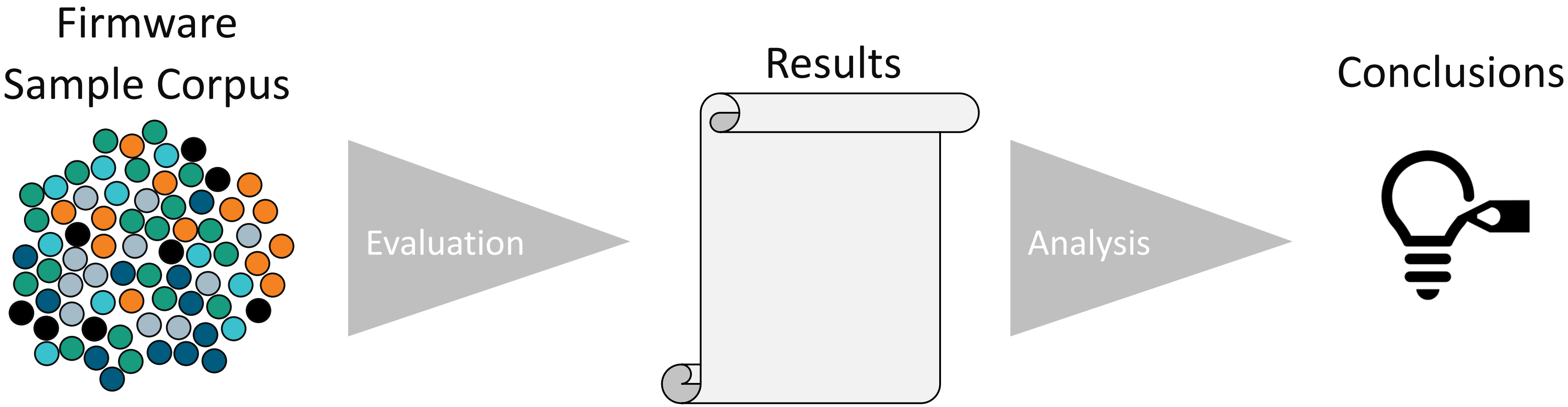
Mens Sana In Corpore Sano

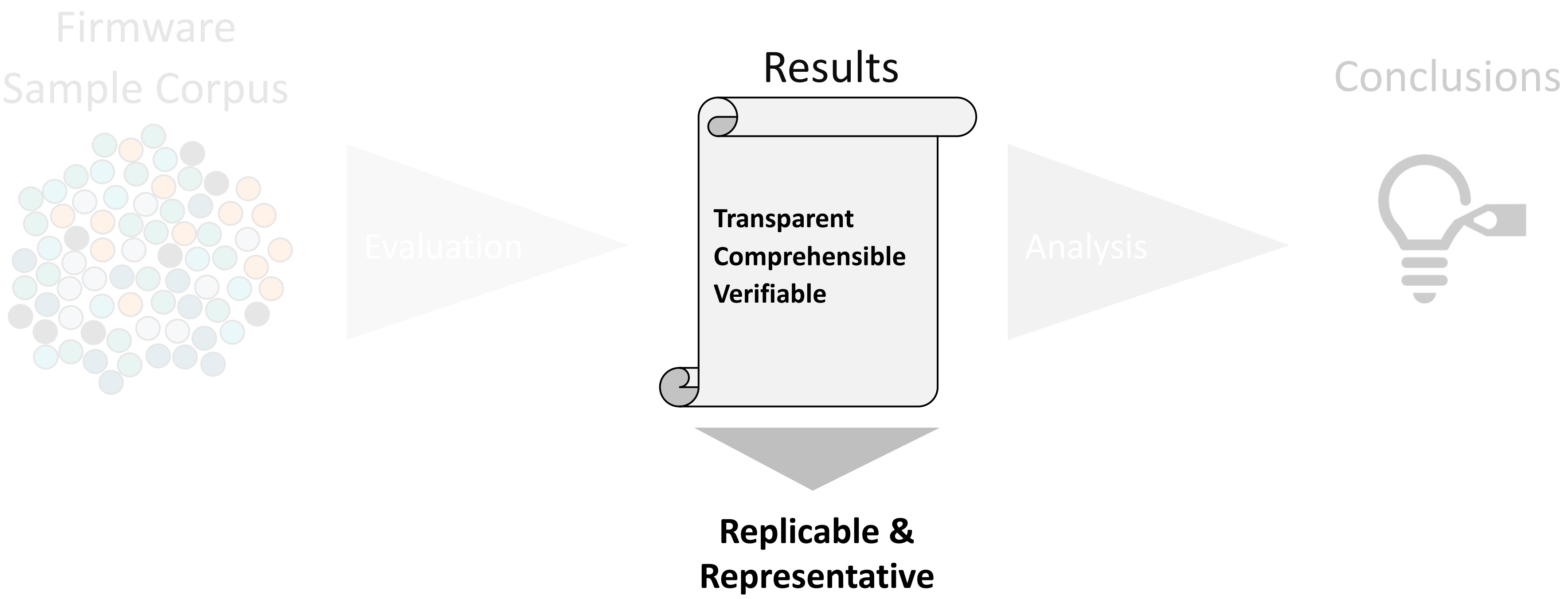
Sound Firmware Corpora for Vulnerability Research

René Helmke, Elmar Padilla,
& Nils Aschenbruck
Fraunhofer FKIE & University of Osnabrück
Germany

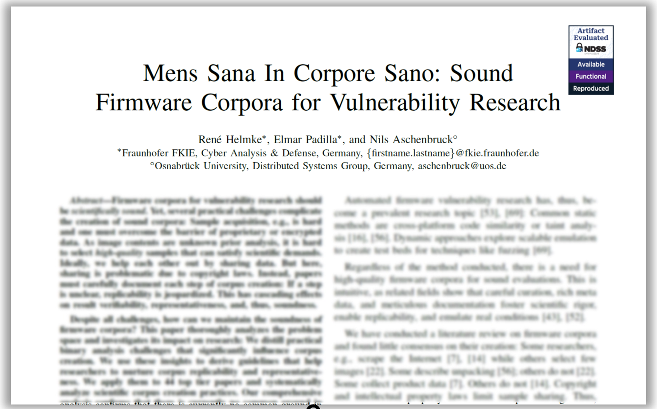
Building, sharing, and documenting evaluation datasets.



Scientific Soundness.



Analysis: How can we help researchers to build scientifically sound firmware corpora?



Artifact Evaluated NDSS
Available
Functional
Reproduced

Mens Sana In Corpore Sano: Sound Firmware Corpora for Vulnerability Research

René Helmke*, Elmar Padilla*, and Nils Aschenbruck°
*Fraunhofer FKIE, Cyber Analysis & Defense, Germany, {firstname.lastname}@fkie.fraunhofer.de
°Osnabrück University, Distributed Systems Group, Germany, aschenbruck@uos.de

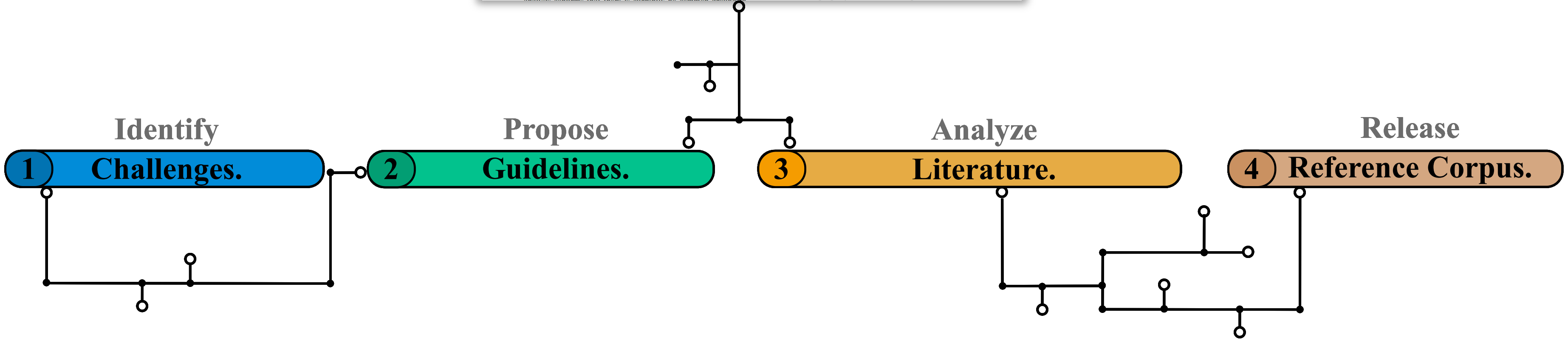
Abstract—Firmware corpora for vulnerability research should be scientifically sound. Yet, several practical challenges complicate the creation of sound corpora: Sample acquisition, e.g., is hard and one must overcome the barrier of proprietary or encrypted data. As image contents are unknown prior analysis, it is hard to select *high-quality* samples that can satisfy scientific demands. Ideally, we help each other out by sharing data. But here, sharing is problematic due to copyright laws. Instead, papers must carefully document each step of corpus creation: If a step is unclear, replicability is jeopardized. This has cascading effects on result verifiability, representativeness, and, thus, soundness.

Despite all challenges, how can we maintain the soundness of firmware corpora? This paper thoroughly analyzes the problem space and investigates its impact on research: We distill practical binary analysis challenges that significantly influence corpus creation. We use these insights to derive guidelines that help researchers to nurture corpus replicability and representativeness. We apply them to 44 top tier papers and systematically analyze scientific corpus creation practices. Our comprehensive analysis confirms that there is currently no common ground in

Automated firmware vulnerability research has, thus, become a prevalent research topic [53], [69]. Common static methods are cross-platform code similarity or taint analysis [16], [56]. Dynamic approaches explore scalable emulation to create test beds for techniques like fuzzing [69].

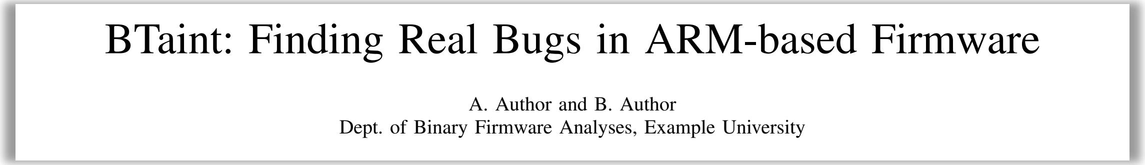
Regardless of the method conducted, there is a need for high-quality firmware corpora for sound evaluations. This is intuitive, as related fields show that careful curation, rich meta data, and meticulous documentation foster scientific rigor, enable replicability, and emulate real conditions [43], [52].

We have conducted a literature review on firmware corpora and found little consensus on their creation: Some researchers, e.g., scrape the Internet [7], [14] while others select few images [22]. Some describe unpacking [56]; others do not [22]. Some collect product data [7]. Others do not [14]. Copyright and intellectual property laws limit sample sharing. Thus,



Understanding the problem space: Why is it hard to create sound corpora?

Example (not a real paper)



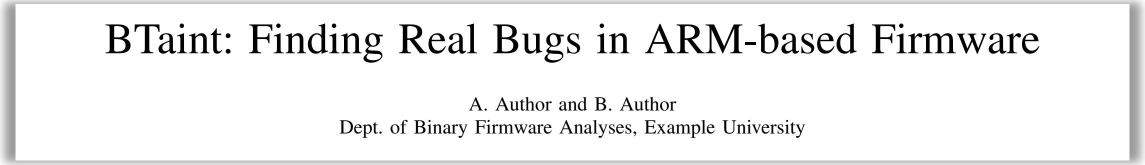
BTaint: Finding Real Bugs in ARM-based Firmware

A. Author and B. Author
Dept. of Binary Firmware Analyses, Example University

Goal: Create firmware corpus with 1000 samples.

Understanding the problem space: Why is it hard to create sound corpora?

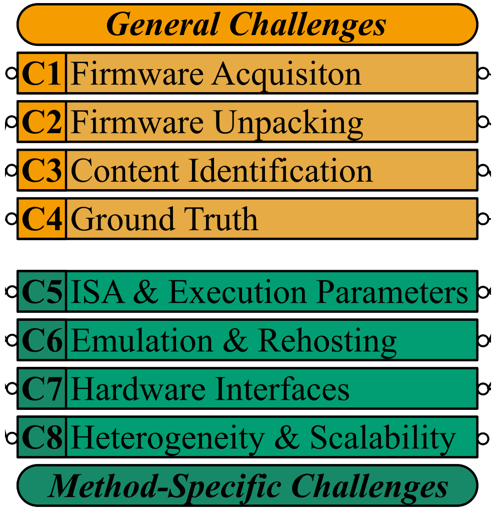
Example (not a real paper)



BTaint: Finding Real Bugs in ARM-based Firmware

A. Author and B. Author
Dept. of Binary Firmware Analyses, Example University

Goal: Create firmware corpus with 1000 samples.



General Challenges

C1 Firmware Acquisiton

C2 Firmware Unpacking

C3 Content Identification

C4 Ground Truth

C5 ISA & Execution Parameters

C6 Emulation & Rehosting

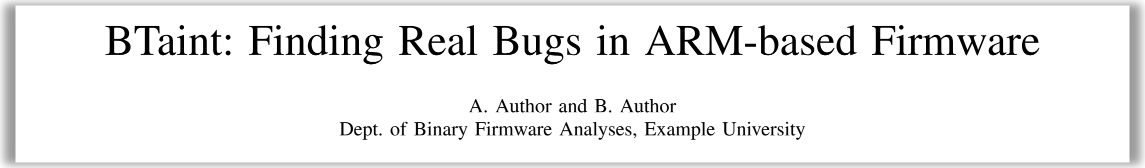
C7 Hardware Interfaces

C8 Heterogeneity & Scalability

Method-Specific Challenges

Understanding the problem space: Why is it hard to create sound corpora?

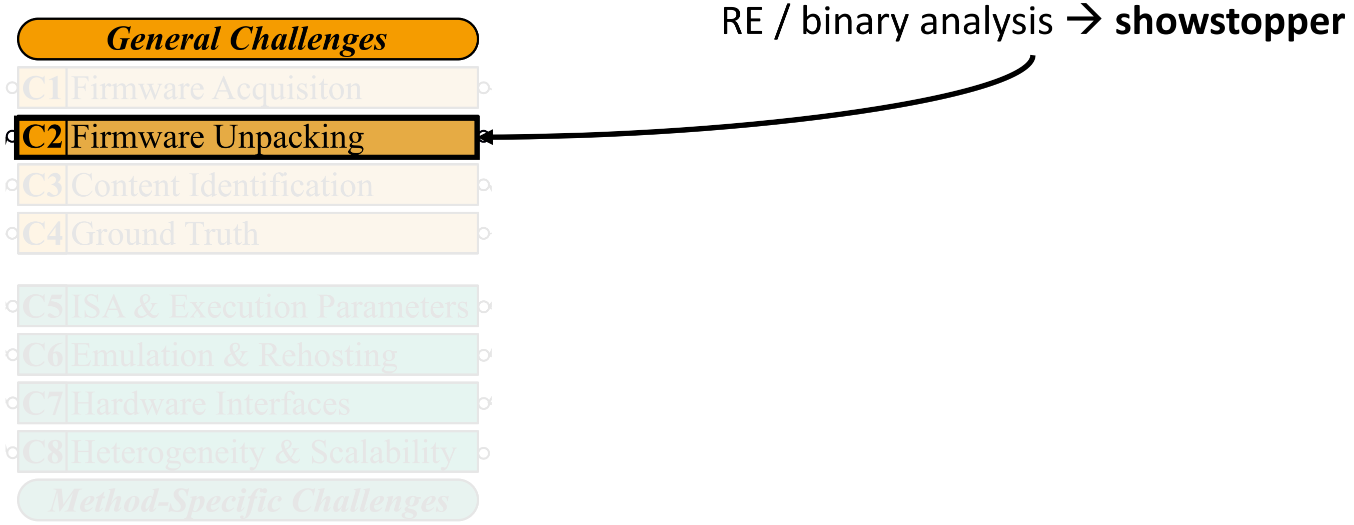
Example (not a real paper)



BTaint: Finding Real Bugs in ARM-based Firmware

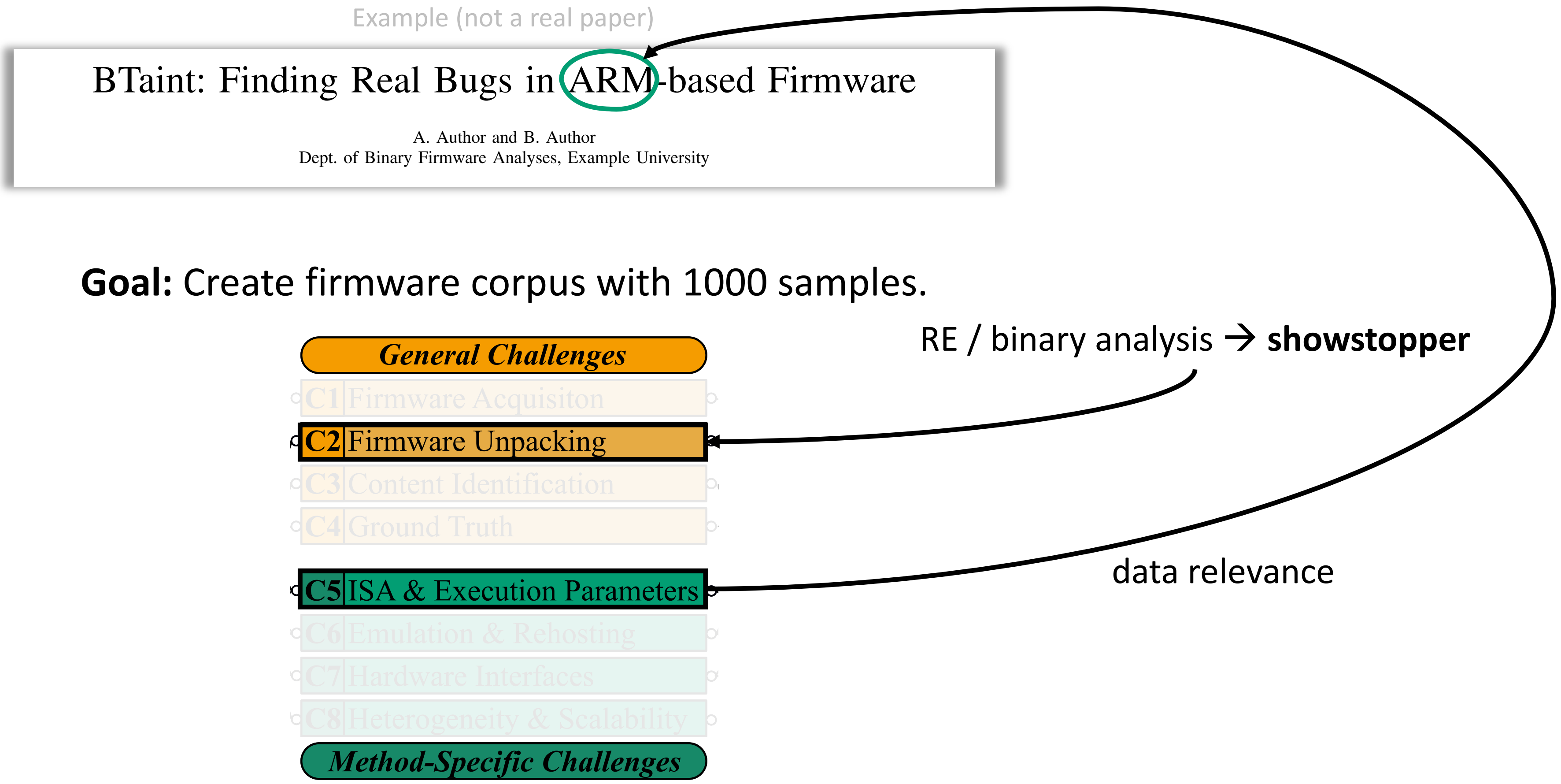
A. Author and B. Author
Dept. of Binary Firmware Analyses, Example University

Goal: Create firmware corpus with 1000 samples.



RE / binary analysis → **showstopper**

Understanding the problem space: Why is it hard to create sound corpora?



Example (not a real paper)

BTaint: Finding Real Bugs in ARM-based Firmware

A. Author and B. Author
Dept. of Binary Firmware Analyses, Example University

Goal: Create firmware corpus with 1000 samples.

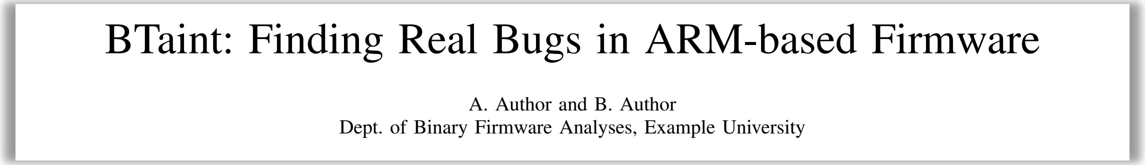
RE / binary analysis → **showstopper**

data relevance

R. Helmke et al., *Mens Sana In Corpore Sano: Sound Firmware Corpora for Vulnerability Research*

Understanding the problem space: Why is it hard to create sound corpora?

Example (not a real paper)



BTaint: Finding Real Bugs in ARM-based Firmware

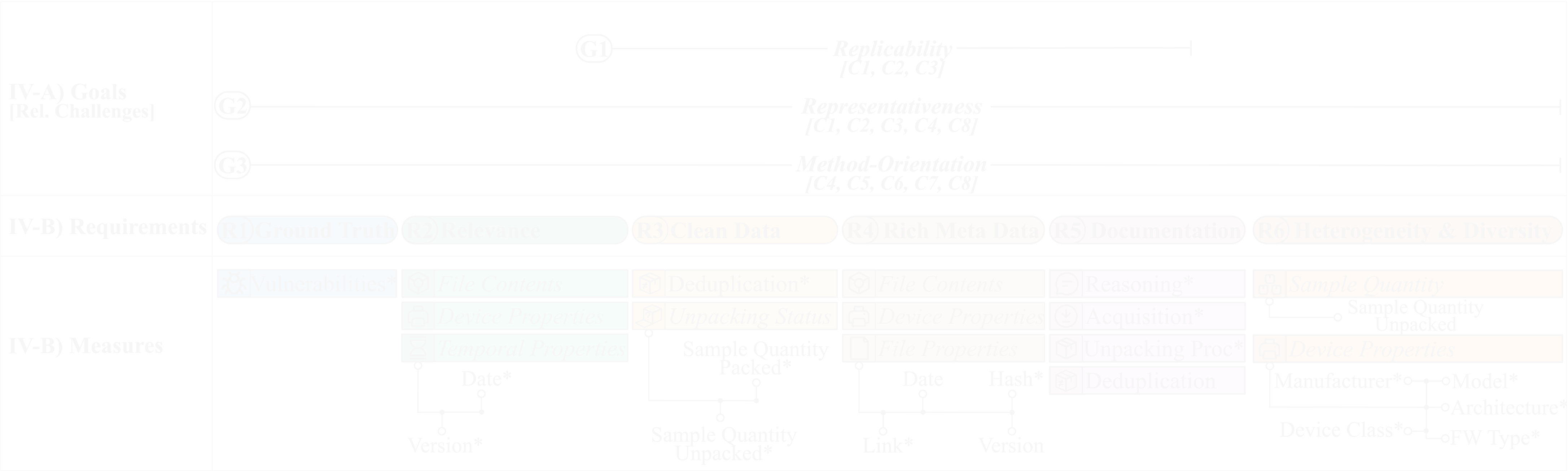
A. Author and B. Author
Dept. of Binary Firmware Analyses, Example University

Goal: Share firmware corpus with 1000 samples.

Illegal: Copyright in firmware images.

Preserve data replicability: Document everything.

C2: Guidelines to create scientifically sound firmware corpora.



IV-A) Goals
[Rel. Challenges]

G1 *Replicability* *[C1, C2, C3]*

G2 *Representativeness* *[C1, C2, C3, C4, C8]*

G3 *Method-Orientation* *[C4, C5, C6, C7, C8]*

IV-B) Requirements

R1 Ground Truth

R2 Relevance

R3 Clean Data

R4 Rich Meta Data

R5 Documentation

R6 Heterogeneity & Diversity

IV-B) Measures

Vulnerabilities*

File Contents
Device Properties
Temporal Properties
Date*
Version*

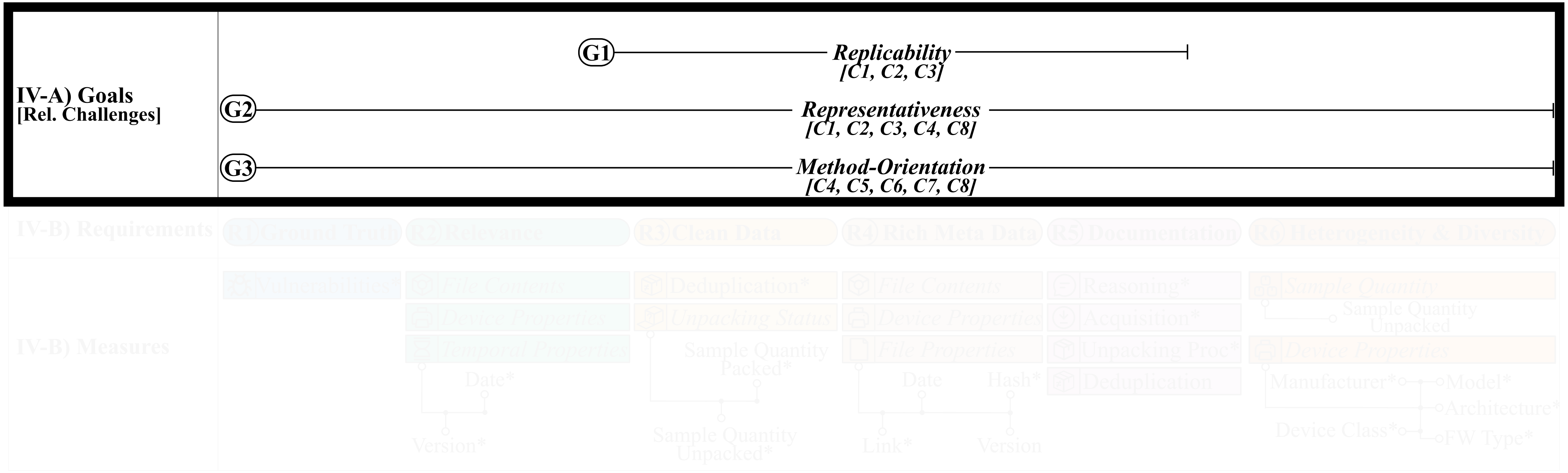
Deduplication*
Unpacking Status
Sample Quantity Packed*
Sample Quantity Unpacked*

File Contents
Device Properties
File Properties
Date
Hash*
Link*
Version

Reasoning*
Acquisition*
Unpacking Proc*
Deduplication

Sample Quantity
Sample Quantity Unpacked
Device Properties
Manufacturer*
Model*
Architecture*
Device Class*
FW Type*

Layer 1: Abstract corpus goals to improve soundness.



IV-A) Goals
[Rel. Challenges]

G1 ***Replicability***
[C1, C2, C3]

G2 ***Representativeness***
[C1, C2, C3, C4, C8]

G3 ***Method-Orientation***
[C4, C5, C6, C7, C8]

IV-B) Requirements

R1 **Ground Truth**
R2 **Relevance**
R3 **Clean Data**
R4 **Rich Meta Data**
R5 **Documentation**
R6 **Heterogeneity & Diversity**

IV-B) Measures

Vulnerabilities*

File Contents
Device Properties
Temporal Properties
Date*
Version*

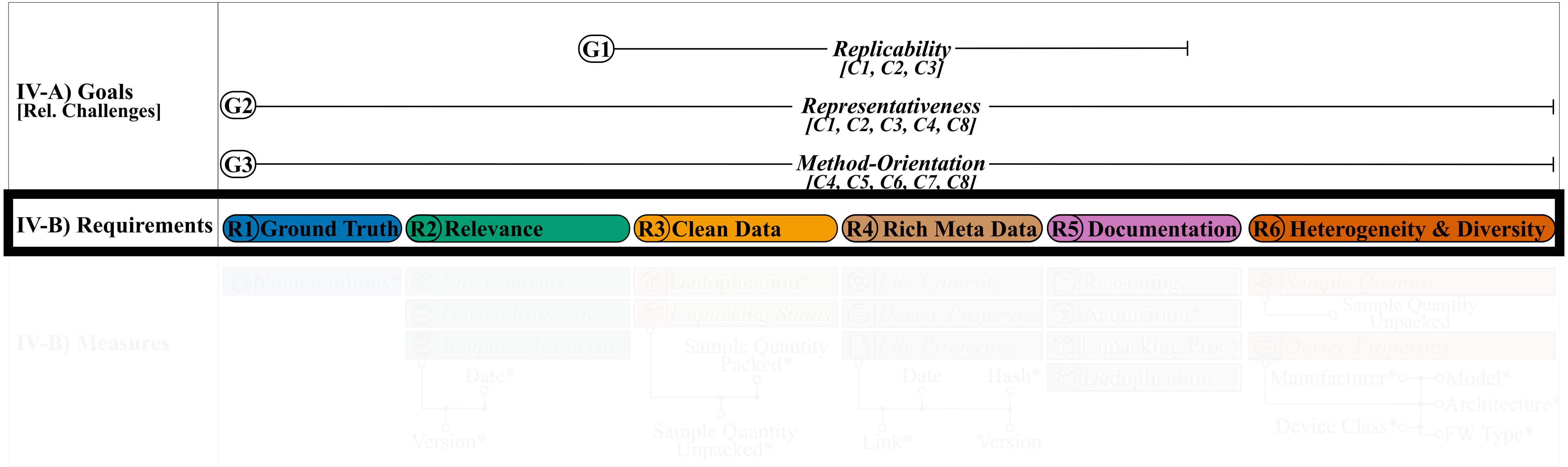
Deduplication*
Unpacking Status
Sample Quantity Packed*
Sample Quantity Unpacked*

File Contents
Device Properties
File Properties
Date
Hash*
Link*
Version

Reasoning*
Acquisition*
Unpacking Proc*
Deduplication

Sample Quantity
Sample Quantity Unpacked
Device Properties
Manufacturer*
Model*
Architecture*
Device Class*
FW Type*

Layer 2: Key requirements that nurture the three goals.



IV-A) Goals
[Rel. Challenges]

G1 *Replicability*
[C1, C2, C3]

G2 *Representativeness*
[C1, C2, C3, C4, C8]

G3 *Method-Orientation*
[C4, C5, C6, C7, C8]

IV-B) Requirements

R1 Ground Truth

R2 Relevance

R3 Clean Data

R4 Rich Meta Data

R5 Documentation

R6 Heterogeneity & Diversity

IV-B) Measures

Vulnerabilities*

File Contents
Device Properties
Temporal Properties
Date*
Version*

Deduplication*
Unpacking Status
Sample Quantity Packed*
Sample Quantity Unpacked*

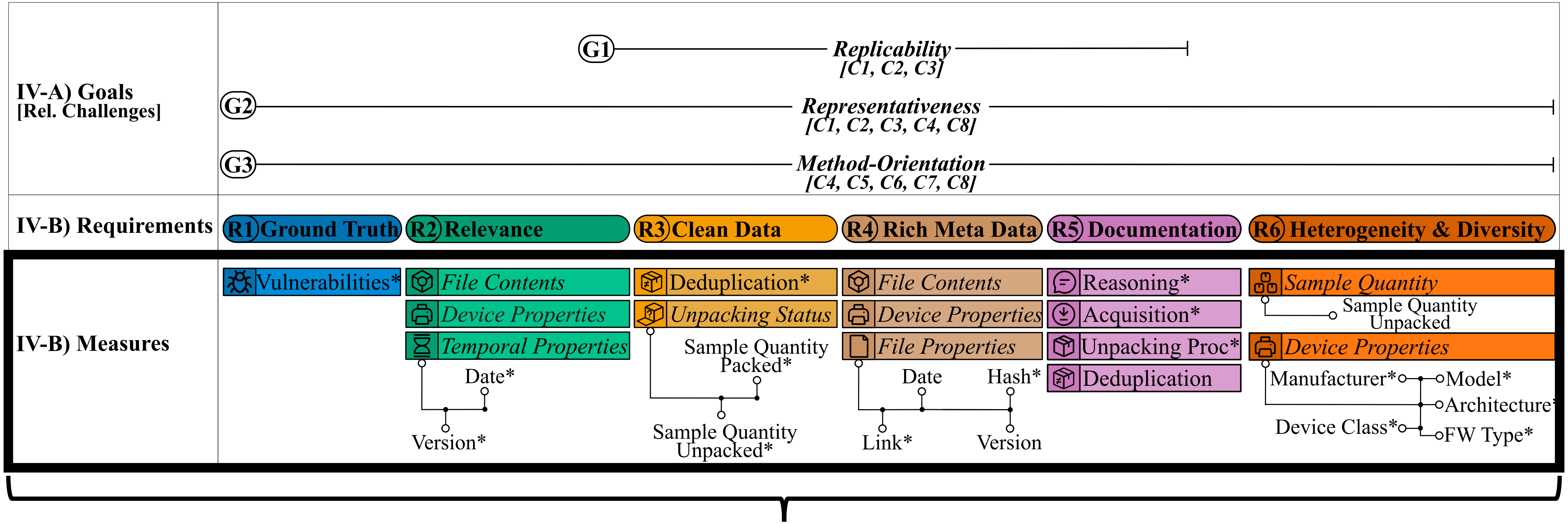
File Contents
Device Properties
File Properties
Date
Hash*
Link*
Version

Reasoning*
Acquisition*
Unpacking Proc*
Deduplication

Sample Quantity
Sample Quantity Unpacked
Device Properties
Manufacturer*
Model*
Architecture*
Device Class*
FW Type*

R. Helmke et al., *Mens Sana In Corpore Sano: Sound Firmware Corpora for Vulnerability Research*

Layer 3: Concrete measures to estimate requirement fulfillment.



IV-A) Goals
[Rel. Challenges]

G1 *Replicability* *[C1, C2, C3]*

G2 *Representativeness* *[C1, C2, C3, C4, C8]*

G3 *Method-Orientation* *[C4, C5, C6, C7, C8]*

IV-B) Requirements

R1 Ground Truth | R2 Relevance | R3 Clean Data | R4 Rich Meta Data | R5 Documentation | R6 Heterogeneity & Diversity

IV-B) Measures

R1: Vulnerabilities*

R2: *File Contents*; *Device Properties*; *Temporal Properties* (Date*, Version*)

R3: Deduplication*; *Unpacking Status* (Sample Quantity Packed*, Sample Quantity Unpacked*)

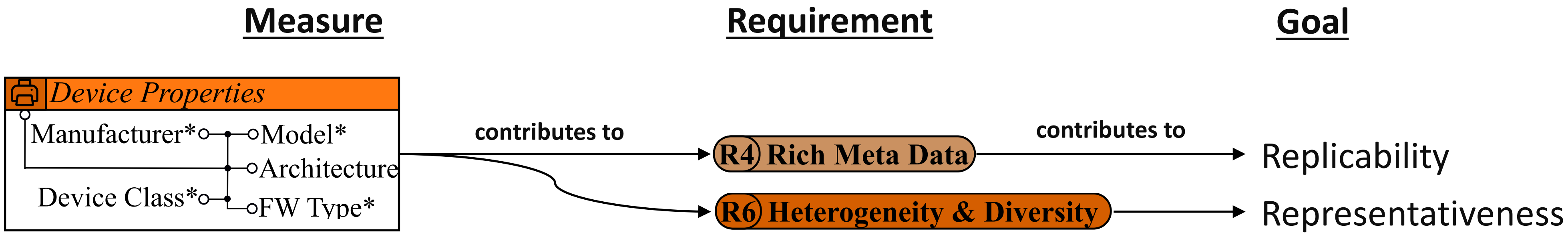
R4: *File Contents*; *Device Properties*; *File Properties* (Date, Hash*, Link*, Version)

R5: Reasoning*; Acquisition*; Unpacking Proc*; Deduplication

R6: *Sample Quantity* (Sample Quantity Unpacked); *Device Properties* (Manufacturer*, Model*, Architecture*, Device Class*, FW Type*)

16 unique measures

Layer 3: Measure examples.



Measure

Requirement

Goal

Device Properties

Manufacturer*

Model*

Architecture

Device Class*

FW Type*

contributes to

R4 Rich Meta Data

contributes to

Replicability

R6 Heterogeneity & Diversity

Representativeness

C3: An analysis of state of the art corpus creation practices in current research.

collected

44 papers

from

NDSS, S&P, USENIX Security, CCS

(and few others, referenced by A papers)*

published

2013 – 2023

criterion

create/use firmware corpus for vulnerability research

C3: An analysis of state of the art corpus creation practices in current research.

collected

44 papers

from

NDSS, S&P, USENIX Security, CCS

(and few others, referenced by A papers)*

published

2013 – 2023

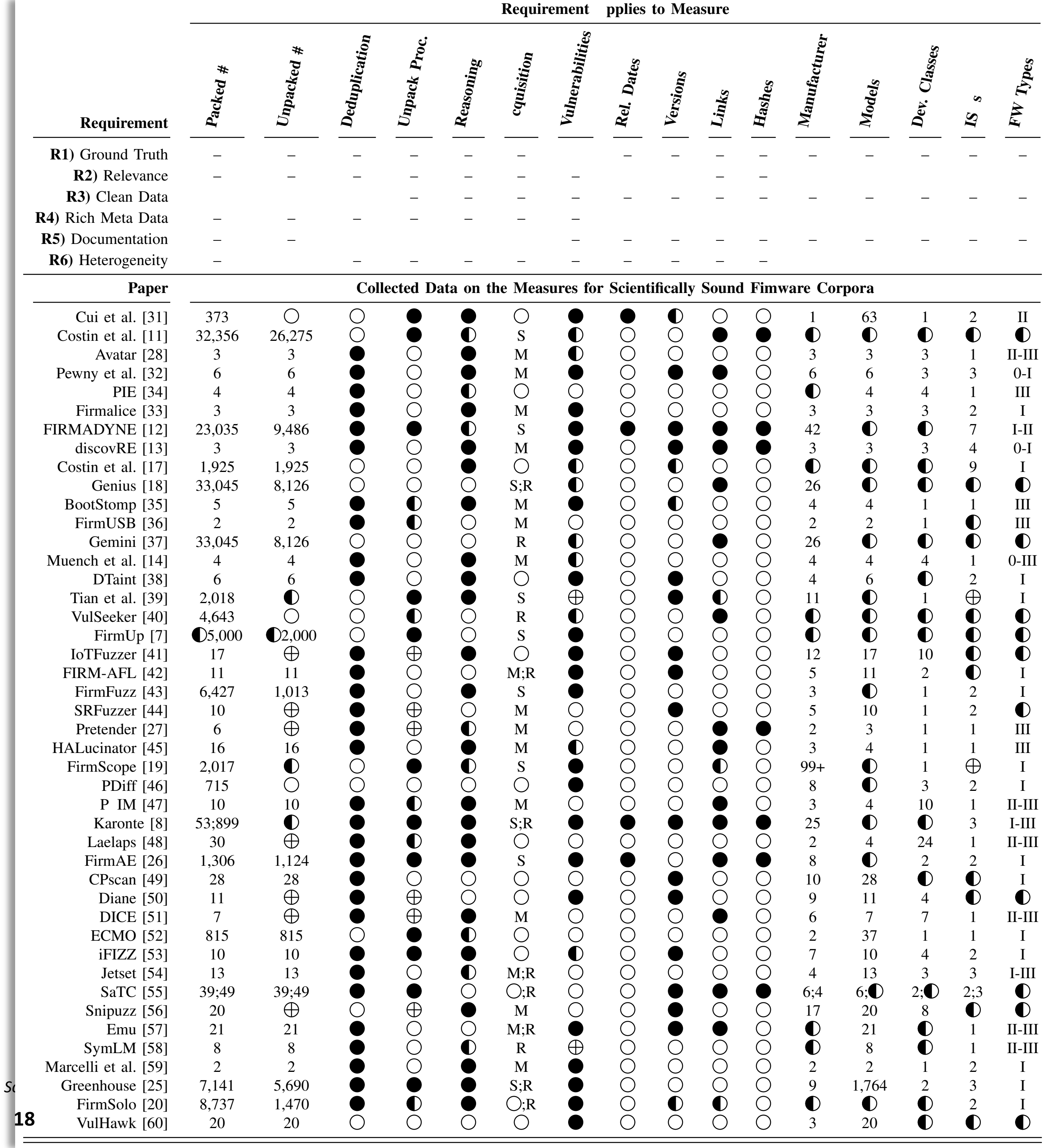
criterion

create/use firmware corpus for vulnerability research

→

Read,
analyze artifacts,
& collect data.

C3: An analysis of state of the art corpus creation practices in current research.



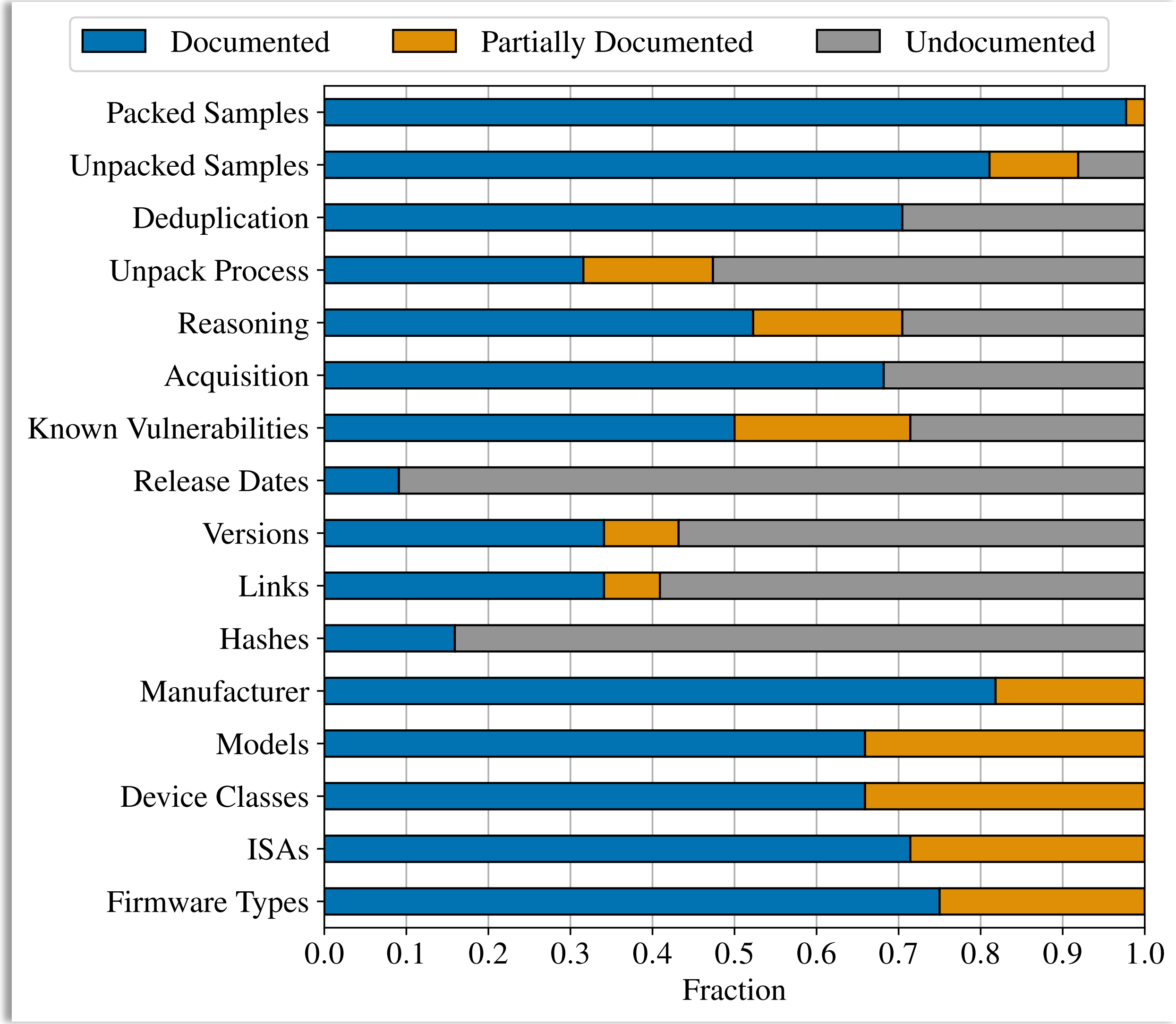
	Requirement pplies to Measure															
Requirement	***Packed #***	***Unpacked #***	***Deduplication***	***Unpack Proc.***	***Reasoning***	***cquisition***	***Vulnerabilities***	***Rel. Dates***	***Versions***	***Links***	***Hashes***	***Manufacturer***	***Models***	***Dev. Classes***	***IS s***	***FW Types***
R1) Ground Truth	–	–	–	–	–	–		–	–	–	–	–	–	–	–	–
R2) Relevance	–	–	–	–	–	–	–			–	–					
R3) Clean Data				–	–	–	–	–	–	–	–	–	–	–	–	–
R4) Rich Meta Data	–	–	–	–	–	–	–									
R5) Documentation	–	–					–	–	–	–	–	–	–	–	–	–
R6) Heterogeneity	–		–	–	–	–	–	–	–	–	–					
Paper	**Collected Data on the Measures for Scientifically Sound Fimware Corpora**															
Cui et al. [31]	373	○	○	●	●	○	●	●	◐	○	○	1	63	1	2	II
Costin et al. [11]	32,356	26,275	○	●	◐	S	◐	○	○	●	●	◐	◐	◐	◐	◐
Avatar [28]	3	3	●	○	●	M	◐	○	○	○	○	3	3	3	1	II-III
Pewny et al. [32]	6	6	●	○	●	M	●	○	●	●	○	6	6	3	3	0-I
PIE [34]	4	4	●	○	◐	○	○	○	○	○	○	◐	4	4	1	III
Firmalice [33]	3	3	●	○	●	M	●	○	○	○	○	3	3	3	2	I
FIRMADYNE [12]	23,035	9,486	●	●	◐	S	●	●	●	●	●	42	◐	◐	7	I-II
discovRE [13]	3	3	●	○	●	M	●	○	●	●	●	3	3	3	4	0-I
Costin et al. [17]	1,925	1,925	○	○	●	○	◐	○	◐	○	○	◐	◐	◐	9	I
Genius [18]	33,045	8,126	○	○	○	S;R	◐	○	○	●	○	26	◐	◐	◐	◐
BootStomp [35]	5	5	●	◐	●	M	●	○	◐	○	○	4	4	1	1	III
FirmUSB [36]	2	2	●	◐	○	M	○	○	○	○	○	2	2	1	◐	III
Gemini [37]	33,045	8,126	○	○	○	R	◐	○	○	●	○	26	◐	◐	◐	◐
Muench et al. [14]	4	4	●	○	●	M	◐	○	○	○	○	4	4	4	1	0-III
DTaint [38]	6	6	●	○	●	○	●	○	●	○	○	4	6	◐	2	I
Tian et al. [39]	2,018	◐	○	●	●	S	⊕	○	●	◐	○	11	◐	1	⊕	I
VulSeeker [40]	4,643	○	○	◐	○	R	◐	○	○	●	○	◐	◐	◐	◐	◐
FirmUp [7]	◐5,000	◐2,000	○	●	○	S	●	○	○	○	○	◐	◐	◐	◐	◐
IoTFuzzer [41]	17	⊕	●	⊕	●	○	●	○	●	○	○	12	17	10	◐	◐
FIRM-AFL [42]	11	11	●	○	○	M;R	●	○	●	○	○	5	11	2	◐	I
FirmFuzz [43]	6,427	1,013	●	○	●	S	●	○	○	○	○	3	◐	1	2	I
SRFuzzer [44]	10	⊕	●	⊕	○	M	○	○	●	○	○	5	10	1	2	◐
Pretender [27]	6	⊕	●	⊕	◐	M	○	○	○	●	●	2	3	1	1	III
HALucinator [45]	16	16	●	○	●	M	◐	○	○	●	○	3	4	1	1	III
FirmScope [19]	2,017	◐	○	●	◐	S	●	○	○	◐	○	99+	◐	1	⊕	I
PDiff [46]	715	○	○	○	○	○	●	○	○	○	○	8	◐	3	2	I
P IM [47]	10	10	●	◐	●	M	○	○	○	●	○	3	4	10	1	II-III
Karonte [8]	53;899	◐	●	●	●	S;R	●	●	●	●	●	25	◐	◐	3	I-III
Laelaps [48]	30	⊕	●	◐	●	○	○	○	○	○	○	2	4	24	1	II-III
FirmAE [26]	1,306	1,124	●	●	●	S	●	●	○	●	●	8	◐	2	2	I
CPscan [49]	28	28	●	○	○	○	○	○	●	○	○	10	28	◐	◐	I
Diane [50]	11	⊕	●	⊕	○	○	●	○	●	○	○	9	11	4	◐	◐
DICE [51]	7	⊕	●	⊕	●	M	○	○	○	●	○	6	7	7	1	II-III
ECMO [52]	815	815	○	●	◐	○	○	○	○	○	○	2	37	1	1	I
iFIZZ [53]	10	10	●	●	●	○	◐	○	●	○	○	7	10	4	2	I
Jetset [54]	13	13	●	○	◐	M;R	○	○	○	○	○	4	13	3	3	I-III
SaTC [55]	39;49	39;49	●	●	○	○;R	○	○	●	●	●	6;4	6;◐	2;◐	2;3	◐
Snipuzz [56]	20	⊕	○	⊕	●	M	○	○	●	○	○	17	20	8	◐	◐
Emu [57]	21	21	●	○	○	M;R	●	○	●	●	○	◐	21	◐	1	II-III
SymLM [58]	8	8	●	○	◐	R	⊕	○	○	○	○	◐	8	◐	1	II-III
Marcelli et al. [59]	2	2	●	○	●	M	●	○	○	○	○	2	2	1	2	I
Greenhouse [25]	7,141	5,690	●	●	●	S;R	●	○	○	○	○	9	1,764	2	3	I
FirmSolo [20]	8,737	1,470	●	◐	●	○;R	●	○	◐	◐	○	◐	◐	◐	2	I
VulHawk [60]	20	20	○	○	○	○	●	○	○	○	○	3	20	◐	◐	◐

R. Helmke et al., *Mens Sana In Corpore Sano: S*

C3: An analysis of state of the art corpus creation practices in current research.

Cluster by measure

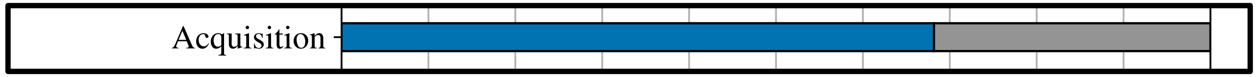
“How many papers documented this data?”



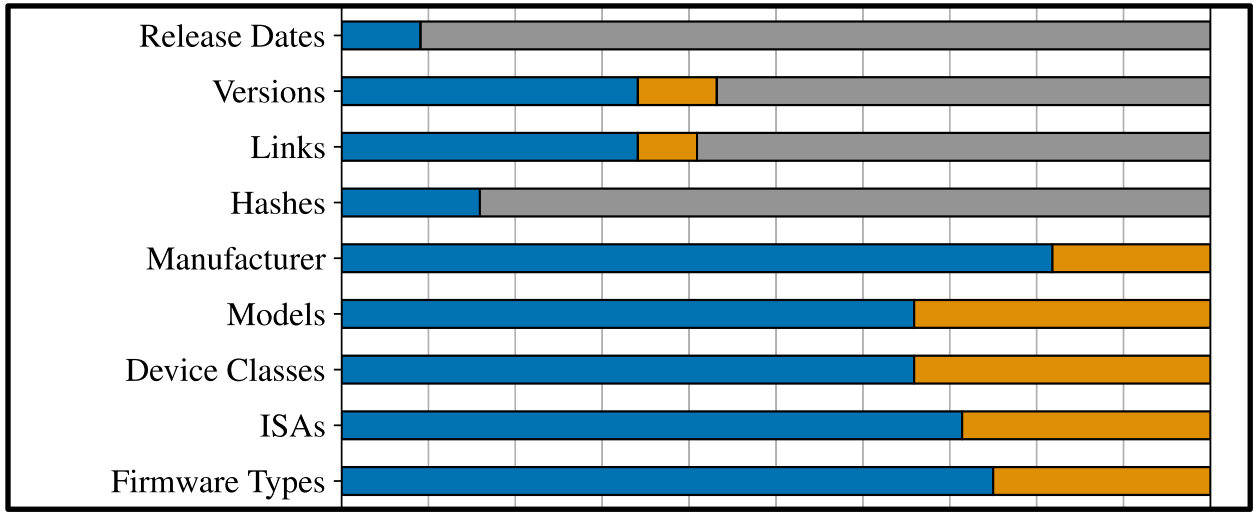
R. Helmke et al., *Mens Sana In Corpore Sano: Sound Firmware Corpora for Vulnerability Research*

Missing meta data & documentation threatens soundness.

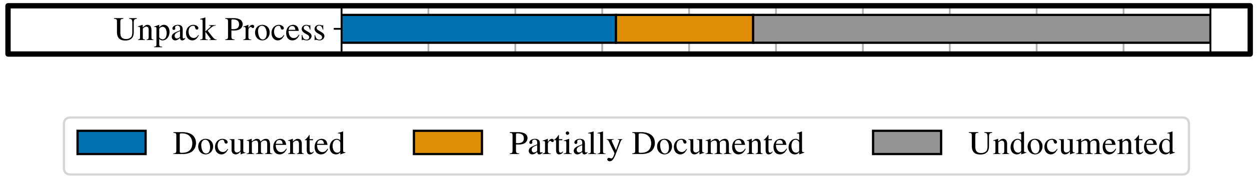
Acquisition steps are often documented.



There is few or incomplete meta data.



Most papers do not fully describe unpacking.



Corpus replicability

Gone.

Result verifiability

Hard.

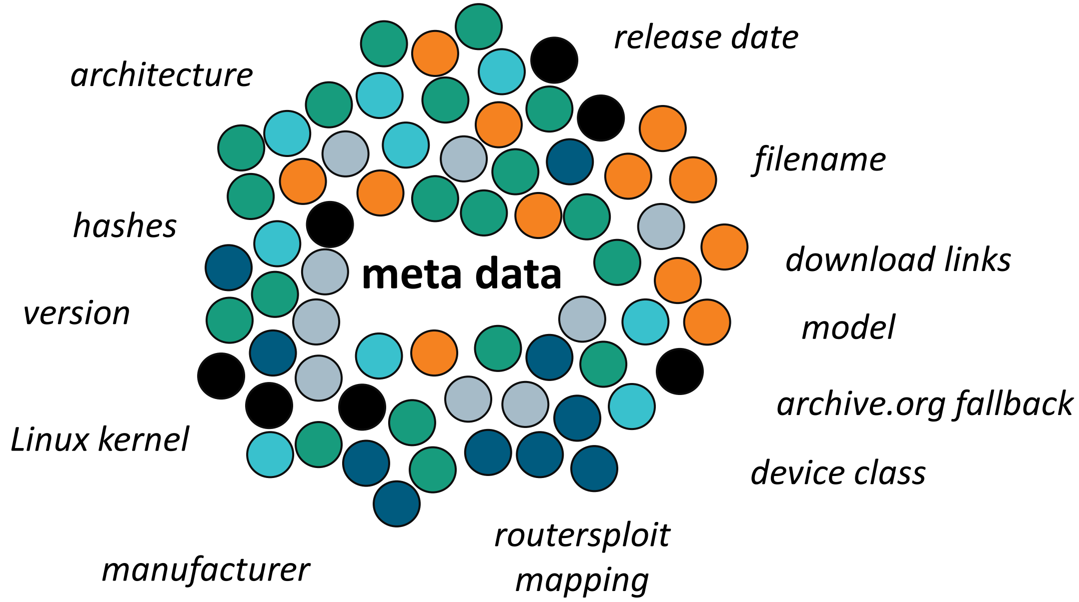
Representativeness

Hard to assess.

R. Helmke et al., *Mens Sana In Corpore Sano: Sound Firmware Corpora for Vulnerability Research*

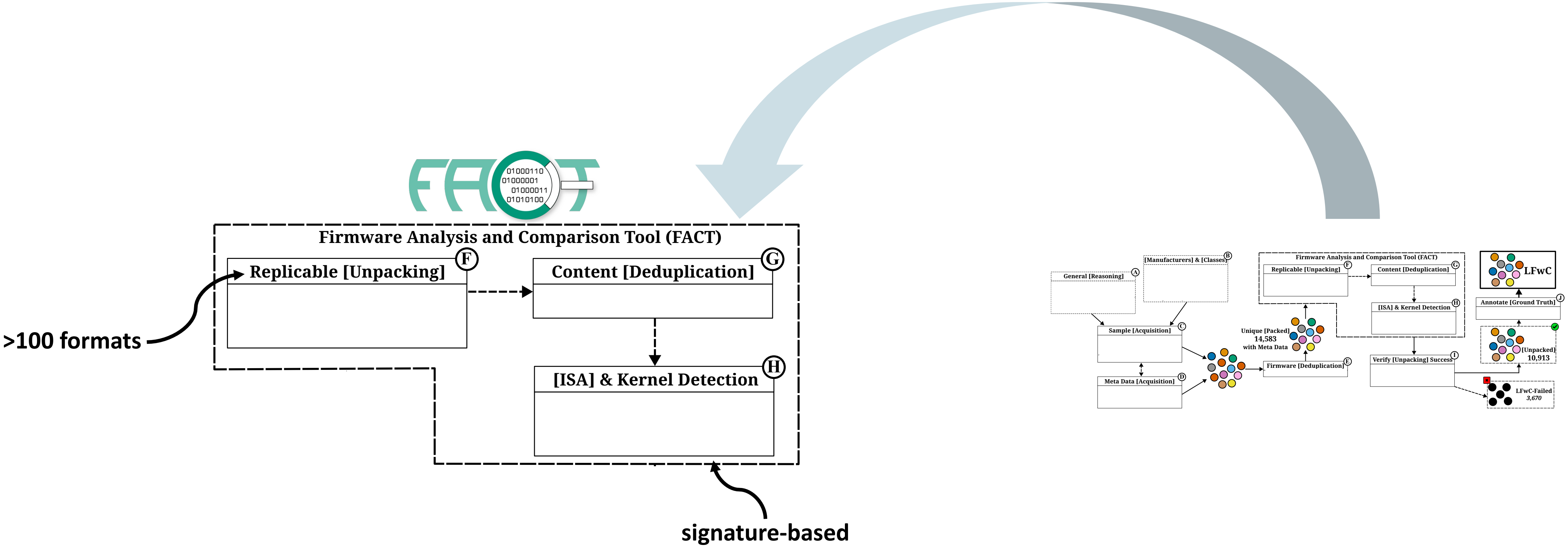
C4: A reference **L**inux **F**irm**w**are **C**orpus (LFwC).

~10,900	unpacked samples
~2,350	devices
22	device classes
10	manufacturers
2005-2023	version history



R. Helmke et al., *Mens Sana In Corpore Sano: Sound Firmware Corpora for Vulnerability Research*

C4: A reference **L**inux **F**irm**w**are **C**orpus (LFwC).



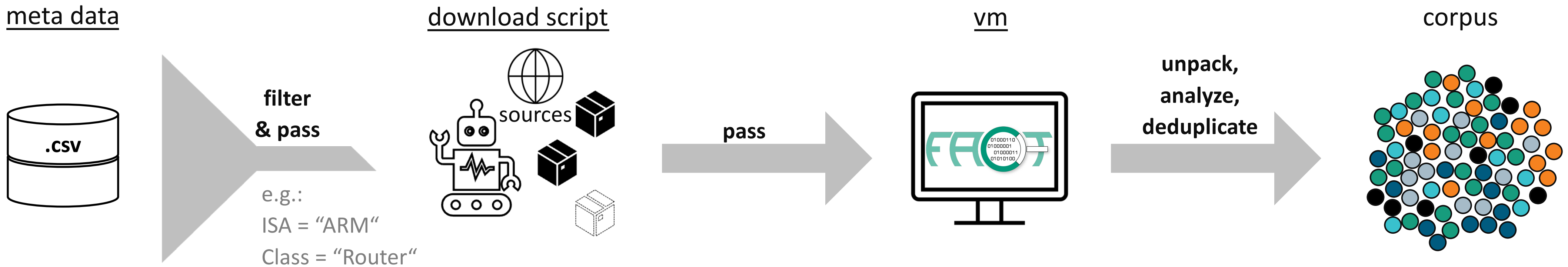
>100 formats

signature-based

R. Helmke et al., *Mens Sana In Corpore Sano: Sound Firmware Corpora for Vulnerability Research*

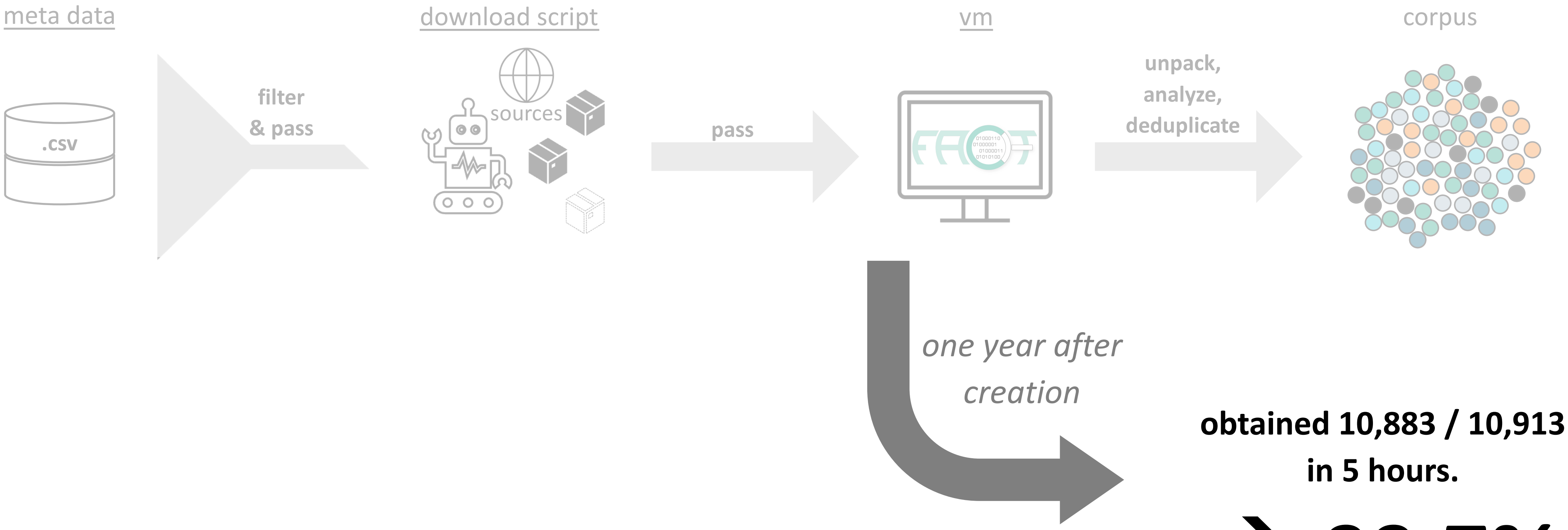
Replicate LFwC.

Artifact Evaluated NDSS
Available
Functional
Reproduced



Replicate LFwC.

Artifact Evaluated NDSS
Available
Functional
Reproduced



obtained 10,883 / 10,913
in 5 hours.

→ 99.7%

Summary.

1 Corpus Creation Challenges: *What are the problems?*

2 Creation Guidelines: *What are some properties of sound corpora?*

→ 16 Measures Towards Sound Corpora.

3 Research Paper Analysis: *How do we currently create corpora?*

→ More Documentation, More Meta Data.

4 Release LFwC Reference Corpus: *Are these guidelines feasible?*

→ Yes. Proven Replicability.

More information, analyses, and case studies in our paper.

Artifact Evaluated NDSS Available Functional Reproduced

Mens Sana In Corpore Sano: Sound Firmware Corpora for Vulnerability Research

René Helmke*, Elmar Padilla*, and Nils Aschenbruck°
*Fraunhofer FKIE, Cyber Analysis & Defense, Germany, {firstname.lastname}@fkie.fraunhofer.de
°Osnabrück University, Distributed Systems Group, Germany, aschenbruck@uos.de

Artifact Evaluated NDSS

Available

Functional

Reproduced

Artifacts, contributions, and contact.

request LFwC.

contribute.

raw data.

replication scripts.

FACT vm.

original scrapers.

setup tutorial.

Jupyter notebooks.
(explore data, gen. paper results)

https://github.com/fkie-cad/linux-firmware-corpus

Contact: rene.helmke@fkie.fraunhofer.de