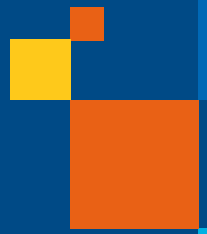
FOSDEM 2025

A Novel Ontology for Enhanced SBOM Data Modeling with TOSCA

Alexios Zavras



intel®

SPDX concepts

- Software:
  - Package
  - File
  - Snippet

- Package definition (in both SPDXv2 and SPDXv3)

  any unit of content that can be associated with a distribution of software

Problem

any unit of content that can be associated with a *distribution* of software

- Sometimes we want to refer to something more abstract
  - “OpenSSL”
  - “zlib”
  - “log4j”
  - “llama”

- Introducing: **Component**

TOSCA

The Open Source Component Aggregator

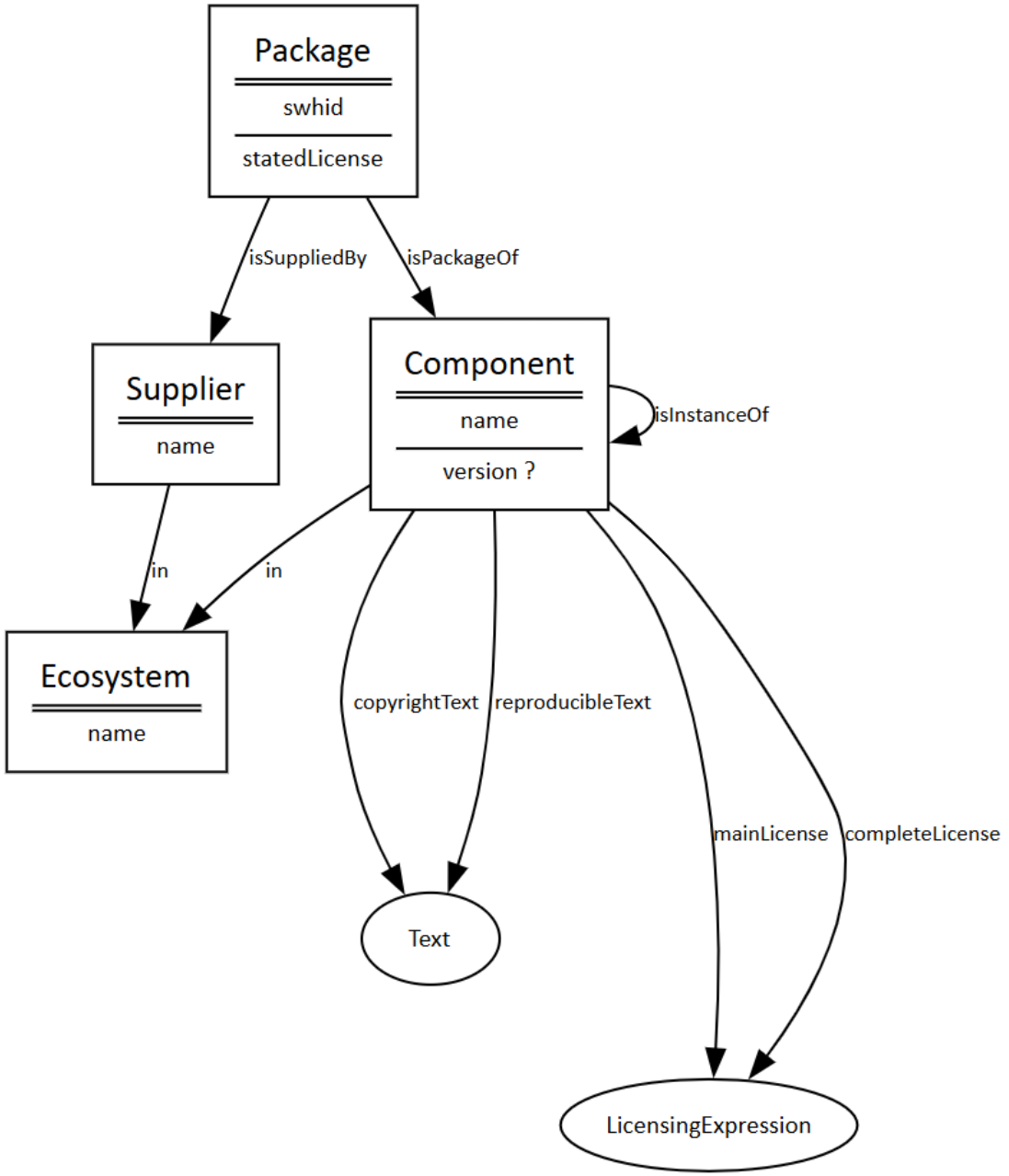
Software Component

- An abstract representation of a specific software component
- Might refer to a specific version or not
- *Package* is a *Component* with a specific *Version* that is provided by a *Supplier*

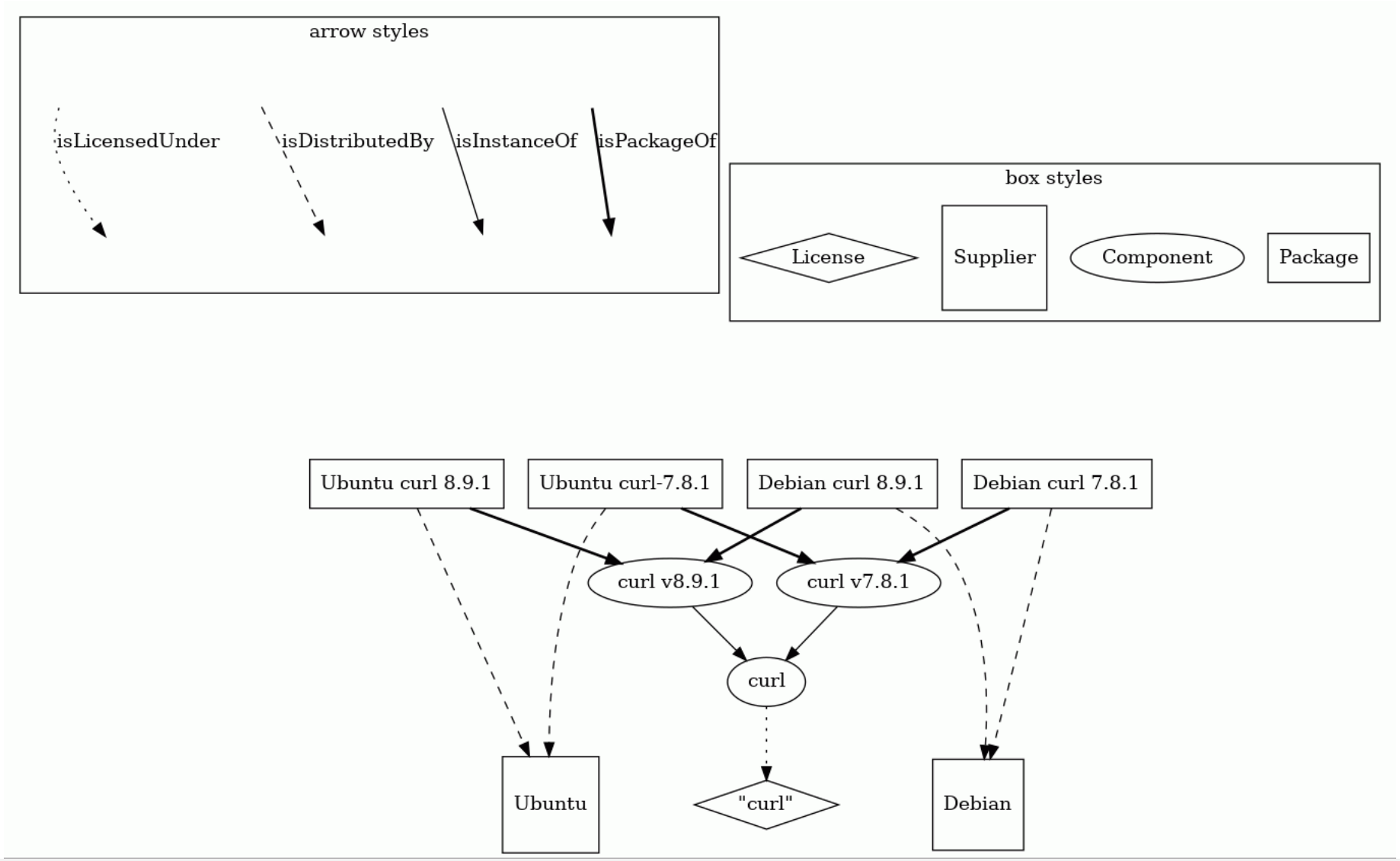
Supplier

- The one who provides the software
- Ecosystems:
  - Deb: Debian, Ubuntu
  - Python: PyPi
  - Java: MavenCentral
  - Npm: Node
  - Rust: Crates.io
  - Go: —
- Projects
  - e.g., OpenSSL, from https://openssl.org

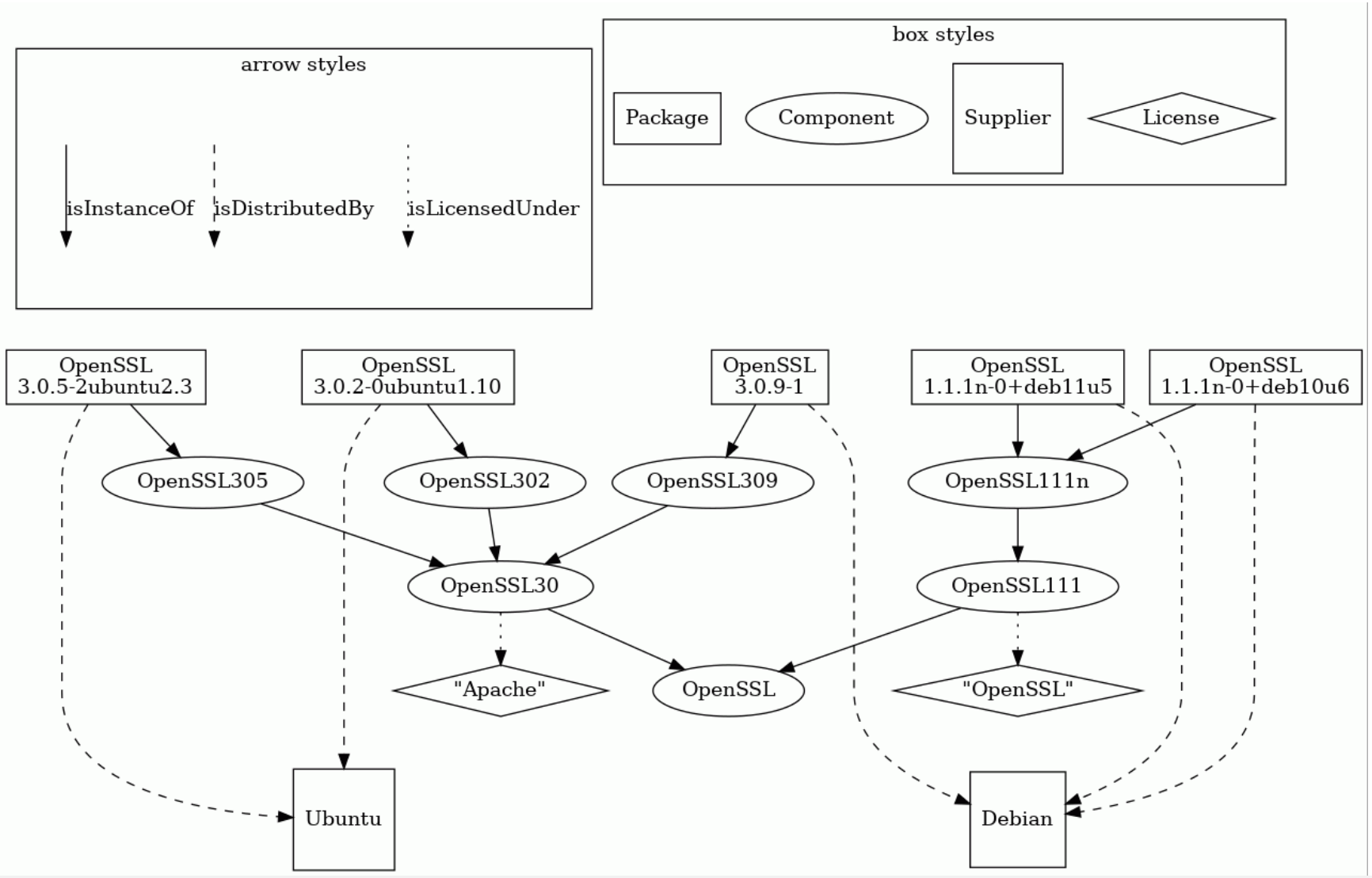
Ontology (fragment)



Example – curl



Example -- OpenSSL



Advantages of using Component

- Refer in abstract to a software component
  - Independent of version
  - Independent of supplier

- Typical example: licensing information
  - But also other, including security

Real-World Data

Real-world use

- Internally designed and developed
- Initial in mid-2023
- Incremental refinements of the model
  - Additional functional requirements
- A number of software implementations
  - Various stages: PoC, Prototype, Pilot, Production
  - Various technologies/infrastructure: Web/API, SQL/RDF, ...
- Continuous data collection
  - Initially ad-hoc; then more structured
  - Adapted to the data model

Supplier data

- Too many variations
- Publishes list of components?
- List includes
  - Latest version?
  - All versions?
  - Supplier stated license?
  - Source link?
  - Other meta-information?
  - All in one or query repeatedly?

Statistics (indicative, after few months)

Ecosystem / Supplier	Data collection	Packages	Components	Savings
Debian	1/mo	234,755	54,367	77%
Ubuntu	LTS, 1/3mo	482 540	182 882	62%
Rust	1/mo	662,172	171,430	74%
PyPi		13,368,874	624,417	95%


intel