From: **Mo McRoberts** Mo.McRoberts@bbc.co.uk
Subject: Position Statement - W3C Workshop on Web Standardization for Graph Data
Date: 11 January 2019 at 12:42
To: group-data-ws-pc@w3.org

Hi workshop committee,

Please find below my position statement to accompany the application form submission for the Berlin workshop. Do get in touch if I've missed anything obvious.

My role at the BBC is concerned with solutions for managing all of the knowledge the BBC has: information about our output programmes, news stories and web pages; information about the topics they relate to, including people, places, events, musical artists, contributors and production staff, and how they all relate to each other; technical metadata at every point in the production chain from conception right through to licensing to Amazon and Netflix or selling Blu-Ray discs; information about our archive holdings (which range in format from wax cylinders to digital files on disk, tape, and other media); and of course information about how and where people consume our output. We have a diverse information management estate with nearly a century's worth of data stored in thousands of different systems. My goal is to make as much of that as possible easily accessible to the right people and services, so that our staff can work efficiently, that we can build new products that bring more of what we have to our audiences, and develop high-performance data pipelines that can be used for enrichment and other machine learning processes. Necessarily, we need to support both high-performance audience-facing services that can scale up to thousands of requests per second without costing our entire operating budget in some contexts, and also the ability to perform complex graph traversal and manipulation in others.

Unsurprisingly, this is not a small set of challenges. RDF is frequently a good fit (indeed, for a lot of the work I do, RDF is the only logical fit short of reinventing it, badly) — in principle — for the kind of data we have. Indeed, one of the things we're building is an internal linked data ecosystem to provide a unified data access layer across a dizzying array of data systems, and that is relatively straightforward in and of itself. However, the management, storage, indexing and querying of that data is a much tougher nut to crack. The breadth and maturity of available RDF-based database systems is limited, particularly compared to RDBMSes, but often also when compared to property graph products. SPARQL is not an especially pleasant query language to use. In the past we've found reasoning easier to avoid altogether than to make work efficiently at our scale. In some of our work, the most flexible path has been to eschew graph databases altogether, and instead store sets of RDF graphs in basic object storage buckets, referencing it from a simple set of relational database tables; for more complex query needs, we can then feed subsets of the data into more specialised databases.

This is workable, but far from ideal. However, it is indicative of my position: I do a lot of work with RDF because it's the right tool for the job, but much of the ecosystem around it isn't, and I wouldn't be unhappy if much of it was consigned to history. I'm actively in favour of exploring the feasibility of adapting the best data management systems we have today — whether they began life as SQL databases or property graphs or something else — to be able to use them effectively with RDF data. In terms of query languages, I consider it a priority issue that a new standards-track graph query language be able to represent meaningful queries across RDF graphs (or if not, there are well-defined ways *reliably* *round-trip* arbitrary sets of RDF quads into property graphs which can then be queried in the same way in those forms, and so achieving the same by proxy).

As an aside, as it was mentioned in the topic list - I have in the past prototyped a hashing algorithm for RDF graphs (taking bNodes properly into an account, including dealing with the entire graph being bNodes), and the issue of digitally-signing graphs is one of some considerable professional interest, but I'm far from expert.

All the best,

M.

--
Mo McRoberts — Lead Technical Architect — Data,
BBC Technology Strategy & Architecture
SMS/iMessage/WhatsApp/Telegram/Signal: +447740177832
https://neva.li/whereami

Further communication will signify your consent to this.
