

Tool Discovery

Maarten van Gompel, KNAW HuC

Introduction

We want to *automatically* harvest, unify and make available **software metadata**. We develop a *data provisioning* pipeline that can be used by portals (e.g. Ineo) which end-users (researchers) interact with to find suitable tools for their needs.

Guiding Principles

- ▶ **Do Not Repeat Yourself**; reuse existing software metadata practices and existing standards
- ▶ **Metadata at the source**
 - ▶ developers themselves are best suited to describe their tool
 - ▶ keep metadata alongside the source code
 - ▶ all software derives from sort of source code
 - ▶ bottom-up, provenance
- ▶ **Short automatic update cycle**; ensure information is always up-to-date
- ▶ **Linked Open Data**; highly structured and interconnected data
- ▶ **Validation**; ensure information is correct and warn if it isn't
- ▶ Distinguish **Software** and **Software as a Service**

Harvester and Converter



- codemeta
- schema.org
- repostatus.org
- spdx.org



Linked Open Data

Software Components

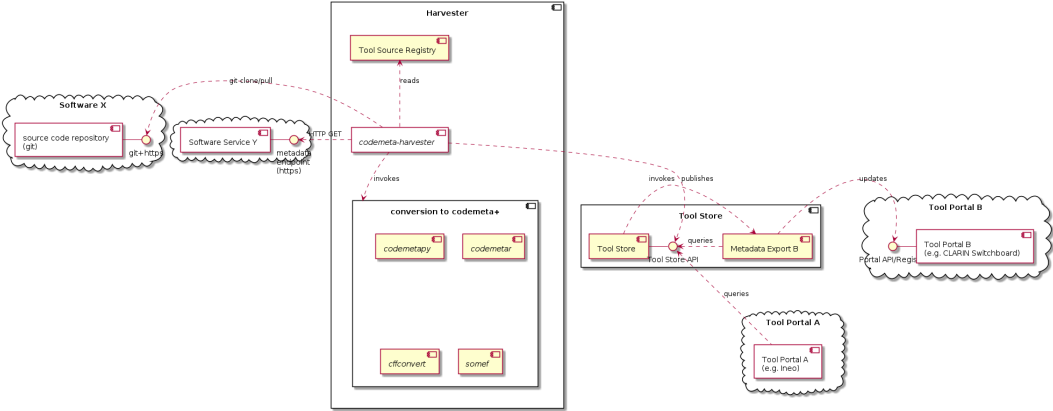


Figure 2: Tool Discovery Component Diagram

Tool Store

- ▶ Simple in-memory RDF triple store
- ▶ Offers JSON-LD and Turtle serialisations (RDF) for every resource
- ▶ SPARQL endpoint (plus YASGUI interface for interactive use)
- ▶ Simple web-interface for end-users; gain some insight into the data

Implementation

- ▶ codemeta-server - Tool Store API (Python)
 - ▶ <https://github.com/proycon/codemeta-server>
- ▶ codemeta-harvester - Harvester (POSIX shell)
 - ▶ <https://github.com/proycon/codemeta-harvester>
- ▶ codemetapy - Converter (Python)
 - ▶ <https://github.com/proycon/codemetapy>

3rd party:

- ▶ cffconvert - CITATION.cff to codemeta
- ▶ somef - Software metadata extraction framework (optional)

Towards prescriptive metadata

Automatic harvesting is not always enough:

- ▶ Exotic software, non-standardized input
- ▶ Additional prescriptive domain-specific vocabulary demanded by the project (e.g. CLARIAH)
- ▶ Harvesting errors

Tool developers can use the codemeta-harvester to generate a `codemeta.json` for their software, then edit/amend it and add it to their source code repository.

Demo

<https://tools.dev.clariah.nl>