Standardized usage patterns for Linked Data

Martin Nally, Steve Speicher, Arnaud Le Hors – IBM Corporation

There is interest in Linked Data technologies for more than one purpose. We have seen interest for the purpose of exposing information – for example public records – on the Internet in a machine-readable format. We have also seen interest in the use of Linked Data for inferring new information from existing information, for example in pharmaceutical applications or IBM Watson. The IBM Rational team has been using Linked Data as an architectural model and implementation technology for application integration.

IBM Rational is a vendor of software development tools, particularly those that support the general software development process such as bug tracking, requirements management and test management tools. Like many vendors who sell multiple applications, we have seen strong customer demand for better support of more complete business processes - in our case software development processes - that span the roles, tasks and data addressed by multiple tools. This demand has existed for many years, and our industry has tried several different architectural approaches to address the problem. Here are a few:

1. Implement some sort of application programming interface (API) for each application, and then, in each application, implement "glue code" that exploits the APIs of other applications to link them together.
2. Design a single database to store the data of multiple applications, and implement each of the applications against this database. In the software development tools business, these databases are often called "repositories".
3. Implement a central "hub" or "bus" that orchestrates the broader business process by exploiting the APIs described in option 1.

While a discussion of the failings of each of these approaches is outside the scope of this document it is fair to say that although each one of them has its adherents and can point to some successes, none of them is wholly satisfactory. So, as an alternative, we have been exploring over the last 5 years the use of Linked Data as an application integration technology. We have shipped a number of products using this technology and are generally pleased with the result. We have more products in development that use these technologies and we are also seeing a strong interest in this approach in other parts of our company.

We are pleased – even passionate – about the results we have seen using Linked Data as an integration technology but we have found successful adoption to be difficult. It has taken us a number of years of experimentation to achieve the level of understanding that we have today, we have made some costly mistakes along the way, and we see no immediate end to the challenges and learning that lie before us. As far as we can tell,

there is only a very limited number of people trying to use Linked Data technologies in the ways we are using them, and the little information that is available on best practices and pitfalls is widely dispersed. We believe that Linked Data has the potential to solve some important problems that have frustrated the IT industry for many years, or at least make significant advances in that direction, but this potential will only be realized if we can establish and communicate a much richer body of knowledge on how to exploit these technologies. In some cases, there also are gaps in the Linked Data standards that need to be addressed. To help with this process, we would like to share information on how we are using these technologies, the best practices and anti-patterns we have identified, and the specification gaps we have had to fill ourselves. We hope that the workshop will give us a forum in which to present these ideas, and that it will also lead to the formation of a community that will build the body of experience, guidance and specifications to make Linked Data a successful technology for addressing application integration problems.

The best practices and anti-patterns can be categorized (but are not limited) to the following:

- Resources - a summary of the HTTP and RDF standard techniques and best practices that you should use, and anti-patterns you should avoid, when constructing clients and servers that read and write linked data.
- Containers - defines resources that allow new resources to be created using HTTP POST and existing resources to be found using HTTP GET
- Validation - defines a simple mechanism for describing the properties that a particular type of resource must or may have
- Paging - defines a mechanism for splitting the information in large resources into pages that can be fetched incrementally

We think that these areas would provide a good basis for which new standards could be developed, possibly in the form of profiles. We have provided a basis for what these pattern definitions in a publication titled ["Towards a Basic Profile for Linked Data"](). Additional guidance and best practices around resource identity stability will ensure the success of these patterns.

We propose the workshop address the follow questions:

1. How do we define a standardized model based on the World Wide Web that meets the needs of application and tool integration based on the core Web-principles of loosely-coupled and scalable?

2. How do we define a recommended set of standards and best practices for utilizing HTTP with RDF for resource access?

3. How can we build the simplest model possible that supports these usage patterns, even through extension?

4. How can we build from existing standards, such as HTTP, RDF, and SPARQL, a framework such that the existing ecosystem of Linked Data applications can evolve to support a standardized set of patterns?

5. How can we clearly define an approach and solution that can be specified and implemented in a timely fashion so the community can get immediate value from it?

We plan to work with this workshop, any subsequent working group and the community to ensure that the answers to these questions provide a solid architectural approach to the scope of the problems stated. We have invested a large amount effort in exploring various alternatives and delivered successful solutions. Our experiences with these approaches will enable us to develop and deliver on solutions. We are committed to working through these questions, feedback, resolution, solutions to provide the needed scalable and robust definitions for standardization and success.