Analyst's Guide to Cyber Risk Data Sources

What Data to Use to Measure, Monitor, and Manage Cyber Risk

May 30, 2025 | DRAFT FOR COMMENT

Authors:

Todd Tucker
Managing Director
FAIR Institute
TTucker@FAIRInstitute.org

Table of Contents

Executive Summary

Cyber risk decisions are only as good as the information that informs them. The FAIR (Factor Analysis of Information Risk) model provides a structured, quantifiable way to analyze risk, and data, especially when timely and relevant, can greatly improve the quality of that analysis.

This guide maps each factor in the FAIR model to commonly available data sources, including telemetry, threat intelligence, industry reports, and financial records. It clarifies where and how data supports not only initial analysis but also ongoing risk monitoring and management.

Importantly, the guide emphasizes that perfect data is not required. Risk analysts can combine available data with structured expert judgment to produce credible and defensible results. It also highlights the role of the cyber risk engineer in integrating telemetry and improving data pipelines across the organization. Whether you're launching or scaling a FAIR-based program, this guide offers practical insight for using data to make better, faster, and more transparent cyber risk decisions.

We welcome feedback and questions about this paper. Please email us at feedback@fairinstitute.org.

Key Takeaways

- **Data Reduces Uncertainty**: In the FAIR model, risk is about uncertainty. Data—especially measured and telemetry-based—helps reduce that uncertainty to support better decisions.
- **Not All Data Is Equal**: FAIR supports a spectrum of data types—measured, estimated, and derived. Calibrated expert judgment is valid and often necessary, especially when empirical data is sparse.
- **Telemetry Is Transformational**: System-generated data (telemetry) is a key enabler for scaling cyber risk analysis and transitioning from point-in-time assessments to ongoing monitoring.
- **Integration Matters**: Embedding FAIR into a cyber risk management system with automated data flows enables real-time visibility and faster response to changes in risk.
- **Cyber Risk Engineers Are Critical**: These professionals serve as the bridge between data owners and risk teams, ensuring that data pipelines are accurate, aligned, and operationalized effectively.
- **Imperfect Data Is Not a Barrier**: Most FAIR analyses begin with incomplete data. What matters most is a consistent and transparent approach to estimation and documentation.
- **FAIR Is Flexible**: Analysts don't always need to estimate at the most granular levels (e.g., Contact Frequency or Probability of Action). The model accommodates estimation at the level best suited to the available data and scenario complexity.
- **Public Data Has Value**: Industry reports (e.g., Verizon DBIR, Cyentia IRIS) and open-source threat intelligence offer valuable benchmarks that can inform estimates and improve defensibility.

The Role of Data in FAIR-Based Cyber Risk Management

The FAIR (Factor Analysis of Information Risk) model is designed to help organizations understand, quantify, monitor, and manage information risk in a consistent, repeatable, and defensible way. At the heart of FAIR is the idea that a combination of frequency and magnitude of loss events drives risk. While the model is inherently flexible and can be applied using informed judgment alone, data, when available and relevant, can greatly enhance the accuracy and credibility of FAIR-based practices.

Data as a Tool to Reduce Uncertainty

Risk, in FAIR, is fundamentally about uncertainty. Each factor in the FAIR model (such as Threat Event Frequency, Vulnerability, or Loss Magnitude) represents an uncertain quantity. The goal of risk analysis is not to predict the future with certainty, but to reduce uncertainty to a level that supports informed decision-making. Data helps in this reduction by anchoring our estimates in observed patterns, measurements, or structured knowledge.

The FAIR Approach to Using Data

FAIR does not require high-precision measurements to be effective. Instead, it embraces the idea that estimates—especially those made through calibrated expert judgment—can be just as useful as empirical measurements *when appropriately framed*. The use of BetaPERT distributions and Monte Carlo simulation allows analysts to model uncertainty explicitly, incorporating data in a way that reflects its confidence level and range.

FAIR distinguishes between:

- **Measured Data**: System logs, financial records, event counts—objective and directly observed.
- **Estimated Data**: SME input, calibrated expert judgment—subjective but structured.
- **Derived Data**: Values computed by combining other inputs (e.g., Loss Event Frequency = Threat Event Frequency x Vulnerability).

Telemetry—data automatically collected from systems, networks, and security tools—is a key source of measured data. When integrated into a cyber risk management system, telemetry can serve as a robust foundation for both initial FAIR-based analyses and ongoing monitoring. Real-time and historical telemetry data enables organizations to:

- Populate and update risk models with observed metrics
- Track changes in threat activity, control effectiveness, and exposure
- Continuously assess and recalibrate risk over time

All three forms of data—measured (including telemetry), estimated, and derived—play valid roles in FAIR-based risk analysis, quantification, monitoring, and management.

Data Challenges in Cyber Risk Programs

Many organizations assume they need perfect data to conduct FAIR analyses. In reality, perfect data doesn't exist; risk decisions are routinely and confidently made with data of varying degrees of quality. Common challenges include:

- Lack of internal incident data or breach history
- Unstructured or inconsistent event logging
- Difficulty mapping technical telemetry to FAIR factors
- Variability in external threat intelligence quality
- Wide ranges of loss magnitude amounts in public data sources

This guide provides practical strategies for aligning available data, including telemetry, expert input, and external reports, with each FAIR factor. The goal is not perfection, but transparency, consistency, and better-informed risk decisions that support ongoing monitoring and strategic risk management.

The Value of Transparency and Defensibility

Data does not eliminate uncertainty, but it provides a defensible basis for estimating uncertain outcomes. FAIR encourages documenting sources, assumptions, and confidence levels. This not only enhances the credibility of the analysis but also improves its utility for decision-makers, auditors, and regulators.

In the following sections, we will map specific types of data to the FAIR model's components, helping practitioners identify what's useful, what's not, and where to focus data collection or estimation efforts for continuous improvement in risk analysis, quantification, monitoring, and management.

Mapping FAIR Factors to Data Sources

This section provides a structured overview of how different types of data—internal and external telemetry and non-telemetry—can be aligned with the components of the FAIR model. It includes guidance on where to find data, how to assess its relevance, and how it supports both point-in-time analysis and continuous monitoring.

Table: Data Sources by Mapped to FAIR Model

FAIR Factor	Internal Sources	External Sources	Usage Notes
Loss Event Frequency (LEF)	Incident logs, risk register trends	Industry-specific ISACs[1], DBIR, sector breach reports	Can be estimated directly or derived from TEF x Susc
Threat Event Frequency (TEF)	IDS/IPS logs, SIEM alerts, past incident logs	Threat intel feeds (e.g., Mandiant, Recorded Future, Crowdstrike), Verizon DBIR	Filtered by threat community relevance
Contact Frequency (CF)	Auth logs, firewall logs, access records, identity and access management (IAM) systems	Architecture models, MITRE ATT&CK mappings	Often inferred from design/configuration
Probability of Action (PoA)	Internal threat profiling	Threat profiling, adversary behavior reports	Based on attacker incentives and capability
Susceptibility (Susc)	Derived: TCap vs. RS comparison	Industry breach reports[2]	Lower RS compared to TCap implies higher Susc

[1] Information Sharing and Analysis Centers, such as

[2] It may be possible to estimate susceptibility from the experience of others in your industry. For example, for a new threat vector or threat actor, you may be able to perform a high-level assessment from reports of other breaches in your industry when internal RS data is not yet available.

FAIR Factor	Internal Sources	External Sources	Usage Notes
Threat Capability (TCap)	Red team results, incident forensics	MITRE ATT&CK, threat intel feeds (e.g., Mandiant, Recorded Future, Crowdstrike), Infragard bulletins, industry ISACs	Estimate median or 90th percentile threat capability for community
Resistance Strength (RS)	Pen test results, vulnerability scan outputs, third-party risk assessments (including outside-in scans)	Control maturity benchmarks, audit reports	Must be assessed relative to TCap
Primary Loss (PL)	Business impact assessments (BIAs), finance systems, asset valuation, impact logs	IBM Cost of a Breach, Cyentia IRIS, Ponemon Institute, HMITH[3]	Align to asset and scenario type
Secondary Loss Event Frequency (SLEF)	Legal case tracking, PR response history	SEC disclosures, crisis reports	Stakeholder reaction likelihood
Secondary Loss Magnitude (SLM)	Legal settlements, PR/media spend, churn models	IBM Cost of a Breach, Cyentia IRIS, Ponemon Institute, HMITH	Often underestimated in early models

[3] www.HowMaterialIsThatHack.org from the FAIR Institute

Loss Event Frequency (LEF)

Loss Event Frequency represents the expected number of times per year that an organization will experience a loss event. It is typically derived by multiplying Threat Event Frequency by Susceptibility, but it can also be estimated directly using empirical data.

Internal/Telemetry Sources:

- Incident response data showing actual security events resulting in loss
- Security operations center (SOC) logs indicating confirmed successful exploits
- Historical loss event logs stored in risk registers or ticketing systems

External/Non-Telemetry Sources:

- Industry-specific breach databases (e.g., FS-ISAC for financial services, H-ISAC for healthcare)
- Public reports such as the Verizon DBIR and Cyentia IRIS Risk Retina
- Sector risk benchmarks and aggregated loss data from insurers or regulators

Usage Guidance:

- When strong internal data is available, LEF can be estimated directly, supporting historical trend analysis.
- In the absence of internal loss data, external sources can provide reference points for similar organizations or industries.
- Whether derived or estimated directly, LEF should be expressed as a range to reflect inherent uncertainty and variability.

Threat Event Frequency (TEF)

Threat Event Frequency represents the frequency at which a threat community is expected to act against an asset. This includes attempts to breach, disrupt, or misuse the asset in a manner that could result in loss.

Internal/Telemetry Sources:

- SIEM platforms tracking intrusion attempts, scanning activity, and unauthorized access attempts
- IDS/IPS logs showing frequency of threat actor behaviors mapped to specific assets or asset types
- Incident management systems capturing attempted attacks

External/Non-Telemetry Sources:

- Threat intelligence feeds from commercial (e.g., Mandiant, CrowdStrike, Recorded Future) and open sources

- Industry and government reports such as Verizon DBIR, ENISA Threat Landscape, and ISAC bulletins
- Reports mapping threat campaigns to specific threat communities

Usage Guidance:

- TEF should be scoped to a well-defined threat community (e.g., cybercriminals, insiders, APT groups).
- It is often helpful to model TEF as a range informed by both external intelligence and internal telemetry.
- Overlap with Contact Frequency must be avoided—CF is about *opportunity*, TEF about *intent + opportunity*.

Contact Frequency (CF)

Contact Frequency represents how often a threat community interacts with or has access to an asset. Unlike TEF, CF does not require intent to cause harm—it reflects opportunity alone.

Most analysts do not estimate CF directly in practice. Instead, they typically estimate Threat Event Frequency at a higher level, which includes both CF and PoA. However, for advanced use cases or specific scenarios where additional fidelity is needed, estimating CF separately may be warranted.

Internal/Telemetry Sources:

- Firewall logs indicating attempted connections from external or internal IPs
- Network flow data showing communication patterns with critical assets
- VPN and endpoint access logs showing authorized and unauthorized touches
- Identity data to show the number of users, both privileged and unprivileged, and details such as location, role, and reporting relationships

External/Non-Telemetry Sources:

- Threat modeling diagrams (e.g., STRIDE, DFDs)
- MITRE ATT&CK mapping of potential access vectors
- Architecture documentation and asset exposure assessments

Usage Guidance:

- CF provides a baseline for estimating how *accessible* an asset is to a given threat community.
- High CF does not necessarily mean high risk unless combined with a capable and motivated adversary.
- CF is particularly relevant for assets exposed to the internet, third-party systems, or high-volume internal environments.

Probability of Action (PoA)

Probability of Action represents the likelihood that a threat actor will take harmful action when presented with the opportunity (i.e., once contact has been made). It reflects adversary intent and motivation.

As with Contact Frequency, most analysts do not estimate PoA directly. Instead, TEF is often calculated as a composite of CF and PoA. Estimating PoA separately may be valuable in scenarios involving detailed threat modeling or advanced simulations.

Internal/Telemetry Sources:

- None typically available at scale; adversary decision-making is not directly observable from internal logs

External/Non-Telemetry Sources:

- Threat actor profiles from intelligence vendors or law enforcement
- MITRE ATT&CK and other behavioral frameworks mapping threat actions and tendencies
- Industry reports on threat motivations, such as financial gain, disruption, or espionage

Usage Guidance:

- PoA is most relevant when analysts need to distinguish between frequent benign contact and actual adversary action.
- Motivation may vary significantly across threat communities (e.g., insiders vs. cybercriminals).
- In most cases, analysts estimate TEF directly based on observable attack patterns, which implicitly includes PoA.

Susceptibility (Susc)

Susceptibility represents the probability that an asset will be affected if a threat event occurs. It is a function of the threat actor's capability and the strength of resistance provided by controls protecting the asset.

Internal/Telemetry Sources:

- Penetration test findings identifying exploitable weaknesses
- Security control telemetry (e.g., endpoint detection, logging failures)
- Patch and configuration management data

External/Non-Telemetry Sources:

- Control maturity assessments (e.g., NIST CSF implementation levels)
- Benchmarking data from peers or industry groups
- Threat actor capabilities and behavior from industry reports

Usage Guidance:

- Susceptibility is derived from comparing Threat Capability and Resistance Strength.
- In FAIR, lower Resistance Strength relative to Threat Capability implies higher Susceptibility.
- Quantifying Susceptibility can support sensitivity testing and help prioritize control improvements.

Threat Capability (TCap)

Threat Capability represents the skill and resources of the threat actor community targeting an asset. It is a crucial component in determining the likelihood that an attacker can circumvent existing controls.

Internal/Telemetry Sources:

- Red team or purple team exercise results that demonstrate adversary simulation outcomes
- Incident response reports detailing techniques and tools used in past breaches

External/Non-Telemetry Sources:

- MITRE ATT&CK mappings of adversary capabilities by group
- Threat intelligence provider reports profiling APTs, cybercriminal gangs, or insider threats
- Sector-specific advisories and community intelligence (e.g., ISAC bulletins)

Usage Guidance:

- TCap should be estimated for a specific threat community, not a general population.
- The FAIR model supports estimating TCap using a percentile approach (e.g., 90th percentile attacker).
- Use consistent calibration methods across scenarios to support comparative analysis.
- TCap helps determine susceptibility when compared to Resistance Strength.

Resistance Strength (RS)

Resistance Strength represents the ability of an asset or environment to withstand an attack by a threat community. It is assessed in the context of Threat Capability to determine the asset's Susceptibility.

Internal/Telemetry Sources:

- Vulnerability scanning and patch management metrics
- Endpoint protection logs, firewall rule effectiveness, and SIEM control coverage
- Penetration test outcomes and red team assessments
- Third-party scans (outside-in and/or outside-in assessments)

External/Non-Telemetry Sources:

- Industry benchmarks (e.g., CIS Controls, NIST CSF maturity assessments)
- Third-party audits and control effectiveness reviews
- External control libraries or security frameworks with scoring systems

Usage Guidance:

- RS is always relative to the Threat Capability of a specific adversary community.
- Higher RS implies a lower probability of successful exploitation, reducing Susceptibility.
- Control telemetry can provide valuable empirical support for RS estimates, but calibration to threat skill level remains critical.
- Gaps in coverage or ineffective implementation can significantly weaken Resistance Strength.

Primary Loss (PL)

Primary Loss represents the direct financial impact to the organization from a loss event. This includes immediate costs such as detection, response, remediation, lost productivity, and direct revenue losses.

Internal/Telemetry Sources:

- Financial and accounting records detailing historical incident costs
- Business impact assessments (BIAs) from business continuity planning
- Time-tracking or case management systems used during incident response
- Ticketing systems logging resource hours and resolution times

External/Non-Telemetry Sources:

- Industry reports like IBM Cost of a Data Breach and Cyentia IRIS
- Peer benchmarking data or data from cyber insurance providers
- Academic and consulting firm studies analyzing incident cost distributions

Usage Guidance:

- Align PL estimates with the specific asset and scenario being analyzed (e.g., data breach, ransomware, denial-of-service).
- Categorize primary loss components (e.g., IT reconstitution, legal fees, business interruption) to improve granularity.
- Use ranges and distributions to capture the wide variability in primary loss outcomes across different events and contexts.
- Calibrated expert judgment is often necessary to estimate cost magnitudes for scenarios with limited historical precedent.

Secondary Loss Event Frequency (SLEF)

Secondary Loss Event Frequency represents how often a secondary stakeholder (such as regulators, customers, or partners) will impose additional consequences after a primary loss event. This factor helps capture reputational, legal, or contractual aftershocks.

Internal/Telemetry Sources:

- Legal or compliance team records of past follow-up actions or fines
- Communications logs from customer support and public relations
- Historical records of secondary stakeholder impacts following incidents

External/Non-Telemetry Sources:

- Regulatory disclosures and enforcement databases (e.g., GDPR, SEC)
- Public breach databases documenting lawsuits or settlements
- Reports from Cyentia, Deloitte, and legal analysis firms on breach follow-ons

Usage Guidance:

- SLEF should be estimated for relevant stakeholder categories depending on the scenario (e.g., customers, regulators, partners).
- Historical precedent and industry sector can significantly influence the frequency of secondary loss events.
- Regulatory environments with high sensitivity or strict penalties tend to increase SLEF.

Secondary Loss Magnitude (SLM)

Secondary Loss Magnitude represents the financial impact of secondary stakeholder reactions. This includes legal judgments, regulatory fines, reputation damage leading to churn, and increased cost of capital.

Internal/Telemetry Sources:

- Legal department cost tracking from prior settlements or defense efforts
- Customer churn analytics and reputational damage assessments
- Insurance claims documentation

External/Non-Telemetry Sources:

- Industry loss studies from Ponemon, Cyentia, and Forrester
- Publicly disclosed fines or class action settlements
- Market research on brand impact and consumer trust metrics

Usage Guidance:

- SLM should reflect the magnitude of each stakeholder reaction, aligned to the scenario context.
- Certain industries (e.g., healthcare, financial services) have higher typical SLM due to stringent regulation and customer expectations.
- It's common to underestimate SLM, so using reference data from public incidents or legal benchmarks can improve accuracy.
- Document the assumptions behind secondary loss drivers to improve transparency and repeatability.

The Role of Asset Data in FAIR Modeling

Assets are foundational to every cyber risk scenario. According to the FAIR Cyber Risk Scenario Taxonomy, an asset is "anything of meaningful business value that can be affected in a manner that results in loss"—whether it be data, business processes, digital services, physical infrastructure, or financial instruments. Because FAIR scenarios are defined in terms of the potential for loss to these assets, a clear understanding of assets and their attributes is essential for scoping, modeling, and quantifying risk.

Asset Data as a Scoping Anchor

Accurate asset data helps define the scope of analysis by grounding it in the real-world inventory of what the business values. This data enables risk teams to:

- Identify which business assets are most critical or exposed
- Filter telemetry and control data by asset relevance
- Prioritize scenarios based on business impact rather than purely technical concerns

A clear understanding of asset types (e.g., sensitive personal data, revenue-generating business processes, cash equivalents) provides context that informs Threat Event Frequency (TEF), Susceptibility, Resistance Strength (RS), and both primary and secondary loss magnitude.

Data about technical assets such as servers, databases, workstations, mobile devices, and more, work best when they are related to their business function (i.e., what business capabilities or processes they support). This provides the context necessary to link asset vulnerabilities (Susc) to loss magnitude.

Common Asset Data Sources

FAIR-aligned programs often rely on data sources such as:

- **CMDBs (Configuration Management Databases)** for identifying infrastructure and software assets and their interdependencies
- **Cloud management tools (e.g., AWS Config, Azure Resource Graph)** for inventorying cloud-based services, containers, and data stores
- **Enterprise Asset Management (EAM) platforms** for tracking operational technology (OT), facilities, and physical assets
- **Data classification and discovery tools** for mapping where sensitive data resides and how it flows
- **Identity management systems** for associating assets with owners, access levels, and privileges

Asset Telemetry and Risk Modeling

Telemetry data from these sources provides dynamic, real-time insights that enhance not only initial risk quantification but also continuous monitoring. For example:

- Changes in asset exposure (e.g., cloud misconfigurations, new data stores) can drive updates to Susceptibility or Threat Event Frequency.
- Inventory aging and asset end-of-life data can influence assessments of Resistance Strength or control effectiveness.
- Usage patterns and asset criticality metrics can guide the selection of Primary Loss (PL) categories and amounts.

By integrating asset telemetry into a cyber risk management system, organizations can automate the identification of relevant scenarios, streamline analysis updates, and support risk intelligence at scale.

Considerations for Integrating Data into a Cyber Risk Management System

Integrating data into a cyber risk management system (CRMS) enables organizations to operationalize FAIR-based risk practices. Depending on the CRMS used, effective integration allows continuous monitoring, provides real-time insights, and facilitates the automation of key elements of the analysis and reporting process. However, realizing these benefits requires thoughtful design and collaboration across stakeholders.

Key Considerations:

- **Cyber Risk Engineer Role:** A cyber risk engineer plays a critical role in bridging the gap between technical system owners and risk analysts. This role involves identifying data sources, developing ingestion pipelines, validating data integrity, and aligning telemetry with

FAIR model factors. Professionals familiar with integrating cybersecurity tools, such as engineers working with SIEM tools, would likely possess the necessary skills.

- **Engagement with System Owners:** Successful integration requires collaboration with owners of telemetry-producing systems, such as SIEMs, CMDBs, ticketing platforms, and financial systems. These stakeholders help ensure data is accessible, interpretable, and updated.
- **Mapping and Transformation:** Many telemetry sources require transformation before they can be used in FAIR models. This may involve mapping log entries to threat event types, translating technical metadata into frequency counts, or normalizing financial impacts for loss estimation. Many cybersecurity teams create a data warehouse to store data from various cybersecurity tools; such warehouses often contain data that is already mapped and transformed.
- **Data Quality and Governance:** Continuous improvement in cyber risk analysis depends on high-quality data. Establishing data quality metrics, conducting routine validation, and documenting lineage and assumptions are all key elements of effective governance. AI continues to make data quality management easier to perform.
- **Feedback Loops**: Build feedback loops so that risk insights prompt action, and those actions are then reflected in updated data. For example, if vulnerability telemetry is used to assess risk and the business remediates findings, that same telemetry should reflect the mitigations. This creates a self-updating cycle: action leads to measurable change, which then informs the next round of risk analysis, eliminating the need for separate tracking.
- **Automation and AI:** Where feasible, automate data updates into the risk platform to reduce manual effort. Consider the use of AI, which can ingest and transform unstructured data such as documents (e.g., third-party contracts, completed questionnaires, audit reports) and spreadsheets (e.g., asset lists) in ways that weren't feasible just a few years ago.

Usage Guidance:

- Integrating data into a cyber risk management system is a strategic investment that enables scale, agility, and defensibility.
- Prioritize data sources that are closest to the FAIR model's inputs and most relevant to the organization's top risk scenarios.
- Assign ownership of data pipelines and ensure regular validation and recalibration practices are in place.

An integrated system enables the risk team to shift from periodic, snapshot-based analyses to a living model of cyber risk that evolves with the environment.

