First impressions of the W3C workshop on Graph Data

SDI.Next: Linked Spatial Data in Europe
March 12, 2019
Amersfoort, The Netherlands

Ivan Herman, W3C
Warning: these are really the first impressions

- The workshop took place last week, ended on Wednesday…
- …and I has a few days off after the event 😊
The facts

- The Workshop on “Web Standardization for Graph Data”:
  - took place in Berlin, 4-6 March 2019
  - there were ≈100 participants
  - one keynote (from Amazon), ≈20 presentations, and a bunch of short presentations
  - lots of discussions, panels
  - program, submissions, etc, are available via: https://www.w3.org/Data/events/data-ws-2019/
### Lots of sessions...

<table>
<thead>
<tr>
<th>Time</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00</td>
<td><strong>Chairs/PC Synch</strong></td>
<td><strong>Interoperation</strong></td>
<td>09:00</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Graph Data interchange</strong></td>
<td><strong>Introduction &amp; Reports from</strong></td>
</tr>
<tr>
<td>09:30</td>
<td></td>
<td><strong>Easier RDF and next steps</strong></td>
<td><strong>Tuesday’s sessions</strong></td>
</tr>
<tr>
<td>10:00</td>
<td></td>
<td><strong>SQL and QQL</strong></td>
<td>10:00</td>
</tr>
<tr>
<td>10:30</td>
<td></td>
<td><strong>Break</strong></td>
<td>10:30</td>
</tr>
<tr>
<td>11:00</td>
<td></td>
<td></td>
<td>11:00</td>
</tr>
<tr>
<td>11:30</td>
<td></td>
<td><strong>Graph query interoperation</strong></td>
<td>11:30</td>
</tr>
<tr>
<td>12:00</td>
<td></td>
<td>**Composition, patterns and</td>
<td>12:00</td>
</tr>
<tr>
<td></td>
<td>OPEN</td>
<td><strong>tractability</strong></td>
<td>**Extending, incubating,</td>
</tr>
<tr>
<td>13:00</td>
<td><strong>Intro &amp; Keynote</strong></td>
<td><strong>Lunch</strong></td>
<td><strong>Initial:</strong></td>
</tr>
<tr>
<td>13:30</td>
<td><strong>Intro &amp; Keynote</strong></td>
<td></td>
<td><strong>12:30 CLUB</strong>:</td>
</tr>
<tr>
<td>14:00</td>
<td><strong>Venues &amp; Vendors</strong></td>
<td><strong>Specifying a Standard</strong></td>
<td><strong>CLOSE:</strong></td>
</tr>
<tr>
<td>14:45</td>
<td><strong>Break</strong></td>
<td><strong>Queries and computation</strong></td>
<td></td>
</tr>
<tr>
<td>15:00</td>
<td><strong>Coexistence or Competition</strong></td>
<td><strong>Break</strong></td>
<td></td>
</tr>
<tr>
<td>15:30</td>
<td><strong>Break</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16:00</td>
<td><strong>Lightning Talks</strong></td>
<td><strong>Graph models and schema</strong></td>
<td></td>
</tr>
<tr>
<td>16:30</td>
<td><strong>Lightning Talks</strong></td>
<td>**Temporal, spatial and</td>
<td></td>
</tr>
<tr>
<td>17:00</td>
<td><strong>Preview of next day</strong></td>
<td><strong>Outreach and education</strong></td>
<td></td>
</tr>
<tr>
<td>17:45</td>
<td><strong>Preview of next day</strong></td>
<td><strong>Preview of next day</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Posters</strong></td>
<td></td>
</tr>
</tbody>
</table>
Issues leading to the Workshop 1.

- Increasing importance of graph-based data and databases in general (used in machine learning, Internet of Things, Big Data, etc.)
- The concept of Property Graphs has come to the fore, alongside RDF, for graphs
  - there is a need to find a way to see how these technologies coexist
    - discussions are ongoing on the pro-s and cons of RDF vs. PG
    - *PG is part of the graph data landscape for good!*
- RDB/SQL is also very much present in this area, too
Issues leading to the Workshop 1.

In theory…

- SQL could be extended to do everything for graphs
- SPARQL could be extended to do everything for PG and tables
- A property graph GQL that handles tables and graphs could do everything SQL can do

Issues leading to the Workshop 1.

In practice...

• That would lead to paralysis, or endless war
• Data communities have very deep social and product roots, and large to huge user bases
• Like humans, they can’t get personality transplants

Issues leading to the Workshop 2.

• There are also concerns with RDF
  • general acceptance is still relatively slow (although there are great successes)
  • there are many minor (or major…) technical issues with RDF & Co that need housekeeping

(“RDF”, in the presentation, is a shorthand for full RDF suite, i.e., RDF, RDFS, OWL, SPARQL, SHACL, etc.)
A few words about Property Graphs

Relationships can have properties (name/value pairs)

Employee
- name: Amy Peters
- date_of_birth: 1984-03-01
- employee_ID: 1

Company
- has CEO
- start_date: 2008-01-20

City
- located_in

Relationships are directional

Relationships connect nodes and represent actions (verbs)

Nodes represent objects (nouns)
Property Graphs

- Framework for representing data and metadata with a graph of nodes and links
  - both nodes and links may have name/value pairs
    - otherwise referred to as “properties”
  - nodes are “just” nodes, not necessarily URL-s
- Link annotations are very useful to assign temporal, spacial, provenance, etc, information easily

Source: neo4j text on PG: https://neo4j.com/developer/graph-database/#property-graph
Property graphs have a real success

- Non-SQL database vendors, like Neo4j, base their business on this
- Major database providers (Oracle, Amazon’s Neptune,…) incorporate PG as well as RDF stores
- There are a number of smaller (including open source) implementations (e.g, TinkerPop)
- There are a number of query languages (declarative and imperative), but not one winner (yet)
- There is work in the ISO/SQL community to incorporate PG, and define query languages
PG can be represented in RDF

- For example:
  - using reification
  - some sort of an intermediate node (usually BNode) to represent the link
  - use a named graph with a single triple
  - extend RDF to include, somehow, a triple as an entity (e.g., “RDF*”)

Source: presentation of David Booth, [http://tinyurl.com/EasierBerlin](http://tinyurl.com/EasierBerlin)
PG can be represented in RDF: RDF*

• Instead of something like:

```rdfs
Company has_ceo Employee .
s rdfs:type df:Statement .
s rdfs:subject Company .
s rdfs:predicate has_ceo .
s rdfs:object Employee .
s :start_date "2008-01-20"^^xsd:date .
```

• Use

```rdfs
<<Company has_ceo Employee>> :start_date "2008-01-20"^^xsd:date .
```

• To be seen whether this is a syntactic sugar, or RDF should be extended

Adapted from the presentation of Olaf Hartig
PG can be represented in RDF

- All these representations do exist in real products
- All have pros and cons
- There is *no* generally accepted way of doing that, i.e., none of those solutions are interoperable.
Why are PG-s interesting for the RDF community?

• They are around on the market…
• They represent, in some ways, a level of abstraction that is easier to understand:
  • by collapsing the “properties” into some sort of labels, the real, “core” aspect of a graph becomes more visible
  • helps in grasping the “essence” of a dataset without being lost in details (date, provenance, tags, etc.)
• adopting a “PG style” would be actually helpful to make RDF more understandable

“…historically, property graphs were somewhat of a reaction to the complexity of RDF. A complex standard will not be accepted by the developer community” (Juan Sequeda)
• The value of RDF is well proven, but…

• The value of RDF is well proven, but…
• Too hard for **average** development teams

Source: presentation of David Booth, [http://tinyurl.com/EasierBerlin](http://tinyurl.com/EasierBerlin)
Issues leading to the Workshop: “EasierRDF” initiative

• Email and github discussion initiated by David Booth
• See: https://github.com/w3c/EasierRDF
Over 600 messages!

EasierRDF github site: 50+ issues

Some result of the email/github discussion

- Technical issues
  - lack of n-ary relations
  - blank nodes
  - missing canonicalization/signature of graphs
  - RDF is too low ("assembly") level
  - no generally accepted and simple rule system

- Non-technical issues
  - lack of beginner level good tutorials
    - no equivalence to, say, MDN
  - no (not yet?) proper integration with Javascript
  - moribundity of tools, registries, lots of abandonware

- Connection to Property Graphs?
Results of
the Workshop
Results of the Workshop: many ideas came up

- RDF*/SPARQL* documentation to be submitted to W3C, with possible incubation towards a standard
- extend JSON-LD with PG
- Standards work around PG
  - an abstract (standard) model for Property Graphs†
  - standard mapping between Property Graphs and RDF
  - standard mapping between Property Graphs and Relational Data†
  - W3C Community Group for Graph Query Language (GQL)†
- RDF improvements
  - RDF for stream processing
  - RDF for time, for geographical data, …
  - solve all the technical and outreach problems in RDF 😊

† Final work probably not at W3C
But... this can lead to chaos

• It would lead to unstructured, unrelated work, not necessarily in the right order

• Final decision and further actions:
  • a workshop report should come out in 3-4 weeks
  • set up a W3C Business Group:
    • look at the bigger story around data: data is strategic asset for companies. What are the features and mappings that are of importance?
    • derive a prioritized list of technical issues to be solved to fulfill those needs
    • spin off task forces, community groups, etc, to look at the technical issues that are of major importance
    • liaise with other organizations (e.g., ISO) for the activities that are to be done elsewhere
    • look at outreach possibilities in general
Watch this space, interesting things will happen!
Some links

- Workshop home page:
  - https://www.w3.org/Data/events/data-ws-2019/

- All submissions
  - https://www.w3.org/Data/events/data-ws-2019/papers.html

- Workshop agenda with links to slides
  - https://www.w3.org/Data/events/data-ws-2019/schedule.html

- These slides:
  - https://www.w3.org/2019/Talks/Amersfoort-IH/Presentation.pdf
Thank you for your attention