Artificial intelligence thrives on structure. Agency-specific FDTA rules need to make sure the data they collect is AI-ready.

AI shines when given access to richly contextualized, standardized, structured data and identifiers. While Large Language Models (LLMs) can source any kind of data, structured or unstructured, for financial analysis and research that demands a high degree of accuracy, *structure and standardization are crucial.*

Here's why open data standards and identifiers drive better AI.

Context matters. Contextual information like definitions, data types, dimensional characteristics and understanding the relationships between facts, give AI the information needed to go beyond surface-level processing. It helps machines understand the intent and behavior of information.

Structure counts. Consistently defined information is easier for AI to understand and process. By contrast, unstructured information, by its nature, is more difficult for anyone to understand.

Standards communicate. The meaning of reported data is captured when it is expressed in a uniformly designed semantic data model like a taxonomy or ontology.

The Legal Entity Identifier (LEI, ISO 17442)) and eXtensible Business Reporting Language (XBRL) together power improved results for regulatory disclosure programs.

As U.S. regulators develop digital asset market infrastructure regulation, GLEIF and XBRL US encourage agencies to incorporate statutory language on the use of identifiers and open data standards that express semantics. This approach is critical to promote efficiency and transparency, and to set agencies up to capitalize on the increasing capabilities of AI to boost efficiency and reduce costs.

Incorporating references to XBRL and to the LEI or the Financial Data Transparency Act's (FDTA) common data standard/identifier language into rulemaking will give regulators direction on the intersection between the legislation and the FDTA. Clearly naming the standards and identifiers required will also give digital market participants, from reporting entities to data consumers to software providers that support the ecosystem, clarity from the start on how to prepare and effectively comply.

XBRL US

GLEIF

The Legal Entity Identifier

The LEI system offers a consistent and reliable dataset of legal entities, which is crucial for training and testing AI models to ensure they produce accurate and reliable results across different jurisdictions.

Improves entity identification: AI models can use the LEI to accurately identify and distinguish between entities, even those with similar names. The LEI is linked to key reference data like the official name, registered address, and country of formation, which helps in identifying entities for due diligence and verification.

Enhances fraud detection and risk assessment: By linking LEIs with other datasets, like beneficial ownership and sanctions data, AI can create more effective systems to detect financial crime and risk.

Automates compliance and due diligence: LEIs significantly speed up AI processes that require entity identification, such as know-your-customer (KYC) and anti-money laundering (AML) checks, by automating due diligence and compliance tasks.

Simplifies complex data analysis: The LEI system helps AI models analyze complex ownership structures. For example, by integrating with LLMs, AI can extract and analyze information from unstructured documents to map ownership structures more accurately.

Enables AI tools with standardized inputs: AI can use the Entity Legal Forms standard (ELF, ISO 20275) to automate the assignment of a unique code to an entity's specific legal form, ensuring consistency and reducing the need for local knowledge.

The XBRL Data Standard

XBRL is an open (free) semantic data model that unambiguously and reliably defines the meaning of data, and relationships between data. U.S. regulators have collected XBRL data from banks, utilities, investment management and public companies for 20 years, a resource optimized for AI.

Increases accurate machine-learning and analysis: Quantitative research[1] proves that structured (XBRL) data generates more accurate results than unstructured when performing financial analysis.

Boosts the ability to detect fraud, and track and monitor risk: Uniform reporting frameworks that standardize fraud data enable sharing, storing and interoperability of data across agencies and businesses; provide consistent data for AI tools.

Semantic data model reduces ambiguity and hallucinations: The consistent structure and context enabled by XBRL taxonomies makes it easier for AI to do its job faster, at lower cost, and more efficiently.

Encourages trust: XBRL prepared data is more reliable. It is based on agreed upon definitions established through authoritative accounting or reporting standards, can be automatically validated, and is produced/vetted by the reporting entity. Unstructured third-party sources are ubiquitous across the Internet, but not as dependable.

Enables the creation of knowledge graphs: Structured, standardized data produces machine-readable knowledge graphs that assist in complex data analysis and provide the most reliable framework for AI systems.

[1] AI determinants of success and failure: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5316518.

XBRL and the LEI are a powerful combination that fuel more robust and trustworthy AI analysis, removing the guesswork and inconsistencies that arise from unstructured data.

XBRL: email info@xbrl.us LEI: email info@gleif.org