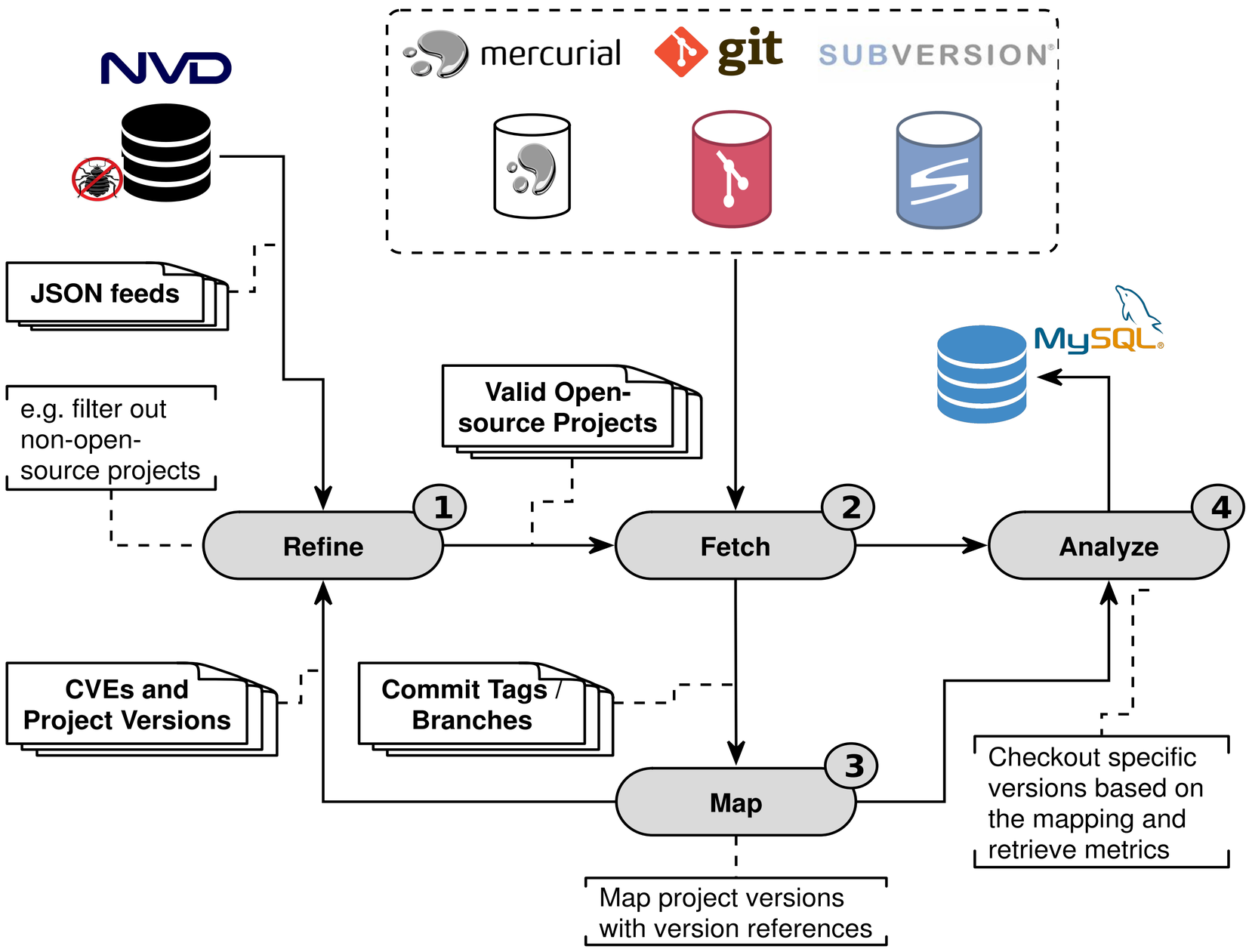
VulinOSS: A Dataset of Security Vulnerabilities in Open-source Systems

Dataset Construction Architecture

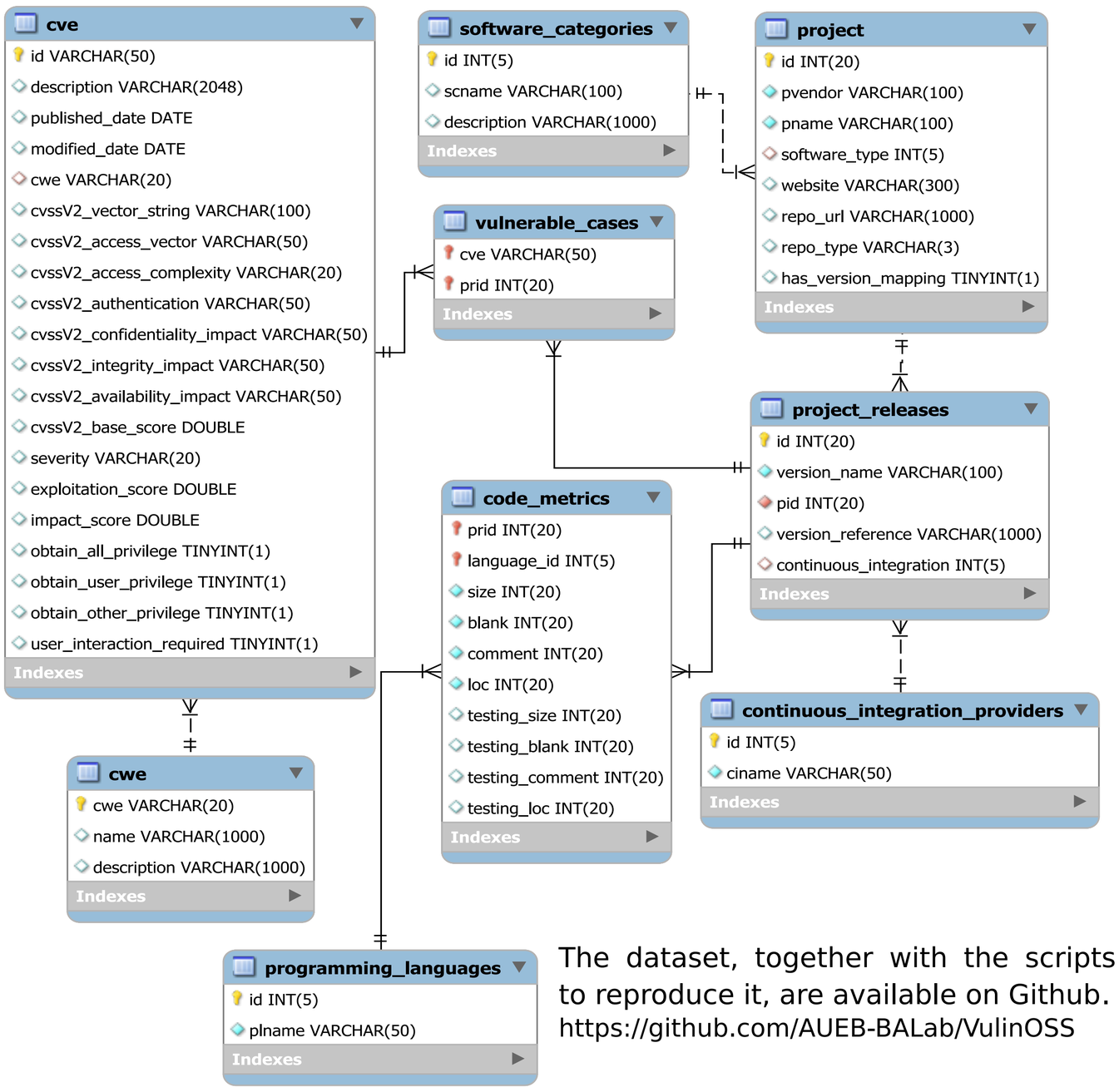


1. Parsed the NVD vulnerability reports and created a list of projects and their vulnerabilities. Manually inspected the list and selected the projects that are open source and have an available version control system.
2. Cloned locally the repositories from Git, Mercurial and Subversion that were selected at step 1.
3. Associated the project versions as reported in NVD with the version references (commit tags and branches) found in their version control system history.
4. With Python workers we checked out each version reference and retrieved code metrics and continuous integration evidence for this specific version. The findings are stored in a MySQL database.

Dataset Descriptive Statistics

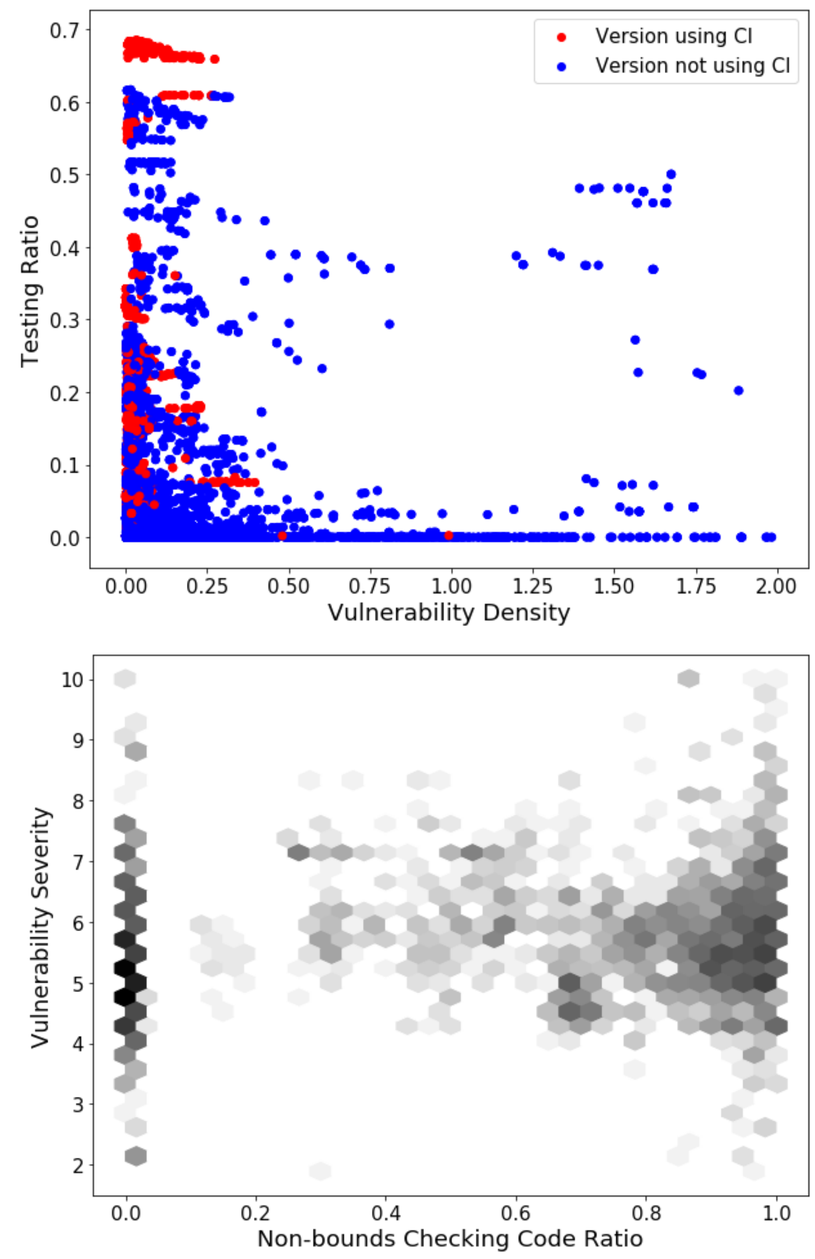
Measurement	Value
Projects	153
Project Versions	23884
Mapped Versions	8694
Number of Vulnerabilities	17738
Project Versions employing CI	1538

Database Schema



The dataset, together with the scripts to reproduce it, are available on Github.
https://github.com/AUEB-BALab/VulinOSS

Application Examples



Antonios Gkortzis, Dimitris Mitropoulos and Diomidis Spinellis
Athens University of Economics and Business, Greece
{antoniosgkortzis,dimitro,dds}@aueb.gr

*The research presented in this paper is supported by the European Union within the H2020 MSCA-ITN-EID Project “SENECA”.