tcgdigital | mcube by tcgdigital

Knowledge Graphs used to Transform & get Siloed Data AI Ready

Knowledge Graphs, Semantic Integration, and AI-Driven Master Data

CONTENT

Industry Challenge

Life sciences organizations generate vast volumes of regulatory, clinical, and operational data. Yet, this data is often fragmented across multiple systems—LIMS, CTMS, EDC, ERP, SOP repositories, and document stores. Variability in metadata, inconsistent identifiers, and siloed documentation creates operational bottlenecks:

- Regulatory teams manually extract product attributes, indications, and references from SmPCs, protocols, and validation reports, increasing risk of errors and delays in submissions.
- Clinical and R&D teams struggle to reconcile assay results, patient cohorts, and biomarker data across trials, slowing evidence evaluation and decision-making.
- Manual reconciliation and lack of semantic integration prevent efficient reuse of historical knowledge and limit AI-driven insights.

These challenges result in slower regulatory submissions, higher compliance risk, operational inefficiencies, and limited ability to leverage AI for actionable insights.

Solution

mcube™ addresses these challenges by unifying fragmented life sciences data into a connected, AI-ready semantic fabric. The platform combines semantic knowledge graphs, ontology integration, GenAI, NLP, and centralized data architecture to provide a single source of truth:

1 Technology Foundation (Bottom Layer)

- **Semantic Knowledge Graph:** Captures entities, relationships, and context across structured and unstructured datasets.
- **GenAI & NLP:** Extracts actionable insights from regulatory documents, protocols, and clinical content.
- **Ontology Integration / Creation:** Standardizes concepts for FAIR metadata, ensuring consistent interpretation across datasets.
- **Data Lakehouse:** Centralizes structured and unstructured data for scalable access.
- **Conversational Interface:** Enables AI-driven exploration, automated master data generation, and traceable workflows.

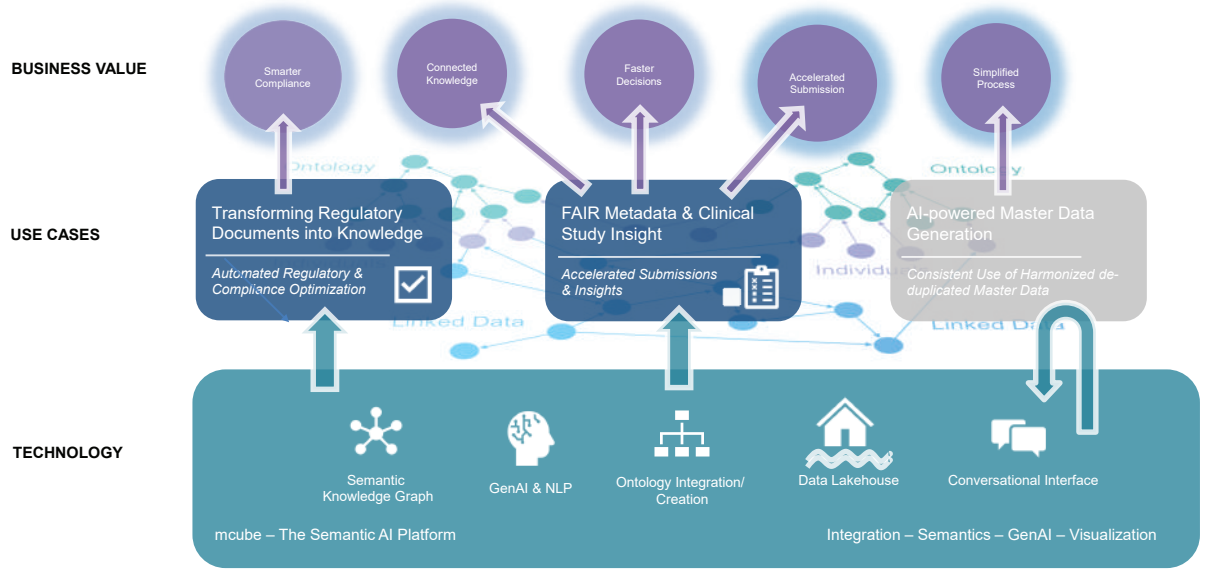
2 Use Cases (Middle Layer)

- **Automated Regulatory & Compliance Optimization:** Transform unstructured regulatory documents into structured knowledge. Harmonize terminology with IDMP, MedDRA, and SNOMED. Enable traceable, audit-ready lineage.
- **Accelerated Submissions & Clinical Insight:** Harmonize metadata across studies, connect patient cohorts, assays, and outcomes, and support FAIR principles for metadata reuse.
- **AI-Powered Master Data Generation:** Generate, deduplicate, and harmonize master data automatically. Ensure consistent terminology across internal and external systems.

3 Business Outcomes (Top Layer)

- **Smarter Compliance:** Semantic knowledge graphs make regulatory documents traceable, actionable, and aligned with global standards.
- **Connected Knowledge & Faster Decisions:** Ontology-driven integration accelerates submissions, improves cross-study insights, and reduces manual metadata reconciliation.
- **Simplified Processes:** Conversational AI reduces operational friction and ensures consistent master data generation.

One Platform – Infinite Possibilities *(Ecosystem of Semantic AI Technologies)*



Use Cases

1 Structured Regulatory Intelligence

Challenge

Regulatory teams face fragmented repositories of SmPCs, protocols, and validation reports. Extracting product attributes and ensuring traceability is manual and error-prone. Maintaining audit-ready documentation and terminology alignment is difficult.

Solution

- mcube™ ingests unstructured documents and extracts structured information using LLM & GenAI.
- Ontology-based mapping ensures alignment with regulatory standards.
- Alerts flag missing fields or inconsistencies.
- Dashboards enable cross-product, cross-region analysis with full lineage.

Outcome

- Automated extraction and harmonization of regulatory data.
- Audit-ready, traceable documentation.

Use case1: Transforming SmPCs & Regulatory Documents Into Structured Knowledge

SmPCs trapped in PDFs; manual extraction	Inconsistent terminology across products & regions	No search or filtering across large SmPC datasets	Difficult to view complete product profile	Siloed regulatory knowledge	Time-consuming compliance checks & variations
Upload doc → LLM Extraction	**Semantic Enrichment** (ontology mapping, synonyms, multilingual labels)	**Graph Linking** (product → indication → class → substance)	**Product Detail Page**	**Knowledge Graph Construction**	**Semantic Search & Comparison**
Automated extraction of product, indication, dosage, safety fields	Standardized disease, vaccine, chemical descriptors	Browsable product catalog with filters: vaccine type, indication, dosage form, year	Unified view: key fields, ontology mapping, GenAI abstract, related documents	Connected regulatory knowledge base for small molecules, antibodies, vaccine etc	Faster impact analysis across SmPC versions and products

2 Cross-Study Clinical Metadata Integration

Challenge

Clinical and R&D teams struggle to integrate data across multiple trials, assays, patient cohorts, and biomarkers. Manual reconciliation delays insight generation and decision-making.

Solution

- mcube™ consolidates clinical metadata and protocols into a unified semantic knowledge graph.
- Ontologies standardize endpoints, sample types, and assay definitions.
- Automated alerts surface missing or inconsistent data.

Outcome

- Instant access to cross-study insights.
- FAIR-compliant, consistent metadata usable across teams.

Use case 2: FAIR Clinical Metadata & Study Intelligence

Ingest→ Extract → Normalize	Knowledge Graph Linking (drug-arm-patient →biomarker sample)	Study Search (filter by condition sample type, assy	Graph-Powered Comparison	Info Access Control	Advanced Analytics (heat maps, dashboards, study reports)
Clinical data scattered across CTMS, EDC, UMS PDFs	No visibility into relationships between study components	Slow, manual study discovery	Hard to compare trials or reuse design elements	Restricted patient data access	Complex biomarker/clinical data hard to interpret
Unified clinical dataset spanning studies, arms, patients, biomarkers	Contextual insights and association discovery	Find relevant studies instantly (e.g. urine-sample trials)	Compare study arms, endpoints, biomarkers: reuse protocol elements	Protected patient-level access; others see aggregated summaries	Interactive visualizations for cohorts, biomarkers, outcomes

Conclusion

Life sciences organizations face high volumes of disconnected data, impeding compliance, operational efficiency, and AI adoption.

By transforming fragmented data into a connected intelligence fabric, mcube™ enables organizations to accelerate submissions, streamline operations, reduce manual effort, and make faster, smarter decisions.

01 A unified, FAIR semantic foundation connecting regulatory, clinical, and operational domains.

02 Consistent, reusable master data across multiple systems.

03 Automated regulatory alignment with global standards.

04 End-to-end traceability for audit and compliance.

05 AI-powered insights grounded in explainable, semantic context.

TCG Digital is the digital & AI arm of The Chatterjee Group (TCG), a multi-billion dollar conglomerate with a diverse portfolio in Aviation, Pharmaceuticals, Biotech, Petrochemicals, and Real Estate across the US, EU, and South Asia. Our umbrella includes companies such as LabVantage, Lummus Technology, and TCG LifeSciences. At TCG Digital, we are driven by our mantra of delivering "Velocity to Value", helping enterprises transform faster and smarter. Our AI Analytics platform mcube™ is at the heart of these transformations. We enable organizations unlock the full potential of their data, and by seamlessly integrating AI/ML capabilities into their business processes, we empower businesses to accelerate their digital transformation journey, enhancing agility and driving impactful results.

www.tcgdigital.com

mcube™ is TCG Digital's flagship Data, AI, and Analytics platform—engineered at the intersection of deep industry knowledge and digital innovation. Built on a domain-driven design, it seamlessly navigates the most complex and disparate data landscapes. With AI 2.0 at its core, mcube™ fuses advanced models with real-world context to solve high-impact business challenges. By integrating mcube.data, mcube.ai, and mcube.agents, it transforms enterprise data into intelligent ecosystems. mcube™ drives semantic discovery through ontologies, and empowers agentic applications for autonomous decision-making and continuous learning. From harmonizing data to deploying intelligent agents, mcube™ accelerates 'velocity to value'—delivering clarity, intelligence, and outcomes at scale.

www.tcgdigital.com/tcg-mcube

Contact Us

contact@tcgdigital.com
www.tcgdigital.com

TCG DIGITAL SOLUTIONS LLC
265 Davidson Ave, Suite 220 Somerset
New Jersey 08873