

Statistical Commission

Fifty-seventh session

New York, 3 - 6 March 2026

Item 5(h) of the provisional agenda

Items for information: coordination of statistical programmes

Background document

Available in English only

AI-Ready Official Statistics: Opportunities, Challenges, and Recommendations

Prepared by the World Bank

AI-Ready Official Statistics: Opportunities, Challenges, and Recommendations

December 2025

Acknowledgments and Process

The production of this document was led by the Office of the World Bank Group Chief Statistician. The document is based on a World Bank paper shared as a [room document](#) at the 56th session of the United Nations Statistical Commission in 2025.

The document was first presented for review and discussion at the Committee for the Coordination of Statistical Activities (CCSA) meeting in Luxemburg on 2-3 October 2025. Member agencies of the CCSA shared their comments and suggestions on the draft. CCSA member agencies approved the revised document without objections for submission as a background paper to the United Nations Statistical Commission (UNSC) 2026.

Contact

For feedback or inquiries regarding this document, please contact Olivier Dupriez, Deputy Chief Statistician, World Bank (odupriez@worldbank.org).

Use of Artificial Intelligence (AI)

AI tools, including Azure OpenAI, Google Gemini, and Microsoft Copilot's *Auto Rewrite* tool, were utilized in the writing of this document. The initial content was drafted by the author, with AI providing suggestions and subsequently enhancing or editing the text for clarity and consistency. All AI-generated content was reviewed for accuracy and relevance.

Abbreviations

AI	Artificial intelligence
API	Application programming interface
BIS	Bank of International Settlements
CC BY	Creative Commons Attribution
CCSA	Committee for the Coordination of Statistical Activities
CES	Conference of European Statisticians
DCAT	Data Catalog Vocabulary
DCAT-AP	Data Catalog Vocabulary – Application Profile
DDI	Data Documentation Initiative
DSD	Data Structure Definition
ESS	European Statistical System
FAIR	Findable, Accessible, Interoperable, Reusable
FAO	Food and Agriculture Organization of the United Nations
GIS	Geographic Information System
GSBPM	Generic Statistical Business Model
GSIM	Generic Statistical Information Model
HTML	Hypertext Markup Language
IFC	Irving Fisher Committee
IMF	International Monetary Fund
ISO	International Organization for Standardization
IT	Information technology
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation for Linked Data
LLM	Large language model
MCP	Model context protocol
MOU	Memorandum of understanding
MSD	Metadata Structure Definition
NIST	National Institute for Standards and Technology (United States)
NQAF	National quality assurance framework
NSO	National statistical organization
ODK	Open Data Kit
OECD	Organization for Economic Co-operation and Development
PCN	Proof Carrying Numbers
QF4SA	Quality Framework for Statistical Algorithms
SDK	Software Development Kit
SDMX	Statistical Data and Metadata Exchange
SIMS	Single Integrated Metadata Structure (Eurostat)
SIS-CC	Statistical Information System Collaboration Community
SPEED	State, Policy, Expertise, Engine, and Delivery
SWG	Standards Working Group
TDO	Trusted Data Observatory
UI	User interface
UNECE	United Nations Economic Commission for Europe
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
WIPO	World Intellectual Property Organization

Contents

Executive Summary.....	i
1 Introduction	1
2 Generative AI: Opportunities for Official Statistics	1
3 Challenges of Data Proliferation and Generative AI	3
3.1 Multiplicity of Data Sources: A Challenge to Authority and Relevance	3
3.2 Transformation of Internet Search and the Rise of AI Chatbots	4
3.3 Limitations of Generative AI Technology	5
3.4 Lack of Strategic Partnerships and a Missed Opportunity	5
3.5 New Requirements for IT Infrastructure, Skills, and Legal and Ethical Frameworks	7
4 Roles, Responsibilities, and Recommendations	7
4.1 Recommendations for National Statistical Organizations.....	9
4.1.1 Metadata Standards and Metadata Optimization	9
4.1.2 Data Quality Control and Data Augmentation	10
4.1.3 Data and Metadata Dissemination.....	11
4.1.4 Skills and Infrastructure for an Enabling Environment.....	13
4.1.5 Collaborations and Partnerships	13
4.2 Recommendations for International and Regional Organizations.....	15
4.2.1 Metadata Standards and Metadata Quality Assurance	15
4.2.2 Tools and Guidance for Data Quality Control and Augmentation	16
4.2.3 Tools and Guidance for Data and Metadata Dissemination	17
4.2.4 Build AI Capacity in Low- and Middle-income Countries.....	18
4.2.5 Catalyze Collaborations and Partnerships	19
4.3 Recommendations for Redistributors of Official Statistics	19
4.3.1 Prioritize Authoritative Sources and Respect Licensing Terms.....	20
4.3.2 Ensure Statistical Integrity in AI-enabled Dissemination	20
4.3.3 Establish Feedback and Collaboration Mechanisms with Data Producers.....	21
4.4 Recommendations for the Development of AI Technology	22
5 Conclusion	23
Annex: The Model Context Protocol	24

Executive Summary

This document examines the transformative impact of artificial intelligence (AI), particularly generative AI, on the dissemination and use of official statistics. As AI-driven platforms operated by official and non-official entities increasingly mediate access to data, NSOs must evolve to ensure their products remain accessible, accurately contextualized, and responsibly applied.

The document introduces a framework for AI-readiness in statistical products. Key components include comprehensive, machine-readable metadata; rigorous quality assurance processes; delivery through accessible APIs and interoperable formats; open access to data and metadata; and improved discoverability, reinforced by explicit signals of authoritativeness.

Achieving this readiness requires cohesive, strategic action. Regardless of whether they adopt generative AI for dissemination, NSOs should modernize data management and dissemination practices, strengthen quality control, enhance governance, and invest in new skills and partnerships. International organizations should lead in establishing standards, providing technical guidance and open-source tools, facilitating capacity building, and acting as intermediaries between national organizations and technology firms. Third-party data redistributors and AI developers must prioritize authoritative sources, maintain data and metadata integrity, and support the sustainability of the public-interest official data ecosystem.

Immediate, coordinated action is required. Without it, the visibility and trustworthiness of official statistics may be diminished within the evolving AI-driven information environment.

1 Introduction

For decades, national statistical organizations (NSOs) have served as the primary providers and custodians of official statistics used to design, implement, and monitor social and economic policies. Today, that landscape has fundamentally changed. Data and statistics are increasingly generated and disseminated by a diverse set of actors, including non-official sources and advanced digital technologies. As a result, the role of NSOs is evolving, from producing official statistics alone to overseeing and guiding a broader and more complex data ecosystem. This expanded mandate includes assessing non-official data sources, providing methodological leadership, and promoting technical and ethical standards, transparency, and interoperability.

Artificial intelligence (AI), and generative AI in particular, is accelerating this transformation. AI now affects every stage of the data lifecycle, from collection and processing to dissemination, discovery, interpretation, and use. The pace of change is rapid and disruptive: AI is reshaping how statistics are accessed and consumed faster than most official systems can adapt. At the same time, AI offers NSOs a powerful opportunity to make official statistics more accessible, relevant, and useful to a wider range of users, including those previously constrained by technical barriers. Realizing this potential requires NSOs to become AI-ready by strengthening governance frameworks, skills, data infrastructure, legal safeguards, and partnerships to ensure the responsible use of AI.

Many national statistical systems are already experimenting with new technologies while upholding the Fundamental Principles of Official Statistics. However, these efforts alone are not sufficient to ensure that official statistics remain visible, trusted, and responsibly used. Official producers now operate in an environment where intermediaries increasingly rely on AI to distribute information in ways that better align with users' expectations, yet do not consistently prioritize authoritative sources or clearly signal data quality.

The current moment is decisive. NSOs must ensure both their own AI preparedness and the readiness of their data outputs for integration into AI-driven systems. This entails improving transparency, discoverability, machine readability, and usability of data products and systems so that official statistics can be reliably used in the digital environments where users seek information.

The document examines the AI readiness of official statistics, identifies key risks and opportunities associated with generative AI, and clarifies the roles and responsibilities of four principal stakeholder groups. It provides actionable recommendations and calls for new partnerships to sustain official statistics as a trusted public good in a rapidly evolving data ecosystem.

2 Generative AI: Opportunities for Official Statistics

In a crowded and rapidly evolving data environment shaped by AI, high-performing NSOs must stand out as trusted providers of reliable evidence. Misinformation, AI-generated synthetic content, and data from unvetted sources are increasingly undermining confidence in public information. In this

context, users are seeking stronger assurances of data quality, provenance, and integrity. Where statistical systems operate professionally and independently, in line with the Fundamental Principles of Official Statistics and robust quality-assurance frameworks, official statistics provide a critical anchor of trust. As impartial producers of high-quality data, NSOs remain essential to both policymakers and citizens.¹

AI offers significant opportunities to strengthen this role by enhancing every stage of the statistical lifecycle.² For example, AI can:

- **Increase productivity** by automating repetitive tasks, translation, and code generation, and by integrating diverse data sources to produce more timely and disaggregated statistics.
- **Strengthen quality assurance** through the detection of inconsistencies and non-compliance across data and metadata, improving accuracy, comparability, consistency, and timeliness.
- **Modernize dissemination and engagement** via natural-language interfaces, improved data discovery and interpretability, tailored products, and more inclusive outputs, including automated reporting to regional and international organizations.

Generative AI is already reshaping how most users discover and consume information. AI-enabled search engines and chatbots operated by large technology companies have become unavoidable intermediaries in the dissemination of official statistics, despite well-documented limitations such as hallucinations and misinterpretation. Increasingly, users will encounter official statistics through these platforms rather than directly from NSO websites. NSOs should therefore proactively engage with such intermediaries and establish partnerships to ensure that authoritative data is accurately surfaced, clearly attributed, and responsibly contextualized. Collaboration with redistributors can help distinguish verified statistics from unverified information and mitigate the risks of misinformation.

Realizing these opportunities requires more than deploying new tools. For most NSOs, AI-readiness is primarily a strategic challenge that involves modernizing data and metadata management, strengthening quality-assurance practices, updating governance and legal frameworks, and building organizational skills. It is a progressive and adaptive process rather than a single technological milestone.³ While many NSOs face resource and capacity constraints, and may lack specialized AI talent, they do not need to develop complex systems such as large language models themselves.

¹ See United Nations Economic Commission for Europe, [The Future of National Statistics Offices: A Call to Action](#), ed. Osama Rahman (May 2025)

² Opportunities, challenges and recommendations related to AI have also been discussed in the context of central banks which function both as data users and producers. See Irving Fisher Committee on Central Bank Statistics (IFC) of the Bank for International Settlements (BIS) (2025): “[Governance and implementation of artificial intelligence in central banks](#)”, IFC Report, no 18, April.

³ See for example the [framework developed by PARIS21](#), which proposes strengthening five interconnected pillars to foster AI readiness: State, Policy, Expertise, Engine, and Delivery (SPEED).

Instead, they should focus on the responsible adoption and adaptation of existing AI technologies to meet their specific needs.

Finally, the transition to AI-enabled official statistics must be inclusive and globally equitable. Without deliberate action, AI adoption risks widening gaps between data-rich and data-poor systems. International collaboration, open sharing of solutions, and pooled resources supported by common standards and good practices are essential to ensure that all NSOs, including the most resource-constrained, can benefit from AI.

3 Challenges of Data Proliferation and Generative AI

A surge of non-official data sources and the rapid diffusion of generative AI are reshaping how information is produced, disseminated, redistributed, and consumed. While these shifts create new opportunities, they also weaken the traditional position of NSOs as the primary providers of trusted social and economic statistics. To respond effectively, NSOs must address five interlinked challenges:

- **The multiplicity of data sources**, which fragments the attention of data users and blurs distinctions between official and non-official, vetted and unvetted content.
- **The transformation of internet search engines and the rise of AI-enabled chatbots**, which concentrate power among third-party data redistributors and increasingly mediate access to official statistics.
- **The inherent limitations of generative AI**, which threaten data quality, contextual integrity, and trust.
- **Weak formal partnerships and collaboration** between official producers and dominant redistributors, leading to inconsistent surfacing, attribution, and interpretation of authoritative data.
- **Insufficient infrastructure, skills, and legal frameworks** within NSOs to deploy and govern AI responsibly.

Grasping these dynamics is essential to crafting a forward-looking response that preserves trust, strengthens data integrity, and ensures that official statistics remain discoverable, usable, and authoritative.

3.1 Multiplicity of Data Sources: A Challenge to Authority and Relevance

The rise of private and citizen-generated data has forced official statistics to compete for attention in an increasingly crowded information environment. Many non-official sources lack the transparency, methodological rigor, and quality assurance that characterize official statistics. However, they often offer features that users value, such as faster updates, finer geographic or thematic granularity, convenient access, and seamless integration into digital products. These factors are reshaping user expectations and placing sustained pressure on NSOs. The challenge is twofold:

- To uphold and clearly demonstrate elevated standards of quality, integrity, and professional independence, reinforcing the “trustworthiness advantage” of official statistics.
- To better meet user demands for timeliness, granularity, and usability by appropriately leveraging emerging technologies and a wider range of data sources.

Maintaining authority and relevance in this environment requires that official statistics be prepared for integration into AI-driven systems. This entails ensuring that data are user-focused, structured, machine-readable, and accessible through APIs; thoroughly documented with transparent provenance and metadata; and aligned with open standards to enable interoperability with modern technologies. Improving discoverability, usability, and interpretability across platforms is essential to making official statistics more accessible, integrable, and trustworthy for a wide range of users.

3.2 Transformation of Internet Search and the Rise of AI Chatbots

For many years, web search engines functioned primarily as gateways, presenting users with links to source websites (including official NSO portals), allowing users to assess provenance, compare sources, and interpret information independently. This model is changing rapidly. While links remain available, they now play a more supplementary role. AI-enabled search engines and leading chatbots (such as Gemini by Google or ChatGPT by OpenAI) increasingly aim to retain users within their own interfaces by delivering instant, synthesized answers. At the same time, search results increasingly surface sponsored or commercially prioritized content.

These platforms, operated by large technology companies with substantial resources, are becoming dominant and largely unavoidable intermediaries for the discovery and consumption of official statistics. While they play an important role in informing societies, there are currently limited legal or regulatory obligations requiring them to proactively ensure the quality, provenance, or accuracy of their outputs.

This transformation creates three interrelated challenges for official statistics:

- **Weak prioritization of authoritative data:** Search and AI algorithms may optimize for engagement, commercial relationships, or popularity rather than methodological rigor. As a result, authoritative sources are not consistently surfaced or clearly attributed, and official statistics may be crowded out by alternative data.
- **Opaque and error-prone synthesis:** AI-generated responses often lack clear citations, provenance, and versioning. They may combine accurate figures with outdated, misinterpreted, or incorrect information, making verification difficult and obscuring limitations.
- **Reduced direct access to NSO channels:** User attention increasingly shifts from NSO websites and tools to third-party platforms. As a result, users may consume statistics without access to original metadata, quality statements, or methodological explanations. Although AI chatbots increasingly provide citations and links—a positive development—the referenced sources are not always authoritative or current, and may point to secondary

aggregators or outdated materials. Ensuring that AI systems consistently surface canonical, verified sources remains a critical challenge.

The cumulative effect of these changes is diminished visibility and reduced control over how official statistics are presented, interpreted, and trusted. Without proactive engagement and improved machine-readability, official data risk being misrepresented or de-prioritized within dominant AI-enabled discovery ecosystems.

3.3 Limitations of Generative AI Technology

Generative AI has advanced rapidly, but its current design and deployment introduce significant risks for how official statistics are accessed, interpreted, and presented.

- **Non-determinism, hallucination, and overconfidence:** Large language models (LLMs) are stochastic. They do not produce identical outputs for the same input and may fabricate plausible-sounding figures or summaries—particularly when authoritative data are missing or ambiguous—and present them with undue certainty. This variability and overconfidence undermine reproducibility, auditability, and trust.
- **Weak integration with authoritative data:** In most current deployments, LLMs do not directly query structured datasets maintained by statistical agencies. Instead, they rely on secondary sources (such as articles, dashboards, or third-party websites), creating several risks:
 - **Outdated or inaccurate figures:** Outputs may reflect static or stale content and present numbers that have since been revised or superseded.
 - **Loss of metadata:** Official statistics require definitions, units, caveats, and methodological notes, which AI-generated responses often omit, truncate, or misstate.
 - **Misattribution and unclear provenance:** Systems may attribute official data to unofficial sources, or conversely attribute unofficial or outdated content to NSOs, eroding credibility and confusing users.

These limitations are evolving with emerging approaches such as the Model Context Protocol (MCP; see annex), but they will not disappear in the near term. Addressing them requires both technological progress and targeted action by NSOs, including improving the machine-readability and quality of data and metadata, strengthening provenance information, and implementing modern search services and APIs.

3.4 Lack of Strategic Partnerships and a Missed Opportunity

Open data has made official statistics easier to share, adapt, and reuse. As a result, many users now access official statistics through third-party platforms such as AI-enabled search engines and chatbots, rather than directly via NSO websites. However, this openness has not been matched by formal, systematic collaboration between NSOs, acting as a community, and dominant redistributors.

In the absence of such collaboration, AI-mediated redistribution introduces persistent risks:

- **Alteration and loss of context:** Official data may be selectively presented, transformed, or taken out of context, leading to hallucinations, inconsistencies, loss of essential metadata, and misrepresentation of methods. This undermines credibility and reduces the quality and trustworthiness of services to users.
- **License gaps and weak compliance:** While open licenses often permit transformation and redistribution, obligations related to citation, provenance, and versioning are frequently interpreted loosely or inconsistently. Weak compliance may discourage some producers from adopting licenses that allow commercial reuse, which some redistributors require.
- **Reduced insight into data use:** As user interactions shift to third-party platforms, NSOs lose visibility into how their data are discovered and used. Portal analytics no longer reflect actual reach, limiting evidence-based prioritization.

This collaboration gap represents a missed opportunity. Proactive and coordinated engagement between NSOs and third-party redistributors could yield tangible benefits for all parties:

- NSOs provide machine-actionable data and structured metadata, (including standardized ontologies and taxonomies, persistent identifiers, source-of-truth URLs, and versioning) enabling more efficient acquisition and redistribution.
- Redistributors improve the visibility, discoverability, and ranking of authoritative official statistics within search engines and AI interfaces.
- Joint good practices are developed, including guidance on attribution and guardrails against distortion during summarization or transformation of official statistics.
- Feedback loops are established, giving NSOs insight into how their statistics are found, cited, and interpreted, and enabling more user-centered product development.

The stakes extend beyond technical and reputational concerns to legal and ethical dimensions. Legally, incorrect attribution may breach licensing terms, impede source verification, and expose both redistributors and NSOs to liability where decisions rely on misleading data. Weak governance over personal or sensitive information also increases risks related to privacy, re-identification, and unauthorized use. Ethically, selective presentation and AI-generated summaries may introduce bias, remove essential context, or mislead users. NSOs, and, arguably third-party redistributors, have a responsibility to exercise reasonable care to avoid disseminating information that could foreseeably cause harm. While the scope and enforceability of such duties vary by jurisdiction, international principles such as the United Nations Fundamental Principles of Official Statistics provide widely recognized benchmarks for responsible data production and dissemination.

Partnerships are therefore essential to uphold integrity, protect rights, and ensure the responsible reuse of official statistics. Active collaboration between official data producers and third-party redistributors can help shape shared infrastructure and practices for trustworthy redistribution. At the same time, it must be recognized that major intermediaries are private, for-profit entities whose

incentives will not always fully align with those of NSOs. Effective partnerships will thus require clear governance arrangements that balance commercial interests with public-interest obligations.

3.5 New Requirements for IT Infrastructure, Skills, and Legal and Ethical Frameworks

NSOs face growing expectations to leverage generative AI in their dissemination activities. While responsible AI adoption promises efficiency gains and expanded reach, it also requires IT infrastructure, skills, and safeguards that many NSOs are still developing.

Generative AI introduces confidentiality and reputational risks when open or third-party models and tools are used within statistical environments. Because many AI platforms operate on external servers or cloud infrastructures beyond NSO control, uploading, referencing, or processing internal datasets may expose sensitive, restricted, or confidential information to unauthorized access or leakage.

Before adopting such tools, NSOs should establish robust governance arrangements, data-protection policies, and risk-assessment procedures. These should clearly define which data may be shared with external systems and ensure secure, anonymized processing environments.

Key challenges to responsible AI-enabled dissemination by NSOs include:

- **Legacy IT systems and limited compute capacity**, which constrain the integration of AI solutions; in resource-constrained contexts, secure hosting environments may be unavailable.
- **Skills shortages** in areas such as AI engineering, data governance, and model oversight, limiting the ability to design, maintain, and audit AI-enabled dissemination tools and making it difficult to attract and retain specialized talent.
- **The need for new validation and assurance mechanisms**, including systematic testing, red-teaming, and monitoring of AI-generated outputs to mitigate risks of hallucination, misinterpretation, and reputational harm.
- **Weak or uneven institutional oversight**, as many countries, particularly lower-income ones, lack mature data protection authorities or cybersecurity agencies.
- **Outdated legal and ethical frameworks**, requiring updates to statistical laws and mandates to address cloud use, data sharing, access to administrative and big data, and AI-specific risks.

Addressing these challenges is essential for ensuring that AI-enhanced dissemination strengthens, rather than undermines, the integrity and trustworthiness of official statistics.

4 Roles, Responsibilities, and Recommendations

As third-party search engines and platforms powered by generative AI increasingly mediate how users discover and consume information, deliberate coordination and partnerships with the private

sector are required to protect the integrity and public value of official statistics. Keeping official and authoritative data visible, accessible, correctly interpreted, and responsibly reused in AI-enabled environments is a shared responsibility.

Major chatbots and search engines acknowledge the limitations of generative AI, especially the risk of hallucinations, by displaying caveats such as “*ChatGPT can make mistakes. Check important info.*” (OpenAI) or “*AI can make mistakes, so double-check responses*” (Google Gemini). Such disclaimers put much of the burden of quality assurance onto users, many of whom may lack the time, expertise, or motivation to validate results. Tools designed to streamline access should not depend on high levels of user verification. Responsibility for quality must shift toward the systems that produce, deliver, and redistribute data.

This chapter defines the roles and responsibilities of four stakeholder groups and offers targeted recommendations. Each has a distinct, complementary role in adapting the data ecosystem to new technological realities:

- **Official data producers** must ensure that official statistics remain authoritative and useful as AI reshapes how information is discovered, interpreted, and reused. Becoming AI-ready means aligning infrastructure and governance with widely accepted standards so data and metadata are truly machine-actionable and accessible, cultivating transparent data quality practices, and opening programmatic access to data and metadata in ways that are clear, predictable, and responsible. It also calls for careful integration of AI into NSO dissemination platforms, while building skills and collaborations that sustain these practices over time.
- **International and regional organizations** that support, coordinate, or represent statistical systems must accelerate AI-readiness of official statistics by harmonizing metadata standards, providing open tools for data and metadata quality assessment and enhancement, modernizing dissemination, building capacity in low- and middle-income countries, and convening partnerships that ensure responsible reuse and visibility of trustworthy data.
- **Third-party redistributors** (including search engines, chatbots, and data aggregators) should prioritize authoritative sources, respect licenses and interpretive guidance, and anchor all AI outputs in canonical data with strong provenance safeguards to allow users to effectively assess the trustworthiness of the data. They should maintain structured collaboration and feedback loops with official producers to ensure accurate, trustworthy, and well-presented official statistics across platforms.
- **AI technology developers** must take responsibility for how their tools source, attribute, and present data and statistics. They should prioritize authoritative integrations in the training of models, support the development of low-resource models, embed statistical integrity and safety checks, and be open to collaborating with the official statistical community.

The recommendations that follow aim to build a more equitable, transparent, and trustworthy AI-enabled data dissemination system, one in which official statistics continue to function as a foundational public good.

4.1 Recommendations for National Statistical Organizations

Official statistics must remain findable and trustworthy in AI-enabled environments. Becoming AI-ready requires modernization across technology, data management and dissemination, governance, and partnerships anchored in machine-actionable metadata, robust quality control, open licensing, responsible AI-enabled access, skills development, and structured collaboration. The recommendations below outline practical steps for NSOs.

4.1.1 Metadata Standards and Metadata Optimization

For NSOs, high-quality, machine-actionable metadata is the foundation of safe, effective AI use. Generative systems rely on metadata to understand meaning, provenance, structure, and fitness for purpose before interacting with the data itself. To maximize semantic discoverability, reliable attribution, and accurate interpretation, NSOs should adopt fit-for-purpose metadata standards and expose metadata in web-optimized formats that AI systems can efficiently ingest and reuse.

Adopt fit-for-purpose metadata standards and ontologies across data types

- **R.1.** Produce comprehensive descriptive and structural metadata that conform to recognized standards, such as the Statistical Data and Metadata Exchange (SDMX) for indicators, Data Documentation Initiative (DDI) for microdata, and ISO 19115/19119/19110 for geospatial datasets and services. Additionally, publish a web-ready subset of these canonical metadata mapped to the Data Catalog Vocabulary (DCAT), schema.org, and Croissant standards (with links to the full canonical records) to enhance online discoverability and machine use.⁴
- **R.2.** Adopt standard taxonomies and domain ontologies when possible, to enable semantic interoperability and consistent query expansion in AI systems. Maintain authoritative versioned code lists, and expose them via stable APIs with clear identifiers to support automated integration.
- **R.3.** Store and disseminate metadata in open, machine-readable formats such as JSON or JSON-LD, including persistent identifiers, licensing, and provenance fields, and ensure endpoints support both human-readable views and programmatic access.
- **R.4.** Implement semantic versioning ⁵ for metadata with human- and machine-readable changelogs, maintain synchronization between metadata and data versions, and provide deprecation notices and migration guidance for downstream users.

⁴ See recommendations to international and regional organizations R.40 to R.43.

⁵ Semantic versioning for datasets uses a MAJOR.MINOR.PATCH scheme where MAJOR increments denote breaking changes to schema or semantics, MINOR increments denote backward-compatible additions (e.g., new fields or records), and PATCH increments denote backward-compatible corrections to data or metadata. See <https://semver.org/>

Framework and AI solutions to assess and enhance metadata quality

- **R.5.** Collaborate with international and regional organizations to define practical quality criteria, benchmarks, and workflows for metadata quality assurance, covering content elements (e.g., indicator names, definitions, methods, relevance, limitations) and format conformance, to improve discoverability, interpretability, clarity, and transparency across statistical domains.
- **R.6.** Co-develop and deploy open-source, agentic AI tools that automatically assess metadata against agreed standards and quality criteria (including specificity, clarity, consistency, completeness, provenance, and licensing), provide actionable suggestions, and integrate into routine curation and dissemination pipelines for continuous improvement.⁶

4.1.2 Data Quality Control and Data Augmentation

To ensure official statistics are interpreted and used accurately by people and AI systems, NSOs should modernize data quality assurance and quality control workflows, automate anomaly detection, enrich datasets with machine-readable context, and transparently document limitations.

Anomaly detection in time series and data augmentation

- **R.7.** Build and deploy quality control pipelines that combine rule-based validation with statistical and machine learning methods to detect inconsistencies and anomalies in time series (sudden level shifts, spikes and drops). Classify detected anomalies as either errors or legitimate outliers. Use AI-assisted tools for classification but always require human review for final decisions.
- **R.8.** Correct the observation values confirmed as true errors by human verification.⁷ For anomalies confirmed as legitimate outliers, document the underlying cause by adding observation-level annotations to the indicators database as part of the data augmentation process.⁸ Use AI to draft annotations and link events (e.g., policy changes, disasters) to specific observations, with validation by subject-matter experts. Make these validated annotations available via APIs and connect them to visualizations so both AI systems and users can access contextual information alongside the data values.

Domain-specific quality control for microdata and geospatial data

- **R.9.** Strengthen microdata quality assurance by running cross-record and within-record checks (ranges, skip patterns, duplicates, imputation audits), verifying sampling weights and

⁶ International organizations should lead the development of such open solutions. See recommendations R.44 and R.45.

⁷ If the error originates from data sourced externally, the issue will have to be communicated to the primary producer in a constructive manner.

⁸ In SDMX-compliant databases, this would consist of adding *observation flags*, which are coded attributes that provide qualitative information about a data point, such as its accuracy, reliability, or status.

design variables, assessing nonresponse and measurement error, conducting disclosure risk assessments as needed, and maintaining reproducible cleaning scripts with systematic documentation of identified issues and limitations.

- **R.10.** Improve quality assurance for geospatial datasets by validating geometry and topology, confirming coordinate reference systems, checking attribute consistency, assessing spatial accuracy and completeness, applying automated topology rules, and verifying temporal consistency for time-stamped layers used in modeled estimates and small-area statistics.

Make quality assurance and quality control processes and known limitations transparent

- **R.11.** Publish machine-readable quality metadata that covers relevance, accuracy, timeliness, coherence, comparability, and accessibility, explicitly documenting known limitations, caveats, compliance issues, and fitness-for-use statements using elements available in specialized metadata standards to inform both automated consumers and end users.

Update National Quality Assurance Frameworks (NQAFs)

- **R.12.** Define and track AI-aware data quality metrics, including metadata completeness and semantic accuracy, conformance to standards and ontologies, lineage and provenance integrity, time-series anomaly detection coverage and annotation rates, and other indicators relevant to AI-assisted production and dissemination.
- **R.13.** Establish governance for statistics generated or assisted by AI, including model documentation, testing and validation protocols, change control, performance monitoring, and fairness and bias assessment, with clear roles and decision thresholds.⁹
- **R.14.** Update NQAFs to incorporate the new AI-aware quality metrics and governance arrangements, ensuring that AI-assisted processes are covered by routine audits and transparent reporting.

4.1.3 Data and Metadata Dissemination

To ensure official statistics are easy for people and AI to find, interpret, and reuse responsibly, NSOs should modernize dissemination for both data and metadata. Priorities include clear, permissive licensing of content; machine-readable access to data and metadata via robust, versioned APIs and convenient bulk downloads; metadata exposed in web-optimized profiles; semantic search and carefully governed chatbot integration; and predictable, machine-readable advanced release calendars and change logs so redistributors can preserve provenance and keep content current.

⁹ This can be inspired by frameworks such as the 2023 [Risk Management Framework](#) by the United States National Institute for Standards and Technology (NIST) , and the [Quality Framework for Statistical Algorithms \(QF4SA\)](#) by the United Nations Economic Commission for Europe (UNECE)

Openness and licensing

- **R.15.** Make data and metadata as openly accessible as permitted by law, apply standard permissive licenses (e.g., Creative Commons Attribution - CC BY 4.0 or applicable government licenses), display license terms prominently at dataset pages, distributions, and API endpoints, and provide straightforward guidance for licensees, noting that many redistributors require licenses that explicitly permit commercial use.
- **R.16.** Publish machine- and human-readable licensing and attribution guidance, including required citation formats, provenance and version references, and examples, so redistributors can implement obligations correctly and consistently.¹⁰

Metadata access: APIs and web profiles

- **R.17.** Expose complete, standards-compliant metadata through well-documented APIs to support AI-enabled discovery, interpretation, and validation, and publish web-ready subsets mapped to common profiles (DCAT, schema.org, Croissant) with resolvable links back to canonical records for search engine optimization and crawler ingestion (see also R.1).
- **R.18.** Distribute metadata alongside data in formats convenient to users, providing human-readable HTML catalog pages that clearly describe provenance, version history, and license terms, and offering downloads in open, machine-readable formats such as JSON/JSON-LD.

Modern dissemination platforms with semantic search and ethical guardrails

- **R.19.** Upgrade platforms that rely on keyword search by adding semantic search that leverages rich metadata, ontologies, and embeddings to improve discoverability, and provide access to these capabilities through both user interfaces and APIs.
- **R.20.** Deploy chatbots only when they can reliably access canonical databases under strict guardrails—such as source citation, retrieval transparency, and Model Context Protocol integration—and adopt an ethical framework for AI-enabled dissemination that addresses inclusion, potential algorithmic bias, transparency, accountability, and redress mechanisms.
- **R.21.** Publish eligible datasets in reputable data commons that offer authoritative semantic dissemination to expand visibility and uptake of NSO data products.¹¹

Data access: APIs and bulk download

- **R.22.** Provide official statistics via stable, versioned, and well-documented APIs with clear rate limits and pagination, offer bulk downloads to support large-scale reuse, discourage scraping by making API and bulk access reliable, and furnish sample queries, SDKs, and code snippets in common languages such as R and Python.

¹⁰ International organizations should provide standardized guidance; see R.59.

¹¹ For example, the International Monetary Fund is proposing StatGPT as an AI/SDMX-enabled data common. The [SDMX Sponsor Organizations' Statement on AI Readiness](#) also recommends the establishment of a Global Trusted Data Commons.

Release calendars and machine-readable change logs

- **R.23.** Maintain a machine-readable advanced release calendar, expose update notifications at dataset and API levels, and publish versioned change logs for data and metadata (including revision histories and deprecation notices) with subscription options so users and partners can track updates automatically.

4.1.4 Skills and Infrastructure for an Enabling Environment

To build AI readiness, NSOs need both the right capabilities and modern, secure, interoperable technology. This goes beyond adopting new tools: it requires multidisciplinary teams, continuous learning, and robust platforms that support safe, scalable dissemination, observability, and sustainable operations across on-premises and cloud environments.

Infrastructure and platforms

- **R.24.** Modernize core infrastructure to support secure, scalable data dissemination by implementing well-managed APIs, standards-compliant data catalogs, and end-to-end data observability for pipelines across the data lifecycle.
- **R.25.** Adopt cloud-ready or hybrid architectures with robust security and compliance controls, including identity and access management, encryption at rest and in transit, and vulnerability scanning.
- **R.26.** Provide development sandboxes, testing and staging environments, and reproducible pipelines to validate changes safely before production and enable rapid, controlled iteration.
- **R.27.** Plan for sustainability and cost by leveraging open-source tooling where appropriate, designing vendor-neutral architectures to reduce lock-in, and tracking costs of operation to inform decisions.

Talent and skills

- **R.28.** Invest in reskilling and targeted recruitment across data science, machine learning, metadata management, data engineering, cloud operations, and cybersecurity, and establish continuous learning programs that combine on-the-job training, certifications, and peer learning communities.
- **R.29.** Create multidisciplinary teams that pair statisticians and subject-matter experts with engineers and product leads to design AI-ready data products, automate curation, and operate reliable dissemination workflows.

4.1.5 Collaborations and Partnerships

Building an AI-ready data ecosystem is a shared effort; NSOs should cultivate partnerships that expand access to tools, expertise, and infrastructure, especially for resource-constrained offices.

Leverage international collaboration and shared resources

- **R.30.** Partner with international and regional organizations and peer NSOs to co-develop and share open-source solutions and common components (e.g., data and metadata validators, metadata tooling, semantic services), prioritizing interoperable, standards-based tools that support internationalization while reducing costs and vendor lock-in.¹²
- **R.31.** Engage actively in regional and international forums to exchange practices and co-create tools and technical guidance for AI implementation, ensuring representation of diverse capacity contexts.¹³
- **R.32.** Collaborate on reference implementations of metadata standards and optimization methods,¹⁴ including validation test suites and metrics to assess discoverability of metadata across domains and languages.
- **R.33.** Pilot joint projects to improve semantic discoverability of official statistics by evaluating embedding models, building shared test sets of user queries, and publishing benchmarks for search quality assessment.
- **R.34.** Contribute case studies and code, to shared repositories or knowledge commons to accelerate adoption and support inclusivity for lower-capacity NSOs.¹⁵

Establish feedback and redress mechanisms with redistributors of official statistics

- **R.35.** When establishing formal agreements with redistributors, include detailed provisions for attribution and provenance, covering source citation, version control, update dissemination, and, where appropriate, the sharing of usage monitoring data, to ensure transparency and timely propagation of corrections.
- **R.36.** Invoke Principle 4 of the Fundamental Principles of Official Statistics by publicly correcting misinterpretation or misuse and issuing notices as needed, and implement clear reporting and escalation mechanisms for misuse, inadequate attribution, or misleading presentation on external platforms, tracking both issues and remediations.

Engage in multi-stakeholder dialogue with private platforms and AI developers

- **R.37.** Establish structured collaboration mechanisms such as working groups or memoranda of understanding (MOUs) with data aggregators, search engines, and AI assistants to improve

¹² Internationalization is the process of designing and building an IT system so it can be easily adapted for different languages, regions, and cultural conventions without requiring code changes.

¹³ This includes for example the [High-Level Group for the Modernization of Official Statistics \(HLG-MOS\)](#) of the United Nations Economic Commission for Europe, the Standards Working Group (SWG) coordinated by Eurostat for countries of the European Union, the [Statistical Information System Collaboration Community \(SIS-CC\)](#) coordinated by OECD, and the [SDMX.IO](#) project established under the Bank of International Settlements (BIS) Open Tech Strategy.

¹⁴ This would include, among others, the development of standardized metadata profiles: data and metadata structure definitions (DSDs/MSDs) for SDMX, and templates for DDI or ISO19100 series. See R.42.

¹⁵ For example, the PARIS21 Task Team on AI is maintaining an online [Catalogue of Generative AI Use Cases](#).

surfacing, ranking, and citation of official statistics and to define guardrails for summarization, transformation, and presentation on third-party platforms.

- **R.38.** Participate in technical exchanges on topics such as Model Context Protocol implementation and API harmonization, and co-create feedback loops and issue-resolution processes with success metrics like citation completeness, response accuracy, and time-to-correction.
- **R.39.** Advocate for equitable access to AI by promoting affordable tooling and API pricing models suitable for resource-constrained NSOs, including tiered pricing, volume-based discounts, and subsidized access for public-interest use cases.

4.2 Recommendations for International and Regional Organizations

International and regional organizations can accelerate AI-readiness across official statistical systems by convening stakeholders around common metadata standards, providing practical tools that raise baseline quality, and demonstrating responsible, cost-effective pathways for AI-enabled dissemination. Tools and reference implementations should be released under permissive open-source licenses and developed in line with global best practices for open collaboration, security, accessibility, and sustainability.¹⁶ Equally important, these organizations should invest in capacity building in low- and middle-income contexts, curate shared knowledge and repositories, and broker collaborations among NSOs, technology firms, trusted redistributors, and civil society to ensure equitable adoption and sustainability.

4.2.1 Metadata Standards and Metadata Quality Assurance

International and regional organizations should convene, align, and equip the official statistics community to raise metadata quality system-wide. This entails curating authoritative standards, providing ready-to-use templates and open-source tools, and embedding robust quality assurance so that metadata are interoperable with statistical business processes, and reliably usable by both humans and AI.

Promote and support the adoption of metadata standards

- **R.40.** Maintain an authoritative catalogue of specialized metadata standards (e.g., DDI, SDMX, DCAT-AP, ISO 19115),¹⁷ mapped to international frameworks such as the Generic

¹⁶ See for example the [Statistical Open-Source Software \(SOSS\) Guiding Principles](#) endorsed by the Conference of European Statisticians (CES), June 2025.

¹⁷ For example, Eurostat's SDMX-native data dissemination systems promote interoperability and straightforward consumption by generative AI systems. Thanks to the SDMX structure and the machine-actionable metadata it provides, Eurostat datasets can be integrated with a variety of generative AI solutions. Metadata for Eurostat datasets is also available in Semantic Web formats. The metadata can be mapped to the schema.org data standard using the DCAT-AP extension of the W3C data catalogue vocabulary for better indexing by search engines and data aggregators, such as data.europa.eu.

Statistical Business Model (GSBPM) and Generic Statistical Information Model (GSIM) and aligned with the FAIR principles, with guidance on applicability by data type.¹⁸

- **R.41.** Represent the official statistics community or support the participation of NSOs in relevant standard-setting bodies to ensure the official statistics community's priorities are reflected in standards maintenance and evolution.¹⁹ Include metadata standards adoption in financing and statistical capacity building projects.
- **R.42.** Publish ready-to-use, versioned metadata templates for efficient implementation of the metadata standards, with recommended taxonomies covering core statistical domains and processes.²⁰ Offer multilingual variants and guidance for national extensions to meet local legal and domain requirements without breaking interoperability. Include validation rules, conformance levels, and quality checklists in the templates to promote consistency and completeness. Distribute the templates through a maintained public repository and a registry service with APIs.
- **R.43.** Build and maintain open-source solutions with related guidance and training materials to enable the efficient creation and management of standards-compliant metadata (including free metadata editors and responsibly designed AI-based assistants to generate and enhance metadata).

Issue metadata quality assurance guidelines and tools

- **R.44.** In partnership with NSOs, define practical criteria, metrics, and workflows for metadata quality assessment (e.g., completeness, conformity, consistency, accuracy, specificity), aligned to endorsed standards, and operationalize an “AI-ready metadata” accreditation to signal machine-readiness and promote continuous improvement.
- **R.45.** Develop and openly share AI-assisted tools for metadata validation and enhancement, with clear documentation and training materials, enabling producers and curators to assess, remediate, and monitor metadata quality at scale and integrate these checks into routine dissemination workflows.

4.2.2 Tools and Guidance for Data Quality Control and Augmentation

International and regional organizations should lead by example in building, validating, and openly sharing tools that assess and strengthen the quality of official statistics. These solutions ought to be

¹⁸ [GSBPM](#) and [GSIM](#) are standards maintained by the United Nations Economic Commission for Europe. The [FAIR principles](#) are guidelines for making data Findable, Accessible, Interoperable, and Reusable.

¹⁹ This includes the [SDMX Technical Working Group](#), and possibly the [DDI Alliance](#), the [MLCommons](#) community, and others.

²⁰ Most metadata standards specify a broad set of elements to cover diverse use cases; no single organization needs all of them. Well-designed templates select the subset that is most relevant for a given context, set clear expectations for required and optional fields, and promote consistency, quality, and interoperability across catalogs. They lower the barrier to adoption while preserving alignment with the underlying standard. For instance, ISO 19115 has been operationalized through [INSPIRE](#) in the EU and [GEMINI](#) in the UK, which provide tailored templates for geographic metadata.

applied to the organizations' own datasets and released under open licenses, with clear documentation and reference implementations.

- **R.46.** Develop, validate, and maintain open tools for detecting point and collective anomalies in statistical time series across core NSO domains (macroeconomic, social, price, labor, and other indicators), with multilingual documentation. These tools are intended to be used by data aggregators and producers (see R.7).
- **R.47.** Provide responsible AI tools for anomaly triage, classification, and explanation that operate with human-in-the-loop review to help curators distinguish data errors from legitimate outliers, generate candidate explanations linked to contextual signals with provenance and confidence levels, and store validated explanations as value-level annotations in time-series databases using standard structures such as SDMX annotations to augment interpretability for both AI applications and human users (see also R.8).
- **R.48.** Deliver open-source tools and practical guidelines for quality control of microdata and geospatial datasets that interoperate with common survey platforms (e.g., CSPro, Survey Solutions, ODK) and GIS software.

4.2.3 Tools and Guidance for Data and Metadata Dissemination

Modern dissemination of official statistics demands high discoverability, multi-channel access, and data openness. This includes programmatic access to data and metadata through well-designed APIs and bulk downloads; advanced search that goes beyond keywords to semantic retrieval; and carefully governed AI-enabled interfaces (e.g., chatbots) that are safe, reliable, and cost-effective. International and regional organizations should lead by incubating, validating, and documenting solutions, then making them broadly accessible, especially to low-resource settings.

- **R.49.** Develop and publish practical guidance and open reference tools to improve search in data catalogs with a focus on semantic discoverability, including evaluation tools and benchmarks (e.g., LLM-as-a-judge alongside human assessment), reference implementations embedded in common dissemination software, and affordable options that leverage lightweight or low-resource AI models.
- **R.50.** Issue comprehensive guidance for responsible API design and deployment (including SDMX APIs) for data and metadata dissemination that specifies technical standards, versioning and deprecation, performance and reliability targets, security and privacy controls, licensing and citation practices, and clear human- and machine-readable documentation.
- **R.51.** Incubate AI technologies such as chatbots and MCP servers²¹ and publish actionable recommendations for safe, governed, and cost-effective implementation, including governance roles, human-in-the-loop review, evaluation and monitoring, incident response,

²¹ See the annex to this document for a brief introduction to the Model Context Protocol. Multiple international organizations, including Eurostat and the World Bank, are currently developing MCP infrastructures, and analyzing the feasibility of producing MCP manifests for their main data products.

and open-source MCP-enabled chatbot models tailored to official statistics with instructions for responsible adaptation and deployment.

- **R.52.** Provide practical guidance and modular tooling for data privacy and statistical disclosure control in AI-enabled dissemination, aligning with legal and ethical requirements while supporting privacy-by-design, risk assessment, and transparent documentation of protections and residual risks.
- **R.53.** Maintain a central, standards-based metadata catalog of official and trustworthy data sources with advanced search and resolvable links to authoritative origins, underpinned by clear accreditation criteria for availability, metadata conformance, quality assurance, accessibility, and dissemination practices.²² Also, maintain a public registry of MCP servers relevant to official statistics, including service metadata, supported standards and versions, and conformance indicators, to enable reliable discovery and safe, automated interaction between LLMs and authoritative repositories of official statistics.

4.2.4 Build AI Capacity in Low- and Middle-income Countries

International organizations have a critical role in coordinating and sustaining efforts to close AI-related capacity gaps among NSOs, particularly in low- and middle-income countries and other resource-constrained contexts. Without targeted, long-term support, uneven access to skills, infrastructure, and governance frameworks risks widening global disparities in the ability of statistical systems to benefit from AI and data innovation. International organizations should therefore lead structured, collaborative capacity-building initiatives that strengthen technical skills, institutional readiness, and the sustainable adoption of AI-enabled practices.

- **R.54.** Design, finance, and coordinate dedicated multi-year training and technical assistance programs focused on AI readiness. These programs should cover metadata standards and metadata quality management; data quality assurance and control; modern data dissemination practices including API-based dissemination and improved data discoverability; data licensing; and practical competencies such as data engineering, responsible AI, and data governance.
- **R.55.** Support the creation of open, modular training materials and self-paced courses that can be reused and adapted by NSOs. Where possible, localize content to reflect languages, regional priorities, and relevant examples and case studies, and ensure alignment with international standards and recognized best practices.
- **R.56.** Encourage and support the establishment of regional communities of practice and peer-support mechanisms that enable NSOs to exchange experiences, accelerate implementation, and sustain capacity beyond initial training. Promote mentorship and partnership arrangements that link more advanced NSOs with those still developing their AI and data capabilities.

²² The Swiss Federal Statistical Office is proposing to establish a global [Trusted Data Observatory](#) (TDO) as a central metadata hub, making reliable information easier to find, understand, and use.

4.2.5 Catalyze Collaborations and Partnerships

To strengthen the AI readiness of NSOs, international and regional organizations should actively catalyze collaborations that bring together statistical expertise, technological innovation, and responsible data stewardship. Effective partnerships among NSOs, international and regional bodies, technology firms, research institutions, civil society, and trusted redistributors of official statistics can accelerate solution development and extend the reach of official data. International and regional organizations are uniquely positioned to convene these actors and structure engagement around shared principles, standards, and public-interest objectives.

- **R.57.** Establish a sustained program to capture, validate, and disseminate lessons from NSOs and other distributors implementing AI-enabled dissemination. The program should systematically identify, document, and publicize AI solutions that have demonstrated value and can be responsibly deployed by NSOs, and share case studies, failure analyses, templates, reference architectures, and reusable open-source assets through maintained open repositories and knowledge bases. It should also include a global monitoring framework to track the impact of AI adoption on the visibility, accuracy, and use of official statistics, supporting peer learning and replication across countries and regions (see also R.34).
- **R.58.** Organize regular interactions between NSOs, international organizations, and AI leaders, including major technology firms, through conferences, thematic forums, and technical workshops. Anchor these engagements in advocacy for open, safe, interoperable, and standards-based approaches to data management and dissemination.
- **R.59.** Encourage the responsible redistribution of official statistics by third parties to extend their reach while preserving integrity and trust. Develop and publish technical guidance for redistributors that articulates good practices for reuse, including proper citation and attribution, retention and propagation of metadata, versioning and provenance, transparency about transformations, quality controls, and an explicit duty of care toward official data.²³
- **R.60.** Collaborate with standard-setting organizations responsible for data licensing instruments, such as Creative Commons, to evolve standard licenses by incorporating clearer and more robust provisions that address AI-facilitated redistribution, safeguard licensors' interests, and preserve the integrity and intended use of official statistics.

4.3 Recommendations for Redistributors of Official Statistics

Entities such as internet search engine operators, chatbot platforms, and data aggregators (including certain international organizations) increasingly shape the ways in which individuals access and engage with official statistics. These intermediaries exert substantial influence on the visibility,

²³ See for example *Redistribution of Official Statistics Published Under Open Data Licenses - Permissions, Obligations, and Recommendations* by FAO, WIPO and the World Bank (2025).

interpretation, and perceived reliability of statistical information. Consequently, they share responsibility for upholding the accuracy, integrity, and credibility of the data they disseminate.

4.3.1 Prioritize Authoritative Sources and Respect Licensing Terms

To preserve accuracy, timeliness, and trust, redistributors of official statistics should make authoritative official sources the default wherever such data exist, and ensure that provenance and attribution are maintained across all dissemination channels, including search results, catalogs, visualizations, and AI-generated outputs. Responsible redistribution further requires strict adherence to the letter and the spirit of open data licenses and use policies.

- **R.61.** Give systematic prominence to data and statistics from accredited official producers such as national statistical offices and international statistical organizations across search, catalog, ranking, and recommender systems, and ensure that AI-generated answers rely on trusted official data whenever available. This requires mechanisms to unambiguously identify official and other trusted sources; curated catalogs of accredited providers or data commons maintained by international or regional organizations can support this objective (see R.53).
- **R.62.** Acquire data and metadata directly from official sources through APIs or bulk download services, with automated synchronization to preserve freshness and provenance. Use advance release calendars published by NSOs where available (see R.23). Track and expose dataset versions, timestamps, and synchronization dates, propagate corrections without delay, and clearly indicate when content has been superseded.
- **R.63.** Embed attribution, persistent dataset identifiers, and links to canonical records and documentation across all delivery channels, including dashboards, visualizations, and AI-generated responses. Implement attribution as specified in the applicable license, respect non-endorsement clauses, and comply with any integrity or no-misrepresentation requirements. Do not imply endorsement by official producers, and do not cite a source unless content is obtained directly from that source or an authorized channel, recognizing that second-hand data may be incomplete, altered, or outdated.
- **R.64.** Clearly disclose any processing, harmonization, aggregation, or synthesis applied to official statistics. Provide change logs and, where feasible, reproducible workflows or methodological descriptions to distinguish derived outputs from original official statistics and avoid user confusion.
- **R.65.** Require explicit citations for every numeric claim produced by AI systems and, where technically feasible, provide a “Verify in source” option that allows users to view the exact figure within the corresponding official dataset. This strengthens traceability, supports user verification, and reinforces confidence in AI-enabled dissemination.

4.3.2 Ensure Statistical Integrity in AI-enabled Dissemination

AI-enabled dissemination can significantly expand access to and understanding of official statistics, but it also introduces new risks, including distortion, hallucination, loss of provenance, and

decontextualized interpretation. To safeguard statistical integrity, entities that redistribute official statistics through AI-enabled systems should design their architectures and workflows to anchor generation and summarization in authoritative sources, preserve links to underlying data and metadata, and provide users with transparent context. Robust technical and operational safeguards are essential to prevent misrepresentation, support traceability, and enable timely detection and correction of errors.

- **R.66.** Use retrieval-based or retrieval-augmented generation that queries canonical databases and validated metadata from NSOs and other trusted data providers. Expose persistent identifiers linking AI-generated answers, narratives, or visualizations to the underlying datasets or series, and implement automated validation checks that flag or block outputs containing unverifiable or inconsistent figures. Adopt protocols that provide verifiable assurances of data authenticity and integrity, such as Proof Carrying Numbers (PCN),²⁴ to enable downstream systems and users to validate the origin and integrity of official statistics used in AI-enabled dissemination.
- **R.67.** Make methodological notes, information on limitations, and other relevant quality information visible to users wherever applicable. Ensure that series-level and cell-level annotations provided by data producers are retained and accessible in AI-generated outputs when such metadata exist in the source systems.

4.3.3 Establish Feedback and Collaboration Mechanisms with Data Producers

Effective redistribution of official statistics depends on sustained, two-way collaboration between redistributors of data and NSOs. Continuous feedback loops are essential to preserve data integrity, ensure timely updates, and improve how statistics are interpreted and presented to users. Collaboration between data providers and redistributors will help ensure that official statistics remain accurate, relevant, and appropriately contextualized across third-party platforms.

- **R.68.** Designate points of contact and participate in relevant community forums to maintain regular dialogue with NSOs and international organizations. Invite NSOs to contribute to red-teaming and testing exercises where appropriate. Provide well-defined channels for NSOs to submit feedback or report issues, and implement transparent, timely processes for reviewing and correcting errors or misrepresentations in third-party data dissemination platforms.
- **R.69.** Provide NSOs with periodic summaries of how their data are accessed and used in third-party platforms, including indicators such as query volumes by dataset or indicator, trends in user interests (e.g. topics or geographies), and high-level metadata or search interaction signals. These insights can support NSOs in prioritizing data production, dissemination, and metadata improvements.
- **R.70.** Work with NSOs to enhance how official statistics and their metadata are indexed, interpreted, and displayed on third-party platforms. Redistributors should leverage the

²⁴ See A. Solatorio (2025) *Proof-Carrying Numbers (PCN): A Protocol for Trustworthy Numeric Answers from LLMs via Claim Verification* (<https://arxiv.org/abs/2509.06902>)

extensive experience of NSOs in data dissemination and international best practices to ensure that official statistics are presented accurately, consistently, and with appropriate context by the redistributor.

4.4 Recommendations for the Development of AI Technology

The official statistical community is not a primary actor in the development of generative AI technologies; however, it is a critical stakeholder and potential partner. NSOs and international organizations define authoritative standards, governance frameworks, and statistical methodologies, and they provide high-quality “ground-truth” data and statistics—most often free of charge—that are used to train and inform AI systems. In this context, the following recommendations are directed at entities that develop generative AI models and platforms. They set out practices intended to ensure that the use of official statistics in AI systems is responsible, interoperable, ethical, and sustainable, while enabling AI developers to benefit from the expertise and public-good infrastructure of the official statistical community.

- **R.71.** Collaborate with the official statistical community, e.g., by inviting national statistical organizations and international organizations to advisory programs and co-running pilot projects. This will help ensure that AI guardrails are grounded in established statistical principles and real-world practices, thereby preserving data integrity, methodological soundness, and public trust in AI-generated outputs.
- **R.72.** Align with open metadata and data-exchange standards and governance frameworks adopted by the official statistical community, such as SDMX. These standards ensure interoperability, consistent interpretation of indicators, traceability to authoritative sources, and transparent governance, all of which are essential for trustworthy, reusable, and policy-relevant AI outputs.
- **R.73.** Respect licensing and ethical constraints on official data, implementing privacy-preserving analytics, and document how official statistics are used in AI training. This is required to ensure legal compliance, protect confidentiality, uphold ethical standards, and maintain transparency and trust in AI systems that rely on public data.
- **R.74.** When acquiring official statistics and related metadata to train models, adopt data-collection practices that minimize load on official statistics providers and preserve data quality and provenance. Respect site-specific terms of use and licenses. Prefer provider-supported channels wherever available (including machine-readable endpoints and bulk download packages) over crawling. Consider co-developing or sponsoring the development of high-volume access options that reduce strain on NSO websites.
- **R.75.** Support low-resource AI deployments through lightweight, open models, affordable API pricing tiers for public institutions in low- and middle-income countries, and other forms of assistance to resource-constrained NSOs. This will help ensure equitable access to AI capabilities, sustain the public-good data ecosystem on which AI systems depend, and foster a fair exchange between AI developers and the NSOs that provide free and open access to data and statistics.

5 Conclusion

Advancing towards AI-ready official statistics demands collective engagement. National statistical organizations, international and regional bodies, data redistributors, and technology innovators each have a role to play. Progress will depend less on individual initiatives than on the ability to align standards, infrastructure, and practices across this ecosystem. This calls for immediate, coordinated action to establish an AI-enabled dissemination environment that broadens access to official statistics and safeguards their integrity as a public good.

Annex: The Model Context Protocol

Emerging technologies such as the Model Context Protocol (MCP) introduced by Anthropic²⁵ offer promising solutions to address some of the current limitations of AI dissemination of official statistics. MCP enables large language models to interact more effectively with structured data and tools. To improve their accuracy, LLMs need a way to interact with trustworthy, up-to-date sources such as the official statistics published by NSOs. MCP makes this possible.

MCP is a **protocol** that acts like a universal language to provide AI systems with dynamic access to tools, resources, and prompts. It can describe in a clear and standardized way how a large language **model** can interact with **context** such as databases and tools made available online through APIs. For example, an NSO can use the MCP to describe what their API-based data services can do: how to search a dataset, extract numbers, perform calculations, return visualizations, format data outputs, etc.

The MCP instructions are written and controlled by the data provider, ensuring there is no guesswork and greatly reducing the risk of misinterpretation or misuse. These instructions are published on one or multiple publicly accessible **MCP servers**, where they can be discovered and used by AI systems (the **MCP host application**) that operate **MCP clients** that each interact with one MCP server (Figure 1). The MCP streamlines external context integration for host application developers by eliminating the need to create and maintain separate, customized AI agents for each context provider. Since MCP servers access content (data and tools), they must be implemented securely to prevent risks like data leakage or unauthorized action by AI agents.

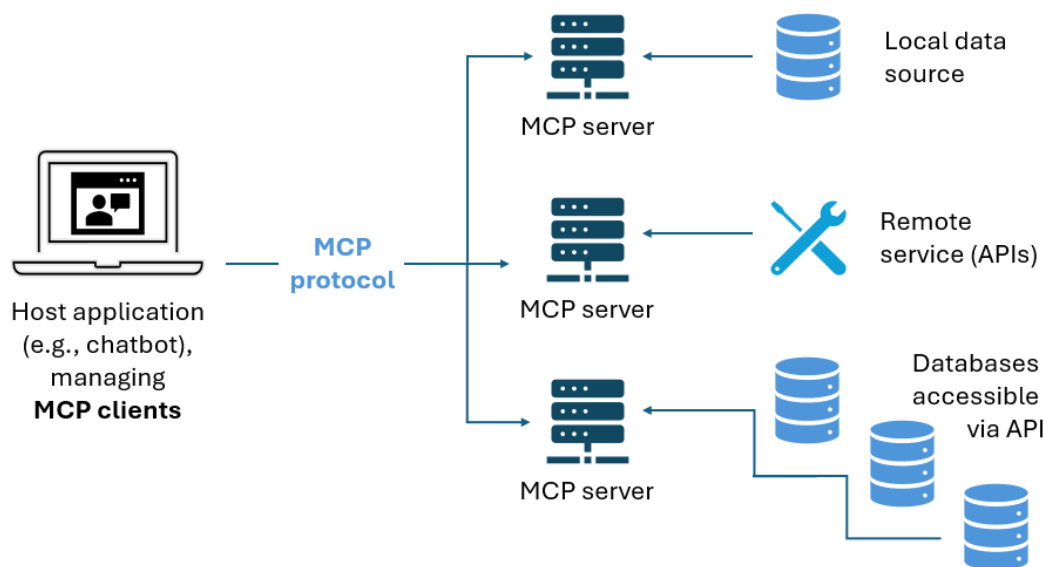


Figure 1 - The architecture of MCP systems

²⁵ See <https://www.anthropic.com/news/model-context-protocol>

When an AI application such as a chatbot receives a user query, it can check the MCP server to see which tools and data services are available and how to use them properly. If the AI system perceives that one of the tools in an MCP server matches the needs of the query (for example, retrieving statistical data), the AI system will:

1. Enhance, if necessary, then break down the user's question into smaller tasks,
2. Use the MCP to generate a step-by-step workflow based on the declared capabilities of the tool,
3. Execute those tasks by calling the tool in the way defined by the MCP instructions,
4. Return the output to the LLM for elaboration and displaying in the appropriate format, where relevant.

If no MCP server tool meets the query's needs, the user will be informed that an answer is unavailable. This process ensures that the data is retrieved directly from official sources, using the correct methods, and that the results are relevant, trustworthy, and up to date.

In short, MCP is an open, decentralized, and tool-agnostic solution. It allows any data provider such as an NSO to publish tools for disseminating data and metadata that LLMs can discover and use at runtime. This makes it particularly well-suited for scalable, trustworthy, and modular interaction between LLMs and the ecosystem of official data. MCPs give statistical agencies a way to make their data "AI-readable", ensuring that future AI systems can access, understand, and use official statistics safely and responsibly, directly from the source.

The deployment and maintenance of MCP servers by NSOs, and the possible creation and maintenance of a central registry of MCPs for trusted data, can contribute to promoting the use of official statistics.

But MCP is only part of the solution. Unlocking the full potential of MCP will require technical investment in the foundations of such systems, i.e., in machine-actionable standardized metadata, and in standardized APIs that make official data more accessible to AI systems. NSOs will also need to modernize their platforms with semantic web technologies to enhance the discoverability of their data in AI-powered environments. For a statistical agency, implementing semantic searchability of its data assets means enabling users to find datasets based on meaning and context, not just exact keywords, by leveraging machine learning techniques like embeddings that represent concepts and relationships in vector space. This requires rich, structured metadata and, ideally, the consistent use of standard taxonomies and ontologies. Together, embeddings and standardized metadata allow queries to match conceptually related content (such as "maternal health" aligning with indicators like maternal mortality rate and skilled birth attendance), and support multilingual and cross-domain discovery, making the NSO's data more discoverable.