III DATAVERSITY

The Great Escape: Liberating 20+ Years of Legacy Enterprise Data

Demo Day July 19, 2023



About Today's Speaker



Eliud Polanco

- Former Head of Analytics at HSBC, DeutscheBank, ScotiaBank
- Former Head of Data Innovation @ Citi
- Current President at Fluree PBC, a start-up focused on secure data-sharing and collaboration at scale using Zero-Trust and Semantics



What Fluree Offers



Intelligent Knowledge Graph

Fluree combines cryptography and semantic data standards to enable secure data collaboration



AI/ML Data Cleansing Pipeline

With **Fluree** it is faster and easier to clean raw data and make it consumable for users and algorithms via semantic Knowledge Graph



Agenda

- I. The Knowledge Graph opportunity. Why Knowledge Graphs are emerging as more critical for Enterprise use cases.
- II. The Legacy Data Problem. Why Knowledge Graphs have historically been difficult to adopt at scale.
- **III. The Innovations.** How JSON-LD and AI/ML are helping bridge the gap between relational and semantic worlds.
- **IV. The Demo.** An How Fluree converts scale legacy data into data ready for semantic Knowledge Graphs.
- V. Getting Started With Knowledge Graphs In Your Organization. A BluePrint for socializing KG use cases in your company.





I. The Knowledge Graph Opportunity



Why Look At Knowledge Graphs?

- Data sharing is becoming an imperative for competitive differentiation
 - Inside a company's four walls
 - Between companies and regulatory agencies who have adopted semantic standards for how they want to receive data
 - Between companies and suppliers, partners, even competitors using common industry ontologies
- Al, Algos and Machines are becoming the primary consumer of vast stores of data inside a company
 - 1. Generative AI and Large Language Models (LLM) shook the world
 - 2. Internet of Things: devices create data, and devices consume data

Semantic interoperability is no longer a niche use case... it is an economic imperative

1. Large Language Models (LLMs)



- The vast quantity of data consumed into LLMs are public data that is either unstructured, or using semantic Web standards (RDF)
- Some market data (data available for a price) is opportunistically being integrated into LLMs (at a price)
- Most internal data is out of the reach, because:
 - It is proprietary.
 - It lacks semantic context.
- Large enterprises today are experimenting in how to **build their own internal LLMs** using a mix of their own structured and unstructured corporate internal data.
- This will drive **competitive differentiation inside companies** in how they create products and optimize their own business processes

fluree sense

Internet of Things (IoT)



- Data being created by sensors and actuators...
- ...and being distributed via APIs...
- ...to be consumed by algorithmic models (e.g., predictive maintenance, or personalization recommendation)...
- ...with little man-in-the-middle intervention.
- Many IoT devices use JavaScript Object Notation (JSON) as the defacto lightweight data interchange format. It is easy for machines to parse and generate.
- This means, that over time, a vast amount of Operational Technology (OT) data will be created in JSON.
- There is a desire to link Information Technology (IT) or relational world data to OT data to create even better models.







Common Conceptual Data Architecture Pattern in Most Large Enterprises





Barrier 1: Contextless Data Proliferation



- Data created by app, for the app. Mostly for Operational use cases
- Semantic context for Data Layer embedded inside the application logic layer

- Data saved for a consumer other than the originating app
- But data is flat and has no context!!

- The copy has a different schema than the original
- Any consumer can create their own consumption schema on a use case by use case basis!



Example: Data from an ERP System

mandt 0	matnr 0	ersda 🌣	ernam ¢	laeda 🌣	aenam 🌣	vpsta 🗘	pstat ≎	lvorm 0	mtart 0	mbrsh 0	matki ¢	bism
T) (v	(T	τ.	τ	T	T		τ	τ) [T] [T) T
400	0000000000000000	2001-06-04T00:00:0	DNOBLE	2020-05-21T00:00:0	PHOU	KDELBSVQGCZX	KDELBSVQGC	null	Z1H	Р	1005	22:
400	000000000000000	2001-06-04T00:00:0	DNOBLE	2019-06-20T00:00:0	MSMITH3	KDELBSVQGCZX	KDELBSVQGC	х	Z1H	Ρ	1000	103
400	000000000000000000000000000000000000000	2001-06-04T00:00:0	DNOBLE	2019-07-03T00:00:0	BSAFINEJ	KDELBSVQGCZX	KDELBSVQGC	null	Z1H	Ρ	1000	101
400	0000000000000000000072	2001-06-04T00:00:0	DNOBLE	2020-02-04T00:00:0	RJACKAN	KDELBSVQGCZX	KDELBSVQGC	null	Z1H	Ρ	1000	106
400	00000000000000000	2001-06-04T00:00:0	DNOBLE	2019-11-29T00:00:00	LSANTACR	KDELBSVQGCZX	KDELBSVQGC	null	Z1H	Ρ	1000	107
400	00000000000000000	2001-06-04T00:00:0	DNOBLE	2019-01-25T00:00:0	MSMITH3	KDELBSVQGCZX	KDELBSVQGC	x	Z1H	Р	1000	11 C
400	0000000000000000	2001-06-04T00:00:0	DNOBLE	2019-07-03T00:00:0	BSAFINEJ	KDELBSVQGCZX	KDELBSVQGC	null	Z1H	Р	1005	221
400	0000000000000000	2001-06-04T00:00:0	DNOBLE	2020-05-21T00:00:0	PHOU	KDELBSVQGCZX	KDELBSVQGC	null	Z1H	Р	1005	22!
400	000000000000000000000000000000000000000	2001-06-04T00:00:0	DNOBLE	2019-07-03T00:00:0	BSAFINEJ	KDELBSVQGCZX	KDELBSVQGC	null	Z1H	Ρ	1000	101
400	000000000000000000000000000000000000000	2001-06-04T00:00:0	DNOBLE	2019-06-20T00:00:0	MSMITH3	KDELBSVQGCZX	KDELBSVQGC	х	Z1H	Ρ	1000	104
400	000000000000000000000000000000000000000	2001-06-04T00:00:0	DNOBLE	2019-07-03T00:00:0	BSAFINEJ	KDELBSVQGCZX	KDELBSVQGC	null	Z1H	Ρ	1000	11C
400	000000000000000000	2001-06-04T00:00:0	DNOBLE	2020-05-21T00:00:0	PHOU	KDELBSVQGCZX	KDELBSVQGC	null	Z1H	Ρ	1005	221
400	0000000000000000	2001-06-04T00:00:0	DNOBLE	2019-07-03T00:00:0	BSAFINEJ	KDELBSVQGCZX	KDELBSVQGC	null	Z1H	Ρ	1000	105
400	00000000000000000	2001-06-04T00:00:0	DNOBLE	2019-07-03T00:00:0	BSAFINEJ	KDELBSVQGCZX	KDELBSVQGC	null	Z1H	Р	1000	11C
400	000000000000000000000000000000000000000	2001-06-04T00:00:0	DNOBLE	2019-07-03T00:00:0	BSAFINEJ	KDELBSVQGCZX	KDELBSVQGC	null	Z1H	Р	1005	221
400	000000000000000	2001-06-04 <mark>T</mark> 00:00:0	DNOBLE	2019-07-03T00:00:0	BSAFINEJ	KDELBSVQGCZX	KDELBSVQGC	null	Z1H	Ρ	1005	221
400	0000000000000000	2001-06-04T00:00:0	DNOBLE	2019-07-03T00:00:0	BSAFINEJ	KDELBSVQGCZX	KDELBSVQGC	null	Z1H	Ρ	1000	107
400	000000000000000	2001-06-04T00:00:0	DNOBLE	2019-07-03T00:00:0	BSAFINEJ	KDELBSVQGCZX	KDELBSVQGC	null	Z1H	Ρ	1000	106
400	00000000000000000	2001-06-04T00:00:0	DNOBLE	2019-07-03T00:00:0	BSAFINEJ	KDELBSVQGCZX	KDELBSVQGC	х	ZOH	Р	1100	103
<100	0000000000000000	2001-06-04700-00-0		0010-07-00T00-00-0	DCAEINE I			null	71⊔	n	1005	22.

- What is this...?!!
- What do these columns mean...?

ERP System Data Dictionary



	CONCEPTS (239)										
	Concepts ¢	Definitions 0									
	٣	[τ									
	Add Concept	Add definition									
:	MANDT	Client									
:	MATNR	Material Number									
:	ERSDA	Created On									
:	ERNAM	Name of Person who Created the Object									
:	LAEDA	Date of Last Change									
:	AENAM	Name of Person Who Changed Object									
:	VPSTA	Maintenance status of complete material									
	PSTAT	Maintenance status									
÷	LVORM	Flag Material for Deletion at Client Level									
:	MTART	Material Type									
:	MBRSH	Industry Sector									
:	MATKL	Material Group									
:	BISMT	Old material number									
:	MEINS	Base Unit of Measure									
:	BSTME	Purchase Order Unit of Measure									
:	ZEINR	Document number (without document management system)									
:	ZEIAR	Document type (without Document Management system)									

- OK, this table does have a meaning...
- It's just that the context resides in some document somewhere else apart from the data (e.g., some catalog system or document repository)
- Now multiply this problem for hundreds of apps and tens of thousands of database tables



Barrier 2: Mental Model and Syntax



- Two completely different worlds
- · Very few people in an organization can operate proficiently in both
- Years of legacy investment, time, training in Relational World







Two Technical Innovations

- JSON Linked Data Standard (JSON-LD): a nice go-between Relational World and Semantic World
- Robust data classification Al Models: a method for layering semantics to flat data at the massive scale required



What is JSON Linked Data? (JSON-LD)

- A method for encoding linked data ("semantic context") using JSON
- Easy to link ontologies to JSON documents and map to RDF

```
{ "@context": {
    "name": "http://xmlns.com/foaf/0.1/name";
    "homepage": {
        "@id": "http://xmlns.com/foaf/0.1/workplaceHomepage";
        "@type": "@id"
    },
    "Person": <u>http://xmlns.com/foaf/0.1/Person</u>
},
    "@id": "https://me.example.com",
    "@type": "Person",
    "name": "John Smith",
    "homepage": "https://www.example.com/"
}
```



Why does JSON-LD matter?

- IoT transacts in JSON
- Lots of APIs transact in JSON
- Very easy to ETL from CSV to JSON
- JSON-LD is not a huge leap from JSON
- Lots of websites use JSON-LD for Google Search Engine Optimization
- Generative AI uses it a lot!!! Easy fit for LLMs

JSON-LD as a bridge to RDF / Triple Stores



fluree sense



Using Data Classification Models to add semantics at scale

- Machine learning models can be trained how to tag columns to vocabulary terms and then infer vocabulary to vocabulary synonym terms
- Once the model is trained, you can run an automated process to scan all columns and then classify them to their respective terms.
- Once complete, you can convert the discovered mappings and semantic relationships into json-Id files that can be transacted into a Knowledge Graph.
- There are many ML classification models; for this talk we will go through an example using a Neural Network



AI/ML Data Classification Methodology

Training Data



Classifiers (Vocabulary for an Entity Called "Fruit")

ID	Term	Definition
FR1	Fruit Name	The name of the fruit
FR2	Fruit Color	The hue or color of the fruit
FR3	Characteristic	Defining feature whether shape or texture used to best describe the fruit





Scaling the Process Through Crowdsourcing

- The power of the approach is to distribute the training to as many people as possible
- Tens of people can individually classify 1-2 tables, or validate the machine prediction and give it reinforcement
- This will generate thousands or tens of thousands of new features to further refine the model





Fluree's Approach to Legacy Data Conversion





Step 1. Connect to Data Sources

Offure	Classify ▼ Start by Default search				œ	¢ 2 0 ₿	PDU Pharma I Sign Out	Demo User
	ata Sources							
	Create New Data Source							
00	E DATA SOURCES (10)							0
	Data Source Name 🔅	Data Source Type	• Created By	 Last Modified 	# of Datasets	• # of Group Entitlements	 # of User Entitlemen 	nts [÷]
Ħ	Ϋ́	Υ	T	T	T	T		
	IDMP Mapping Full	🎲 Hadoop	pharma_user	04/11/2023		10	1	6
	DMP Mapping	🤣 Hadoop	pharma_user	04/11/2023		14	1	6
	SAPECC	👘 Hadoop	pharma_user	03/29/2023		2	1	6
E	CSRS registration system for the ingredients in medicinal products	Hadoop	pharma_user	04/7/2023		2	1	6
₹≡	IQVIA Market intelligence of pharmaceutical customers, products and clinical inf	Hadoop	pharma_user	04/11/2023		5	1	6
闘	GEN Pharma Default Datasource Default Datasource created while creating a tenant	🎲 Hadoop	59_user	04/7/2023		13	1	6
\$								
Oflure	Copyright © 2023 Fluree - All rights reserved					V	ersion: v0.2.1 Env	vironment: fsqa

Step 2. Generate Data Set Profiles

Oflur	📽 Classify 👻 🗹 st	tart by Default search	٩			ଓ 🗘 🖓 🛞 😳 Sign Out
	Data Sets 🗲 📄 MARA	í.				
ជ	Data Set Description SAP Material data u Object Tags G MARA Medicin	sed to retain pharma drug mate Ø	Data Source SAP ECC Data Set Type Parquet	Total Records 273,105 Total of Columns 22	Last Modified On 03/29/2023 Ga Classify Set	Overall quality score 100%
₫\/	Data Set Attributes	Data Set Sample Data Set Relationships Data Refresh	Data Entitlements Related Projects Dat	a Quality		
Ħ	Attribute 0	Data Type \diamond Classification \diamond	Profiler Result		Classification Result	
0	٣	T	Total Size for Profiling: 273,105 Rows Total Rows: 273,105	Total Sampled Rows: 273,105 03/29/2023		
	ManufacturerCode	STRING	Desults	Date Times		
	ManufacturerDesc	STRING	uniques 13 (0.0%)	NUMBER 100.0%		
	ManufacturerDesc		Nulls 252974.0 (92.63%)			
	MAKTX	STRING	Distribution Of Values	Regex Patterns		
Ē	mara_ZZ_PLANT	LONG	Min. Value 2	9 72.89%		
	STRU_DESC	STRING	Max, Value 99	99 27.11%		
¥=	FAMI_DESC	STRING				
•	SUBF_DESC	STRING				
æ	STRE DESC	STRING				
H	DAGA DECO	CTDING				
3	PAFM_DESC	STRING				
5.5	PASI_DESC	LONG				
	MATNR	LONG	Top 10 Most Frequent 252974	94 15		
	mara_MATNR	LONG	2 8794	12 96		
	WRKST	STRING	8 4694	97 105 96 136		
	ZZ STRUCTURE	LONG	99 1240	11 168		
	77 DEAMORD	STRING	9 1186	16 357 98 396		
			98 396	95 536		
	mara_ZZ_PFAMGRP	STRING	16 357	9 1186		
	ZZ_PRODFM	STRING	MedicinalProductName	00 1240		
	MTART	STRING				
	UMREN	STRING				
	LAEDA	DATE				Run Model 😰
Oflure	e Copyright © 2023 Flur	ee - All rights reserved				Version: v0.2.1 Environment: fsqa



Step 3. Import Ontologies

Oflur	Se Classify - Start by Default search	٩								2 ⁹ (?)	ø 😶	Pharma Demo Use Sign Out
	Data Catalogs											
	Create New Data Catalog +											
00	DATA CATALOGS (6)											5
11	Data Catalog 🗘	Data Catalog Description 0	Created By \circ	Last Modified 0	Semantic Objects	• Concepts	 Mapped Data Sources 	 Mapped Data Sets 	 Mapped Data Set Columns 	• QI RI	of Data uality ules	Measured Data Quality
	T	T	т	T	۲	T	٣	T	۲.		r	T
	Automation_catalogtest	null	Pharma Demo User	05/2/2023		0	0	0	0	0	Add	n/a
P	E GEN Pharma Catalog	Business information model for Generics Pharmaceutical company	Pharma Demo User	03/30/2023		3	32	3	3	51	Add	n/a
0	Global Substance Registration System (GSRS)	Registration system for the ingredients in medicinal products. It makes it easier for regulators and other stal	c. Pharma Demo User	02/27/2023		34 7	86	2	2	25	2	100%
	Identification of Medicinal Products (IDMP)	IDMP is a suite of five standards developed within the ISO to facilitate the unique identification of medicinal	. Pharma Demo User	03/9/2023		3	90	3	3	32	2	100%
	SAP ECC Data Dictionary	Data glossary for SAP ECC R3 system	Pharma Demo User	04/7/2023		75 47	84	1	1	11	Add	n/a
	zFlureeSense System Catalog	Out of the box system catalog and semantic entities	Fluree Sense	02/27/2023		2 2	62	1	1	2	Add	n/a
Offure	€ Copyright © 2023 Fluree - All rights reserved										Versio	n: v0.2.1 Environment; f

fluree sense

Step 3. Import Ontologies (cont'd)

Offlure	Classify 👻 🗹 Start by Default	search Q					¢ ¢	0 🕸 😶	Pharma De Sign Out	mo Use
	Data Catalogs > Identification of Medi	cinal Products (IDMP) > Pharmaceutical Products					(Technical View	Business	View
ය i N	Data elements and structures for unit 3 Data Sources As of 5/6/2023	que identification and exchange of regulated pharmaceutical product information A Data Sets 60 As of 5/b/2023	Data Quality Rules 60 As of \$(6)/2023	,	n/a Meas As of 5	sured Data (/8/2023	Quality			
	* CONCEPTS (19)							1		٥
R	Concepts 0	Definitions ©	Tags \circ	Actions	Synonyms	Mapped Data Columns	• Train Model	# of Data Quality Rules	Measured Data Quality	¢
0	τ	Υ	٣			T		T	Ŧ	
	Add Concept	Add definition		+						
_	Reference Strength	strength of an active substance(s) and/or specified substance(s) used as a reference from which the strength of an inv	+ Add Tag	Ø	<i>8</i> 1	0	*	Add	n/a	
E	Presentation Strength	quantity or range of quantities of the substance/specified substance present per unitary volume (or mass) expressed in	+ Add Tag	Ø	81	1	8	Add	n/a	
~	Pharmaceutical Product Identifier	unique identifier for a pharmaceutical product	+ Add Tag	Ø	8 2	3	1	Add	n/a	
Ϋ́=	Placebo	inactive substance, treatment or procedure that is intended to provide baseline measurements for the experimental pro	+ Add Tag	Ø	81	0	i≡ (1)	Add	n/a	
R	Packaged Medicinal Product	medicinal product in a container being part of a package, representing the entirety that has been packaged for sale or s	+ Add Tag	Ø	8 1	0	合	Add	n/a	
=	Investigational Medicinal Product I	unique identifier allocated to an investigational medicinal product supplementary to any existing identifier as ascribed b	+ Add Tag	Ø	80	0	18	Add	n/a	_
63	Clinical Trial	investigation in human subjects intended to discover or verify the clinical, pharmacological and/or other pharmacodyna	+ Add Tag	Ø	<i>8</i> 2	0	ŝ	Add	n/a	
-0-	Package Item		+ Add Tag	Ø	81	1	ŝ	Add	n/a	_
	Outer Packaging		+ Add Tag	0	81	0	1	Add	n/a	
	Intermediate Packaging		+ Add Tag	Ø	81	0	ŝ	Add	n/a	_
	Immediate Container		+ Add Tag	Ø	81	0	畲	Add	n/a	
	Investigational Medicinal Product		+ Add Tag	Ø	80	0	1	Add	n/a	_
	Medicinal Product Identifier	unique identifier allocated to a medicinal product supplementary to any existing authorization number as ascribed by a	+ Add Tag	0	80	0	ŝ	Add	n/a	- 1
	Medicines Regulatory Agency	institutional body that, according to the legal system under which it has been established, is responsible for the grantin	+ Add Tag	Ø	81	0	â	Add	n/a	
	Mass Based Strength		+ Add Tag	0	81	0	畲	Add	n/a	
	Concentration		+ Add Tag	0	81	0	ŝ	Add	n/a	
	Activity Based Strength		+ Add Tag	0	81	0	ŝ	Add	n/a	
Oflure	e Copyright © 2023 Fluree - All rights rese	erved						Version	v0.2.1 Enviro	nment: fsi

Step 4. Train Classification Model

Oflu	Start by Default	arch	Q			() L <mark>9</mark> () () () () () () () () () () () () ()	Pharma Demo User Sign Out
	Data Catalogs > Identification of Medicinal Pr	oducts (IDMP) > Pharmaceutical Products	> Pharmaceutical Product Identifier > Train Mo	odel			
	Review Prior Training Data	Select Data Sets	Map Data Columns		Finalize Training Data		
ជា	Map the column from the data sets selected tha	t accurately represent the Concept that is be	ing trained.				
/11 <	TM TOTAL DATA ATTRIBUTES (8)			\$	PHARMACEUTICAL PRODUCTS SEMANTIC OBJ	IECT MODEL MAPPINGS (0 FIELDS MAPPED)	۵
	Data set 🌣	Attribute 0	Mapping(0)		Concept ¢	Mapping	
Ш	т	τ.			T		
P	IDMP Medicinal Products - IMS	Manufacturer			Pharmaceutical Product Identifier	Drag & drop data set attibute here.	
	IDMP Medicinal Products - IMS	MedicinalProductIdentifier					
B	IDMP Medicinal Products - IMS	MedicinalProductName			IDMP Medicinal Products - IMS		
E	IDMP Medicinal Products - IMS	PackageItem			PharmaceuticalProductIdentifier		
	IDMP Medicinal Products - IMS	PharmaceuticalProductIdent	ifier		0		
U	IDMP Medicinal Products - IMS	PresentationStrength					
ž	IDMP Medicinal Products - IMS	RegistrationNumber					
	IDMP Medicinal Products - IMS	id					
BBB							
.0.							
23							
						Cancel 🛪 🤄 Previous	Step Next Step 🌖
Oflur	ee Copyright © 2023 Fluree - All rights reserved					v	ersion: v0.2.1 Environment: fsqa

Step 5. Review results and calibrate model

View of the second s	Oflu	😝 Classify 👻 🖌 s	tart by Default Searc	ch	Q			🛞 🗘 🖓 🕸 🔁 Pharma Demo User Sign Out
		Data Sets > 🗐 MARA	κ					
No. Data Statungen Data Statungen <th>仚</th> <th colspan="3">Data Set Description SAP Material data used to retain pharma drug mate Object Tags 60 MARA Medicinal Products MedicinalProduct</th> <th>Data Source SAP ECC Data Set Type Parquet</th> <th>Last Modified On 03/29/2023 Ga Classify Set</th> <th>Overall quality score 100%</th>	仚	Data Set Description SAP Material data used to retain pharma drug mate Object Tags 60 MARA Medicinal Products MedicinalProduct			Data Source SAP ECC Data Set Type Parquet	Last Modified On 03/29/2023 Ga Classify Set	Overall quality score 100%	
Image: series National series N		Data Set Attributes	Data Set Sample	Data Set Relationships Data Refresh	Data Entitlements Related Projects Data	a Quality		
V V	\blacksquare	Attribute ¢	Data Type 🗢	Classification 0	Profiler Result		Classification Result	
No. No. <th></th> <th>Υ</th> <th>T</th> <th>T</th> <th>Total Size for Profiling: 273,105 Rows</th> <th>Total Columns: 22 Last Extract on Total Sampled Powe: 273105 03/29/2023</th> <th></th> <th></th>		Υ	T	T	Total Size for Profiling: 273,105 Rows	Total Columns: 22 Last Extract on Total Sampled Powe: 273105 03/29/2023		
Manufacture/Beil: STRING Manufacture/Meil: Man		ManufacturerCode	STRING	ManufacturerId	Beguite	Data Tupor	Semantic Concept Mapping	g Confidence Level o
Notice Notice State State Notice State <		ManufacturarDasc	STRING	Manufacturer ManufacturerName	uniques 13 (0.0%)	NUMBER 100.0%	Object	
MAR /k S HUNC Mademaka Preduct Mande WAR /k S HUNC		MananactorerDese	STRING		Nulls 252974.0 (92.63%)		<u> </u>	
Image:ZZZ:PLANT LONG ZZ:PLANT LONG Agent Zo:PLANT LONG Agent Zo:PLANT LONG Agent Zo:PLANT LONG Agent Zo:PLANT LONG String Zo:PLANT LONG MATNR LONG Mathematical Poduct Name Different To Piceas Frequent		MAKIX	STRING	Medicinal Product Name	Distribution Of Values	Regex Patterns	MARA ZZ_PLANT HIG	H 100% 💼 1 😱
STRU_DESC STRINO FMU_DESC STRINO SUBF_DESC STRINO SUBF_DESC STRINO MATINR LONG MATINR LONG MATINR LONG Z_SFRU/DERC STRINO TOP I MOSE Frequent Normal Participane Product Medicina Pr	E	mara_ZZ_PLANT	LONG	ZZ_PLANT	Min. Value 2	9 72.89%	MedicinalProduct Pharmaceutical.	N 35% 🖒 🗘
YAMULDESC STRING Agent Subtance Name SUBF_DESC STRING Subtance Name Classifier		STRU_DESC	STRING		Max. Value 99	99 27.11%	Medicinal Products Pharmaceutical.	N 31% 🖞 🕠
SUBF_DESC STRN0 Substance Name Classifier STRE_DESC STRN0 Strength NAMTAR CN0 MATINR NAMTAR CN0 MATINR T2	*=	FAMI_DESC	STRING	Agent Substance Name			MedicinalProduct MedicinalProdu.	N 0% 🖆 👎 1
STRE_DESC STRING Strength NPM_DESC STRING	•	SUBF_DESC	STRING	Substance Name Classifier			MedicinalProduct MedicinalProdu	N 0% 🖒 🖷 1
PARALDESC STRING PARALDESC LONG MATNE PARALDESC LONG MATNE MATNE LONG MATNE MATNE LONG MATNE VIKST STRING Madianal Product Name Top 10 Lest Frequent ZZ_STRUCTURE LONG Matinal Product Name Top 10 Lest Frequent Tage Bin Conditional Product Name Dial String Matinal Product Name Tage String Madianal Product Name Madianal Product Name Dial String Madianal Product Name Matrix String Matinal Product Name Medianal Product Name Medianal Product Name Medianal Product Name Matrix String Matinal Product Name Medianal Product Name <		STRE_DESC	STRING	Strength				010 83
Number Number PASLESC LONG MARK LONG MARK LONG mar.MATNR LONG MKST STRING VKST STRING ZZ,STRUCTURE LONG Market ZZ,STRUCTURE MRGCOR STRING Market Medicinal Product Name VKST STRING Market STRING Market Medicinal Product Name MART STRING Medicinal Product Name Market STRING		PAEM DESC	STRING					
PAS_LESC LING MATNR LONG MATNR MATNR Image MATNR LONG MATNR MATNR Top 10 Most Frequent 1 mara_JMATNR STRING Matninal Product Name 1 1 1 XZ_STRUCTURE LONG Matinal Product Name 1 1 1 1 ZZ_STRUCTURE STRING Medicinal Product Name 1	53	Dial Dias	LONG					
MATNR LONG MATNR Important (Construction) mara_MATNR LONG MATNR (Diportant (Construction)) (Diportant (Construction)) mara_MATNR LONG MATNR (Diportant (Construction)) (Diportant (Construction)) WRKST STRING Mathina (Construction) (Diportant (Construction)) (Diportant (Construction)) ZZ_STRUCTURE LONG Mathina (Construction) (Diportant (Construction)) (Diportant (Construction)) ZZ_STRUCTURE STRING Medicinal Product Name (Diportant (Construction)) (Diportant (Construction)) (Diportant (Construction)) mara_ZZ_PFANGRP STRING Medicinal Product Name (Diportant (Construction)) (Diportant (Construction)) (Diportant (Construction)) mara_ZZ_PFANGRP STRING Medicinal Product Name (Diportant (Construction)) (Diportant (Construction)) (Diportant (Construction)) MATR STRING Medicinal Product Name (Diportant (Construction)) (Diportant (Construction)) (Diportant (Construction)) MARK STRING STRING (Mathina) (Diportant (Construction)) (Diportant	5	PASI_DESC	LONG		Top 10 Most Frequent	Top 10 Least Frequent		
mara_MATNR LONG MATNR LONG MATNR Page 879 12 96 WRKST STRING Medicinal Product Name 464 97 105 Z2_STRUCTURE LONG ZZ_STRUCTURE LONG ZZ_STRUCTURE 11 188 Z2_FAMORP STRING Medicinal Product Name 99 124 11 188 Mara_ZZ_FAMORP STRING Medicinal Product Name 95 536 95 386 Z2_PROOFM STRING Medicinal Product Name Medicinal Product Name 16 357 95 186 MTAR STRING Medicinal Product Name 16 357 95 186 386 MARN STRING Medicinal Product Name 16 93 140 140 140 MARN STRING Matri 16 93 140 140 140 140 140 140 140 140 140 140 140 140 140 140 <td></td> <td>MATNR</td> <td>LONG</td> <td>MATNR</td> <td>252974</td> <td>94 15</td> <td></td> <td></td>		MATNR	LONG	MATNR	252974	94 15		
WRKST STRING Medicinal Product Name Addia 67 105 Z2_STRUCTURE LONG ZZ_STRUCTURE IONG ZZ_STRUCTURE 99 1240 11 188 SZ_ZPFAMGRP STRING Medicinal Product Name 99 1240 11 188 mara_ZZ_PFAMGRP STRING Medicinal Product Name 95 536 386 396 ZZ_PRODFM STRING Medicinal Product Name 16 357 9 116 ZZ_PRODFM STRING Medicinal Product Name 16 99 1240 124 MTART STRING Material Product Name Medicinal Product Name 16 1240 UNREN STRING MEDA MEDA Medicinal Product Name 16 1240 LAEDA DATE LAEDA MEDA MEDA 1240 1240 1240		mara_MATNR	LONG	MATNR	2 8794	12 96		
ZZ_STRUCTURE LONG ZZ_STRUCTURE 99 140 1 168 ZZ_PFAMGRP STRING Medicinal Product Name 95 536 68 336 mara_ZZ_PFAMGRP STRING Medicinal Product Name 98 366 556 556 ZZ_PRODFM STRING Medicinal Product Name 16 357 9 1166 ZZ_PRODFM STRING Medicinal Product Name 168 99 1240 1240 MTART STRING Matri Medicinal Product Name 168 99 1240 UMREN STRING LAEDA DATE LAEDA LAEDA LAEDA Medicinal Product Name 168 99 1240		WRKST	STRING	Medicinal Product Name	18 2408	9/ 105 96 136		
Number Num Num Number		ZZ STRUCTURE	LONG	ZZ_STRUCTURE	99 1240	11 168		
K2_FRANCIAR STRING Medicinal Product Name No No No No mara_ZZ_PFANGRP STRING Medicinal Product Name 0 0 0 0 0 ZZ_PRODFM STRING Medicinal Product Name Medicinal Product Name 188 9 1186 ZZ_PRODFM STRING Medicinal Product Name Medicinal Product Name 188 9 1240 MTART STRING MTAT Medicinal Product Name Medicinal Product Name 188 9 1240 UMREN STRING LAEDA DATE LAEDA Medicinal Product Name Medicinal Product Name Medicinal Product Name Medicinal Product Name			STRING	Medicinal Broduct Namo	9 1186	16 357 98 396		
mara_22_PAMORP STRING Medicinal Product Name 16 357 9 1186 Z2_PRODFM STRING MedicinalProductNierarchyG MedicinalProductName 166 99 1240 MTAR STRING MTAR 166 99 1240 UMREN STRING LAEDA LAEDA 166 99 1240		ZZ_FFAMORP	STRING	Wedena Product Name J	98 396	95 536		
ZZ_PRODFM STRING MedicinalProductNienarChyG MedicinalProductName 168 99 1240 MTART STRING MTART JUMREN STRING Image: String in the		mara_ZZ_PFAMGRP	STRING	Medicinal Product Name	16 357	9 1186		
MTART STRING MTART UMREN STRING LAEDA DATE		ZZ_PRODFM	STRING	MedicinalProductHierarchyG	MedicinalProductName 168	99 1240		
UMREN STRING LAEDA DATE LAEDA Run Model ®		MTART	STRING	MTART				
LAEDA DATE LAEDA		UMREN	STRING					
		LAEDA	DATE	LAEDA				Run Model 😰
	08		and the states are seen as a local					

Step 5. Review results and calibrate model (cont'd)

Oflun	😵 Classify 👻 💽 sta	art by Default search		Q					¢ ¢°0 ¢	Pharma Demo User Sign Out
	Data Sets 🔰 🖨 MARA									
습	Data Set Description SAP Material data used to retain pharma drug mate Object Tags 60 MARX Medicinal Products MedicinalProduct			Data S SAP I Data S Parqu	iource ECC iet Type iet	Total Records 273,105 Total of Columns 22	Last Mod 03/29/2 Classify	fied On 023 Ga Set	Overall quality score	100%
	Data Set Attributes	Data Set Sample Data	Set Relationships	Edit Classification Tag : Mar	nufacturerid			×		
⊞	Attribute ¢	Data Type 🗢	Classification	 You've given this mapping a Keep my selected mapping 	thumbs down.			JIL		
	T ManufacturerCode	STRING	Manufacturerld	- Or select a better mapping:				Concept ©	Mapping Confidence Level	
Ŀ	ManufacturerDesc MAKTX	STRING	Manufacturer Medicinal Produ	CLASSIFICATION TAGS				© T		
E	mara_ZZ_PLANT	LONG	ZZ_PLANT	T	Semantic Object ©	Concept c	T	Presentation S	st. Low 47% 🖞 🗘	
¥=	STRU_DESC FAMI_DESC	STRING	Agent Substan	GEN Pharma Catalog	MedicinalProduct	Strength Descentation Strength	HIGH 100%	SubstanceAm	o. LOW 47% 的 印 LOW 42% 的 印	
	SUBF_DESC	STRING	Substance Nan Strength	GEN Pharma Catalog	PharmaceuticalProduct	SubstanceAmount	LOW 47%	MedicinalProd	u. LOW 41% 🖒 🗘	
	PAFM_DESC	STRING		GEN Pharma Catalog GEN Pharma Catalog	MedicinalProduct MedicinalProduct	Manufacturer MedicinalProductHierarchyGr	LOW 42%	Manufacturen	d LOW 35% ເປັ ພາ Iu. LOW 33% ເປັ ພົ	
255	PASI_DESC MATNR	LONG	MATNR	GENI Dharma Catalon Pick a Catalog	Pick a Semantic Object	Manufacturarid Pick a Concept	1.0W 25%	Presentation S	St. LOW 0% 🖒 🏴 1	
	mara_MATNR WRKST	LONG	MATNR Medicinal Produ	O I'm Not Sure						
	ZZ_STRUCTURE	LONG	ZZ_STRUCTUR				Cancel × Save & Close			
	mara_ZZ_PFAMGRP	STRING	Medicinal Prod	ict Name		252996 5MG/100ML	1	_		
	ZZ_PRODFM	STRING	MedicinalProdu	ctHierarchyG	MG MG MG	1823 1/500MG 1160 325MG/10MG/6.25/5M 1076 G	1			
	UMREN	STRING		20 10	IOMG OMG IMG	892 1.0MG/0.2 MG/ML 866 2.5MG/20MG 822 400/57MG/5ML	1			
	LAEDA	DATE	LAEDA	50	OMG	803 4MG/100ML	1			Run Model
Oflure	Copyright © 2023 Flure	e - All rights reserved								

fluree sense

Step 5. Review results and calibrate model (cont'd)



fluree sense

Step 6. Validate Discovered Synonyms

Synonyms SYNONYM NAME: SUBSTANCE IDENTIFICATION:(15) From Original Catalog: Identification of Medicinal Products (IDMP) From Other Catalogs: Global Substance Registration System (GSR5) Catalog * Semantic Object * Concept * Synonym Catalog * Synonym Object * Synonym Concept * Sy	G C G C G G G G G G G G G G G G G G G G
SYNONYM NAME: SUBSTANCE IDENTIFICATION:(15) Form Original Catalog: Identification of Medicinal Products (IDMP) Form Other Catalogs: Global Substance Registration System (GSRS) Image: Standard Sta	×
From Original Catalog: Identification of Medicinal Products (IDMP) From Other Catalogs: Global Substance Registration System (GSRS) Image: Second	
Catalog é al Semantio Object é al Cencept é al Synonym Catalog é al Synonym Object é al Synonym	
Image: Control of Me. Substance Identification Molecyl Identifier Global Substance Registration System (GSRS) Substance Approval ID High Identification of Me Substance Identification Molecular Structure Global Substance Registration System (GSRS) Structure Molety Formula High Identification of Me Substance Identification Molecular Structure Global Substance Registration System (GSRS) Structure Molety Formula High Identification of Me Substance Identification Substance Code Global Substance Registration System (GSRS) Substance Approval ID High Identification of Me Substance Identification Nucleic Acid Substance Registration System (GSRS) Substance Approval ID High	Confidence ©
Identification of Me. Substance Identification Molecyl Identifier Global Substance Registration System (GSRS) Substance Approval ID Height Identification of Me. Substance Identification Molecular Structure Global Substance Registration System (GSRS) Structure Molecyl Formula Height Identification of Me. Substance Identification Substance Code Global Substance Registration System (GSRS) Substance Approval ID Height Identification of Me. Substance Identification Substance Code Global Substance Registration System (GSRS) Substance Approval ID Height Identification of Me. Substance Identification Substance Code Global Substance Registration System (GSRS) Substance Approval ID Height	
Identification of Me Substance Identification Molecular Structure Global Substance Registration System (GSRS) Structure Molety Formula High Identification of Me Substance Identification Substance Code Global Substance Registration System (GSRS) Substance Approval ID High Identification of Me Substance Identification Nucleic Acid Substance Registration System (GSRS) Substance Approval ID High	H 100% 🐽 1 🐶
Identification Substance Code Global Substance Registration System (GSRS) Substance Approval ID High Identification of Me. Substance Identification Nucleic Acid Substance Registration System (GSRS) Nucleic Acid UUID High	H 100% ы 1 😡
Identification of Me., Substance Identification Nucleic Acid Substance Global Substance Registration System (GSRS) NucleicAcid NucleicAcid UUD	H 100% 👍 1 🐶
	H 100% 🐽 1 😳
Identification of Me Substance Identification Polymer Substance Global Substance Registration System (GSRS) Polymer Polymer UUID HIGH	H 100% 👍 1 😱
Identification of Me Substance Identification Protein Substance Registration System (GSRS) Protein Protein UUD HIGH	H 100% 🐽 1 😳
E Identification of Me Substance Identification Substance Name Clas Global Substance Registration System (GSRS) SubstanceReference Agent Substance Name HIGH	H 100% ы 1 😱
Identification of Me Substance Identification Excipient GEN Pharma Catalog Substance d	H 100% 💼 1 😱
Identification of Me Substance Identification Molety GEN Pharma Catalog Substance SubstanceId HIGH	H 100% 🐽 1 😱
dentification of Me Substance Identification Substance Code GEN Pharma Catalog Substance d	H 100% 🐽 1 😱
E Identification of Me Substance Identification Substance Name Clas GEN Pharma Catalog Substance Marce SubstanceName HIG	H 100% ы 1 😱
dentification of Me Substance Identification Substance Name Clas GEN Pharma Catalog Substance Substance SynorymName HIG	H 100% ы 1 💭
Ventification of Me Substance Identification Substance Code GEN Pharma Catalog PharmaceuticalProduct SubstanceId HIG	H 100% 🐽 1 😱
Identification of Me Substance Identification Substance Name Clas GEN Pharma Catalog Substance OfficialName HIGH	H 100% ы 1 😡
Identification of Me Substance Identification Molecular Structure GEN Pharma Catalog Substance Substance Molecular StructureId HIG	H 100% ы 1 😱



fluree

Step 7. Publish Semantic Data Sets







V. Getting Started In Your Organization

Define Your Use Case

• Which business use case?

• Which ontology?

• Who are the users?

• Which data sources to start with?

Driver	Common Use Cases	Users	Ontology	Data Sources
LLM	 Internal knowledge mgmt Business process optimization 	 Internal transformation teams 	Business process models	 KM systems Project management systems Accounting systems
IT / OT Data Integration	 Supply Chain Optimization Smart Factory Enablement 	 Factory Operations 	 Standardized Bill of Materials 	 Sensors ERP SCADA PLCs
Internal Data Sharing	 Marketing (Upsell / Cross-sell) Corporate Functions data sharing (Risk, Finance, Compliance) 	 Business End Users Corporate Functions End Users 	 Internal Business Information Models Upper Ontologies 	 Customer Product Account Transaction
External Data Sharing	Supply Chain OptimizationIndustry Data Sharing	 Operations across external Partnerships 	 Industry Standard Ontologies 	Manufacturing data



Run a FailFast PoC

Phase I: Data Acquisition	Phase II: Classification Training	Phase III: Data Consumption
 Connect to existing systems that you can easily access Start with existing data warehouses and data lakes Stand-up the analytic infrastructure 	 Collect business information models and ontologies Collect as much Source to Target mapping information, data catalogs, documented ETL logic Import as training material for classification model Conduct focused 2-week training with Human Subject Matter Experts 	 Ingest data into Knowledge Graph Generate test queries and results that demonstrate the business use case objectives
3-6 weeks	4 weeks	3 weeks



In Closing

- We're at this interesting juxtaposition where the world is both waking up to the need for semantics, AND the tools and capabilities for creating it at scale are better than ever
- We hope we've shared practical concepts on how it can be done
- Please do not hesitate to reach out to me if you are interested further in pursuing this journey!!
- Eliud Polanco <u>epolanco@flur.ee</u>

Questions?

