperability levels. This minimization of domain knowledge facilitates interoperation among systems in different domains due to requiring low domain expert knowledge to develop an interoperability middleware. Thus, domain-neutral standards are needed to cover as many levels of interoperability as possible. The underlying architecture to achieve crossdomain interoperability uses service-oriented architecture that also adopts web services technology as well as the domain-neutral and widely-used WS-\* family of specifications. In the followings, we describe the four levels of interoperability within our framework. We have also provided two case studies, one for technical level and one for semantic level of interoperability, with refer to two standards ACORD and HL7 v3 in insurance and healthcare domains, respectively.

## 4.1. Technical level

The technical level refers to the data transportation aspects such as: security, reliability, and authentication. There is no need to apply specific domain knowledge at this level. For example, WS-\* family have a set of



# Figure 1. Cross-domain interoperability technical framework based on WS-\* family of technologies with minimal use of domain specific knowledge.

specifications for messaging that are developed and widely used by industry and can be applied to cover all the required message passing requirements. As shown in Figure 1, existing protocols such as SOAP and HTTP can handle technical interoperability and WS-\* family specifications can be developed on top of SOAP to add more capability to message passing. In the followings, we will discuss to what extent technical issues have been covered by domain specific standards. In this case, HL7 v3 transmission wrappers are studied and we conclude that these wrappers have been completely handled by the WS-\* family technology, and hence it is efficient to leave such responsibilities for WS-\* family to implement message passing technical infrastructure.

#### Over specifying in HL7 v3

In our approach, "over specifying" refers to a case where a domain-specific standard violates its defined boundaries by specifying standards of another interoperability level. We argue that HL7 v3 (as a domain specific standard) should not specify the requirements for technical level of interoperability infrastructure for message transmission between systems. However, such an "over specifying" makes it difficult to use these message types in service oriented architectures. In HL7 v3 messaging, transmission wrappers are the outermost layer and they have been designed to cover transmission issues such as: acknowledgement messages, packaging and routing messages, identification of sender and receiver, transport specifications, and attributes that address the message handling of the receiver counterpart. These transmission wrappers consist of the attributes shown in Table 1. However, all these features (except those related to payload of the transmission) can be handled by the corresponding web service protocols and WS-\* family facilities. Table 1

HL7 v3 Transmission Wrapper Attribute	Goal	Web Service Facility
id	Transmission Identification	WS-Addressing
creationTime	The Creation Time of the Transmission	WS-Security
interactionId	Unique Identifier of the Interaction	WS-Addressing
responseModeCode	Explanation for Time of the Required Response	WS-Addressing
securityText	Extra Security Features of a	WS-Addressing and WS-Security
versionCode	Transmission Identification	WS-Addressing
Sender	Identification of the Sender	SOAP
Receiver	Identification of the Receiver	SOAP
respondTo	Identification of the Application to respond to for this	SOAP
	Transmission	
AttentionLine	Representation of a Technology Specific Data	N/A
ControlActProcess	Body of the HL7 v3 Message	Message Body
AttachmentText	Attachments to the Payload of The Transmission	SOAP-Attachments
	which is Refereed to By Message Content	
typeCode	Acknowledgement Details	WS-Reliability
code	Acknowledgement Details	WS-Reliability
text	Acknowledgement Details	WS-Reliability
location	Acknowledgement Details	WS-Reliability

#### Table 1. Covering HL7 v3 Transmission Wrapper attributes with Web Service Facilities

represents a goal-based mapping between WS-\* family facilities and classes provided by the transmission wrappers.

## 4.2. Syntactic level

At this level of interoperability, the main concern is the structure and format of the data that are exchanged. The abstract data types specified by HL7 v3 documents [4] are used in our framework, and ISO (International Standard Organization) data types are used as an implementation of these abstract data types. The XML format that is used to generate messages is specific to XML platform and the message contents are derived from serialization of each message content in cross-HMD step of refinement that is defined in the semantic interoperability. The XML schema is generated from the tabular representation of the standard message contents which allows both the sender and receiver to refer to the same schema by the means of the standard message identifier. The detailed implementation specification of data types can be found in [9].

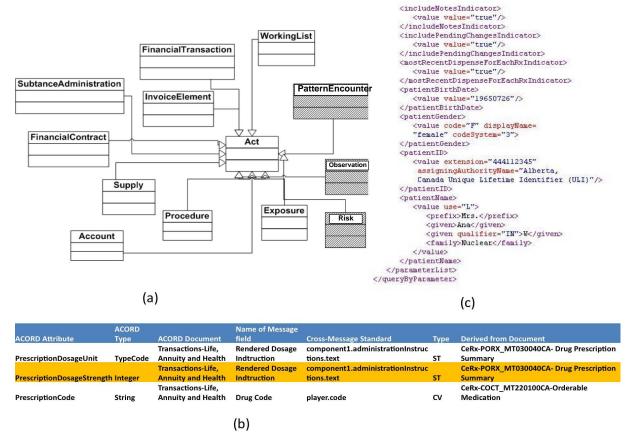
WSDL standard are applied both to expose services and to define signatures that should be used to invoke the services among the SOA components. WSDL is an XML document that describes services in terms of a series of communication endpoints and ports to expose them to service consumers. Abstract definitions of service components are separate from their concrete network protocols and data format binding. This binding is the focus to achieve syntactic interoperability. This feature is illustrated in the syntactic level of Figure 1. Therefore, at the syntactic level the required domain-neutral interoperability has been provided by the communication abstraction of web service description languages (WSDL) which provides a common method of accessing the required domain-specific data types and a common grammar to parse the XML messages.

#### 4.3. Semantic Level

At this level of interoperability, domain specific knowledge is widely needed. We propose a framework to ensure semantic interoperability between systems in different application domains with minimal effort to use each domain's specific knowledge or standard. At this level of interoperability, we require an information model, a terminology system, and a shared set of data types. Our framework for interoperability follows the Hub-and-spoke pattern [17] as opposed to existing point-to-point solutions. Point-to-point patterns are complicated and inefficient when data sources grow over the time. In the followings, each component of this framework is discussed in detail.

#### Shared information model

In order to achieve cross-domain interoperability, we propose to use the same process of refinement as in HL7 v3 information model [7] to build a consistent information model between different domains. We adopt a Core-RIM that represents the common set of classes, attributes and relationships between classes among all the existing domains. The Core-RIM is derived from HL7-RIM and consists of classes that are not specific to healthcare. Examples include: WorkingList, Procedure, and Exposure. For each set of scenarios to perform information exchange between two domains, there exist a cross-DMIM which is a clone of classes of Core-RIM that are constrained to the requirements of that set of scenarios. Further refinement is performed to generate cross-RMIM for each transaction of a scenario and cross-HMDs for their required interactions. This framework has a bottom up approach, where the steps



# Figure 2. (a) Extended RIM from the Core-RIM. (b) Mapping between ACORD message fields and cross-domain message fields. (c) One message sample: Response to Prescription Summary Query.

for building the shared information model are discussed below.

*i) Scenario definition*: first we describe a set of scenarios that require data flow between the domains. For example: "the insurance domain may want to receive pharmaceutical information of a person from a pharmacy". This step should generate a set of use case diagrams.

*ii) Transaction extraction*: for each scenario we extract its use cases as separate transactions, where each transaction can be represented by an interaction diagram.

*iii) Interaction extraction*: each transaction is realized by one or more interactions, each of which is a single data flow from one application to another.

*iv)* Information definition: for each interaction we define a set of information to be exchanged. These information should be restricted only to those needed for that specific data exchange. The output of this step is a set of class diagrams, attributes and their relationships.

v) Mapping to Core-RIM: the information from the previous step will be mapped to the Core-RIM classes, attributes and relationships. vi) Extension points: after the mapping step, if there is any further information remained (i.e., classes, attributes or associations), we extend the Core-RIM and generate a new class diagram for RIM. For the new generated classes that should be specializations of the foundation classes of RIM (namely, Act, ActRelationship, Role, RoleLink, Participation and Entity or the subclasses) we find the type of the attributes from the shared data types. As an output a new class diagram for Core-RIM is provided.

*vii) Cross-DMIM*: for each pair of application domains (e.g., healthcare-insurance) we clone all classes that are needed to communicate between the domains and perform a refinement in terms of cardinality, relationship names, etc.

*viii) Cross-RMIM*: for each transaction extracted in step ii, we develop an R-MIM and finalize the information that we should put in each message for each interaction.

*ix)* Serializing the information for each message (cross-HMD): in this step we develop a tabular representation of data in each message independent of any implementation technology.

*x) Generate message schemas*: according to the tables produced in previous step, message schemas for each inter-

action is generated.

*xi) Mapping*: a mapping between the cross-domain information model to domain-specific information model is provided to make the system process the received information properly.

In this framework two steps iv and v are meant to minimize domain-specific knowledge. We have a Core-RIM that guarantees all the produced messages are derived from the same information model. The refinement process for each scenario (also can be considered as information categorization for cross-domain interaction) is used to manage vocabularies, class associations, and mandatory attributes in each interaction. For a detailed description of refinement process refer to HL7 v3 Ballot [7]. Future trends may use this information model to semantically annotate WSDL and expose it to have a complete set of semantic and syntactic interoperability.

#### Shared terminology system

The shared terminology system possesses the same architectural style as SNOMED CT terminology system. It consists of concepts that are logically defined by relationship to one or more other concepts. Formal rules for *postcoordinated* expressions are used to make this terminology system precise in terms of relationships between concepts. Any concept can be refined using this formal rule. Concepts are represented in a *compositional grammar* [8].

In our case study to achieve exchanging pharmaceutical information across two different systems in different domains, we accepted SNOMED CT vocabulary system architecture and added concepts needed to be exchanged to the whole terminology system. To expand terminology system to include insurance specific concepts, we also added concepts that are used for exchanging pharmaceutical information in ACORD *Life and Annuity Standards Licensing and Appointments Implementation guide V2.1* Lookup section [6].

#### **Shared Data Types**

To have a meaningful data exchange, definition of the values that are exchanged is inevitable. Any data element within a data flow between two systems has a data type. HL7 v3 messaging standard uses a complete set of external data type systems and different implementation technologies can be employed as mentioned in HL7 v3 Ballot [9]. For our case study we used HL7 v3 data type system (as shared data types) due to its comprehensive coverage of all data types defined by ACORD.

# 4.4. Pharmaceutical interoperability between insurance and healthcare

We proposed a framework to achieve semantic crossdomain interoperability between closely related domains. As a real world case study we present the core of the information model to generate a set of messages to exchange pharmaceutical information between insurance and healthcare systems. As stated by ACORD standard documents, the life insurance industry is quickly moving to explore other data sources in their underwriting and decision process. We have reviewed the following documents for this case study: ACORD Life, Annuity and Health standard documents [6] and CeRx (Canadian Electronic Drug) messaging documents specifically COCT-MT220100CA- Orderable Medication and CeRx PORX-MT030040CA-Drug Prescription Summary [2]. We applied the framework steps that were defined in Subsection 4.3 and the details are presented below:

*i)* Scenario definition: the scenario is to explore the external pharmaceutical databases by an insurance application which provides inter-domain interoperability with ACORD standard.

*ii) Transaction extraction*: different transactions are explored but the one that is selected for this case study is *Pharmaceutical Information Transmittal* tc=1601 from ACORD that directly involves exchanging information with a pharmacy or a healthcare system. This transaction includes two interactions *TXLifeRequest Data Stream Requirements* and *TXLifeResponse Data Stream Requirements* [6].

*iii) Interaction extraction:* the selected interaction is *TXLifeRequest Data Stream Requirements* which is the response to the request to get *Drug Prescription Summary* from a pharmacy or healthcare system.

*iv)* Information definition: the required information for this transaction is selected and its class diagram is generated with its attributes and relationships. The information that is needed to be exchanged between healthcare and insurance parties were extracted from the CeRx documents on *Pharmacy Drug Summary* and *Life, Annuity and Health* documents in insurance. We first selected the intersection of all data that are shared between these two domains and then we added the data that are required by one of the parties (either healthcare or insurance).

*v) Mapping to Core-RIM*: we mapped the class diagram generated in the previous step to the existing Core-RIM. The white-colored classes within Figure 2(a) illustrate Core-RIM classes and the gray-colored classes are the extended classes.

*vi)* Extension points: Figure 2(a) illustrates the comparison between Core-RIM Acts Subject Area with the very basic classes that we selected to be the Core-RIM and the specific extensions for the whole scenario of getting pharmaceutical information by an insurance party. The Observation and PatientEncounter classes (already in HL7 v3 RIM) are extended for healthcare requirements and the class Risk is for insurance side.

vii) Cross-DMIM: following the refinement process, we

select the classes from the above extended Core-RIM that correspond to the scenario and apply cardinality, vocabulary, and type constraints. The produced Cross-DMIM include the class from the Acts Subject Area: *PatientEncounter, Observation, Risk, Exposure, Supply*, and *SubstanceAdministration* in Figure 2(a).

*viii) Cross-RMIM:* for transaction code tc=1601 in LAH ACORD standards we group the classes into Pharmaceutical Information Transmittal R-MIM and refine the message information for each of the interactions.

*ix)* Serializing the information for one message (cross-HMD): using a tabular representation of the serialized data, a spreadsheet form serialized data for each of the messages is generated. The output of this step is two Excel files, one for the query interaction and one for the response.

*x) Generate message schemas*: using XML technology for message passing and XMLSpy tool, we generated the schemas for the request and response messages; one instance is shown in Figure 2(c). These messages are HL7 v3 compatible and syntactically and semantically are interoperable with HL7 v3 compliant healthcare systems.

*xi)* Mapping: to allow these messages work properly with ACORD standard, a mapping between these message fields and the ACORD message fields for tc=1601 is generated and presented in Figure 2(b).

#### **Pragmatic Level**

At this level, domain specific knowledge is required to provide a set of business processes between two or more domains. Furthermore, these business processes can be specified using WS-\* family facilities listed in Figure 1 such as BPEL. Anzbck et al. [18] provide a semi-automatic tool to generate these web services and BPEL files. To have pragmatic interoperability at the organizational level, Service Level Agreements (SLAs) are used to define responsibilities and agreements between industries to use their mutual services.

# 5. Conclusion

In this paper, we proposed a technical framework with a focus on minimization of domain knowledge to achieve cross-domain interoperability. We discussed, different levels of interoperability along with the list of WS-\* family specifications. For cross-domain interoperability, a core information model is adopted consisting of generic classes of the HL7 v3 RIM which are further extended to provide the means for representing shared and communicable information among domains. Furthermore, the semantic interoperability requires adoption of shared terminology and data type systems that allow seamless interpretation of the messages across the incorporating domains. As a proof of concept, two case studies were presented one at technical level to deal with over-specification of domain specific standards, and one in semantic level to signify the minimized knowledge for interoperability. One of the main features of cross-domain interoperability is that it does not need expert knowledge about the other domains to understand the cross-domain standards and communication steps.

## References

- Association for cooperative operations research and development. URL = http://www.acord.org/.
- [2] Cerx specifications
  URL = http://forums.infoway-inforoute.ca/CeRx/.
- [3] Corba. URL = http://www.corba.org/.
- [4] Data types abstract specification. url = http://www.hl7.org/v3ballot/
  . html/infrastructure/datatypes/datatypes.htm.
- [5] Healthcare transaction base . URL = http://www.oracle.com/industries/healthcare/htb.html.
- [6] Life and annuity and health standards programURL = http://www.acord.org/Standards/LifeMain.aspx/.
- [7] Refinement, constraint and localization, release 2
  URL = http://www.hl7.org/v3ballot/html/infrastructure/.
- [8] Snomed ct access in canada: August 2007 webcasts. URL = http://sl.infoway-inforoute.ca/.
- [9] Xml implementation technology specification, release 2
  URL = http://www.hl7.org/v3ballot/html/infrastructure.
- [10] L. Guijarro. Semantic interoperability in egovernment initiatives. Computer Standards and Interfaces, November 2007.
- [11] M. Janssen and H. J. J. Scholl. Interoperability for electronic governance. In *ICEGOV '07*, pages 45–48. ACM, 2007.
- [12] AMD Committee Meeting. Airport mapping standards integration. September 2006.
- [13] Grace A. Lewis, et al. Why standards are not enough to guarantee end-to-end interoperability. In *IEEE ICCBSS '08*, pages 164–173, 2008.
- [14] Hamid R. Motahari Nezhad, et al. Web services interoperability specifications. *Computer*, 39(5):24–32, May 2006.
- [15] J. A. Mykkänen and M. P. Tuomainen. An evaluation and selection framework for interoperability standards. *Inf. Softw. Technol.*, 50(3):176–197, 2008.
- [16] P. C. Donachy, et al. Finance sector: Requirements for high assurance within spatial soa based grid infrastructures. In *IEEE High As*surance Systems Engineering Symposium, 2007.
- [17] Mei Y. Selvage, et al. Achieve semantic interoperability in a SOA. IBM Inc., 2006.
- [18] S. D. Rainer Anzbek. Semi-automatic generation of Web services and BPEL processes - A Model-Driven approach, volume 3649/2005 of Lecture Notes in Computer Science, chapter Business Process Management, pages 64–79. Springer Berlin / Heidelberg, 2005.
- [19] M. R. Ulrich Homann and A. Wimmer. Flexible value structures in banking. In *Communications of the ACM*, volume 47, May 2004.