aarth
MEANINGFUL DATA

The Pharma Revolution

How graph technology is shifting the paradigm by digitizing knowledge

From Prescriptions to Predictions: How Graph Technology is Transforming Healthcare & Life Sciences

"Knowledge has a beginning but no end."

Geeta Iyengar

INDEX

Introduction

Graph technology, also known as graph database or graph computing, is a rapidly growing field that is revolutionizing the way we store, manage, and analyze data. With the ability to handle complex relationships and connections, graph technology is being used to solve some of the most challenging problems.

Graph technology is revolutionizing the way the pharmaceutical industry approaches drug repurposing, discovery and development with the ability to handle large amounts of data and identify relationships and connections between different data points, graph technology enables the discovery of new drug targets, the identification of new drug interactions, and the analysis of complex patient data to optimize clinical trials.

The pharmaceutical industry is undergoing a revolution with the use of knowledge graphs. By representing complex data as a network of interconnected entities and relationships, knowledge graphs provide a powerful tool for drug discovery and development. They enable researchers and pharmaceutical companies to gain new insights and make more informed decisions by linking disparate data sources and identifying patterns and connections that would be difficult to see with traditional methods. Knowledge graphs can be used to represent everything from drug molecules and proteins to clinical trial results and patient data, providing a holistic view of the drug development process. They also enable the integration of external data sources such as scientific literature, patents and clinical trial data, providing a broader perspective on the field.

With the ability to handle large amounts of data and update in real-time, knowledge graphs are becoming an essential tool for the pharmaceutical industry to speed up drug development, optimize clinical trials and bring new drugs to market faster. Additionally, graph databases enable pharma companies to streamline their operations and gain a competitive edge by providing real-time visibility into their supply chains and enabling the tracking of drugs through their lifecycle. Overall, graph technology is a boon to the pharma industry, helping companies to develop new drugs faster and more efficiently while providing a wealth of new insights into the drug development process.

> "A new horizon of data driven decisions will be focused on hybrid intelligence which is data fusing with knowledge"

KNOWLEDGE refers to reasoning, process, rules, business, cognition, objects, events, concepts & so on...

Representing knowledge in form of graphical networks consisting of nodes & relationships is **GRAPHICAL DATA**

Context   Experience

Data ⟷ Information ⟷ Knowledge
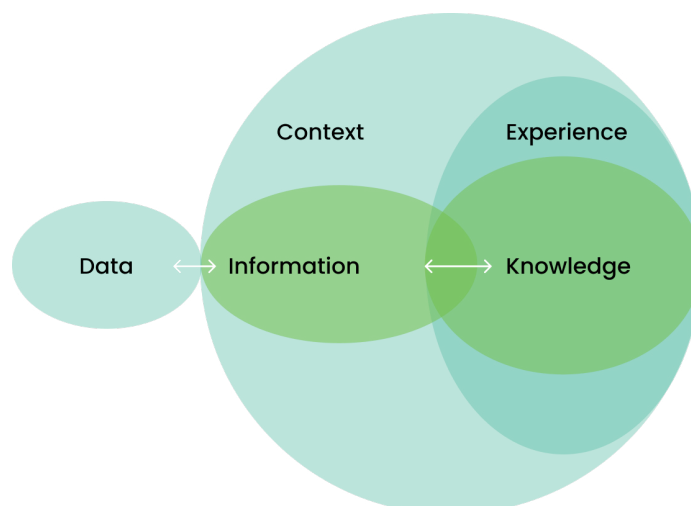
Knowledge graphs Empower Datawarehouse / Data lakes

Data warehouses continue to be valuable for providing analytics for a wide range of use case. You may be among the many who are completely rethinking your data warehouse or using different in-house technologies to tackle the task with increasing data & existing databases which are not scalable to large extent.

Today's trend is for pharma companies to consider new deployment options such as the cloud, new licensing models like subscriptions, open source, community and even new paradigms for analyzing data that can save time, cost and resources.

Here is a solution, Graph tech are schema-free & will work with structured & unstructured data. They enable complex analytics against the data and, can operate in real-time. Graph databases, to be used in conjunction with traditional data warehouses. This in turn reinforces the whole concept of the logical data warehouse which, in turn, means consideration of what you need to support a distributed, heterogeneous warehousing environment required by any time for users.
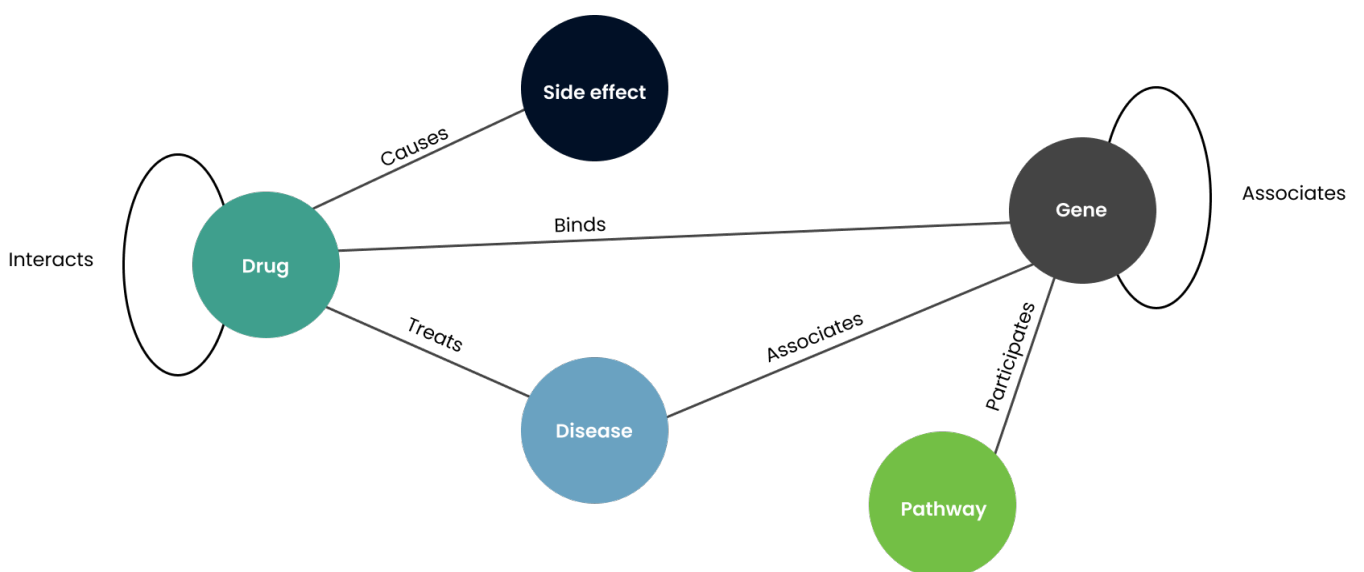
Graph data helps in simplifying Metadata & Master data, usually data team have less time to manage schema for incoming data from multiple sources. Handling schemas & potential changes are time consuming challenges; however, NoSQL databases are popular for their ease of use due to flexible schemas. Graph databases can power up data warehouses & data lakes simplifying meta data management, by configuring all your data with unlimited scalability & limit your need to set up complicated ETL & data transformation.
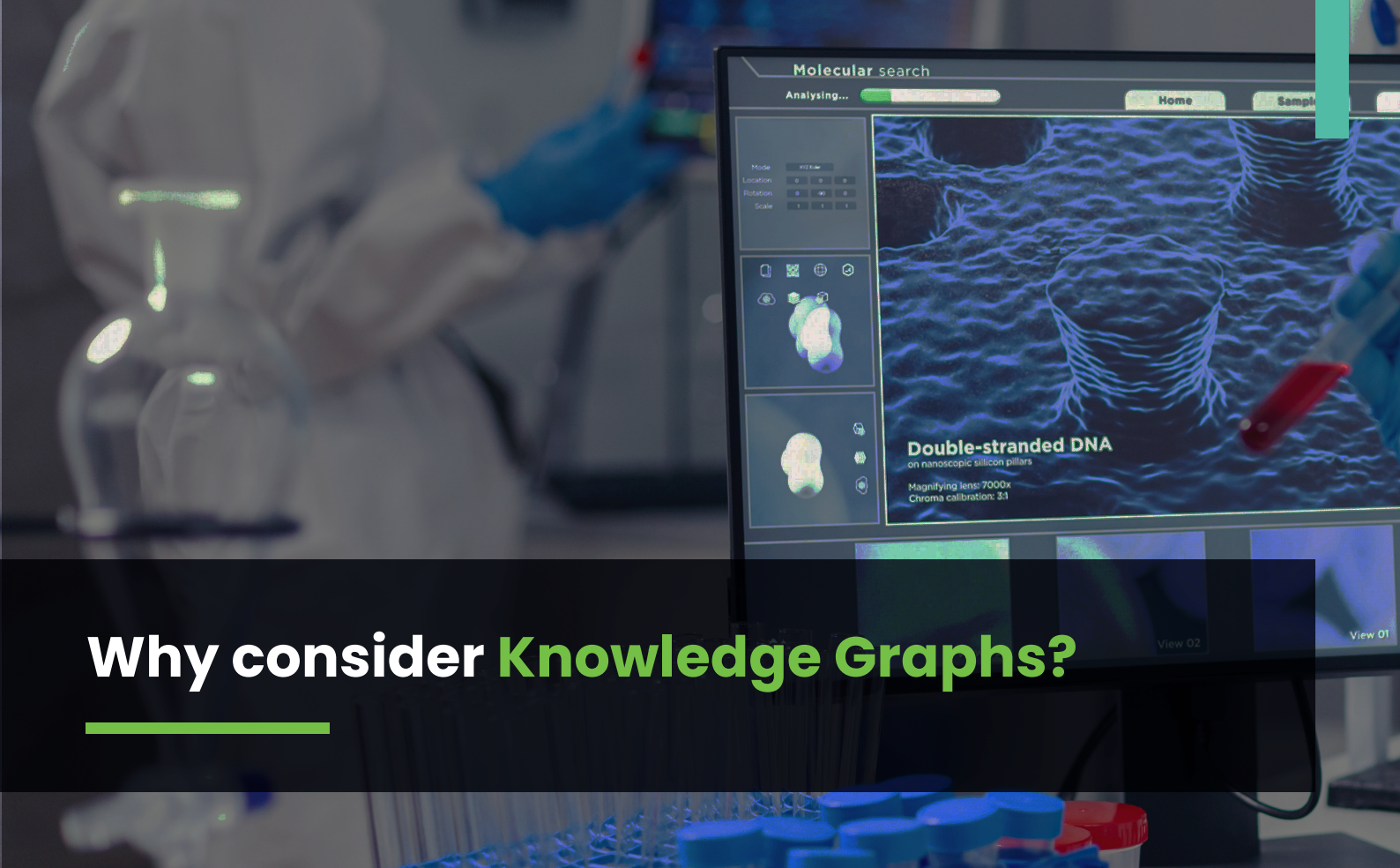
A key component of master data management (MDM) is to supply meaningful views of disparate data. The fact/property underpinnings of graph databases are designed to optimize those views. You can easily model both hierarchical and nonhierarchical master data in a graph, making it easier than an RDBMS for visualizing data relationships.

Below is a simple representation of Graph database in Drug repositioning

Side effect

Causes

Interacts

Drug

Binds

Gene

Associates

Treats

Associates

Participates

Disease

Pathway

Why consider Knowledge Graphs?

One of the most valuable and fundamental uses of Knowledge Graphs is to express relationships between data. Knowledge Graphs visualize the nature of underlying relationship and how they are interconnected to gain sufficient understanding of the landscape to improve intelligent strategic business decisions.

On a high level, Knowledge Graphs helps pharma industry with following benefits:

1. Combined siloed data.
2. Combine structured & unstructured data.
3. Help businesses & operations make more informed decisions.
4. Summarize relationships.
5. Insights from hierarchical data.
6. Revealing communities.
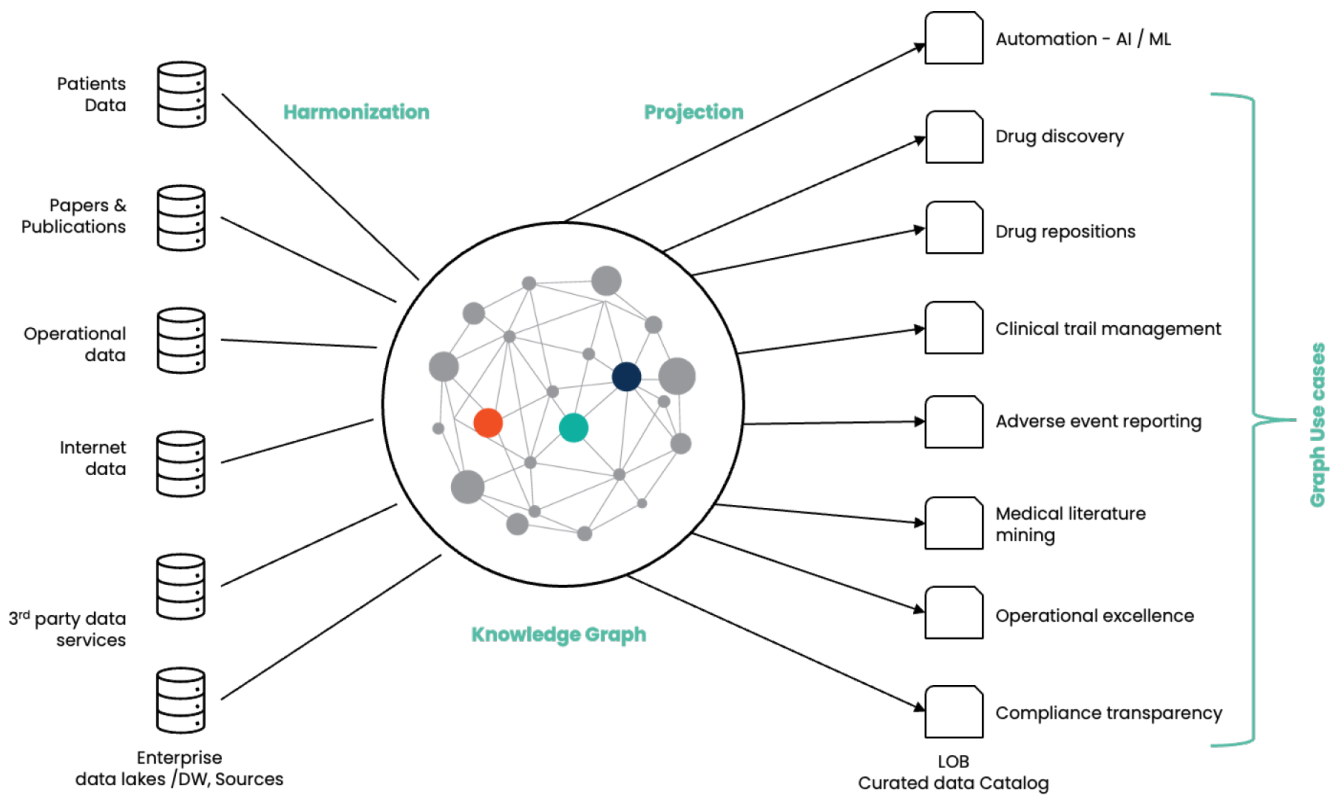7. Visualizing a flow of information.
8. Network data.

Data is increasing! The problem with existing traditional databases are not scalable upto large scale which in fact organizations need to have huge infrastructures again which will be built on pipelines. Where knowledge graphs can make difference in data use cases listed few below,

- Knowledge graphs helps facilitate end to end analytics enabling maximum utilization of AI & ML.
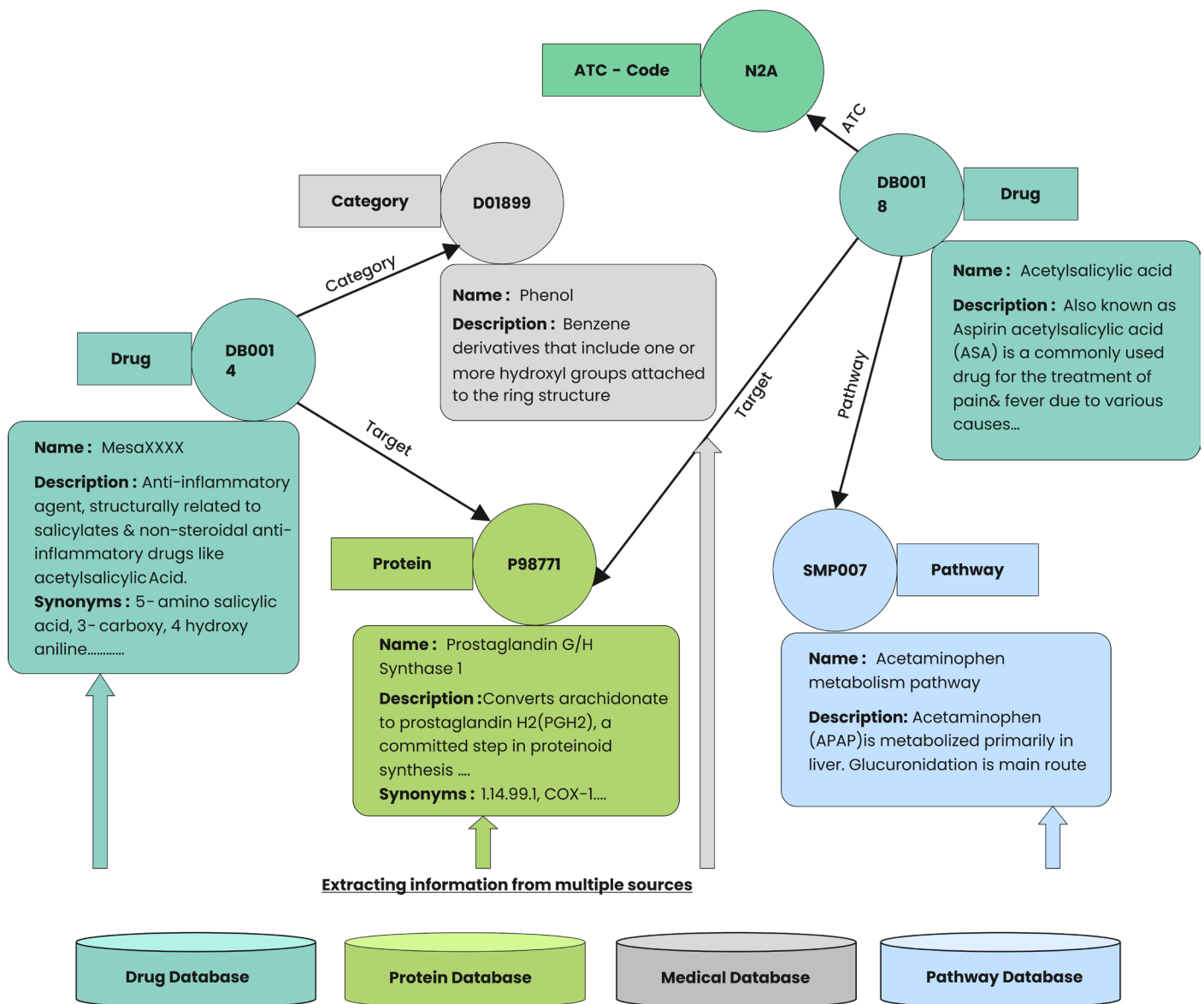- Prevent missing relationships enabling users to

answer questions which were not solved from long time.

- Semantic Graphs helps to capture & connect data outside operational use case.
- Graph databases connect in a multi-direction with all users from producers to consumers.
- Less expenditure on data infrastructures.
- Overcome end user discovery challenges within seconds.
- Clear ownership of core concepts.
- Less involvement of data team to fix the mess.
- Come out of technical debt easily with long term solution.
- One time investment on increasing query complexity.
- Actual real time solutions for all your data needs.
- No investment on multiple products to build solutions.

Patients Data

Papers & Publications

Operational data

Internet data

3ʳᵈ party data services

Enterprise data lakes /DW, Sources

Harmonization

Projection

Knowledge Graph

Automation - AI / ML

Drug discovery

Drug repositions

Clinical trail management

Adverse event reporting

Medical literature mining

Operational excellence

Compliance transparency

Graph Use cases

LOB
Curated data Catalog

Some Illustrations of Knowledge Graphs in Pharma

Extracting information from multiple sources
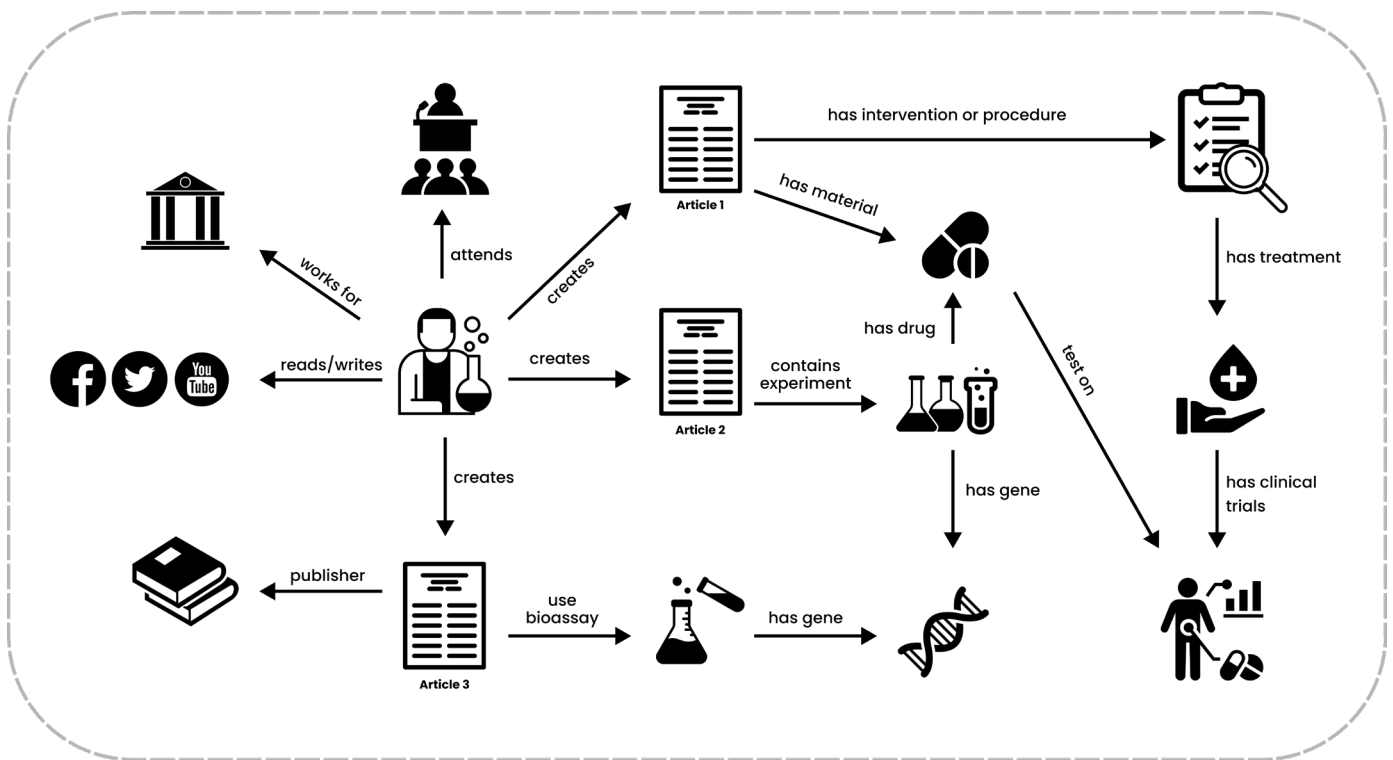
Use Cases for Pharma Industry

Use case 1:

Drug Repositioning / Repurposing

Drug repositioning, also known as drug repurposing, is the process of identifying new therapeutic uses for existing drugs. Knowledge graphs can be used to support this process by storing and analyzing large sets of data on drugs, diseases, and the relationships between them.

In drug repositioning, knowledge graphs can be used to represent the complex relationships between drugs, target proteins, disease symptoms, and other related entities. For example, a node representing a drug can be connected to nodes representing the target protein it binds to, the disease it is currently used to treat, and any related side effects. This allows researchers to easily explore the connections between different entities and identify potential new therapeutic uses for the drug.

Graph databases can also be used to store and analyze large sets of data from high-throughput experiments, such as genomics and proteomics data. This can help researchers understand the underlying biology of a disease and identify potential new therapeutic targets for a drug.

Additionally, graph databases can be used to integrate data from various sources such as scientific literature, patent information, and external databases. This can help researchers stay up to date with the latest developments in their field and identify potential new therapeutic uses for drugs. Use of knowledge graphs in drug repositioning can help researchers more effectively analyze and understand complex data sets, identify potential new therapeutic uses for existing drugs, and accelerate the drug development process.

Business benefits:

- **Cost savings:** Repositioning existing drugs can be significantly less expensive than developing new drugs from scratch. Graph technology can help researchers quickly and efficiently identify potential new therapeutic uses for existing drugs, which can save a lot of time and money.
- **Faster time to market:** Repositioning drugs that have already been through clinical trials can significantly reduce the time it takes to bring a new drug to market. By using graph technology to identify potential new therapeutic uses for existing drugs, companies can bring new treatments to patients more quickly.
- **Increased revenue:** Bringing new therapeutic uses for existing drugs to market can increase revenue for pharmaceutical companies. By repositioning drugs, companies can extend the life of a drug and generate additional revenue from existing products.
- **Reduced risk:** Developing new drugs is a risky and expensive process. By repositioning existing drugs, companies can reduce the risk of failure and invest in more promising opportunities.
- **Improved data analysis:** Graph technology allows for more efficient and accurate data analysis. With graph technology, researchers can more easily identify patterns, connections and insights in the data, which can help them make better decisions.
- **Better IP protection:** As graph technology can be used to integrate data from various sources such as scientific literature, patent information, and external databases, it can help researchers to identify potential IP opportunities that have been previously overlooked.
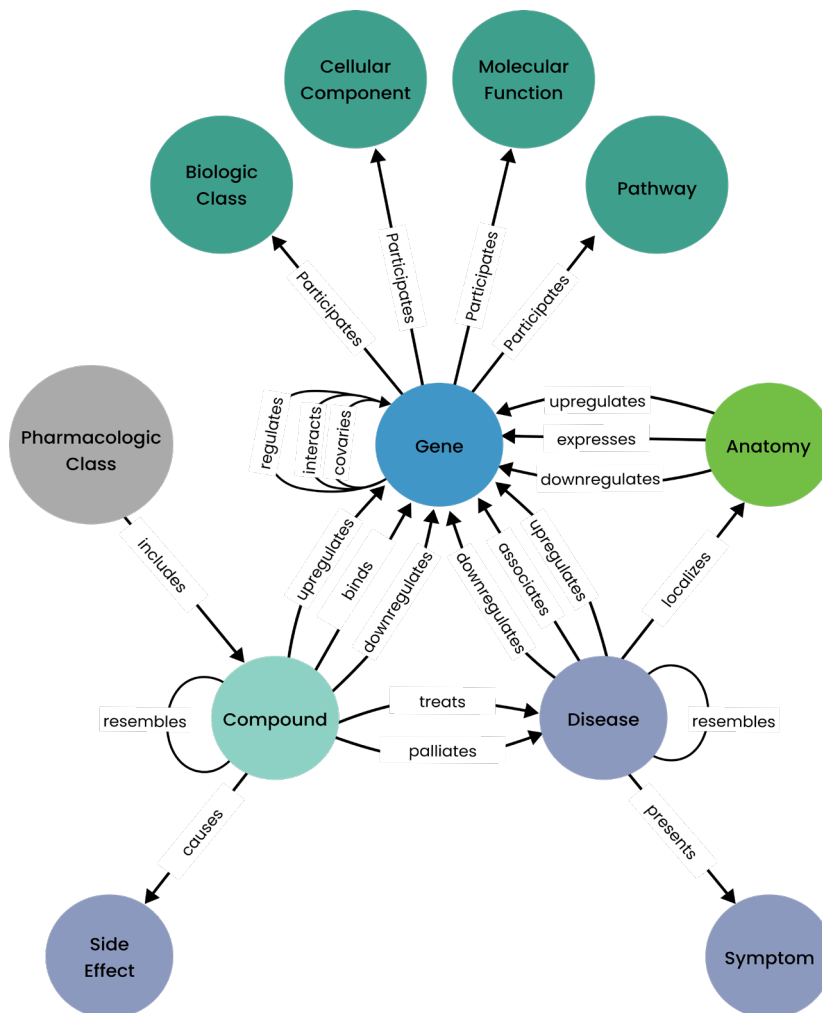
Use case 2:
Clinical Trail Management

Clinical trials are studies aiming at determining the safety and efficacy of interventions, treatments or investigational drugs on human subjects. Effective and successful clinical trials are essential in developing new drugs and advancing new treatments. The average cost of a single phase in clinical trials ranges from 1.4 million up to 52.9 million US dollars. In addition, the success rate of the clinical trials is considerably low.

As reported for certain therapeutic groups like Oncology, the overall success rate of clinical trials could be as low as 3.4%. The high cost and low success rate of clinical trials motivate deliberate analysis of existing clinical trials, inferring knowledge from them, utilizing existing clinical trials in innovative ways, and accordingly carefully designing future clinical trials., KG has been proven to be an effective representation for knowledge inference purposes. Constructing a KG over clinical trial data is vital for advancing the analysis and research of clinical trials.

Any data representing medical entities such as studies, drugs and conditions, drugs used in studies, adverts events, outcomes. Are fused with knowledge to gain rich information using Graph tech enabling more biomedical applications (e.g., adverse drug event prediction, outcome prediction) than the existing knowledge base in clinical trials.

This results in demonstrating its potential utilities in various applications such as drug repurposing and similarity search, embedded analysis & other application uses like drug event prediction & outcome prediction much more.
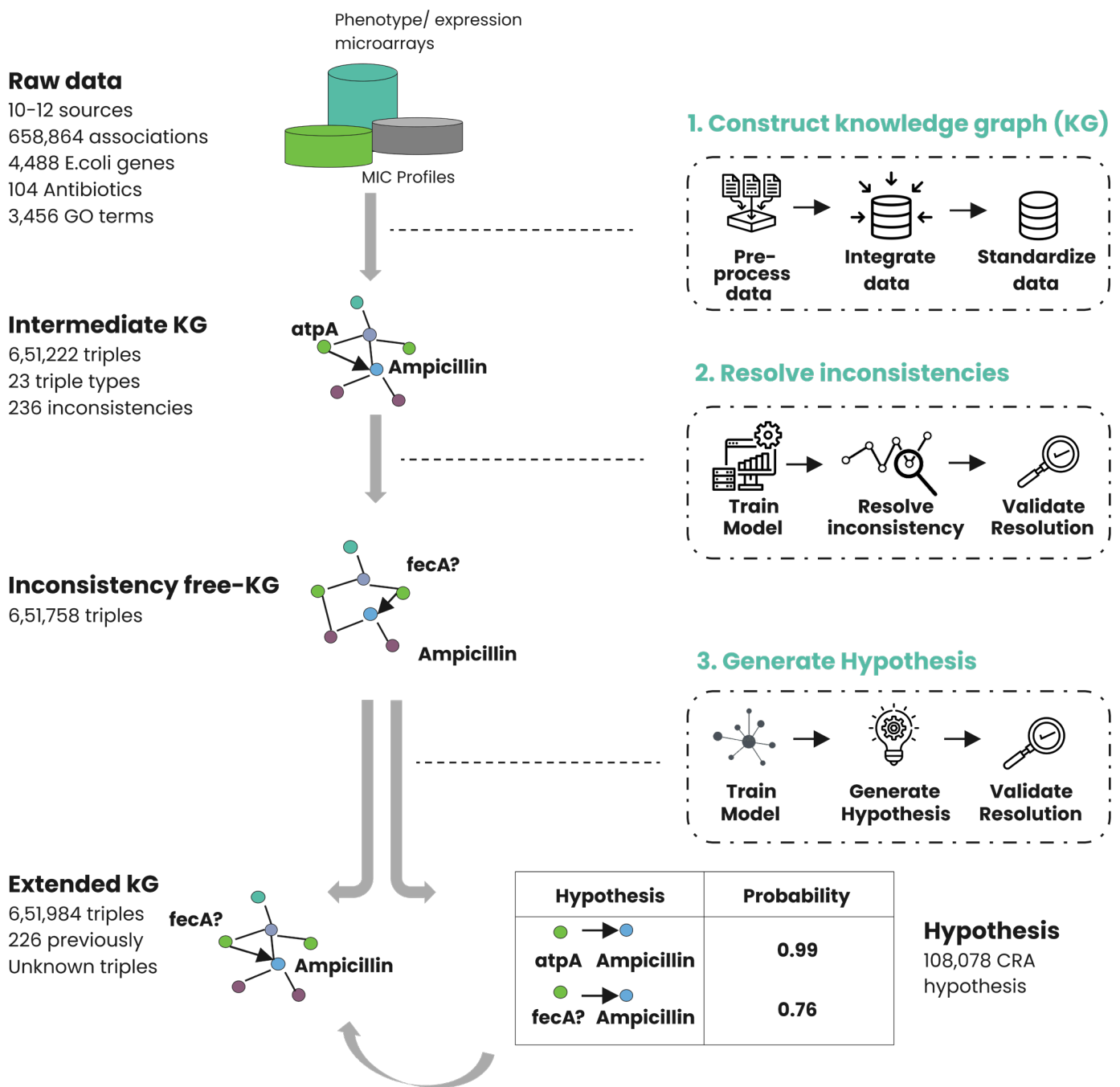
14

Business benefits:

- **Increased efficiency:** Knowledge graphs can help researchers quickly and efficiently analyze large sets of data from clinical trials, which can save a lot of time and money. This can also help to speed up the drug development process and get new drugs to market faster.

- **Improved data analysis:** Graph technology allows for more efficient and accurate data analysis. With graph technology, researchers can more easily identify patterns, connections and insights in the data, which can help them make better decisions.

- **Better patient safety:** Knowledge graphs can help researchers identify potential safety issues with a drug more quickly and effectively. This can help to ensure that drugs are safe for patients and can reduce the risk of adverse events.

- **Better decision making:** By using knowledge graphs to analyze data from clinical trials, researchers can make more informed decisions about the development of a drug. This can help to ensure that resources are being used effectively and that only the most promising drugs are being developed.

- **Better compliance:** Knowledge graphs can help researchers stay on top of regulatory requirements and ensure that trials are conducted in compliance with relevant guidelines.

- **Post analysis of failed clinical trails:** Graph technology has been successfully explored in recent years for many use cases among which big data mining is one & helps to create large interconnected knowledge databases failure analysis for demonstrating their utility to store, index, retrieve and analyze complex datasets from drug development pipelines.
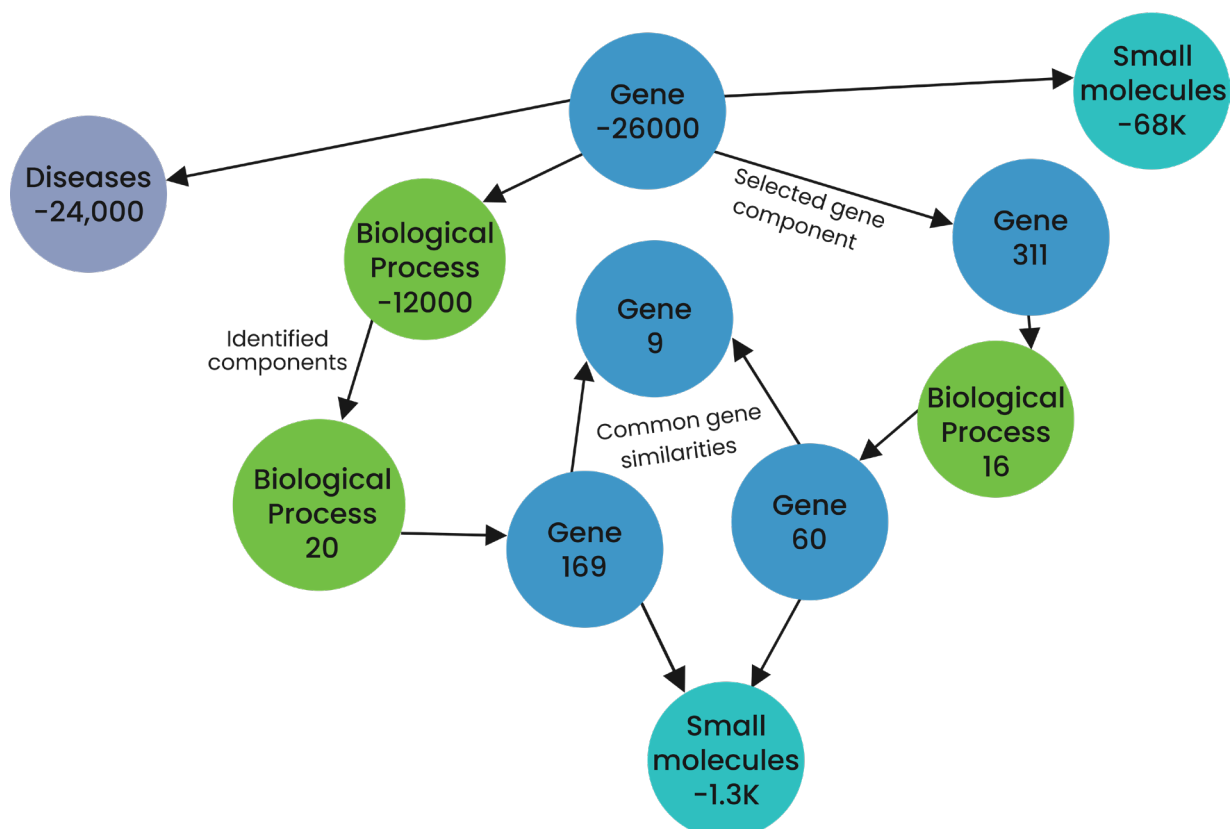
Sample process figure - Fusing knowledge & integrating decision support for clinical trail management to show discovery of antibiotic resistance genes

Phenotype/ expression microarrays

Raw data

10-12 sources
658,864 associations
4,488 E.coli genes
104 Antibiotics
3,456 GO terms

MIC Profiles

1. Construct knowledge graph (KG)

Pre-process data → Integrate data → Standardize data

Intermediate KG

6,51,222 triples
23 triple types
236 inconsistencies

atpA
Ampicillin

2. Resolve inconsistencies

Train Model → Resolve inconsistency → Validate Resolution

Inconsistency free-KG

6,51,758 triples

fecA?
Ampicillin

3. Generate Hypothesis

Train Model → Generate Hypothesis → Validate Resolution

Extended kG

6,51,984 triples
226 previously
Unknown triples

fecA?
Ampicillin

Hypothesis	Probability
atpA → Ampicillin	0.99
fecA? → Ampicillin	0.76

Hypothesis

108,078 CRA hypothesis

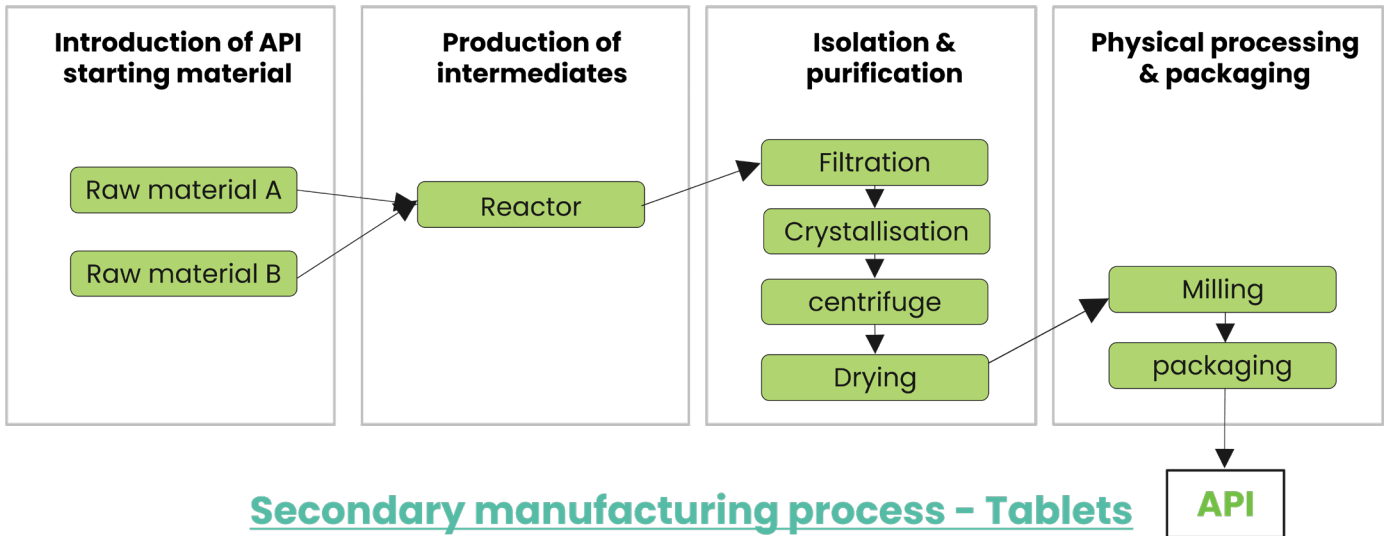Use case 3:

API Manufacturing Process

API (Active Pharmaceutical Ingredient) manufacturing is the process of producing the active ingredient in a pharmaceutical drug. The API is the part of the drug that provides the intended therapeutic effect. The manufacturing process for an API can vary depending on the type of drug and its intended use. API manufacturing is a complex and highly regulated process that requires a significant investment in equipment, facilities, and personnel. The quality of the API is crucial for the safety and efficacy of the drug, leaving less time for in-depth analysis.

It includes several steps for API manufacturing like synthesis, purification, characterization, formulation, quality control, packaging. Here, use of knowledge graphs in API manufacturing can help pharmaceutical companies improve the efficiency and effectiveness of their manufacturing processes, ensure regulatory compliance, improve supply chain management, and support knowledge management.
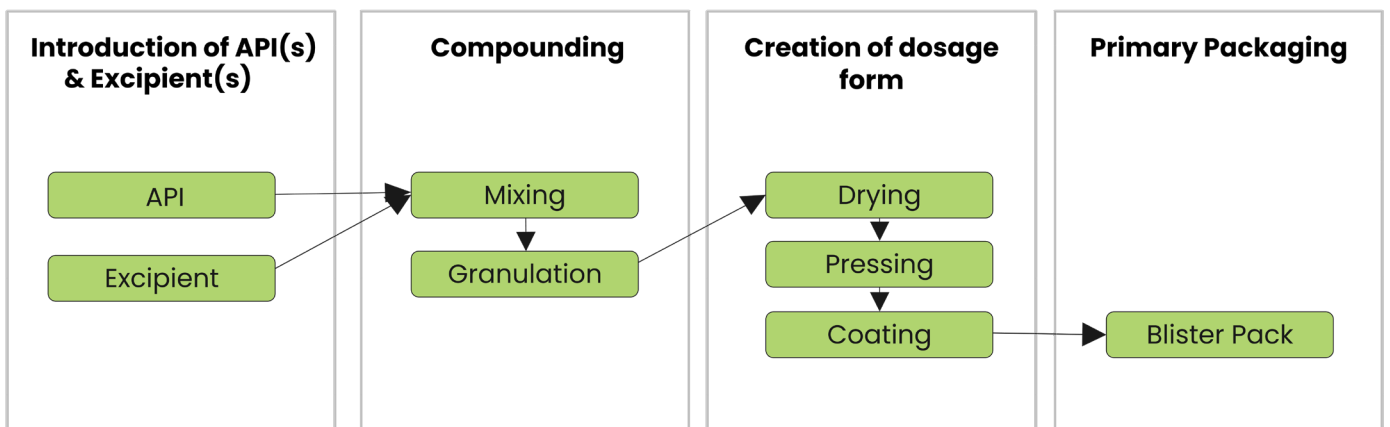
Business benefits:

- **Process optimization:** Graph tech can be used to store and analyze data from process development experiments, such as synthesis conditions, reaction kinetics, and process parameters. This can help researchers identify patterns and connections in the data and optimize the manufacturing process to improve the yield and purity of the API.

- **Quality control:** knowledge graphs can be used to store and analyze data from quality control tests, such as purity, potency, and stability. This can help companies ensure that the API meets the required specifications and regulatory requirements.

- **Supply chain management:** Graph tech are utilized to track the movement of raw materials and finished goods through the supply chain. This can help companies ensure that they are using high-quality raw materials, and that finished products are being shipped to the correct destinations.

- **Data integration:** Graph databases can be used to integrate data from various sources, such as electronic health records, sensor data, and external databases. This can help companies gain a more complete understanding of the manufacturing process and identify potential issues that may have been previously overlooked.

- **Predictive modeling:** Graph databases can be used to analyze data and build predictive models to identify patterns, trends and potential issues in the manufacturing process. This can help to optimize production and reduce costs, while also ensuring compliance with regulation.

- **Knowledge management:** Graph databases can be used to store and organize information about the API manufacturing process, such as process development and optimization, quality control, and regulatory requirements. This can help to ensure that knowledge is shared and reused across the company and can be used to train new employees.

- **Regulatory compliance:** Knowledge graphs can be used to store and analyze data related to regulatory compliance, such as inspection reports, certificates of analysis, and adverse event reports. This can help companies ensure that they are complying with relevant regulations and can respond quickly to any issues that arise.
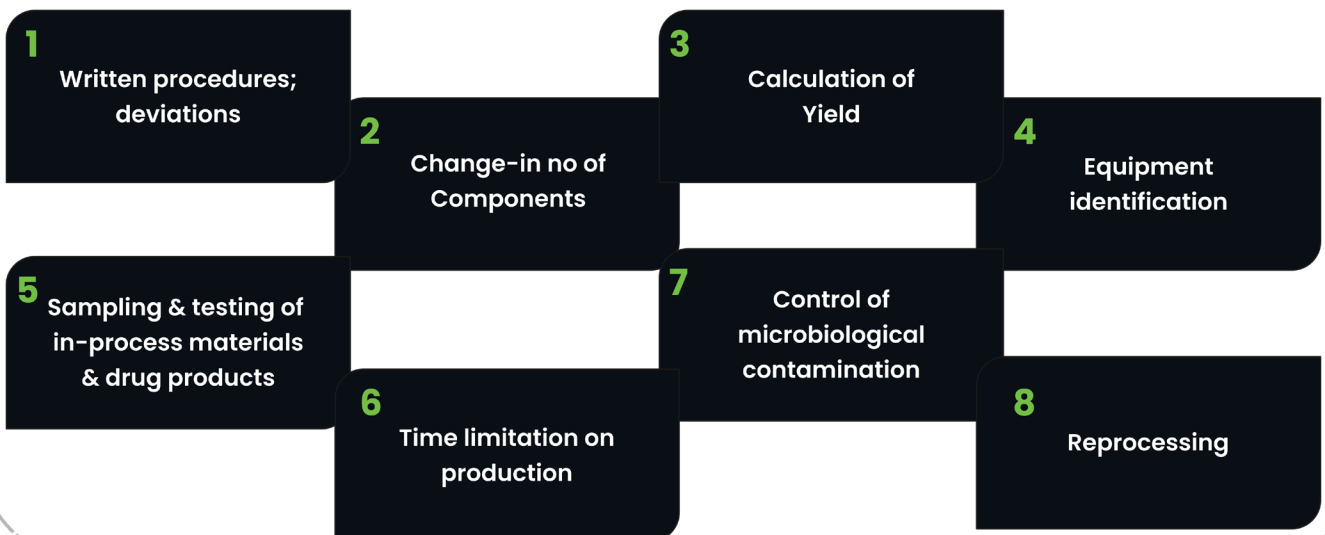
API manufacturing process

Introduction of API starting material

Raw material A

Raw material B

Production of intermediates

Reactor

Isolation & purification

Filtration

Crystallisation

centrifuge

Drying

Physical processing & packaging

Milling

packaging

API

Secondary manufacturing process – Tablets

Introduction of API(s) & Excipient(s)

API

Excipient

Compounding

Mixing

Granulation

Creation of dosage form

Drying

Pressing

Coating

Primary Packaging

Blister Pack

Production & Process Control

1 Written procedures; deviations

2 Change-in no of Components

3 Calculation of Yield

4 Equipment identification

5 Sampling & testing of in-process materials & drug products

6 Time limitation on production

7 Control of microbiological contamination

8 Reprocessing

Use case 4:
Supply Chain visibility & Improve Efficiency

Supply chain visibility refers to the ability to track the movement of goods and materials through the supply chain, from the manufacturer to the end consumer. Knowledge graphs can be used to support this process by storing and analyzing large sets of data on the movement of goods, suppliers, and the relationships between them. In supply chain visibility, knowledge graphs can be used to represent the complex relationships between products, suppliers, transportation, and other related entities.

Graph databases can also be used to store and analyze large sets of data from the supply chain, such as inventory levels, shipping information, and production data. This can help companies understand the flow of goods through the supply chain, identify bottlenecks, and make decisions about where to invest in improvements.

Graph databases can be used to integrate data from various sources such as sensor data, electronic records, and external databases. Knowledge graphs can be used to support the implementation of SCOR by storing huge multi dimensional data. Using knowledge graphs in SCOR can help companies more effectively analyze and understand complex data sets, identify potential issues in the supply chain, and make more informed decisions about supply chain operations. This can help to improve supply chain efficiency and support compliance with regulatory requirements.

Pharma Supply Chain Structure

Source	Make	Deliver

Demand

Suppliers — Manufacturers — Wholesale/distributors — Hospitals/Pharmacies — Patients

Supply

→ Flow of Supplies (Drugs)
--→ Flow of Return (Drugs)
↔ Flow of Data
← Flow of funds
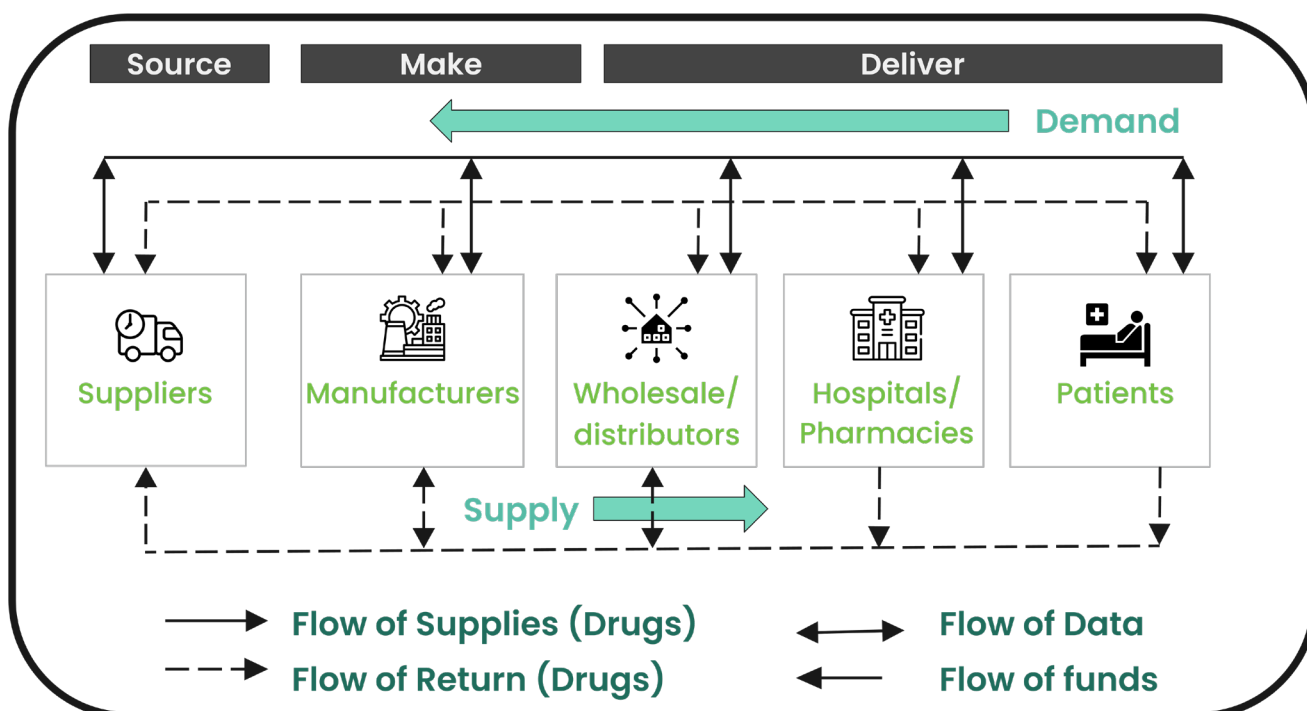
- **Improved supply chain visibility:** Knowledge graphs can help companies gain a better understanding of the flow of goods through the supply chain and identify bottlenecks or other issues that may be impacting performance.
- **Increased efficiency:** Knowledge graphs can help companies optimize the supply chain by identifying areas where performance can be improved. This can lead to reduced lead times, lower inventory levels, and improved delivery performance.
- **Better decision-making:** Knowledge graphs can help companies make more informed decisions about supply chain operations by providing a more complete view of the supply chain and allowing them to identify patterns and connections that may not be immediately apparent.
- **Reduced costs:** Knowledge graphs can help companies identify opportunities for cost savings, such as by reducing inventory levels or streamlining logistics operations.
- **Better compliance:** Knowledge graphs can help companies stay on top of regulatory requirements and ensure that supply chain operations are conducted in compliance with relevant guidelines.
- **Better risk management:** Knowledge graphs can help companies identify potential risks in the supply chain, such as by identifying suppliers who may be at risk of disruption.
- **Improved collaboration:** Knowledge graphs can help to improve collaboration between different entities in the supply chain, such as suppliers and logistics providers, by providing a common platform to share and access information.
- **Unambiguously** identify the parties, transactions, markets and instruments involved in the activity along with Visual information consumption and investigative interfaces, replicating and enhancing the workflow.

Pharma Supply Chain graph data overview
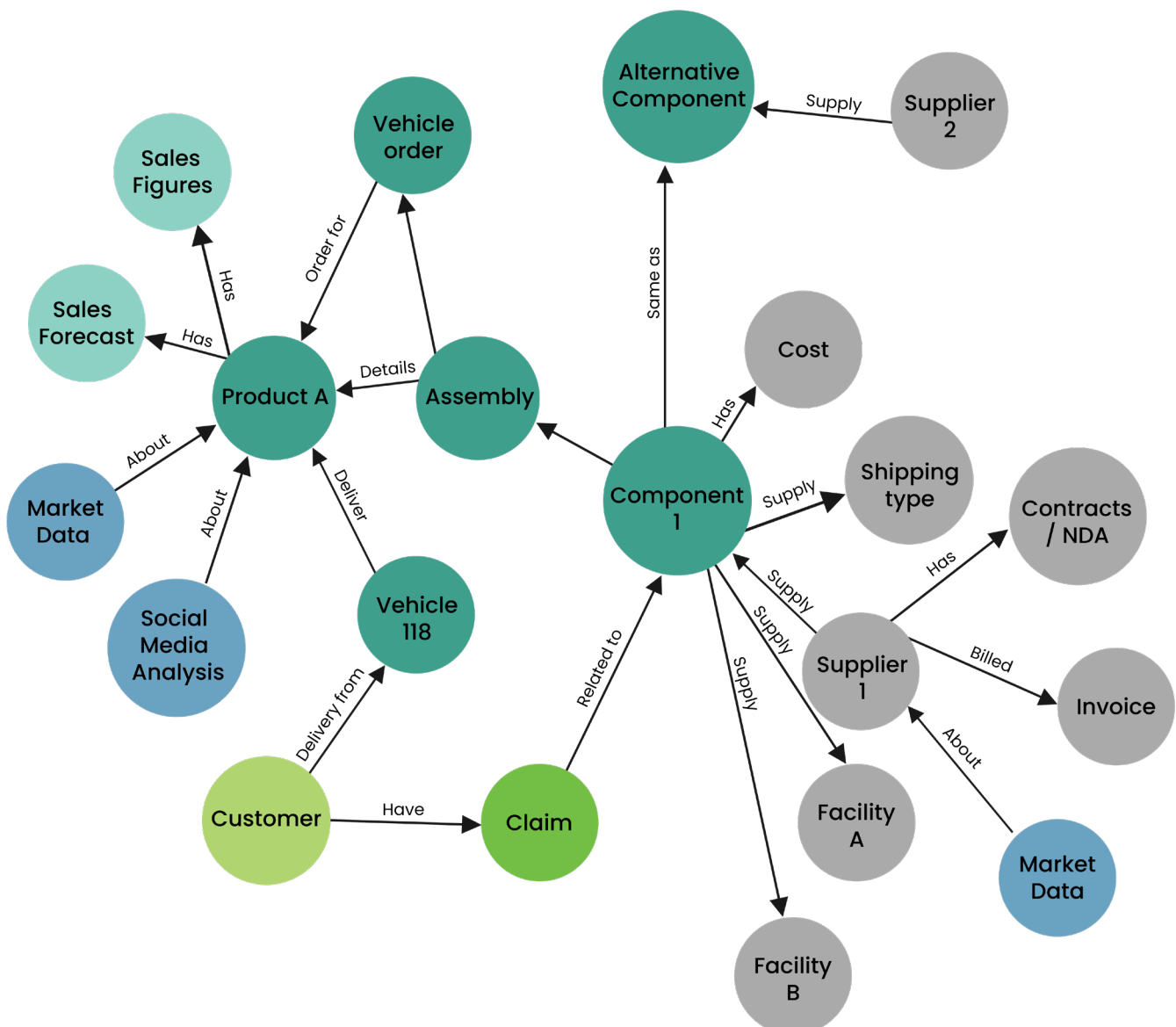
Pharma Supply Chain data captures from multi sources

Operational

- Documentation
- Process
- Facilities
- Organisational Hierarchy
- KPI & reports
- Systems & Databases

Product

- Documentation
- Process
- Product details
- Product Hierarchy
- Bill of materials

Customer

- Individual
- Corporate
- Customer Relations

Events

- Sensors
- Warranty Claims
- Customer Contact

3rd party

- Dealers
- Social media
- Market Data

Supplier

- Supplier/ Vendor
- Logistics
- Inventory

Use case 5:

Adverse Event Reporting / Advanced Drug Discovery Safety Analysis

Adverse event reporting in the pharmaceutical industry refers to the process of collecting and reporting information about any negative or unintended effects of drugs on patients. In adverse event reporting, knowledge graphs can be used to represent the complex relationships between drugs, patients, and any adverse events that occur. Graph databases are used to store and analyze large sets of data on adverse event reporting, such as patient demographics, treatment details, and outcome data. This can help companies identify potential safety risks associated with drugs and make decisions about how to address them.

Graph databases can also be used to perform advanced analytics such as network analysis, which can help in identifying key influencers or important nodes in the graph, like specific drugs or patient subgroups that are more susceptible to certain adverse events. Using graph database technology for advanced drug safety analysis can provide a more powerful and effective way to analyze complex data sets, whereas decision enablement system or platforms like RDF help every type of user in organization to look at probability in each case as RDF are the only solutions for reasoning across

industries and the good part is RDF never save data, platform like ECCENCA works on "in cache system" instead of in-memory giving more power on performance & other parameters.

In the other hand, RDF can be used as a single unified platform to represent the complex relationships between drugs, patients, and adverse events, by connecting the drug resource to the resources representing the patients who have taken it and the adverse events that have been reported. RDF are easy to integrate data from various sources, such as electronic medical records, clinical trials data, and external databases. This can help in gaining a more complete understanding of the safety profile of drugs and identifying potential issues RDF can also be used to perform advanced analytics such as querying and reasoning, which can help in identifying key influencers or important nodes in the graph, like specific drugs or patient subgroups that are more susceptible to certain adverse events.

Business benefits:

- **Improved understanding of drug interactions:** Knowledge graphs can be used to represent the complex relationships between drugs and their interactions, allowing companies to identify potential interactions and side-effects.

- **Enhanced decision making:** Knowledge graphs can be used to perform advanced analytics such as network analysis, which can help in identifying key influencers or important nodes in the graph, like specific drugs or patient subgroups that are more susceptible to certain adverse events. This can improve the decision-making process and reduce the risk of safety issues.

- **Better data integration:** Knowledge graphs can be used to integrate data from various sources such as electronic medical records, clinical trials data, and external databases. This can help companies gain a more complete understanding of the safety profile of drugs and identify potential issues that may have been previously overlooked.

- **Compliance:** Adverse event reporting is a regulatory requirement, and knowledge graphs can help companies in meeting those requirements by providing a more efficient way to collect, store and analyze data.

- **Single Unified platform:** Connection to presentation to decision, yes connected data from internal & external data & integrate within the system, without storing data in platform. Real time updates & visualization on your data. These platforms are enablers for AI & ML which include missing data. Values & parameters for decision.
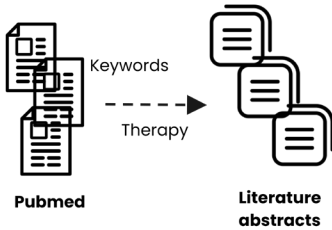
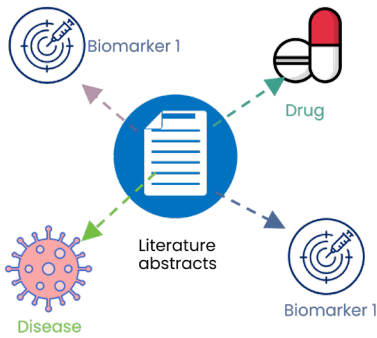## Predictions using KG of unknown adverse drug reactions
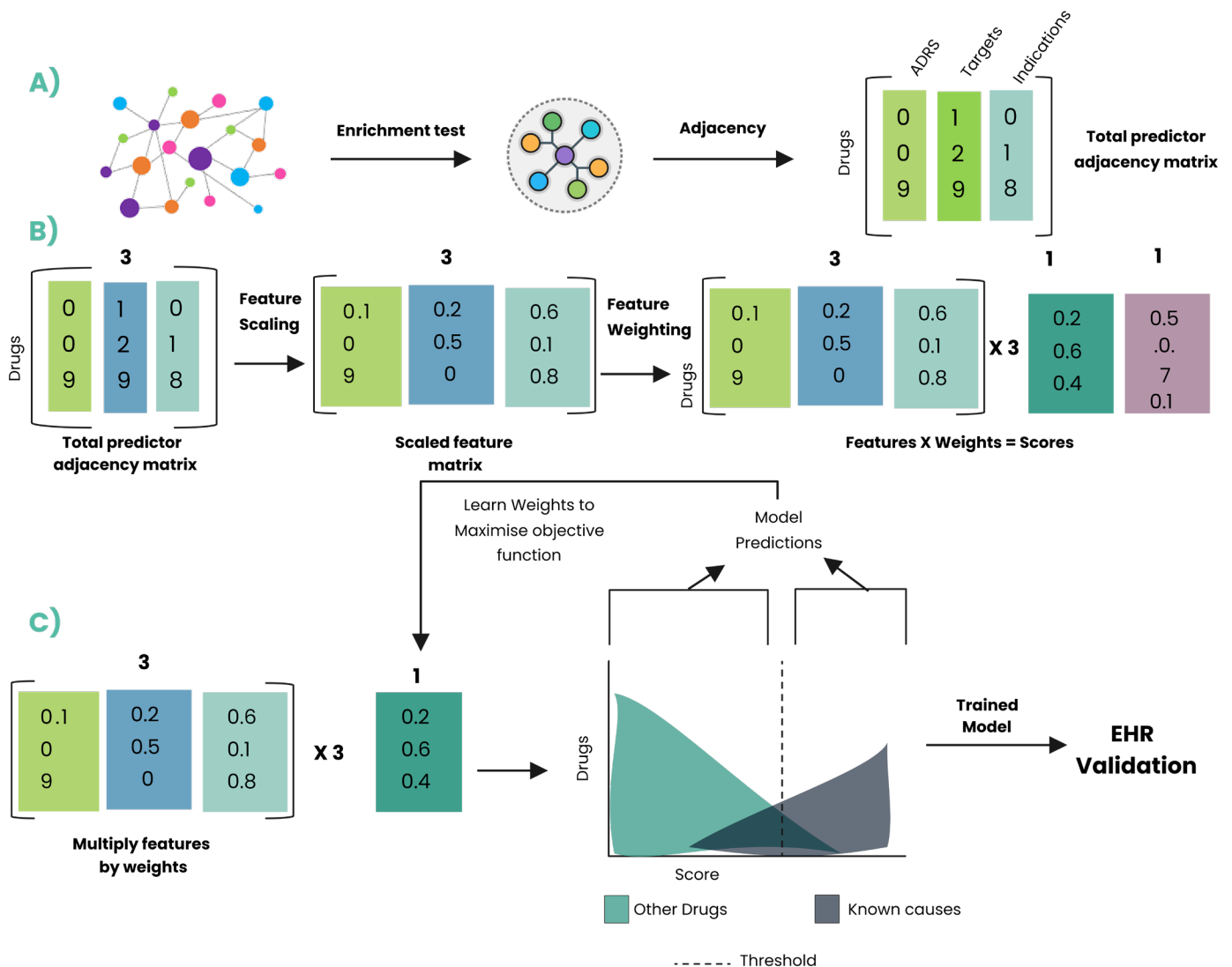
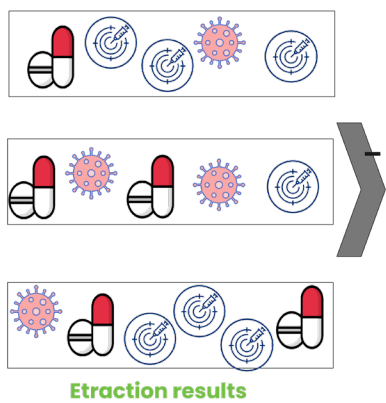# Process Representation of Adverse Drug Reaction using Knowledge Graphs

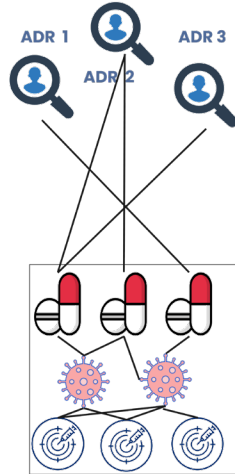## Graph Building

(i) Literature summary acquisition

Pubmed → Keywords → Therapy → Literature abstracts

(ii) Entity Extraction

Biomarker 1, Drug, Literature abstracts, Disease, Biomarker 1

(iii) Graph building
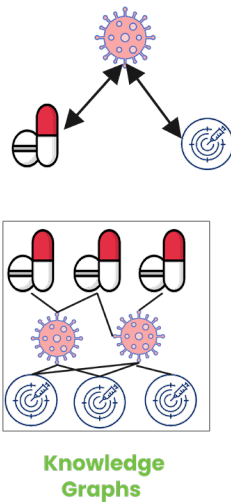
Etraction results

ADR Discovery

(i) Discovery of adverse drug reactions

ADR 1 ADR 2 ADR 3

(ii) Explanation of path-biomarkers which link the drugs with ADRs were found in KG

Mode

Knowledge Graphs

Graph Verification

(i) Three cross Validation

Data

Train | Test

Test

Test

(i) Model Validation

Co-occurrence method

Our KG based Model
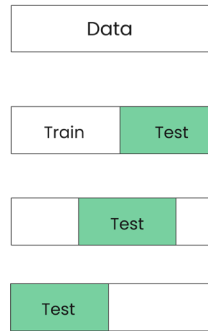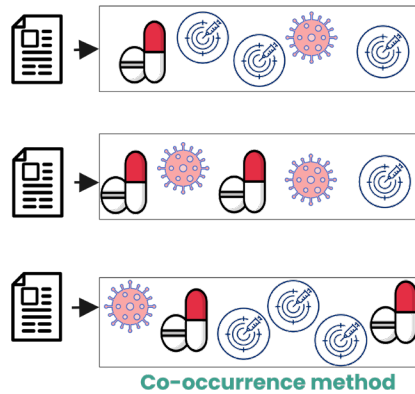
Clinical Validation

(i) Compared with the reported ADRs from anti-disease drug's manual

(i) Compared with reported ADRs of clinical cases

Some Complex Questions Solved & Answered Rapidly

1 — Show me all the studies where subjects/groups were treated first with dosed compound 1 and subsequently by compound.

2 — What have we done on osteoarthritis in the last 10 years? At which sites?

3 — Were there patients in clinical trials with a profile like a field safety report?

4 — Which set of compounds are creating a similar effect or which compound has been tested in similar conditions and similar treatments?

5 — Could some gene expression be used as a biomarker to understand whether some drug is delivering some effect?

6 — Does my clinical trial comply with regulatory and industry standards?

7 — Was my standards-compliant clinical trial set up simply and quickly?

8 — Were trial metadata changes easy to make, govern and manage?

9 — Can I quickly answer change control and impact analysis questions?

10 — Did I eliminate or reduce data management issues during data collection, analysis and reporting?

Current Day Data Challenges

1

<u>Data is not uniform:</u>

This data is highly varied — IoT data, social media posts, doctor's notes, clinical trial results, and more. This data is ultimately a mix of structured, semi-structured, and unstructured data, and most data integration platforms are not designed to accommodate data from such a diversity of sources.

2

<u>Complex Relationships:</u>

The interplay between genes, targets, adverse effects, diseases, symptoms, trials, sites, and more is incredibly complex. Not to mention added complexity once a drug goes into production. A molecule may be renamed in trials, branded upon drug release, and even rebranded later on. How then can you tie a complaint about a certain drug back to the manufacturing site of a particular drug ingredient? A holistic data solution for pharma must be able to address this complexity with precision.

3

<u>Inconsistent Naming:</u>

Any given gene will have different ID numbers in different labs. Multiply this by all the entities required for R&D — genes, expressions, adverse effects, and more. This becomes a monumental problem at the scale of global pharma, who may be evaluating data from dozens or hundreds of labs. It's impossible to enforce terminology consistency across each lab, which operates with some level of independence, not to mention external labs and datasets.

4

<u>Hard Navigation::</u>

Unfortunately, even if you bring all your data together, it doesn't solve the problem. Data lakes alone have not led to better clinical outcomes. Data scientists are still mired in exploring and preparing data for each new analysis.

ROI with Knowledge Graphs

Businesses care about results, and knowledge graphs deliver unmatchable ROIs. Knowledge graphs deliver more than 500% ROI in any specific use cases. Organizations that use these Graph technologies enjoy impressive ROI due to:

Reduce research time 40%- 50% due to integrated search.

Reduce R&D cost up to 10% with reusable models.

Increase drug success rate averagely 11%.

Bring down clinical trails results from Years to Months.

Be 6–9 months early in market taking competitive advantage.

Data infra cost reduced up to 68% on your other data products.

Increase reliability with single platform for any type of Use case, LOB, Analytics.

Greater scalability from TB to PB supporting billions of nodes, relationships & properties.

Increased productivity of 70% with data team & 60% with business team.

20x faster & better decision making.

Accelerate innovation 10x faster.

USA

1340 S De Anza Blvd
San Jose
CA – 95129
+1 510 509 9368

INDIA

Level 4, Vasavi MPM Grand
Ameerpet, Hyderabad
Telangana – 500073
+91 770 223 3685

www.aarth.io