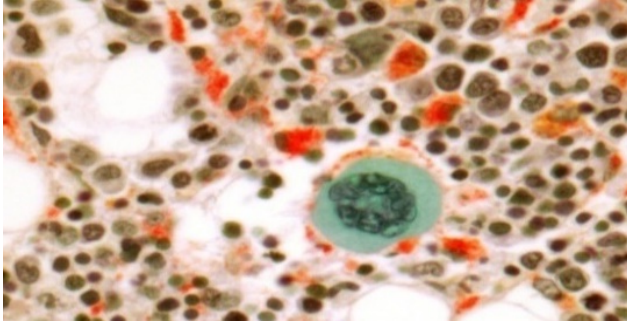


Semantic Technology Assisted Data Harmonizing and Joining Optimizes Transportation

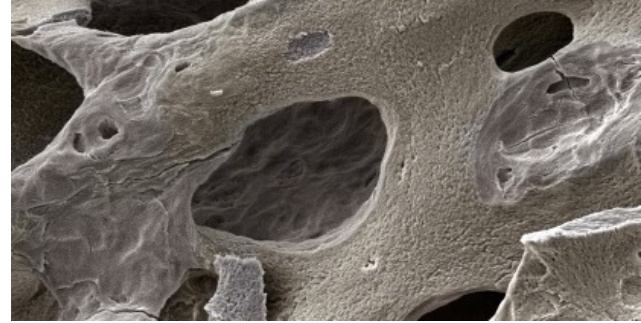
Alice Clara Augustine Ph.D.
Taxonomy Management Lead
Amgen Inc

Amgen Focus Areas

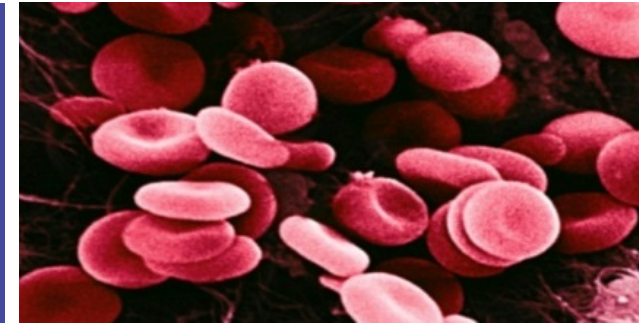
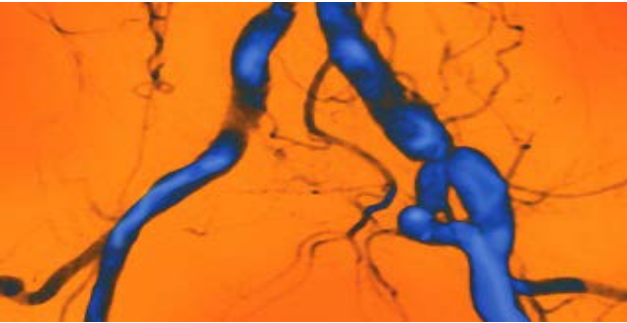
**Oncology/
Hematology**



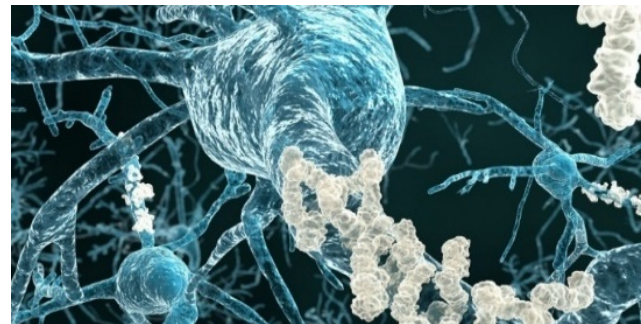
Bone Health



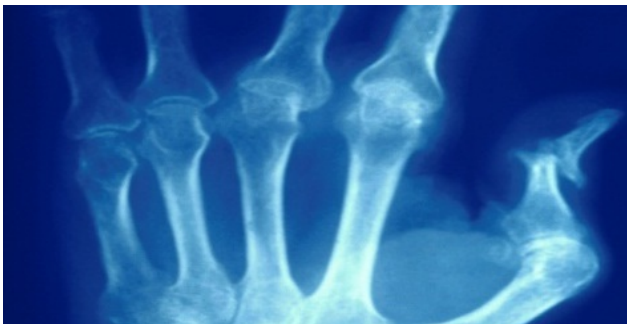
**Cardiovascular
Disease**



Nephrology



Inflammation

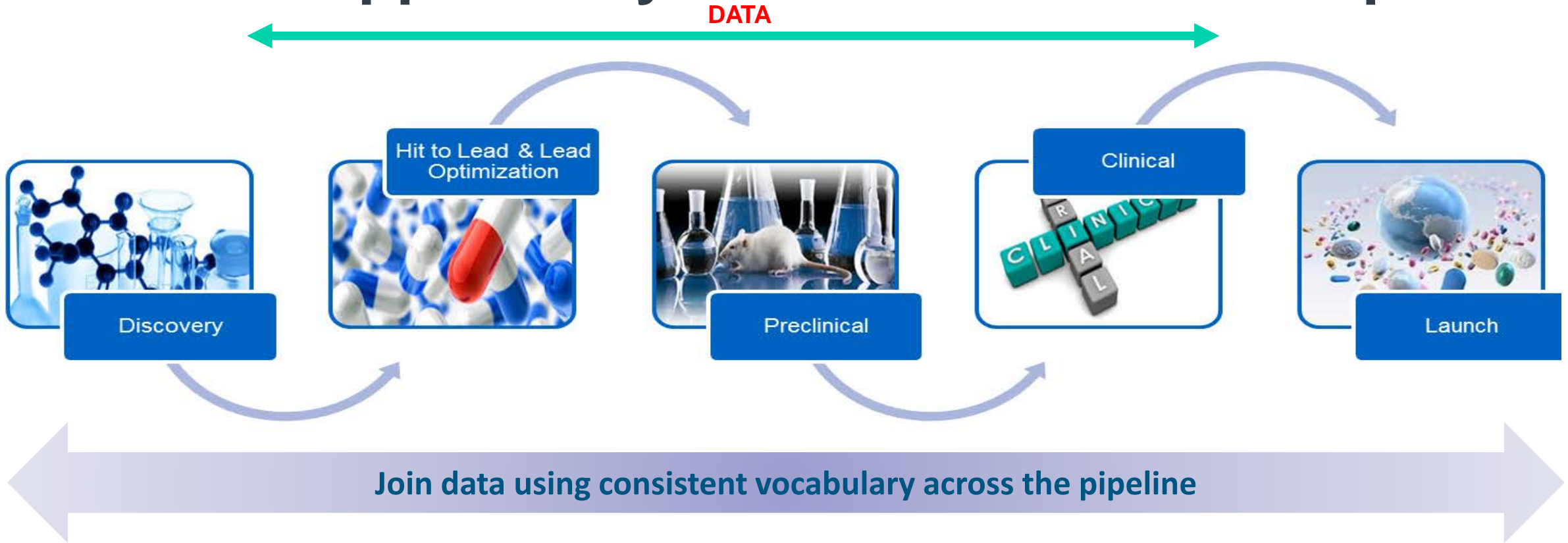


Neuroscience

Reference data – a problem across all industries

- Every industry has a reference data problem
 - Keeping reference data synchronized across all systems is difficult, time consuming and expensive
 - Different systems use different representations of reference data, leading to further confusion
 - Industry bodies work on standard reference data vocabularies, but integrating them into enterprise systems is the responsibility of individual organizations

Business Opportunity: Connected Pharma Pipeline



Bench to Bedside connectivity of data enables faster time to market

Amgen's Challenges




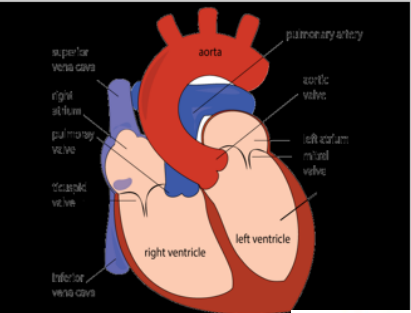
How are we at Amgen trying to change the experience of

- Finding the Data/Right Data
- Wrangling the Data
- Joining the Data

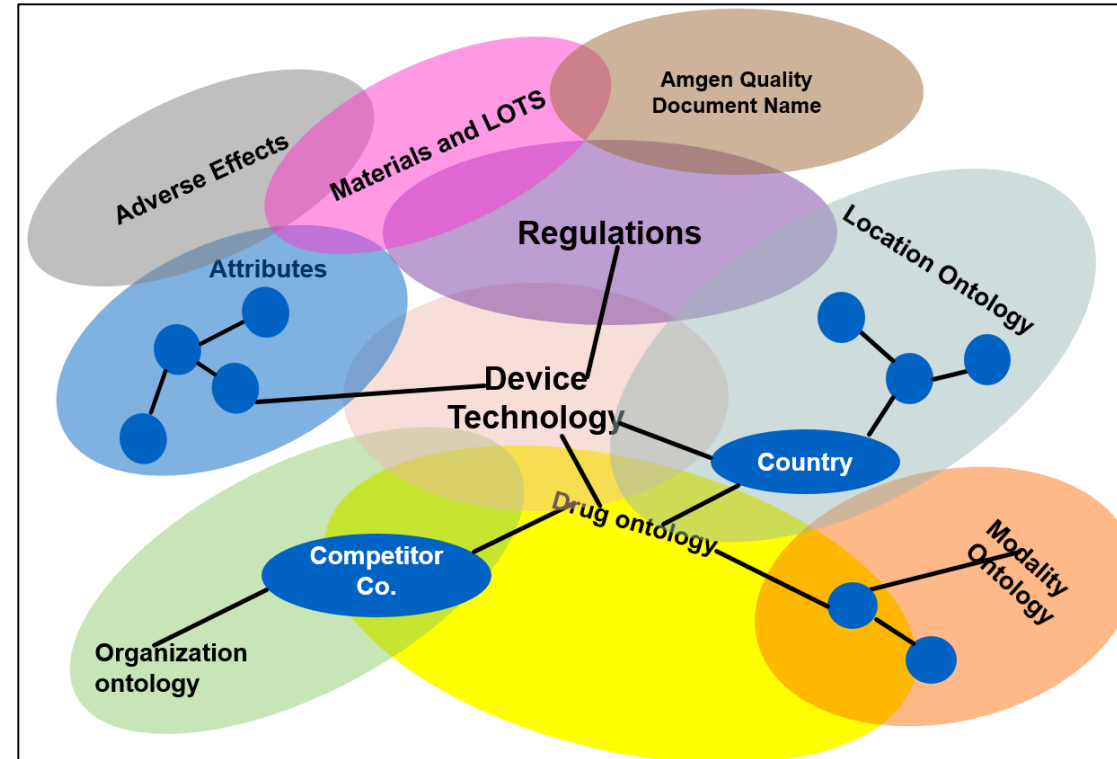
Is it a “NON FRUSTRATING
EXPERIENCE”?

Challenge statement: to provide a “Non-frustrating Data Use experience”

WHY IS IT SO DIFFICULT?

<p>So many sources, so little time...</p>	<p>Textual, Streaming and data at rest, email, mobile, image, intranet, internet</p>	
<p>Different words describe the same thing</p>	<ul style="list-style-type: none"> • Sodium bicarbonate • NaHCO₃ • Alkali • Salt • E500 	
<p>Same words describe different things</p>	<ul style="list-style-type: none"> • Site (Facility) • Site (Chromosome Location) • Site (Website) 	
<p>Domain knowledge and context are not included</p>	<p>When you are searching for something about the AORTA, it should mean you are talking about the heart.</p>	

Amgen's Novel Approach – Modular and Linked

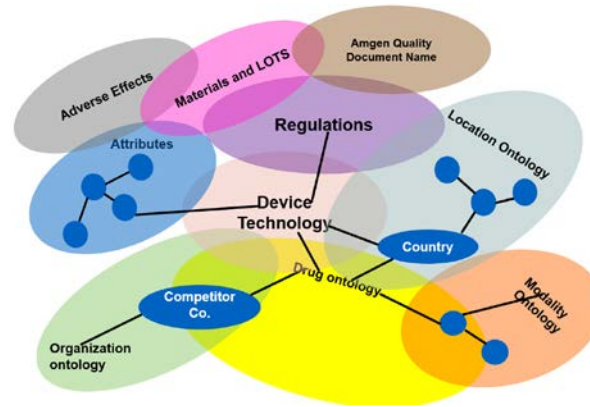


The Modular & Linked Methodology:

- Maximizes cross-functional and public data joins
- Enables cross-domain search functionality
- Provides consistent vocabulary via independent modules benefiting a localized function (when desired)
- Knowledge captured in a way that is useful for machine learning
- Empowers users to do sophisticated querying and higher quality analytics

Alice Clara Augustine@ IS ACADEMY 2017

Value of The MODULAR Linked REFERENCE DATA STRATEGY



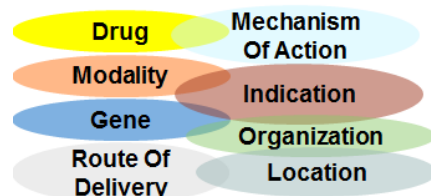
Drug Delivery Technology

What drugs are delivered using this delivery device and which company manufactures them?



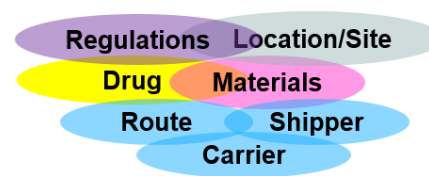
Product Portfolio Management

Are there uncovered pathways that specifically impact a particular trait?



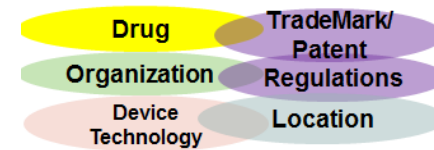
Logistics

Right package, Right time, Right quality, Right location



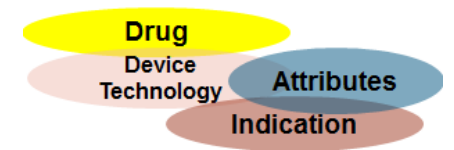
Regulatory/LAW

What are all the regulations and patents that are relevant to this drug using this delivery technology?



Translational Sciences

What are all the indications for this drug using a specific delivery technology and what are the attributes?



How Reference Data (a semantic framework) was used to join Logistics Core Data

Lane Data (Route)

- Origin
- Destination
- Carrier Type
- Carrier
- Shipper (package)
- Temperature requirement
- Duration

- *When merged with Order data,*
- *Temperature Excursion data, Stability data*
- *Security data*

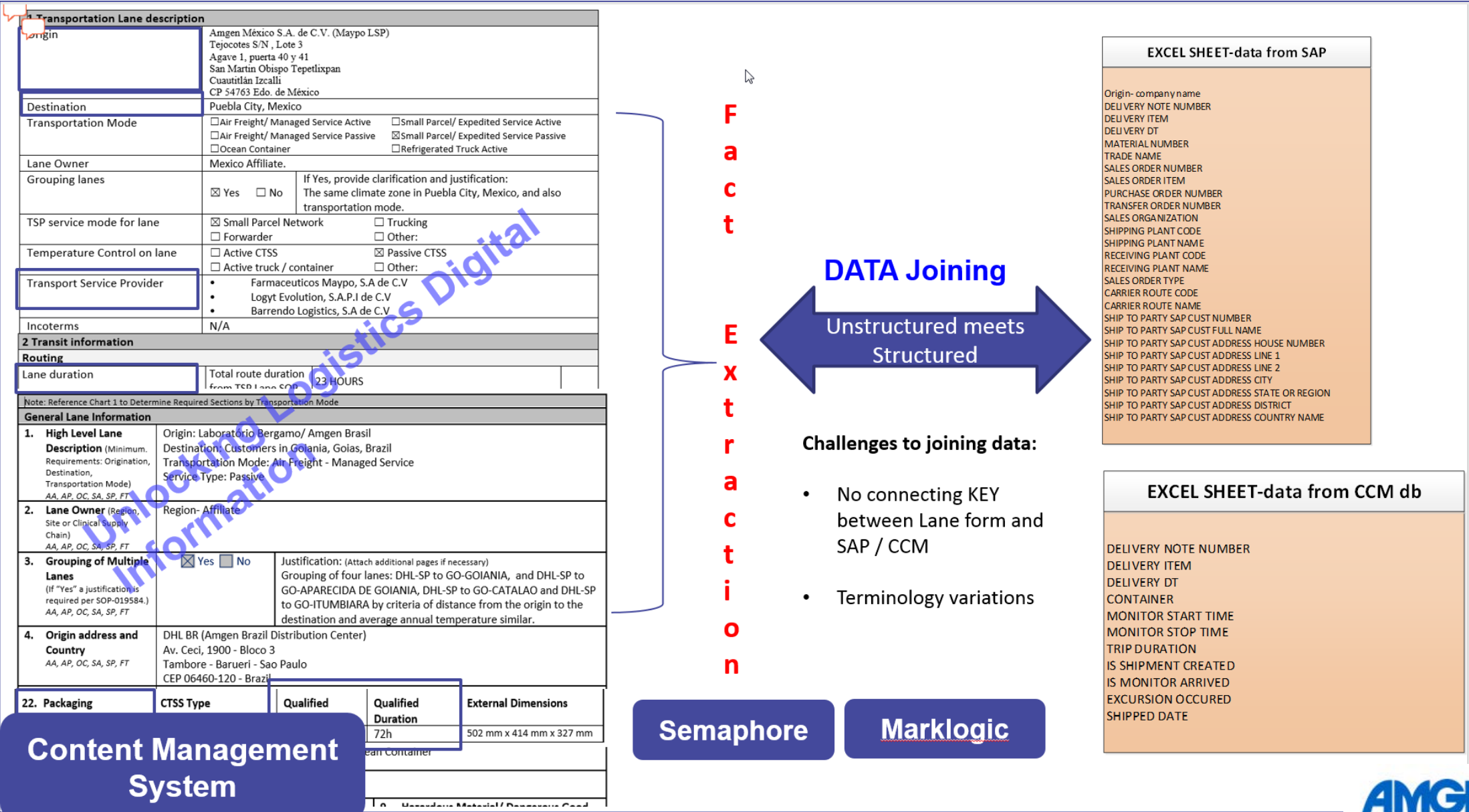
More insights on Efficiency, Cost and Risk.



Data Challenges

- Valuable treasure trove of **data lies in documents** (data for re-use)
- There are challenges **connecting different data** due to semantic gaps (i.e. different structures, inconsistent terminology, descriptions and IT-systems)
- Lack of or limited **master data oversight** across platforms resulting in poor data quality. (SAP, temperature db, Lane)
- There is **no platform to connect** Lane, Order, Temperature, Spend, or Performance information to facilitate meaningful analytics

How Amgen uses Semantic Technologies to Unlock logistics digital information



FACT EXTRACTION: CREATION OF AMGEN LOGISTICS REFERENCE DATA TO ENABLE JOINING OF DATA

Filter Model

Filter any text...

Relationships

+ Add Relationship filter

hasOriginCompany

☒ Set ☐ Not Set

Amgen Europe B.V. (ABR)

hasDestinationCompany

☒ Set ☐ Not Set

Medison Pharma Ltd

usesCarrier

☒ Set ☐ Not Set

UPS-SCS

usesShipper

☒ Set ☐ Not Set

Envirotainer RKN e1 Container

hasTransportMode

☒ Set ☐ Not Set

Air Freight

hasTemperatureControl

☒ Set ☐ Not Set

Active

Reset Submit

Filter Results:

FORM-104425a

Details History Visualizer

Class:

- Logistics Lane Reference Data
- Logistics Shipping Site
- Logistics Carrier Company
- Concept
- Shipper Model

Shipper

Carrier

Origin, Destination

```
graph TD; FORM-104425a((FORM-104425a)) -- usesCarrier --> UPS-SCS((UPS-SCS)); FORM-104425a -- usesShipper --> Envirotainer((Envirotainer RKN e1 Container)); FORM-104425a -- hasDestinationCompany --> MedisonPharmaLtd((Medison Pharma Ltd)); FORM-104425a -- hasOriginCompany --> AmgenEuropeABR((Amgen Europe B.V. (ABR))); FORM-104425a -- hasTransportMode --> AirFreight((Air Freight)); FORM-104425a -- hasTemperatureControl --> Active((Active)); FORM-104425a -- has broader --> LaneFormName((Lane Form Name)); UPS-SCS --> Carrier[Carrier]; MedisonPharmaLtd --> OriginDestination[Origin, Destination]; AmgenEuropeABR --> OriginDestination;
```

FORM-104425a

Comments Share

Class

Preferred Label

Metadata

Understanding Data, Data Mapping, Semantic joins

SAP DATA	DELIVERY NOTE NUMBER	DELIVERY ITEM	DELIVERY DT	MATERIAL NUMBER	TRADE NAME	SALES ORDER NUMBER	SALES ORDER ITEM	PURCHASE ORDER NUMBER	TRANSFER ORDER NUMBER	SALES ORGANIZATION	SHIPPING PLANT CODE	SHIPPING PLANT NAME	RECEIVING PLANT CODE	RECEIVING PLANT NAME	SALES ORDER TYPE	CARRIER ROUTE CODE	CARRIER ROUTE NAME	SHIP TO PARTY SAP CUST NUMBER
	86500600	10	1/5/2017	9001405	ARANESP	5364826	10	5364826	2662960	S030	P001	Amgen Europe BV	N/A	Medison Pharma	ZOR	SCSACU	?	10070921
CCM DATA	DELIVERY NOTE NUMBER	DELIVERY ITEM	DELIVERY DT	CONTAINER	MONITOR START TIME	MONITOR STOP TIME	TRIP DURATION	IS SHIPMENT CREATED	IS MONITOR ARRIVED	EXCURSION OCCURED	SHIPPED DATE							
	86500600	10	1/5/2017	ENVIROTAIRER RKN, SOP-014363	#####	1/8/2017 13:28	4 Days 0 Hrs 20 Minutes	T	T	No	1/4/2017							
LANE FORMS	EDMQ#	Origin-company name	Origin Country	Origin address and Country	Destination Country	Destination - company name	Destination address and Country	Transportation Mode	Lane Owner (data from Form)	TSP service mode for lane (data from Form)	Temperature Control	Transportation Service Provider	Incoterms (data from Form)	Lane Duration (data from Form) hrs	Preferred ship date	Calendar restrictions- DESTINATION	CTSS type / also	Shipper Name
	FORM-104425a	Amgen Europe B.V. (ABR)	NL	Amgen Breda Minervum 7061 4817 ZK BREDA The Netherlands	IL	Medison Pharma Ltd	Medison Pharma Ltd 20 Hamagshimim	Airfreight Managed Services	Marga Eestermans	Forwarder	Active truck / container	UPS-SCS	CIP (Cost Insurance Paid) to	37	Tue-Wed-Thur	Destination Israel, weekends :	Envirotainer RKN e1	Envirotainer RKN e1 Container
	FORM-104425b	Amgen Europe B.V. (ABR)	NL	Amgen Breda Minervum 7061 4817 ZK BREDA The Netherlands	IL	Medison Pharma Ltd	Medison Pharma Ltd 20 Hamagshimim	Airfreight Managed Services	Marga Eestermans	Forwarder	Active truck / container	UPS-SCS	CIP (Cost Insurance Paid) to	52	Tue-Wed-Thur	Destination Israel, weekends :	Envirotainer RKN e1	Envirotainer RKN e1 Container
	FORM-092940	Amgen Europe B.V. (ABR)	NL	Amgen Breda Minervum 7061 4817 ZK BREDA The Netherlands	IL	Medison Pharma Ltd	Medison Pharma Ltd (DOOR TO DOOR)	Air Freight/ Managed Service	Region EU	Forwarder	Passive : CTSS	DSV	Door to Airport DTA TEL AVIV (MEDISON)	52			Kalibox All sizes	
	FORM-105077	Amgen Europe B.V. (ABR)	NL	Amgen Europe B.V. Minervum 7061, Breda 4817 ZK	IL	Trialog Clinical Trials Ltd.	Trialog Clinical Trials Ltd. Department of Cardiology	Airfreight/ Managed Service	ABR Clinical Supply Chain	Forwarder	Passive CTSS	DSV	dap	87	Friday and Monday	Business days only. (Sunday till Thursday)	Chronos 12L Chronos 90L	

- In-consistent data field labels
- In-consistent Reference Data
- Data gaps

Origin	
Destination	
Carrier	
Shipper	



Complex Key

REFERENCE DATA CREATION TO ENABLE JOINS

ups-s

sub-contracted carrier RTT %
TAMRO %
Tee Hai Chem Pte Ltd %
Thermomed %
TOTE %
Unitrans %
UPD %
UPI-Internat(Not PR) %
UPP- Red OvernightPR %
UPS %
UPS (AHEC) %
UPS (ATS) %
UPS (CEVA) %
UPS (Loomis) %
UPS (Purolator) %
UPS Parcel %
UPS WW EXPRES EXPEDI %
UPS-SCS %
Van Dijk %
Vilar Trucking / TOTE Maritime CRST-Pegasus %
World Courier %
Zuellig Pharma Korea %

12

Logistics Document Model

Logistics Lane Model

International Commercial Terms

Lane Form Name

123...18

FORM-088396
FORM-089974a
FORM-089974b
FORM-089974c

UPS-SCS

Concept Class

Logistics Carrier Company

Preferred Labels

[Create a preferred label](#)

UPS-SCS en

Alternative Labels

[Create an alternative label](#)

alternative label > SCS - UPS - Cooling en

alternative label > SCS - UPS SCS en

alternative label > SCS UPS Cargo (INT.) en

alternative label > SCSACU en

alternative label > UPS SCS en

alternative label > UPS Standard ABR en

alternative label > UPS_SCS_Clinical en

Metadata

[Add metadata field](#)

Top Concept Of

[Select a concept scheme](#)

Related Concepts ³²

[Select a related concept](#)

123

isCarrierUsedBy > FORM-093898b

isCarrierUsedBy > FORM-094176

isCarrierUsedBy > FORM-095508

isCarrierUsedBy > FORM-095777b

isCarrierUsedBy > FORM-096111a

isCarrierUsedBy > FORM-096111b

isCarrierUsedBy > FORM-096111c

isCarrierUsedBy > FORM-096111d

isCarrierUsedBy > FORM-096111e

isCarrierUsedBy > FORM-096112a

isCarrierUsedBy > FORM-096112b

isCarrierUsedBy > FORM-096112c

isCarrierUsedBy > FORM-096112d

123

Broader Concepts ¹

[Select a broader concept](#)

has broader > Logistics Carrier Company

Narrower Concepts ⁰

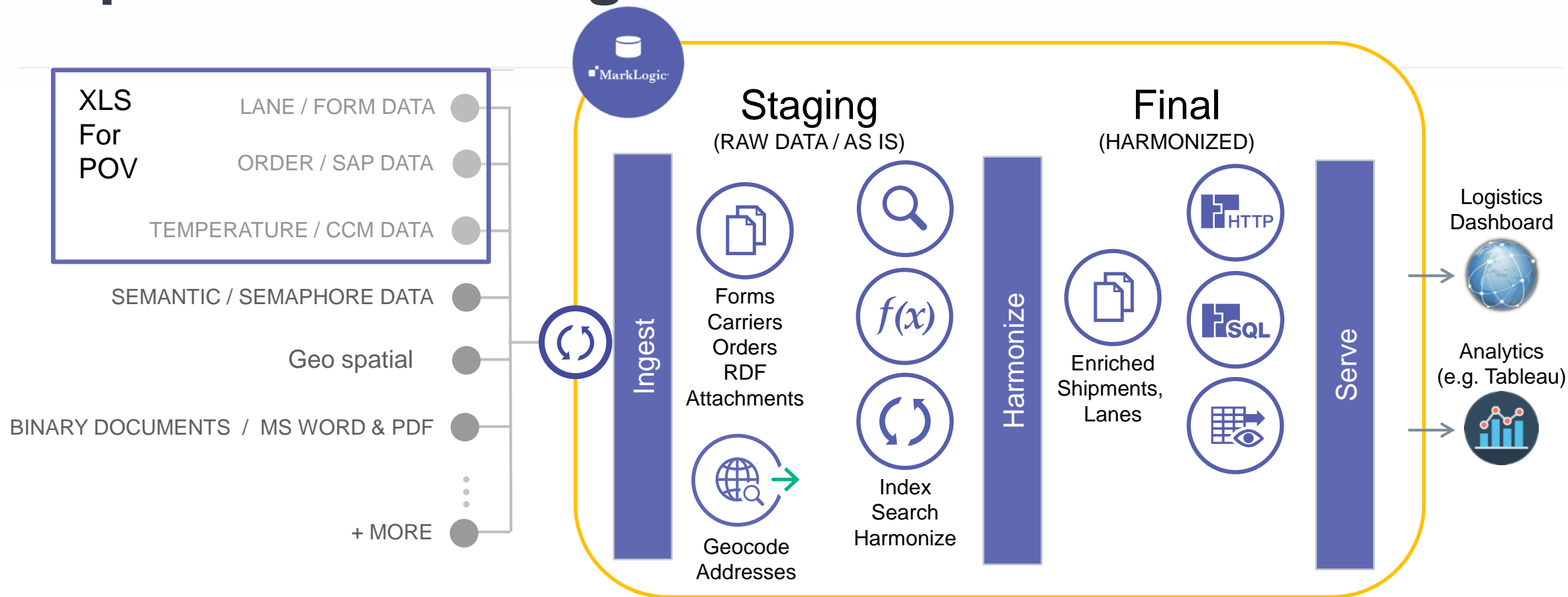
DATA JOINING, HARMONIZING

Score based indexing used for matching (a feature of Smart Mastering) in MarkLogic

Description	Weight	Use reference data
match originName; SAP - "SHIPPINGPLANTNAME" = Lane - "Origin- company name", "Origin address and Country"	2	yes
match destination name SAP - SHIPTOPARTYSAPCUSTFULLNAME = LANE - Destination -company name	2	yes
match carrier route code; SAP - Carrier Route Code = LANE - "Transportation Service Provider"	4	yes
match container; CMS - Container = LANE - "Shipper Name" + "EDMQ-SOP"	4	no
match SAP destGeo to Lane destGeo, geospatial query, match if within 5 miles	14	no
match SAP originGeo to Lane originGeo, geospatial query, match if within 5 miles	14	no
	40- total	

Score	Label
>=40	
>=34	
>=28	
>=24	
23 or less	Reject

Representative Target State



■ Changes often

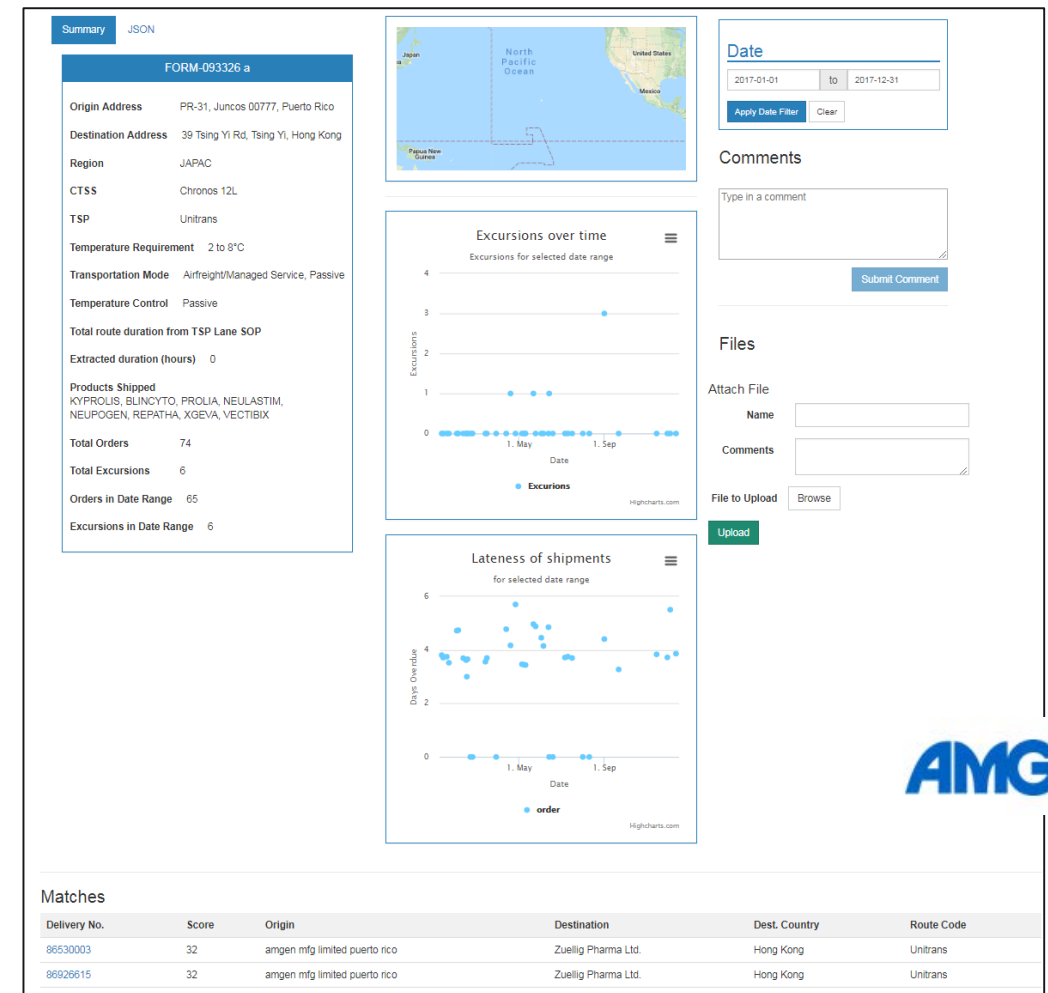
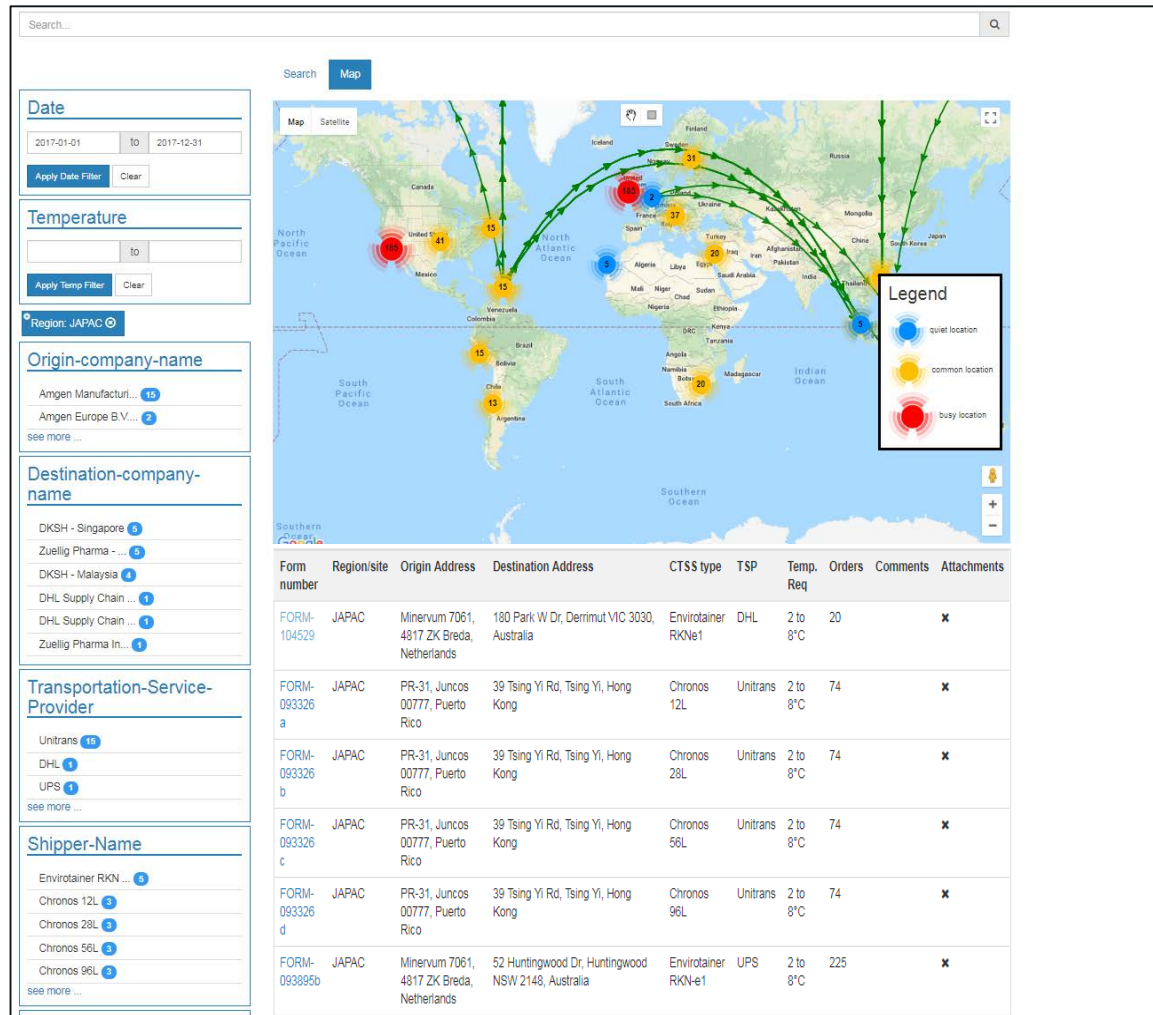
■ This part needs to respond to rapid changes and remain agile

■ Changes often

Technical solution design highlighting the sources of data, challenges encountered, how the gap/pain points were addressed using a combination of MarkLogic and Smartlogic

CHALLENGE	Technology solution	POV data examples
Challenge of multiple data silos, formats, ownership	Ingest any data source “as-is” using MarkLogic (mlcp) and open source (Apache NIFI) tools Multi-model (different types of data models)	SAP, temp data and Lane form data
Challenge of data synchronization and semantic relationship	Enrichment Services – Normalizing language- Semaphore taxonomy/ontology creation and management. MarkLogic triplestore, Semantic Triple creation to normalize table names, SPARQL/Javascript/XQuery programmatic access	Taxonomies for Carriers, Shippers, Amgen Sites, Transport Service Provider, Transport type, etc Semantic triples created for labels of columns (Carrier Route Code SameAS Transport Service provider)
Master data challenges- Customer (sites, names, addresses)	Enrichment Services – via Geoservices	Challenge of Addresses (since we didn’t have uniform company names, nor did we have proper addresses)
Fact Extraction	Semaphore	Extracted all fields from EDMQ FORMS holding textual information using a document Schema.
Matching / JOINING various fields across SAP, CCM and Lane Data	Score based indexing used for matching (a feature of SmartMastering)	We had to match data across 6 data points (each data type had a score) e.g. origin match geo-address within 5 miles 14pt e.g. destination match address within 5 miles14 pt e.g. Carrier Name 4 e.g. Shipper Name 4

THE PoV PROVIDES THE SEMANTIC LAYER, HARMONIZATION OF DATA, VISUALIZATION DASHBOARD, AND GOOGLE LIKE SEARCH



Logistics Intelligence Platform

Search

Semantic Indexing

Geospatial Integration

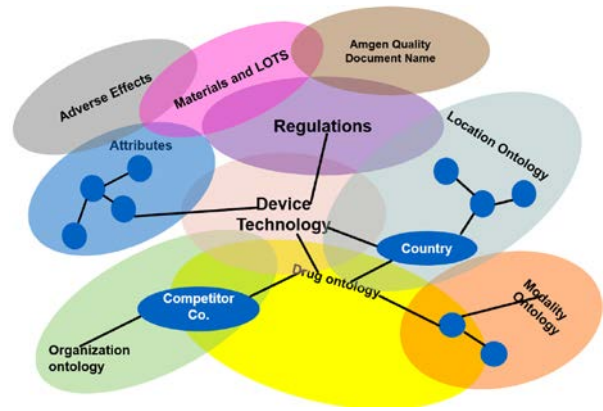
Integration with a reporting tool Tableau

Upload of Documents/Photos (Complaints etc.)

Lane Score Card

Shipper Score Card

Solving multiple problems - Using Multi-model approach and Modular Linked Reference Data



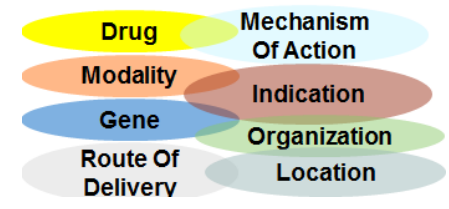
Drug Delivery Technology

What drugs are delivered using this delivery device and which company manufactures them?



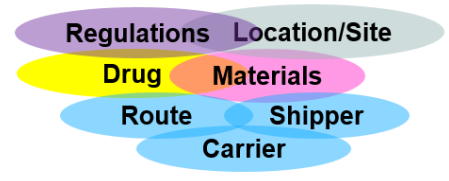
Product Portfolio Management

Are there uncovered pathways that specifically impact a particular trait?



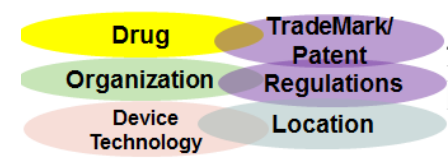
Logistics

Right package, Right time, Right quality, Right location



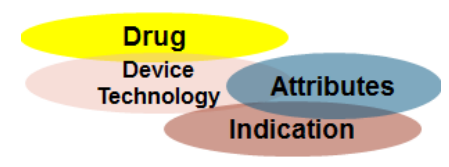
Regulatory/LAW

What are all the regulations and patents that are relevant to this drug using this delivery technology?



Translational Sciences

What are all the indications for this drug using a specific delivery technology and what are the attributes?



THE TEAM:

Clare Augustine	Semantic technologist
Kelley VanArsdale	Supply Chain Senior Manager
Lisa Baker	Director Supply Chain
Nick Haycocks	QA Sr Manager

Mark Ferneau	MarkLogic team (super brilliant, awesome collaborators)
Greg McFaul	
Nish Bakshi	
Michael Malgeri	