# UnIT-NET IEDI: Ukrainian National Infrastructure for Electronic Data Interchange

**Sections de rattachement: ...**
**Secteur : ...**

***RÉSUMÉ**. Electronic Data Interchange (EDI) is the computer-to-computer exchange of highly structured business data between one application and another within a trading community, in this case – among the universities and state organizations of Ukraine. The paper presents the reference architecture of IEDI which is specified in full detail in [BGE04]. IEDI is the software infrastructure supporting EDI. More precisely, IEDI is the multi-layered distributed information system comprising the software servers, services, components and tools for providing intelligent ontology-driven information retrieval from distributed, heterogeneous, legally and physically autonomous Information Resources (IRs) in the organizational framework of the National Higher Education System. Presented are the principles of architectural design, the processes which define IEDI functionality, the architectural layering and the components of IEDI, IEDI ontologies. It is planned in the UnIT-NET[1] project that IEDI will further on be used as the core to provide the services of data interchange for University management and for distributed research groups working on joint projects.*

***MOTS-CLÉS** : UnIT-Net, IEDI, architecture, mediator, wrapper, information system, ontology*

[1] **UnIT-NET**: IT in University Management Network. TEMPUS/TACIS multiplier project MP JEP 23010-2003, http://www.unit-net.org.ua/.

# UnIT-NET IEDI: Ukrainian National Infrastructure for Electronic Data Interchange

**Vadim Ermolayev*, Alexander Spivakovsky**, Grygoriy Zholtkevych*****

** Zaporozhye State University, Information Technology Department*
*66 Zhukovskogo st. 69063 Zaporozhye UKRAINE*

*** Kherson State University*
*27, 40 years October st. 73000 Kherson UKRAINE*

**** V. Karazin Kharkiv National University*
*4 Svobody sq. 61077 Kharkiv UKRAINE*

***eva@zsu.zp.ua; spivakovsky@ksu.ks.ua; Gregory.N.Zholtkevych@univer.kharkov.ua***

## Sections de rattachement :
## Secteur :

***RÉSUMÉ**. Electronic Data Interchange (EDI) is the computer-to-computer exchange of highly structured business data between one application and another within a trading community, in this case – among the universities and state organizations of Ukraine. The paper presents the reference architecture of IEDI which is specified in full detail in [BGE04]. IEDI is the software infrastructure supporting EDI. More precisely, IEDI is the multi-layered distributed information system comprising the software servers, services, components and tools for providing intelligent ontology-driven information retrieval from distributed, heterogeneous, legally and physically autonomous Information Resources (IRs) in the organizational framework of the National Higher Education System. Presented are the principles of architectural design, the processes which define IEDI functionality, the architectural layering and the components of IEDI, IEDI ontologies. It is planned in the UnIT-NET[2] project that IEDI will further on be used as the core to provide the services of data interchange for University management and for distributed research groups working on joint projects.*

***MOTS-CLÉS** : UnIT-Net, IEDI, architecture, mediator, wrapper, information system, ontology*

[2] **UnIT-NET**: IT in University Management Network. TEMPUS/TACIS multiplier project MP JEP 23010-2003, http://www.unit-net.org.ua/.

## 1. Introduction

To achieve and sustain dynamic improvement, service-oriented organizations like universities, need an infrastructure that underpins flexible and robust management of their activities and decision making support. To a large extent the activities within Universities as well as their coordination and control at national level involve the processing of enterprise data and knowledge. As far as the organizations involved in the Educational framework are rightfully independent, they own and maintain their data and knowledge sources autonomously – i.e. independently from each other and, to a high degree, from the coordination body, like a National Ministry. The fact that these information resources are autonomous implies serious complications for their integration. IEDI is the software infrastructure providing for the Electronic Data Interchange between the Universities and the State Bodies of Ukraine. More precisely, IEDI is the multi-layered distributed software system comprising the software servers, services, components and tools for providing intelligent ontology-driven information retrieval from distributed, heterogeneous, legally and physically autonomous IR in the frame of the organizational network of the National Higher Education System.

The reminder of the paper is structured as follows. Section 2 outlines the related work and the principles used in IEDI architectural design. Section 3 sketches out the architecture of UnIT-NET IEDI. Section 4 focuses on the family of ontologies which drive query decomposition, query translation information retrieval and query results mark-up in IEDI. Section 5 gives concluding remarks and outlines the directions of the future work.

## 2. Related Work

In the outlined context the genre of the IEDI falls down to the Distributed Intelligent Information Retrieval (I2R) domain within the broader area of Intelligent Information Integration (I3). The research activities within this domain have been intense in the past decade, especially within the Information Society Technologies Key Action Line of the EU FP6 and similar national and international frameworks. Examples of R&D projects developing the formal, algorithmic, architectural frameworks, deploying software prototypes for I2R from distributed, heterogeneous IR-s and Intelligent Information Integration (I3) are BUSTER [STU00], DOME ([CJO01], [CJO02]), InfoSleuth [BAY97], KRAFT [GRA97], MOMIS [BCD98], OBSERVER [KS00], Ontobroker [DEF99], PICSEL [LR00], SIMS [AKS96], TSIMMIS [GAM95], and others. A good survey of ontology-based approaches to I2R and I3 may be found in [WAC01].

Although all these projects use different techniques, approaches and software paradigms for the task, they identify similar pitfalls for the domain. The first group of possible pitfalls is the way in which semantic heterogeneity is resolved in the processes of ontology-based information integration. As outlined in [CJO01], this includes the questions of developing ontologies (bottom-up and top-down approaches), mapping between ontologies, and relationships between ontologies and information resources as data providers.

Most projects adopt one of the following approaches to using ontologies [WAC01]: single ontology (SIMS), multiple ontology (OBSERVER), hybrid approach (BUSTER, DOME). Mapping between ontologies is necessary when the ontologies architecture of the system works with several ontologies either “horizontally” (as in multiple ontologies approach) or “vertically” (as in hybrid approach). Mappings between ontologies within the system provide links between equivalent or related in elements of ontologies, thus ensuring re-use of ontologies. Mappings between ontologies and information resources schemas maintain correspondences between ontology elements and elements of the data schemas. As stated in [CJO01], the reasons for these mappings are:

– Data schema definitions are not always a good source of domain knowledge for people querying the system, they often play technical role;
– Queries posed to the system are expressed in the ontology-oriented query language not from the data schemas Thus a mapping between ontology elements and data schema elements makes for transparent execution of user queries within the system;

Other reasons for using mappings between ontology elements and data schemas of information resources are the requirements for information resource autonomy and openness of the system as a whole.

The second group concerns the questions of supplying autonomy and dynamic nature of the open system elements. The solutions here advocate one of the mediator architectures: centralized and decentralized. A centralized mediator architecture provides for one centre, which stores all the information about ontologies, information resources, mappings between them, and which controls the query formulation and execution. A known realization of this approach is TSIMMIS. A decentralized mediator architecture provides for each information resource a separate agent/wrapper, which stores mappings between global/shared ontology (-ies) and the underlying information resource (RACING [EKP03]). The resource broker communicates with resource agents/wrappers and determines relevant and accessible resources for every query personally (InfoSleuth, SIMS, KRAFT).

The third group of possible pitfalls is formed by the tasks of query formulation, effective query decomposition without loss of information and query results merging and refinement.

Known approaches for solving these tasks are:

– Use knowledge from ontologies (hypernym/hyponym relationships) to reformulate queries containing terms which do not exist in the ontology(-ues) to construct query plans with no loss of information (OBSERVER)

- Use some rewriting techniques together with mapping techniques to produce queries on information resources that most effectively satisfy the input query (PICSEL)

Some of the problems mentioned have received only partial solution, for example, the problem of semantic interoperability is typically partially solved by committing the participating nodes to a kind of a convention, providing the framework for semantic representations. These partial solutions evidently constrain the application domain and the functionality of the deployed software prototypes for I2R. The constraints for IEDI are as follows:

- IEDI is built on the principles of the mediator-wrapper architecture [WIE92] with the centralized mediator
- IEDI exploits the hybrid approach [WAC01] for knowledge representation
- IEDI uses information resource registration to allow the resource to become available for querying
- IEDI does not provide full automation for ontologies' mapping and alignment
- IEDI components use rewriting techniques with mappings to produce, process, and perform queries

The solutions for IEDI are not aimed to broaden the horizons of the current state of the art in I3 or, more specifically, in I2R. The task is to design the software prototype to demonstrate the feasibility of the ontology-driven approach to I2R and, further on in EDI between the Universities and the State Bodies at National level.

## 3. IEDI Reference Architecture

The main purpose of IEDI is to provide for performing queries over the set of pre-registered, but independent, distributed and semantically heterogeneous IRs. This implies that IEDI is naturally a distributed system. A query may demand to retrieve data from several geographically distributed IRs which belong to different legal owners and are physically stored in different places. This is why IEDI processes are composed of a number of tasks and activities performed at distributed nodes. These tasks should of course be executed in a controlled and ordered way. A process normally involves both automated activities performed by the IEDI software and human activities, like ontology merge and alignment, supplied with appropriate methodologies and software tools. Human activities are performed by various user roles: authorized user (AU), mediator ontology engineer (MOE), IR ontology engineer (IROE), IR provider (please refer to [BGE04] for more details).

An important factor which seriously influenced the design of IEDI architecture is semantic heterogeneity of the IRs which are registered to IEDI mediator. This implied the use of the hierarchy of ontologies which actually drive the performance of distributed queries to different IRs. The tasks of merging and alignment of the

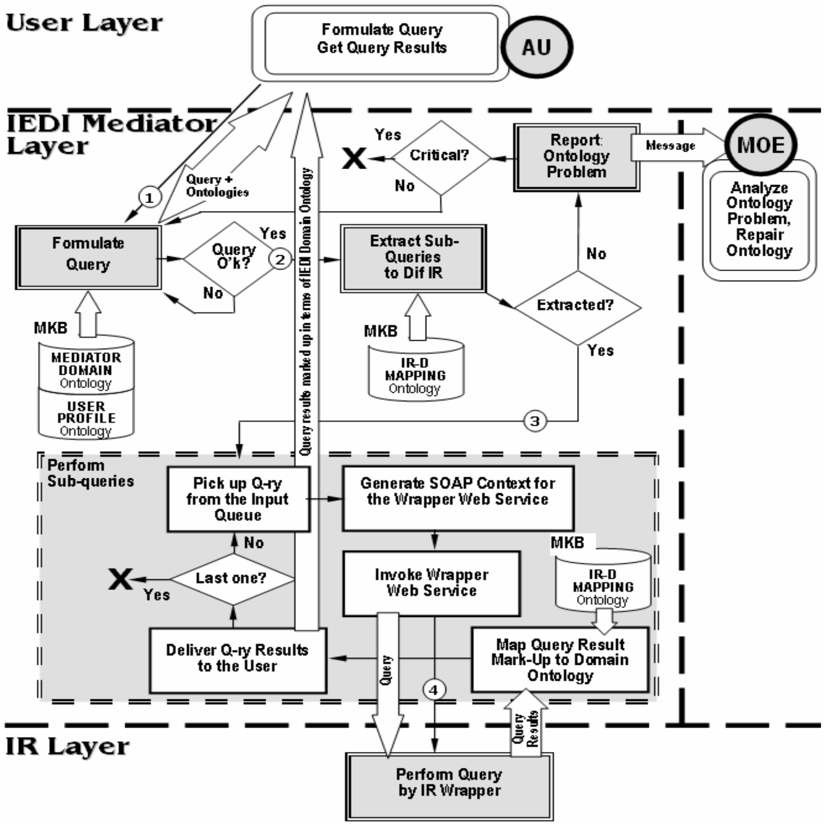


**Figure. 1.** *Process Diagram for IEDI query performance scenario.*

ontologies describing the semantics of the IRs and the common ontology of the mediator - Mediator Domain Ontology (MDO) are performed manually. IEDI provides reference ontologies and tools for this ontology engineering activities. However, this thorough preparation work allows to further perform query formulation, sub-query extraction, sub-query execution tasks in a straightforward manner and almost automatically. The diagram of IEDI query performance scenario is given in Fig. 1. The diagrams for IR Registration and Ontology Coherence Maintenance are omitted here because of space limitations (please refer to [BGE04]).

IEDI Architectural layering is defined according to the analysis of the IEDI processes and tasks and reflects the mediator-wrapper type of IEDI architecture. The layering represents the overall organization of the IEDI and is outlined according to the following points of view:

– What are the Components, the Tools and the User Roles at the specific IEDI layers?

– How do IEDI Clients and Servers interoperate across the layers of its architecture?

IEDI User Layer is the environment for AUs and AU Clients. IEDI IR Wrapper and IR Layers represent autonomous, heterogeneous, and distributed IR holders. IEDI Mediator Layer is the holder for the components and the tools providing the means for mediation between the AU-s formulating queries and retrieving the results from the registered IR-s and respective IR Wrappers to provide the relevant information. IEDI architectural layering is given in Fig. 2.

The software components of IEDI are split into two categories of Clients and Servers according to their functionality. IEDI Clients are related to IEDI AU-s and provide the interfaces for their activities. AU client provides IEDI interfaces for an AU. It functions in generic Web Browser environment (+ Java Virtual Machine) at the User Layer of IEDI Architecture (Fig. 10) and provides the interfaces for the tasks of: User Query Formulation, User Query Approval, Browsing Query Results. AU Client interoperates with the IEDI Query Formulation Tool and with the following IEDI components: IEDI Mediator Access Server and Query Formulation Server (the component of IEDI Mediator Server). MOE Client provides IEDI interfaces for the MOE. It functions in Java Virtual Machine (JVM) environment at the Mediator Layer of IEDI Architecture and provides the interfaces for the tasks of IEDI Ontologies Discussion, Merge, Alignment, Editing and Repair. IROE Client provides IEDI interfaces for an IROE and is similar to MOE Client. It functions at the Mediator and the IR Wrapper Layers of IEDI Architecture and provides the interfaces for the tasks of IRO Ontology Discussion, Editing and Repair as well as for the Negotiation on IRO – MDO Merge within the IR Registration Process. MOE and IROE Clients interoperate with the following IEDI tools: Ontology Discussion and Alignment (under development in UnIT-NET), Ontology Editor (Protégé [NSD01]). MOE and IROE Clients interoperate with the following IEDI components: IEDI Mediator Access Server

IEDI Clients and Servers are listed in Fig. 2.

## 4 IEDI Ontologies

IEDI by its role is the distributed mediator system providing some kind of semantic integration of the information retrieved from distributed, heterogeneous, and autonomous information resources. This is why the implementation and the proper usage of semantic descriptions of this information is the critical problem for the overall IEDI system implementation. It is assumed that semantic descriptions within IEDI are formalized and maintained as OWL [OWL03] ontologies at different layers of the architecture. IEDI architecture uses hybrid [WAC01] approach to explicit description of the information resource semantics. Provided are the four types of ontologies: top-level ontology, domain ontology, resource ontology and reference ontology.

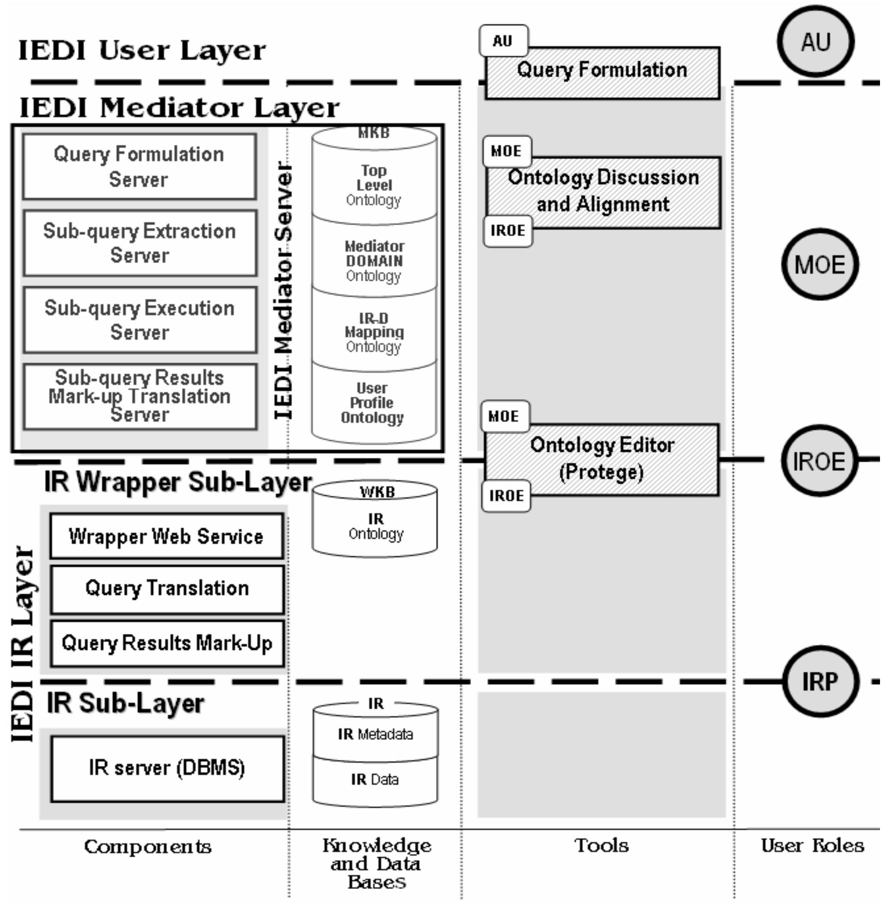


**Figure 2.** *IEDI Clients and Servers along the layers of the architecture.*

**Top-level ontology** (TLO) defines basic top-level elements. These elements according to their definitions are used in the process of mapping resource ontology elements to domain ontology elements. Top-level ontology serves as the foundation for discussion on each concept between MOE and IROE. Top-level ontology allows any two IEDI ontologies to be comparable. IEDI top-level ontology design is based on DOLCE [MBG02].

**Domain ontology** (MDO) represents particular domain knowledge. There are several reasons to explore domain ontology in UNIT-NET IEDI mediator. First one is that domain ontology provides the AUs with the opportunity to formulate their queries using concepts, agreed within domain community and to store correspondences between personal user knowledge on the domain and agreed domain ontology in their user profiles (User Profiles Reference Ontology - UPRO). Another reason is that domain ontology presents a vision of the community on the domain, and therefore plays an educational role.

**Information Resource ontology** (IRO) is a kind of domain ontology, which is constructed at the resource side independently of other resources as well as from mediator ontologies. It presents the vision of IROE on the domain. Resource ontology is used in the process of resource registration at the mediator. Each registered information resource should have its own resource ontology.

**Reference ontologies** (IR – Domain Mapping Ontology (IRDMO), UPRO) are mediator ontologies, which store the knowledge on correspondences between concepts in two or more ontologies. IRDMO contains axioms on equivalence/subsumption between concepts/slots. The function of the UPRO is to represent the semantics of AU profiles (refer to [EKP03] for more details).

Table 1 summarizes the involvement of IEDI mediator ontologies in IEDI processes.

**Table 1.** *Use of Ontologies in IEDI processes*

| **Ontologies** / **Processes** | **TLO** | **MDO Core** | **MDO** | **IRDMO** | **IRO** | **UPRO** |
|---|---|---|---|---|---|---|
| Query distributed autonomous semantic-ally heterogeneous IRs | -- | R | R | R | R | R/U |
| Register new informa-tion resource | R | R | R/U | R/U | R | -- |
| Maintain coherence in semantic descriptions | R | R/U | R/U | R/U | R/U | R/U |

Legend: R – used for reference purposes only, R/U – used as a reference and is updated, -- – not used.

## 4 Concluding Remarks

The paper has presented the reference architecture of IEDI. IEDI is the multi-layered distributed software system comprising software servers, services, components and tools for providing intelligent ontology-driven information retrieval from distributed, heterogeneous, legally and physically autonomous IRs for the organizational network of the Ukrainian National Higher Education System in frame of UnIT-Net project. This architecture is built upon the following principles:

– It is mediator-wrapper architecture with the centralized mediator
– It exploits the hybrid approach for knowledge representation

– It uses information resource registration to allow the resource to become available for querying
– IEDI combines processes performed both automatically (ontology driven distributed query processing) and manually (ontology discussion, merging, mapping, and alignment during IR Registration and Ontology Coherence Maintenance)
– Its components use rewriting techniques with mappings to produce, process, and perform queries

From the semantic point of view IEDI exploits the hierarchy of ontologies which are replenished incrementally while new IRs are registered to the mediator. These ontologies are in turn used to drive the main function of the mediator: to assist in query formulation, to decompose the query into the set of sub-queries (one per relevant IR), to convey the sub-queries to the respective IR wrappers, to translate the sub-query at the IR wrapper level, to mark-up query results.

The short-term plans for future work comprise the implementation of IEDI research prototype and its evaluation on the initial collection of the IRs of UnIT-Net Consortium member Universities.

## Bibliography


[AKS96] **Arens, Y., Knoblock, C.A., Shen, W.**: Query Reformulation for Dynamic Information Integration. *Journal of Intelligent Information Systems*. 1996.

[BAY97] **Bayardo et al.**: InfoSleuth: Semantic Integration of Information in Open and Dynamic Environment. In Proceedings of the 1997 ACM International Conference on the Management of Data (SIGMOD), Tucson, Arizona, May 1997.

[BCD98] **Bergamaschi, S., Castano, S., De Capitani di Vimercati, S., Montanari, S. Vincini, M.**: An Intelligent Approach to Information Integration. In: Proc. Of Formal Ontology in Information Systems (FOIS-98), June, 1998.

[BGE04] **Bulat, A., Gray, E., Ermolayev, V., Keberle, N., Plaksin, S., Shapar, V., Vladimirov, V., Zholtkevich, G.**: The Infrastructure for Electronic Data Interchange. Reference Architecture Specification. Version 1.0. UNIT-NET Deliverable No D2.2.D.1. URL: http://www.compscipreprints.com/

[CJO01] **Cui, Z., Jones, D., O'Brien, P.**: Issues in Ontology-based Information Integration. In: (A. Gomez-Perez, M. Gruninger,H. Stuckenschmidt,M. Uschold) Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing, Seattle, USA, August 4-5, 2001, pp.141-146.

[CJO02] **Cui, Z., Jones, D., O'Brien, P.**: Semantic B2B Integration: Issues in Ontology-based Applications. SIGMOD Record, Vol.31, No.1, March 2002. Pp.43-48

[DEF99] **Decker, S., Erdmann, M., Fensel, D., Studer, R.**: Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information. In **R. Meersman et al. (eds.):** Semantic

Issues in Multimedia Systems. Proceedings of DS-8. Kluwer Academic Publisher, Boston, 1999, 351-369.

[EKP03] **Ermolayev, V., Keberle, N., Plaksin, S., Vladimirov, V.** (2003) Capturing Semantics from Search Phrases: Incremental User Personification and Ontology-Driven Query Transformation. In: Proc. of the 2-nd Int. Conf. on Information Systems Technology and its Applications (ISTA'2003), Kharkiv, Ukraine, June 19-21, 2003, pp. 9-20, LNI Series, ISBN 3-88579-359-8

[GAM95] **Garcia-Molino, H. et. al.**: The TSIMMIS Approach to Mediation: Data Models and Languages. In : Proc. Next Generation Information Technologies and Systems (NGITS), June 1995.

[GRA97] **Gray. P., Preece A., Fiddian N., Gray W., Bench-Capon T., Shave M., Azarmi N., Wiegand M.**: KRAFT: Knowledge Fusion From Distributed Databases and Knowledge Bases. In: Proc. 8th Intl. Workshop on Database and Expert System Applications (DEXA-97), IEEE Press, pp. 682-691.

[KS00] **Kashyap V., Sheth A**.: Information Brokering across Heterogeneous Digital Data: A Metadata-based Approach. Kluwer Academic Publishers, 2000

[LR00] **Lattes V., Rousset M.-C.**: The Use of CARIN Language and Algorithms for Information Integration: The PICSEL System. *International Journal of Cooperative Information Systems*, Vol.9, No.4, 2000, pp.383-401.

[MBG02] **Masolo, C., Borgo, S., Gangemi A., Guarino, N., Oltramari, A., Schneider, L.**: WonderWeb Project Deliverable D17. The WonderWeb Library of Foundational Ontologies and the DOLCE ontology. URL:http://www.compscipreprints.com/comp/Preprint/stborgo/20021127/2/WondeWebD17v2.0.pdf

[NSD01] **F.-Noy, N., Sintek, M., Decker, S., Crubezy, M., Fergerson, R. W., Musen, M. A.**: Creating Semantic Web Contents with Protege-2000. IEEE Intelligent Systems 16(2):60-71, 2001.

[OWL03] OWL Web Ontology Language Reference. W3C Proposed Recommendation 15 December 2003. URL: http://www.w3.org/TR/owl-ref/

[STU00] **Stuckenschmidt H., Wache H., Voegele T., Visser U**.: Enabling technologies for interoperability. In: (**Visser, U., Pundt H.** Eds.) Workshop on the 14th International Symposium of Computer Science for Environmental Protection, Bonn, Germany, 2000, pp. 35-46.

[WAC01] **Wache, H., Vogele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., Hubner, S.**: Ontology-Based Integration of Information - A Survey of Existing Approaches. In: (A. Gomez-Perez, M. Gruninger, H. Stuckenschmidt, M. Uschold) Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing, Seattle, USA, August 4-5, 2001, pp.108-118, URL: http://www.cs.vu.nl/~heiner/public/ois-2001.pdf

[WIE92] **Wiederhold, G.:** Mediators in the Architecture of Future Information Systems. *IEEE Computer*, 25, 3 (March), 1992, 38–49.