Image from https://unsplash.com/photos/vBvfXlqC4E4

Getting Continuous Testing Done Right with CD-Linter

Carmine Vassallo
University of Zurich

DevOps Institute, *Continuous Testing SKILup Day*, November 19, 2020

@ccvassallo

Who Am I

I'm on the Job Market!

My name is **Carmine Vassallo**

Research intern in the Continuous Delivery team at ING Nederland (2015)

PhD Graduate from the University of Zurich (2020), where I am currently a postdoctoral researcher

My research goal is to facilitate the adoption of DevOps practices

http://tiny.uzh.ch/WV

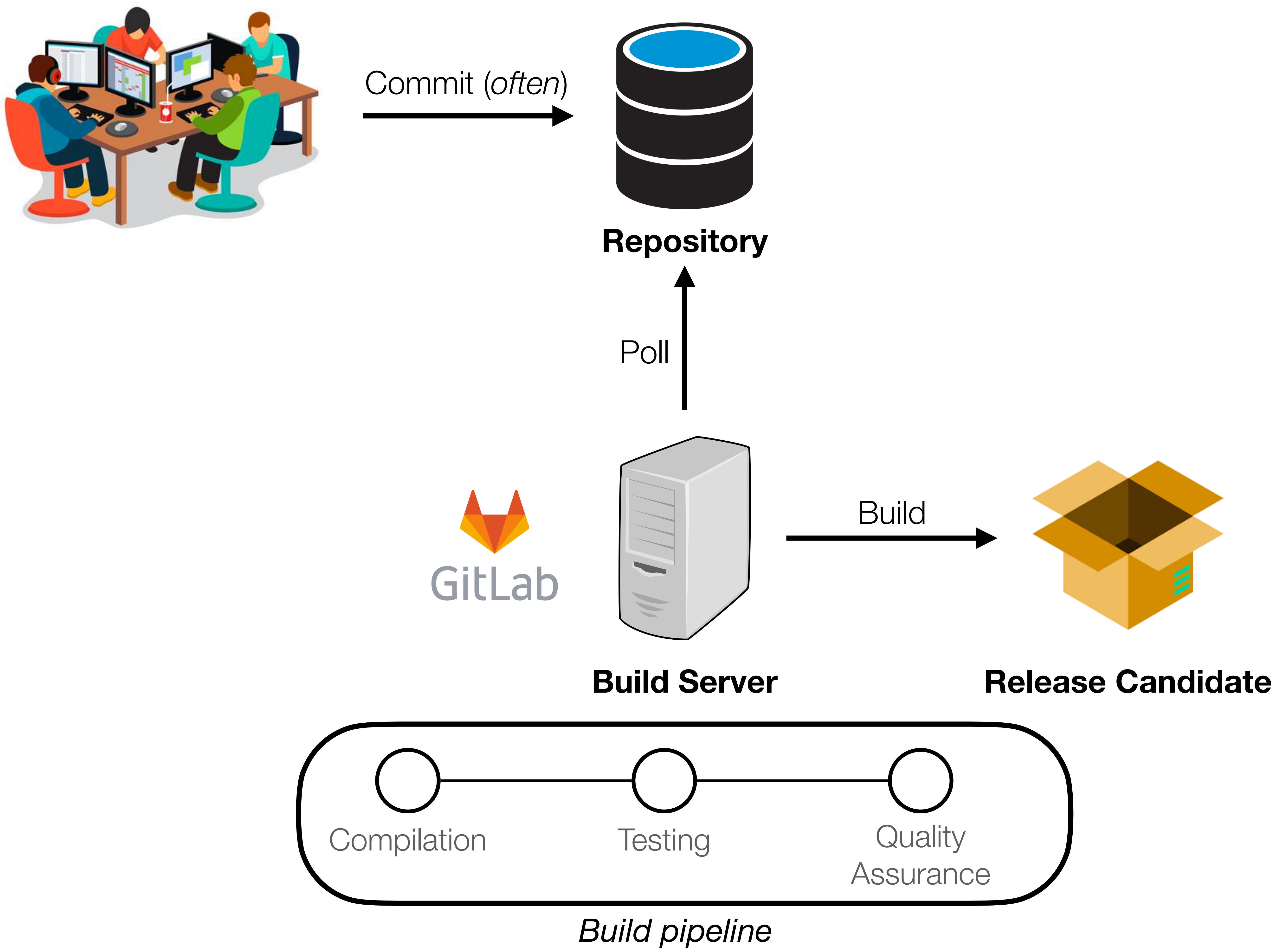
@ccvassallo

The Addison-Wesley Signature Series
CONTINUOUS
A MARTIN FOWLER SIGNATURE BOOK
Foreword by Martin Fowler

Continuous Testing is a foundation of Continuous Delivery

(Humble et Farley, 2010)

Continuous Delivery (CD)



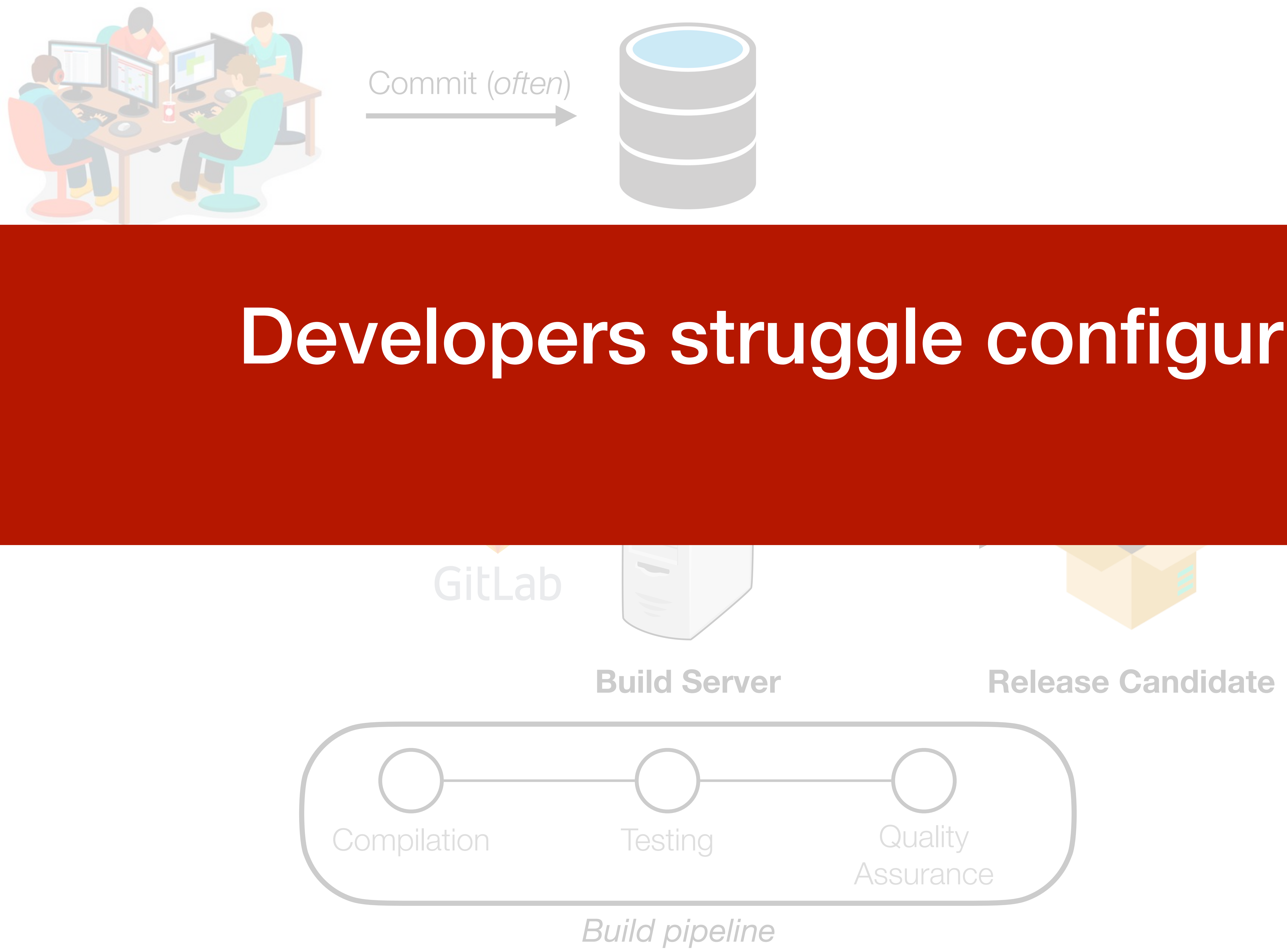
```
stages:
    -  compilation
    -  testing
    -  qa

variables:
    POSTGRES_USR: user
    POSTGRES_PWD: password

compile_production_code:
    stage: compile
    script: “mvn compile”
    when: manual
    allow_failure: false

compile_test_code:
    stage: compilation
    script: “mvn test”
    retry: 3
…
```

.gitlab-ci.yml

Icons from https://vitalitychicago.com/blog/top-reasons-agile-didnt-work-for-us-1-we-couldnt-co-locate-teams/, https://www.flaticon.com/authors/roundicons, https://www.pinclipart.com/pindetail/hbTb_clipart-info-server-png-transparent-png/

@ccvassallo

Continuous Delivery (CD)



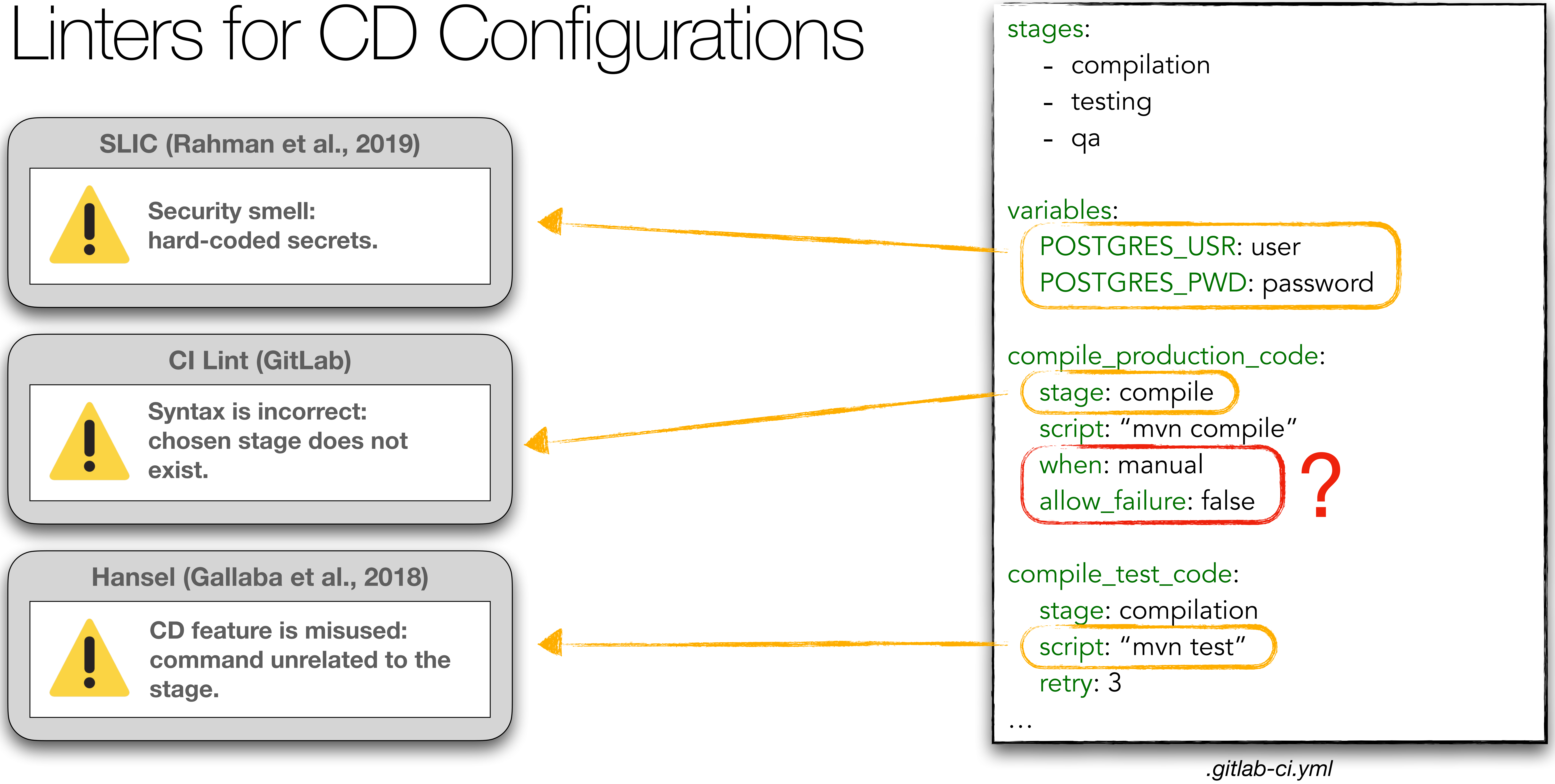
```
stages:
    - compilation
    - testing
    - qa

variables:
    POSTGRES_USR: user
    POSTGRES_PWD: password

compile_test_code:
    stage: compilation
    script: "mvn test"
    retry: 3
...
```

.gitlab-ci.yml

Developers struggle configuring build pipelines

(Hilton et al., 2017)

Linters for CD Configurations



SLIC (Rahman et al., 2019)

⚠ **Security smell: hard-coded secrets.**

CI Lint (GitLab)

⚠ **Syntax is incorrect: chosen stage does not exist.**

Hansel (Gallaba et al., 2018)

⚠ **CD feature is misused: command unrelated to the stage.**

```
stages:
    - compilation
    - testing
    - qa

variables:
    POSTGRES_USR: user
    POSTGRES_PWD: password

compile_production_code:
    stage: compile
    script: "mvn compile"
    when: manual
    allow_failure: false        ?

compile_test_code:
    stage: compilation
    script: "mvn test"
    retry: 3
...
```

.gitlab-ci.yml

Linters for CD Configurations

SLIC (Rahman et al., 2019)

⚠ Security smell: hard-coded secrets.

```
stages:
   -  compilation
   -  testing
   -  qa

variables:
  POSTGRES_USR: user
  POSTGRES_PWD: password
```

Developers typically lack awareness of CD principle (e.g., Continuous Testing) violations that threaten expected benefits

(Vassallo et al., 2019)

⚠ CD feature is misused: command unrelated to the stage.

```
  stage: compilation
  script: "mvn test"
  retry: 3
…
```

.gitlab-ci.yml

CD-Linter: Detecting violations of CD principles

Fake Success

Retry Failure

Manual Execution

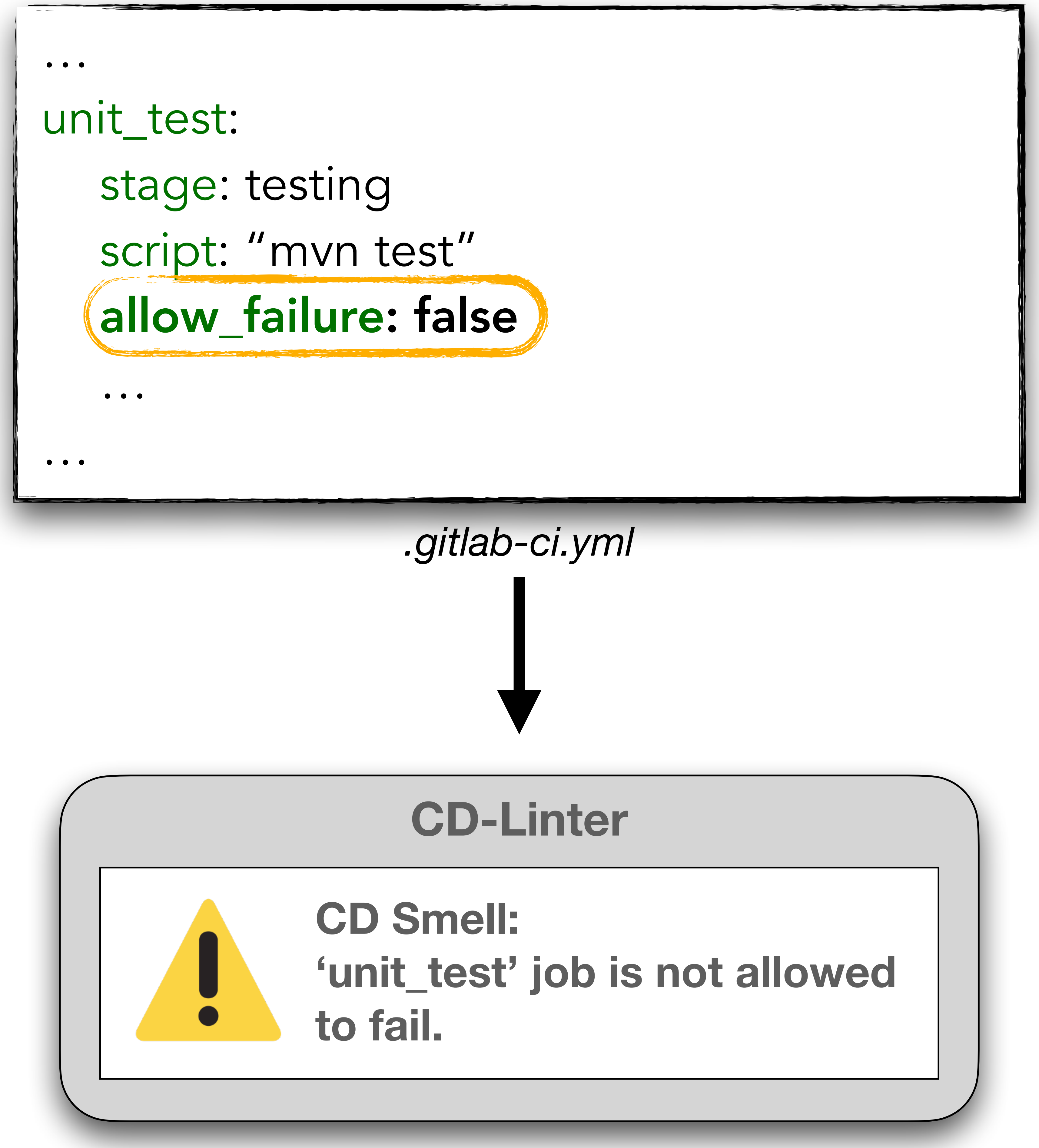
Fuzzy Version

Carmine Vassallo, Sebastian Proksch, Anna Jancso, Harald C. Gall, Massimiliano Di Penta.
Configuration Smells in Continuous Delivery Pipelines: A Linter and A Six-Month Study on GitLab. In ESEC/FSE, 2020.

Fake Success

👍 Fail the build in presence of defects

👎 Prevent job failures from failing the build



```yaml
...
unit_test:
   stage: testing
   script: "mvn test"
   allow_failure: false
   ...
...
```

.gitlab-ci.yml

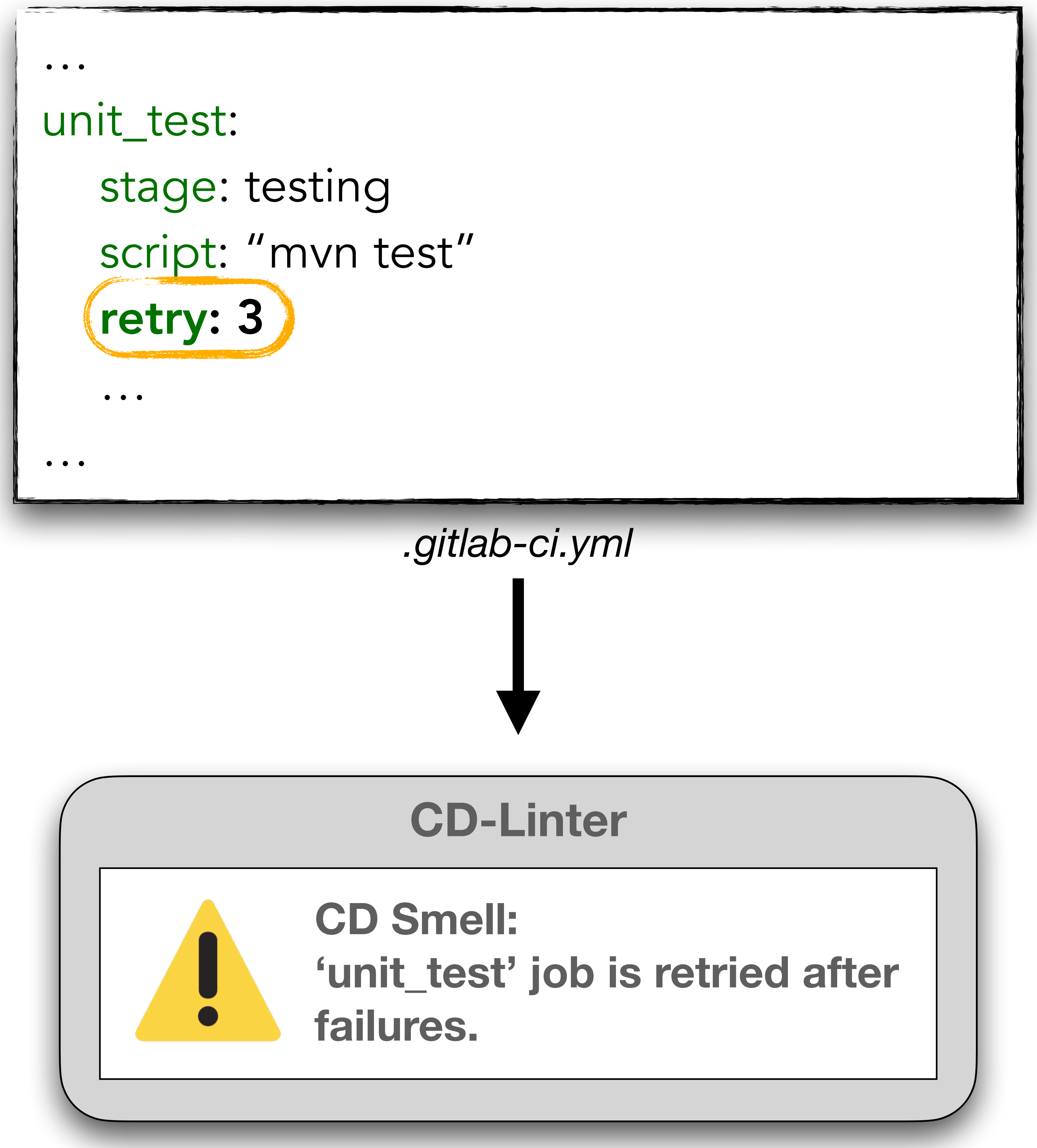
CD-Linter

CD Smell:
'unit_test' job is not allowed to fail.

Retry Failure

👍 The build process has to be deterministic

👎 Hiding flakiness by rerunning a job multiple times after failures.



```
...
unit_test:
   stage: testing
   script: “mvn test”
   retry: 3
   ...
...
```

.gitlab-ci.yml

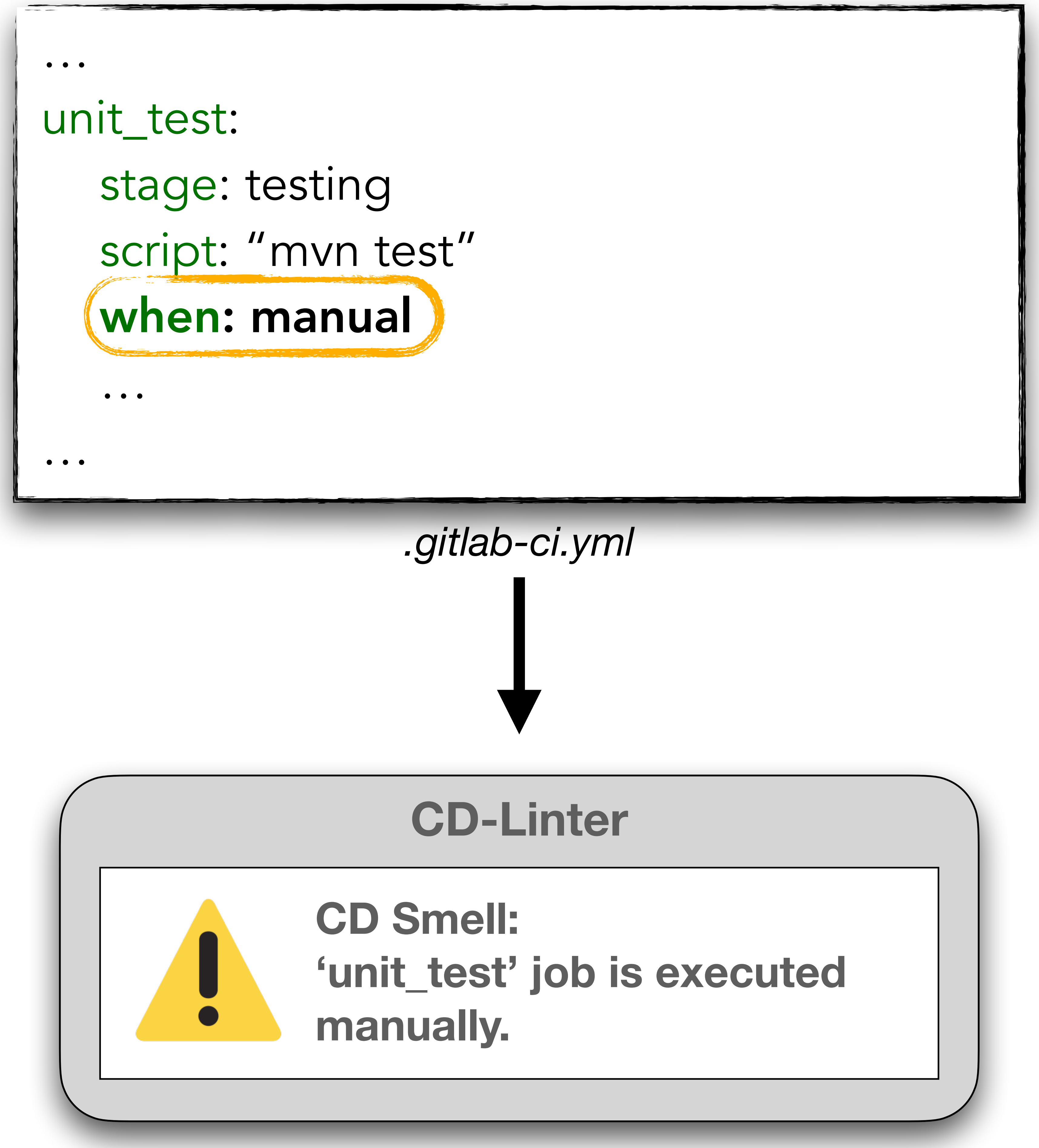
CD-Linter

CD Smell:
‘unit_test’ job is retried after failures.

Manual Execution

👍 The pipeline has to be fully automated

👎 Some jobs are triggered manually



```
...
unit_test:
    stage: testing
    script: “mvn test”
    when: manual
    ...
...
```

.gitlab-ci.yml

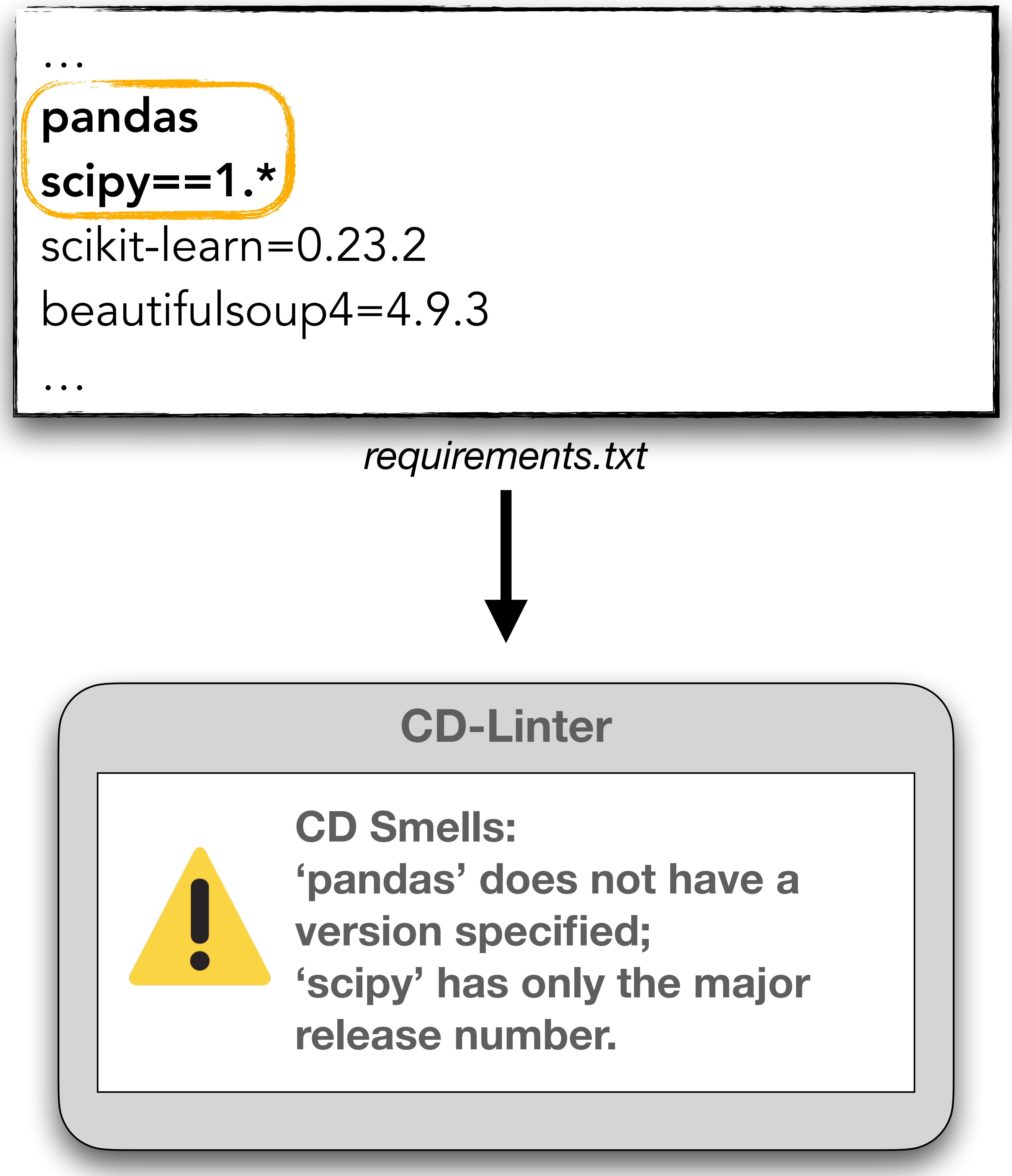
CD-Linter

CD Smell:
‘unit_test’ job is executed manually.

Fuzzy Version

👍 The build needs to be reproducible

👎 Do not specify the exact version of dependencies



```
...
pandas
scipy==1.*
scikit-learn=0.23.2
beautifulsoup4=4.9.3
...
```

requirements.txt

CD-Linter

CD Smells:
‘pandas’ does not have a version specified;
‘scipy’ has only the major release number.

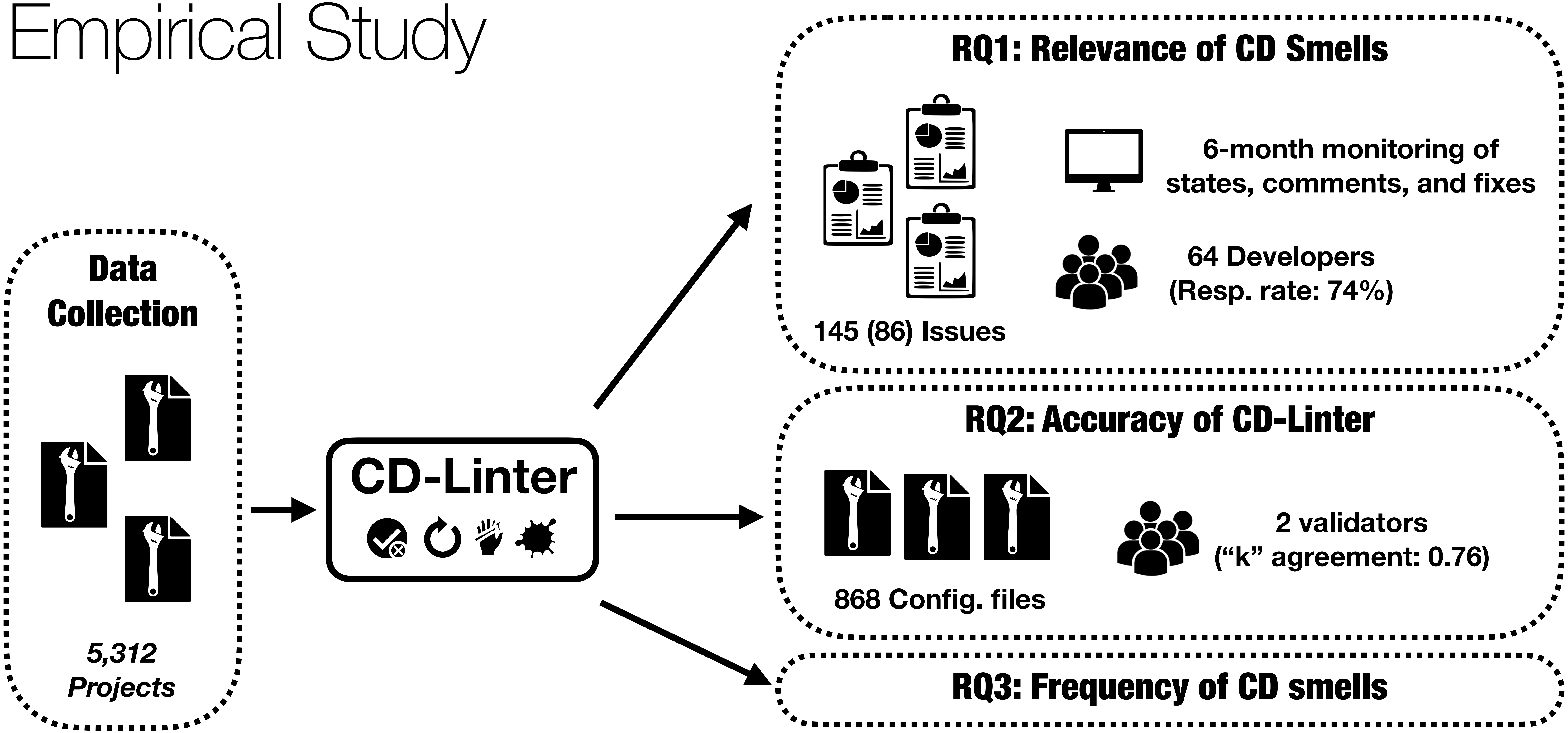
Evaluation of CD-Linter

RQ1: Are the CD Smells Detected by CD-Linter Relevant to Developers?

RQ2: How Accurate Is CD-Linter?

RQ3: How Frequent Are the Investigated CD Smells in Practice?

Empirical Study



Data Collection

5,312 Projects

CD-Linter

RQ1: Relevance of CD Smells

145 (86) Issues

6-month monitoring of states, comments, and fixes

64 Developers (Resp. rate: 74%)

RQ2: Accuracy of CD-Linter

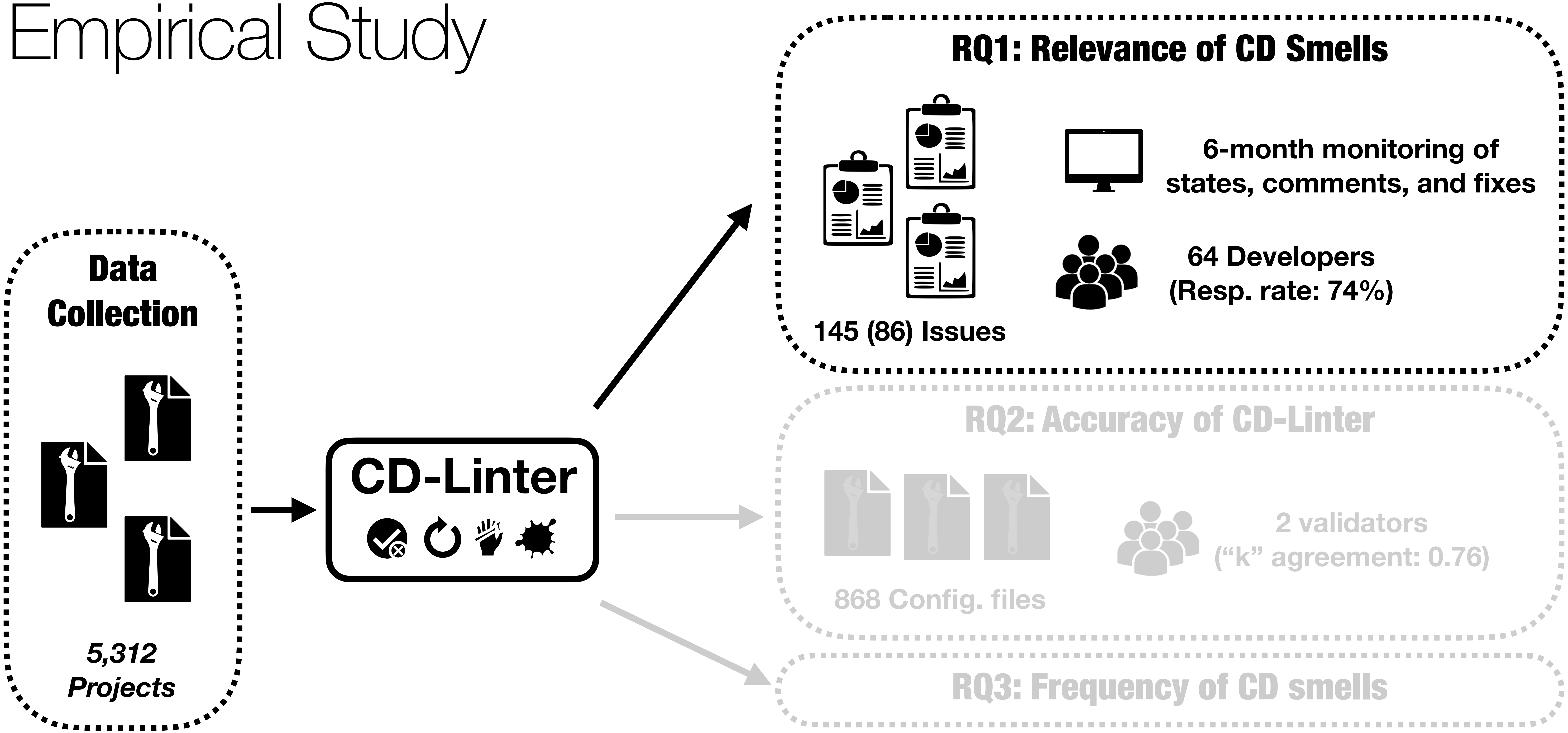
868 Config. files

2 validators (“k” agreement: 0.76)

RQ3: Frequency of CD smells

Icons from: https://www.flaticon.com/authors/freepik

Empirical Study



Data Collection

5,312 Projects

CD-Linter

RQ1: Relevance of CD Smells

145 (86) Issues

6-month monitoring of states, comments, and fixes

64 Developers (Resp. rate: 74%)

RQ2: Accuracy of CD-Linter

868 Config. files

2 validators (“k” agreement: 0.76)

RQ3: Frequency of CD smells

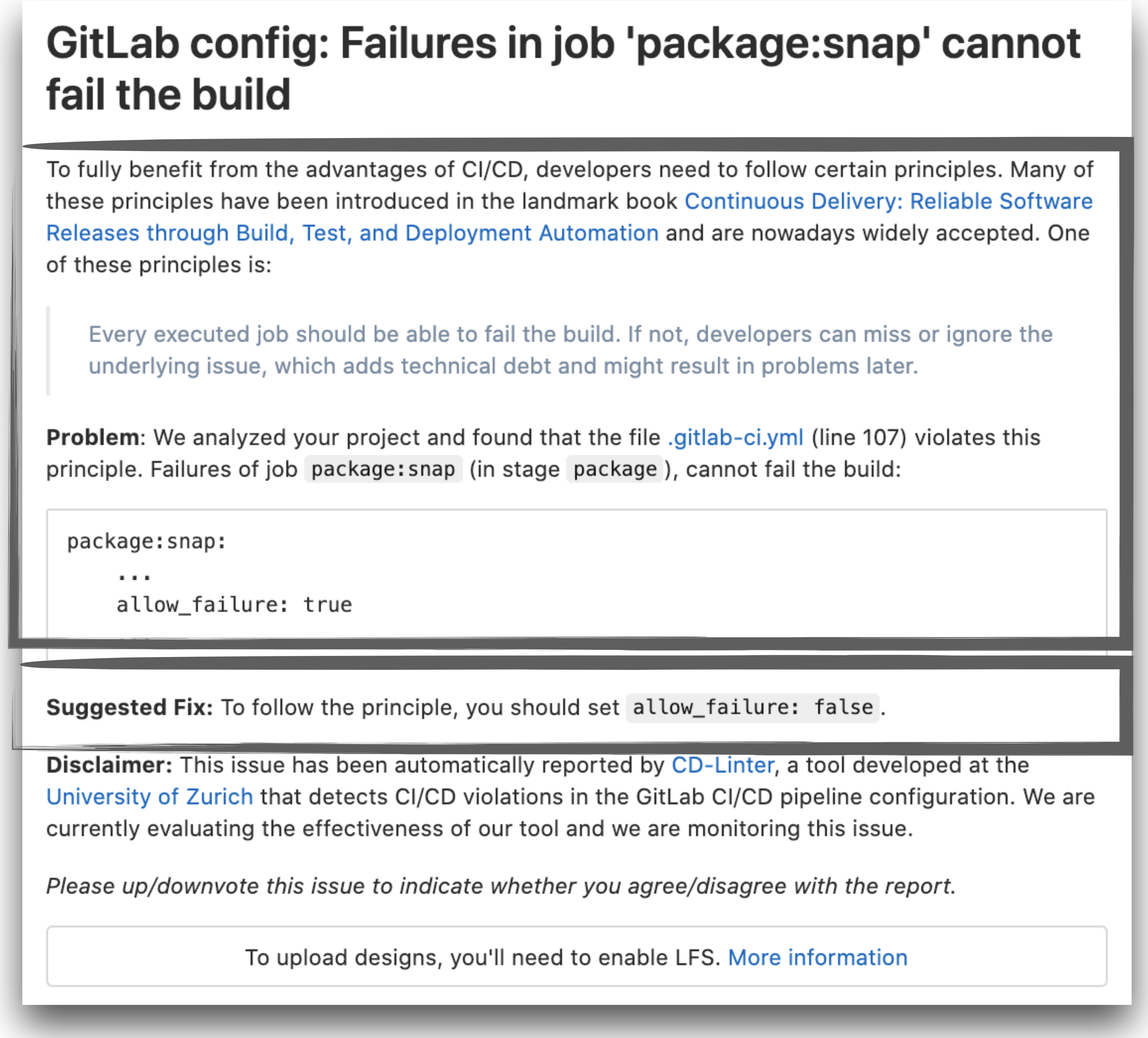
Icons from: https://www.flaticon.com/authors/freepik

RQ 1: GitLab issues reporting CD smells

```
stages:
   - build
   - package
…
package:snap:
  image: ubuntu:18.04
  stage: package
  script:
    - snapcraft
    - echo $SNAPCRAFT_LOGIN_FILE | base64
--decode --ignore-garbage > snapcraft.login
    - snapcraft login --with snapcraft.login
    - snapcraft push *.snap --release beta

  allow_failure: true
…
```

Fake Success

https://gitlab.com/bitseater/meteo/blob/master/.gitlab-ci.yml#L107



GitLab config: Failures in job 'package:snap' cannot fail the build

To fully benefit from the advantages of CI/CD, developers need to follow certain principles. Many of these principles have been introduced in the landmark book Continuous Delivery: Reliable Software Releases through Build, Test, and Deployment Automation and are nowadays widely accepted. One of these principles is:

> Every executed job should be able to fail the build. If not, developers can miss or ignore the underlying issue, which adds technical debt and might result in problems later.

Problem: We analyzed your project and found that the file .gitlab-ci.yml (line 107) violates this principle. Failures of job `package:snap` (in stage `package`), cannot fail the build:

```
package:snap:
    ...
    allow_failure: true
```

Suggested Fix: To follow the principle, you should set `allow_failure: false`.

Disclaimer: This issue has been automatically reported by CD-Linter, a tool developed at the University of Zurich that detects CI/CD violations in the GitLab CI/CD pipeline configuration. We are currently evaluating the effectiveness of our tool and we are monitoring this issue.

Please up/downvote this issue to indicate whether you agree/disagree with the report.

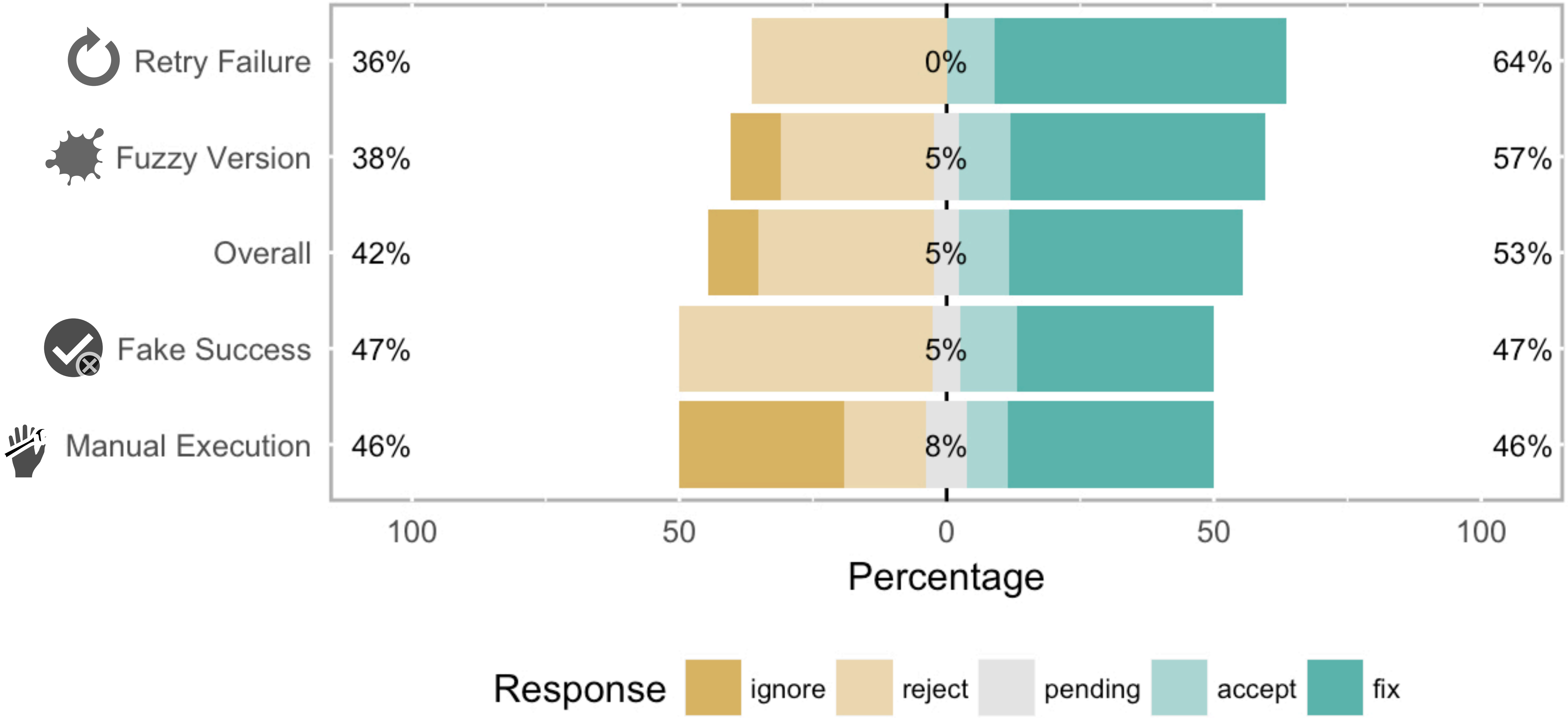
To upload designs, you'll need to enable LFS. More information

Problem

Fix

https://gitlab.com/bitseater/meteo/-/issues/125

RQ 1: Reactions to issues



@ccvassallo

RQ 1: Reasons for rejecting issues

Fake Success

- Warned jobs are not essential or not fully implemented yet
- The CD smell is contained in a template

Retry Failure

- Warned jobs are executed on out-of-control machines

Manual Execution

- Lack of trust in automated issue reporting
- Warned jobs are not fully integrated yet

Fuzzy Version

- Tools should be automatically updated to the latest version

@ccvassallo

RQ 1: Reasons for rejecting issues

Fake Success

- Warned jobs are not essential or not fully implemented yet
- The CD smell is contained in a template

Retry Failure

- Warned jobs are executed on out-of-control machines

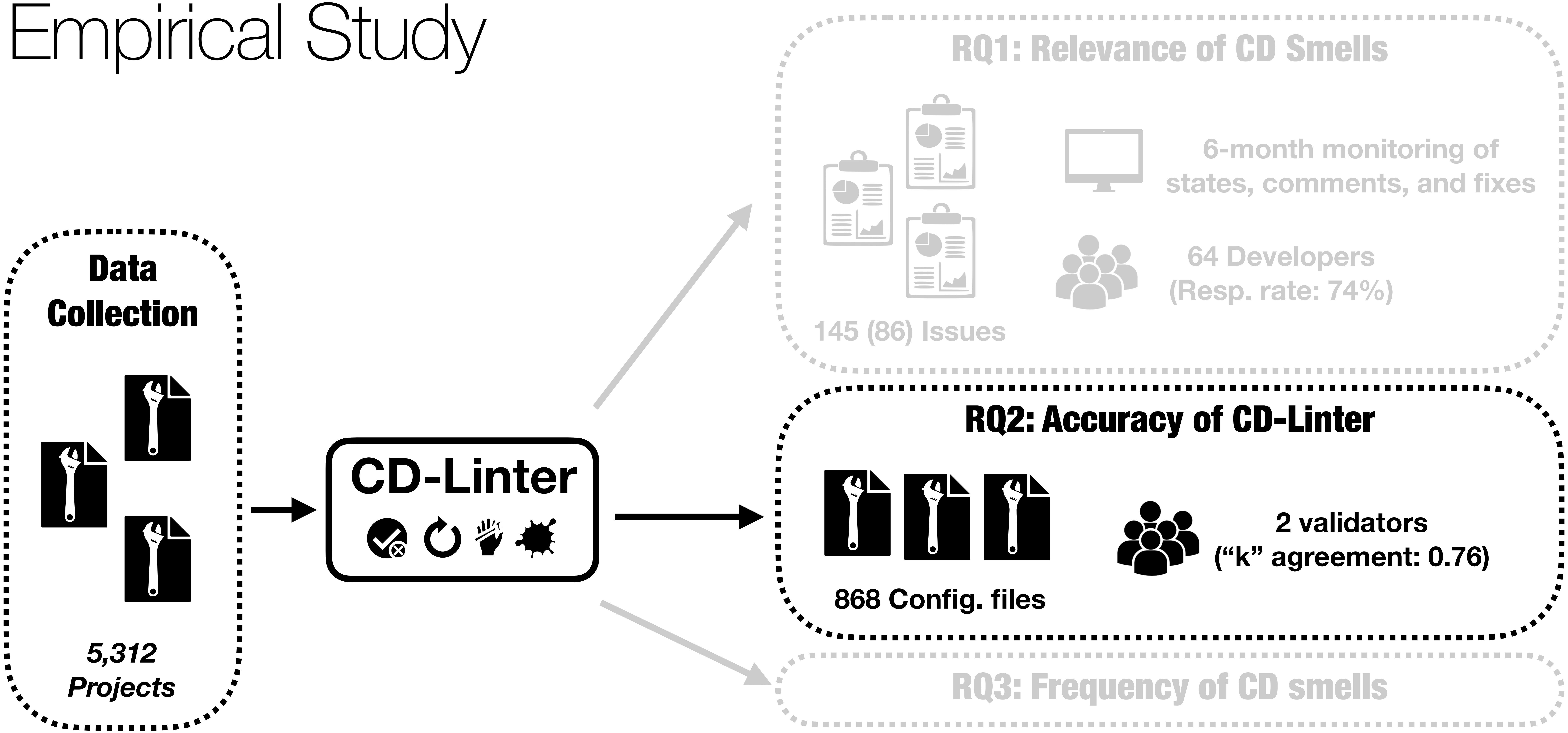
Manual Execution

- Lack of trust in automated issue reporting
- Warned jobs are not fully integrated yet

Fuzzy Version

- **Tools should be automatically updated to the latest version**

Empirical Study



Data Collection

5,312 Projects

CD-Linter

RQ1: Relevance of CD Smells

145 (86) Issues

6-month monitoring of states, comments, and fixes

64 Developers (Resp. rate: 74%)

RQ2: Accuracy of CD-Linter

868 Config. files

2 validators (“k” agreement: 0.76)

RQ3: Frequency of CD smells

RQ 2: Accuracy of CD-Linter

Precision: 87%

False positives:

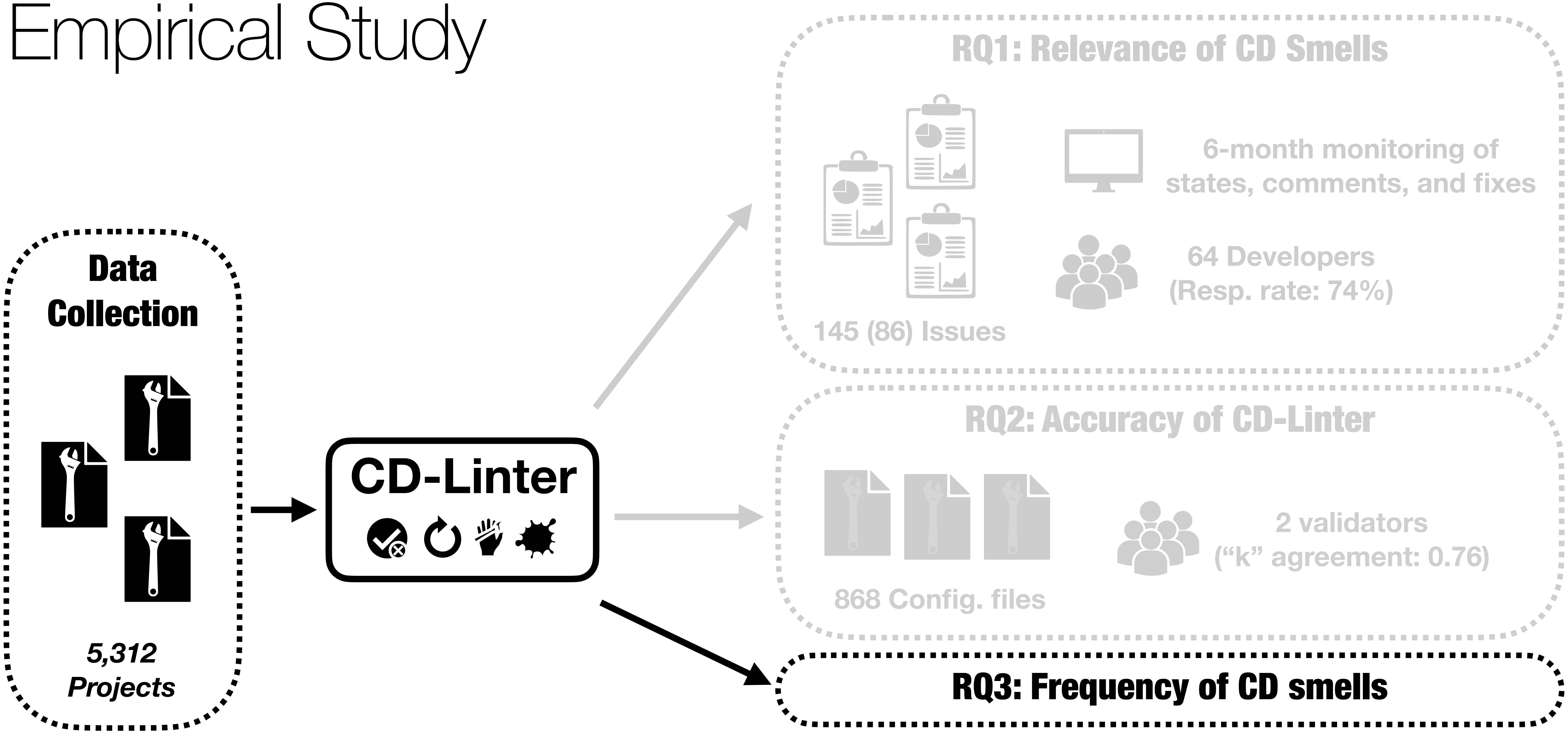
- Jobs (with unconventional names) executed in a release stage (*Manual Execution*)
- Tool dependencies without versions (*Fuzzy Version*)

Recall: 94%

False negatives:

- Dependencies specified in a .pip file (*Fuzzy Version*)
- Jobs with release-related names (*Manual Execution*)

Empirical Study



Data Collection

5,312 Projects

CD-Linter

RQ1: Relevance of CD Smells

145 (86) Issues

6-month monitoring of states, comments, and fixes

64 Developers (Resp. rate: 74%)

RQ2: Accuracy of CD-Linter

868 Config. files

2 validators (“k” agreement: 0.76)

RQ3: Frequency of CD smells

RQ 3: Frequency of CD smells

The majority of detected smells (70%) affect projects with long configuration files

- 31% of them are affected by at least one CD smell

Fake Success	Retry Failure	Manual Execution	Fuzzy Version
⚠ 17% of projects	⚠ 6% of projects	⚠ 4% of projects	⚠ 40% of projects

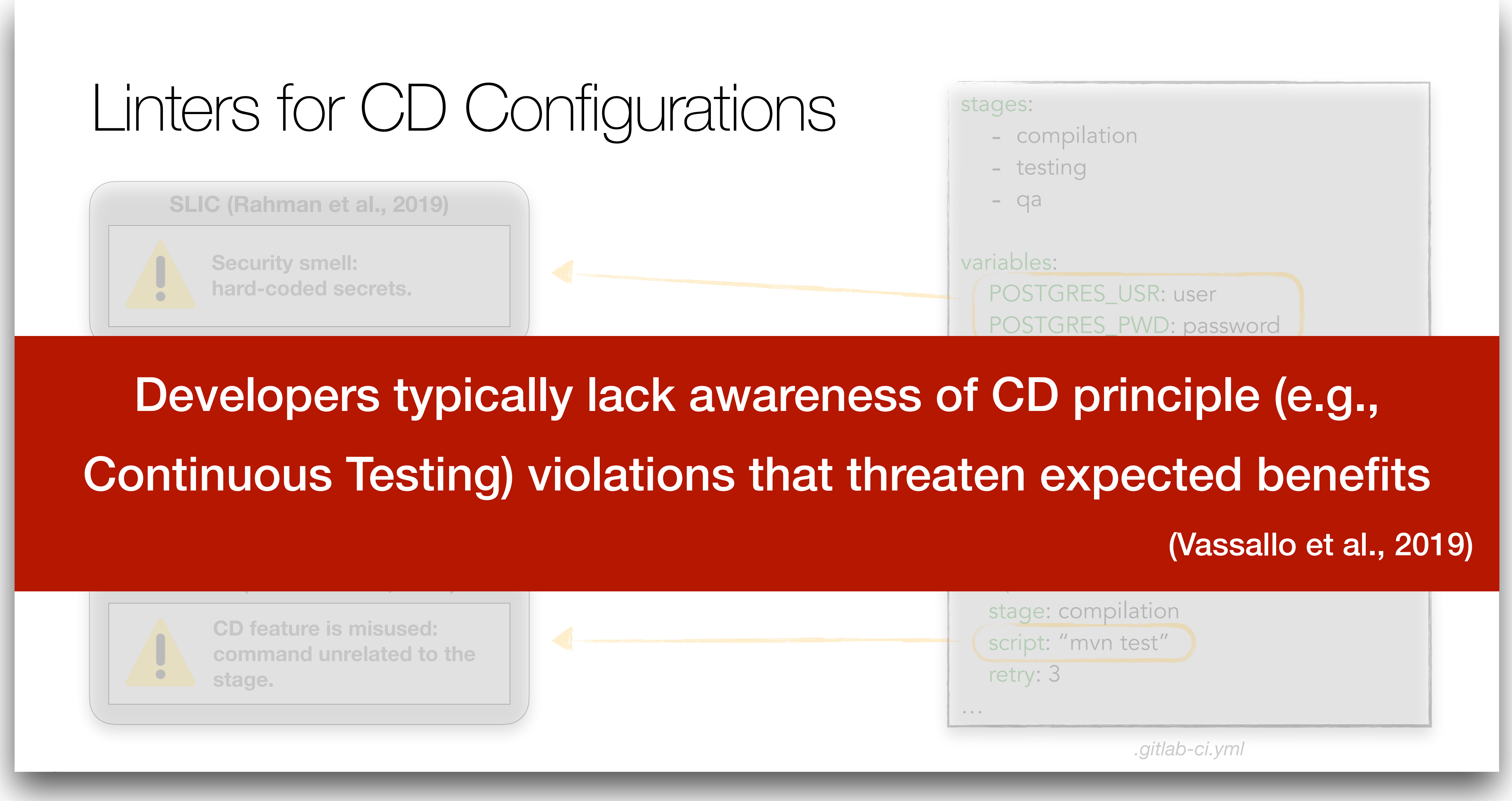
Implications

CD-Linter as a *mentor*
when configuring CD pipelines

Linting rules have to be approved
by developers

Long and complex CD
configurations are often smelly

@ccvassallo



Linters for CD Configurations

SLIC (Rahman et al., 2019)

Security smell: hard-coded secrets.

CD feature is misused: command unrelated to the stage.

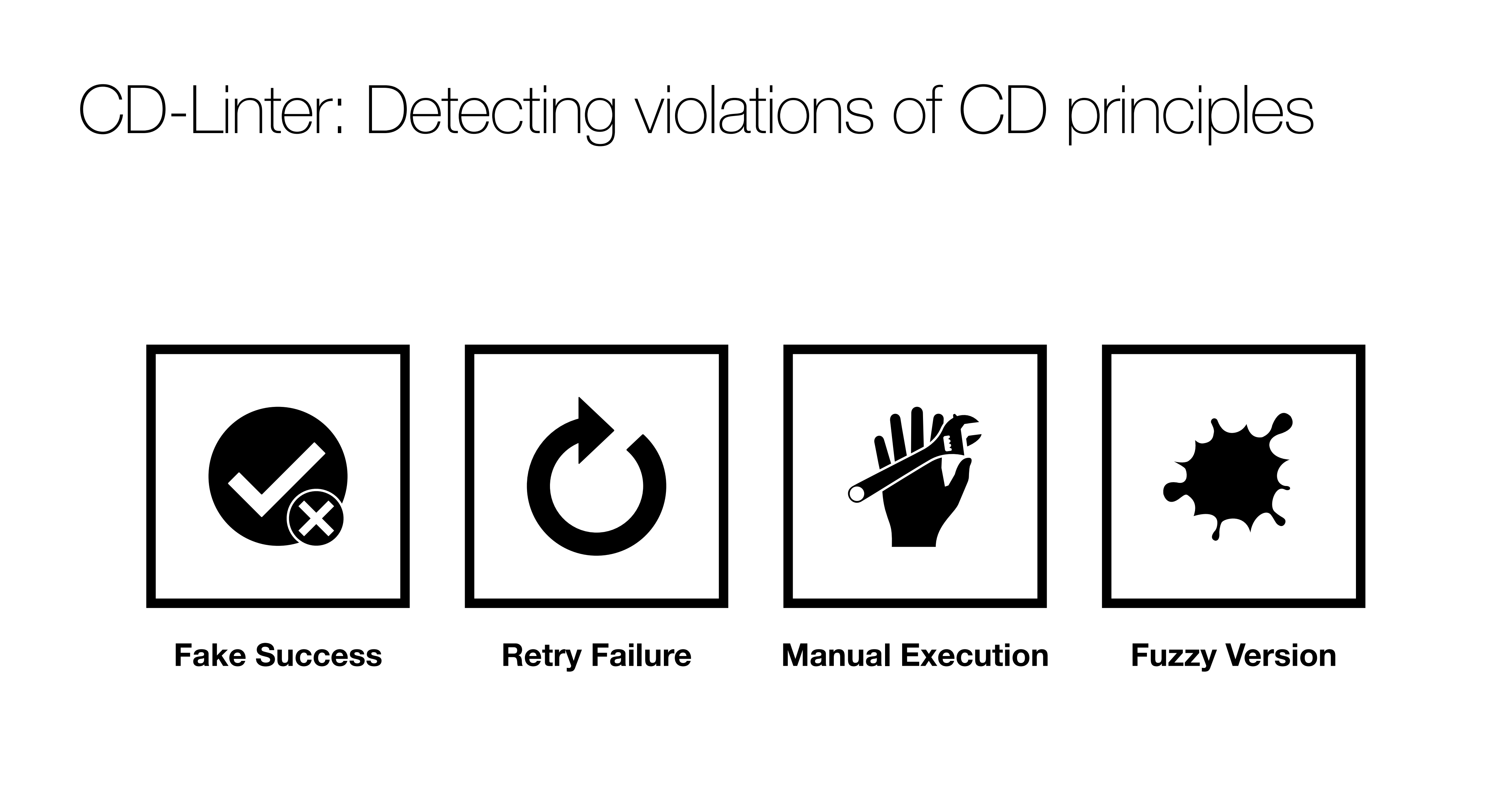
```
stages:
  - compilation
  - testing
  - qa

variables:
  POSTGRES_USR: user
  POSTGRES_PWD: password
```

```
  stage: compilation
  script: "mvn test"
  retry: 3
...
```

.gitlab-ci.yml

Developers typically lack awareness of CD principle (e.g., Continuous Testing) violations that threaten expected benefits

(Vassallo et al., 2019)



CD-Linter: Detecting violations of CD principles

Getting Continuous Testing Done Right with CD-Linter

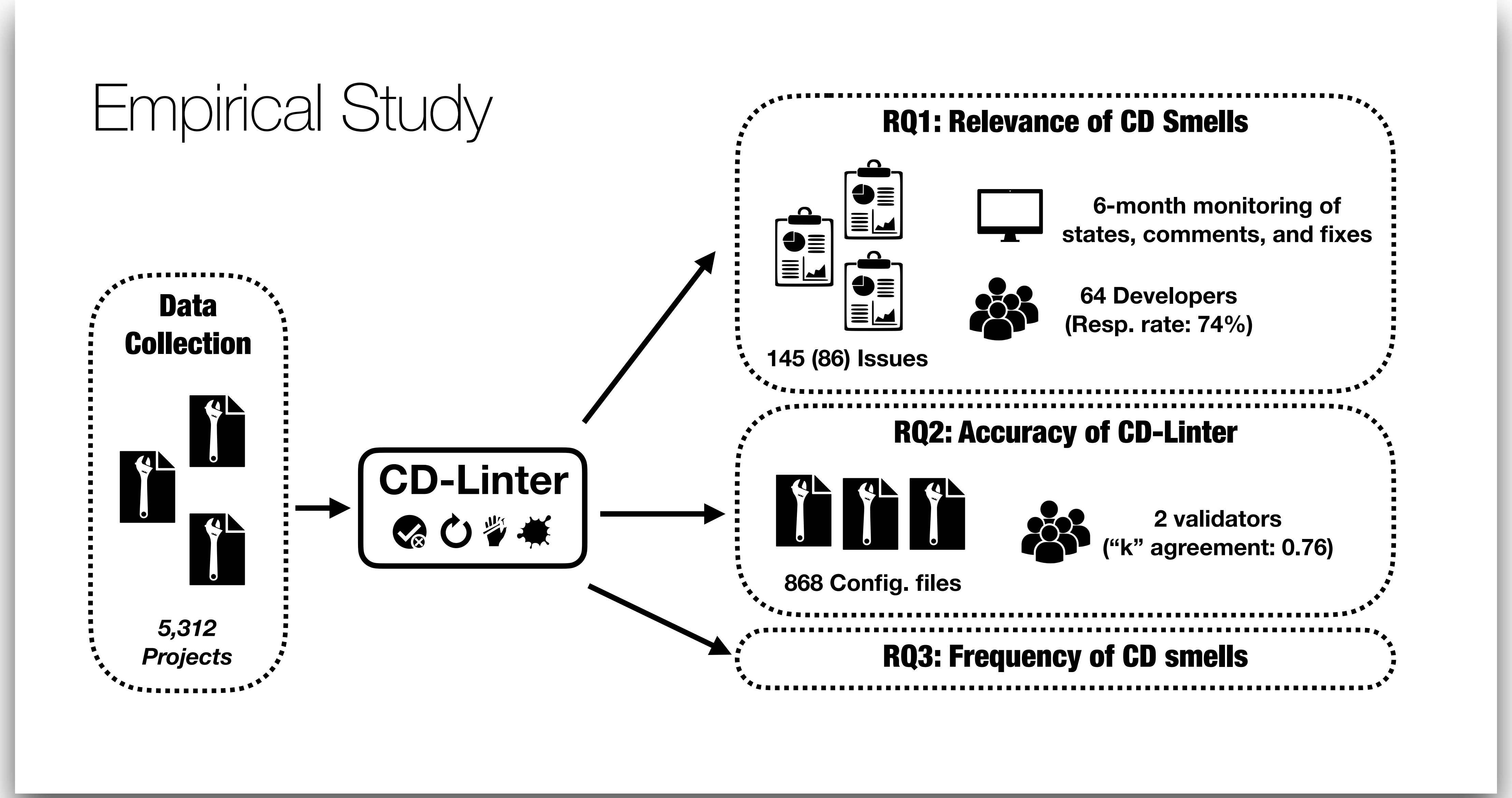
Carmine Vassallo

vassallocarmine@gmail.com

@ccvassallo

http://tiny.uzh.ch/WV

I'm on the Job Market!



Empirical Study

Data Collection
5,312 Projects
CD-Linter
RQ1: Relevance of CD Smells
145 (86) Issues
6-month monitoring of states, comments, and fixes
64 Developers (Resp. rate: 74%)
RQ2: Accuracy of CD-Linter
868 Config. files
2 validators ("k" agreement: 0.76)
RQ3: Frequency of CD smells