intuit turbotax quickbooks mint

Data Mesh at Intuit

May 13, 2021

Tristan Baker - linkedin.com/in/tristanbaker
Suresh Raman - linkedin.com/in/ramansuresh
Allison Bellah (in absentia) - linkedin.com/in/allisonbellah

Agenda

Arriving at data mesh

Our vision and four part strategy

Now with 25% more parts!

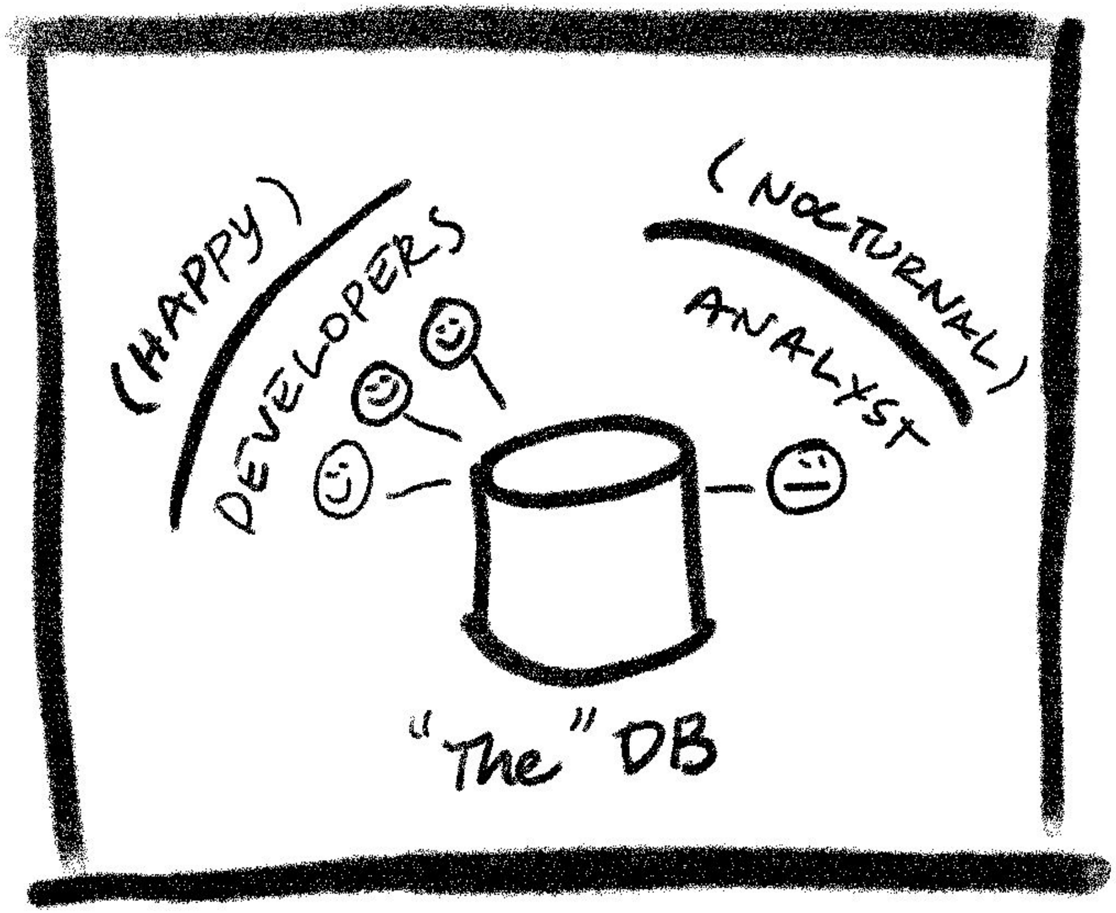
Q&A

Fire away!

Arriving at data mesh

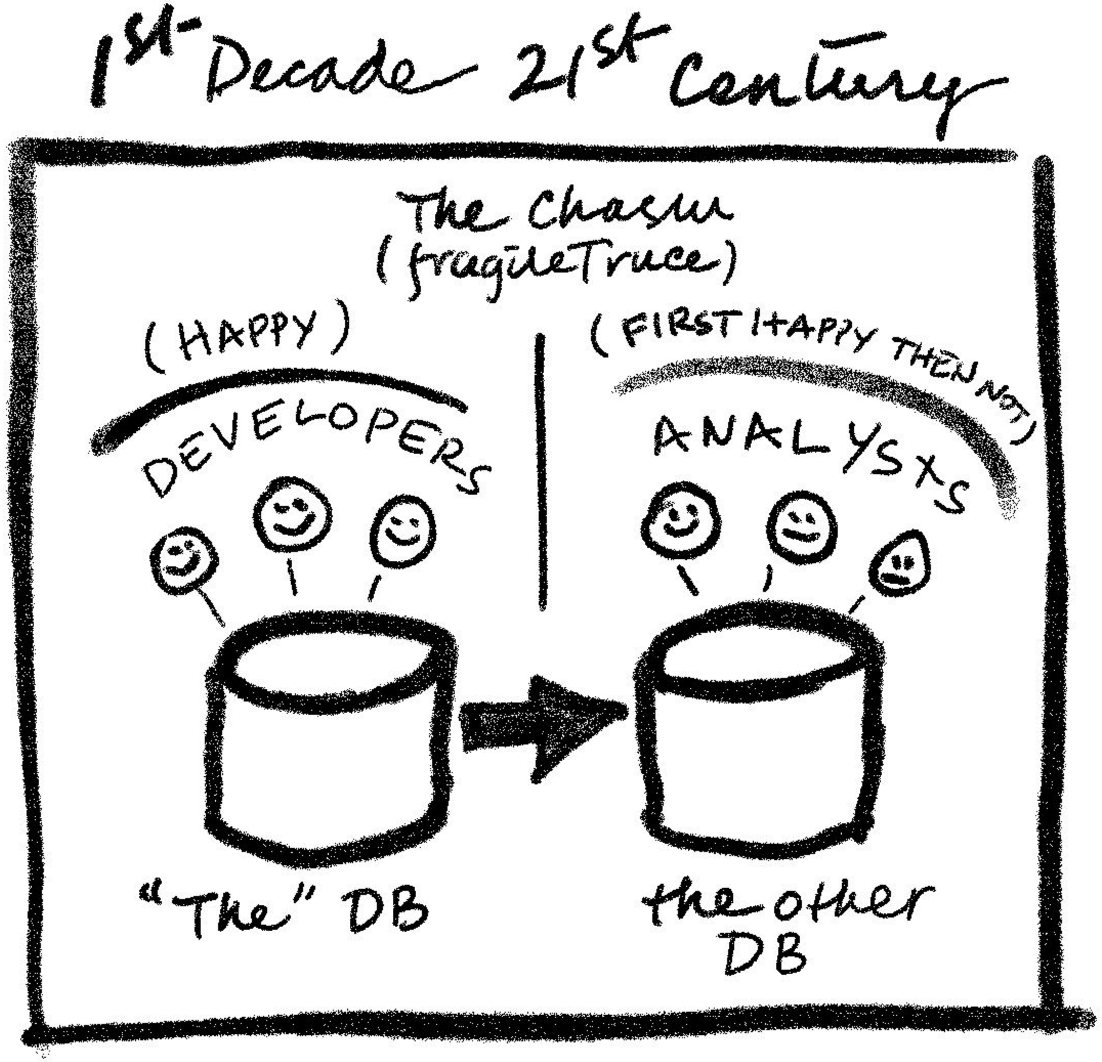
A brief history of data infrastructure

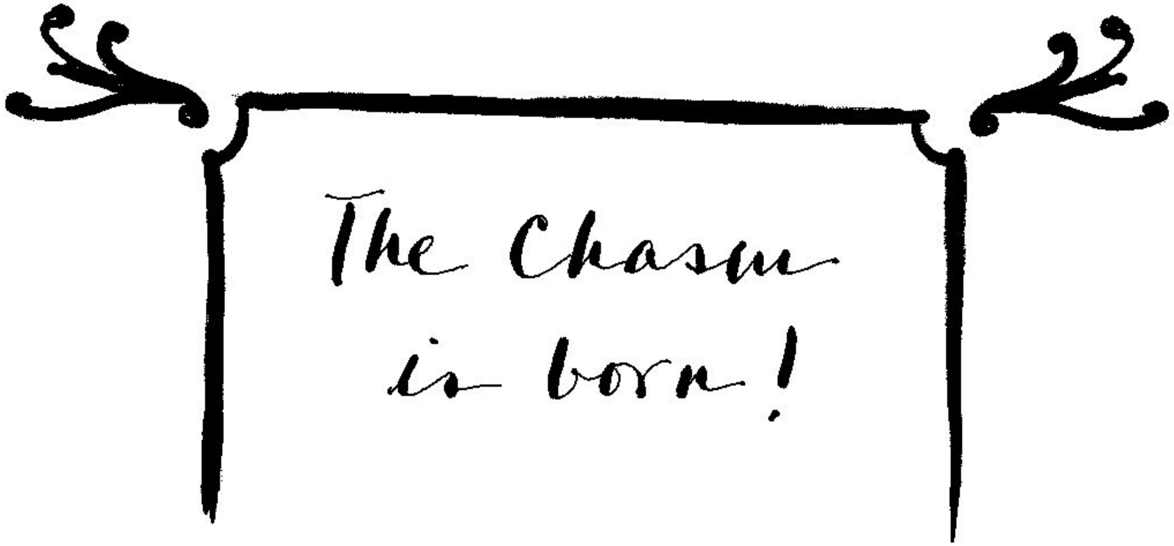
Late 20th Century



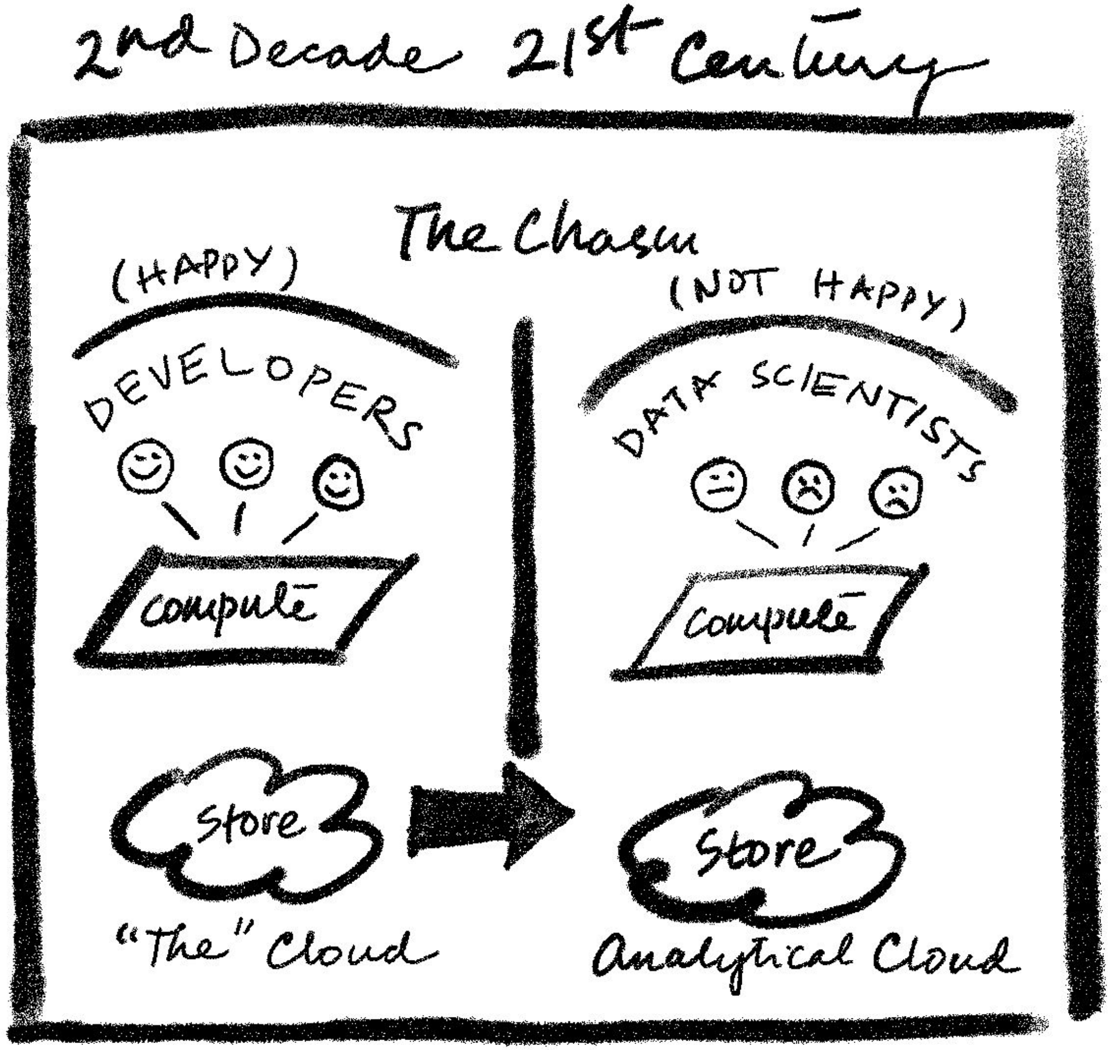
Had no idea what was about to hit us !!

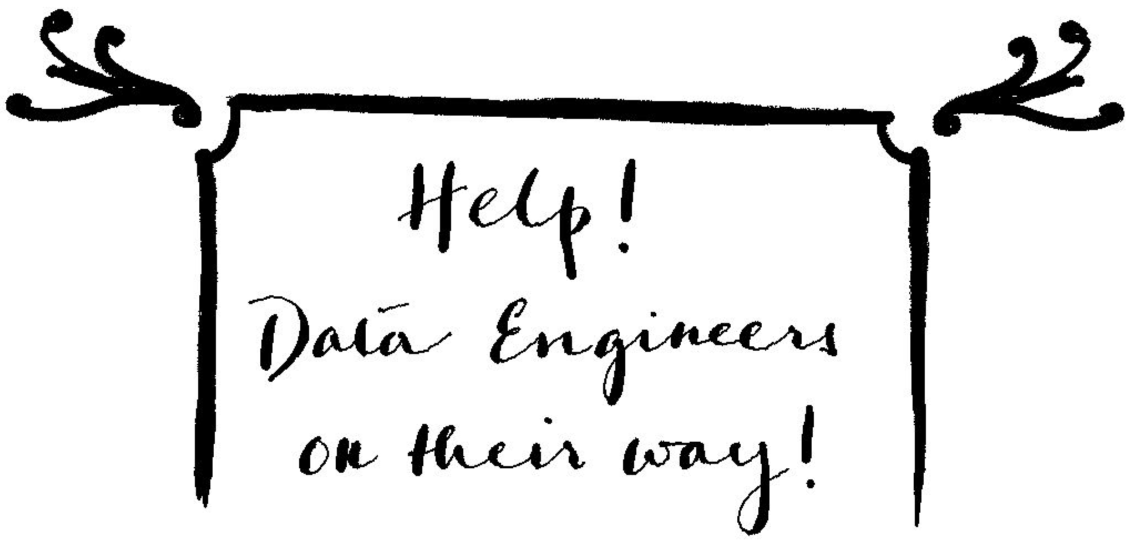
A brief history of data infrastructure





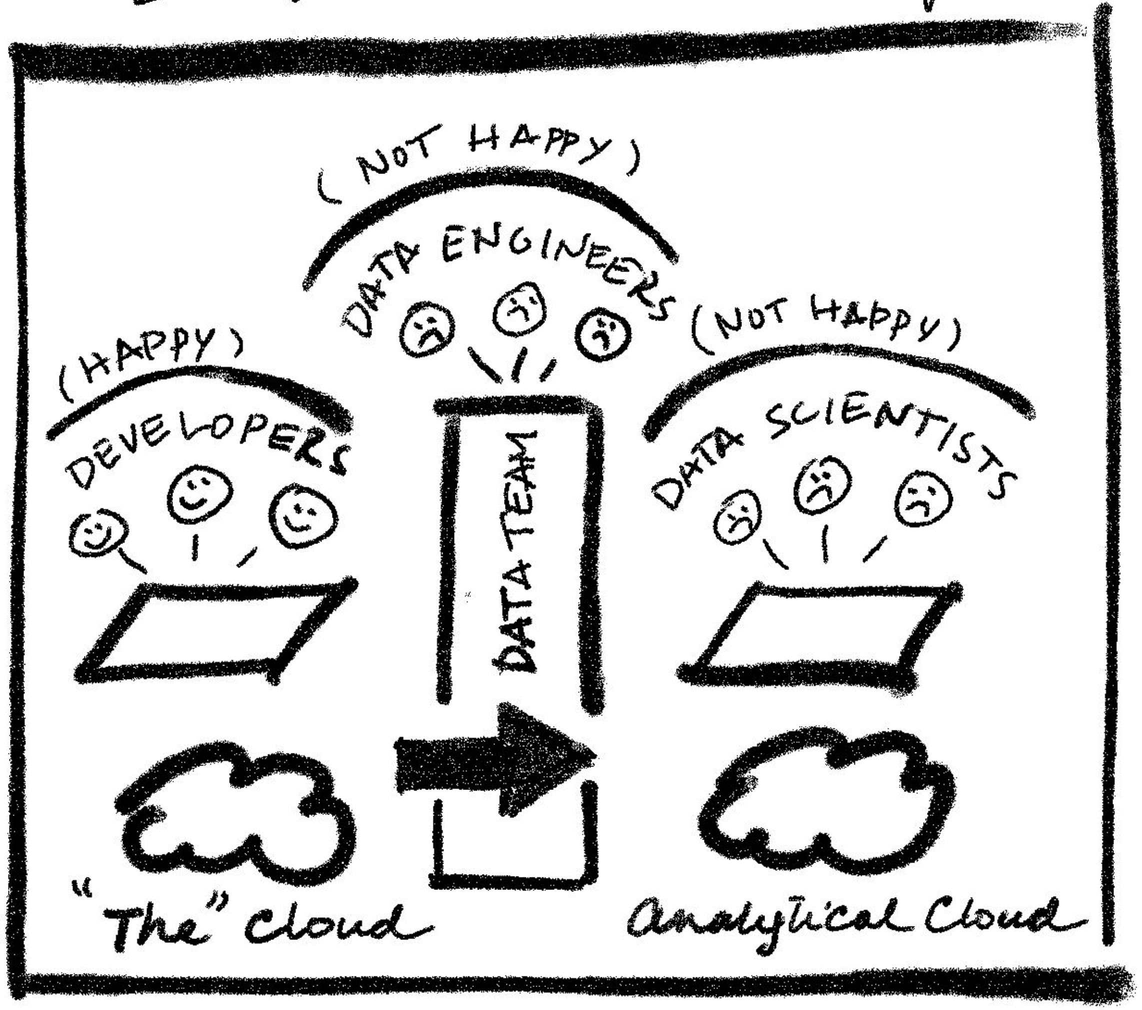
A brief history of data infrastructure

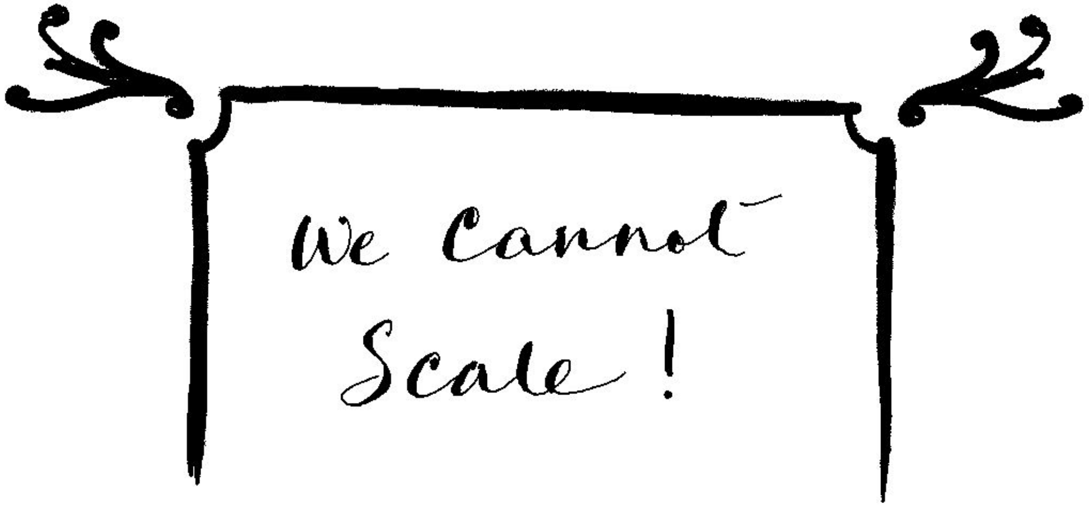




Today

2nd Decade 21st Century




We Cannot Scale!

What “we cannot scale” sounds like from our users

Discovering Data

- Where can I find data about a particular thing (customer, company, etc)?
- Where can I find the data sourced from a particular product or service?

Understanding Data

- Who can approve my access so that I can see samples of the data?
- What is the schema of the data?
- What is the business meaning and context of the data?
- Is this data related to other concepts? Is it joinable to other data? What is the meaning of the relationship?

Trusting Data

- What system produces this data and at what latency?
- What other systems use this data?
- What is the quality of this data? Is it ‘clean’?
- Which team supports this data if it breaks?

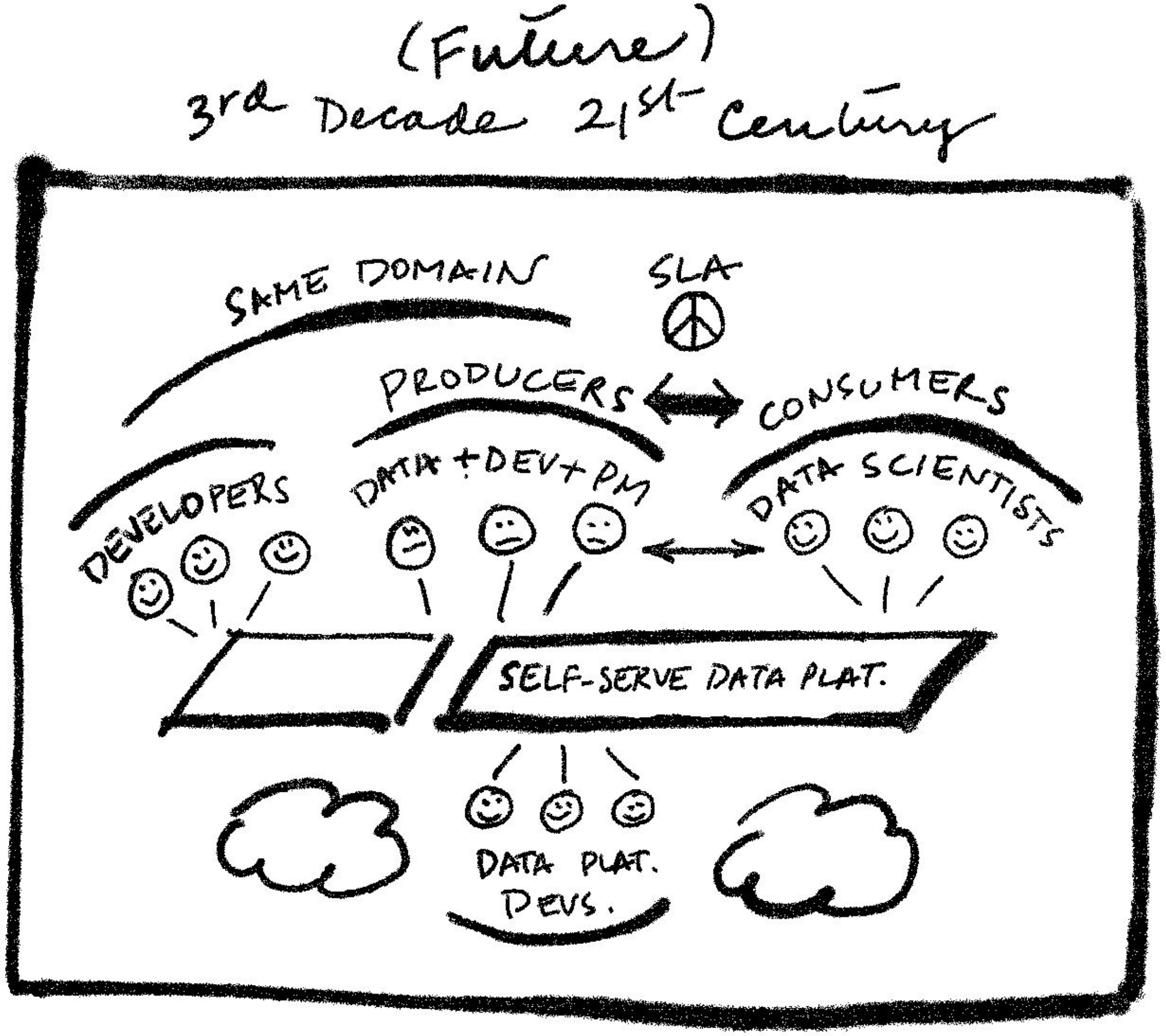
Consuming Data

- How is this table/topic partitioned?
- Who can approve my production system to access it?
- Will I get alerted if the schema changes?

Publishing Data

- How do I describe my data so that others understand what it means and how to use it?
- Where do I host my data so that other systems can access it?
- Data systems are complicated, how can I build and operate my process on top of one?
- What are my operational responsibilities once my process/data is in production?
- How do I meet my compliance requirements for processing/storing/publishing data?
- Am I duplicating processing/data that already exists?

The future of data infrastructure



The provocation

- Data treated as code
- Data service as a facet of a product
- Data responsibility decentralized
- Producers take responsibility for data
- Producers serve consumers
- Data platform provides the ecosystem to govern and manage the lifecycle of data and machine learning

Data Mesh is born

Our vision and four part strategy

Enable more Intuit teams to more easily use and create data

Four part strategy

- Stewardship
  - ensures accountability for a set of defined responsibilities in building and managing their solutions; including adherence to a set of defined best practices to produce only high quality data.
- Organizing people, code and data
  - A systematic approach to organizing the people, code and data which clearly identifies the owners of a business problem and its solution.
- Self serve products
  - A rich suite of self serve products that enable teams to more easily author, deploy, govern and operate their own solutions, aided by automation and processes that support best practices and high quality as a precondition for deployment.
- Rationalizing data definitions
  - A process for rationalizing all critical data definitions at the company so that data concepts like Customer, Product and Entitlement are unique, re-usable and non-conflicting.

Stewardship

One Intuit Account Management CROSS ECOSYSTEM

FROM

- Customers must update their account info across products
- Customers can't view all the Intuit products they manage in a single place
- Frustrated customers make several MM unnecessary customer support calls
- Central data engineering team tries to put it all back together with little involvement form Intuit Account services team.

TO

DATA API

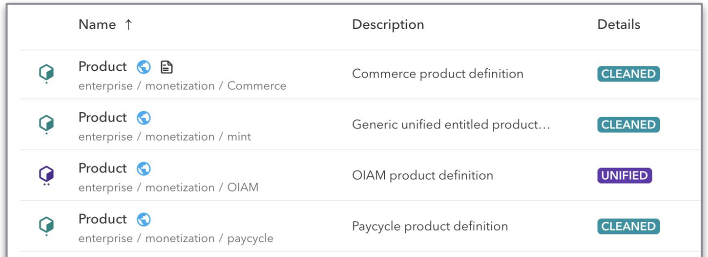
Billing statements | Product | Payments | Product feature set | Wallet reference | Refunds | Charge item | Customer account

DOMAINS

- Billing
- Monetization

TEAM

- Account
- Data Eng



Name ↑	Description	Details
Product enterprise / monetization / Commerce	Commerce product definition	CLEANED
Product enterprise / monetization / mint	Generic unified entitled product…	CLEANED
Product enterprise / monetization / OIAM	OIAM product definition	UNIFIED
Product enterprise / monetization / paycycle	Paycycle product definition	CLEANED

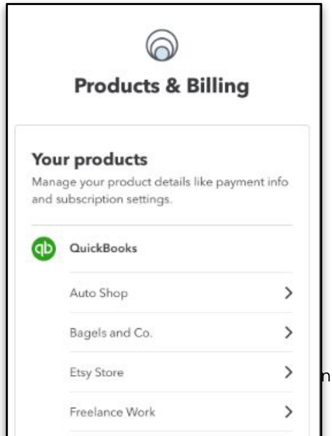
IMPACT

CUSTOMER

Single ecosystem experience

BUSINESS

- Increased developer productivity
- Accurate customer records
- In FY21, $$ savings from fewer customer support calls



Products & Billing

Your products

Manage your product details like payment info and subscription settings.

QuickBooks

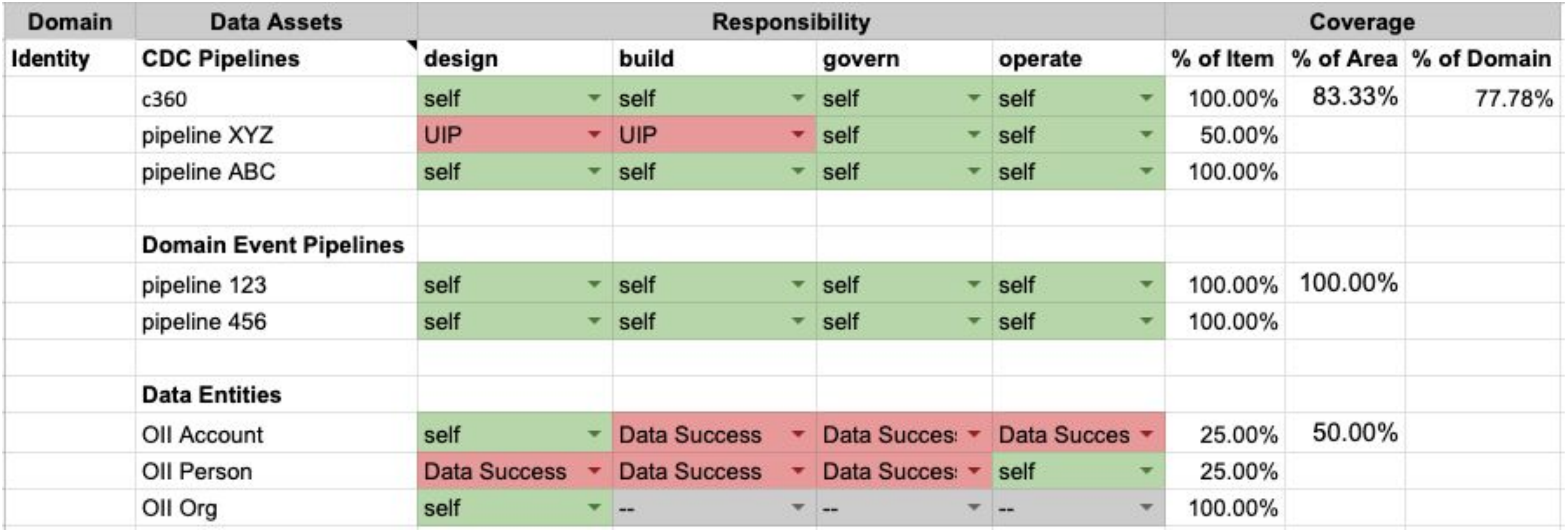
Auto Shop ›

Bagels and Co. ›

Etsy Store ›

Freelance Work ›

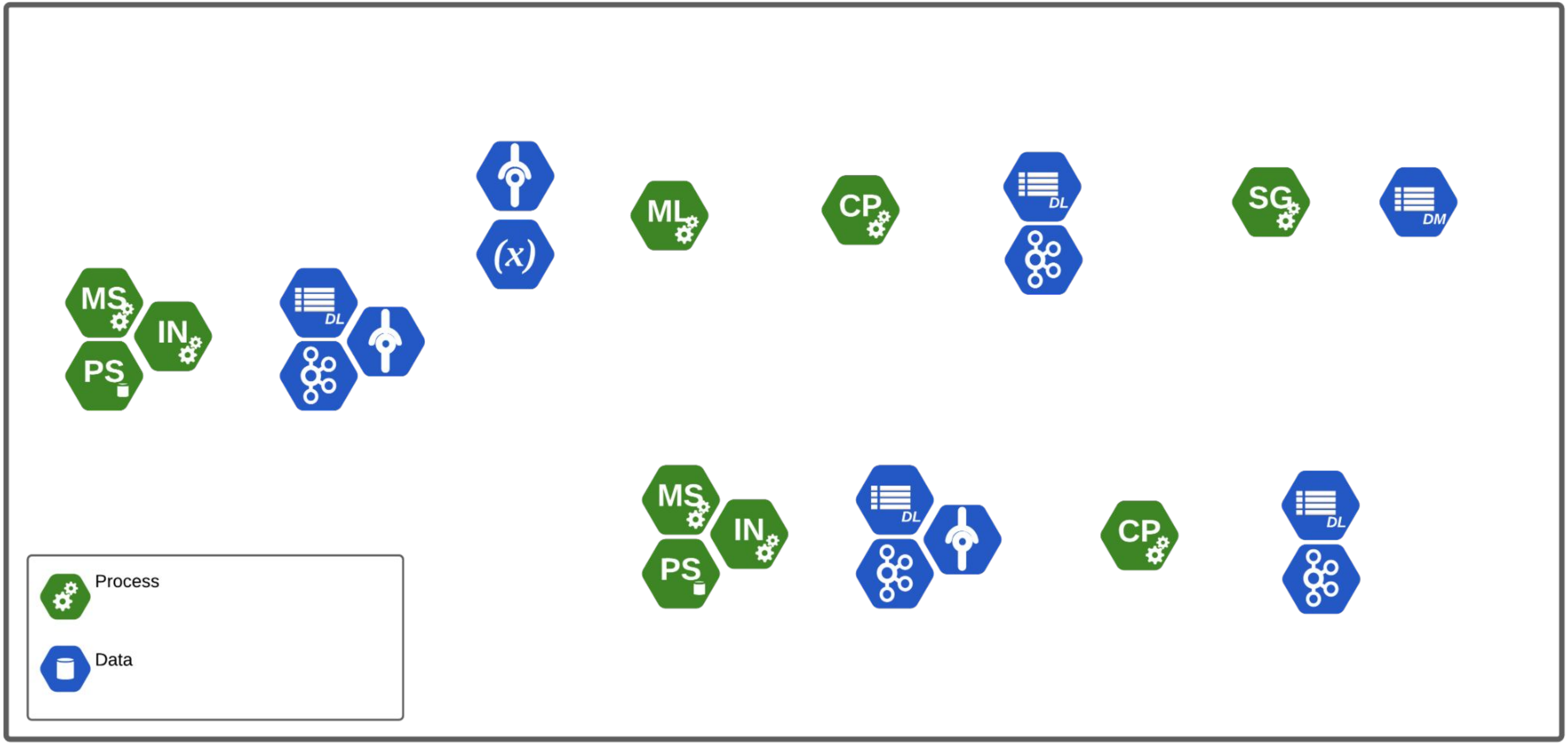
Stewardship goals for next year



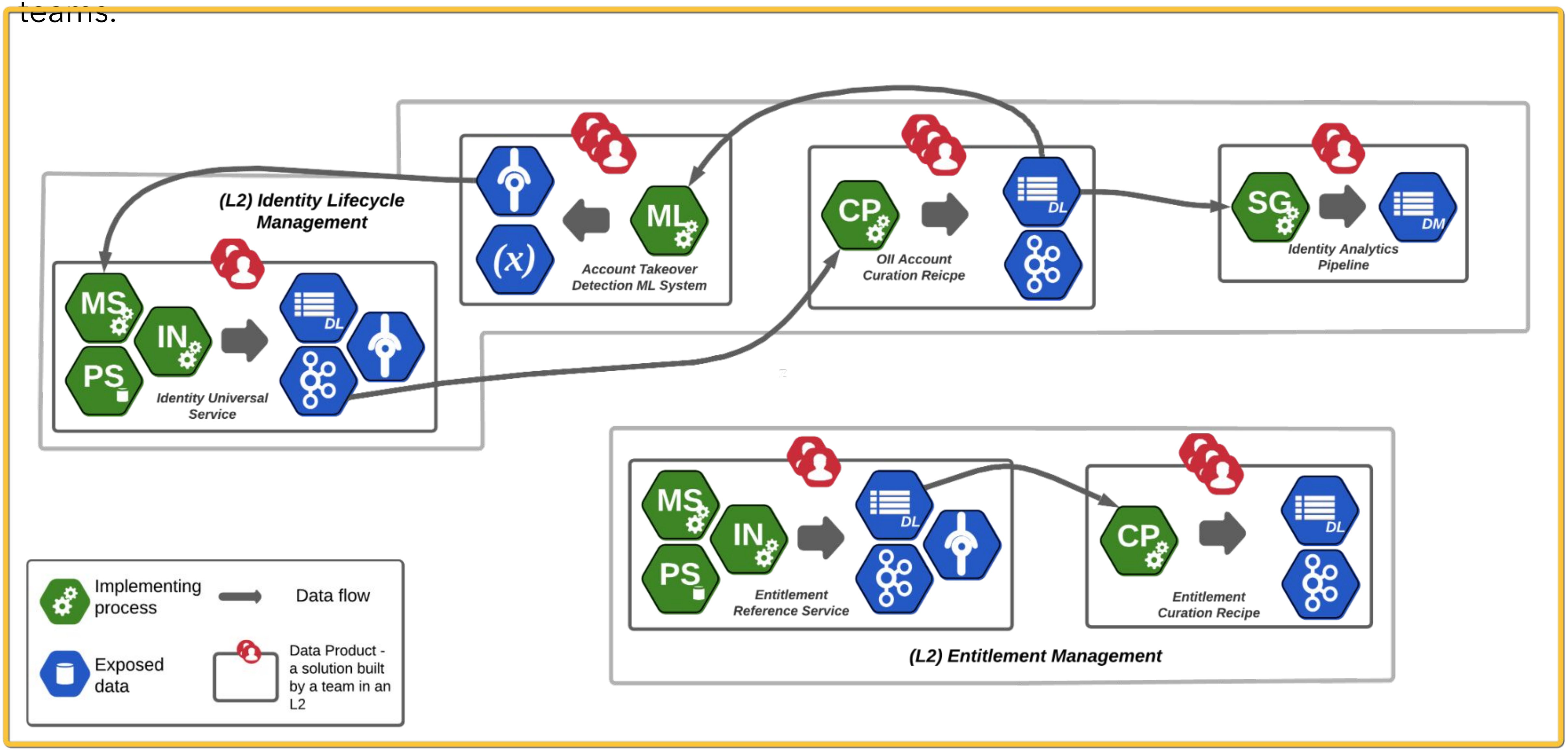
Domain	Data Assets	Responsibility				Coverage		
Identity	CDC Pipelines	design	build	govern	operate	% of Item	% of Area	% of Domain
	c360	self	self	self	self	100.00%	83.33%	77.78%
	pipeline XYZ	UIP	UIP	self	self	50.00%		
	pipeline ABC	self	self	self	self	100.00%		
	Domain Event Pipelines							
	pipeline 123	self	self	self	self	100.00%	100.00%	
	pipeline 456	self	self	self	self	100.00%		
	Data Entities							
	OII Account	self	Data Success	Data Succes:	Data Succes	25.00%	50.00%	
	OII Person	Data Success	Data Success	Data Succes:	self	25.00%		
	OII Org	self	--	--	--	100.00%		

Organizing People, Code, and Data

Raw information about physical systems that describes where the data is stored and where code is executing. This describes where data is physically located so that it can be accessed.



Basic dependency, ownership and classification information provides additional context about physical data and code locations so that data can be better governed, secured and operated by the owning teams.



(L2) Identity Lifecycle Management
MS
IN
PS
DL
Identity Universal Service
(x)
ML
Account Takeover Detection ML System
CP
DL
OII Account Curation Reicpe
SG
DM
Identity Analytics Pipeline
MS
IN
PS
DL
Entitlement Reference Service
CP
DL
Entitlement Curation Recipe
(L2) Entitlement Management
Implementing process
Data flow
Exposed data
Data Product - a solution built by a team in an L2

Why organizing people, code and data matters

Private vs Public
~50% tables are either temp/sandbox/staging/test/backup tables

- Messes up search & discovery
- Teams consume data not meant for external use

Data Ownership
~50% tables don't have clearly identified owners

- Erodes Trust
- Copies proliferate
- Operational, Governance risk

Self Serve Products

Data Processing Capabilities

- SG SuperGlue ETL
- MS MSaaS Service
- IN Ingestion
- CP Curation Process
- PS Persistence Service
- SP Streaming Process
- ML ML Pipeline

Data Serving Capabilities

- S3 S3 Bucket
- API Gateway
- DL Data Lake Table
- (x) Feature Store
- EB Topic
- DM Data Mart Table

Self Serve goals for next year

100% of Top 20 tasks in the Data lifecycle are Self Serve

Infra Provisioning

- Transactional Persistence
- Compute for stream, batch processing
- Monitor, Debug Infra
- Cost

Data Authoring

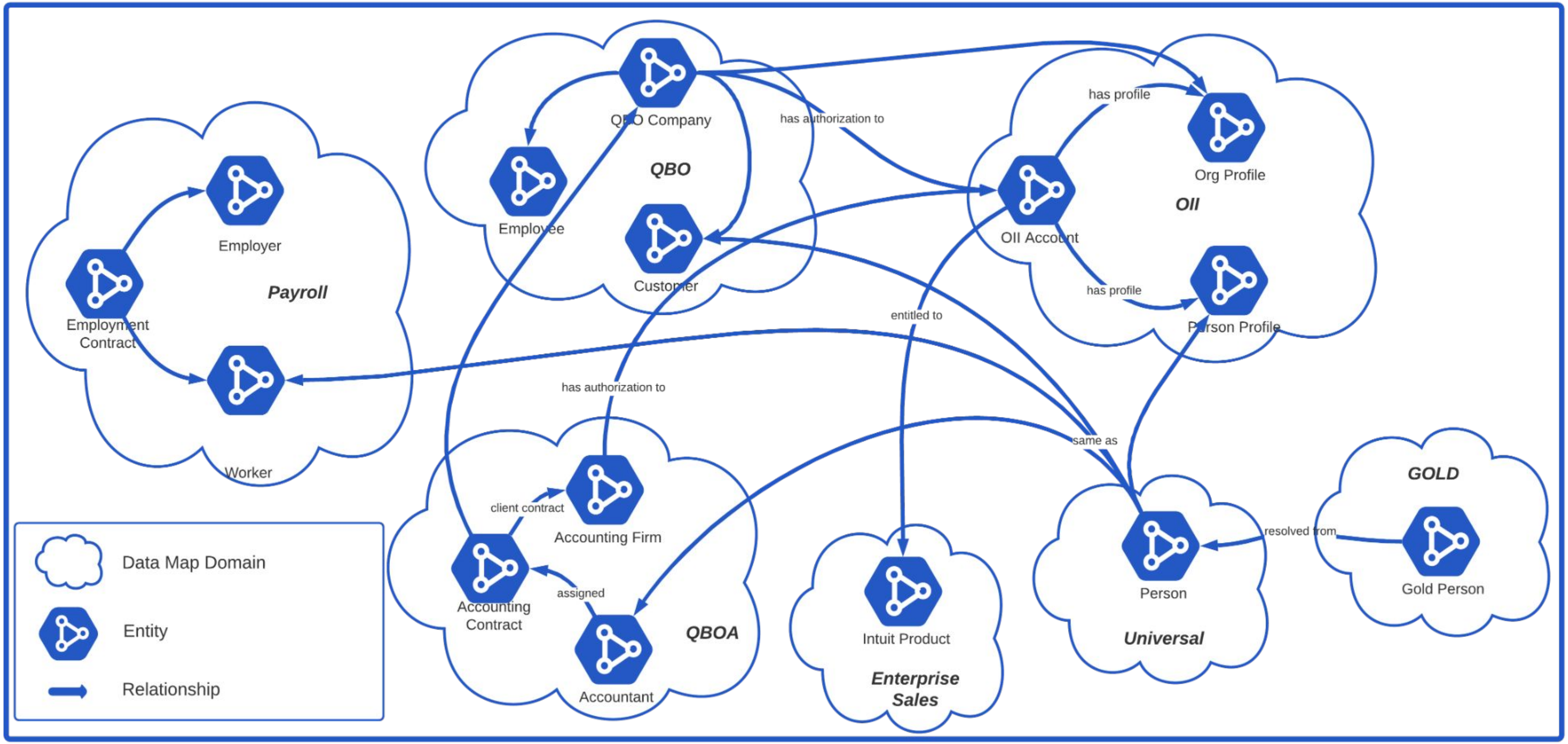
- Events, Schemas
- Ingestion
- Transformations
- Entities
- ML Features
- Data Quality, Observability
- Orchestration

Data Governance

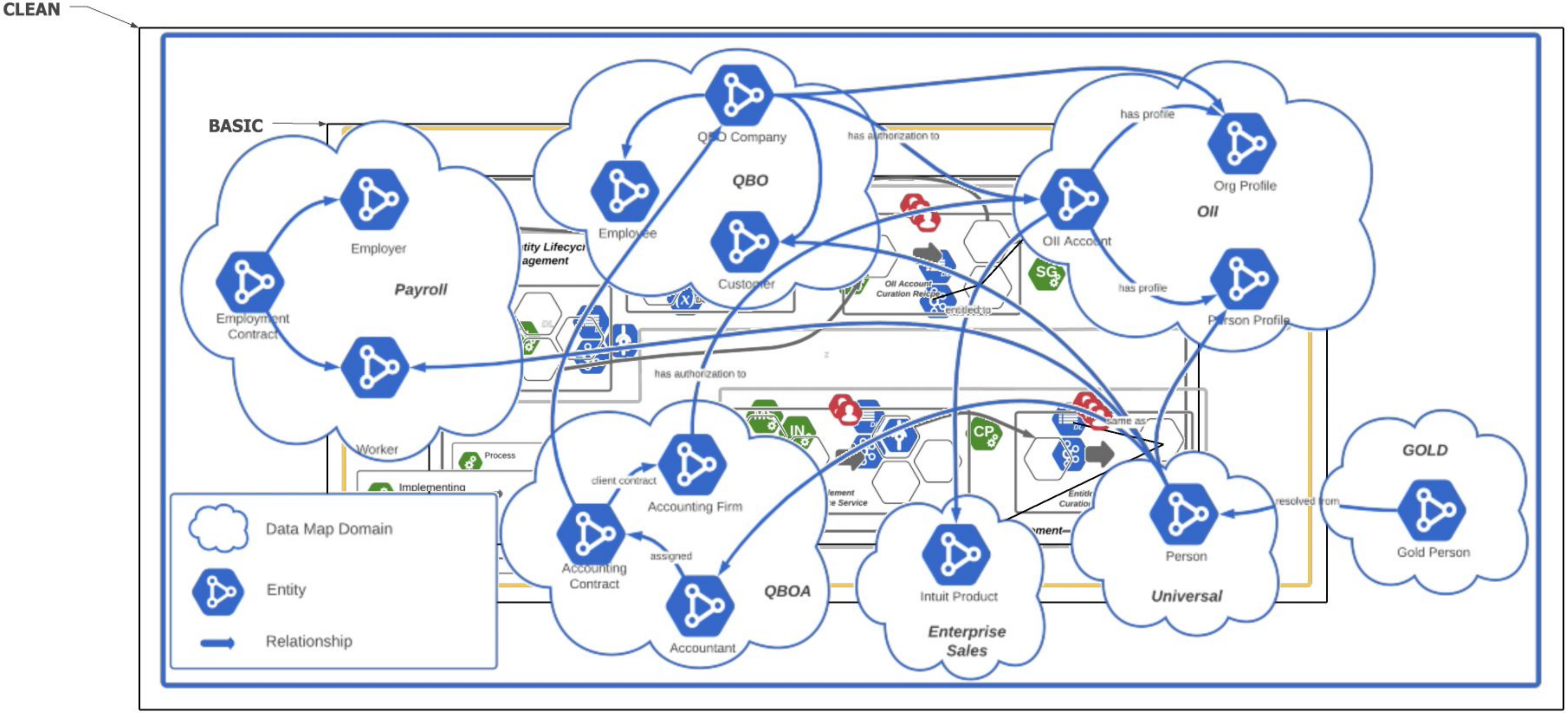
- Access Management
- Key management
- Compliance Controls & Audit
- Privacy

Rationalizing Data Definitions

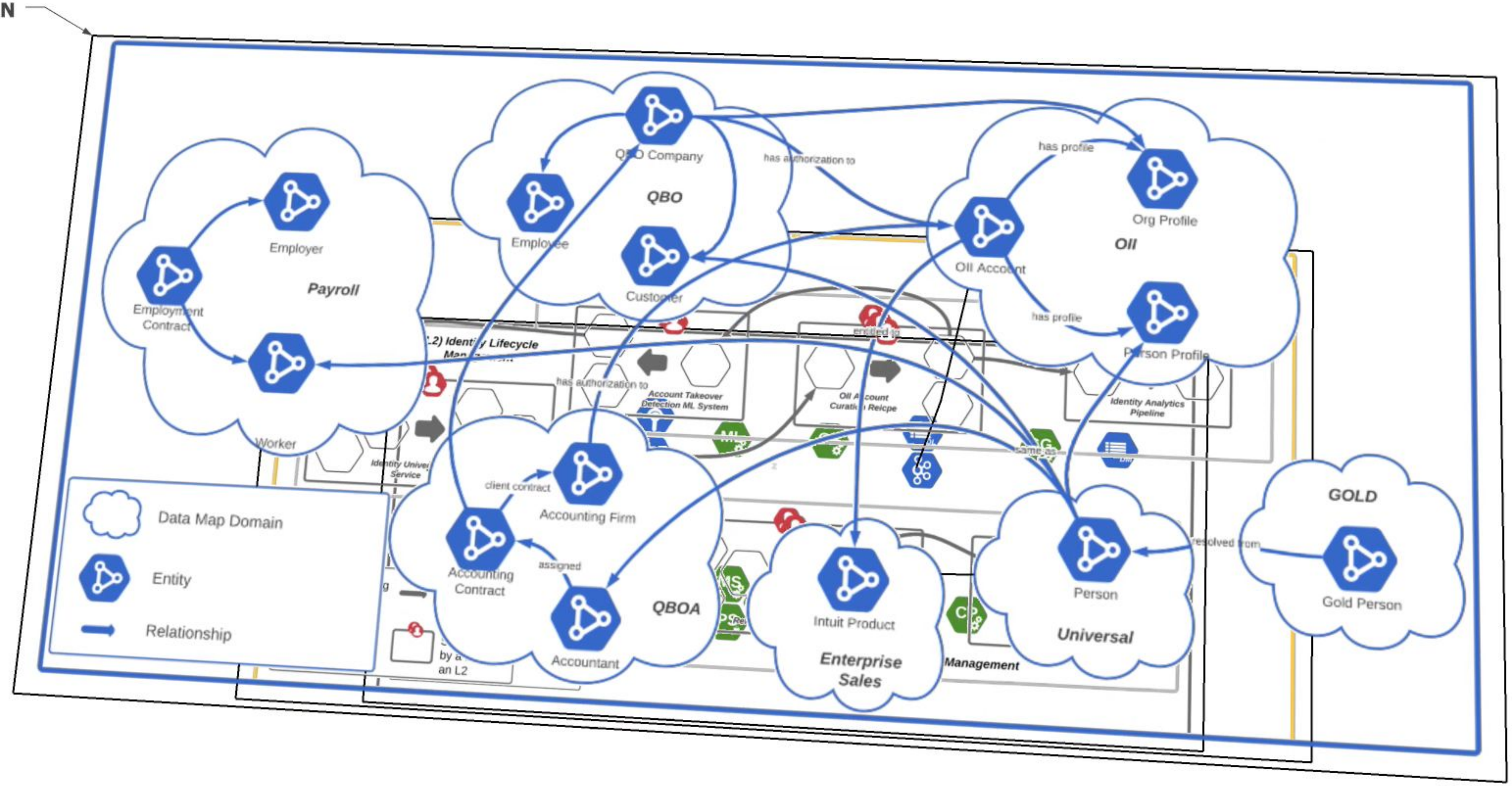
Clean entity information with formally defined meaning and relationships enables better data understanding. This is the purpose of entity definitions. They ensure that data is clean, organized, connected, discoverable and documented in a formal way.

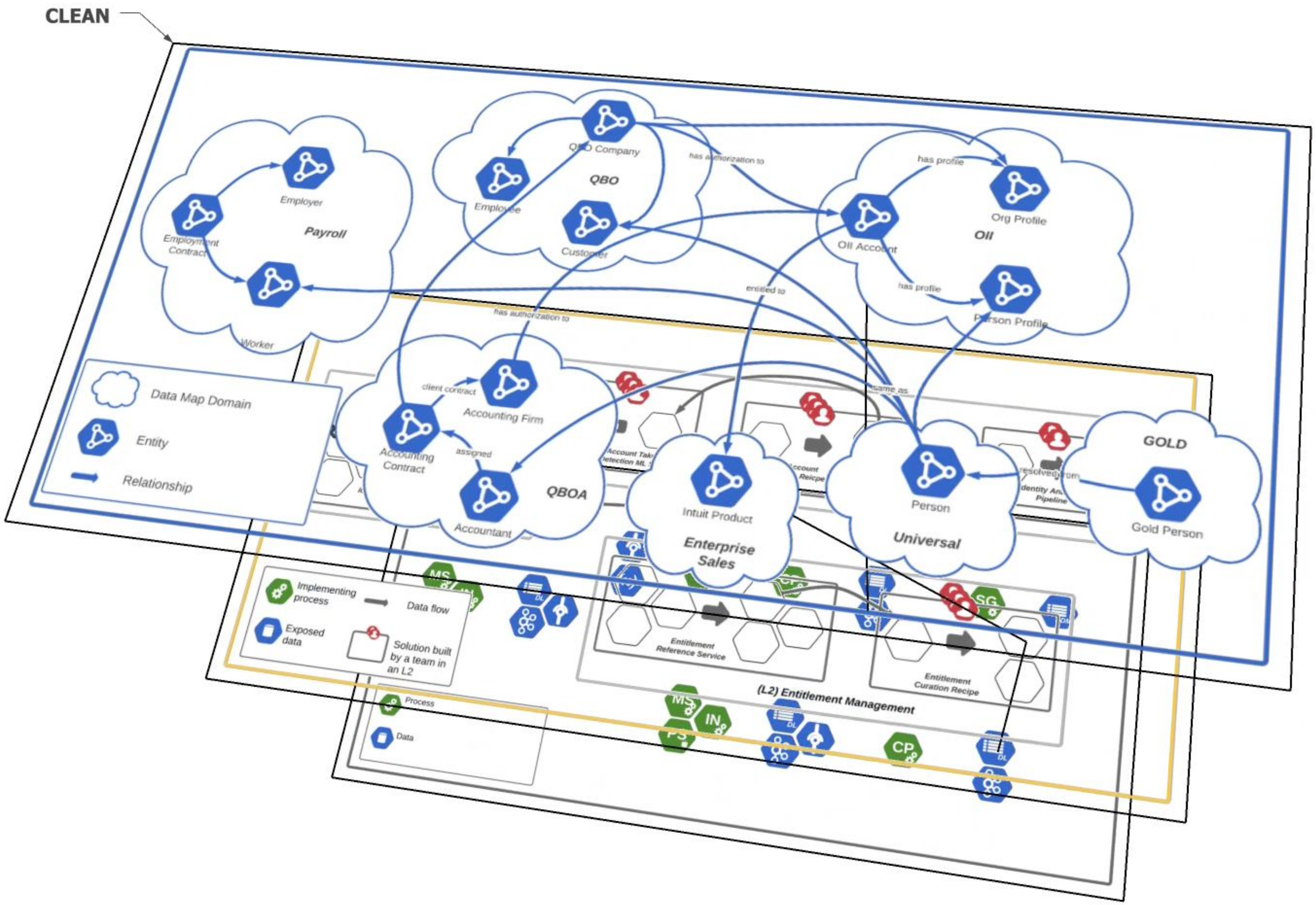


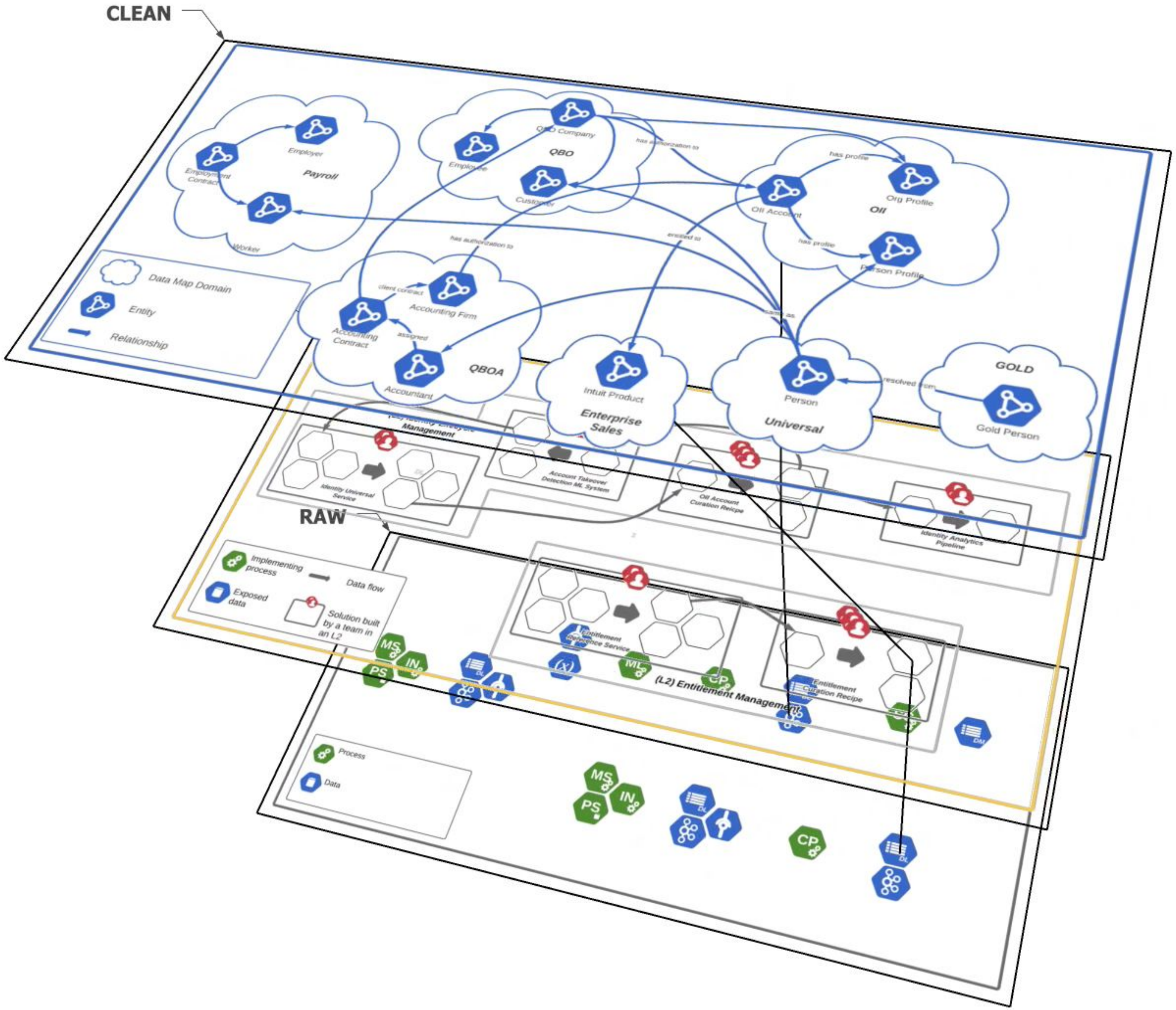
When you bring it all together, you get Intuit’s Data Mesh


CLEAN
BASIC
Payroll
Employer
Employment Contract
Worker
QBO
QBO Company
Employee
Customer
has authorization to
Oll
Org Profile
Oll Account
Person Profile
has profile
has profile
entitled to
has authorization to
QBOA
Accounting Firm
Accounting Contract
Accountant
client contract
assigned
Enterprise Sales
Intuit Product
Universal
Person
same as
resolved from
GOLD
Gold Person
Oll Account Curation Reicpe
Process
Implementing
Data Map Domain
Entity
Relationship

CLEAN


Payroll
Employer
Employment Contract
Worker
QBO
QBO Company
Employee
Customer
has authorization to
OII
OII Account
has profile
Org Profile
has profile
Person Profile
entitled to
has authorization to
same as
QBOA
client contract
Accounting Firm
Accounting Contract
assigned
Accountant
Enterprise Sales
Intuit Product
Universal
Person
resolved from
GOLD
Gold Person
Data Map Domain
Entity
Relationship
(L2) Identity Lifecycle Management
Account Takeover Detection ML System
OII Account Curation Reicpe
Identity Analytics Pipeline
Identity Universe Service
Management

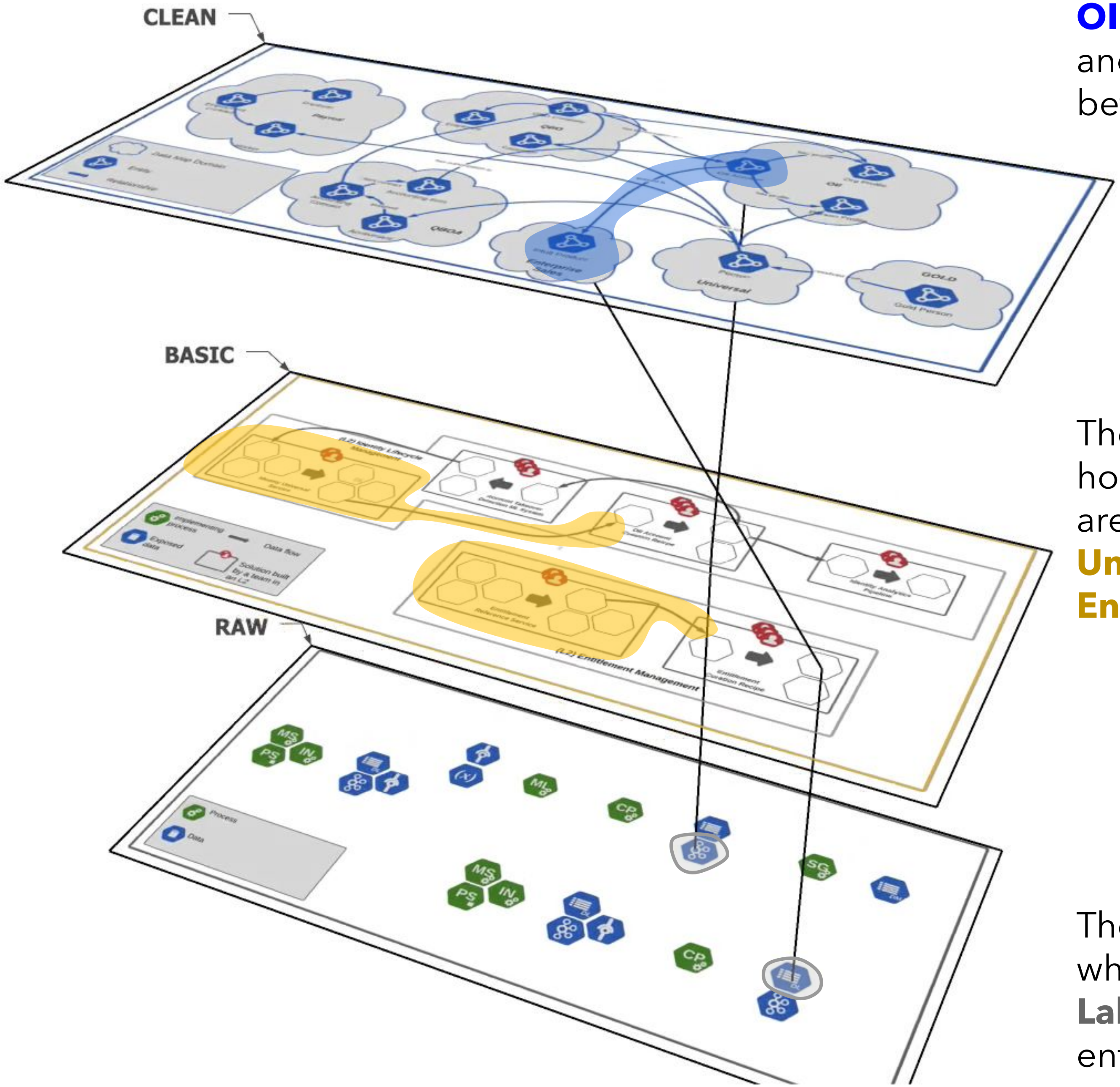

CLEAN
QBO Company
QBO
Employee
Customer
has authorization to
Employer
Employment Contract
Payroll
Worker
has profile
Org Profile
OII Account
OII
has profile
Person Profile
entitled to
has authorization to
client contract
Accounting Firm
Accounting Contract
assigned
Accountant
QBOA
same as
Intuit Product
Enterprise Sales
Person
Universal
resolved from
GOLD
Gold Person
Data Map Domain
Entity
Relationship
Account Tak
etection ML
Account
Reicpe
dentity An
Pipeline
Implementing process
Data flow
Exposed data
Solution built by a team in an L2
MS
IN
PS
DL
CP
SG
Entitlement Reference Service
Entitlement Curation Recipe
(L2) Entitlement Management
Process
Data


CLEAN
Employer
Employment Contract
Payroll
Worker
QBO Company
Employee
QBO
Customer
has authorization to
has profile
Org Profile
OII Account
OII
has profile
Person Profile
entitled to
has authorization to
client contract
Accounting Firm
Accounting Contract
assigned
Accountant
QBOA
same as
Intuit Product
Enterprise Sales
Person
Universal
resolved from
GOLD
Gold Person
Data Map Domain
Entity
Relationship
(L2) Identity Lifecycle Management
Identity Universal Service
Account Takeover Detection ML System
OII Account Curation Reicpe
Identity Analytics Pipeline
RAW
Implementing process
Data flow
Exposed data
Solution built by a team in an L2
Entitlement Reference Service
Entitlement Curation Recipe
(L2) Entitlement Management
MS
IN
PS
ML
CP
DL
DM
Process
Data

In this example, the clean information describes entities **OII Account** and **Intuit Product** and the **Entitled To** relationship between them.

Capturing meaning, relationship, ownership, and system dependencies builds a full, rich picture for everyone.

No tribal knowledge needed!


CLEAN
Enterprise Sales
Universal
GOLD
BASIC
(L2) Identity Lifecycle Management
(L2) Entitlement Management
RAW
MS
IN
PS
ML
CP
SG

The basic information describes how the data for these entities are sourced from the **Identity Universal Service** and the **Entitlement Reference Service**.

The raw information describes which **Event Bus topic** and **Data Lake table** the data for these entities can be found in.

Q&A

Tristan Baker - linkedin.com/in/tristanbaker
Suresh Raman - linkedin.com/in/ramansuresh
Allison Bellah (in absentia) - linkedin.com/in/allisonbellah