Use case

Unlock the full potential of ELN data
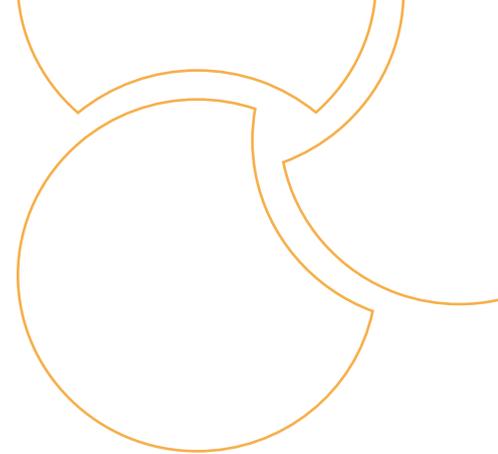
Unlock the full potential of ELN data

Most pharmaceutical companies will have, at some point, deployed an Electronic Laboratory Notebook (ELN) with the goal of centralising R&D data. ELNs have become an important source of not only key experimental data points but also the development of methods and SOPs. However, much of the information stored within an ELN is captured as qualitative free text or as a collection of attachments and, as a result, the ability to mine it is typically limited to rudimentary text and keyword searches.

Most ELNs only enable users to search for exactly what was written by the author of an experiment. The inconsistent use of synonyms during data entry makes it difficult to identify and collate all relevant data for a disease or target of interest. For example, an experiment describing work on 'muscarinic acetylcholine receptor M1' will not be found by a scientist who performs a search using the commonly used synonym 'cholinergic receptor muscarinic 1'. Even for more defined entries, the meaning of a particular field or its contents may be ambiguous, imprecise or contain multiple different data types, such as Gene, Assay Type and Species.

SciBite provides organisations with the opportunity to augment their ELN by semantically enriching the data stored within it, opening up new possibilities to mine the data more effectively and derive valuable insights.

Semantic enrichment of ELN data

At the core of the SciBite platform are the established controlled vocabularies, or VOCabs, which apply an explicit, unique meaning and description to scientific terms. This enables complex experimental text to be contextualised so that it can be understood and used as high quality, actionable data, irrespective of its source.

SciBite's ontology management platform, CENtree, enables organisations to maintain up to date ontologies representing evolving scientific language. Subject matter experts can easily contribute to keeping things current and augment our manually curated standard reference vocabularies (such as Cell, Gene and Indication) with proprietary information, such as project codes and IDs used to track materials such as compounds and cell lines, to create a single, authoritative, universally applied terminology.

When used in combination with CENtree, SciBite's Named Entity Recognition (NER) engine, TERMite, applies standard, well established ontologies and controlled vocabularies to ELN experiments, generating a semantic index and endowing text entries with an explicit, specific meaning. This transforms unstructured experimental text, including embedded supporting files such Word documents, PowerPoint presentations and PDFs, into a structure that can be queried in a simple fashion to answer questions that would otherwise require time-consuming, error-prone manual aggregation.

As illustrated below, the semantically enriched data can be surfaced through SciBite's easy to use interface, via 3rd party search and visualization tools such as Spotfire and Linkurious, using the user interface of the ELN interface itself, or a combination of these.
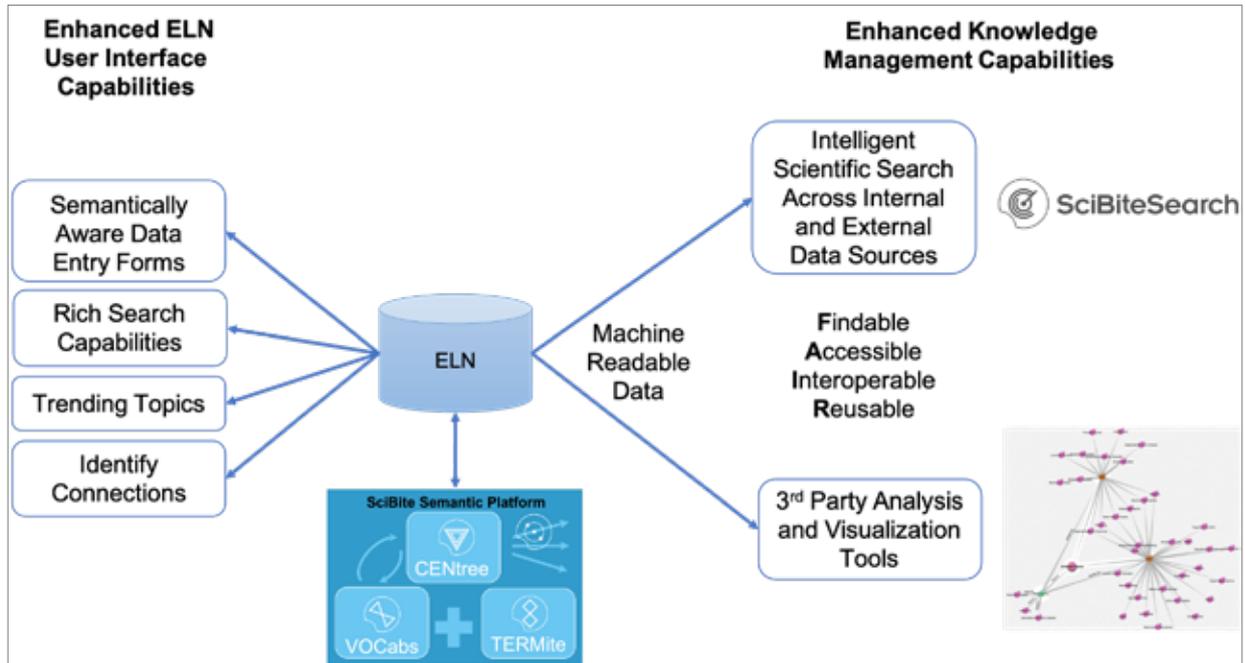
Figure 1: Semantic enrichment of ELN data

The benefits of a semantically enriched ELN

Most ELNs only have rudimentary search capabilities. For example, a search of a typical ELN for the Alzheimer's related gene, PSEN1, would miss references to synonyms such as Presenilin-1, AD3 and PSNL1.

Through semantic enrichment, SciBite ensures that all relevant data is found, regardless of which synonym is used as the search term. SciBite not only makes it simpler to interrogate ELN data, it also facilitates more complex ontology-based questions.
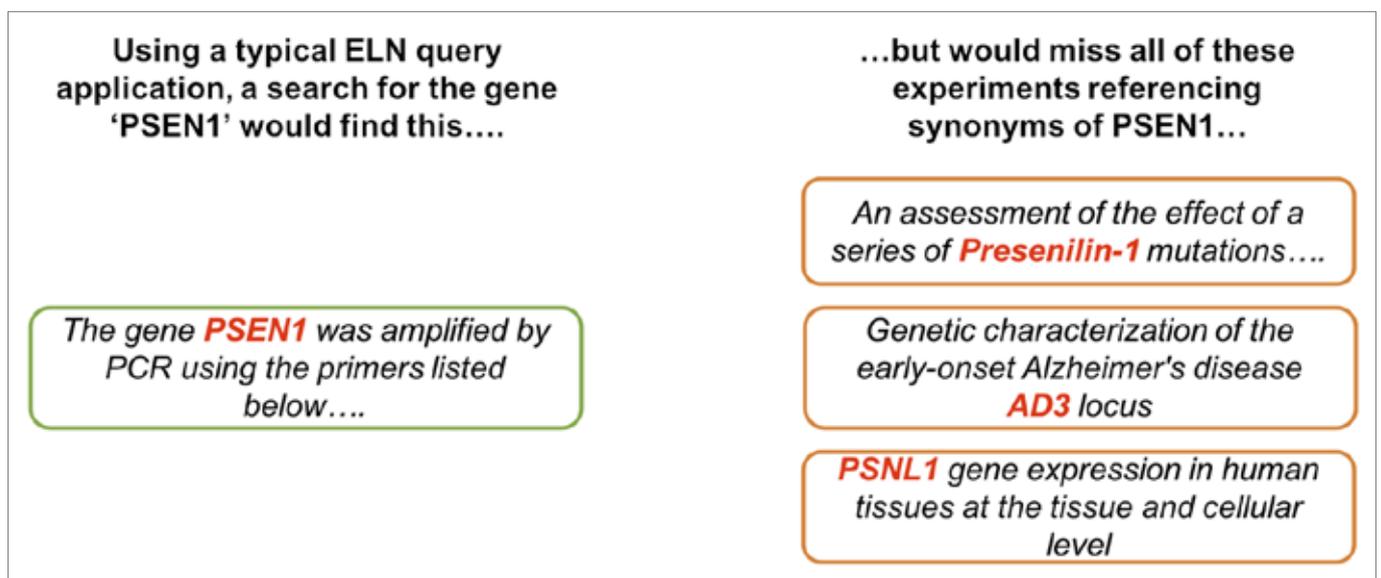

Figure 2: ELN search tools miss synonyms of search terms of interest

Questions that can be rapidly answered with semantically enriched ELN data

- Find all references to project ABC-101, regardless of the syntax used by the author (e.g. ABC-101, ABC101 and ABC 101)
- Find all experiments that reference a compound of interest used in combination with one or more other compounds of interest.
- Find all experiments for a specific target across the organisation, regardless of which synonym was used by the author of the experiment.
- Which projects are investigating potential biological therapeutics?

- Which targets have we studied that are associated with inflammatory disorders?
- Which diseases have we studied for both a target of interest and other targets in the same class and what were the outcomes?
- Which pre-clinical studies have utilised a specified mouse model?
- Which experimental techniques are growing across the organisation and would benefit from a core facility?

The benefits of a semantically enriched ELN

Most ELNs only have rudimentary search capabilities. For example, a search of a typical ELN for the Alzheimer's related gene, PSEN1, would miss references to synonyms such as Presenilin-1, AD3 and PSNL1.

Through semantic enrichment, SciBite ensures that all relevant data is found, regardless of which synonym is used as the search term. SciBite not only makes it simpler to interrogate ELN data, it also facilitates more complex ontology-based questions.
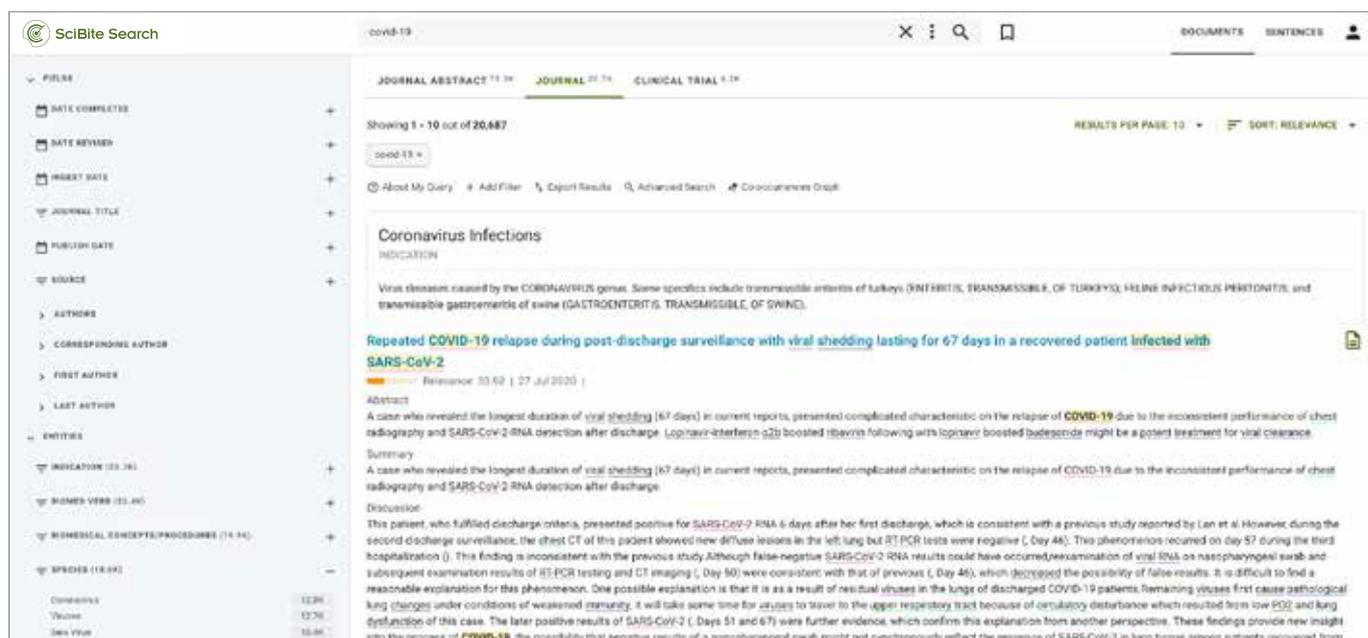
Figure 3: Summarising all relevant terms identified within an experiment (left side of screen)

Identify important themes

SciBite accurately marks-up all relevant terms and concepts present within an experiment, without being limited by the indexing terms used by the ELN. Scientists can, as a result, rapidly identify the topics covered in an experiment, easily interpret the text and instantly get a feel for what the experiment is about.

This can be expanded beyond individual experiments to provide a summary of all the content from one or more studies and present the information to users in Spotfire dashboards and Linkurious network views. This delivers quick insight into what a study is about, without having to read all the experimental records within it. Similarly, it is possible to analyse all the experiments within an ELN and, for example, identify the most frequently mentioned genes or biomarkers.

The results of such analyses deliver valuable business insight and enable companies to understand their research landscape, such as identifying trending topics within the organisation and reveal how the volume of information associated with the current 'hot' target has changed over time. Similarly, it can help with planning activities by assessing the amount of work associated with a disease or target of interest compared to its relative importance to the organisation.

Angiotensin I Converting Enzyme 2	**Erb-B2 Receptor Tyrosine Kinase 4**	**Galanin Receptor Type 1**	**cGMP Phosphodiesterase V**
	Interleukin-1 Receptor Type 1	**AT-Rich Interaction Domain 1A**	**Triggering Receptor Expressed on Myeloid Cells 2** / **Phospholipase C Gamma 2**
		GATA Binding Protein 4	**ABI Family Member 3** / **Apolipo-protein E**
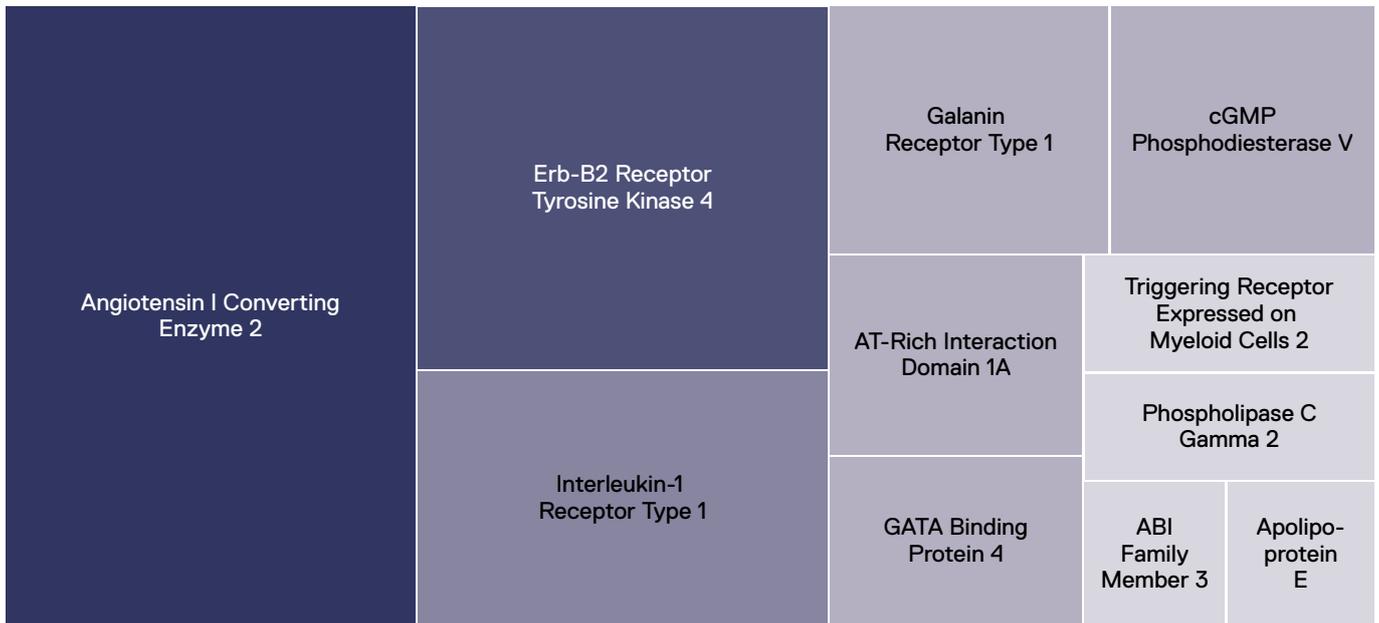
Figure 4: Visualising all frequently mentioned genes across multiple studies

Explore connections

When different experiments use different terms to describe the same thing, such as 'heart attack' and 'myocardial infarction', it can be difficult to make connections between them. SciBite provides the "semantic glue" to enable cross-linking between equivalent concepts described by different experiments, revealing otherwise hidden associations.

When writing or reviewing an experiment, scientists can be presented with a list of experiments similar to their own. SciBite can also reveal associations between individual scientists and the topics addressed in their experiments, which helps identify experts in a particular disease or methodology, regardless of what is written in their profile.

Individual studies are shown as green circles. Other circles indicate the main concepts (coloured by type) shared with other experiments, thus highlighting relationships between studies. Visualisation is of a neo4j database using the Linkurious browser tool.
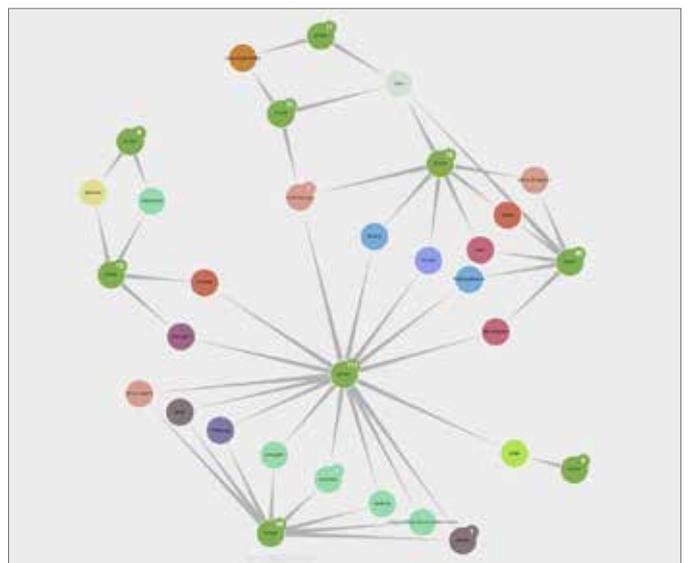
Figure 5: Interactive relationship network of experiments
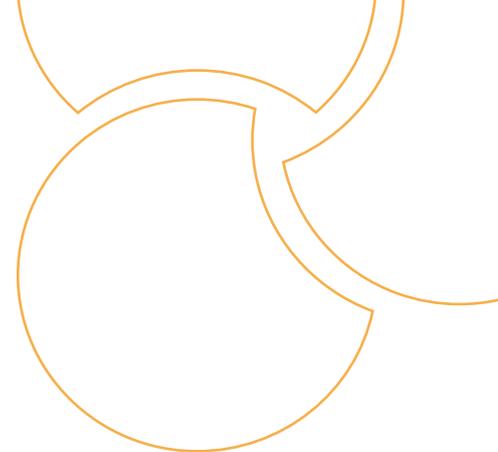
Intelligent data entry

Semantic enrichment should not be limited to the retrospective analysis of existing data. CENtree's APIs can be leveraged to make any browser-based ELN data entry form semantically intelligent, enabling organisations to enrich new data at the point of capture and ensure that it is 'born semantic'.. For example, by leveraging this capability, a field to capture 'Species' can be made both semantically aware and computationally accessible without adding unnecessary burden to scientists who subsequently enter data. Instead of being presented with restrictive and lengthy drop-down menus, users can enter text into semantically aware fields and have relevant terms suggested to them as they type.

Figure 6: Semantically-intelligent data entry forms

Gain a holistic view

In addition to ELN data, SciBite can also apply semantic enrichment to other internal and thirdparty data sources such as internal document repositories, scientific literature, patents and commercially available databases. Since the resulting data will be as well-structured and interoperable as public data, it becomes facile to integrate multiple disparate sources and gain a holistic view of everything that is known, both internally and externally, about any compound, target or disease and understand the trends associated with any topic of interest.

This enables users to ask questions across data sources that would have otherwise been time consuming or impossible to answer. For example, it becomes easy or a company to define its internal strengths and understand its competitive position relating to specific targets or diseases and assess if it is working on things that its competitors aren't (and vice versa). Similarly, it is possible to identify which companies or institutions are working in which disease or technology area of interest to explore options for collaboration. In each case, alerts can be setup to ensure information is highlighted to the right people in a timely manner.
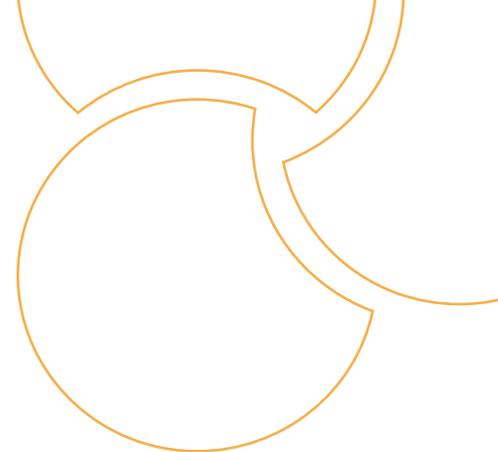
Summary

Most pharmaceutical companies are unable to realise the true value of the data stored in their ELN. SciBite's unique combination of retrospective and prospective semantic enrichment immediately brings scientific intelligent search to any ELN platform, enabling the data within it to be readily integrated with other sources and making it computationally accessible for automated analysis. SciBite delivers a cost-effective solution to unlock the wealth of information managed within ELN databases, opening up new possibilities to mine the data more effectively and derive valuable scientific and business insights.

About SciBite

SciBite is an award-winning semantic software company offering an ontology-led approach to transforming unstructured content into machine-readable clean data. Supporting the top 20 pharma with use cases across life sciences, SciBite empowers customers with a suite of fast, flexible, deployable API technologies, making it a critical component in scientific data-led strategies. Contact us to find out how we can help you get more from your data.

To learn how SciBite can unlock the value of your data, speak to one of our experts today or email us at contact@scibite.com

SciBite

an ELSEVIER company

SciBite's data-first, semantic analytics software is for those who want to innovate and get more from their data. At SciBite we believe data fuels discovery and we are leading the way with our pioneering infrastructure that combines the latest in machine learning with an ontology-led approach to unlock the value of scientific content. Supporting the world's leading scientific organisations with use-cases from discovery through to development, SciBite's suite of fast, flexible, deployable API technologies empower our customers, making it a critical component in scientific, data-led strategies. Contact us to find out how we can help you get more from your data.

To learn how SciBite can unlock the value of your data, speak to one of our experts today or email us at contact@scibite.com