MarkLogic®

MARKLOGIC: MIFID II – SOLVING THE DATA CHALLENGE

MARKLOGIC WHITE PAPER · NOVEMBER 2016

This white paper addresses one of the key regulations the top executives at financial institutions have on their agenda: MiFID II. What are the key data challenges and what technology is needed to solve them? These and many other questions are answered in our publication.

EXECUTIVE SUMMARY

The regulatory landscape is becoming more and more complex with new mandates being introduced across various jurisdictions. One of the most significant regulations in financial markets is the MiFID II directive coming into effect on the 3rd January 2018, following a 12 month delay agreed by legislators in June 2016 from the original date of 3rd January 2017.

MiFID II is an EU regulation that is designed to harmonize regulations for investment services, securities trading and processing in primary and secondary markets. It is largely focused on market integrity and encourages fair and orderly market functioning to reduce risk and protect everyone involved.

In terms of data governance and compliance readiness, this regulation impacts transaction reporting of all financial instruments traded in Europe, post-trade transparency and real-time data delivery. It also requires better execution with a specific need to be able to reconstruct past events and provide all types of communications related to a transaction.

In this whitepaper we will look at the wider regulatory landscape and what this means for financial market participants, explain key data challenges related to MiFID II as well as propose a solution architecture which will help address the needs of affected institutions and expedite compliance.

REGULATORY COMPLIANCE IN CONTEXT

The financial services industry is under constant pressure from financial regulation bodies globally to ensure the integrity of the financial system. Reporting challenges can put enormous pressure on financial institutions, and consume considerable resources internally. It is therefore vital to establish a sound application framework that will enable your organisation to respond to regulators in an efficient, timely, and responsive manner. The volume and variety of data to be processed and analysed, the complexity of the analytical processes, and the urgency of the reporting requirements necessitate a step change in technology.

The primary policy goals are stability in the global markets, integrity and solvency of financial institutions, mitigation of capital risks, optimality of capital allocation, consumer and investor protection and fraud prevention across jurisdictions. Secondarily, standardization and compatibility of regulatory processes and methodologies and reporting hierarchies are needed to facilitate global collaboration. The tools and levers of enforcement vary but are limited and include process-specific procedural guidelines, data governance and validation criteria and metrics alongside tighter reporting rules that include specification of content and format.

> " The prospect of adjusting swiftly and easily to new situations would mark a significant and welcome change from banking as usual. Firms have tended to organize themselves as collections of mostly autonomous operating units, so-called silos, rather than as single, cohesive entities. This necessarily creates a constrained, limited view of the world that has led them to repeat the same mistakes in assessing the operating environment and especially the risk outlook at what turn out to be the most perilous times."

Wolters Kluwer Financial, 'Intelligence: A Tech Revolution for the Evolution in Compliance'

> " When technology risks materialize, the financial, regulatory, and reputational implications can be severe… Regulators penalize firms for noncompliance—from data breach–related fines to mandated remediation activities… To manage these risks, many banks simply deploy their considerable IT expertise on patching holes, maintaining systems, and meeting regulations."

McKinsey Quarterly, "The Ghost in the Machine: Managing Technology Risk", July 2016

The first steps to compliance involve addressing data management – specifically, the flow of the data, validation of the data and including a variety of content in the flow of information to the decision makers and outside regulators. There are two significant challenges organisations face in this regard:

- The evolution of the IT infrastructure in the industry since the client-server revolution has led to proliferation of systems and fragmentation of data.

- The rapid rise of social media, instant messaging, forum usage, unstructured data as a source of new content and trader behaviour analytics have increased the amount of information that grows outside transactional system but must be included for both business planning and compliance.

Hence, the response to the waves of regulatory requirements will have to include tackling the data management infrastructure – and this will have to start with moving the data out of the silos and also incorporating the content that is generated by social media. The acute data intensive challenge from a MiFID II point of view consists of not just increase in data volumes and rates of updates but of completeness, accuracy and timeliness.

COMPLIANCE READINESS CHECKLIST

Can your trade/transaction reporting platform report across multiple sources of data in real time?

Can it track the provenance of information through your reporting regime?

Are you able to explore and query the data you have aggregated, to answer business questions as well as new regulatory requests?

Does your platform offer the flexibility to adapt to evolving regulatory requirements?

Does your MiFID II reporting solution represent another data silo within the bank?

How reusable and extensible is the logic that you have built into the reporting system?

Does your solution empower the IT staff or Business Analysts?

MarkLogic

KEY DATA CHALLENGES AND REQUIREMENTS IN THE NEW REGULATORY REGIME

DATA MANAGEMENT: NO DATA LEFT BEHIND

Siloed controls and inflexible, legacy technologies are in most cases unable to handle multiple sources of data. And, under MiFID II the number of fields for transaction reporting has significantly increased. Organisations who select newer generation database technology – that allows them to ingest the data as-is and reconcile the data in place – will find significant advantages and expedite the route to compliance.

COMMUNICATION DATA SUPERVISION & RECONSTRUCTION

Companies must demonstrate effective oversight and control over policies and procedures which govern all communications. Moreover, MiFID II requires organisations to supply regulators with communications associated with a specific trade, and to be able to reconstruct history of trade cycle events.

Knowing what information you knew and when and how it has changed over time – a *bitemporal* view of data lineage – becomes a critical component of your regulatory reporting infrastructure.
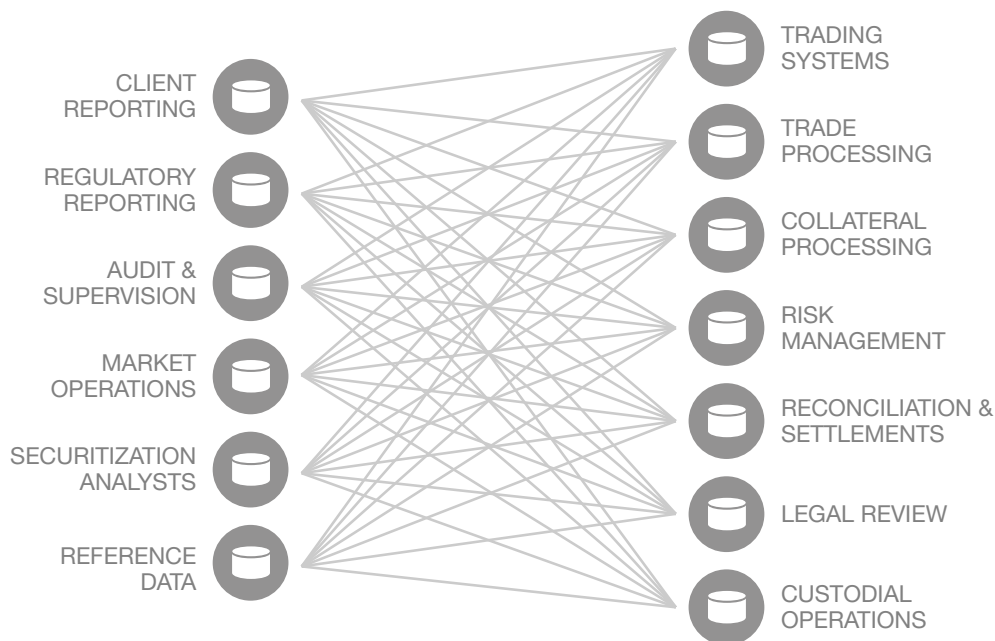
DATA STORAGE

Companies must make records available to clients for a retention period of five years and for up to seven years for regulators. The need for cost-effective and secure data storage has never been greater especially when jurisdictional limitations need to be considered. Some jurisdictions may enforce confidentiality rules and restrict information sharing among parties; in those cases identifiers will need to be used.

DATA QUALITY AND DATA LOGIC CHECKS

Financial institutions have to be able to generate the right information for the regulator, creating the need for a compliance library and logic which can be reused. With many consumers of information internally and externally, logic which defines fields for reporting is often duplicated across reports.

Rather than encourage an army of business analysts to re-process information for each data delivery, a better approach is to store and re-use methods and program code and associate both to the data sets and results used with timestamps. The methods and code can be updated over time independently of data updates.

CLIENT REPORTING

REGULATORY REPORTING

AUDIT & SUPERVISION

MARKET OPERATIONS

SECURITIZATION ANALYSTS

REFERENCE DATA

TRADING SYSTEMS

TRADE PROCESSING

COLLATERAL PROCESSING

RISK MANAGEMENT

RECONCILIATION & SETTLEMENTS

LEGAL REVIEW

CUSTODIAL OPERATIONS

MAKING YOUR COMPLIANCE SOLUTION A BUSINESS ANALYST-LED FRAMEWORK

It is important to institute the processes in place which can be managed by non-technical analysts. This will have significant cost-saving implications as well as ensure operational efficiency without the need to re-architect the regulatory data centre when the regulatory rules evolve.

The right technology can provide a number of benefits, including:

- Consolidation of all data, structured and unstructured, with indexing and searchability
- Bitemporality for auditing and tracking
- Discovery of new and emerging relationships

But these benefits only become realized when users can access the data they need, in the many ways they need to see it – using tools that they are comfortable with. So, the right platform has to be easy to integrate into any front-office business intelligence applications.

SOURCING THE DATA JUST ONCE

Within a distributed architecture, calculations can be duplicated in multiple databases and applications. Accuracy can be achieved by placing all data into one operational database.

Taking a risk calculation as an example, customer reference data stored in a single location can be enriched with the products of multiple business processes. In risk analytics, the same data set on a traded portfolio may be used by several users with different models.

The data serviced by a shared data platform that marks the record of reference set for each method and time would be prudent governance. The method information and the results may also be returned to the same database with annotations and associations establishing time history of analytical results, reports and lineage inbound data as one unit of work subject to review, audit and certify.

QUERYING THE DATA AND EXPORTING THE INFORMATION

A traditional data warehouse is extremely difficult to query. Arguably, only a multi-model technology

can ensure the data can be explored without losing the richness of the original source. Only by storing information in the highly optimised models of the source systems can we maintain the flexibility to explore the richness of the original form of the data.

In traditional approaches, in-bound data is converted across formats and source models are flattened so they can be retained; this creates vast flat tabular structures that may be conducive to spreadsheet operations and may be filtered. However, with these traditional approaches we cannot ask new questions, nor can we identify what we discarded in the 'flattening' process.

A multi-model approach provides the ability to explore information without necessarily having to map and transport these flattened models.

E-DISCOVERY, ARCHIVING AND RECORD RETENTION ACROSS ALL SORTS OF DATA, STRUCTURED AND UNSTRUCTURED.

Retail banks are very familiar with the issue of discovery, but investment banks have a serious challenge, with many messaging and data formats and the need to identify information across not just business silos but information silos (email, messages, texts, office documents etc.). How do I really identify what I knew when, and how can I prove it? Traditional databases store unstructured data as BLOBs which, as a unit of information, is not sufficiently granular and cannot resolve complex search and discovery questions.

Even after the search problem is solved, managing this information as part of an information life cycle management (ILM) process becomes singularly important to prevent operational costs escalating. The retention requirements vary; trade transactions, depending on asset class and venue, may need to be retained and rendered accessible online for five years. Audit trails, reports, settlement and clearance instructions, specific investigations, corrective actions and contractual information and counterparty data affiliated with transactions are required for retention and regulatory access for seven years.

EVOLVING REPORTING REQUIREMENTS

We recommend a design approach for MiFID II

The ability to handle much more information much more quickly, and to improvise as urgent, out-of-the-blue requests and unanticipated changes to supervisory frameworks are made, is a hallmark of the technological solutions being developed."

Wolters Kluwer Financial, 'Intelligence: A Tech Revolution for the Evolution in Compliance'

---

reporting solutions that ensures agility and flexibility. The solution should deliver a regulatory reporting platform that incorporates best practices and operational effectiveness and allow for adaptive growth in scope and scale. The design goal should not be to remedy one-off reporting requests but to build in a capability to respond to emerging requirements with relative ease and cost efficiency.

SOLUTION ARCHITECTURE FOR REGULATORY REPORTING

MarkLogic has a long and successful history of solving complex data challenges for financial services companies globally.

Our solution architecture for regulatory reporting is based on years of experience working with top investment banks that have deployed the operational and transactional MarkLogic® Enterprise NoSQL

database platform to integrate, store, manage, and search their mission-critical data.

DATA INGESTION AND OPTIMISATION

Information is ingested in its native form from multiple sources. On ingest the information is validated. When information is invalid or damaged it is still ingested, but exceptions are raised for later reconciliation in-situ. This allows for a holistic data management and ensures that no data is left behind.

Optimisations can take place here; we may decide to add metadata (materialise core attributes/complex attributes). These may be extremely common attributes, or attributes that are expensive to calculate at run-time. This metadata is also stored as RDF – Resource Description Framework, or *triples* – that express relationships between 'facts'. These facts can be wired into a map – an enterprise data model (often referred to as an *ontology*).
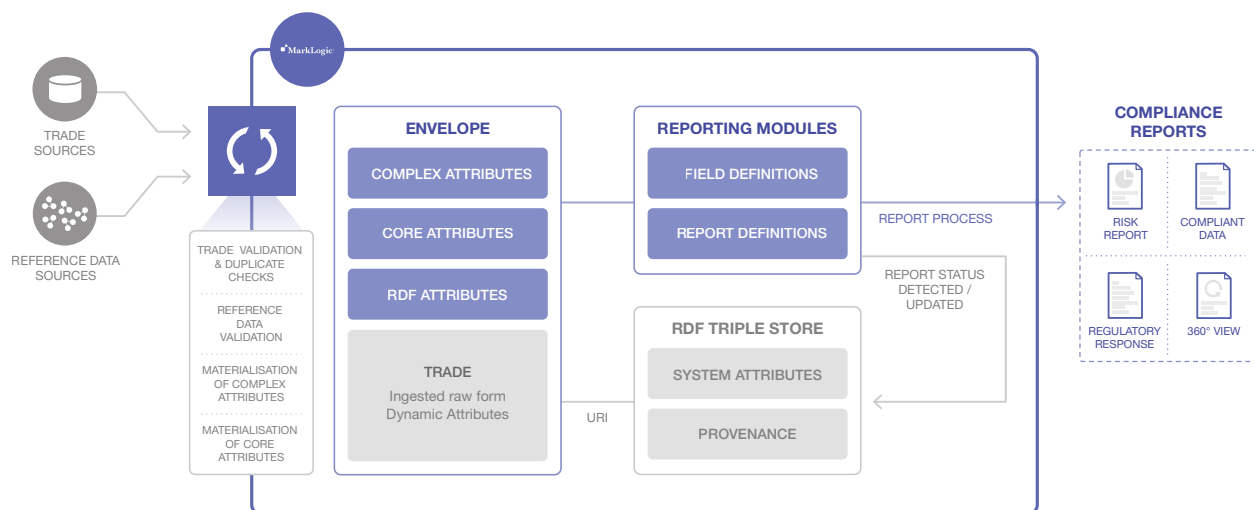
Using these attributes for enterprise class reports, derived from multiple source systems in real-time, provides full lineage and provenance. In addition, the bitemporal management of these trades and associated metadata allows for the consistent view of the trade landscape at any point in time regardless of evolving data shape and sources, assisting in the governance process.

DATA STORAGE

The Repository itself is now populated with some materialised data, some metadata (RDF) as well as the original data. Consider this a three-level system where I can ask questions of original data, optimise access to materialised attributes, or use the RDF meta-data to deliver enterprise-wide reporting and e-discovery.

REPORTING

A key component is the reusable reporting modules. These libraries can be re-used to define fields in new reports for new consumers. This offers flexibility to cater for evolving regulatory requirements.

MONITORING

The flow of information is tracked through the system within the transaction meta-data itself: when was it ingested, reported, re-reported, are all tracked with a simplistic ease.

KEY FEATURES

Three key features of the MarkLogic Enterprise NoSQL database platform are also worth emphasizing:

1. MarkLogic Semantics provides a new approach to modeling data that focuses on relationships and context. This is particularly relevant for the correlation of data, such as legal documentation and trade data. It also extends MarkLogic's built-in search capability, providing the ability to expand searches to include related terms or to show users the connection between related entities.

2. The MarkLogic Bitemporal feature allows the querying of data across both system time and event time axes. This is of particular importance for regulatory requirements, to avoid the increasingly harsh downside consequences from not adhering to government and industry regulations, particularly in financial services and insurance, and for audits— to preserve the history of all data, including the changes made to it, so that clear audits can be conducted without having to worry about lost data, data integrity, or cumbersome extract, transform, and load (ETL) processes with archived data.

3. MarkLogic Tiered Storage lets you manage your data at different tiers of storage that's the most appropriate for the business, whilst still under the control of MarkLogic – allowing a cost-effective approach to the increasing regulatory demands of data storage volume for compliance purposes.

NEXT STEPS

Contact MarkLogic to help you review your data readiness for MiFID II and expedite your regulatory compliance processes.

- **Visit our website**
  www.marklogic.com

- **Watch our recent MiFID II webinar**
  http://www.marklogic.com/resources/mifid-ii-its-about-data-not-reporting/

- **Attend a Waters Technology webinar**
  "Managing Risk and Compliance, Through the Flow of Data."
  http://www.waterstechnology.com/reference-data-data-management/2473186/managing-risk-and-compliance-through-the-flow-of-data

MarkLogic®