Data marketplace
A required infrastructure for
business of scale

Problem statement

1. Drivers own cars and produce data
2. Data is stored by the entity who created/deployed the device to the car
3. New data driven businesses need to access the data

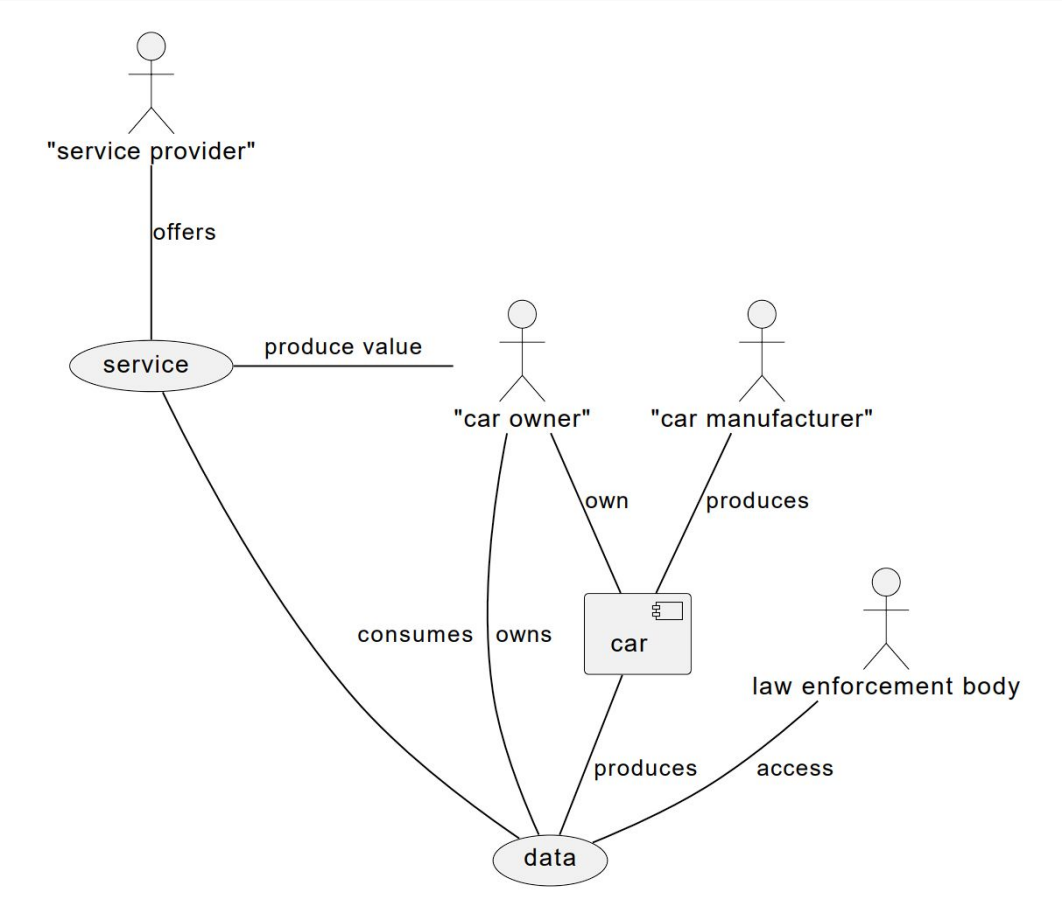
As of today, there is no unified way of handling consent, and data access across all parties. Everything is handled individually by the parties involved.

Eco system participants (initial)

Service provider (data consumer but also data provider)

Car owner (also data producer)

Car manufacturer (data provider, aggregator, also possibly data consumer)



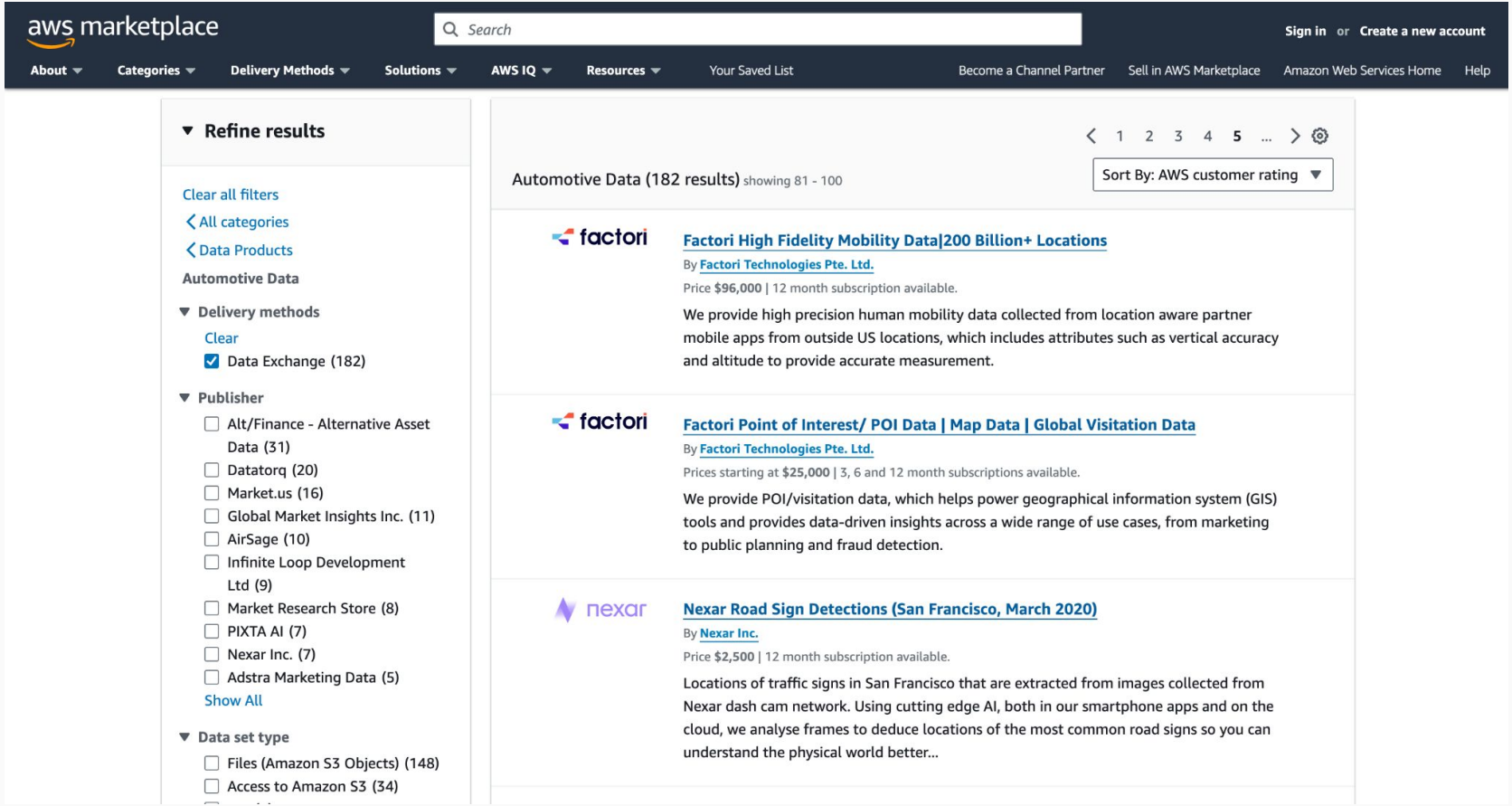
Desired state

1. Drivers and owners produce data based on consent
2. Data is stored by the entity who created/deployed the device to the car if consented by the owner
3. This entity defines terms and conditions to access the stored data
4. Interfaces allow the access to data protected by state of the art security measures and access control
5. We create a goto-place where all participants can go and explore
6. Everyone can onboard and register for accessing the data according to the predefined terms and conditions
7. Curated/processed datasets, models, and other things can be made available again

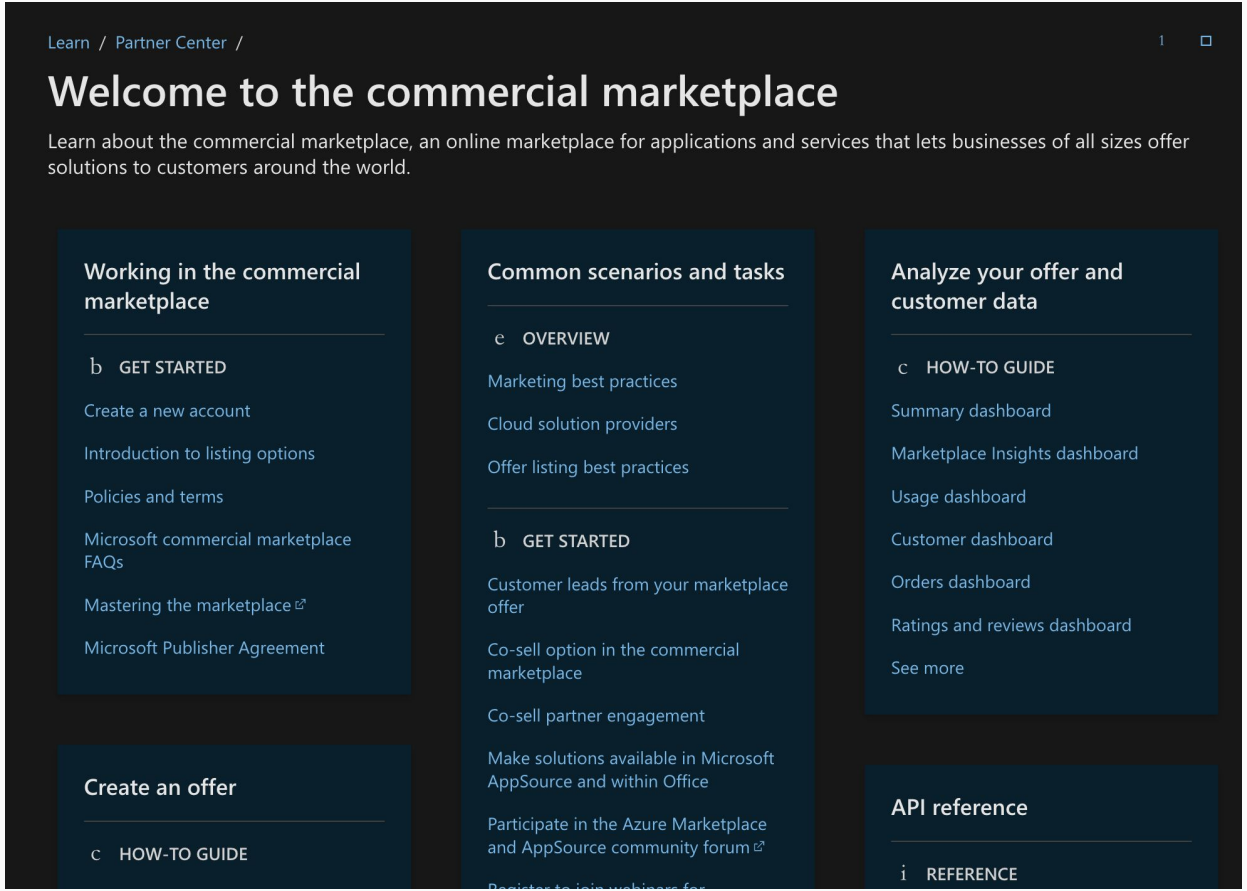
examples

(Data) marketplaces are omnipresent in todays web based ecosystems, some examples you may know

AWS data exchange



Examples: MS azure marketplace



Examples: Kaggle

kaggle

Search datasets

Filters

Create

Home

Competitions

Datasets

Models

Code

Discussions

Learn

More

All datasets | Computer Science | Education | Classification | Computer Vision | NLP | Data Visualization | Pre-Trained Model

Salaries 2024

Emre Öksüz · Updated 3 days ago

Usability **10.0** · 23 kB

1 File (CSV)

19

⭐ Stars Dataset ⭐

Waqar Ali · Updated 6 days ago

Usability **10.0** · 41 kB

1 File (CSV)

7

Dataset from TikTok

Ramin Huseyn · Updated 18 days ago

Usability **10.0** · 813 kB

1 File (CSV)

8

Analysis

Jenil hareshbhai ghori · Updated 3…

Usability **10.0** · 1 MB

4 Files (other, CSV)

14

Music

See All

1990s Classic Hits (with Spotify Data)

The Bumpkin · Updated 15 days ago

Usability **10.0** · 42 kB

1 File (CSV)

Small Guitar Model Dataset (113 Models)

The Bumpkin · Updated 16 days ago

Usability **10.0** · 9 kB

1 File (CSV)

1980s Classic Hits (with Spotify Data)

The Bumpkin · Updated 16 days ago

Usability **10.0** · 45 kB

1 File (CSV)

Ozzy Osbourne Studio Tracks (with Spotify Data)

The Bumpkin · Updated 9 days ago

Usability **10.0** · 7 kB

1 File (CSV)

Architecture

Guiding principles

It is not the goal to re-invent the wheel

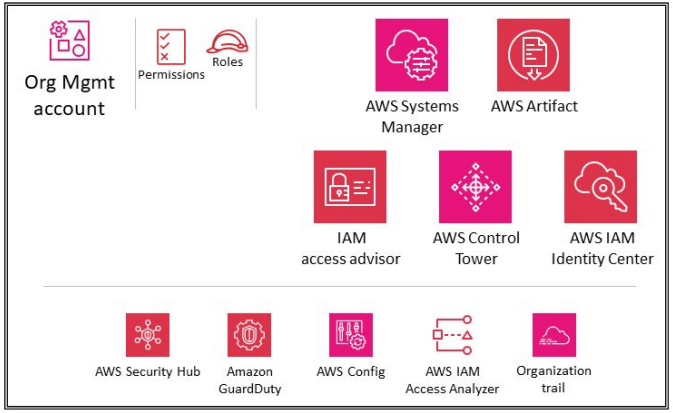
We need a fast enabler to accelerate the business growth in the service space.

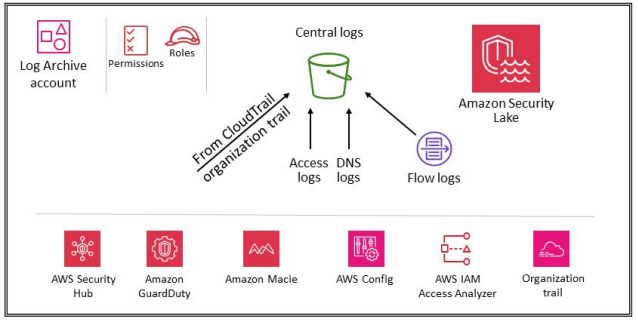
1. Resilient, low maintenance costs and highly available
2. Secure and cost efficient
3. Self services (onboarding, resource creation)
4. Multitenancy
5. Worldwide reach
6. (Almost) limitless scale
7. Ease of use
8. Easy to extend

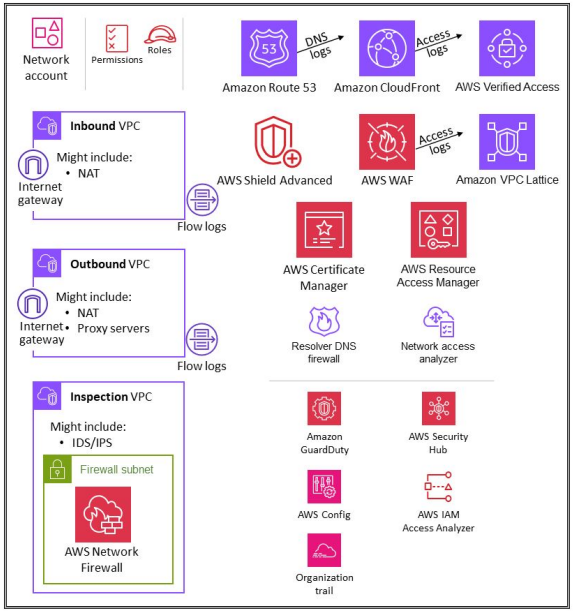
Parts

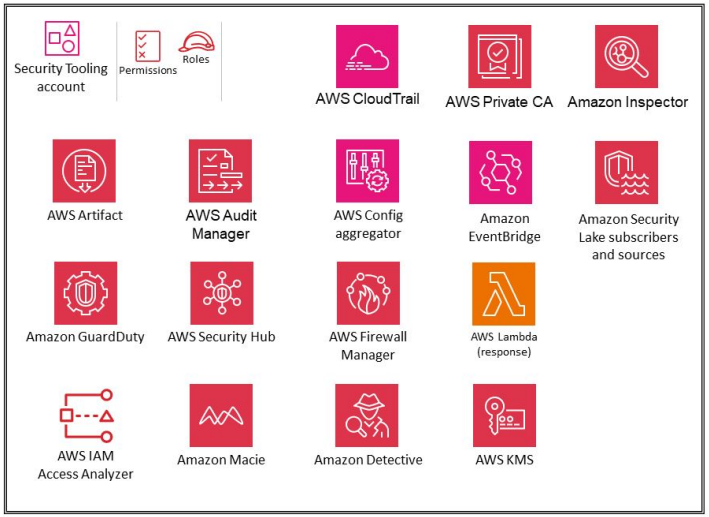
Parts of the ecosystem, separated by concern and loosely coupled for data transmission between them

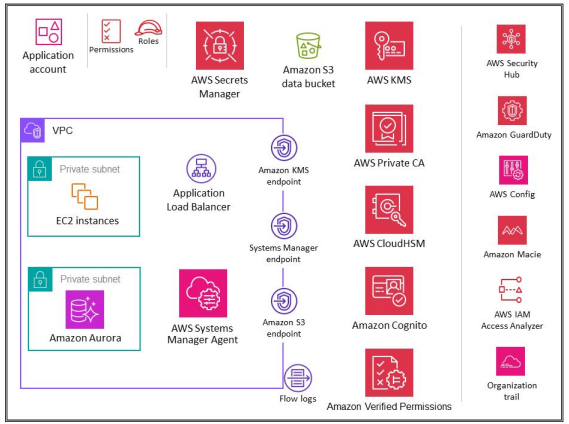
1. User and role management
2. Access control
3. Usage measuring
4. Billing
5. User onboarding (self services)
6. Terms and condition store
7. dataset /data source metadata store
8. Consent store
9. Onboarding strategies for data sources (streaming, blocks), data access mechanisms
10. Data transportation (source to sink)
11. Fulfillment, payment
12. And more

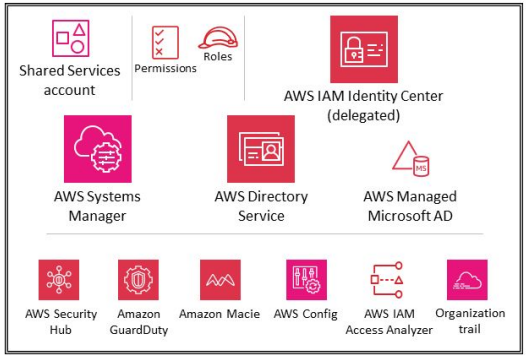

Org Mgmt account
Permissions
Roles
AWS Systems Manager
AWS Artifact
IAM access advisor
AWS Control Tower
AWS IAM Identity Center
AWS Security Hub
Amazon GuardDuty
AWS Config
AWS IAM Access Analyzer
Organization trail


Log Archive account
Permissions
Roles
Central logs
From CloudTrail and organization trail
Access logs
DNS logs
Flow logs
Amazon Security Lake
AWS Security Hub
Amazon GuardDuty
Amazon Macie
AWS Config
AWS IAM Access Analyzer
Organization trail


Network account
Permissions
Roles
Amazon Route 53
DNS logs
Amazon CloudFront
Access logs
AWS Verified Access
Inbound VPC
Might include:
• NAT
Internet gateway
Flow logs
AWS Shield Advanced
AWS WAF
Access logs
Amazon VPC Lattice
Outbound VPC
Might include:
• NAT
• Proxy servers
Internet gateway
Flow logs
AWS Certificate Manager
AWS Resource Access Manager
Resolver DNS firewall
Network access analyzer
Inspection VPC
Might include:
• IDS/IPS
Firewall subnet
AWS Network Firewall
Amazon GuardDuty
AWS Security Hub
AWS Config
AWS IAM Access Analyzer
Organization trail


Security Tooling account
Permissions
Roles
AWS CloudTrail
AWS Private CA
Amazon Inspector
AWS Artifact
AWS Audit Manager
AWS Config aggregator
Amazon EventBridge
Amazon Security Lake subscribers and sources
Amazon GuardDuty
AWS Security Hub
AWS Firewall Manager
AWS Lambda (response)
AWS IAM Access Analyzer
Amazon Macie
Amazon Detective
AWS KMS


Application account
Permissions
Roles
AWS Secrets Manager
Amazon S3 data bucket
AWS KMS
VPC
Private subnet
EC2 instances
Application Load Balancer
Amazon KMS endpoint
Systems Manager endpoint
Amazon S3 endpoint
Flow logs
Private subnet
Amazon Aurora
AWS Systems Manager Agent
AWS Private CA
AWS CloudHSM
Amazon Cognito
Amazon Verified Permissions
AWS Security Hub
Amazon GuardDuty
AWS Config
Amazon Macie
AWS IAM Access Analyzer
Organization trail

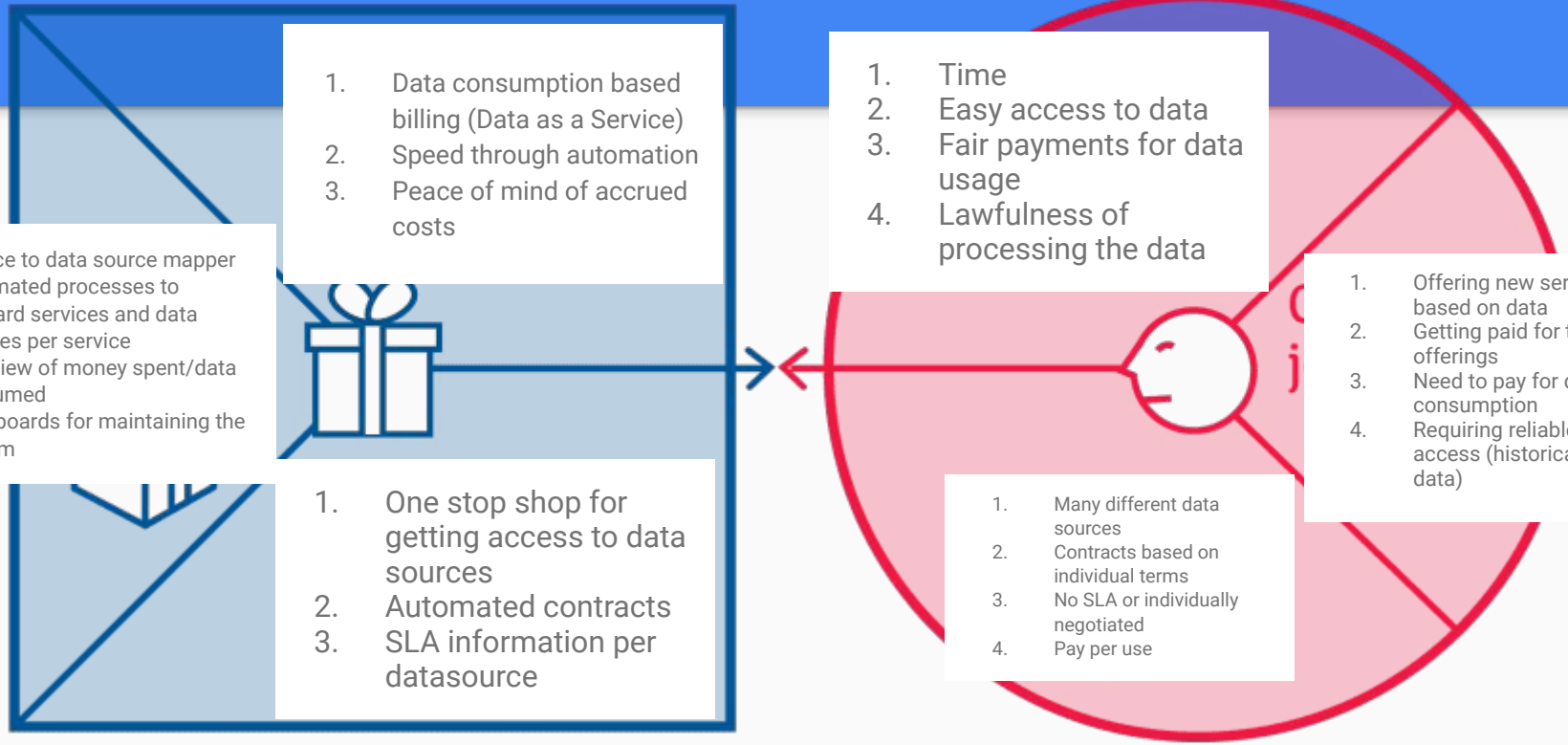

Shared Services account
Permissions
Roles
AWS IAM Identity Center (delegated)
AWS Systems Manager
AWS Directory Service
AWS Managed Microsoft AD
AWS Security Hub
Amazon GuardDuty
Amazon Macie
AWS Config
AWS IAM Access Analyzer
Organization trail

What is in for the parties?

What are the benefits of this solution for the participants of the ecosystem?

Value Proposition

Data consumer



1. Data consumption based billing (Data as a Service)
2. Speed through automation
3. Peace of mind of accrued costs

1. Service to data source mapper
2. Automated processes to onboard services and data sources per service
3. Overview of money spent/data consumed
4. Dashboards for maintaining the system

1. One stop shop for getting access to data sources
2. Automated contracts
3. SLA information per datasource

1. Time
2. Easy access to data
3. Fair payments for data usage
4. Lawfulness of processing the data

1. Offering new services based on data
2. Getting paid for their offerings
3. Need to pay for data consumption
4. Requiring reliable data access (historical or live data)

1. Many different data sources
2. Contracts based on individual terms
3. No SLA or individually negotiated
4. Pay per use

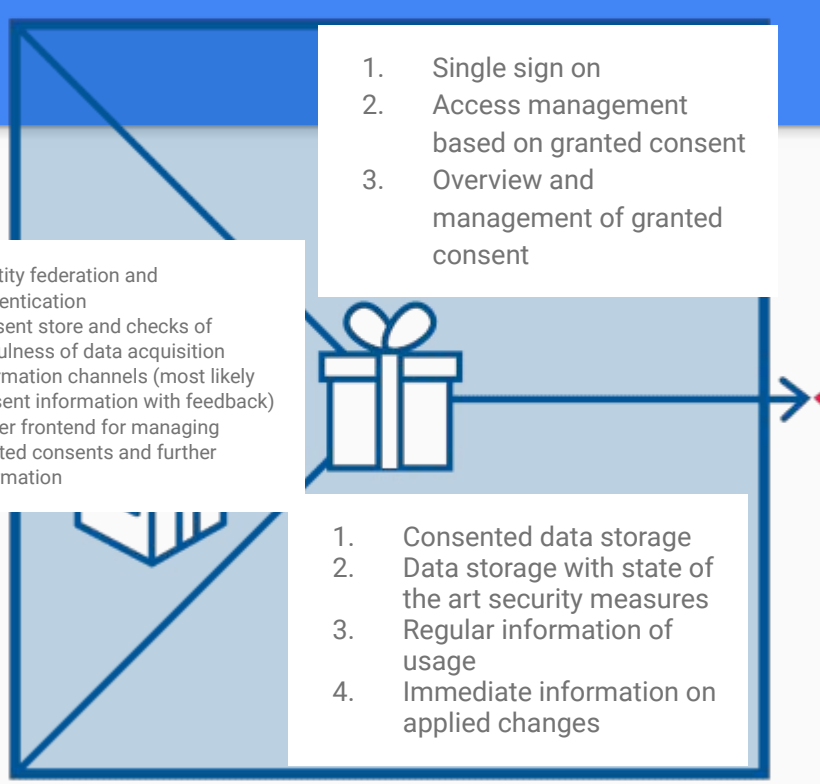
Value Proposition

Data producer

1. Identity federation and authentication
2. Consent store and checks of lawfulness of data acquisition
3. Information channels (most likely consent information with feedback)
4. A user frontend for managing granted consents and further information

1. Single sign on
2. Access management based on granted consent
3. Overview and management of granted consent

1. Consented data storage
2. Data storage with state of the art security measures
3. Regular information of usage
4. Immediate information on applied changes



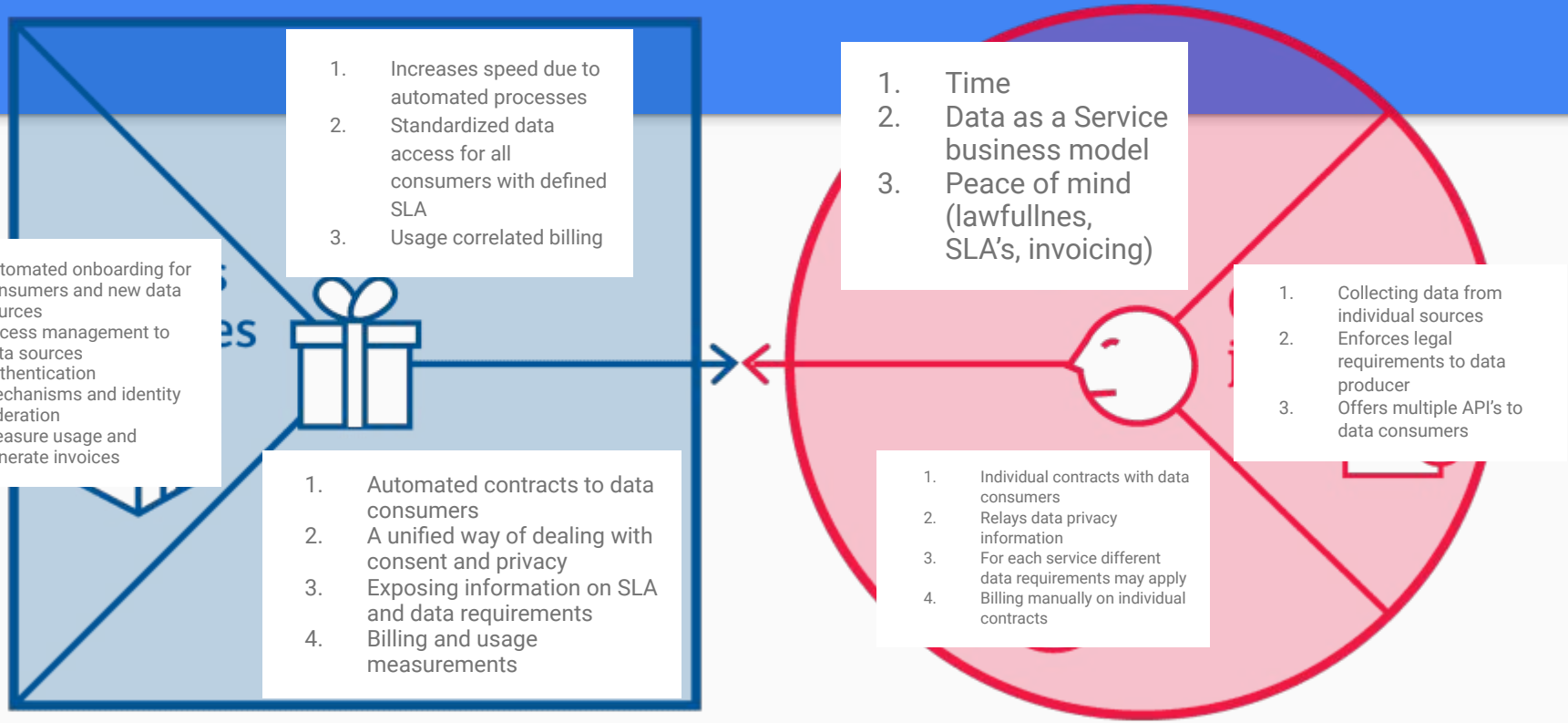
1. One login process only (car manufacturer)
2. Peace of mind wrt. Consent and data access management
3. Time due to automated onboarding processes

1. Wanna use new services
2. Is producing data that can be collected by the data aggregator (aka provider)
3. Enjoying riding a great car
4. Executing data privacy rights

1. Is my data protected
2. Multiple logins and passwords for multiple services
3. enrolling for new services
4. Lawfulness of data acquisition

Value Proposition

Data provider



1. Increases speed due to automated processes
2. Standardized data access for all consumers with defined SLA
3. Usage correlated billing

1. Automated onboarding for consumers and new data sources
2. Access management to data sources
3. Authentication mechanisms and identity federation
4. Measure usage and generate invoices

1. Automated contracts to data consumers
2. A unified way of dealing with consent and privacy
3. Exposing information on SLA and data requirements
4. Billing and usage measurements

1. Time
2. Data as a Service business model
3. Peace of mind (lawfullnes, SLA's, invoicing)

1. Collecting data from individual sources
2. Enforces legal requirements to data producer
3. Offers multiple API's to data consumers

1. Individual contracts with data consumers
2. Relays data privacy information
3. For each service different data requirements may apply
4. Billing manually on individual contracts

The team

Answer the question, “Why are we the ones to solve the problem we identified?”

Wendy Writer

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor

Ronny Reader

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor

Abby Author

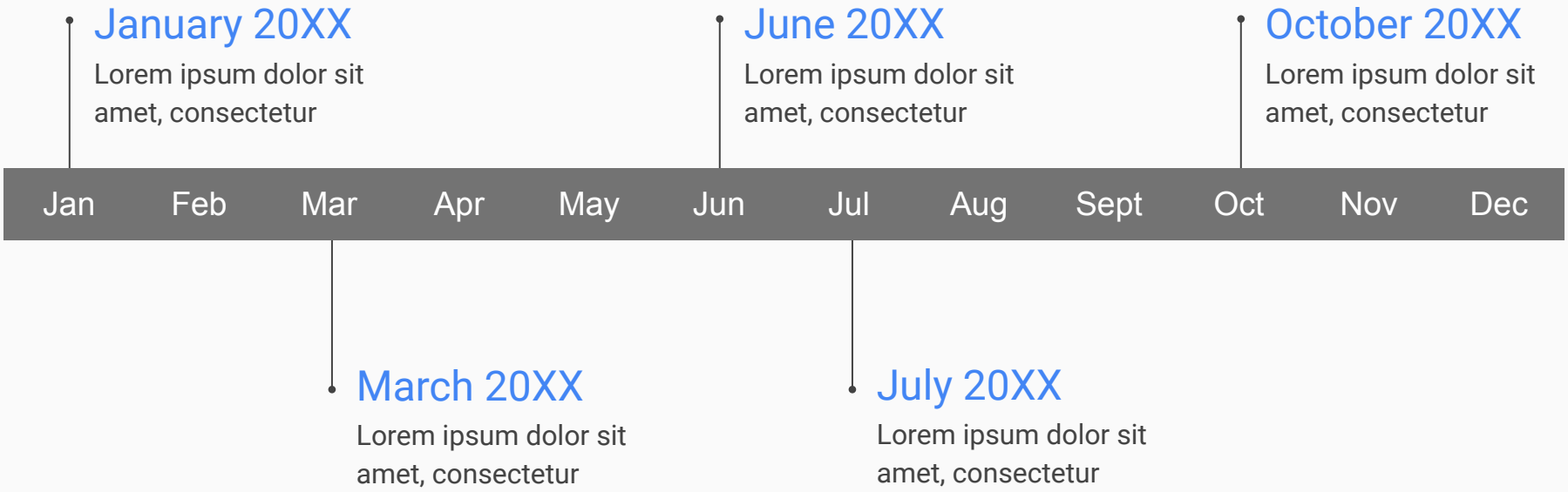
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor

Berry Books

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor

Milestones

Show where you are in the process and what's left to tackle



Appendix

Show the audience you anticipated their questions.

Leave room for Q&A, but use the Appendix as a way to show that you both thought about those questions and have solid answers with supporting information. Let the audience test their understanding of the problem and the solution you've outlined - questions give them a chance to talk themselves into your approach, and give you a chance to show mastery of the subject.

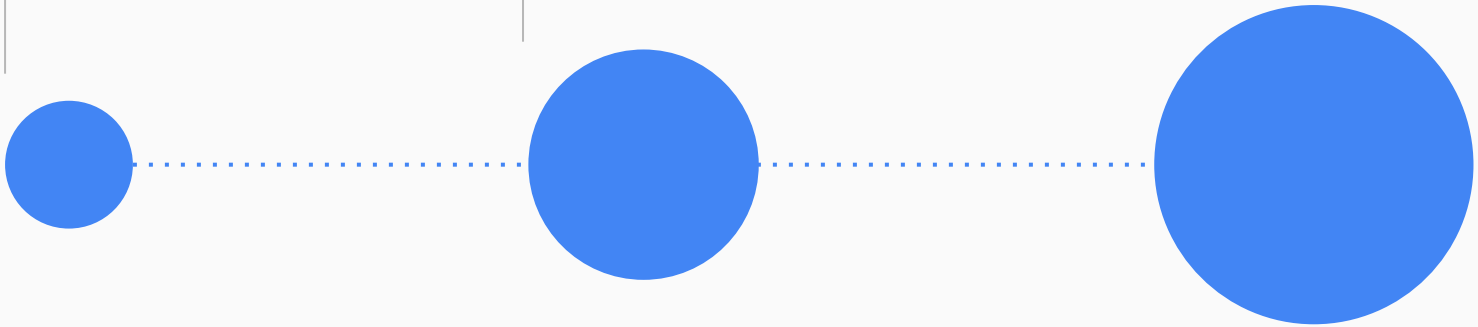
How it works

Step 1

Lorem ipsum dolor sit amet, consectetur

Step 2

Lorem ipsum dolor sit amet, consectetur

Step 3

Lorem ipsum dolor sit amet, consectetur