Integration “In The Large”

Ian Oliver & Ora Lassila
Nokia

September 16, 2011

The original goals of the Semantic Web [2] – and in some degree also those of its “younger cousin” Linked Data [3] – are to provide novel solutions for the integration and interoperability of systems, and to move towards a situation where more information work can be automated. These solutions are well understood “in the small”, as we can relatively easily define and expose the semantics of any system of limited scope, but larger-scale integration still eludes us. This is due to the fact that ultimately reconciling the semantic differences of many individual systems *at the level of data* is destined to suffer from issues of achieving consistent global agreements (and as such we are dealing not only with technical issues, but also – and perhaps more importantly – with social and organizational challenges).

This leads us to an “existential crisis” of the Semantic Web that derives from the *angst* of trying to integrate everything *at the data level*. The plurality of ontologies drives the impossibility of mapping all these together to produce a single, coherent and consistent “global ontology”. After all, ontologies are sets of arbitrarily defined terms used for some form of reasoning and for some specific purpose (not to mention the various kinds of reasoning over ontologies as dictated by the different description logics ALC, ALCH, ALCHI, SHOIQ(D), etc.).

More recently, *cloud computing* has emerged as a new paradigm[1] of distributed information systems, and brings with it a host of interesting interoperability challenges [4], many of which can be seen as directly related to data semantics.

Our proposal, therefore, is to look at the problem of integration at a higher level, at the level of entire systems and information flows between these. For example, Seligman and Barwise [1] provide frameworks for describing information flow and the integration of these flows. In particular, the ideas of local logics from their theory of information flow provide us with the required structure to understand how we may better look at and define the “macroscopic structure” of the Semantic Web [14].

[1]Whether cloud computing really is a *new* paradigm is of course debatable.

We have through experience noticed and constructed a "hierarchy of information scales".[2] Individual, "small" pieces of information can be easily circumscribed and made internally self-consistent with regard to their content and semantics [13], e.g., dates, times, addresses, names. Mapping between these structures is relatively straightforward with the ability to preserve semantics over the mapping and, additionally, to preserve the ability to consistently map back and forth as well as compose mappings together: we call this *the principle of bijections and transitivity* of semantic mapping.

As these smaller structures get embedded in larger structures, say, calendar entries, the structure as a whole becomes more difficult to map to other comparable structures. While individual elements of a calendar entry can be successfully mapped to another ontology, the information contained in the structure as a whole is often partially lost.[3] Such mappings are over subsets of the source information, and as consequences of this

- some source information is not mapped,
- the interpretation by the receiver is a subset of the interpretation of the sender,
- the interpretation of some information by the receiver may not be compatible with the original interpretation.

As mappings at this level are composed we see repeated "subsetting" of information as well as corruption (and subsetting) of the original, intended semantics. Certainly it can be seen that as soon as subsetting of information occurs then forming bijections between sets of information becomes impossible. Additionally, under some circumstances, transitivity is also lost.

As users become more accustomed to creating their own corpora of information, we soon see a third level of hierarchy forming. Often at this level, a particular user will spread their information over many systems (e.g., Nokia Ovi, Facebook, Google+) with the effect that *their* particular interpretation is consistent, but interpretation by other users (biological or artificial) in constituent systems is different. This leads neatly to the next level of hierarchy: Within larger systems such as social network providers we have consistent information but at a larger scale of information chunking. Currently this appears to be the highest level of granularity, though a higher level may again be forming based on national or cultural interpretations [8]; this is under investigation.

Given that we can identify a number of levels of information hierarchy we also see that each level has looser mapping properties such as preservation of information consistency and semantics. In our experience it is not clear whether this is because the difficulty of mapping and preserving properties such as bijection and transitivity are hard (in the mathematical

[2] Much like the 'clumping' of structures that has been observed in cosmology [7, p.71].

[3] We note that commendable work has been undertaken over several years in the area of *ontology mapping* [5]; without this work much of the promise of the Semantic Web would be an even more distant goal.

sense) or whether this is because of the human construction of such mappings. Our experience suggests the former, and work is progressing on understanding the ramifications of not being able to preserve certain properties over said mappings.

An alternative formulation of this is to construct deeper levels of semantics (meta-, meta-meta-, etc. semantics) and construct the mappings there. Indeed this takes its inspiration from linguistic semantics which appears to provide a *natural* mechanism for encoding these relationships, and as we progress through our semantic hierarchy, we witness the *breaking* or *deconstruction* of the relationships and properties between the lower-levels in the aforementioned semantic hierarchy. As the discussion in [1] shows, as well as other work on the relationship between representation and semantics has shown [11, 12], this will be difficult.

At the time of writing, we cannot offer a fully thought out solution but rather the framework and concepts in which we are formulating our work and ideas. Our earlier work on applying Semantic Web to *information spaces* [9, 10, 6] does suggest, however, that a macroscopic metastructure over "semantic systems" (with individual, different semantics) is a promising direction and one that does appear to successfully capture what we are seeing in our work (and particularly the concrete implementations of that work).

It is our belief that the current Web architecture and W3C's "arsenal" of technology specifications offers an excellent starting point to pursue a practical solution in the broader context of cloud computing. Specifically, the current set of Semantic Web specifications (RDF, OWL, SPARQL, GRDDL and RIF) should be considered as indispensable building blocks for the next generation of integrated information architectures.

Our contribution supporting this, as we have attempted to outline here, is to place and possibly extend this "arsenal" of technology within a much larger and more formalized hierarchy and superstructure that clearly explains how these technologies interact and fit together.

References

[1] J. Barwise and J. Seligman. *Information Flow: The Logic of Distributed Systems*. Cambridge University Press, 1997.

[2] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, May 2001.

[3] C. Bizer, T. Heath, and T. Berners-Lee. Linked Data – The Story So Far. *International Journal of Semantic Web and Information Systems*, 5(3):1–22, July-September 2009.

[4] S. Boldyrev, D. Kolesnikov, O. Lassila, and I. Oliver. Data and Computation Interoperability in Internet Services. In *Proceedings of the Fifth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM 2011)*, Lisbon, Portugal, 2011 (to appear).

[5] J. Euzenat and P. Shvaiko. *Ontology Matching*. Springer-Verlag, 2007.

[6] D. Khushraj, O. Lassila, and T. Finin. sTuples: Semantic Tuple Spaces. In *First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (Mobiquitous '04)*, pages 268–277, Boston, Massachusetts, USA, Aug. 2004. IEEE Computer Society Press.

[7] R. P. Kirshner. *The Extravagant Universe: Exploding Stars, Dark Energy and the Accelerating Cosmos*. Princeton University Press, 2002.

[8] O. Lassila. *Sharing Meaning Between Devices, Systems, Users, and Cultures*. Keynote address at the French-Finnish Symposium on Digital Semantic Content Across Cultures, Le Louvre, Paris, France, May 2006.

[9] I. Oliver. Information Spaces As a Basis for Personalising The Semantic Web. In *Proceedings of the International Conference on Enterprise Information Systems (ICEIS'09)*, Milan, Italy, 2009.

[10] I. Oliver. M3 Information SmartSpaces Technology Overview. Technical Report NRC-TR-2009-M3-1, Nokia Research Center, Helsinki, Finland, 2009.

[11] C. Sayward. The Received Distinction Between Pragmatics, Semantics and Syntax. Foundations of Language 11, 1974

[12] P. Smith. The Galois Connection between Syntax and Semantics. University of Cambridge, 2010.

[13] B. Cantwell-Smith. On the Origin of Objects. The MIT Press 9-780262-692090. 1996

[14] P. Gärdenfors. Conceptual Spaces: The Geometry of Thought. MIT Press/Bradford Books. ISBN 978-0585228372. 2000