Reliable Context

RELIABLE CONTEXT / RESEARCH & PRACTICAL GUIDANCE

NIST AI RMF Explained: Functions and Data Provenance

> Reliable Context is a kernel of truth: a system of record for organizations that fuse private and public information and must defend every value they act on. One canonical record per real thing, a cited source and an observation date behind every value, unknowns declared rather than left blank, and any past state of the record reproducible.

Published by Reliable Context · 2026-10-29 · nist ai rmf · nist ai risk management framework · ai rmf govern map measure manage · nist ai rmf playbook · ai rmf data provenance

Original article: https://reliablecontext.com/articles/nist-ai-rmf-govern-map-measure-manage-data-provenance



In this article

1. What the AI RMF is, and what it is not
2. The trustworthiness characteristics
3. Where data provenance appears in the text
4. GOVERN: policies and accountability
5. MAP: context and categorization
6. MEASURE: evaluation and tracking
7. MANAGE: treatment, response and monitoring
8. The evidence matrix
9. A worked example
10. Using the Playbook
11. Sources

What the AI RMF is, and what it is not

The NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0) is a United States government publication, NIST AI 100-1, dated January 2023 and released on 26 January 2023 according to the NIST AI RMF page. It was written, as the document says, "as directed by the National Artificial Intelligence Initiative Act of 2020", to help organizations that design, develop, deploy or use AI systems manage AI risks.

Three properties determine how to read it:

- **Voluntary.** NIST describes the framework as "voluntary, rights-preserving, non-sector-specific, and use-case agnostic" (AI 100-1, Executive Summary). It is not a regulation and carries no penalties of its own. Contracts, procurement rules or sector regulators may refer to it; the framework itself imposes nothing.
- **Outcome-based.** The core is a set of outcomes ("Legal and regulatory requirements involving AI are understood, managed, and documented"), not prescribed controls. NIST states that the actions "do not constitute a checklist, nor are they necessarily an ordered set of steps".
- **Under revision.** As of 28 October 2026 the NIST landing page states that "AI RMF 1.0 is being revised as part of the White House AI Action Plan". The same page lists the Generative AI Profile (NIST AI 600-1, 26 July 2024) and an April 2026 concept note for a profile on trustworthy AI in critical infrastructure. Everything below refers to version 1.0, the published text.

This article reads the framework as an evidence plan. For each function, it asks what records an organization would need to show a reviewer that the outcome was achieved, rather than asserting it. That reading is this article's, not NIST's; NIST does not prescribe records. The focus is on data and its provenance, because that is where many outcomes quietly depend on information that is never kept.

Two definitions used throughout:

- **AI system**, in AI 100-1: "an engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments", adapted from the OECD and ISO/IEC 22989.

- **Data provenance**: the documented origin and history of a piece of data, including where it came from, when, by what process and under whose responsibility. The W3C formalizes this with entities, activities and agents (see the provenance glossary entry).

The trustworthiness characteristics

Part 1 of AI 100-1 (Section 3) names seven characteristics of trustworthy AI systems:

Characteristic	What NIST says about it (paraphrased)	Data question it raises
Valid and reliable	A "necessary condition of trustworthiness", shown as the base of the others; validation is confirmation through objective evidence that requirements for an intended use are fulfilled (ISO 9000:2015)	What evidence shows the system works for this use, on data like the deployment data?
Safe	Should not, under defined conditions, lead to a state endangering human life, health, property or the environment (ISO/IEC TS 5723:2022)	Which inputs and conditions were tested?
Secure and resilient	Withstand adverse events; maintain confidentiality, integrity and availability	Can tampering with training or reference data be detected?
Accountable and transparent	Shown as a vertical box because it "relates to all other characteristics"	Who decided what, when, on which data?
Explainable and interpretable	Representation of mechanisms; meaning of outputs in context	Can an output be traced to its inputs?
Privacy-enhanced	Safeguarding autonomy, identity and dignity through norms and practices	Where did personal data come from, and on what basis?
Fair with harmful bias managed	Equality and equity; NIST names three bias categories: systemic, computational and statistical, and human-cognitive	Is the data representative, and how do we know?

The figure caption in Section 3 matters for data teams. Validity and reliability sit underneath everything else, and validation is defined as confirmation "through the provision of objective evidence". Evidence is a record. A characteristic that cannot be demonstrated from records is, for a reviewer, a claim.

Where data provenance appears in the text

AI 100-1 mentions provenance sparingly, but in a load-bearing place. In Section 3.4, "Accountable and Transparent", the framework states that "maintaining the provenance of training data and supporting attribution of the AI system's decisions to subsets of training data can assist with both transparency and accountability" (AI 100-1, Section 3.4). The same section says transparency's scope "spans from design decisions and training data to model training, the structure of the model, its intended use cases, and how and when deployment, post-deployment, or end user decisions were made and by whom".

Read as requirements for a data layer, those two sentences ask for three things:

1. **Provenance of training data**: for each data set, its source, acquisition date and the process that prepared it.

2. **Attribution**: the ability to connect an output or decision back to the subset of data that influenced it. For a classical model trained once, this is hard and remains a research problem. For a retrieval-based system that answers from documents at run time, it is tractable: the retrieved passages are the subset.
3. **Decision history**: "how and when" decisions "were made and by whom", which is a dated log of human and system actions.

Section 3.3 adds a boundary: resilience "goes beyond the provenance of the data to encompass unexpected or adversarial use". Provenance is necessary for several characteristics and sufficient for none.

The Generative AI Profile is more explicit. NIST AI 600-1 uses the word provenance throughout its suggested actions. Examples include GV-1.6-003, which suggests that inventory entries for generative AI systems include "data provenance information (e.g., source, signatures, versioning, watermarks)"; GV-6.1-008, maintaining "records of changes to content made by third parties to promote content provenance, including sources, timestamps, metadata"; and MP-2.2-001, documenting how the system relies on upstream data sources. AI 600-1 also defines confabulation, its term for hallucination, as "the production of confidently stated but erroneous or false content", one of twelve risks it lists as novel to or exacerbated by generative AI.

GOVERN: policies and accountability

NIST designs GOVERN as "a cross-cutting function to inform and be infused throughout the other three functions" (AI 100-1, Figure 5). It has six categories, from policies and accountability structures (GOVERN 1 and 2) to workforce diversity (GOVERN 3), culture (GOVERN 4), engagement with external AI actors (GOVERN 5) and third-party software, data and supply chain (GOVERN 6).

The subcategories that produce or depend on records are numerous. Selected examples, quoted from the core tables:

- GOVERN 1.1: "Legal and regulatory requirements involving AI are understood, managed, and documented."
- GOVERN 1.5: ongoing monitoring and periodic review of the risk management process are planned, with roles and review frequency defined.
- GOVERN 1.6: "Mechanisms are in place to inventory AI systems".
- GOVERN 1.7: processes for decommissioning and phasing out AI systems safely.
- GOVERN 2.1: roles, responsibilities and lines of communication "are documented".
- GOVERN 4.2: teams "document the risks and potential impacts" of the AI they design, develop, deploy, evaluate and use.
- GOVERN 6.1: policies address risks "associated with third-party entities, including risks of infringement of a third-party's intellectual property".
- GOVERN 6.2: contingency processes for failures or incidents "in third-party data or AI systems deemed to be high-risk".

The evidence implication is that GOVERN is mostly about who is accountable and how the organization knows what it runs. An AI system inventory (GOVERN 1.6) that does not record which data sets each system depends on cannot support GOVERN 6.2: when a third-party data source turns out to be wrong, nobody can

list the affected systems.

MAP: context and categorization

The MAP function "establishes the context to frame risks related to an AI system" (AI 100-1, Section 5.2). NIST notes that AI actors in charge of one part of the lifecycle "often do not have full visibility or control over other parts". MAP is where the organization writes down intended purpose, users, risk tolerance and dependencies.

Data-relevant subcategories include:

- MAP 1.1: intended purposes, context-specific laws, norms and expectations, and deployment settings "are understood and documented".
- MAP 1.5: "Organizational risk tolerances are determined and documented."
- MAP 2.2: information about the system's "knowledge limits and how system output may be utilized and overseen by humans is documented".
- MAP 2.3: scientific integrity and TEVV (test, evaluation, verification and validation) considerations are documented, "including those related to experimental design, data collection and selection (e.g., availability, representativeness, suitability)".
- MAP 4.1: approaches for mapping risks of components, "including the use of third-party data or software", are in place, followed and documented.

MAP 2.2 deserves attention. "Knowledge limits" is a data property: what the system was given to know, as of what date, from which sources. A system that answers questions about a product label cannot document its knowledge limits unless the organization knows which version of the label the system was given and when.

MEASURE: evaluation and tracking

The MEASURE function "employs quantitative, qualitative, or mixed-method tools, techniques, and methodologies to analyze, assess, benchmark, and monitor AI risk and related impacts" (AI 100-1, Section 5.3). NIST says AI systems "should be tested before their deployment and regularly while in operation".

Subcategories that produce evidence directly:

- MEASURE 1.1: approaches and metrics are selected, and "the risks or trustworthiness characteristics that will not – or cannot – be measured are properly documented".
- MEASURE 1.3: internal experts who were not front-line developers, or independent assessors, are involved in regular assessments.
- MEASURE 2.1: "Test sets, metrics, and details about the tools used during TEVV are documented."
- MEASURE 2.3: performance is measured and demonstrated "for conditions similar to deployment setting(s)".
- MEASURE 2.4: functionality and behavior "are monitored when in production".
- MEASURE 2.5: the system "is demonstrated to be valid and reliable", and limits of generalizability are documented.
- MEASURE 3.3: feedback processes let end users and affected communities "report problems and appeal system outcomes".

MEASURE 1.1 contains an unusual requirement: documenting what is not measured. That is the framework's version of declaring an unknown. MEASURE 2.1 and 2.5 together imply that test sets must be versioned and preserved; a validity claim that cannot be re-run against the same test set is not reproducible. For generative systems, faithfulness to sources (whether each claim in an output is supported by the retrieved material) is one measurable property; the site's protocol for calibrating an LLM judge of groundedness covers how to measure it without trusting the judge blindly.

MANAGE: treatment, response and monitoring

The MANAGE function "entails allocating risk resources to mapped and measured risks on a regular basis and as defined by the GOVERN function" (AI 100-1, Section 5.4). Risk treatment "comprises plans to respond to, recover from, and communicate about incidents or events".

Subcategories with record implications:

- MANAGE 1.1: a determination is made "as to whether the AI system achieves its intended purposes and stated objectives and whether its development or deployment should proceed".
- MANAGE 1.3: responses to high-priority risks are "developed, planned, and documented"; options include mitigating, transferring, avoiding or accepting.
- MANAGE 1.4: negative residual risks to downstream acquirers and end users "are documented".
- MANAGE 2.4: mechanisms exist "to supersede, disengage, or deactivate AI systems" whose outcomes are inconsistent with intended use.
- MANAGE 3.2: pre-trained models used for development are monitored as part of regular monitoring and maintenance.
- MANAGE 4.1: post-deployment monitoring plans include "appeal and override, decommissioning, incident response, recovery, and change management".
- MANAGE 4.3: processes for tracking, responding to and recovering from incidents and errors "are followed and documented".

MANAGE 1.1 is a go or no-go decision. Months later, the question a reviewer asks is not whether the decision was made but on what evidence: which evaluation results, on which data version, approved by whom, on which date.

The evidence matrix

The table below translates selected outcomes into records a data layer would need to produce. It is this article's reading of the framework, not a NIST requirement, and it is deliberately limited to data and provenance.

Function	Outcome (subcategory)	Record that demonstrates it	Fails when
GOVERN	1.6 inventory of AI systems	System register listing, per system, the data sets and external sources it depends on, with versions	A data source changes and affected systems cannot be listed

Function	Outcome (subcategory)	Record that demonstrates it	Fails when
GOVERN	2.1 documented roles	Named owner per system and per data set, with effective dates	Ownership is a team alias with no history
GOVERN	6.1, 6.2 third-party data	Source register: supplier, licence, date obtained, contingency if withdrawn	Third-party data is copied in with no acquisition record
MAP	1.1 intended purpose and context	Dated purpose statement tied to a system version	The purpose changed but the statement did not
MAP	2.2 knowledge limits	Snapshot of what the system can draw on (corpus version, cut-off date)	Answers rely on documents with unknown dates
MAP	2.3 data collection and selection	Data set datasheet: origin, selection criteria, representativeness checks	Training data provenance is "internal sources"
MEASURE	1.1 what is not measured	Explicit list of unmeasured risks, dated and owned	Silence is read as "no risk"
MEASURE	2.1, 2.5 test sets and validity	Versioned test sets with hashes; results linked to model and data versions	The test set was edited after the result
MEASURE	2.4 production monitoring	Per-output log: input, retrieved sources, output, time, system version	Outputs cannot be traced to sources
MEASURE	3.3 problem reports and appeals	Ticket linked to the specific output and its sources	Complaints refer to "the chatbot" generically
MANAGE	1.1 go or no-go	Decision record citing evaluation results and approver	The decision exists only in meeting notes
MANAGE	2.4, 4.1 deactivate, change management	Change log of model, prompt, corpus and configuration versions	A rollback cannot reproduce the prior state
MANAGE	4.3 incidents and errors	Incident record with affected outputs, root-cause data, correction date	Corrections overwrite the erroneous value without trace

Three patterns run through the table. First, almost every record needs a date and a version, because the framework's outcomes are about states that change. Second, records must link: an output to its sources, a source to its supplier, a decision to its evidence. Third, corrections must be additive. If fixing an error destroys the evidence of what the system said before, MANAGE 4.3 and MEASURE 3.3 cannot be demonstrated.

A worked example

Halden Therapeutics AG is a fictional pharmaceutical company used for illustration. It deploys a retrieval-based assistant that drafts answers to medical-information enquiries from healthcare professionals, using approved product labels and internal standard response documents.

A healthcare professional reports that the assistant gave an outdated maximum daily dose. The organization works through the framework:

Step	Framework outcome	What Halden needs to retrieve
1. Log the report	MEASURE 3.3	The exact output, timestamp and session
2. Find the source	MEASURE 2.4	The passages retrieved for that output, with document IDs and versions
3. Diagnose	MAP 2.2	Whether the corpus held the current label on that date, or only the superseded one
4. Assess reach	GOVERN 1.6	Which other systems use the same document set
5. Treat	MANAGE 1.3, 2.4	Decision record: correct corpus, suppress affected answers, or deactivate
6. Close	MANAGE 4.3	Incident record listing every affected output and the correction date

Steps 2 and 3 decide the outcome. If the retrieval log shows the superseded label was retrieved because the current one was never ingested, the fault is in the source register (GOVERN 6.1 and MAP 2.3), not in the model. If the current label was retrieved and the model still produced the old dose, the fault is in generation, and MEASURE 2.5 validity claims need revisiting. Without per-output source records, the two cases are indistinguishable, and the organization can only retrain or re-prompt and hope.

Using the Playbook

NIST publishes a companion AI RMF Playbook that "provides suggested actions for achieving the outcomes laid out in the AI Risk Management Framework", aligned to each subcategory. The Playbook page states that it "is neither a checklist nor set of steps to be followed in its entirety" and that its suggestions are voluntary. It is available as PDF, CSV, Excel and JSON, and NIST says updates are released approximately twice per year and that the Playbook will be updated after the AI RMF is revised.

Practical use: export the Playbook in CSV or JSON, filter to the subcategories in scope, and add a column for the record that will evidence each one. That turns the Playbook into a control matrix whose rows can be checked.

AI 100-1 also defines **profiles**: use-case profiles implement the functions, categories and subcategories for a specific setting ("for example, an AI RMF hiring profile"), and temporal profiles describe a Current Profile and a Target Profile so that the gap can be managed (AI 100-1, Section 6). The Generative AI Profile, AI 600-1, is the main published profile; its suggested actions are where most of the explicit provenance language lives.

Reliable Context is a system of record in which each value carries a cited, dated source, which is one way to produce the provenance records this framework's outcomes depend on.

Sources

1. National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1, January 2023. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf
2. National Institute of Standards and Technology, "AI Risk Management Framework" (landing page), accessed 28 October 2026. https://www.nist.gov/itl/ai-risk-management-framework

3. National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*, NIST AI 600-1, July 2024. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf
4. National Institute of Standards and Technology, Trustworthy and Responsible AI Resource Center, "NIST AI RMF Playbook", accessed 28 October 2026. https://airc.nist.gov/airmf-resources/playbook/

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf
2. https://www.nist.gov/itl/ai-risk-management-framework
3. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf
4. https://airc.nist.gov/airmf-resources/playbook

THE PUBLISHER

About Reliable Context

Reliable Context is a system of record for organizations that fuse private and public information and must defend every value they act on.

The guarantees

One canonical record per real thing: duplicates resolve by rule, ambiguity is shown rather than hidden, and related organizations relate without ever being merged. A verbatim, cited source and an observation date sit behind every value. Unknowns are declared as not found, blocked or not applicable with a reason, never left as an empty cell. Any past state of the record is reproducible, together with what changed since.

Where it runs

It runs inside the customer's perimeter, one database per organization. Permitted sources, model routing, reader clearance and retention are set as the customer's own policies, and every value is exportable with its chain of custody for an auditor or regulator.

What it is not

It is not a vector database or a RAG framework, not a data catalog, MDM or data-quality suite, not a CRM or data-enrichment vendor, and not a chatbot or AI assistant. Where a reader needs one of those, the honest answer is to say so and point them there.

Visit

See Reliable Context.

Reliable Context is in production use in two sectors; more than six thousand of its identity decisions were made by rule and none has been reversed. Articles published here explain general practice, cite their primary and academic sources, and name the jurisdiction of any regulatory rule they describe. They are not legal, regulatory or compliance advice for a specific organization, and requirements change.

Website: https://reliablecontext.com

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.