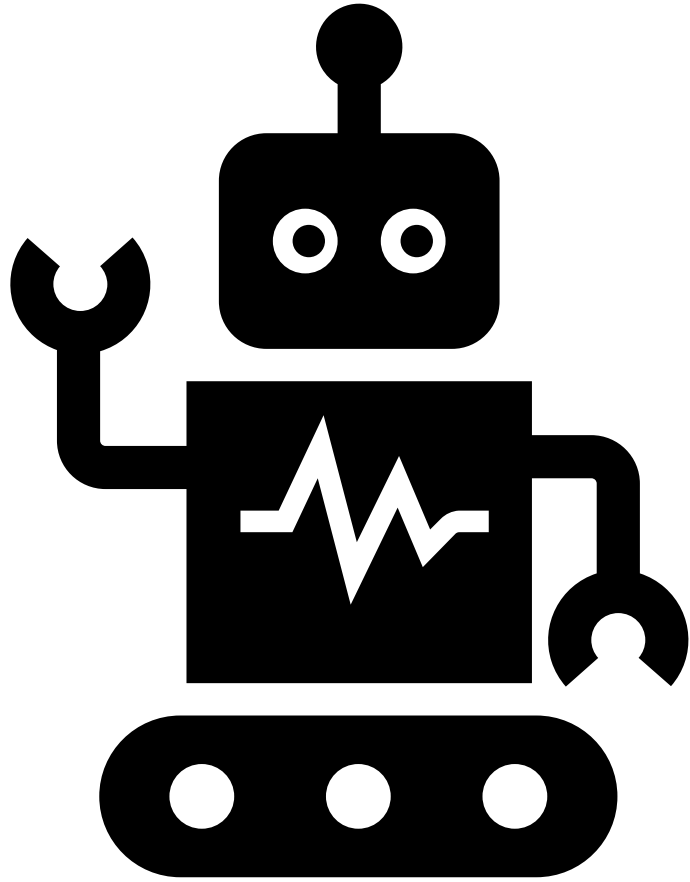# SciBite

The #cleandata company

THE LANGUAGE OF SCIENCE

Machine Learning/AI initiatives?

Harvard Business Review

If Your Data Is Bad, Your Machine Learning Tools Are Useless

by Thomas C. Redman

APRIL 02, 2018

Harvard Business Review

Only 3% of Companies' Data Meets Basic Quality Standards

by Tadhg Nagle, Thomas C. Redman, and David Sammon

SEPTEMBER 11, 2017

Pistoia Alliance

Data Quality in support of AI/ML: Pistoia Alliance CoE for AI Webinar Series

Date: 10 Dec 2018: 4pm - 5pm GMT, 11 am ET/5pm CET

Ontologies Are The Key To Unlocking FAIR

Findable **A**ccessible **I**nteroperable **R**eusable

Image credit: Sungya Pundir, Wikimedia Commons CC BY-SA 4.0

€10.2bn Can you afford not to adopt FAIR?

European Commission

Cost-benefit analysis for FAIR research data

….Therefore, we are confident that the true cost of not having FAIR research data is much higher than the estimated **€10.2bn** per year...

http://bit.ly/fairify

Our Purpose

Combining world class ontologies & machine learning to revolutionise the access to and utilisation of scientific information

Transforming unstructured text into contextualised, machine readable data suitable for new discovery

VOCabs : Beyond Ontologies

at the heart of everything!

- Maintained by our dedicated team using expert curation and machine learning

- Comprehensive coverage

- Aligned to industry standards to maintain interoperability

- Enriched with synonyms and rules to manage the complexity of scientific language

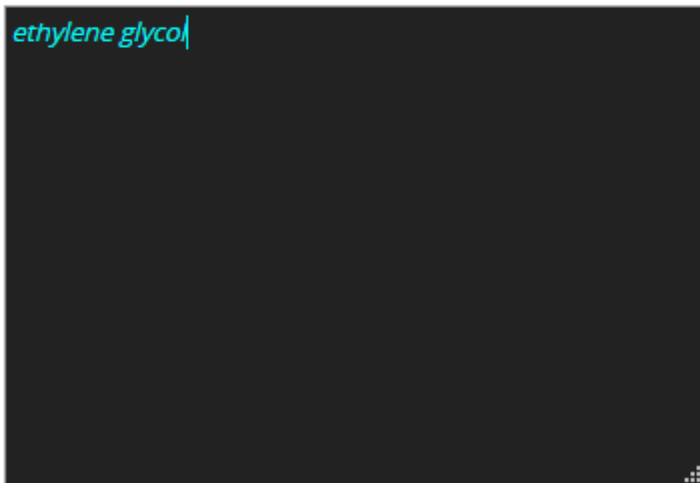- Customize, augment our existing or deploy your own vocabularies

ChemAxon integrated as VOCab
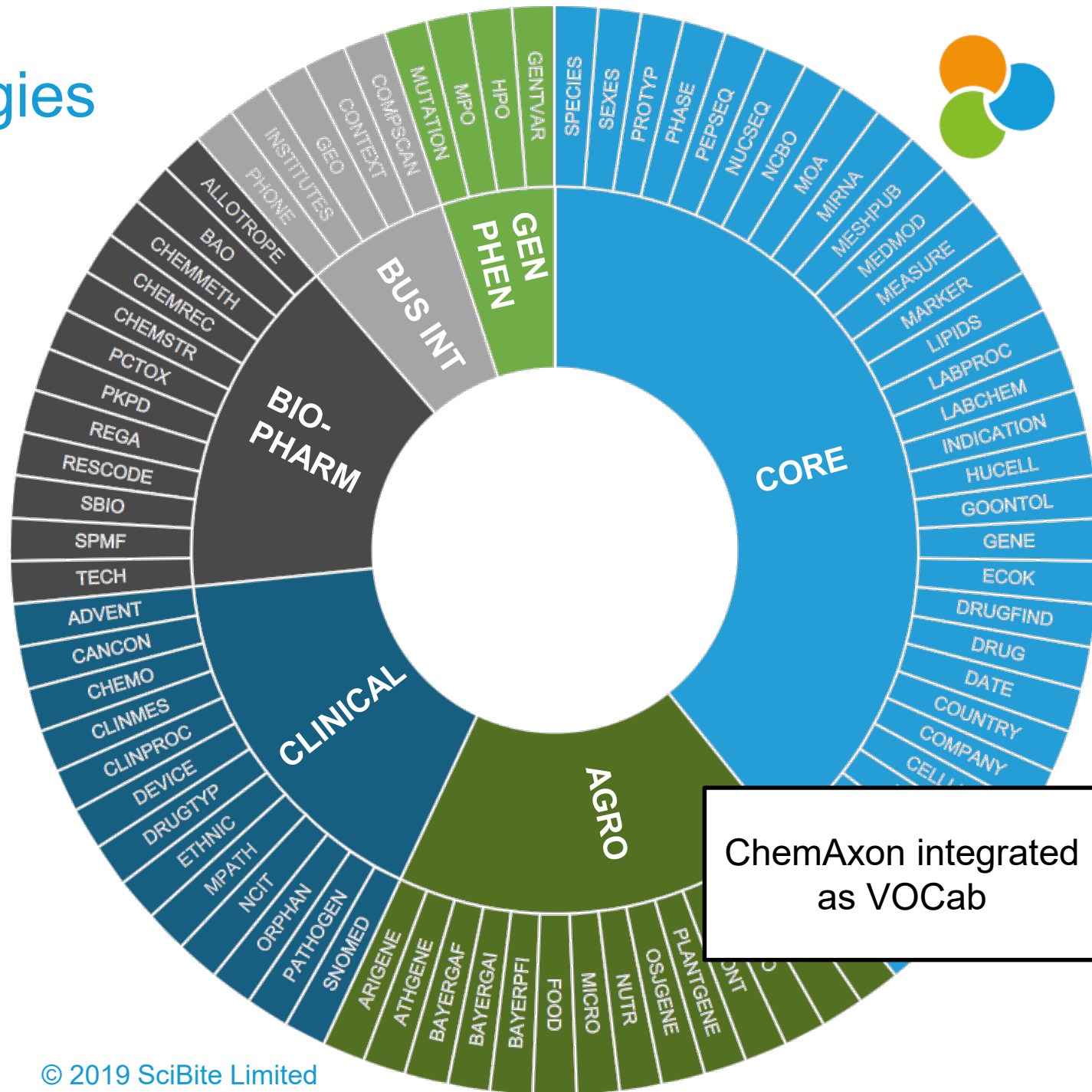
ChemAxon Integration - Example

ChemAxon lookup service seamlessly integrated into TERMite, SciBite's Biomedical Named Entity Recognition tool

Explore TERMite [Show/Hide]

Just Type Below! *(Example)*

ethylene glycol

As-You-Type Results

Summary

ethylene glycol

Termite Hits

Poor quality (ambiguous) hits shown in red/grey and should be treated with much caution, often users filter these out of their results as they are of low accuracy.

Type	Name	ID	Hit Synonyms	Poor
CA_LOCAL	ethylene glycol	ethylene glycol	ethylene glycol	N

SciBite Semantics Empower Scientific Infrastructure

Data Stewardship | **F.A.I.R & Search** | **Data Mining & AI**

| Semantic MDM | Smart Data Entry | Improve Integrity | F.A.I.R Data Catalogues | Semantic Search | Analytics & ML/A.I | Semantic Q&A | Knowledge Graphs |

Ontology Management

Electronic Lab Notebooks

Genomic Data Processing

Assay Registration

Asset Management

Portfolio Analytics

L.I.M.S.

Clinical Data Search

Enterprise Search

Departmental Search

Pharmacovigilance

Horizon Scanning

Drug Repurposing

Phenotype Triangulation

Outcomes Prediction

Diagnostic prediction

Phenotype triangulation
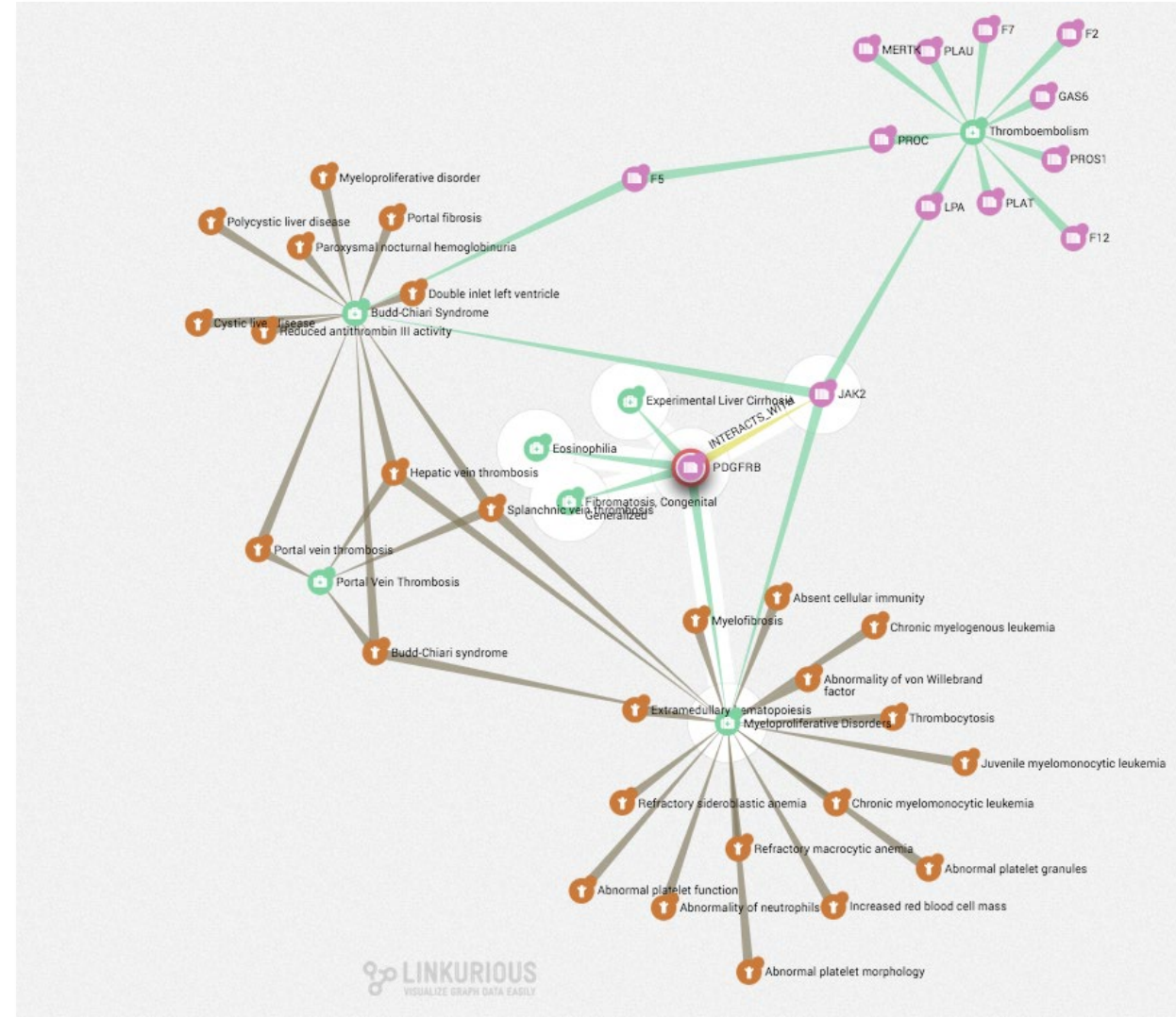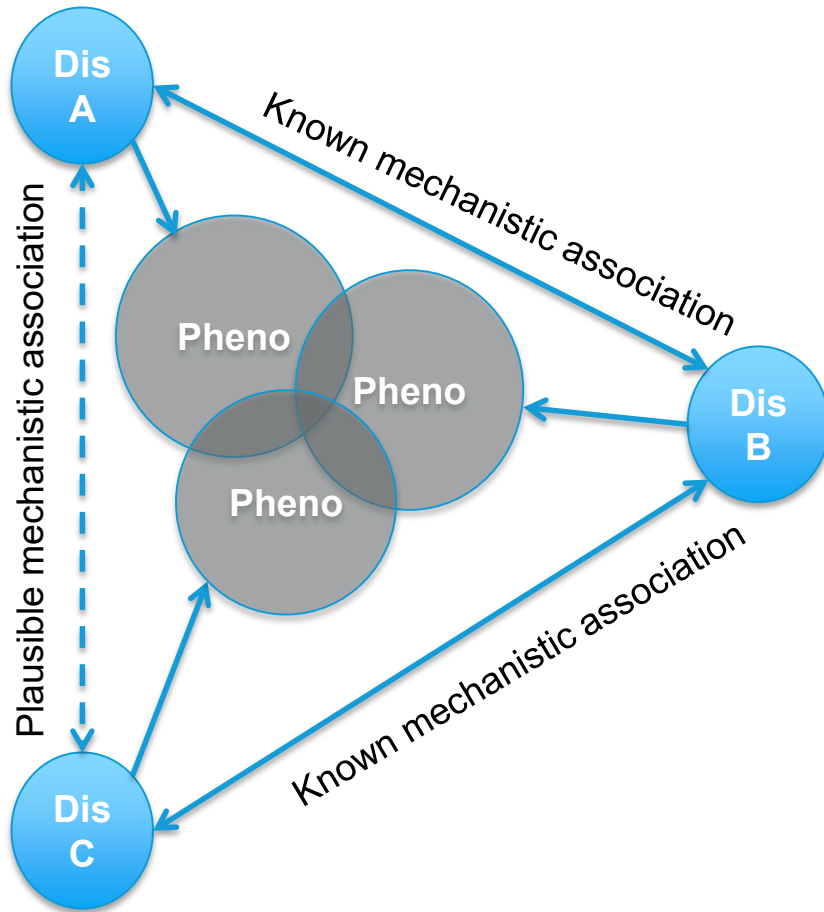
Building network views of disease-phenotype mappings

© 2019 SciBite Limited

Adding Best-in-Class Chemistry Enables Repurposing

By layering in ChemAxon's Named Chemical Entity Extraction, we expand the use of Phenotype Triangulation to Drug Repurposing at-scale