Reflecting on the Semantic Web Vision

A Personal View

Dr. Ora Lassila

Principal Technologist, Amazon Neptune

Background

Principal Technologist, Amazon Neptune graph database team

Some accomplishments

- co-authored the original W3C RDF specification
- co-authored the seminal article on the Semantic Web
- designed and implemented the frame-based KR subsystem that flew on NASA's "Deep Space 1" probe past the Asteroid Belt in 1998
- Grand Prize Winner, Usenix 1989 Obfuscated C Code Contest
- currently: co-chair of the W3C RDF-star WG

Education:

- Ph.D CS, Helsinki University of Technology

aws

Background

My whole professional career has been spent with (symbolic) AI

- founding member of the Finnish AI Society (in 1986, as an undergrad)
- planning + agents + ontologies
- M.Sc thesis on frame-based KR (this work flew past the Asteroid Belt)

Some 25+ years ago I got involved with W3C

- metadata and KR for the Web → RDF + Semantic Web vision
- Ph.D on graph queries, reasoning, and Semantic Web application development

Lately:

- knowledge graphs, graph databases, helping people make sense of it all…
- worrying about RDF/LPG ”alignment”

aws

Outline

1. Talking about old stuff
2. Talking about new stuff
3. Talking about the future

Ennustaminen on vaikeaa, erityisesti tulevaisuuden ennustaminen.

- one of my Dad's favorite sayings

(Making predictions is difficult, especially making predictions about the future.)

aws

A brief history of graphs and ontologies

3rd Century BCE: Categories & logic (Aristotle)

1730s: Graph theory (Euler)

1950s and onwards: Graphs as the essential underpinning of computer science

1960s: Social networks, “small-world experiment”, Erdős number (Milgram et al)

1960s-1970s: Network databases (CODASYL), semantic networks (Quillian et al)

1730s: Taxonomical classification of plants and animals (Linnaeus)

1870s: Library classification (Dewey)

1900: Semantics, ontology and logic (Husserl)

1970s-1990s: Predicate logic as the foundation of Knowledge Representation (Hayes et al)

1997 and onwards: The Semantic Web, RDF, OWL, etc. (Lassila et al)

Today: Modern knowledge graphs and graph databases

aws

A brief history of graphs and ontologies

1730s: Graph theory (Euler)

1950s and onwards: Graphs as the essential underpinning of computer science

1960s: Social networks, “small-world experiment”, Erdős number (Milgram et al)

1960s-1970s: Network databases (CODASYL), semantic networks (Quillian et al)

3rd Century BCE: Categories & logic (Aristotle)

1730s: Taxonomical classification of plants and animals (Linnaeus)

1870s: Library classification (Dewey)

1900: Semantics, ontology and logic (Husserl)

1970s-1990s: Predicate logic as the foundation of Knowledge Representation (Hayes et al)

This is what I will talk about

1997 and onwards: The Semantic Web, RDF, OWL, etc. (Lassila et al)

Today: Modern knowledge graphs and graph databases

aws

Original SemWeb vision

The “old” Semantic Web

Some early history

Late 1996: Tim Berners-Lee asks me a fateful question 😱

1997: Brainstorming on metadata and “KR on the Web”, start of the RDF work

1999: RDF W3C Recommendation

2000-2001: Tim, Jim Hendler, and me write down our vision for the SW

Early 2000s: DARPA DAML program, proliferation of W3C SW specifications


THE SEMANTIC WEB

A new form of Web content
that is meaningful to computers
will unleash a revolution of new possibilities

by
TIM BERNERS-LEE,
JAMES HENDLER and
ORA LASSILA

PHOTOILLUSTRATIONS BY MIGUEL SALMERON

www.sciam.com

SCIENTIFIC AMERICAN 35

Semantic Web: original 2001 vision

	Old World:	New World:
Content:	Web of documents (implicit/hidden semantics)	Web of data (explicit & accessible semantics)
Actors:	Humans	Humans & Agents
Standards:	Anticipate & standardize everything up front	"Delayed semantic commitment"
KR:	Centralized KR	Decentralized KR

aws

Semantic Web

is not a new vision of the Web

is not even “KR for the Web”

it is KR using Web technologies!

Web technologies:

- well understood, lots of software support, widely deployed
- “networking friendly” (HTTP goes through firewalls, etc.)
- prevailing mindset of distributed systems and sharing

aws

Semantic Web

Represents a different take on standardization

- semantics: specify “how to say it”, not “what to say”

The key aspect of the Semantic Web is serendipity

- solution for use cases yet to be articulated (seriously)
- “delayed semantic commitment”

aws

Semantic Web: original 2001 vision – retrospective view

Good:

Strong, global IDs

Graph merging

Simple schema language

Self-describing data

- embedded or referenced ontology

Not so good (still not fixed):

Services & publishing

- SPARQL is pricey to run; no working business model

SPARQL (success and failure)

Upper tiers of the ”layer cake”

- trust, proofs never happened

No usable composite datatypes

aws

Semantic Web: original 2001 vision – retrospective view

Good:

Strong, global IDs

Graph merging

Simple schema language

Self-describing data

- embedded or referenced ontology

All this is about sharing

aws

What still needs to be fixed…?

Technical

- Simple schema language
- SPARQL
- Composite datatypes

Not technical

- Business model?
- Ready for enterprise?

and most importantly:

- PERCEPTION

aws

Simple schema language? Really?

What about OWL?

- experience shows OWL is difficult for users and poorly understood
- OWA was the right choice, but we also need CWA: enter SHACL

End result: we really have 3 schema languages, and lots of confusion

- (maybe 4 if you count R2RML)

Why is SPARQL both a success and failure?

SPARQL 1.0 was an utter failure

- I told the WG that they should have support for paths.
  Their response: “paths are not a use case for graphs” (really).

SPARQL 1.1 is a lot better

- some support for paths
- federated queries
- update queries

But:

- no path discovery (we have an idea how to fix this)
- developers want nothing to do with SPARQL (our fix takes care of this too)

aws

Composite datatypes

In RDF, composite datatypes are constructed using the graph structure

Instead, we should have provided abstractions and interfaces

Programming with RDF composite datatypes is very hard

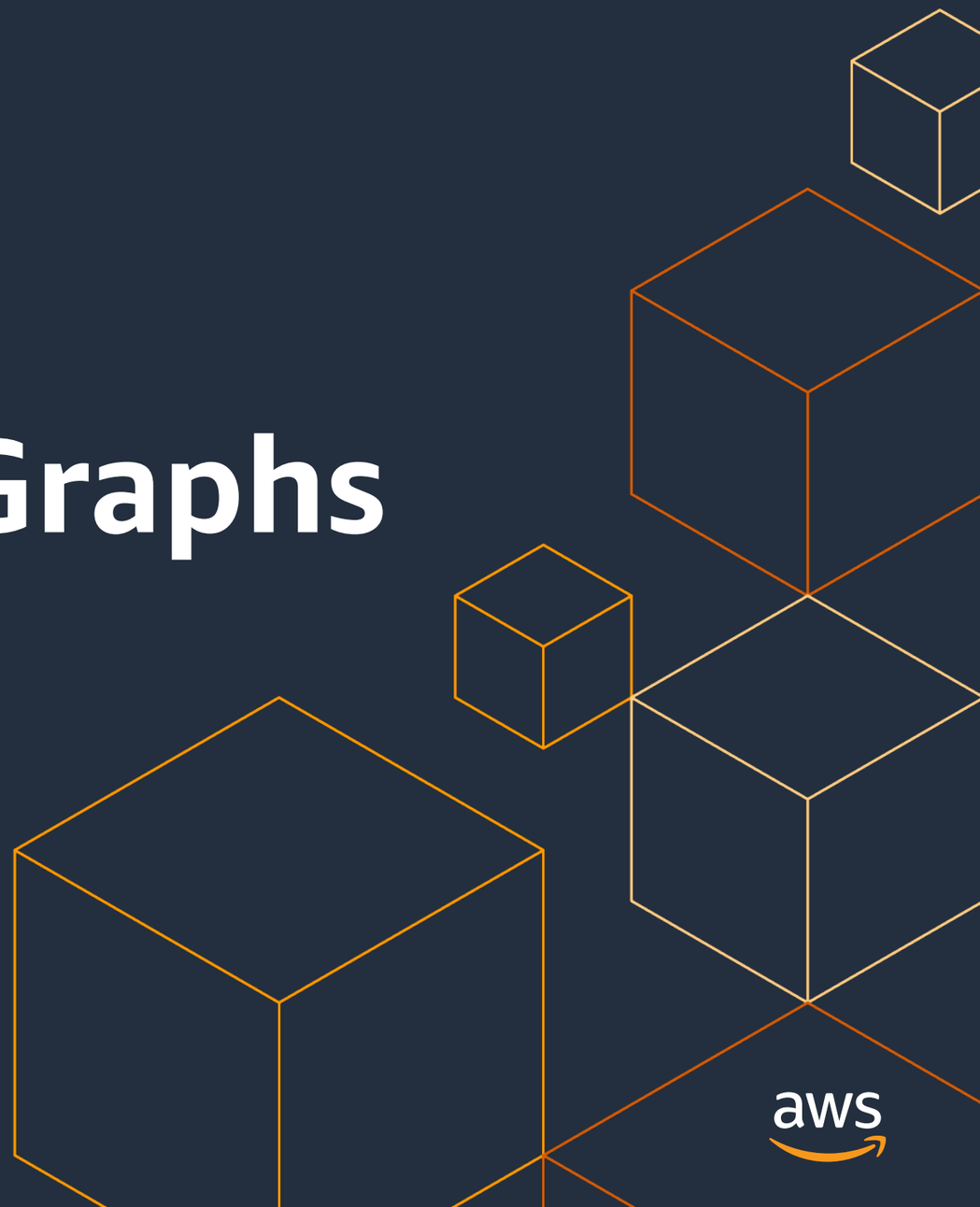
I think we really messed this one up…

(This week, my colleague Olaf Hartig will present our work at ESWC
on how we think RDF composite datatypes could be fixed)

aws

Modern Knowledge Graphs

The “new” Semantic Web



The “new” Semantic Web

We see adoption in all kinds of industries

- graph databases have entered the mainstream
- (new technologies: Labeled Property Graphs)

Many popular use cases:

- data integration, “digital twins”, organizational/collective memory…

Integration of non-symbolic AI techniques

- a broader interpretation of “reasoning”

And of course LLMs + KGs

aws

Modern knowledge graphs

Typically organizational context

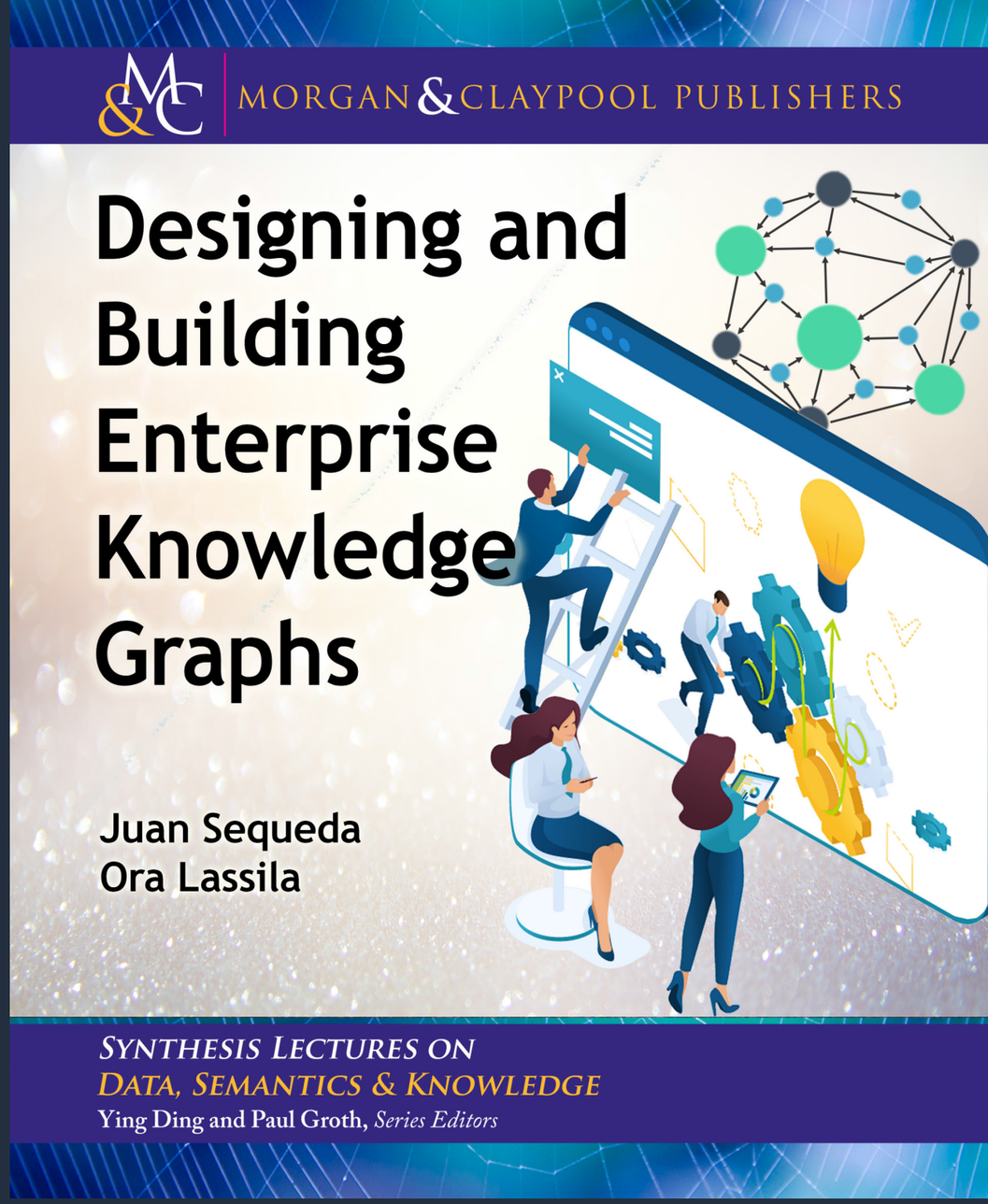
- “enterprise knowledge graphs”

Often introduced to clean the mess that exists with data

- (i.e., the mess created with the old, non-SemWeb technologies…)
- data integration, “democratization of data”
- organizational upper ontologies

Sometimes these are closed systems

- (I think this is usually a mistake…)



aws

The “new” Semantic Web → Knowledge Graphs

	Old:	New:
Scope:	WWW	Enterprise
Actors:	Humans & Agents	Humans & Cloud Services
Sources:	WWW (?)	Enterprise data
AI:	Symbolic	Symbolic & non-symbolic

aws

Success?

YES!

The word “semantics” is now commonplace

…and everybody wants a “semantic layer”

According to **Gartner’s Top 10 Trends in Data**:

Trend 6: Add Semantic Data Integration & Knowledge Graphs

“W3C standards such as RDF, OWL, SPARQL enable your data to speak a common universal language.”

“A knowledge graph acquires and integrates data into an ontology (or many) and then makes it available to applications.”

I think of this as finally having reached the mainstream.

aws

What’s next?

The future of the Semantic Web

The “new” Semantic Web → Knowledge Graphs

	Old:	New:	Future?
Scope:	WWW	Enterprise	
Actors:	Humans & Agents	Humans & Cloud Services	
Sources:	WWW (?)	Enterprise data	
AI:	Symbolic	Symbolic & non-symbolic	

aws

The “new” Semantic Web → Knowledge Graphs + GenAI

	Old:	New:	Future:
Scope:	WWW	Enterprise	Enterprise & more
Actors:	Humans & Agents	Humans & Cloud Services	Humans & Cloud Services & Agents
Sources:	WWW (?)	Enterprise data	Enterprise data & more
AI:	Symbolic	Symbolic & non-symbolic	Symbolic & non-symbolic

aws

Modern knowledge graphs + LLMs

I am still highly skeptical of GenAI, and I don't think of it as "magic"

It is incredibly important that we have the right expectations about GenAI and that we understand how it can be used with KGs

1. Knowledge graphs can support GenAI applications
   - by focusing LLMs to relevant material: "GraphRAG"
2. GenAI can support knowledge graph applications
   - where LLMs generate graph queries, but KGs are the source of answers
   - (cf. dual-process models of cognition, "thinking fast & slow", etc.)

Intelligent agents!?

aws

Modern knowledge graphs and standardization

W3C still working on the Semantic Web stack

- RDF-star (aka “RDF 1.2”): better support for reification and “edge properties”

New kid in town: Labeled Property Graphs (LPGs)

- originally, no standards…
- now standards are starting to emerge (e.g., GQL)
- LPGs come from the database tradition, unlike the Semantic Web

What is the relationship between the two?

aws

Labeled Property Graphs

LPGs do not have this:

Strong, global IDs

Graph merging

Simple schema language

Self-describing data

- embedded or referenced ontology

Note that this is exactly what is so good about the Semantic Web technologies and RDF

aws

A graph is a graph is a graph…?

For knowledge graphs, you typically need what the Semantic Web technologies offer

Other graph applications often treat the graph as a very large, potentially complex data structure

Graph as a logical representation vs. graph as a data structure

RDF & friends

LPGs

aws

A graph is a graph is a graph…?

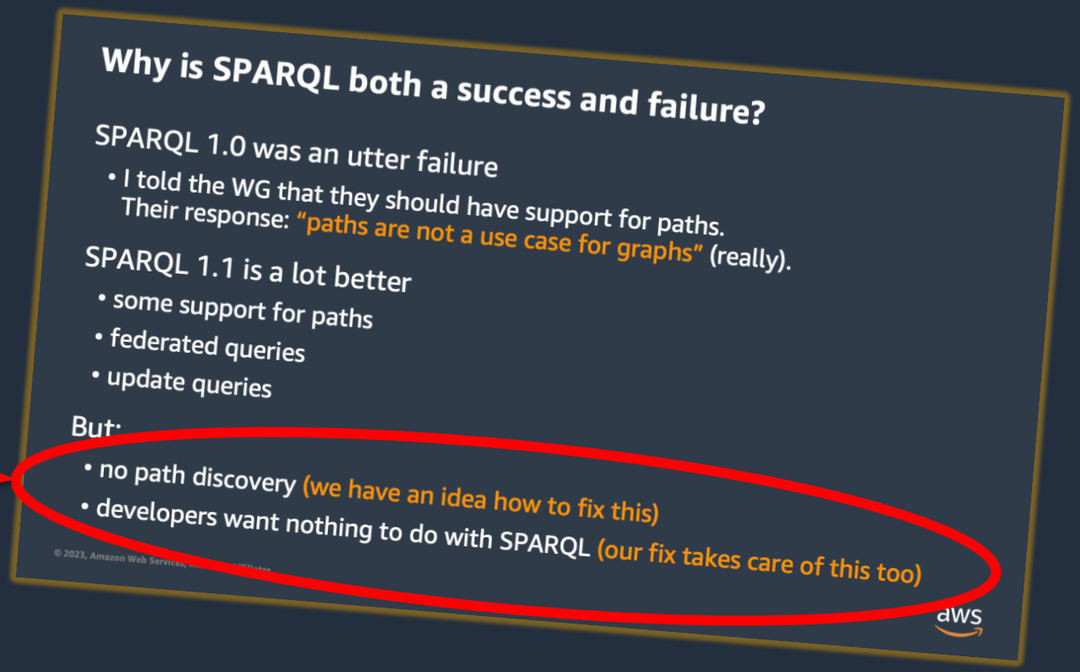
Amazon Neptune currently supports both RDF and LPGs

- either or, not simultaneously, forcing customers to choose

The Neptune team is working to mitigate the rift: Project OneGraph will offer “graph interoperability”

- cross-use of query languages
- single graph representation

OneGraph will offer help with existing SPARQL problems



aws

Summary

Original Semantic Web vision: some things worked, others not so much

Modern knowledge graphs are the new Semantic Web

New technologies, great expectations, much confusion

aws

Thank you!

Contact:

- ora@amazon.com
- Twitter: @oralassila

aws