

Using Natural Language Processing to transform real world data

- Introduction
- Linguamatics & NLP Overview
- Novo Nordisk and Team Overview
- NLP Projects
- Integrated Advanced Analytics
- Rybelsus® Dashboard
- Technical Landscape: i2e and Amazon Web Services (AWS)

About Linguamatics



Software

Consulting

Enterprise & Cloud

- Agile, scalable, real-time Natural Language Processing (NLP)-based text analytics
- Rapid relationship extraction and knowledge synthesis using unique and powerful blend of methods
- Commercial out-of-the-box, deep domain expertise

Pharma/Biotech

Healthcare

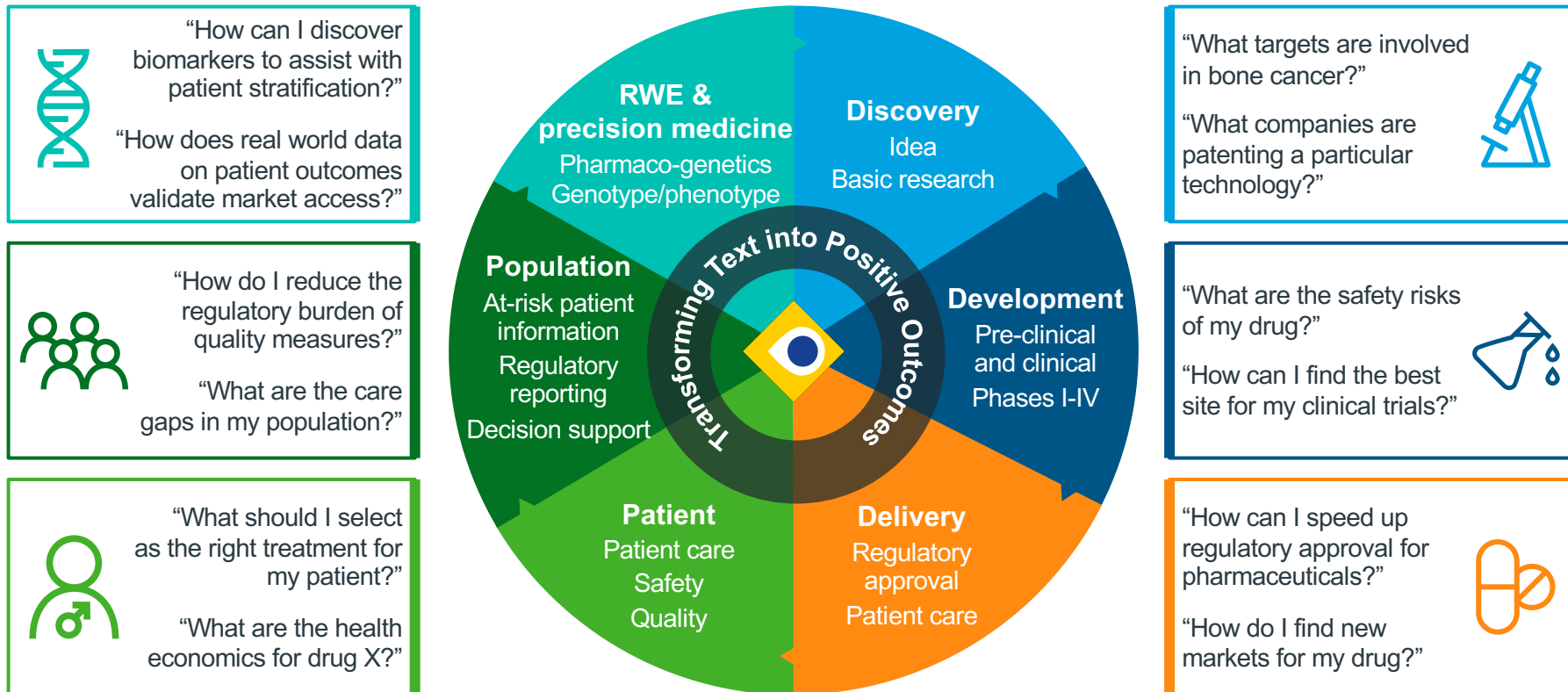
Government

Including 18 of the top 20

Including major cancer centres,
health systems, academic medical
centres

Including FDA, NCI

Molecule to Market with Linguamatics NLP

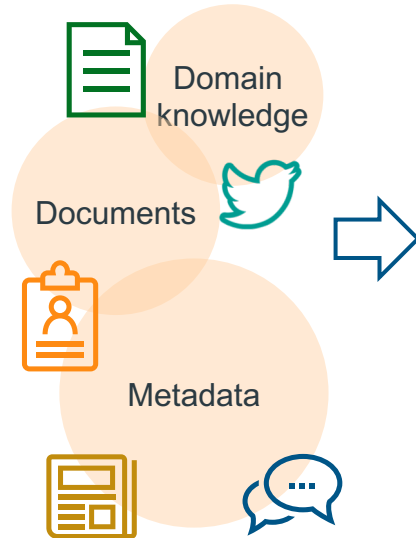


From document to data centric decision support

Turn text ...

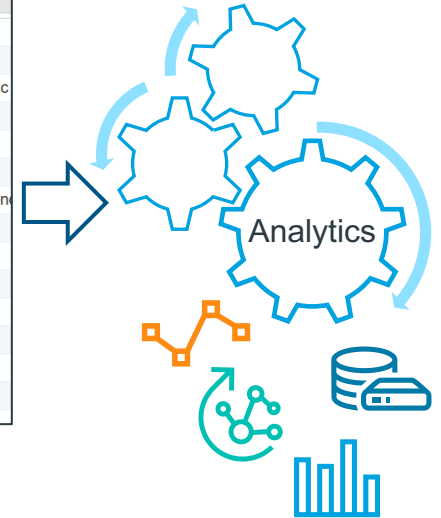
... into **structured data** using **powerful queries** ...

... to drive **analytics** and **outcomes**

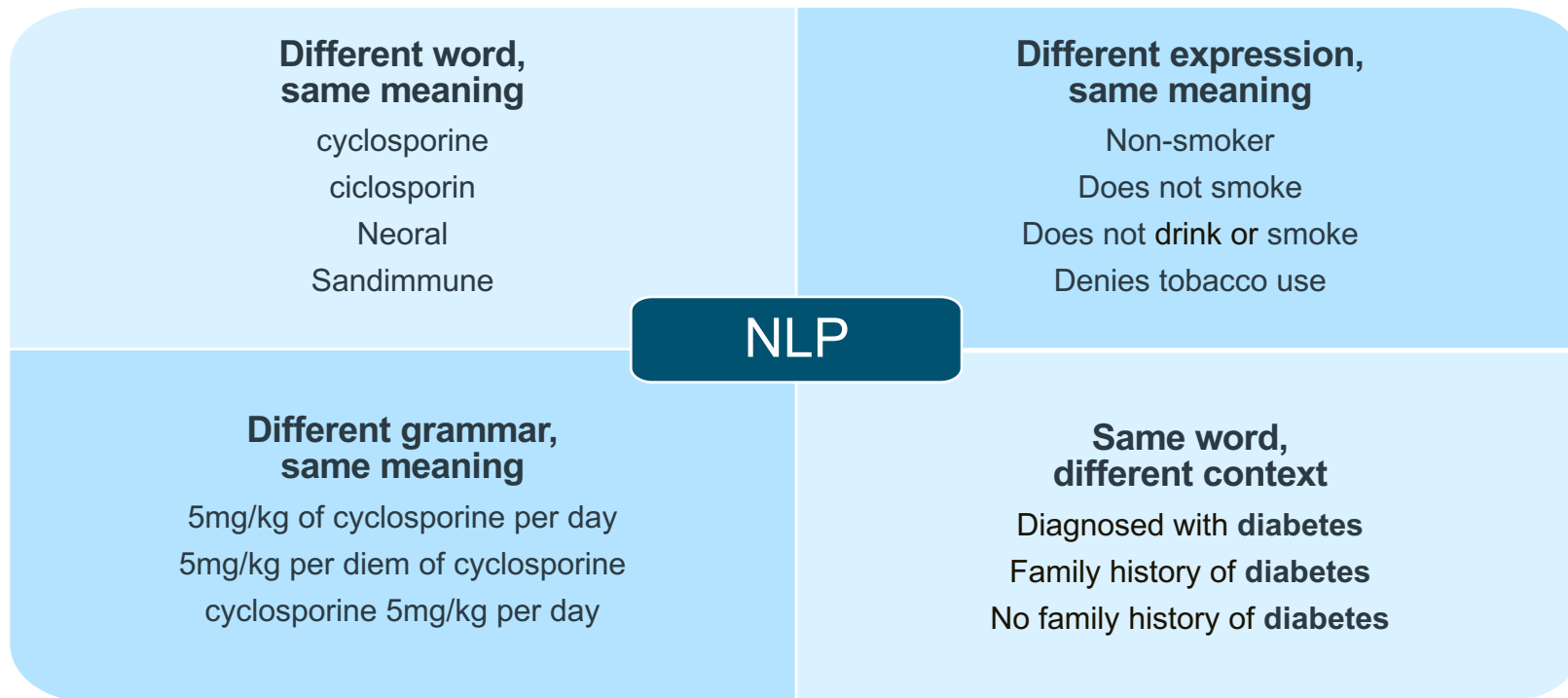


Ejection Fraction	Type	#Docs	Doc	#Hits	Hit
60 %	>	17	M912	2	... motion and wall thickening, EF of 60 %.
40 %	>	5	M2332	2	Left ventricular ejection fraction is 40 % by left ventriculogram.
55 %	>	4	M714	1	... tamponade, normal left ventricular ejection fraction at 55 % with mild concentric
30 %	>	2	M1519	1	Heart failure, ejection fraction of 30 %.
50 %	>	4	M150	1	Her ejection fraction is 50 % and she has paradoxical septal ...
70 %	>	2	M65	1	... was somewhat hyperdynamic with an ejection fraction of 70 %.
[20, 25] %	>	4	M1727	2	... , congestive heart failure with EF of 20 % - 25 %, hypertension, renal insufficien
65 %	>	2	M1836	2	The LV systolic ejection fraction is normal at 65 %.
[45, 50] %	> Estimated	2	M2340	2	... with mild global hypokinesis and estimated ejection fraction of 45 to 50 %.
69 %	> Calculated	2	M281	1	Ejection fraction was calculated to be 69 %.
(65, inf] %	> Estimated	2	M2400	1	Estimated left ventricular ejection fraction is greater than 65 %.
[40, 45] %	>	2	M347	1	... to be mildly reduced with ejection fraction between 40 % and 45 %.
[50, 55] %	>	2	M1101	1	Ejection fraction was 50 % to 55 %.
35 %	>	1	M2543	2	... a 2-D echocardiogram revealed an ejection fraction of 35 %.
58 %	>	1	M1102	2	... revealed normal wall motion, ejection fraction of 58 %.

Natural Language Processing – Ontologies – Statistical
 Methods – Machine Learning – Chemistry – Regular
 Expressions – etc.

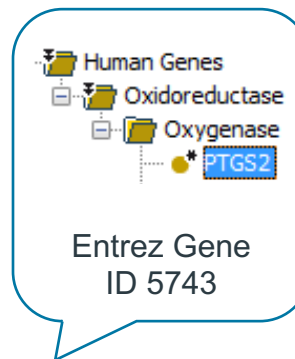
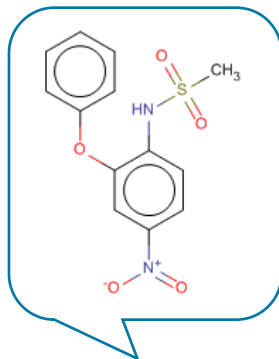


Natural Language Processing finds information however it is expressed



Linguistic Processing: From Words to Meaning

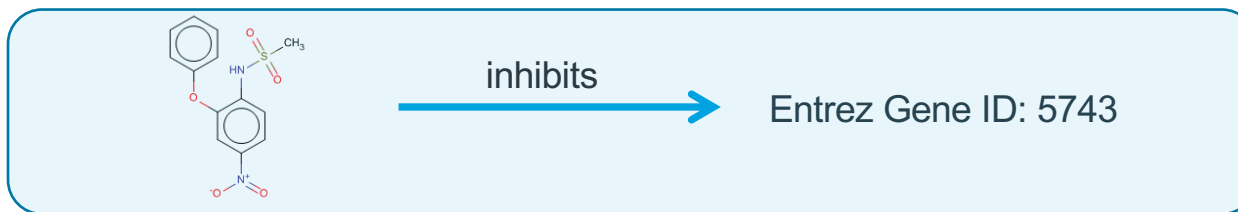
Identifying
entities and
relations



“Among them, nimesulide, a selective COX2 inhibitor, ...”

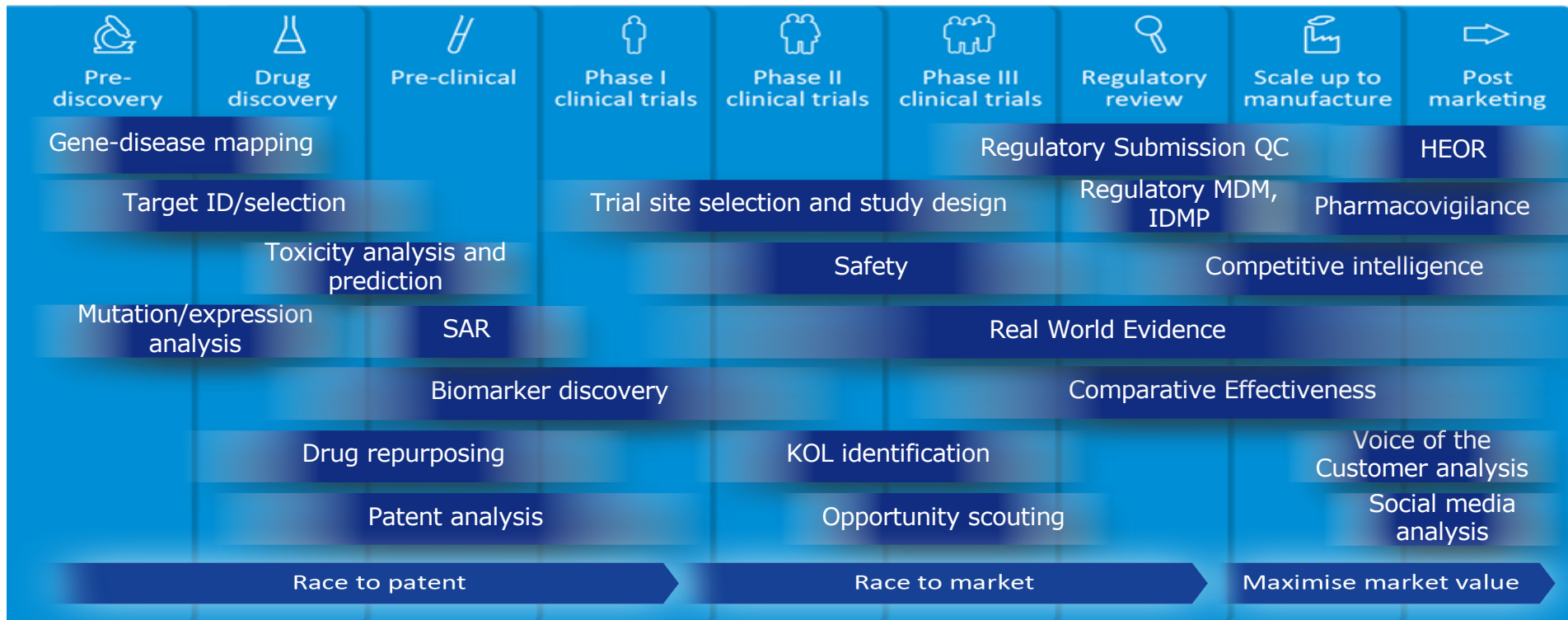
Linguistics to establish relationships

inhibits



Advanced text analytics delivers value

From discovery to market, NLP transforms text for decision support



Novo Nordisk in the U.S.

For 95 years, Novo Nordisk, a global leader in diabetes care, has been committed to discovering and developing innovative medicines to help people with chronic diseases lead longer, healthier lives.



ESTABLISHED IN 1923
**COPENHAGEN,
DENMARK**



U.S. HEADQUARTERS IN
PLAINSBORO, NJ WITH
LOCATIONS
IN 5 STATES



 EMPLOYS APPROXIMATELY
6,000 PEOPLE

**\$8.5
BILLION**
TOTAL U.S. SALES
IN 2017

**WORKING WITHIN
DIABETES, HEMOPHILIA,
GROWTH DISORDERS, AND OBESITY**



APPROXIMATELY
250,000
SHAREHOLDERS



17 MEDICINES
MARKETED
ACROSS
THERAPEUTIC
AREAS

**EMPLOYS NEARLY
582 PEOPLE** WORKING
IN R&D IN THE U.S.

Driving change to improve patient lives

Clinical Development, Medical & Regulatory Affairs (CMR)

Aspires to be the best at understanding and meeting the needs of our patients

MARY EDWARDS
USA
Mary's BMI is 44

Analytics Team Overview

Information Analytics turns information into action by helping stakeholders extract deep insights from their data

Text Mining & Deep Learning

Tools and analyses for discovery, interpretation, and communication of meaningful patterns in unstructured data – including NLP, predictive analytics, machine learning, etc.

Automation

Leveraging cutting edge tools and technologies to provide automated and scalable insights



Data Visualization

Visual data discovery for quick access to relevant business insights

Insights & Analysis

Partner with stakeholders to interpret and communicate trends and insights

NLP Projects

- Publication Gap Analysis
- HEOR Ethnographic insights
- SoMe: Digital Opinion Leaders
- Clinical Trial Protocol Deviations
- Medical Patient Dashboard



Integrated Advanced Analytics

- Aligning with Global Big Data CoE
- Leveraging advanced analytics platform, tools & data lake
- Successfully moved i2e to our OASIS/AWS environment



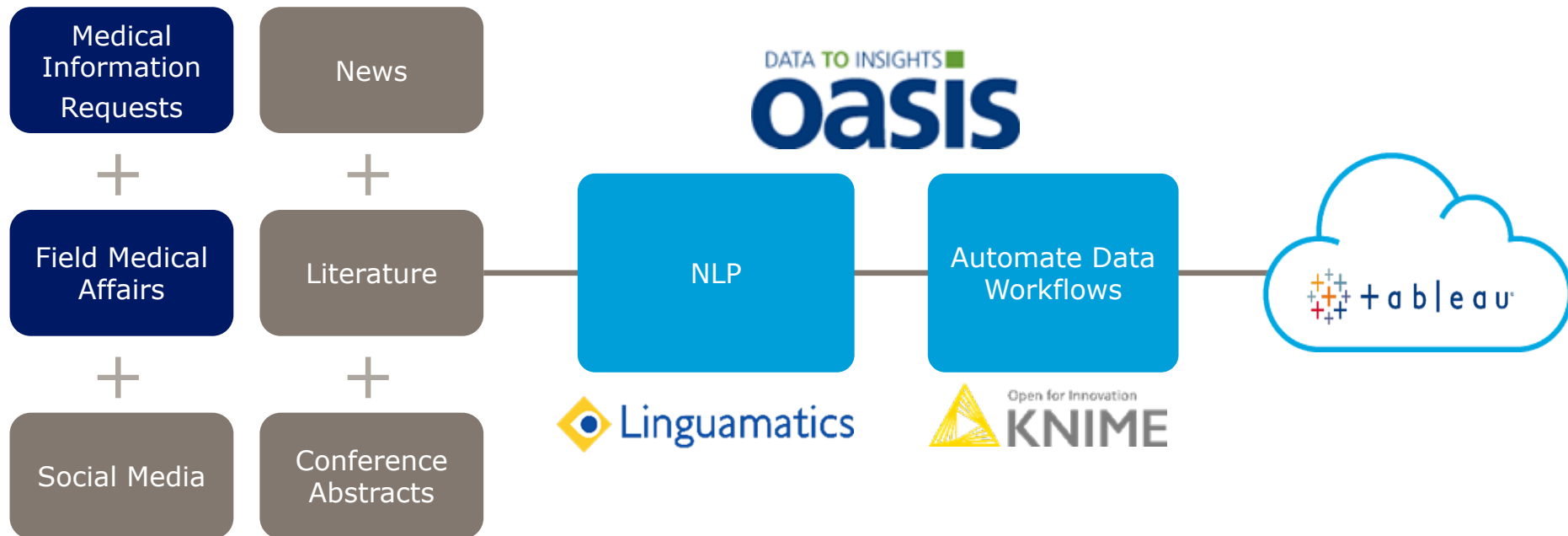
Rybelsus® Insights Dashboard

First Steps

- Multiple Streams of Data:
 - Internal: Databases Capturing Medical Information Requests and ML Interactions
 - Manual Reporting
 - Hard to Search
 - Information Presented Staticly
 - External: News, Social Media, Publications, Conferences
- Insights:
 - What topics are being discussed?
 - Where are conversations happening?
 - What types of Healthcare Professionals are submitting questions?
 - Where are conversations happening nationally?
- Opportunity to Leverage i2e and Tableau

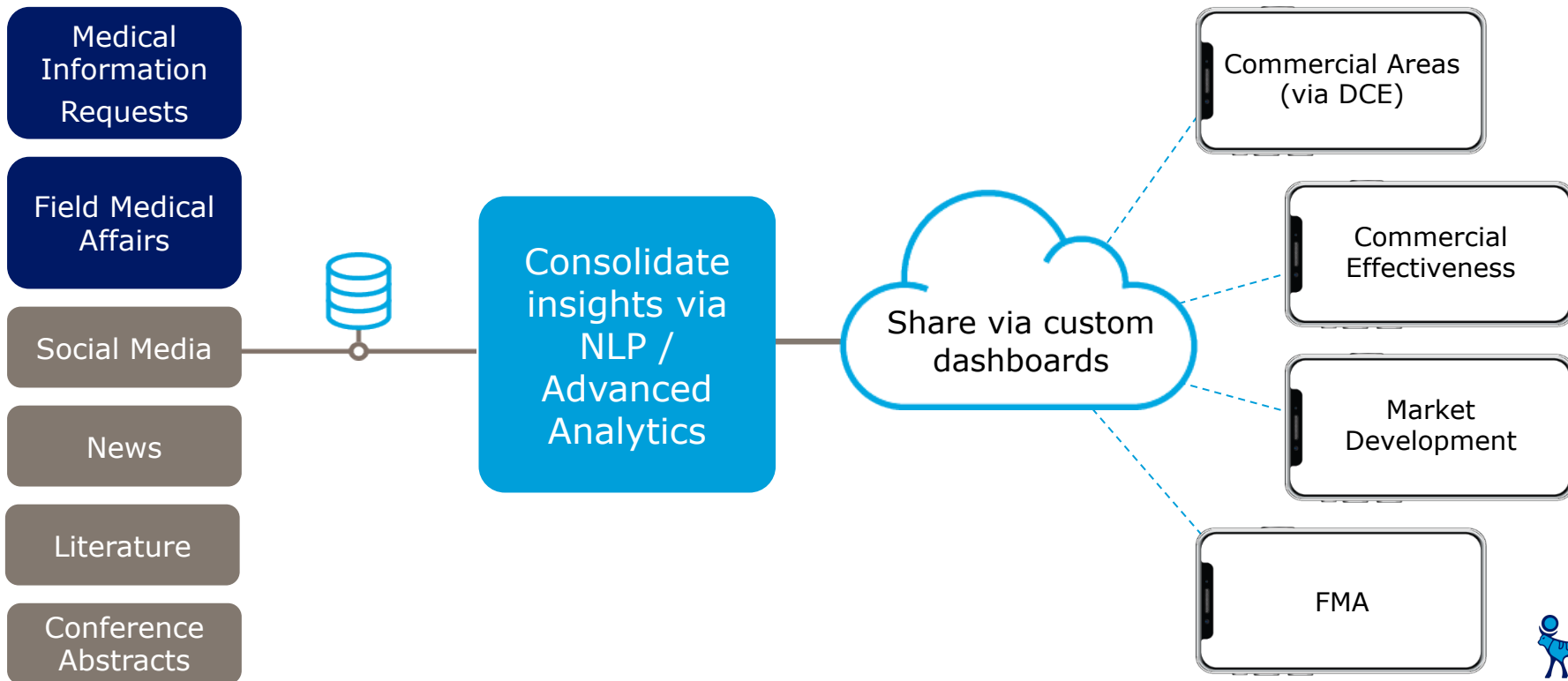
RyBELsus Dashboard

Insight Generation from Six Streams of Data



Medical & Patient Dashboard

Making medical / patient data & insights broadly actionable



Query Development



- Querying topics related to:
 - Safety
 - Efficacy
 - PK/PD
 - Randomized Controlled Trials
 - Patient Populations
 - Dosing
 - Devices
- MeSH, NCI, MedDRA and in-house ontologies used
- Development and Refinement
 - Input from Team Pharmacists

Query Development



Found 2026 assertions from 7435 hits in 1552 docs.

Examined all 1552 docs.

Took 1.74 secs (CPU 1.2).

[\[more details\]](#)

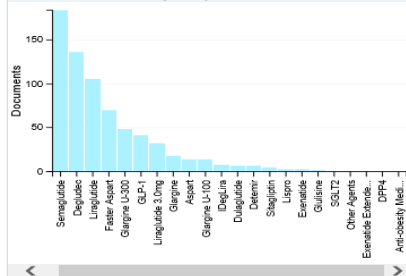
I2E Query: tzh final ePIR query.i2qy

[Clear Selection](#)

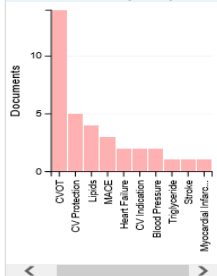
[Filtering Help](#)



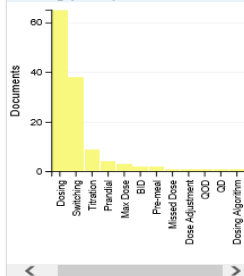
Additional Products (23/23)



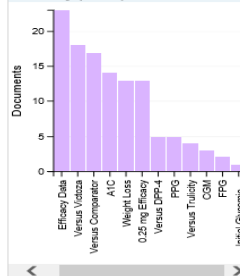
Cardiovascular (11/11)



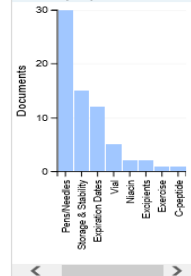
Dosing (13/13)



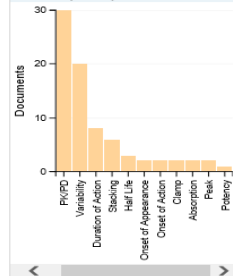
Efficacy (13/13)



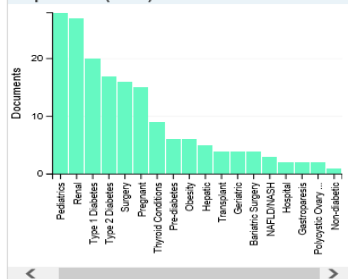
Other (9/9)



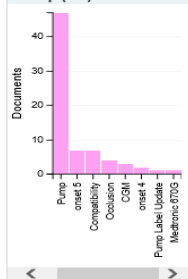
PK/PD (12/12)



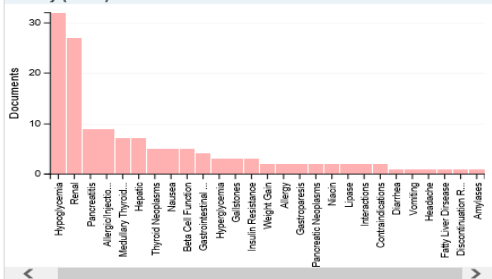
Populations (19/19)



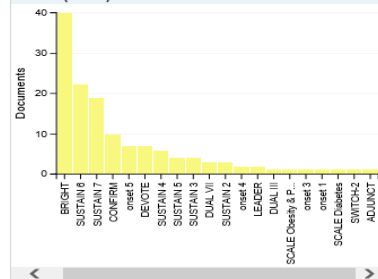
Pump (9/9)



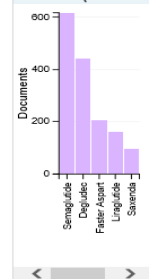
Safety (28/28)



Trials (21/21)



Product (9/9)



Medical Information



Sample:

What HCP
questions are being
asked for
Rybelsus®?

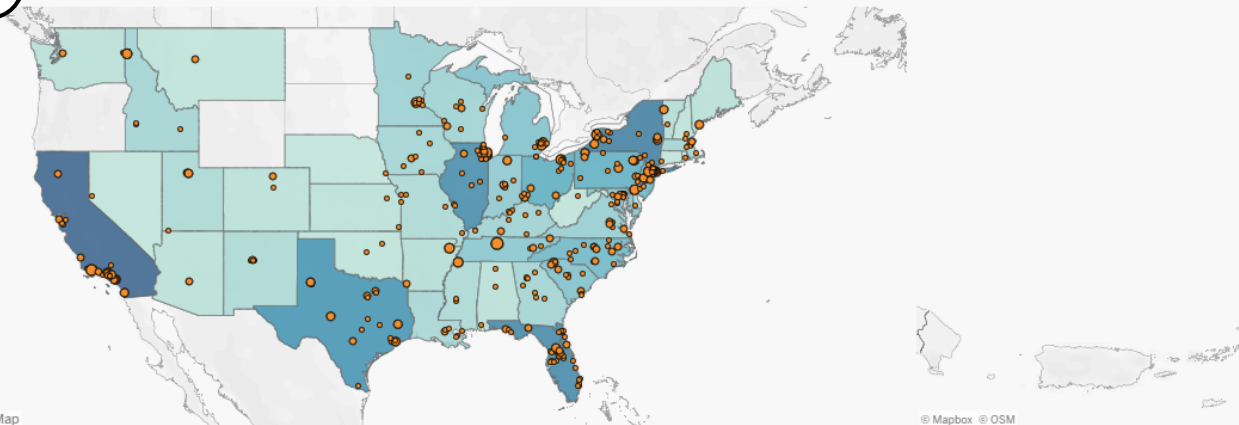
Date

10/14/2019

10/31/2019



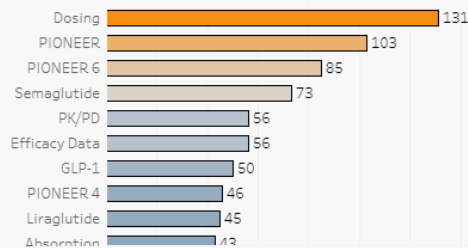
© 2019 Mapbox © OpenStreetMap



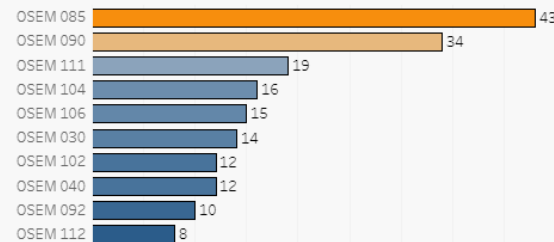
Insights

[Map]

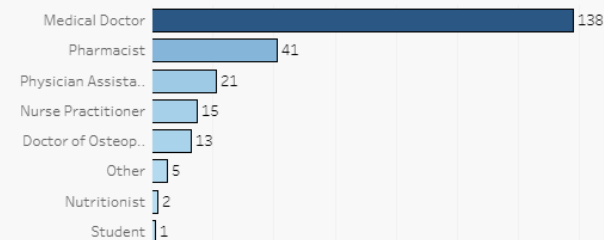
Insights



Letters

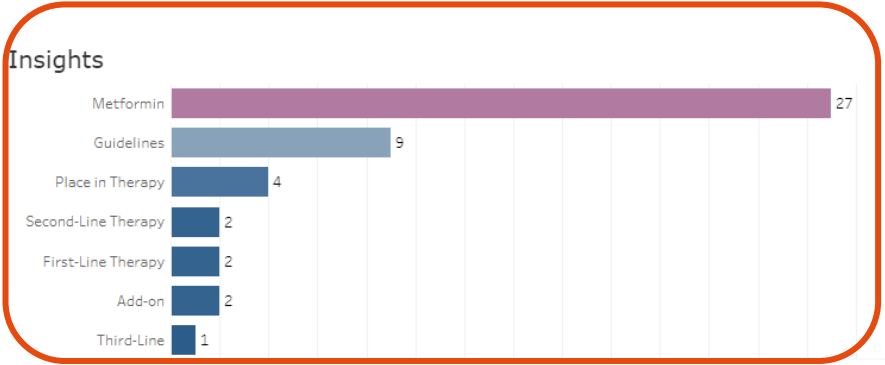
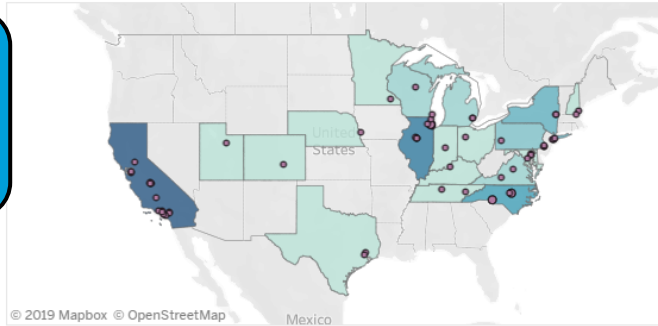


Degree

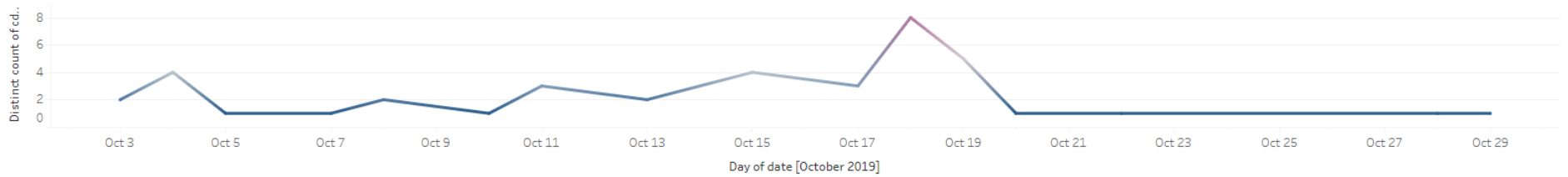




Sample:
What HCP are asking about Place in therapy for Rybelsus®?



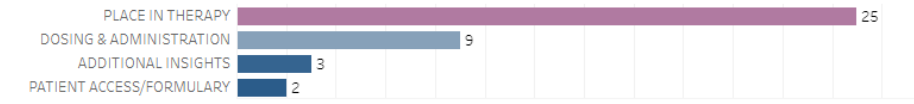
Over Time



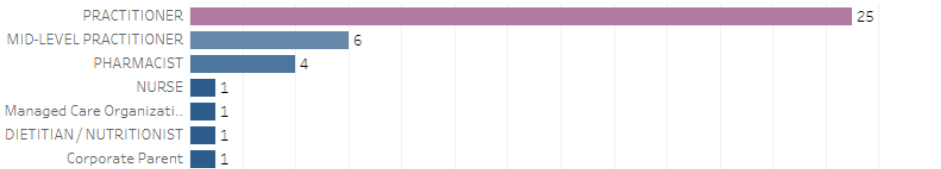
HCP Specialty
(All)

HCP Subtype
(All)

Subcategory



HCP Type



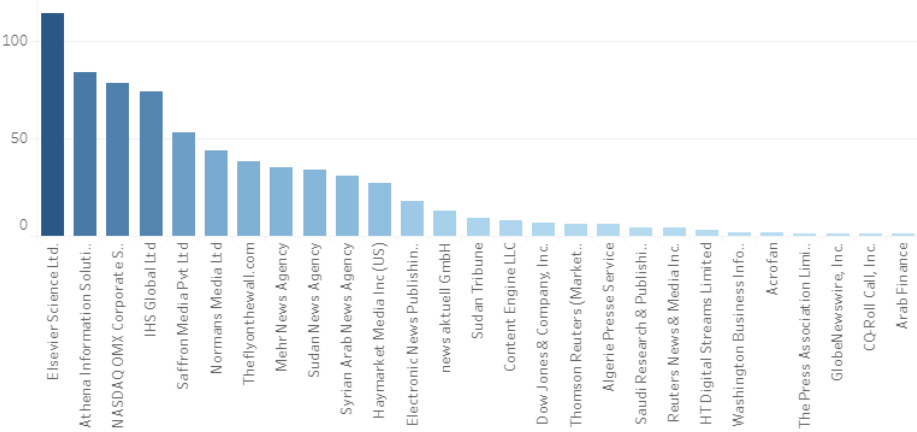
News

Top Articles

Novo Nordisk files for EU regulatory approval of oral semaglutide for the treatment of type 2 diabetes	34
Novo Nordisk files for EU regulatory approval of oral semaglutide for the treatment of type 2 diabetes.	32
Oral semaglutide provides superior HbA1c and weight reductions versus placebo in people with type 2 diabetes and renal impairment in the PIONEER 5 trial	26
Novo Nordisk announces positive results from PIONEER 5 trial with oral semaglutide in adults with type 2 diabetes & moderate renal impairment	26
Novo Nordisk's oral semaglutide beats Victoza and MSD's Januvia in treatment of diabetes	25

Sample:
What and where are stories being published about Rybelsus®?

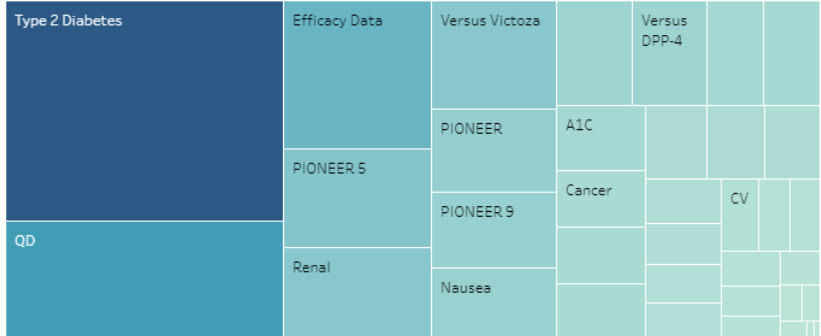
Publisher



Insight Categories

populations	548
Trials	533
efficacy	448
Dosing	290
safety	247
Retinopathy	62
Cardiovascular	43
PK/PD	14

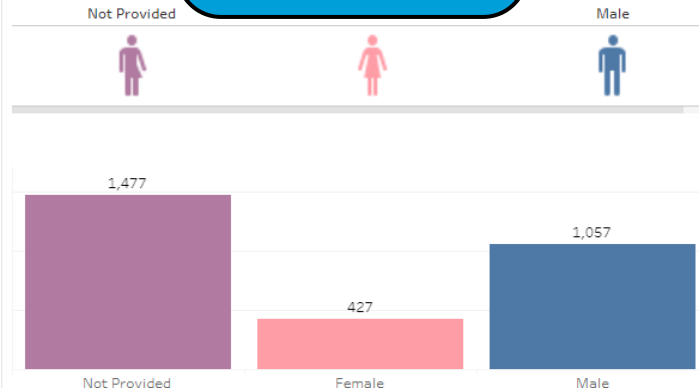
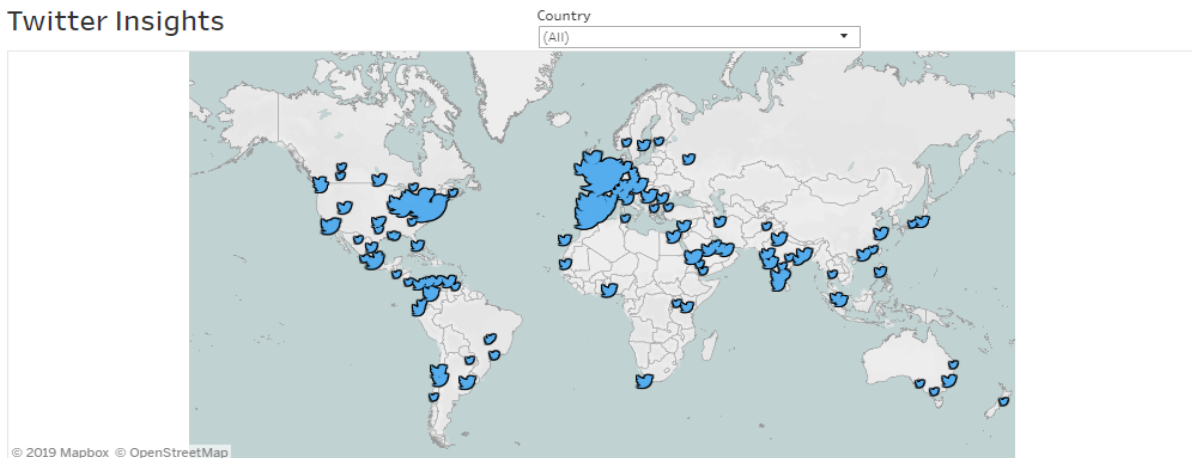
Insights



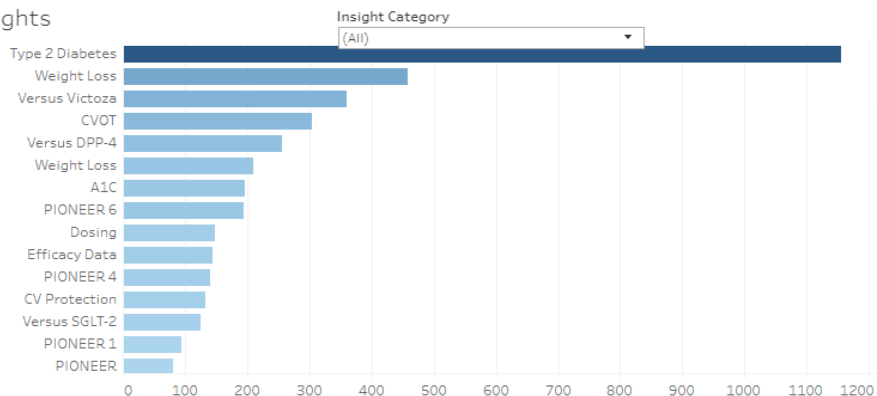
What's the conversation about Rybelsus® on Twitter?

Where are people tweeting from?

Twitter Insights



Insights



Tweet

1 of the biggest potential US FDA approvals in Sept is Novo Nordisk's oral semaglutide in type 2 diabetes. EvaluatePharma Vision data expects it to be their 2nd biggest growth driver, with \$3.2bn sales in 2024. See the full 2024 top 5 for type 2 diabetes <https://t.co/21s7NwOqIH> <https://t.co/gi3NXL6cEJ>

1. Oral #Semaglutide NDAs Seeking #T2D Indication, #CV Risk Reduction. 2. #Dapagliflozin Reduces MACE, #CV Death in Patients With #T2D and Prior #MI. 3. #FDA Places Clinical Hold on #Venetoclax Trials for Multiple Myeloma. <https://t.co/C8iA52Elrw>

4.8/5 stars at #EASD2019 - had an amazing time attending ...a huge THANK YOU to the entire @EASDnews team! A super busy week, capped off with @novonordisk Rybelsus approval (oral semaglutide)! See you in #Vienna

. @busediabetes previews results from the PIONEER Program Trials, which will be presented at #ADA2019 on Tuesday, June 11 @ADA_DiabetesPro @AmDiabetesAssn <https://t.co/jRL1aPjHp8>

Conference Abstracts

Abstracts

Month..F	Conference	Title	Authors
September 23, 2019	2019 Boulder Peptide Symposium of the Boulder Peptide Society (BPS)	Effect of Octreotide-Bile Salt Interactions on Octreotide Flux Through Polymeric Membranes	Dening, Tahr
September 16, 2019	55th Annual Meeting of the European Association for the Study of Diabetes (EASD)	Efficacy of oral semaglutide according to baseline HbA_{1c}: An exploratory subgroup analysis of the PIONEER trial programme	Meier, J.J.(1, T.M.(2); Ling M.B.(2); Zin J.(5)
		Efficacy of oral semaglutide according to diabetes duration: An exploratory subgroup analysis of the PIONEER trial programme	Haluzik, M.(Eriksson, J.V Kallenbach, Buse, J.B.(5)
		Oral semaglutide improves postprandial glucose and lipid metabolism and delays first-hour gastric	Dahl, K.(1); C.(2); Brook

Abstract

been developed by co-formulation with the absorption enhancer, sodium N-(8-[2-hydroxybenzoyl] amino) caprylate (SNAC). The site of absorption of oral semaglutide was investigated in a series of clinical and non-clinical studies. Twenty-four healthy males in the fasted state received single-dose oral semaglutide (10 mg with 300 mg SNAC) with 50 mL water followed by 4 h post-dose fasting (ClinicalTrials.gov ID NCT01619345). Complete tablet erosion (CTE) occurred in the stomach with an estimated mean time to CTE [95% confidence interval (CI)] of 85 min [62;118], as assessed by gamma scintigraphy. Semaglutide plasma exposure confirmed early systemic absorption (AUC_{0-24h} of 119 nmol·h/L [59;243]) with a median t_{max} of 90 min. When dosed in the fed state, limited or no semaglutide exposure was seen after once-daily oral semaglutide dosing for 10 days in 25 healthy subjects (5 days of 5 mg/300 mg SNAC, then 5 days of 10 mg/300 mg SNAC) (ID NCT02172313). Complementary studies in dogs showed that prevention of intestinal absorption by pyloric ligation followed by intragastric semaglutide dosing resulted in semaglutide exposure comparable to that seen after oral dosing in normal dogs (p=0.83). In another study, semaglutide exposure (AUC_{0-0.5h}) after intragastric dosing in non-ligated dogs was higher in the splenic vein (draining the gastric cavity) than in the portal...

has been developed by co-formulation with the absorption enhancer sodium N-(8-[2-hydroxybenzoyl] amino) caprylate (SNAC). Data suggest that SNAC facilitates the absorption of semaglutide in the stomach after oral administration. As the solubility of SNAC is pH dependent, and the mode of action of SNAC involves local buffering of the acidic pH, the absorption enhancing effect of SNAC on semaglutide may be affected by concomitant use of

Medline Publications

Month, Da..	Title	
February 1, 2019	Oral semaglutide for the treatment of type 2 diabetes.	
March 1, 2019	Cardiovascular safety of oral semaglutide in patients with type 2 diabetes: Rationale, design and patient baseline characteristics for the PIONEER 6 trial.	
April 16, 2019	Effect of Additional Oral Semaglutide vs Sitagliptin on Glycated Hemoglobin in Adults With Type 2 Diabetes Uncontrolled With Metformin Alone or With Sulfonyleurea: The PIONEER 3 Randomized Clinical Trial.	
June 1, 2019	Safety and Pharmacokinetics of Single and Multiple Ascending Doses of the Novel Oral Human GLP-1 Analogue, Oral Semaglutide, in Healthy Subjects and Subjects with Type 2 Diabetes.	
July 1, 2019	Efficacy and safety of oral semaglutide in patients with type 2 diabetes and moderate renal impairment (PIONEER 5): a placebo-controlled, randomised, phase 3a trial.	
	Efficacy and safety of oral semaglutide with flexible dose adjustment versus sitagliptin in type 2 diabetes (PIONEER 7): a multicentre, open-label, randomised, phase 3a trial.	
July 6, 2019	Oral semaglutide versus subcutaneous liraglutide and placebo in type 2 diabetes (PIONEER 4): a randomised, double-blind, phase 3a trial.	
August 29, 2019	Oral Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes.	
September 1, 2019	Effect of Oral Semaglutide on the Pharmacokinetics of Lisinopril, Warfarin, Digoxin, and Metformin in Healthy Subjects.	
	PIONEER 1: Randomized Clinical Trial of the Efficacy and Safety of Oral Semaglutide Monotherapy in Comparison With Placebo in Patients With Type 2 Diabetes.	
	Quantifying the Value of Orally Delivered Biologic Therapies: A Cost-Effectiveness Analysis of Oral Semaglutide.	
October 1, 2019	Cardiovascular, mortality, and kidney outcomes with GLP-1 receptor agonists in patients with type 2 diabetes: a systematic review and meta-analysis of cardiovascular outcome trials	

To view in Pubmed:
Click on the Publication
you wish to see and
select the link at the
bottom of the tooltip

Dashboard Workflow



Why Automate?

- Length of workflow
- Frequency of updates
- Data from multiple sources (not centralized)
- Increasing amount of legacy data
- Global demand from other Novo Nordisk affiliates
- What steps can be taken to speed up this process without sacrificing quality?

Technical Landscape

OASIS: Novo Nordisk's Cloud Based Global & Analytics Platform

Data Lake

Availability of a wide variety of cross-functional data in one central repository

Data Engineering

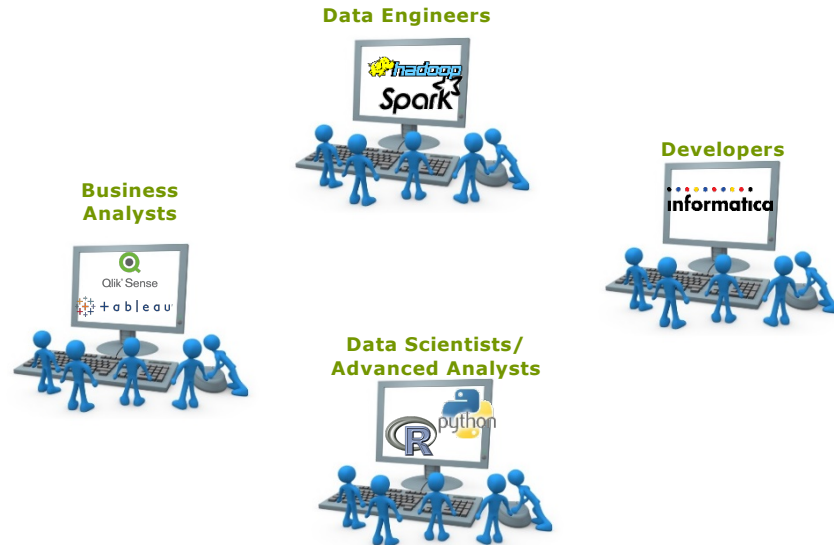
Tools to prepare blend and integrate various data assets

Data Science

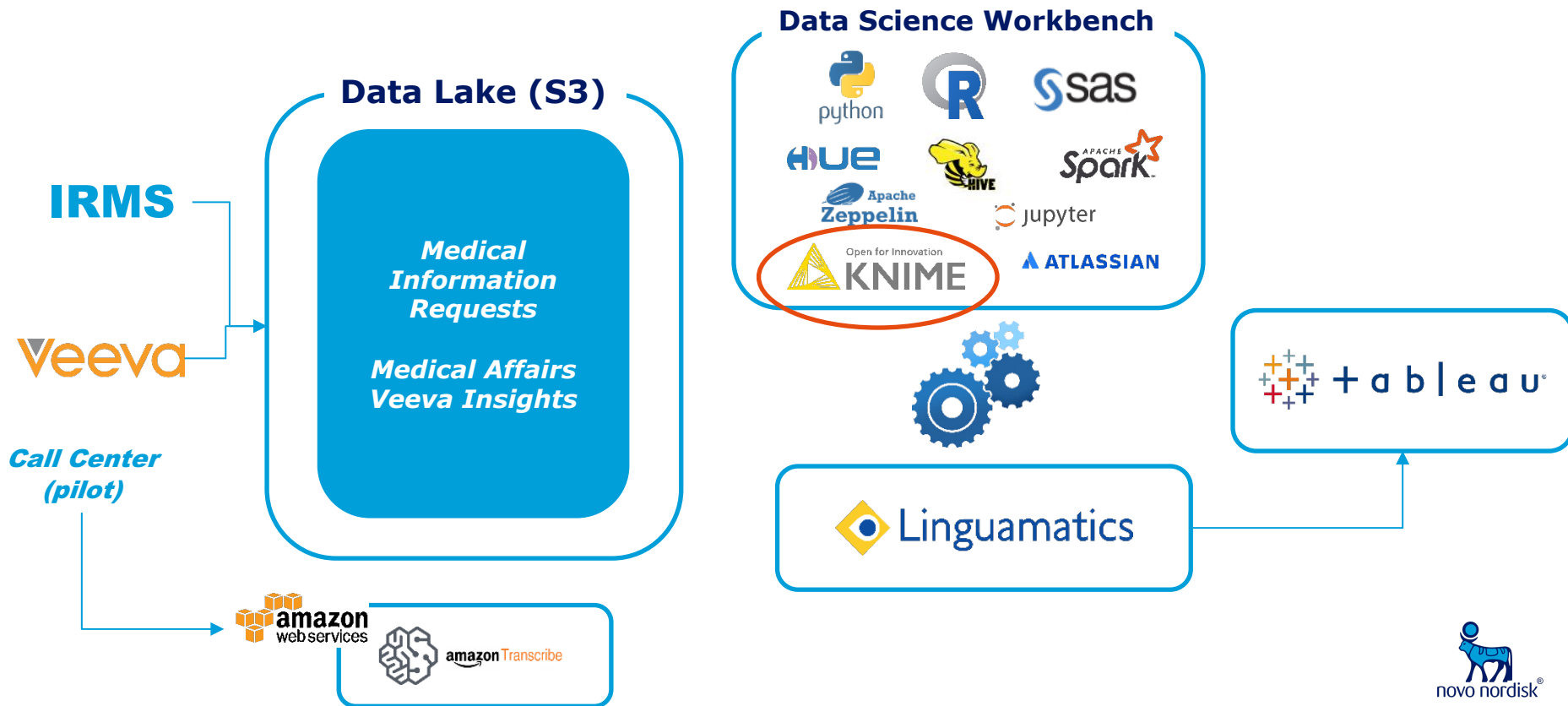
Tools to use analysis-ready data sets for advanced analytics & data science

Data Visualization

Connectivity to data in the data lake form BI tools visualize data.



OASIS Architecture



What we've done

- Installed I2E on a server in AWS
 - Migrated all queries, indexes, ontologies, etc. from preexisting environment to new server
- Automated six processes
 - Medical Information Requests (MIRs)
 - Medical Affairs Interactions
 - Social Media
 - News
 - Conference presentations
 - Publications
- Pilot
 - AWS Transcribe for Call Center

Process Enhancement: Medical Affairs

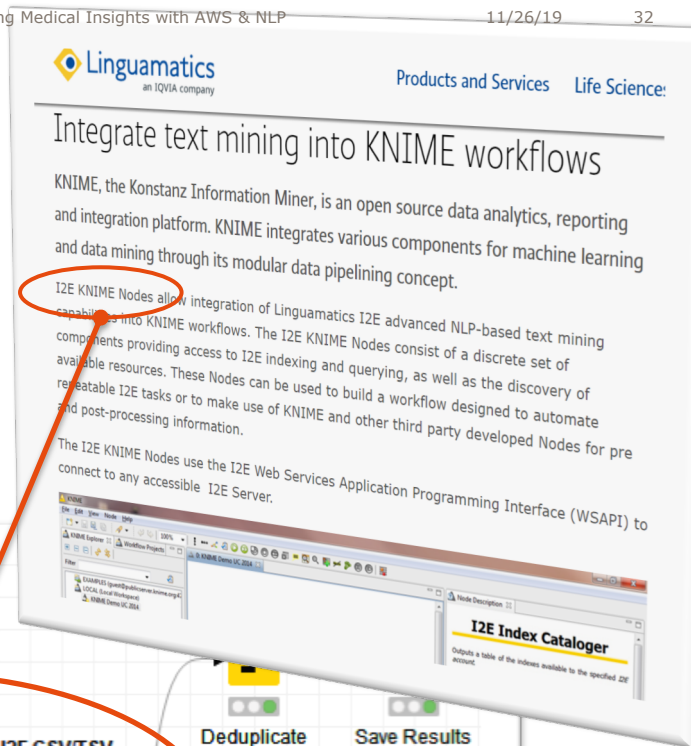
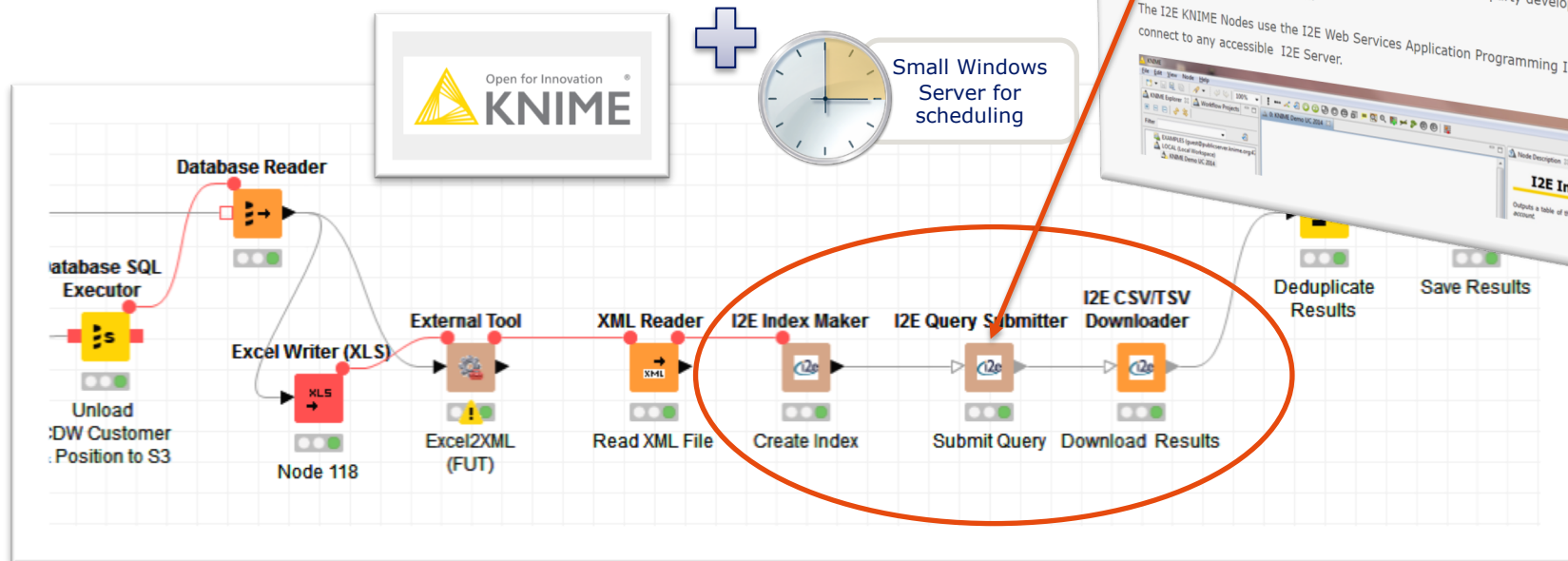
- Framework is reusable

Prior Process	Enhancement
Deloitte was used for Insight Analysis	Starting in 2019, process moved in house.
Data coming from multiple disparate sources	Using data available in a centralized location (data lake)
Geography data manually extracted from source systems.	Geography and customer attributes sourced directly from data lake.
Data refreshed Monthly	Data refreshed weekly
Manual indexing and querying in I2E	No manual involvement
Tableau Dashboard data sources updated manually	Tableau Dashboard updates automatically

Data Pipeline – Medical Affairs Insights



Automating Using KNIME



Speech to Text Pilot

- 75 Recordings from Call Center
- Used KNIME to submit recordings in batches to AWS Transcribe service.
- Python Program to beautify results



Towards the Future

- Automation of insights/scaling our offerings to meet global needs using existing framework.
- Build a data pipeline for text mining speech to text transcription results in I2E
- Expand topic areas to disease state landscapes
- Natural Language Generation (NLG) – building narratives within dashboards

Questions?

SHELLEY BRIGSTOCK, LENE HOLM JAKOBSEN &
LOLITA SOLE PEDERSEN
Global Development
Denmark