

# Towards a RESTful Architecture for Managing a Global Distributed Interlinked Data-Content-Information Space

Maria Chiara Pettenati, Lucia Ciofi, Franco Pirri, and Dino Giuli

Electronics and Telecommunications Department, University of Florence, Via Santa Marta, 3  
50139 Florence, Italy

{mariachiara.pettenati, lucia.ciofi, franco.pirri,  
dino.giuli}@unifi.it

**Abstract.** The current debate around the future of the Internet has brought to front the concept of “Content-Centric” architecture, lying between the Web of Documents and the generalized Web of Data, in which explicit data are embedded in structured documents enabling the consistent support for the direct manipulation of information fragments. In this paper we present the InterDataNet (IDN) infrastructure technology designed to allow the RESTful management of interlinked information resources structured around documents. IDN deals with globally identified, addressable and reusable information fragments; it adopts an URI-based addressing scheme; it provides a simple, uniform Web-based interface to distributed heterogeneous information management; it endows information fragments with collaboration-oriented properties, namely: privacy, licensing, security, provenance, consistency, versioning and availability; it glues together reusable information fragments into meaningful structured and integrated documents without the need of a pre-defined schema.

**Keywords:** Web of Data; future Web; Linked Data; RESTful; read-write Web; collaboration.

## 1 Introduction

There are many evolutionary approaches of the Internet architecture which are at the heart of the discussions both in the scientific and industrial contexts: Web of Data/Linked Data, Semantic Web, REST architecture, Internet of Services, SOA and Web Services and Internet of Things approaches. Each of these approaches focus on specific aspects and objectives which underlie the high level requirements of being a driver towards “a better Internet” or “a better Web”.

Three powerful concepts present themselves as main drivers of the Future Internet [1][2]. They are: a user-centric perspective, a service-centric perspective and a content-centric perspective. The user-centric perspective emphasizes the end-user experience as the driving force for all technological innovation; the service-centric perspective is currently influenced in enterprise IT environment and in the Web2.0 mashup culture, showing the importance of flexibly reusing service components to build efficient applications. The Content-Centric perspective leverages on the importance of creating, pub-

lishing and interlinking content on the Web and providing content-specific infrastructural services for (rich media) content production, publication, interlinking and consumption. Even if it is very difficult to provide a strict separation of approaches because either they are often positioned or have evolved touching blurred areas between Content, Services and User perspectives, a rough schema in Table 1 can provide highlights the main, original, driving forces of such approaches.

**Table 1.** Rough classification of main driving forces in current Future Network evolutionary approaches

	<b>Content-centric</b>	<b>Service-centric</b>	<b>Users-centric</b>
<b>Approaches</b>	Web of Data/ Linked Data	Internet of Services WS-*	Web 2.0, Web 3.0, Semantic Web
	REST	SOA	Internet of Things

The three views can be interpreted as emphasizing different aspect rather than expressing opposing statements. Hence, merging and homogenizing towards an encompassing perspective may help towards the right decision choice for the Future Internet. Such an encompassing perspective has been discussed in terms of high-level general architecture in [1] and has been named “Content-Centric Internet”. At the heart of this architecture is the notion of Content, defined as “any type and volume of raw information that can be combined, mixed or aggregated to generate new content and media” which is embedded in the Content Object, “the smallest addressable unit managed by the architecture, regardless of its physical location”. In such an high-level platform, Content and Information are separate concepts [3] and Services are built as a result of a set of functions applied to the content, to pieces of information or services. As a consequence of merging the three views (user, content, service-oriented) the Future Internet Architecture herewith described essentially proposes a Virtual Resources abstraction required for the Content-Centric approach. Another view of “Content-centric Internet architecture” is elaborated in [2] by Danny Ayers, based on the assumption that “what is missing is the ability to join information pieces together and work more on the level of knowledge representation”. Ayers’ proposal is therefore a “Transitional Web” lying between the Web of Documents and the generalized Web of Data in which explicit data are embedded in documents enabling the consistent support for the direct manipulation of information as data without the limitation of current data manipulation approaches. To this end, Ayers identifies the need to find and develop technologies allowing the management of “micro-content” i.e. sub-document-sized chunks (information/document fragments), in which content being managed and delivered is associated with descriptive metadata.

Abstracting from the different use of terms related to the concepts “data”, “content” and “information” which can be found in literature with different meanings [4], the grounding consistency that can be highlighted is related to the need of providing an evolutionary direction to the network architecture hinging on the concept of a small, Web-wide addressable data/content/information unit which should be organized according a specific model and handled by the network architecture so as to

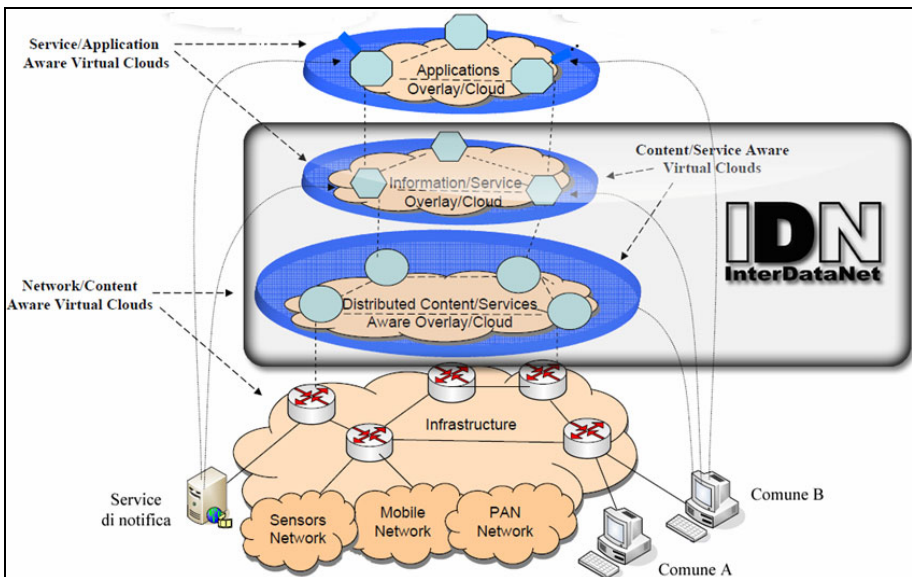
provide basic Services at an “infrastructural level” which in turn will ground the development of Applications fulfilling the user-centric needs and perspectives. Among the different paths to the Web of Data the one most explored is adding explicit data to content. Directly treating content as data has instead had little analysis.

In this paper we discuss evolution of InterDataNet (IDN) an high-level Resource Oriented Architecture proposed to enable the Future Internet approaches (see [5] [6] and references therein).

InterDataNet is composed of two main elements: the IDN-Information Model and the IDN-Service Architecture. Their joint use is meant to allow:

1. the management of addressable and reusable information fragment
2. their organization into structured documents around which different actors collaborate
3. the infrastructural support to collaboration on documents and their composing information fragments
4. the Web-wide scalability of the approach.

The purpose of this paper is to show that InterDataNet can provide a high-level model of the Content-Centric Virtualized Network grounding the Future Internet Architecture. For such a purpose InterDataNet can provide a Content-Centric abstraction level (the IDN-Information Model) and a handling mechanism (the IDN Service Architecture), to support enhanced content/information-centric services for Applications, as highlighted in Figure 1.



**Fig. 1.** InterDataNet architecture situated with respect to the Future Internet architecture envisaged in [7].

Referring to Table 1, current Content-centric network approaches and architectures, though aiming at dealing with distributed granular content over the Web, suffer from a main limitation: the more we get away from the data and move into the direction of information, the fewer available solutions are there capable of covering the following requirements:

- the management of addressable and reusable information fragment
- their organization into structured documents around which different actors collaborate
- the infrastructural (i.e. non application-dependent) support to collaboration on above documents and their composing information fragments
- the uniform REST interaction with the resources
- the Web-wide scalability of the approach.

This consolidates the need to look for and provide solutions fitting the visionary path towards a content-information/abstraction levels as illustrated in Figure 1, to which we provide our contribution, with the technological solution described in the following section.

## 2 The InterDataNet Content-Centric Approach

InterDataNet main characteristics are the following:

1. IDN deals with globally identified, addressable and reusable information fragments (as in Web of Data)
2. IDN adopts an URI-based addressing scheme (as in Linked Data)
3. IDN provides simple a uniform Web-based interface to distributed heterogeneous data management (REST approach)
4. IDN provides - at an infrastructural level - collaboration-oriented basic services, namely: privacy, licensing, security, provenance, consistency, versioning and availability
5. IDN glues together reusable information fragments into meaningful structured and integrated documents without the need of a pre-defined schema.

This will alleviate application-levels of sharing arbitrary pieces of information in ad-hoc manner while providing compliancy with current network architectures and approaches such as Linked Data, RESTful Web Services, Internet of Service, Internet of Things.

### 2.1 The InterDataNet Information Model and Service Architecture

IDN framework is described through the ensemble of concepts, models and technologies pertaining to the following two views (Fig. 2):

**IDN-IM (InterDataNet Information Model).** It is the shared information model representing a generic document model which is independent from specific contexts and technologies. It defines the requirements, desirable properties, principles and structure of the document to be managed by IDN.

**IDN-SA (InterDataNet Service Architecture).** It is the architectural layered model handling IDN-IM documents (it manages the IDN-IM concrete instances allowing the users to “act” on pieces of information and documents). The IDN-SA implements the reference functionalities defining subsystems, protocols and interfaces for IDN document collaborative management. The IDN-SA exposes an IDN-API (Application Programming Interface) on top of which IDN-compliant Applications can be developed.

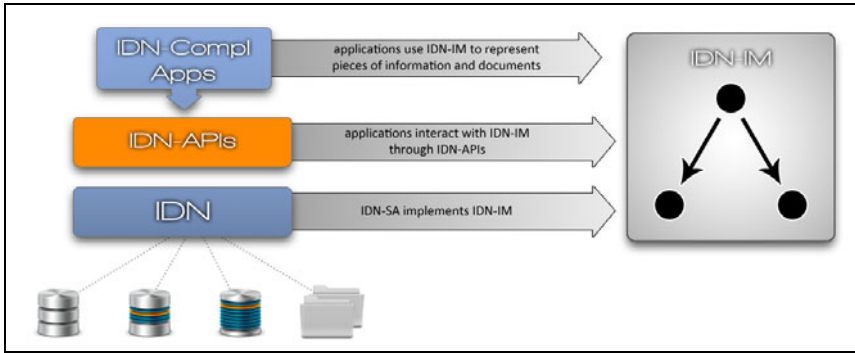


Fig. 2. InterDataNet framework

**The InterDataNet Information Model.** The Information Model is the graph-based data model (see Figure 3) to describe interlinked data representing a generic document model in IDN and is the starting point from which has been derived the design of IDN architecture.

Generic information modeled in IDN-IM is formalized as an aggregation of data units. Each data unit is assigned at least with a global identifier and contains generic data and metadata; at a formal level, such data unit is a node in a Directed Acyclic Graph (DAG). The abstract data structure is named IDN-Node. An IDN-Node is the “content-item” handled by the “content-centric” IDN-Service Architecture. The degree of atomicity of the IDN Nodes is related to the most elementary information fragment whose management is needed in a given application. The information fragment to be handled in IDN-IM compliant documents, has to be inserted into a container represented by the IDN-Node structure, i.e. a rich hierarchically structured content. The IDN Node is addressable by an HTTP URI and its contents (name, type, value and metadata) are under the responsibility of a given entity. The node aggregated into an IDN-IM document is used (read, wrote, updated, replicated, etc.) under the condition applied/specified on it by its responsible. An IDN-document structures data units is composed of nodes related to each other through directed “links”. Three main link types are defined in the Information Model:

- *aggregation links*, to express relations among nodes inside an IDN-document;
- *reference links*, to express relations between distinct IDN-documents;
- *back links*, to enable the mechanism of notification of IDN-nodes updates to parent-nodes (back propagation).

The different types of links envisaged in IDN-IM are conceived to enable collaboration extending the traditional concept of *hyperlink*. Indeed the concept of link expressed by the “href” attribute in HTML tags inherently incorporate different “meanings” of the link: either inclusion in a given document, such as it is the case of the tag <img> image, or reference to external document, such as it is the case of the <a> anchor tag. Explicitly separating these differences allow to make the meaning explicit and consequently enable the different actions, e.g. assign specific authorizations on the resources involved in the link. Actually, if two resources are connected through an aggregation link then the owner of each resources has the capability to write and read both of them. Instead the reference link implies only the capability to read the resource connected. In Figure 3 aggregation, reference and back links are illustrated.

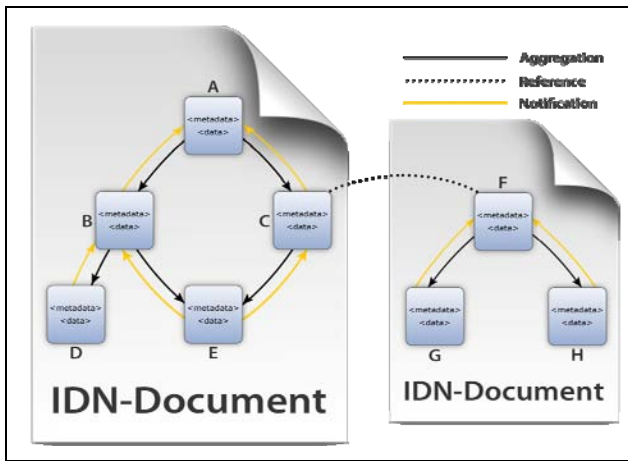


Fig. 3. InterDataNet Information-Model

**InterDataNet Service Architecture.** IDN Service Architecture (IDN-SA) is made up of four layers: Storage Interface, Replica Management, Information History and Virtual Resource. Each layer interacts only with its upper and lower level and relies on the services offered by IDN naming system. IDN-Nodes are the information that the layers exchange in their communications. In each layer a different type of IDN-Node is used: SI-Node, RM-Node, IH-Node and VR-Node. Each layer receives as input a specific type of IDN-Node and applies a transformation on the relevant metadata to obtain its own IDN-Node type. The transformation (adding, modifying, updating and deleting metadata) resembles the encapsulation process used in the TCP/IP protocol stack. The different types of IDN-Nodes have different classes of HTTP-URI as identifiers.

*Storage Interface (SI)* provides the services related to the physical storage of information and an interface towards legacy systems. It offers a REST interface enabling CRUD operations over SI-Nodes. SI-Nodes identifiers are URLs.

*Replica Management (RM)* provides a delocalized view of the resources to the upper layer. It offers a REST interface enabling CRUD operations over RM-Nodes. RM-Nodes identifiers are HTTP-URI named PRI (Persistent Resource Identifier). RM presents a single RM-Node to the IH layer hiding the existence of multiple replicas in the SI layer.

*Information History (IH)* manages temporal changes of information. It offers a REST interface enabling CRUD operations over IH-Nodes. IH Nodes identifiers are HTTP-URI named VRI (Versioned Resource Identifier). The IH layer presents the specific temporal version of the Node to the VR layer.

*Virtual Resource (VR)* manages the document structure. It offers to IDN compliant applications a REST interface enabling CRUD operations on VR-Nodes. VR-Nodes identifiers are HTTP-URIs named LRIs (Logical Resource Identifiers).

On top of the four layers of the Service Architecture, the IDN-compliant Application layer uses the documents' abstraction defined in the Container-Content principle. Interfacing to the VR layer, the application is entitled to specify the temporal instance of the document handled.

The communications between IDN-SA layers follows the REST [8] paradigm through the exchange of common HTTP messages containing a generic IDN-Node in the message body and IDN-Node identifier in the message header.

IDN architecture envisages a three layers naming system (figure 5): in the upper layer are used Logical Resource Identifier (LRI) to allow IDN-application to identify IDN-nodes. LRI are specified in a human-friendly way and minted by the IDN-Applications. Each IDN-Node can be referred to thanks to a global unique canonical

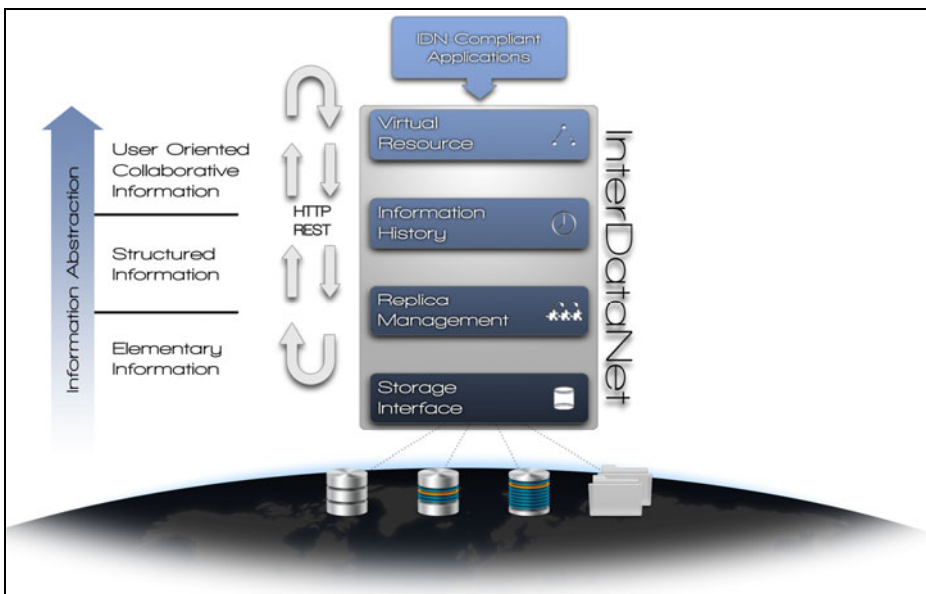


Fig. 4. InterDataNet Service Architecture

name and one or more “aliases. In the second layer are used Persistent Resource Identifiers (PRI) in order to obtain a way to unambiguously, univocally and persistently identify the resources within IDN-middleware independent of their physical locations; in the lower layer are used Uniform Resource Locators (URL) to identify resource replicas as well as to access them. Each resource can be replicated many times and therefore many URLs will correspond to one PRI. The distributed pattern adopted in the IDN naming system resembles the traditional DNS system in which the load corresponding on a single hostname is distributed on several IP addresses. Analogously we distributed a PRI (HTTP-URI) on a set of URLs (HTTP-URIs), using a DNS-like distributed approach.

The implementations of IDN-SA are a set of different software modules, one module for each layer. Each module, implemented using an HTTP server, will offers a REST interface. The interaction between IDN-compliant applications and IDN-SA follows the HTTP protocol as defined in REST architectural style too. CRUD operations on application-side will therefore be enabled to the manipulation of data on a global scale within the Web.

REST interface has been adopted in IDN-SA implementation as the actions allowed on IDN-IM can be translated in CRUD style operations over IDN-Nodes with the assumption that an IDN-document can be thought as an IDN-Node resources collection. The resources involved in REST interaction are representations of IDN-Nodes. As introduced earlier in this Section, there are several types of Nodes all of which are coded in an “IDN/XML format” (data format defined with XML language). Every resource in such format must be well formed with respect to XML syntax, and valid with respect to a specific schema (coded in XML Schema) defined according to this purpose. The schema for “IDN/XML format” uses both an ad-hoc built vocabulary to describe terms which are peculiar to the adopted representation, as well as standard vocabularies from which it borrows set of terms. Each IDN-Node resource is identified by an HTTP-URI. Through its HTTP-URI it is possible to interact with the resource in CRUD style, using the four primitives POST, GET, PUT, DELETE.

It is worth highlighting that the IDN-Service Architecture is designed to allow inherent scalability also in the deployment of its functions. According to the envisaged steps the architecture can offer and deploy specific functionalities of the architecture

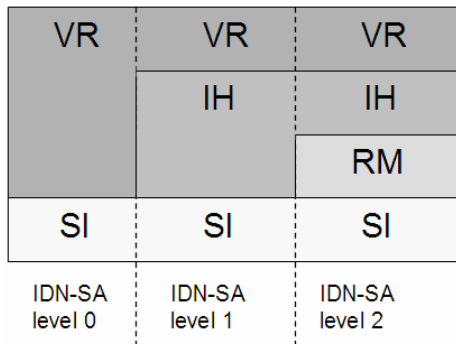


Fig. 5. InterDataNet Service Architecture scalability features



without the need to achieve the complete development of the architecture before its adoption. In this way, specific IDN-Applications developed on top of IDN-SA level 0/1/2 would thus leverage on a subset of IDN characteristics, and functions while keeping the possibility to upgrade to the full IDN on a need or case basis and once further releases will be available, as it is illustrated in Figure 5.

### 3 Conclusion

The kind of Content-centric interoperability we aim to provide in IDN is related to the exchange and controlled use (discovery, retrieve, access, edit, publish, etc.) of the distributed information fragments (contents) handled through IDN-IM documents. The presented approach is not an alternative to current Web of Data and Linked Data approaches rather it aims at viewing the same data handled by the current Web of Data from a different perspective, where a simplified information model, representing only information resources, is adopted and where the attention is focused on collaboration around documents and documents fragments either adopting the same global naming convention or suggesting new methods of handling data, relying on standard Web techniques.

InterDataNet could be considered to enable a step ahead from the Web of Document and possibly grounding the Web of Data, where an automated mapping of IDN-IM serialization into RDF world is made possible using the Named Graph approach [9]. Details on this issue are beyond the scope of the present paper.

The authors are aware that the IDN vision must be confronted with the evaluation of the proposed approach. Providing demonstrable contribution to such a high level goal is not an easy task, as it is demonstrated by the state of the art work defending this concept idea which, as far as we know, either do not provide concrete details about the possible implementation solutions to address it [1][3] or circumvent the problem adopting a mixed approach, while pinpointing the main constraint of the single solutions. However, several elements can be put forward to sustain our proposal, with respect to three main elements: a) the adoption of layered architecture approach to break down the complexity of the system problem [10]; b) using HTTP URIs to address information fragments to manage resources “in” as well as “on” the Web [11]; c) the adoption of a RESTful Web Services, also known as ROA – Resource Oriented Architecture to leverage on REST simplicity (use of well-known standards i.e. HTTP, XML, URI, MIME), pervasive infrastructure and scalability. The current state of InterDataNet implementation and deployment, is evolving along two directions: a) the infrastructure; the Proof of Concept of the implementation of the full IDN-Service Architecture is ongoing. Current available releases of the Architecture implement all the layers except for the Replica Management layer, while the implementation of the three-layers naming system is being finalized; b) application/working examples: an IDN-compliant applications has been developed on top of a simplified instance of the IDN-SA, implementing only the IDN-VR and IDN-SI layers. This application is related to the online delivery of Official Certificates of Residence. The implemented Web application allows Public Officers to assess current citizens’ official residence address requesting certificates to the entitled body, i.e. the Municipality.

InterDataNet technological solution decreases the complexity of the problems at the Application level because it offers infrastructural enablers to Web-based interoperation without requiring major preliminary agreements between interoperating parties thus providing a contribution in the direction of taking full advantage of the Web of Data potential.

**Acknowledgments.** We would like to acknowledge the precious work of Davide Chini, Riccardo Billero, Mirco Soderi, Umberto Monile, Stefano Turchi, Matteo Spanpani, Alessio Schiavelli and Luca Capannesi for the technical support in the implementation IDN-Service Architecture prototypes.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

1. Zahariadis, T., Daras, P., Bouwen, J., Niebert, N., Griffin, D., Alvarez, F., Camarillo, G.: Towards a Content-Centric Internet. In: Tselentis, G., Galis, A., Gavras, A., Krco, S., Lotz, V., Simperl, E., Stiller, B., Zahariadis, T. (eds.) *Towards the Future Internet - Emerging Trends from European Research*, pp. 227–236. IOS Press, Amsterdam (2010)
2. Ayers, D.: From Here to There. *IEEE Internet Comput* 11(1), 85–89 (2007)
3. European Commission Information Society and Media. *Future Networks The way ahead!* European Communities: Belgium (2009)
4. Melnik, S., Decker, S.: A Layered Approach to Information Modeling and Interoperability on the Web. In: *Proceedings ECDL'00 Workshop on the Semantic Web*, Lisbon (September 2000)
5. Pettenati, M.C., Innocenti, S., Chini, D., Parlanti, D., Pirri, F. (2008) *Interdatanet: A Data Web Foundation For The Semantic Web Vision*. *Iadis International Journal On Www/Internet* 6(2) (December 2008)
6. Pirri, F., Pettenati, M.C., Innocenti, S., Chini, D., Ciofi, L.: *InterDataNet: a Scalable Middleware Infrastructure for Smart Data Integration*, in D. In: Giusto, D., et al. (eds.) *The Internet of Things: 20th Tyrrhenian Workshop on Digital Communications*, Springer, Heidelberg (2009), doi: 10.1007/978-1-4419-1674-7\_12
7. Zahariadis, T., Daras, P., Bouwen, J., Niebert, N., Griffin, D., Alvarez, F., Camarillo, G.: *Towards a Content-Centric Internet Plenary Keynote Address*. Presented at *Future Internet Assembly (FIA) Valencia, SP*, 15-16 April (2010)
8. Richardson, L., Ruby, S.: *RESTful Web Services*; O'Reilly Media, Inc.: Sebastopol, CA, USA (2007)
9. Carroll, J.J., Bizer, C., Hayes, P., Stickler, P.: *Named graphs, provenance and trust*. In: *Proceedings of the 14th international conference on World Wide Web - WWW '05*. Presented at the 14th international conference, Chiba, Japan, p. 613. Chiba, Japan (2005), doi:10.1145/1060745.1060835
10. Zweben, S.H., Edwards, S.H., Weide, B.W., Hollingsworth, J.E.: *The Effects of Layering and Encapsulation on Software Development Cost and Quality*. *IEEE Trans. Softw. Eng.* 21(3), 200–208 (1995)
11. Hausenblas, M.: *Web of Data*. "Oh – it is data on the Web" posted on April 14, 2010; accessed September 8, 2010, <http://webofdata.wordpress.com/2010/04/14/oh-it-is-data-on-the-web/>