

Size Inflation of PingER Data for use in PingER LOD

Author

Cristiane Ceia UFRJ, cristianeceia@gmail.com

Abstract

This quantifies the Inflation in size of PingER data as it is prepared for Linked Open Data (LOD) access. The size of the [PingER hourly data for 2005-Sep 2014 archived via FTP in text form](#) amounts to ~ 3.12GB and this corresponds to 15.66×10^9 (billion) triples. Then using 5 triples for each measurement and using Turtle without compression gives us 685 Gbytes or an inflation factor of ~ 200.

Method

In order to have the number of PingER triples, I processed the quantity of measurement values on PingER hourly data from 1998 to September 2014 (packet size: 100 bytes).

Below, we can see how many measurement values we have per year.

	#Measurements
1998	6,740,974
1999	8,617,718
2000	11,617,057
2001	13,137,702
2002	7,247,257
2003	14,690,615
2004	36,060,787
2005	32,745,602
2006	38,461,602
2007	89,549,322
2008	115,999,447
2009	150,312,565
2010	203,265,500
2011	441,150,811
2012	697,272,874
2013	733,745,502
2014	531,572,876
Total	3,132,188,211

These measurement values generate 15,660,941,055 triples.

I am considering a basic description of a measurement following Renan's PingER LOD ontology, in which a measurement is minimally defined by 5 triples. Here is an example:

@prefix : <<http://www-iepm.slac.stanford.edu/pinger/ld/resource#>> .

@prefix o: <<http://www-iepm.slac.stanford.edu/pinger/ld/ontology/PingEROntology.owl#>> .

:EDU.SLAC.STANFORD.N3-BR.UFRJ.PINGER-AverageRTT-15Feb03H23 a o:Measurement ;

o:measuresMetric :AverageRTT ;

o:hasSourceDestinationNodes :EDU.SLAC.STANFORD.N3-BR.UFRJ.PINGER ;

o:hasDateTime :Time15Feb03H23 ;

o:hasValue 233.926 .

The volume in bytes to define one measurement, stored as plain text, utilizing RDF turtle format, WITHOUT any compression or indexing techniques (which commonly reduces size of data and is dependent on the Triple Store we are going to use) gives us 235 bytes. Hence, the estimate for the total triplified data volume in bytes is $235 * \text{\#Measurements} = 736,064,229,585$ bytes (about 685.5 GB).