

# SBOM Challenges for Developers: From Analysis of Stack Overflow Questions



Wataru Otoda<sup>†</sup> Tetsuya Kanda<sup>††</sup>

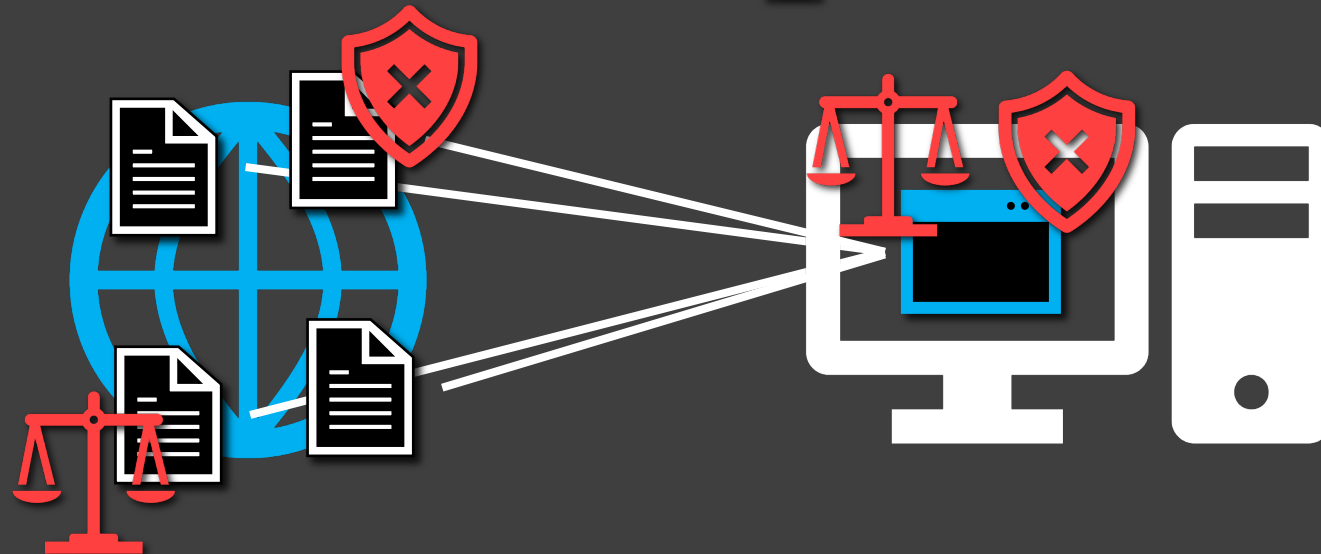
Yuki Manabe<sup>†††</sup> Katsuro Inoue<sup>††††</sup> Yoshiki Higo<sup>†</sup>

<sup>†</sup> Osaka University <sup>††</sup> Notre Dame Seishin University

<sup>†††</sup> The University of Fukuchiyama <sup>††††</sup> Nanzan University

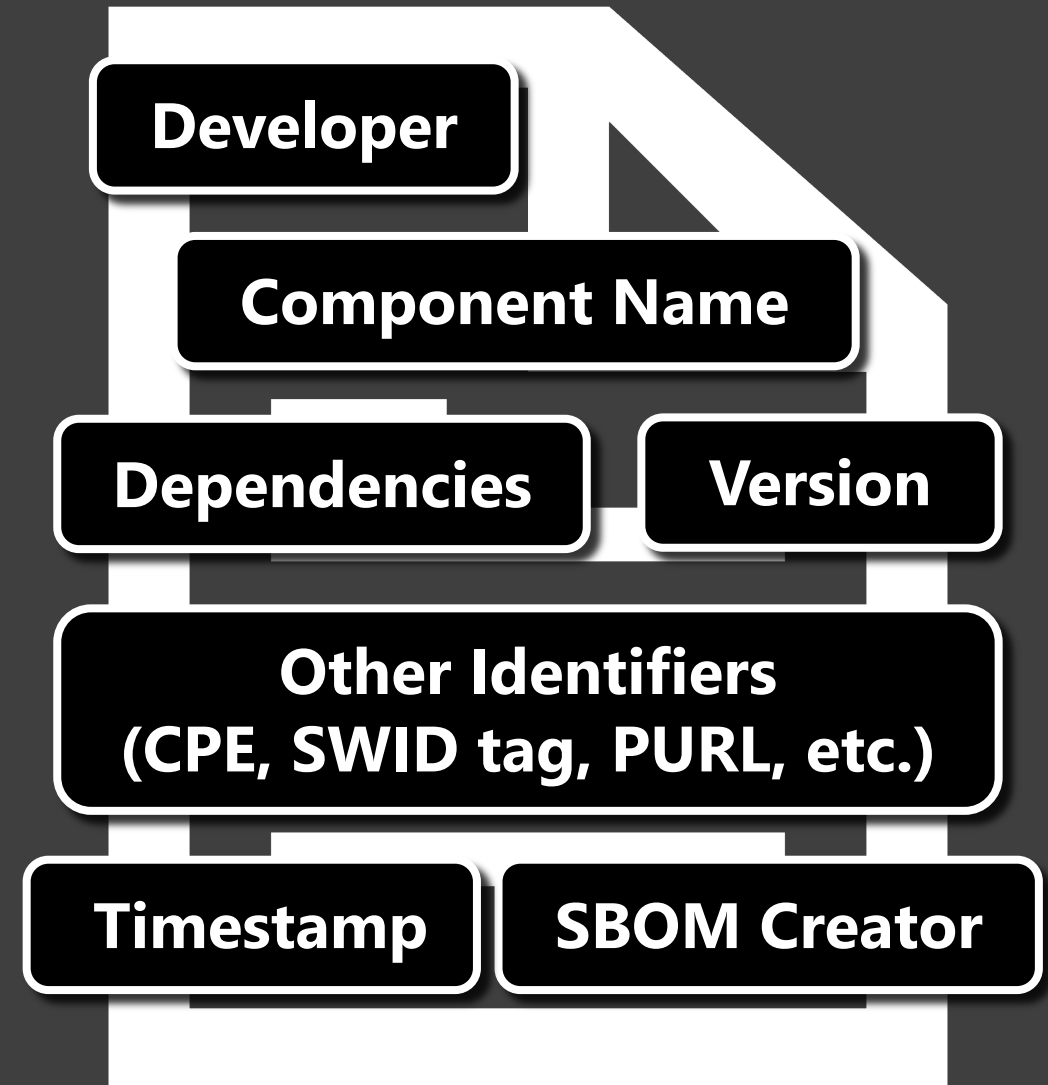
# Background: Software Supply Chain Risks

- Current software development takes advantage of many **external libraries**
  - Reduced dev costs, short dev cycles, and more robust & sophisticated software
  - **96%** codebases contain open-source code [1]
- However, it entails **risks**
  - **Cyber Security (Supply Chain Attack)** 
    - Cyberattacks on the U.S. govts and private sectors via SolarWinds Orion software (2020)
  - **Copyright Infringement (License Violation)** 



# Software Bill of Materials (SBOM)

- Formal and **machine-readable** record of software components [2]
  - Components: Libraries, etc.
- **Quick response to supply chain risks**
  - Executive Order on Improving the Nation's Cybersecurity
    - US Exec. Order, 2021
  - Cyber Resilience Act
    - EU legislation proposal
    - Expected to be effective on 2024

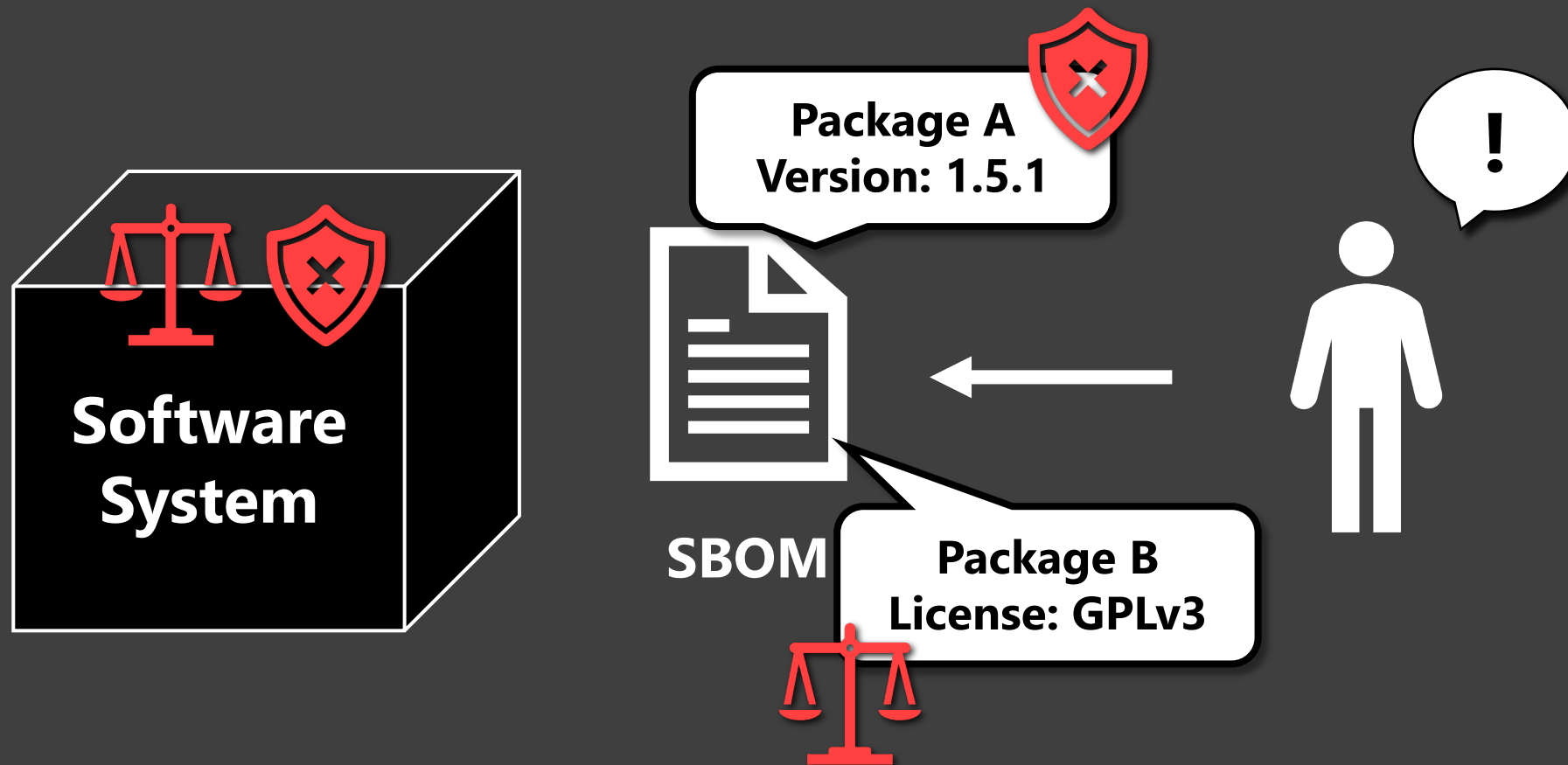


[2] The United States Department of Commerce, "The Minimum Elements For a Software Bill of Materials (SBOM)"

# SBOM Advantages

Easy acquisition of software component information

→ **Detect & mitigate** vulnerability/licensing issues **before** they get problematic



# Primary SBOM Formats



## Linux Foundation ISO/IEC Standard

Compliance and Transparency



## OWASP Standard

Security and Supply Chain Analysis

SBOM adoption is still inadequate

## Readiness Survey (Linux Foundation, 2022) [3]

Studied the SBOM readiness of 412 organizations in various types and sizes from various aspects

## Questionnaire (Xia et al., 2023) [4]

Questioned online SBOM practitioners and presented their views and future goals

## Interview (Stalnaker et al., 2024) [5]

Issues that stakeholders faced when creating and consuming SBOMs and ways to handle them

## Adoption Study (Nocera et al., 2023) [6]

Quantitatively analyzed OSS projects hosted on GitHub

No study revealed **concrete challenges** that **ordinal devs** faced

[3] S. Hendrick and J. Zemlin, "The State of Software Bill of Materials (SBOM) and Cybersecurity Readiness"

[4] B. Xia et al., "An Empirical Study on Software Bill of Materials: Where We Stand and the Road Ahead"

[5] T. Stalnaker et al., "BOMs Away! Inside the Minds of Stakeholders: A Comprehensive Study of Bills of Materials for Software Systems"

[6] S. Nocera et al., "Software Bill of Materials Adoption: A Mining Study from GitHub"

# Stack Overflow

- Developer Q&A website where **many ordinal developers ask questions**
- #Users > 23 million
- #Posts > 60 million
- #New questions per day > 2,400

1 Answer

**Answer**

Sorted by: Highest score (default)

I think you are looking for [sum.golang.org](https://sum.golang.org) which is

2 an auditable checksum database which will be used by the go command to authenticate modules.

you can read more on how it works on [this post](#) from go blog

**Accepted Sign**

answered Sep 9, 2021 at 15:54

NuLo 1,378 ● 1 ● 11 ● 16

**Get ShaXXX of a GO Package** **Title** **Question**

Asked 2 years, 5 months ago Modified 2 years, 5 months ago

**Discussions LABS** A space to share your insight, advice, and perspective. [Try Discussions →](#)

**Body**

I Need to get the Sha512 or similar from a Golang package for SBOM purposes. For example, the hash for package <https://pkg.go.dev/encoding/json>

I can't found any information or api to get it. If possible I need it with source code.

**Tags** go sha

**Timestamp** asked Sep 9, 2021 at 15:43

Share Improve this question Follow

1,128 ● 7 ● 17

<https://stackoverflow.com/questions/69121175>

# Analysis of Stack Overflow Questions

Our expectation:

**Developers facing challenges in utilizing SBOM are likely to seek solutions on Stack Overflow by asking questions**

**Stack Overflow analysis should reveal SBOM challenges that ordinal developers faced**





## Analysis 1: Answered and Resolved Rate of SBOM Questions

Can Stack Overflow resolve SBOM challenges?



## Analysis 2: Trends in the Number of SBOM Questions

Are there any effects on the SBOM question count from the U.S. Exec. Order requesting SBOM use and ISO/IEC standardization of SPDX?

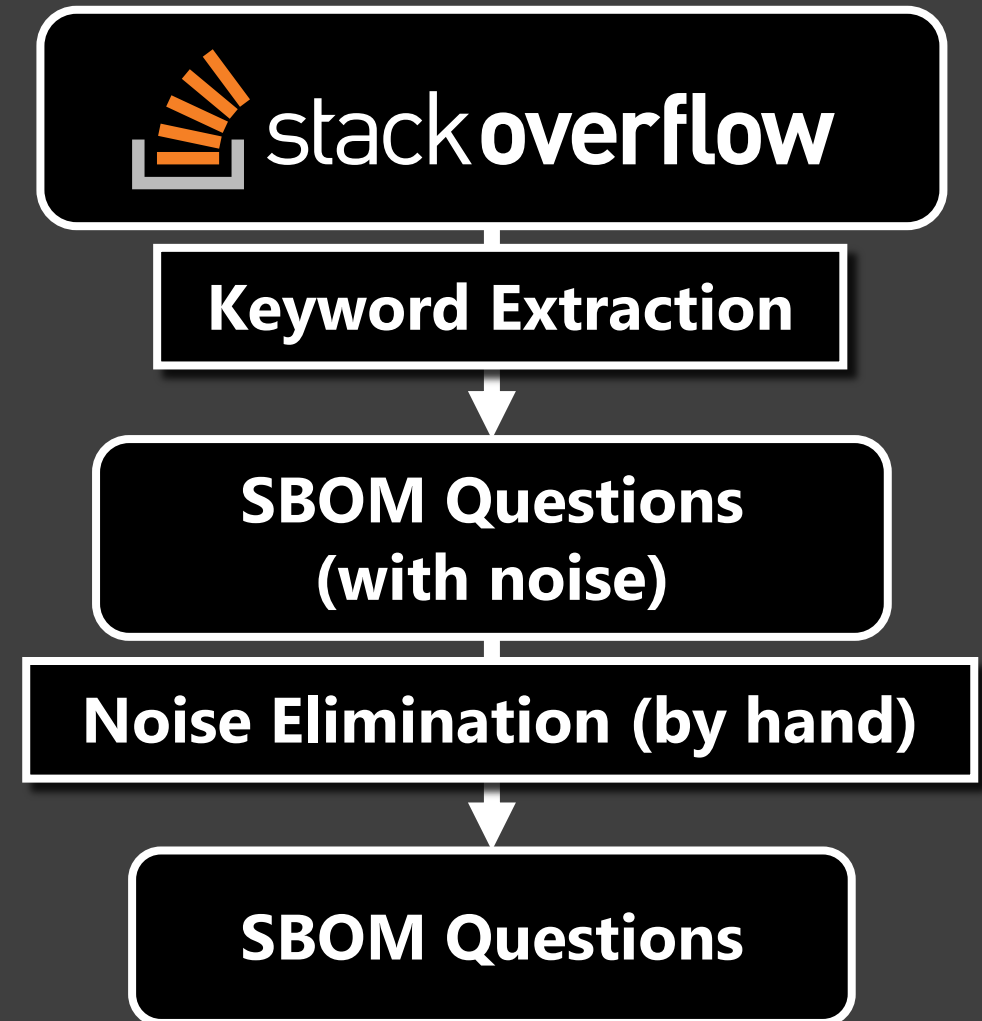


## Analysis 3: Challenges Faced by SBOM Users

What are the concrete challenges that ordinal developers faced?

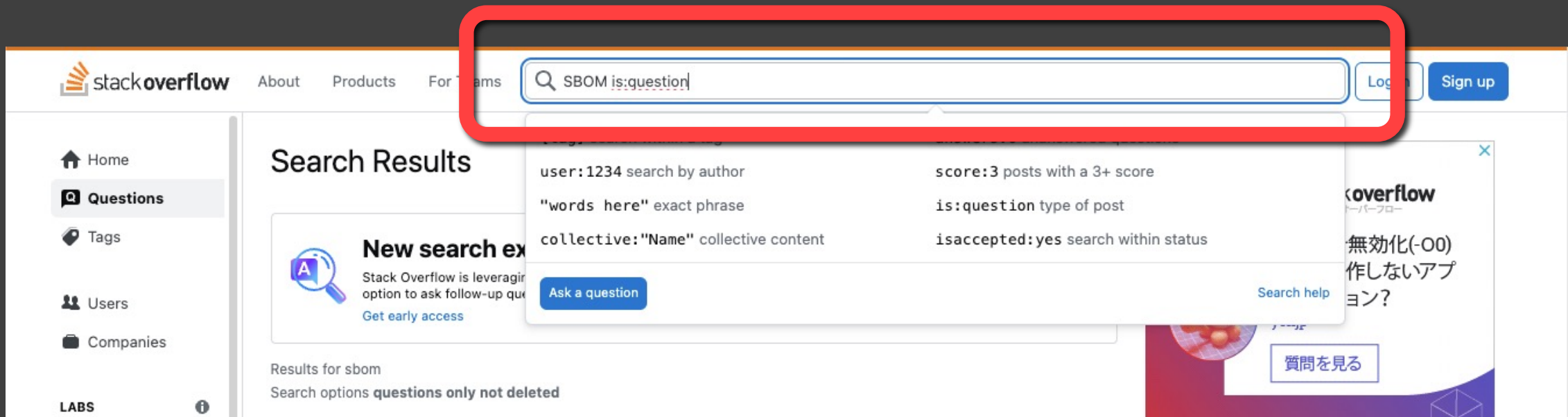
# Question Extraction – Overview

- Keywords: *SBOM*, *CycloneDX*, *SPDX*
  - Case-insensitive
  - Appears in title or body (whole-word)
- Tags are useless
  - SBOM tags are not widely used



# Question Extraction – SBOM, CycloneDX

- Used the search bar on the Stack Overflow website
  - `is:question` allows us to extract only questions
- Excluded noise by hand
- Extracted only questions before Sep. 2023 to keep the condition same to *SPDX*



# Question Extraction – *SPDX*

- Search result exceeded the upper limit of 500 imposed by Stack Overflow
- However, they were mostly noise in some certain patterns



- Filter out noise from *Stack Exchange Data Dump 2023-09-12*
  - Automatically remove questions with a *noise pattern string*
  - Manually exclude remaining noise



## Stack Exchange Data Dump 2023-09-12

by [Stack Exchange, Inc.](#)

Publication date [2023-09-12](#)  
Topics [Stack Exchange Data Dump](#)  
Contributor [Stack Exchange Community](#)

This is a copy of the 2023-09-12 dump in [stackexchange](#). It contains all files present in that item at the time of mirroring on 2023-11-19.

### Why?

The stackexchange item gets overwritten periodically, and the old dumps become hard to impossible to access. Cf. the Meta SE discussion: [All Stack Exchange data dumps](#)



73 Views

### DOWNLOAD OPTIONS

<a href="#">7Z</a>	371 files
<a href="#">ITEM TILE</a>	1 file
<a href="#">PNG</a>	1 file

# Question Extraction – Noise Pattern Strings

- Frequent text in noise verified to not appear in an SPDX SBOM
- 41 questions extracted
  - SBOM: 27 out of 58
  - CycloneDX: 28 out of 31
  - SPDX: 6 out of 1432

Noise Pattern Strings	Count
SPDX-License-Identifier	1229
License should be a valid SPDX license expression	30
spdy	51
spdx-correct	
spdx-expression-parse	62
spdx-exceptions	
spdx-license-ids	



## Analysis 1: Answered and Resolved Rate of SBOM Questions

Examine the presence of answers and the resolution status of each question. Questions with an accepted answer were considered resolved.



## Analysis 2: Trends in the Number of SBOM Questions

Sum up the number of new questions posted each year.



## Analysis 3: Challenges Faced by SBOM Users

Extract technical issues from each question, identifying FAQs and those seemingly unresolved based on the subjective judgment of the lead author.

# Results – Analysis 1 (Answered and Resolved Rate)

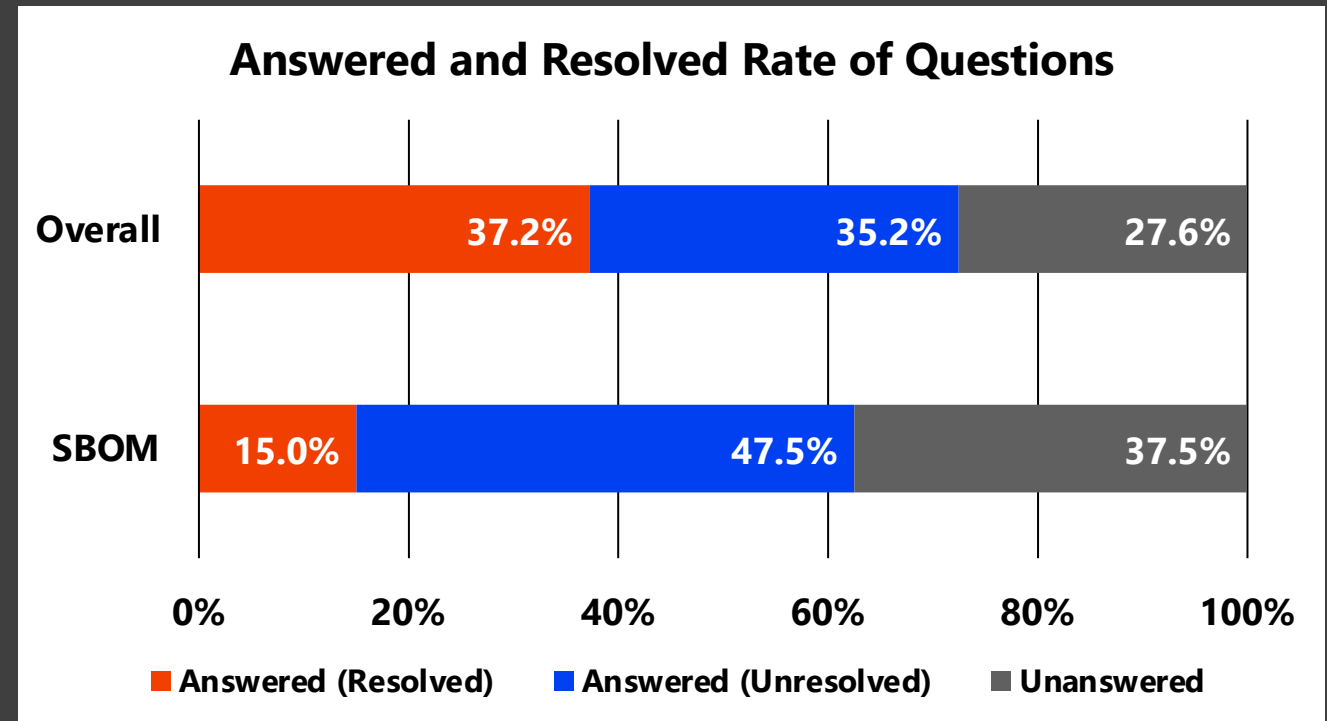
- Among SBOM questions
  - Answered (Resolved) was 6
  - Answered (Unresolved) was 19
  - Unanswered was 15
  - (2021 onwards for fair comparison)



**Answer Rate: Average**

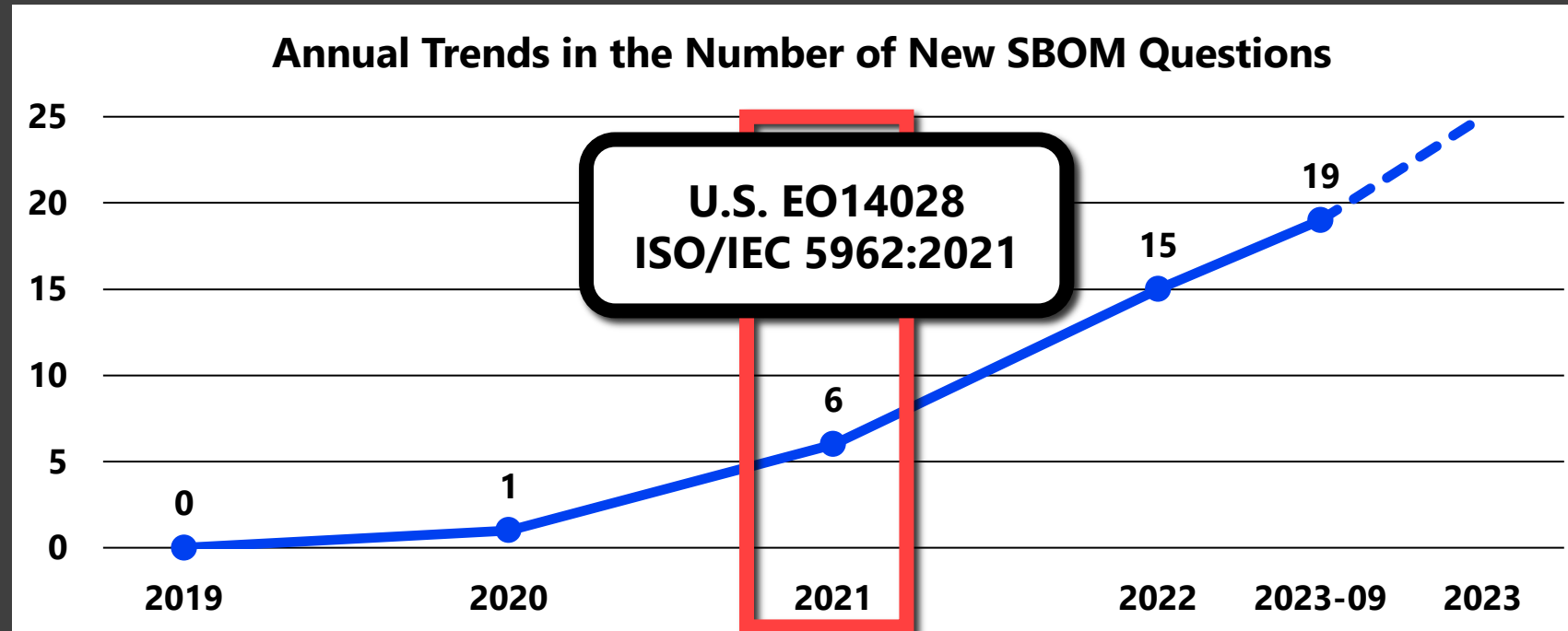


**Resolve Rate: Low**



- **It is challenging to find solutions for SBOM questions on Stack Overflow**
  - Shortage of software developers knowledgeable enough to answer SBOM questions
  - SBOM tools being so immature that many SBOM issues are currently unsolvable

# Results – Analysis 2 (Trends in the Number of Questions)



- Small over the four years
  - Before 2019: 1 in 2012 and 0 for the rest
- Steady increase from 2020 to 2023, with a notable acceleration from 2021
  - Correlation with the U.S. exec. order and the ISO/IEC standardization of SPDX?



# Results – Analysis 3 (SBOM Challenges)



**Insufficient Coverage of Use Cases by SBOM Tools (7)**



**Inability of SBOM Tools to Meet Requirements (4)**



**Immaturity of SBOM Tools or Unclear Usages (19)**

# Results – Analysis 3 – Challenge 1

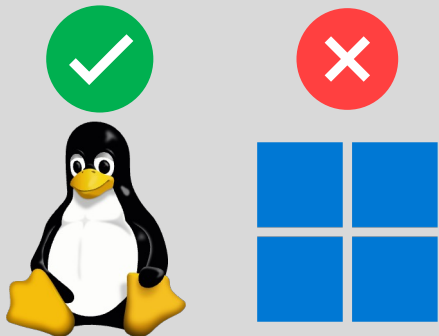


## Insufficient Coverage of Use Cases by SBOM Tools



### Insufficient tool support for older project management systems

*"How do I generate a Cyclonedx bom for a Java project built with Ant?"*<sup>1</sup>



### Absence of tool support for Microsoft Windows

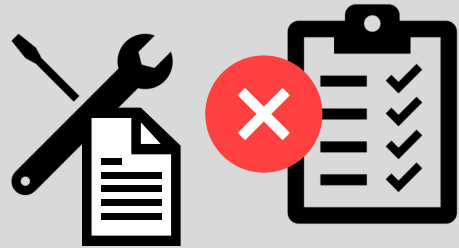
*"Is there any tool through which we can generate SBOM report (SPDX / CycloneDX) for Windows programs?"*<sup>2</sup>

<sup>1</sup> <https://stackoverflow.com/questions/71605182> (March 2022)

<sup>2</sup> <https://stackoverflow.com/questions/73648096> (September 2022)



## Inability of SBOM Tools to Meet Requirements



**Gap between SBOM tools' ability and compliance requirements, or challenges in obtaining all the necessary information as specified in the guideline from the perspective of SBOM tool developers**

*"NTIA minimum SBOM requirement tool" <sup>1</sup>*

(NTIA released it in July 2021 following the U.S. Exec. Order)



**Limitations of GitHub's SBOM functionality for software license compliance**

*"How to include Open Source license in GitHub SBOM export?" <sup>2</sup>*

(GitHub generates an SBOM from info available within the repo)

<sup>1</sup> <https://stackoverflow.com/questions/76103711> (April 2022)

<sup>2</sup> <https://stackoverflow.com/questions/76962834> (August 2023)



## Immaturity of SBOM Tools or Unclear Usages

- **Immature SBOM tools are making correct SBOM generation difficult**
  - defects
  - insufficient features
  - lack of comprehensive docs or accumulated knowledge among devs on basic usage
- Questions on **errors** and **unclear usages** by SBOM tool users were prevalent
- ***How to create a BOM file in a Flutter project***<sup>1</sup>
  - Encountered an error on CycloneDX SBOM generation for a Flutter project



<sup>1</sup> <https://stackoverflow.com/questions/76515339> (June 2023)

# Results – Analysis 3 – Discussion

- Ordinal developers seek:



**Accumulated Knowledge on Existing SBOM Tools**



**Refinement and Extension to Existing SBOM Tools**

- Questions requiring fundamentally new approaches or systems were scarce
- Most were revolving around usages, issues, and minor missing functionalities on existing tools

# Conclusion: SBOM Challenges for Developers on Stack Overflow



## Analysis 1: Answered and Resolved Rate of SBOM Questions

Resolve rate is exceptionally low, indicating that Stack Overflow is still not a satisfactory venue to ask SBOM questions to obtain solutions.



## Analysis 2: Trends in the Number of SBOM Questions

Notable increase from 2021, possibly due to the U.S. executive order mandating SBOM use and the standardization of SPDX by ISO/IEC.



## Analysis 3: Challenges Faced by SBOM Users

"Insufficient Coverage of Use Cases by SBOM Tools", "Inability of SBOM Tools to Meet Requirements", and "Immaturity of SBOM Tools or Unclear Usages".

Future work: Keep tracking the latest trends in SBOM use

# Appendix

# NTIA-defined Minimum Elements for an SBOM [2]

Data Field	Description
Supplier Name	The name of an entity that creates, defines, and identifies components.
Component Name	Designation assigned to a unit of software defined by the original supplier.
Version of the Component	Identifier used by the supplier to specify a change in software from a previously identified version.
Other Unique Identifiers	Other identifiers that are used to identify a component, or serve as a look-up key for relevant databases (e.g., CPE, SWID tag, PURL).
Dependency Relationship	Characterizing the relationship that an upstream component X is included in software Y.
Author of SBOM Data	The name of the entity that creates the SBOM data for this component.
Timestamp	Record of the date and time of the SBOM data assembly.

[2] The United States Department of Commerce, "The Minimum Elements For a Software Bill of Materials (SBOM)"



# Question Extraction – Details on Noise Pattern Strings

- **SPDX-License-Identifier**
  - Notation for declaring licenses at the beginning of program code
  - Part of the SPDX standard but not an SBOM
- **License should be a valid SPDX license expression**
  - Warning output from the Yarn package manager
- **spdy**
  - Variable name represents "speed  $y$ " often accompanied by the variable name **spdx** (speed  $x$ )
- **spdx-correct, spdx-expression-parse, spdx-exceptions, spdx-license-ids**
  - NPM package names as they appear in the log outputs for NPM errors

# Question Extraction – Noise Example (SBOM)

- Docker question with a `docker info` result to show the environment  
→ Matches the search condition by including “SBOM” message

```
• Here is the output of docker info:  
  
Client:  
Context: default  
Debug Mode: false  
Plugins:  
  buildx: Docker Buildx (Docker Inc., v0.10.0)  
  compose: Docker Compose (Docker Inc., v2.15.1)  
  dev: Docker Dev Environments (Docker Inc., v0.0.5)  
  extension: Manages Docker extensions (Docker Inc., v0.2.17)  
  sbom: View the packaged-based Software Bill Of Materials (SBOM) for an image (.  
  scan: Docker Scan (Docker Inc., v0.23.0)  
  
Server:  
Containers: 2  
  Running: 2  
  Paused: 0  
  Stopped: 0  
Images: 1  
Server Version: 20.10.22  
Storage Driver: overlay2  
  Backing Filesystem: extfs
```

# Question Extraction – Noise Example (CycloneDX)

- General coding question with a project configuration file attached  
→ Matches the search condition because of a dependent CycloneDX SBOM tool

and here is my pom:

```
    </configuration>
  </plugin>
  <plugin>
    <groupId>org.cyclonedx</groupId>
    <artifactId>cyclonedx-maven-plugin</artifactId>
    <version>2.7.0</version>
    <executions>
      <execution>
        <phase>package</phase>
        <goals>
          <goal>makeBom</goal>
        </goals>
      </execution>
    </executions>
  </plugin>
</plugin>
```

# Question Extraction – Noise Example (SPDX)

- General coding question with a program code attached  
→ Matches the search condition because of a license info (SPDX License List URL)

and this is my set up code in my index.js

```
const port = 3050;
const options = {
  definition: {
    openapi: "3.0.0",
    info: {
      title: "LogRocket Express API with Swagger",
      version: "0.1.0",
      description:
        "This is a simple CRUD API application made with Express and documented",
      license: {
        name: "MIT",
        url: "https://spdx.org/licenses/MIT.html",
      },
      contact: {
        name: "LogRocket",
        url: "https://logrocket.com",
        email: "info@email.com",
      },
    },
  },
},
```

# Example of an SPDX SBOM

```
{
  "SPDXID": "SPDXRef-DOCUMENT",
  "spdxVersion": "SPDX-2.3",
  "creationInfo": {
    "created": "2024-02-26T15:42:46Z",
    "creators": ["Tool: GitHub.com-Dependency-Graph"]
  },
  "name": "com.github.watamario15/formsclient",
  "dataLicense": "CC0-1.0",
  "documentDescribes": ["SPDXRef-com.github.watamario15-formsclient"],
  "documentNamespace": "https://github.com/watamario15/formsclient/dependency_graph/sbom-dc5516f2edc68207",
  "packages": [
    {
      "SPDXID": "SPDXRef-com.github.watamario15-formsclient",
      "name": "com.github.watamario15/formsclient",
      "versionInfo": "",
      "downloadLocation": "git+https://github.com/watamario15/formsclient",
      "licenseDeclared": "CC0-1.0",
      "filesAnalyzed": false,
      "supplier": "NOASSERTION",
      "externalRefs": [
        {
          "referenceCategory": "PACKAGE-MANAGER",
          "referenceType": "purl",
          "referenceLocator": "pkg:github/watamario15/formsclient"
        }
      ]
    }
  ], (Omitted)
  "relationships": [
    {
      "relationshipType": "DEPENDS_ON",
      "spdxElementId": "SPDXRef-com.github.watamario15-formsclient",
      "relatedSpdxElement": "SPDXRef-actions-actions-checkout-3"
    }, (Omitted)
  ]
}
```

# SQL Query for Noise Elimination

```
create table posts_spdx
select * from Posts
where PostTypeId = 1 and
    (Body regexp '[[:<:]]spdx[[:>:]]' or Title regexp '[[:<:]]spdx[[:>:]]');

select Id from posts_spdx
where not Body like '%SPDX-License-Identifier%' and
    not Body like '%spdy%' and
    not Body like '%License should be a valid SPDX license expression%' and
    not Body like '%spdx-correct%' and
    not Body like '%spdx-expression-parse%' and
    not Body like '%spdx-exceptions%' and
    not Body like '%spdx-license-ids%'
order by Id desc;
```

# Threats for Validity

## Noise Elimination

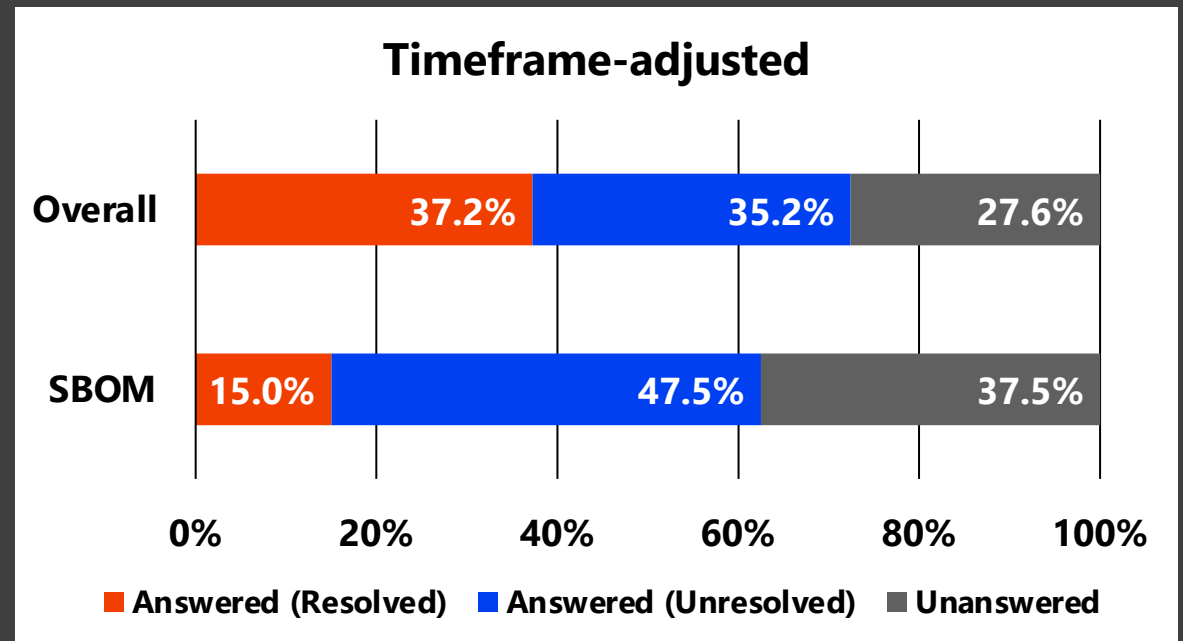
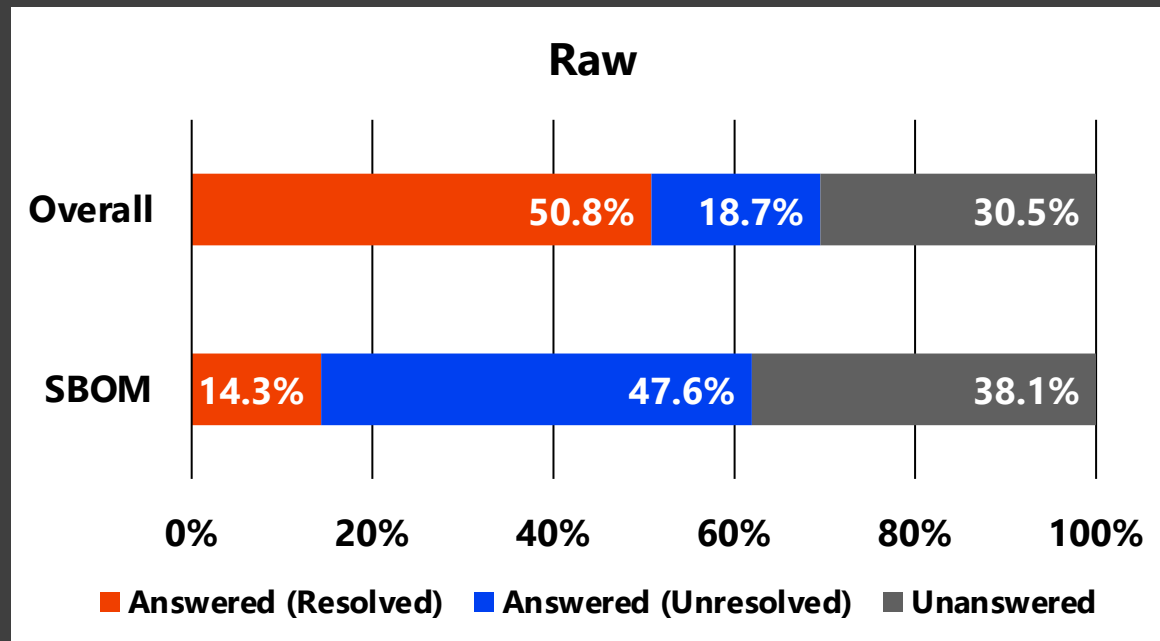
- We automatically filtered out *SPDX* questions
  - **Might excluded some SBOM questions which were actually not noise**
- They were likely noise as we investigated multiple such posts
  - **The threat to validity caused by this exclusion is considered small**

## Analyzed Question Count

- The number of questions extracted as meeting the criteria was only 42
  - **Might affect the accuracy of the analysis results**
- We observed no active usage of General Q&A sites like Quora apart from Stack Overflow by developers or managers involved in SBOM adoption

# Timeframe Adjustment for Analysis 1 (Answered/Resolved Rate)

- Most SBOM questions are posted within 3 years from now
  - Might be just that we analyzed too early for them to get answers
- Limited the timeframe to 2021-01-01 - 2023-09-03 for both overall and SBOM Qs
  - Both the raw and timeframe-adjusted result shows the similar tendency





# Future Works

- There should be more hidden struggles and questions
- **We should keep tracking the latest trends in SBOM use**



Continue investigating Stack Overflow, which is showing an increasing trend in the number of SBOM questions



Conduct analyses on websites other than Q&A sites, such as Reddit and Issues in the SBOM-related projects on GitHub