Linking Data on the Future Internet
Semantic networks and intelligent objects

Muriel Foulonneau[1], Gérald Arnould[1], Karl Devooght[1], Yannick Naudet[1]

[1] Tudor Research Centre, Luxembourg
{muriel.foulonneau, gerald.arnould, karl.devooght, yannick.naudet}@tudor.lu

Abstract.

Future Internet resources include many physical and conceptual resources, which can be described and documented by structured data. They can even publish themselves structured data on the Web. These new actors will raise significant challenges to adapt linkage mechanisms between datasets on a larger scale and implement them as part of the Future Internet infrastructure. We illustrate the potential role of Linked Data in the Future Internet, together with Content Centric Networks and Content Objects. We emphasize the importance of linkage aspects and suggest new possibilities to support data linkage, in particular by embedding linkage behaviors as properties of Future Internet resources.

Keywords: Future Internet, Linked Data, Content Object, Content Centric Networks, Usage data

1 Introduction

The current visions on the Future Internet make resources actors of the Internet, able to publish (i.e. provide information) and interact on the Web (Bleeker, 2006, Missikof et al., 2010). The Semantic Web can bring solutions to the challenges of the Future Internet, by providing identifiers for instance to the digital representations of those resources. Moreover, it can shape the communication of Future Internet resources.
On the Semantic Web, everything can become a resource, i.e. traditional document-like resources such as Web pages, as well as new types of resources, such as concepts (e.g., Geography), people (e.g., Albert Einstein), places (e.g., Berlin), things (e.g., a chair), companies … Those are considered *non-information resources*. They are identified by URIs and described with RDF statements.
We believe that the Future Internet tends to blur the frontiers between traditional infrastructure layers. The data layer of the Future Internet is not only dedicated to ensuring interoperability between applications. It is present at different levels of the architecture. We illustrate this idea by discussing the addition of semantics at the

network level and the addition of interactivity rules at the content level. We argue that this will both raise new challenges and create new opportunities for data linkage.
We present key concepts developed for the Future Internet, namely Content Centric Networks and Content Objects. We then show how the Future Internet infrastructure can be enriched by Linked Data and how Linked Data can benefit from Future Internet technologies.

2 Content Centric Networks: adding semantics at the network level

In Content Centric Networks, the network does not address packets but content. Semantics is added at the level of data packets. While in traditional networks, each node is represented by a unique identifier (IP address), in Content Centric Network each packet is uniquely represented by its name and each node can be a data provider. There is a shift from the client – server paradigm to a new one, where the data itself is at the center of the network. Van Jacobson et al. (2009) define a straightforward content naming scheme that allows encoding metadata in data packets. These metadata include the content producer as well as keywords. They are encoded in a hierarchical form. Hence, each keyword added to the name of a data packet gives a more precise definition of the content of the packet. This allows basic semantic content retrieval to be performed at the network level (through the broadcasting of interest packets) without impairing too much the efficiency of the underlying network.

If the naming scheme was to include too much semantics, this would be at the expense of the overall performance. However, the addition of semantics at the network level aims to make content addressing more efficient, by blurring the traditional separation between infrastructure layers (OSI7). This is especially the case for the Future Internet where mobile terminals have a larger place: the same piece of data can be retrieved from several different nodes, allowing the use of optimized data dissemination and retrieval techniques. Indeed, Content Centric Networks suggest that semantics can make the Internet more efficient across infrastructure layers.

3 Making content intelligent

Zahariadis et al. (2010) have proposed a content model suitable for a Content Centric Internet. Although they do not refer to Van Jacobson's framework for Content Centric Networks, the concept is similar. The latter exposes a network perspective, while the former takes the perspective of multimedia resources. Indeed, Zahariadis et al. (2010) focus on the way in which content should be represented if the content centricity concept is disseminated across other layers of the architecture. In the content model presented, i.e. Content Objects, different types of properties are assigned to content, including classic descriptive properties (*Characteristics*; e.g., dcterms:Creator), linkage properties (*Relations*), as well as properties which direct the interaction of content with external objects and services (i.e., *Rules* and *Behaviour*).

In those property categories, it is possible to encode actual rules to inform the behavior of a search engine for instance, although the application can use or ignore these rules.

Part of the behaviour of applications and services can therefore be encoded at the level of content rather than at the level of applications. For example, in order to support targeted content recommendations, the target audience of a resource can be defined as content properties (Naudet et al., 2010). Whereas the recommender system can use or ignore that information, it is possible to record that the content is relevant to people who are located close to Le Louvre museum.

We have implemented the Content Object model described by Zahariadis et al. (2010) as simple RDF metadata to support the targeted delivery of resources, according to both the user profile and his context (Naudet et al., 2010). We have extended this model (Figure 1) with two main categories, to support content provenance and usage data, which are the basis of recommendation mechanisms (Wolpers et al., 2007). We plan to enrich the implementation of interactivity mechanisms in the RDF representation of Content Objects, based on our previous work in the multimedia environment (Renault et al., 2006).

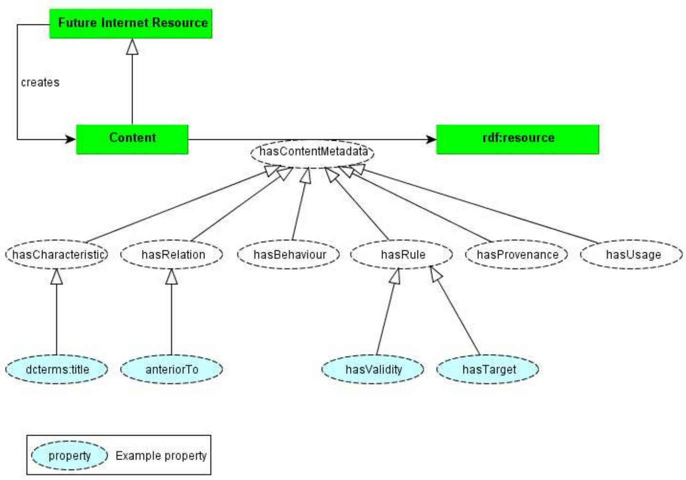

Figure 1 – Extended Content Object structure, using RDFS sub-properties

Whereas the Content Object model initially focuses on media content (Zahariadis et al., 2010), it provides a model for any content that is transmitted across the Future Internet networks.

Search engines, recommender systems, and the content matching mechanism of CCNs can support content retrieval based on keywords and particularly keywords extracted from descriptive properties. Nevertheless, other properties can take the form of rules or be interpreted as rules to express the way in which applications should use content, following a similar pattern as smart digital objects (e.g., smart graphics; Piombo et al., 2010). This offers an opportunity for applications to handle content more efficiently, whereas at the level of the application layer, content is usually passive.

4 Linking resources on the Semantic Web

Linked Data represent resources on the Web as structured data (RDF). They describe both document-like resources and non digital resources, such as places and concepts. Each resource is identified by a URI. The datasets thus created are interlinked through one of the following mechanisms: either through the use of a URI from a distinct dataset (e.g., instead of creating a new concept with a local URI, it is possible to use http://dbpedia.org/page/Category:Geography from the DBPedia dataset), or through the identification of similar concepts in other datasets (e.g., my Geography concept http://mynamespace/Geography is the same as - *owl:sameAs* - the Geography concept in the DBPedia dataset http://dbpedia.org/page/Category:Geography). These mechanisms allow navigating across datasets created in different environments. Linked Data enable machines to access the semantics of the Web and reason on it. RDF data are valuable because they are interlinked, i.e. they use resources from different datasets. Nevertheless, linkage raises issues in the current Semantic Web environment.

A major drawback of creating new URIs and linking them through an *owl:sameAs* property is that it is necessary to run inferences in order to navigate through the Web of data. However, using directly an existing resource defined by a third party represents a risk. The data source should be considered reliable and persistent. It is necessary to be certain that the resource represents the exact same concept as the one we refer to, and that this concept will not evolve. The category http://dbpedia.org/page/Category:Geography for instance can evolve over time according to the resources to which it is applied or the descriptive statements associated to it. As a result of those constraints in particular, most links are performed to a limited number of datasets (such as DBPedia) (Daniel, 2010).

In addition, the linkage is currently triggered by one party, i.e., the data author who decides to reuse an existing URI from a different dataset or to create a relation to that URI in his dataset). A person creating RDF data about a resource can look for an existing resource in different datasets through indexes (such as Sindice[1]) and SPARQL interfaces of specific datasets (e.g., http://dbpedia.org/sparql). The author of an existing resource does not however receive the information that a new resource was created in another dataset, which might be related to his existing resource. If none of the parties takes the initiative of linking resources, two URIs may exist with no relation, for instance http://mynamespace/Einstein and http://dbpedia.org/resource/Albert_Einstein.

Future Internet infrastructures can provide new solutions to this issue. We illustrate in the next sections how Future Internet resources can use Linked Data as part of their communication mechanism and how the Future Internet can support data linkage, through proactive linking mechanisms based on the Content Centric infrastructure.

[1] http://sindice.com/

5 Linked Data as part of the Future Internet resources' communication mechanisms

The Content Centric infrastructure should be based on CCNs for data transport, Content Objects to structure the content to be transported, and semantics composed of interactivity rules and metadata, to support the effective communication of content by Future Internet resources.

All types of devices can publish what they sense on the Internet (e.g. Blogjects, Bleecker, 2006). An RFID tag for instance (passive device) can allow documenting a user context. Alternatively, the user mobile phone (sensing device) can sense his location and all events, which happen around him. In both cases, these data can be published. Future Internet resources can be active and generate content for the Web. They will natively produce structured content (as opposed to content originally created by humans). They have the potential to create an enormous mass of structured data on the Web (e.g., from sensor datastreams). Data provenance[2], including the context under which data were created, is key to the publication of usable data on the Future Internet, where all resources will be potential data authors.

However, the creation of large quantities of data can create a challenge for data linkage mechanisms. The current tools and interfaces for data linkage may not be suitable for such massive data generated automatically by Future Internet resources, acting more or less autonomously. The development of automated or semi-automated approaches can support the growth of data on the Future Internet. The automatic identification of relations between ontology concepts or instances can be performed based on their properties (Giunchiglia et al., 2007), using linguistic processing on labels or comparing different properties. The relation created depends on the properties used to establish the linkage and the proximity of the concepts. Such mechanisms however need to be fully integrated in Future Internet infrastructures and the lifecycle of resources.

6 Content centricity to support proactive linking mechanisms

The content centricity concepts developed for the Future Internet can also provide the means for a different linking mechanism, based on the definition of linking behaviors. If a content provenance is documented (including the context in which it was created), a Content Object containing the content can have a Rule specifying a linking mechanism to any other content, which would have similar context properties for instance. This can take the form of a new RDF statement created as Relation or a *owl:sameAs* statement added to a specific resource. The Content Object can even have a Rule to indicate to network nodes (from a Content Centric Network) that they could broadcast an Interest Packet (i.e., a CCN request) on a regular basis to search for other Content Objects having the same creation context properties. Network nodes would then execute or not such Rules, according to their own capabilities and priorities.

---

[2] http://www.w3.org/2005/Incubator/prov/wiki/W3C_Provenance_Incubator_Group_Wiki

If another resource description has to be modified (e.g. the presenter of a talk has already been represented in another dataset), the linking suggestion mechanism should also be expanded to that resource. Since content naming schemes can be built from descriptive content properties (such as Characteristics), the network itself can contribute to the retrieval of potential links. The Future Internet can improve linkages, based on Content Objects interactivity mechanisms and their interpretation by semantic network nodes.

7 Towards a data linkage infrastructure

We have illustrated that the traditional infrastructure layers will be blurred in the Future Internet. Networks use semantics at the level of data packets to make data transport more efficient. In turn, content properties can provide indications to applications on how they can and should interact with the content. Links between resources are needed across different infrastructure layers.

Future Internet resources are expected to produce huge amounts of structured data on the Web, which will need to be integrated with other datasets. This will raise a number of challenges related to the scalability of existing tools and mechanisms. The linkage of those data will be a core challenge of the Future Internet. Mechanisms based on Content Objects and Content Centric Networks for instance can be implemented to proactively link data. Future research should investigate the implementation of the linkage mechanisms in the interactivity properties of Content Objects and the experimentation of semi-autonomic execution of those mechanisms on CCN networks.

Acknowledgments. This paper has benefited from the work carried out in the scope of the CLAIRVOYANT project (Context-Aware Personalized Mobile Services in Self-Organised Hybrid Networks, FNR C09/IS/12), carried out as a collaboration between the Tudor Research Centre in Luxembourg, the Technical University of Berlin, the University of Metz, the University of Evry, and the INSA engineering school in France. It was supported by the National Research Fund, Luxembourg. In addition, it benefits from the work carried out by Ralph Esber from Antonine University (Lebanon).

5 References

1. Bleecker, J. A Manifesto for Networked Objects — Cohabiting with Pigeons, Arphids and Aibos in the Internet of Things. (2006). Retrieved November, 2010, from http://www.nearfuturelaboratory.com/files/WhyThingsMatter.pdf
2. Daniel, R.: Six Step SAFARI from the Dublin Core to the Semantic Web. Tutorial at the Dublin Core conference 2010, Pittsburgh. (2010). Retrieved November 2010 from http://dublincore.org/resources/training/dc-2010/Tutorial4_SAFARIDaniel.pdf.
3. Giunchiglia, F., Yatskevich, M., Shvaiko, P.: Semantic matching: Algorithms and implementation. Journal on Data Semantics (2007).

4. Jacobson, V., Smetters, D.K., Thornton, J.D., Plass, M.F., Briggs, N.H., Braynard, R.L.: Networking named content. Proceedings of the 5th international conference on Emerging networking experiments and technologies (2009). pp 1-12.
5. Missikoff, M., Drissi, S., Giesecke, R., Grilo, A., Li, M.-S., Mehandjiev, N., Werth, D. (FInES Research Roadmap Task Force): Future Internet Enterprise Systems (FInES) Research roadmap. European Communities (2010).
6. Naudet, Y., Schwartz, L., Mignon, S., Foulonneau, M.: Applications of user and context aware recommendations using ontologies. In Proceedings of the Human Machine Interface (IHM) conference 2010, Luxembourg. (2010).
7. Piombo, C., Charvillat, V., Grigoras, R.: Adapting Smart Graphics' Behaviour to Users' Characteristics. Proceedings of the 12th International Workshop of the Multimedia Metadata Community, Kosch, H., Klamma, R., Lux, M., Spaniol, M., Stegmaier, F. (eds.) CEUR, vol. 680. (2010).
8. Renault, S., Gateau, B., Mathevon, D., Naudet, Y.: Implementation of RAMO-based Multimedia Applications on MPEG-4 Platform. International Conference on Consumer Electronics. (2006).
9. Wolpers, M., Najjar, J., Verbert, K., Duval, E.: Tracking Actual Usage: the Attention Metadata Approach. International Journal Educational Technology and Society, ISSN: 1436-4522 (2007).
10. Zahariadis, T., Daras, P., Bouwen, J., Niebert, N., Griffin, D., Alvarez, F., Camarillo, G.: Towards a Content-Centric Internet. In G. Tselentis et al. (Eds.) Towards the Future Internet - A European Research Perspective. IOS Press, Amsterdam (2010).