CIRCL
Computer Incident
Response Center
Luxembourg

GCVE.eu

Digital Sovereignty Starts with Vulnerability Data

Can we build vulnerability intelligence that we actually own?

https://www.vulnerability-lookup.org – https://gcve.eu

info@gcve.eu - @gcve@social.circl.lu

GCVE Workshop Vulnopticon 2026

CIRCL https://www.circl.lu

TLP:CLEAR

Contents

The Uncomfortable Question

Where the previous talk left us

GCVE, just presented

- **Sovereignty**, **autonomy**, **independence**, **fast innovation**.
- Applied to the **name** of a vulnerability and its **publication path**.
- Who controls timeline, revisions and endpoint? The GNA, with attribution.

This talk: the announced deep dive, consumer side

- Allocation and namespaces: **taken as given**.
- APIs, dumps, directory, interoperability: **open data, federated consumption**.
- KEV conversion: **exploitation assertions**.
- cpe.gcve.eu, patch2vuln: **open catalogues and models**.
- One layer not covered yet: **sensors**.

An identifier you own is not yet **intelligence** you own.

Vulnerability data is critical infrastructure

- Security teams use it every day to **identify affected software**, **prioritise** remediation, **understand exploitation** and **decide about risk**.
- Regulation now depends on it: NIS2, the Cyber Resilience Act, the EU vulnerability database (EUVD).
- Every scanner, SBOM tool, SOC playbook and patch policy is built on top of it.

Who owns vulnerability data?
Who controls the identifiers, the databases, the enrichment, the scoring, and increasingly the **models** we rely on to understand vulnerabilities? And who **pays** for the CVE Program, on which all of this depends, to operate?

April 2025: the day the CVE Program almost stopped

Date	Event
15 April 2025	MITRE warns that its U.S. government contract to run the CVE Program expires **the next day**.
16 April 2025	CISA grants an **11-month** extension, hours before expiration. The CVE Foundation is announced.
16 April 2025	09:46 CEST: first commit of `gcve.eu`. **GCVE is born**, the same morning.
13 May 2025	ENISA launches the **European Vulnerability Database (EUVD)**.
21 Jan. 2026	The CVE Board is told there will be "no funding cliff in March": CVE moves from leftover discretionary money to a **core** budget line.
16 March 2026	The 11-month extension ends. CISA states the program is **fully funded**.

- **The whole planet's** identifiers depended on one funding decision, in one country.
- Not a people problem, an **architecture** problem: one namespace, one operator, one budget line.

The layers we depend on

Identifiers – CVE, GHSA, OSV, CNVD, GCVE, . . .	*Who allocates? Who can dispute?*
Databases – NVD, EUVD, vendor CSAF feeds, national CERT feeds, . . .	*Who hosts? Who mirrors?*
Enrichment & scoring – CVSS, EPSS, CWE, CPE, KEV catalogues, . . .	*Who decides it is "exploited"?*
Observations – exploitation sightings, PoCs, scanner templates, social signals	*Who is the sensor?*
Models – classification, extraction, forecasting, "AI" summaries	*Can I run it? Can I retrain it?*

Sovereignty has to be asked at **every layer**. Losing one layer is enough to lose the stack.

What sovereignty means here (and what it does not)

It means being able to

- **access** the data without asking permission,
- **understand** how it was produced,
- **process** it on infrastructure we control,
- **reproduce** the results,
- **extend** and **disagree** openly,
- keep working when **someone else stops**.

It does not mean

- building a European silo,
- one organisation controlling everything,
- stopping to share with the rest of the world.

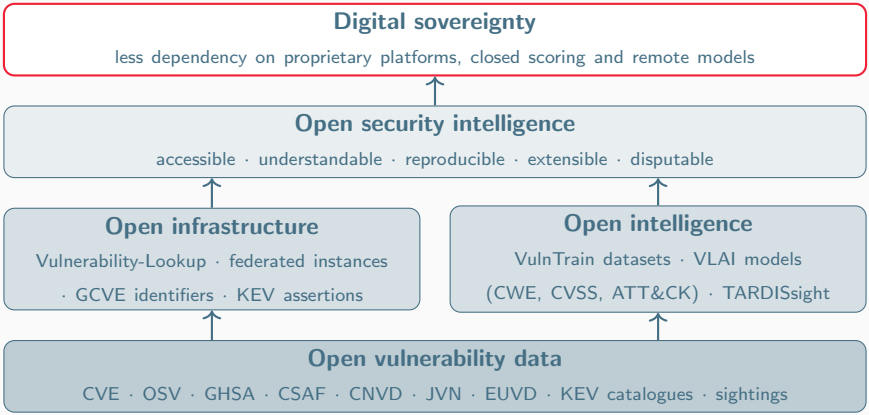
Sovereignty is not isolation. It is **optionality**: no single actor has to control everything.

The central argument

Digital sovereignty is not only about **where** your servers or data are hosted.
It is also about whether you can **independently understand, process, enrich and act upon** the security information your infrastructure depends on.

- Vulnerability information is the ideal case study: **almost every organisation depends on it**, every day.
- Most of it comes from a **single** identifier namespace, a **single** exploitation list, and a growing number of **closed** enrichment services.
- If we cannot own this layer, sovereignty of the layers above it is an illusion.

Vulnerability data as the foundation



Open Data: Vulnerability-Lookup

Who is behind Vulnerability-Lookup?

Vulnerability-Lookup[1] is an **Open Source** project led by **CIRCL**, the Computer Incident Response Center Luxembourg.
It is co-funded by **CIRCL** and the **European Union**[2].
Used by many organisations including CSIRTs and ENISA (EUVD).
Reference implementation of the **GCVE** standards. `db.gcve.eu` is a Vulnerability-Lookup instance.

[1] https://www.vulnerability-lookup.org
[2] https://github.com/ngsoti

Why a public CSIRT builds its own vulnerability database

- Started with `cve-search`[3] in 2012: keep a **local** copy of CVE, because the network, the API or the funding can disappear.
- Vulnerability sources have **diversified**. The NVD is no longer the sole source, and it is no longer sufficient.
- A CSIRT cannot outsource its understanding of a vulnerability to a vendor dashboard.
- We wanted the data **on disk**, **in the original format**, with **provenance**, queryable **without a licence**.

The first sovereignty question is trivial: **do you have a copy?**

[3] `https://github.com/cve-search/cve-search`

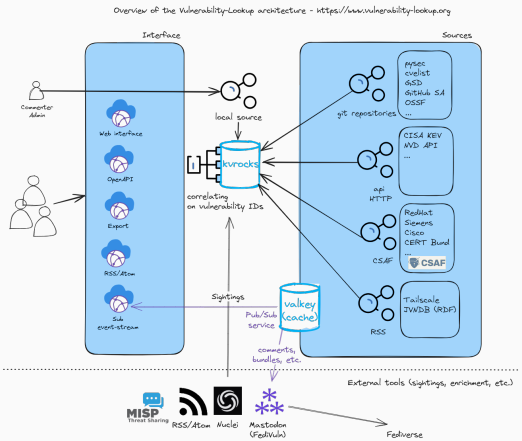
Where the data comes from: no single provider

Source family	Examples
Core vulnerability identifiers	CVE List v5, NVD, Fraunhofer FKIE NVD, CISA Vulnrichment, GCVE GNAs
Ecosystem advisories	GitHub Advisory Database, GSD, PySec, Bitnami, Drupal, Tailscale, OSV (AlmaLinux, Haskell, OCaml, OSS-Fuzz, RustSec)
CSAF 2.0 producers (40+)	CERT-Bund, CISA, Cisco, Microsoft, Red Hat, SUSE/openSUSE, Siemens, ABB, Schneider Electric, Phoenix Contact, WAGO, SICK, Nozomi, ...
National / regional feeds	CERT-FR (FR), CNVD (CN), JVNDB (JP), VARIoT (PL), FSTEC (RU), Moksha (DK), EUVD (EU)
Exploitation and risk signals	EPSS, CISA KEV, ENISA / EU CSIRTs Network KEV, Shadowserver, Previdian, GCVE GNA KEV feeds
Knowledge and supply chain	CWE, CAPEC, EMB3D, CleanStart, OpenSSF malicious packages
Observations (sightings)	MISP, Nuclei, Metasploit, Exploit-DB, Sploitus, GitHub Gists, Mastodon, Bluesky, Telegram

- **75+ feeds**, more than **2.3 million** advisories, from the U.S., Europe, China, Japan, Russia, and open-source ecosystems.
- Different jurisdictions, different views, different **biases**. Aggregation makes the differences **visible**.

Design principles that matter for sovereignty

- **Never alter the original record.** Harmonisation is a layer on top, the source stays verbatim.
- **Provenance everywhere.** Each record carries where it came from and when.
- **Open dumps**[a] and an **unauthenticated public API**.
- **Self-hostable.** One instance for a CSIRT, a vendor, a university, a country.
- **Correlation across identifiers**: CVE, GHSA, OSV, GCVE, ... are views on the same vulnerability.



[a] https://vulnerability.circl.lu/dumps/

Open Identifiers: Building on GCVE

What we take as given from GCVE

- A vulnerability can get a **name** from any of 46 autonomous GNAs, without asking a central authority.
- **GNA 0** keeps every CVE usable: `CVE-2025-53770` is `GCVE-0-2025-53770`. Nothing is thrown away.
- Records are published from the GNA's **own endpoint**, discovered through a **signed directory** (BCP-01, BCP-03).
- Disagreement is handled by **attribution and cross-references**, not by authority.
- Best Current Practices are **running code**, reviewed in public.

With the naming problem addressed, the uncomfortable question moves one layer up: **what do you actually know about a vulnerability once it has a name?**

A name is not intelligence

The identifier does not tell you	Who answers today	Sovereign answer
Is it **exploited**?	CISA KEV, vendor blogs	Federated KEV assertions (GCVE-BCP-07)
How **severe** is it, right now?	NVD CVSS, often days late	Open severity models (VLAI), local inference
Which **products** are affected?	CPE dictionary, vendor CSAF	cpe.gcve.eu, community-curated CPE/PURL
Is anyone **talking** about it?	Commercial threat feeds	Sightings from open sensors (GCVE-BCP-12)
Does anyone **disagree**?	Nobody, in a list	Attributed, contradictable assertions

- Every row brings back the same dependency on a **single external actor**.
- The rest of this talk walks through the rows: **federation and KEV**, **open models**, **open sensors**.

Open Infrastructure: Federated Consumption and KEV

Publication is solved. Consumption is the next problem

Settled by GCVE (previous talk)

- Each GNA publishes from its **own** endpoint (BCP-05).
- A signed directory says **where** to fetch (BCP-01, BCP-03).
- The publisher controls timeline, revisions and endpoint.

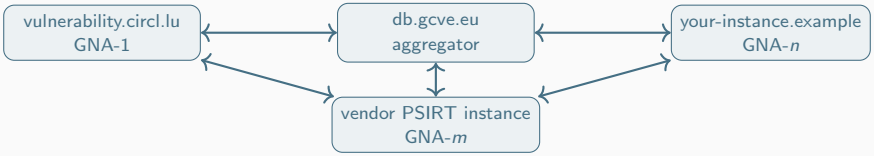
Still open, on the consumer side

- **Which** GNAs, catalogues and sensors do I follow?
- How do I **aggregate** 46 endpoints plus CVE, CSAF, OSV, CNVD without losing provenance?
- What when two sources **contradict** each other?
- How do instances **share** comments, sightings, exploitation claims?

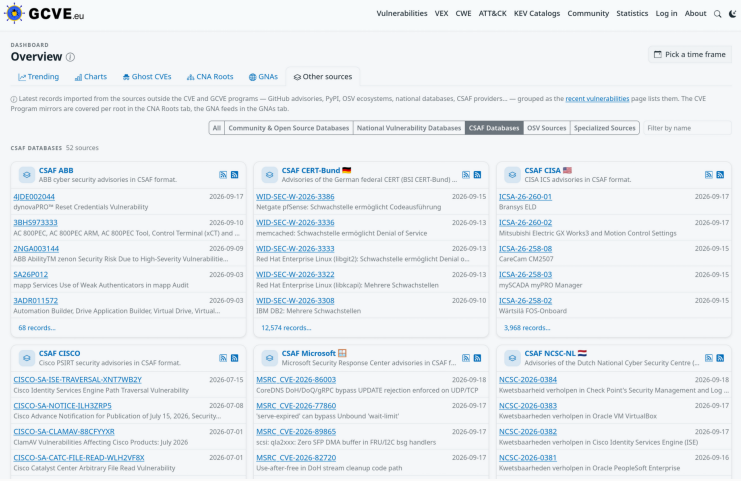
Publication became a **local decision with global reach**. Consumption must become a local decision too.

Federation in practice: Vulnerability-Lookup instances talk to each other

- Any public instance can **publish** its own advisories (as GNA or CNA) and **synchronise** advisories from other instances.
- Since version 4.0 (February 2026): **remote instance synchronisation** of **comments**, **bundles**, **sightings** and **KEV entries**.
- Retrieved records keep their **provenance** (`GNA-1`, `GNA-2`, ...): you always know who said what.
- Version 5.x (2026): multi-organisation **publication workflows** for coordinated vulnerability disclosure.



What a federated instance looks like: db.gcve.eu



`https://db.gcve.eu` – `https://vulnerability.circl.lu` – same software, different operators, same data model.

TLP:CLEAR

How to publish with GCVE in 4 steps

- **1.** Install `vulnerability-lookup`, or generate a feed compliant with `GCVE-BCP-05`.
- **2.** Apply to become a GCVE Numbering Authority (GNA) and submit your feed URL.
- **3.** Your feed is now discoverable and consumable by all software in the GCVE network.
- **4.** Make some tea.

The same four steps make you a **consumer**: install, point the instance at the directory, pull the GNAs you trust, make more tea. Publishing and consuming are the same software, in both directions.

The other shared catalogue: naming the product

- A vulnerability is only actionable once you know **which product** it affects.
- The CPE dictionary is one U.S.-maintained list, and a mess: `oracle:java` versus `sun:java`, missing vendors, no package URLs.
- **cpe.gcve.eu**[a] (June 2026): a collaborative catalogue of vendors, products, CPEs and PURLs.
- Discover, propose, moderate, distribute: reviewed in the open, exported as data and API.

[a] https://cpe.gcve.eu/

Why it matters for sovereignty

- Product naming decides what your scanner **matches**. Whoever owns the dictionary owns your blind spots.
- GNAs and vendors align their own product names instead of waiting for a dictionary update.

Same pattern as identifiers and KEV: a shared catalogue open to proposals and disputes.

Known Exploited Vulnerabilities: the most important list you do not control

- Tens of thousands of CVEs per year. Only a small fraction is ever **exploited in the wild**.
- "Is it exploited?" is the single most useful signal for **prioritisation**: patch deadlines, reporting, board decisions.
- The de facto reference is the **CISA KEV** catalogue: a U.S. federal list, built primarily for U.S. federal agencies, from evidence **we cannot see**.
- Existing KEV publications are **list-based and opaque**: they assert exploitation without saying **who** made the claim, **when** it was observed versus declared, on **what evidence**, **where**, and with **what confidence**.

A European hospital is told to patch within 48 hours because a vulnerability appeared on a list. Who decided? Based on what? Was it exploited in Europe? Can a European CSIRT say "we see it too", or "we do not"?

GCVE-BCP-07: from KEV lists to KEV assertions

- A KEV entry is an **assertion** made by an **observer** at a point in time. It is **not ground truth**, and it is **not part of the identity** of the vulnerability.
- Assertions can be **revised**, **withdrawn** or **contradicted** by other observers.
- One JSON object per assertion, linked to any identifier (CVE, GCVE, GHSA, . . .).
- Anyone can publish a catalogue: a CSIRT, a vendor, a honeypot operator, a researcher.
- Since September 2026: a **KEV directory** so that catalogues can be discovered, not just known by word of mouth.

What an assertion carries

- **status**: exploited true/false, with a reason: `confirmed`, `suspected`, `disputed`, `historical`, `withdrawn`
- **timestamps**: `first_seen_at`, `asserted_at`, `recorded_at`, `last_seen_at`
- **evidence[]**: type (observation, report, . . .), **confidence** 0.0–1.0, provenance
- **scope**: regions, victim countries, sectors, exposure
- **characteristics**: RCE, authentication required, severity
- **uuid** of the assertion, so it can be referenced and superseded

`https://gcve.eu/bcp/gcve-bcp-07/`

A KEV assertion, in short

```
{
  "vulnerability": {"id": "CVE-2025-53770"},
  "status": {"exploited": true, "reason": "confirmed"},
  "timestamps": {"first_seen_at": "2025-07-18T00:00:00Z",
                 "asserted_at":   "2025-07-20T09:12:00Z"},
  "scope": {"regions": ["EU"], "sectors": ["public administration"]},
  "evidence": [
    {"type": "observation", "confidence": 0.9,
     "source": "incident response, CIRCL"},
    {"type": "report", "confidence": 0.7,
     "source": "https://example.org/advisory"}
  ]
}
```

Illustrative example. **Who** asserted it, **when**, **why**, and **how sure** they are is now data, not folklore.

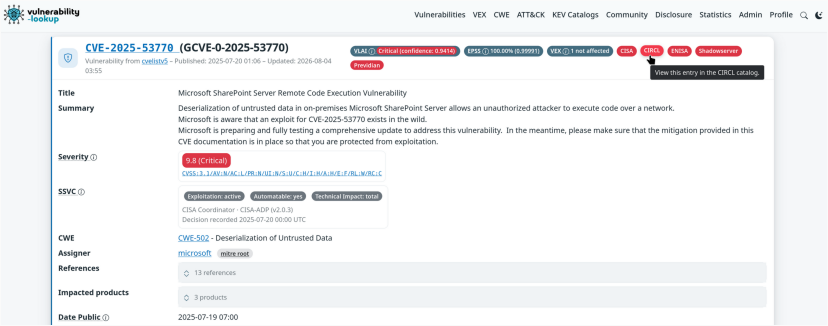
Five federated KEV catalogues, one format, side by side

- Vulnerability-Lookup aggregates five catalogues on its public instance[4], all as BCP-07 assertions: **CISA KEV**, **ENISA / EU CSIRTs Network**, **Shadowserver** (honeypots), **Previdian** (aggregation), **CIRCL** (hand-curated).
- About **2,900 vulnerabilities** tracked. A **coverage matrix** shows which catalogue references what, and where they **disagree**.
- Each catalogue is essentially **one kind of sensor**: vendor reports, CSIRT reports, honeypots, public reports.

	CISA	ENISA	Shadowserver	Previdian	CIRCL
CVE-2025-53770 (SharePoint)	✓	✓	✓	✓	✓
Many entries	✓	–	–	✓	–
GCVE-1-2026-0020	–	–	–	–	✓

[4] https://vulnerability.circl.lu/known-exploited-vulnerabilities-catalog/

One vulnerability, five independent exploitation claims



- CVE-2025-53770 (SharePoint): one badge per catalogue asserting exploitation: **CISA**, **CIRCL**, **ENISA**, **Shadowserver**, **Previdian**.
- Each badge opens the assertion: who, when, on what evidence. VLAI, EPSS and SSVC sit next to it, each attributed to its source.

What do KEV catalogues actually know?[5]

Catalogue	Median days	Before CVE
CIRCL	9	1
ENISA	117	2
CISA	252	60
Previdian	277	80
Shadowserver	532	15

Median days between CVE publication and listing, and number of entries listed before the CVE existed. Public instance, September 2026.

- **102 vulnerabilities** were listed as exploited **before** their CVE record was published: about one listing in thirty.
- For CVEs published in 2026, the median time to first listing is **7 days**; 51% are listed within a week.
- Only 42% of entries are backed by **both** observation and reports. No single catalogue corroborates itself.

> "A team that waits for the CVE record to act has, in cases like this one, been the last to know."

[5] https://www.vulnerability-lookup.org/2026/09/13/what-kev-ca...

The sovereignty lesson from KEV

- A single exploitation list is a **single sensor** with a **single bias**. You cannot audit it, and you cannot add to it.
- With an open assertion format, a **European CSIRT**, a **vendor** or a **honeypot operator** can publish their own view, and be **compared** against the others.
- Disagreement becomes structured, attributable **information**: `disputed` is a valid status.
- The same applies to **sightings** (GCVE-BCP-12): observations of a vulnerability in the wild, from any sensor, in a standard format.

Owning vulnerability intelligence does not mean having the biggest list. It means being able to **say what you see**, **show your evidence**, and let others **verify or contradict** you.

Open Models: VulnTrain and VLAI

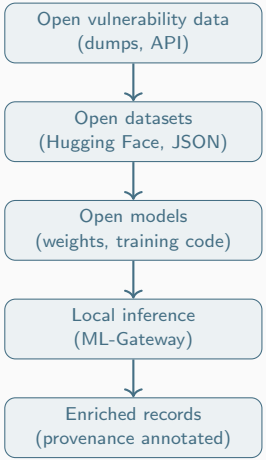
The next dependency: the intelligence layer

- Open data is useful. But increasingly, the **interpretation** is delegated to models: severity guesses, CWE classification, summaries, "AI-powered" prioritisation.
- If those models are **closed**, served by **proprietary inference APIs**, trained on **undisclosed data**, we have simply moved the dependency one layer up.
- Every vulnerability description you send to a remote model is also **data leaving your perimeter**. Sometimes before publication.

What happens when the data is open, but the **understanding** of it is not?

VulnTrain: from open data to open datasets and open models

- **VulnTrain**[a] turns Vulnerability-Lookup data into **datasets** and trains **models**, published on Hugging Face.
- Tasks: **severity** prediction (VLAI), **CWE** classification, CVSS characteristics extraction, mapping to **MITRE ATT&CK**.
- Small models (about 125M parameters) trained **locally** in hours; inference through **ML-Gateway**, **on premises**, no external call.
- Enrichment stored as a **traceable extension** (BCP-05-X-01): original record untouched, model version recorded.



[a] https://github.com/vulnerability-lookup/VulnTrain – https://huggingface.co/CIRCL

Why small, open, reproducible models

- **Reproducible**: dataset, training script, hyper-parameters and weights are public. Anyone can retrain and get comparable results.
- **Understandable**: a classifier with a confidence score, not an oracle. The output is **non-authoritative** by design.
- **Affordable**: a CSIRT budget, not a hyperscaler budget. Sovereignty that costs a billion is not sovereignty.
- **Composable**: exposed to agents through an open protocol (VulnMCP), so the "AI assistant" can also stay local.
- **Generative, still local**: patch2vuln[6] turns a git patch into a draft advisory with a local model, provenance recorded (BCP-05-X-02).

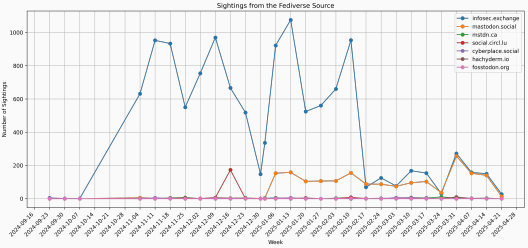
We enhance existing workflows with models we can **run, read and retrain**. We do not replace judgement with a remote black box.

[6] https://github.com/gcve-eu/gcve-lab-patch2vuln

Open Intelligence: Decentralised Sensors

The Fediverse as a sensor

- Vulnerabilities **emerge** in conversations before they are catalogued: researchers, admins, vendors, responders.
- **FediVuln**[a] follows public Mastodon streams, extracts vulnerability identifiers and pushes **sightings** to Vulnerability-Lookup.
- Same for Bluesky, Telegram, GitHub Gists, Nuclei, Metasploit, Exploit-DB, MISP.
- A federated network is a **decentralised sensor**: no single platform decides what we may observe.



Weekly sightings collected from the Fediverse.

[a] https://github.com/CIRCL/FediVuln

What the sensors see

In **August 2026** alone, the public instance collected **43,008 sightings** on **13,594** distinct vulnerabilities.

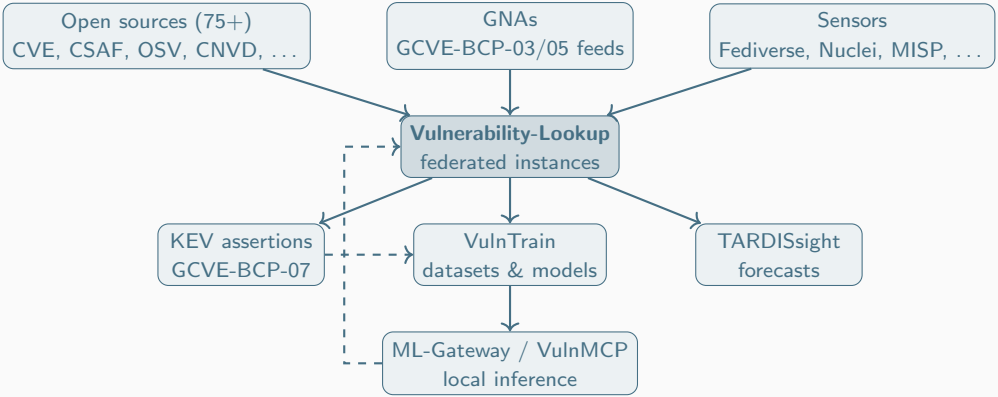
Sighting type	Sightings	Vulns
Seen (mentioned, discussed)	31,697	12,495
Published proof of concept	7,474	2,985
Exploited	3,595	883
Confirmed	242	70

- Every sighting keeps its **source** and **type**. Weighting by source is a **local policy**, not a vendor secret.
- **TARDISsight**[a] forecasts whether a wave of sightings is growing or fading. Open code, open method.
- GCVE-BCP-12 standardises the sighting format: **your** sensors can feed **anyone's** instance.

[a] https://github.com/vulnerability-lookup/TARDISsight – https://arxiv.org/abs/2604.16038

Intelligence We Actually Own

The architecture, end to end



open data → open identifiers → open infrastructure → open models → open intelligence

A sovereignty checklist for your vulnerability intelligence

Question	What it looks like when the answer is yes
Do I have a **copy**?	Full dumps on disk, in the original format, refreshed automatically.
Can I **name** a vulnerability?	I can allocate and publish an identifier without asking a central authority.
Can I **publish** on my terms?	My advisories and KEV assertions are fetched from my endpoint, with my policy.
Can I **disagree**?	I can dispute a record or an exploitation claim, and the dispute is visible to others.
Can I **verify**?	Directory signatures, provenance on every record, evidence on every assertion.
Can I **understand** the enrichment?	Models with public weights, training data and code; confidence scores; annotations marked as such.
Can I **run** it alone?	Everything still works if the upstream provider, API key or contract disappears.
Can others **build on it**?	Open licence, open format, open standard, open community.

Honest limitations

- We still **consume CVE**, NVD and CISA KEV, and we want to keep doing so. GNA 0 exists on purpose. Sovereignty is about **not depending only on them**.
- Federation moves the trust question to **you**: which GNAs, which catalogues, which sensors do you follow? This requires **curation**.
- Open infrastructure is not free. Vulnerability-Lookup and GCVE are funded by a public CSIRT and **EU co-funding**. Funding remains a sovereignty question, on this side of the Atlantic too.
- Open models are **smaller** than commercial ones. They do fewer things, but we know exactly what they do.
- Standards are only worth the number of implementations. The BCPs are living documents and need **reviewers and implementers**.

How to take part

If you run a CSIRT, a PSIRT, or an open-source project

- Become a **GNA**: gna@gcve.eu.
- Publish what you see as **BCP-07 KEV assertions**, even a handful a year.
- Run a **Vulnerability-Lookup** instance and synchronise.

If you are a developer, researcher, or citizen of the Fediverse

- Add a **source** or a **sensor** (a feeder is a small Python module), or review a **BCP**.
- Reuse the **datasets and models**, retrain, publish.
- Talk about vulnerabilities in public. You are already a sensor.

Sovereignty is not only about **where** our systems run, but about whether we can independently **understand**, **process**, **enrich** and **act upon** the security information they depend on.

References

https://www.vulnerability-lookup.org – https://gcve.eu

https://vulnerability.circl.lu – https://db.gcve.eu – https://cpe.gcve.eu

https://gcve.eu/bcp/gcve-bcp-07/ – KEV Assertion Format

https://github.com/vulnerability-lookup – https://github.com/gcve-eu – https://github.com/CIRCL/FediVuln

https://huggingface.co/CIRCL

https://social.circl.lu/@circl – https://social.circl.lu/@gcve

Thank you for your attention

- Issues, new sources of advisories or ideas:
  `https://github.com/vulnerability-lookup/vulnerability-lookup`
- Discussion: `https://discourse.ossbase.org/c/gcve/14` –
  `https://discourse.ossbase.org/c/vulnerability-lookup/6`
- For support and questions, contact: info@circl.lu
- If you want to become a GNA: gna@gcve.eu

Can we build vulnerability intelligence that we actually own?

We think so. It is already running, and it is open.