

Adopting Linked Open Data in Product Security

A Modular Knowledge Graph

David Sastre Medina

Principal Product Security Engineer

What we'll discuss today

- ▶ AI relevance today
- ▶ LLMs just a piece of the puzzle
- ▶ KGs, LLMs and the importance of RAG
- ▶ Proudly Found Elsewhere
- ▶ How to get from A to B
- ▶ As a ProdSec engineer, what's in it for me?





“Passionate Red Hatters like you, coupled with our unique brand and **open culture**, will help Red Hat continue **leading the way in AI** and beyond”, Kicking off We Are Red Hat Week, memo-list@

Matt Hicks
President & CEO, Red Hat



Our Mission, Our Vision



To be the **catalyst** in communities of customers, contributors, and partners creating **better technology** the **open source** way.

Red Hat's Mission Statement

<https://www.redhat.com/en/book-of-red-hat>



We believe that everyone, everywhere, is entitled to **quality information** needed to mitigate security and privacy risk as well as the access to do so. We strive to protect communities of customers, contributors and partners from digital security threats. We believe **open source principles** are the best way to achieve this.

Product Security Team Vision



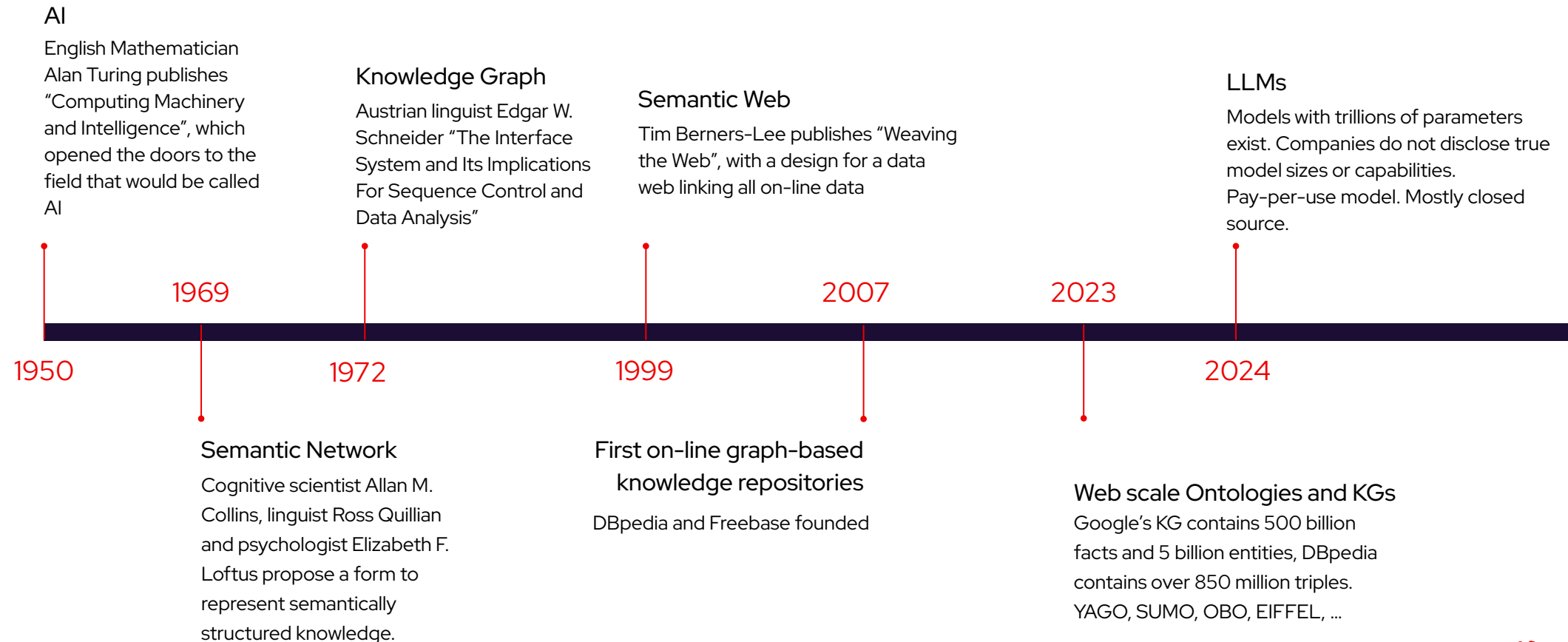
To **lead the forefront of secure software development**, ensuring robust protection and integrity throughout Red Hat's software development lifecycle. Safeguarding our software ecosystem, fostering a culture of security excellence, and fortifying Red Hat's reputation as a trusted industry leader.

PSRD Team Vision

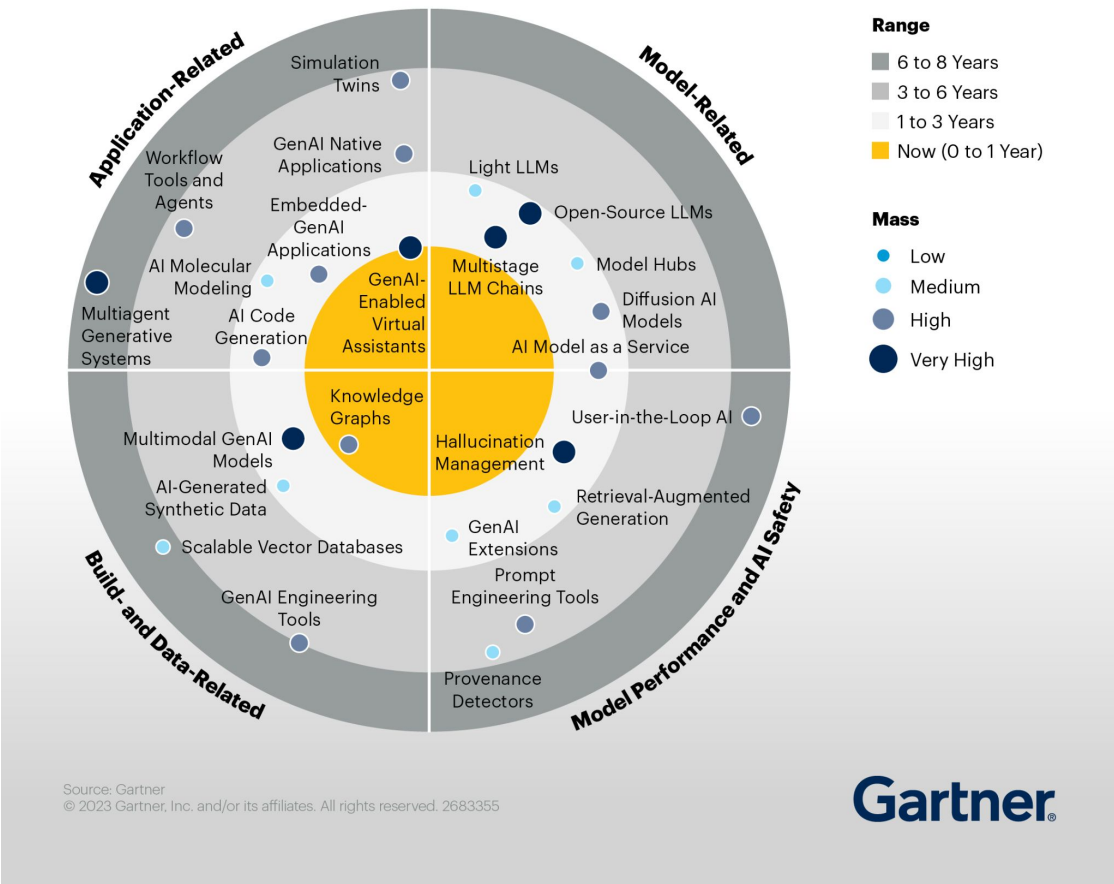


Bleeding Edge Technology

Or is it?



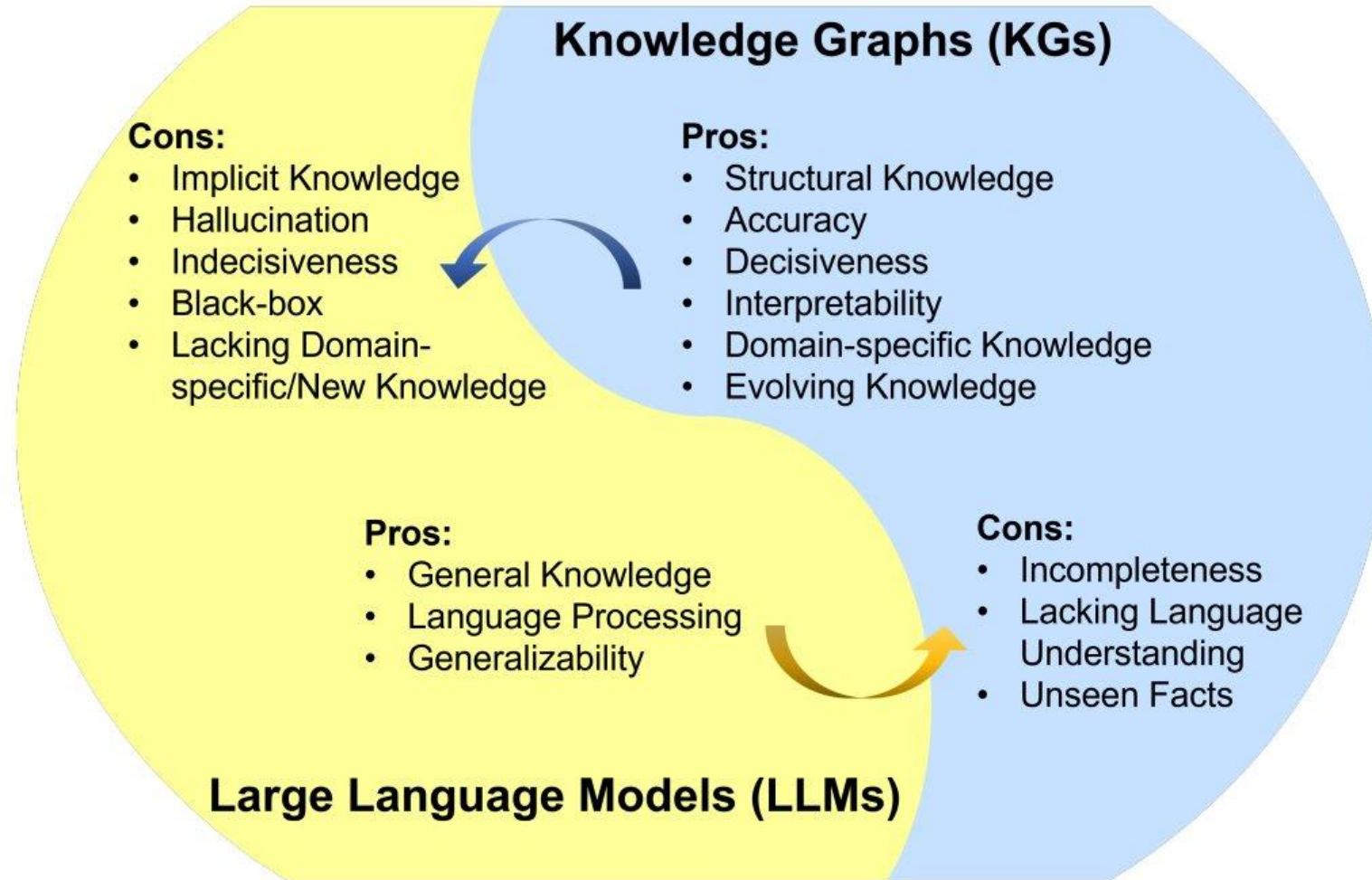
Impact Radar for Generative AI



Model build and data-related

Critical steps and decisions in building and advancing a GenAI model. KGs are machine-readable data structures that represent knowledge of the physical and digital worlds, including entities and their relationships, which adhere to a graph data model.





KGs and LLMs cooperation

LLMs are black-box models, which often fall short of capturing and accessing factual knowledge. In contrast, KGs are structured knowledge models that explicitly store rich factual knowledge. KGs can enhance LLMs by providing external knowledge for inference and interpretability.

Fig. 1. Summarization of the pros and cons for LLMs and KGs. LLM



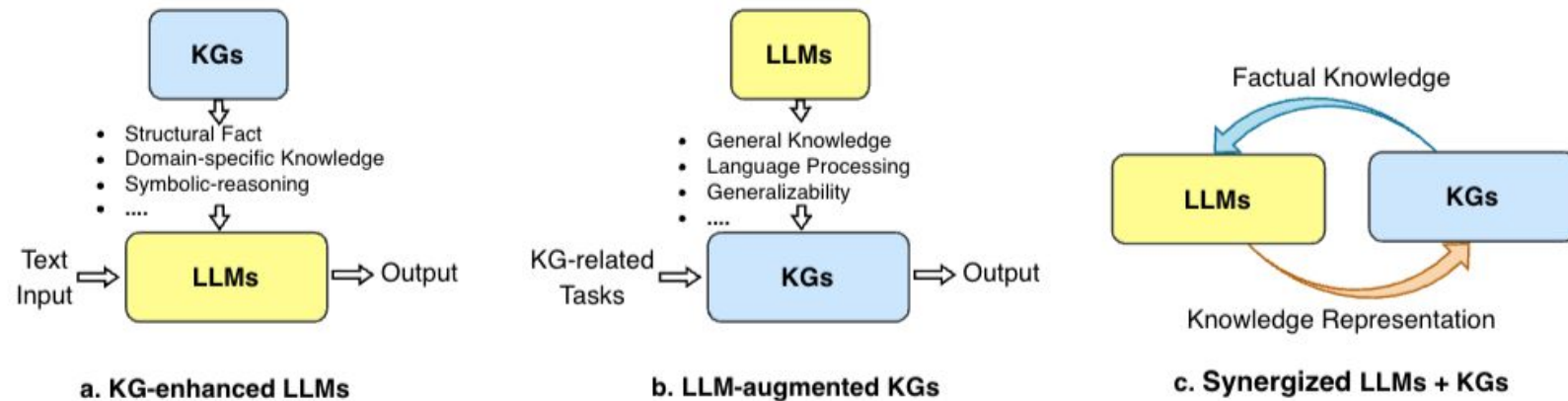


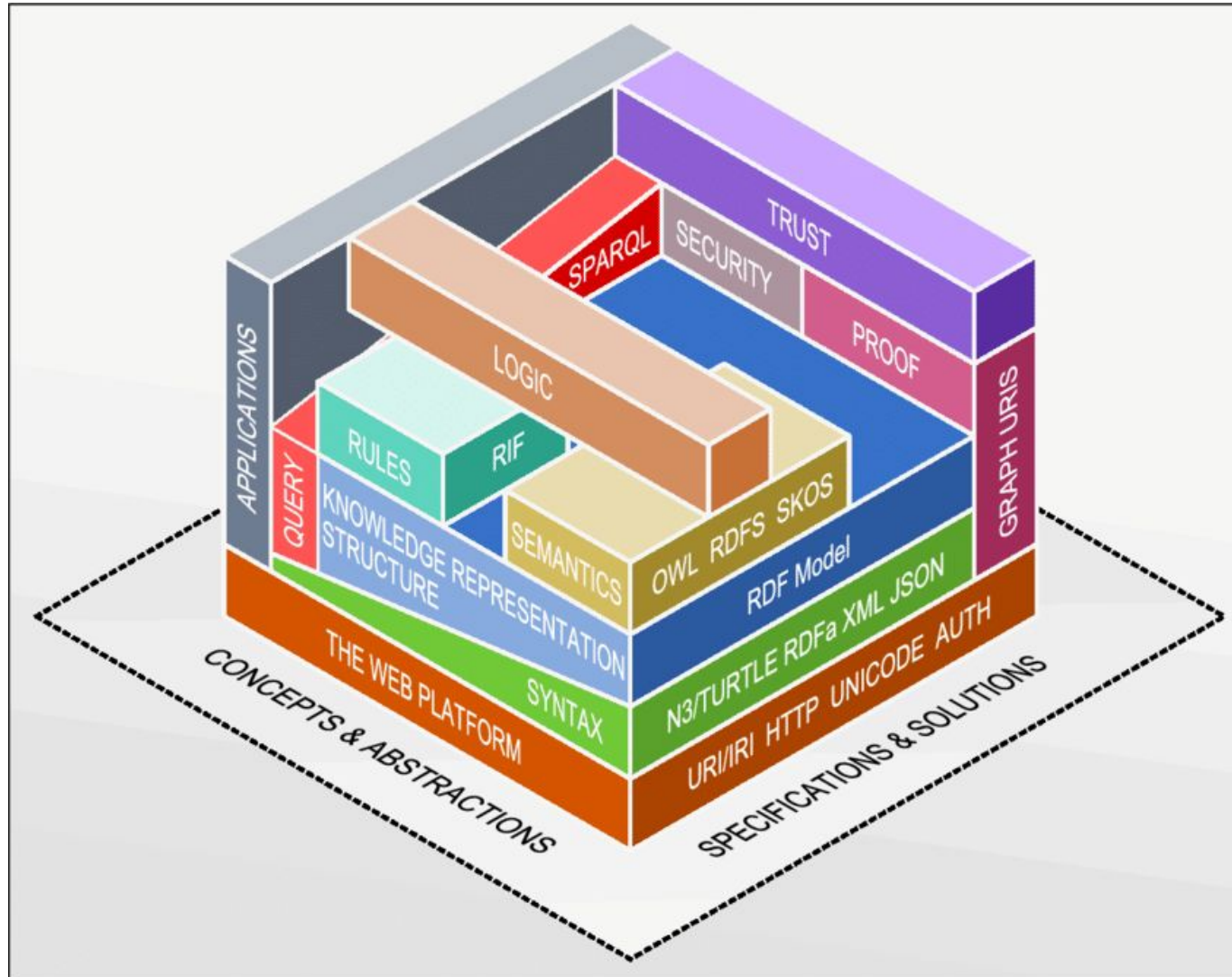
Fig. 6. The general roadmap of unifying KGs and LLMs. (a.) KG-enhanced LLMs. (b.) LLM-augmented KGs. (c.) Synergized LLMs + KGs.

What Is Retrieval-Augmented Generation?

RAG is a technique for enhancing the accuracy and reliability of generative AI models with facts fetched from external sources.

GraphRAG uses KGs as a source of context and factual information.





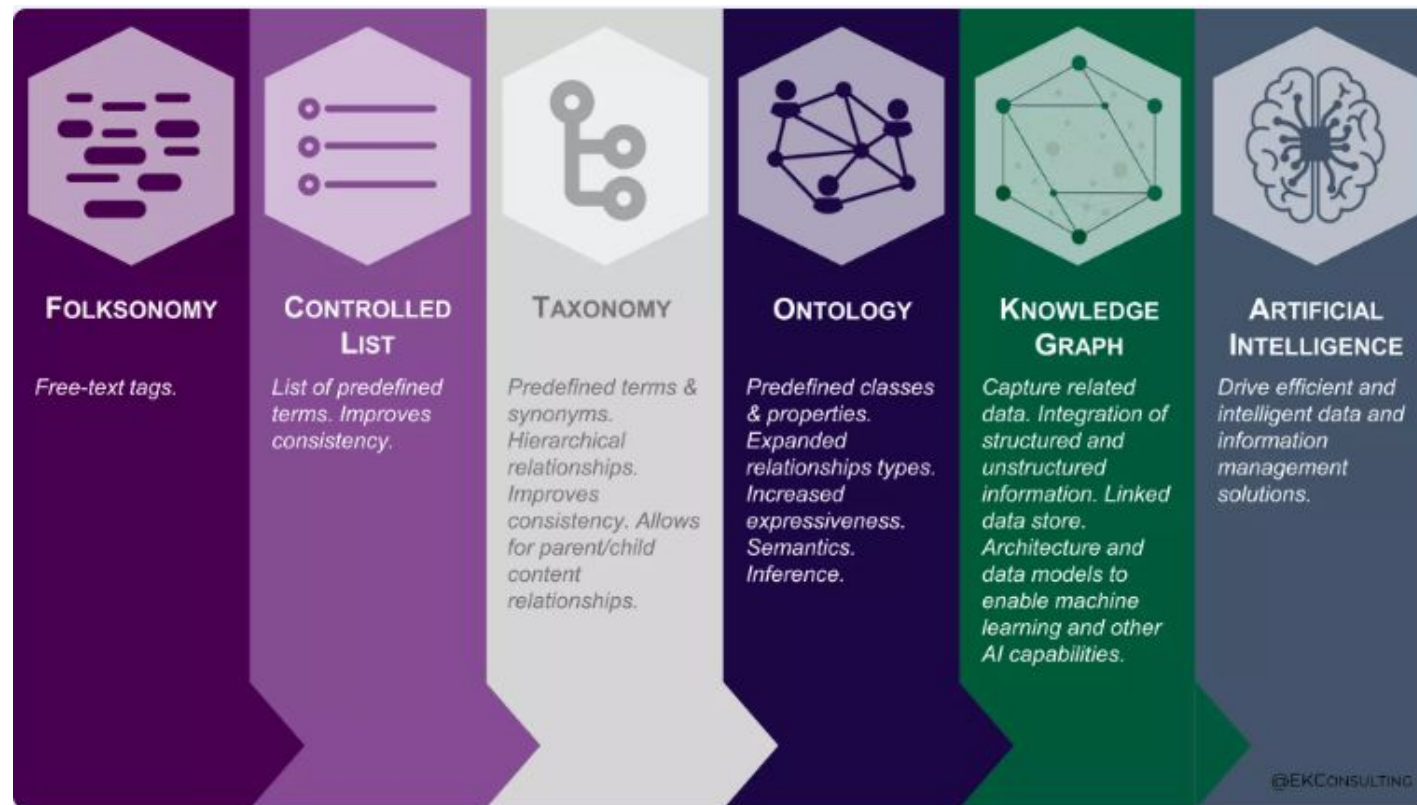
Innovation model

Keep an attitude of
resistance to innovations
"Not Invented Here" and
enthusiasm for those
"Proudly Found Elsewhere."

Source:

<https://web.archive.org/web/20220628120341/http://bnode.org/blog/2009/07/08/the-semantic-web-not-a-piece-of-cake>
<https://web.archive.org/web/20220126000041/http://www.cs.rpi.edu/~hendler/presentations/LayercakeDagstuhl-share.pdf>
<https://hbswk.hbs.edu/archive/pg-s-new-innovation-model>
<https://decentralized-id.com/web-standards/w3c/verifiable-credentials/data-integrity-Id-proofs/>
<https://lists.w3.org/Archives/Public/public-credentials/2021May/att-0082/2021-Linked-Data-Security-WG-Charter.pdf>





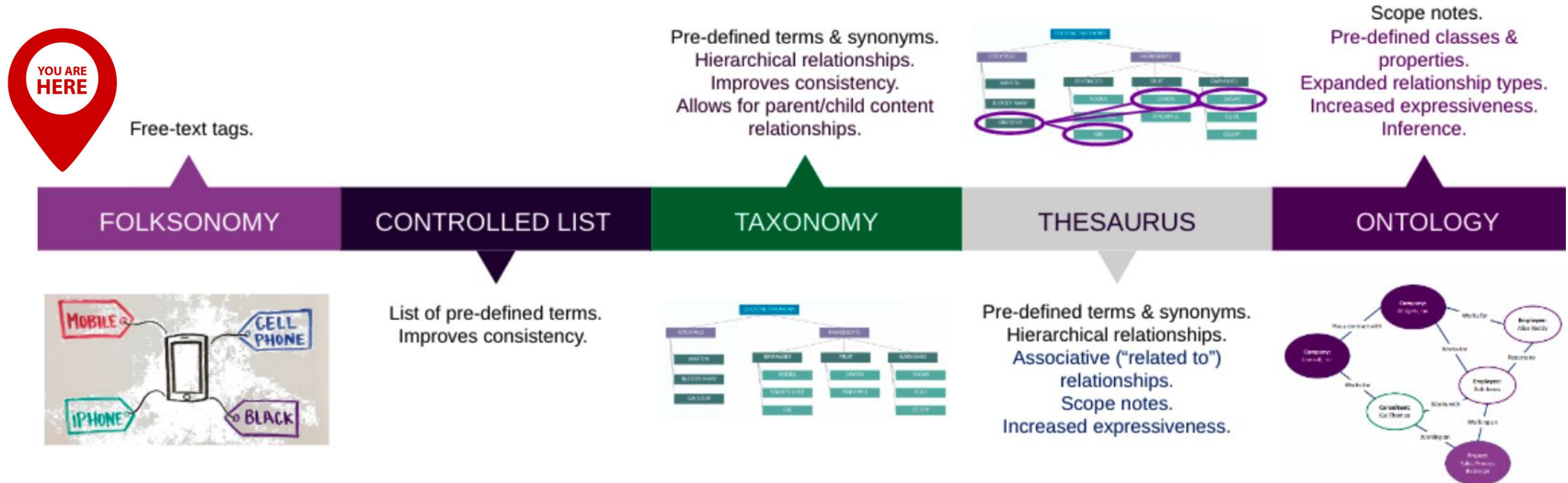
Semantic Spectrum

The semantic or ontology spectrum, smart data continuum, or semantic precision, is a series of increasingly precise and semantically expressive definitions for data elements in knowledge representations, especially for machine use.



How to get from A to B

...and why



Current status of knowledge management in our industry

Duplicated efforts, non-normative, non-searchable, no semantic relationships, no disambiguation, not comprehensive enough, not machine-readable

- ▶ <https://www.ncsc.gov.uk/section/advice-guidance/glossary>
- ▶ <https://cwe.mitre.org/documents/glossary/>
- ▶ <https://csrc.nist.gov/glossary/>
- ▶ <https://github.com/mdn/content/tree/main/files/en-us/glossary>

“Get CSV/JSON” at best. Some may be real vocabularies/taxonomies, but the source is hidden

- ▶ <https://www.ukcybersecuritycouncil.org.uk/glossary/>
- ▶ <https://www.globalknowledge.com/us-en/topics/cybersecurity/glossary-of-terms/>
- ▶ <https://www.sans.org/security-resources/glossary-of-terms/>
- ▶ <https://niccs.cisa.gov/cybersecurity-career-resources/vocabulary>



Common Platform Enumeration	CPE	Red Hat identifier assigned to a particular product or product version. A product's CPE identifier is publicly used and can be found in numerous places to identify content.
CPE		Common Platform Enumeration Community Platform Engineering Customer Premises Equipment or Customer Provided Equipment CEE Partner Engagement

Common Vulnerabilities and Exposures (CVE)

A list of publicly disclosed computer security flaws. When someone refers to a CVE, they

Common Vulnerability Scoring System (CVSS)

An industry standard for assessing the severity of security vulnerabilities.

Common Weakness Enumeration (CWE)

A community-developed list of common weaknesses in software and hardware that they are part of the vuln



common platform enumeration (CPE)



Abbreviations / Acronyms / Synonyms:


[CPE](#) [show sources](#)

Definitions:

 A nomenclature and dictionary of hardware, operating systems, and applications.

Sources:

[CNSSI 4009-2015](#) from [NIST SP 800-126 Rev. 2](#)

 A SCAP specification that provides a standard naming convention for operating systems, hardware, and applications for the purpose of providing consistent, easily parsed names that can be shared by multiple parties and solutions to refer to the same specific platform type.

Sources:

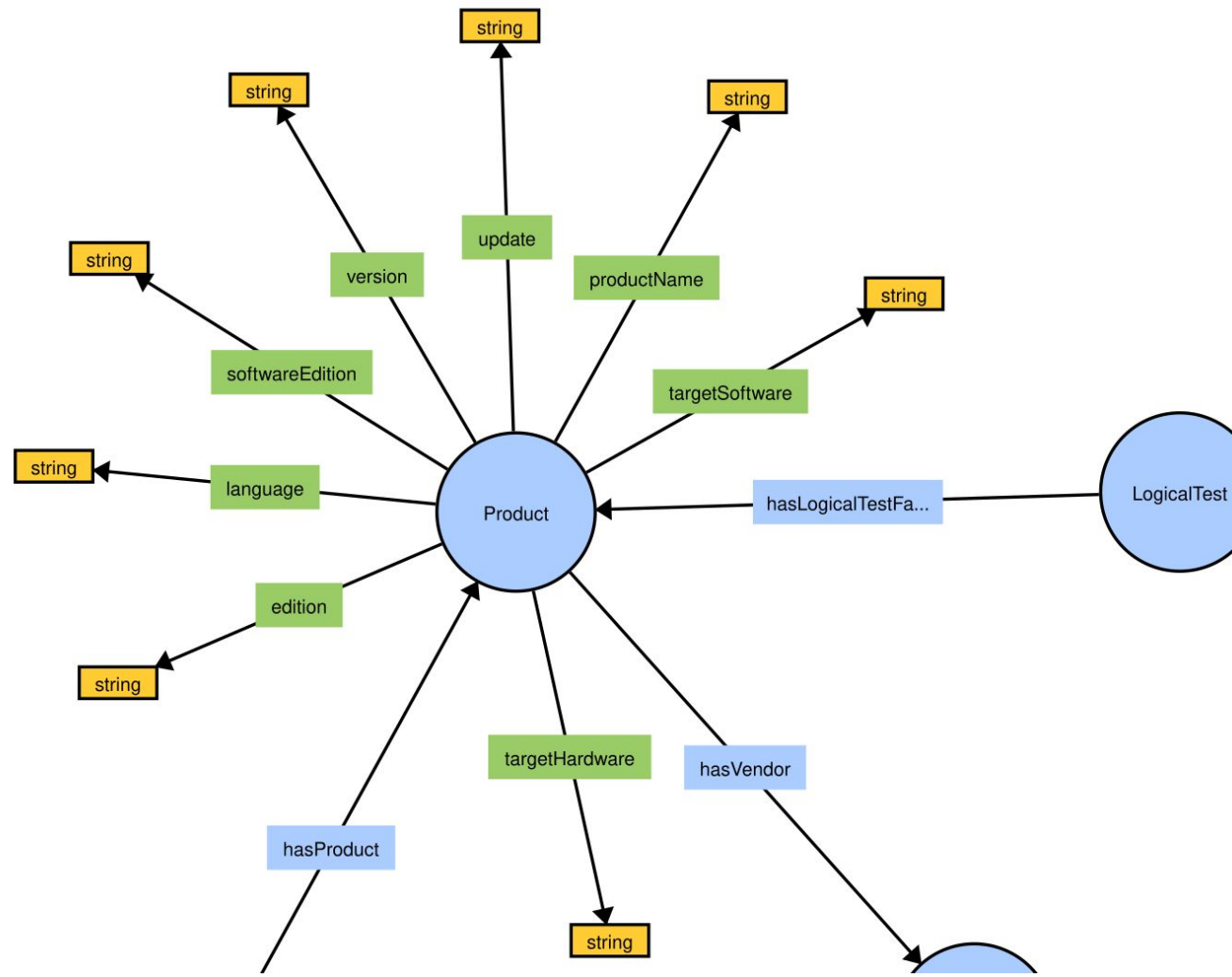
[NIST SP 800-128](#) under Common Platform Enumeration (CPE)

[NIST SP 800-128](#)

Why a normative taxonomy?

- To collect, disseminate and manage terminology
- To promote uniformity in term format and assignment
- To provide a way to organize knowledge for subsequent search and retrieval
- To meet the needs of machine interoperability

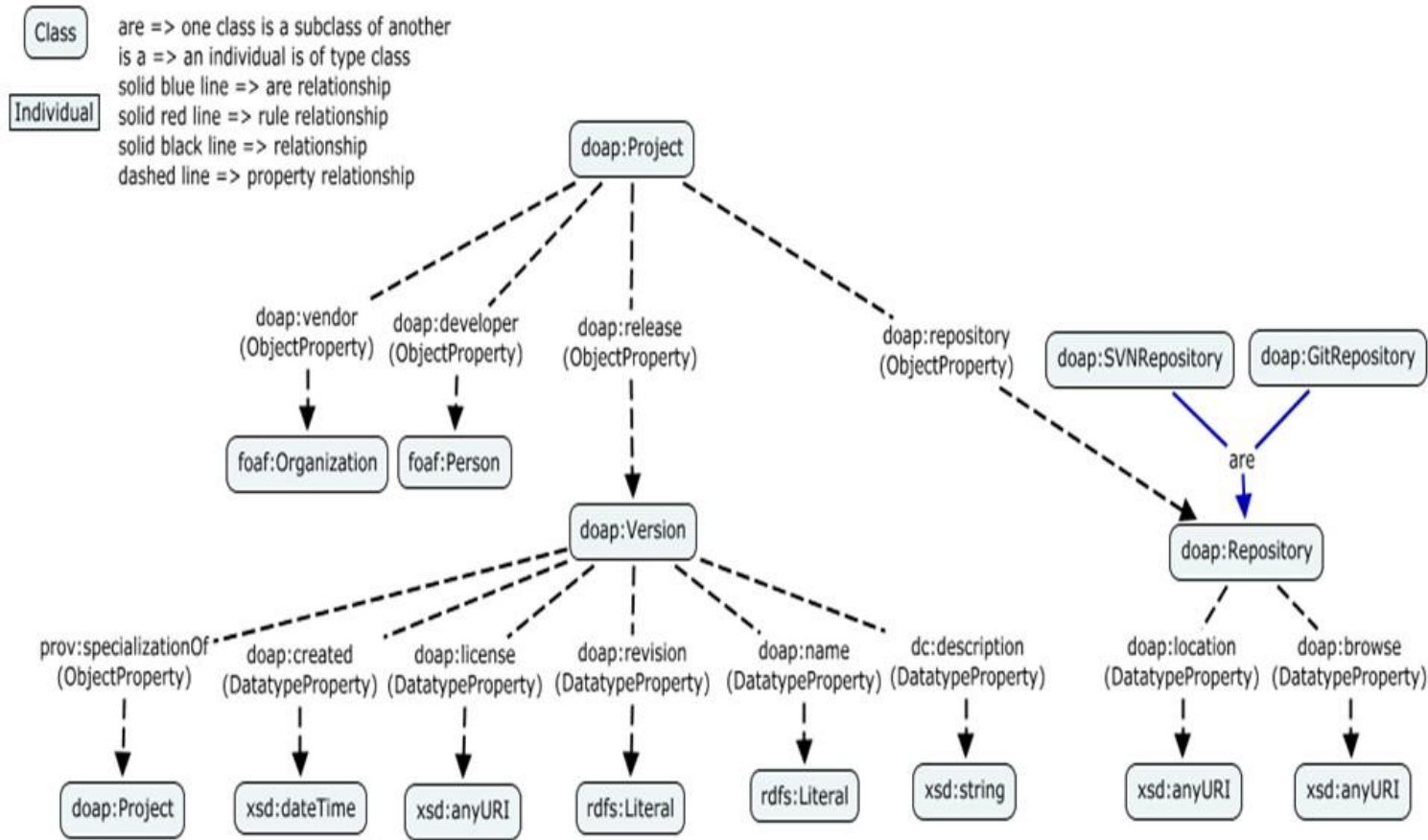




Ontology modularization

A methodological principle in ontology engineering. The idea is that an ontology is built in a modular manner, i.e. developed as a set of small modules and later composed to form, and be used as, one modular ontology.

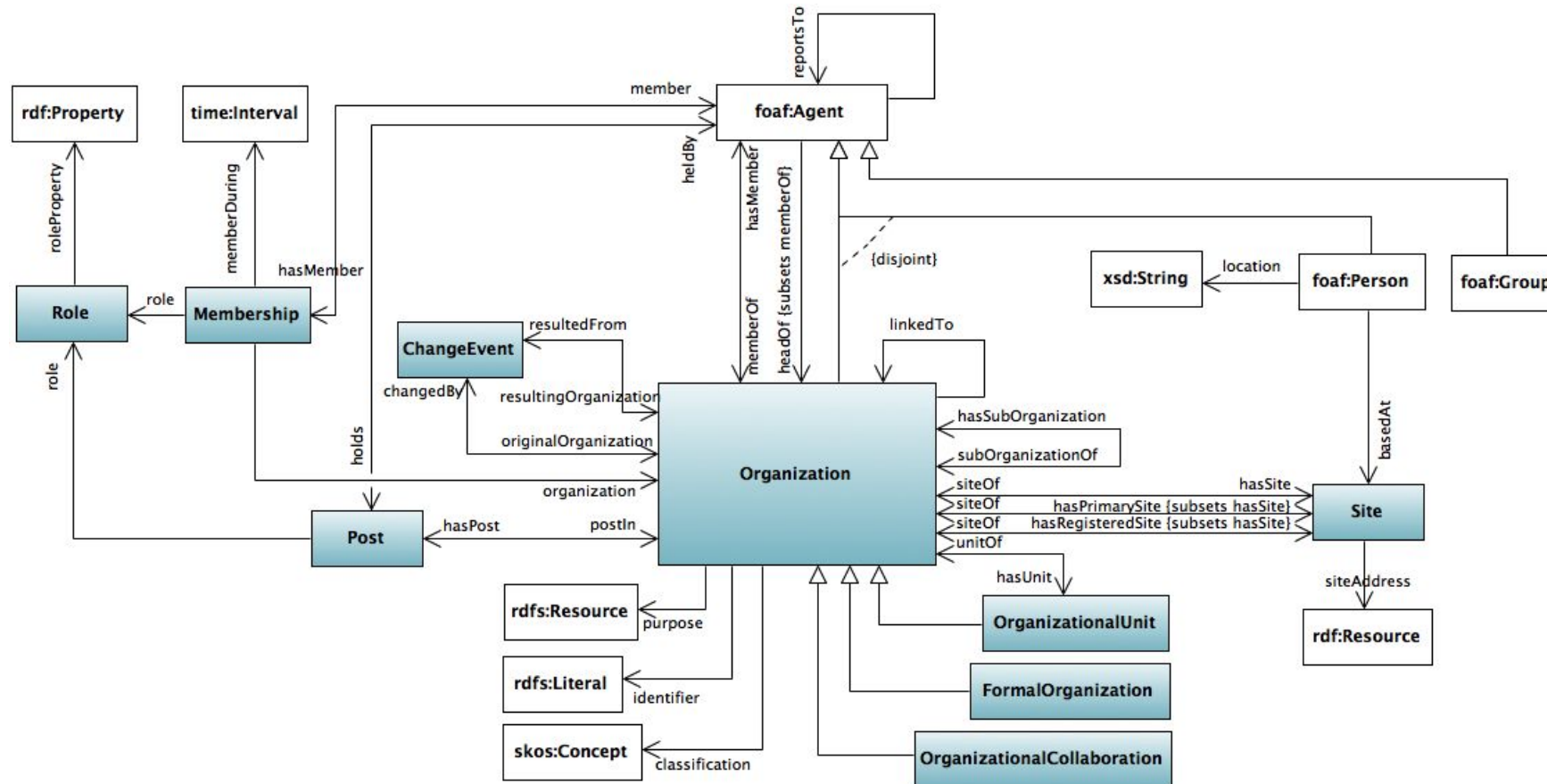




The DOAP Ontology

DOAP (Description of a Project) is an RDF Schema and XML vocabulary to describe software projects, in particular free and open source software.

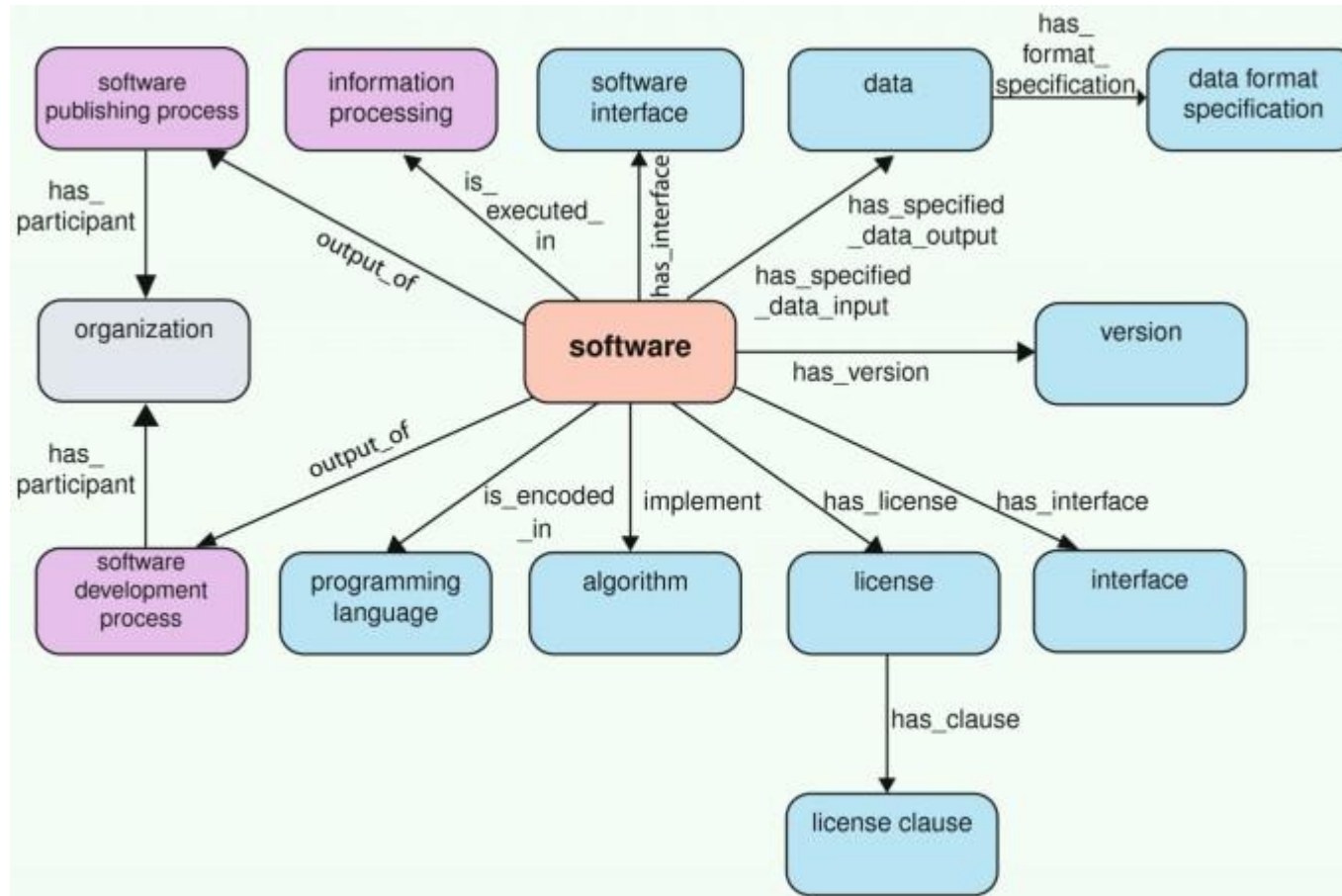




The Organization Ontology

Core ontology for organizational structures, aimed at supporting linked data publishing of organizational information across a number of domains. It is designed to allow domain-specific extensions to add classification of organizations and roles, as well as extensions to support neighbouring information such as organizational activities.



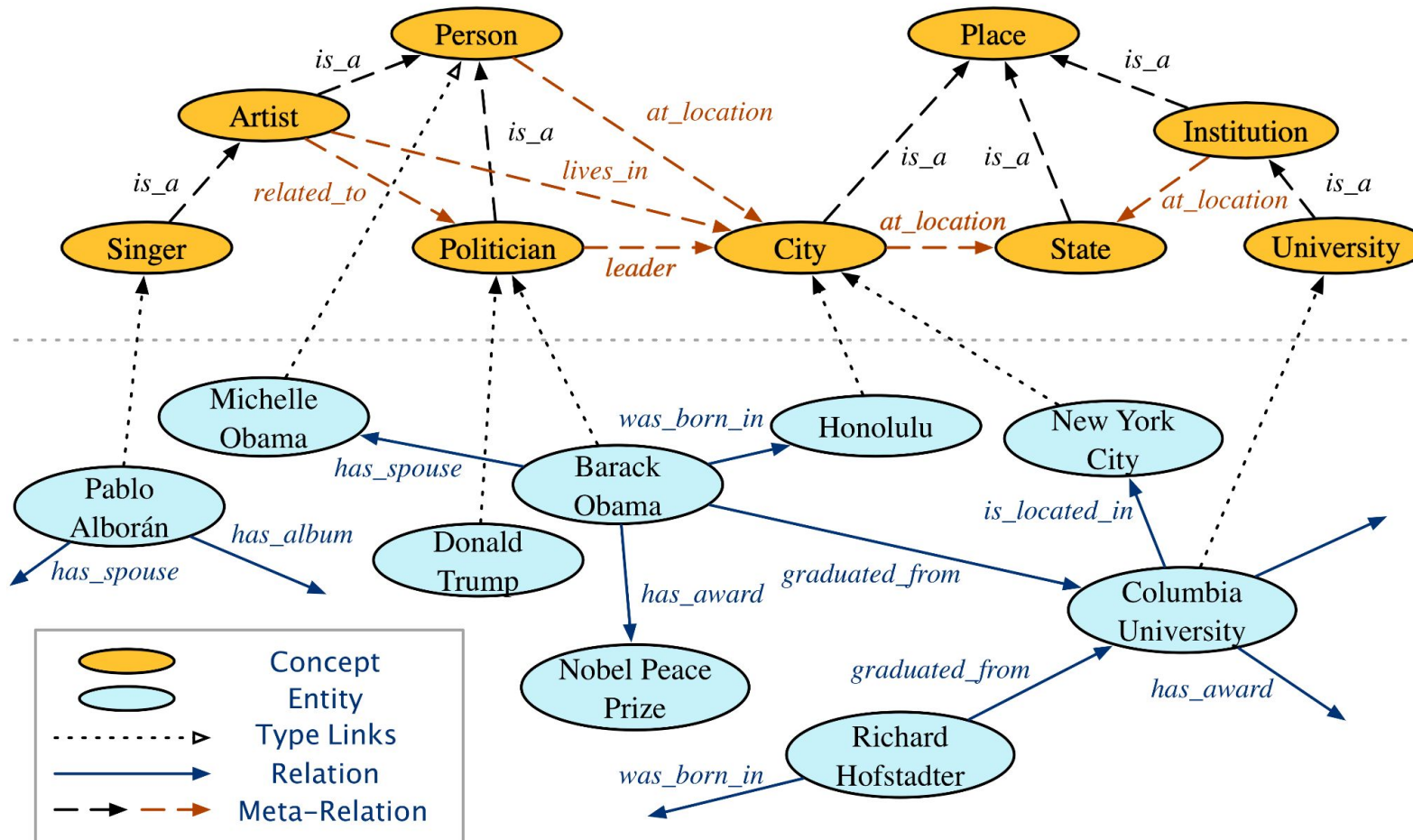


The Software Ontology

A description of software used to store, manage and analyze data. Recently, the SWO has incorporated EDAM, a vocabulary for describing data and related concepts in bioinformatics.



Ontology-view Knowledge Graph

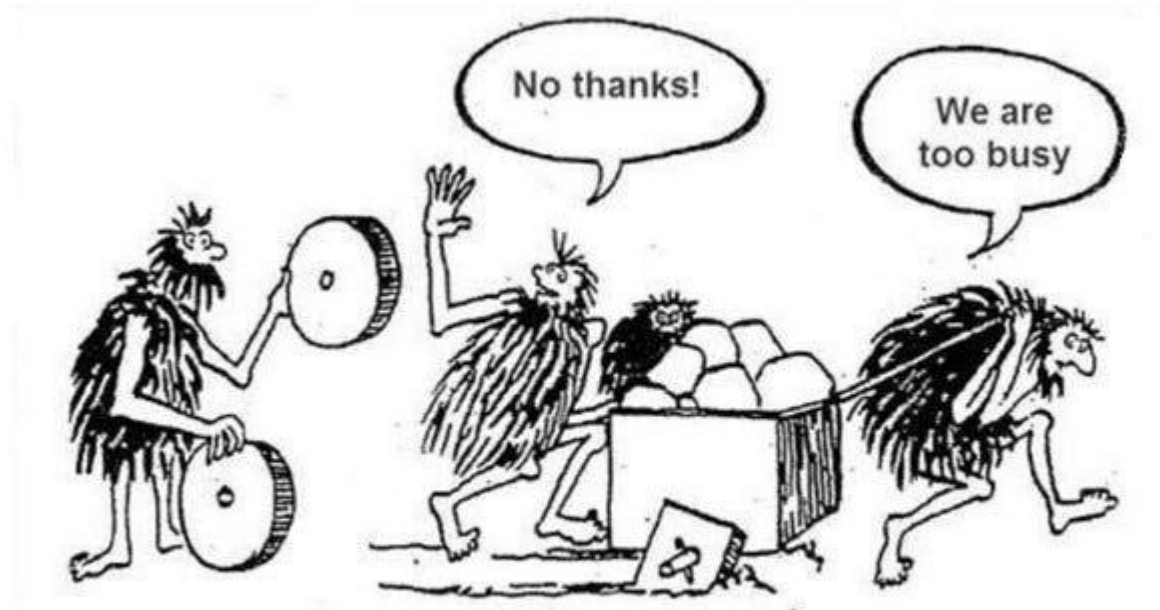


Instance-view Knowledge Graph

KG Embeddings

Knowledge bases simultaneously represent two views of KGs: an ontology view for abstract and commonsense concepts, and an instance view for specific entities that are instantiated from ontological concepts. KEGs enable entity classification, link prediction to measure the plausibility of a triple (e.g. 'Columbia University', 'is located in', 'NYC').





Red Hat's Mission

To be the **catalyst** in communities of customers, contributors, and partners creating **better technology** the **open source** way.



Thank you

Red Hat is the world's leading provider of enterprise open source software solutions. Award-winning support, training, and consulting services make Red Hat a trusted adviser to the Fortune 500.



linkedin.com/company/red-hat



youtube.com/user/RedHatVideos



facebook.com/redhatinc



x.com/RedHat