ICT-MobileSummit 2009 Conference Proceedings
Paul Cunningham and Miriam Cunningham (Eds)
IIMC International Information Management Corporation, 2009
ISBN: 978-1-905824-12-0

Semantic Unified Access to Traffic Measurement Systems for Internet Monitoring Service

Ángel FERREIRO[1], Thomas FICHTEL[2], Jorge LÓPEZ DE VERGARA[3], Peter MÁTRAY[4], Felix STROHMEIER[2], Giuseppe TROPEA[5], Udi WEINSBERG[6]
[1]*Telefonica Investigación y Desarrollo, c/ Emilio Vargas, 6, 28043-Madrid, Spain*
Tel: +34 913374820, Email: olivo@tid.es
[2]*Salzburg Research, Jakob-Haringer-Straße 5, Salzburg, 5020, Austria*
Tel: +43 662 2288442, Email: felix.strohmeier@salzburgresearch.at
[3]*Universidad Autonoma de Madrid, C/ Francisco Tomás y Valiente, 11*
Campus de Cantoblanco, Madrid, 28049, Spain
Tel: +34 91 497 22 46, Email: jorge.lopez_vergara@uam.es
[4]*Eötvös Loránd University, Egyetem tér 1-3, Budapest, 1053, Hungary*
Tel: +36 1 3722860, Email: matray@complex.elte.hu
[5]*CNIT research unit of Rome Tor Vergata, Via del Politecnico, 1, Rome, 00133, Italy*
Tel: +39 328 5363676, Email: giuseppe.tropea@cnit.it
[6]*Tel-Aviv University, Ramat Aviv 1, Tel Aviv, 69978, Israel*
Tel: +972 3 6405605, Email: udiw@eng.tau.ac.il

Abstract: The increase of traffic in IP networks, leads operators to search for efficient tools to monitor it in order to support QoS. Several infrastructures to measure traffic and monitor routing have been deployed but SLA can not rely on them because of their lack of interoperability. MOMENT project is currently developing a mediator to access most of available traffic measurement systems. Thus a unified platform is being built so that clients can retrieve complex traffic data in a standardized format through an integrated and user friendly interface. MOMENT design is not a simple interface converter but it develops a new abstraction layer, based on a measurement-specific ontology to allow the semantic presentation and retrieval of traffic information. This paper describes its key features as well as its architecture and some implementation aspects.

Keywords: Ontology**,** Future Internet, IP, QoS, SLA, Network Monitoring, Semantic Middleware.

1. Introduction

QoS in Future Internet depends strongly on the ability of network operators to cope with customers demands in terms of bandwidth and parameters like jitter, delay, bit error rate, that can be specified in the Service Level Agreement (SLA). So far, the rapid growth of the Internet and the continuous convergence of services in the IP layer have been solved by increasing bandwidth, even in the users' access segment. Besides, malicious users or unintended configuration errors collapse the network and turn connectivity service out of SLA [1]. Traffic engineering and network control is then required to remove bottlenecks and get a fair usage of the networks.

Several projects have developed measurement infrastructures to track Internet functioning at various levels (collected details are available in [2]); their objectives being different, they employ different measurement techniques and store their data in various formats [3]. Thus those implementations of monitoring and measurement (Mo/Me) infrastructures are vendor-specific and lack interoperability; there is no means for researchers, operators and service providers to get simultaneously data from several infrastructures to monitor or analyze the network behaviour, aside from the inconvenience derived from dealing with different interfaces, since leveraging combined queries may be extremely difficult.

An open Pan-European platform that unites legacy Mo/Me systems and provide the possibility of extracting data, through a single semantic interface, making intelligent combinations possible, represents a qualitative step towards getting efficient network monitoring. This paper presents a novel ontology-based architecture of a unified layer for Internet measurement infrastructures, namely a mediator that acts as unified interface towards different data sources. Its design is based on a Service Oriented Architecture (SOA) approach with the help of Web-services, to keep maximum flexibility and extensibility for applications as well as for future measurement infrastructures. The introduction and usage of a measurements-oriented ontology allows the system to overcome the vast differences between data bases and tools. A similar approach has been reported as Metric Service in [4], but without the advantages provided by semantic technologies.

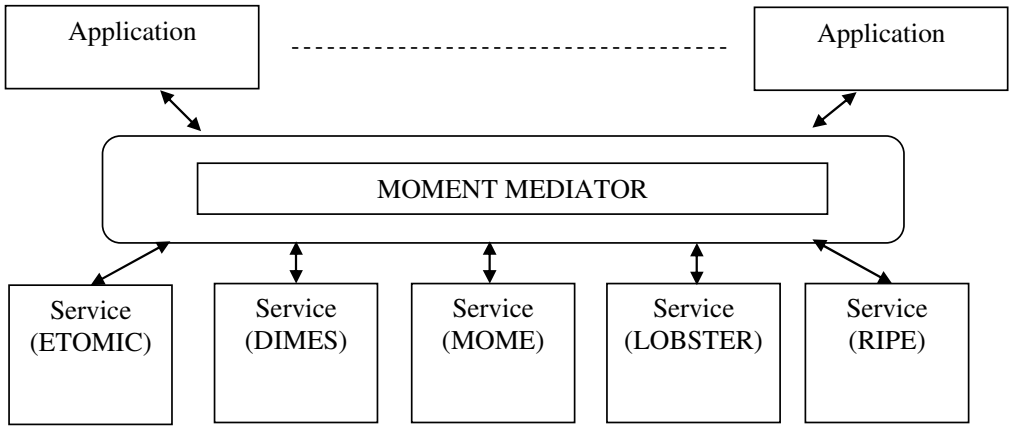

Figure 1: Schematic diagram of MOMENT mediator

Another important aspect of the MOMENT mediator is the conceived taxonomy of possible anonymization approaches. Generic anonymization schemes are difficult to design ([5]-[7]) because different organizations have different requirements. So, the main effort was focused on the dependencies between the possible anonymization strategies and the role and purpose of data consumers that query data from the mediator. The design and identification of all possible values for the various anonymization components of the ontology was the most important task.

2. Heterogeneous data sources

The main concern for an integrated interface to different monitoring and measurement infrastructures, which store data in arbitrary formats, is to be flexible in order to provide solutions for three different problems of heterogeneity: 1) heterogeneous data semantics, 2) heterogeneous data access, and 3) heterogeneous data formats. First, the semantics of data from mediated infrastructures needs to be unified, as different measurement infrastructures use different names (e.g. delay, one-way-delay or owd) and units (e.g. second, millisecond, nanosecond) for the same things. Second, data access is provided by different means (e.g. direct SQL, FTP or Web Services). Finally the data format can differ between

infrastructures and needs to be harmonized (e.g. XML, binary or CSV). The mediator middleware design also takes into account the importance of preserving confidential data of Internet users. Controlling the processing of data prior to releasing them is a strong requirement for the mediation. Thus, aside from the obvious mechanism to characterize the mediator users and assigning them different permissions, a more sophisticated procedure is needed. This is necessary because user roles and their data usage-patterns are tightly connected to anonymization strategies that must be applied to the data. Being as open as possible for every kind of data consumer is the goal of the mediation layer, although without breaking the obfuscation rules suggested by the data issuer.

MOMENT mediator will provide interoperability with the existing infrastructures and will be open for new ones to ensure enhanced functionality of the platform: The mediator architecture and the unified interface development are generic and can be extended to future tools provided some standard interfacing be agreed. This is achieved without major efforts, since the mediator architecture and unified interface is based on measurement specific ontology that allows semantic queries.

Ontologies are well known in other fields [8] for providing a vocabulary of classes and relations to describe a certain domain, stressing knowledge sharing and knowledge representation. Figure 2 shows the concept applied to traffic measurement as developed in MOMENT (a complete description has been published in [9]):

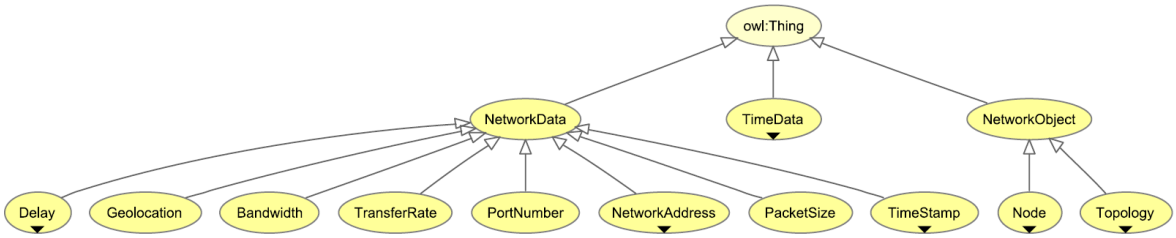

Figure 2: Fragment of the data measurement ontology implemented in MOMENT mediator

The advantages of using ontologies are manifold: The information is modelled in a more flexible way than using tables; their semantic definition of information enables a classification of knowledge (e.g. a tool that performs active measurement is an active tool) and inference (e.g. if a measurement is over a threshold then the network is overloaded); at the same time it is possible to query this knowledge (e.g. obtain all measurements with a given destination address). On the other hand, the ontology can be downloaded from the web and browsed by anyone, which allows standardization procedures to be simplified.

A major goal of the work in MOMENT is to achieve that its interfaces and ontology for traffic measurements become standard. This task is in progress.

3. Architecture of the mediator

The requirements of the proposed system, to be capable of integrating both passive and active monitoring and measurement methods via a common web services interface that allows semantic queries, are not limited to ensure continuous monitoring of the macroscopic status of the network but also allows it to perform network tomography techniques in individual domains and is able to collect information about a temporal variation of reachability and other QoS parameters without trespassing citizen rights.

The platform is non-intrusive as ISPs concern: Monitoring probes are placed outside the commercial network domains, data obtained by passive monitoring are anonymized and middleware solutions has been deployed inside the system to properly control and manage access to data subject to privacy issues. But research groups can also use the mediator and flexibly design both, passive and active measurements, so that they perform their own experiments provided their user profile meets the expertise level required. This user

classification, together with an incremental repository of queries is a relevant feature of the mediator. Nevertheless, metadata repositories (like IMDC [10] from CAIDA) are used to work with raw measurement data (like NMVO [11]-[12] from ETOMIC). If not provided by the producer of the measurement data, additional metadata can also be extracted from the measurement tools or produced by the mediator data analysis components. Besides, this approach saves computation resources.

To get acquainted with the mediator building blocks, Figure 3 provides the high-level architecture followed by short descriptions of each building block. The infrastructures below the architecture are current or future measurement systems willing to share their data using the MOMENT mediator. The selected examples are typically used access interfaces to their data. Many of the studied measurement infrastructures store their data in SQL databases, some of them provide a web service interface using REST or SOAP, and others already have semantically enriched data and provide them via a SPARQL interface (SPARQL is a query language specified at W3C to express queries for RDF data sources, providing capabilities to request information stored in an ontology [13]). Even data sources storing their data on FTP servers can be integrated. Achieving this, the MOMENT mediator allows the integration of all different kinds of data sources from very straightforward up to very sophisticated.

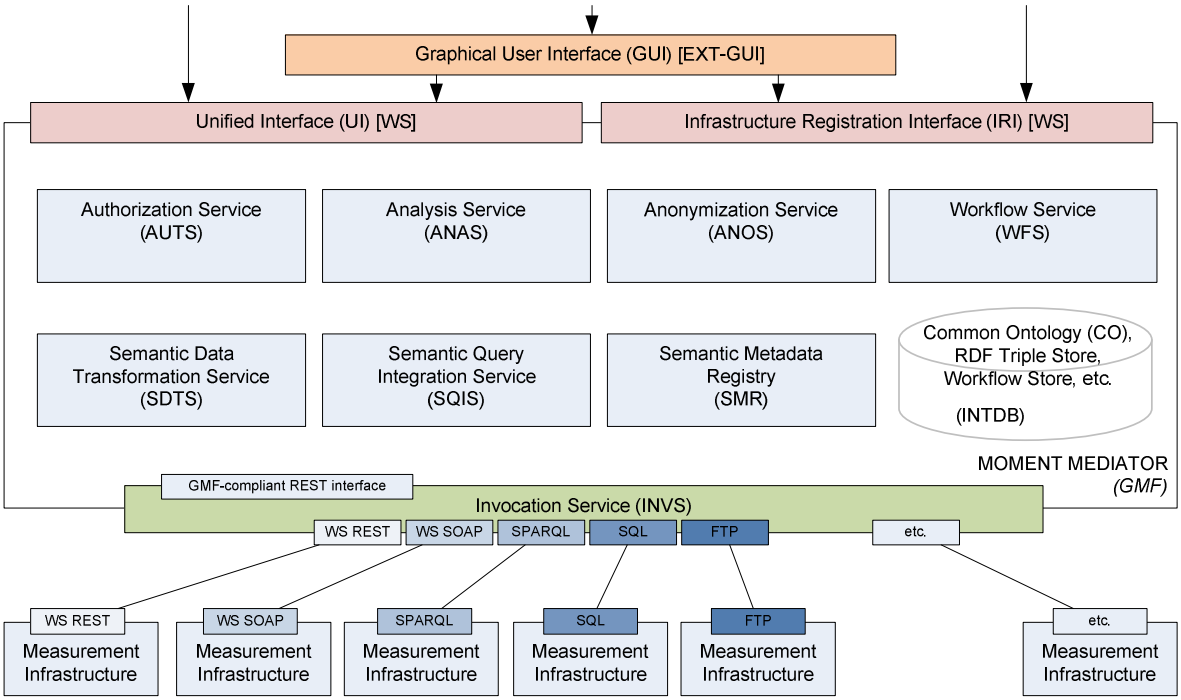

Figure 3: High-level architecture diagram showing the main components of MOMENT mediator

As shown in Figure 3, the Generalized Mediator Framework (GMF) provides interfaces towards the clients (upper part), and requests data sources from the servers of the measurement infrastructures (lower part). The core services within the GMF are independent modules interacting by means of web service technologies, and able to access the internal database (INTDB). In direction to the clients, the mediator provides two web service (WS) interfaces. First, the Unified Interface (UI) provides the uniform access to the functions of the mediator via a web service (WS) interface. This allows invocation of the mediator through external software programs. Second, the Infrastructure Registration Interface (IRI) is responsible for receiving semantically enhanced metadata information from the infrastructures. It provides an interface for the measurement infrastructures to register their metadata into the system in order to make their measurement data findable and

accessible. Infrastructures can publish their metadata in RDF triples which are defined in the MOMENT ontology.

Both of the above mentioned web service interfaces can also be accessed via a Graphical User Interface (GUI) to enable interaction with the mediator in a user-friendly way. Supported by the Authorization Service (AUTS), authentication and differentiated treatment of users are features provided by the GUI as well as the interfacing to other modules so as to compose the queries and present the results. It also presents the facilities to record the usage of the mediator, keep a repository of its functionality and help beginners to use it as well as enable discussions about its results, namely monitoring the Internet.

The Semantic Metadata Registry (SMR) is a central component of the system: Its innovative approach is to store semantically enhanced information, based on the ontology model, including measurement infrastructures description (metrics it deals with, data format, period of a measurement, etc.), data about workflows (its description, involved components, etc.) and data about analysis methods (objective of method, input and output format). All data in the metadata registry are based on a common ontology and therefore semantically interpretable. The metadata registry also stores information about the method to access measurement data in a mediated infrastructure. This information is passed on to the invocation service, e.g. the SQL URL, username, password and query.

The Semantic Query Integration Service (SQIS) is the module responsible for executing queries that access infrastructures directly, allowing a fine grained search for measurement data stored in query-enabled sources. These queries are requested by an application on top of the unified interface using semantic technologies such as SPARQL and the answer is integrated from existing sources, using a common vocabulary taken from the network data ontology.

The Analysis Service (ANAS) is responsible for further analysis of measurement results. This service can process data from one ore more infrastructure(s). Information about available analysis services is stored in the metadata registry.

The Anonymization Service (ANOS) provides methods to anonymize measurement data in order to preserve privacy. The service can be invoked if anonymization of measurement data is necessary and it can be also invoked, if needed, in those situations when data has already been anonymized, for internal policy reasons, directly at the source. This might be useful, and also needed, depending on the final user rights and privileges within the mediator. Measurement infrastructures that already have a well structured privacy policy will map their policies onto the anonymization ontology so that the correct strategy will be enforced over each mediated user. This module will also represent the interface with existing platforms that don't have encoded specific policies for data release, providing default mappings and policies based on semantic reasoning.

The Workflow Service (WFS) helps to automate recurring processes. In a workflow a user can define a sequence of MOMENT operations, e.g. get data from infrastructure A and B and analyze this data with method X.

The Semantic Data Transformation Service (SDTS) is able to transform the data in cases, where measurement data provided by an infrastructure is not exactly the data a user or application requires, e.g. transforming measurement units.

Finally, the Invocation Service (INVS) is used to retrieve specific measurement data from an infrastructure. To retrieve specific measurement data a user first has to query or browse the registry in order to identify the data he wants. For this data the mediator can then retrieve the reference to the data and pass it on to the invocation service. Depending on the type of the reference (SQL, WS, FTP etc.) the invocation service can instantiate an appropriate handler for the reference and retrieve the data from the mediated infrastructure.

4. Functional description of the mediator

A service orchestration can be performed by the MOMENT mediator as shown in Figure 4. MOMENT mediator provides a unified interface to the application in order to actually *command* the monitoring service, including the ability to discover and select resources (and combine their data if needed) for the specific task the client has asked for. This exploits two key features of the MOMENT mediator, which is first to hide complexity of a measurement infrastructure from the end user, and second to get value added information from multiple measurement infrastructures compared to a single one. In the given figure SPARQL is used as query language, but the same queries can be provided via user-friendly web interfaces.

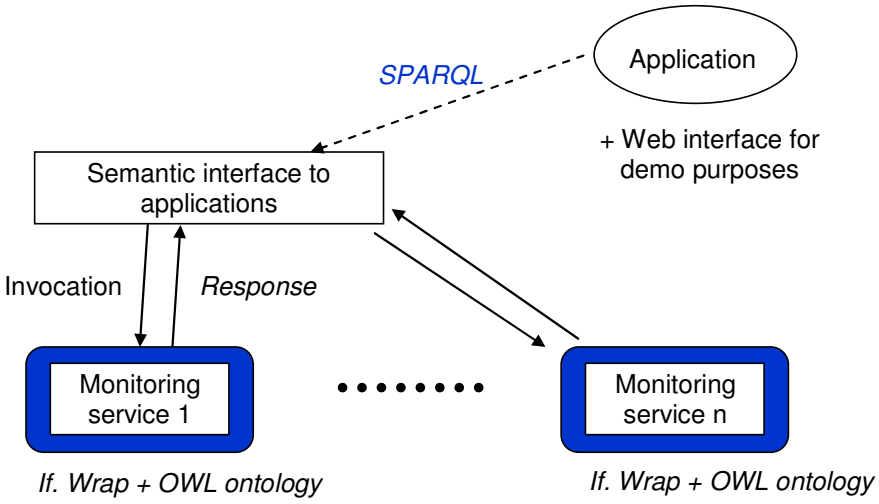

Figure 4: Resource discovery

Figure 5 illustrates the process of retrieving data from multiple sources by the mediator to perform a complex query for a client who only enters a high level query without needing precise information about the infrastructures that finally provide the measurements: The user gets the results he from the GUI/UI, that forwards the request to the SQIS; subsequently, the SQIS asks the SMR if matching metadata are available. Depending on these metadata, the infrastructures hosting the parts of the requested measurement results are contacted (e.g. via SQL). The SQIS combines these replies into the integrated result (internal memory of the mediator), which is given back to the user.

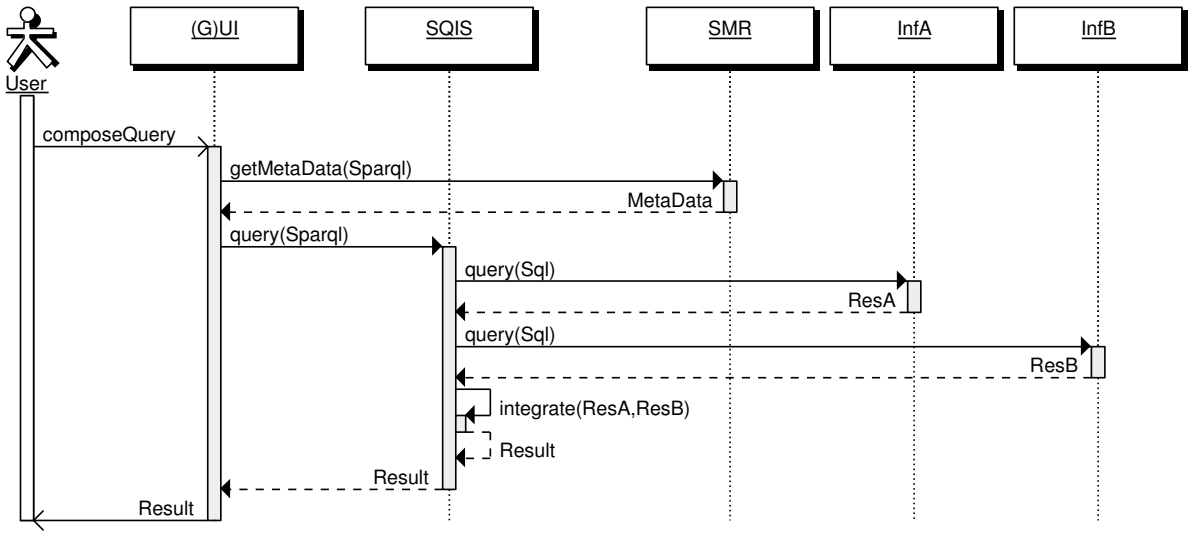

Figure 5: Process of retrieving integrated measurement results from multiple infrastructures

Not included in Figure 5, but optionally possible on several places during this process, the results can be sent through the ANOS module, to be obfuscated. In such case a skeletal Acceptable Use Policy document is generated by the ANOS to be shown to the user. It represents an informative document, although structured, about what the provider expects from the user regarding the usage of the data that the provider itself is willing to release. It can be accepted and signed by the user. Although this does not imply any technical enforcement by the mediator, it can be regarded as a kind of End User Legal Agreement. Details about the semantic framework to anonymize network data are presented in [14].

6. Conclusions and outlook

The MOMENT mediator, a unified interface to measuring and monitoring infrastructures has been presented in this paper. Its novel architecture relies on a common ontology that is based upon several different taxonomies: A Monitoring Tool ontology, a Metadata ontology, a Measurement ontology and an Anonymization ontology to build a semantic unifying layer for Internet measurement infrastructures access. Using such unified interface to measurement infrastructures, researchers and network operators can easily merge data and gain a more complete and accurate view of the Internet, as required to manage and operate networks on which it is supported and guarantee a fair evolution towards the future Internet. In addition, it must be highlighted that anonymization process included in the mediator and usage of metadata guarantees the appropriate respect to laws that preserve privacy.

The design and implementation of the Mediator sets the foundations for an even more advanced system, to establish plausible standard QoS references and mechanisms to govern future Internet. The mediator architecture and the unified interface are not limited to the already available monitoring and measurement infrastructures, but are generic and can be extended to the future similar infrastructures. The unified interface can be shown applicable to interface with such infrastructure without major efforts, since the mediator architecture and unified interface are based on the measurement specific ontology that allows semantic queries. The MOMENT mediator will be complete when its interfacing and ontology for traffic measurements will become a standard. To move this activity forward, an ETSI ISG is considered to be launched.

Acknowledgments

The authors thank the partial support of the EU ICT MOMENT Collaborative Project (Grant Agreement No.215225). P. Matray thanks the partial support of the National Office for Research and Technology (NAP 2005/ KCKHA005) and the ELTE eScience RET

References

[1] Mahajan, R., Wetherall, D., Anderson, T. “Understanding BGP Misconfiguration”. ACM SIGCOMM Computer Communications Review. (2002) p3-16.

[2] MOMENT related projects: http://www.fp7-moment.eu/index.php?option=com_content&view=article&id=45&Itemid=59

[3] Database schemas of MOMENT related data sources: http://www.fp7-moment.eu/index.php?option=com_content&view=article&id=81&Itemid=58

[4] Kumar, A. Agarwal, V. “A Customizable Mediation Engine for Metrics Collection, Aggregation and Composition”. The 10th IEEE/IFIP Network Operations and Management Symposium, 2006 (NOMS 2006).

[5] Alllman, M., Paxson, V. “Issues and etiquette concerning use of shared measurement data”. In Proceedings of the 7th ACM SIGCOMM conference on Internet measurement (IMC '07), New York, NY, USA, ACM (2007) p135-140.

[6] Koukis, D., Antonatos, S., Antoniades, D., Markatos, E., Trimintzios, P. “A generic anonymization framework for network traffic”. In Proceedings of the IEEE International Conference on Communications (ICC), vol. 5, Istanbul, Turkey: IEEE Computer Society Press, (June 2006) p2302-2309.
[7] Paxson, V. “Strategies for sound internet measurement”. In: IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement, New York, NY, USA, ACM (2004) p263-271.
[8] Gruber, T.R. “A translation approach to portable ontology specifications”. Knowledge Acquisition 5(2) (1993) p199-220.
[9] Salvador, A. et al. “Ontology Design and Implementation for IP Networks Monitoring”, submitted to the 1st International workshop on Web services & Semantic technology, Wrexham, North East Wales, UK (8-11 September 2009, pending to be published).
[10] Shannon, C., Moore, D., Keys, K., Fomenkov, M., Huffaker, B., and claffy, k. “The internet measurement data catalog”. ACM SIGCOMM Computer Communications Review 35, 5 (Oct. 2005) p97-100.
[11] Morato, D. et al., “ETOMIC: A testbed for universal active and passive measurements”. Best Testbed Award. In Proceedings of IEEE TRIDENTCOM 2005, Trento, Italy, (Feb. 2005) p283-289.
[12] Matray, P. et al., “Building a Prototype for Network Measurement Virtual Observatory”. In Proceedings of ACM SIGMETRICS MineNet Workshop 2007, San Diego, USA, (Jun. 2007)
[13] Prud'hommeaux, E., Seaborne, A., Eds., “SPARQL Query Language for RDF”, W3C Recommendation, 15 January 2008, available at http://www.w3.org/TR/rdf-sparql-query/
[14] Tropea, G., Scibilia, F., Blefari-Melazzi, N., “A Semantic Framework to Anonymize Network Data and Define Their Acceptable Use”, accepted Poster presentation ICT-MobileSummit 2009, Santander, Spain, (10-12 June 2009, to be published).

