Investigating Leaked Sensitive Information in Version Control Systems with the Kraulhorizon Framework

Felix Wilkening, Lars Stiemert, Matthias Schopp, Daniela Pöhn, Wolfgang Hommel

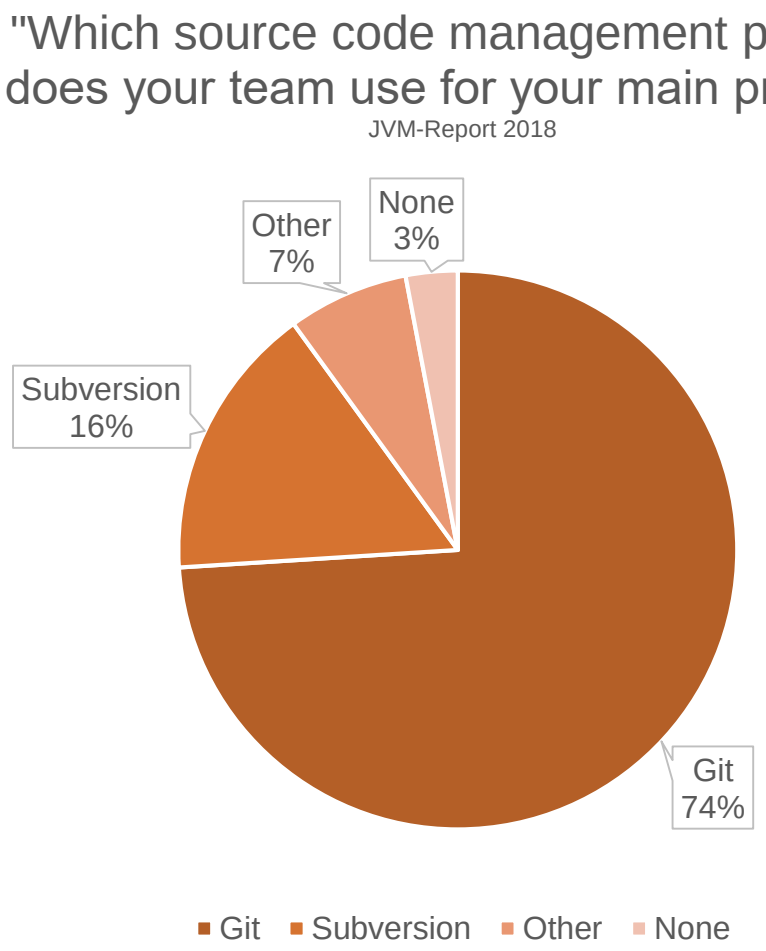
FI CODE
Forschungsinstitut
Cyber Defence
Universität der Bundeswehr München

https://www.unibw.de/code
felix.wilkening@unibw.de

Agenda

- Einführung und Motivation
- Anforderungen
- Konzeptionierung
- Evaluation
- Zusammenfassung und Ausblick

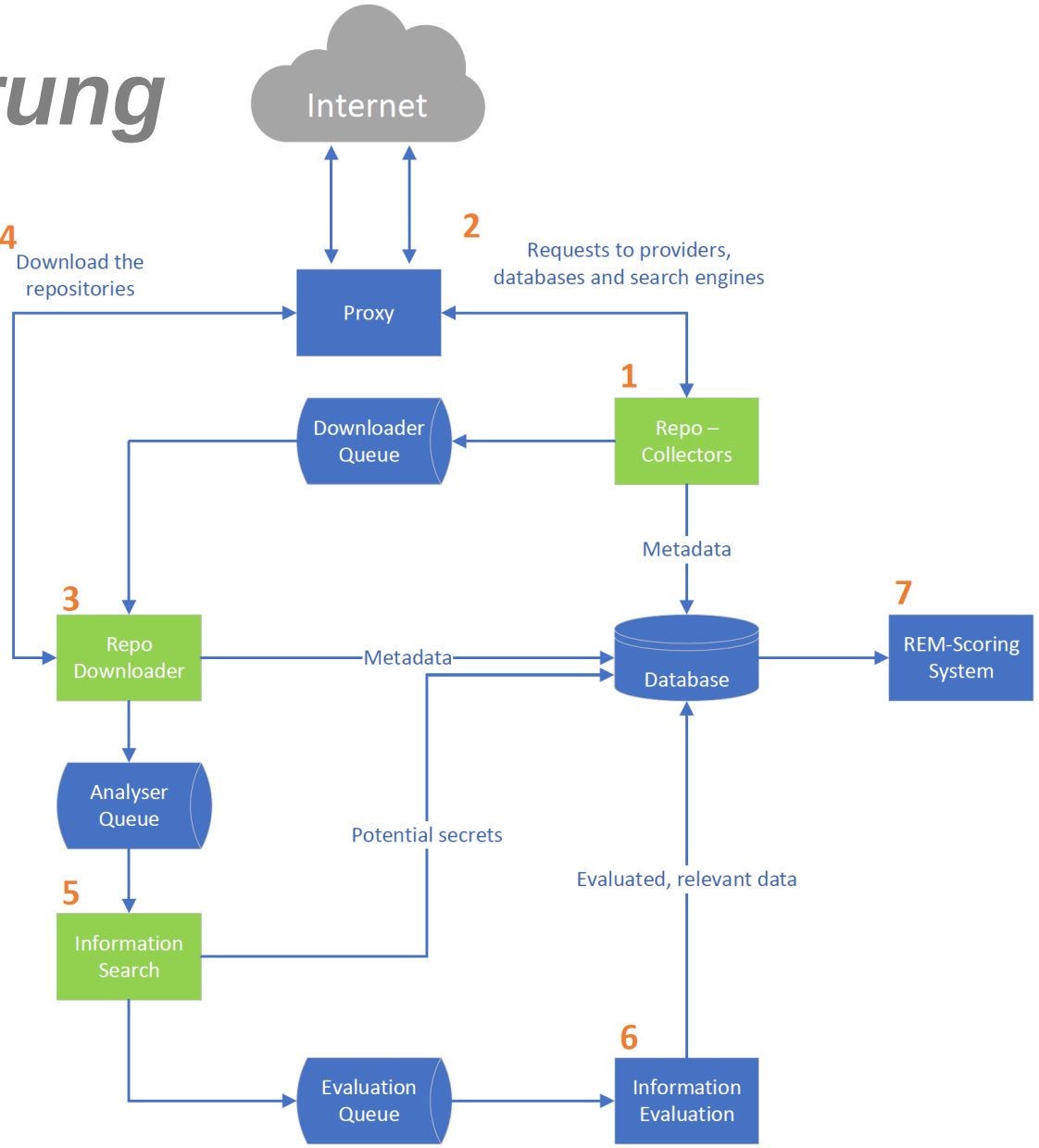
Einführung und Motivation



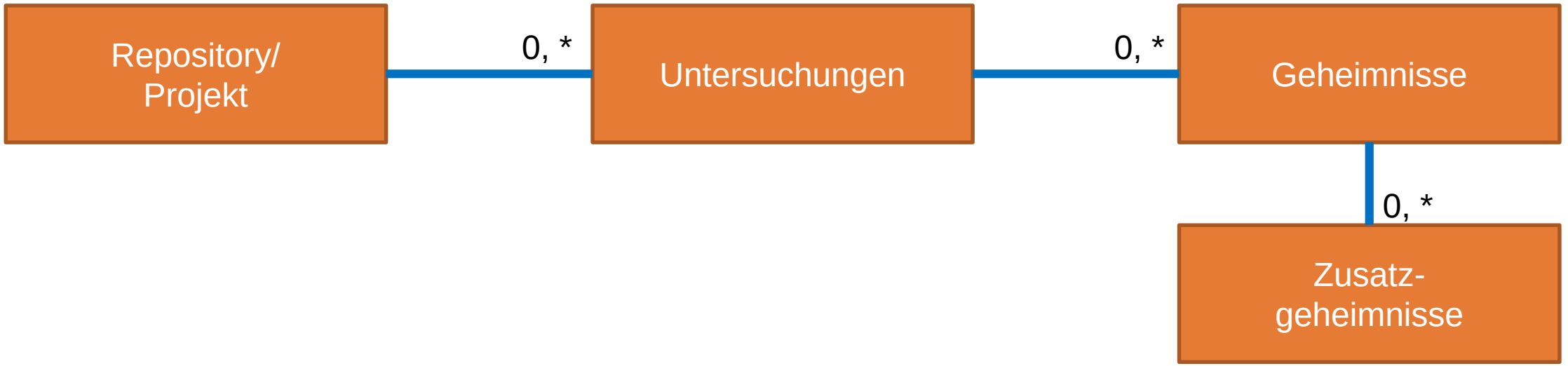
Anforderungen

- Automatische Suche nach Repositories
- Automatisches Herunterladen der Daten
- Analyse und Beurteilung der Daten
- Verschleierung der eigenen Identität
- Skalierbarkeit
- Modularität

Konzeptionierung



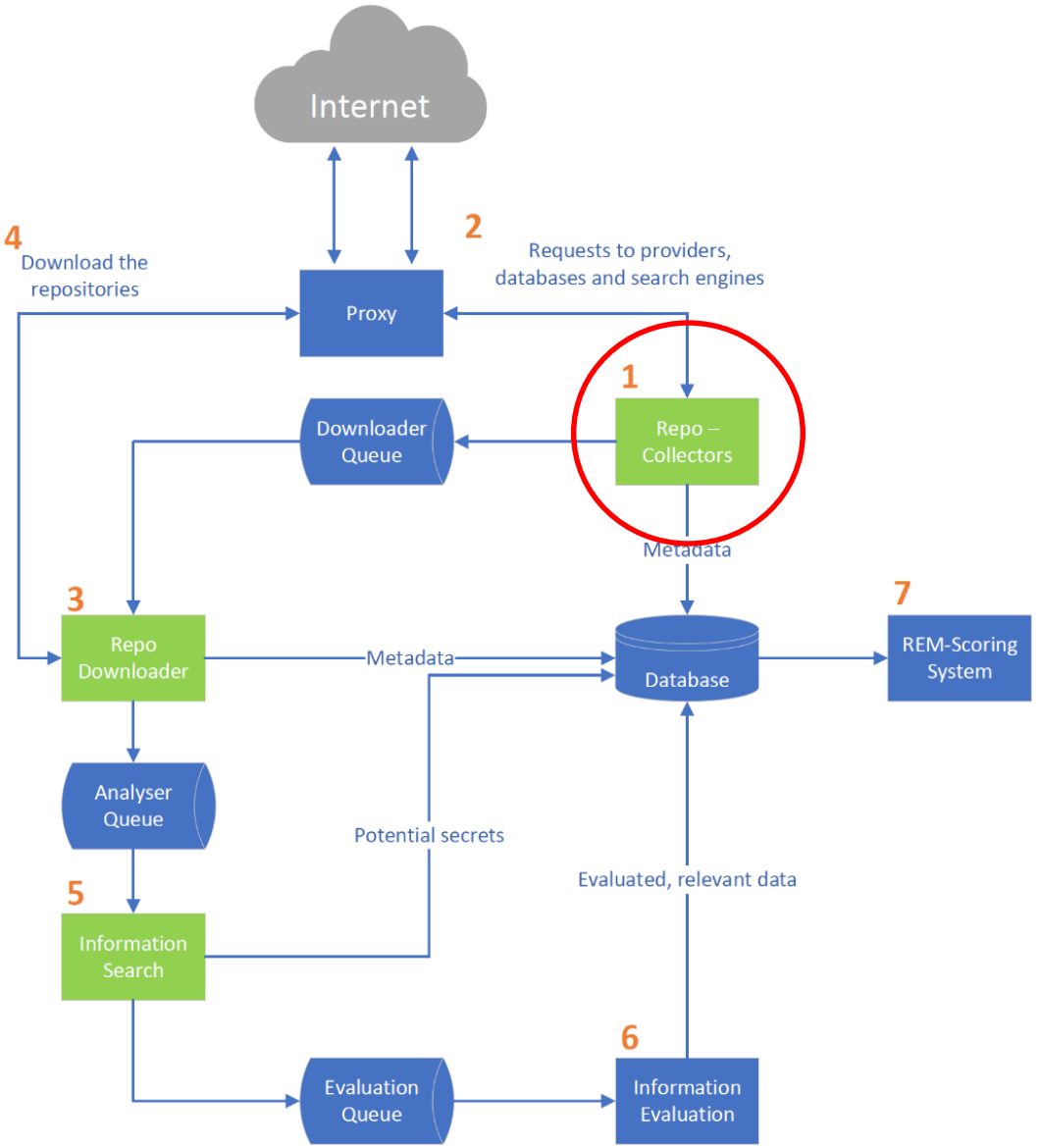
Konzeptionierung



Konzeptionierung

Repo-Kollektoren:

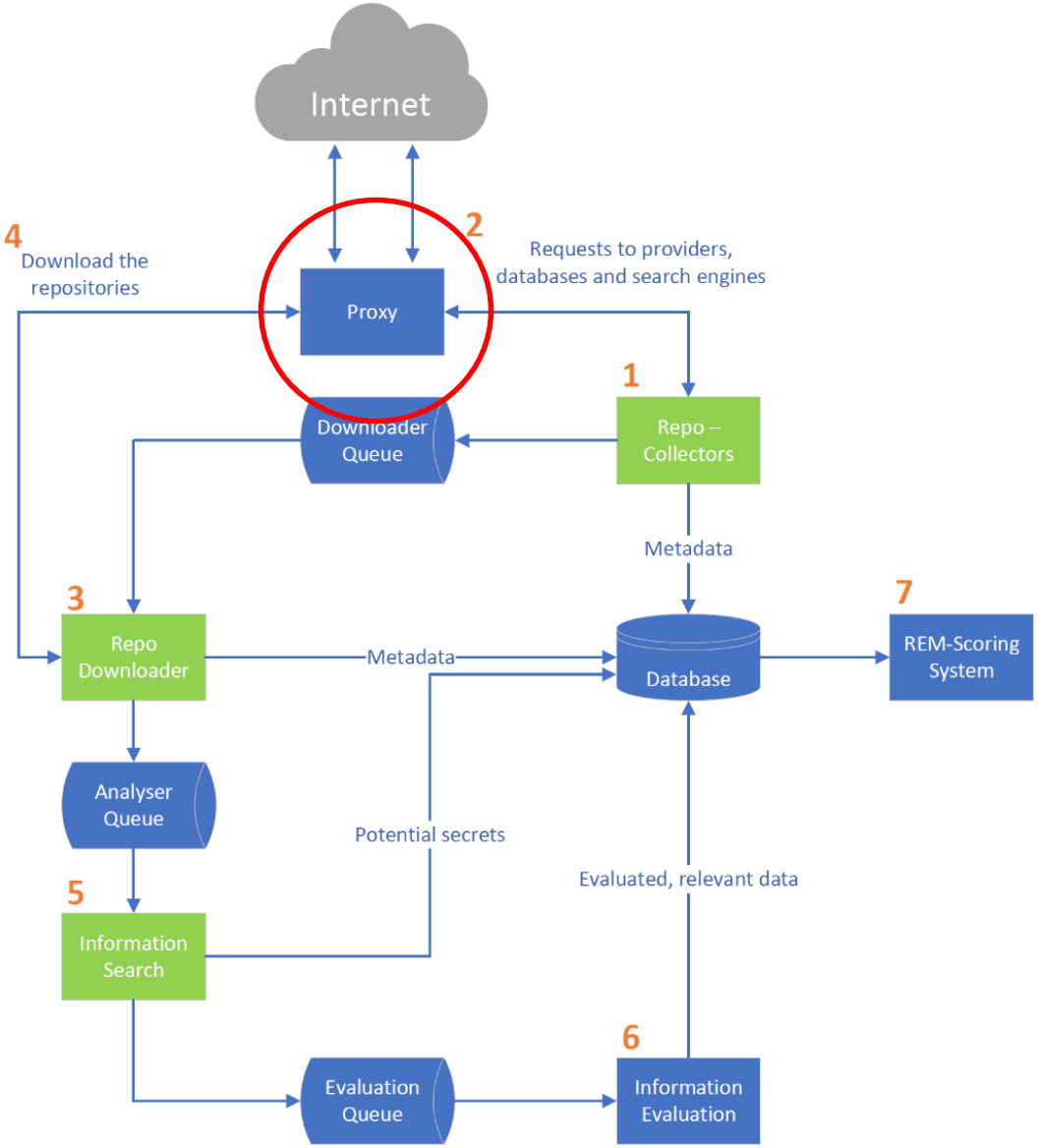
- *Sammeln Repositories für spätere Untersuchungen*
- *Verschiedene Implementierungen für unterschiedliche Ziele*



Konzeptionierung

Proxy:

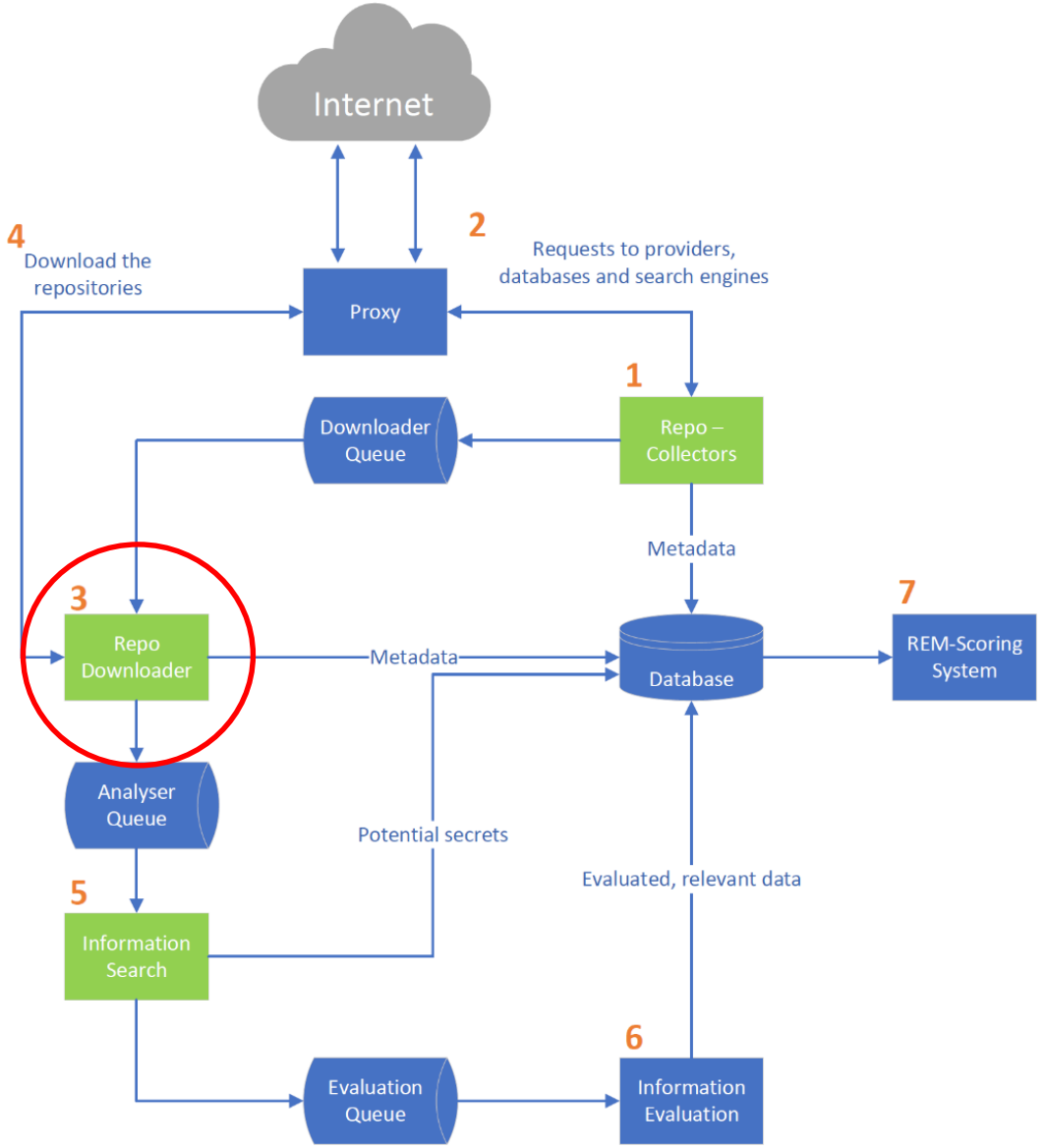
- *Verschleiert die eigene Identität*
- *Zugang zu internen Systemen*



Konzeptionierung

Repo-Downloader:

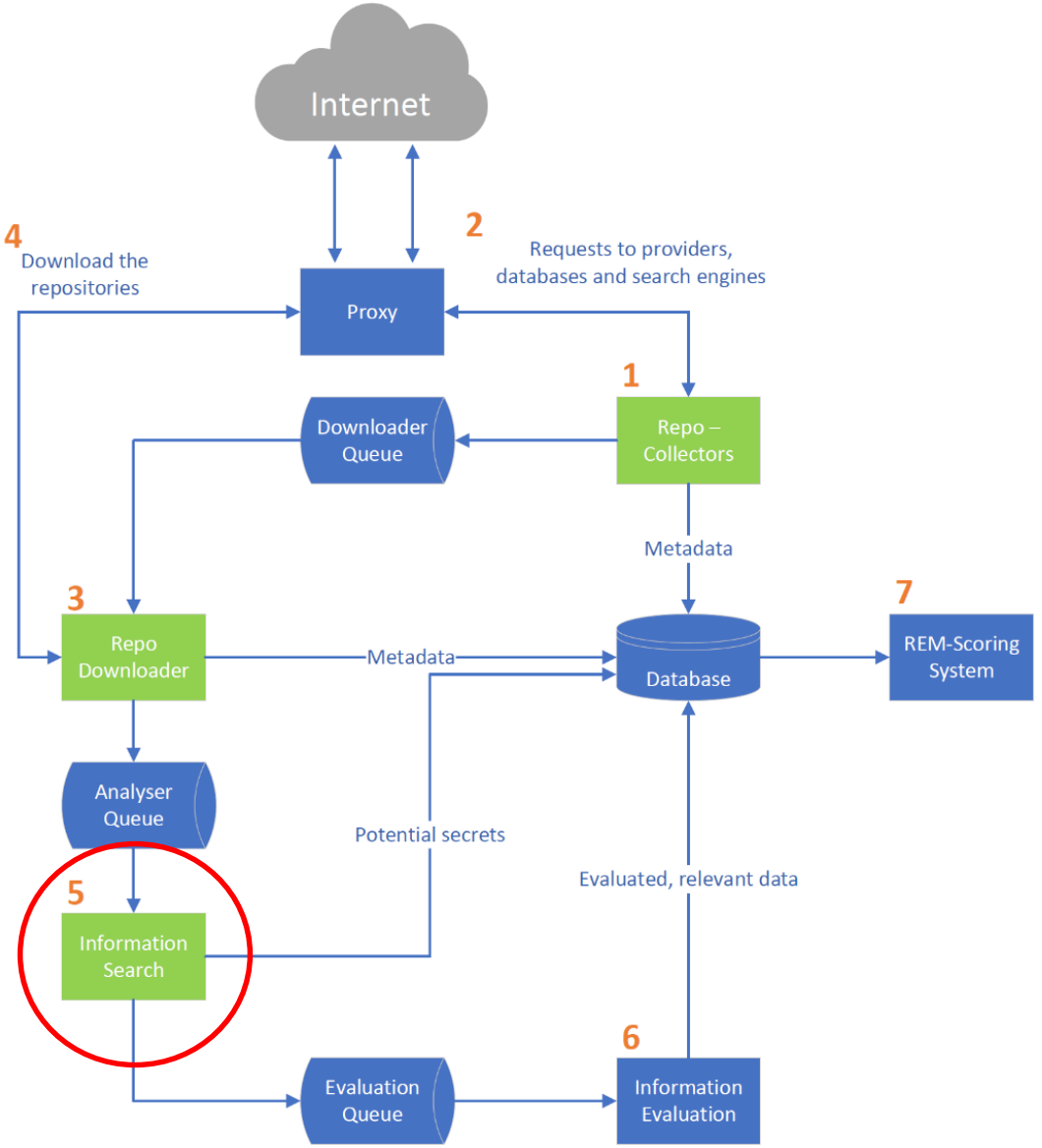
- *Lädt Repositories herunter und speichert die Metadaten ab*



Konzeptionierung

Information Search:

- *Analysiert Repositories mittels Regex auf potentielle Geheimnisse*

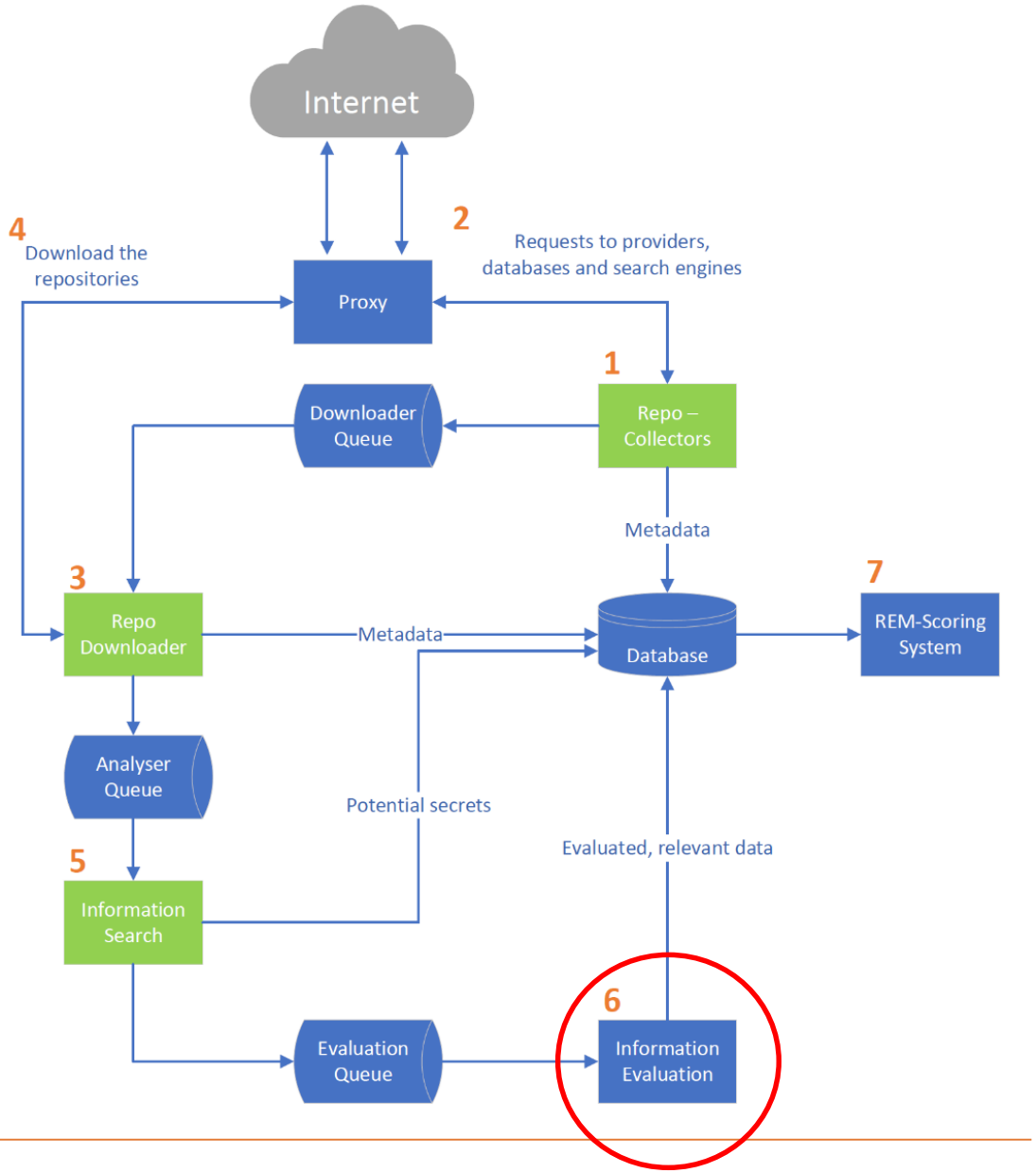


Konzeptionierung

Information Evaluation:

- *Analysiert die gefundenen Geheimisse*

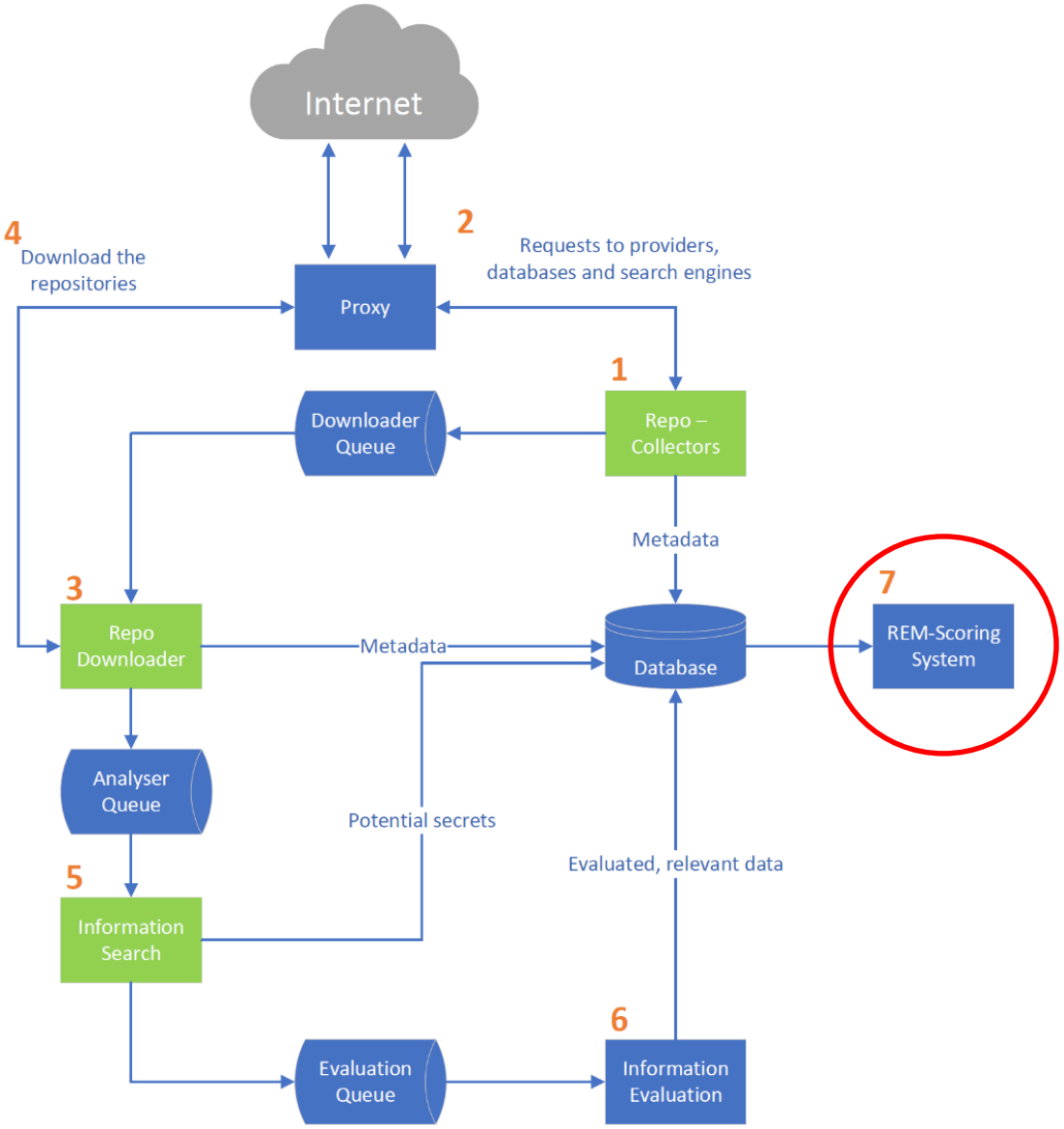
- *Word-Filter*
- *Entropy-Filter*
- *Pattern-Filter*



Konzeptionierung

REM-Scoring System:

- *Bewertung der Repositories mittels der Bewertungsvorschrift*



Konzeptionierung

Domain	Platform/API	Key Type	Single-factor or Multi-factor	Primary Risks			
				Monetary Loss	Privacy	Data Integrity	Message Abuse
Social Media	Twitter	Access Token	M		X	X	X
	Facebook	Access Token	S		X	X	X
	YouTube[a]	API Key	S	X	X		
		OAuth ID	M		X	X	X
	Picatic	API Key	S		X	X	X
Finance	Stripe	Standard API Key	S	X		X	
		Restricted API Key	S				
	Square	Access Token	S	X		X	
		OAuth Secret	S				
	PayPal Braintree	Access Token	S	X		X	
	Amazon MWS	Auth Token	M	X	X	X	X
Communications	Gmail	(same as YouTube)[a]	(same as YouTube)[a]		X	X	X
	Twilio	API Key	S		X	X	X
	MailGun	API Key	S		X	X	X
	MailChimp	API Key	S		X	X	X
Storage	Google Drive	(same as YouTube)[a]	(same as YouTube)[a]		X	X	
IaaS	Amazon AWS	Access Key ID	S	X	X	X	
	Google Cloud Platform	(same as YouTube)[a]	(same as YouTube)[a]	X	X	X	
Private Keys	RSA	Cryptographic key	M	X	X	X	X
	EC	Cryptographic key	M	X	X	X	X
	PGP	Cryptographic key	M	X	X	X	X
	General	Cryptographic key	M	X	X	X	X

Aus Meli et. al – „How Bad Can It Git?“

Konzeptionierung

$$ISS = 1 - ((1 - Geldverlust) \cdot (1 - Privacy) \cdot (1 - Integritaet) \cdot (1 - Falschmeldungen)) \tag{5.1}$$

$$Auswirkungen = 5 \cdot ISS \tag{5.2}$$

$$Grundwert = 4 \cdot Typ \cdot Multifaktor \tag{5.3}$$

$$Kennwert = 0(wennAuswirkungen = 0) \tag{5.4}$$

$$Kennwert = Round(Minimum(Veraltet \cdot (Grundwert + Auswirkungen), 10)) \tag{5.5}$$

$$Gesamtkennwert = min(\sum_{s=0}^{n} Kennwert_s, 20)$$

Konzeptionierung

Domain	Platform/API	Key Type	Single-factor or Multi-factor	Primary Risks			
				Monetary Loss	Privacy	Data Integrity	Message Abuse
Social Media	Twitter	Access Token	M		X	X	X
	Facebook	Access Token	S		X	X	X
	YouTube[a]	API Key	S	X	X		
		OAuth ID	M		X	X	X
	Picatic	API Key	S		X	X	X
Finance	Stripe	Standard API Key	S	X		X	
		Restricted API Key	S				
	Square	Access Token	S	X		X	
		OAuth Secret	S				
	PayPal Braintree	Access Token	S	X		X	
	Amazon MWS	Auth Token	M	X	X	X	X
Communications	Gmail	(same as YouTube)[a]	(same as YouTube)[a]		X	X	X
	Twilio	API Key	S		X	X	X
	MailGun	API Key	S		X	X	X
	MailChimp	API Key	S		X	X	X
Storage	Google Drive	(same as YouTube)[a]	(same as YouTube)[a]		X	X	
IaaS	Amazon AWS	Access Key ID	S	X	X	X	
	Google Cloud Platform	(same as YouTube)[a]	(same as YouTube)[a]	X	X	X	
Private Keys	RSA	Cryptographic key	M	X	X	X	X
	EC	Cryptographic key	M	X	X	X	X
	PGP	Cryptographic key	M	X	X	X	X
	General	Cryptographic key	M	X	X	X	X

Aus Meli et. al – „How Bad Can It Git?“

Konzeptionierung

Eigenschaft	Metrik-Wert	Symbol	Wert	Definition
Geldverlust (Gv) Privacy (Pr)	Keiner	K	0	Wenn die entsprechende Eigenschaft ungefährdet ist.
Integrität (I) Falschmeldungen (Fm)	Hoch	H	0.4	Wenn durch Missbrauch der Datei die entsprechende Eigenschaft gefährdet ist.
Multifaktor (Mf)	Unsicher	U	1	Es handelt sich um ein allein stehendes Geheimnis, oder ein Multifaktorgeheimnis, bei welchem alle Geheimnisse bekannt sind.
	Sicher	S	0.2	Bei einem Multifaktorgeheimnis sind nicht alle Geheimnisse bekannt und die Daten können nicht ausgenutzt werden.
Typ (Ty)	Zugangsdaten	Z	1	Standartwert für alle Datentypen, die keine kryptografischen Schlüssel sind.
	Kryptodaten	K	1.5	Wenn es sich um kryptografische Daten, wie z. B. Schlüssel handelt.
Veraltet (Ve)	Gültig	G	1	Standartwert für alle gefundenen sensiblen Daten.
	Veraltet	V	0	Der Wert ist nachweislich veraltet oder ungültig. Dies tritt ein, wenn z. B. ein Zugangstoken oder Schlüssel abgelaufen ist.

Konzeptionierung

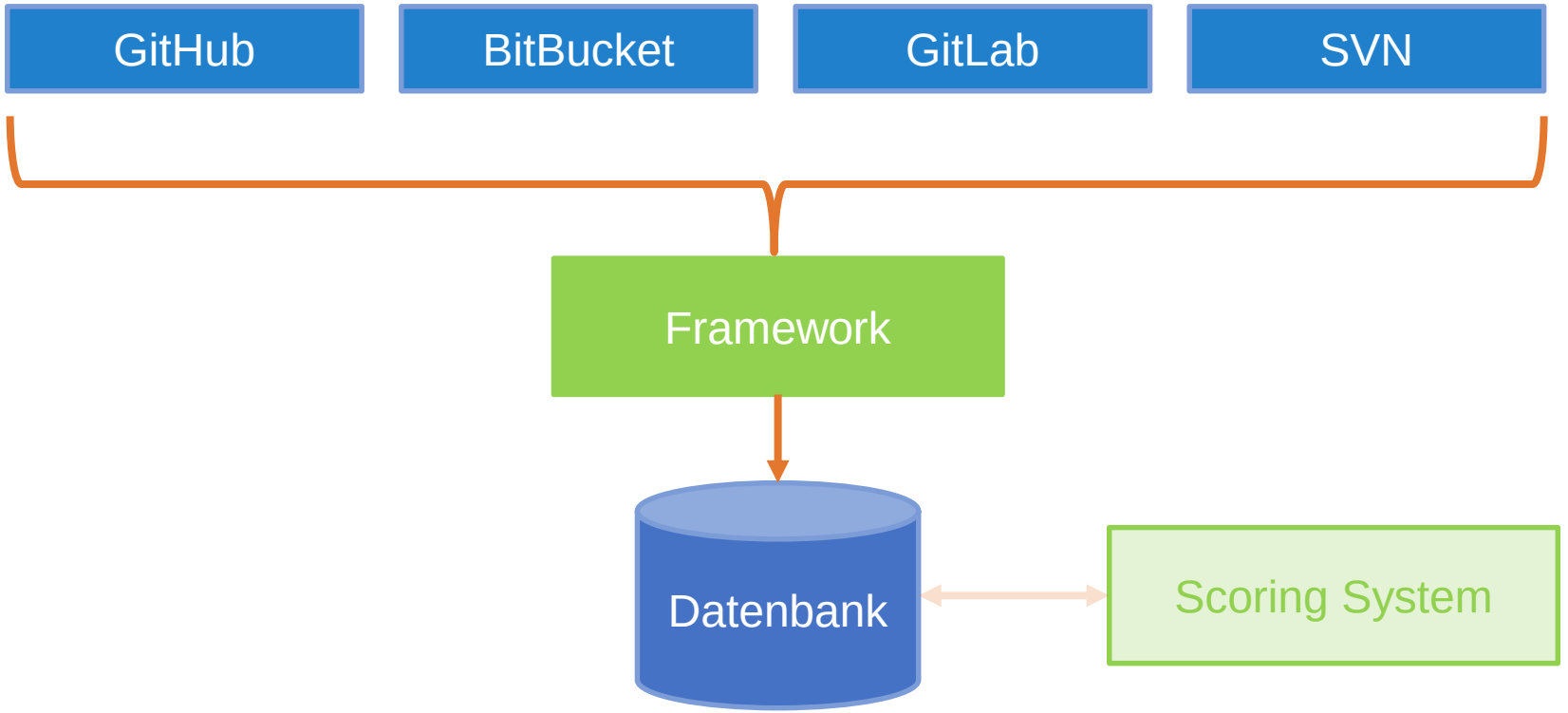
Beispielrechnung für einen Amazon MWS-Token (inklusive Zusatzgeheimnis)

$$
\begin{aligned}
0.8704 &= 1 - ((1 - 0.4) \cdot (1 - 0.4) \cdot (1 - 0.4) \cdot (1 - 0.4)) \\
4.352 &= 5 \cdot 0.8704 \\
4 &= 4 \cdot 1 \cdot 1 \\
8.4 &= Round(Minimum(1 \cdot (4 + 4.352), 10))
\end{aligned}
$$

$$
\text{REMv1.0} = \{\text{GV:H/Pr:H/I:H/Fm:H/Mf:U/Ty:Z/Ve:G}\}
$$

Evaluation

Erster Test: Präparierte Daten von verschiedenen SVN und Git-Repositories herunterladen und analysieren



Evaluation

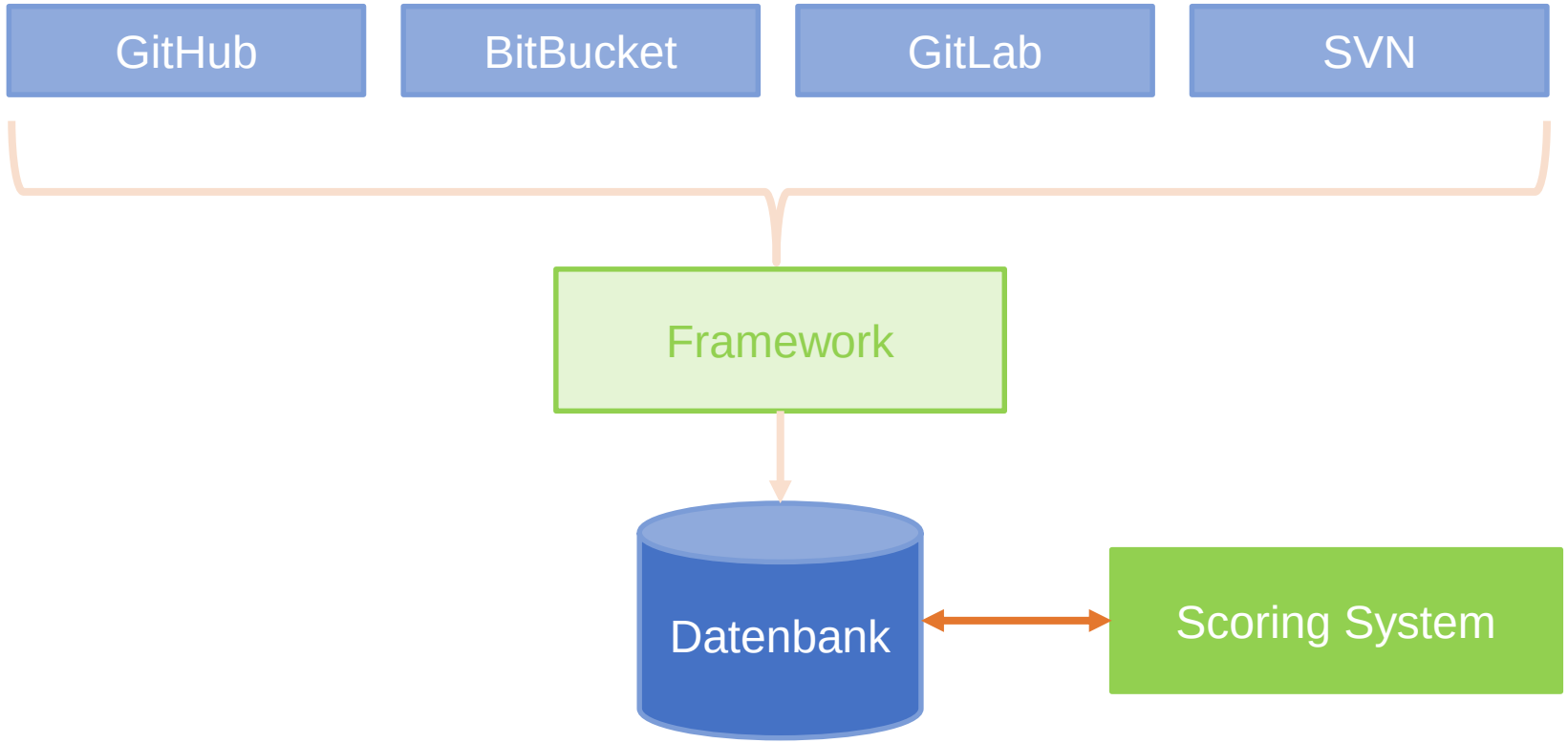
Erster Test: Präparierte Daten von verschiedenen SVN und Git-Repositories herunterladen und analysieren

Resultate:

- *Verschiedene Provider können angefragt werden*
- *Daten werden lokal gespeichert*
- *Bei Git-Repositories werden alle gestreuten Geheimnisse erkannt,*
- *Bei SVN lediglich die letzte Revision*
- *Doppelte Erkennungen durch Regex sind möglich*

Evaluation

Zweiter Test: Auswerten der Repositories mittels des Scoring Systems



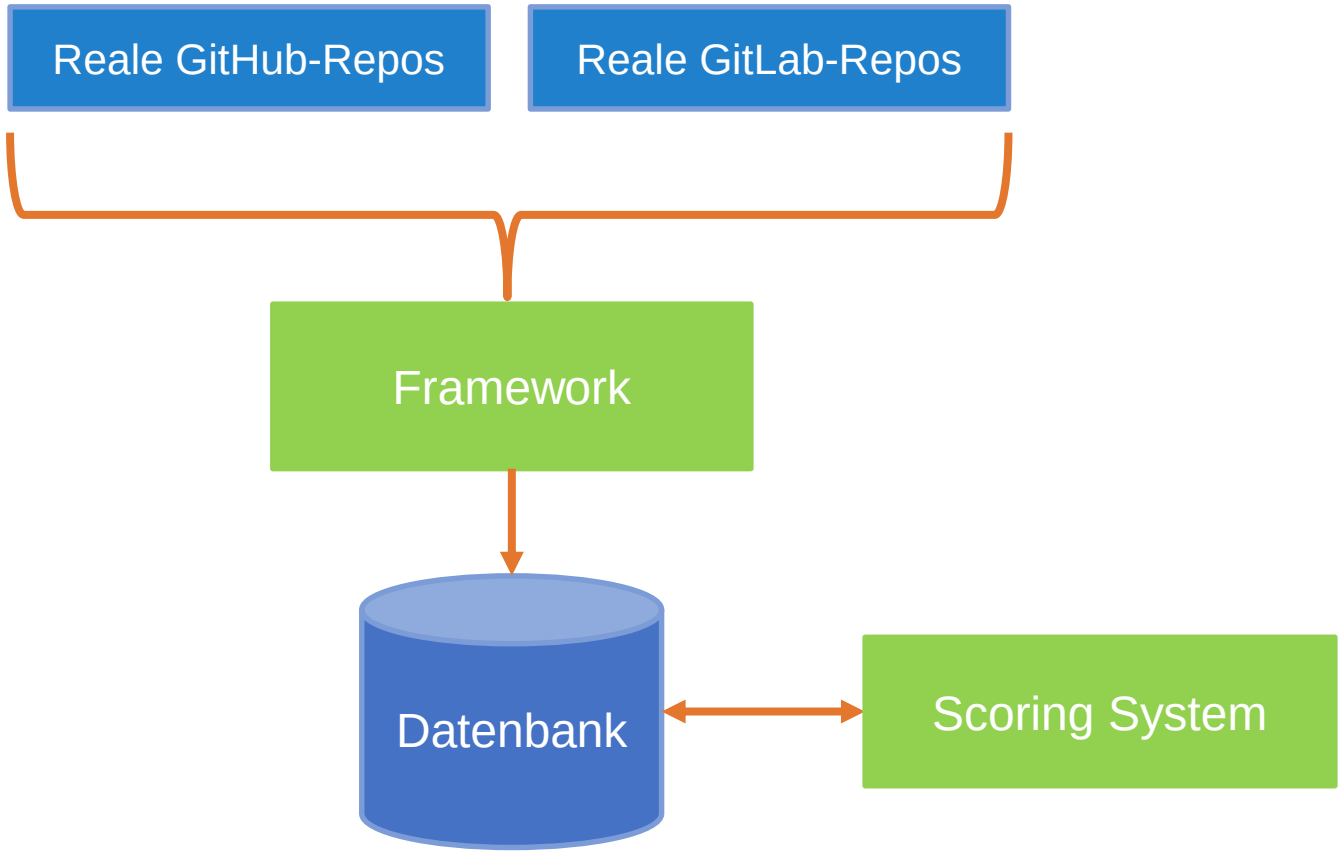
Evaluation

Zweiter Test: Auswerten der Repositories mittels des Scoring Systems


```
{
        "_id" : ObjectId("60edf7419d363d75c1dbea76"),
        "repo_storage_id" : ObjectId("60edf73b13dd88222b33fa23"),
        "type" : "Git",
        "date_time" : "13-07-2021 22:27:44",
        "size" : 3683,
        "duration_download" : 0.8720064440003625,
        "filepath" : "/home/worker/Repositories/gitlab.com/justm4ster1t/uselessstuff4testing/13-07-2021 22:27:44",
        "found_secrets" : [
                ObjectId("60edf7413e152c513dc32334"),
                ObjectId("60edf7413e152c513dc32335"),
                ObjectId("60edf7413e152c513dc32336")
        ],
        "duration_file_analyser" : 0.08544835199973022,
        "score" : 13.1,
        "score_vector" : "REMv1.0{Gv:H/Pr:H/I:H/Fm:H/MF:H/Ty:Z/Ve:G}{Gv:H/Pr:H/I:H/Fm:H/MF:K/Ty:Z/Ve:G}"
}
{
        "_id" : ObjectId("60edf74332fbc704c10f765e"),
        "repo_storage_id" : ObjectId("60edf73b13dd88222b33fa23"),
        "type" : "Git",
        "date_time" : "13-07-2021 22:27:46",
        "size" : 4781,
        "duration_download" : 1.4344857480000428,
        "filepath" : "/home/worker/Repositories/Bitbucket/W1Fe1483984@bitbucket.org/W1Fe1483984/uselssstuff4demo/13-07
-2021 22:27:46",
        "found_secrets" : [
                ObjectId("60edf7445aa7b619b1bbfef7"),
                ObjectId("60edf7445aa7b619b1bbfef8")
        ],
        "duration_file_analyser" : 0.0883312020005178,
        "score" : 9.4,
        "score_vector" : "REMv1.0{Gv:H/Pr:H/I:H/Fm:H/MF:K/Ty:Z/Ve:G}{Gv:H/Pr:H/I:H/Fm:H/MF:K/Ty:Z/Ve:G}"
}
```

Evaluation

Dritter und vierter Test: Auswerten von GitHub und GitLab Repositories



Evaluation

Dritter Test: Auswerten von GitLab Repositories

Resultate:

- *12.500 GitLab Repositories erfasst*
- *9.039 Repositories heruntergeladen*
- *4.717 Repositories analysiert*
- *Keine Funde von sensiblen Informationen*

Evaluation

Vierter Test: Auswerten von GitHub Repositories. 1.000 Repositories jeweils via Git und SVN.

Resultate:

- *991 heruntergeladene Repositories*
- *734 analysiert*
- *Git: 22 gefundene private RSA-Schlüssel in 15 Repositories*

Zusammenfassung und Ausblick

- *Framework für die Untersuchung verschiedener Versionsverwaltungssysteme*
- *Bewertungssystem für geleakte Informationen*
- *Nachweis der Funktionsfähigkeit des Prototypen an realen Repositories*

- *Weitere Verbesserungen*
- *Umfassendere Studie mit Realdaten*

https://www.unibw.de/code
felix.wilkening@unibw.de

Fragen?

https://www.unibw.de/code
felix.wilkening@unibw.de

Bildquellen

Bitcoin (Folie 3 – Einführung und Motivation)

https://pixabay.com/de/vectors/bitcoin-digitale-w%c3%a4hrung-4130299 (Zugriff am 30.01.2022 16:15)