SBOM利用ツールの評価に向けた JavaにおけるSBOMデータセットの構築

肥後研究室

M2 岸本 理央

Department of Computer Science, Graduate School of Information Science and Technology, Osaka University.
Software Engineering Laboratory. - https://sel.ist.osaka-u.ac.jp

ソフトウェアの管理の必要性

ライブラリが広く利用されている

- ソフトウェアの開発期間の短縮や開発費用の削減
- 推移的な依存も含めると多くのライブラリに依存

ソフトウェアの依存の管理は不十分である

- 脆弱性への対応の遅れや対応漏れ
- ライセンス違反の発生

依存関係を把握し，適切な管理を行うために
ソフトウェア部品表（SBOM）の利用が推奨される

ソフトウェア部品表（Software Bill of Materials, SBOM）

ソフトウェアの情報を記述したドキュメント

依存するライブラリ等を一覧して記述する

SBOMの記述形式とツールの必要性

JSONやXML等で
記述された機械可読なデータ

手作業での

- SBOMの作成
- SBOMの情報に基づいた
  ソフトウェアの管理作業

は手間がかかる

SBOMの利用を支援する
ツールが存在する

SPDX形式で記述されたSBOMの抜粋（1ライブラリの情報）

```
{
  "SPDXID": "SPDXRef-Package-4924B7C27DC9F6AF4D00 ... ",
  "name": "org.apache.logging.log4j.log4j-slf4j-impl",
  "versionInfo": "2.4.1",
  "packageFileName": "log4j-slf4j-impl-2.4.1.jar",
  "supplier": "Organization: Apache Software Foundation",
  "downloadLocation" : “https:// ... /log4j-slf4j-impl-2.4.1.jar",
  "filesAnalyzed": false,
  "checksums": [],
  "homepage": "http://logging.apache.org/log4j/2.x/",
  "licenseConcluded": "Apache-2.0",
  "licenseDeclared": "Apache-2.0",
  "copyrightText": "Copyright 1999-2014 Apache Software Foundation",
  "externalRefs": [
    {
      "referenceCategory": "PACKAGE-MANAGER",
      "referenceType": "purl",
      "referenceLocator": "pkg:maven/ ... /log4j-slf4j-impl@2.4.1"
    }
  ]
},
```

SBOMに関するツールの分類

SBOM生成ツール

- ソフトウェアの情報からSBOMを生成する
- 生成ツールの評価を行った研究は多く存在する[1, 2]

SBOM利用ツール

- SBOMを用いたソフトウェアの管理を支援する
  - 脆弱性やライセンス違反の検出
  - SBOMの閲覧や編集
- 利用ツールの評価を行った研究は存在しない

[1] On the Way to SBOMs: Investigating Design Issues and Solutions in Practice (Bi et al, 2024, TOSEM)
[2] Challenges of Producing Software Bill of Materials for Java (Balliu et al, 2023, IEEE Security & Privacy)

SBOMのデータセット

ソフトウェア工学の他分野ではデータセットが存在

- ツールやアルゴリズムの評価におけるベンチマーク
- 例：Semantic Clone Bench（コードクローン群）

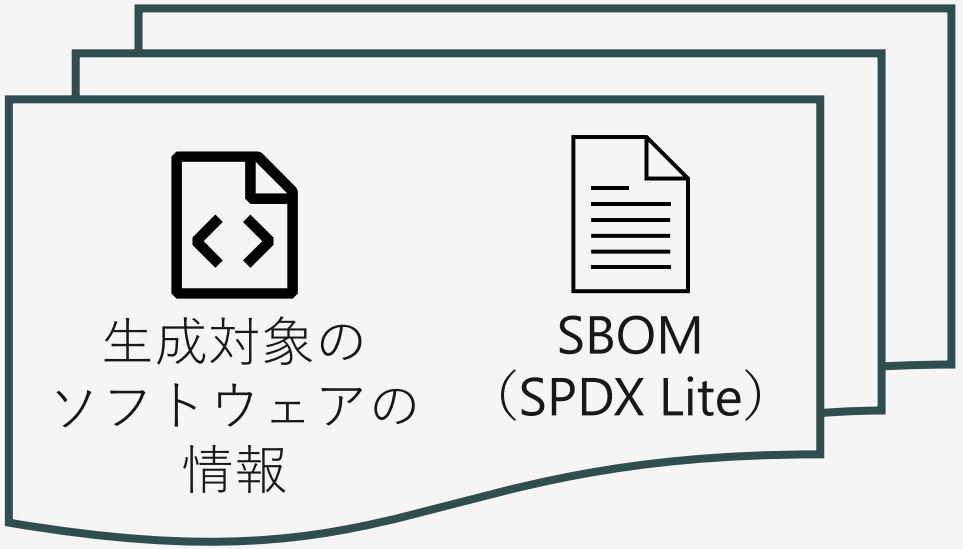
SBOM分野にはデータセットが存在しない

- SBOMのサンプルファイルは存在するが,
  データセットとして整理はされていない
- SBOM利用ツールの評価に使用できる
  品質が保証されたデータセットの存在が望ましい

作成するデータセットの概要

SBOM生成対象のソフトウェアの情報とSBOMの組の集まり

GitHub上のJavaプロジェクトを対象としてSBOMを作成する



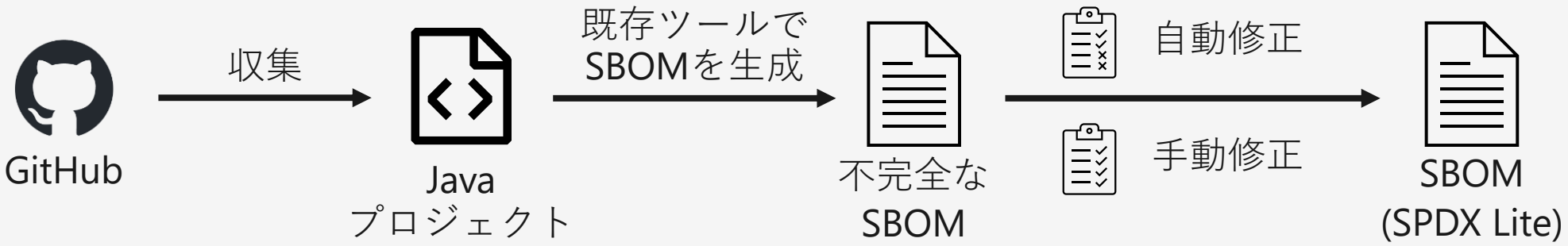
SBOMに記述する項目はSPDX Liteに従う

- SBOMの主要なフォーマットの1つである SPDXのサブ規格
- SBOMの主な用途に必要な項目のリスト （脆弱性診断とライセンス管理）

様々なツールの評価に利用可能な
SBOMを作成できると考えられる

データセットを構成するSBOMの作成手順

1. 作成対象のソフトウェアをGitHubから収集
2. 既存ツールを用いたSBOMの生成
3. 生成されたSBOMを確認して情報の修正・追加
    a. 自動修正
    b. 手動修正



SBOMの作成対象となるソフトウェアの条件

1. Javaで記述され，Mavenを使用している
2. Git上で条件を満たすタグが存在する
   名前にsnapshot，alpha，beta，rcを含まないタグ

→ 手順1で作成対象のソフトウェアを収集するときに確認

3. Maven以外によって管理される依存関係を持たない
4. Maven Central Repositoryに存在するライブラリにのみ依存する

→ 手順3で修正作業を行うときに確認

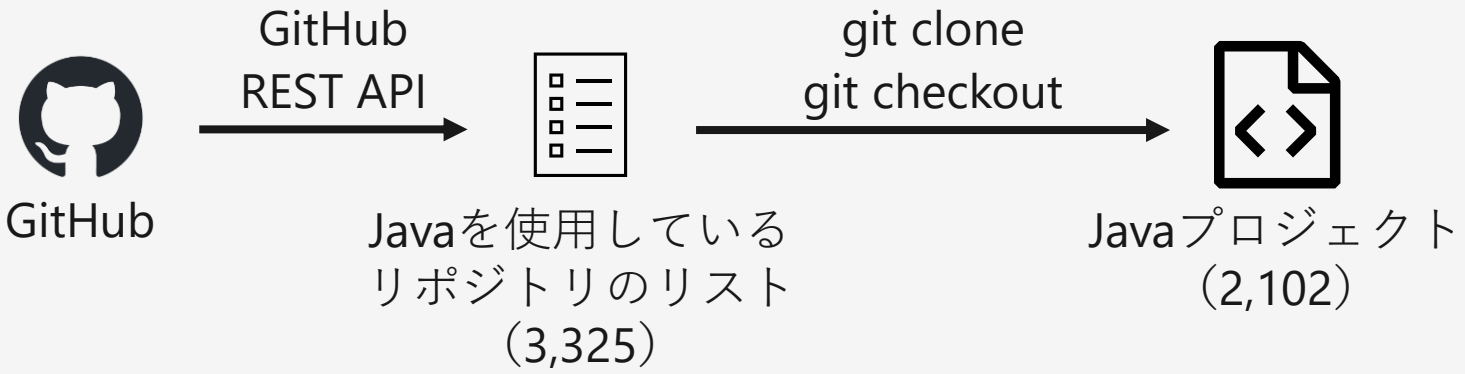
1. 生成対象のソフトウェアの収集

GitHubのREST APIを用いて条件を満たすリポジトリを取得

Javaを使用しており，Star数が50以上

リポジトリをクローンし，条件を満たすタグをチェックアウト

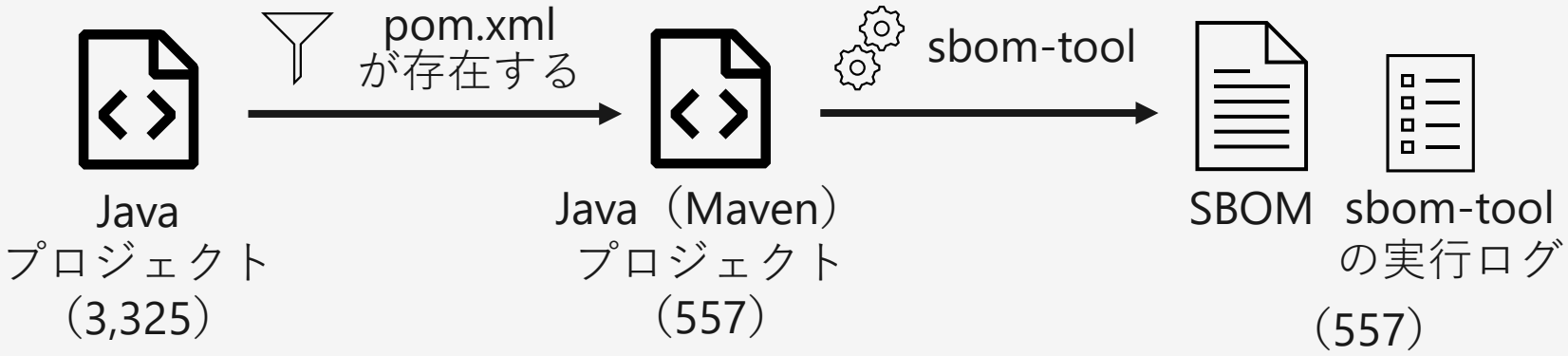
名前にsnapshot，alpha，beta，rcを含まないタグ



2. 既存ツールを用いたSBOMの生成

SBOMの生成には，sbom-toolを使用した

1. で収集したプロジェクトの内，Mavenを使用している557プロジェクトを対象にSBOMを生成した



3. SBOMの修正

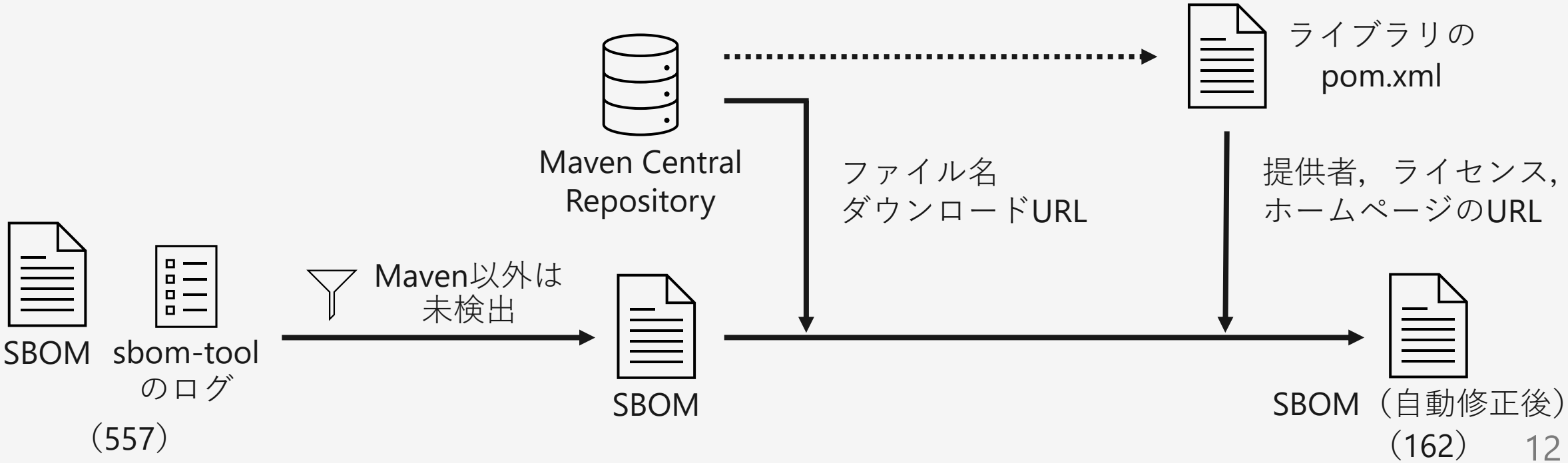
SPDX Liteに準拠したSBOMを作成するために,
sbom-toolが生成したSBOMに不足する情報を追加する

ライブラリの情報	SBOM（sbom-tool）	SBOM（自動修正後）	SBOM（SPDX Lite）
	✗ ファイル名	✓ ファイル名	✓ ファイル名
	✗ 提供者	✗ 提供者	✓ 提供者
	✗ ダウンロードURL	✓ ダウンロードURL	✓ ダウンロードURL
	✗ ホームページのURL	✗ ホームページのURL	✓ ホームページのURL
	✗ ライセンス	✗ ライセンス	✓ ライセンス

3.a SBOMの修正｜自動修正

Maven以外のパッケージ管理システム（npm，pipなど）の設定ファイルが検出されたプロジェクトは除外する

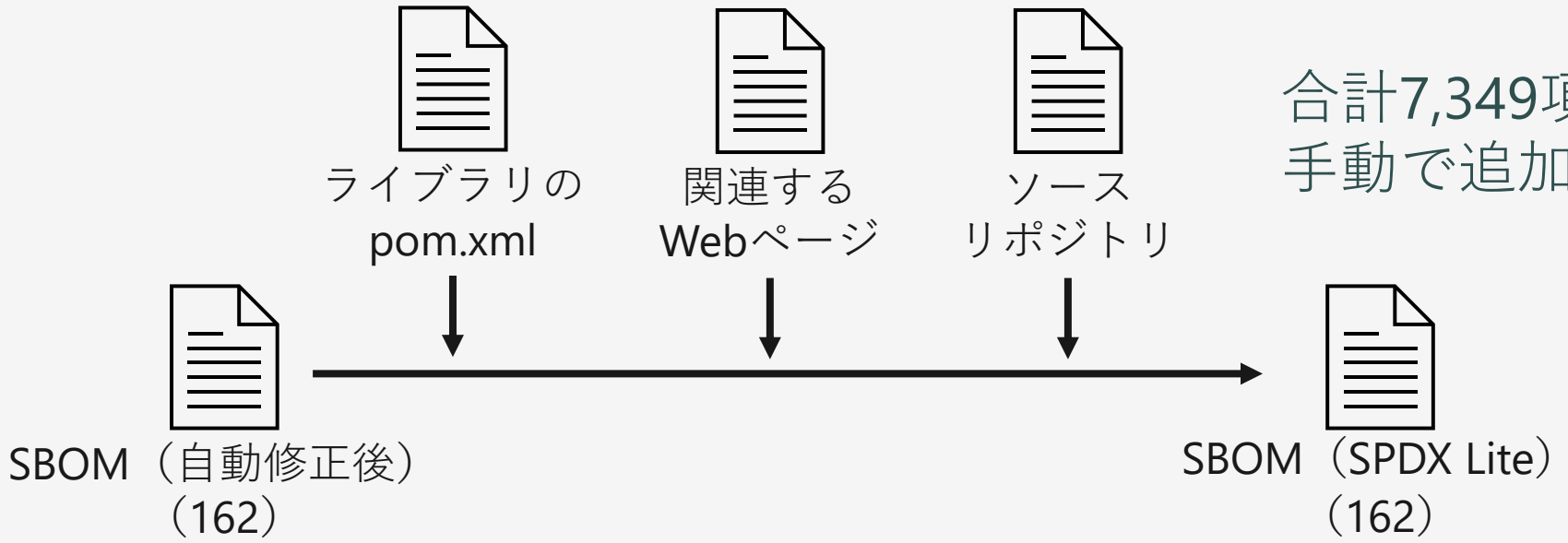
Maven Central Repository上で公開されていないライブラリに依存するプロジェクトは除外する



3.b SBOMの修正｜手動修正

SPDX Liteに準拠したSBOMとなるように情報を追加・修正する

- ライブラリのpom.xml
- pom.xmlからのリンク先（ホームページ，Issue Tracker，ソースリポジトリ）
- Web上をライブラリ名で検索して見つけたソースリポジトリ



データセットの評価

フォーマットの正しさの検証

- SPDXが公式に提供するフォーマットの検証ツールを使用
- データセットに含まれる162個すべてのSBOMが，
  SPDXの規格に従った正しい形式であることを確認

SBOM利用ツールの評価実験

- データセットを用いて既存のSBOM利用ツールを実行し，
  既存ツールの問題点を調査する
- データセットがSBOM利用ツールの評価に役立つことを確認する

SBOM利用ツールの評価実験（1/2）

評価対象

SPDXの公式サイトに掲載されているSBOMの閲覧機能を持つ5個のツール

評価に使用するSBOM

7個のSBOMを選択

GitHub上でスター数が多い，複雑な依存関係を持つ，ライセンスの記法が特殊

広く利用されている一般的なプロジェクトと，
特に問題を引き起こしやすいと考えられるプロジェクトのSBOMを使用

SBOM利用ツールの評価実験（2/2）

評価実験で明らかになった既存ツールの問題点

- 表示される情報量の多さと読みやすさを両立したツールは存在しない
- 5個中4個のツールが，特殊な記法のライセンス情報の表示に非対応
- 一部のSBOMでしか実行できないツールが存在した

評価対象のツールの問題点が明らかになった

作成したデータセットはSBOM利用ツールの評価と改善に役立つものであると考えられる

まとめ

SBOMのデータセットを作成

GitHub上のJavaプロジェクトを対象に，162個のSBOMを作成

既存のSBOM生成ツールを用いて生成し，自動修正・手動修正を行った

品質が保証されたSBOMのサンプルを提供し，
SBOM利用ツールの評価を容易にする

作成したデータセットの評価

フォーマットの正しさの検証

SBOM利用ツールの評価実験