



CLINICAL INFORMATICS
RESEARCH PROGRAMME

PhD Introductory Presentation

John Booth
Data Engineer CIRP

December 2021

Great Ormond Street 
Hospital for Children
NHS Foundation Trust

Who am I and what is my current role?

- I have worked at GOSH since June 2015
 - Applied while undertaking a MSc Health Informatics (Kingston)
 - Research Data Manager – Histopathology
 - Data Analyst – Infection Control
 - Senior Data Steward, Lead Data Engineer – DRE
 - MSc Data Science (Birkbeck)
 - Data Engineer and PhD student with CIRP.
- Previously
 - Owner/founder of various small software companies
 - Database centric, multitudinous disparate clients, finally specialising in high end audio firmware which we sold under licence.
 - Construction Engineer, BSc (Eng) Civil Engineering (Imperial College)
 - Worked on various projects in the Middle East and the West Indies.

Introduction to my PhD

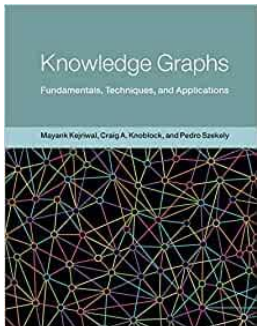
Exploring the role of graph databases in the interrogation and visualisation of routinely collected Electronic Patient Record (EPR) data

Presentation

- The World of Graphs and Graph Analytics.
- What is a Graph?
- What is a Knowledge Graph?
- What is a Property Graph?
- What is a Semantic Graph?
- An example of a Semantic Graph.
- Where do Graph Databases fit in?
- Why does this matter for EPR data?

The World of Graphs and Graph Analytics.

- Nodes or Vertices
 - Edges or Links
 - Directed or Un-directed
 - Triples
 - RDF Triples and Turtle
 - Ontologies and OWL
 - Wikipedia, DBpedia and Wikidata
- Graph Types
 - Property Graphs
 - Knowledge, Semantic and Conceptual graphs
 - Graph Databases
 - Standards?
 - Query languages
 - GraphQL
 - SparQL
 - Cypher



References:

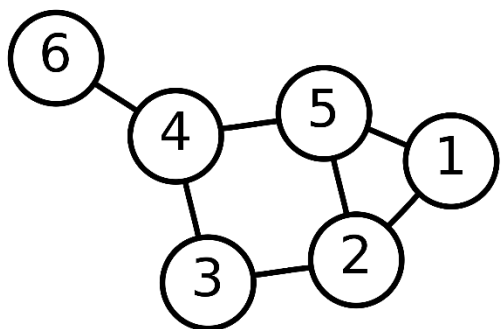
Kejriwal, M., Knoblock, C.A. and Szekely, P., 2021. *Knowledge Graphs: Fundamentals, Techniques, and Applications*. MIT Press.

Wikipedia. Available at: <https://en.wikipedia.org> (Accessed: September 2019)



What is a Graph?

- A Graph is a structure amounting to a set of objects in which some pairs of the objects are in some sense “related”.
- The objects correspond to mathematical abstractions called vertices (also called nodes or points) and each of the related pairs of vertices is called an edge (also called link or line).
- Edges may be directed or un-directed.
- A triple consists of a pair of distinct nodes connected by an edge.



A graph of six nodes and seven un-directed edges.

“In computer science, directed graphs are used to represent knowledge”

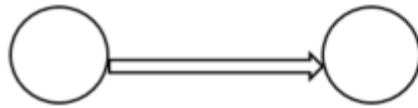
What is a Knowledge Graph?

- A Knowledge Graph is a labelled multirelational, directed graph.



Node 1 Edge 2 Node 3

A graph with no knowledge.



Patient Is administered Drug

A graph with knowledge.

What is a Property Graph?

A Property Graph is a set of nodes and edges using local identifiers. Each node and edge can be a structure of keys and values, i.e. properties.



Patient			Is Administered			Drug	
Key	Value		Key	Value		Key	Value
Name	John Doe		Route	Oral		Name	Amikacin
MRN	9999999		Datetime	2021-09-23 15:19		Sub-class	Antibiotic
Sex	Male					Class	Antimicrobial
Birth date	2000-01-01						

What is a Semantic Graph?

A semantic Graph utilises the Resource Description Framework (RDF) as a graph data model that formally describes the semantics, or meaning of information. It also represents metadata, that is, data about data.

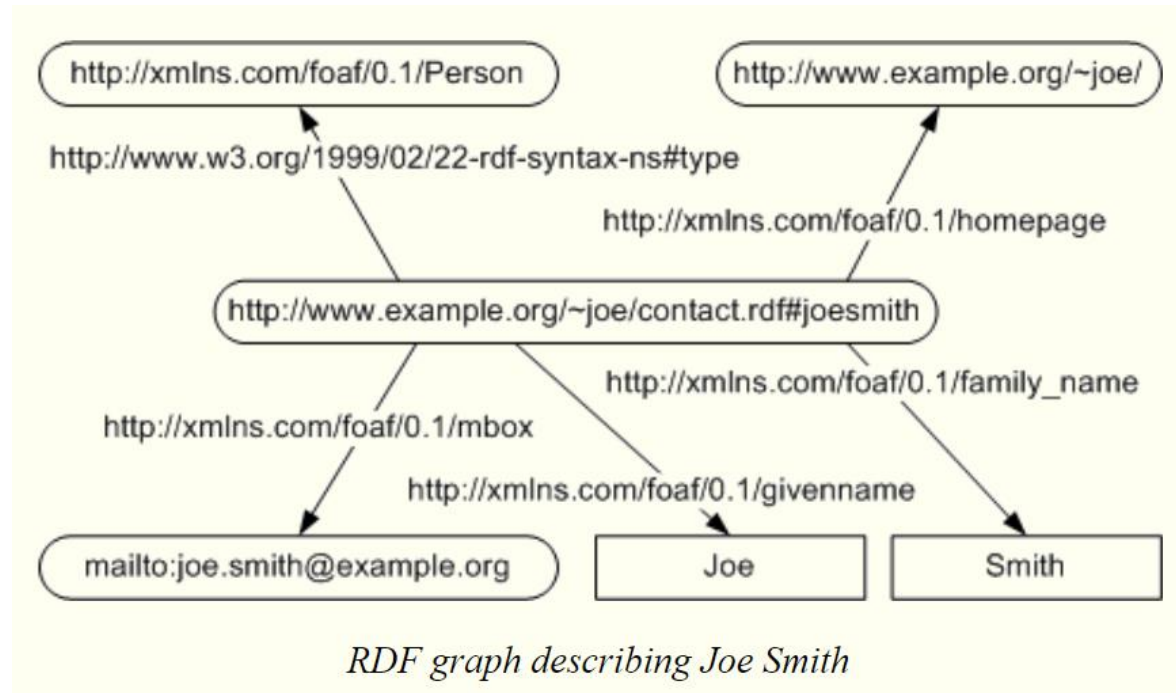
RDF consists of triples. These triples are based on an Entity Attribute Value (EAV) model, in which the subject is the entity, the predicate is the attribute, and the object is the value. Each triple has a unique identifier known as the Uniform Resource Identifier, or URI.

RDF Schema (RDFS) defines a metamodel of concepts like Resource, Literal, Class, and Datatype. RDFS provides a means for defining the classes, properties, and relationships in an RDF model and organizing these concepts and relationships into hierarchies.

sparQL is an RDF query language able to retrieve and manipulate data stored in RDF format.

Defined by the Semantic Web Standards and published by the W3C.

An example of a Semantic Graph.



<https://www.obitko.com/tutorials/ontologies-semantic-web/rdf-graph-and-syntax.html>

FOAF (an acronym of friend of a friend) is a machine-readable ontology describing persons, their activities and their relations to other people and objects. FOAF is a descriptive vocabulary expressed using the Resource Description Framework (RDF) and the Web Ontology Language (OWL).

Resource Description Framework Schema (RDFS)

<rdf:RDF

```
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns="http://www.example.org/~joe/contact.rdf#"
<foaf:Person rdf:about=
  "http://www.example.org/~joe/contact.rdf#joesmith">
  <foaf:mbox rdf:resource="mailto:joe.smith@example.org"/>
  <foaf:homepage
    rdf:resource="http://www.example.org/~joe/">
  <foaf:family_name>Smith</foaf:family_name>
  <foaf:givenname>Joe</foaf:givenname>
</foaf:Person>
</rdf:RDF>
```

Same but in TURTLE (Terse RDF Triple Language)

```
@prefix : <http://www.example.org/~joe/contact.rdf#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

:joesmith a foaf:Person ;
  foaf:givenname "Joe" ;
  foaf:family_name "Smith" ;
  foaf:homepage <http://www.example.org/~joe/> ;
  foaf:mbox <mailto:joe.smith@example.org> .
```

Where do Graph Databases fit in?

A graph database is a database that uses graph structures with nodes, edges, and properties to represent and store data. A variant of the noSQL database but with defined relationships.

- Pros:
 - Graph databases hold the relationships between data as a priority.
 - Querying relationships is fast because they are perpetually stored in the database.
 - Relationships can be intuitively visualized using graph databases, making them useful for heavily inter-connected data.
- Cons:
 - No agreed standards.
 - Graph databases that support the Semantic model should follow W3 standards
 - Neo4J is fundamentally a Property Graph.
 - No common querying language.
 - Semantic Graphs should support sparQL (sparkle)
 - Neo4J has it's own proprietary query language Cypher.
 - graphQL is not a graph query language, it's a query language for API's developed by Facebook, should have been called treeQL but it didn't sound so good.

Why does this matter for EPR data?

EPR is 'Big Data' based on the 5 'V's rule - Volume, Velocity, Variety, Veracity and Value. We need all the tools available to help us make sense of it. And Graphs and Graph Analytics are one of those tools which I believe is currently being underutilised.

- Analytics – Property graphs
- Interoperability – Semantic graphs
- Enhanced Natural Language Querying – Semantic graphs
- Visualisation
- FHIR and Semantic Graphs
 - FHIR Release 4 supports RDF, in TURTLE format, as a primary storage schema along side JSON and XML.
 - Support from both HL7 FHIR and W3C Semantic Web Health Care and Life Sciences Interest Group.

Outline Plan for PhD

Three phases starting on **Monday**:

- What are graphs and graph analytics?
 - Types
 - Databases
 - Analytics
- What are graphs good for in clinical informatics?
 - Define POCs using real EPR data
- How do you communicate with graphs?
 - To a non-data scientist i.e. a clinician