Exposure Signal Intelligence Platform

Why ESIP Signals Can Be Trusted

Deterministic Intelligence for Explainable Cybersecurity Decisions

---

Version 1.0 / June 2026 / Classification: External / Audience: Security architects, VM leads, CTEM practitioners, CISOs, procurement reviewers, technical evaluators

This document answers one question: why should a security team trust an ESIP signal enough to make a real-world remediation decision?

> **The core principle** ESIP does not ask consumers to trust a score. It enables consumers to verify why that score exists.

It is not an API reference and it is not an integration guide. Those documents exist separately. This document explains the philosophical and architectural foundations that make ESIP output trustworthy.

This document applies to all ESIP outputs, whether accessed through the Free Data Feed or the Commercial API. Commercial services provide additional Temporal Context and decision support fields, but the same principles of evidence, determinism, and explainability apply throughout the platform.

1. Introduction

Security teams are not failing to find vulnerabilities. They are failing to decide which ones matter.

The tools that exist today produce prioritised lists based on static scores. Those scores are calculated at the time of disclosure and do not update as the threat landscape changes. A CVE with a CVSS score of 9.8 received that score the day it was published. Whether that score reflects the current reality six months later, when attackers have built weaponized tools and exploit code has appeared in public repositories, is a different question entirely.

Traditional vulnerability intelligence often produces verdicts without context. Security teams are expected to make prioritisation decisions without understanding what changed, why the score is what it is, or what evidence supports it. When a tool says 'this is Critical,' practitioners are expected to act on that without knowing whether the score reflects a theoretical flaw or confirmed attacker activity.

ESIP was designed differently. Every ESIP output is evidence-based, deterministic, explainable, and reproducible. This document explains what that means and why it matters for the security decisions you make.

> **The question ESIP answers** What is becoming dangerous right now? Not what vulnerabilities exist. Not what the theoretical severity is. What is the current state of the threat landscape for this CVE, and is it getting better or worse?

2. What Explainability Means

Explainability in ESIP is not a feature. It is a structural property of the signal model.

In the context of AI-generated content, the word explainable often describes a system's attempt to describe its own reasoning in human terms. That is not what this means here. ESIP does not use AI reasoning to produce its outputs. There is no language model involved. There is no probabilistic inference that generates a verdict and then attempts to explain it afterward.

ESIP explainability means something specific and verifiable: every output can be traced back to the observable evidence that produced it, using deterministic rules that do not change between computations.

> **Explainability is the guarantee that a consumer can always answer six questions** What changed? Why did it change? What evidence supports it? Which sources contributed? When did it change? Can it be reproduced?

The answers to those six questions are not generated on demand by reasoning through the question. They are recorded at the moment the computation occurs and attached to every signal output. The evidence that produced a verdict exists as a permanent, immutable record. The rules that were applied exist in a versioned, auditable form. The verdict produced is identical every time the same evidence is evaluated against the same rules.

This is the foundation. Everything in ESIP's trust model follows from it.

3. From Evidence to Decision Support

ESIP transforms globally observable public data into decision-ready signals through a deterministic pipeline. Understanding each stage clarifies why the output can be trusted.

Each stage has a defined contract with the next. Observations must pass validation before becoming evidence. Evidence must satisfy threshold conditions before triggering a signal. Signal verdicts must satisfy deterministic scoring rules before producing a score. Temporal context is computed from the history of evidence events, not from the current score. Nothing passes from one stage to the next without being recorded.

The result is a complete audit trail. Not as an afterthought or a compliance feature, but as an architectural property. At any stage in the pipeline, the record of what happened, when it happened, and what produced it is preserved.

4. Every ESIP Signal Has Evidence

> **No evidence, no signal** ESIP never generates a signal in the absence of observable evidence. Every score, every verdict, every temporal context classification is the deterministic product of evidence that exists in the observable data. If no evidence has crossed a qualifying threshold, there is no signal.

No ESIP verdict is produced without observable evidence. There is no inference from historical patterns without a triggering event. There is no AI-generated risk assessment. There is no score adjustment based on undisclosed internal factors.

Every score reflects weighted evidence: which sources observed something, what they observed, how trusted those sources are, and how long ago the observation occurred. The score changes when new evidence arrives or when existing evidence ages. It does not change spontaneously.

What is backing each part of an ESIP verdict

Verdict element	What is backing it
Lifecycle stage	*The specific evidence events that advanced or held the stage. A Confirmed stage means a government agency (CISA or ENISA) has confirmed real-world exploitation. That is a factual event, not an assessment.*
Severity score	*A weighted sum of active evidence, adjusted for source trust tier and evidence age. Every input is named and recorded. No hidden factor contributes to the score.*
Confidence level	*Derived from the trust tiers of the contributing sources and how many independent sources are corroborating the same CVE. Verified means government-confirmed exploitation. That is the highest trust evidence available.*
Attack relevance	*Derived from mapped ATT&CK technique associations. The specific technique and tactic are named, not inferred.*
Temporal context	*Derived from the history of evidence events. Whether the threat is Escalating, Stable, or Declining reflects what has actually happened in the observable data over the last 14 days.*
What Changed	*A human-readable description of the most recent evidence event. The source, the observation type, and the threshold crossed are named explicitly.*

> **Absence of signal is accurate, not incomplete** A CVE with no active ESIP signal has not been assessed as safe. It means ESIP has not observed globally meaningful evidence for that class. The signal model does not produce outputs in the absence of evidence. Silence is accurate, not a gap.

5. Temporal Context

A score is a snapshot. A signal is a story.

Scores describe state. Signals describe change. Temporal Context explains that change.

Most vulnerability intelligence tools answer the question: how dangerous is this vulnerability? ESIP also answers: is this getting better or worse?

A score alone cannot tell you this. A CVE held at Critical by a government exploitation floor can simultaneously have new weaponized exploit activity appearing from independent sources. The score does not move because the floor is already applied. But the threat picture is actively worsening. Without Temporal Context, both that CVE and a stable, long-running Critical CVE look identical in your prioritisation queue. They are not.

Temporal Context field	What it tells you
Current trend	*Whether the threat is Escalating, Stable, or Declining. Derived from the pattern of evidence events in the last 14 days. Not from the score. A score can be flat while a trend is Escalating.*
Trend drivers	*Why the current trend has the value it does. Each driver is a named, specific evidence event: a new weaponized source appeared, an additional organization corroborated the same CVE, exploit probability increased sharply. Not an AI summary: a list of what actually happened.*
Latest material change	*When something significant last happened. The most recent exploitation confirmation, weaponized exploit detection, or significant exploit probability movement. Lets you distinguish a CVE that changed yesterday from one that has been static for six months.*
Time since last material change	*How many days since the last meaningful evidence event. Computed at response time, not stored as a prediction.*
Signal velocity	*How rapidly evidence is accumulating. High velocity means three or more significant evidence events in the last 14 days. This is a rate-of-change indicator, not a severity indicator: a CVE with a low severity score can have High velocity if evidence is suddenly appearing.*
Threat age	*How long this CVE has been active in the global threat landscape. Useful for distinguishing a new emergent threat from a long-running confirmed one that your VM programme has already addressed.*

Temporal Context is computed from evidence history. It is deterministic, versioned, and auditable. The same evidence history always produces the same trend classification. It is not

generated by AI or editorial judgment. A practitioner who disagrees with a trend classification can trace the specific evidence events that produced it and understand exactly why.

6. Deterministic by Design

This is not a marketing statement. It is a verifiable architectural property. ESIP signals are computed by a deterministic scoring engine using versioned rules. When the same source observations are evaluated against the same scoring rules at the same point in time, the result is always identical. There are no stochastic elements, no random sampling, no probabilistic reasoning that introduces variance between runs.

> **What determinism means in practice** If you receive an ESIP signal showing a CVE at severity 70 with current_trend Escalating and two weaponized sources contributing, that output can be reproduced from the source observations that produced it. Any authorized technical reviewer can verify the calculation by re-evaluating the same evidence against the same rules and confirming the same verdict. This is the difference between explainable intelligence and a black box.

Determinism enables four properties that matter for security decision-making:

- Auditability. A security team can explain every prioritisation decision to a CISO or a compliance auditor without saying 'the tool said so.' The evidence is there. The rules are versioned. The verdict follows from both.
- Reproducibility. Any past verdict can be reconstructed from its source observations. This means a signal from three months ago can be verified against the evidence available at that time. Historical decisions have a traceable basis.
- Correction transparency. When a signal is corrected, the correction is recorded, versioned, and visible. Consumers monitoring their integration will detect corrections automatically. There are no silent updates.
- Trust under scrutiny. In a regulatory review, a partner integration, or a security incident, ESIP outputs can be presented with their evidence chain intact. The verdict is not an opinion. It is a deterministic conclusion from observable facts.

7. Explainability in Practice

Note: some of the examples in this section use commercial ESIP fields such as Temporal Context and Trend Drivers. These are used to illustrate the explainability model at its fullest. The underlying principles of evidence, determinism, and traceability apply equally to the free and commercial tiers.

Here is what explainability looks like for a real CVE. CVE-2026-41940 is a cPanel/WHM authentication bypass. It appears in CISA KEV: real-world exploitation confirmed.

Two practitioners receive this signal in the same week.

Without Temporal Context

Both practitioners see: Confirmed, Critical, Verified.

That tells them exploitation is confirmed. The government has verified it. The score is at the Critical floor. Both practitioners treat it as a high-priority remediation item and put it in their next patching cycle.

With Temporal Context

The first practitioner receives the signal on Monday. current_trend is Stable. No material changes in the last 14 days. Threat age 200+ days. This is a long-running confirmed threat their team has likely already evaluated.

The second practitioner receives the signal on Wednesday, two days later. current_trend is now Escalating. The trend_drivers list shows:

Y	New weaponized exploit detected in ExploitDB (4 days ago)
Y	New weaponized exploit detected in Metasploit (13 days ago)
Y	EPSS exploit probability increased significantly (4 days ago)
Y	Three independent sources now corroborating the same CVE simultaneously

signal_velocity: High. The threat picture is actively worsening.

These are not interpretations or AI-generated risk assessments. They are the named, specific evidence events that occurred. Each one is traceable to a source, a timestamp, and an observation.

The second practitioner escalates. They notify their team that this CVE needs immediate attention, not because a score changed, but because the evidence shows active, accelerating attacker interest. They can explain every part of that decision to their CISO using named evidence. Not 'the tool flagged it.' 'These specific things happened in the last four days.'

> **This is the difference** Without temporal context, both CVEs look identical: Confirmed Critical Verified. With temporal context, one is a stable long-term threat and the other is actively getting worse. Same score. Completely different situation.

8. What ESIP Does Not Do

Defining what ESIP does not do is as important as explaining what it does. A trust model requires clear boundaries. ESIP's explainability guarantees apply to what ESIP produces. They do not extend to inferences ESIP does not make.

ESIP does not...	Why this boundary exists
Know your asset inventory	*ESIP signals describe global class-level threat state. Whether CVE-2026-41940 is dangerous in your environment depends on whether you run cPanel/WHM. ESIP does not know what you run.*
Know your business criticality	*A Critical CVE affecting a test server is different from a Critical CVE affecting a payment processor. ESIP does not know which systems in your environment are business-critical.*
Estimate exploit probability	*EPSS scores (from FIRST.org) are ingested as evidence inputs. ESIP does not generate exploit probability estimates. It uses EPSS delta as a signal trigger.*
Estimate time remaining before exploitation	*ESIP does not predict when a CVE will be exploited in your environment. It describes the current and recent state of the global threat landscape for that CVE class.*
Recommend patch order	*ESIP produces the global intelligence half of a prioritisation decision. You supply the asset context half. The two together produce a prioritisation decision. ESIP alone does not.*
Replace human judgment	*ESIP provides deterministic threat intelligence to inform decisions. The decision itself: what to patch, in what order, with what urgency, given your business context, requires human judgment that ESIP cannot and should not replace.*
Generate AI narratives	*Temporal context fields are computed by a deterministic classification engine. Current trend, trend drivers, and velocity are not AI-generated summaries. They are structured outputs from deterministic rules. The same evidence always produces the same classification.*

ESIP provides the global intelligence half of a prioritisation decision. You provide the asset context half. Together, those two things answer the question practitioners actually need answered.

9. Consumer Guarantees

ESIP makes seven explicit guarantees to every consumer. These are not aspirational commitments: they are structural properties of the system that can be verified by any authorized technical reviewer.

	Guarantee	What it means
1	**Evidence-backed**	Every verdict is linked to the specific observations that produced it. The evidence is not inferred, estimated, or generated. It is recorded at the time of ingestion and preserved permanently.
2	**Deterministic**	Given the same source observations and the same scoring rules, ESIP always produces identical verdicts. There is no randomness, no AI variance, no black-box factor.
3	**Versioned**	Every schema change, every scoring rule change, and every temporal context rule change is versioned. Consumers know exactly what rules produced the output they received. When rules change, the version changes.
4	**Explainable**	Every verdict carries a human-readable explanation of what changed and why. Every temporal context field carries a structured list of the specific evidence events that produced it.
5	**Auditable**	The evidence chain from raw observation to final verdict is preserved and accessible. Any authorized reviewer can trace the path from a source observation to the score it contributed to.
6	**Reproducible**	Any past verdict can be reconstructed from its source observations. Historical security decisions made using ESIP signals have a traceable, verifiable basis.
7	**Transparent corrections**	When a signal is corrected, the correction is recorded and versioned. Consumers monitoring their integration detect corrections automatically. No silent updates.

> **The responsibility boundary** ESIP provides global class-level threat intelligence. Consumers provide asset context and business criticality. The prioritisation decision emerges from both. ESIP's guarantees apply to its half of that decision.

10. Conclusion

Traditional vulnerability intelligence tells you that a vulnerability exists. CVSS tells you how severe it could theoretically be. KEV tells you whether the government has confirmed exploitation. These are valuable inputs.

ESIP tells you what changed, why it matters, and how the threat landscape is evolving. Every output is grounded in observable evidence and deterministic rules that can be explained, audited, and reproduced.

When a security team uses ESIP to escalate a CVE to immediate remediation, they are not acting on a risk score produced by an algorithm they cannot see. They are acting on named evidence: these specific sources observed this specific attacker activity, at these specific times, which satisfies these specific threshold conditions, which produced this specific verdict.

That is a security decision that can be explained to a CISO, defended in a regulatory review, and traced back to its evidence chain in a post-incident investigation.

The verdict is not an opinion. It is a deterministic conclusion from observable facts.

That is what it means to trust a signal.

The ESIP documentation set covers three dimensions of this product:

Document	What it answers
ESIP Overview v1.6	*What is ESIP, how does it work, and how do practitioners use it?*
Public Output Contract v1.5	*What will I receive, what does it mean, and how stable is it?*
Why ESIP Signals Can Be Trusted (this document)	*Why should I trust an ESIP signal enough to make a security decision?*

For access to technical integration documentation, contact ZenzizenSec or register for a free tier API key at exposuresignal.io.

Why ESIP Signals Can Be Trusted / June 2026