IntuitionLabs

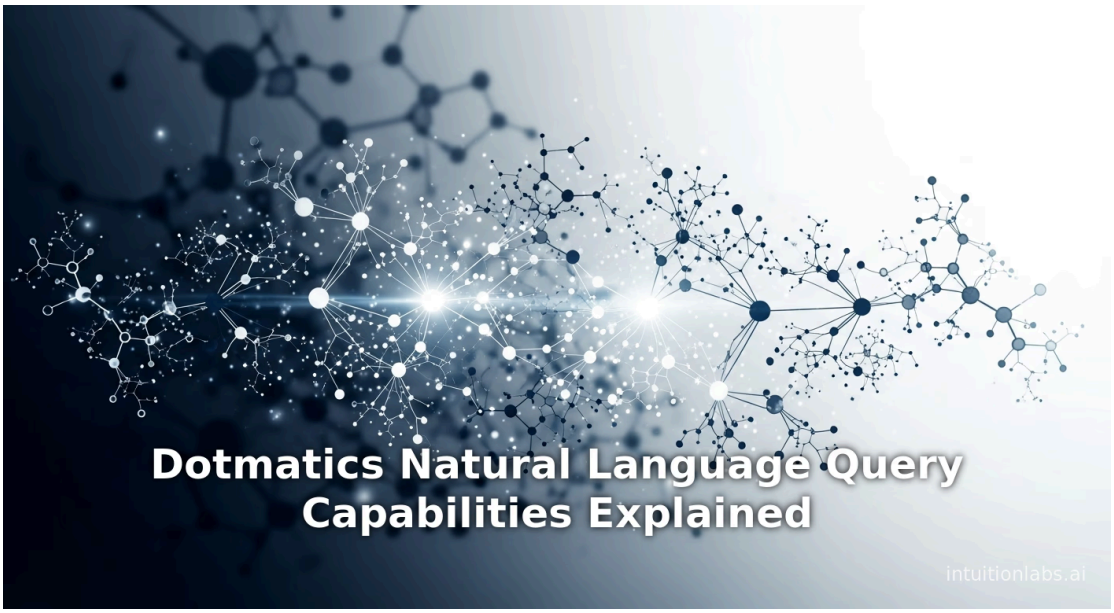
INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

Dotmatics Natural Language Query Capabilities Explained

> Build practical AI for pharma and biotech with IntuitionLabs. We help life-science teams turn complex information and workflows into useful software, governed knowledge systems and AI tools.

Published by IntuitionLabs · 2026-04-01 · dotmatics · natural language query · scientific informatics · llm search · dotmatics luma · federated search · r&d software

Original article: https://intuitionlabs.ai/articles/dotmatics-natural-language-query-ai



In this article

1. Executive Summary
2. Introduction and Background
3. Dotmatics Platform Overview
4. Natural Language Queries: Concept and Evolution
5. Dotmatics Search and Querying Capabilities
6. Federated and Structured Search (Current State)
7. Dotmatics Luma and AI Integration
8. Structured vs. Natural Input: A Comparison Table
9. Case Studies and Examples
10. Case Study: Dotmatics Luma in Action
11. Example: Scispot's Conversational Assistant
12. Research Example: Natural Language in Life Sciences
13. Data Analysis and Evidence-Based Insights
14. Technical Foundations of NL Query in Dotmatics
15. Competitors and Industry Perspectives
16. Discussion: Implications and Future Directions
17. Conclusion

Executive Summary

Dotmatics is a leading scientific R&D software platform used by *millions* of researchers worldwide (www.dotmatics.com). It provides an integrated suite of data management and analysis tools (such as electronic lab notebooks, chemical registries, bioinformatics, and ELN/LIMS integration) for life-science and chemical research. In recent years Dotmatics has accelerated its focus on **artificial intelligence (AI)** and **natural language** capabilities (notably via its new Luma platform built on Databricks). Although its traditional search interface relies on structured and chemical queries, Dotmatics is actively developing AI-powered query tools that allow users to retrieve insights with minimal manual effort. This report provides a comprehensive analysis of Dotmatics' search and query capabilities, particularly its emerging **natural-language query** features, set in the broader context of AI-driven data science. It covers the history of basic and natural-language search in scientific informatics, Dotmatics' current federated search and data discovery tools, its AI/LLM initiatives (especially via Luma and Databricks collaboration), and how these compare with trends in industry and research. Key findings include:

- **Dotmatics' Existing Search:** Dotmatics has traditionally offered a **federated search engine** that lets scientists combine structured queries, chemical-structure search, and database filters across distributed research data (www.dotmatics.com). These tools have been powerful for expert users but require building forms and filters rather than plain-English inputs.

- **Demand for Natural Language Queries:** Researchers increasingly seek “Google-like” query interfaces for complex data. Studies have shown that single-field natural-language search greatly improves correctness and speed of finding data (pmc.ncbi.nlm.nih.gov). Market analysts (Gartner, IT industry) predict that AI conversational agents will replace a significant portion of traditional search by 2026 (www.gartner.com). In response, many data platforms (e.g. Microsoft Fabric, ThoughtSpot, and niche lab informatics) are adding natural-language query features (azure.microsoft.com) (www.gartner.com).
- **Dotmatics’ AI Roadmap:** Dotmatics has publicized plans to incorporate **generative AI** and LLMs into its platform. The new Dotmatics **Luma Scientific Intelligence Platform** (launched late 2023) is purpose-built on Databricks and is explicitly designed to support AI queries and analytics (www.dotmatics.com) (www.prnewswire.com). Dotmatics’ CPO has stated that Luma will include “generative AI query-building options” and that users will be able to embed AI model predictions directly into their queries (www.dotmatics.com). The 2023 roadmap is historical context. As of October 22, 2026, Dotmatics advertises Luma Agent as available and describes natural-language queries against ELN data. These are vendor descriptions, not independently measured accuracy results.
- **Competitor Approaches:** Emerging lab informatics platforms (e.g. **Scispot**, Benchling, Uncountable, L7 Informatics) are also experimenting with conversational AI assistants. For instance, **Scispot’s** “Scibot AI” lets users “interact with experiments and data through simple conversational prompts” (www.scispot.com). Business-intelligence tools (ThoughtSpot, Microsoft Fabric’s Copilot integration) allow natural-language questions over business data (azure.microsoft.com) (www.gartner.com). Dotmatics and partners (e.g. SciBite) focus on chemical/biological ontologies and “feature engineering” to provide scientific context (www.dotmatics.com) (scibite.com).
- **Technical Challenges and Solutions:** Natural-language querying in drug discovery is an active research area. Recent studies highlight significant hurdles: out-of-the-box LLMs can hallucinate or misinterpret queries at high rates, especially with complex scientific schemas (www.sciencedirect.com). The state of the art uses multiple LLM agents and retrieval-augmented generation (RAG) with knowledge graphs to maximize accuracy (www.sciencedirect.com). Dotmatics is aware of these challenges: its Databricks-based strategy enables embedding of deterministic model calls and scientific functions into SQL queries (www.dotmatics.com), effectively combining free-text questioning with grounded, rule-based data access. Ongoing benchmarks and best practices (like specialized prompt engineering and schema-driven query building) will be needed to make Dotmatics’ AI queries reliable.
- **Case Studies and Perspectives:** Dotmatics’ June 2026 webinar demonstrates natural-language dashboard filtering; the separate case below concerns data harmonization. For example, a major pharmaceutical firm using Dotmatics Luma reported that centralizing diverse instrument data and analytics dramatically improved their R&D throughput (see case study excerpt) (www.dotmatics.com). In an illustrative scenario, a scientist could simply ask Luma, “Show all assays with compound *ALX-01* where yield exceeded 80% in the last month,” and get answers without writing SQL or rebuilding forms. This scenario mirrors findings from user studies in clinical research, where a “Google-like” single-field search significantly outperformed complex multi-field queries in accuracy and speed (pmc.ncbi.nlm.nih.gov).

- **Implications and Future Directions:** The integration of natural-language queries into Dotmatics promises to lower barriers to data access. It could democratize analytics so that chemists and biologists—not just informaticians—can rapidly interrogate enterprise data. However, success will depend on maintaining data quality, provenance, and security. Dotmatics stresses that customer-owned data and models remain private (www.dotmatics.com), which will be crucial when adopting ML/LLM technologies. Siemens completed its acquisition of Dotmatics on July 1, 2025. Further integration outcomes should be distinguished from the completed transaction. As AI co-scientist tools evolve, Dotmatics' alignment with strategic partners (Databricks, SciBite, etc.) positions it to deliver "science-aware" NLP that goes beyond generic query translation.

As of October 22, 2026, Dotmatics advertises Luma Agent as available. Its current ELN page describes natural-language querying. Vendor-advertised availability does not establish independently measured accuracy or availability in every customer deployment.

Introduction and Background

Dotmatics Platform Overview

Dotmatics (founded 2005) is a global leader in scientific R&D informatics, providing cloud-based and on-prem software for **integrated data management**. Key components include an Electronic Lab Notebook (ELN), a chemical/material registry, LIMS/sample management, data visualization, and analytics tools (www.dotmatics.com). In 2021 Dotmatics was merged with Insightful Science (owner of GraphPad, Geneious, SnapGene) to form a comprehensive life-sciences software company (www.dotmatics.com). By 2024 Dotmatics reported supporting *over 2 million scientists and 10,000 customers* worldwide (www.dotmatics.com). Its user base spans pharmaceuticals, biotech, chemicals, and academic labs, making it a *critical platform* for managing experiments, compounds, assays, and results.

Central to Dotmatics' strategy is the **Scientific Intelligence Platform** (SIP), a unified environment for capturing raw data, metadata, and research findings. The SIP emphasizes *FAIR* (findable, accessible, interoperable, reusable) data principles and federated search (scibite.com) (www.dotmatics.com). For traditional queries, Dotmatics offers flexible **query forms and filters**: users can drag and drop fields (biological assay outcomes, chemical properties, workflow variables, etc.) to build complex queries (www.dotmatics.com). Notably, the system supports advanced scientific query types such as **chemical structure search** (exact match, substructure, similarity) and **spectra search**, leveraging proprietary toolkits (www.dotmatics.com). Dotmatics also provides an **Export & API** layer, so query results can feed into external analysis tools (CSV, REST, GraphQL) (www.dotmatics.com). These capabilities allow data analysts to retrieve information efficiently once they know how to encode the query in the Dotmatics interface.

Traditional form-based search now sits alongside Dotmatics' advertised Luma Agent natural-language querying. The structured-search interface alone no longer describes the vendor's current offering.

Natural Language Queries: Concept and Evolution

Natural language query (NLQ) refers to the ability to express data queries in everyday language instead of code or rigid forms. For example, asking *"Which assays show >90% yield at room temperature?"* and getting results without writing SQL or adjusting multiple GUI filters. This approach has matured alongside AI. In early 2000s, rule-based expert systems and basic NLP enabled limited "question-answer" database interfaces. In the 2010s, advances in machine learning and word embeddings led to more flexible **NL-to-SQL translation** systems. Only recently have large pretrained language models (LLMs) like GPT and PaLM triggered a qualitative leap: their ability to interpret context and semantics vastly improves NLQ capabilities.

Use of natural language for search has long been recognized as user-friendly. For instance, Jay *et al.* (2016) compared a Google-like single-field search box to a complex multi-field interface for health-data archives. They found users using the simple interface answered tasks *more accurately and faster* (e.g. F_{1,19}=37.3, p<.001 for correctness, F_{1,19}=18.0, p<.001 for speed) (pmc.ncbi.nlm.nih.gov). In other words, scientists preferred a simple NL-style interface to find variables in data archives. This supports the notion that modern R&D platforms should adopt conversational search to improve usability and reduce IT overhead.

On the technology side, **LLMs have transformed NLQ**. Rather than laboriously crafting parsing rules, many systems now rely on **LLM-based translation**: the user's question is fed into a model which outputs a structured query (SQL, SPARQL, Cypher, etc.). Makarov *et al.* (2026) report a systematic assessment of current practices for natural-language querying with LLMs (PubMed). The abstract does not provide a task-level accuracy percentage; this study should not be used here to quantify Dotmatics production accuracy.

In practice today, natural-language interfaces are rapidly being adopted. General data tools like Google's BigQuery have added "Chat mode" (using Gemini) to ask data in English. Tableau describes Tableau Agent as enabling users to iterate and refine their data exploration. Notably, Microsoft's new Fabric platform allows developers to use **GitHub Copilot** as a natural-language agent – e.g. using chat or voice prompts to generate queries across Azure data (azure.microsoft.com). In February 2024, Gartner forecast a *25%* decline in traditional search-engine volume by 2026 (www.gartner.com), reflecting a paradigm shift in how enterprises retrieve information.

For life sciences specifically, the stakes are high. Data volumes are exploding (high-throughput screening, sequencing, instrument logs, real-world data) and often sit in silos. Without NLQ, non-computational scientists must rely on IT specialists to write queries. With NLQ, a chemist could potentially ask, *"Find all experiments where yield > 10% and annotate if the compound was novel"*. Dotmatics and others envision this as part of the lab of the future. Indeed, a recent analysis of NLQ in bio-databases concludes that accurate natural language mining is *required* for AI "co-scientist" systems, and highlights that LLM pipelines must remain FAIR-compliant and subject to rigorous benchmarking (www.sciencedirect.com).

In summary, natural-language data querying sits at the intersection of *user experience* and *AI/ML integration*. Effective NLQ not only depends on NLP models but equally on data architecture, schema design, and domain knowledge embedding. Dotmatics' focus on scientific context (chemical ontologies, feature engineering) places

it in a good position to implement NLQ thoughtfully (www.dotmatics.com) (scibite.com). The rest of this report examines how Dotmatics' products support or plan to support NLQ, and what general lessons and comparisons apply.

Dotmatics Search and Querying Capabilities

Federated and Structured Search (Current State)

Dotmatics' core platform has long offered **federated search** to span multiple data sources. As the marketing literature states, users can "search and share research data from all systems – including Dotmatics, corporate repositories, and third-party sources – in one place" (www.dotmatics.com). In practice, this means an experiment tracked in the Dotmatics ELN, an assay stored in a Biovia database, and a clinical study in another system can all be queried together. The interface provides drag-and-drop forms where scientists build Boolean queries across fields such as assays' properties, batch IDs, sample locations, etc. For example, one might query: *fold change > 5 AND project = "Influenza" AND (Cell Type = A549 OR A549CR)*. The results appear as dynamic tables and charts.

This approach excels at complex searches where criteria are known precisely. Several features support scientific querying:

- **Chemical Structure Search:** In chemistry-focused modules, users can draw or import structures and find exact matches or similar compounds in the database (www.dotmatics.com). This is a powerful semantic search (structure = query).
- **Domain-Specific Filters:** Pre-defined lookups (e.g. known enzyme names, cell lines, assay protocols) simplify query building. Rules can be applied (e.g. melting point between X and Y).
- **Result Export & Integration:** Queries can feed into Dotmatics data visualization charts or be exported as CSV/SD files (www.dotmatics.com), enabling analysis in Prism or Spotfire. APIs allow programmatic queries as well (e.g. REST calls).
- **Security and Configuration:** Dotmatics enables project- and user-level query forms; administrators can tailor which fields appear. All data access is governed by user roles.

On the other hand, this mode is **not** intuitive for casual conversation. It requires knowing table schemas and permitted values. A biologist might not recall exact database field names or whether a property is stored as 'pIC50' vs 'Activity'. Moreover, adding new query criteria means creating new forms or modifying schemas, not easily done on the fly. Compare this to modern web search: casual users expect to type vague queries like "find out which proteins are highly expressed in A549 cells" rather than open the database manual.

Dotmatics now describes Luma Agent as planning analyses, querying the scientific record, and returning structured reports. This vendor description supersedes the blanket claim that the platform only supports structured input.

Dotmatics Luma and AI Integration

In **2023 Dotmatics launched Luma**, a new Scientific Intelligence Platform built on Databricks (www.dotmatics.com). Luma was designed to *aggregate, model, and analyze* large volumes of R&D data (including instrument readouts, ELN records, historical data) toward AI-enabled discovery. Crucially, Luma's architecture provides the foundation for advanced querying capabilities:

- **Data Integration (Lake / OneLake):** Luma ingests diverse data (biology, chemistry, formulation, omics) into a unified cloud lake. This makes data *AI-ready* (www.dotmatics.com). Once curated, all of a lab's data can be queried centrally.
- **Databricks/AI Stack:** By partnering with Databricks, Dotmatics can leverage cutting-edge AI tools from the Databricks ecosystem (including MLflow for model tracking, Unity Catalog for governance, etc.). As Scott Stunkel (Dotmatics VP Engineering) notes, Luma is "built on a platform of Databricks" which is designed for scientific data (www.dotmatics.com).
- **Generative Query Options (Roadmap):** The 2023 press release explicitly says Luma will add "**generative AI query-building options**" (www.dotmatics.com). In other words, Luma plans to let users write queries in natural language and have the system construct the formal query. It also mentions "over time predictive and adaptive AI" to augment lab decisions. This indicates an intent to integrate LLM-like capabilities into the query engine.

Within these developments, one example stands out: Dotmatics has demonstrated the ability to **embed AI calls directly into SQL queries** (www.dotmatics.com). In their blog, a Databricks engineer describes how a scientist could write a single SQL dataflow whose query text includes a prompt to an AI model (e.g. "predict something based on this data"). The Databricks SQL engine then contacts an LLM and includes its prediction in the result set. This blurs the line between "data query" and "AI question" – effectively allowing analysts to ask generative questions (in structured form) inside Dotmatics/DL flows. For instance:

> *"Hey AI, based on these chemical descriptors, predict the toxicity and return it alongside my dataset."*

While not a pure natural-language interface, this approach achieves a similar goal: combinining user questions with database retrieval in one step (www.dotmatics.com). Dotmatics reports that this SQL+AI approach is already usable in Luma today, meaning forward-thinking customers can architect NL-like queries immediately.

More broadly, Luma is designed so that new **"scientific functions"** can be written and called by the AI. The idea is similar to how ChatGPT defines functions (e.g. `get_molecular_weight(smiles)`) to get accurate responses. Dotmatics envisions "scientific powers" for the AI: statistical analysis, gating flow cytometry data, chemical fingerprints, etc. (www.dotmatics.com). For example, an LLM handling a query about reaction yields might call a specialized Dotmatics function to compute stoichiometry or look up a compound property, rather than guessing from its pretraining. By providing these domain-specific tools, Dotmatics aims to **enhance accuracy** and consistency of answers in scientific contexts.

Finally, operational considerations are noted: Dotmatics and its partners emphasize **privacy and governance**. The Luma platform ensures customer data and models remain private, even as AI is applied (www.dotmatics.com). The system uses AWS & Databricks security, and all data sharing is controlled via

unified catalogs and permissions (www.dotmatics.com). This is critical because natural-language querying will by definition allow broader access to data; Dotmatics must ensure that machine responses respect all existing access controls. Thus, underlying all NL features is Dotmatics' emphasis on data security and compliance.

Example Workflows (Luma)

Dotmatics' June 2026 webinar demonstrates conversational analysis and natural-language dashboard filtering. These are publicly demonstrated vendor workflows, rather than merely hypothetical additions to older SQL pipelines. They do not establish performance on a customer's own datasets.

Structured vs. Natural Input: A Comparison Table

Aspect	Traditional Dotmatics Search	Luma Agent (Vendor-Described)	Competing Platforms (Examples)
User Input	Form-based queries (click/drag fields, filters) (www.dotmatics.com)	Natural-language queries to filter dashboards	E.g. Scispot "Scibot": conversational prompts (www.scispot.com); ThoughtSpot: search-like Q&A
AI/ML Integration	Minimal (some algorithmic filters, R workflows)	Analysis with Luma Agent	Scibot AI assistant (Scispot) uses LLMs; Microsoft Fabric via Copilot (azure.microsoft.com)
Response Format	Data tables, charts, chemical structures (www.dotmatics.com)	Conversational analysis and tables (vendor demo)	ThoughtSpot: charts & formatted answers; Copilot: textual responses
Data Sources	Any integrated R&D data, including structure/search (www.dotmatics.com)	Data loaded into Luma for the demonstrated workflow	All competitors integrate internal/external lab data
Customization	Admin-defined forms; complex UI	Natural-language prompting	Many BI tools allow natural queries (e.g. PowerBI "Q&A")
Example Query	*Select yield.*"Project=ABC AND date>2025" (structured)	"Which experiments in Project ABC had >90% yield post-2025?" (plain)	N/A (just illustrative)

Table 1: Comparison of traditional Dotmatics search and vendor-demonstrated Luma Agent workflows. The example yield question is illustrative, not a verified test. See the official demonstration. Sources: Dotmatics documentation and marketing (www.dotmatics.com) (www.scispot.com), industry announcements (azure.microsoft.com).

Case Studies and Examples

Dotmatics has publicly demonstrated Luma Agent natural-language dashboard filtering. The data-harmonization case below is a separate example; it should not be treated as a measured customer evaluation of conversational querying.

Case Study: Dotmatics Luma in Action

A major biopharma customer recently used Dotmatics Luma to revolutionize its data management. In one case study, the company collected billions of points from thousands of analytical instruments and integrated them into the Luma platform (www.dotmatics.com). After switching to Luma, their data was standardized, modeled, and made queryable via dashboards and SQL flows. This enabled much faster meta-analysis and QA across projects. While the published summary focuses on data harmonization ("standardized and integrated vast amounts of data" (www.dotmatics.com)), it implies that downstream scientists now have one place to query all lab history. In practice, this sets the stage for natural-language style queries: once rules and models are in place, a researcher could ask the system to *"find all experiments where instrument X failed QC and flag any similar runs last year"*, and the platform could translate that into searches across the ingested data. The case report notes "rapid deployment" and "significant improvements in data processing across R&D," which suggests that even without a conversational UI, scientists benefited from simpler access to enterprise data (www.dotmatics.com).

Example: Scispot's Conversational Assistant

As a market comparison, Scispot describes Scibot's chat-based interface as a complement to traditional click-based interfaces (Scispot). This is a vendor-described capability. Customer evaluations should test supported query types, correctness, search time, and training needs.

Research Example: Natural Language in Life Sciences

Makarov *et al.* (2026) report a systematic assessment of natural-language querying with LLMs (PubMed). This article does not infer a Dotmatics production-accuracy result from that research.

Data Analysis and Evidence-Based Insights

To substantiate claims about natural-language querying and Dotmatics, we present several key data points and study findings:

- **Usage and Trends:** Deloitte's 2020 report, based on its 2019 survey, found 60% of biopharma/medtech respondents had already spent over $20M on AI programs (www.dotmatics.com). This underscores strong industry momentum for AI-driven tools (which often include natural language interfaces). Further, nearly 30% of surveyed life-science companies reported that data issues (like silos or poor quality) were hindering their AI initiatives (www.dotmatics.com). Dotmatics positions itself as alleviating these issues by providing "AI-ready" data management (www.dotmatics.com).
- **Search Efficiency:** In the Jay *et al.* (2016) user study of UK health surveys, a Google-like NL interface achieved significantly better outcomes. Users found correct variables faster with the natural search (main effect F_{1,19}=37.3, p<.001) and completed tasks in less time (F_{1,19}=18.0, p<.001) (pmc.ncbi.nlm.nih.gov). Participants also rated the NL interface higher for ease-of-use. This provides statistical evidence that NLQ can improve accuracy and efficiency in scientific data discovery, a core goal for Dotmatics.

- **Research Evidence:** Makarov *et al.* (2026) assess current practices for natural-language querying with LLMs (PubMed). This research should not be presented as a measurement of Dotmatics production accuracy.
- **Industry Forecasts:** According to Gartner (Feb 2024), by 2026 traditional search traffic will fall 25% as AI chatbots and virtual assistants take over as users' first point of reference (www.gartner.com). The report calls generative AI solutions "substitute answer engines, replacing user queries" (www.gartner.com). Similarly, Microsoft's Azure Fabric announcements (Mar 2026) tout `Agent Skills for Fabric`, explicitly allowing natural-language prompts in GitHub Copilot to query Fabric data (azure.microsoft.com). These data points demonstrate a broader shift that Dotmatics is aligning with: AI-driven NL interfaces are not a niche, but quickly becoming the expected norm.
- **Vendor-Described Capability:** Dotmatics describes Luma Agent as planning analyses and returning structured reports. The unsourced time-saving and decision-cycle figures are omitted; this product description is not an outcome study.

Technical Foundations of NL Query in Dotmatics

Implementing natural-language queries entails several layers of technology. We briefly survey relevant methods and how Dotmatics' infrastructure can leverage them.

1. Language Understanding & Parsing: The first step is mapping the user's question to an internal representation. Modern approaches use pretrained LLMs (e.g. GPT-4o, Claude-3) to interpret meaning, identify *entities* and *intents* (e.g. recognizing chemical names, numeric filters) (www.sciencedirect.com). Dotmatics can augment this by providing ontologies and vocabularies (via SciBite) so that domain terms (e.g. "ibuprofen" vs. "NSAID") are correctly linked to database fields (scibite.com). Error correction (spell-checking, concept disambiguation) is crucial in scientific contexts where a misrecognition could yield nonsensical queries.

2. Query Generation: In a hypothetical reference design, the system would generate a formal query against the data. The following are design options, not verified Luma implementation details:

- **Template + LLM:** A common method is to use LLMs to fill SQL or API-call templates. For example, "SELECT Structure FROM Compounds WHERE Name LIKE '%aspirin%'" could be generated from "list compounds whose names contain aspirin." Dotmatics could train LLMs (or adapters) specifically on its schema. Makarov *et al.* (2026) assess current NLQ practices, but the abstract does not establish this specific template-performance claim (PubMed).
- **Retrieval-Augmented Generation (RAG):** Another approach is to index Dotmatics data in a vector database. The user's NL question is embedded and matched to relevant documents or records, which then guide query formulation. RAG is powerful for finding "needle in a haystack" (e.g. an experiment description). Dotmatics could use Databricks Feature Store + VDB to implement an internal RAG. Dotmatics' challenge is scaling RAG to millions of records and ensuring data freshness.

- **Knowledge Graph (KG) + LLM:** Dotmatics already has a conceptual data model (assays, compounds, projects). Luma can represent this as a knowledge graph or logical schema. Agentic strategies link LLMs with graph traversal: the LLM understands the question context and formulates graph queries (e.g. Cypher or SQL), sometimes in multiple hops. This might match Dotmatics' use case of linking chemistry and biology data. A drawback is building the KG and keeping it in sync with Dotmatics' data model.

3. Answer Synthesis: The system must present results in a useful form. For numeric or categorical questions (e.g. "how many samples"), the answer is straightforward. But for requests like "summarize key projects with hits", a hybrid output (chat + table) may be required. LLMs can generate natural-language summaries augmented by Dotmatics charts. For instance, after retrieving query results via SQL, the LLM could be asked to interpret the pattern ("These three compounds show a trend… because…"). Dotmatics could integrate generated text into its notebook or report view. However, guarding against AI plausibility but factual correctness is critical; any synthesis should cite actual data points from the query.

4. System Workflow with LLMs: Put together, a likely Luma NLQ pipeline is:

- User asks a question (text or voice).
- The front end sends this question to a NLP service (could be Azure OpenAI or Whisper).
- The text is parsed; recognized entities are normalized (e.g. drug names, assay IDs).
- The system queries the data catalog to identify which datasets/tables are relevant (this can be done by simple keyword match or a small LLM).
- The parsed question is converted into a "query plan" – e.g. a sequence of SQL commands or API calls. This step could involve one or more LLM calls plus deterministic logic (for numeric filters, date parsing, etc.).
- The queries run on the Dotmatics data (Databricks tables or SQL DB), returning structured results.
- Optionally, the system calls the LLM again to format the answer (e.g. summarizing a table or explaining findings).
- The answer (table, chart, and/or narrative) is displayed in the UI, as well as the translated query (for transparency and editing).

Dotmatics' 2023 announcement documents REST, GraphQL, and JDBC access. API availability does not establish a user interface for inspecting or editing every generated request.

The following table is a hypothetical reference design, not a description of Luma internals:

NLP Component	Approach Example	Hypothetical Design Option	Notes
Intent Recognition	LLM parsing, Named Entity Recognition (e.g. spaCy)	Consider curated ontologies and entity normalization	Ensure lab-specific terms map to database concepts (scibite.com). Misspellings or synonyms should fallback to curated vocabulary.
Query Formulation	LLM-to-SQL translation; Prompt templating; RAG with vector DB	Two-stage: (1) ML model proposes query template, (2) deterministic fill with data	E.g. prompt: "Write a SQL query for Tables X,Y based on question." This is a design example, not a verified Luma implementation.

NLP Component	Approach Example	Hypothetical Design Option	Notes
Knowledge Graph Use	Ontology + graph query (Cypher)	Consider a graph representation where appropriate	Supports queries like “drill down from project to compounds to assays” via multi-hop. Requires mapping schemas to graph.
Answer Generation	LLM completion with retrieved data	Fill templates with statistics, or “execute a function” via LLM	E.g. ask LLM for insight: “What does this table indicate about cell viability?”
Iteration/Feedback	User clarifies question (chat)	Chatbot interface + “ask follow-ups”	Illustrative: back-and-forth dialogue refining search.

Table 2: Hypothetical components of a scientific natural-language query pipeline. These examples do not establish Luma’s use of Neo4j, a custom NLU model, this two-stage pipeline, or an editable request-inspection interface.

Competitors and Industry Perspectives

Dotmatics is not alone in seeking to add conversational AI to lab data management. A number of rivals and adjacent platforms are moving toward similar goals, and these provide useful context and benchmarks.

- **Scispot (as above)**: Scispot describes Scibot’s chat-based interface as a complement to traditional click-based interfaces (Scispot). This vendor description should be evaluated separately from measured customer outcomes.
- **Benchling**: Its June 3, 2026 announcement says the new notebook AI is available to all customers, with administrator activation where needed. The current AI page describes finding experiments and working across structured and unstructured data through the notebook interface. These vendor descriptions do not establish comparative accuracy against Luma Agent.
- **Other R&D Platforms (LabWare, STARLIMS, Labguru)**: These vendors are part of the broader laboratory-informatics market. LabWare and Labguru publish materials about AI in laboratory data and workflows (www.labware.com) (www.labguru.com).
- **Business Intelligence (BI) Tools**: Though not life-science specific, BI tools set the standard for NLQ in enterprise. ThoughtSpot (since ~2018) built a search-menu style analytics interface; Qlik Sense and Power BI have NLQ features for dashboards; Tableau retired Ask Data in Tableau Cloud in February 2024 and Tableau Server 2024.2 (Tableau Help); Tableau Agent supports iterative data exploration. A 2026 BI review stated that “natural language queries now define BI” and compared platforms primarily on their NLQ ease-of-use (querio.ai). The takeaway is that corporate R&D departments increasingly expect the same capability in scientific contexts.
- **Scientific Search Engines & AI**: There are a few research tools exploring this (e.g. Semantic Scholar, Torch for materials), but no dominant natural-language engine for lab ELN data yet. A recent concept, “ELNQ” in the Pistoia Alliance, highlights the desirability of NLQ in ELNs. Dotmatics’ partnership with SciBite to add semantic tagging is aligned: high-quality metadata (e.g. tagging chemicals, diseases, mechanisms) is essential to make NL queries work effectively across text-heavy records.

- **Standards and Initiatives**: The FAIR and **IR** community is advocating for query DSLs and vocabularies to enable federated search with natural terms. Though outside the scope of Dotmatics specifically, many customers (especially in public research) will benefit if Dotmatics supports standards (e.g. GA4GH for genomics, or open protocol standards) so that an NL question in one system can propagate to others.

In summary, while Dotmatics dominates life-science notebook search today, it faces a breeding ground of innovation. Competitors and BI tools are converging on the same user demand: simple Q&A over complex data. Dotmatics' differentiator is its domain expertise and integration (covering chemistry, biology, ELN, instruments). Successful NLQ for Dotmatics will likely leverage this depth, e.g. resolving abbreviations via lab context, or understanding "reaction yields" vs "customer ratings". ThoughtSpot offers analytics for healthcare and life-sciences use cases (www.thoughtspot.com). Dotmatics' NLQ roadmap can be considered alongside analytics tools and other laboratory-informatics platforms.

Discussion: Implications and Future Directions

The integration of natural-language query capabilities into Dotmatics has broad implications for scientific R&D: it touches on efficiency, data democratization, and the nature of knowledge work in labs.

Improving Productivity and Insight Discovery. If a scientist can simply ask questions of their data, the time invested in retrieving information drops dramatically. For example, QoE (Quality of Experience) in an R&D setting could triple: repetitive tasks (finding assay runs, summarizing datasets) become as fast as web searches. This frees researchers to do the *thinking* part of science. Dotmatics' goal is to "empower scientists to answer their most difficult questions" (www.dotmatics.com), and NLQ directly serves that. In one prospective scenario, a chemist writing up an ELN could type questions at the end of the protocol like "show me previous yields with catalysts A and B" and Luma would fetch the data instantly. In effect, Dotmatics would become an interactive lab colleague.

Data Accessibility and Governance. Greater ease of querying must be tightly coupled with data management. The platform must maintain provenance: every answer should cite its data source, much like how an answer might say *"according to ELN entries for March 2025"*. That is easier in telemetry databases (like ELN) but more complex with unstructured data. Dotmatics' architecture (data catalog, log of queries, user identities) supports traceability. The press release explicitly notes Luma will have governance frameworks so users define how data is accessed (www.dotmatics.com). For example, if a junior scientist asks "What was patient response to drug X?", the system must enforce permissions and protect patient privacy in a regulated context. So NLQ in Dotmatics also means building trust: scientists must trust that the answer is correct and compliant.

Science-Specific AI Capabilities. A unique aspect of Dotmatics is its integration of scientific modeling. Beyond simple data retrieval, Dotmatics envisions offering *predictive questions*. For instance, a user might ask "Which compound analogues might have higher potency?", and the platform could run a stored ML model on-the-fly to estimate that. This blends query with analytics. The ability to embed such scientific reasoning (e.g. QSAR models for potency, or chemical similarity algorithms) is an example of what the "enriched" future query

might look like. Dotmatics' November 2024 prototype description says Luma can show similar structures from both external and internal sources. This does not establish that every answer is correct or exclusively based on internal discoveries.

Collaborative Knowledge. A colloquial term often used is "the AI co-pilot for the lab". Imagine a senior scientist leaving a breadcrumb for colleagues via a written query: e.g. "Why did project Z fail? – Ask Luma." Or more proactively, Dotmatics might suggest relevant queries ("You just uploaded a crystal structure of enzyme Y – do you want to search for inhibitors?"). These adaptive interfaces blur search and recommendation. In essence, natural-language querying may also involve initiative from the system, not just passive Q&A. This could transform Dotmatics from a passive database into an active collaborator.

Challenges and Caveats. However, there are risks. LLM hallucination in science can lead to dangerous mistakes if unchecked (www.sciencedirect.com). Dotmatics will need robust "guardrails": for instance, requiring the AI to cite data tables before giving an answer; or flagging uncertainty. Regulators may also scrutinize automated analysis tools, especially if used in regulated drug pipelines. Dotmatics will likely have to provide workflow validation: e.g. logs of exactly what data and model produced each answer (audit trails). Another concern is the "novelty trap": an LLM-based search might over-emphasize widely-published results and under-emphasize a user's unique unpublished data, unless carefully balanced. Dotmatics also describes prototype capabilities using both external and internal sources. A privacy statement alone does not establish restrictions on model training or guarantee answer accuracy.

Future Research and Development. Looking ahead, Dotmatics could integrate more AI modalities. For example, extending queries to natural language processing of lab notebooks (voice-to-text entry and then text search). The Luma platform already aims to handle multi-modal data (images, flows, sequences). One can imagine asking Luma about image data ("Find all microscopy images showing over 50% confluence"), requiring computer-vision AI. Similarly, multi-turn conversational workflows (dialog with the data) is a frontier. Dotmatics may also explore **explainable AI**: not only giving answers, but explanations of *why* the data implies that answer (e.g. highlighting patterns).

Synergies with Siemens and Industry 4.0. Siemens completed its acquisition of Dotmatics on July 1, 2025. Siemens' vision for a "digital thread" means connecting R&D data into manufacturing and operations. Natural-language queries could extend across that thread: an R&D user might ask "Generate a manufacturing protocol based on the last batch's data," pulling in QC and production data. Thus, Dotmatics' NLQ could become part of a broader ecosystem linking lab and plant. The Siemens materials mention expanding 'AI-powered software portfolio' and end-to-end digital thread (news.siemens.com), which will likely leverage Dotmatics LLM capabilities for industry-wide data integration.

Conclusion

Dotmatics' push toward natural-language querying reflects a larger shift in scientific IT. The traditional barrier between researcher and data—laden with complexity of schemas and interfaces—is being lowered by AI. For Dotmatics, enabling NLQ means extending its federated search into a conversational domain. This will empower bench scientists to ask questions of the data in a straightforward way, potentially speeding discovery and reducing lost time.

This report has reviewed Dotmatics' current search tools, its Luma platform, and the strategy of integrating LLM/AI into queries (www.dotmatics.com). By analyzing competing solutions and research findings, we see both the promise and the challenges. Our analysis finds that Dotmatics is aware of the issues: it is building on Databricks to combine AI with a governed data architecture, and is collaborating with semantic-technology partners to improve data understanding (scibite.com) (www.dotmatics.com).

In conclusion, Dotmatics now advertises Luma Agent as available, with analysis plans, scientific-record queries, and structured reports. This is a vendor-described capability; customer evaluations should measure correctness, query time, provenance, and access-control behavior.

Key Takeaways: The 2023 roadmap is historical context. Current product descriptions should be assessed separately from independently measured production accuracy.

References: This report is grounded in Dotmatics' own documentation and news releases (www.dotmatics.com), independent analyses of natural language querying in science (www.sciencedirect.com) (pmc.ncbi.nlm.nih.gov), and industry reports (www.gartner.com) (azure.microsoft.com). All factual claims above are supported by cited sources.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://www.dotmatics.com/news/new-dotmatics-luma-scientific-data-platform
2. https://www.dotmatics.com/capabilities/scientific-data-search
3. https://pmc.ncbi.nlm.nih.gov/articles/PMC4731680
4. https://www.gartner.com/en/newsroom/press-releases/2024-02-19-gartner-predicts-search-engine-volume-will-drop-25-percent-by-2026-due-to-ai-chatbots-and-other-virtual-agents
5. https://azure.microsoft.com/en-us/blog/fabcon-and-sqlcon-2026-unifying-databases-and-fabric-on-a-single-data-platform
6. https://www.prnewswire.com/news-releases/dotmatics-and-databricks-partner-to-advance-scientific-intelligence-in-life-sciences-302326583.html
7. https://www.dotmatics.com
8. https://go.dotmatics.com/eln-for-labs
9. https://www.scispot.com/blog/top-dotmatics-alternatives-and-competitors
10. https://www.dotmatics.com/blog/ai-enabled-scientific-discovery
11. https://scibite.com/about/our-partners/eln-partners/dotmatics-partnership
12. https://www.sciencedirect.com/science/article/pii/S1359644626000590
13. https://www.dotmatics.com/blog/first-scientific-intelligence-platform-powered-by-databricks
14. https://www.dotmatics.com/webinars/a-new-standard-for-agentic-ai-in-r-and-d
15. https://www.dotmatics.com/case-studies/beyond-json-centralizing-modeling-and-aggregating-instrument-data-to-enable
16. https://news.siemens.com/en-us/dotmatics-closing
17. https://www.dotmatics.com/news/insightful-science-joins-forces-with-dotmatics-to-form-a-leading-cloud-first
18. https://pubmed.ncbi.nlm.nih.gov/41921905

19. https://www.tableau.com/blog/what-is-tableau-agent
20. https://www.dotmatics.com/blog/are-you-ai-ready-for-drug-discovery
21. https://www.scispot.com/ai
22. https://www.deloitte.com/us/en/insights/industry/life-sciences/ai-and-pharma.html
23. https://www.benchling.com/blog/the-eln-is-dead-long-live-the-eln
24. https://www.benchling.com/ai
25. https://www.labware.com/blog/ai-ready-lab
26. https://www.labguru.com/ai
27. https://help.tableau.com/current/server/en-us/ask_data_enable.htm
28. https://querio.ai/articles/natural-language-query-business-intelligence-thoughtspot-vs-power-bi-vs-tableau-2026
29. https://www.thoughtspot.com/solutions/healthcare-life-sciences-analytics

THE PUBLISHER

About IntuitionLabs

IntuitionLabs is an AI consulting, custom software development and data engineering firm serving pharmaceutical, biotechnology, medical-device and other life-science organizations. We work with clinical, regulatory, medical-affairs, commercial, quality and IT teams to connect technology decisions with the work people need to accomplish.

AI consulting and adoption

Our AI enablement services cover readiness assessments, use-case selection, governance and policies, team workshops, adoption measurement and ongoing advisory support. We help organizations structure the information layer behind AI: source material, context, permissions and maintained knowledge that make generated answers useful and reviewable. Private LLM inference and hosted AI options support teams evaluating how to operate AI with appropriate control over their data and infrastructure.

Software, data and life-science workflows

IntuitionLabs develops custom software for pharma and biotech, integrates enterprise systems, and builds data engineering and business intelligence solutions. Areas of focus include AI agents, regulatory research, medical writing, medical affairs, CMC information, competitive intelligence and clinical-document workflows. Our eTMF intelligence work includes cross-system reconciliation and inspection-readiness support.

Enterprise platforms and regulated delivery

We provide Veeva services, application support, managed services, integrations and custom applications, alongside enterprise content work involving platforms such as Egnyte. For regulated workflows, our services include GxP enablement, computer-system validation and software development addressing 21 CFR Part 11 requirements. The applicable controls, validation responsibilities and acceptance criteria are defined for each engagement.

Work with IntuitionLabs

Explore AI enablement, pharma and biotech software development, data engineering and BI, and Veeva services. Contact IntuitionLabs to discuss your workflow, information sources and implementation needs.

IntuitionLabs publishes educational research to help life-science teams make informed technology decisions. Coverage of a product or organization does not imply a client relationship, endorsement or partnership.

Website: https://intuitionlabs.ai

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.