

Quality Assessment on Enterprise Linked Data

Authors: Carlos Eduardo Montoya Sanchez, Rauf Aghayev, Vugar Aghayev

Supervisor: Debattista Jeremy

08.10.2014

- *Problem Definition*
- *Metrics*
- *Architecture*
- *Performance Test*

• Problem Definition

- The quality term means in Semantic Web is fitness of use.
- During the development of Web Semantic technologies huge volume of data has being published on the web as a Linked Open Data. In order to use LOD for specific purpose, we should check its quality in advance.

- **Metrics**

1. **Verifiability** – *checks Authenticity and Digital signature*

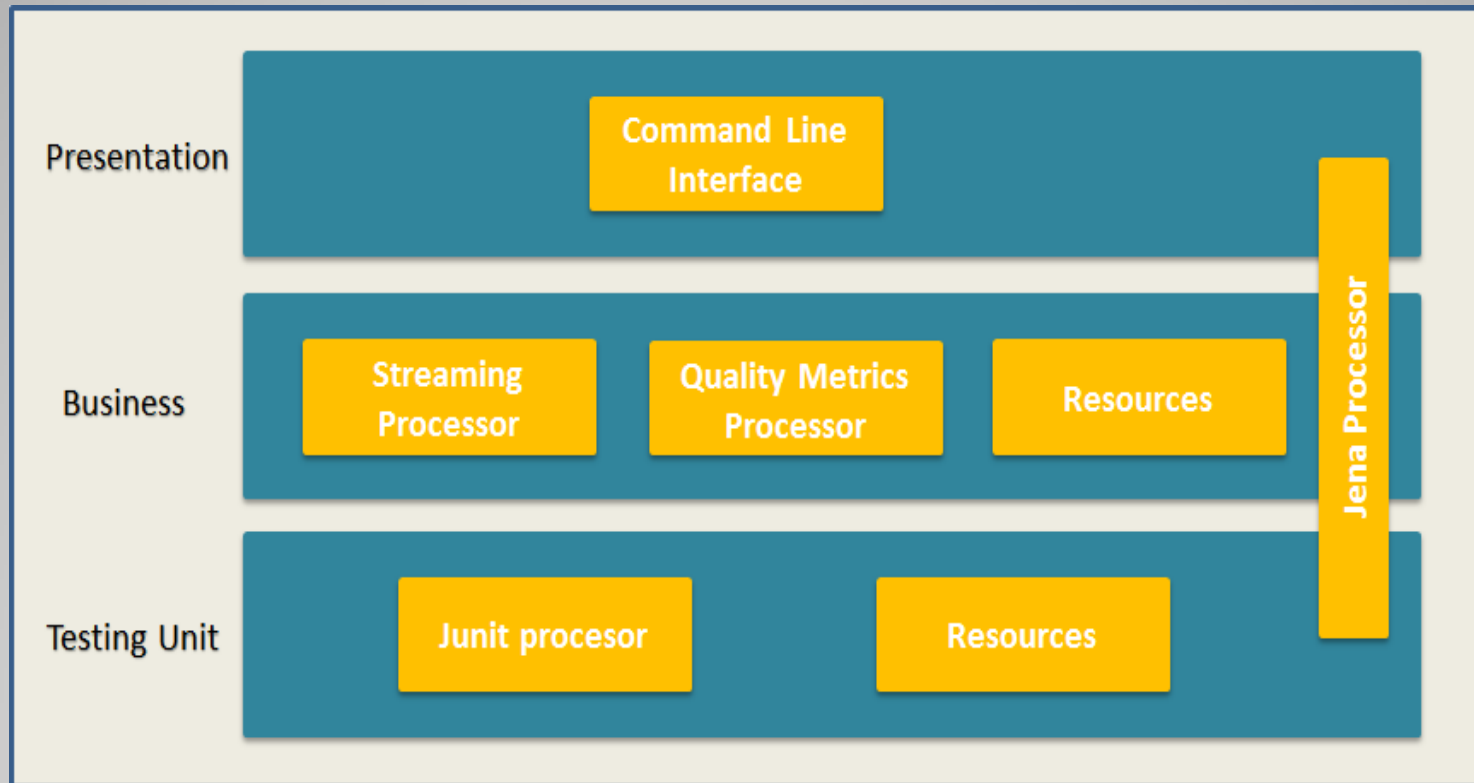
1.1. Authenticity of data set

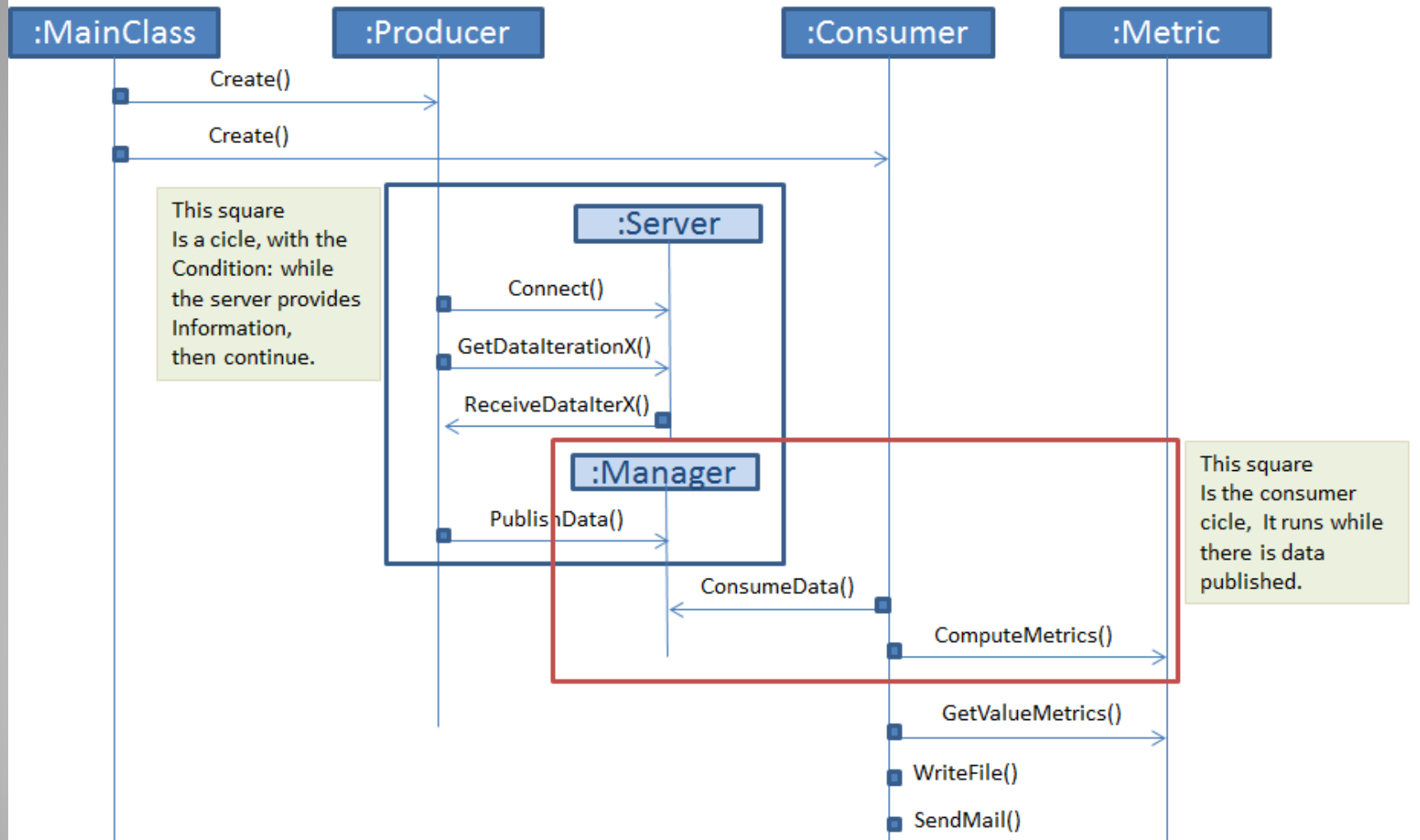
1.2. Digital Signature

2. **Free Of Error** – *Checks if triples syntactically correct.*

3. **Measurability** – *Checks if number of triples is greater than threshold and SPARQL end point works*

- **Architecture**





• Performance Test

<u>Source</u>	<u>Number Of Triples</u>	<u>Running Time</u>
http://localhost:8080/openrdf-sesame/repositories/Data_Set Data dump was taken from- https://developers.google.com/freebase/data	In data set there are 2.1M number of triples but for testing 500.000 of them was loaded.	35 seconds

Global Store

<u>Source</u>	<u>Number Of Triples</u>	<u>Running Time</u>
http://es.dbpedia.org/sparql	In data set there are more than 10M number of triples but for testing 500.000 of them was loaded.	53 seconds

