Cataloguing Software Ecosystems with swid-reg

Alex J. Nelson, Ph.D.
Computer Scientist
NIST

Software Supply Chain Assurance Forum
2023-09-13

Disclaimer

The views and opinions expressed in this presentation are those of the authors and do not necessarily reflect the official policy or position of any agency of the U.S. government. Any mention of a vendor or product is not an endorsement or recommendation. Logos and trademarks are copyright their respective owners.

whoami

Ph.D., Computer Science, 2016
Emphases: File Systems, Digital Forensics, and Information Retrieval

Computer Scientist

UCO Ontology Committee Chair

Ontology Engineer

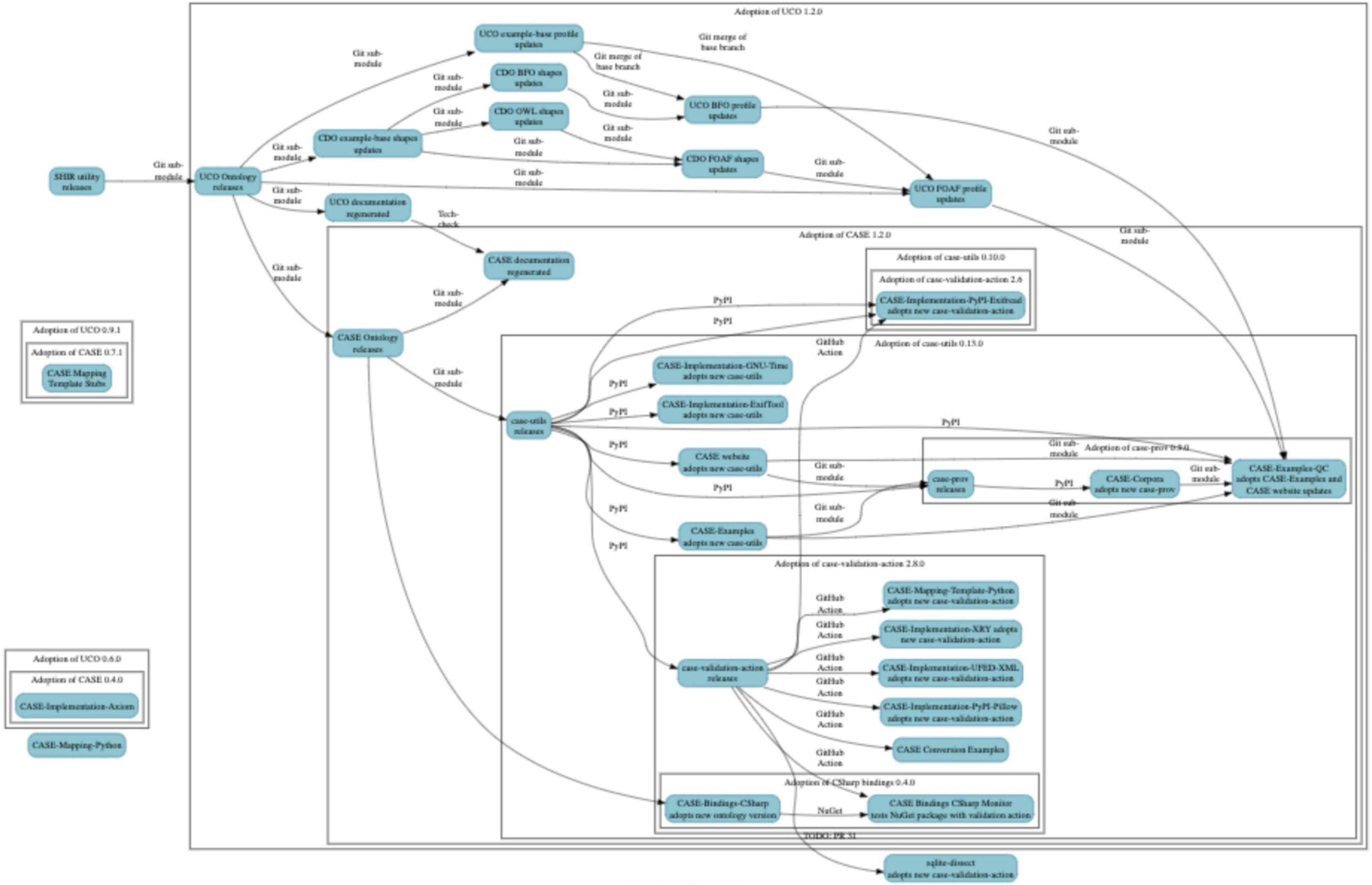
An ontology engineer’s perspective on supply chain

The chart on the right is the Cyber Domain Ontology’s (CDO) “Release Flow” diagram.

Each teal node is a public GitHub repository, providing an ontology, software, or example data.

Each arrow shows how updates propagate between repositories.

Experience has encouraged keeping update procedures small in human effort, including from external dependencies (e.g. Python code formatters pinned to latest versions).


Adoption of UCO 1.2.0
SHIR utility releases
Git sub-module
UCO Ontology releases
UCO example-base profile updates
CDO BFO shapes updates
CDO OWL shapes updates
CDO example-base shapes updates
Git merge of base branch
UCO BFO profile updates
CDO FOAF shapes updates
UCO FOAF profile updates
UCO documentation regenerated
Tech-check
CASE documentation regenerated
Adoption of CASE 1.2.0
CASE Ontology releases
Adoption of case-utils 0.10.0
Adoption of case-validation-action 2.6
CASE-Implementation-PyPI-Exifread adopts new case-validation-action
PyPI
GitHub Action
Adoption of case-utils 0.13.0
case-utils releases
CASE-Implementation-GNU-Time adopts new case-utils
CASE-Implementation-ExifTool adopts new case-utils
CASE website adopts new case-utils
CASE-Examples adopts new case-utils
Adoption of case-prov 0.9.0
case-prov releases
CASE-Corpora adopts new case-prov
CASE-Examples-QC adopts CASE-Examples and CASE website updates
Adoption of case-validation-action 2.8.0
case-validation-action releases
CASE-Mapping-Template-Python adopts new case-validation-action
CASE-Implementation-XRY adopts new case-validation-action
CASE-Implementation-UFED-XML adopts new case-validation-action
CASE-Implementation-PyPI-Pillow adopts new case-validation-action
CASE Conversion Examples
Adoption of CSharp bindings 0.4.0
CASE-Bindings-CSharp adopts new ontology version
NuGet
CASE Bindings CSharp Monitor tests NuGet package with validation action
TODO: PR 31
sqlite-dissect adopts new case-validation-action
Adoption of UCO 0.9.1
Adoption of CASE 0.7.1
CASE Mapping Template Stubs
Adoption of UCO 0.6.0
Adoption of CASE 0.4.0
CASE-Implementation-Axiom
CASE-Mapping-Python

Release flow of CDO ontologies

Nodes link to the respective Git repository.

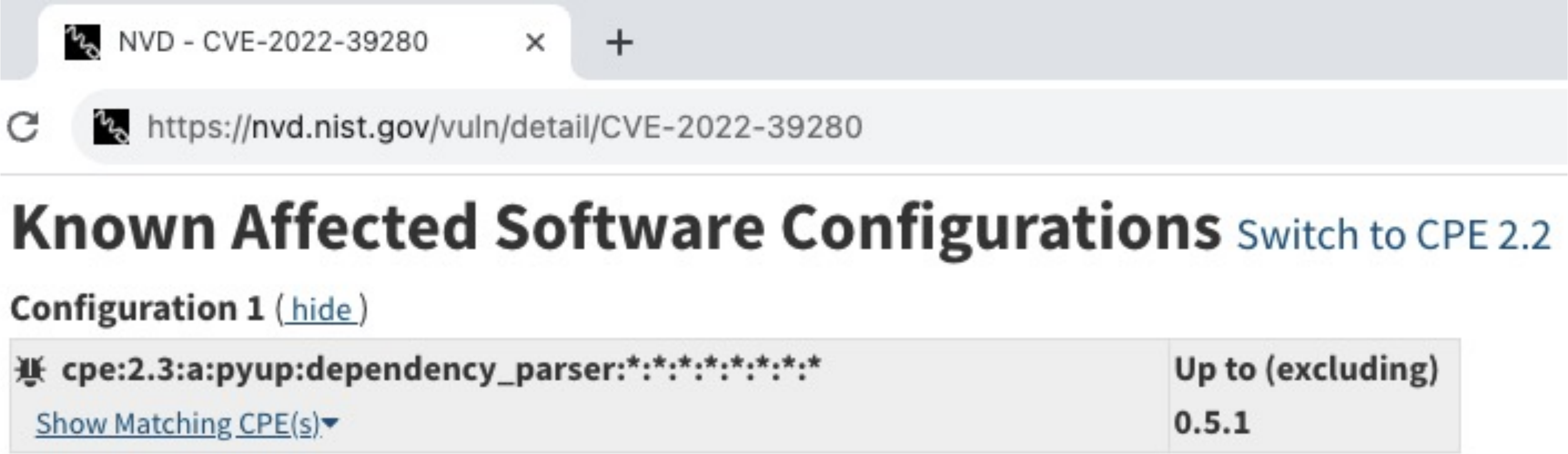
https://cyberdomainontology.org/resources/project_release_flow.html

Outline

- Background: NVD and CPE
- Package managers and swid-reg
- A light touch of ontology, tailored to software supply chain

NVD and CPE

Problem:
CPEs are a significant labor point in NVD.



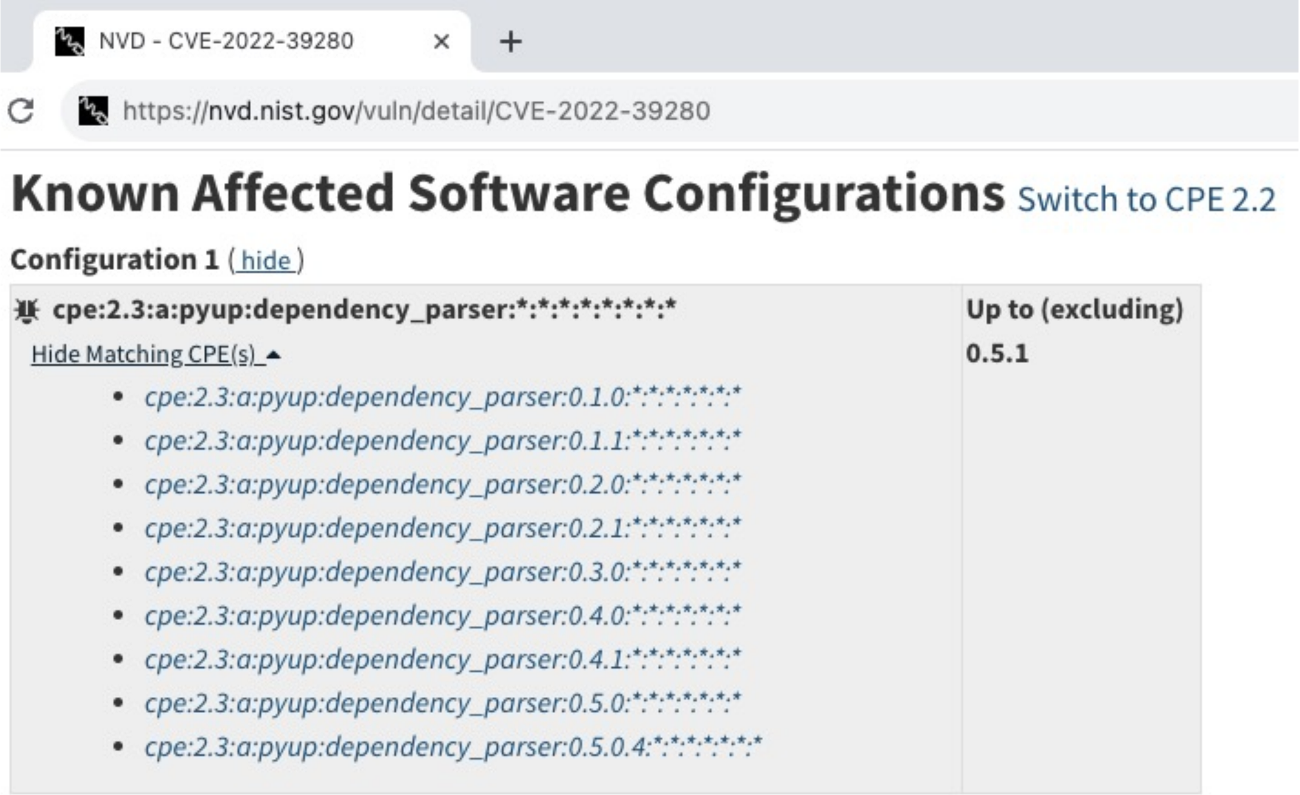
Known Affected Software Configurations Switch to CPE 2.2

Configuration 1 (hide)

cpe:2.3:a:pyup:dependency_parser:*:*:*:*:*:*:*:* Show Matching CPE(s)	Up to (excluding) 0.5.1

From “dependency_parser, all versions <0.5.1”, NVD enumerates affected versions.

Problem:
CPEs are a significant labor point in NVD.



NVD - CVE-2022-39280

https://nvd.nist.gov/vuln/detail/CVE-2022-39280

Known Affected Software Configurations Switch to CPE 2.2

Configuration 1 (hide)

cpe:2.3:a:pyup:dependency_parser:*:*:*:*:*:*:*:*	**Up to (excluding)**
Hide Matching CPE(s) ▲	**0.5.1**

- *cpe:2.3:a:pyup:dependency_parser:0.1.0:*:*:*:*:*:*:**
- *cpe:2.3:a:pyup:dependency_parser:0.1.1:*:*:*:*:*:*:**
- *cpe:2.3:a:pyup:dependency_parser:0.2.0:*:*:*:*:*:*:**
- *cpe:2.3:a:pyup:dependency_parser:0.2.1:*:*:*:*:*:*:**
- *cpe:2.3:a:pyup:dependency_parser:0.3.0:*:*:*:*:*:*:**
- *cpe:2.3:a:pyup:dependency_parser:0.4.0:*:*:*:*:*:*:**
- *cpe:2.3:a:pyup:dependency_parser:0.4.1:*:*:*:*:*:*:**
- *cpe:2.3:a:pyup:dependency_parser:0.5.0:*:*:*:*:*:*:**
- *cpe:2.3:a:pyup:dependency_parser:0.5.0.4:*:*:*:*:*:*:**

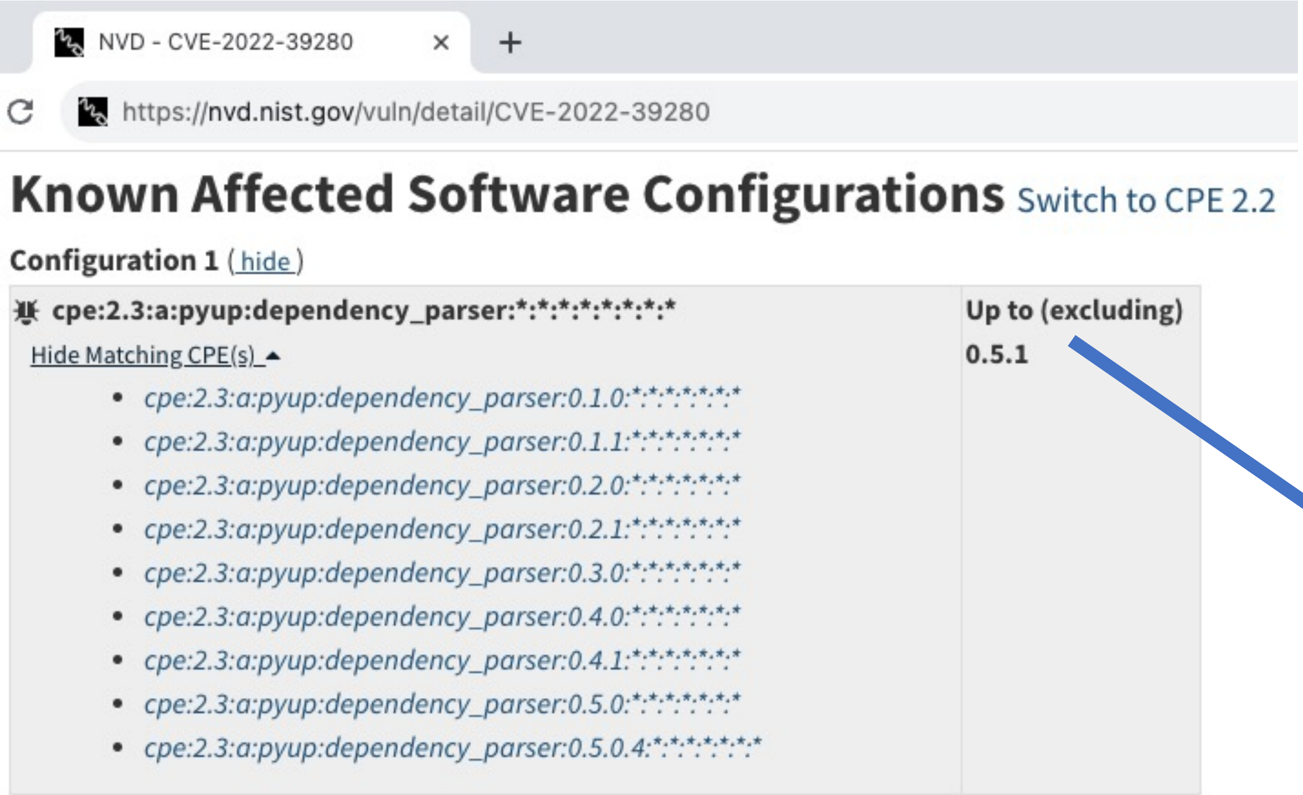
From “dependency_parser, all versions <0.5.1”, NVD enumerates affected versions.

Step 1: Enumerate all versions, or most in range’s ballpark.

Step 2: Identify affected subset and define CPEs.

For open source software…

…why not crawl?



https://nvd.nist.gov/vuln/detail/CVE-2022-39280

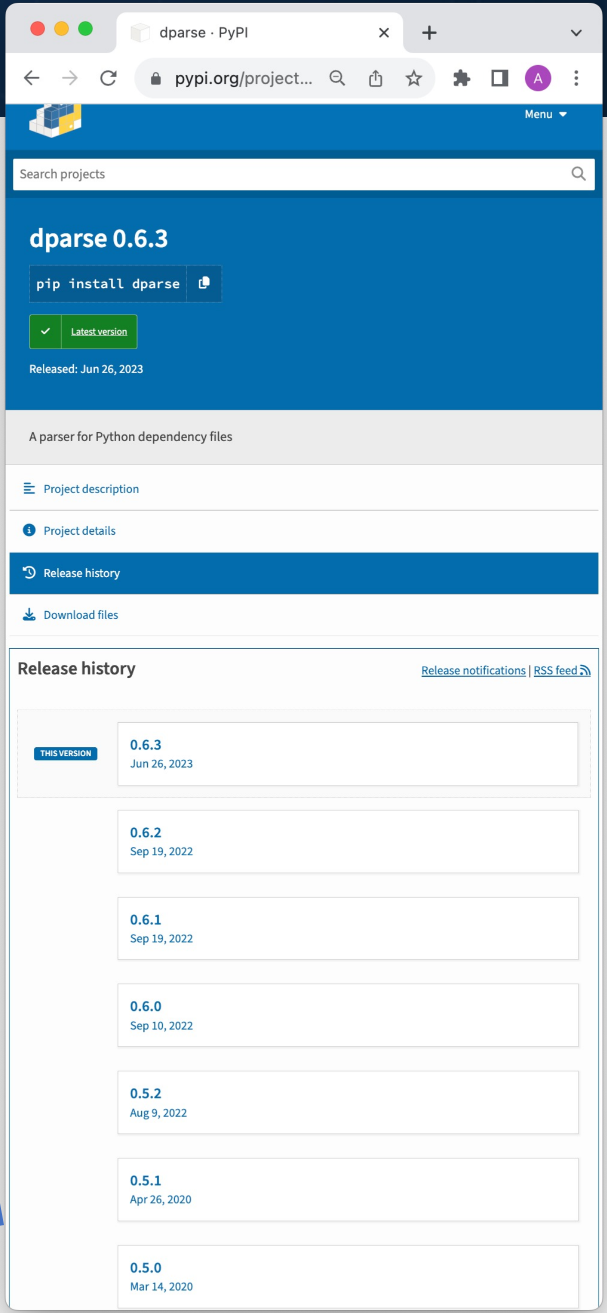
Known Affected Software Configurations Switch to CPE 2.2

Configuration 1 (hide)

cpe:2.3:a:pyup:dependency_parser:*:*:*:*:*:*:*:*	**Up to (excluding)** **0.5.1**

Hide Matching CPE(s)

- *cpe:2.3:a:pyup:dependency_parser:0.1.0:*:*:*:*:*:*:**
- *cpe:2.3:a:pyup:dependency_parser:0.1.1:*:*:*:*:*:*:**
- *cpe:2.3:a:pyup:dependency_parser:0.2.0:*:*:*:*:*:*:**
- *cpe:2.3:a:pyup:dependency_parser:0.2.1:*:*:*:*:*:*:**
- *cpe:2.3:a:pyup:dependency_parser:0.3.0:*:*:*:*:*:*:**
- *cpe:2.3:a:pyup:dependency_parser:0.4.0:*:*:*:*:*:*:**
- *cpe:2.3:a:pyup:dependency_parser:0.4.1:*:*:*:*:*:*:**
- *cpe:2.3:a:pyup:dependency_parser:0.5.0:*:*:*:*:*:*:**
- *cpe:2.3:a:pyup:dependency_parser:0.5.0.4:*:*:*:*:*:*:**



dparse 0.6.3

```
pip install dparse
```

Released: Jun 26, 2023

A parser for Python dependency files

Release history

0.6.3	Jun 26, 2023
0.6.2	Sep 19, 2022
0.6.1	Sep 19, 2022
0.6.0	Sep 10, 2022
0.5.2	Aug 9, 2022
0.5.1	Apr 26, 2020
0.5.0	Mar 14, 2020

SWID, swid-reg, and CPE mapping

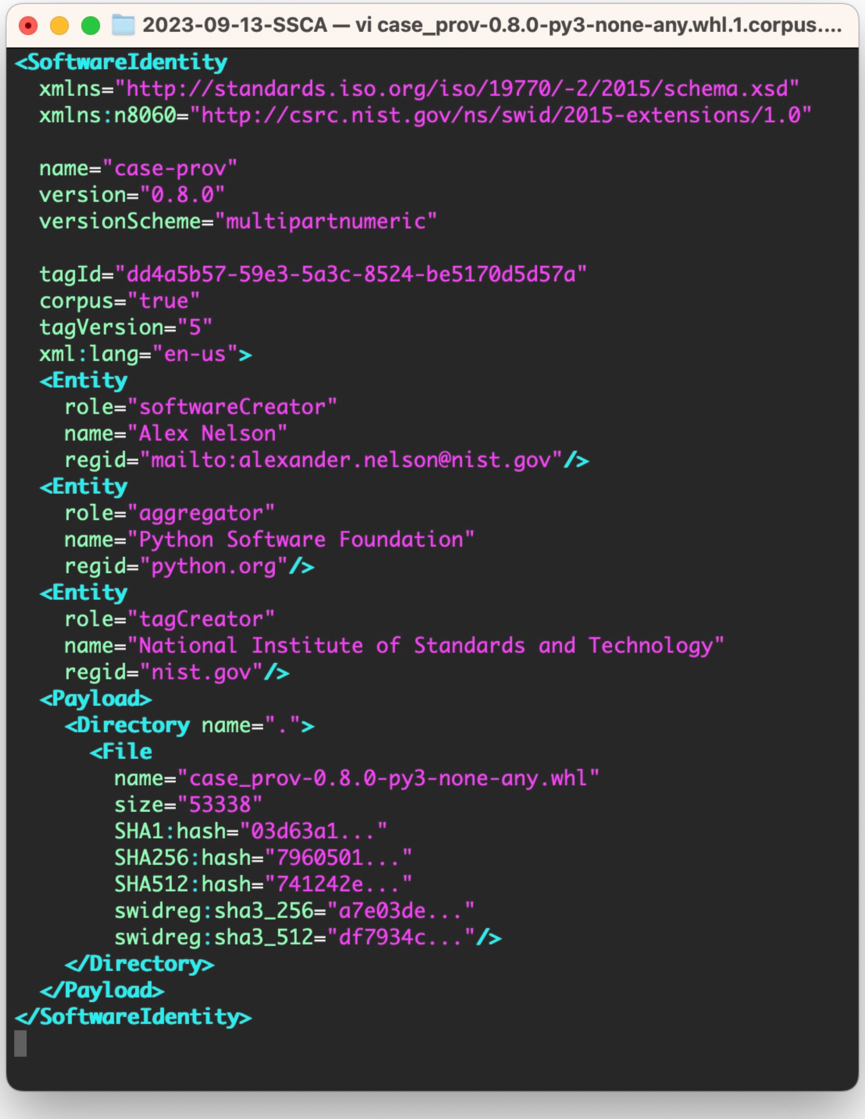
SWID is a metadata format for software.

A SWID Tag, XML, can map to CPE.

For example, this CPE …

```
cpe:2.3:a:alex_nelson:case-prov:0.8.0:*:*:*:*:*:*:*
```

… generates from that:



2023-09-13-SSCA — vi case_prov-0.8.0-py3-none-any.whl.1.corpus....

```xml
<SoftwareIdentity
  xmlns="http://standards.iso.org/iso/19770/-2/2015/schema.xsd"
  xmlns:n8060="http://csrc.nist.gov/ns/swid/2015-extensions/1.0"

  name="case-prov"
  version="0.8.0"
  versionScheme="multipartnumeric"

  tagId="dd4a5b57-59e3-5a3c-8524-be5170d5d57a"
  corpus="true"
  tagVersion="5"
  xml:lang="en-us">
  <Entity
    role="softwareCreator"
    name="Alex Nelson"
    regid="mailto:alexander.nelson@nist.gov"/>
  <Entity
    role="aggregator"
    name="Python Software Foundation"
    regid="python.org"/>
  <Entity
    role="tagCreator"
    name="National Institute of Standards and Technology"
    regid="nist.gov"/>
  <Payload>
    <Directory name=".">
      <File
        name="case_prov-0.8.0-py3-none-any.whl"
        size="53338"
        SHA1:hash="03d63a1..."
        SHA256:hash="7960501..."
        SHA512:hash="741242e..."
        swidreg:sha3_256="a7e03de..."
        swidreg:sha3_512="df7934c..."/>
    </Directory>
  </Payload>
</SoftwareIdentity>
```

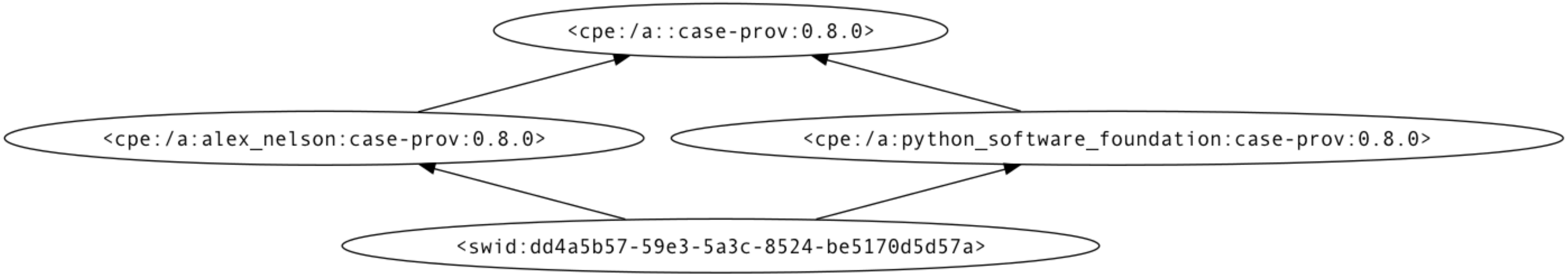
SWID and CPE differ in precision

One SWID tag can induce many CPEs.

E.g. 1 per involved entity (aside from `tagCreator`). Each “CPE Vendor” variant could assist with finding distributor-modified software.

```
cpe:2.3:a:alex_nelson:case-prov:0.8.0:*:*:*:*:*:*:*
cpe:2.3:a:python_software_foundation:case-prov:0.8.0:*:*:*:*:*:*:*
```

Could be more helpful to consider as classes. These CPEs describe all software named case-prov, versioned 0.8.0, vended by Alex Nelson (or separately, vended by the Python Software Foundation).



Classes…?

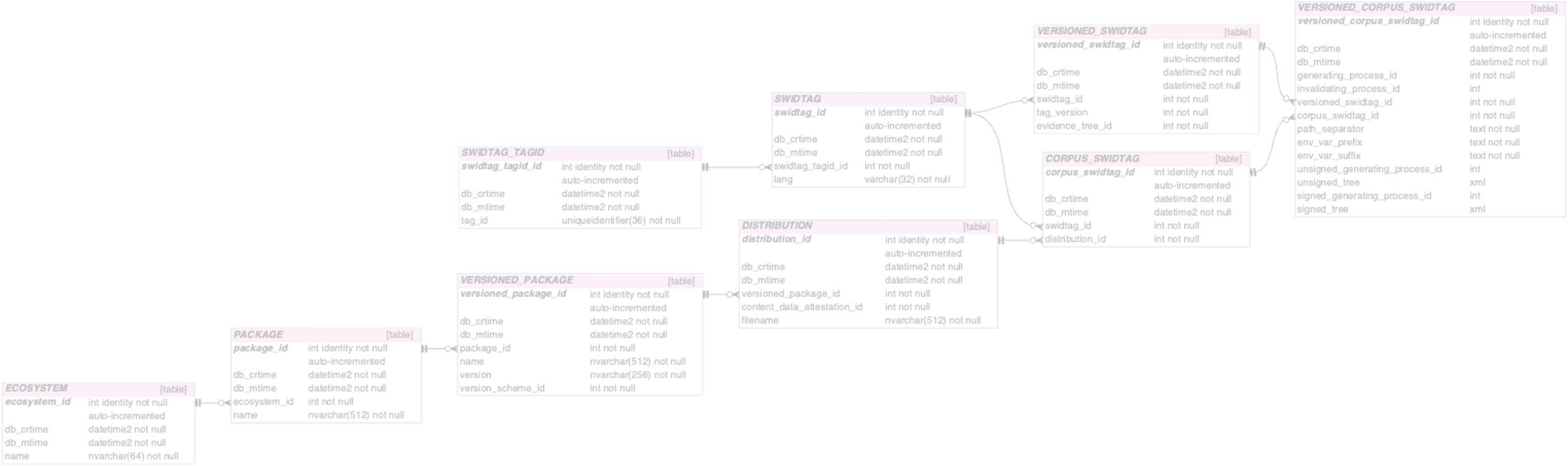
Treat classes synonymous with sets.

A key ITAM objective: Knowledge of assets.
“How big is the set of computers in my org?”
“How big is the set of software licenses in use right now?
“…Versus how many we’ve paid for?”

CPE describes sets of software.
CPE *name* serializes the set description.
CPE *URI* identifies the set. E.g. all software: `<cpe:/a>`
SWID describes smaller sets.

“Never underestimate the power of a theorem that counts something.”

Crawling package managers with swid-reg



ECOSYSTEM	[table]
ecosystem_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
name	nvarchar(64) not null

PACKAGE	[table]
package_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
ecosystem_id	int not null
name	nvarchar(512) not null

VERSIONED_PACKAGE	[table]
versioned_package_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
package_id	int not null
name	nvarchar(512) not null
version	nvarchar(256) not null
version_scheme_id	int not null

SWIDTAG_TAGID	[table]
swidtag_tagid_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
tag_id	uniqueidentifier(36) not null

DISTRIBUTION	[table]
distribution_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
versioned_package_id	int not null
content_data_attestation_id	int not null
filename	nvarchar(512) not null

SWIDTAG	[table]
swidtag_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
swidtag_tagid_id	int not null
lang	varchar(32) not null

VERSIONED_SWIDTAG	[table]
versioned_swidtag_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
swidtag_id	int not null
tag_version	int not null
evidence_tree_id	int not null

CORPUS_SWIDTAG	[table]
corpus_swidtag_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
swidtag_id	int not null
distribution_id	int not null

VERSIONED_CORPUS_SWIDTAG	[table]
versioned_corpus_swidtag_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
generating_process_id	int not null
invalidating_process_id	int
versioned_swidtag_id	int not null
corpus_swidtag_id	int not null
path_separator	text not null
env_var_prefix	text not null
env_var_suffix	text not null
unsigned_generating_process_id	int
unsigned_tree	xml
signed_generating_process_id	int
signed_tree	xml

Package managers

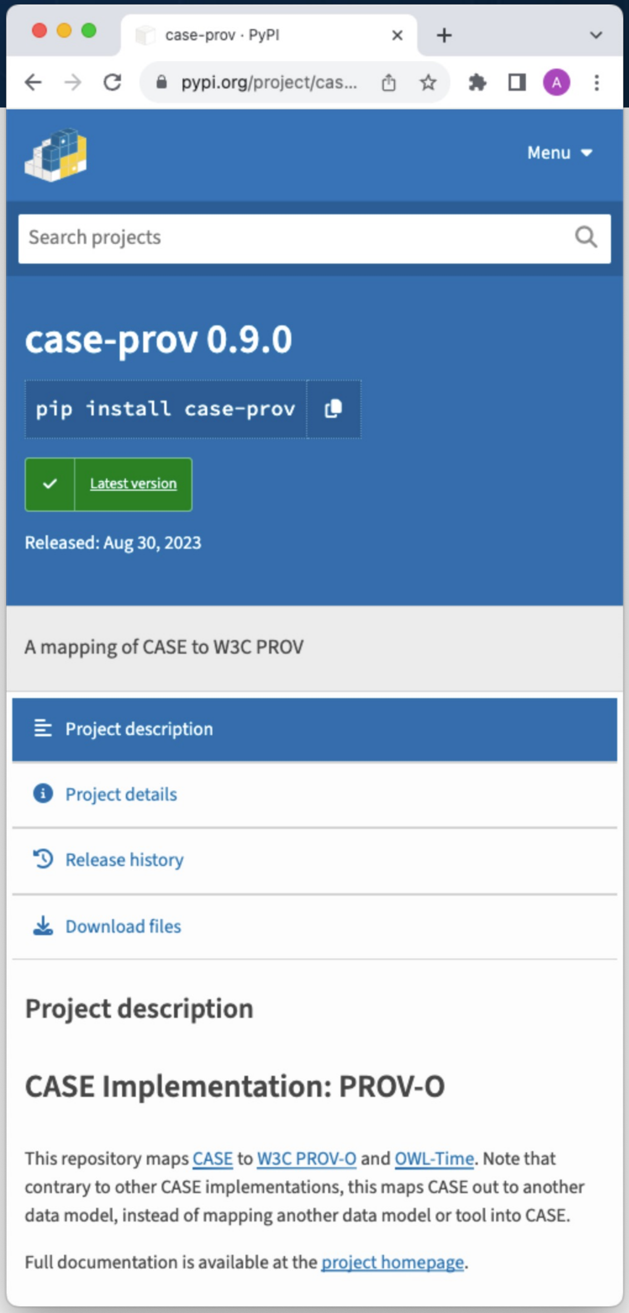
Package managers provide:

- Authorship information
- Package update discovery
- Dependency graphs
- Payload files
- Project pages
  https://pypi.org/project/case-prov/

swid-reg calls package managers **ecosystems**,

and crawls them to produce SWID tags**.**

(And from SWID tags, CPEs will be generated.)



case-prov 0.9.0

```
pip install case-prov
```

Released: Aug 30, 2023

A mapping of CASE to W3C PROV

Project description

CASE Implementation: PROV-O

This repository maps CASE to W3C PROV-O and OWL-Time. Note that contrary to other CASE implementations, this maps CASE out to another data model, instead of mapping another data model or tool into CASE.

Full documentation is available at the project homepage.

Package managers and swid-reg

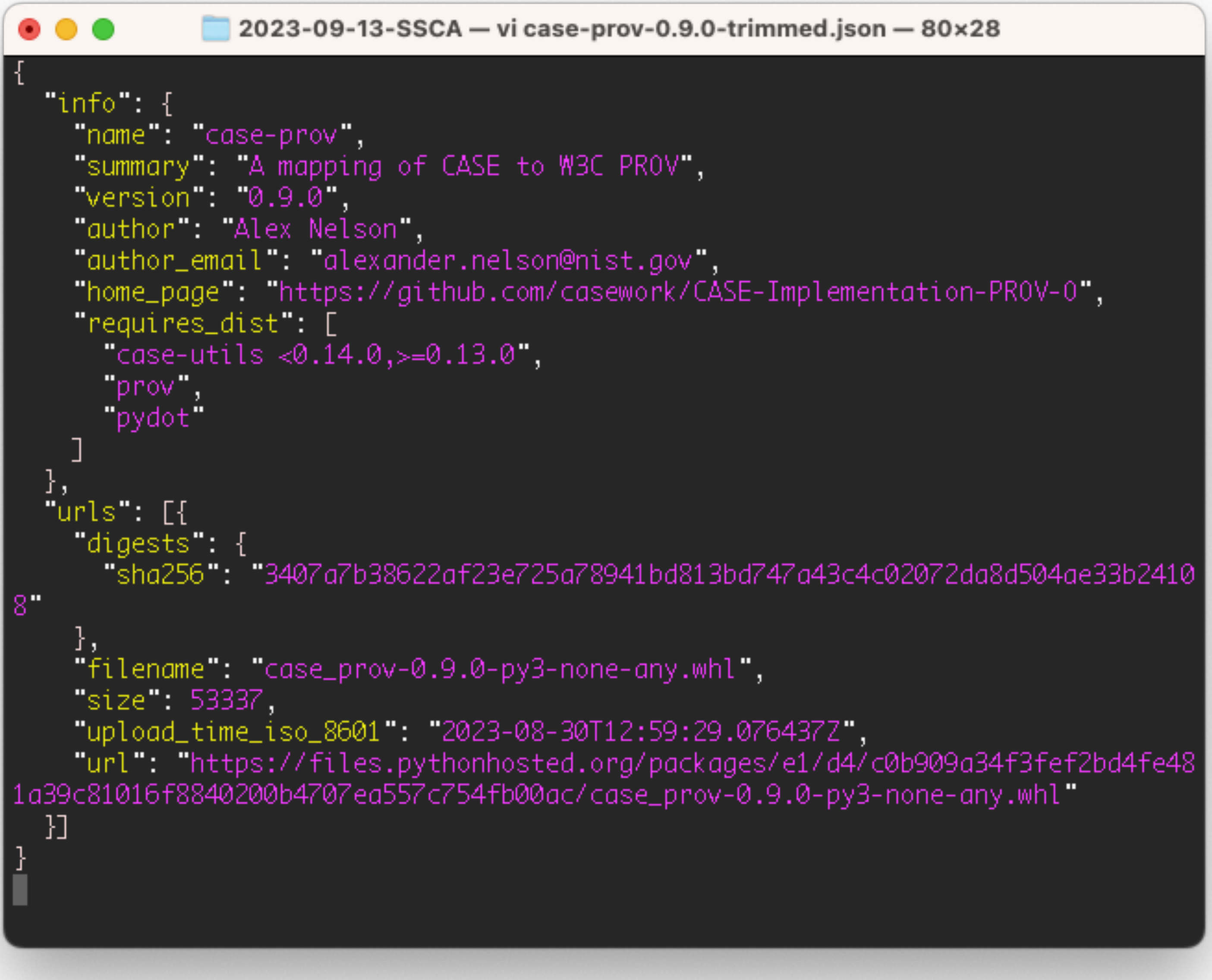
Package managers provide:

- Authorship information
- Package update discovery
- Dependency graphs
- Payload files
- Metadata feeds
  https://pypi.org/pypi/case-prov/0.9.0/json

swid-reg calls package managers **ecosystems**,

and crawls them to produce SWID tags**.**

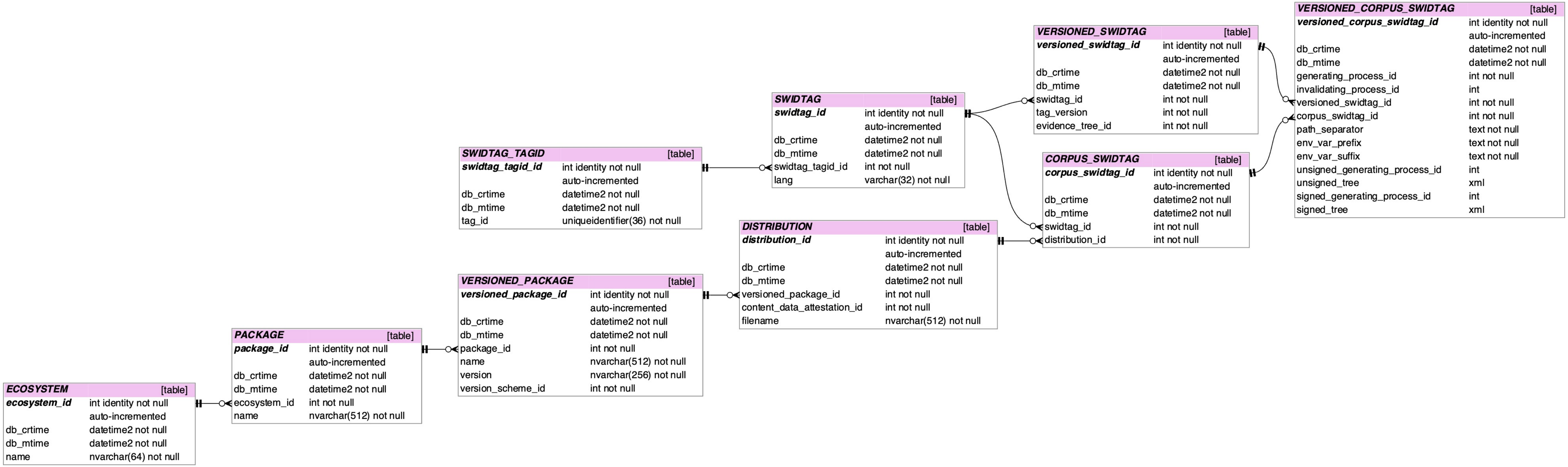
(And from SWID tags, CPEs will be generated.)



2023-09-13-SSCA — vi case-prov-0.9.0-trimmed.json — 80×28

```
{
  "info": {
    "name": "case-prov",
    "summary": "A mapping of CASE to W3C PROV",
    "version": "0.9.0",
    "author": "Alex Nelson",
    "author_email": "alexander.nelson@nist.gov",
    "home_page": "https://github.com/casework/CASE-Implementation-PROV-O",
    "requires_dist": [
      "case-utils <0.14.0,>=0.13.0",
      "prov",
      "pydot"
    ]
  },
  "urls": [{
    "digests": {
      "sha256": "3407a7b38622af23e725a78941bd813bd747a43c4c02072da8d504ae33b2410
8"
    },
    "filename": "case_prov-0.9.0-py3-none-any.whl",
    "size": 53337,
    "upload_time_iso_8601": "2023-08-30T12:59:29.076437Z",
    "url": "https://files.pythonhosted.org/packages/e1/d4/c0b909a34f3fef2bd4fe48
1a39c81016f8840200b4707ea557c754fb00ac/case_prov-0.9.0-py3-none-any.whl"
  }]
}
```

swid-reg data model:
From ecosystem to SWID tag



ECOSYSTEM	[table]
ecosystem_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
name	nvarchar(64) not null

PACKAGE	[table]
package_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
ecosystem_id	int not null
name	nvarchar(512) not null

VERSIONED_PACKAGE	[table]
versioned_package_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
package_id	int not null
name	nvarchar(512) not null
version	nvarchar(256) not null
version_scheme_id	int not null

SWIDTAG_TAGID	[table]
swidtag_tagid_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
tag_id	uniqueidentifier(36) not null

DISTRIBUTION	[table]
distribution_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
versioned_package_id	int not null
content_data_attestation_id	int not null
filename	nvarchar(512) not null

SWIDTAG	[table]
swidtag_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
swidtag_tagid_id	int not null
lang	varchar(32) not null

VERSIONED_SWIDTAG	[table]
versioned_swidtag_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
swidtag_id	int not null
tag_version	int not null
evidence_tree_id	int not null

CORPUS_SWIDTAG	[table]
corpus_swidtag_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
swidtag_id	int not null
distribution_id	int not null

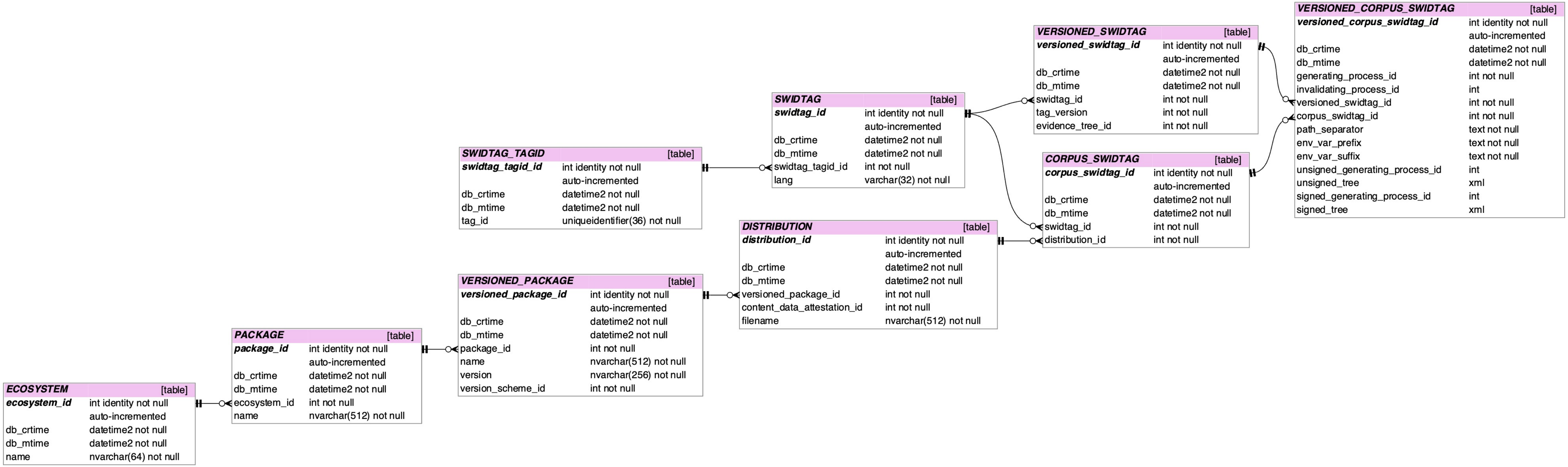
VERSIONED_CORPUS_SWIDTAG	[table]
versioned_corpus_swidtag_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
generating_process_id	int not null
invalidating_process_id	int
versioned_swidtag_id	int not null
corpus_swidtag_id	int not null
path_separator	text not null
env_var_prefix	text not null
env_var_suffix	text not null
unsigned_generating_process_id	int
unsigned_tree	xml
signed_generating_process_id	int
signed_tree	xml

What does one download from a package manager?
– A **distribution**.

What bears the version?
– A **versioned package**.

What do swidtags describe in swid-reg?
– A **distribution**.

swid-reg data model:
From ecosystem to SWID tag



ECOSYSTEM	[table]
ecosystem_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
name	nvarchar(64) not null

PACKAGE	[table]
package_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
ecosystem_id	int not null
name	nvarchar(512) not null

VERSIONED_PACKAGE	[table]
versioned_package_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
package_id	int not null
name	nvarchar(512) not null
version	nvarchar(256) not null
version_scheme_id	int not null

SWIDTAG_TAGID	[table]
swidtag_tagid_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
tag_id	uniqueidentifier(36) not null

DISTRIBUTION	[table]
distribution_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
versioned_package_id	int not null
content_data_attestation_id	int not null
filename	nvarchar(512) not null

SWIDTAG	[table]
swidtag_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
swidtag_tagid_id	int not null
lang	varchar(32) not null

VERSIONED_SWIDTAG	[table]
versioned_swidtag_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
swidtag_id	int not null
tag_version	int not null
evidence_tree_id	int not null

CORPUS_SWIDTAG	[table]
corpus_swidtag_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
swidtag_id	int not null
distribution_id	int not null

VERSIONED_CORPUS_SWIDTAG	[table]
versioned_corpus_swidtag_id	int identity not null auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
generating_process_id	int not null
invalidating_process_id	int
versioned_swidtag_id	int not null
corpus_swidtag_id	int not null
path_separator	text not null
env_var_prefix	text not null
env_var_suffix	text not null
unsigned_generating_process_id	int
unsigned_tree	xml
signed_generating_process_id	int
signed_tree	xml

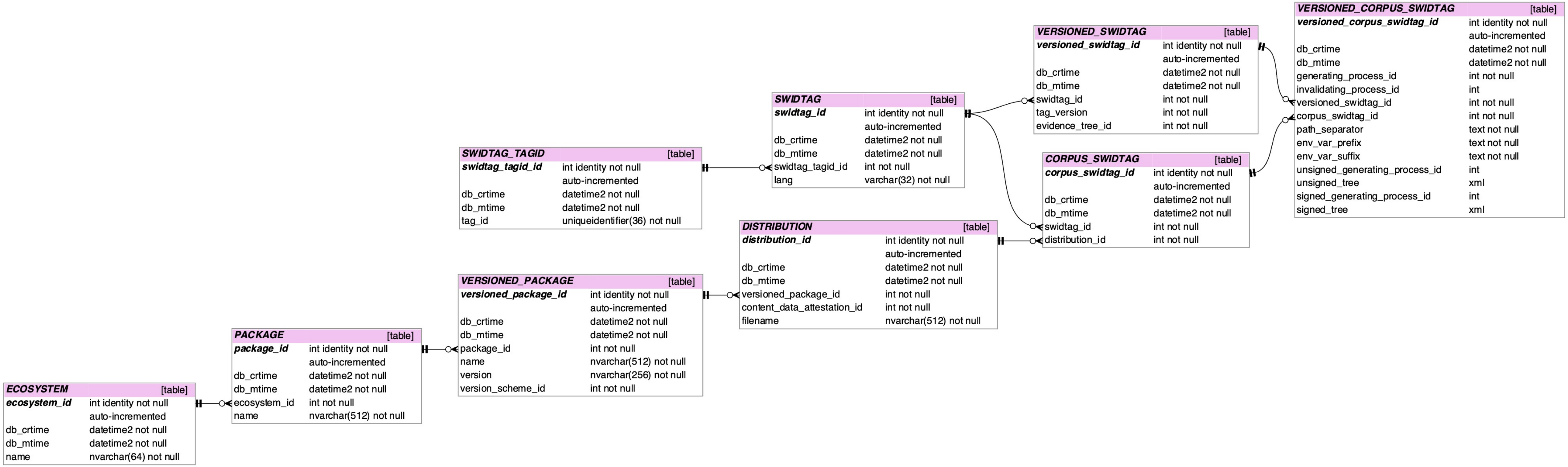
What does one download from a package manager?
– A **distribution**.

What bears the version?
– A **versioned package**.

What do swidtags describe in swid-reg?
– A **distribution**.

But…

Difference between Projects and Packages: Packages are in ecosystems.



ECOSYSTEM	[table]
ecosystem_id	int identity not null
	auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
name	nvarchar(64) not null

PACKAGE	[table]
package_id	int identity not null
	auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
ecosystem_id	int not null
name	nvarchar(512) not null

VERSIONED_PACKAGE	[table]
versioned_package_id	int identity not null
	auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
package_id	int not null
name	nvarchar(512) not null
version	nvarchar(256) not null
version_scheme_id	int not null

SWIDTAG_TAGID	[table]
swidtag_tagid_id	int identity not null
	auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
tag_id	uniqueidentifier(36) not null

DISTRIBUTION	[table]
distribution_id	int identity not null
	auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
versioned_package_id	int not null
content_data_attestation_id	int not null
filename	nvarchar(512) not null

SWIDTAG	[table]
swidtag_id	int identity not null
	auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
swidtag_tagid_id	int not null
lang	varchar(32) not null

VERSIONED_SWIDTAG	[table]
versioned_swidtag_id	int identity not null
	auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
swidtag_id	int not null
tag_version	int not null
evidence_tree_id	int not null

CORPUS_SWIDTAG	[table]
corpus_swidtag_id	int identity not null
	auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
swidtag_id	int not null
distribution_id	int not null

VERSIONED_CORPUS_SWIDTAG	[table]
versioned_corpus_swidtag_id	int identity not null
	auto-incremented
db_crtime	datetime2 not null
db_mtime	datetime2 not null
generating_process_id	int not null
invalidating_process_id	int
versioned_swidtag_id	int not null
corpus_swidtag_id	int not null
path_separator	text not null
env_var_prefix	text not null
env_var_suffix	text not null
unsigned_generating_process_id	int
unsigned_tree	xml
signed_generating_process_id	int
signed_tree	xml

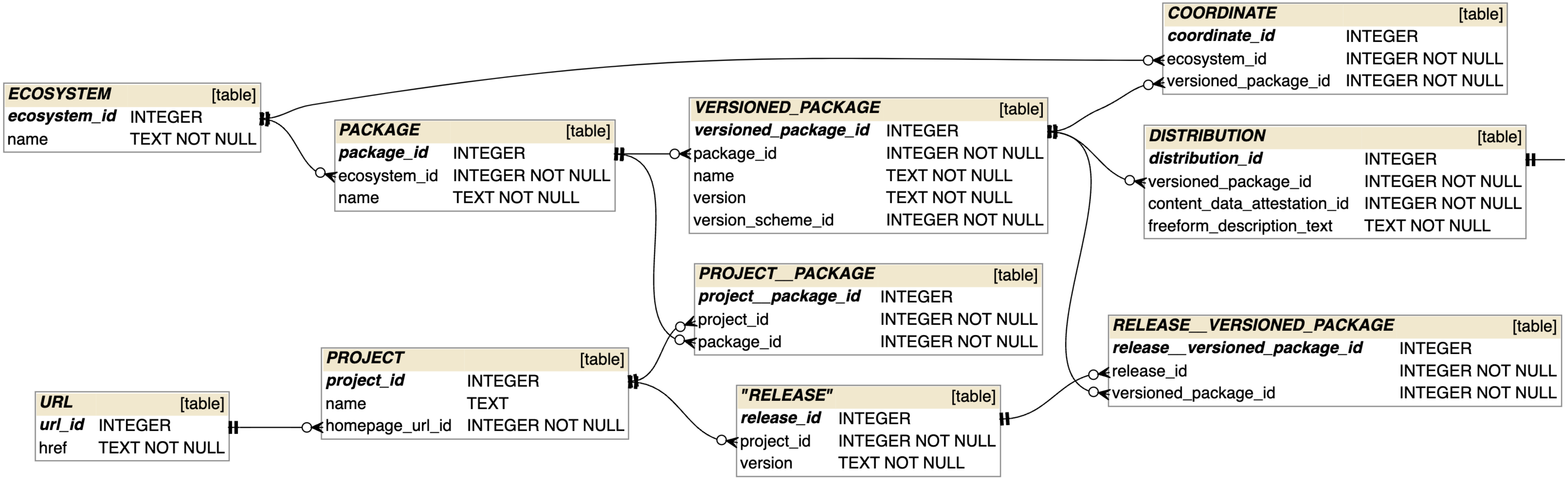
What does one download from a package manager?
– A **distribution**.

What bears the version?
– A **versioned package**.

What do swidtags describe in swid-reg?
– A **distribution**.

But…
What are vulnerabilities reported against?
– A **release**, bearing a version of a **project**.
(Not depicted.)

Difference between Projects and Packages: Packages are easier to crawl.



COORDINATE	[table]
coordinate_id	INTEGER
ecosystem_id	INTEGER NOT NULL
versioned_package_id	INTEGER NOT NULL

ECOSYSTEM	[table]
ecosystem_id	INTEGER
name	TEXT NOT NULL

PACKAGE	[table]
package_id	INTEGER
ecosystem_id	INTEGER NOT NULL
name	TEXT NOT NULL

VERSIONED_PACKAGE	[table]
versioned_package_id	INTEGER
package_id	INTEGER NOT NULL
name	TEXT NOT NULL
version	TEXT NOT NULL
version_scheme_id	INTEGER NOT NULL

DISTRIBUTION	[table]
distribution_id	INTEGER
versioned_package_id	INTEGER NOT NULL
content_data_attestation_id	INTEGER NOT NULL
freeform_description_text	TEXT NOT NULL

PROJECT__PACKAGE	[table]
project__package_id	INTEGER
project_id	INTEGER NOT NULL
package_id	INTEGER NOT NULL

RELEASE__VERSIONED_PACKAGE	[table]
release__versioned_package_id	INTEGER
release_id	INTEGER NOT NULL
versioned_package_id	INTEGER NOT NULL

PROJECT	[table]
project_id	INTEGER
name	TEXT
homepage_url_id	INTEGER NOT NULL

URL	[table]
url_id	INTEGER
href	TEXT NOT NULL

"RELEASE"	[table]
release_id	INTEGER
project_id	INTEGER NOT NULL
version	TEXT NOT NULL

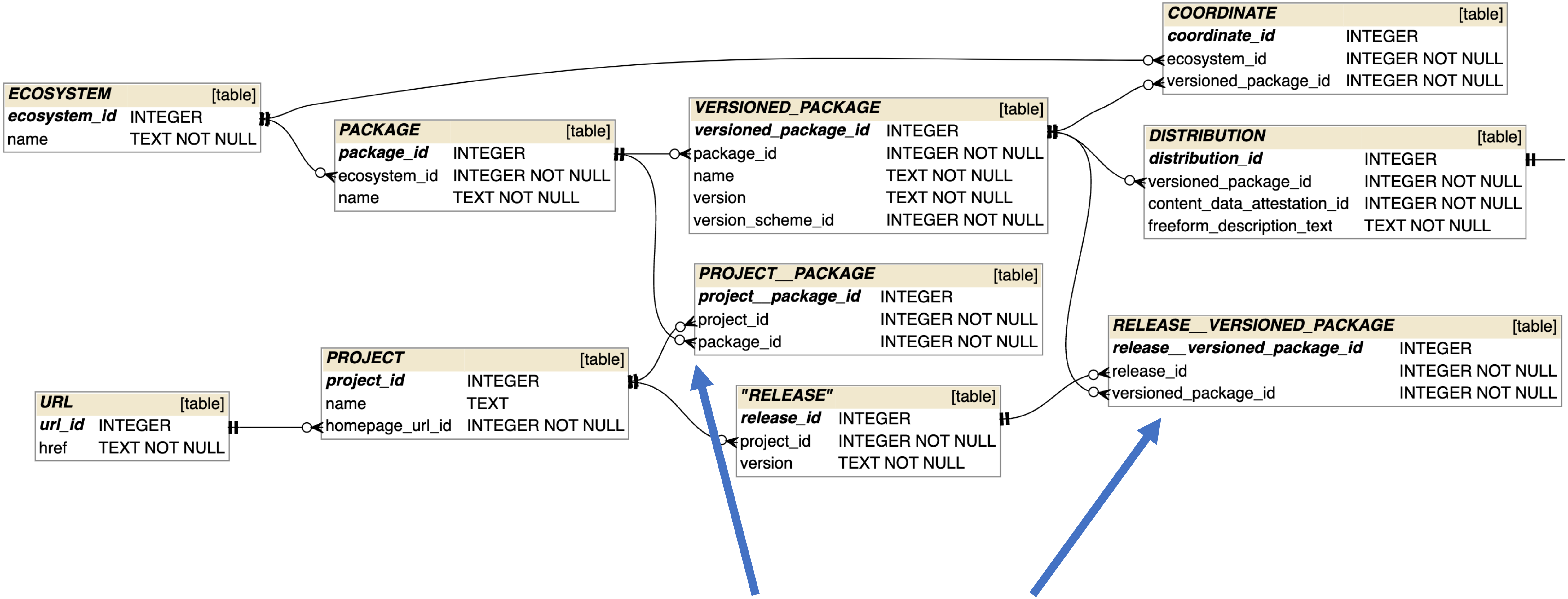
What does one download from a package manager?
– A **distribution**.

What bears the version?
– A **versioned package**.

What do swidtags describe in swid-reg?
– A **distribution**.

But…
What are vulnerabilities reported against?
– A **release**, bearing a version of a **project**.

Bridging Projects and Packages: A challenge.



COORDINATE	[table]
coordinate_id	INTEGER
ecosystem_id	INTEGER NOT NULL
versioned_package_id	INTEGER NOT NULL

ECOSYSTEM	[table]
ecosystem_id	INTEGER
name	TEXT NOT NULL

PACKAGE	[table]
package_id	INTEGER
ecosystem_id	INTEGER NOT NULL
name	TEXT NOT NULL

VERSIONED_PACKAGE	[table]
versioned_package_id	INTEGER
package_id	INTEGER NOT NULL
name	TEXT NOT NULL
version	TEXT NOT NULL
version_scheme_id	INTEGER NOT NULL

DISTRIBUTION	[table]
distribution_id	INTEGER
versioned_package_id	INTEGER NOT NULL
content_data_attestation_id	INTEGER NOT NULL
freeform_description_text	TEXT NOT NULL

PROJECT__PACKAGE	[table]
project__package_id	INTEGER
project_id	INTEGER NOT NULL
package_id	INTEGER NOT NULL

RELEASE__VERSIONED_PACKAGE	[table]
release__versioned_package_id	INTEGER
release_id	INTEGER NOT NULL
versioned_package_id	INTEGER NOT NULL

PROJECT	[table]
project_id	INTEGER
name	TEXT
homepage_url_id	INTEGER NOT NULL

URL	[table]
url_id	INTEGER
href	TEXT NOT NULL

"RELEASE"	[table]
release_id	INTEGER
project_id	INTEGER NOT NULL
version	TEXT NOT NULL

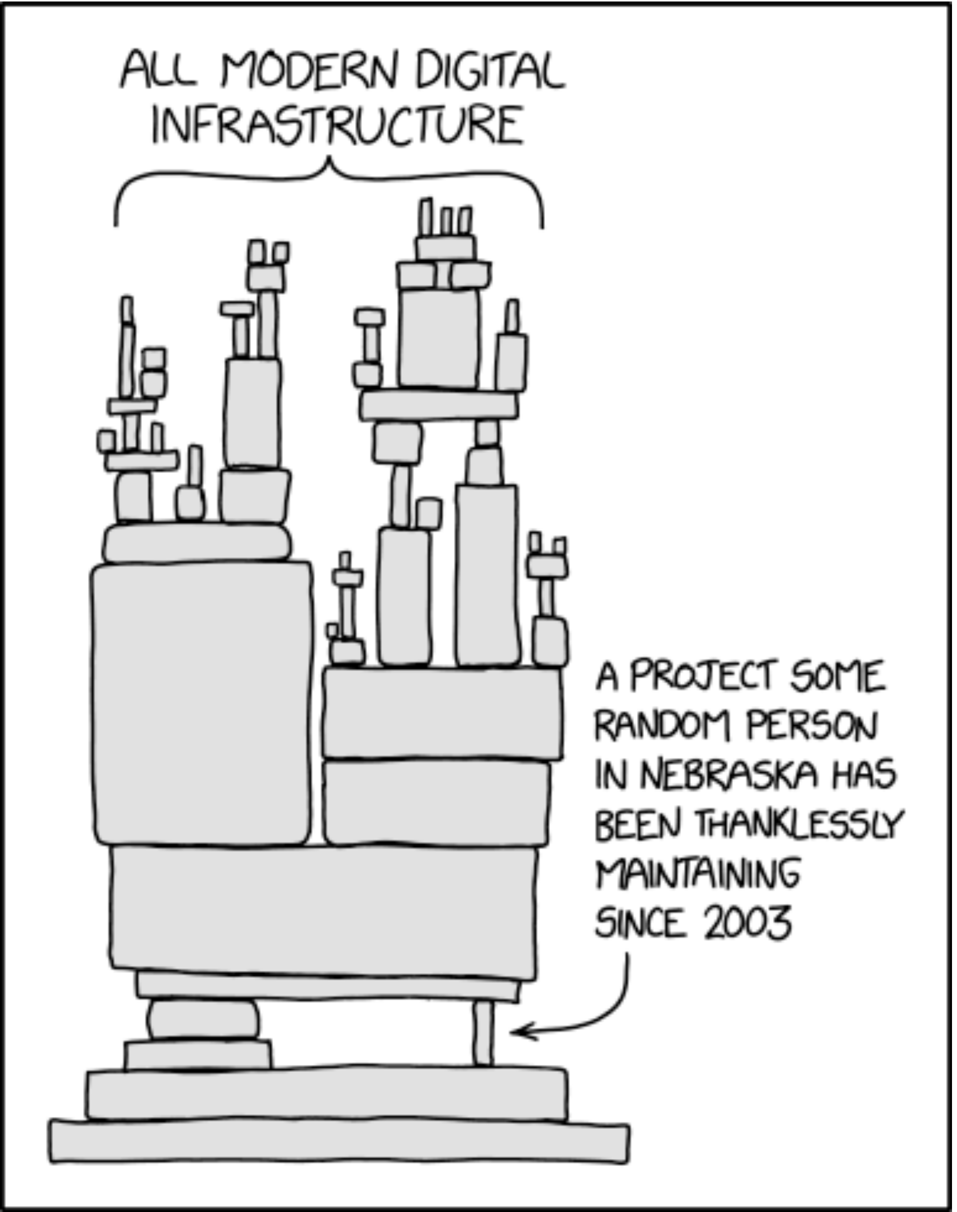
These links require much customization to populate.

Planned open-source ecosystems

Crawlers have been designed for:

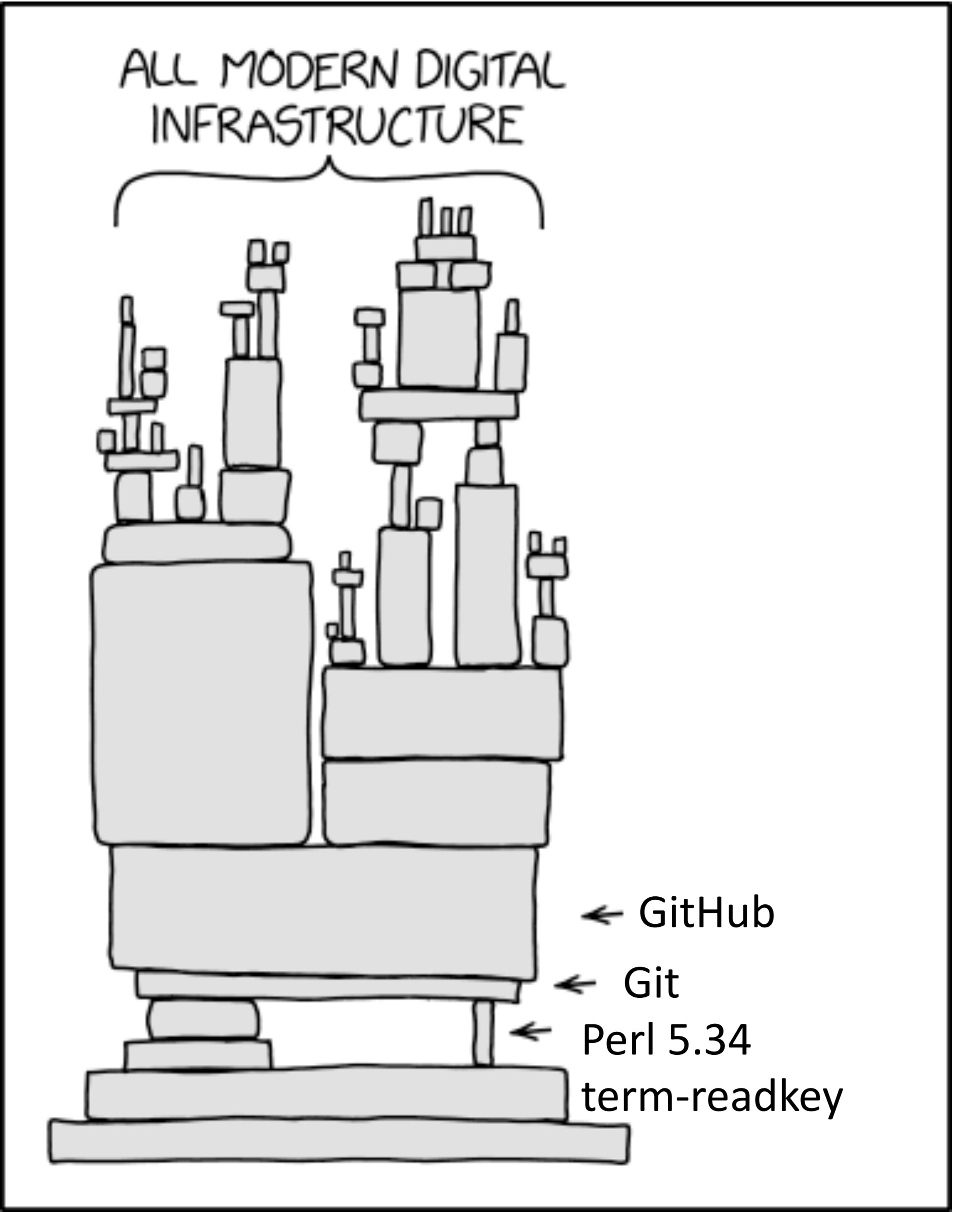
- PyPI (activating)
- Maven
- Debian
- CPAN
- RubyGems
- NPM

…did he say CPAN?


ALL MODERN DIGITAL INFRASTRUCTURE
A PROJECT SOME RANDOM PERSON IN NEBRASKA HAS BEEN THANKLESSLY MAINTAINING SINCE 2003

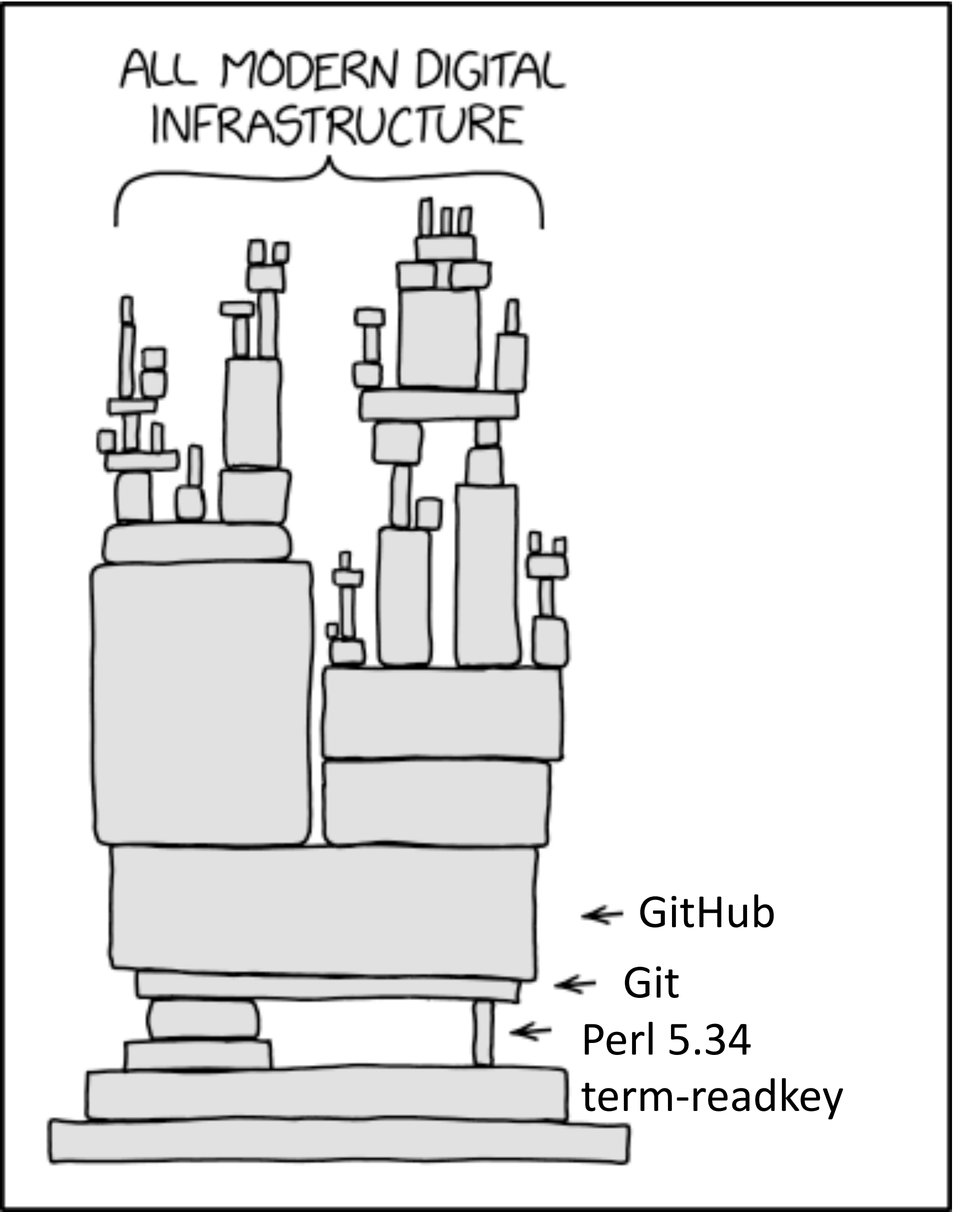
https://xkcd.com/2347/

…did he say CPAN?

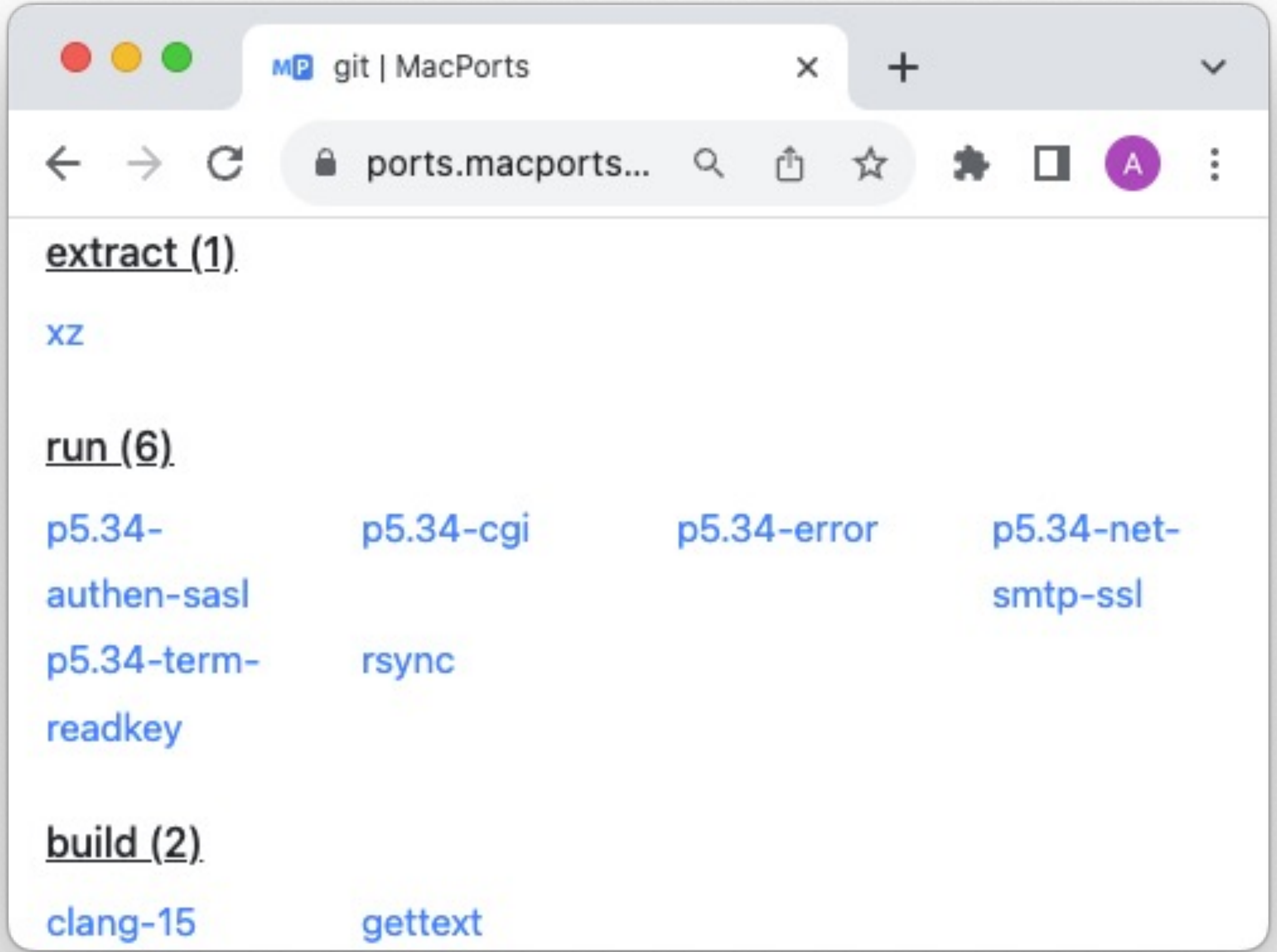


https://xkcd.com/2347/

…did he say CPAN?



Per: https://ports.macports.org/port/git/details/



extract (1)

xz

run (6)

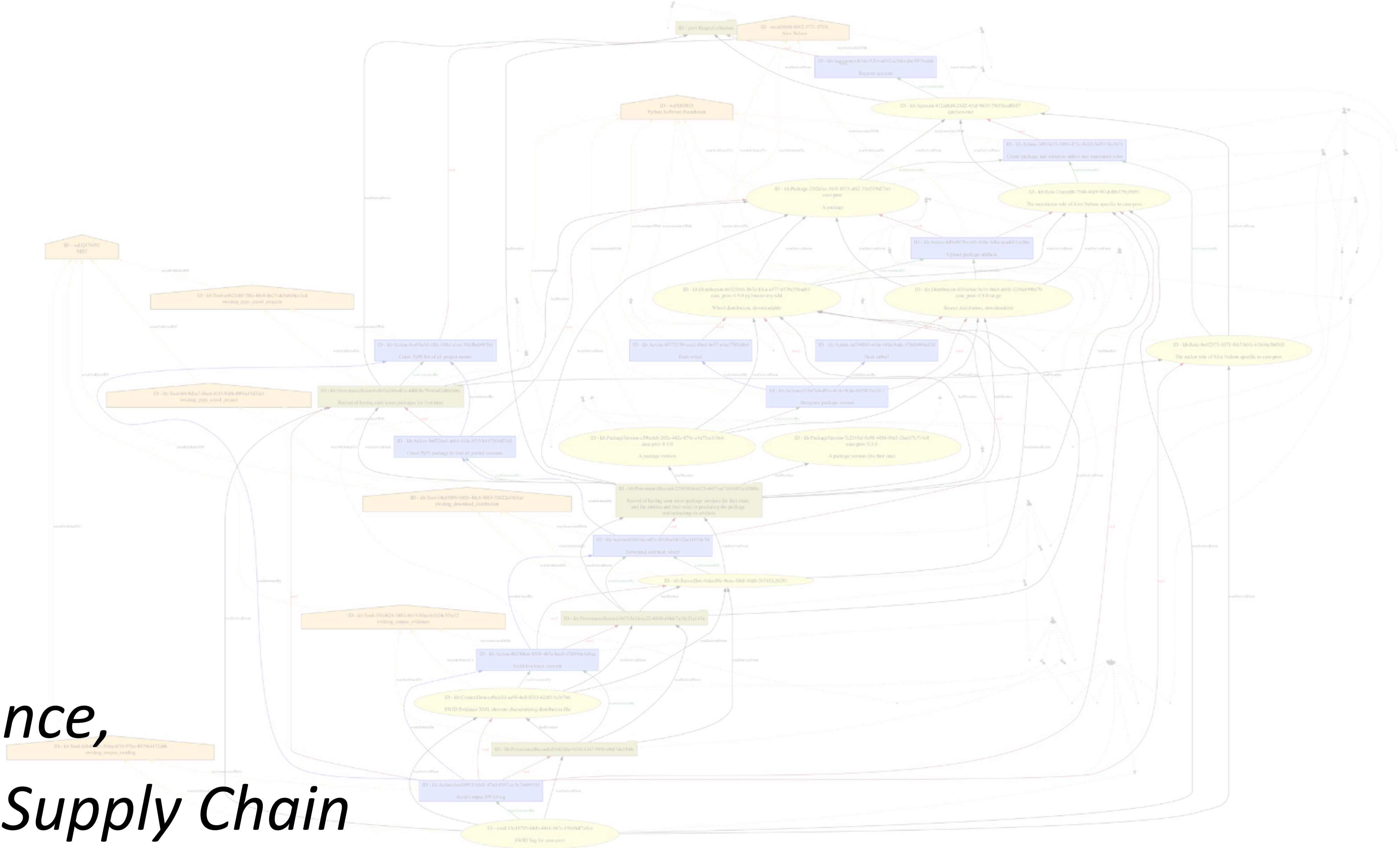
p5.34-authen-sasl p5.34-cgi p5.34-error p5.34-net-smtp-ssl p5.34-term-readkey rsync

build (2)

clang-15 gettext

https://xkcd.com/2347/

A light touch of ontology



ID - prov:EmptyCollection
Alex Nelson
Register account
ID - wd:Q83818
Python Software Foundation
jgnelson-nist
Create package and initialize author and maintainer roles
case-prov
A package
The maintainer role of Alex Nelson specific to case-prov
ID - wd:Q175091
NIST
swidreg_pypi_crawl_projects
Upload package artifacts
case_prov-0.9.0-py3-none-any.whl
Wheel distribution, downloadable
case_prov-0.9.0.tar.gz
Source distribution, downloadable
Crawl PyPI list of all project names
Hash wheel
Hash tarball
The author role of Alex Nelson specific to case-prov
swidreg_pypi_crawl_project
Record of having seen some packages for first time
Designate package version
Crawl PyPI package to find all posted versions
case-prov 0.9.0
A package version
case-prov 0.3.0
A package version (the first one)
Record of having seen some package versions for first time, and the entities and their roles in producing the package and uploading its artifacts
swidreg_download_distribution
Download and hash wheel
swidreg_corpus_evidence
Build Evidence element
SWID Evidence XML element characterizing distribution file
swidreg_corpus_swidtag
Build Corpus SWID tag
SWID Tag for case-prov
wasDerivedFrom
wasGeneratedBy
wasAssociatedWith
wasAttributedTo
wasInformedBy
used
hadMember

Linking,

Time,

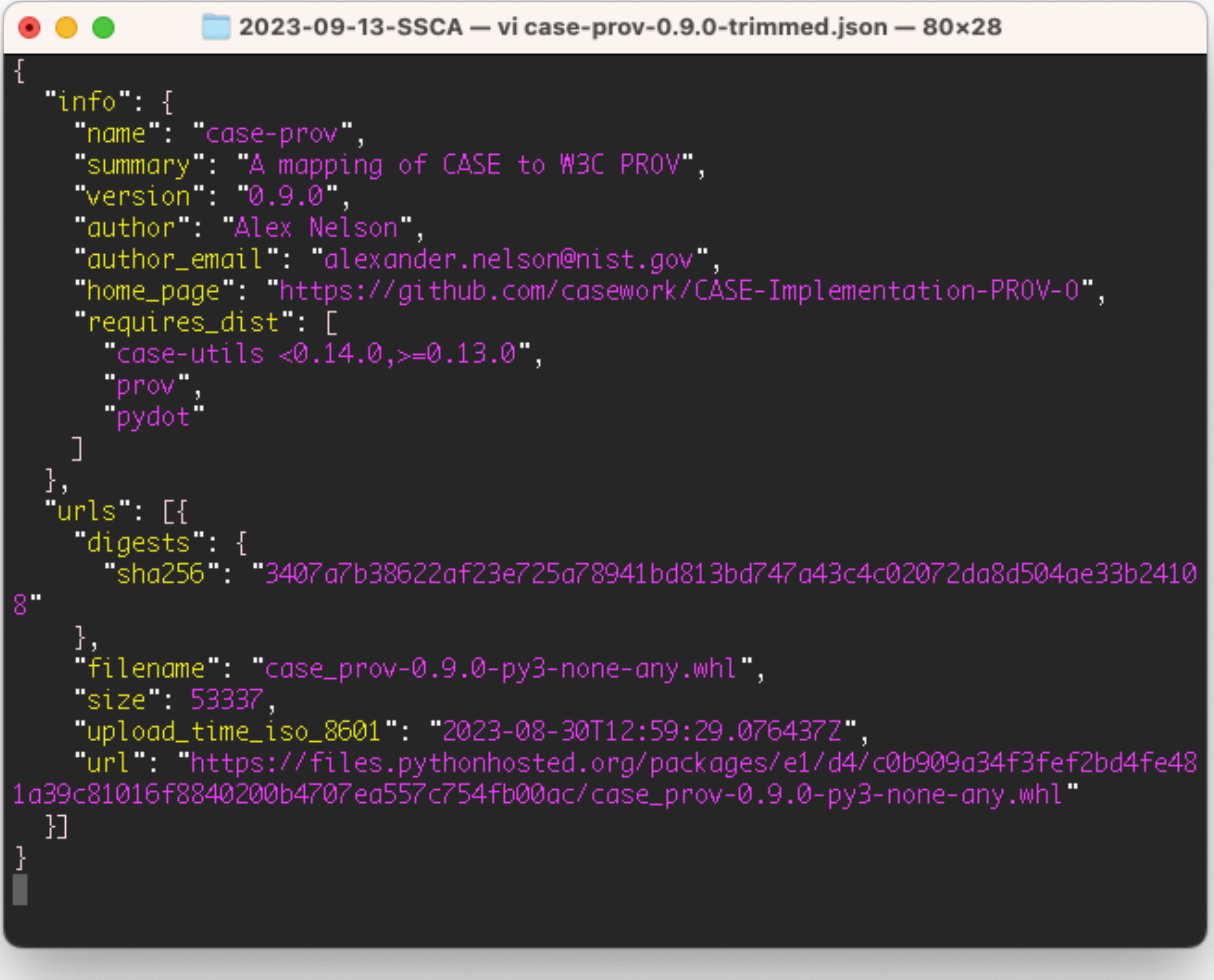
Provenance,

and the Supply Chain

Supply chain review is relationship analysis.

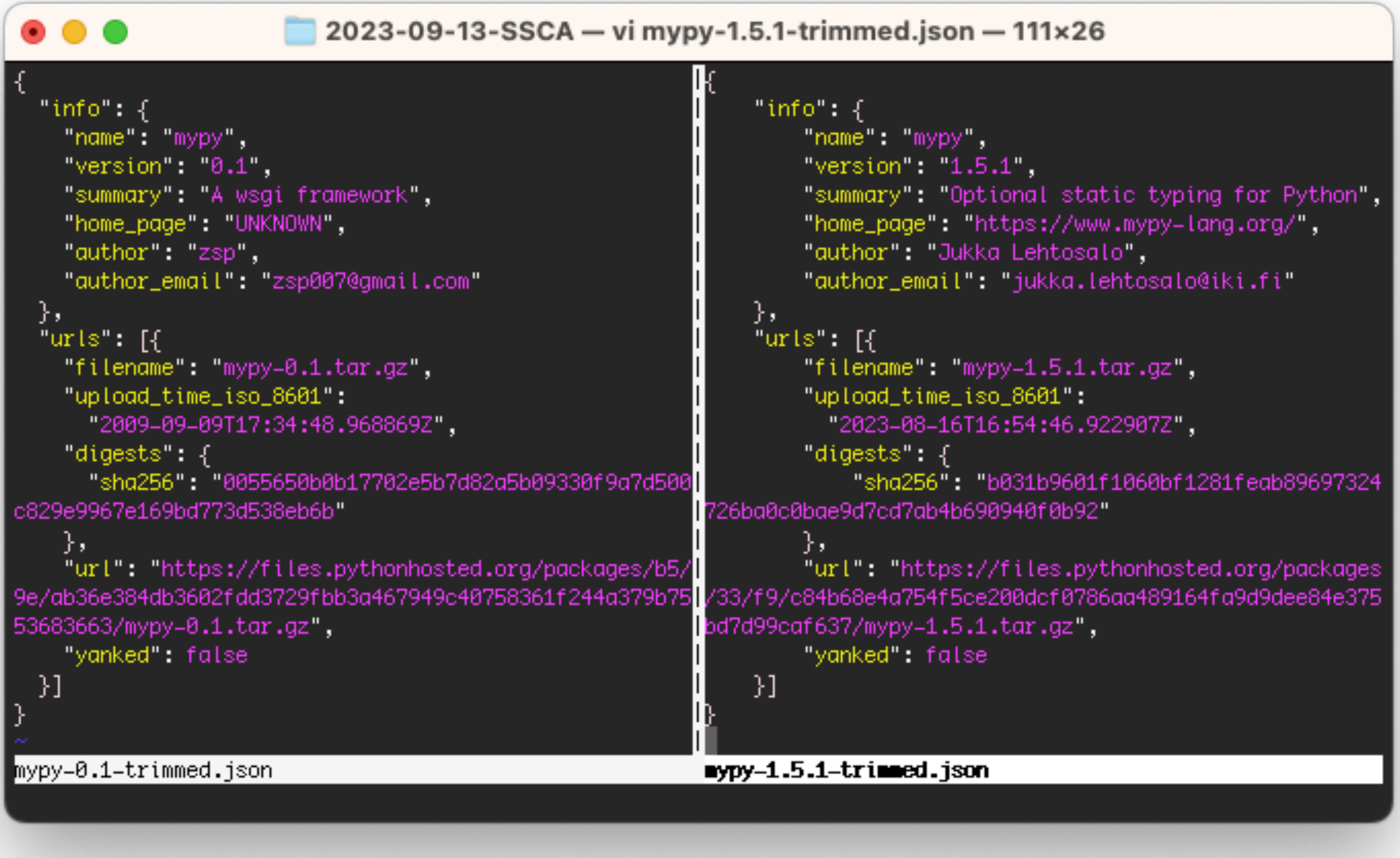
Package managers provide:

- Authorship information
- Package update discovery
- Dependency graphs
- Payload files
- Metadata feeds
  https://pypi.org/pypi/case-prov/0.9.0/json



```
{
  "info": {
    "name": "case-prov",
    "summary": "A mapping of CASE to W3C PROV",
    "version": "0.9.0",
    "author": "Alex Nelson",
    "author_email": "alexander.nelson@nist.gov",
    "home_page": "https://github.com/casework/CASE-Implementation-PROV-O",
    "requires_dist": [
      "case-utils <0.14.0,>=0.13.0",
      "prov",
      "pydot"
    ]
  },
  "urls": [{
    "digests": {
      "sha256": "3407a7b38622af23e725a78941bd813bd747a43c4c02072da8d504ae33b2410
8"
    },
    "filename": "case_prov-0.9.0-py3-none-any.whl",
    "size": 53337,
    "upload_time_iso_8601": "2023-08-30T12:59:29.076437Z",
    "url": "https://files.pythonhosted.org/packages/e1/d4/c0b909a34f3fef2bd4fe48
1a39c81016f8840200b4707ea557c754fb00ac/case_prov-0.9.0-py3-none-any.whl"
  }]
}
```

Are these the same project?



```
{
  "info": {
    "name": "mypy",
    "version": "0.1",
    "summary": "A wsgi framework",
    "home_page": "UNKNOWN",
    "author": "zsp",
    "author_email": "zsp007@gmail.com"
  },
  "urls": [{
    "filename": "mypy-0.1.tar.gz",
    "upload_time_iso_8601":
      "2009-09-09T17:34:48.968869Z",
    "digests": {
      "sha256": "0055650b0b17702e5b7d82a5b09330f9a7d500
c829e9967e169bd773d538eb6b"
    },
    "url": "https://files.pythonhosted.org/packages/b5/
9e/ab36e384db3602fdd3729fbb3a467949c40758361f244a379b75
53683663/mypy-0.1.tar.gz",
    "yanked": false
  }]
}
~
mypy-0.1-trimmed.json
```

```
{
    "info": {
        "name": "mypy",
        "version": "1.5.1",
        "summary": "Optional static typing for Python",
        "home_page": "https://www.mypy-lang.org/",
        "author": "Jukka Lehtosalo",
        "author_email": "jukka.lehtosalo@iki.fi"
    },
    "urls": [{
        "filename": "mypy-1.5.1.tar.gz",
        "upload_time_iso_8601":
          "2023-08-16T16:54:46.922907Z",
        "digests": {
            "sha256": "b031b9601f1060bf1281feab89697324
726ba0c0bae9d7cd7ab4b690940f0b92"
        },
        "url": "https://files.pythonhosted.org/packages
/33/f9/c84b68e4a754f5ce200dcf0786aa489164fa9d9dee84e375
bd7d99caf637/mypy-1.5.1.tar.gz",
        "yanked": false
    }]
}
mypy-1.5.1-trimmed.json
```

Two versions of the project parked at “mypy” on PyPI: The first (0.1), and today’s (1.5.1), 14 years apart.

- Summary is different.
- Home page now recorded.
- Author-role now held by someone else.
- Was never yanked (retracted).

Absent deep review of home page’s blog, we instead consider:

What are **properties**?

What are **qualities**?

What are **independent and related objects**?

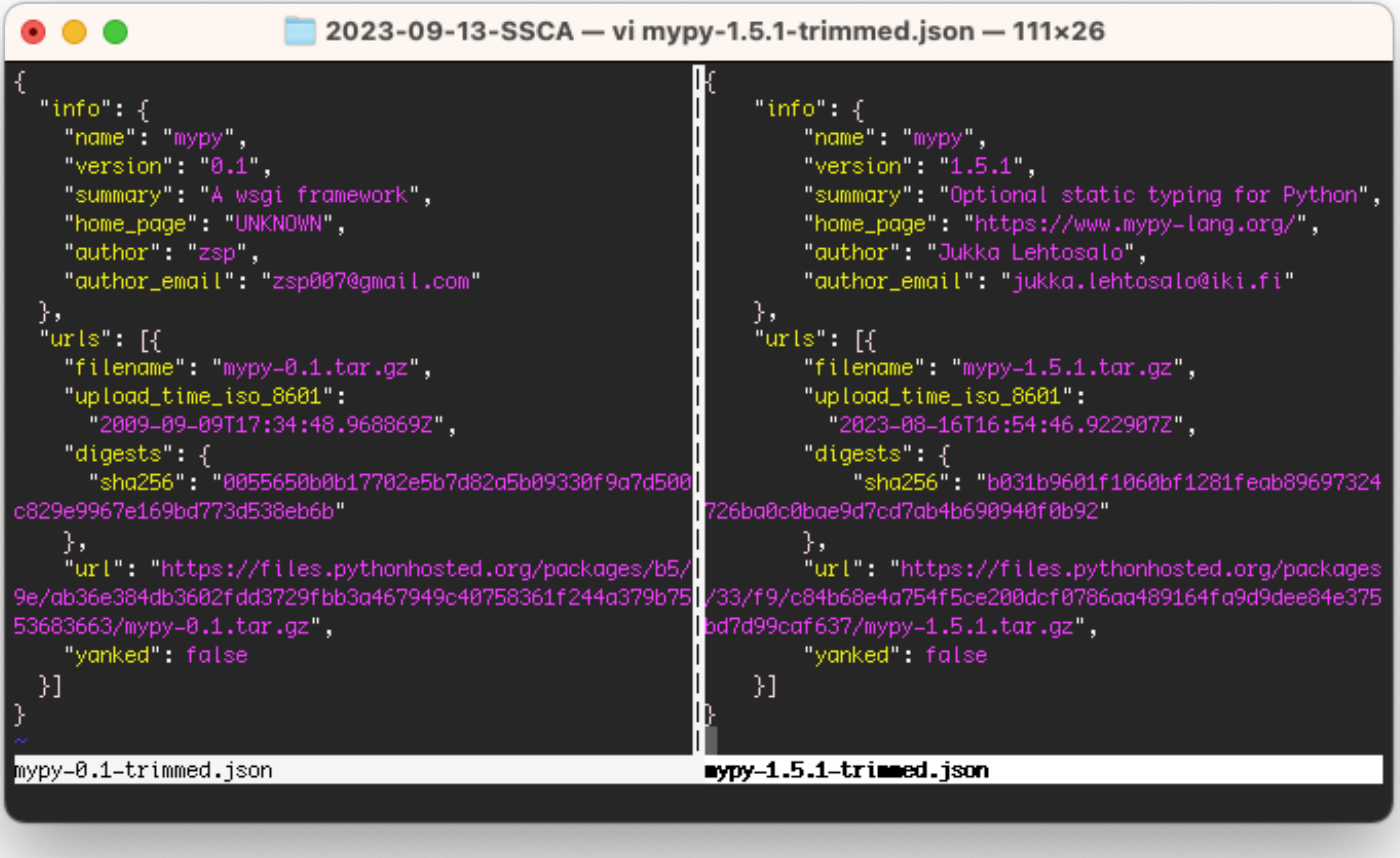
Graphs link objects in several ways.

Three ways to relate two objects, O1 and O2, are **properties**, **qualities**, and **relationships**.

(In some cases, O2 is a literal-data value, like a string or integer.)

- **Property** – The linked thing is fundamental to the identity of O1.
  E.g. A package in an ecosystem has a name as an identifier. Changing the name creates a new package.
- **Quality** – The linked thing is mutable.
  E.g. A package's download count does not change the identity of the package when it ticks up.
- **Relationship** – Neither O1 nor O2 need each other to exist. A relationship ties them together.
  E.g. a package's maintainer can change from release to release.
  - The relationship can end without inducing O1 or O2 to also end.

Are these the same project?



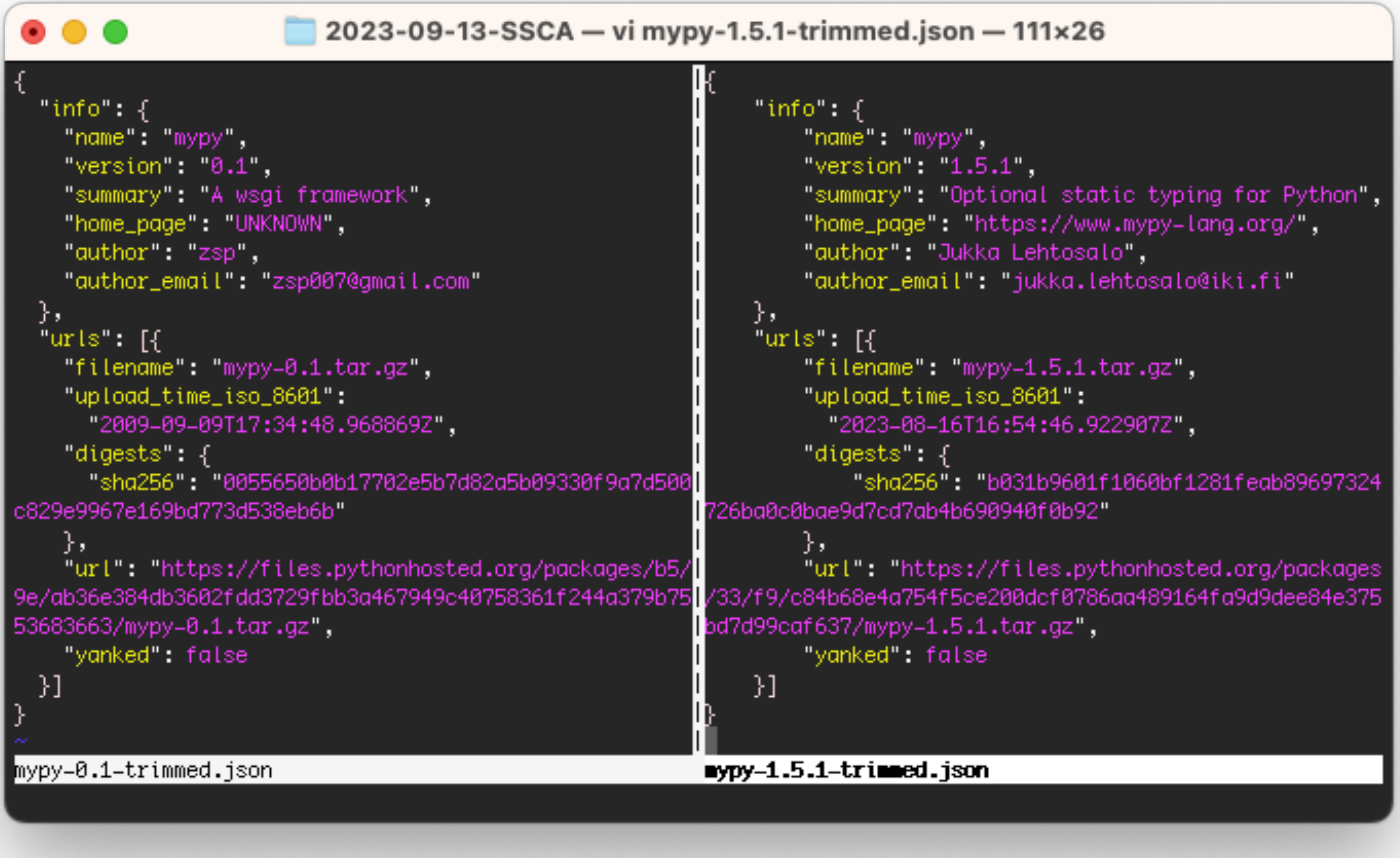
```
{
  "info": {
    "name": "mypy",
    "version": "0.1",
    "summary": "A wsgi framework",
    "home_page": "UNKNOWN",
    "author": "zsp",
    "author_email": "zsp007@gmail.com"
  },
  "urls": [{
    "filename": "mypy-0.1.tar.gz",
    "upload_time_iso_8601":
      "2009-09-09T17:34:48.968869Z",
    "digests": {
      "sha256": "0055650b0b17702e5b7d82a5b09330f9a7d500
c829e9967e169bd773d538eb6b"
    },
    "url": "https://files.pythonhosted.org/packages/b5/
9e/ab36e384db3602fdd3729fbb3a467949c40758361f244a379b75
53683663/mypy-0.1.tar.gz",
    "yanked": false
  }]
}
~
mypy-0.1-trimmed.json
```

```
{
    "info": {
        "name": "mypy",
        "version": "1.5.1",
        "summary": "Optional static typing for Python",
        "home_page": "https://www.mypy-lang.org/",
        "author": "Jukka Lehtosalo",
        "author_email": "jukka.lehtosalo@iki.fi"
    },
    "urls": [{
        "filename": "mypy-1.5.1.tar.gz",
        "upload_time_iso_8601":
          "2023-08-16T16:54:46.922907Z",
        "digests": {
            "sha256": "b031b9601f1060bf1281feab89697324
726ba0c0bae9d7cd7ab4b690940f0b92"
        },
        "url": "https://files.pythonhosted.org/packages
/33/f9/c84b68e4a754f5ce200dcf0786aa489164fa9d9dee84e375
bd7d99caf637/mypy-1.5.1.tar.gz",
        "yanked": false
    }]
}
mypy-1.5.1-trimmed.json
```

Two versions of the project parked at "mypy" on PyPI: The first (0.1), and today's (1.5.1), 14 years apart.

- Summary is different.
- Home page now recorded.
- Author-role now held by someone else.
- Was never yanked (retracted).

Are these the same project?



2023-09-13-SSCA — vi mypy-1.5.1-trimmed.json — 111×26

```
{
  "info": {
    "name": "mypy",
    "version": "0.1",
    "summary": "A wsgi framework",
    "home_page": "UNKNOWN",
    "author": "zsp",
    "author_email": "zsp007@gmail.com"
  },
  "urls": [{
    "filename": "mypy-0.1.tar.gz",
    "upload_time_iso_8601":
      "2009-09-09T17:34:48.968869Z",
    "digests": {
      "sha256": "0055650b0b17702e5b7d82a5b09330f9a7d500
c829e9967e169bd773d538eb6b"
    },
    "url": "https://files.pythonhosted.org/packages/b5/
9e/ab36e384db3602fdd3729fbb3a467949c40758361f244a379b75
53683663/mypy-0.1.tar.gz",
    "yanked": false
  }]
}
~
mypy-0.1-trimmed.json
```

```
{
        "info": {
                "name": "mypy",
                "version": "1.5.1",
                "summary": "Optional static typing for Python",
                "home_page": "https://www.mypy-lang.org/",
                "author": "Jukka Lehtosalo",
                "author_email": "jukka.lehtosalo@iki.fi"
        },
        "urls": [{
                "filename": "mypy-1.5.1.tar.gz",
                "upload_time_iso_8601":
                  "2023-08-16T16:54:46.922907Z",
                "digests": {
                        "sha256": "b031b9601f1060bf1281feab89697324
726ba0c0bae9d7cd7ab4b690940f0b92"
                },
                "url": "https://files.pythonhosted.org/packages
/33/f9/c84b68e4a754f5ce200dcf0786aa489164fa9d9dee84e375
bd7d99caf637/mypy-1.5.1.tar.gz",
                "yanked": false
        }]
}
mypy-1.5.1-trimmed.json
```

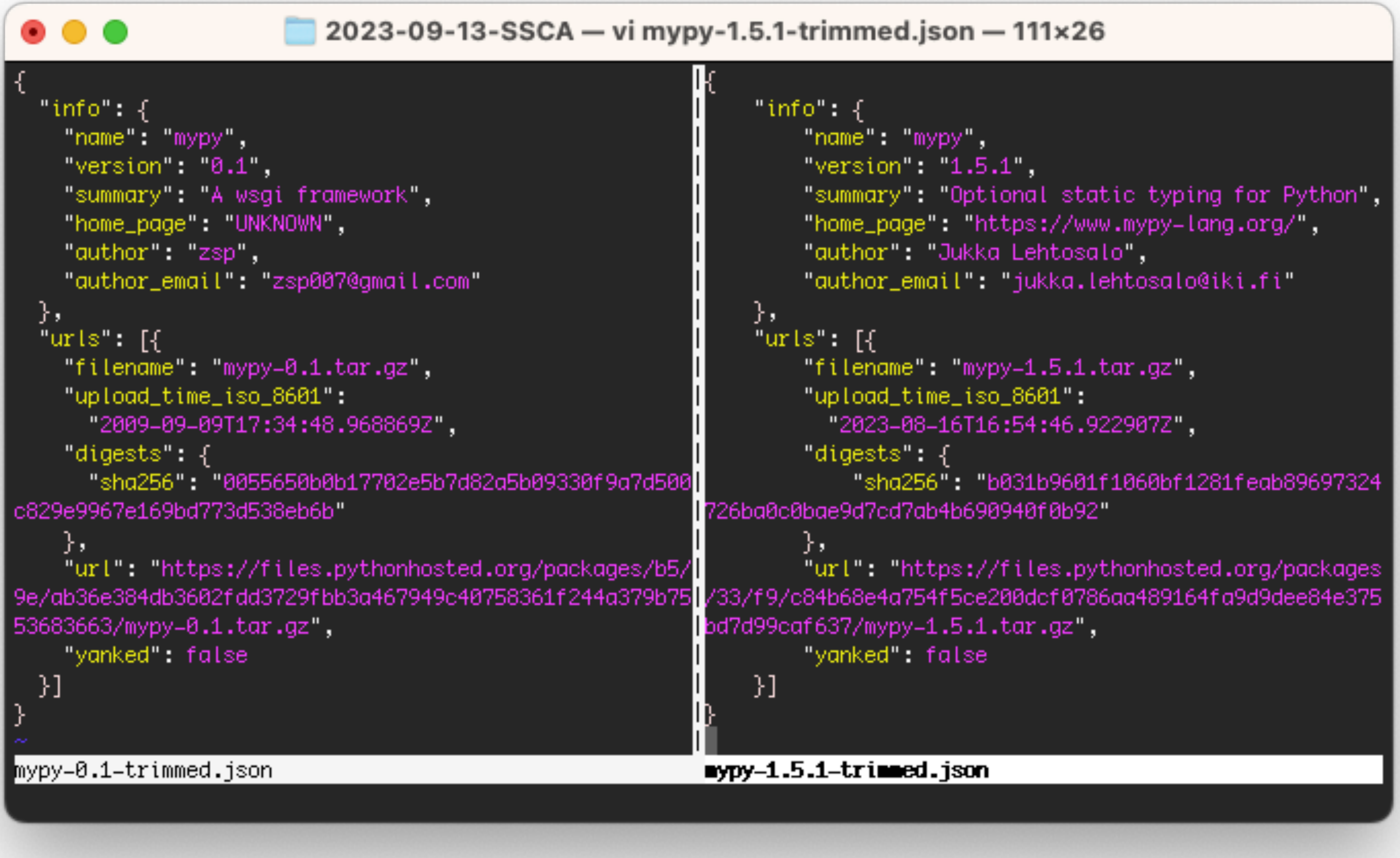
Two versions of the project parked at "mypy" on PyPI: The first (0.1), and today's (1.5.1), 14 years apart.

- Summary is different.
- Home page now recorded.
- Author-role now held by someone else.
- Was never yanked (retracted).

What are **properties**?

- Name

Are these the same project?



2023-09-13-SSCA — vi mypy-1.5.1-trimmed.json — 111×26

```
{
  "info": {
    "name": "mypy",
    "version": "0.1",
    "summary": "A wsgi framework",
    "home_page": "UNKNOWN",
    "author": "zsp",
    "author_email": "zsp007@gmail.com"
  },
  "urls": [{
    "filename": "mypy-0.1.tar.gz",
    "upload_time_iso_8601":
      "2009-09-09T17:34:48.968869Z",
    "digests": {
      "sha256": "0055650b0b17702e5b7d82a5b09330f9a7d500
c829e9967e169bd773d538eb6b"
    },
    "url": "https://files.pythonhosted.org/packages/b5/
9e/ab36e384db3602fdd3729fbb3a467949c40758361f244a379b75
53683663/mypy-0.1.tar.gz",
    "yanked": false
  }]
}
~
mypy-0.1-trimmed.json
```

```
{
        "info": {
                "name": "mypy",
                "version": "1.5.1",
                "summary": "Optional static typing for Python",
                "home_page": "https://www.mypy-lang.org/",
                "author": "Jukka Lehtosalo",
                "author_email": "jukka.lehtosalo@iki.fi"
        },
        "urls": [{
                "filename": "mypy-1.5.1.tar.gz",
                "upload_time_iso_8601":
                    "2023-08-16T16:54:46.922907Z",
                "digests": {
                        "sha256": "b031b9601f1060bf1281feab89697324
726ba0c0bae9d7cd7ab4b690940f0b92"
                },
                "url": "https://files.pythonhosted.org/packages
/33/f9/c84b68e4a754f5ce200dcf0786aa489164fa9d9dee84e375
bd7d99caf637/mypy-1.5.1.tar.gz",
                "yanked": false
        }]
}
mypy-1.5.1-trimmed.json
```

Two versions of the project parked at “mypy” on PyPI: The first (0.1), and today’s (1.5.1), 14 years apart.

- Summary is different.
- Home page now recorded.
- Author-role now held by someone else.
- Was never yanked (retracted).

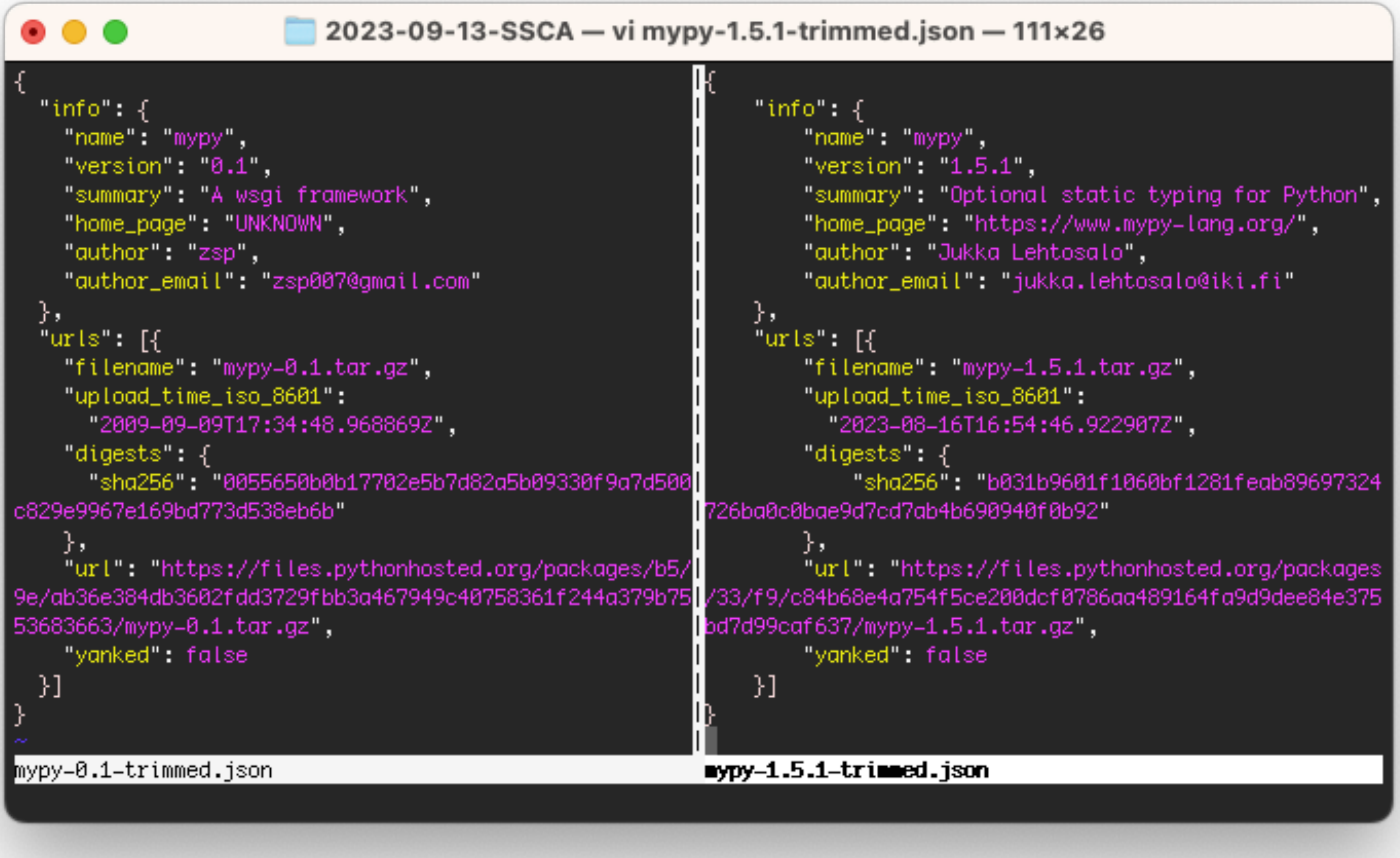
What are **properties**?

- Name

What are **qualities**?

- Version
- Summary

Are these the same project?



```
{
  "info": {
    "name": "mypy",
    "version": "0.1",
    "summary": "A wsgi framework",
    "home_page": "UNKNOWN",
    "author": "zsp",
    "author_email": "zsp007@gmail.com"
  },
  "urls": [{
    "filename": "mypy-0.1.tar.gz",
    "upload_time_iso_8601":
      "2009-09-09T17:34:48.968869Z",
    "digests": {
      "sha256": "0055650b0b17702e5b7d82a5b09330f9a7d500
c829e9967e169bd773d538eb6b"
    },
    "url": "https://files.pythonhosted.org/packages/b5/
9e/ab36e384db3602fdd3729fbb3a467949c40758361f244a379b75
53683663/mypy-0.1.tar.gz",
    "yanked": false
  }]
}
~
mypy-0.1-trimmed.json
```

```
{
    "info": {
        "name": "mypy",
        "version": "1.5.1",
        "summary": "Optional static typing for Python",
        "home_page": "https://www.mypy-lang.org/",
        "author": "Jukka Lehtosalo",
        "author_email": "jukka.lehtosalo@iki.fi"
    },
    "urls": [{
        "filename": "mypy-1.5.1.tar.gz",
        "upload_time_iso_8601":
          "2023-08-16T16:54:46.922907Z",
        "digests": {
            "sha256": "b031b9601f1060bf1281feab89697324
726ba0c0bae9d7cd7ab4b690940f0b92"
        },
        "url": "https://files.pythonhosted.org/packages
/33/f9/c84b68e4a754f5ce200dcf0786aa489164fa9d9dee84e375
bd7d99caf637/mypy-1.5.1.tar.gz",
        "yanked": false
    }]
}
mypy-1.5.1-trimmed.json
```

Two versions of the project parked at “mypy” on PyPI: The first (0.1), and today’s (1.5.1), 14 years apart.

- Summary is different.
- Home page now recorded.
- Author-role now held by someone else.
- Was never yanked (retracted).

What are **properties**?

- Name

What are **qualities**?

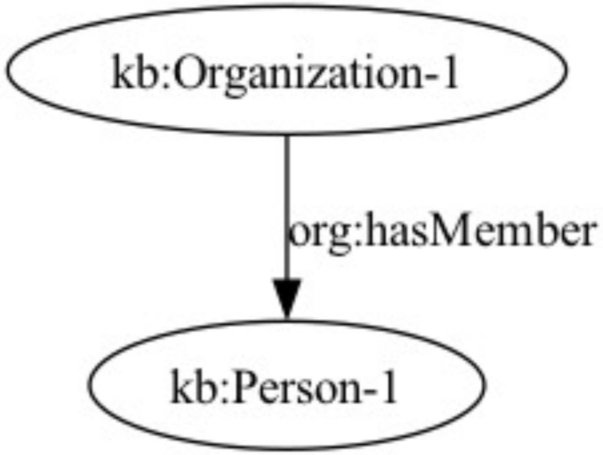
- Version
- Summary

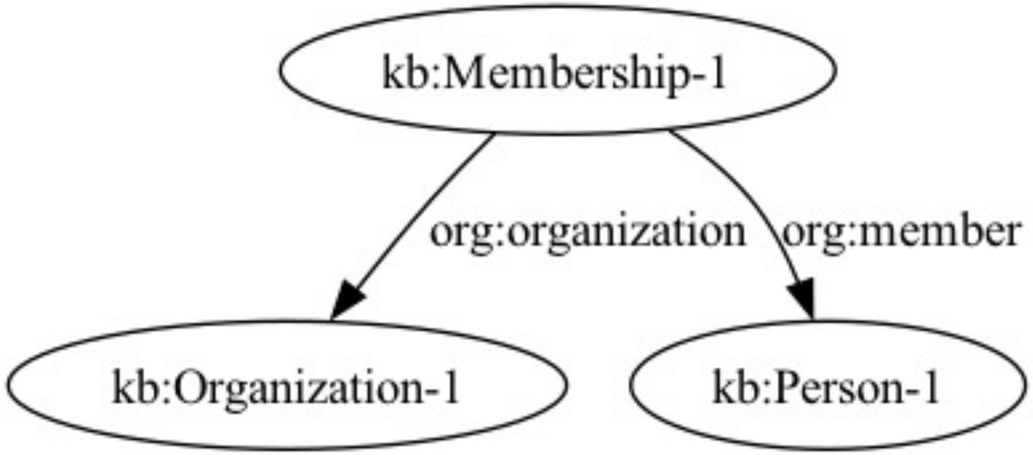
What are **independent and related**?

- Person in author role
- Home page

Example:
W3C ORG demonstrates two linking styles

Compare:





Flat (implemented with Property)	**Reified (implemented with Relationship)**
How are these maintained when faced with new facts? (E.g. Person-1 no longer in org.)	
Remove the outdated statement.	*Declare there exists an end* of the Membership.
What is the influence of time on the questions you can ask?	
Is Person-1 in the org? (Implicit: *Right now*.)	Was Person-1 ever in the org? Was Person-1 in the org last year?

When reviewing deployed software configurations, time information is essential.

A detour on time, for consistency review

The W3C's *OWL-Time* is an OWL-based ontology.

Defines Intervals, Instants, interval-relating algebra (right), plus more.

Timeline *consistency review* can use interval predicates, such as `time:intervalDuring`.

Example: All actions requiring a PKI signature SHOULD take place during the certificate's interval of validity.

Else, the knowledge base is *inconsistent*.

Some predicates make strong implications: "*i* before *j*" means *i* has a definite end, even if the specific timestamp is not known.

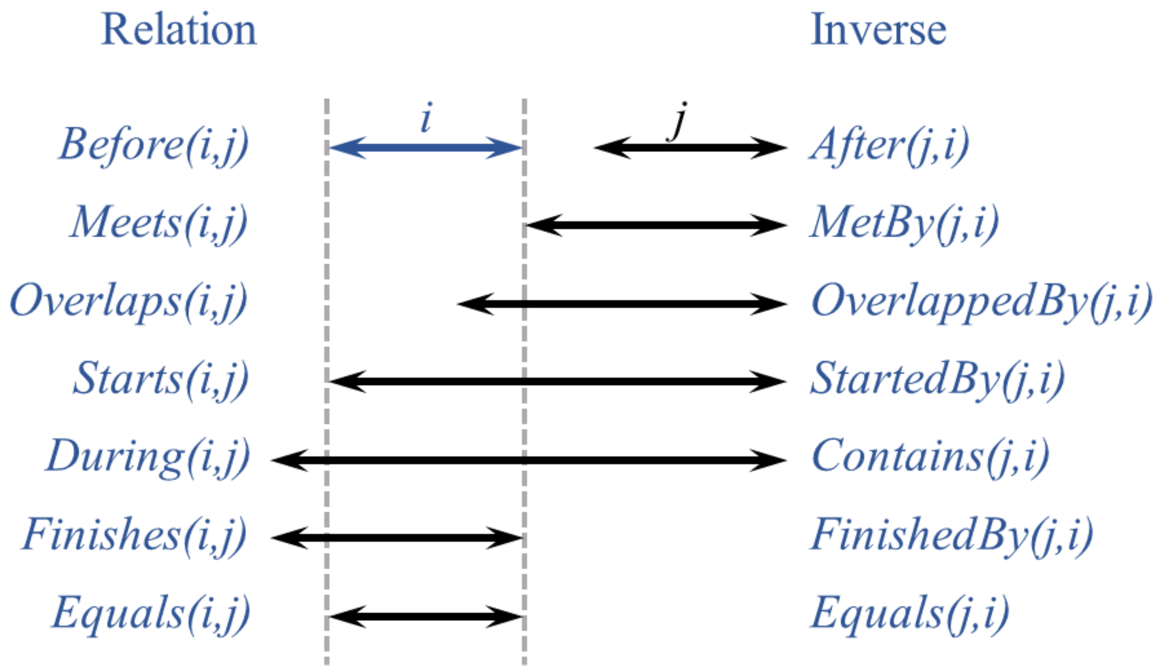


Figure source: http://dx.doi.org/10.1007/978-0-585-28322-7_7, via Figure 2 of https://www.w3.org/TR/owl-time/

A detour on provenance, for history description

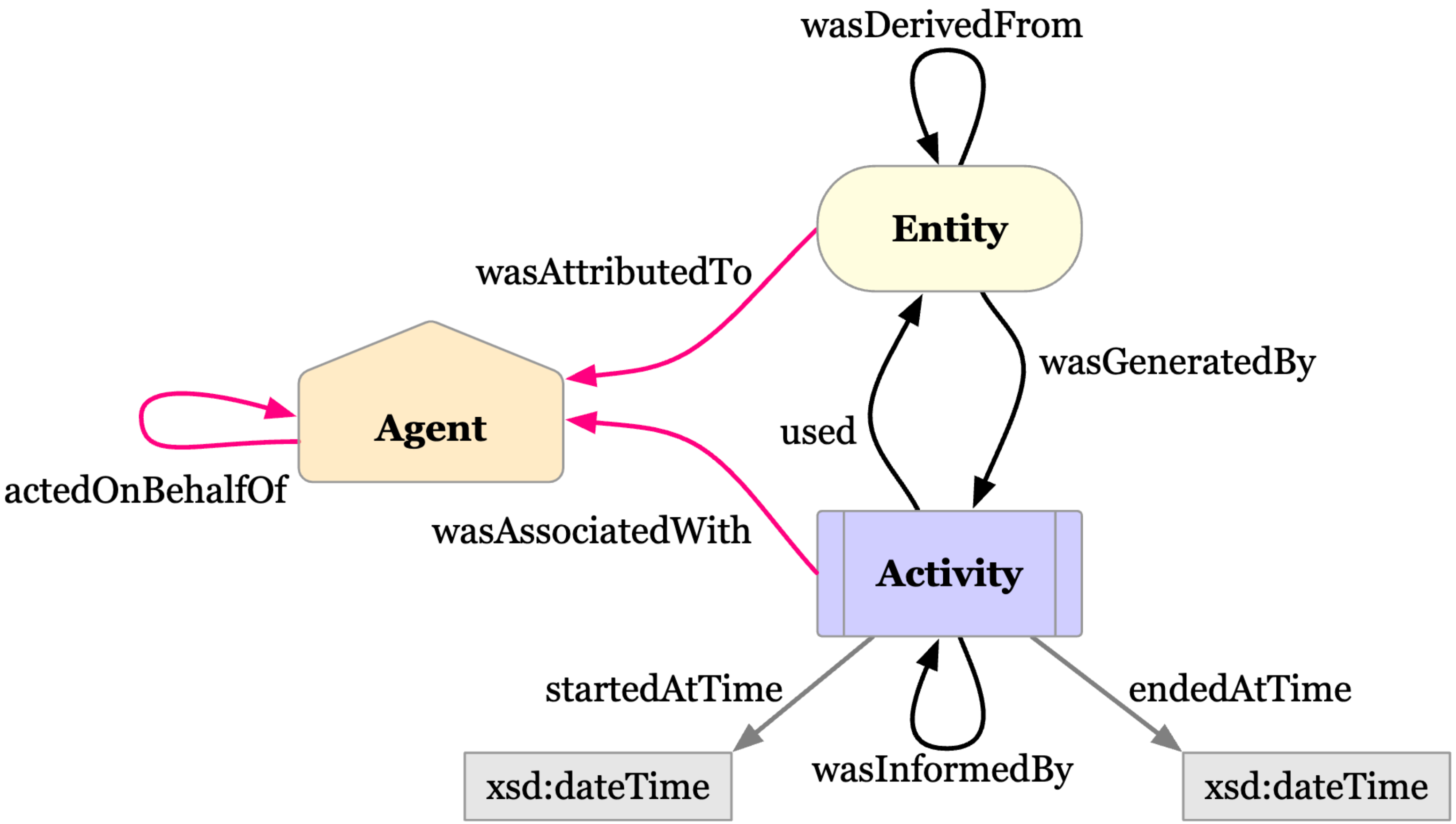
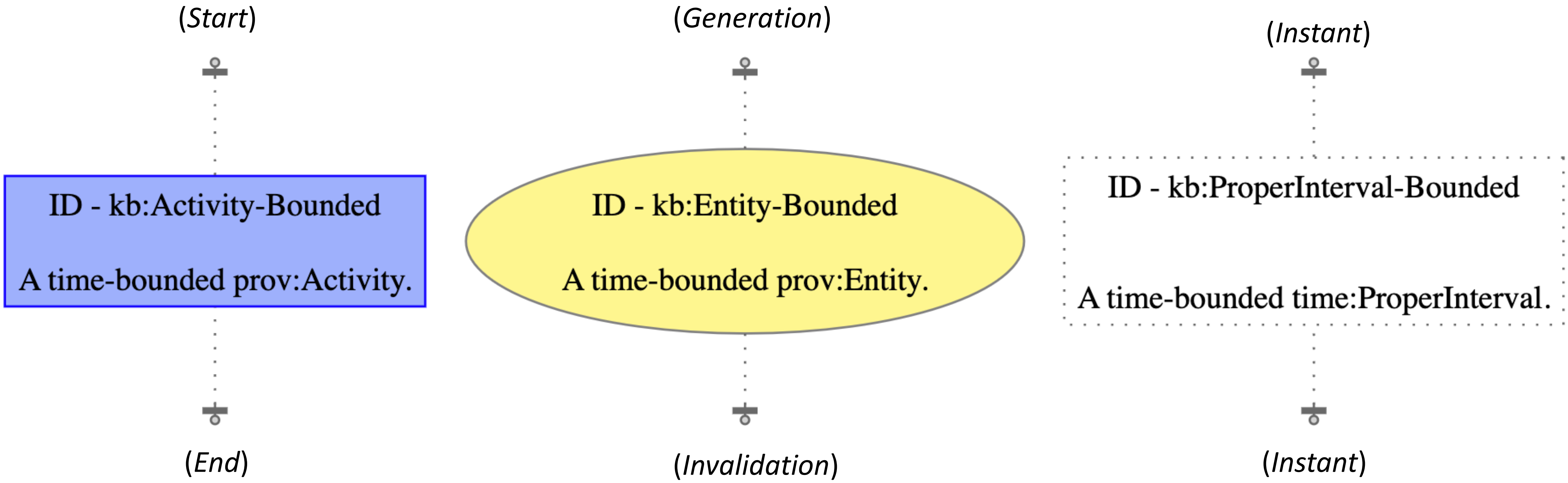


Figure 1. The three Starting Point classes and the properties that relate them.
The diagrams in this document depict Entities as yellow ovals,
Activities as blue rectangles, and Agents as orange pentagons.
The responsibility properties are shown in pink.

https://www.w3.org/TR/prov-o/

PROV concepts can align with OWL-Time

OWL-Time defines instants and intervals. PROV-O specializes these.



(*Start*)

ID - kb:Activity-Bounded

A time-bounded prov:Activity.

(*End*)

(*Generation*)

ID - kb:Entity-Bounded

A time-bounded prov:Entity.

(*Invalidation*)

(*Instant*)

ID - kb:ProperInterval-Bounded

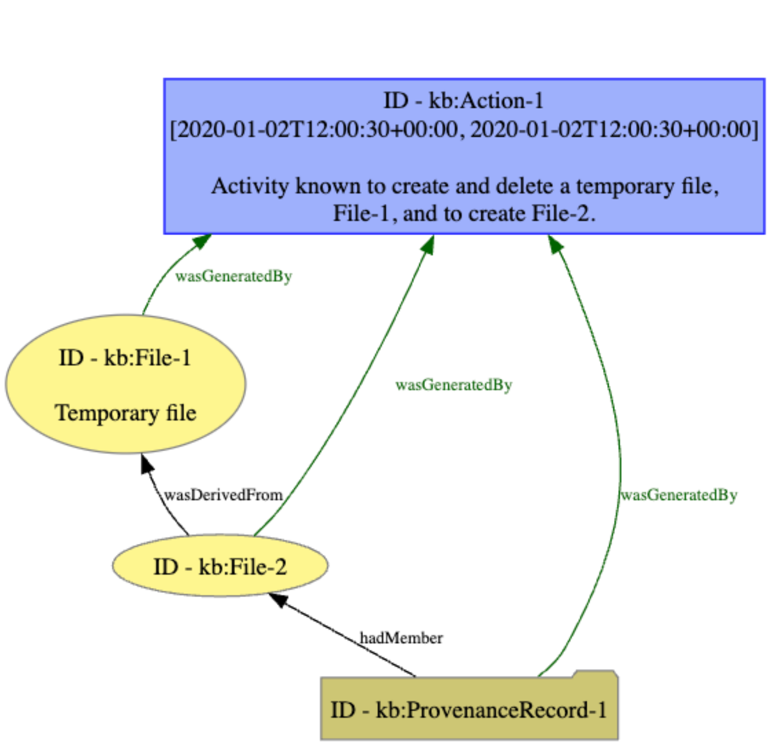
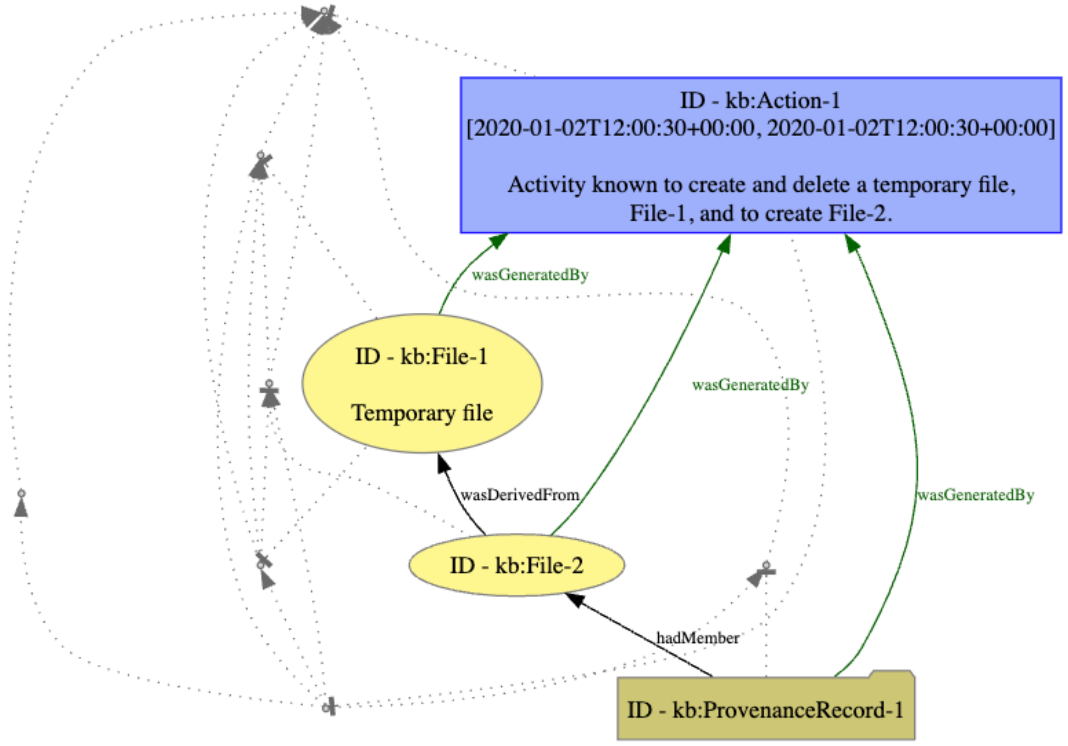
A time-bounded time:ProperInterval.

(*Instant*)

Figure source: `case-prov`'s README

Provenance analysis uses links and/or time

(Left and right displays only toggle time object visibility.)

Default display	Display with time links
	

ID - kb:Action-1
[2020-01-02T12:00:30+00:00, 2020-01-02T12:00:30+00:00]
Activity known to create and delete a temporary file, File-1, and to create File-2.
wasGeneratedBy
ID - kb:File-1
Temporary file
wasGeneratedBy
wasDerivedFrom
wasGeneratedBy
ID - kb:File-2
hadMember
ID - kb:ProvenanceRecord-1

ID - kb:Action-1
[2020-01-02T12:00:30+00:00, 2020-01-02T12:00:30+00:00]
Activity known to create and delete a temporary file, File-1, and to create File-2.
wasGeneratedBy
ID - kb:File-1
Temporary file
wasGeneratedBy
wasDerivedFrom
wasGeneratedBy
ID - kb:File-2
hadMember
ID - kb:ProvenanceRecord-1

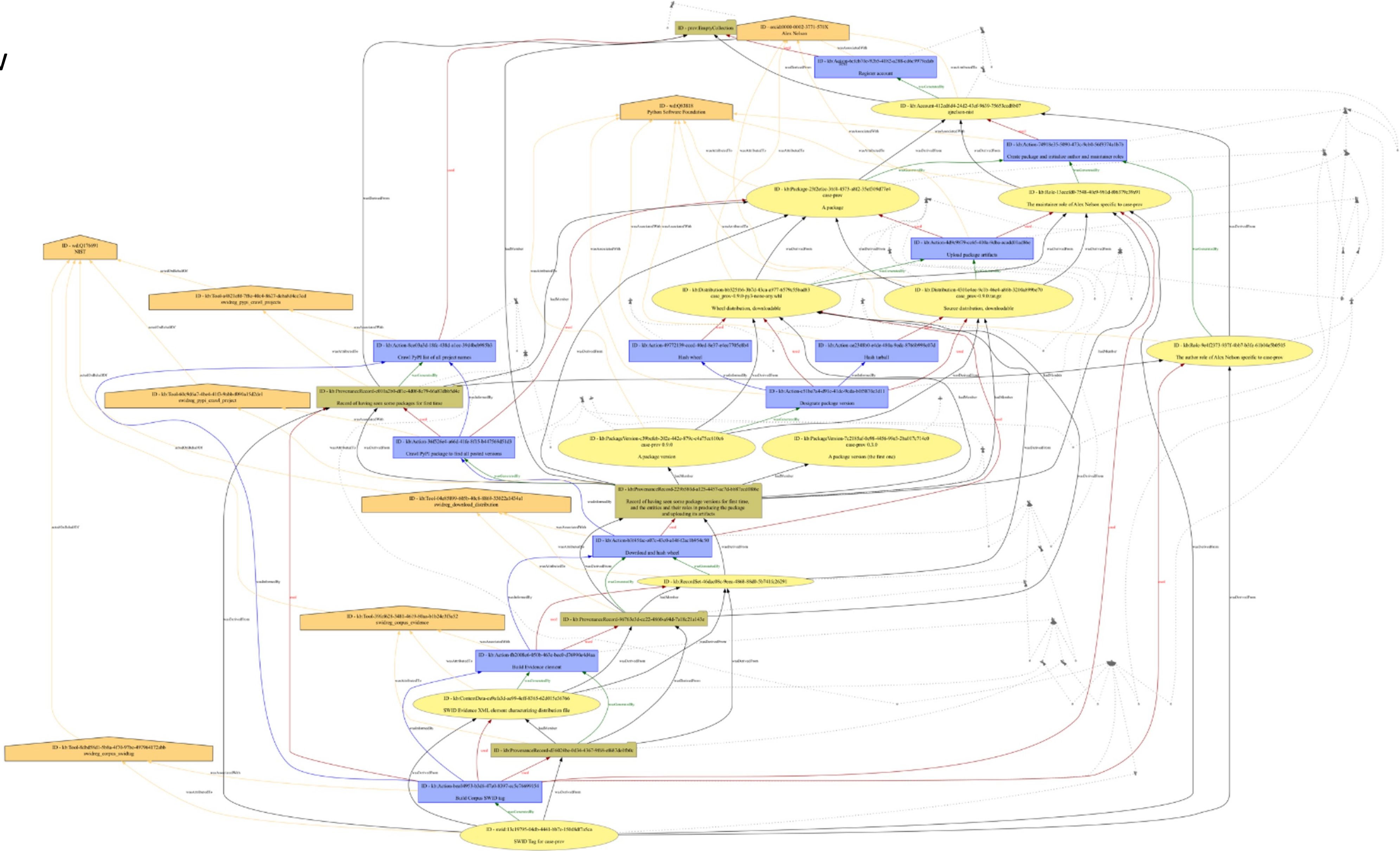
Figure source: `case-prov`’s README

Provenance example: History of swid-reg generated tag for case-prov@0.9.0

Provenance graphs show interwoven chains of:

- *Derivation*:
  entities from entities
  (yellow)
- *Communication*:
  Activities sharing
  entities
  (blue)
- *Delegation*:
  Agents acting on
  behalf of agents
  (orange)

Time flows **downward**.



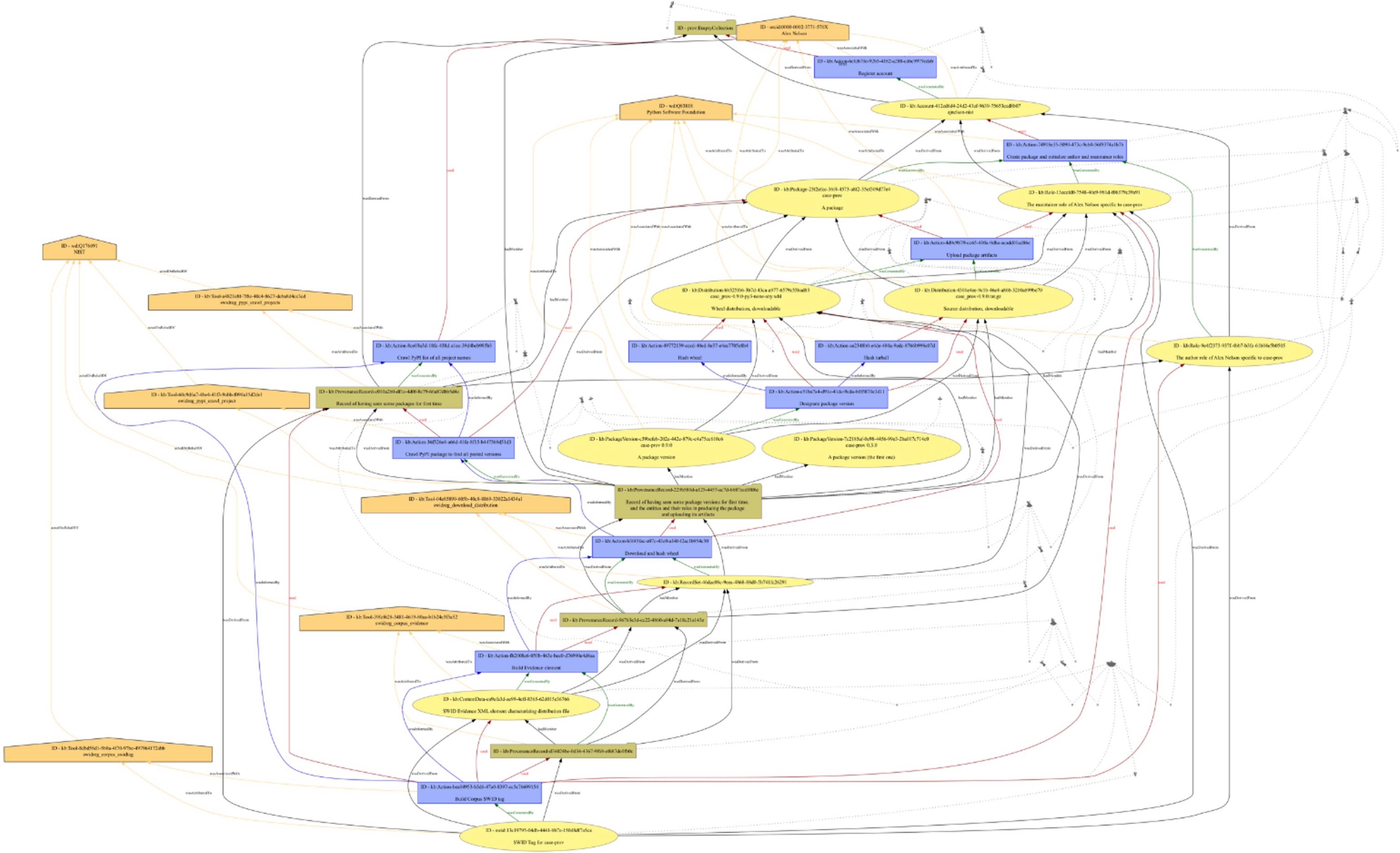
ID - prov:EmptyCollection
ID - orcid:0000-0002-3771-570X
Alex Nelson
ID - kb:Action-bc1cb70e-92b5-4182-a288-cd6c9979cdab
Register account
ID - wd:Q83818
Python Software Foundation
ID - kb:Account-412ed6d4-24d2-43cf-9639-75653ccd8b07
ajnelson-nist
ID - kb:Action-74918e35-5090-473c-9cb0-56f9374a1b7b
Create package and initialize author and maintainer roles
ID - kb:Package-23f2efce-3618-4573-a8f2-35ef319d77e4
case-prov
A package
ID - kb:Role-13eccfd0-7548-40e9-981d-f06379c39a91
The maintainer role of Alex Nelson specific to case-prov
ID - kb:Action-4d9f9679-cc65-410a-9dba-acadd01ad86e
Upload package artifacts
ID - wd:Q176691
NIST
ID - kb:Tool-a4821e8f-7f8e-48e4-8627-deba8d4ec3ed
swidreg_pypi_crawl_projects
ID - kb:Distribution-b0325f66-3b7d-43ca-a977-6579c55bad83
case_prov-0.9.0-py3-none-any.whl
Wheel distribution, downloadable
ID - kb:Distribution-4310e4ee-9c1b-46e4-a86b-3210a899be70
case_prov-0.9.0.tar.gz
Source distribution, downloadable
ID - kb:Action-8ce03a3d-18fc-438d-a1ee-39d4beb985b3
Crawl PyPI list of all project names
ID - kb:Action-49772139-eced-40ed-8e37-e6ee7705e8b4
Hash wheel
ID - kb:Action-ae2348b0-e4de-484a-9edc-8766b998c07d
Hash tarball
ID - kb:Role-9e4f2373-937f-4bb7-b3fa-61b04e5b0505
The author role of Alex Nelson specific to case-prov
ID - kb:Tool-60c9d6a7-4be4-41f3-9abb-f090a15d2de1
swidreg_pypi_crawl_project
ID - kb:ProvenanceRecord-c011a2b0-df1c-4d08-8c79-66a87d8b5d4c
Record of having seen some packages for first time
ID - kb:Action-c51ba7a4-d91c-41de-9cda-b0f5870e3d11
Designate package version
ID - kb:Action-36f526e4-a66d-41fa-8f15-b4475694d51d3
Crawl PyPI package to find all posted versions
ID - kb:PackageVersion-c39befeb-202e-442e-879c-c4a75cc610c6
case-prov 0.9.0
A package version
ID - kb:PackageVersion-7c2185af-8e98-4456-99e3-2ba017c714c0
case-prov 0.3.0
A package version (the first one)
ID - kb:Tool-04e85899-685b-40c8-8868-33022a1434a1
swidreg_download_distribution
ID - kb:ProvenanceRecord-229b580d-a125-4457-ac7d-b687ccd080be
Record of having seen some package versions for first time, and the entities and their roles in producing the package and uploading its artifacts
ID - kb:Action-b3f45fac-a87c-43c0-a146-f2ac1b954c50
Download and hash wheel
ID - kb:RecordSet-46dac08c-9eea-4868-88d0-5b7415c26291
ID - kb:Tool-39fc0628-3481-4619-80aa-b1b24e3f5a52
swidreg_corpus_evidence
ID - kb:ProvenanceRecord-96763e3d-ec22-4860-a94d-7a18c21a143e
ID - kb:Action-fb2008e6-050b-463e-bec0-d76990e4d4aa
Build Evidence element
ID - kb:ContentData-ca9a1a3d-ae99-4eff-8315-62d015e36766
SWID Evidence XML element characterizing distribution file
ID - kb:ProvenanceRecord-d16024be-0d34-4367-9f68-e8683de0fb0c
ID - kb:Tool-8dbd58d1-5b8a-4f30-97bc-4979641 72abb
swidreg_corpus_swidtag
ID - kb:Action-bea04953-b3d8-47a0-8397-ec5e76699154
Build Corpus SWID tag
ID - swid:13c19795-04db-4441-bb7c-150d8df7a5ca
SWID Tag for case-prov

Swid-reg separates ecosystem's posting history from crawler's observation provenance.

Upper-right:
History of case-prov, scoped to PyPI and the project's author.

Lower-left:
swid-reg actions observing, downloading, hashing artifacts.

Bottom:
The SWID tag for `case-prov@0.9.0`.



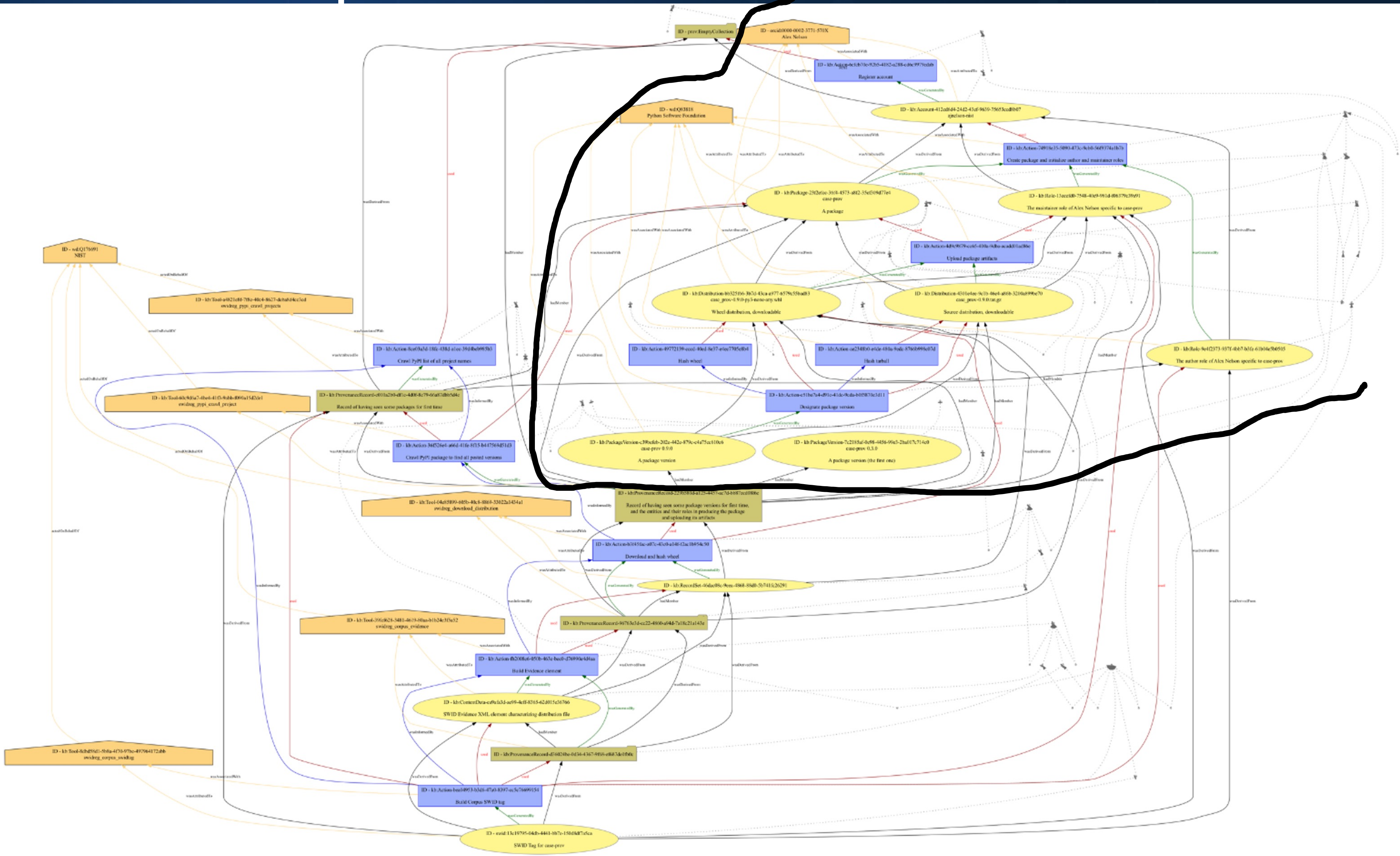
ID - prov:EmptyCollection
ID - orcid:0000-0002-3771-570X
Alex Nelson
Register account
ID - wd:Q83818
Python Software Foundation
ID - kb:Account-412ed6d4-2442-43ef-9639-75653ced6b07
ajnelson-nist
Create package and initialize author and maintainer roles
ID - kb:Package-23f2efee-3fc8-4573-a8f2-35ef309d77e4
case-prov
A package
The maintainer role of Alex Nelson specific to case-prov
Upload package artifacts
case_prov-0.9.0-py3-none-any.whl
Wheel distribution, downloadable
case_prov-0.9.0.tar.gz
Source distribution, downloadable
Hash wheel
Hash tarball
The author role of Alex Nelson specific to case-prov
Designate package version
ID - wd:Q176691
NIST
swidreg_pypi_crawl_projects
Crawl PyPI list of all project names
swidreg_pypi_crawl_project
Record of having seen some packages for first time
Crawl PyPI package to find all posted versions
case-prov 0.9.0
A package version
case-prov 0.3.0
A package version (the first one)
swidreg_download_distribution
Record of having seen some package versions for first time, and the entities and their roles in producing the package and uploading its artifacts
Download and hash wheel
swidreg_corpus_evidence
Build Evidence element
SWID Evidence XML element characterizing distribution file
swidreg_corpus_swidtag
Build Corpus SWID tag
SWID Tag for case-prov
wasAssociatedWith
wasAttributedTo
wasDerivedFrom
wasGeneratedBy
wasInformedBy
used
hadMember
actedOnBehalfOf

Swid-reg separates ecosystem’s posting history from crawler’s observation provenance.

Upper-right:
History of case-prov, scoped to PyPI and the project’s author.

Lower-left:
swid-reg actions observing, downloading, hashing artifacts.

Bottom:
The SWID tag for `case-prov@0.9.0`.



ID - prov:EmptyCollection
ID - orcid:0000-0002-3771-570X
Alex Nelson
Register account
ID - wd:Q83818
Python Software Foundation
ajnelson-nist
Create package and initialize author and maintainer roles
case-prov
A package
The maintainer role of Alex Nelson specific to case-prov
Upload package artifacts
case_prov-0.9.0-py3-none-any.whl
Wheel distribution, downloadable
case_prov-0.9.0.tar.gz
Source distribution, downloadable
Hash wheel
Hash tarball
The author role of Alex Nelson specific to case-prov
Designate package version
case-prov 0.9.0
A package version
case-prov 0.3.0
A package version (the first one)
ID - wd:Q176691
NIST
swidreg_pypi_crawl_projects
Crawl PyPI list of all project names
swidreg_pypi_crawl_project
Record of having seen some packages for first time
Crawl PyPI package to find all posted versions
swidreg_download_distribution
Record of having seen some package versions for first time, and the entities and their roles in producing the package and uploading its artifacts
Download and hash wheel
swidreg_corpus_evidence
Build Evidence element
SWID Evidence XML element characterizing distribution file
swidreg_corpus_swidtag
Build Corpus SWID tag
SWID Tag for case-prov

Provenance-oriented model enables flexible swid-reg augmentation of hashes.

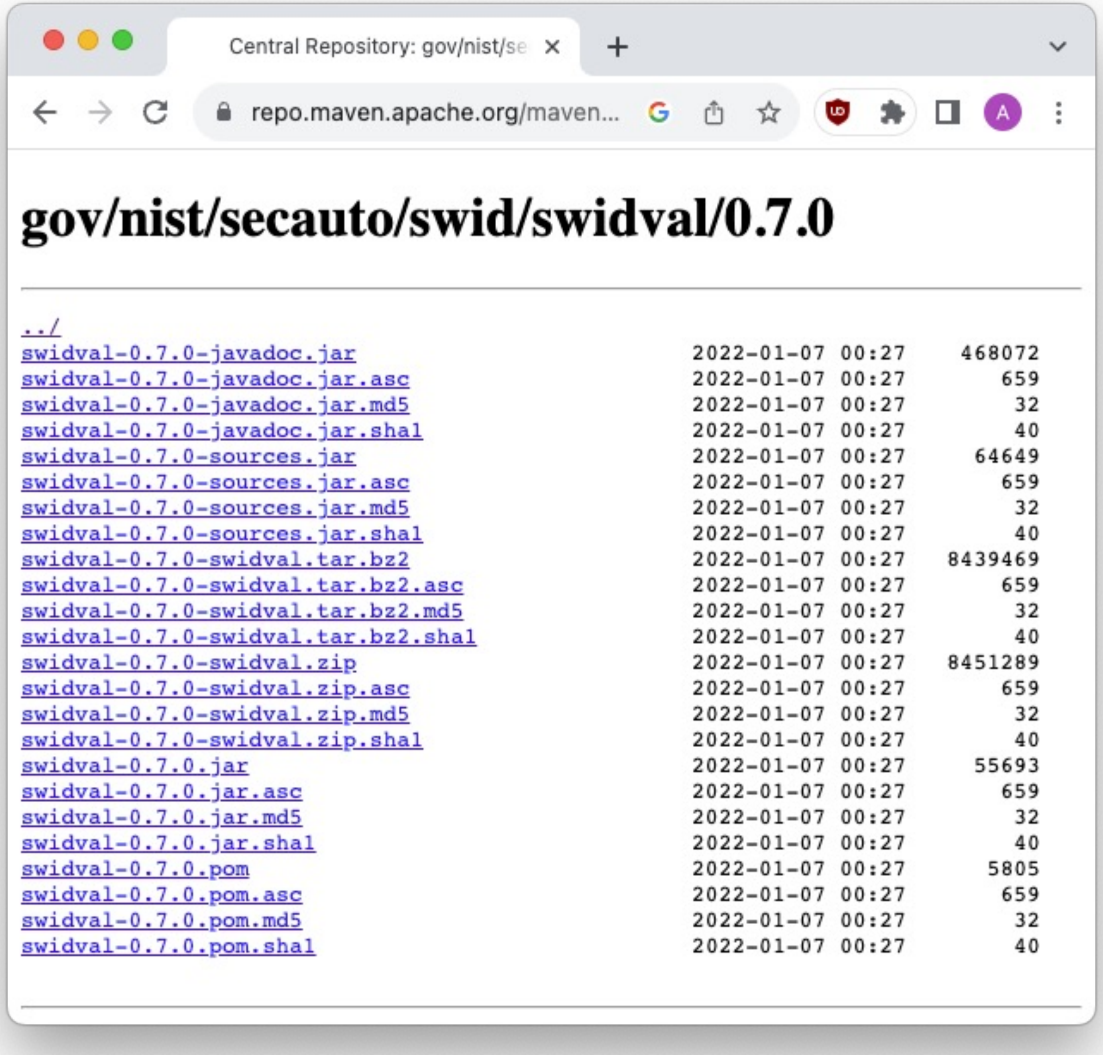
Compare PyPI’s JSON feed to Maven’s detached signature files.

swid-reg confirms provided file measurements (size, hashes), records as “attestation” from ecosystem.

Time of signatures’ observation is recorded, in case of later change.

Then, augments hashes to include:

- MD5*, SHA-1*
- SHA2-256 and -512
- SHA3-256 and -512
- File size
- Modification time of distribution file (if appears stable)


Central Repository: gov/nist/se
repo.maven.apache.org/maven...

gov/nist/secauto/swid/swidval/0.7.0

../		
swidval-0.7.0-javadoc.jar	2022-01-07 00:27	468072
swidval-0.7.0-javadoc.jar.asc	2022-01-07 00:27	659
swidval-0.7.0-javadoc.jar.md5	2022-01-07 00:27	32
swidval-0.7.0-javadoc.jar.sha1	2022-01-07 00:27	40
swidval-0.7.0-sources.jar	2022-01-07 00:27	64649
swidval-0.7.0-sources.jar.asc	2022-01-07 00:27	659
swidval-0.7.0-sources.jar.md5	2022-01-07 00:27	32
swidval-0.7.0-sources.jar.sha1	2022-01-07 00:27	40
swidval-0.7.0-swidval.tar.bz2	2022-01-07 00:27	8439469
swidval-0.7.0-swidval.tar.bz2.asc	2022-01-07 00:27	659
swidval-0.7.0-swidval.tar.bz2.md5	2022-01-07 00:27	32
swidval-0.7.0-swidval.tar.bz2.sha1	2022-01-07 00:27	40
swidval-0.7.0-swidval.zip	2022-01-07 00:27	8451289
swidval-0.7.0-swidval.zip.asc	2022-01-07 00:27	659
swidval-0.7.0-swidval.zip.md5	2022-01-07 00:27	32
swidval-0.7.0-swidval.zip.sha1	2022-01-07 00:27	40
swidval-0.7.0.jar	2022-01-07 00:27	55693
swidval-0.7.0.jar.asc	2022-01-07 00:27	659
swidval-0.7.0.jar.md5	2022-01-07 00:27	32
swidval-0.7.0.jar.sha1	2022-01-07 00:27	40
swidval-0.7.0.pom	2022-01-07 00:27	5805
swidval-0.7.0.pom.asc	2022-01-07 00:27	659
swidval-0.7.0.pom.md5	2022-01-07 00:27	32
swidval-0.7.0.pom.sha1	2022-01-07 00:27	40

Future work

- Augmentation of NVD vulnerability feeds with more than CPE
- Setting up feed for NIST-produced SWID tags
- Accepting submissions of SBOMs from partnering organizations to expand software knowledge base beyond open source ecosystems
- Researching “at-scale” association of Packages with Projects
- Better versioning:
  Using Git-based Projects’ histories to establish stronger partial-order package version graphs, improving “Affected versions” vulnerability associations

Questions?

https://github.com/usnistgov/swid-reg/

alexander.nelson@nist.gov