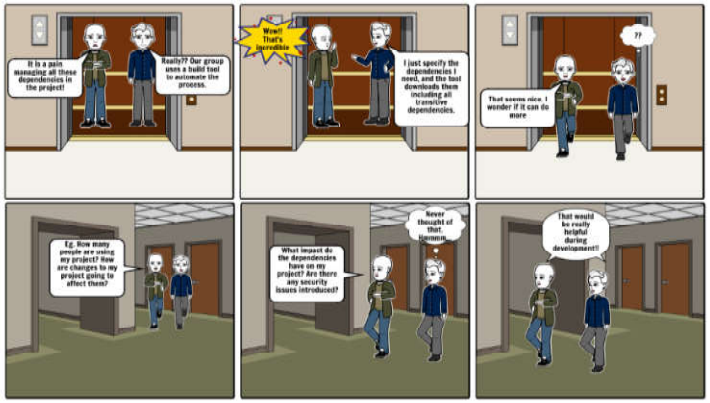
A Semantic Web Enabled Approach for Dependency Management

Ellis E. Eghan (e_eghan@encs.concordia.ca), Sultan S. Alqahtani (s_alqaht@encs.concordia.ca), Juergen Rilling (juergen.rilling@concordia.ca)

Concordia University, Canada

Ambient Software Evolution Group aseg.cs.concordia.ca | SERC | NSERC CRSNG | People. Discovery. Innovation. | Concordia University

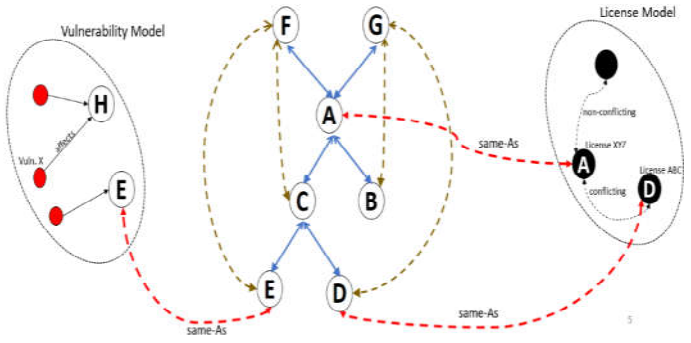
MOTIVATION



Current build tools provide support for automatic dependency management and analysis at the individual project level – **using only project-specific dependencies.**

In our approach, we extend this dependency analysis:

- to **cross-projects dependencies** – by creating a "global" dependency graph.
- integrating this graph with facts from **other software knowledge sources** to provide **bi-directional traceability.**
- support for **novel applications,** such as detecting license violations and to perform security vulnerability analysis within and across project boundaries.

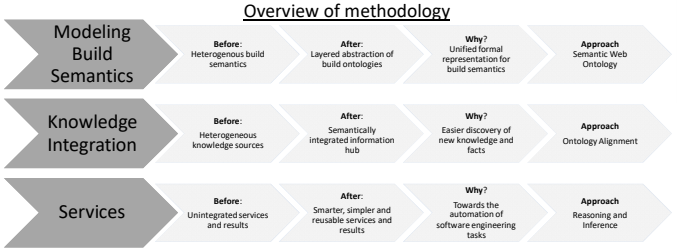


Illustrative example of an integrated knowledge graph

APPROACH

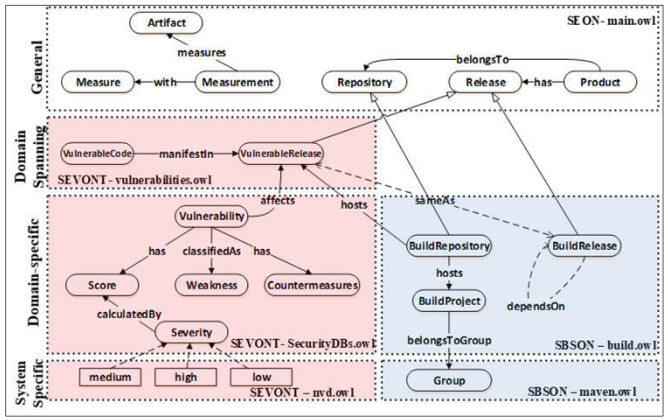
In our research, we introduce a novel approach that takes advantage of the Semantic Web and its technology stack (e.g., ontologies, Linked Data, reasoning services) to establish a unified knowledge representation of build semantics. We further automatically integrate the build model with other knowledge models to eliminate existing information silos and support new types of dependency analysis at a global scale.

Overview of methodology



Dataset Facts

Maven	NVD	Licenses
178,763 Projects 1,849,756 Releases 5,143 Organizations 66,777,338 direct dependencies 410,943 releases with Licenses	82,415 Vulnerabilities 29,354 Affected Projects 186,212 Affected Releases 16,017 Patched vulnerabilities	346,553 Apache-2.0 releases 25,511 MIT releases 7,971 LGPL-2.1 releases 6,690 EPL-1.0 releases 6,272 GPL-3.0 releases 6,069 BSD-3-Clause releases



Layered Abstraction of a subset of Integrated Ontologies

APPLICATIONS/SERVICES

Services supported by our integrated knowledge model include:

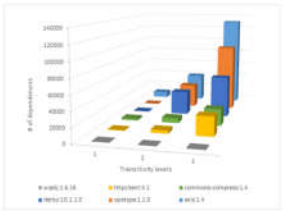
- Vulnerability impact analysis
- License violation analysis
- Identification of potential failures due to breaking changes in dependencies
- Assessing overall quality of build dependencies

INTERESTING RESULTS

Vulnerability Impact Analysis

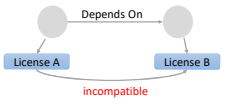
```
select ?secontProject ?mvnProject ?vulnerability
where{
   #identify the individuals from MAVON that have owl:sameAs property with SECONT individuals.
   ?secontProject owl:sameAs ?mvnProject.
   #identify all the CVE-IDs for all the vulnerable projects that satisfied the property "owl:sameAs"
   ?mvnProject secont:hasVulnerability ?vulnerability.
}
```

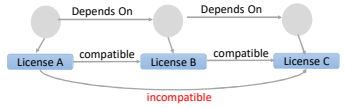
- 0.062% of all Maven projects contain known security vulnerabilities.
- 48.8% of the identified vulnerable project releases suffer from multiple security vulnerabilities (e.g. PostgreSQL 7.4.1 contains 25 known vulnerabilities)



As transitivity levels increase, the number of potentially affected dependent projects increase drastically.

License Violation Analysis





Top Direct License Violations in Maven	
Apache-2.0 -> GPL-2.0	2964
EPL-1.0 -> GPL-3.0	737

Top Indirect License Violations in Maven	
EPL-1.0 -> Apache-2.0 -> GPL-3.0	3784
GPL-3.0 -> LGPL-2.1 -> GPL-2.0	176

FUTURE WORK

- Identification of potential failures due to breaking changes in dependencies
- Impact analysis with build configuration

- Alqahtani, S. S., **Eghan, E. E.**, & Rilling, J. (2016). Recovering Semantic Traceability Links between APIs and Security Vulnerabilities: An Ontological Modeling Approach. In *10th IEEE International Conference on Software Testing, Verification and Validation (ICST).*
- Alqahtani, S. S., **Eghan, E. E.**, & Rilling, J. (2016). SV-AF — A Security Vulnerability Analysis Framework. In *2016 IEEE 27th International Symposium on Software Reliability Engineering (ISSRE)* (pp. 219–229). IEEE. http://doi.org/10.1109/ISSRE.2016.12
- Alqahtani, S. S., **Eghan, E. E.**, & Rilling, J. (2016). "Tracing Known Security Vulnerabilities in Software Repositories – A Semantic Web Enabled Modeling Approach." *Science of Computer Programming*, February. doi:10.1016/j.scico.2016.01.005.