Chapter 2
Bridging AI Into Standards Without Compromise

Tanja Auge
http://orcid.org/0009-0006-2150-9713
University of Regensburg, Germany

Otmane Azeroual
http://orcid.org/0000-0002-5225-389X
FernUniversität in Hagen, Germany

Laura Waltersdorfer
University of Business and Economics, Vienna, Austria

ABSTRACT

Artificial intelligence, particularly generative AI, LLMs, and autonomous AI agents, is transforming organizational processes, scientific infrastructures, and digital systems. However, successful adoption requires more than technological innovation; it depends on integrating AI into standards, governance frameworks, and interoperable infrastructures while preserving reliability, transparency, and rigor. This chapter examines how AI can be embedded into technical, organizational, and scientific standards without disrupting existing systems. It addresses challenges related to interoperability, data quality, traceability, governance, provenance, metadata consistency, and reproducibility. Special attention is given to integrating AI into Current Research Information Systems and scientific workflows while maintaining transparency, validation, and responsible decision-making. The chapter proposes

DOI: 10.4018/979-8-2600-1227-7.ch002

Copyright © 2027, IGI Global Scientific Publishing. Copying or distributing in print or electronic forms without written permission of IGI Global Scientific Publishing is prohibited. Use of this chapter to train generative artificial intelligence (AI) technologies is expressly prohibited. The publisher reserves all rights to license its use for generative AI training and machine learning model development.

a framework for standards-compliant AI adoption that balances innovation with interoperability and positions AI as an evolutionary extension of sustainable and trustworthy digital infrastructures.

1. INTRODUCTION

AI has become one of the most transformative technological developments of the last decade, fundamentally reshaping digital infrastructures, organizational processes, and scientific ecosystems. The emergence of generative AI, LLMs, and autonomous decision-support systems has accelerated the integration of AI into domains such as healthcare, education, public administration, research management, and industrial innovation (Bommasani et al., 2021; Dwivedi et al., 2023). While these technologies create significant opportunities for automation, knowledge discovery, and intelligent decision-making, they also introduce substantial challenges related to interoperability, governance, transparency, and standardization.

One of the central tensions in contemporary AI adoption lies in the relationship between innovation and standards. Existing technical and organizational standards provide the foundation for interoperability, reliability, data exchange, and long-term sustainability across digital ecosystems (ISO, 2023). Particularly in scientific and research infrastructures, standards play a critical role in ensuring data quality, reproducibility, metadata consistency, and methodological transparency (Wilkinson et al., 2016). However, many AI systems, especially generative AI applications, operate probabilistically and dynamically, often producing outputs that are difficult to reproduce, validate, or systematically govern (Bender et al., 2021). As a result, organizations increasingly face the challenge of integrating AI into established infrastructures without creating fragmentation, incompatibility, or loss of trust.

The growing complexity of AI-driven environments further intensifies concerns regarding data governance and provenance. AI systems depend heavily on large-scale heterogeneous data sources whose origins, transformations, and quality are not always transparent (Sambasivan et al., 2021). In this context, provenance emerges as a crucial concept for maintaining traceability, reproducibility, and accountability within AI-supported systems. Provenance mechanisms enable organizations to systematically document how data are collected, processed, transformed, and utilized within AI pipelines, thereby supporting transparency and governance across complex digital ecosystems (Belhajjame et al., 2013).

These challenges are particularly relevant in research and scholarly infrastructures, where AI increasingly influences scientific workflows, research evaluation, metadata enrichment, and *current research information systems (CRIS)*[1]. The integration of AI into research environments raises critical questions about scientific

rigor, explainability, bias, and responsible automation in knowledge production and evaluation processes (Azeroual & Schöpfel, 2024). Without appropriate governance structures and standards-aligned integration approaches, AI may unintentionally undermine the very principles of transparency, reproducibility, and methodological reliability that scientific systems are designed to protect.

Against this background, this chapter investigates how AI can be integrated into existing standards and infrastructures without compromising their foundational principles. From this perspective, AI should be embedded within established infrastructures in a manner that preserves institutional control, methodological rigor, and system compatibility while enabling new analytical and generative capabilities.

Existing AI governance frameworks provide important principles for risk management, accountability, transparency, and regulatory compliance. However, they offer limited guidance on how AI components should be technically and organizationally embedded into information infrastructures that already rely on established metadata models, interoperability protocols, provenance mechanisms, and scientific validation procedures. This limitation is particularly relevant for CRIS, where AI-generated metadata, classifications, recommendations, and analytical outputs must coexist with standards such as CERIF, Dublin Core, DataCite, FAIR, persistent identifiers, and institutional governance processes.

The specific gap addressed in this chapter is therefore not the absence of AI governance principles. Rather, it is the lack of an integrated implementation perspective that connects governance requirements with system architecture, metadata interoperability, provenance evidence, data quality controls, scientific reproducibility, and human validation. The chapter addresses this gap by treating AI as an augmentation layer within existing standards-based infrastructures rather than as an autonomous replacement system.

1.1. Contributions

This chapter makes three main contributions:

1. It identifies a practical integration gap between general AI governance frameworks and the operational requirements of standards-based information infrastructures.
2. It proposes a layered conceptual framework in which data and metadata standards form the foundation, existing information systems provide the operational core, AI functions as an augmentation layer, and governance and human oversight control the resulting processes.
3. It applies and illustrates this framework to CRIS and research information management by linking AI integration to CERIF, FAIR, W3C PROV, Dublin Core, DataCite, ISO/IEC 42001, the NIST AI Risk Management Framework, and the

EU AI Act, and by translating these relationships into practical implementation patterns.

1.2. Methodological Approach

This chapter adopts a structured conceptual review and synthesis approach. The analysis focuses on literature and standards published in the fields of AI governance, interoperability, provenance, data quality, scientific reproducibility, research information management, and human-centered AI. Sources were selected when they fulfilled at least one of the following criteria:

1. they represent an internationally recognized governance or regulatory framework;
2. they define established data, metadata, or provenance standards;
3. they address AI integration in research or standards-based information environments; or
4. they provide concepts relevant to model documentation, reproducibility, human oversight, or lifecycle monitoring.

The analysis included ISO/IEC 42001, the NIST AI Risk Management Framework, the EU AI Act, the OECD AI Principles, FAIR, W3C PROV, CERIF, Dublin Core, and DataCite. The framework was derived by comparing the scope and limitations of these instruments and synthesizing their complementary functions into a layered architecture for standards-compliant AI integration. The chapter is conceptual and does not claim empirical validation of the proposed framework.

2. UNDERSTANDING STANDARDS IN THE AGE OF AI

The rapid diffusion of AI across organizational, industrial, and scientific domains has fundamentally transformed the relationship between technological innovation and standardization. AI is increasingly applied across multiple lifecycle phases, including data collection, data integration, knowledge extraction, decision support, process automation, monitoring, and reporting. In scientific environments, AI supports activities such as literature discovery, research evaluation, metadata enrichment, expert recommendation, and research management, while in industrial contexts it is used for predictive maintenance, quality control, digital twins, and operational optimization. As AI becomes embedded throughout these lifecycle stages, the

need for interoperable standards, governance mechanisms, and traceable processes becomes increasingly important.

Historically, standards have served as foundational mechanisms for ensuring interoperability, consistency, reliability, and trust across digital ecosystems. However, the emergence of AI-driven systems, particularly *machine learning (ML)* and generative AI technologies, introduces new complexities that challenge traditional approaches to standardization. Unlike conventional information systems that operate according to predefined rules and deterministic processes, AI systems often exhibit adaptive, probabilistic, and context-dependent behavior. As a result, organizations increasingly face the challenge of integrating AI capabilities into existing technical, organizational, and regulatory frameworks while preserving compatibility, transparency, and long-term sustainability.

Understanding the role of standards in AI-driven environments requires a multidimensional perspective that encompasses technical infrastructures, data governance, research methodologies, and compliance mechanisms. At the same time, it is necessary to recognize AI not only as a technological innovation but also as a transformative force that reshapes digital ecosystems and introduces new risks related to fragmentation, interoperability, and governance.

2.1 The Role of Standards in Digital Ecosystems

Standards provide the foundation for interoperability, data exchange, governance, and quality assurance across digital ecosystems. They establish common specifications, procedures, and frameworks that enable heterogeneous systems, organizations, and stakeholders to interact effectively across technological and institutional boundaries. Without standards, digital ecosystems would be characterized by incompatibility, duplication of effort, and reduced efficiency, ultimately limiting innovation, scalability, and collaboration (ISO, 2023).

Technical Standards

Technical standards establish the protocols, architectures, and communication mechanisms that enable interoperability among information systems. They define how systems exchange information, authenticate users, process transactions, and communicate across distributed environments. Examples include internet communication protocols such as TCP/IP, web standards developed by the World Wide Web Consortium (W3C), and service-oriented architectures based on application programming interfaces (APIs). These standards create stable technological foun-

dations that allow organizations to integrate diverse systems while maintaining operational consistency and scalability.

In AI-driven environments, technical standards become increasingly important because AI systems often operate across multiple infrastructures, cloud environments, and data platforms. Standardized interfaces facilitate the integration of AI models into existing enterprise architectures and support interoperability between AI services, data repositories, and decision-support systems.

Data and Metadata Standards

Data constitute the primary resource upon which AI systems depend. Consequently, standards governing data structures, metadata schemas, and semantic interoperability play a critical role in ensuring the quality and usability of AI-driven processes. Metadata standards provide structured descriptions of digital objects and enable systems to interpret, exchange, and reuse information consistently across organizational boundaries.

Within research and scientific infrastructures, metadata standards such as Dublin Core (Dublin Core Metadata Initiative, n.d.), DataCite Metadata Schema (DataCite Metadata Working Group, 2026), the Common European Research Information Format (euroCRIS CERIF Task Group, 2012) (CERIF), W3C Web Standards (World Wide Web Consortium, n.d.), and FAIR data principles contribute significantly to data discoverability, accessibility, interoperability, and reusability (Wilkinson et al., 2016). For AI systems, standardized metadata support data lineage tracking, provenance documentation, model transparency, and explainability. Without consistent metadata management, AI applications risk operating on incomplete, inconsistent, or poorly documented datasets, thereby reducing reliability and trustworthiness.

Research and Methodological Standards

Scientific progress depends heavily on methodological rigor, transparency, and reproducibility. Research standards establish common practices for data collection, analysis, reporting, and validation, thereby ensuring the credibility and reliability of scientific findings. In recent years, concerns regarding reproducibility crises in several scientific disciplines have reinforced the importance of standardized research methodologies and transparent documentation practices (Munafò et al., 2017; Freese et al., 2022).

The integration of AI into research environments introduces additional challenges because AI models often operate as complex *black-box systems* whose internal decision-making processes are difficult to interpret. Consequently, emerging methodological standards increasingly emphasize reproducible computational

workflows, transparent model documentation, dataset versioning, and explainable AI approaches. These requirements are particularly important in CRIS, scholarly analytics platforms, and AI-assisted research evaluation systems, where automated analyses may directly influence scientific decision-making processes.

Governance and Compliance Frameworks

Beyond technical and methodological considerations, standards also serve as governance instruments that support accountability, risk management, and regulatory compliance. Governance frameworks define organizational responsibilities, oversight mechanisms, and ethical principles that guide the development and deployment of digital technologies.

In the context of AI, governance frameworks are becoming increasingly important due to concerns regarding fairness, transparency, bias, privacy, and accountability. International initiatives such as the OECD AI Principles, the European Union AI Act, and ISO/IEC 42001 provide structured approaches for managing AI-related risks and ensuring responsible deployment practices (OECD, 2019; European Commission, 2024). These frameworks establish governance structures that enable organizations to balance innovation with societal expectations and regulatory obligations.

Limitations of Existing Standards

Although existing standards provide essential guidance for AI deployment, each addresses only a specific aspect of AI integration. Technical standards support interoperability, governance frameworks address accountability and risk management, provenance standards provide traceability, and metadata standards ensure semantic consistency. Regulatory frameworks such as the EU AI Act complement these standards by defining legal obligations for AI systems but do not prescribe how AI should be embedded into existing information infrastructures. The challenge therefore lies in combining these complementary perspectives within a single integration architecture, which is introduced in the following section.

Table 1 illustrates that existing standards and initiatives address complementary aspects of AI adoption, including governance, regulatory compliance, interoperability, provenance, data stewardship, and research information management. While each of these standards provides valuable guidance within its respective domain, none offers a comprehensive perspective on how these dimensions interact when AI is integrated into existing organizational and scientific infrastructures. Rather than proposing a new standard, this chapter synthesizes these complementary perspectives to provide a holistic understanding of standards-compliant AI adoption. This integrated perspective forms the conceptual basis for the subsequent discussion of

AI as a transformative force and the challenges of embedding AI into established standards-based ecosystems.

Table 1. Scope of Existing Standards/Initiatives and Their Relevance for AI Integration

Standard/ Initiative	Primary Focus	Scope and Limitations	Relevance for AI Integration
Standards			
ISO/IEC 42001	AI management systems	Focuses on organizational AI governance	Governance and AI management
Dublin Core	Metadata description	General metadata standard without AI-specific governance	Metadata interoperability and semantic consistency
Frameworks and Principles			
NIST AI RMF	AI risk management	Provides a general lifecycle-oriented risk management framework but does not prescribe domain-specific CRIS integration architectures	Risk identification, governance, measurement, and management
FAIR Principles	Data stewardship	Focuses on research data	Data quality and interoperability
Recommendations and Specifications			
W3C PROV	Provenance and traceability	Covers traceability only	Provenance and auditability
CERIF/CRIS	Research information interoperability	Focuses on research information interoperability	AI integration into research information systems
Regulations			
EU AI Act	Regulatory compliance	Defines legal obligations	Regulatory requirements and risk management

Table 1 also clarifies the distinct contribution of the proposed framework. ISO/IEC 42001 addresses organizational management structures, the National Institute of Standards and Technology (NIST) AI RMF(Tabassi, 2023) supports risk identification and lifecycle governance, and the EU AI Act defines legal obligations. FAIR, Dublin Core, DataCite, CERIF, and W3C PROV address data stewardship, metadata representation, research information interoperability, and traceability. The framework proposed in this chapter does not replace these instruments. Instead, it explains how their functions can be combined within one layered implementation architecture. Its distinctive contribution lies in connecting policy-level governance with the technical and scientific requirements of existing information infrastructures.

2.2 AI as a Transformative Force

AI has evolved from a specialized technological capability into a general-purpose technology with the potential to reshape virtually every sector of society. Similar to previous transformative technologies such as electricity or the internet, AI acts as an enabling infrastructure that influences economic processes, organizational structures, and knowledge production mechanisms (McAfee & Brynjolfsson, 2017).

From Automation to Generative AI

Early AI applications primarily focused on automation and rule-based decision support. Traditional expert systems relied on predefined rules and structured knowledge representations to automate repetitive tasks and support operational decision-making. Advances in machine learning, particularly deep learning, expanded AI capabilities by enabling systems to learn patterns from large-scale datasets without explicit programming.

The recent emergence of generative AI and LLMs represents a significant shift in AI capabilities. Unlike traditional predictive systems, generative AI can create novel content, including text, images, software code, and scientific summaries. Models such as GPT, Claude, and Gemini demonstrate unprecedented abilities in natural language processing, content generation, and knowledge synthesis (Bommasani et al., 2021). This evolution significantly expands the scope of AI applications across education, research, healthcare, public administration, and business environments.

AI-Driven Decision-Making and Adaptive Systems

Modern AI systems increasingly support or automate decision-making processes in complex environments. Machine learning models are used for predictive analytics, recommendation systems, fraud detection, resource allocation, and strategic planning. In many cases, AI systems continuously adapt to changing conditions through ongoing learning processes and dynamic feedback loops.

Adaptive AI systems offer significant advantages by improving responsiveness, efficiency, and scalability. However, their dynamic nature also creates governance challenges because system behavior may evolve over time in ways that are difficult to predict or explain. The transition from static information systems to adaptive AI-driven environments therefore requires new approaches to monitoring, validation, and accountability.

Opportunities and Systemic Risks

The transformative potential of AI creates significant opportunities for innovation, productivity improvements, and knowledge discovery across organizational, industrial, and scientific domains. AI technologies increasingly support data analysis, decision-making, automation, prediction, and knowledge generation at scales that would be difficult to achieve through traditional approaches. In research environments, AI enables applications such as literature discovery, metadata enrichment, expert identification, research intelligence, funding analytics, and scientific knowledge extraction. More recently, generative AI, LLMs, and emerging agent-based systems have expanded these capabilities by supporting content generation, conversational interfaces, workflow automation, and intelligent decision support.

AI can also contribute directly to the implementation and operationalization of standards themselves. Rather than being viewed only as a technology that must be governed through standards, AI can support organizations in assessing, maintaining, and improving alignment with established requirements. In research information infrastructures, for example, AI can assist with automated metadata enrichment, semantic classification, identification of missing or inconsistent metadata, and mapping between different metadata schemas. Such capabilities can reduce manual effort while improving the consistency and completeness of institutional information.

AI can further support standards-oriented quality assessment and assurance. For example, AI-assisted tools can help assess datasets and metadata against FAIR-related criteria, identify missing persistent identifiers, detect incomplete provenance information, or flag potential interoperability problems. Similarly, AI can support the continuous monitoring of governance requirements by identifying deviations from predefined rules, documentation requirements, or quality thresholds. In this context, AI functions not only as an application consuming standards-compliant data, but also as an instrument for operationalizing and monitoring standards compliance.

These opportunities are particularly relevant for complex standards ecosystems in which compliance cannot be reduced to a single technical check. AI can assist in translating formal requirements into practical checks, recommendations, and workflow actions, while human experts remain responsible for interpreting ambiguous cases and making final decisions. This combination of automated assessment and human validation can support a more continuous and scalable approach to standards implementation.

At the same time, the use of AI for standards assessment introduces an additional governance challenge, particularly because defining what constitutes an AI system has itself become increasingly difficult. The term AI encompasses a broad range of technologies, including rule-based systems, machine learning algorithms, generative AI, large language models, and autonomous agents. Different organizations and

regulatory frameworks, including the OECD, NIST, and the European Union AI Act, apply varying definitions and classifications of AI systems. The rapid evolution of AI technologies further complicates regulatory and governance efforts, as policymakers and legislators often struggle to establish clear boundaries between conventional software, automation, and artificial intelligence. This conceptual ambiguity can also affect AI-assisted standards assessment, as such assessments may be incomplete, biased, or based on an incorrect interpretation of the underlying requirements. AI-generated compliance assessments should therefore not be treated as authoritative decisions, but as recommendations or evidence that remain subject to appropriate human validation. This creates a recursive governance challenge: AI can support standards implementation, while its own use and outputs must remain subject to appropriate standards and human oversight.

Despite its considerable benefits, AI also introduces significant systemic risks. These include algorithmic bias, discrimination, privacy violations, misinformation, hallucinations, cybersecurity vulnerabilities, and the excessive automation of critical decisions (Bender et al., 2021). In addition, concerns arise regarding the transparency of training datasets, model architectures, fine-tuning processes, and the tools used to develop and deploy AI systems. Many advanced AI models are provided as proprietary commercial services, limiting insight into their internal mechanisms and making independent validation, auditing, and explainability more difficult.

Further challenges relate to governance and digital sovereignty. The growing dependence on a small number of technology providers may create structural dependencies that affect competition, organizational autonomy, and long-term strategic flexibility. Vendor lock-in, changing pricing models, restricted access to model internals, and dependence on external cloud infrastructures may introduce operational and financial risks for organizations. These concerns are particularly relevant in public-sector, scientific, and critical infrastructure environments, where transparency, accountability, and long-term accessibility are essential requirements.

The increasing availability of open-source and open-weight AI models offers alternative approaches that may enhance transparency, auditability, and technological independence. However, these models may not always provide the same performance, scalability, security guarantees, or support ecosystem as leading proprietary solutions. Organizations therefore face the challenge of balancing innovation, performance, transparency, governance requirements, and digital sovereignty when selecting AI technologies.

Institutions can reduce vendor dependency through several architectural and contractual measures. These include the use of model-independent APIs, portable metadata and prompt formats, exportable audit logs, internally controlled vector databases, open identity standards, documented fallback procedures, and contractual rights to retrieve generated data and system records. For high-dependency services,

institutions should also define exit strategies and periodically test whether critical workflows can be transferred to an alternative provider or an internally hosted model.

Consequently, the successful adoption of AI requires more than technological capability alone. It demands robust governance mechanisms, transparent standards, effective risk management, and clear accountability structures that ensure AI systems remain trustworthy, explainable, and aligned with organizational and societal objectives.

2.3 The Risk of Fragmentation

While AI offers substantial opportunities for innovation and transformation, its rapid adoption also increases the risk of fragmentation across digital ecosystems. Fragmentation occurs when technologies, data structures, governance mechanisms, or organizational practices evolve independently without sufficient coordination or standardization.

Incompatibility Between AI Systems and Existing Infrastructures

Many AI applications are developed independently of existing enterprise architectures and information systems. As a result, organizations frequently encounter integration challenges when attempting to embed AI capabilities into established infrastructures. Differences in data formats, metadata schemas, APIs, and governance models may create interoperability barriers that limit the effectiveness of AI deployments.

In scientific environments, incompatibilities between AI tools and CRIS can undermine data quality, reproducibility, and traceability. Without standardized integration approaches, AI systems may create isolated information silos that reduce transparency and hinder collaboration across institutions.

Vendor Lock-In and Isolated AI Ecosystems

A significant challenge associated with contemporary AI adoption is the growing dependence on proprietary platforms and cloud-based AI services. Many organizations rely on commercial AI providers that offer powerful capabilities but operate within closed technological ecosystems. Such dependencies may create vendor lock-in situations in which organizations face high switching costs and limited flexibility regarding future technology choices.

Vendor lock-in also affects governance and transparency because proprietary AI systems often restrict access to model architectures, training data, and decision-making processes. These limitations may hinder independent auditing, explainability, and

regulatory compliance efforts. Consequently, organizations increasingly seek open standards and interoperable architectures to reduce dependency risks and maintain technological sovereignty.

Challenges for Interoperability and Sustainability

Long-term sustainability of AI-enhanced ecosystems depends on interoperability across technical, organizational, and regulatory dimensions. However, the rapid pace of AI innovation often outpaces standardization processes, creating gaps between emerging technologies and existing governance frameworks.

Addressing these challenges requires a proactive standardization strategy that promotes *interoperability-by-design*, *governance-by-design*, and *provenance-by-design principles*. Such approaches ensure that AI systems remain compatible with existing infrastructures while supporting transparency, accountability, and sustainable digital transformation. Rather than allowing AI to create isolated technological ecosystems, organizations should integrate AI within broader standards-based frameworks that preserve trust, interoperability, and long-term resilience.

The challenge, therefore, is not the absence of standards themselves but the lack of an integrated architectural perspective that explains how existing standards can jointly support AI adoption across organizational and scientific infrastructures. The following section introduces the proposed layered integration framework, which operationalizes this architectural perspective.

These fragmentation risks demonstrate that standards must be translated into concrete architectural relationships. Section 3 therefore examines how AI can function as an extension layer within established information infrastructures.

3. INTEGRATING AI INTO EXISTING STANDARDS

The increasing adoption of AI across organizations and research institutions has raised fundamental questions regarding its relationship with existing information infrastructures. Although AI is often portrayed as a disruptive technology capable of replacing established systems and processes, practical implementation scenarios reveal a different reality. Organizations typically introduce AI into mature socio-technical environments comprising enterprise systems, CRIS, digital repositories, data warehouses, established governance structures, and standardized workflows. These infrastructures represent substantial organizational investments, support mission-critical processes, and are embedded within regulatory and operational frameworks that have evolved over many years. Consequently, the primary challenge is not to replace existing infrastructures but to integrate AI capabilities into

them without compromising their stability, interoperability, or governance. For this reason, recent research increasingly emphasizes AI augmentation rather than AI substitution (Dellermann et al., 2019).

From an augmentation perspective, AI contributes additional analytical and decision-support capabilities without fundamentally altering the underlying infrastructure. Machine learning models may enhance information retrieval, support predictive analyses, automate metadata generation, or identify hidden patterns in complex datasets. However, the surrounding information systems continue to provide the operational and governance framework within which these capabilities are deployed. In this sense, AI functions as an additional layer of intelligence that complements existing systems rather than replacing them. For example, a university does not replace its CRIS with an LLM. Instead, an AI service is connected via standardized APIs to enrich publication metadata by suggesting keywords, research topics, or links to persistent identifiers. Existing CERIF structure, identifiers such as DOI and ORCID, and established governance processes remain unchanged, while AI provides recommendations that can be reviewed and validated before being incorporated into the system.

3.1 Positioning the Proposed Framework

The proposed framework is positioned between general AI governance and domain-specific information architecture. It does not introduce a new legal, ethical, or technical standard. Instead, it provides an integration model for combining existing instruments in operational environments.

ISO/IEC 42001 contributes organizational management requirements. The NIST AI RMF contributes lifecycle-oriented risk management. The EU AI Act contributes legal and risk-based obligations. FAIR, Dublin Core, DataCite, CERIF, and W3C PROV contribute requirements for data stewardship, metadata interoperability, research information exchange, and traceability. The proposed framework connects these dimensions through a four-part architecture:

1. standards-based data and metadata foundations;
2. established operational information systems;
3. AI-based augmentation services; and
4. governance, human oversight, and evidence mechanisms.

This architecture is particularly relevant where AI outputs must be incorporated into controlled information systems rather than used as standalone content. In a CRIS, for example, an AI-generated classification is not merely a model output. It becomes a candidate institutional record that must be linked to a source, aligned

with a metadata schema, reviewed by a responsible person, and documented for later verification.

3.2 AI as an Extension Rather Than a Replacement

AI integration should therefore be understood as an evolutionary process in which new computational capabilities are added to existing standards-based infrastructures. This perspective is particularly important in domains where reliability, traceability, and institutional accountability are central requirements. In research institutions, public administrations, healthcare organizations, and regulated industries, digital infrastructures are not merely technical systems. They are socio-technical arrangements that connect data, people, processes, standards, and governance mechanisms.

Preserving Existing Infrastructures

Replacing these infrastructures with fully AI-driven alternatives would risk weakening established mechanisms of control, validation, and accountability. By contrast, an extension-oriented approach allows organizations to benefit from AI while preserving the stability of existing systems. AI can support the automation of repetitive processes, the enrichment of metadata, the identification of semantic relationships, the generation of analytical insights, and the improvement of decision support. However, these capabilities should remain embedded within established data models, workflows, and governance structures.

Figure 1 illustrates this layered integration approach. The framework positions AI between existing information systems and governance structures while maintaining the foundational role of data and metadata standards. Such an architecture reflects the observation that AI systems depend heavily on standardized data resources, interoperable infrastructures, and clearly defined governance mechanisms. Without these foundations, the benefits of AI cannot be realized in a reliable and sustainable manner.

Figure 1. Layered framework for integrating AI into existing standards-based ecosystems. Data and metadata standards provide the semantic and technical foundation. Existing information systems remain the operational core. AI services are introduced as modular augmentation components, while governance, compliance, and human oversight control their use and validate their outputs.

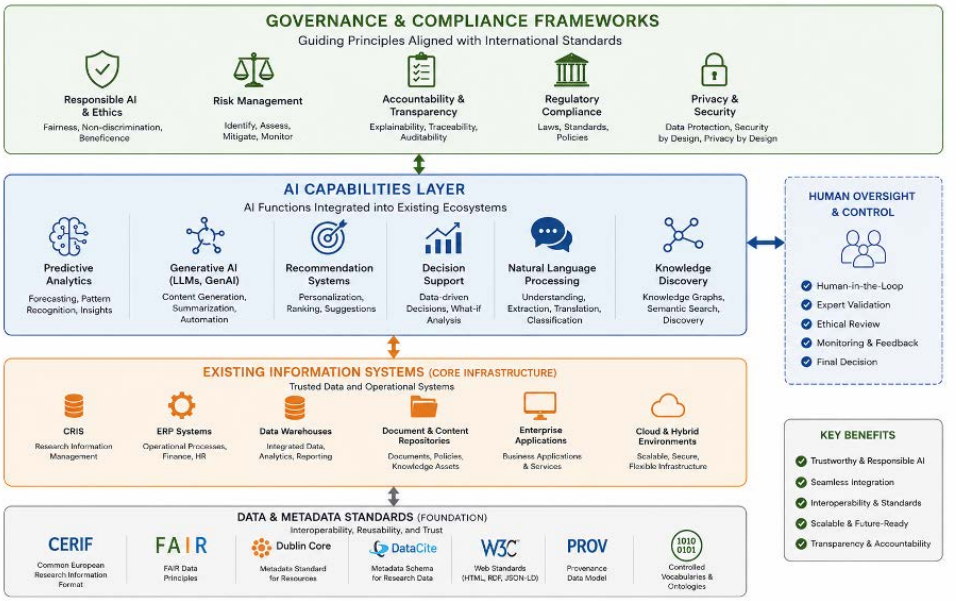


Figure 1 should be read from bottom to top. The foundation layer contains the standards and identifiers that structure and connect data, including CERIF, FAIR, Dublin Core, DataCite, W3C standards, provenance models, and controlled vocabularies. The second layer consists of the existing operational systems in which institutional records and workflows are maintained. These systems remain authoritative. The third layer contains modular AI capabilities such as metadata enrichment, predictive analytics, natural language processing, recommendation systems, and generative AI. These services access information through controlled interfaces and return suggestions, classifications, summaries, or analytical results. They do not automatically replace authoritative records. The upper layer defines governance and compliance requirements. It defines organizational responsibilities, regulatory controls, and the human validation points through which AI-generated outputs become authoritative institutional records.

Importantly, the layered representation should not be interpreted as implying that standards and governance operate only at individual architectural levels. While data and metadata standards such as CERIF, Dublin Core, DataCite, and W3C PROV provide the semantic and technical foundation, governance and compliance require-

ments operate across the architecture. They guide the integration and operation of AI capabilities, existing information systems, and their interactions, while human oversight provides an additional control mechanism for consequential AI-supported processes.

For example, consider AI-supported metadata enrichment in a CRIS. A publication record stored in the existing CRIS is structured according to established metadata standards and linked to persistent identifiers such as DOI and ORCID. An AI service analyzes the publication and proposes additional keywords or research classifications. These suggestions are returned to the CRIS through a controlled interface and remain identifiable as AI-generated candidates. Depending on the applicable governance rules, human validation may be required before the enriched metadata becomes part of the authoritative institutional record. Provenance information documents the source data, generated suggestion, validation step, and responsible actor. This example illustrates how the architectural layers interact in a concrete research information management process.

Human-Centered and Standards-Aligned AI Integration

This layered perspective aligns with current approaches to human-centered AI, which emphasize maintaining human control, accountability, and institutional responsibility in AI-supported environments (Shneiderman, 2022). AI is therefore understood as an evolutionary capability that expands the potential of established information infrastructures rather than as a disruptive replacement mechanism. By preserving interoperability, governance, and evidential integrity, the framework supports the sustainable integration of AI into complex institutional contexts.

3.3 Compatibility with Existing Protocols and Architectures

The integration of AI into existing standards also depends on the technical compatibility between AI components and established information architectures. Many organizations operate complex environments consisting of legacy systems, databases, repositories, enterprise platforms, APIs, and domain-specific applications. These systems frequently use different data models, exchange formats, authentication mechanisms, and communication protocols. As a result, AI integration cannot be reduced to the deployment of a model or algorithm. It requires careful architectural design that enables AI services to interact with existing infrastructures without creating additional fragmentation.

Legacy Systems and AI Integration

Legacy systems represent a particular challenge in this context. Although they may be technically outdated, they often remain essential for organizational operations because they store critical data, support established workflows, and reflect institutional knowledge accumulated over many years. Replacing them is rarely feasible in the short term. Therefore, AI integration often requires intermediary layers that connect legacy infrastructures with modern AI services. These may include APIs, middleware, data pipelines, semantic mapping tools, and modular service architectures.

APIs, Interoperability Layers, and Modular Architectures

APIs play a central role in this process because they provide standardized interfaces through which AI systems can access data, exchange information, and return analytical results. Well-designed APIs reduce dependency between system components and allow organizations to introduce AI services incrementally. Instead of directly modifying core systems, organizations can expose selected functionalities or datasets through controlled interfaces. This supports modularity, maintainability, and security while reducing the risk of disrupting existing operations. For example, an institutional repository may expose publication metadata through a REST API. An AI service retrieves these records, performs tasks such as entity resolution and keyword extraction, and returns enriched metadata to the repository through the same standardized interface. The repository itself remains unchanged, while AI capabilities are added as a loosely coupled service. This architecture preserves interoperability, facilitates incremental adoption, and minimizes the risks associated with modifying existing systems.

Interoperability layers further support AI integration by translating between different data formats, harmonizing metadata structures, and coordinating interactions between heterogeneous systems. Such layers are particularly important in multi-system environments where AI applications must process data from various sources. Without interoperability mechanisms, AI systems may operate on inconsistent or incomplete information, which can negatively affect model performance, reliability, and trustworthiness.

Semantic Interoperability and Metadata Harmonization

Semantic interoperability is especially important for standards-compliant AI integration. Technical connectivity alone does not guarantee that systems interpret data in the same way. For example, different systems may use different labels,

classifications, identifiers, or metadata structures for the same concept. In CRIS, this problem is particularly relevant because publications, projects, researchers, organizations, grants, and research outputs are often represented across multiple platforms and databases. Semantic harmonization helps ensure that these entities can be interpreted consistently across systems.

Metadata standards, ontologies, controlled vocabularies, persistent identifiers, and knowledge graphs provide important mechanisms for achieving semantic interoperability. In research environments, standards such as CERIF, Dublin Core, DataCite, ORCID, ROR (Research Organization Registry, n.d.), and FAIR principles contribute to the structured representation of scholarly information (Zendulková et al., 2025). These standards enable AI systems to process research information in a more reliable and transparent manner. For example, AI-supported metadata enrichment can only be effective if the underlying metadata structures are sufficiently consistent and machine-interpretable.

Knowledge graphs are increasingly relevant in this context because they allow heterogeneous information sources to be connected through explicit semantic relationships. By representing entities and relationships in a structured form, knowledge graphs can support AI applications in information retrieval, recommendation, research analytics, and explainable decision support. They also contribute to transparency because the relationships used by AI systems can be inspected, validated, and documented.

From an architectural perspective, compatibility with existing protocols and standards requires modular and layered integration. **AI services should not be embedded as isolated or opaque components, but connected through transparent interfaces, documented data flows, and standardized metadata structures.** Such an approach enables organizations to combine innovation with continuity. It allows AI to enhance existing infrastructures while maintaining interoperability, governance, and long-term sustainability.

MLOps and Versioning

Technical compatibility must also extend beyond data exchange to the operational lifecycle of AI models. Institutions should maintain versioned records of datasets, model configurations, code, evaluation results, and deployment environments. A model update should be treated as a controlled system change rather than as a routine software patch, particularly when the model influences metadata, classifications, recommendations, or institutional reports. Model registries, reproducible pipelines, approval gates, rollback mechanisms, and post-deployment monitoring therefore form part of standards-compliant integration (Sculley et al., 2015).

3.4 Generative AI and the Provenance Gap

Generative AI introduces additional challenges for standardization because its outputs are dynamic, probabilistic, and context-dependent. Traditional information systems usually operate according to deterministic rules: given the same input and system configuration, they are expected to produce the same output. Generative AI systems, especially large language models, do not follow this logic. Their outputs depend on model architecture, training data, prompts, system instructions, decoding parameters, retrieval contexts, and model versions. As a result, identical or similar inputs may produce different outputs under different conditions.

Dynamic and Probabilistic Outputs

This probabilistic behavior complicates traditional approaches to validation, reproducibility, and quality assurance. In standards-based environments, system outputs must often be verifiable, auditable, and reproducible. Generative AI challenges these assumptions because the generated content may vary and may not always be directly traceable to a specific source. This becomes particularly problematic in scientific, administrative, and regulatory contexts where decisions must be justified and documented.

Explainability and Reproducibility Issues

A central challenge is explainability. Many generative AI models operate as complex systems whose internal reasoning processes cannot be easily inspected. Although they may produce fluent and convincing outputs, these outputs do not necessarily reflect factual accuracy, methodological validity, or transparent reasoning. This creates risks for organizations that rely on AI-generated summaries, recommendations, classifications, or evaluations. Explainability mechanisms are therefore necessary to clarify how outputs are generated, which data sources were used, and what limitations or uncertainties are associated with the result.

Reproducibility is another major challenge. In scientific contexts, reproducibility is a core methodological principle. However, generative AI systems complicate reproducibility because outputs may be influenced by changing model versions, stochastic generation processes, updated training data, and external retrieval mechanisms. Even when prompts are documented, the same system may not always generate identical outputs. Therefore, standards for generative AI should not only focus on reproducing exact outputs, but also on documenting the conditions under which outputs were produced.

This requires systematic provenance documentation. Relevant information may include the model version, prompt, system instructions, input data, retrieval sources, parameters, date of generation, human modifications, and validation steps. Such documentation enables contextual reproducibility, meaning that the conditions of generation can be reconstructed even if the exact output cannot be reproduced. This approach connects directly to provenance-by-design and auditability principles discussed in Section 4.3.

Managing Uncertainty and Model Drift

Generative AI also creates risks related to hallucinations and misinformation. AI-generated outputs may include plausible but incorrect statements, fabricated references, or misleading interpretations. These risks are particularly serious in research environments, where inaccurate metadata, incorrect citations, or distorted summaries may affect scholarly evaluation and decision-making. Standards-compliant AI integration therefore requires validation mechanisms that combine automated checks with human oversight.

Model drift and uncertainty further complicate standardization. AI systems may change over time as models are updated, retrained, fine-tuned, or connected to new data sources. This means that the behavior of an AI system cannot be treated as fixed. Continuous monitoring is required to detect changes in performance, bias, reliability, or output quality. In this sense, AI standardization cannot be limited to initial system certification. It must include lifecycle governance, continuous evaluation, and periodic reassessment.

Managing these challenges requires standards-aligned governance practices tailored to generative AI. These include prompt and model-version documentation, source verification, output validation, audit trails, continuous monitoring, human review, and clearly assigned responsibility for the acceptance or rejection of generated content.

A concrete CRIS-related example illustrates this risk. An LLM used for metadata enrichment may generate a plausible project title, funding relationship, author affiliation, or publication reference that is not present in the source record. If the generated value is written directly into the CRIS, the error can propagate into institutional reports, researcher profiles, funding statistics, and external data exports. The problem is therefore not limited to an incorrect answer displayed in a conversational interface. It can become a persistent institutional metadata error.

A standards-compliant workflow would require the generated value to remain marked as an AI suggestion, linked to its source documents, accompanied by the model and prompt version, and validated before it is accepted as an authoritative record.

Documented cases in scholarly communication demonstrate that generative systems may fabricate references, attribute incorrect bibliographic details, or produce unsupported summaries. Such errors are especially consequential when generated information is transferred into persistent scholarly records or used in research evaluation. These incidents reinforce the need for source-linked generation, bibliographic verification, and human approval before AI-generated information is incorporated into CRIS or publication workflows (Walters & Wilder, 2023).

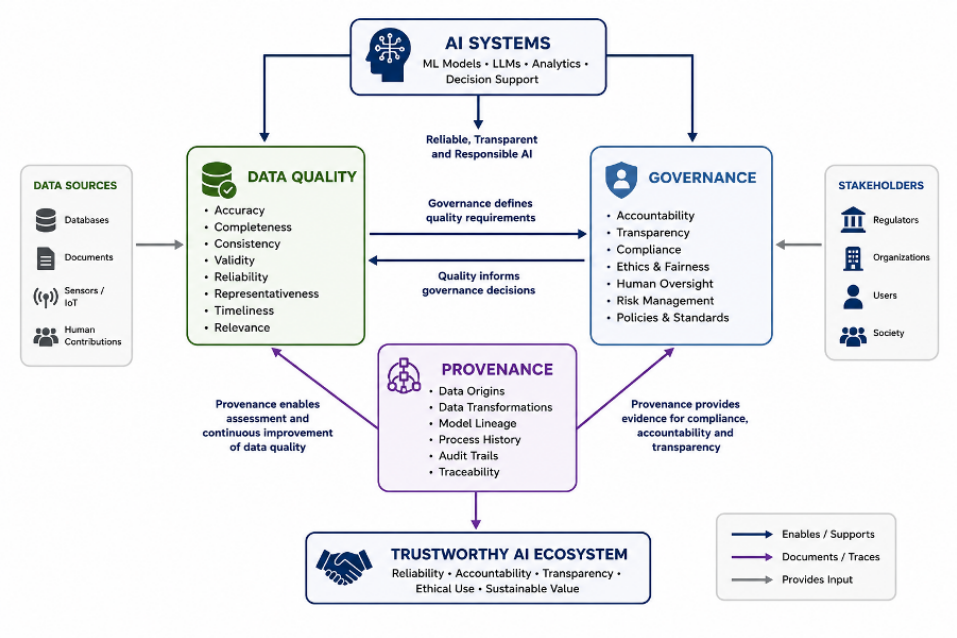
Overall, generative AI does not eliminate the need for standards; rather, it increases their importance. The more dynamic and probabilistic AI systems become, the more organizations require structured mechanisms for documentation, validation, interoperability, and governance. Standards provide the necessary foundation for ensuring that generative AI can be integrated into existing infrastructures without undermining reliability, transparency, or accountability.

The integration of AI into standards-based environments therefore extends beyond technical interoperability and system architecture. As AI systems become increasingly data-driven and adaptive, questions of data quality, governance, and traceability become central prerequisites for trustworthy and standards-compliant deployment. These aspects are examined in the following section.

4. DATA QUALITY, GOVERNANCE, AND PROVENANCE

Building on the layered integration framework introduced in Section 3, this section operationalizes the three architectural control mechanisms: data quality, governance, and provenance. Together, they provide the technical and organizational evidence required to verify AI-supported processes throughout the lifecycle.

Figure 2. Relationship between Data Quality, Governance, and Provenance in Trustworthy AI Ecosystems.



Source: Author's own illustration.

Figure 2 illustrates the interdependent relationship between data quality, governance, and provenance within AI-driven ecosystems. While data quality determines the reliability of inputs and outputs, governance establishes accountability and oversight mechanisms. Provenance acts as the connecting layer that documents data origins, transformations, and model processes, thereby enabling transparency, auditability, and trust. Together, these dimensions contribute to the development of trustworthy and standards-compliant AI systems.

4.1 Data Quality in AI-Driven Systems

Data quality constitutes a foundational requirement for reliable AI systems and directly influences the performance, robustness, fairness, and trustworthiness of machine learning models. In contrast to traditional data management environments, AI-driven systems do not treat data quality as a static attribute of datasets but as a dynamic property emerging from continuously evolving data pipelines, feedback loops, and deployment contexts.

Accuracy, Consistency, Completeness, and Reliability

Classical dimensions of data quality include accuracy, consistency, completeness, validity, and reliability (Ehrlinger and Wöß, 2022). In AI-centric environments, these dimensions must be extended by additional aspects such as representativeness, temporal relevance, and distributional stability. These extensions are necessary because ML-systems are highly sensitive to shifts in input distributions, noisy annotations, and structural biases embedded in large-scale heterogeneous datasets (Sambasivan et al., 2021).

Data-quality problems propagate throughout the AI lifecycle. Errors introduced during data collection, labeling, preprocessing, or integration may be amplified during model training and inference, resulting in systematic distortions in downstream outputs. This effect is particularly pronounced in distributed AI environments where data pipelines span multiple organizational actors and technical systems.

Bias Propagation and Hallucinations

In addition to these classical challenges, AI systems are increasingly affected by data-centric failure modes such as dataset shift, label noise accumulation, feedback loops, and contamination of training corpora through synthetic or model-generated data. In generative AI systems, these issues may also contribute to hallucination phenomena, where model outputs are not grounded in the underlying data distribution or reliable factual source. These dynamics may lead to recursive degradation effects when AI-generated content is reintegrated into training datasets without sufficient validation mechanisms.

Data Validation Mechanisms

To address these challenges, contemporary AI engineering increasingly relies on structured data quality assurance practices. These include automated validation pipelines, dataset versioning and lineage tracking, continuous monitoring of data distributions, anomaly detection mechanisms, human-in-the-loop validation, and benchmarking against reference datasets or established gold standards. These mechanisms reflect a broader shift toward data-centric AI, in which improvements in data quality are recognized as a primary driver of system performance (Zha et al., 2023).

Interoperability-by-Design

Interoperability-by-design requires that data structures, metadata models, identifiers, and exchange mechanisms are considered during system architecture develop-

ment rather than addressed retrospectively. In research information infrastructures, this principle enables heterogeneous systems to exchange and interpret information consistently through shared semantic models and established standards.

Beyond technical connectivity, interoperability-by-design contributes directly to data quality by preserving semantic consistency, contextual meaning, and traceability across distributed information environments. Standards and controlled vocabularies support this approach by providing common structures for representing, exchanging, and interpreting research information. In AI-driven systems, such foundations are essential because model performance and reliability depend not only on the volume of available data but also on the consistency and interpretability of the underlying information.

Data Observability and Continuous Monitoring

Recent developments in data-centric AI increasingly emphasize the concept of data observability (Kumar et al., 2024). Unlike traditional data quality management, which primarily assesses static dataset characteristics, data observability continuously monitors data pipelines, schema changes, distribution shifts, and data freshness across operational environments. This enables the early detection of anomalies, pipeline failures, and concept drift before they affect AI model performance, making data observability an important complement to traditional data quality assurance.

This systemic nature of data quality provides the conceptual basis for governance and provenance mechanisms, which extend data quality management by introducing organizational accountability and traceable evidence across the AI lifecycle. These mechanisms complement the architectural integration principles introduced in Section 3.

4.2 Governance Frameworks for Responsible AI

Within the proposed framework, governance operationalizes the organizational control layer by assigning responsibilities, defining decision authority, and establishing continuous oversight across the AI lifecycle. In contrast to traditional IT governance, AI governance must address additional complexity introduced by probabilistic system behavior, model opacity, continuous learning processes, and dynamically evolving data environments (OECD, 2019).

Accountability and Transparency

Accountability refers to the ability to trace AI-supported decisions back to responsible actors, data sources, and system configurations. Transparency refers to

the capacity to inspect and understand system behavior across different levels of abstraction. AI governance can be understood as a multi-layered structure comprising technical, organizational, regulatory, and ethical dimensions that jointly shape the operation of AI systems within digital ecosystems.

Governance operates across multiple layers that translate these principles into concrete mechanisms: technical governance addresses system-level controls, while organizational governance establishes responsibilities and decision structures.

Technical governance refers to mechanisms embedded within AI infrastructures that ensure observability and controllability of system behavior. This includes model versioning, logging infrastructures, audit trails, explainability tools, monitoring systems, and reproducibility mechanisms. These elements enable post-hoc analysis of system decisions and support transparency at the technical level.

Organizational governance defines roles, responsibilities, and accountability structures within institutions deploying AI systems. This includes governance boards, data stewardship functions, model risk management teams, and formalized accountability frameworks. Such structures ensure that AI systems are embedded within clearly defined decision-making and oversight processes rather than operating as isolated technical artifacts.

These internal governance structures must operate within a broader ecosystem of regulatory, ethical, and normative expectations that shape how AI systems are designed, deployed, and monitored.

Regulatory and Ethical AI Governance

Regulatory governance encompasses external legal and normative frameworks such as the EU AI Act, GDPR, and ISO/IEC AI management standards. These frameworks introduce risk-based classifications for AI systems and define requirements related to transparency, documentation, and human oversight (European Commission, 2024).

Ethical governance addresses normative dimensions that extend beyond legal compliance. It includes principles such as fairness, non-discrimination, harm prevention, safety, and broader societal and environmental considerations. Ethical governance ensures that AI systems remain aligned not only with regulatory requirements but also with societal values and expectations (Floridi et al., 2018).

Lifecycle-Oriented Governance

AI governance should not be understood as a one-time compliance activity limited to system deployment. Instead, governance mechanisms must accompany the entire AI lifecycle, including design, development, validation, deployment,

monitoring, maintenance, and retirement phases. Each stage introduces distinct risks and governance requirements that must be addressed through appropriate controls and oversight mechanisms.

Lifecycle-oriented governance approaches have become increasingly important in response to the adaptive and evolving nature of AI systems. Frameworks such as the NIST AI Risk Management Framework and ISO/IEC 42001 emphasize continuous monitoring, periodic reassessment, and ongoing risk management throughout the operational lifetime of AI systems.

Governance-by-Design

Governance-by-design extends lifecycle-oriented AI governance by embedding accountability, oversight, risk management, and compliance requirements directly into AI-enabled workflows and system architectures. Rather than treating governance as an external control layer applied after system development, this approach considers governance capabilities as inherent properties of responsible AI infrastructures.

In practice, governance-by-design requires that roles, decision authorities, monitoring mechanisms, documentation requirements, and human validation points are defined during the design and implementation phases of AI systems. This enables organizations to operationalize regulatory and ethical requirements while maintaining continuous oversight throughout the AI lifecycle. This governance layer corresponds to the upper architectural layer presented in Figure 1 (Section 3.2).

Compliance and Regulatory Considerations

Across all governance dimensions, accountability and transparency represent core principles. However, the effectiveness of governance frameworks depends critically on the availability of structured traceability information. Without systematic documentation of data origins, transformations, and model processes, governance remains partially declarative and limited in its ability to enforce compliance. This dependency establishes the conceptual transition toward provenance as an enabling layer that connects data quality and governance through verifiable system-level evidence.

Privacy-by-Design

Privacy requirements should be addressed before data are transferred to an AI component. In CRIS environments, records may contain personal data relating to researchers, reviewers, applicants, project partners, or research participants. AI integration should therefore apply purpose limitation, data minimization, role-

based access control, retention rules, and documented legal bases. Where possible, institutions should separate publicly available research metadata from confidential administrative records and avoid sending unnecessary personal information to external model providers. Privacy impact assessments should also examine whether prompts, logs, embeddings, or generated outputs may reveal personal or sensitive information.

Environmental Sustainability

Governance decisions should also consider the environmental effects of AI deployment. Large-scale training and repeated inference may require substantial computing resources, while smaller domain-specific models, retrieval-based systems, caching, and task-specific pipelines may provide sufficient performance at lower computational cost. Institutions should therefore include resource consumption, model size, expected usage frequency, hosting location, and reuse of existing services in technology assessments. Environmental considerations should be treated as part of system proportionality: the computational intensity of an AI solution should be justified by the importance and expected benefit of the use case (Strubell et al., 2020).

Multilingual and Linguistic Bias

Research information infrastructures operate across languages, countries, and disciplinary cultures. AI models trained predominantly on English-language resources may perform less reliably when classifying publications, extracting entities, generating summaries, or matching research topics in other languages (Blodgett et al., 2020). Linguistic bias may therefore affect metadata quality, researcher visibility, and evaluation outcomes. CRIS operators should test model performance across the languages represented in their institution, preserve original-language metadata, document translation processes, and involve domain experts when multilingual outputs influence reporting or assessment.

4.3 Provenance as a Core Principle

Within the proposed framework, provenance provides the evidential foundation that enables governance decisions to be verified. Rather than functioning as an isolated documentation mechanism, provenance connects data, models, users, and institutional decisions through traceable evidence. Rather than being limited to the documentation of dataset origins, provenance increasingly encompasses the full lifecycle of data, models, system configurations, prompts, and outputs within complex AI pipelines.

Tracking Data Origins and Transformations

Conceptually, provenance refers to structured information describing the creation, transformation, movement, and use of digital artifacts across computational systems. It captures relationships between entities, activities, and agents involved in data and model processing, thereby enabling the reconstruction of system behavior beyond traditional logging approaches (Herschel et al., 2017). In AI-driven contexts, this extends toward supporting transparency and auditability in end-to-end machine learning and generative AI pipelines, where system behavior must often be reconstructed post hoc rather than deterministically reproduced (Zha et al., 2023).

A clear distinction can be made between data provenance and data lineage. While *(data) provenance* focuses on the origin and transformation history of datasets, *data lineage* emphasizes data flow and dependencies within engineering pipelines. Beyond this distinction, provenance in AI systems also includes information about model training procedures, hyperparameters, evaluation processes, prompt engineering strategies, and inference-time configurations, which are particularly relevant in generative AI systems (Zha et al., 2023).

Provenance in AI Pipelines and Model Outputs

In AI-driven infrastructures, provenance operates across multiple interconnected layers, including the data layer (dataset origin, preprocessing, labeling, versioning), the model layer (training procedures, hyperparameters, evaluation metrics), the interaction layer (prompts, system instructions, retrieval contexts, supporting system components), and the output layer (linking generated results to contributing data and model components). This multi-layered structure is essential because generative models do not produce deterministic outputs but operate in probabilistic and context-dependent ways.

Model updates, stochastic decoding processes, retrieval-augmented architectures, and dynamic system configurations introduce variability that limits strict reproducibility. This challenge has been conceptualized as a *provenance gap*, referring to the limitations of traditional traceability approaches in fully reconstructing generative model behavior (Ahmed et al., 2026).

In a CRIS-based workflow, the provenance gap appears when an AI-generated record cannot be linked unambiguously to the evidence that supported it. For example, a generated project classification may depend simultaneously on a publication abstract, a project description, retrieved external documents, a prompt template, and a proprietary model version. A conventional database log may record only that a field was changed, but not why the model proposed the value or which retrieved sources influenced the result. Closing this gap requires provenance records that connect

source entities, retrieval activities, model configurations, generated suggestions, human review, and the final accepted CRIS record.

In such environments, provenance does not aim to guarantee identical reproduction of outputs but instead enables traceable reconstruction of the conditions under which outputs were generated. This shift from deterministic reproducibility toward contextual reproducibility is a defining characteristic of modern AI systems.

Reproducibility and Traceability in Complex Systems

From a governance perspective, provenance provides the evidential foundation required for accountability, auditability, and compliance. It enables organizations to trace system behavior across interconnected layers of data, models, and processes, thereby transforming governance principles into verifiable mechanisms. Recent research highlights the importance of provenance-based audit trails for accountability in large language model systems (Ojewale et al., 2026).

In scientific and research infrastructures, provenance is essential for ensuring methodological transparency and reproducibility of computational workflows. Standardization initiatives such as the W3C PROV model formalize provenance as a machine-readable representation of entities, activities, and agents, enabling interoperability across systems and domains (W3C Working Group, 2013). However, AI-driven environments require extended provenance models capable of capturing model evolution, prompt histories, and dynamic system configurations (Schlegel et al., 2025).

A key architectural principle is provenance-by-design, in which traceability mechanisms are embedded directly into system architectures rather than added retrospectively. This includes automated logging of data transformations, structured recording of model training and inference processes, and embedding provenance metadata into APIs and AI system interfaces. Recent MLOps research confirms that such embedded provenance systems significantly improve debugging, auditability, and compliance readiness in machine learning pipelines (Schlegel et al., 2025).

Ultimately, provenance constitutes the connective infrastructure between data quality and governance. While data quality ensures input reliability and governance defines system control structures, provenance provides the evidential layer that enables continuous verification of both. In this sense, provenance functions as a core mechanism for ensuring trustworthy, interoperable, and standards-compliant AI systems across digital and scientific ecosystems.

Provenance-by-Design

Together, interoperability-by-design, governance-by-design, privacy-by-design, and provenance-by-design represent complementary architectural principles for trustworthy AI integration. A promising approach for operationalizing provenance in AI-driven environments is the concept of provenance-by-design (Miles et al., 2011, Moreau et al., 2023). Similar to privacy-by-design (Cavoukian, 2009) and security-by-design (Howard & Lipner, 2006) principles, provenance-by-design advocates the systematic integration of traceability mechanisms into system architectures from the earliest stages of development.

This approach includes automated recording of data transformations, model training activities, inference processes, prompt interactions, and system configuration changes. By embedding provenance mechanisms directly into AI infrastructures, organizations can improve auditability, reproducibility, regulatory compliance, and long-term maintainability. Provenance-by-design therefore represents a practical strategy for translating transparency and accountability requirements into operational system capabilities (Waltersdorfer et al., 2025).

In research and scientific infrastructures, these mechanisms become particularly critical, as AI-generated outputs increasingly influence scholarly communication, evaluation, and knowledge production processes. The application of these principles in research environments is discussed in Section 5.

5. AI INTEGRATION IN RESEARCH AND SCIENTIFIC INFRASTRUCTURES

5.1 AI in CRIS and Scholarly Ecosystems

The integration of AI into CRIS and the broader scholarly ecosystem represents one of the most significant developments in the digital transformation of research management. Throughout this section, the term *CRIS* is used following the terminology established by the international euroCRIS community. Consistent with the layered integration framework introduced in Section 3, research institutions increasingly integrate AI into existing CRIS while preserving established infrastructures, governance mechanisms, and information standards. Within this architecture, AI provides analytical services such as metadata enrichment, entity resolution, and decision support without changing the authoritative role of the CRIS.

AI-Supported CRIS Environments

Contemporary CRIS environments are built upon mature standards and interoperability frameworks such as CERIF, Dublin Core, ORCID, ROR, DataCite, and DOI-based identification systems. These standards provide the structural foundation for the exchange and management of research information across institutions and national research ecosystems. Within the layered integration framework, AI services support metadata generation, entity resolution, classification, and semantic enrichment while remaining compatible with established standards and workflows.

For example, after a publication has been imported from Crossref, an NLP model extracts research topics, identifies and links authors to their ORCID identifiers, proposes CERIF classifications, and suggests additional keywords based on the publication content. Rather than updating the CRIS automatically, these recommendations are presented to a research information manager or librarian for review before they are incorporated into the system. This workflow illustrates how AI can enhance metadata quality while preserving existing standards, governance processes, and human oversight.

Metadata Enrichment and Research Analytics

A key application area is metadata enrichment. CRIS increasingly manage large volumes of heterogeneous data including publications, project descriptions, research datasets, funding records, and institutional repositories. Natural language processing and machine learning methods enable the automatic extraction of keywords, topics, research domains, and inter-entity relationships. By linking generated metadata to controlled vocabularies and domain ontologies, such as ACM CCS (Association for Computing Machinery, n.d.), MeSH, or discipline-specific taxonomies, AI-generated metadata contribute to improved discoverability and interoperability of research outputs. The enriched information can be integrated into existing CERIF- or Dublin Core-based infrastructures while maintaining structural consistency.

Beyond metadata processing, AI contributes to research analytics and knowledge discovery. Machine learning and graph-based analytical approaches support the identification of collaboration networks, emerging research topics, funding patterns, and interdisciplinary connections that are difficult to detect through traditional methods. Knowledge graphs based on RDF, OWL, and linked-data principles provide a suitable foundation for integrating AI-generated insights into scholarly ecosystems. By combining structured metadata with semantic relationships, these approaches support the FAIR principles, particularly regarding findability, accessibility, and interoperability of research information (Wilkinson et al., 2016). Beyond institutional CRIS environments, AI-generated metadata and semantic linking mechanisms

increasingly enable interoperability across publishers, research data repositories, funding agencies, and open science infrastructures.

AI-Assisted Knowledge Management

AI also plays an important role in knowledge management and retrieval. The growing volume of scientific literature and research data creates substantial challenges for researchers, administrators, synthesis, and decision-makers. *Retrieval-Augmented Generation (RAG)* systems, semantic search systems, and domain-specific language models support users by summarizing content, answering queries, and identifying relevant research outputs. However, the reliability of such systems strongly depends on the quality and governance of the underlying data sources. Therefore, AI-generated outputs must remain linked to their original sources and be accompanied by provenance information that enables users to verify how results were produced. This requirement underscores the importance of provenance as a core principle in AI-enhanced scholarly ecosystems.

Provenance is particularly important in research contexts because information is frequently used for reporting, evaluation, funding allocation, and strategic decision-making. As discussed in Section 4.3, provenance has evolved into a foundational infrastructure component for ensuring traceability, accountability, and contextual reproducibility in AI-supported environments. Standards such as the W3C PROV model provide mechanisms for documenting datasets, processes, and responsible agents involved in information generation and transformation. Within CRIS environments, provenance can be extended beyond traditional data processing activities to cover AI-specific processes, including metadata enrichment, entity resolution, model-based classification, retrieval contexts in RAG systems, and recommendation generation.

By systematically capturing entities, activities, agents, model versions, prompts, and outputs, CRIS can establish transparent audit trails across the entire AI-supported information lifecycle. This strengthens governance, supports compliance, and enhances trust in AI-generated insights used for institutional reporting and evaluation. Provenance therefore functions as a critical bridge between automated knowledge generation and scientific requirements for transparency, accountability, and reproducibility.

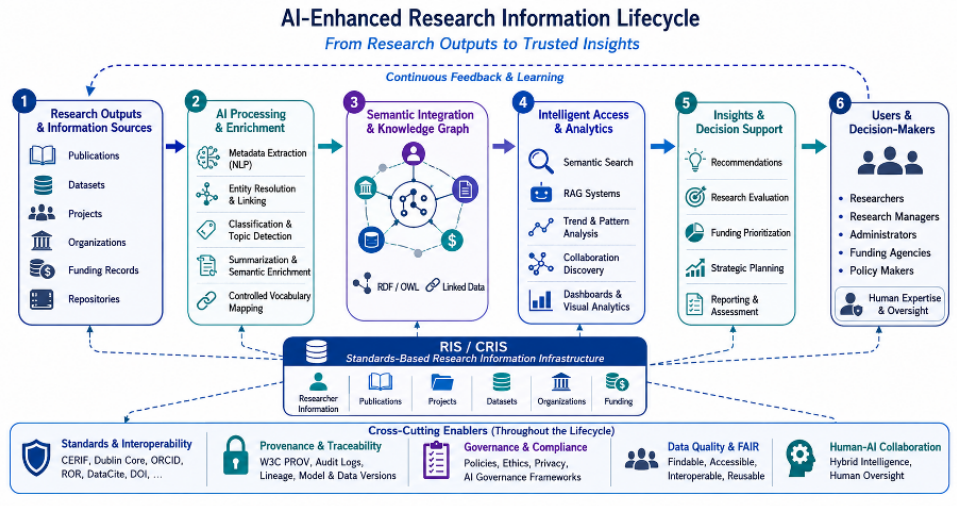
Building on these functional dimensions, AI integration can also be understood at a systemic level as an end-to-end lifecycle that extends from heterogeneous research outputs to trusted knowledge and decision support. As illustrated in Figure 3, research information is progressively transformed through AI-supported enrichment, semantic integration, and analytical processing into structured knowledge representations and actionable insights. Consistent with the integration architecture presented in Section

3, AI operates as a cross-cutting layer that enhances interoperability, discoverability, and analytical depth across the scholarly ecosystem.

This lifecycle perspective is supported by cross-cutting enablers such as research information standards, provenance frameworks, governance mechanisms, and human–AI collaboration principles. These ensure that AI-generated outputs remain interpretable, auditable, and aligned with scientific requirements for transparency, accountability, and methodological rigor. The lifecycle model thus provides a conceptual bridge between infrastructural foundations and the normative requirements of scientific practice discussed in the following sections.

Figure 3 presents the lifecycle in six stages. First, publications, datasets, projects, organizational records, funding information, and repository content enter the research information environment. Second, AI services support enrichment activities such as entity resolution, classification, summarization, and controlled-vocabulary mapping. Third, the enriched records are connected through semantic models, identifiers, and knowledge graphs.

Figure 3. AI-enhanced research information lifecycle. Heterogeneous research records are enriched, semantically connected, analyzed, and transformed into decision-support outputs. The CRIS remains the authoritative standards-based infrastructure, while governance, provenance, data quality, and human oversight operate across the complete lifecycle.



Source: Authors' own illustration.

Fourth, users access the integrated information through semantic search, retrieval-augmented generation, dashboards, and analytical tools. Fifth, these services produce

recommendations and decision-support outputs for reporting, collaboration analysis, research evaluation, and strategic planning. Sixth, researchers, research managers, administrators, funding organizations, and policymakers review and use these outputs.

The CRIS remains the standards-based institutional core throughout the lifecycle. Provenance, data quality controls, governance, and human oversight operate across all six stages rather than appearing only at the final decision point.

5.2 Preserving Scientific Standards

The increasing integration of artificial intelligence into research processes raises fundamental questions regarding the preservation of scientific standards. While AI offers substantial opportunities to accelerate data analysis, automate routine tasks, and enhance knowledge discovery, its use must not compromise the methodological foundations of scientific inquiry. Scientific validity depends on rigor, transparency, validation, and reproducibility (Munafò et al., 2017), which remain essential regardless of technological context.

Methodological Rigor

Methodological rigor remains a fundamental requirement of scientific work, whether research is conducted by humans, machines, or hybrid systems. AI-supported methods should therefore be understood as extensions of established research methodologies rather than as substitutes for scientific reasoning. This perspective aligns with broader concepts of hybrid intelligence, in which AI augments human expertise while researchers retain responsibility for methodological design and interpretation (Dellermann et al., 2019).

The increasing use of machine learning models, large language models, and automated analytical systems introduces additional methodological considerations. Researchers must carefully evaluate training data quality, model selection criteria, parameter configurations, and potential sources of bias (Mehrabi et al., 2022; Sambasivan et al., 2021; Kumar et al., 2024). Similar to traditional statistical methods, AI-based approaches require explicit documentation of assumptions, limitations, and uncertainties. The use of advanced AI techniques does not eliminate the need for rigorous research design; instead, it often increases the importance of methodological oversight due to the complexity and opacity of many AI systems.

Particularly in generative AI applications, methodological rigor requires researchers to critically assess AI-generated outputs before incorporating them into scientific workflows. Literature summaries, data classifications, coding suggestions, or generated hypotheses should be treated as analytical support rather than authoritative scientific results. The responsibility for methodological soundness

remains with researchers, who must evaluate whether AI-supported outputs align with established disciplinary standards, research objectives, and accepted scientific practices (Dwivedi et al., 2023).

Transparency and Validation

Transparency constitutes a central mechanism for maintaining trust in AI-supported research. Scientific knowledge relies on the ability of other researchers to understand how conclusions were reached and which methods contributed to the generation of results. As AI systems increasingly participate in data processing, analysis, interpretation, and content generation, transparency must extend beyond traditional methodological reporting to include information about the AI components involved in research activities (Floridi et al., 2018).

Transparent AI-supported research requires documentation of datasets, model versions, training procedures, prompts, retrieval mechanisms, parameter settings, and validation processes (Floridi et al., 2018). Such documentation enables researchers, reviewers, and other stakeholders to assess the reliability and limitations of AI-generated results. In this context, transparency is closely linked to the provenance principles discussed in Section 4.3, which provide structured mechanisms for documenting the relationships between data, models, processes, and outputs (Belhajjame et al., 2013; Herschel et al., 2017). Emerging documentation frameworks such as model cards and datasheets further support transparency by providing standardized descriptions of models, datasets, intended use cases, and known limitations (Mitchell et al., 2019; Pushkarna et al., 2022). Standardized dataset documentation can further record dataset composition, collection procedures, recommended uses, limitations, and potential biases (Gebru et al., 2021).

Validation remains equally important. AI-generated results should be subjected to the same critical scrutiny as any other scientific output (Bommasani et al., 2021; Bender et al., 2021). Automated classifications, predictions, recommendations, or generated content require systematic verification through expert review, benchmarking, statistical evaluation, or comparison with established reference data. Human validation is particularly important in domains where AI systems may produce plausible but incorrect outputs, commonly referred to as hallucinations in generative AI systems.

A practical approach adopted by many research organizations involves multi-stage validation processes in which AI-generated outputs are first assessed automatically through consistency checks and subsequently reviewed by domain experts. Such hybrid validation mechanisms combine computational efficiency with scientific judgment and reflect broader human-centered AI approaches that emphasize col-

laboration between human expertise and machine intelligence (Dellermann et al., 2019; Shneiderman, 2022).

Reproducibility in AI-Supported Research

Reproducibility has long been regarded as a cornerstone of scientific practice. Research findings gain credibility when independent researchers can verify results using the same methods and evidence (Munafò et al., 2017; Freese et al., 2022). However, AI-supported research introduces new challenges for reproducibility, particularly in the context of machine learning and generative AI systems.

Unlike many traditional computational methods, contemporary AI systems often exhibit non-deterministic behavior. Variations in training data, model updates, parameter settings, stochastic inference processes, retrieval contexts, or prompt formulations can influence generated outputs (Bommasani et al., 2021; Bender et al., 2021). As discussed in Section 4.3, strict replication of AI-generated results may therefore not always be achievable. This does not imply that scientific reproducibility becomes unattainable; rather, it requires an expanded understanding of reproducibility that incorporates the contextual conditions under which AI-supported results are produced. As a result, recent discussions increasingly distinguish between computational reproducibility, which focuses on documenting and reconstructing the computational environment and workflow (Peng, 2011; Munafò et al., 2017), functional reproducibility, which emphasizes the ability to obtain scientifically consistent conclusions despite differences in implementation or system configuration (Freese et al., 2022; Munafò et al., 2017), and semantic reproducibility, which is increasingly discussed in the context of generative AI systems and refers to the preservation of meaning or scientific interpretation even when outputs are not identical at the surface level (Bommasani et al., 2021). Together, these concepts reflect a broader understanding of reproducibility that extends beyond exact output replication toward the reproducibility of methods, evidence, and scientifically valid conclusions.

To address these challenges, research infrastructures increasingly adopt practices such as model versioning, dataset versioning, prompt logging, workflow documentation, and provenance recording (Schlegel et al., 2025; Ahmed et al., 2026; Ojewale et al., 2026). These mechanisms allow researchers to reconstruct the conditions under which analyses were performed and outputs were generated, even when identical reproduction of results is not possible. Provenance-by-design approaches are particularly relevant because they embed traceability directly into research workflows and AI infrastructures, thereby supporting transparency, auditability, and methodological accountability.

Furthermore, open science practices provide important support for reproducibility in AI-enabled research environments. The publication of datasets, source code,

model documentation, evaluation procedures, and methodological protocols enables independent verification and fosters transparency across the research lifecycle. The FAIR principles provide a widely accepted framework for improving the accessibility and reusability of research outputs (Wilkinson et al., 2016). Emerging standards for machine learning documentation, such as model cards and datasheets for datasets, further contribute to making AI-supported research more understandable, comparable, and reproducible

Ultimately, preserving scientific standards requires implementing the layered integration approach through governance, provenance, validation, and human oversight. Under these conditions, AI can strengthen rather than undermine scientific reliability. Methodological rigor, transparency, validation, and reproducibility remain essential foundations of trustworthy research. When supported by appropriate governance mechanisms, provenance frameworks, human oversight, and emerging standards for responsible AI, AI can strengthen these principles and contribute to a more efficient, transparent, and accountable scientific ecosystem

5.3 Responsible AI in Scientific Decision-Making

As artificial intelligence becomes increasingly integrated into research infrastructures, its role extends beyond information management and analytical support toward influencing decision-making processes. Research institutions, funding organizations, publishers, and policy makers increasingly employ AI-supported systems for evaluating research performance, identifying emerging trends, allocating resources, and supporting strategic planning. While these applications offer significant opportunities for improving efficiency and analytical capacity, they also raise important questions regarding accountability, fairness, transparency, and scientific integrity. Within scientific decision-making, the governance mechanisms introduced earlier ensure that AI recommendations remain subject to documented human review and institutional accountability.

Human experts remain essential for interpreting context, resolving ambiguity, validating outputs and taking responsibility for critical decisions. Hybrid intelligence therefore requires more than simply adding human oversight, but appropriate workflow designs for the task at hand.

Risks of Automation in Research Evaluation

One of the most significant challenges associated with AI adoption in research environments concerns the increasing automation of evaluation processes. Research assessment traditionally involves complex judgments regarding scientific quality, societal relevance, methodological rigor, innovation potential, and disciplinary

impact. These dimensions are difficult to quantify and often depend on contextual factors that extend beyond measurable indicators.

AI systems can process large volumes of publication data, citation networks, funding records, collaboration patterns, and institutional metrics more efficiently than human evaluators. Consequently, they are increasingly used to support activities such as grant pre-screening, institutional benchmarking, reviewer selection, trend analysis, and performance monitoring. For example, an AI system may recommend reviewers for grant applications by analysing publication networks, research expertise, and potential conflicts of interest. These recommendations are subsequently reviewed by a programme officer before reviewer invitations are issued, ensuring that AI supports rather than replaces human judgment. However, excessive reliance on automated assessments may introduce systematic risks that undermine the quality and fairness of scientific evaluation.

A central concern is the amplification of existing biases embedded within historical research data. This phenomenon is often described as algorithmic bias or biased learning, where machine learning systems reproduce patterns contained in their training data rather than independently assessing scientific quality. Models trained on publication records, citation databases, or funding histories may therefore perpetuate structural inequalities related to geography, language, institutional prestige, gender, or disciplinary representation (Mehrabi et al. (2022).

Furthermore, quantitative indicators often fail to capture essential aspects of scientific quality. Citation counts, publication volumes, journal metrics, and network centrality measures may provide useful signals but cannot fully represent originality, methodological robustness, societal relevance, or long-term scientific contribution. An overreliance on automated metrics risks encouraging optimization toward measurable outputs while neglecting less quantifiable dimensions of research quality. Similar concerns have long been raised regarding metric-based research evaluation (Hicks et al., 2015) and are further intensified when AI systems automate the interpretation of such indicators.

More recent reform initiatives in research assessment similarly caution against the inappropriate use of narrow quantitative indicators (CoARA, 2022). The Agreement on Reforming Research Assessment emphasizes qualitative judgment, diversity of research contributions, transparency, and responsible use of metrics. These principles are directly relevant to AI-supported assessment systems, which may otherwise automate and amplify limitations already present in conventional indicator-based evaluation.

Generative AI introduces additional challenges because AI-generated recommendations or assessments may appear authoritative despite being based on incomplete information or flawed assumptions. Without appropriate validation mechanisms, decision-makers may attribute excessive credibility to AI-generated outputs, a

phenomenon sometimes described as automation bias. Responsible AI adoption therefore requires explicit recognition that AI-generated recommendations constitute decision support rather than decision authority.

Explainable AI for Academia

To address these concerns, explainability has emerged as a key requirement for AI systems operating within scientific environments. *Explainable AI (XAI)* seeks to make the behavior, reasoning processes, and outputs of AI systems understandable to human users. In academic contexts, explainability is particularly important because scientific decision-making depends upon justification, peer review, and critical evaluation rather than solely on predictive accuracy.

Research administrators, funding agencies, ethics committees, and evaluation panels must be able to understand why a system generated a particular recommendation, ranking, classification, or prediction. Black-box models that provide outputs without meaningful explanations may create significant challenges for accountability and trust. As a result, increasing attention is being devoted to developing explainability mechanisms that can communicate relevant factors influencing AI-supported decisions.

Explainability in research environments can take multiple forms. Feature importance analyses, model interpretability techniques, confidence scores, provenance information, and transparent documentation of training datasets all contribute to improving the understandability of AI-generated outputs. Particularly in high-impact contexts such as funding allocation, hiring decisions, promotion procedures, or research assessment exercises, explainability enables stakeholders to critically evaluate whether AI-generated recommendations are justified and consistent with institutional objectives.

The importance of explainability extends beyond technical transparency. It also supports procedural fairness by allowing affected individuals to understand how decisions were influenced and to challenge or review potentially problematic outcomes. This aligns with broader governance principles emphasizing transparency, accountability, and contestability in AI-supported systems (Floridi et al., 2019; European Commission, 2024).

Human Oversight and Hybrid Intelligence Models

In the context of scientific decision-making, human oversight operationalizes the governance layer by ensuring that AI-generated recommendations are reviewed, documented, and formally approved before influencing institutional decisions.

This perspective is consistent with the *concept of hybrid intelligence*, which emphasizes the collaborative interaction between human and artificial intelligence systems. Hybrid intelligence models recognize that humans and AI possess complementary strengths. AI systems excel at processing large datasets, identifying complex patterns, and generating analytical insights, whereas humans contribute contextual understanding, ethical reasoning, creativity, and domain-specific judgment. Combining these capabilities allows organizations to benefit from automation while maintaining human responsibility for critical decisions.

In research environments, hybrid intelligence can be operationalized through human-in-the-loop and human-on-the-loop governance models (European Commission, 2019). Human-in-the-loop approaches require direct human review and approval before AI-generated recommendations influence decisions, whereas human-on-the-loop approaches permit greater automation while maintaining human supervision and intervention capabilities. These approaches reflect broader concepts of human–automation interaction and supervisory control that have long been discussed in automation research (Parasuraman et al., 2000). The appropriate model depends on the significance and potential impact of the decision under consideration. High-stakes activities such as grant allocation, promotion decisions, ethics review, or strategic research prioritization typically require stronger forms of human oversight than routine administrative processes.

Effective oversight also requires institutional mechanisms that define responsibilities, escalation procedures, and accountability structures for AI-supported decision-making. Governance frameworks such as ISO/IEC 42001 and emerging regulatory initiatives, including the EU AI Act, increasingly emphasize the importance of human oversight as a core principle of trustworthy AI deployment. Within scientific institutions, these requirements translate into organizational practices that ensure AI-generated recommendations remain subject to expert review, transparent documentation, and continuous evaluation.

Ultimately, responsible AI in scientific decision-making is not achieved through technical solutions alone. It requires a balanced combination of governance, transparency, explainability, and human judgment. By embedding AI within existing scientific norms and decision-making structures, research organizations can leverage the analytical capabilities of AI while preserving accountability, fairness, and methodological integrity. In this way, AI becomes a tool for augmenting scientific

expertise rather than replacing the human responsibility that remains central to trustworthy research and evidence-based decision-making.

These challenges highlight that responsible AI in scientific environments cannot be achieved through conceptual principles alone. Instead, they require concrete implementation strategies that translate standards, governance frameworks, and provenance mechanisms into operational system designs. Such strategies are discussed in Section 6.

The preceding sections identify the substantive requirements for responsible AI integration in research infrastructures: standardized data, documented provenance, privacy safeguards, reproducible workflows, human validation, and accountable decision structures. The next section translates these requirements into implementation stages, responsible roles, technical controls, and evidence that institutions can use when designing or procuring AI-supported CRIS services.

6. PRACTICAL APPROACHES AND INTEGRATION STRATEGIES

This section translates the conceptual framework into practical implementation guidance. Standards-compliant AI integration requires coordinated action across technical, organizational, legal, ethical, and scientific dimensions. Technical controls establish interoperable interfaces, versioning, logging, provenance, monitoring, and evaluation. Organizational controls assign ownership and decision authority. Legal controls address applicable regulatory duties, contractual obligations, data protection, and documentation requirements. Ethical and scientific controls ensure that AI-supported processes remain proportionate, fair, transparent, and methodologically defensible.

6.1 Frameworks for Standards-Compliant AI Integration

To adopt AI models into internal processes, different approaches may be chosen depending on the organisational readiness, risk exposure and internal AI expertise: from experimental tests, to proof of concepts, operational integration, scaling-up and finally organization-wide governed application of AI. Existing frameworks such as the NIST AI Risk Management Framework, ISO/IEC 42001 and responsible AI maturity models provide useful foundations to start such an integration process. Nevertheless, a major gap is the translation of policy statements into both technical and organisational matters, taking into account AI risks and opportunities that come

with different strategies (choice of vendor, definition of system and user scope, technical robustness and safety).

AI systems should support interoperability across technical, semantic, procedural, and organizational levels. Technical interoperability requires stable APIs, structured and machine-readable data, documented interfaces, and compatible authentication mechanisms. Semantic interoperability requires shared identifiers, vocabularies, metadata schemas, and mapping rules. Procedural interoperability requires agreed workflows for validation, change management, exception handling, and quality assurance. Organizational interoperability requires defined responsibilities, access rights, escalation procedures, and accountability structures.

Enabling mechanisms include the use of common vocabularies, metadata, versioning, provenance, and standards to document and integrate all these viewpoints in a harmonised way on the semantic level. Procedurally, this requires agreed workflows for data processes, ranging from exchange, change management, documentation, and quality validation. Organisationally, clear responsibilities, compliance requirements, accountability structures and access rights applying standards, best practices and regulations. A related development is the growing relevance of Digital Humanism, emphasising the integration and alignment of human values, democratic principles, participation and a broader benefit to society (Werthner et al., 2022). In the interoperability context, this perspective expands the discussion from a purely technical matter to a social-technical lens.

6.2 Organizational and Technical Best Practices

Organizations should begin with well-scoped and comparatively low-risk use cases. Use cases should be assessed against criteria including expected organizational value, data sensitivity, estimated error impact, technical complexity, availability of human oversight and how well performance can be objectively measured. In this context, suitable entry points mean cases that combine clear and measurable advantages with limited legal, operational and ethical risk, where outputs can be reviewed and revised before inflecting real consequences. Typical examples include semantic search, keyword extraction, metadata suggestions, document classification, and non-sensitive summarization. Before moving toward decision support, automated recommendations, or sensitive-data processing, institutions should establish such evaluation criteria, approval procedures, monitoring mechanisms, and responsible ownership.

Because AI integration affects technical architecture, domain practices, legal compliance, information governance, cybersecurity, and ethics, institutions should establish cross-functional governance teams. These teams should include represen-

tatives from IT, research information management, data stewardship, legal and data protection functions, information security, domain experts, and affected user groups.

Continuous activities, such as AI tool registries, use case rules, impact assessments and evaluations support the ongoing monitoring of AI systems to ensure accountability, transparency and participation. Continuous monitoring and auditing are essential after deployment. They provide evidence about system performance, changing data distributions, recurring errors, security incidents, user behavior, and unintended consequences. Monitoring results should feed into model updates, risk reassessments, user guidance, and the institution's broader AI strategy.

Static and one-off audits are not sufficient, since the performance and risk model of modern AI systems may change over time due to model updates, access rights, changes in the underlying data distribution or also to the revision of relevant standards or regulation.In this context, continuous monitoring and auditing is best understood as a multi-layered activity.

Technically, logging and monitoring tools, evaluation metrics and traces provide the evidence basis. The method and tool landscape for continuous monitoring and auditing is evolving quickly, from MLOps and AI observability tools, including Mlflow (Mlflow, n.d.) to implementation-specific frameworks such as AutoPyTorch (AutoML Group, n.d.), AutoKeras (Autokeras, n.d.), workflow tools including Airflow (Apache Software Foundation, n.d.). Procedurally, regular monitoring cycles, thresholds, harm mitigation methods and routines need to be established. Organisationally, clear distribution of tasks and ownership across relevant artefacts, including models, data, risks are needed. The legal and ethical perspectives are needed to ensure that monitoring supports relevant values.

From an MLOps perspective, institutions should record model versions, dataset versions, code dependencies, evaluation results, approval decisions, and deployment configurations. Every production model should have an identified owner, documented intended use, defined performance thresholds, and a rollback procedure. These controls reduce hidden technical debt and make changes auditable over time (Sculley et al., 2015).

However, in a standards context, technical tools are not sufficient on their own. They must be combined with domain knowledge and appropriate governance controls for bias, provenance, security, human oversight and conformity assessment. They cannot rely on generic approaches, but have to be tied to the application-specific workflow and failure modes and process outcomes in order to be meaningful. This is especially important with LLMs or agents, for which results may look acceptable on first glance even when internal processes, such as retrieval, reasoning, generation might be incomplete, wrong or inconsistent or even dangerous (Ruan et al, 2024), leading to downstream risks and harms. That is why traces, system context and intermediate outputs are important and need to be analysed.

That is why error analysis and monitoring should not be regarded as only a technical process, but a governance mechanism to understand recurring weaknesses, clarify responsibilities or adapt safeguards and other procedures. Outsourcing of these processes may provide good results at the beginning but harms the outcome in the long term, as well as the AI-readiness of the organisation, since error analysis is critical to understand emerging issues and challenges.

Proportional Implementation for Smaller Organizations

These requirements also raise the question of proportionality: not every organization will have the resources or organizational structures required to implement the full set of technical and governance controls described above.

The organizational requirements described above may be difficult to implement in smaller institutions that lack dedicated AI governance, data stewardship, information security, or legal functions. A fully developed governance structure with multiple specialized roles may therefore be disproportionate for low-risk use cases or organizations with limited resources. The framework should consequently be implemented according to the principle of proportionality. Smaller organizations may initially rely on a limited set of essential controls, such as documented use-case ownership, source-linked outputs, basic model and prompt versioning, human validation, and incident escalation procedures. More advanced capabilities, including automated provenance infrastructures, continuous model monitoring, dedicated governance boards, or comprehensive MLOps environments, can be introduced as organizational maturity, AI usage, and risk exposure increase.

This suggests that standards-compliant AI integration should not be equated with maximal implementation complexity. The objective is rather to establish a level of governance and technical control that is appropriate to the potential impact and risk of the respective AI application. Such a scalable approach can make the framework more accessible to organizations with different levels of technical and organizational capacity while preserving its core principles of interoperability, accountability, provenance, and human oversight.

6.3 Lifecycle Matrix for Standards-Compliant AI Integration

The preceding sections have shown that standards-compliant AI integration cannot be reduced to a single technical or governance decision. Requirements emerge throughout the complete AI lifecycle, from the initial definition of a use case to the eventual retirement or replacement of the system. Each lifecycle stage introduces

specific risks, requires different controls, produces distinct forms of evidence, and involves different organizational actors.

Table 2 translates the conceptual framework developed in this chapter into a practical lifecycle matrix. It links seven stages of AI integration—use-case definition and assessment, data preparation, model selection, development and testing, deployment, operation and monitoring, and retirement or replacement—to the risks, standards, controls, documentation requirements, and responsibilities associated with each stage. The matrix does not assume that one standard governs the complete lifecycle. Instead, it demonstrates how complementary frameworks can be combined according to their respective functions.

ISO/IEC 42001 and the NIST AI Risk Management Framework provide overarching structures for organizational governance and risk management. The EU AI Act contributes legally defined obligations and risk-based requirements. FAIR, CERIF, metadata standards, and W3C PROV support data stewardship, semantic interoperability, and traceability. MLOps practices provide mechanisms for model and dataset versioning, controlled deployment, monitoring, rollback, and technical maintenance. The matrix integrates these perspectives by connecting normative requirements with operational controls and verifiable evidence.

A central feature of the matrix is the explicit assignment of responsibility. Standards and automated controls do not by themselves ensure accountable AI use. Every lifecycle stage therefore requires clearly identified actors who are responsible for approving decisions, maintaining documentation, evaluating risks, and responding to failures. In CRIS environments, these responsibilities are distributed across business owners, research information managers, data stewards, domain experts, system architects, data protection and information security officers, AI development teams, and institutional governance bodies.

The matrix should be understood as an adaptable implementation instrument rather than as a universal compliance checklist. Institutions may adjust the individual controls and responsible roles according to the intended use, level of automation, data sensitivity, institutional context, and potential impact of the AI-supported process. Higher-risk applications, particularly those influencing research evaluation, funding decisions, recruitment, or sensitive personal data, require stronger documentation, validation, human oversight, and escalation mechanisms than low-risk applications such as assisted search or metadata suggestions.

Table 2. Lifecycle Matrix Linking AI Integration Stages to Risks, Standards, Controls, Evidence, and Responsible Actors

Lifecycle stage	Main risks	Relevant standards/ frameworks	Required controls	Evidence	Responsible actors
Use-case definition and assessment	unclear purpose, excessive, high-risk automation	NIST AI RMF, ISO/IEC 42001, EU AI Act	purpose definition, risk classification and mitigations, stakeholder review	approved, suitable use-case record	business owner, CRIS manager, governance team
Data preparation	poor quality, privacy breaches, bias	FAIR, GDPR, ISO/IEC 5259	data-quality checks, minimization, access controls, documentation	data sheet, quality report, legal basis	data steward, data protection officer
Model selection	vendor lock-in, opacity, unsuitable model	ISO/IEC 42001, NIST AI RMF	model comparison, exit strategy, performance and transparency criteria	model assessment	system architect, procurement, domain expert
Development and testing	unreliable outputs, undocumented changes	MLOps practices, W3C PROV	versioning, test datasets, prompt documentation, evaluation thresholds	model card, test report, provenance record	AI team, domain experts
Deployment	uncontrolled integration, security risks	ISO/IEC 42001, EU AI Act	API controls, human approval, role-based access, logging	deployment approval, architecture documentation	IT operations, security, system owner
Operation and monitoring	model drift, hallucinations, recurring bias	NIST AI RMF, provenance standards	monitoring, audit trails, incident management, periodic review	monitoring reports, incident records	model owner, governance team
Retirement or replacement	loss of records, vendor dependency	archival and interoperability standards	export, retention, migration, decommissioning plan	retirement record, archived documentation, data export, migration or decommission-ing report	system owner, records manager, IT operations, governance team

Table 2 illustrates that governance must remain continuous rather than being concentrated at the beginning or end of an AI project. Decisions made during use-case definition influence the appropriateness of later data, model, and deployment choices. Similarly, evidence generated during development, testing, and operation enables institutions to reassess risks and determine whether the system may continue to operate under its existing conditions.

The lifecycle stages are therefore interdependent and partly iterative. Monitoring may reveal data drift, systematic errors, new legal requirements, or changes in institutional objectives that require a return to earlier stages. A model may need to be re-evaluated, replaced, or restricted to a narrower scope. In this sense, the matrix operationalizes governance-by-design, interoperability-by-design, privacy-by-design, and provenance-by-design as continuing practices across the AI lifecycle rather than as one-time implementation activities.

6.4 Implementation Examples for Standards-Compliant AI Integration

While the preceding sections have outlined conceptual principles and governance requirements, their practical value becomes more apparent when translated into concrete implementation scenarios. The following examples illustrate how organizations can integrate AI into existing standards-based infrastructures while preserving interoperability, governance, transparency, and accountability. Rather than representing exhaustive case studies, these scenarios demonstrate typical implementation patterns that can be adapted across different organizational contexts.

Example 1: AI-Assisted Metadata Enrichment in Research Information Systems

A university operating a CRIS seeks to improve the quality and completeness of publication metadata. Instead of replacing the existing CERIF-based infrastructure, an AI-based natural language processing service is integrated through standardized APIs. After new publications are imported from institutional repositories or publishers, the AI system automatically extracts keywords, research topics, funding information, and potential links to existing research projects.

Rather than writing directly into the production database, the generated metadata are presented to research administrators for validation before publication. The AI-generated suggestions are stored together with provenance information, including the model version, generation timestamp, confidence scores, and reviewer decisions. Existing metadata standards such as CERIF, Dublin Core, ORCID, and persistent identifiers remain unchanged, while AI functions solely as an additional enrichment layer. This approach improves metadata quality without compromising interoperability or governance.

- Relevant standards: CERIF, Dublin Core, ORCID, FAIR, W3C PROV.
- Main control: mandatory validation before write-back.

- Evidence: model version, source record, confidence score, generated suggestion, reviewer decision.
- Responsible actor: research information manager or librarian.

Example 2: Retrieval-Augmented Generation for Institutional Knowledge Management

A public research organization introduces an internal Retrieval-Augmented Generation (RAG) system to support administrative staff in answering questions regarding funding regulations, internal policies, and research procedures. Rather than relying solely on the language model's internal knowledge, every generated response is based on documents retrieved from the organization's controlled knowledge base.

The implementation combines semantic search with provenance-by-design principles. Each generated answer contains references to the underlying policy documents, while the system records retrieval sources, prompts, model versions, and user interactions for auditing purposes. Human experts remain responsible for validating policy updates and maintaining the knowledge repository. This architecture improves efficiency while ensuring that AI-generated recommendations remain transparent, verifiable, and aligned with institutional governance requirements.

- Relevant standards and frameworks: W3C PROV, institutional information governance, data protection requirements.
- Main control: answers must remain linked to retrieved institutional sources.
- Evidence: retrieved documents, prompt, model version, response, user feedback.
- Responsible actor: knowledge-base owner and relevant policy expert.

Example 3: AI-Supported Research Evaluation with Human Oversight

A national funding agency employs machine learning techniques to assist reviewers during the initial screening of research proposals. The AI system identifies relevant scientific disciplines, suggests suitable reviewers based on expertise networks, and highlights potential conflicts of interest. However, it does not rank proposals or make funding decisions autonomously.

All recommendations are presented together with explanatory information, confidence measures, and links to the underlying data sources. Human reviewers retain full authority for evaluating scientific quality and making funding decisions. Governance mechanisms require that AI-supported recommendations, reviewer interactions, and decision processes are documented through comprehensive audit

trails. This hybrid intelligence approach combines computational efficiency with scientific accountability while reducing the risks associated with automation bias and opaque decision-making.

- Relevant frameworks: ISO/IEC 42001, NIST AI RMF, EU AI Act, responsible research assessment principles.
- Main control: AI may support reviewer matching but must not autonomously determine funding outcomes.
- Evidence: recommendation rationale, conflict-of-interest check, human review, final decision.
- Responsible actor: programme officer and evaluation panel.

6.5 Practitioner's Checklist for CRIS Managers

Before introducing an AI component into a CRIS, institutions should verify:

1. Is the intended use clearly defined and proportionate?
2. Is the authoritative source system identified?
3. Are the relevant metadata schema and persistent identifiers defined?
4. Can every generated output be linked to its input sources?
5. Are model, dataset, prompt, and configuration versions recorded?
6. Is personal or sensitive information minimized and protected?
7. Are multilingual performance and potential bias evaluated?
8. Is human review required before institutional records are changed?
9. Are monitoring thresholds and incident procedures defined?
10. Can the institution export its data, logs, and configurations if the provider changes?
11. Are environmental and operational costs proportionate to the use case?
12. Is there a responsible owner for the complete AI-supported workflow?

6.6 Emerging Standards and Future Developments

Standards are currently being extended, revised and newly created to address the evolving landscape of AI applications. Existing standardization bodies increasingly adopt regulatory frameworks and other governance principles in response to the developments in this field.

Together, the development and publishing of extended frameworks suggest a shift from general principles toward more concrete guidelines covering accountability, risk management, documentation, transparency, and oversight.

The EU AI Act is a risk-based regulation, requiring documentation, training, human oversight, and conformity-related obligations for higher-risk AI systems. The Act entered into force in 2024, while its obligations came into force in multiple phases staged over time. In parallel, international standards such as ISO/IEC 42001 and IEEE's AI ethics and governance standards support the translation of abstract principles into more technical guidance and management processes. The FAIR principles, while universal to data in general, are also central because AI-enabled standardization depends on high-quality, well-described, machine-actionable data and metadata. Without this foundation, AI tools may suffer from the aforementioned challenges and risks.

AI governance standards play an essential part in the operationalisation of trustworthy AI. They support the definition of acceptable uses, classification of risks, management of data, assignment of accountability, documentation of decisions, monitoring of performance, and audit outcomes. While these emerging frameworks provide important guidance for implementation, they also reveal persistent tensions between innovation, governance, and interoperability. These tensions are discussed in Section 7.

7. DISCUSSION

The conceptual analysis presented in this chapter suggests that sustainable AI integration depends on more than model performance or technological innovation. It requires the alignment of AI capabilities with existing standards, institutional responsibilities, provenance mechanisms, and validation practices.

One of the central challenges concerns the balance between innovation and control. AI technologies evolve rapidly, often introducing new functionalities before organizations and regulators have established appropriate governance mechanisms. While excessive regulation may inhibit innovation, insufficient governance may create risks related to transparency, accountability, reliability, and trust. The challenge for organizations is therefore not to choose between innovation and governance, but to develop approaches that allow both objectives to coexist.

A second tension arises between transparency and performance. Many of the most powerful AI systems currently available are proprietary models that offer limited visibility into training data, model architectures, and fine-tuning processes. Although such systems often provide superior performance, they may create challenges for explainability, auditability, and regulatory compliance. Conversely, open-source and open-weight models can improve transparency and technological sovereignty but may not always provide equivalent performance or operational support. Organizations

must therefore evaluate AI technologies not only on technical performance but also on governance, transparency, and long-term sustainability criteria.

The discussion also highlights the growing importance of provenance as a foundational mechanism for trustworthy AI ecosystems. As AI systems become increasingly adaptive, autonomous, and integrated into decision-making processes, provenance provides the evidential basis required for accountability and contextual reproducibility. Future AI infrastructures are likely to depend increasingly on provenance-by-design principles that enable organizations to reconstruct data flows, model behavior, and decision processes across complex socio-technical environments.

Particular challenges emerge in scientific and research infrastructures. AI offers substantial opportunities for improving research information management, metadata enrichment, knowledge discovery, and scholarly analytics. At the same time, excessive automation may conflict with established scientific principles such as methodological rigor, transparency, peer review, and critical evaluation. The framework indicates that hybrid intelligence approaches, combining AI capabilities with human expertise and oversight, provide the most promising pathway for responsible AI adoption in research environments.

Finally, the chapter highlights that AI governance should be understood as a continuous organizational capability rather than a one-time compliance activity. As AI technologies continue to evolve toward increasingly autonomous and agent-based systems, organizations will require adaptive governance frameworks capable of managing new forms of complexity, uncertainty, and risk. Future standards will therefore need to address not only technical interoperability but also issues of explainability, accountability, digital sovereignty, and lifecycle governance.

Overall, the discussion suggests that the long-term success of AI adoption depends less on the sophistication of individual models and more on the ability to integrate AI into trustworthy, interoperable, and standards-based ecosystems. Organizations that successfully combine innovation with governance will be better positioned to realize the benefits of AI while maintaining reliability, transparency, and institutional trust.

The framework also shows that standards-compliant integration must extend beyond conventional technical and governance concerns. Privacy-by-design is required where CRIS data include personal or sensitive information. Multilingual evaluation is necessary to prevent language-dependent metadata and assessment errors. Environmental proportionality should be considered when institutions select between large general-purpose models and smaller task-specific systems. These dimensions should be treated as cross-cutting governance concerns rather than isolated technical issues.

7.1 Limitations and Scope of Applicability

This chapter presents a conceptual integration framework rather than an empirically validated implementation model. The framework was developed through the synthesis of selected standards and literature and has not yet been evaluated through comparative case studies across institutions. Consequently, the proposed architecture should not be interpreted as a universally applicable implementation blueprint or as evidence that all organizations can adopt the full set of proposed controls with comparable effort and outcomes.

A first limitation concerns the transferability of the framework beyond the context of a CRIS. The framework is grounded in research information infrastructures, where interoperability, metadata quality, provenance, persistent identifiers, scientific reproducibility, and human validation are particularly important. Although the four-layer architecture is conceptually applicable to other standards-based information environments, the specific standards, governance requirements, validation mechanisms, and evidence structures will differ substantially across domains. For example, healthcare, public administration, and business environments may involve different regulatory requirements, data models, risk profiles, and accountability structures. The transferability of the framework should therefore be understood at the level of architectural principles rather than through a direct replication of the CRIS-specific implementation patterns.

A second limitation involves organizational capacity and the effort required for implementation. The full implementation of the proposed approach requires access to interoperable information systems, data stewardship capabilities, governance structures, and AI expertise, as well as a certain level of technical maturity. Implementing versioning, provenance tracking, monitoring, human validation, audit trails, and cross-functional governance can introduce considerable technical, organizational, and financial overhead. This may be particularly challenging for smaller institutions or organizations with limited IT resources, specialized personnel, or dedicated AI governance functions. Therefore, a standards-compliant architecture should not imply that every organization must implement all controls with the same level of complexity. Rather, implementation should be proportional to the risk, criticality, sensitivity, and expected organizational value of the respective AI use case.

This limitation also suggests that the framework should be understood as a scalable architecture rather than a fixed implementation package. Organizations with limited resources may initially apply a reduced set of controls to low-risk use cases, such as metadata suggestions or semantic search, and progressively introduce more comprehensive provenance, monitoring, validation, and governance mechanisms as the scope and risk of AI adoption increase. Such proportional implementation is consistent with the chapter's emphasis on incremental adoption, but requires further

empirical investigation to determine which controls are essential at different levels of organizational maturity.

An important tension therefore arises between governance completeness and implementation burden. More comprehensive implementation of provenance, monitoring, versioning, human validation, and cross-functional governance may improve accountability and traceability, but it also increases the technical and organizational effort required. Excessive control requirements may therefore hinder AI adoption, particularly in smaller or less mature organizations. This reinforces the need for risk-proportionate and scalable implementation rather than a uniform governance architecture.

Finally, the framework's effectiveness remains to be demonstrated empirically. Future research should therefore evaluate its implementation effort, organizational acceptance, governance effectiveness, and measurable impact across institutions of different sizes and levels of technical maturity. Comparative studies should also examine whether the proposed architectural principles remain effective when transferred from CRIS to other standards-based information infrastructures. In addition, standards and regulatory requirements continue to evolve. The framework should therefore be interpreted as an adaptable architecture rather than a fixed compliance model, particularly as AI systems develop toward increasingly autonomous and agent-based forms.

8. CONCLUSION

AI is increasingly becoming a core component of organizational, industrial, and scientific infrastructures. Its growing capabilities in data analysis, knowledge discovery, automation, and decision support create significant opportunities for improving efficiency, innovation, and organizational intelligence. However, the integration of AI also introduces new challenges related to interoperability, governance, transparency, reproducibility, and accountability.

This chapter has argued that AI should not be viewed as a disruptive replacement for existing standards and infrastructures. Instead, the successful adoption of AI requires the extension and evolution of established standards, governance frameworks, and interoperability mechanisms. Existing infrastructures already provide essential foundations for data quality, traceability, compliance, and trust, which remain equally important in AI-driven environments.

The chapter proposes that trustworthy AI integration depends on the combined interaction of data quality, governance, and provenance. While data quality ensures reliable inputs and outputs, governance provides accountability and oversight mechanisms, and provenance enables transparency and contextual reproducibility

across complex AI lifecycles. Together, these dimensions form the foundation for standards-compliant and responsible AI deployment.

Particular attention was given to research and scientific infrastructures, where AI increasingly supports CRIS, scholarly analytics, metadata management, and scientific decision-making. The discussion emphasizes that methodological rigor, validation, transparent documentation, and human oversight remain necessary when AI is embedded in research processes. In this context, hybrid intelligence models that combine human expertise with AI capabilities offer a promising approach for balancing efficiency with scientific integrity.

From a practical perspective, organizations should adopt AI through interoperability-by-design, governance-by-design, and provenance-by-design principles. Incremental adoption strategies, cross-functional governance structures, continuous monitoring, and standards-based architectures can help ensure that AI remains transparent, auditable, and aligned with organizational objectives.

As AI technologies continue to evolve toward increasingly autonomous, adaptive, and agent-based systems, the importance of standards will not diminish but rather increase. Future digital ecosystems will require stronger integration between AI governance, provenance frameworks, interoperability standards, and regulatory mechanisms. The long-term value of AI will therefore depend not only on technological progress but also on the ability to embed AI within trustworthy and sustainable infrastructures.

The principal conceptual contribution of the chapter is the layered integration model that connects standards-based data foundations, existing operational systems, AI augmentation services, and governance and human oversight. Its value lies not in replacing existing frameworks, but in showing how regulatory, organizational, technical, and scientific requirements can be combined within one implementation architecture for CRIS and related information infrastructures.

Future research should further investigate governance mechanisms for autonomous AI agents, provenance models for generative AI systems, interoperability standards for multi-agent environments, and methods for balancing transparency, privacy, and digital sovereignty in increasingly AI-driven ecosystems.

REFERENCES

Ahmed, M. S., ... (2026). The Provenance Gap in Clinical AI: Evidence-Traceable Temporal Knowledge Graphs for Rare Disease Reasoning. https://arxiv.org/html/2604.17114v2

Apache Software Foundation. (n.d.). Apache Airflow. https://airflow.apache.org/

Association for Computing Machinery. (n.d.). ACM Computing Classification System. https://dl.acm.org/ccs

Auto, M. L. Group. (n.d.). Auto-PyTorch. AutoML.org. https://www.automl.org/automl-for-x/tabular-data/autopytorch/

AutoKeras. (n.d.). AutoKeras: An AutoML system based on Keras. https://autokeras.com/

Azeroual, O., & Schöpfel, J. (2024). New Developments in Research Data Management-The Potential of AI. *Encyclopedia of Libraries, Librarianship, and Information Science*, 2, 206-211. https://doi.org/DOI: 10.1016/B978-0-323-95689-5.00253-4

Belhajjame, K., ... (2013). PROV-DM: The PROV Data Model. https://www.w3.org/TR/prov-dm/

Bender, E. M., ... (2021). On the dangers of stochastic parrots: Can language models be too big? *ACM Conference on Fairness, Accountability, and Transparency (FAccT),* 610-623. https://dl.acm.org/doi/10.1145/3442188.3445922

Blodgett, S. L., Barocas, S., Daumé Iii, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in NLP. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 5454-5476). https://doi.org/DOI: 10.18653/v1/2020.acl-main.485

Bommasani, R.. (2021). On the opportunities and risks of foundation models. https://arxiv.org/abs/2108.07258

Cavoukian, A. (2009). Privacy by Design: The 7 Foundational Principles. *Information and Privacy Commissioner of Ontario.* https://student.cs.uwaterloo.ca/~cs492/papers/7foundationalprinciples_longer.pdf

CoARA. (2022). Coalition for Advancing Research Assessment. https://www.coara.org/

DataCite Metadata Working Group. (2026). DataCite metadata schema for the publication and citation of research data and other research outputs (Version 4.7). DataCite e.V. https://schema.datacite.org/

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering*, *61*(5), 637–643. DOI: 10.1007/s12599-019-00595-2

Dublin Core Metadata Initiative. (n.d.). Dublin Core. https://www.dublincore.org/

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., & Wright, R. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*(102642), 1–63. DOI: 10.1016/j.ijinfomgt.2023.102642

Ehrlinger, L., & Wöß, W. (2022). A survey of data quality measurement and monitoring tools. *Frontiers in Big Data*, *5*, 850611. https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2022.850611/full. DOI: 10.3389/fdata.2022.850611 PMID: 35434611

euroCRIS CERIF Task Group. (2012). CERIF-DataModel (Version 1.6.1) [Data model].

European Commission. (2019). Ethics Guidelines for Trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Commission. (2024). Artificial Intelligence Act (EU AI Act). https://artificialintelligenceact.eu/

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, *28*(4), 689–707. DOI: 10.1007/s11023-018-9482-5 PMID: 30930541

Freese, J., Rauf, T., & Voelkel, J. G. (2022). Advances in transparency and reproducibility in the social sciences. *Social Science Research*, *107*, 102770. DOI: 10.1016/j.ssresearch.2022.102770 PMID: 36058608

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé, H.III, & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, *64*(12), 86–92. DOI: 10.1145/3458723

GitHub. https://github.com/EuroCRIS/CERIF-DataModel

Herschel, M., Diestelkämper, R., & Ben Lahmar, H. (2017). A survey on provenance: What for? What form? What from? *The VLDB Journal*, *26*(6), 881–906. DOI: 10.1007/s00778-017-0486-1

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429–431. DOI: 10.1038/520429a PMID: 25903611

Howard, M. A., & Lipner, S. (2006). The Security Development Lifecycle. *Microsoft Press.* https://www.microsoftpressstore.com/store/security-development-lifecycle-9780735622142?utm_source=chatgpt.com

International Organization for Standardization. (2024). *ISO/IEC 5259-1:2024 Artificial intelligence—Data quality for analytics and machine learning (ML)—Part 1: Overview, terminology, and examples.* https://www.iso.org/standard/81088.html

ISO. (2023). Artificial Intelligence — Management System Standards. *International Organization for Standardization.* https://www.iso.org/standard/42001

Kumar, S., Datta, S., Singh, V., Singh, S. K., & Sharma, R. (2024). Opportunities and Challenges in Data-Centric AI. *IEEE Access : Practical Innovations, Open Solutions*, *12*, 33173–33189. DOI: 10.1109/ACCESS.2024.3369417

McAfee, A., & Brynjolfsson, E. (2017). Machine, platform, crowd: Harnessing our digital future. *WW Norton & Company.* https://wwnorton.com/books/Machine-Platform-Crowd/

Mehrabi, N., ... (2022). A Survey on Bias and Fairness in Machine Learning. In: *ACM Comput. Surv.*, 54(6), 115:1-115:35. https://doi.org/DOI: 10.1145/3457607

Miles, S., Groth, P., Munroe, S., & Moreau, L. (2011). Prime: A methodology for developing provenance-aware applications. In: *ACM TOSEM*, 20(3), 1-42. https://doi.org/DOI: 10.1145/2000791.2000792

Mitchell, M.. (2019). Model Cards for Model Reporting. *ACM Conference on Fairness, Accountability, and Transparency (FAT)*, 220-229. https://doi.org/DOI: 10.1145/3287560.3287596

MLflow. (n.d.). MLflow: The open source AI platform. https://mlflow.org/

Moreau, L., Hogan, N., & O'Donnell, N. (2023, April). Implementing an Environmental Management System Using Provenance-By-Design. *ACM Web Conference*, 1531-1541. https://doi.org/DOI: 10.1145/3543873.3587560

Munafò, M. R., ... (2017). A manifesto for reproducible science. In: *Nat Hum Behav*, 1(1), 0021. https://doi.org/DOI: 10.1038/s41562-016-0021

OECD. (2019). OECD principles on artificial intelligence. https://oecd.ai/en/ai-principles

Ojewale, V., Suresh, H., & Venkatasubramanian, S. (2026). Audit Trails for Accountability in Large Language Models. https://arxiv.org/abs/2601.20727

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans*, *30*(3), 286–297. DOI: 10.1109/3468.844354 PMID: 11760769

Pushkarna, M., Zaldiva, A., & Kjartansson, O. (2022). Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI. ACM Conference on Fairness, Accountability, and Transparency *(FAccT)*, 1776-1826. https://doi.org/ DOI: 10.1145/3531146.3533231

Research Organization Registry. (n.d.). Research Organization Registry (ROR). https://ror.org/

Ruan, Y.. (2024). Identifying the risks of LM agents with an LM-emulated sandbox. *International Conference on Learning Representations*, Vol. 2024, 27031-27098. https://doi.org/DOI: 10.48550/arXiv.2309.15817

Sambasivan, N., ... (2021). “Everyone wants to do the model work, not the data work”: Data Cascades in High-Stakes AI. *CHI Conference on Human Factors in Computing Systems*, 1-15. https://doi.org/DOI: 10.1145/3411764.3445518

Schlegel, M., & Sattler, K.-U. (2025). Capturing End-to-End Provenance for Machine Learning Pipelines. *Information Systems*, *132*, 102495. DOI: 10.1016/j.is.2024.102495

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. In *Advances in Neural Information Processing Systems 28*, 2503–2511. https://dl.acm.org/doi/10.5555/2969442.2969519

Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press., https://global.oup.com/academic/product/human-centered-ai-9780192845290?cc=de&lang= DOI: 10.1093/oso/9780192845290.001.0001

Strubell, E., Ganesh, A., & McCallum, A. (2020). Energy and policy considerations for modern deep learning research. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 09, pp. 13693-13696). https://doi.org/DOI: 10.1609/aaai.v34i09.7123

Tabassi, E. (2023). *Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1)*. National Institute of Standards and Technology., DOI: 10.6028/NIST.AI.100-1

Tabassi, E.NIST. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). *NIST AI*, *100-1*. Advance online publication. DOI: 10.6028/NIST.AI.100-1

W3C Working Group (2013). W3C PROV. https://www.w3.org/TR/2013/NOTE-prov-overview-20130430/

Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*, *13*(1), 14045. DOI: 10.1038/s41598-023-41032-5 PMID: 37679503

Waltersdorfer, L., Hausler, D., and Auge, T. (2025). Provenance Question-based AI Transparency and Accountable AI Governance. *ProvenanceWeek@SIGMOD*, 94-94. https://doi.org/ .DOI: 10.1145/3736229.3736260

Werthner, H.. (2022). *Perspectives on digital humanism*. Springer Cham., DOI: 10.1007/978-3-030-86144-5

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 1–9. DOI: 10.1038/sdata.2016.18 PMID: 26978244

World Wide Web Consortium. (n.d.). Web standards. https://www.w3.org/standards/

Zendulková, D., Azeroual, O., & Krivjanská, A. (2025). From Data Silos to FAIR Research: Interoperability and Integration in European CRIS. In *Digitalization and Management Innovation IV* (pp. 228–240). IOS Press., DOI: 10.3233/FAIA250723

Zha, D.. (2023). Data-centric AI: Perspectives and Challenges. *SIAM International Conference on Data Mining (SDM)*, 945-948. https://doi.org/DOI: 10.1137/1.9781611977653.ch106

ENDNOTE

[1] In this chapter, the internationally established term *Current Research Information Systems (CRIS)* (euroCRIS) is used throughout instead of *Research Information Systems (RIS)* or the German term *Forschungsinformationssysteme (FIS)*.