Knowledge Graphs & Semantic Web

Foundations for Law and Finance

RDF, RDFS/OWL, SKOS, SPARQL, SHACL, JSON-LD

Michael J Bommarito II · Jillian Bommarito · Daniel Martin Katz

December 21, 2025

Working Draft Chapter

Version 0.1

This chapter introduces core Semantic Web technologies that support durable identifiers, typed relations, validation, and query. We align examples to legal and financial data and connect to the book's Evidence Record (Chapter 3, Structured Outputs).

Contents

0.1 Linked Data Principles

Linked Data rests on four simple practices: use IRIs as names for things, use HTTP IRIs so people and machines can look them up, provide useful information (RDF) when IRIs are dereferenced, and include links to other IRIs to enable discovery. (**tbl-linked-data**)

Practical Guidance for Legal/Financial Data

- Prefer persistent HTTP IRIs for entities (courts, statutes, issuers, instruments).
- Publish minimal RDF descriptions and link to authoritative registries.
- Reuse standard vocabularies; only mint new terms when necessary.

0.2 RDF Data Model and RDFS/OWL

RDF 1.1 defines a graph data model of triples (subject, predicate, object), IRIs, literals with datatypes, and blank nodes. It provides a foundation for interoperable data across systems. (**w3c-rdf11-concepts**)

RDFS supplies basic vocabulary for classes and properties. OWL 2 adds richer semantics for classes, properties, and individuals; its profiles (EL, QL, RL) target tractable reasoning in different scenarios. (**w3c-rdf-schema**; **w3c-owl2-profiles**)

Choosing an OWL 2 Profile

Use **EL** for large taxonomies/hierarchies (e.g., healthcare codes), **QL** for efficient querying over relational backends, and **RL** for rule-engine implementations. Profiles trade expressivity for predictable performance.

Note on RDF 1.2.. Quoted triples and related features are tracked in ongoing RDF 1.2 work; adopt when specifications reach Recommendation status.(**w3c-rdf12-status**)

0.3 Vocabulary Design Basics

Model only what you need. Start with existing terms (e.g., FOAF, PROV, schema.org modules when appropriate) and mint new IRIs in a controlled namespace. Provide labels, definitions, examples, and version IRIs.

0.4 SKOS for Taxonomies and Thesauri

SKOS models controlled vocabularies with concepts, preferred/alternate labels, and hierarchical and mapping relations. It is ideal for legal/financial taxonomies, harmonizing synonyms, acronyms, and crosswalks. (**w3c-skos-reference**)

When to Use SKOS vs. OWL

- **SKOS**: curated terms and relationships among concepts; lightweight semantics.
- **OWL**: domain ontology with class axioms, property characteristics, and reasoning.

0.5 SPARQL Query, Update, and Federation

SPARQL 1.1 standardizes querying RDF graphs, updating graphs, a uniform protocol, and federated querying across endpoints. It is the primary interface to KGs. (**w3c-sparql11-overview**)

Query Essentials

- SELECT/CONSTRUCT/ASK/DESCRIBE forms cover common needs.
- Use IRIs for precise joins; lean on datatypes (`xsd:date`, `xsd:decimal`).
- Prefer parameterized queries and fixed graph names for auditability.

SPARQL 1.2.. Track ongoing updates (e.g., quoted triple support alignment) and adopt once Recommendations are finalized. (**w3c-sparql12-status**)

0.6 Validation with SHACL

SHACL defines shapes—constraints over nodes and properties—that validate graph contents (presence, cardinality, datatype, regex, closed shapes). It bridges your Evidence Record fields and operational data quality. (**w3c-shacl**)

Use in Legal/Financial Graphs

Require dates with timezones, jurisdiction codes, and stable locators for citations. Run shapes on ingestion and nightly to detect drift.

0.7 JSON-LD for Web Interop

JSON-LD 1.1 enables publishing and consuming RDF via familiar JSON, using contexts to map keys to IRIs and optional framing for client-friendly shapes. It is an on-ramp for applications and APIs. (**w3c-jsonld11**)

Publishing Tips

- Serve machine-readable JSON-LD alongside human pages.
- Version contexts and document them; avoid breaking changes.
- Reuse identifiers from authoritative registries (LEI, CIK, ELI/ECLI).

0.8 Domain Identifiers and Vocabularies

Knowledge graphs depend on durable identifiers and shared vocabularies.

- **Entities:** LEI (ISO 17442) for legal entities (GLEIF), SEC CIK for US issuers/filers, ELI/ECLI for EU legislation and case law. (**gleif-lei**; **sec-cik**; **ecli-ejustice**)
- **Financial vocabularies:** FIBO as an OWL reference model; XBRL taxonomies (IFRS 2024, US GAAP 2025) for reporting concepts. (**fibo-spec**; **ifrs-taxonomy-2024**; **fasb-usgaap-2025**)
- **Provenance:** PROV-O for process lineage; align with the Evidence Record (Chapter 3, Structured Outputs). (**w3c-provo**)

0.9 Publishing and Cataloging

Follow Data on the Web Best Practices for discoverability, provenance, quality, and licensing. Use DCAT to catalog datasets and endpoints. (**w3c-dwbp**)

Checklist

- License and access terms; contact points and update cadence.
- Machine-readable metadata (DCAT), JSON-LD contexts, and example queries.
- Versioning policy and deprecation notices.