Data Protection and Trade Secrets in AI-Powered Materials Databases: An Integrated Legal Framework

Bird & Bird LLP, Madrid, Spain
Paseo de la Castellana 7 – floor 7 Madrid (Spain)

Joaquin.munoz@twobirds.com

The advancement of AI-driven materials discovery depends critically on interoperable materials databases containing structural, electronic and performance data for thousands of materials across diverse applications in electronics, energy, biomedicine and quantum computing. However, the creation, sharing and utilisation of these databases raise complex legal challenges at the intersection of data protection, trade secrets, intellectual property and collaborative research governance.

This presentation develops an integrated legal framework addressing these multifaceted challenges for materials science researchers, database developers and industrial users.

First, data protection considerations arise when materials databases contain information that could directly or indirectly identify individuals or organisations, requiring GDPR compliance in data collection, storage and sharing, particularly in collaborative research consortia spanning multiple jurisdictions. The presentation examines anonymisation techniques for proprietary materials data, lawful bases for processing in research contexts and cross-border data transfer mechanisms for international materials science collaborations.

Second, trade secret protection presents a fundamental tension in materials science: researchers must share sufficient data to enable interoperability and reproducibility whilst protecting commercially sensitive information about novel materials, synthesis processes and performance characteristics. How can organisations participate in open materials databases whilst safeguarding competitive advantages? The presentation addresses trade secret identification and classification in materials data, contractual mechanisms for confidential information in data sharing agreements, technical measures for selective disclosure and legal boundaries for reverse engineering from shared materials data.

Third, intellectual property rights in databases themselves require careful consideration: Who owns compiled materials databases? What protection exists for curated datasets under sui generis database rights, copyright and patent law? How should AI models trained on materials databases be licensed? The presentation examines ownership allocation in multi-party database development, licensing strategies for materials data and trained models, and IP due diligence in database acquisitions. Fourth, governance frameworks for interoperable databases must address technical standards, access controls, usage restrictions and dispute resolution mechanisms.

Drawing on practical experience structuring materials science collaborations, the presentation provides model contractual clauses for data sharing agreements, risk assessment frameworks for trade secret disclosure, compliance checklists for GDPR in research databases and governance structures for consortium-managed materials databases.

Case studies illustrate these principles in practice, including multi-institutional materials genome initiatives, industry-academia data sharing partnerships and commercial materials informatics platforms.

This integrated legal framework enables the materials science community to build the robust, interoperable data infrastructure essential for AI-driven discovery whilst protecting legitimate commercial interests, ensuring regulatory compliance and fostering the collaborative innovation that AI4AM2026 seeks to advance.

References

[1] Regulation (EU) 2016/679 (General Data Protection Regulation), Official Journal of the European Union (2016)
[2] Directive (EU) 2016/943 on the protection of undisclosed know-how and business information (trade secrets), Official Journal of the European Union (2016)
[3] Authors' experience structuring data governance in collaborative R&D projects, Bird & Bird LLP (2023-2025)