



WHITE PAPER · DATA GOVERNANCE

Continuous, Risk-Adaptive Data Governance

Continuous risk scoring, active metadata and policy as code

White Paper 2 · Data Governance series · Version 1.0 · October 2026

Bassirah Technology — Innovation & R&D

Continuous, Risk-Adaptive Data Governance

White Paper 2 · Data Governance series · Bassirah Technology — Innovation & R&D · 28 October 2026

Document control

Field	Value
Series	Data Governance, paper 2 of 6
Author	Bassirah Technology — Innovation & R&D
Version	1.0 · 28 October 2026
Scope	A technology and regulatory survey. It describes published research, official legal texts, standards and publicly documented systems; it does not report Bassirah benchmarks or customer results, and it is not legal advice.

Contents

1. Abstract	4
2. Introduction	4
2.1 Why periodic governance falls behind	4
2.2 About the Data Governance series	5
2.3 Scope	5
2.4 Terms used in this paper	5
3. What "continuous" means	5
4. The regulatory and standards baseline	6
4.1 European law	6
4.2 Standards	6
4.3 Frameworks and supervisory principles	7
5. A risk-adaptive model	7
5.1 From static rules to risk scores	7
5.2 Risk signals for data	8
5.3 Proportionality	8
6. Active metadata	8
6.1 Catalogs that act	8
6.2 Machine-readable records	9
6.3 Maturity models	9
7. Automated classification	9
8. Policy as code	9
8.1 Decision and enforcement	9
8.2 Engines and languages	10
8.3 Enforcement in data platforms	10
8.4 Verifying policies	10
8.5 Compliance as code	10

9. Purpose and information flow	11
10. Continuous data quality controls	11
11. The operating model.....	11
12. Comparative analysis.....	12
13. Open challenges and future opportunities.....	12
13.1 Measuring risk.....	12
13.2 Trustworthy labels	12
13.3 Policy correctness at scale	13
13.4 From rules to regulation	13
13.5 Purpose and flow at pipeline scale	13
13.6 Consistency across engines.....	13
13.7 Evidence that continuous governance works	13
13.8 People in the loop.....	13
14. What this means for platform decisions	13
References	14

1. Abstract

Most data governance programs still work in cycles: an annual policy review, a quarterly access review, a data protection impact assessment written once before launch. Data, people and risk change much faster than that. This paper examines the move from periodic governance to continuous, risk-adaptive governance, in which controls are expressed as code, evaluated on every relevant change or request, and tightened or relaxed as risk signals change. It sets out what regulation and standards already expect as of October 2026, including the GDPR, the EU AI Act, DORA, NIS2, the BCBS 239 principles, ISO 31000, ISO/IEC 38505-1, ISO/IEC 27001 and 27701, and the NIST Cybersecurity and Privacy Frameworks. It then describes the building blocks: risk-adaptive frameworks and risk-based access control, active metadata, automated classification, policy engines and languages, enforcement in data platforms, formal verification of policies, compliance as code, purpose and information-flow controls, and continuous data quality checks. It is explicit about the limits, including the probabilistic nature of classification and the immaturity of risk scoring. It compares periodic and continuous approaches by control area, sets out open challenges and closes with notes for platform decisions.

2. Introduction

2.1 Why periodic governance falls behind

Periodic control assumes that the world between two reviews is stable enough for a snapshot to stay true. For data that assumption fails often. New tables arrive daily, pipelines change with each deployment, people change roles, and a data set that was low risk in one use becomes high risk when it is joined to another or used to train a model. Research on auditing makes the same point: traditional periodic or cyclical audits struggle with systems that learn and adapt, and continuous auditing, discussed since the 1980s, is being extended to AI for that reason [1].

Experience in banking shows the cost of governance that is not sustained. The Basel Committee's principles for risk data aggregation, published after the 2007 to 2009 financial crisis exposed banks' inability to aggregate risk data quickly and accurately [2], were followed about a decade later by a progress report finding that the largest banks were at varying stages and all needed further work to reach or keep full compliance [3]. The report asks boards to take full responsibility for overseeing robust data governance and supervisors to act more firmly on long-standing weaknesses [3]. Reviews of data governance in the age of AI and connected devices find persistent challenges, including fragmented frameworks, limited scalability and data quality problems [4].

The alternative is not more frequent audits. It is to put controls where data changes: express policies in machine-readable form, evaluate them automatically when data, metadata or requests change, and use risk signals to decide how strict a control should be. Frameworks proposed for this, such as the dynamic risk-adaptive data governance framework DRADG, combine static policy with sensitivity classification, contextual risk scoring and feedback loops, map lifecycle stages to mitigations in a governance-risk matrix and adjust controls through a rules engine [5]. The evidence for such frameworks is so far conceptual or simulation-based [5], which this paper treats as a reason for care rather than for dismissal.

2.2 About the Data Governance series

This is the second of six papers. Paper 1 followed data through the AI lifecycle and identified the controls each stage needs. This paper describes how to run controls continuously and in proportion to risk. Paper 3 covers data products and data contracts, Paper 4 data sovereignty and cross-border compliance, Paper 5 privacy-enhancing computation and Paper 6 data sharing ecosystems.

2.3 Scope

This paper covers the regulatory and standards basis for continuous, risk-based governance; risk-adaptive models and risk-based access control; active metadata and automated classification; policy as code, its engines, enforcement points and verification; compliance as code; purpose and information-flow controls; and continuous data quality checks. It concentrates on official legal texts, standards, peer-reviewed research and official documentation, cites foundational work where it introduced an idea still in use, and labels preprints and company engineering reports as such. Statements about laws, regulations and standards are general information as of October 2026. They are not legal advice: obligations depend on the facts of each case, on national implementation and on guidance that continues to change, and organizations should take qualified legal advice. Statements about products reflect documentation accessed in October 2026. The paper does not quote performance, accuracy or adoption figures from its sources.

2.4 Terms used in this paper

A **control** is a measure that keeps a risk within tolerance. **Continuous** control means assessing and enforcing it at a frequency sufficient to support risk-based decisions, which may be per request, per change or on a schedule, rather than literally all the time [6]. **Risk-adaptive** governance changes the strength of a control as risk signals change. **Active metadata** is metadata that triggers governance actions when it changes, rather than only describing assets. **Policy as code** means policies written in a machine-readable language, versioned and tested like software, and evaluated by an engine. A **policy decision point** evaluates a policy; a **policy enforcement point** applies the decision [7]. **Compliance as code** applies the same approach to evidence that controls are in place.

3. What "continuous" means

"Continuous" is easily overstated. NIST's control for continuous monitoring requires a monitoring strategy with defined metrics, frequencies of monitoring and assessment, ongoing control assessment, analysis of results, response actions and reporting, and explains that continuous and ongoing mean assessing controls and risks at a frequency sufficient to support risk-based decisions, with different controls monitored at different frequencies [6]. Its predecessor guidance defined information security continuous monitoring as maintaining ongoing awareness of assets, threats, vulnerabilities and control effectiveness, so that controls stay aligned with risk tolerance [8].

Regulation follows the same pattern of a periodic floor with continuous signals on top. DORA requires financial entities to review their ICT risk management framework at least once a year and after major incidents, to improve it continuously, to review the classification of information assets at least yearly, to identify sources of ICT risk on a continuous basis, and to monitor and

control ICT systems continuously [9]. The NIST Cybersecurity Framework 2.0 states that risk management must be a continuous process and that its functions, including the new Govern function, should be addressed concurrently [10]. ISO 31000, the general risk management guideline, describes risk management as an iterative process of establishing context, assessing and treating risk, and monitoring and review [11].

For data governance, three practical consequences follow. First, frequency should follow risk: a column holding health data may need checks on every access, while a reference table may need a weekly scan. Second, triggers matter as much as schedules: a new data set, a schema change, a change of purpose or a new consumer should start an evaluation. Third, periodic reviews remain, but they change role: instead of discovering the state of controls, they review the evidence that continuous controls have produced.

4. The regulatory and standards baseline

4.1 European law

The GDPR is risk-based and ongoing in its own words. Article 24 requires controllers to implement measures that take account of risks of varying likelihood and severity and to review and update them where necessary [12]. Article 25 applies data protection by design both when the means of processing are determined and at the time of the processing itself [12]. Article 32 requires a process for regularly testing, assessing and evaluating the effectiveness of security measures, and Article 35 requires a review of a data protection impact assessment at least when the risk changes [12]. A one-off assessment that is not revisited when processing changes does not match this text.

The EU AI Act describes the risk management system for high-risk AI systems as a continuous iterative process planned and run throughout the lifecycle, with regular systematic review and updating [13]. Following the Digital Omnibus on AI, Regulation (EU) 2026/1744, the high-risk requirements apply from 2 December 2027 for systems listed in Annex III and from 2 August 2028 for systems covered by Annex I [13]. DORA has applied since 17 January 2025 and contains the continuous duties described in Section 3 [9]. The NIS2 Directive requires management bodies to approve and oversee cybersecurity risk-management measures and makes them accountable, and requires appropriate and proportionate measures, including policies and procedures to assess their effectiveness [14]. NIS2 does not itself prescribe continuous monitoring [14], but continuous evidence is one practical way to show that measures are effective.

4.2 Standards

ISO 31000:2018 provides the general risk management process and is under revision [11]. ISO/IEC 38505-1, revised in 2026, applies the governance principles of ISO/IEC 38500 to data and treats data governance as a responsibility of the governing body within organizational governance [15]. ISO/IEC 27001:2022, amended in 2024, requires an information security management system with risk assessment and treatment, monitoring and continual improvement [16]. ISO/IEC 27701 was revised in 2025 as a standalone privacy information management system standard, no longer only an extension of ISO/IEC 27001 [17]. ISO/IEC 23894 applies the ISO 31000 process to AI risk [18].

4.3 Frameworks and supervisory principles

The NIST Cybersecurity Framework 2.0 added Govern as a sixth function covering organizational context, risk management strategy, roles and responsibilities, policy, oversight and supply chain risk, and includes continuous monitoring among its detection outcomes [10]. The NIST Privacy Framework 1.0 treats privacy risk as part of enterprise risk management; a version 1.1 aligned with CSF 2.0 remains a draft [19]. NIST's zero trust architecture requires access to be decided by dynamic policy using the observable state of the requester and environment, describes authentication and authorization as a constant cycle of reassessment, and uses telemetry to improve policy [20]. Its guide to attribute-based access control describes decisions computed from attributes of the subject, object, operation and environment [21]. In banking, the BCBS 239 principles remain the supervisory reference for risk data governance [2], [3].

Table 1 summarizes how each instrument bears on continuous, risk-adaptive governance. It is a simplified orientation, not a compliance checklist.

Instrument	Status as of October 2026	Relevance to continuous governance
GDPR, Arts. 24, 25, 32, 35	Applies	Risk-scaled measures, reviewed when necessary; regular testing; impact assessments reviewed when risk changes
EU AI Act, Art. 9	High-risk requirements from 2 December 2027 (Annex III) or 2 August 2028 (Annex I)	Risk management as a continuous iterative process across the lifecycle
DORA	Applies since 17 January 2025	Yearly reviews plus continuous identification, monitoring and improvement
NIS2	Transposition deadline 17 October 2024	Board accountability; proportionate measures; assessment of effectiveness
BCBS 239	Supervisory principles since 2013	Risk data aggregation and governance as a sustained capability
ISO 31000:2018	Current; under revision	Iterative risk process
ISO/IEC 38505-1:2026	Current	Data governance as a governing-body responsibility
ISO/IEC 27001:2022 and 27701:2025	Current	Management systems with monitoring and continual improvement
NIST CSF 2.0; Privacy Framework 1.0	Current (Privacy Framework 1.1 in draft)	Govern function; continuous risk management
NIST SP 800-53 CA-7; SP 800-137	Current	Continuous monitoring at risk-appropriate frequencies
NIST SP 800-207; SP 800-162	Current	Dynamic, attribute-based access decisions

5. A risk-adaptive model

5.1 From static rules to risk scores

Most deployed access control models apply static policies that return the same result regardless of circumstances. Risk-based access control instead estimates a risk value for each request from contextual and real-time information and uses it to decide whether, and under which conditions, to grant access [22]. Zero trust adopts the same idea for networks and services [20], and surveys of zero trust list the computation of trust and risk as an open

challenge [23]. For data, DRADG proposes combining static policy with data sensitivity classification and contextual risk scoring, adjusting controls when thresholds are crossed and feeding outcomes back into the model [5].

5.2 Risk signals for data

A risk-adaptive control needs signals it can evaluate automatically. Table 2 lists the signals most often available in data platforms, where they come from and how they can change a control.

Signal	Typical source	Example of adaptation
Sensitivity of the data	Classification tags and glossary terms	Mask or filter sensitive columns; require approval for export
Purpose of use	Purpose tags on data and declared purpose of the job or request	Allow analytics but block marketing use of the same data
Requester and context	Identity, role, group, location, device, time	Deny access from unmanaged devices; step up authentication
Lineage and joins	Lineage events and query plans	Raise risk when sensitive data is joined with identifiers
Data quality and freshness	Quality checks and contracts	Block publication when checks fail
Volume and behavior	Query logs and access telemetry	Alert or throttle on unusual extraction
Regulatory context	Residency, consent and retention metadata	Restrict processing to permitted regions and periods

The signals are only as good as the metadata behind them. Section 6 describes how metadata is kept current, and Section 7 how sensitivity labels are produced.

5.3 Proportionality

Risk adaptation is a matter of proportion as much as automation. Controls that are too strict push people toward copies and workarounds, which create new risk; controls that are too loose fail the purpose. A practical design keeps a small set of risk tiers, each with defined controls and monitoring frequency, maps every data asset and use to a tier from its signals, and records why. Human judgment remains responsible for the tiers, the thresholds and the exceptions.

6. Active metadata

6.1 Catalogs that act

Data catalogs give an overview of data assets and their characteristics, and they deliver most value when integrated with data sources and the other tools of data management [24]. A peer-reviewed description of Microsoft Purview, written by its developers, illustrates the problem catalogs are now asked to solve: data spread across on-premises, edge and multi-cloud stores leaves policies applied piecemeal per source in source-specific languages, and a central catalog populated by automated scanning, with managed classification and attribute-based policies, is proposed to enforce them consistently [25]. It is a vendor architecture, not an independent evaluation, but it describes the pattern clearly.

Active metadata turns the catalog from a record into a trigger. The open-source DataHub Actions Framework reacts in real time to change events in the metadata graph through configured pipelines of sources, filters, transformers and actions, so that a new tag, owner

change or schema change can start governance work [26]. DataHub also documents Metadata Tests, rules that select assets, test them against conditions on owners, domains, terms and tags, evaluate on a schedule and can apply tags or terms on the result [27]; the documentation points to DataHub Cloud for their operation [27]. OpenLineage emits lineage as events while jobs run, which supplies the lineage signal in Table 2 [28].

6.2 Machine-readable records

Continuous governance also needs governance records in machine-readable form. Records of processing activities under the GDPR are often kept in spreadsheets or proprietary systems, which blocks automation; DPCat represents them as machine-readable catalogs built on open vocabularies, validated with shape constraints and queried to produce documentation that stays current [29]. The same approach applies to policies, risk registers and control evidence (Section 8.5).

6.3 Maturity models

Maturity models for data governance are well cataloged [30], and new ones continue to appear for big data and AI [4]. They remain useful for planning, but they are periodic assessments by design. In a continuous program, a maturity model can be scored partly from the evidence that controls produce, rather than from interviews alone.

7. Automated classification

Sensitivity labels drive most risk-adaptive controls, and labeling by hand does not keep up with new data. Automated classification fills the gap, with limits that need to be stated plainly. OpenMetadata documents an auto-classification workflow that tags sensitive columns using regular expressions on column names and named-entity recognition on sampled values, with a confidence threshold that decides when to apply a tag [31]. Presidio, a toolkit that began at Microsoft and is now a community-owned project, detects and de-identifies sensitive data in text, images and structured data, and its documentation states that it may not find all sensitive information and should be used alongside other safeguards [32]. DataHub's classification framework remains, but its built-in classifier was removed in version 1.8.0, so users register their own [26].

Research explains why classification of structured data is still maturing. Most work on sensitive data detection has focused on unstructured text; approaches for tables often classify cells independently, and there is no recognized benchmark data set for the task [33]. Column context improves detection [33], and a 2025 preprint argues that sensitivity depends on context beyond the data type, such as the surrounding data set and domain rules, and that explanations grounded in context help human reviewers [34]. The practical conclusion is to treat automated classification as a proposal: apply it at scale, route uncertain and high-impact labels to people for review, and measure how often reviewers change them.

8. Policy as code

8.1 Decision and enforcement

The XACML 3.0 standard set out the reference architecture that most policy systems still follow: a policy administration point where policies are written, a policy decision point that evaluates

them, a policy enforcement point that requests and applies decisions, and a policy information point that supplies attribute values [7]. Attribute-based access control supplies the decision model, computing access from attributes of the subject, object, operation and environment [21]. Separating the decision from its enforcement is what allows one policy to apply across many systems.

8.2 Engines and languages

Open Policy Agent is a general-purpose, open-source policy engine that decouples decisions from enforcement: services send structured input and receive a decision, and policies are written in Rego, a declarative language for querying hierarchical data [35]. It became a graduated project of the Cloud Native Computing Foundation in 2021 [35]. Cedar is an authorization language designed to be expressive and analyzable; its peer-reviewed description shows that it supports role-, attribute- and relationship-based patterns, includes a validator that catches policy errors, and was formally modeled so that properties of policies can be proved [36]. Relationship-based access control, introduced at scale by Google's Zanzibar system for consistent global authorization [37], is available in open source through OpenFGA, a Cloud Native Computing Foundation incubating project whose models can be tested in continuous integration [38].

8.3 Enforcement in data platforms

The pattern that matters most for data is "classify once, enforce everywhere": a sensitivity tag applied to a column drives one policy wherever the column appears. Apache Ranger documents tag-based policies, with tags synchronized from Apache Atlas, evaluated before resource-based policies, and with a deny in a tag policy taking precedence [39]. Snowflake documents masking policies attached to tags, so that every column carrying the tag is masked, including columns added later [40]. Starburst Galaxy documents attribute-based policies that grant privileges, apply row filters or apply column masks to objects selected by tags, with tags inherited from parent objects and a deny overriding an allow [41]. Trino can delegate authorization to Open Policy Agent, which can return row filters and column masks as well as allow or deny decisions [42]. These features make the risk signals of Table 2 enforceable at query time, provided the tags are accurate (Section 7).

8.4 Verifying policies

Policies written as code can be checked by machines, which is what makes continuous control practical at scale. Zelkova encodes the semantics of cloud access policies for an automated reasoning engine, allowing policies to be compared and properties verified, and has been used in production [43]. Cedar was designed with such analysis in mind, for example to prove that a refactored set of policies grants the same permissions as before [36]. Research on network security compliance shows that continuous verification is feasible when the policy language is formal and the verification is fast [44]. Verification of this kind answers a question periodic review struggles with: whether a change to policy has an effect nobody intended.

8.5 Compliance as code

Compliance as code applies the approach to evidence. NIST's Open Security Controls Assessment Language provides machine-readable formats for control catalogs, baselines, mappings, system security plans and assessment results, and NIST presents it as enabling automated assessment and monitoring of control effectiveness [45]. Research describes

architectures that replace manual, document-based compliance management with automated, continuous processes in regulated hybrid clouds [46], and argues for designing systems around explicit, integrated security policies [47]. The term continuous compliance appears in peer-reviewed work on regulated software development from 2013 [48], while a 2026 preprint still describes compliance as code as an emerging idea with limited research and finds that mapping technical rules to regulatory requirements is hard to do consistently [49]. Infrastructure practice offers a useful template: Kubernetes Gatekeeper enforces policies when resources are created or changed and also audits resources that already exist [50]. Data governance needs both halves: preventive checks at change time and a continuous detective audit of the current state.

9. Purpose and information flow

Access control decides who may read data. Many obligations concern why data is used and where it flows. Purpose-based access control attaches intended purposes to data and checks the purpose of each access against them [51], [52]. Two systems show purpose and flow constraints enforced at run time. Riverbed, a peer-reviewed research system, lets users express constraints in a simple policy language and enforces them through information-flow control in a managed runtime, without manual labeling of code [53]. Meta has described, in a company engineering report rather than a peer-reviewed paper, an information-flow control system that labels data assets with permitted purposes, checks flows at run time across services and batch systems, and rolls out by logging first, then remediating, then enforcing [54]. Its stated reason is relevant to this paper: point checks and data set permissions require repeated exhaustive audits as code changes, and lineage detects flows only after they happen [54]. Purpose and flow enforcement at this scale remains rare and is an open problem for most organizations (Section 12).

10. Continuous data quality controls

Quality is a governance control, and it lends itself to continuous checking. Declarative constraints, in effect unit tests for data, can run as part of pipelines and incrementally as data grows, with anomaly detection on the history of quality metrics [55]. Recurring production pipelines accumulate quality problems from upstream changes, and engineers spend substantial effort monitoring them; research shows that suitable constraints can be programmed automatically from a pipeline's own history [56]. In a risk-adaptive program, quality results become signals: a failing check can hold back publication of a data set, lower the trust level shown in the catalog or alert the owner. Paper 3 develops this through data contracts.

11. The operating model

Continuous control changes how people work, not only which tools run. Governance remains a responsibility of the governing body [15]: NIS2 makes management bodies approve and oversee risk measures [14], CSF 2.0 places governance at the center of its framework [10], and the Basel Committee asks bank boards to take full responsibility for data governance [3]. What changes is the evidence they see and how quickly.

Four practices help. First, write policies once, in code, with an owner, tests and a change history, and derive documents from them rather than the reverse [45], [29]. Second, roll out new controls in stages, logging violations before enforcing them, as in the flow-control rollout described above [54] and the dry-run and audit modes of infrastructure policy engines [50]. Third, keep a human review loop for classification, exceptions and threshold changes [32], [34], and record every exception with its owner and expiry. Fourth, treat governance telemetry as data: decisions, denials, exceptions, reviewer corrections and control failures are the inputs to the next risk assessment, as zero trust uses telemetry to improve policy [20].

12. Comparative analysis

Table 3 compares periodic and continuous approaches by control area and gives a qualitative maturity judgment drawn from the sources cited. It describes the state of practice, not any single product.

Control area	Periodic approach	Continuous, risk-adaptive approach	Maturity
Access control	Quarterly access reviews of role assignments	Per-request decisions from attributes, tags and context	Mature engines; risk scoring still open
Sensitivity labeling	Manual inventory campaigns	Automated proposals with human review on change	Maturing; accuracy limits
Policy management	Documents approved annually	Policies as code with tests, versioning and verification	Mature engines; verification maturing outside cloud IAM
Purpose limitation	Assessments before launch	Purpose tags and runtime flow checks	Open problem outside a few large operators
Data quality	Periodic profiling and reports	Checks in pipelines with anomaly detection	Maturing; uneven adoption
Compliance evidence	Collected for audits	Produced by controls in machine-readable form	Maturing; regulatory mappings hard
Risk assessment	Annual risk register update	Triggered reassessment on change, with periodic review of evidence	Conceptual frameworks; limited field evidence

The engines for policy as code are mature and widely available. The weak points are upstream and downstream of them: the accuracy of the metadata that feeds decisions, the scoring of risk, and the mapping between technical controls and legal requirements.

13. Open challenges and future opportunities

13.1 Measuring risk

Challenge. Risk-based access control depends on risk estimates, and methods for computing them are not standardized [22], [23]. **Research direction.** Risk factors and estimation techniques validated in real deployments rather than simulation [5]. **Opportunity.** Shared, explainable risk models would let organizations adapt controls without opaque decisions.

13.2 Trustworthy labels

Challenge. Automated classification of structured data has accuracy limits and no standard benchmark [32], [33]. **Research direction.** Context-aware detection and explanations for

reviewers [33], [34]. **Opportunity.** Benchmarks and measured review loops would make label quality visible and improvable.

13.3 Policy correctness at scale

Challenge. Large policy sets interact in ways people cannot review by hand. **Research direction.** Analyzable policy languages and automated reasoning [36], [43], [44]. **Opportunity.** Routine verification of every policy change would catch unintended access before it reaches production.

13.4 From rules to regulation

Challenge. Mapping technical controls to legal requirements is hard to do consistently [49]. **Research direction.** Machine-readable control catalogs and mappings [45] and machine-readable records of processing [29]. **Opportunity.** Shared mappings would let one control produce evidence for several regimes.

13.5 Purpose and flow at pipeline scale

Challenge. Purpose and flow constraints are enforced at run time only in a few large organizations [54]. **Research direction.** Information-flow control that works with existing code and engines [53]. **Opportunity.** Practical flow control would make purpose limitation enforceable rather than documented.

13.6 Consistency across engines

Challenge. Policies are still applied per source in source-specific languages [25]. **Research direction.** Central decision points and tag-driven enforcement across engines [35], [42]. **Opportunity.** One policy model for many engines would reduce gaps and duplicated effort.

13.7 Evidence that continuous governance works

Challenge. Risk-adaptive frameworks are mostly conceptual or simulated, and maturity models are themselves periodic [5], [30], [4]. **Research direction.** Field studies of continuous programs and continuous auditing of adaptive systems [1]. **Opportunity.** Evidence of which controls reduce risk would focus investment.

13.8 People in the loop

Challenge. Automation shifts work to exceptions, reviews and threshold decisions, and the division of labor between people and machines needs design [1]. **Research direction.** Staged rollout, review interfaces and measurement of reviewer workload [54], [34]. **Opportunity.** Well-designed review loops would keep accountability with people while automation handles volume.

14. What this means for platform decisions

Start by defining a small number of risk tiers with their controls and monitoring frequencies, and map regulatory duties to them using Table 1, with legal advice. Make classification and purpose tags the inputs to policy: propose labels automatically, review uncertain and high-impact ones, and measure corrections. Choose a policy engine and language that can be tested and analyzed, keep policies in version control, and verify changes before they are deployed. Enforce tag-driven policies at the points where data is read, including query engines and warehouses, rather than in each application. Use active metadata so that new

assets, schema changes and ownership changes trigger checks. Put quality checks in pipelines and treat their results as risk signals. Roll out new controls in log-only mode first, and produce compliance evidence from the controls in machine-readable form. Keep periodic reviews, and use them to examine evidence and decide thresholds and exceptions.

For federated query and access control, Bassirah works with Starburst as a Strategic Technology Partner; Starburst Galaxy documents tag-based attribute policies with row filters and column masks, and features should be checked against the specific Starburst edition in use, since the documentation describes them for Galaxy. For metadata and active metadata, Bassirah provides independent support for DataHub Core, the open-source edition of DataHub; this support is Bassirah's own and is not vendor support from DataHub's maintainers. The Actions Framework is open source, while Metadata Tests are documented for DataHub Cloud, and the open-source classification framework now requires a classifier supplied by the user. Other products, projects and standards named in this paper are ecosystem references. In each case, validate the design against your own data, risks, legal advice and versions before relying on it.

Paper 3 turns to data products and data contracts, which give risk-adaptive controls a stable unit to attach to: a product with an owner, a purpose and stated quality expectations.

References

1. Minkinen, M., Laine, J. and Mäntymäki, M. (2022). Continuous Auditing of Artificial Intelligence: a Conceptualization and Assessment of Tools and Frameworks. Digital Society, 1(3). (<https://doi.org/10.1007/s44206-022-00022-2>)
2. Basel Committee on Banking Supervision (2013). Principles for effective risk data aggregation and risk reporting (BCBS 239). Bank for International Settlements, Basel, January 2013. (<https://www.bis.org/publ/bcbs239.htm>)
3. Basel Committee on Banking Supervision (2023). Progress in adopting the Principles for effective risk data aggregation and risk reporting. Bank for International Settlements, Basel, 28 November 2023 (BCBS d559). (<https://www.bis.org/bcbs/publ/d559.htm>)
4. Bena, Y.A., Ibrahim, R.B., Mahmood, J., Al-Dhaqm, A., Alshammari, A., Nasser, M., Yusuf, M.N. and Ayemowa, M.O. (2025). Big Data Governance Challenges Arising From Data Generated by Intelligent Systems Technologies: A Systematic Literature Review. IEEE Access, 13, 12859–12888. (<https://doi.org/10.1109/ACCESS.2025.3528941>)
5. Gharib, J. and Gahi, Y. (2026). DRADG: A Dynamic Risk-Adaptive Data Governance Framework for Modern Digital Ecosystems. Information, 17(1), 102. (<https://doi.org/10.3390/info17010102>)
6. Joint Task Force (2020). Security and Privacy Controls for Information Systems and Organizations. NIST Special Publication 800-53, Revision 5 (Release 5.2.0), control CA-7 Continuous Monitoring. (<https://doi.org/10.6028/NIST.SP.800-53r5>)
7. OASIS (Rissanen, E., ed.) (2013). eXtensible Access Control Markup Language (XACML) Version 3.0. OASIS Standard, 22 January 2013. (<https://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html>)
8. Dempsey, K., Chawla, N.S., Johnson, L.A., Johnston, R., Jones, A.C., Orebaugh, A.D., Scholl, M.A. and Stine, K.M. (2011). Information Security Continuous Monitoring (ISCM) for Federal Information Systems and Organizations. NIST Special Publication 800-137. (<https://doi.org/10.6028/NIST.SP.800-137>)

9. European Parliament and Council of the European Union (2022). Regulation (EU) 2022/2554 on digital operational resilience for the financial sector (DORA), Articles 6, 8, 9 and 13. Official Journal of the European Union, L 333, 27.12.2022, pp. 1–79. (<https://eur-lex.europa.eu/eli/reg/2022/2554/oj>)
10. National Institute of Standards and Technology (2024). The NIST Cybersecurity Framework (CSF) 2.0. NIST Cybersecurity White Paper NIST CSWP 29, 26 February 2024. (<https://doi.org/10.6028/NIST.CSWP.29>)
11. International Organization for Standardization (2018). ISO 31000:2018 Risk management — Guidelines. ISO, Geneva (ISO/TC 262), 2nd edition. (<https://www.iso.org/standard/65694.html>)
12. European Parliament and Council of the European Union (2016). Regulation (EU) 2016/679 (General Data Protection Regulation), Articles 24, 25, 32 and 35. Official Journal of the European Union, L 119, 4.5.2016, pp. 1–88. (<https://eur-lex.europa.eu/eli/reg/2016/679/oj>)
13. European Parliament and Council of the European Union (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), Article 9, as amended by Regulation (EU) 2026/1744. Official Journal of the European Union, L series, 12.7.2024. (<https://eur-lex.europa.eu/eli/reg/2024/1689/oj>)
14. European Parliament and Council of the European Union (2022). Directive (EU) 2022/2555 on measures for a high common level of cybersecurity across the Union (NIS 2 Directive), Articles 20, 21 and 41. Official Journal of the European Union, L 333, 27.12.2022, pp. 80–152. (<https://eur-lex.europa.eu/eli/dir/2022/2555/oj>)
15. ISO/IEC JTC 1/SC 40 (2026). ISO/IEC 38505-1:2026 Information technology — Governance of data — Part 1: Application of ISO/IEC 38500 to the governance of data. ISO/IEC, Geneva, edition 2.0. (<https://webstore.iec.ch/publication/116038>)
16. ISO/IEC JTC 1/SC 27 (2022). ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection — Information security management systems — Requirements. ISO/IEC, Geneva, 3rd edition (amended by ISO/IEC 27001:2022/Amd 1:2024). (<https://webstore.iec.ch/publication/79694>)
17. ISO/IEC JTC 1/SC 27 (2025). ISO/IEC 27701:2025 Information security, cybersecurity and privacy protection — Privacy information management systems — Requirements and guidance. ISO/IEC, Geneva, edition 2.0. (<https://webstore.iec.ch/publication/110123>)
18. ISO/IEC JTC 1/SC 42 (2023). ISO/IEC 23894:2023 Information technology — Artificial intelligence — Guidance on risk management. ISO/IEC, Geneva, edition 1.0. (<https://webstore.iec.ch/publication/82914>)
19. National Institute of Standards and Technology (2020). NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management, Version 1.0. NIST Cybersecurity White Paper, 16 January 2020. (<https://doi.org/10.6028/NIST.CSWP.01162020>)
20. Rose, S., Borchert, O., Mitchell, S. and Connelly, S. (2020). Zero Trust Architecture. NIST Special Publication 800-207. (<https://doi.org/10.6028/NIST.SP.800-207>)
21. Hu, V.C., Ferraiolo, D., Kuhn, R., Schnitzer, A., Sandlin, K., Miller, R. and Scarfone, K. (2014). Guide to Attribute Based Access Control (ABAC) Definition and Considerations. NIST Special Publication 800-162 (includes updates as of 2 Aug 2019). (<https://doi.org/10.6028/NIST.SP.800-162>)
22. Atlam, H. F., Azad, M. A., Alassafi, M. O., Alshdadi, A. A., & Alenezi, A. (2020). Risk-Based Access Control Model: A Systematic Literature Review. Future Internet, 12(6), 103. (<https://doi.org/10.3390/fi12060103>)

23. Syed, N. F., Shah, S. W., Shaghaghi, A., Anwar, A., Baig, Z., & Doss, R. (2022). Zero Trust Architecture (ZTA): A Comprehensive Survey. *IEEE Access*, 10, 57143–57179. (<https://doi.org/10.1109/ACCESS.2022.3174679>)
24. Jahnke, N. and Otto, B. (2023). Data Catalogs in the Enterprise: Applications and Integration. *Datenbank-Spektrum*, 23(2), 89–96. (<https://doi.org/10.1007/s13222-023-00445-2>)
25. Ahmad, S., Arumugam, D., Bozovic, S., Degefa, E., Duvvuri, S., Gott, S., Gupta, N., Hammer, J., Kaluskar, N., Kaushik, R., Khanduja, R., Mujumdar, P., Malhotra, G., Naik, P., Ogg, N., Parthasarthy, K.K., Ramakrishnan, R., Rodriguez, V., Sharma, R., Szymaszek, J. and Wolter, A. (2023). Microsoft Purview: A System for Central Governance of Data. *Proceedings of the VLDB Endowment*, 16(12), 3624–3635. (<https://doi.org/10.14778/3611540.3611552>)
26. DataHub project (2026). DataHub Actions Framework; Classification (metadata ingestion developer guide, v1.8.0). Official documentation, accessed October 2026. (<https://docs.datahub.com/docs/actions>)
27. DataHub project (2026). Metadata Tests. Official documentation, accessed October 2026. (<https://docs.datahub.com/docs/tests/metadata-tests>)
28. OpenLineage project (LF AI & Data Foundation) (2026). OpenLineage documentation — About OpenLineage. Official documentation, accessed October 2026. (<https://openlineage.io/docs/>)
29. Ryan, P., Brennan, R. and Pandit, H.J. (2022). DPCat: Specification for an Interoperable and Machine-Readable Data Processing Catalogue Based on GDPR. *Information*, 13(5), 244. (<https://doi.org/10.3390/info13050244>)
30. Bernardo, B.M.V., Mamede, H.S., Barroso, J.M.P. and dos Santos, V.M.P.D. (2024). Data governance & quality management—Innovation and breakthroughs across different fields. *Journal of Innovation & Knowledge*, 9(4), 100598. (<https://doi.org/10.1016/j.jik.2024.100598>)
31. OpenMetadata project (Collate) (2026). Auto-Classification in OpenMetadata. Official documentation, accessed October 2026. (<https://docs.open-metadata.org/latest/how-to-guides/data-governance/classification/auto-classification>)
32. Presidio project (Data Privacy Stack; originally Microsoft) (2026). Presidio — data protection and de-identification SDK. Official documentation, accessed October 2026. (<https://presidio.dataprivacystack.org/>)
33. Kužina, V., Petric, A.-M., Barišić, M., & Jović, A. (2023). CASSED: Context-based Approach for Structured Sensitive Data Detection. *Expert Systems with Applications*, 223, 119924. (<https://doi.org/10.1016/j.eswa.2023.119924>)
34. Telkamp, L., & Hulsebos, M. (2025). Towards Contextual Sensitive Data Detection. *arXiv preprint arXiv:2512.04120* (v2, March 2026; not peer-reviewed). (<https://arxiv.org/abs/2512.04120>)
35. Open Policy Agent project (CNCF) (2026). Open Policy Agent documentation (Introduction; Policy Language — Rego). Official documentation, accessed October 2026. (<https://www.openpolicyagent.org/docs>)
36. Cutler, J. W., Disselkoen, C., Eline, A., He, S., Headley, K., Hicks, M., Hietala, K., Ioannidis, E., Kastner, J., Mamat, A., McAdams, D., McCutchen, M., Rungta, N., Torlak, E., & Wells, A. M. (2024). Cedar: A New Language for Expressive, Fast, Safe, and Analyzable Authorization. *Proceedings of the ACM on Programming Languages*, 8(OOPSLA1), 670–697. (<https://doi.org/10.1145/3649835>)

37. Pang, R., Cáceres, R., Burrows, M., Chen, Z., Dave, P., Germer, N., Golynski, A., Graney, K., Kang, N., Kissner, L., Korn, J. L., Parmar, A., Richards, C. D., & Wang, M. (2019). Zanzibar: Google's Consistent, Global Authorization System. 2019 USENIX Annual Technical Conference (USENIX ATC 19), 33–46.
(<https://www.usenix.org/conference/atc19/presentation/pang>)
38. OpenFGA project (CNCF) (2026). OpenFGA documentation — Introduction. Official documentation, accessed October 2026. (<https://openfga.dev/docs/fga>)
39. Apache Ranger project (Neethiraj, M.) (2015). Tag Based Policies. Apache Ranger wiki, official documentation, accessed October 2026.
(<https://cwiki.apache.org/confluence/display/RANGER/Tag+Based+Policies>)
40. Snowflake Inc. (2026). Tag-based masking policies. Official documentation, accessed October 2026. (<https://docs.snowflake.com/en/user-guide/tag-based-masking-policies>)
41. Starburst Data (2026). Access control policy types (Starburst Galaxy documentation); Built-in access control overview (Starburst Enterprise documentation). Official documentation, accessed October 2026. (<https://docs.starburst.io/starburst-galaxy/security-and-compliance/manage-data-access/access-control-policy-types.html>)
42. Trino Software Foundation (2026). Open Policy Agent access control (Trino documentation, Security section). Official documentation, accessed October 2026.
(<https://trino.io/docs/current/security/opa-access-control.html>)
43. Backes, J., Bolignano, P., Cook, B., Dodge, C., Gacek, A., Luckow, K., Rungta, N., Tkachuk, O., & Varming, C. (2018). Semantic-based Automated Reasoning for AWS Access Policies using SMT. 2018 Formal Methods in Computer Aided Design (FMCAD), 1–9.
(<https://doi.org/10.23919/FMCAD.2018.8602994>)
44. Lorenz, C., Clemens, V., Schrötter, M. and Schnor, B. (2022). Continuous Verification of Network Security Compliance. IEEE Transactions on Network and Service Management, 19(2), 1729–1745. (<https://doi.org/10.1109/TNSM.2021.3130290>)
45. National Institute of Standards and Technology (2026). Open Security Controls Assessment Language (OSCAL), release v1.2.3. NIST OSCAL project (pages.nist.gov/OSCAL/) and [usnistgov/OSCAL GitHub repository](https://github.com/usnistgov/OSCAL). (<https://pages.nist.gov/OSCAL/>)
46. Agarwal, V., Butler, C., Degenaro, L., Kumar, A., Sailer, A. and Steinder, G. (2022). Compliance-as-Code for Cybersecurity Automation in Hybrid Cloud. IEEE International Conference on Cloud Computing (CLOUD 2022), pp. 427–437.
(<https://doi.org/10.1109/CLOUD55607.2022.00066>)
47. Chuprikov, P., Eugster, P. and Mangipudi, S. (2025). Security Policy as Code. IEEE Security & Privacy, 23(2), 23–31. (<https://doi.org/10.1109/MSEC.2025.3535803>)
48. Fitzgerald, B., Stol, K.-J., O'Sullivan, R., & O'Brien, D. (2013). Scaling agile methods to regulated environments: An industry case study. 2013 35th International Conference on Software Engineering (ICSE), 863–872. (<https://doi.org/10.1109/ICSE.2013.6606635>)
49. Ruohonen, J., Tuli, E. A., & Morita, H. (2026). Compliance as Code: A Study of Linux Distributions and Beyond. arXiv preprint arXiv:2603.01520 (under review; not peer-reviewed). (<https://arxiv.org/abs/2603.01520>)
50. Open Policy Agent Gatekeeper project (2026). Gatekeeper documentation — How to use Gatekeeper (introduction). Official documentation, accessed October 2026.
(<https://open-policy-agent.github.io/gatekeeper/website/docs/>)

51. Byun, J.-W., & Li, N. (2008). Purpose based access control for privacy protection in relational database systems. *The VLDB Journal*, 17(4), 603–619. (<https://doi.org/10.1007/s00778-006-0023-0>)
52. Byun, J.-W., Bertino, E., & Li, N. (2005). Purpose based access control of complex data for privacy protection. *Proceedings of the 10th ACM Symposium on Access Control Models and Technologies (SACMAT '05)*, 102–110. (<https://doi.org/10.1145/1063979.1063998>)
53. Wang, F., Ko, R., & Mickens, J. (2019). Riverbed: Enforcing User-defined Privacy Constraints in Distributed Web Services. *16th USENIX Symposium on Networked Systems Design and Implementation (NSDI '19)*, 615–630. (<https://www.usenix.org/conference/nsdi19/presentation/wang-frank>)
54. Dong, W., Kirti, R., Marsala, D., Hopkins, K., Wayne, L., Roth, H., Bergeron, J., Ponomarenko, A., Brar, A., & Raghunath, A. (2024). How Meta enforces purpose limitation via Privacy Aware Infrastructure at scale. *Engineering at Meta* (company engineering blog, not peer-reviewed), 27 August 2024. (<https://engineering.fb.com/2024/08/27/security/privacy-aware-infrastructure-purpose-limitation-meta/>)
55. Schelter, S., Lange, D., Schmidt, P., Celikel, M., Biessmann, F., & Grafberger, A. (2018). Automating large-scale data quality verification. *Proceedings of the VLDB Endowment*, 11(12), 1781–1794. (<https://doi.org/10.14778/3229863.3229867>)
56. Tu, D., He, Y., Cui, W., Ge, S., Zhang, H., Han, S., Zhang, D., & Chaudhuri, S. (2023). Auto-Validate by-History: Auto-Program Data Quality Constraints to Validate Recurring Data Pipelines. *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23)*, 4991–5003. (<https://doi.org/10.1145/3580305.3599776>)

Product and project names are trademarks of their respective owners. Their mention describes publicly documented technology and does not imply a business relationship, except where Bassirah names a Strategic Technology Partner.