

# ECCBSO Task Force on AI-based Data Extraction (TF AIDE)

**Initial focus:** The Task Force will initially concentrate on AI-supported extraction of ESG-related information from corporate reports, with a view to developing methodologies that can subsequently be extended to other use cases, such as financial indicators, governance information and risk-related disclosures, as well as to other statistical and analytical domains. Depending on members' interests and priorities, the Task Force may decide to work collectively on a common topic or pursue different use cases that deliver the greatest value to participating central banks.

**Proposed launch: Q4 2026 | Indicative completion: Q3 2027**

## 1. Background and rationale

Corporate financial statements, annual reports, sustainability reports, management reports and regulatory filings contain valuable information that remains only partially exploited because conventional extraction is resource-intensive. Advances in natural language processing, machine learning, large language models, OCR and document intelligence can support the identification, extraction and structuring of this information, subject to appropriate validation, traceability, security and human oversight.

ECCBSO institutions are at different stages of experimentation. Coordinated work can help institutions address common challenges, avoid duplication, compare approaches and methodology and accelerate learning through the exchange of experiences, methods, tools, evaluation results and governance practices. The Task Force will provide a framework for identifying areas of common interest, assessing different approaches and exploring opportunities for collaboration where they create value for participating institutions.

The Task Force will start with ESG indicators as a policy-relevant and demanding use case, while retaining the possibility of extension to financial variables, risk, governance, innovation, digitalisation and other business statistics.

## 2. Objective

The TF AIDE will establish a temporary cooperation framework in which participating institutions jointly develop a common, reusable and scalable approach for extracting information from corporate reports using AI. By collaborating, participating institutions can pool expertise, avoid duplicating experimentation, establish common standards and evaluation methods, and create reusable assets that would be costly and less effective to develop individually. To achieve this, where feasible, participating institutions identify opportunities for shared standards, reusable assets, or technical components, while recognizing that implementation is likely to remain institution-specific.

## 3. Lead questions

The lead questions are organised around three dimensions: Goal Setting, Methodology, and Technology & Tooling. This reflects the need to first establish a common objective and priority area, then assess appropriate methodological approaches, and finally explore the technologies and implementation models that can support those objectives.

### A. Goal setting

- Which business or statistical challenges are most relevant and shared across participating institutions?
- Which categories of information from corporate reports would create the greatest value if extracted using AI?

### B. Methodology

- Which extraction approaches provide sufficient accuracy, completeness, traceability, reproducibility and scalability?
- How should reference data, quality metrics, confidentiality, human review and audit trails be designed?
- How should AI-extracted information be validated and integrated into statistical or analytical workflows?

### C. Technology & Tooling

- Which technologies, models and tools are most suitable for different extraction use cases?
- Which legal, confidentiality, copyright, security, model-risk and AI-governance considerations must be addressed?
- Which assets, standards or technical components can realistically be shared?
- How can coordination help reduce duplication and improve compatibility across institutions with different tools and systems?

## 4. Mandate and scope

The Task Force will provide a framework for collaboration among ECCBSO institutions on the application of AI to corporate reporting data. The emphasis will be on defining common priorities, exchanging knowledge, comparing methodological approaches and promoting best practices. Where feasible and beneficial, this collaboration may also lead to the development of reusable assets, shared standards or other forms of technical cooperation.

- Take stock of existing and planned institutional initiatives, source documents, target variables, methodologies, tools, architectures and governance arrangements.
- Prioritize ESG-related use cases as an initial area of work, while identifying other common modules, requirements and use cases of interest to participating institutions, including financial, governance, risk, innovation and digitalisation indicators.
- Compare methodologies and document quality, limitations, failure modes, resource requirements and circumstances requiring human intervention.
- Facilitate the exchange of experiences, lessons learned, governance practices and methodological approaches across participating institutions.
- Explore opportunities for developing common standards, evaluation frameworks, taxonomies, validation protocols and documentation to improve comparability and interoperability.
- Where sufficient overlap exists in technology choices and institutional requirements, assess opportunities for sharing reusable assets such as code, prompts, schemas, test protocols or reference components.
- Assess the benefits and feasibility of centralised, federated and institution-specific implementation models, recognising that adoption pathways may differ across institutions.
- Identify additional use cases and future collaboration opportunities and prepare progress updates and a final report for the ECCBSO.

## 5. Cooperation model

The Task Force will focus on identifying and addressing common challenges related to the use of AI for extracting and analysing information from corporate reports. Participating institutions will collaborate to establish common principles, methodologies and evaluation approaches that can be adapted to their respective environments and use cases. These use cases will serve as a vehicle for methodological learning and collaboration.

## 6. Deliverables

| Ref. | Deliverable     | Content                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Target                                           |
|------|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|
| D0   | Survey          | Prior to the kick-off, a survey will be conducted to identify participating institutions' interests, current initiatives, priorities and needs.                                                                                                                                                                                                                                                                                                                     | Q4 2026                                          |
| D1   | Inventory       | Mapping of institutional needs, ongoing initiatives, methodological approaches, governance practices and collaboration opportunities across participating institutions.                                                                                                                                                                                                                                                                                             | Q1 2027                                          |
| D2   | Reusable assets | Common principles, evaluation criteria, best practices, governance considerations, reference architectures and implementation guidance for AI-supported extraction from corporate reports. This may include, in particular, reusable methodologies, taxonomies and extraction frameworks for ESG indicators and disclosures, as well as, where feasible, reusable technical components and documentation that can be extended to other corporate reporting domains. | Q3 2027                                          |
| D3   | Final report    | Findings on common challenges, methodological approaches, governance considerations, implementation options, and recommendations for future collaboration and knowledge sharing across ECCBSO institutions.                                                                                                                                                                                                                                                         | Q3 2027<br>ECCBSO_IFC_NBB<br>Conference 21/10/27 |

## 7. Indicative timeline

| Period  | Phase                   | Main activities and milestones                                                                                                                                                                                                                                                                                                                      |
|---------|-------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Q4 2026 | Launch                  | Approval of mandate; expression of interest; nomination of experts; in person kick-off; First review mandate together and agree on plan and then launch survey of participating institutions to identify priorities, current initiatives and areas of common interest; And agreement on workstreams and priorities.                                 |
| Q1 2027 | Stocktaking             | Inventory of institutional experiences, reports, use cases, indicators, methods, tools, governance and shareable assets. Delivery of D1.                                                                                                                                                                                                            |
| Q2 2027 | Methodology & framework | comparison of approaches; assessment of quality and portability; governance considerations and implementation requirements; identification of reusable assets. Delivery of D2.                                                                                                                                                                      |
| Q3 2027 | Consolidation           | Final assessment; Development of common guidance, methodological framework and implementation considerations; assessment of collaboration and standardisation opportunities; final recommendations and roadmap for future cooperation. centralised, federated and autonomous models; reusable package; recommendations and roadmap. Delivery of D3. |

## 8. Organisation, principles and success criteria

The TF will be temporary, led by a Chair or Co-Chairs appointed by the ECCBSO. Relevant business, statistical, data-science, IT and governance experts may participate, where data science should be in the core project group and business and data engineers should be included for specific tasks. Work will rely mainly on virtual meetings, written exchanges and practical workstreams. Contributions and sharing will remain subject to institutional rules, access restrictions and licensing conditions.

The work will follow traceability, security, reproducibility and technological-neutrality. Success will be demonstrated by documentating the common methodological agreements based upon the use cases, sharing evidence on portability and scalability, making reusable assets available to the greatest extent feasible, and disseminating actionable recommendations for future ECCBSO cooperation.

**Coordinated experimentation should be transformed, wherever feasible, into reusable collective assets that reduce duplication, accelerate innovation and strengthen synergies across the ECCBSO community.**