ontotext

Making Sense of Text and Data

AI-Powered Target Discovery

Identify New Drug Targets And Promising Drug Repurposing Candidates Quickly, Easily and with high Confidence

Todor Primov

SWAT4LSHC – Basel (CH)
Feb, 2023

Presentation outline

- Introduction
- The challenges
- Use Cases
- Solution
- Case Study
- Key components

ontotext

About Ontotext

23 YEARS

100% SEM. TECH. FOCUS

- Semantic graph database engine combined with Content and Data analytics solutions
- Interlinking text and data to unveil meaning and hidden insights
- Delivering unmatched search and analytics

100+ EMPLOYEES

Sofia R&D

- Global team, local sales & accounts support
- R&D Center in Sofia, Bulgaria
- Serving Elsevier, EBSCO, Roche, IET, Lurie Children's Hospital, SIB, Healthdirect, etc
- Deployment on AWS, Azure, GCP, on-prem...

Offices: New York Sofia London Basel

ontotext

What we do?

Helping Pharma and Biotechs find the next therapeutic target in the sea of disconnected data

We help enterprises get **profound insights** by linking:

- Diverse databases & Unstructured content
- Proprietary & Global data

ontotext

The Status Quo

Only 1 in 5, 000 compounds, which enter drug discovery actually becomes an approved drug[1]

In the process of finding the most successful candidate, companies often

- Fail to quickly build and validate on-target biological hypotheses
- Struggle to derive key insights, e.g. safety and druggability, from structured and unstructured data
- Lack enough confidence in their decision - which are the top targets to be validated
- Fail to adapt in the dynamic drug R&D environment

1. Ezekiel J. Emanuel. "The Solution to Drug Prices". *New York Times*.

Target Discovery Solution – Use Cases

- Utilize large amounts of information from **various internal and external biomedical sources** to speed up the process of discovering and repurposing drugs.
- Discover **previously unknown connections** between targets and diseases, quickly generate and assess new ideas to guide drug research and development.
- Prioritize targets based on the latest research data by using **custom evaluation criteria and algorithms**, and get value from both structured and text data.

ontotext

Target Discovery Solution – Value Drivers

Provides deeper insights
over a highly interlinked knowledge network, in which long sequences of relations can be mined

Facilitates information discovery
as it bridges the siloed data across multiple sources

Drives innovation
with an evolving data model that adapts to the iterative development of use cases

Low cost
of adding new data and maintenance of data updates

Data quality
confidence score and provenance are used as discovery metrics

ontotext

Who is it for?

Researchers /

Translational medicine experts

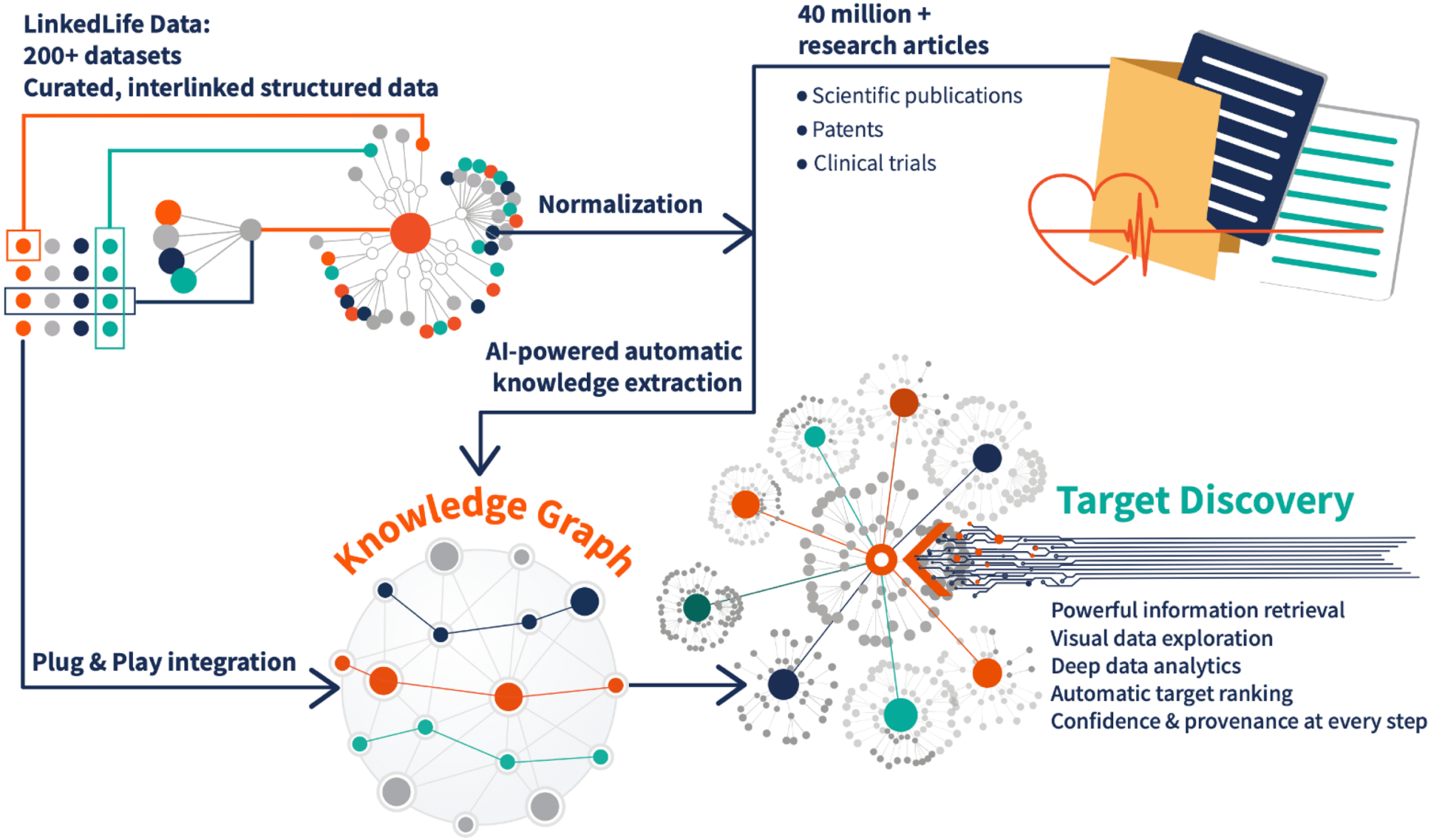
Help develop a consistent strategy for early drug R&D which is innovative, safe and foundational to decision-making

Data scientists /

Computational biologists

Help deliver quality insights, which translate all heterogeneous (internal and external) data into new therapeutic targets

Our Approach



ontotext

Customer Case Study - Key Metrics

Before KG-powered target discovery	**With** KG-powered target discovery	**With NLP**
0 mappings separate redundant, non-curated databases Data dispersed in multiple databases Manually gather, align, process and analyze	**5000** Million facts ~ 35 curated datasets ~ hundreds of mappings All public datasets are interconnected in a single knowledge graph Hidden relationships can be discovered	**40** Million documents ingested **>500** Million Additional relations extracted New information unlocked, otherwise not available publicly We now deliver more information about any biological entity

Several hours/days → **10X** → a couple of minutes

ontotext

Extensive LinkedLife Data Catalogue

200+ curated, mapped, ready-to-use biological, medical and scientific datasets

Genes

NCBI

EMBL

Proteins

UniProt

Drugs & Compounds

chEMBL

DrugCentral 2021
Online drug Compendium- Database Update Oct 2021

Pathways

Pathway Commons

BioGRID

STRING

Clinical trials

ClinicalTrials.gov

Literature

medRxiv

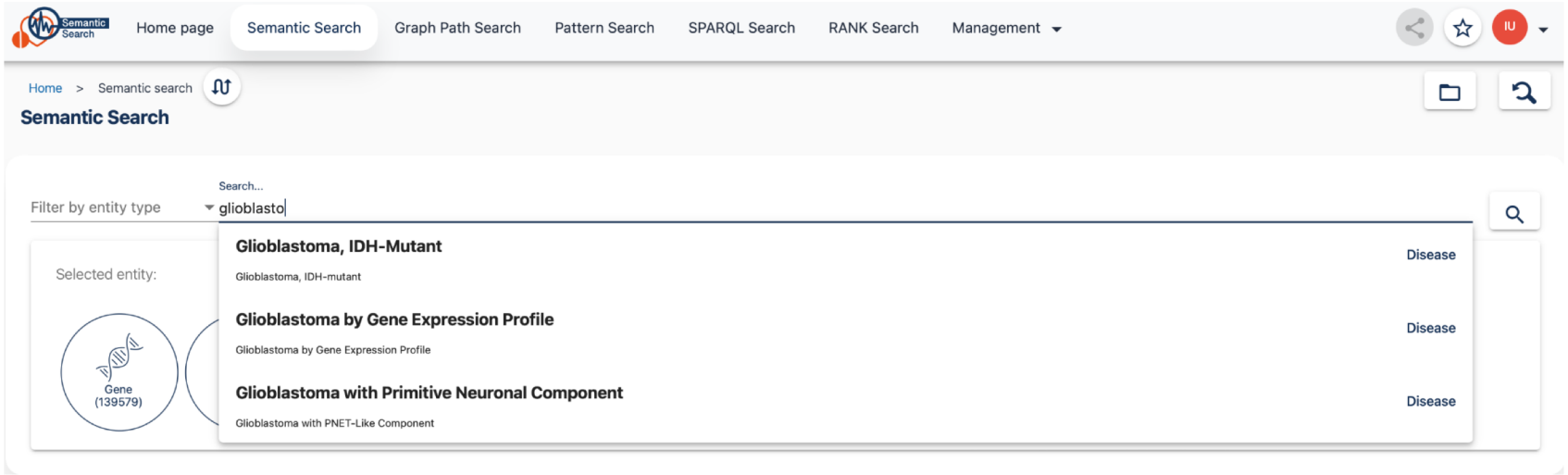
PubMed.gov

Patents

ontotext

Semantic Search Module

- Easily explore related entities in the same system
- Customize search parameters and visual representations
- Increased data quality and traceability
- Powerful result filtering
- Discover predicted relationships



ontotext

Scientific Literature Mining

Published: PLoS One ;10(8):e0134398

Author: Bemanian V, Sauer T, Touma J, Lindstedt BA, Chen Y, Ødegård HP, Vetvik KM, Bukholm IR, Geisler J

The epidermal growth factor receptor (EGFR / HER-1) gatekeeper mutation T790M is present in European patients with early breast cancer.

The epidermal growth factor receptor (EGFR) is one of the major oncogenes identified in a variety of human malignancies including breast cancer (BC). EGFR-mutations have been studied in lung cancer for some years and are established as important markers in guiding therapy with tyrosine kinase inhibitors (TKIs). In contrast, EGFR-mutations have been reported to be rare if not absent in human BC, although recent evidence has suggested a significant worldwide variation in somatic EGFR-mutations. Therefore, we investigated the presence of EGFR-mutations in 131 norwegian patients diagnosed with early breast cancer using real-time PCR methods. In the present study we identified three patients with an EGFR-T790M-mutation. The PCR-findings were confirmed by direct Sanger sequencing. Two patients had triple-negative BC (TNBC) while the third was classified as luminal-A subtype. The difference in incidence of T790M mutations comparing the TNBC subgroup with the other BC subgroups was statistical significant (P = 0.023). No other EGFR mutations were identified in the entire cohort. Interestingly, none of the patients had received any previous cancer treatment. To our best knowledge, the EGFR-T790M-TKI-resistance mutation has not been previously detected in breast cancer patients. Our findings contrast with the observations made in lung cancer patients where the EGFR-T790M-mutation is classified as a typical "second mutation"causing resistance to TKI-therapy during ongoing anticancer therapy. In conclusion, we have demonstrated for the first time that the EGFR-T790M-mutation occurs in primary human breast cancer patients. In the present study the EGFR-T790M mutation was not accompanied by any simultaneous EGFR-activating mutation.

TOC | Controls | Mentions

Annotation types

- Show/hide all annotations
- Protein 4
- FunctionOrProcess 8
- PopulationGroup 10
- Gene 1
- Disease 44
- Substance 1

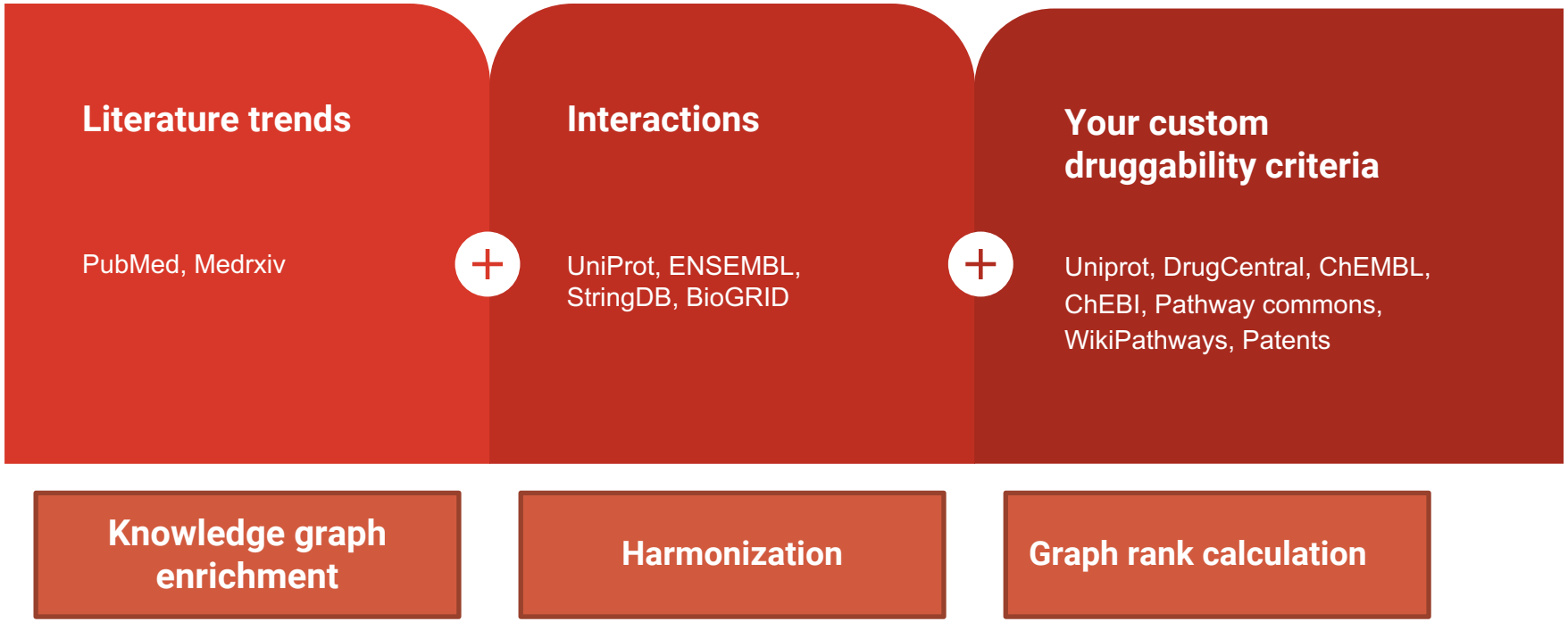
Annotation score

0 — 1

ontotext

Intelligent Target Selection and Ranking

- Streamline identification of potential targets
- Facilitates insights-driven decision making
- Customizable scoring based on predefined criteria



ontotext

Intelligent Target Selection and Ranking

Display fields

Importance, Centrality_top

Selected entity: **Protein**

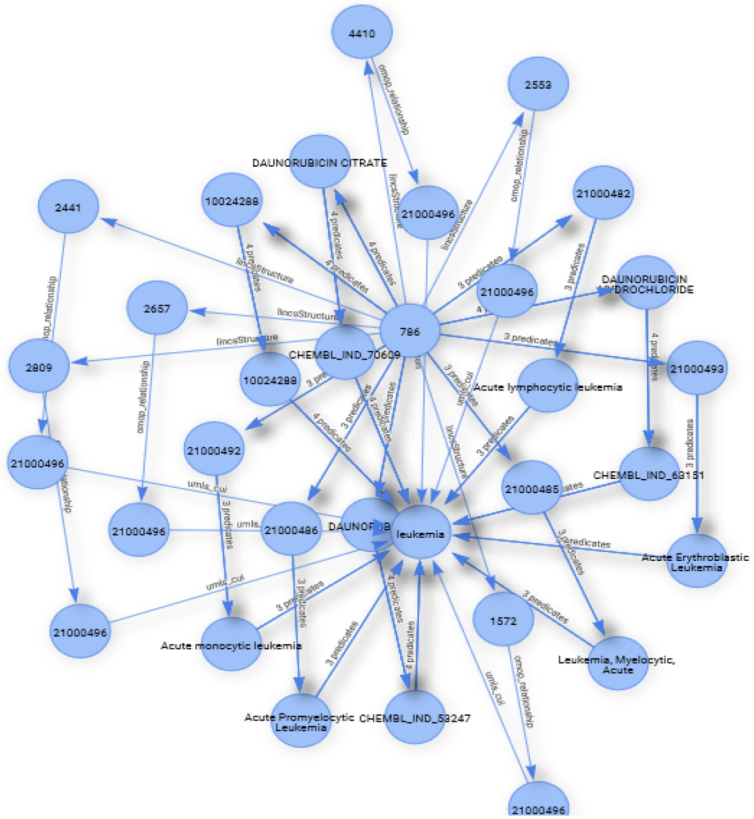
Showing 1 - 10 of 18

Summary	Importance ↑	Centrality_top
Epidermal growth factor receptor ***Preview dashboard*** ***Preview rank calculations***	• Rank: 1 • Calculated score: 1	• Rank: 1 • Calculated score: 1
Cellular tumor antigen p53 ***Preview dashboard*** ***Preview rank calculations***	• Rank: 1 • Calculated score: 1	• Rank: 1 • Calculated score: 1
Albumin ***Preview dashboard*** ***Preview rank calculations***	• Rank: 2 • Calculated score: 0.96874994	• Rank: 2 • Calculated score: 0.96874994

ontotext

Graph Path Search

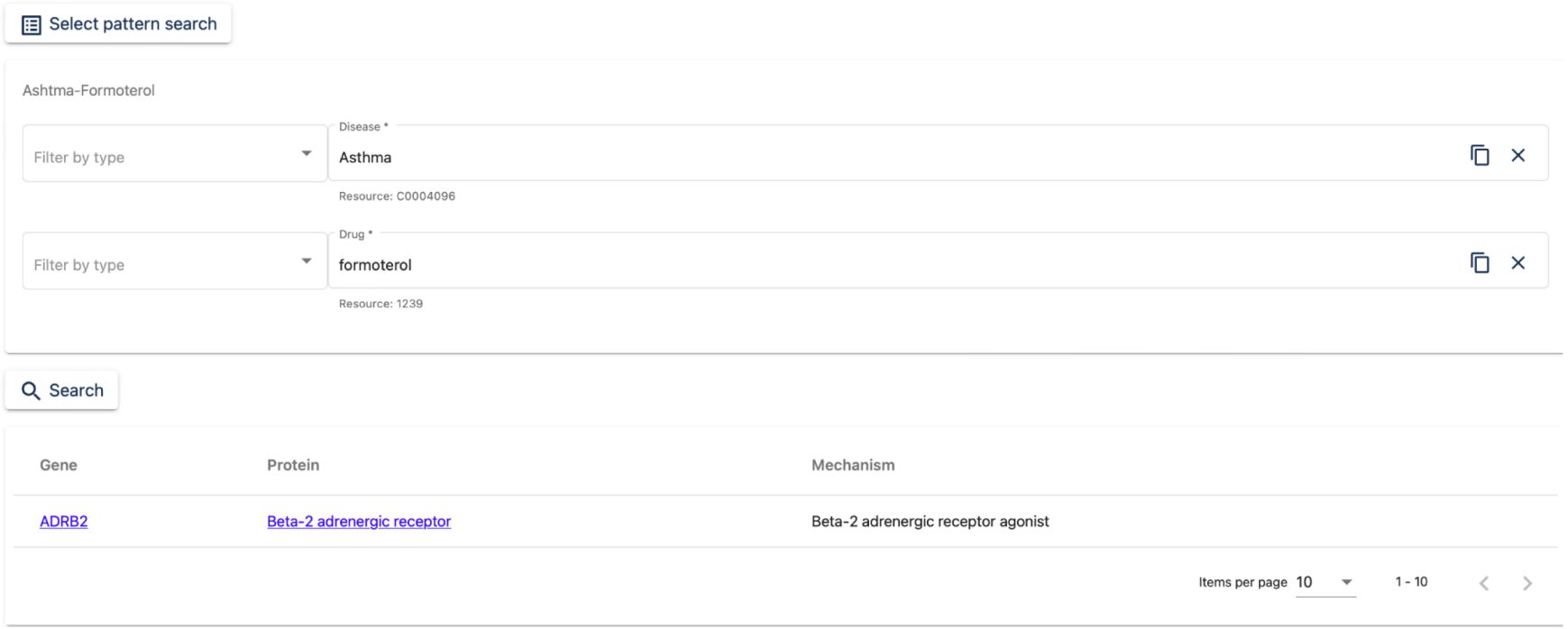
- Uncover hidden relationships in otherwise unavailable data
- Enhance capabilities - facilitate innovation and get deeper insights from diverse data
- Basis for building new hypothesis, shortcut potential R&D pitfalls



ontotext

Pattern Search

- Definition of complex graph patterns that can be easily contextualized by researchers providing base parameters (e.g. gene name, disease, etc)
- End users need to specify the input parameters for each pattern in order to run it and to retrieve the relevant subgraphs



ontotext

Gene Dashboard

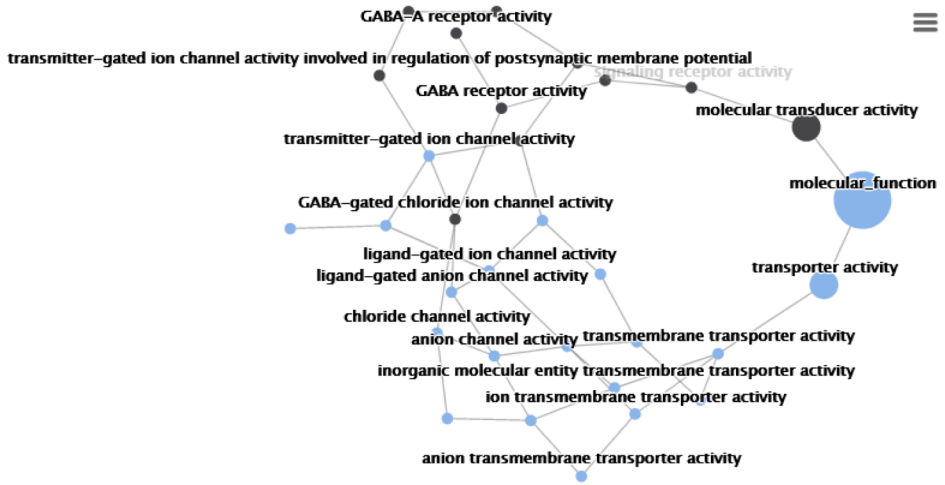
Official Gene Symbol	Gene type	Organism
CTLA4	protein-coding	Human

Encoded proteins

- Cytotoxic T-lymphocyte protein 4

Full name	Gene location
cytotoxic T-lymphocyte associated protein 4	2q33.2

Gene ontology terms



Expression in tissues

vermiform appendix	52
tonsil	40
gall bladder	17
bladder	15
urinary bladder	15
spleen	10
thymus	8
bone marrow	7

Items per page 10 1 - 10

ontotext

Drugs Dashboard

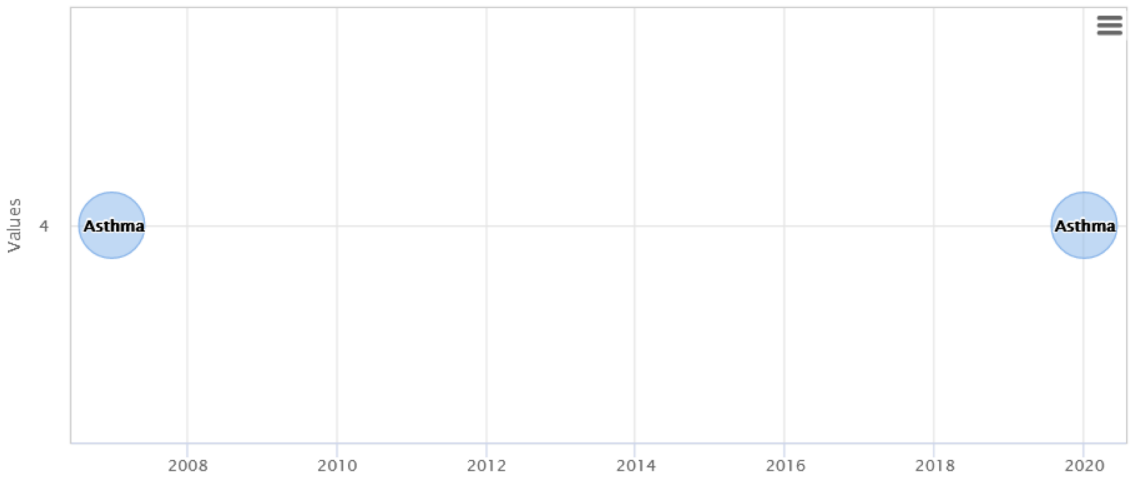
Adverse events (UMLS)

- Product quality issue
- No adverse event
- Product dose omission
- Asthma
- Dyspnea
- Product container issue
- Coughing
- Wheezing

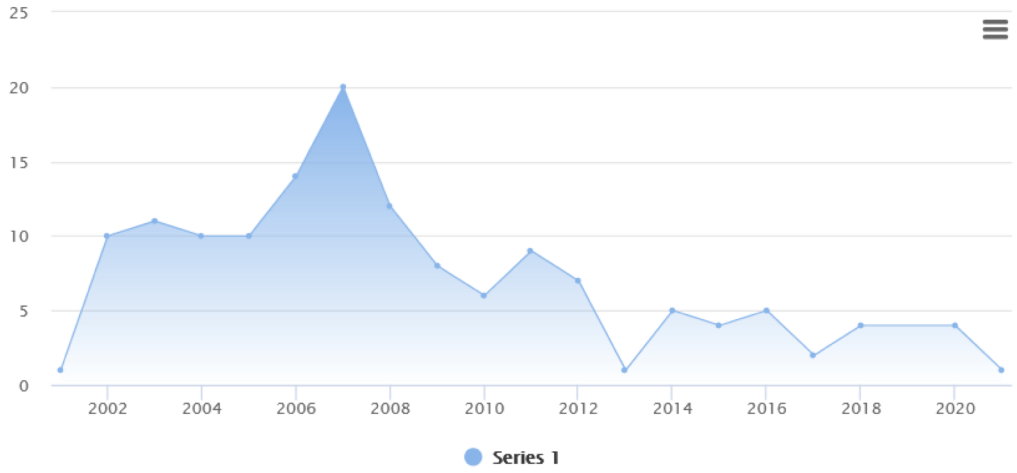
Indications (UMLS)

- Alzheimer's Disease
- Presenile dementia
- Familial Alzheimer Disease (FAD)
- Alzheimer Disease, Late Onset
- Acute Confusional Senile Dementia
- Alzheimer's Disease, Focal Onset
- Alzheimer Disease, Early Onset
- Chronic Obstructive Airway Disease

Active clinical studies with indication

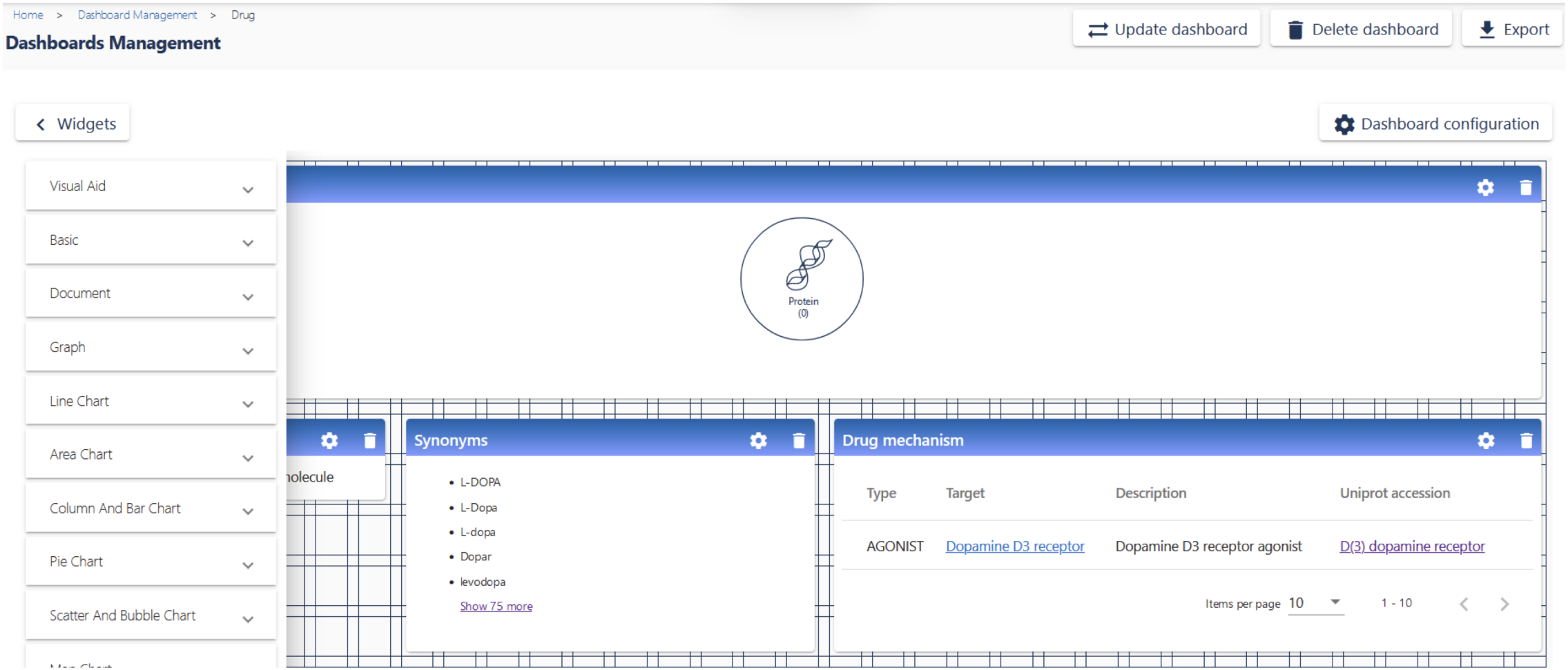


Clinical studies per year



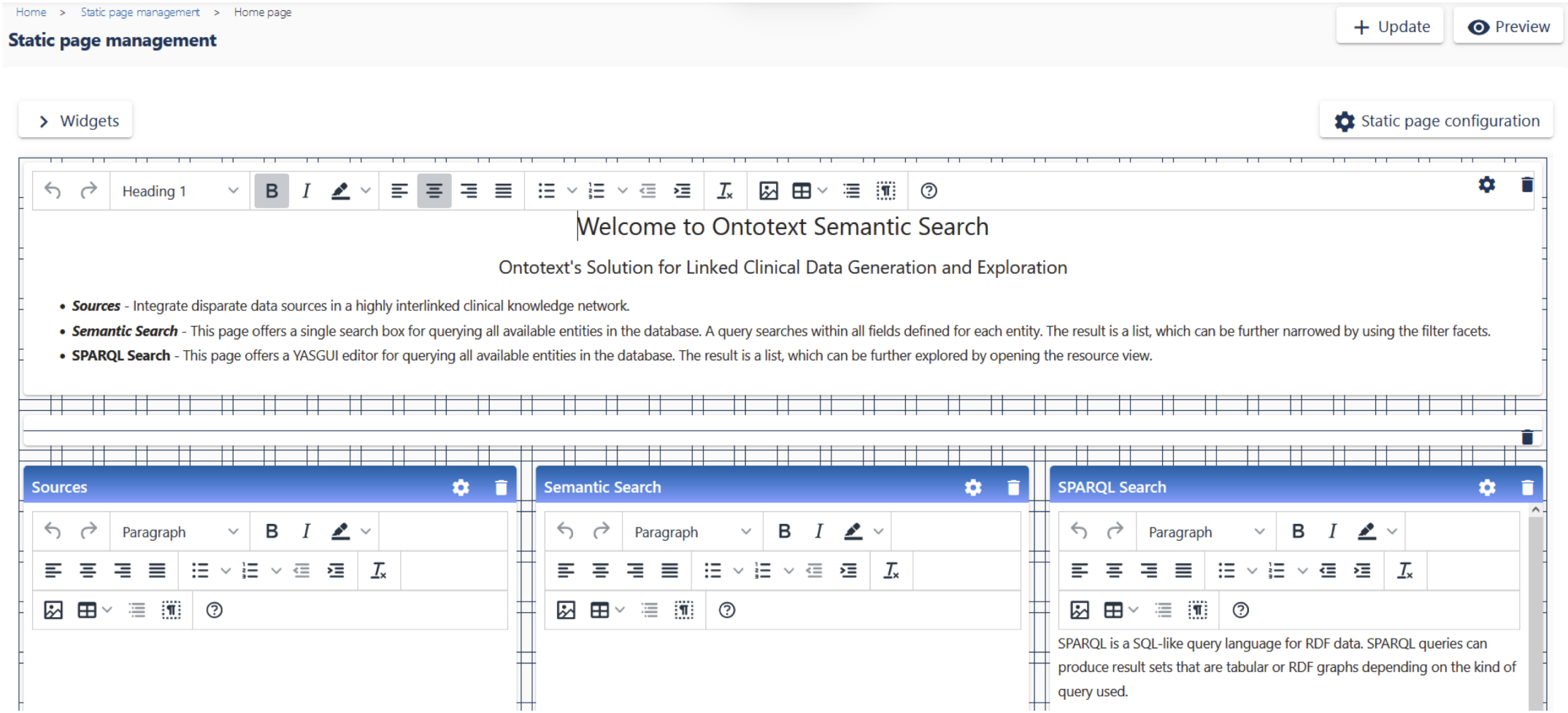
ontotext

Easy visual configuration



ontotext

Easy, Word-like app page configuration



ontotext

Key Differentiators

- Integrate more data sources than competitors leveraging Ontotext LinkedLifeData inventory (200+ relevant datasets) including regular updates. Seamlessly integrate customer's proprietary data in the KG
- Additional insights generated on top of several key resources like Pubmed, CT.gov, Google patents, etc. either using our existing NLP pipelines, or building custom ones addressing customer specific needs
- Graph algorithms and deeper analytics on top of highly interconnected KG. Fully customizable search, dashboards and ranking (customer proprietary ranking methodology)
- Deliver custom-tailored Target Discovery solution (with existing NLP pipelines) in a matter of 6 to 9 weeks

Takeaways

- Selecting the next successful target candidate can be empowered by leveraging public datasets and automatically extracted data from scientific literature
- Researchers and data scientists can both leverage cutting edge technologies, such as KGs and NLP, in a user-friendly and customizable way
- Critical decision-making processes in target selection can be improved by automating target ranking and mitigating risks by leveraging provenance and evidence metrics

Thank you!

Sign up for a free workshop with our SMEs

Identify new drug targets or promising drug repurposing candidates quickly and easily

Target Discovery