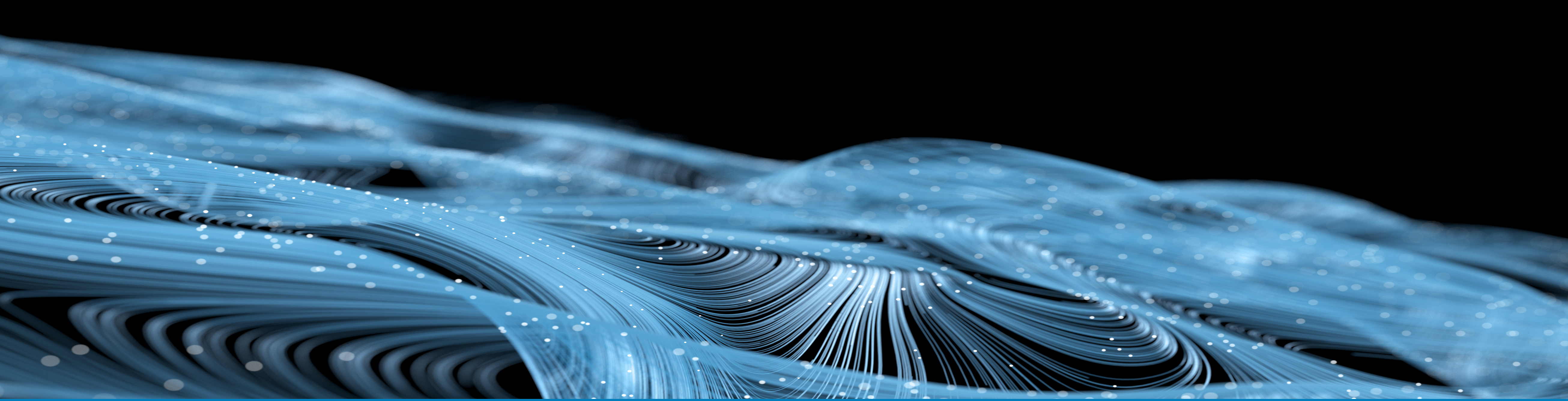
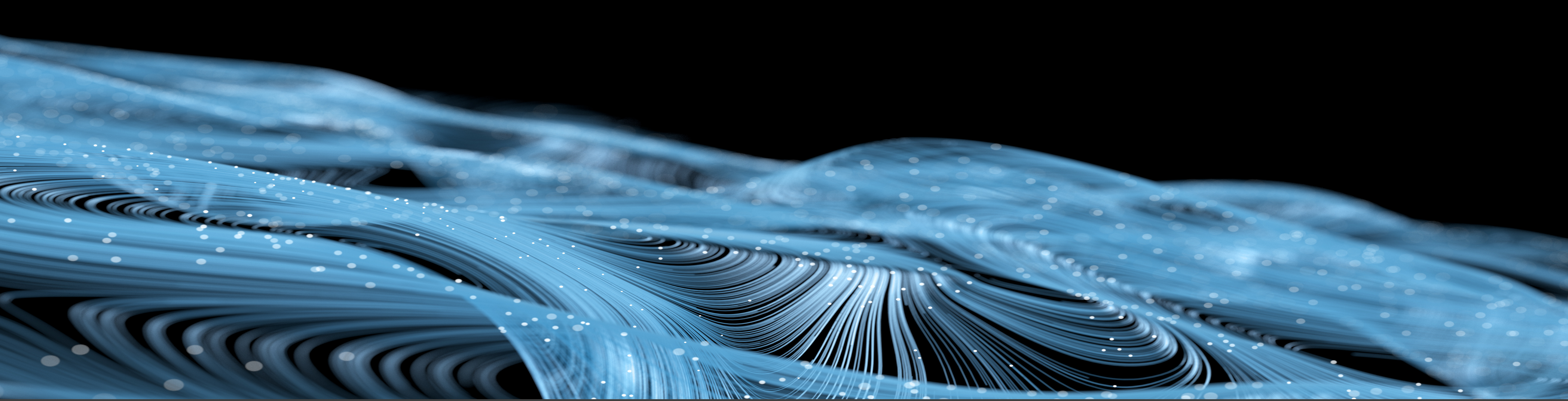


Niteo

Governança de Dados & Analytics

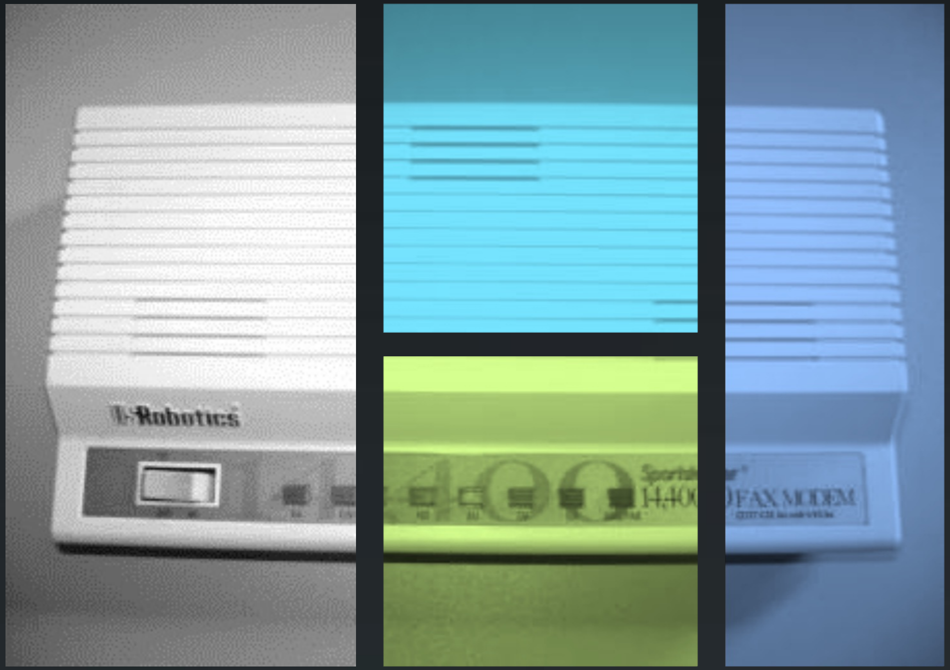
Visão Executiva



transformações atuais


aceleração tecnológica
maior capacidade + baixo custo

1 Maior conectividade

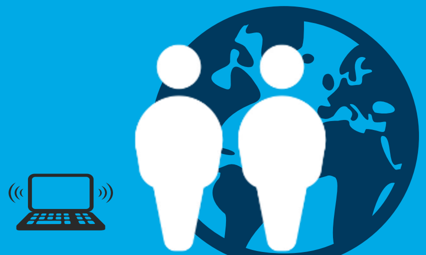


USRobotics Sportster 14,400
Fax modem (1994)

2012 the number of networked devices…

…equals the global population

In **2016**, the number of networked devices…



…is equal **2x** the global population

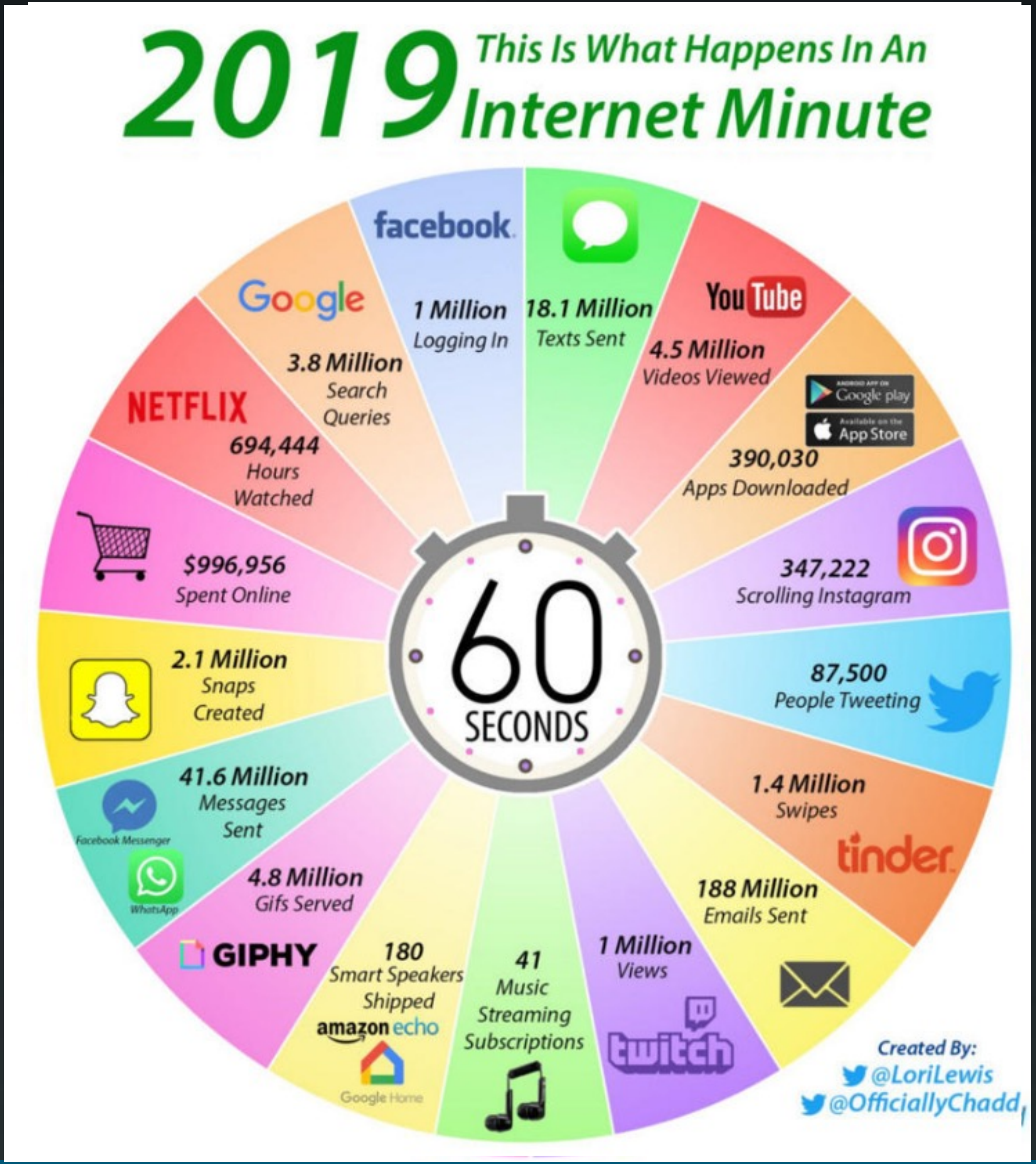
Data source: Intel, "What Happens in an Internet Minute"

2 Mais dados

10x
increase every
five years

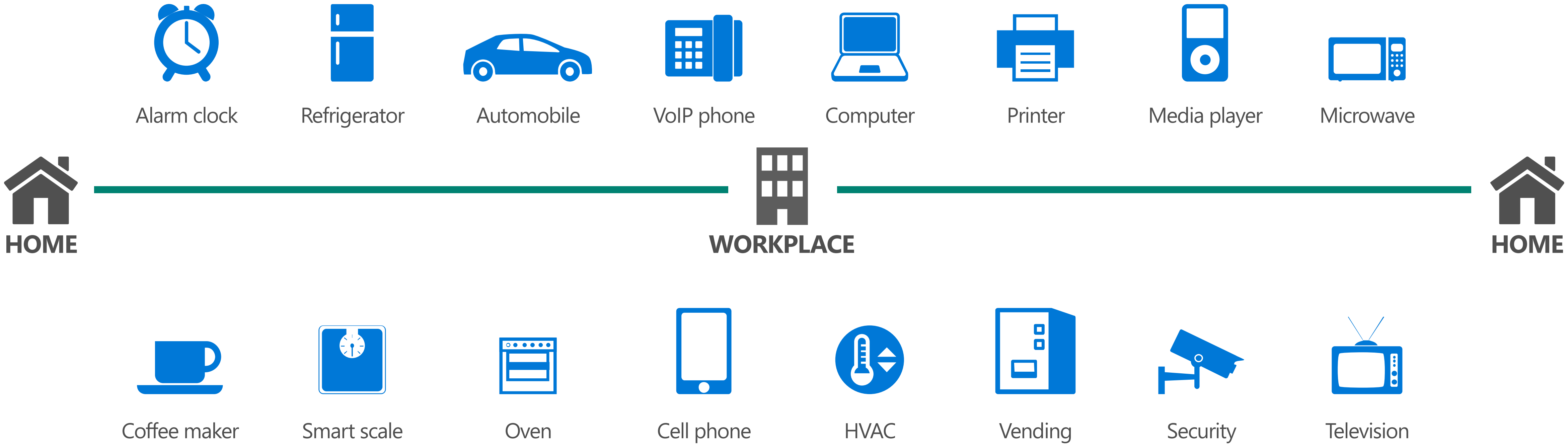
85%
from **new**
data types

volume
velocity
variety



Data source: IDC

IOT 2012

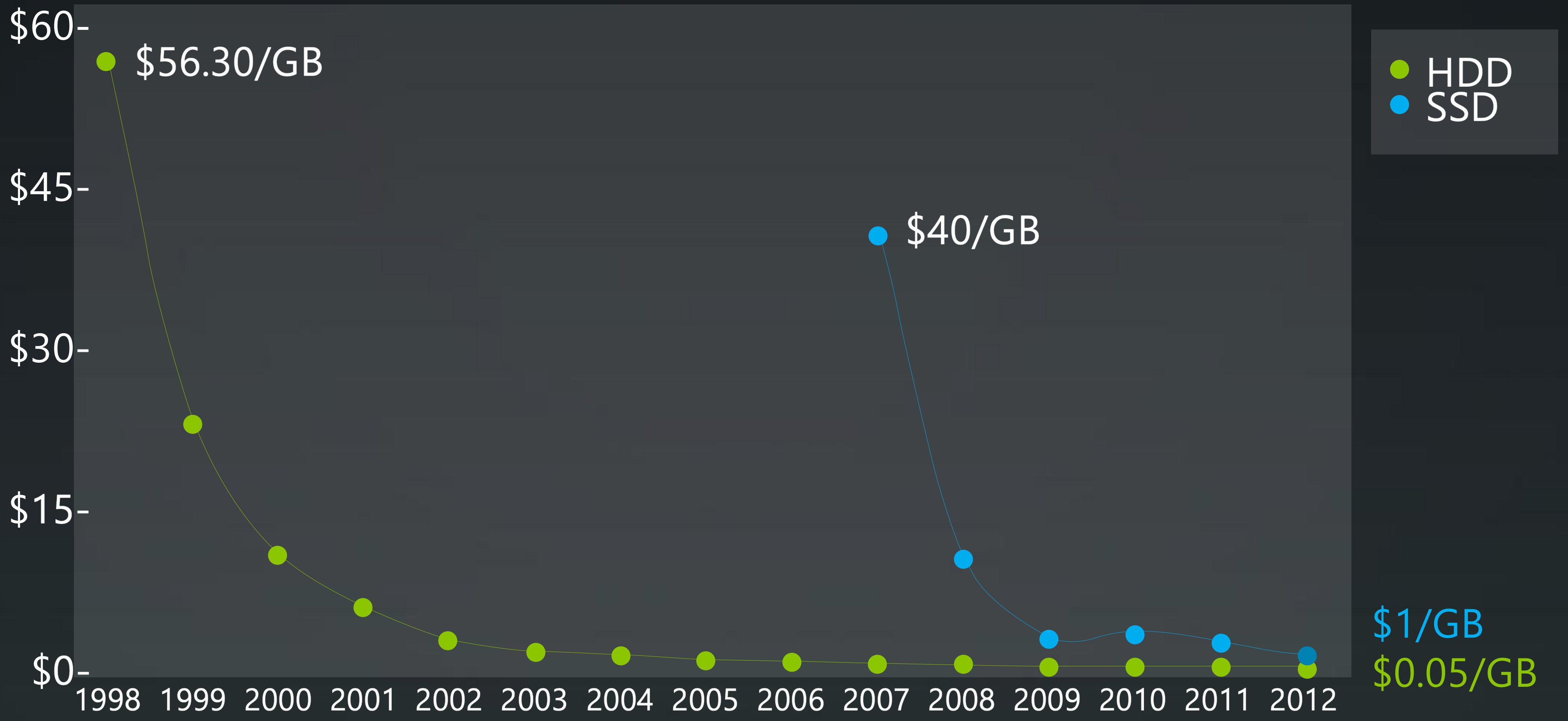

Alarm clock
Refrigerator
Automobile
VoIP phone
Computer
Printer
Media player
Microwave
HOME
WORKPLACE
HOME
Coffee maker
Smart scale
Oven
Cell phone
HVAC
Vending
Security
Television

IOT 2022

Medication adherence
Health monitoring
Pet tracking
Behavior modification
Object tracking
Child and elder monitoring
Sports and fitness
Smart lighting
Indoor navigation
Beacons and proximity
Trip tracking and car health
HOME
COMMUTE
WORKPLACE
COMMUTE
HOME
Smart appliances
Food and nutrition tracking
Identity
Office equipment
Smart vending machines
Bike ride stats and protection
Sleep tracking
Air conditioning and temperature control
Environmental sensors
Information capture
Control
Home security
Home automation
Leak detection
Garden, lawn and plant care
New devices and sensors
Entertainment systems

3 E com menor custo...

average HDD & SSD prices in USD per gigabyte



Data sources: Mkomo.com, Gartner, and Pingdom (December 2011) www.pingdom.com

4 Maior poder computacional

The ‘cloud’ 1990

The ‘cloud’ 2016

5 Disponibilidade da tecnologia como serviço

aceleração para um
mundo de tecnologias
como serviços

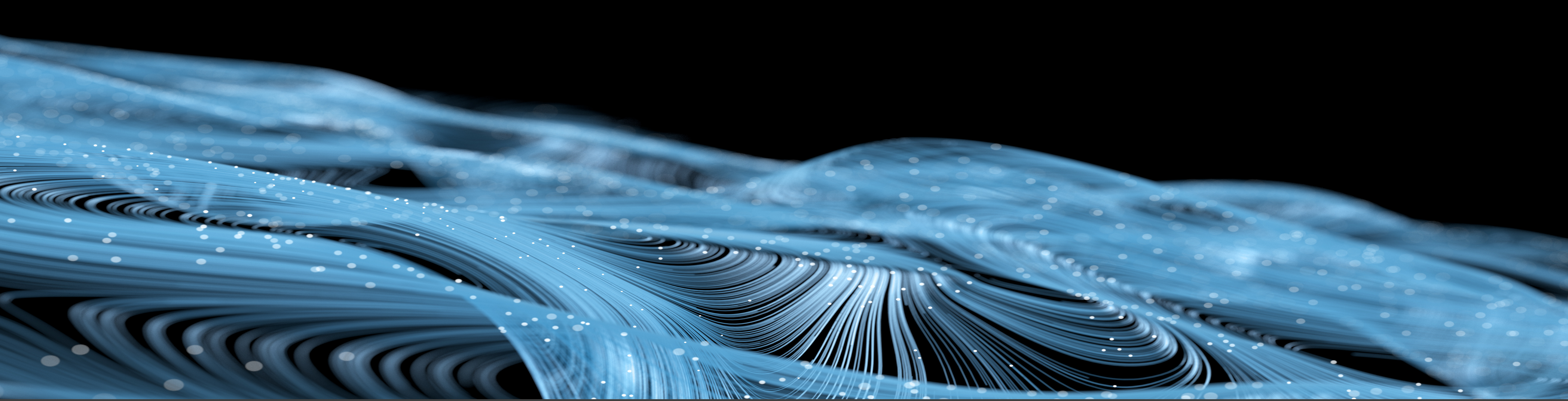
Infra > eletricidade> telefonia > IT > sist. especialistas > IA

menor custo para inovar


IoT
conectividade
Cloud & Big Data
dados
aceleração tecnológica
dispositivos
Robótica,AR, 3D printing
insights
Inteligência Artifical

"Digital transformation is when companies use **technology** to radically change the **performance** or **reach** of an enterprise"

George Westerman – Principal Research Scientist – MIT Sloan



Dados => Informação

Dados x Informações

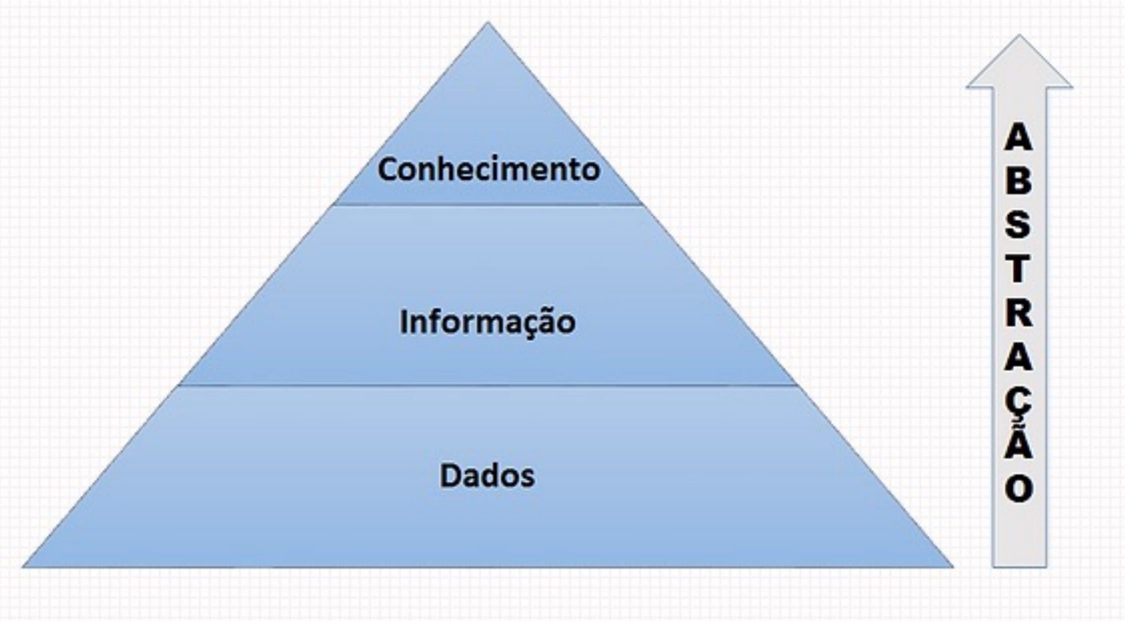
- **Data (Latin datum)** are a set of qualitative or quantitative values and variables about a person, an object, or a location at a given time or across a period of time.
- Data does not depend on information. However, information cannot exist without data
- **Information (Latin 'īnfōrmātiō /informare')** is described as that form of data which is processed, organised, specific and structured, which is presented in the given setting.
- It assigns meaning and improves the reliability of the data, thus ensuring understandability and reduces uncertainty.

BASIS FOR COMPARISON	DATA	INFORMATION
Meaning	Data means raw facts gathered about someone or something, which is bare and random.	Facts, concerning a particular event or subject, which are refined by processing is called information.
What is it?	It is just text and numbers.	It is refined data.
Based on	Records and Observations	Analysis
Form	Unorganized	Organized
Useful	May or may not be useful.	Always
Specific	No	Yes
Dependency	Does not depend on information.	Without data, information cannot be processed.

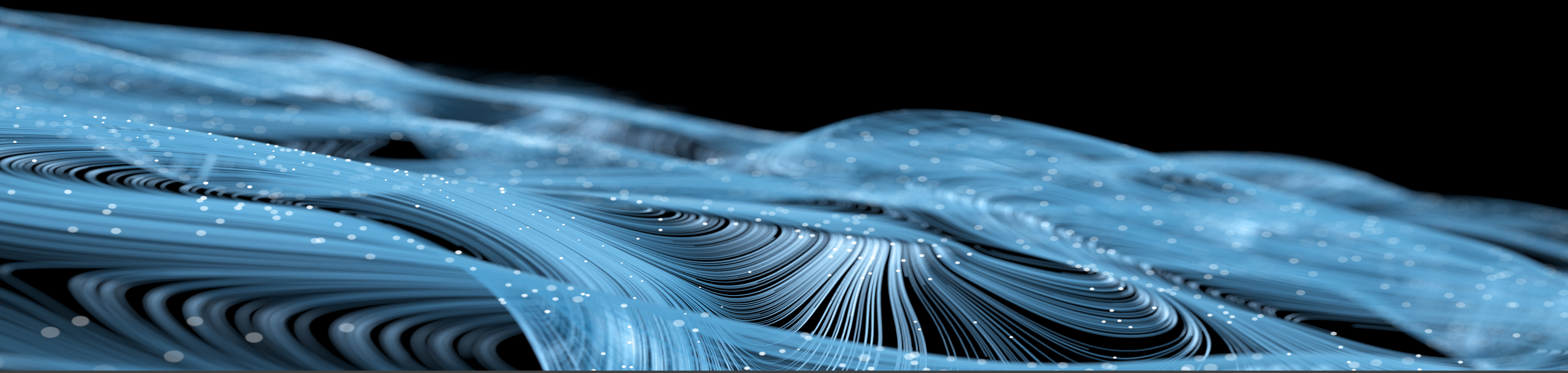
Exemplos

Data	Information
each individual homework and test grade of a student in one class	the student’s average grade for each class
typing the words “cat videos” in your computer search engine (input)	the list of search results that includes a variety of cat videos on the internet (output)
55112377988	a person’s phone number 55(11)237-7988
100, 212, 0, 32	the freezing and boiling points of water in Fahrenheit and Celsius

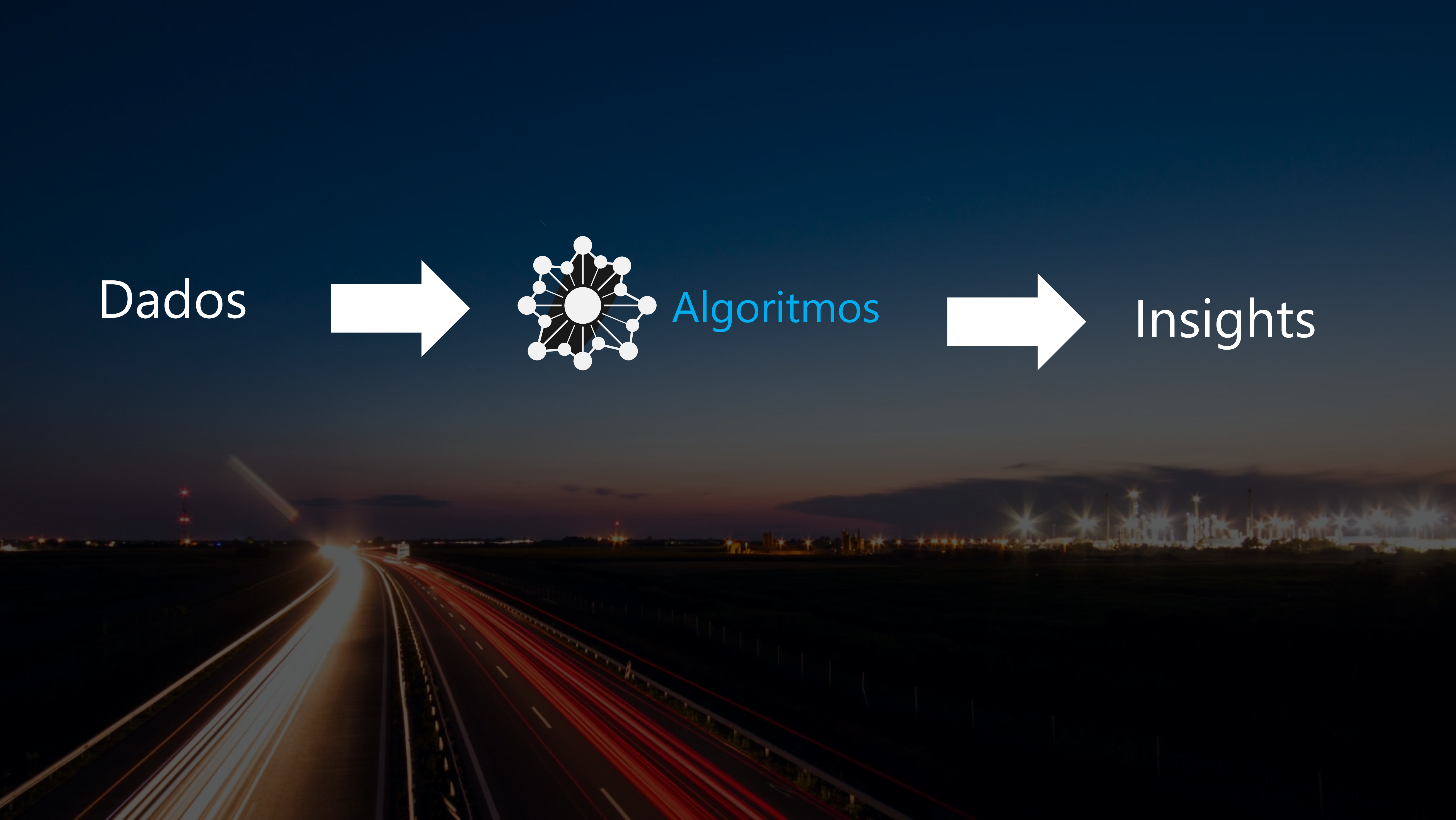
Conhecimento

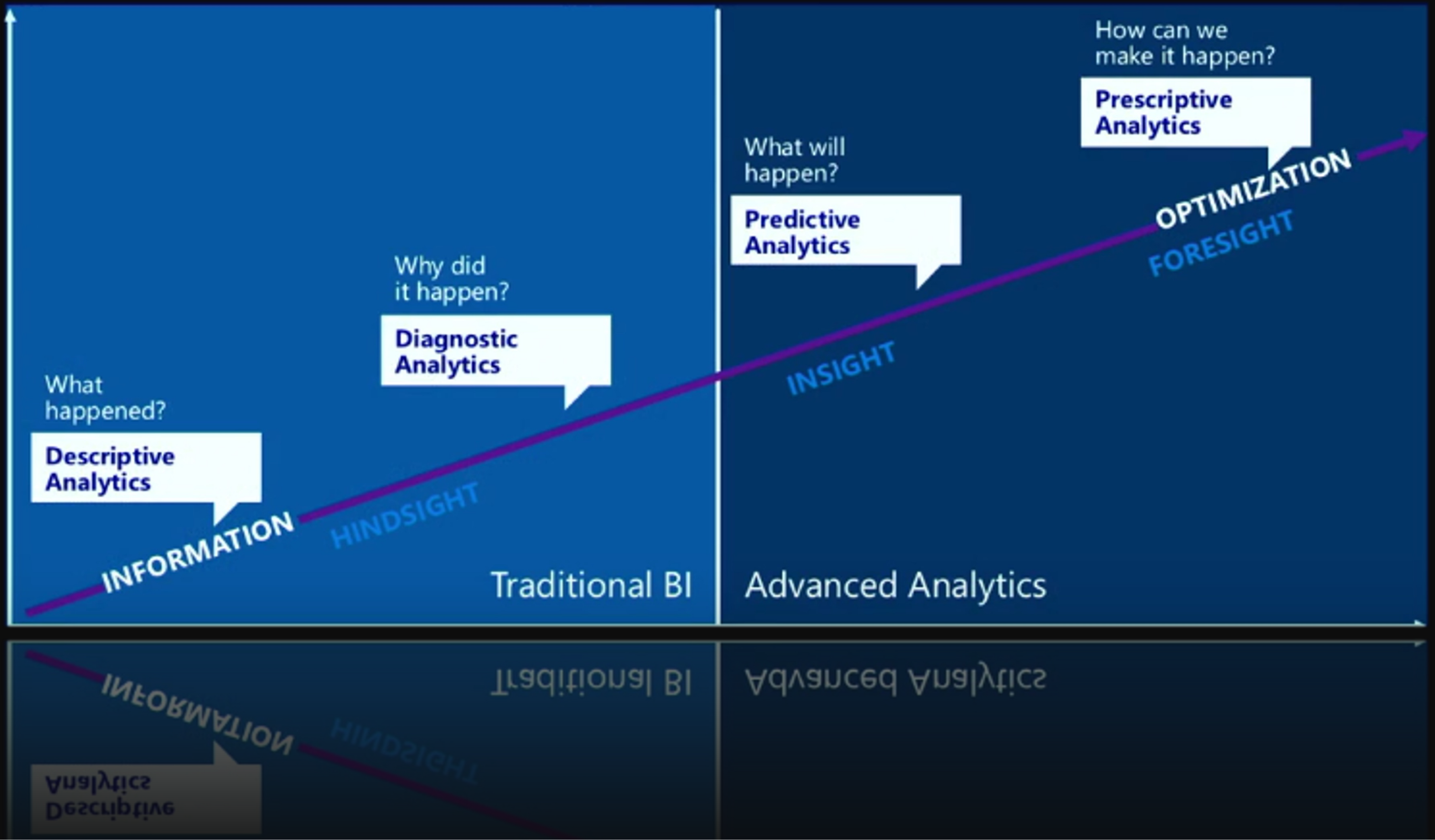


- Data typically comes before information, but it's hard to say which is more useful.
- For example, if the information was processed or organized in a biased manner or incorrectly, it's not useful, but the data still is.



Dados => Informação => Insights


Dados
Algoritmos
Insights


How can we
make it happen?
Prescriptive
Analytics
What will
happen?
Predictive
Analytics
Why did
it happen?
Diagnostic
Analytics
What
happened?
Descriptive
Analytics
INFORMATION
HINDSIGHT
INSIGHT
OPTIMIZATION
FORESIGHT
Traditional BI
Advanced Analytics

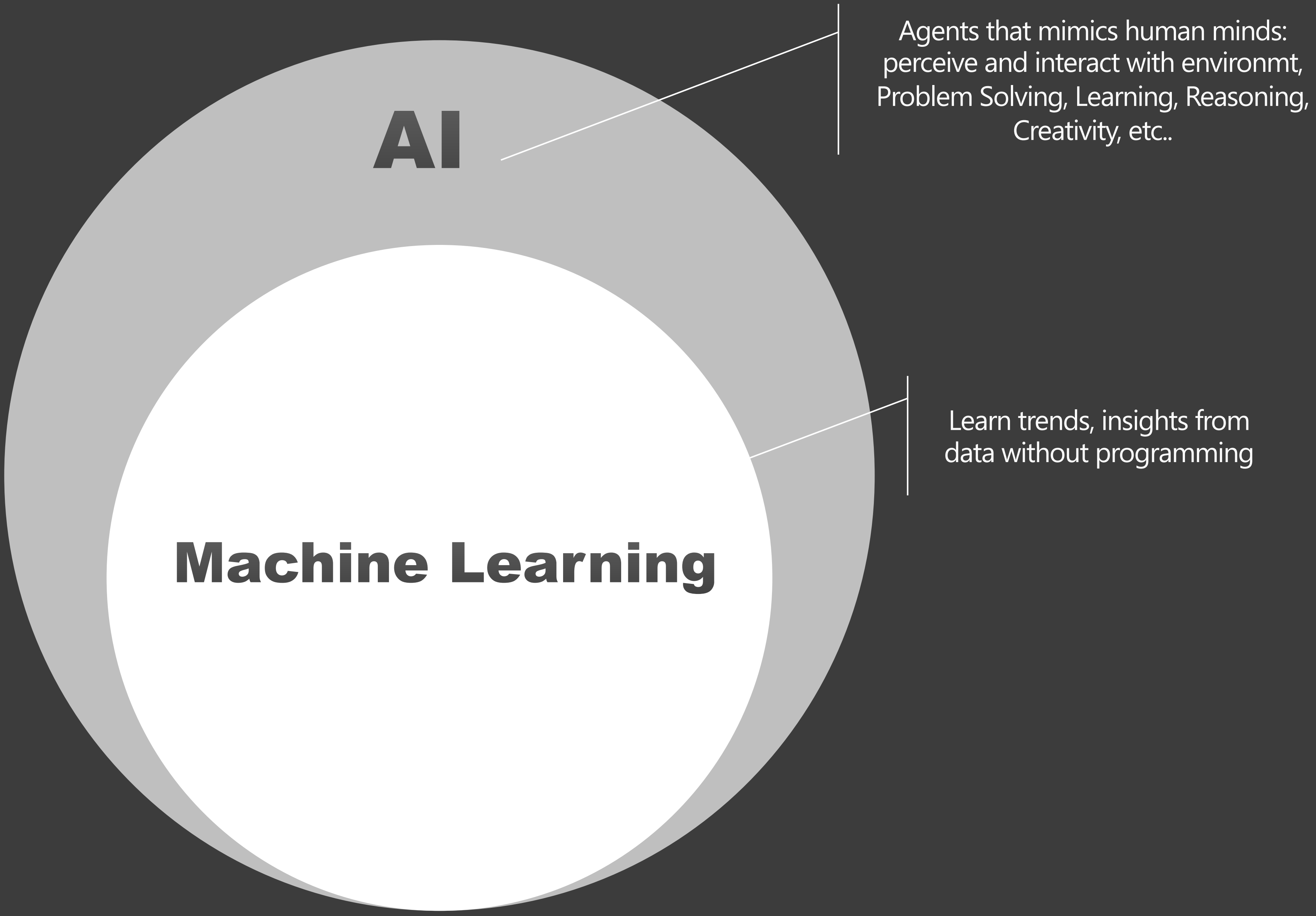
Inteligência Artificial (AI)

"Ability of a machine to perform cognitive functions we associate with human minds, such as **perceiving**, **reasoning**, **learning**, **interacting** with the environment, **problem solving**, and even exercising **creativity**."

machine learning

"Field of study that gives computers the ability to **learn without** being explicitly **programmed**."

Arthur Samuel (MIT engineer) - 1959


AI
Machine Learning
Agents that mimics human minds: perceive and interact with environmt, Problem Solving, Learning, Reasoning, Creativity, etc..
Learn trends, insights from data without programming

machine learning

aprender **com os dados**

machine learning

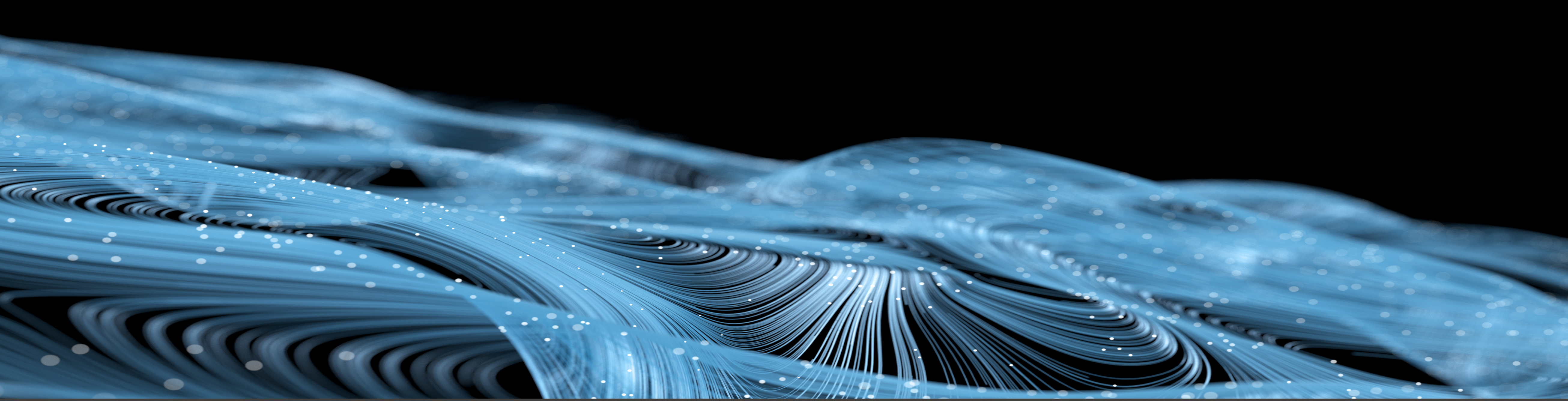
aprender **com os dados históricos**
a fim de identificar tendências

As máquinas deixarão de ser programadas e passarão a ser **treinadas**

Ai is the new electricity

"About 100 years ago, electricity transformed every major industry. AI has advanced to the point where it has the power to transform" every major sector in coming years"

Andrew Ng – Stanford Professor

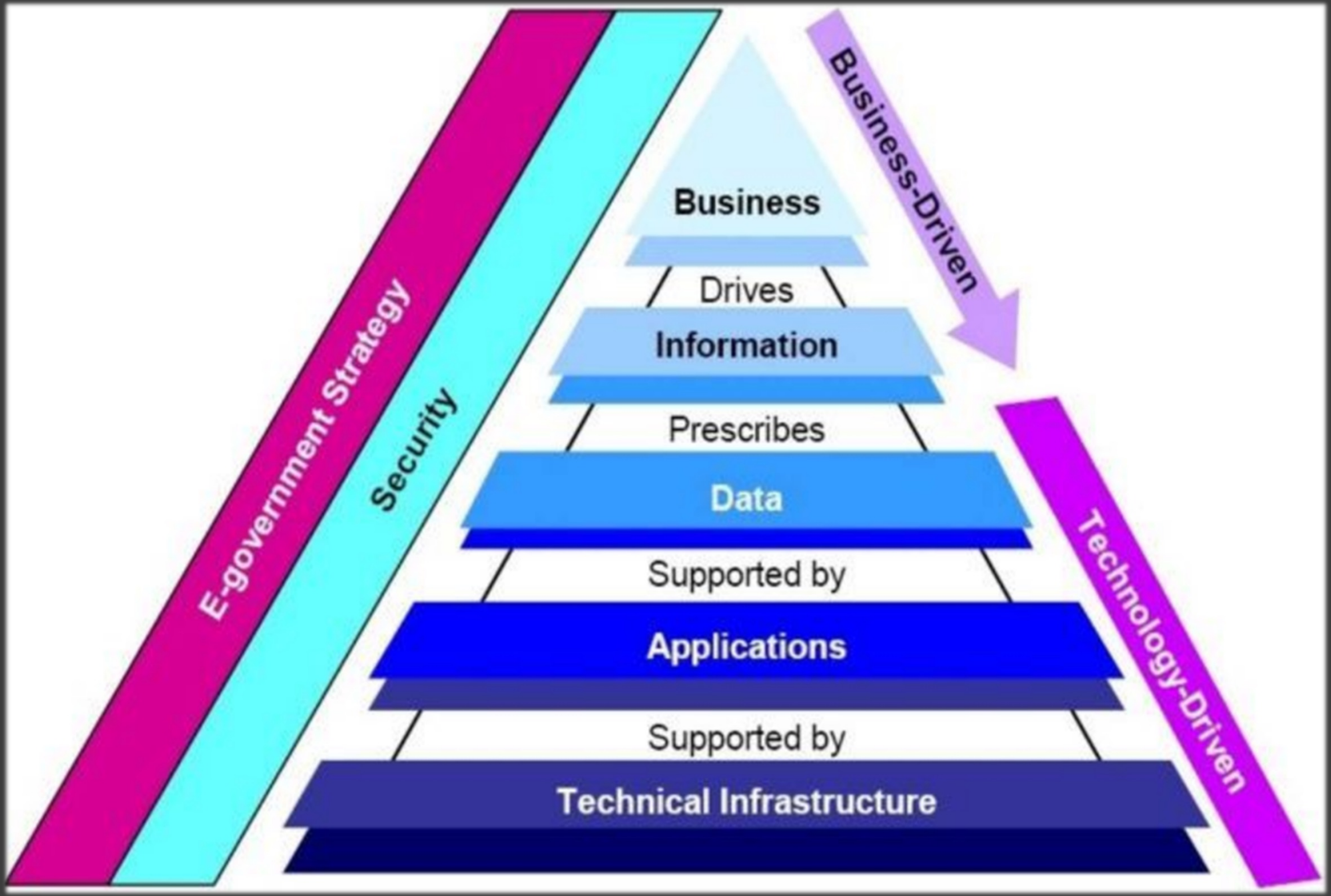


Dados são ativos

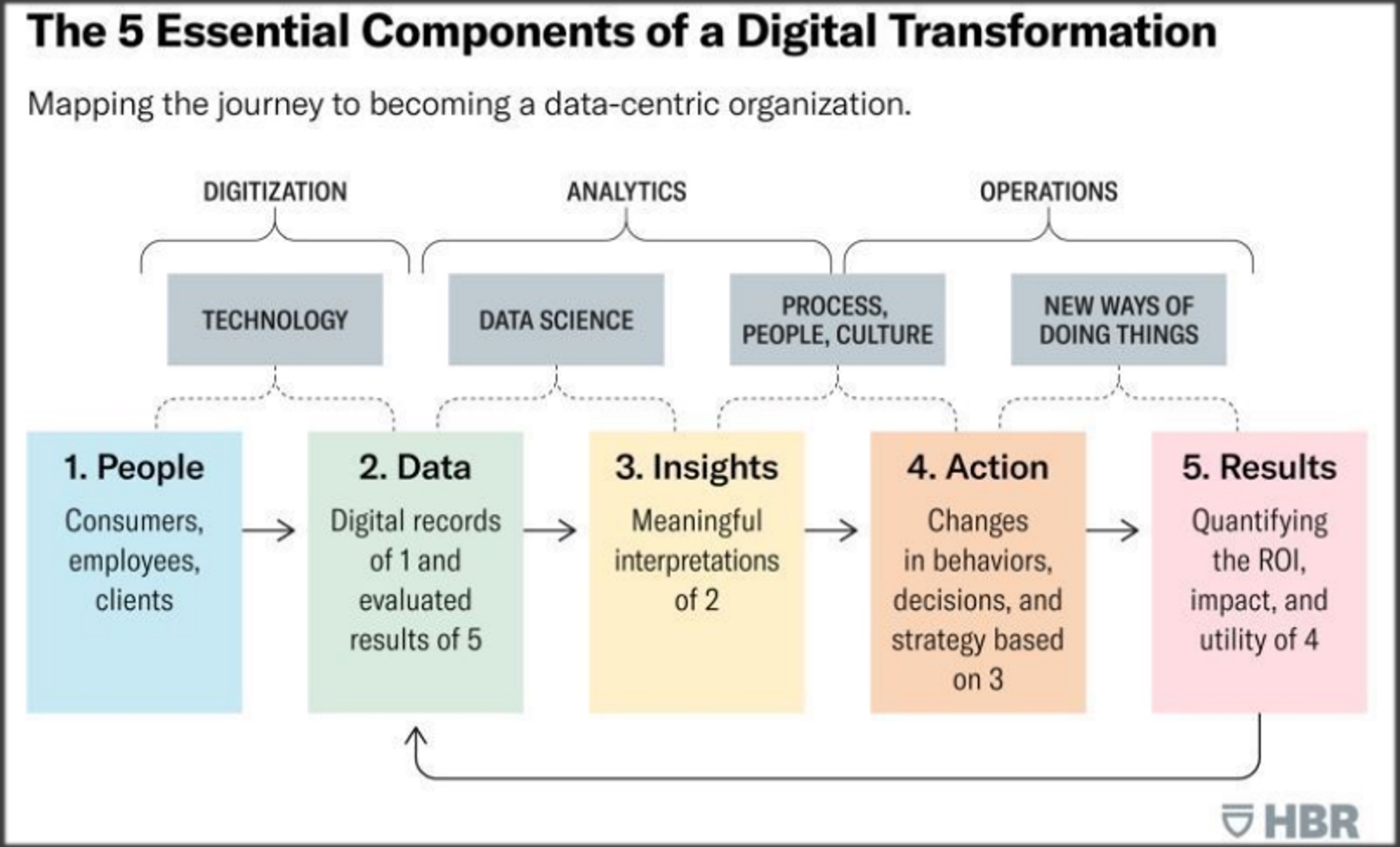
- **Ativo** é um recurso econômico que detém ou produz valor
- São vistos como propriedade e que **podem ser convertidos em valor financeiro**
- Os dados não são apenas necessários para as operações de negócio, mas também podem fornecer informações sobre usuários, produtos e serviços e **criar vantagem competitiva**.
- Gerenciamento de dados é permite que as organizações **obtenham valor com seus dados**, assim como o gerenciamento eficaz de ativos financeiros e físicos permite que as organizações obtenham valor com esses ativos.

- Pesquisas mostram que **poucas organizações tratam seus dados como um ativo** e para muitas eles podem até ser um passivo
- **A falha no gerenciamento de dados é semelhante à falha no gerenciamento de capital**: desperdício e em oportunidades perdidas. Além do mais, dados mal gerenciados apresentam riscos éticos e de segurança
- A obtenção de valor dos dados **não ocorre no vácuo ou por acidente**, pois além de gerenciamento, ela requer compromisso organizacional e liderança

Data is key to innovate



Data is key to innovate



Ciclos rápidos de inovação: vantagem estratégica

[data + analytics + people] @ *speed*

The inhibitors of digital transformation

80%
are unable to **collaborate** on common data

84%
say **fragmented data** gets in the way

more than **9** out of **10**
require **faster data and analytics** to compete

Harvard Business Review

/ business

Workers waste half their time as they struggle with data

As data grows in complexity, data workers waste time searching for and preparing data instead of gaining insights according to a new report.

Written by Eileen Brown, Contributor on July 9, 2019

Data is becoming increasingly important to success in the digital economy and data workers spend the majority of their work week on data activities.

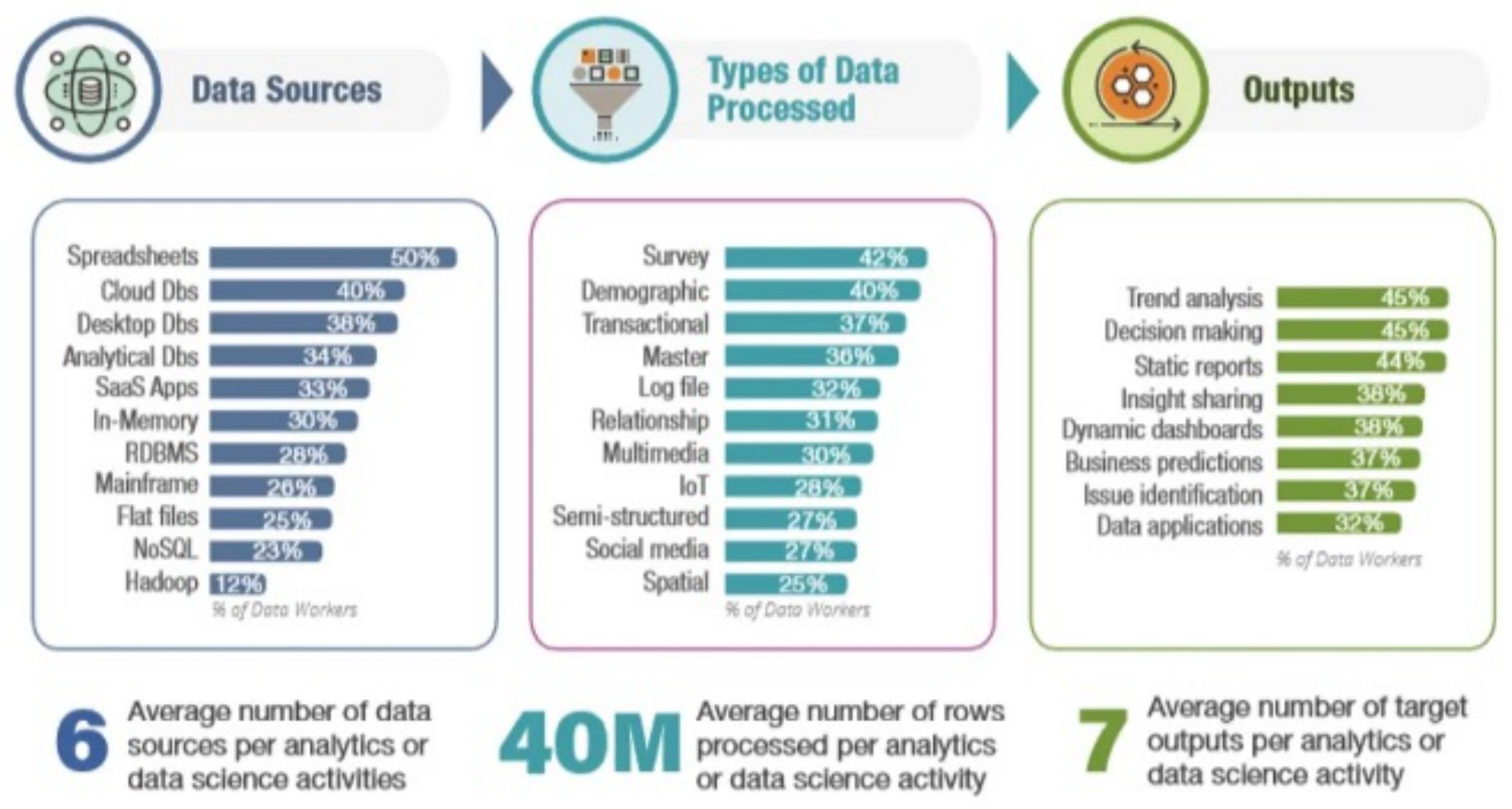
/ must read

It's that time: Here's every major retailer's return policy

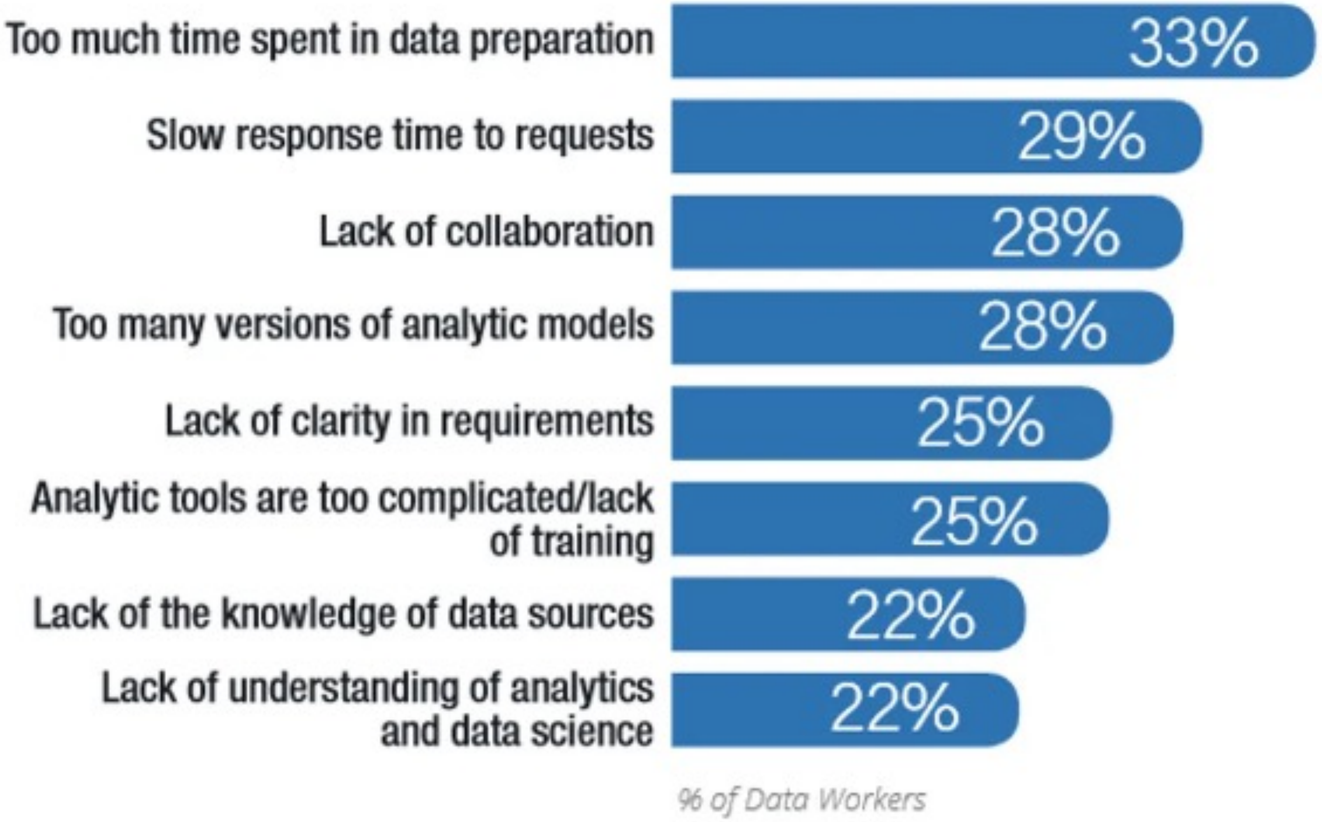
Read now →

/ related

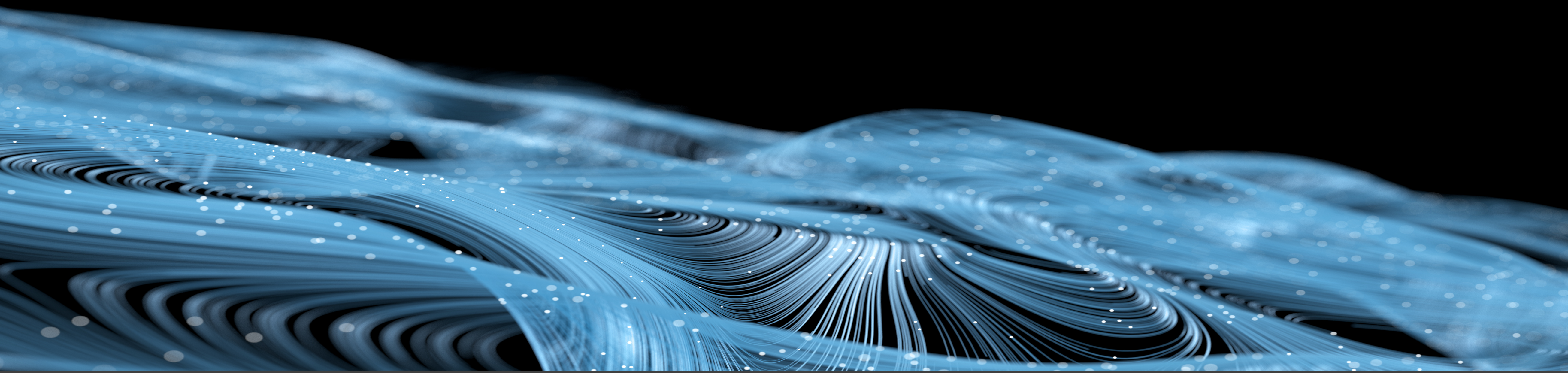
Qualcomm readies Wi-Fi 7 platform for your home mesh networking needs


Data Sources
Types of Data Processed
Outputs
Spreadsheets 50%
Cloud Dbs 40%
Desktop Dbs 38%
Analytical Dbs 34%
SaaS Apps 33%
In-Memory 30%
RDBMS 28%
Mainframe 26%
Flat files 25%
NoSQL 23%
Hadoop 12%
% of Data Workers
Survey 42%
Demographic 40%
Transactional 37%
Master 36%
Log file 32%
Relationship 31%
Multimedia 30%
IoT 28%
Semi-structured 27%
Social media 27%
Spatial 25%
% of Data Workers
Trend analysis 45%
Decision making 45%
Static reports 44%
Insight sharing 38%
Dynamic dashboards 38%
Business predictions 37%
Issue identification 37%
Data applications 32%
% of Data Workers
6 Average number of data sources per analytics or data science activities
40M Average number of rows processed per analytics or data science activity
7 Average number of target outputs per analytics or data science activity

Alteryx


Too much time spent in data preparation
33%
Slow response time to requests
29%
Lack of collaboration
28%
Too many versions of analytic models
28%
Lack of clarity in requirements
25%
Analytic tools are too complicated/lack of training
25%
Lack of the knowledge of data sources
22%
Lack of understanding of analytics and data science
22%
% of Data Workers

Alteryx



Tipos de dados

Tipos de dados

DATA TYPE	DEFINITIONS	IMPORTANCE AND SIGNIFICANCE	RISK
Personal	Also known as Personally Identifiable Information (PII). According to the government of the United States, PII is information which can be used to distinguish or trace an individual's identity when used alone, such as name, social security number, and biometric records, or when combined with other personal or identifying information linked to a specific individual, such as date and place of birth, and mother's maiden name.	PII collected from consumers and end users of a product and service have become a valuable asset for technology vendors focusing on AI. It reveals the external behavior and tendency, as well as inner preference and biases of an individual, a community, or a certain population demographic during all decision-making processes.	Without proper data governance, personal data can be used by malicious actors for unlawful impersonation, micro-targeted advertising, and to potentially manipulate individual preferences. Sensitive PII collected from citizens in one country may be shared or sold to a company in a different country. This grants permission to foreign entities to access valuable political, economic, social, and geographical data, opening doors for foreign entities to identify, exploit, and expose the weak links in the country.
Machine Generated	Any data generated automatically by a sensor, device, or equipment connected to a local server, cloud service, or other. The data can exist as time series, image, video, and audio.	As the Internet of Things (IoT) market continues to grow, the large-scale communications between machines provide deep operational insight into the status of a machine, a production line, and sometimes the entire workflow. These IoT communication data can be used for conditional-based monitoring, predictive maintenance, and overall efficiency enhancement of industrial and manufacturing processes.	When not appropriately managed, sensitive or confidential operational data can be exposed, resulting in the loss of valuable patents, trade secrets, blueprints, or even contracts. Also, if this data falls into the hands of malicious actors, it could be tweaked to create false outcomes or representations, negatively interfering with the overall business and industrial/manufacturing processes and put the entire business at severe risk
Geospatial	Any data related to both space and time, that provides information on specific regions or events. A key feature of geospatial data is that they are not static. Instead, these data are constantly being updated and change over time.	Aside from machine-generated data, geospatial data is seeing massive growth due to the rise of autonomous mobile vehicles, robots, and drones. One autonomous vehicle can generate up to four terabytes of data per day through various sensors, such as high-resolution cameras, radar, LiDAR, and GPS.	Geospatial data can be extremely sensitive as it contains information on real-time traffic, high-definition landscape maps, strategic premises, infrastructure, etc. The collection, storage, and transfer of these data create concerns around proprietary data operation for mission-critical data, data sovereignty, and cybersecurity concerns.
Metadata	Any data that provides information about other data. This includes information on the purpose of the data, time, date, and location of creation, creator of the data, standards, the process used, and data trends.	Companies are major consumers of metadata. The proliferation of AI models witnesses the boom in metadata generated by AI during the data analytics process. When going through a large amount of data, AI generates metadata that informs end-users on the patterns and trends observed from the data.	Similar to machine-generated data, sensitive or confidential operational data can be exposed. This could cost the loss of valuable patents, trade secrets, blueprints, or even contracts. Also, if metadata are tempered with maliciously, they can create serious bias, which could seriously affect the prediction outcomes and lead to compromises and vulnerabilities in the business decision-making process.
Synthetic	Any high-quality data generated programmatically for data science and machine learning training and testing.	For specific ML applications, it's easier to create synthetic data than to collect and annotate actual data. Developers can generate as much synthetic data as needed to develop the models with the desired accuracy.	AI algorithmic bias can be introduced through unvalidated or low-quality synthetic data. This will result in AI algorithms being prone to error with the biased outcome, causing disastrous events for mission and business-critical applications.

Informação direta e indireta

- A informação direta é aquela que permite a imediata individualização da pessoa.
- A informação indireta permite por meio da reunião de informações, chegar à identificação do sujeito.

Exemplos

- identificação direta: um cliente, ao fazer uma compra online, informa seu nome completo e CPF, ou seja, a loja virtual com essas informações consegue identificar o indivíduo que realizou a compra.
- identificação indireta: uma empresa não cadastrou o nome completo ou o CPF de um cliente, a princípio, a companhia não teria como identificá-lo. Porém, utilizando outras informações que possui, é possível descobrir sua identidade. Tais como: profissão, endereço, gênero, ou qualquer outro dado que ajude a identificá-lo.

- **Em outras palavras, a identificação indireta ocorre quando associamos informações, que isoladamente não conseguem identificar um indivíduo, para descobrimos a identidade de uma pessoa**.

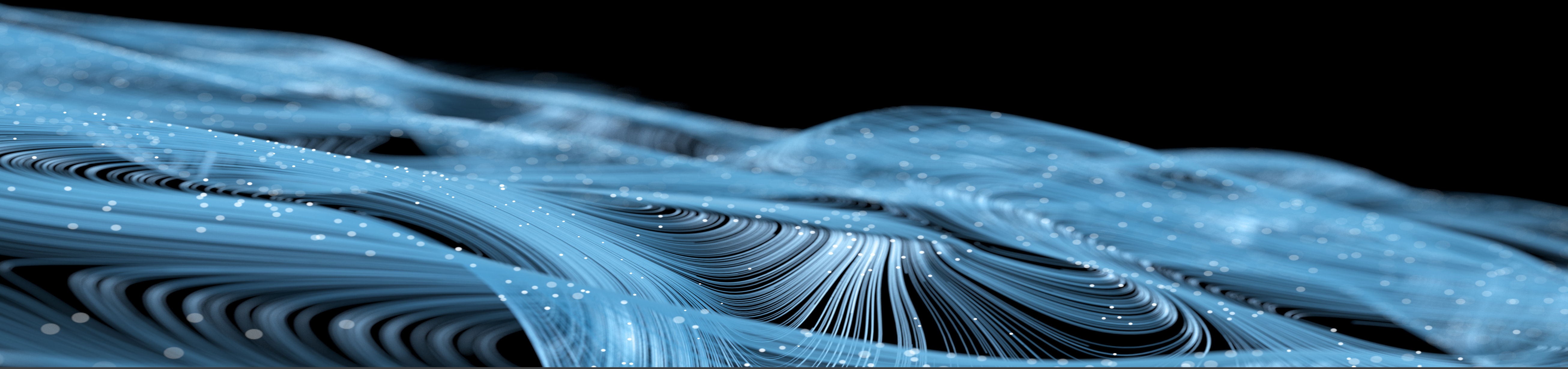
PII- Personally identifiable Information

termo usado em segurança da informação, referem-se a informações que podem ser usadas para identificar, contactar ou localizar uma única pessoa. Também podem ser usadas com outras fontes para identificar um único indivíduo.

- Nome completo (se não for comum)
- Número da Cédula de identidade
- Número do Cadastro de Pessoas Físicas
- Número do Título eleitoral
- Endereço IP (em alguns casos)
- Número de Placa de identificação de veículos
- Número da Carteira Nacional de Habilitação
- Rosto, Impressões digitais, ou manuscritos
- Número do Cartão de crédito
- Numero do celular
- Data de Nascimento
- Nome da mãe
- Local de Nascimento
- Informações Genéticas

EU directive 95/46/EC

Article 2a: 'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, **directly or indirectly**, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity



Informações Indiretas & Privacidade

Exemplo: PII- User data

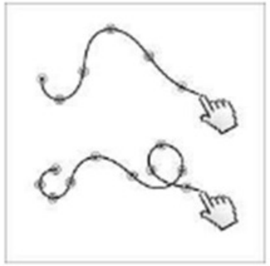
JUNE 23, 2021

Mouse movements reveal your behavior

by University of Luxembourg

Credit: Pixabay/CC0 Public Domain

In two recently published research papers, computer scientists from the University of Luxembourg and international partners show how mouse movements can be used to gain additional knowledge about the user behavior. While this has many interesting applications, mouse movements can also reveal sensitive information about the users such as their age or gender. Scientists want to raise awareness about these potential privacy issues and have proposed measures to mitigate them.



Prof. Luis Leiva from the University of Luxembourg and corresponding author of the two papers explains in more details the key findings.

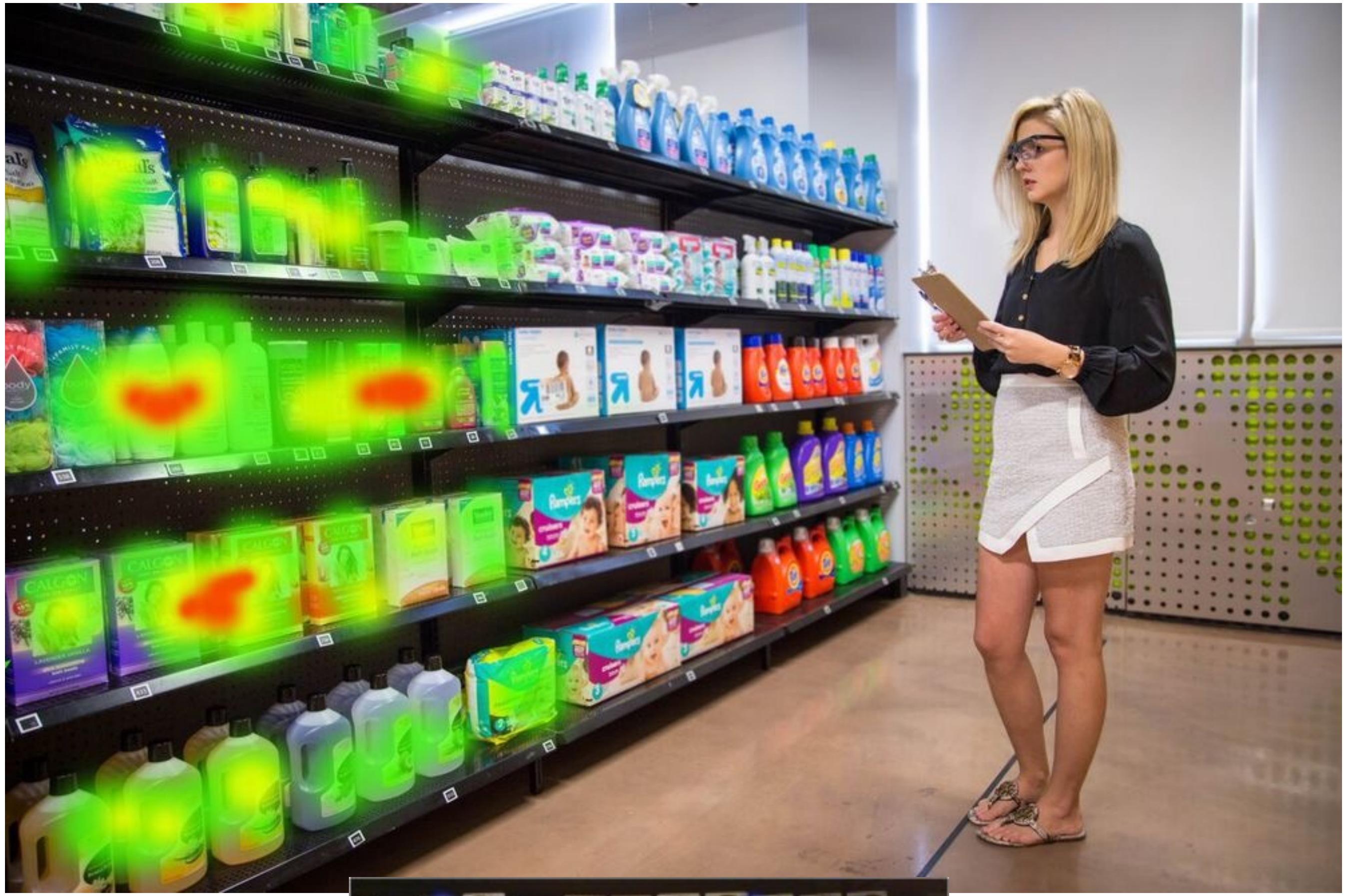
My mouse, my rules

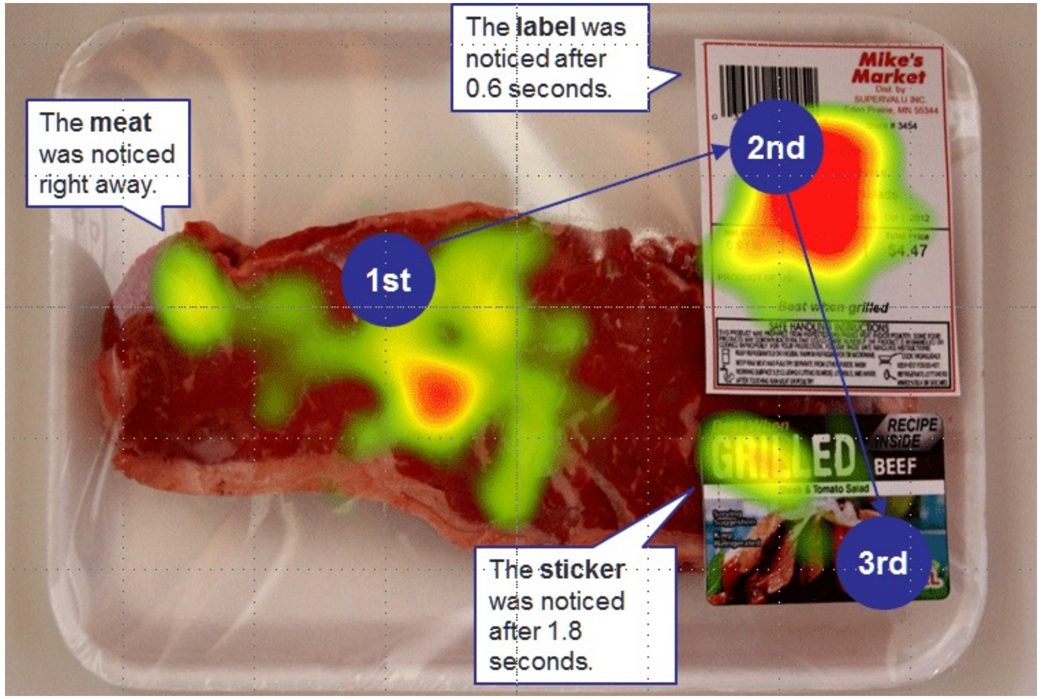
"We have demonstrated how straightforward it is to capture behavioral data about the users at scale, by unobtrusively tracking their mouse cursor movements, and predict user's demographics information with reasonable accuracy using five lines of code. For years, recording mouse movements on websites has been easy, however to analyze them one would need advanced expertise in computer science and machine learning. Today, there are many libraries and frameworks that allows anyone with a minimum of programming knowledge to create rather sophisticated classifiers. This raises new privacy issues and users do not have an easy opt -out mechanism."

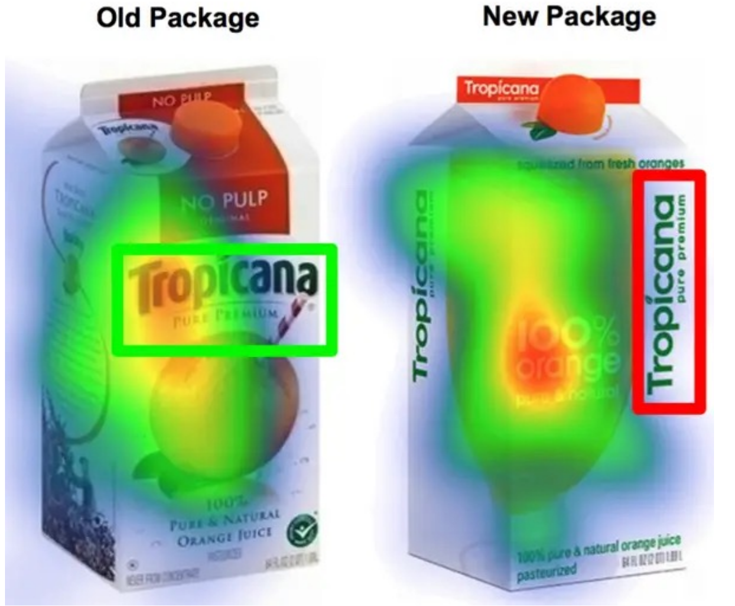
Browser tracking (Direct and meta information)

- Browsing history
- Geo-Location
- Hardware and Software Information
- Social Media Log-ins
- Image data (metafinformation)
- Fonts & Language
- Mouse clicks and hovers ("heat maps" for sales analytics)

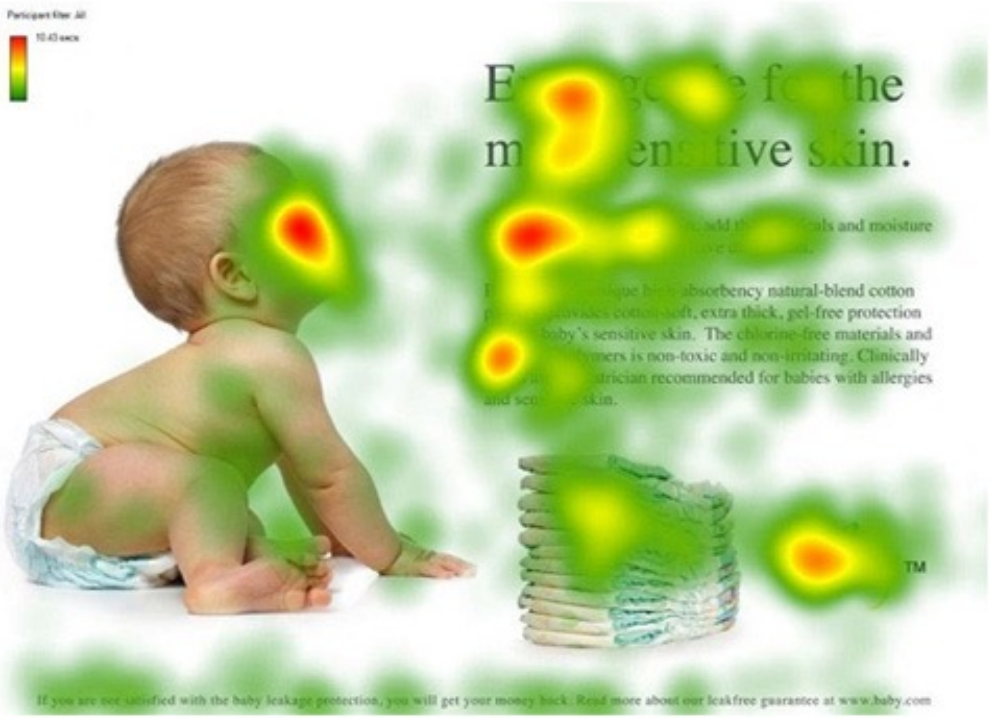
Eye Tracking



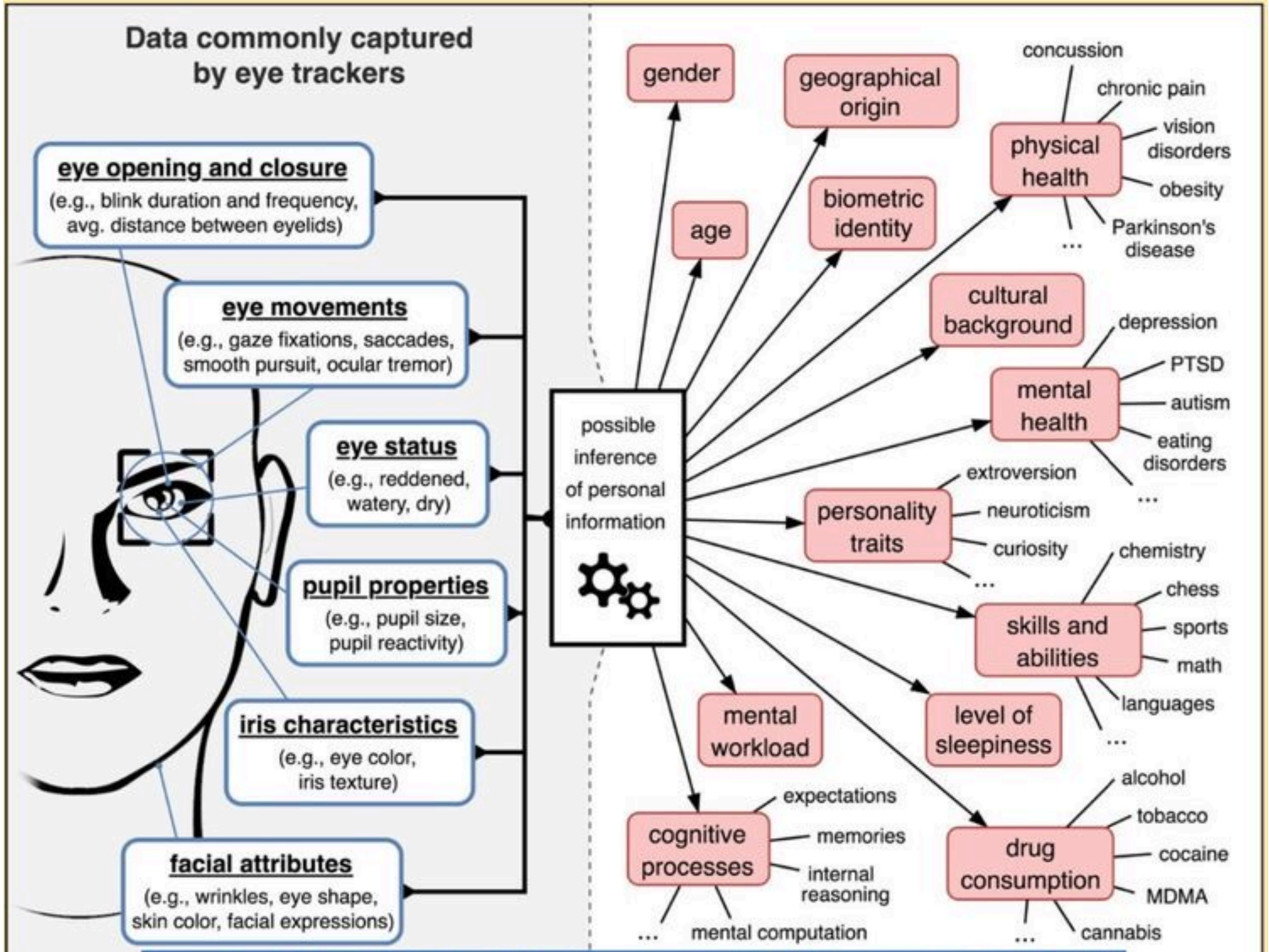




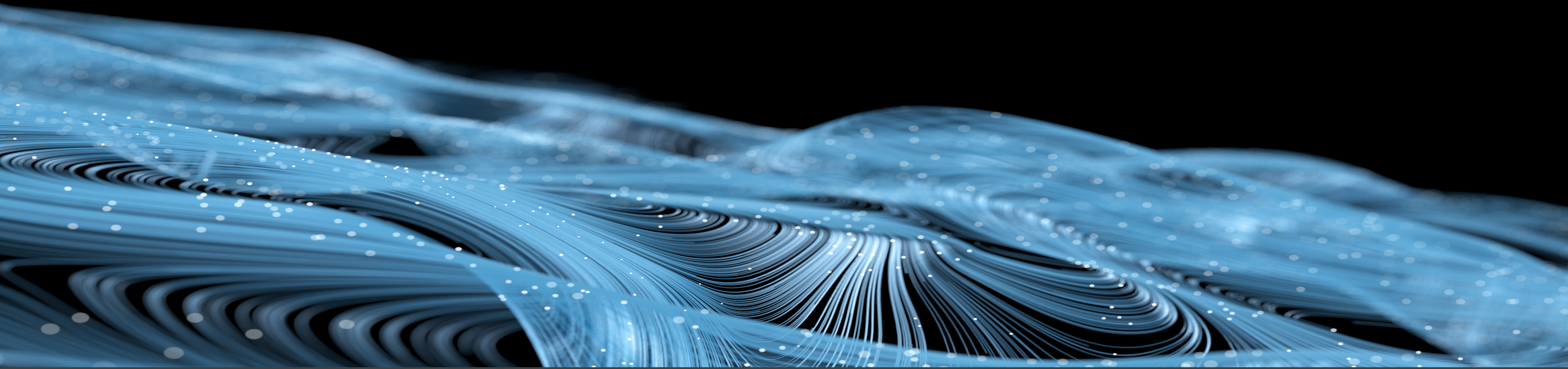
Eye Tracking



Eye Tracking

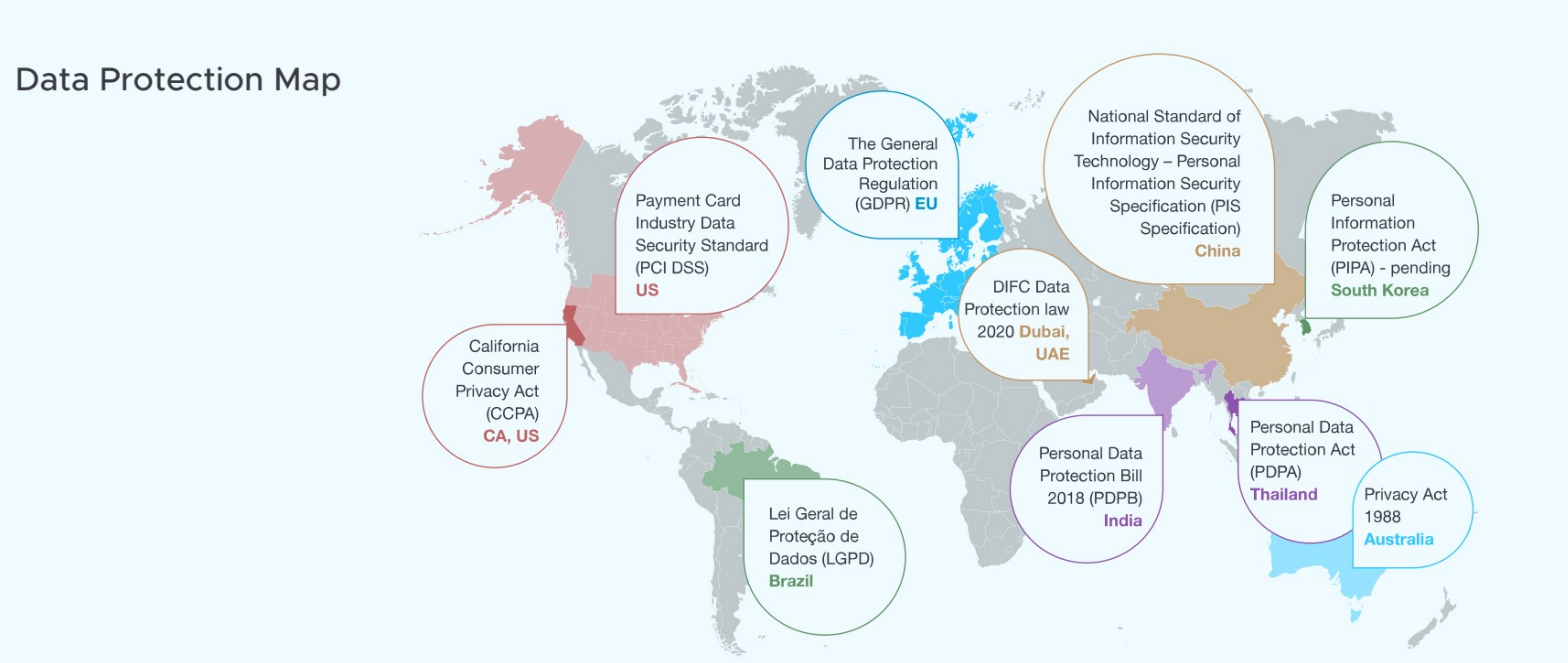


Source: Kröger et al. (2020) „What Does Your Gaze Reveal About You? On the Privacy Implications of Eye Tracking"
In: *Privacy and Identity Management*. Springer Nature - https://doi.org/10.1007/978-3-030-42504-3_15



Regulamentações

Regulations



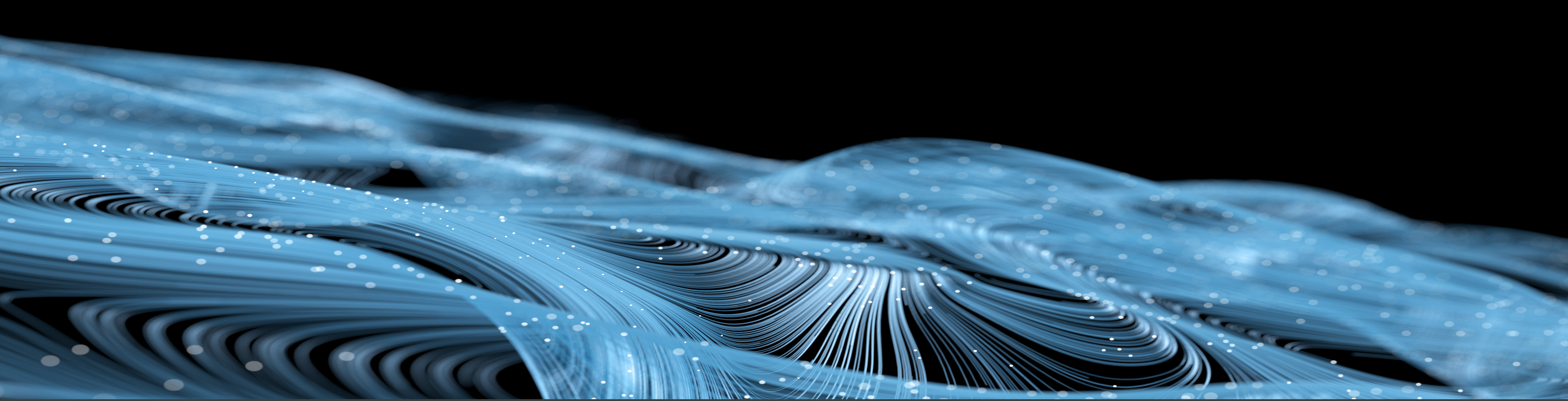
Regulations

Sarbanes Oxley Act	The Sarbanes–Oxley Act of 2002 is a United States federal law that mandates certain practices in financial record keeping and reporting for corporations.
Basel I	Basel I is the round of deliberations by central bankers from around the world, and in 1988, the Basel Committee on Banking Supervision (BCBS) in Basel, Switzerland, published a set of minimum capital requirements for banks.
Basel II	Basel II is the second of the Basel Accords, (now extended and partially superseded by Basel III), which are recommendations on banking laws and regulations issued by the Basel Committee on Banking Supervision.
HIPAA	The Health Insurance Portability and Accountability Act of 1996 (HIPAA or the Kennedy–Kassebaum Act[1][2]) is a United States federal statute enacted by the 104th United States Congress and signed into law by President Bill Clinton on August 21, 1996. It modernized the flow of healthcare information, stipulates how personally identifiable information maintained by the healthcare and healthcare insurance industries should be protected from fraud and theft, and addressed some limitations on healthcare insurance coverage.
GDPR	The General Data Protection Regulation (EU) 2016/679 (GDPR) is a regulation in EU law on data protection and privacy in the European Union (EU) and the European Economic Area (EEA). The GDPR is an important component of EU privacy law and of human rights law, in particular Article 8(1) of the Charter of Fundamental Rights of the European Union.
cGMP	Good manufacturing practices (GMP) are the practices required in order to conform to the guidelines recommended by agencies that control the authorization and licensing of the manufacture and sale of food and beverages,[1] cosmetics,[2] pharmaceutical products,[3] dietary supplements,[4] and medical devices.[5] These guidelines provide minimum requirements that a manufacturer must meet to assure that their products are consistently high in quality, from batch to batch, for their intended use.
LGPD	The General Personal Data Protection Law (Brazil) 13709/2018 (Portuguese: Lei Geral de Proteção de Dados Pessoais, or LGPD), is a statutory law on data protection and privacy in the Federative Republic of Brazil. The law's primary aim is to unify 40 different Brazilian laws that regulate the processing of personal data.[1] The LGPD contains provisions and requirements related to the processing of personal data of individuals, where the data is of individuals located in Brazil, where the data is collected or processed in Brazil, or where the data is used to offer goods or services to individuals in Brazil.[2]

Lei Geral de Proteção de Dados Pessoais (LGPDP)

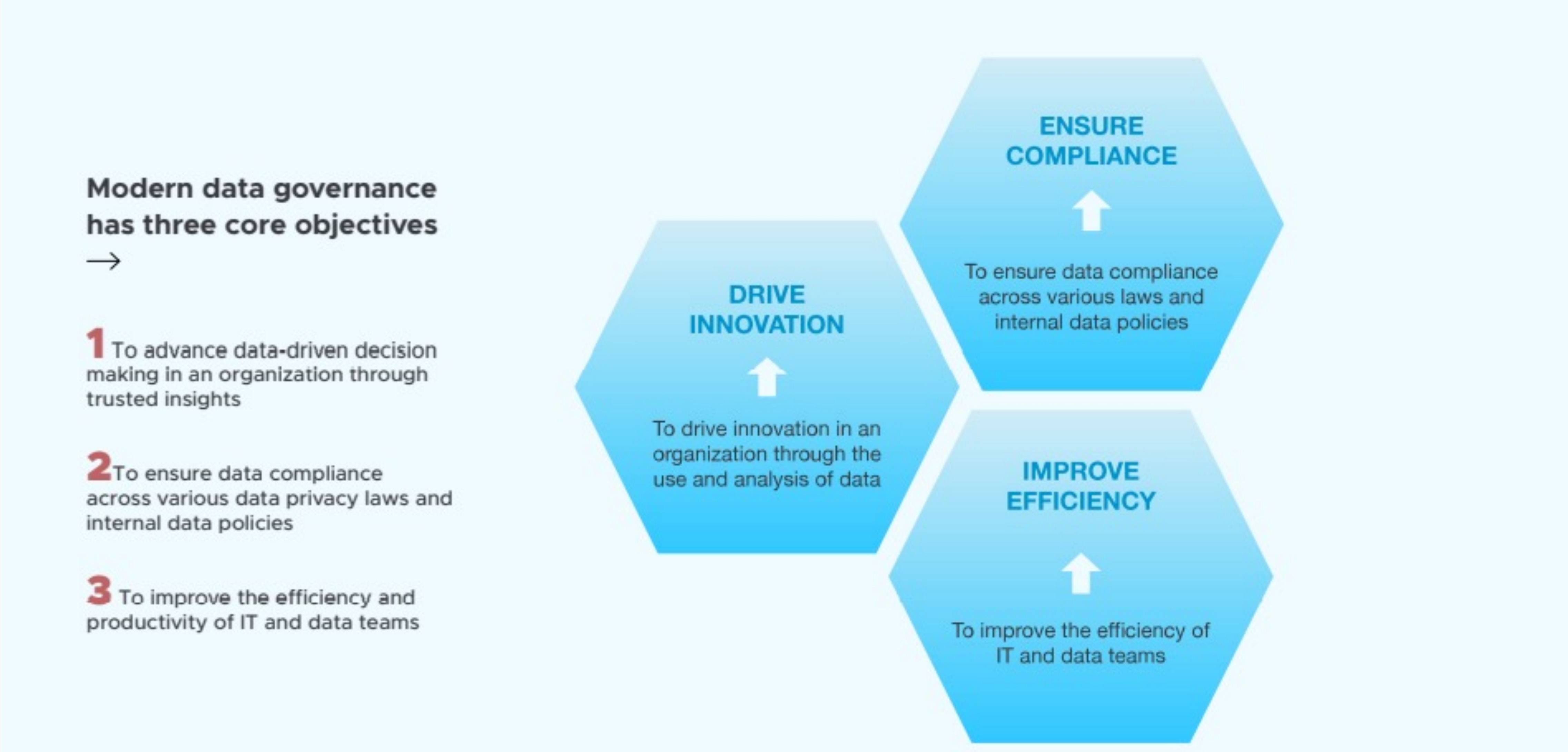
Fornece as diretrizes de como os dados pessoais dos cidadãos podem ser coletados e tratados e altera a Lei 12.965/14 (Marco Civil da Internet).

CICLO DE VIDA DOS DADOS		
Fase do Ciclo	**Antes da LGPD**	**Com a LGPD**
Coleta	Os dados pessoais são coletados indiscriminadamente.	Os dados pessoais coletados devem obedecer ao princípio da necessidade e da finalidade.
Processamento	Os dados podem ser processados sem um tratamento específico.	O processamento de dados só poderá ser realizado se o tratamento estiver enquadrado no Art. 7º da LGPD.
Análise	A análise de dados é feita para entender o mercado, conhecer o perfil das pessoas e definir estratégias para oferecer bens e serviços para o público-alvo.	A análise dos dados deve levar em consideração a finalidade da coleta. Devem ser obedecidos os princípios de tratamento com propósito legítimo específico e explicito.
Compartilhamento	Os dados pessoais são compartilhados sem a necessidade do consentimento de seus titulares.	O compartilhamento de dados deve ser consentido pelos seus titulares* *Ver Inciso II do Art. 3º do Decreto nº 10.046/2019.
Armazenamento	Os dados pessoais são reutilizados sem a necessidade de consentimento de seus titulares.	Os dados pessoais devem ser armazenados e mantidos por prazos definidos, ou seja, até que a finalidade seja alcançada ou deixem de ser necessários ou pertinentes ao alcance da finalidade.
Reutilização	Os dados pessoais são reutilizados sem a necessidade de consentimento de seus titulares.	Um novo consentimento deve ser solicitado sempre que houver mudança de finalidade.
Eliminação	Os dados pessoais são mantidos sem a obrigatoriedade de serem eliminados.	Os dados pessoais devem ser eliminados após o término de seu tratamento.

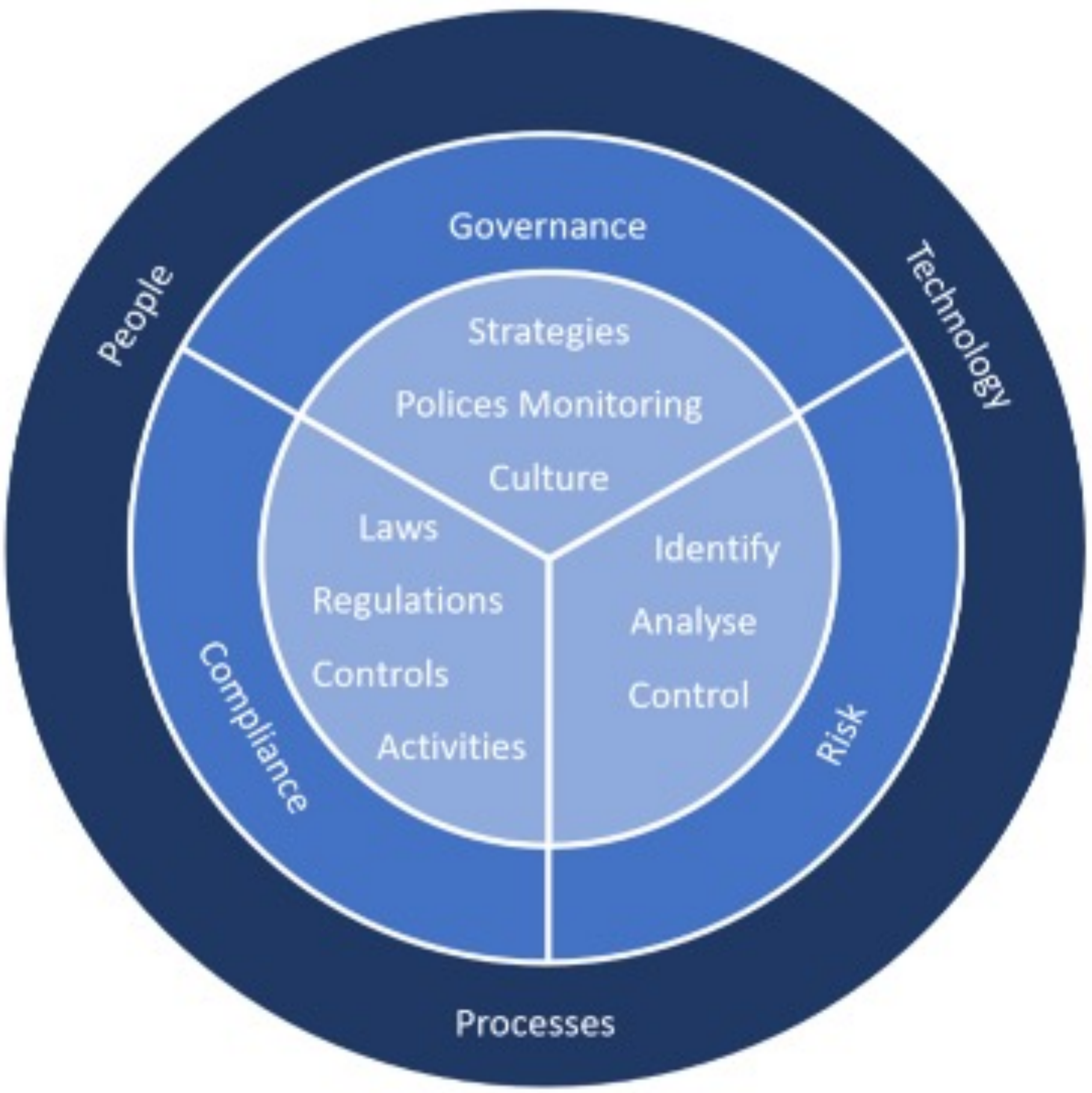


Governança de Dados

Objectives


Modern data governance has three core objectives
→
1 To advance data-driven decision making in an organization through trusted insights
2 To ensure data compliance across various data privacy laws and internal data policies
3 To improve the efficiency and productivity of IT and data teams
ENSURE COMPLIANCE
To ensure data compliance across various laws and internal data policies
DRIVE INNOVATION
To drive innovation in an organization through the use and analysis of data
IMPROVE EFFICIENCY
To improve the efficiency of IT and data teams

Gestão Ativa dos dados



Data Governance

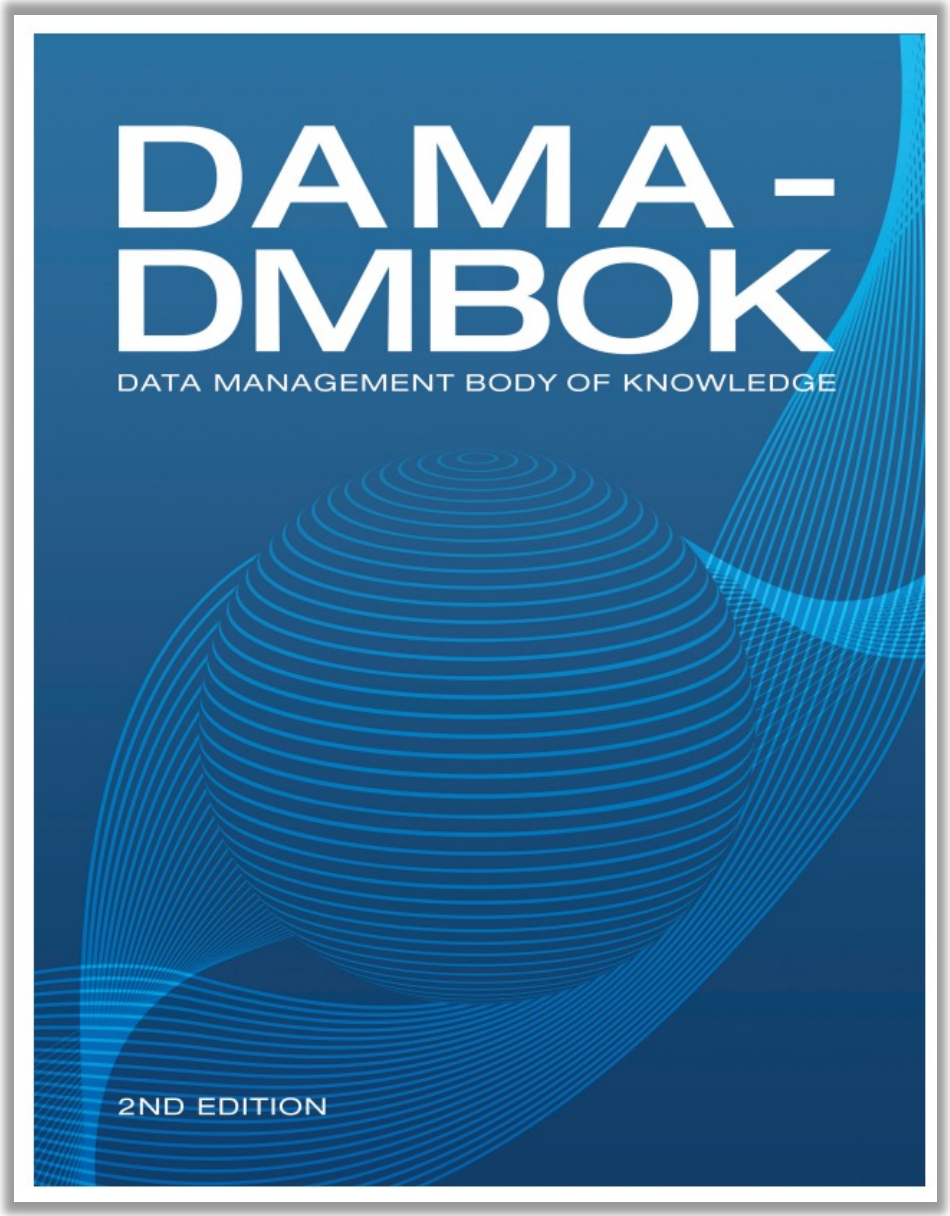
Data governance encompasses
the people, processes, and information
technology required to create a consistent
and proper handling of an organization's
data across the business enterprise.

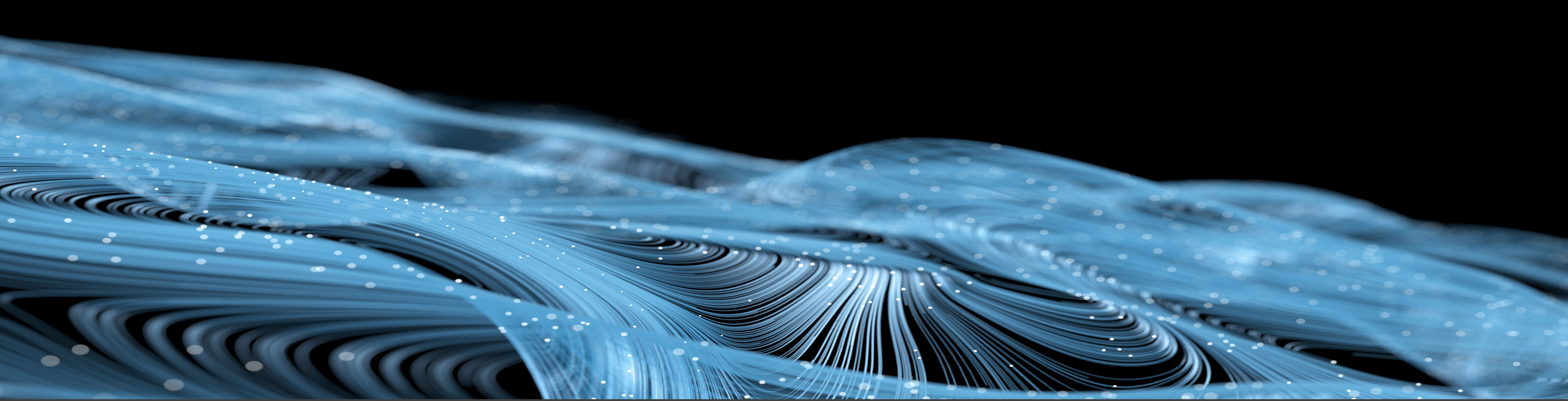
Availability
Usability
Consistency
Integrity
Security and Privacy

DAMA Framework : data wheel

DAMA International

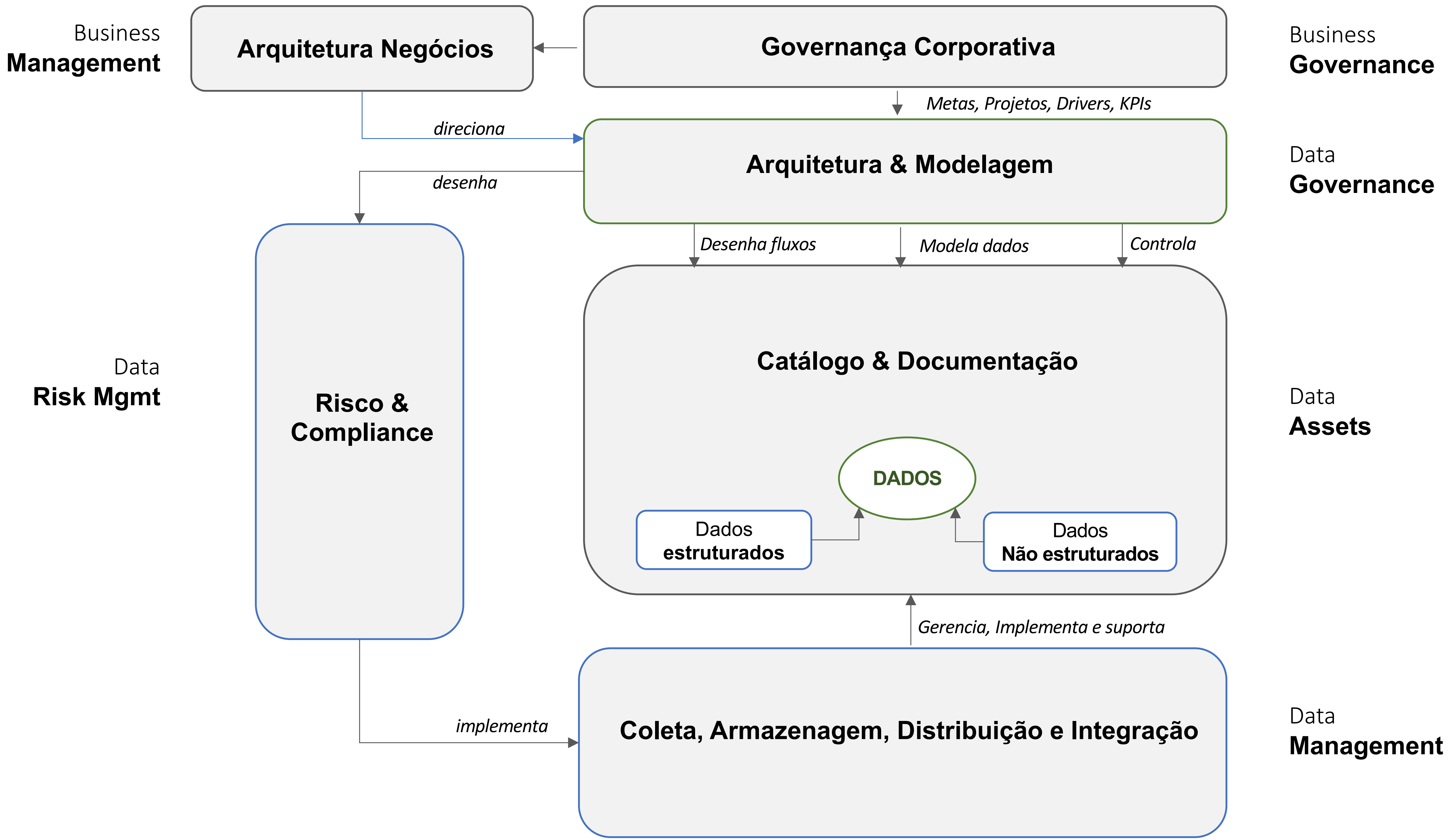
DAMA DMBOK



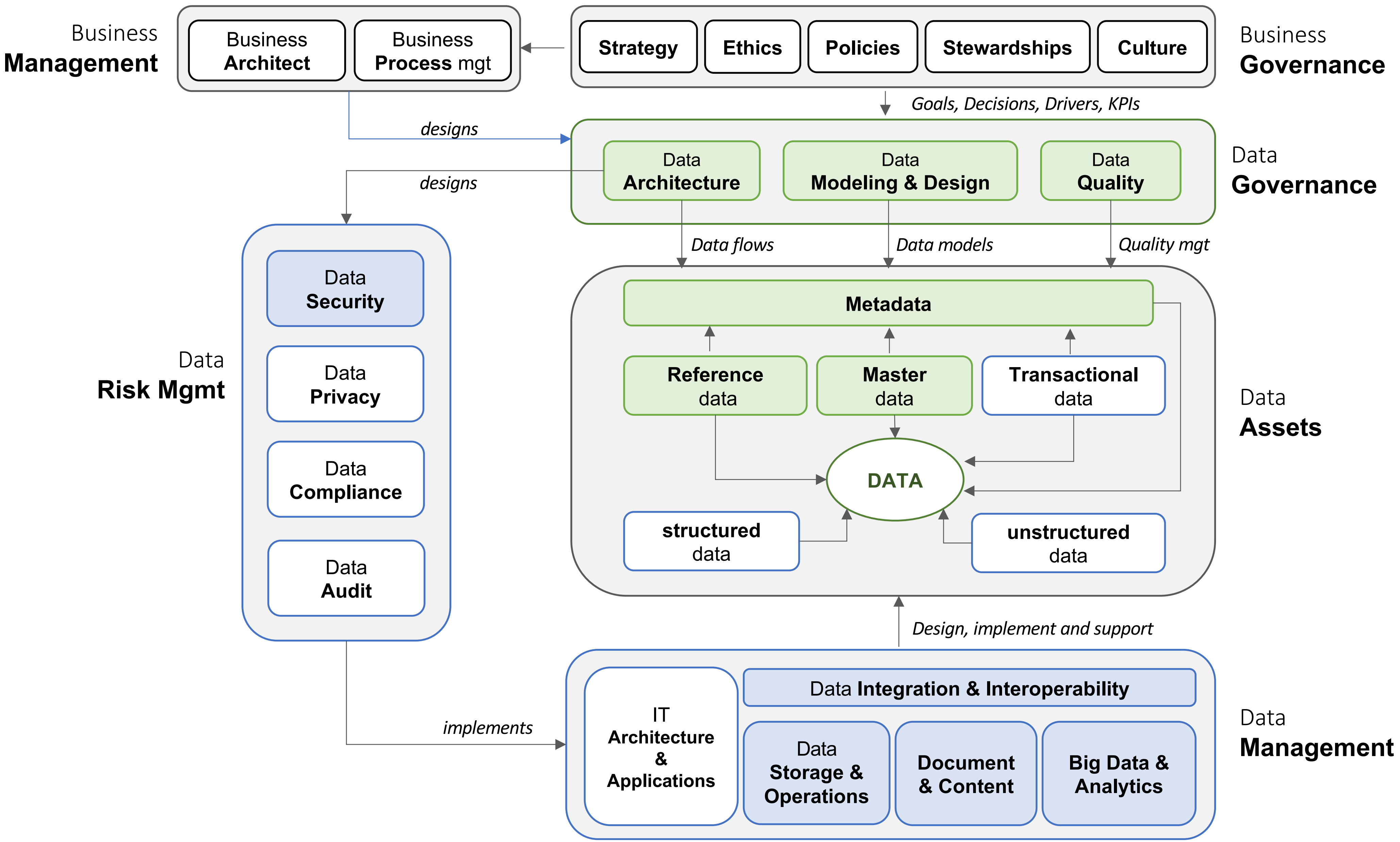


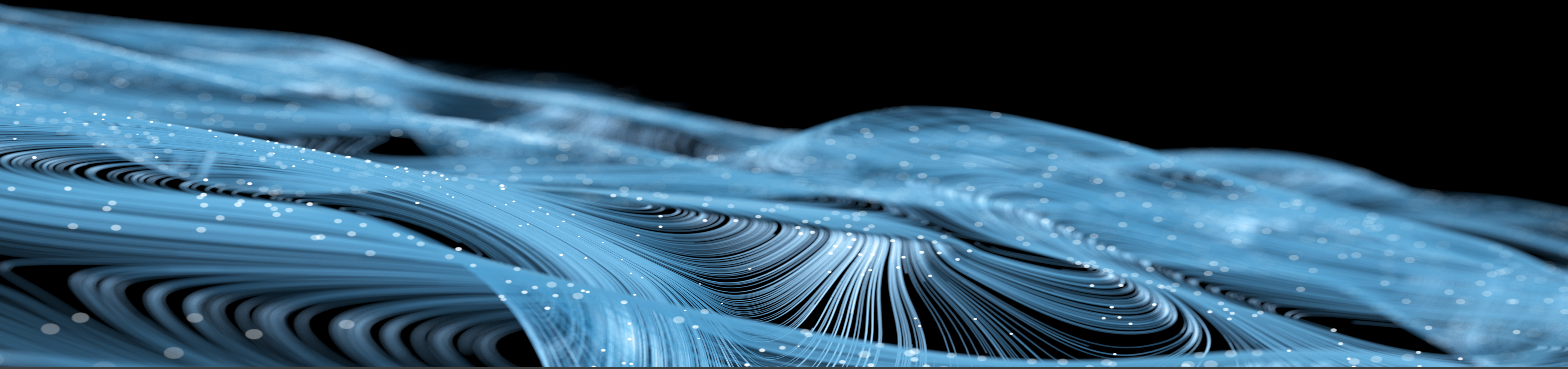
Visão Integrada

Governança de Dados



Data Governance



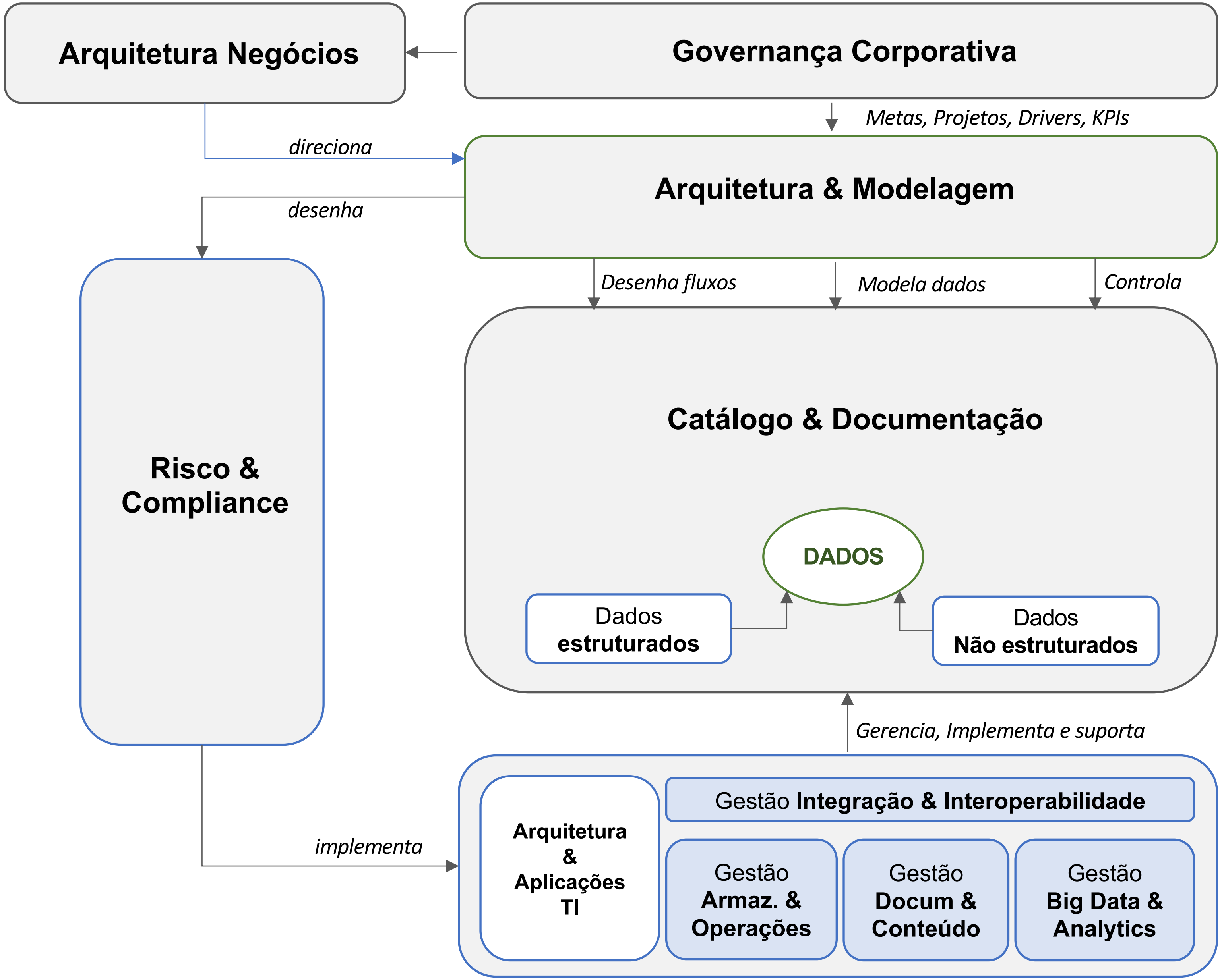


Módulos

modulo 1

Gestão
Armazenagem, Conteúdo e Integração

Armazenagem, Conteúdo e Integração



Armazenagem, Conteúdo e Integração

a) Gestão de Armazenagem e operações

Objetivo

- Visa a melhoria dos processos de gestão de armazenamento de dados, tais como planejamento, desenho, implementação e suporte de storages (SQL, No-SQL, datalakes,etc)
- Também inclui os serviços e processos de manutenção dos aplicativos, backups e implementação dos requisitos de segurança, privacidade e controle necessários durante todo o ciclo de vida
- O objetivo será prover a disponibilidade dos dados e garantir a integridade, confiabilidade e performance.

Plano de atuação

- Levantamento de indicadores e inventário interno atuais
- Elaborar roadmap de ações para implantação dos projetos de melhorias definidos no relatório de melhorias (“gaps”) tais como disponibilidade, monitoramento, governança, backup e auditoria para atendimento dos requisitos de qualidade e gestão de risco.
- Implementar o roadmap de ações e monitorar indicadores

Resultados

- Definição dos processos de governança
- Definição dos papéis e custodiantes (Data Stewardship)
- Execução do plano de melhorias
- Execução da implantação da gestão (inventário, processos, capacitação e monitoramento)

b) Gestão de Documentos e Conteúdo

Objetivo

- Visa a melhoria dos processos de gestão dos dados não estruturados (documentos, imagens, vídeos etc.) durante todo o ciclo de vida
- Dentre estes serviços destacam-se o controle, armazenagem e suporte. Também inclui os serviços de gestão de “Content Management”, backups e implementação dos requisitos de segurança, privacidade necessários.

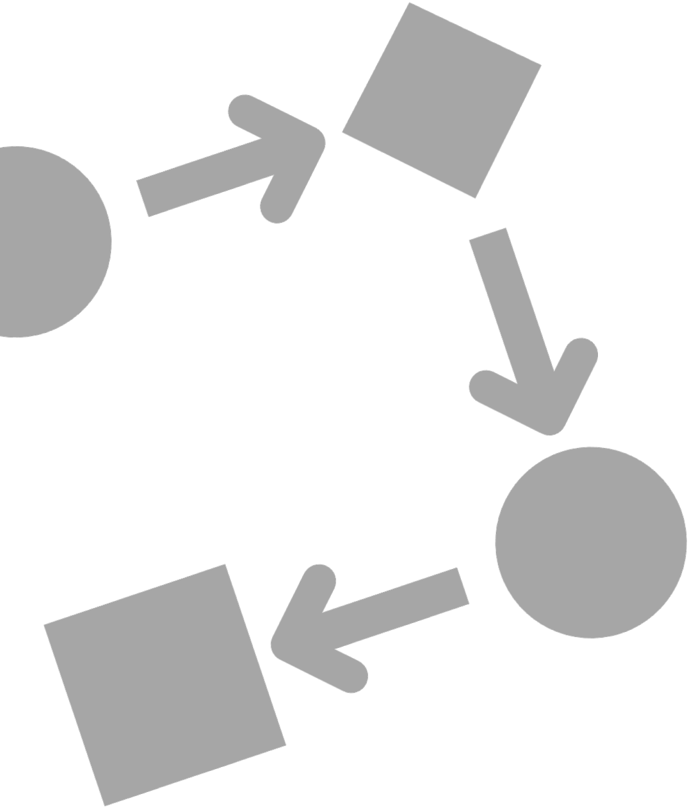
Plano de atuação

- Definição das métricas de inventário e performance tais como número de documentos, classificação, taxonomia, estatísticas transacionais, capacidade, performance, número de requisições e outros indicadores acordados. Caso existente, utilizar referências de SLA já definidos.
- Implementar a estratégia de gestão de conteúdo: definir a estratégia de gestão dos contéudos e documentos tais como ciclo de vida, pesquisa, versionamento, retenção e descarte. Caso aplicável, incluir estratégia para mídias sociais

Resultados

- Inventário de Documentos/Conteúdo e ferramentas
- Documento de Estratégia de Gestão de Conteúdo
- Definição dos processos de governança, definição dos papéis e custodiantes (Data Stewardship)

c) Gestão de Integração & Interoperabilidade



Objetivo

- Visa a melhoria dos processos de gestão de serviços responsáveis pela captura, distribuição, tratamento e consolidação dos dados entre os sistemas de armazenagem, aplicações e áreas internas . Sua missão é prover dados de forma segura, performática e escalável para todos os sistemas e usuários

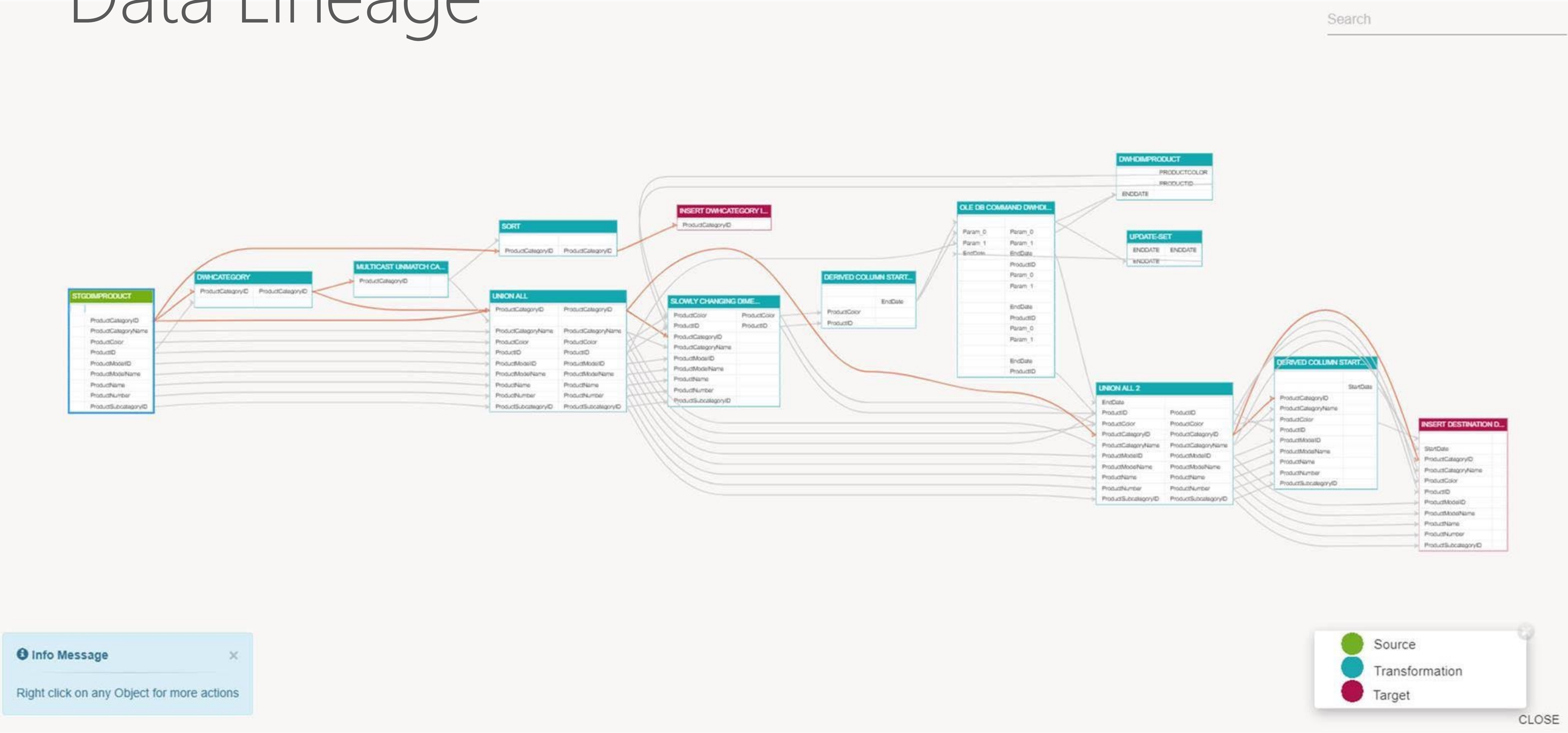
Plano de atuação

- Definição das métricas de inventário e performance tais como número de sistemas, classificação, taxonomia, estatísticas transacionais, capacidade, performance, número de requisições e outros indicadores acordados. Caso existente, utilizar referências de SLA já definidos.
- Arquitetura de integração - definir, a partir da arquitetura de dados e modelagem, a arquitetura de integração dos dados e ciclo de vida
- Linhagem de dados (Data Lineage) - Gerenciar o ciclo de transformação dos dados linhagem de dados. Levantar e documentar como os dados trafegam, são transformados e se interligam nos processos internos

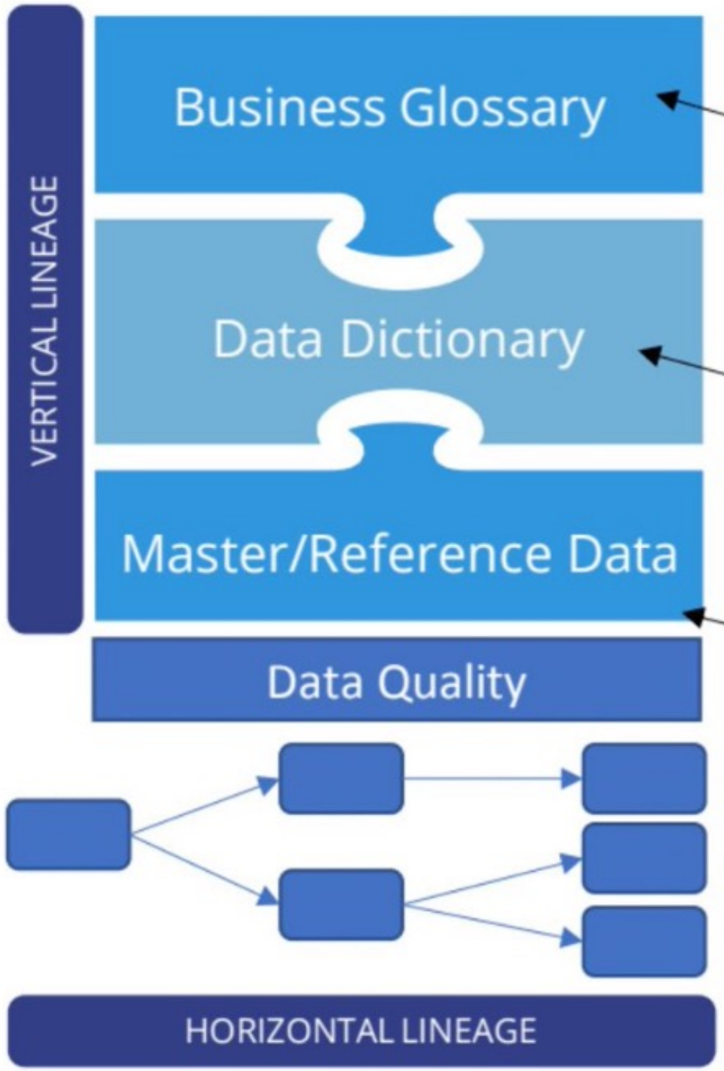
Resultados

- Arquitetura de integração (lógica e técnica)
- Linhagem de Dados (Data Lineage)
- Definição dos processos de governança, definição dos papéis e custodiantes (Data Stewardship)

Data Lineage



Horizontal / Vertical Data Lineage



How legal entities used/transformed in our business processes?

- What is the definition of legal entity?
- How (and where) was legal entity implemented?
- What are our legal entities (golden records)?
- What is our standard reference data?

Nearly everyone is familiar with horizontal lineage, or how data elements are used and evolve, across system-to-system interactions within organizational business process.

For their project, the bank created a new kind of lineage: vertical lineage.

Vertical lineage is how conceptual, business definitions map to system implementations, and data values.

From a governance perspective, vertical lineage answers several key questions.

d) Gestão de Big Data & Analytics

Objetivo

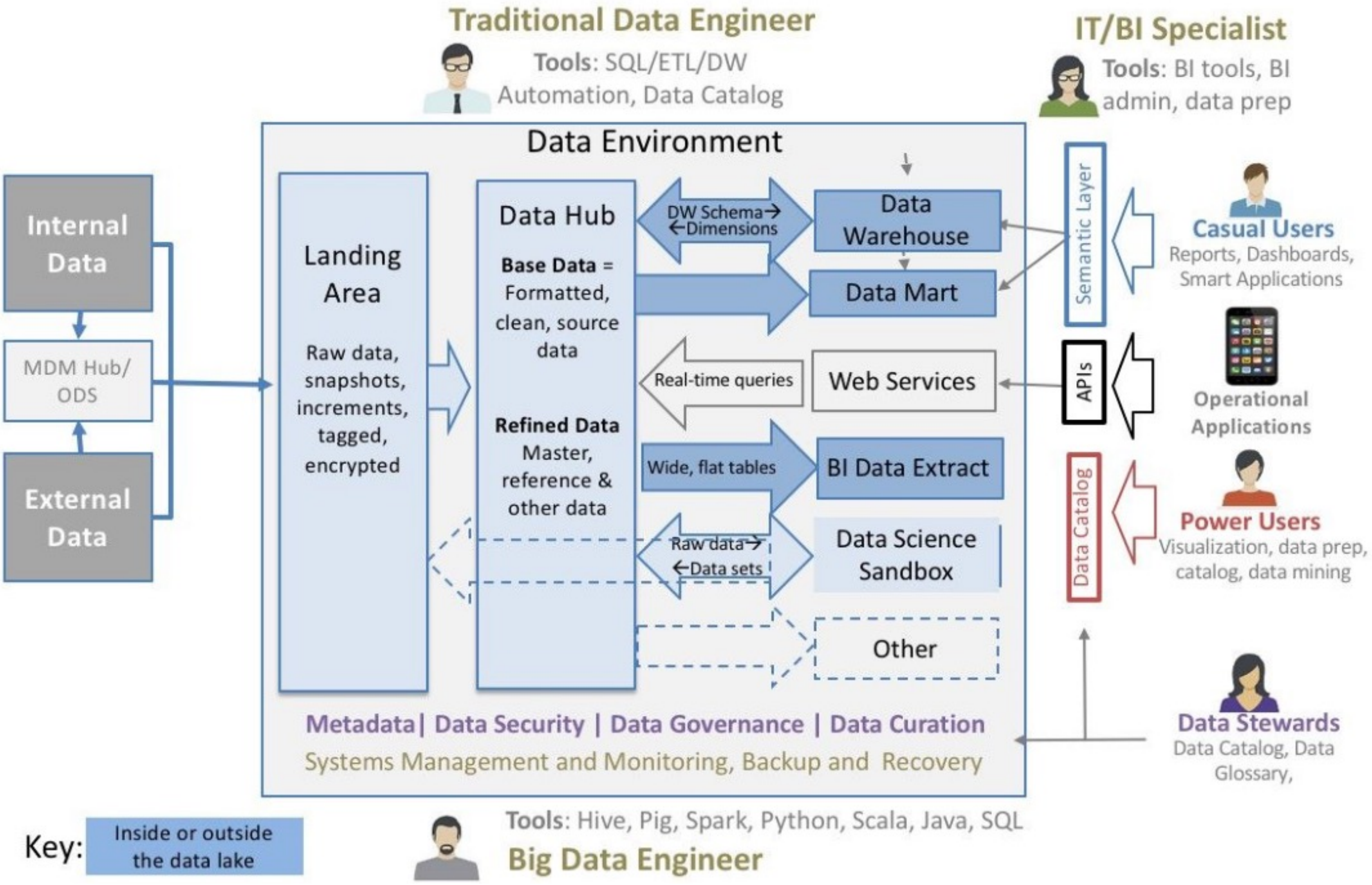
- Visa a melhoria dos processos de gestão dos grandes volumes de dados disponibilizados para consumo interno na forma de datawarehouses e cubos bem como os processos de elaboração, construção e distribuição de painéis e relatórios BI e modelos preditivos (Data Science).

Plano de atuação

- Inventário de Big Data & Analytics - conduzir o inventário dos cubos, painéis, modelos
- Plano Estratégico de Big Data & Analytics - Definir a estratégia de gestão de BI e os processos de melhorias necessários
- Arquitetura de serviços Big Data & Analytics - Definição dos requerimentos e arquitetura de produção e consumo de dados , relatórios BI e modelos

Resultados

- Inventário de Big Data & Analytics
- Plano Estratégico de Big Data & Analytics
- Arquitetura de serviços Big Data & Analytics
- Definição dos processos de governança, definição dos papéis e custodiantes (Data Stewardship)


Traditional Data Engineer
Tools: SQL/ETL/DW Automation, Data Catalog
IT/BI Specialist
Tools: BI tools, BI admin, data prep
Data Environment
Internal Data
MDM Hub/ ODS
External Data
Landing Area
Raw data, snapshots, increments, tagged, encrypted
Data Hub
Base Data = Formatted, clean, source data
Refined Data Master, reference & other data
DW Schema→
←Dimensions
Data Warehouse
Data Mart
Real-time queries
Web Services
Wide, flat tables
BI Data Extract
Raw data→
←Data sets
Data Science Sandbox
Other
Semantic Layer
APIs
Data Catalog
Casual Users
Reports, Dashboards, Smart Applications
Operational Applications
Power Users
Visualization, data prep, catalog, data mining
Data Stewards
Data Catalog, Data Glossary,
Metadata| Data Security | Data Governance | Data Curation
Systems Management and Monitoring, Backup and Recovery
Tools: Hive, Pig, Spark, Python, Scala, Java, SQL
Big Data Engineer
Key: Inside or outside the data lake

DataOps

Objetivo

- Visa a melhoria dos processos de gestão dos grandes volumes de dados disponibilizados para consumo interno na forma de datawarehouses e cubos bem como os processos de elaboração, construção e distribuição de painéis e relatórios BI e modelos preditivos (Data Science).

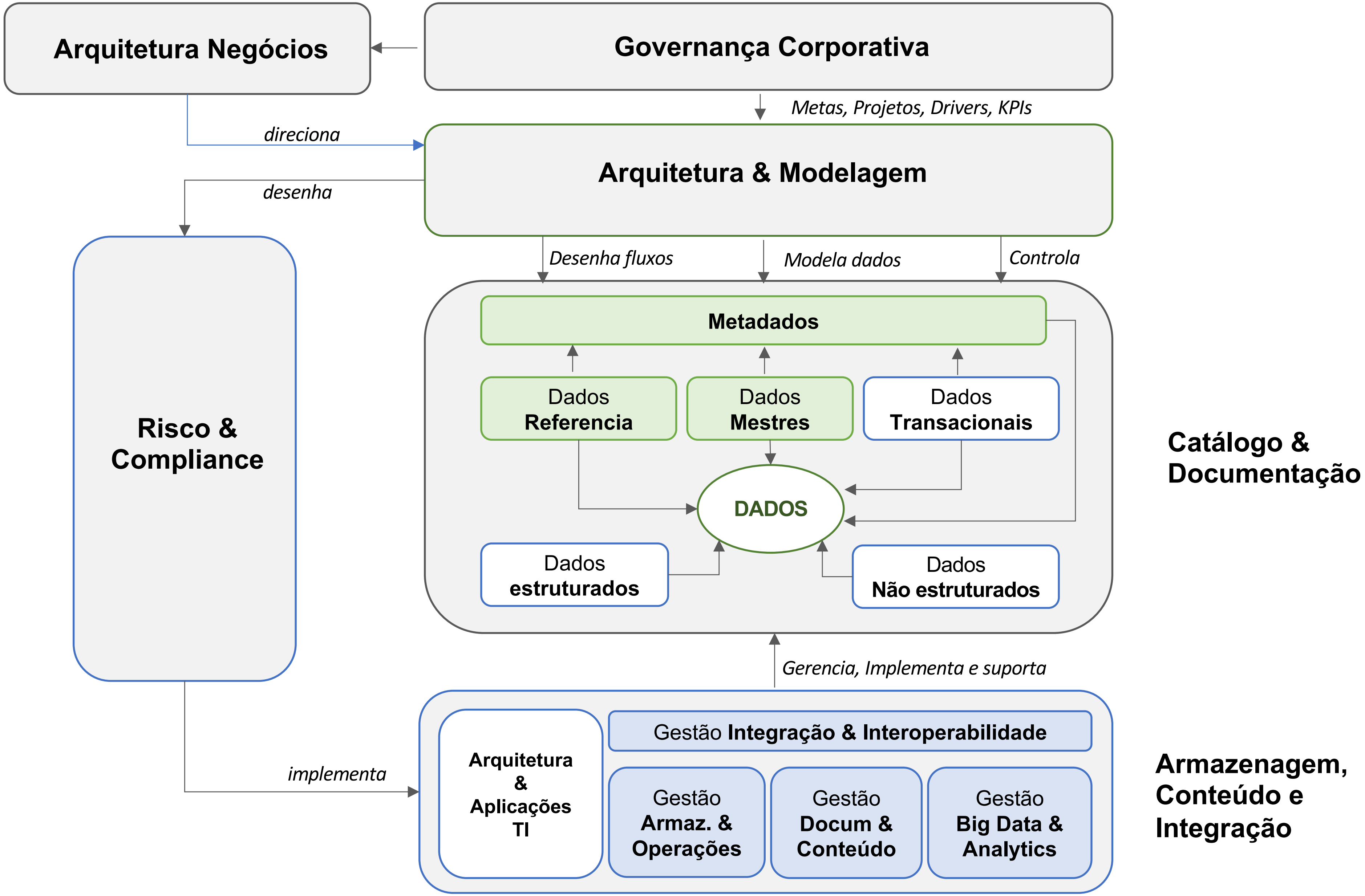
Plano de atuação

- DataOps é uma metodologia automatizada, orientada a processos, usada por equipes analíticas e de dados, para melhorar a qualidade e reduzir o tempo de ciclo da análise de dados.
- Aplica a todo o ciclo de vida dos dados, da preparação dos dados aos relatórios, e reconhece a natureza interconectada da equipe de análise de dados e das operações de tecnologia da informação
- DataOps se inspirou no DevOps, que se concentra na entrega contínua, aproveitando os recursos de TI sob demanda e automatizando o teste e a implantação de análises. Essa fusão do desenvolvimento de software e das operações de TI melhorou a velocidade, a qualidade, a previsibilidade e a escala da engenharia e implantação de software.
- DataOps utiliza o controle estatístico do processo (CEP) para monitorar e controlar o pipeline de análise de dados. Com o SPC instalado, os dados que fluem através de um sistema operacional são constantemente monitorados e verificados como funcionando. Se ocorrer uma anomalia, a equipe de análise de dados poderá ser notificada por meio de um alerta automatizado

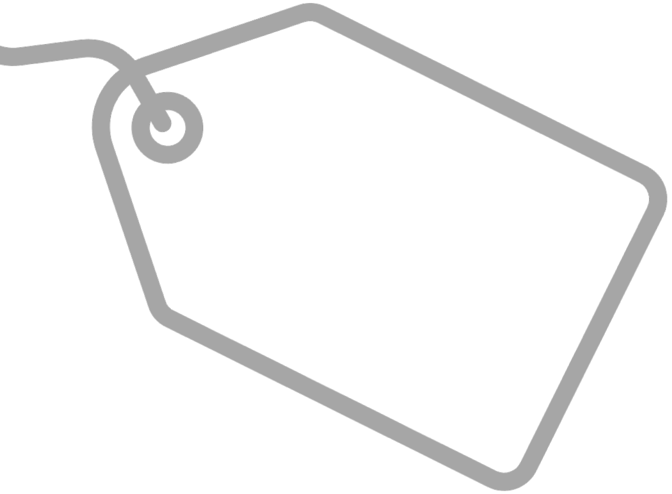
modulo 2

Gestão
Catálogo & Documentação

Catálogo & Documentação



a)Gestão de Dados Mestres, Referência, Metadata



Objetivo

- Visa a melhoria dos processos de gestão de serviços responsáveis pelo inventário, conceituação e ajuste dos dados cadastrais, referência e metadata visando a melhoria e consumo destes dados.
- Como são dados muito utilizados internamente em sistemas e relatórios, possuem grande impacto na performance e qualidade final nos clientes internos e externos

Plano de atuação

- Data Profiling- Definição das categorias de classificação dos dados, eventos e tipos
- Data Models: desenho e elaboração da arquitetura de dados bem como modelos de dados (Data Models) conceituais e lógicos.
- Geração e Gestão dos Catálogos de dados (Data Catalog) e Glossário de dados (Data Glossary)
- Linhagem de dados (Data Lineage) - Gerenciar o ciclo de transformação dos dados linhagem de dados. Levantar e documentar como os dados trafegam, são transformados e se interligam nos processos internos

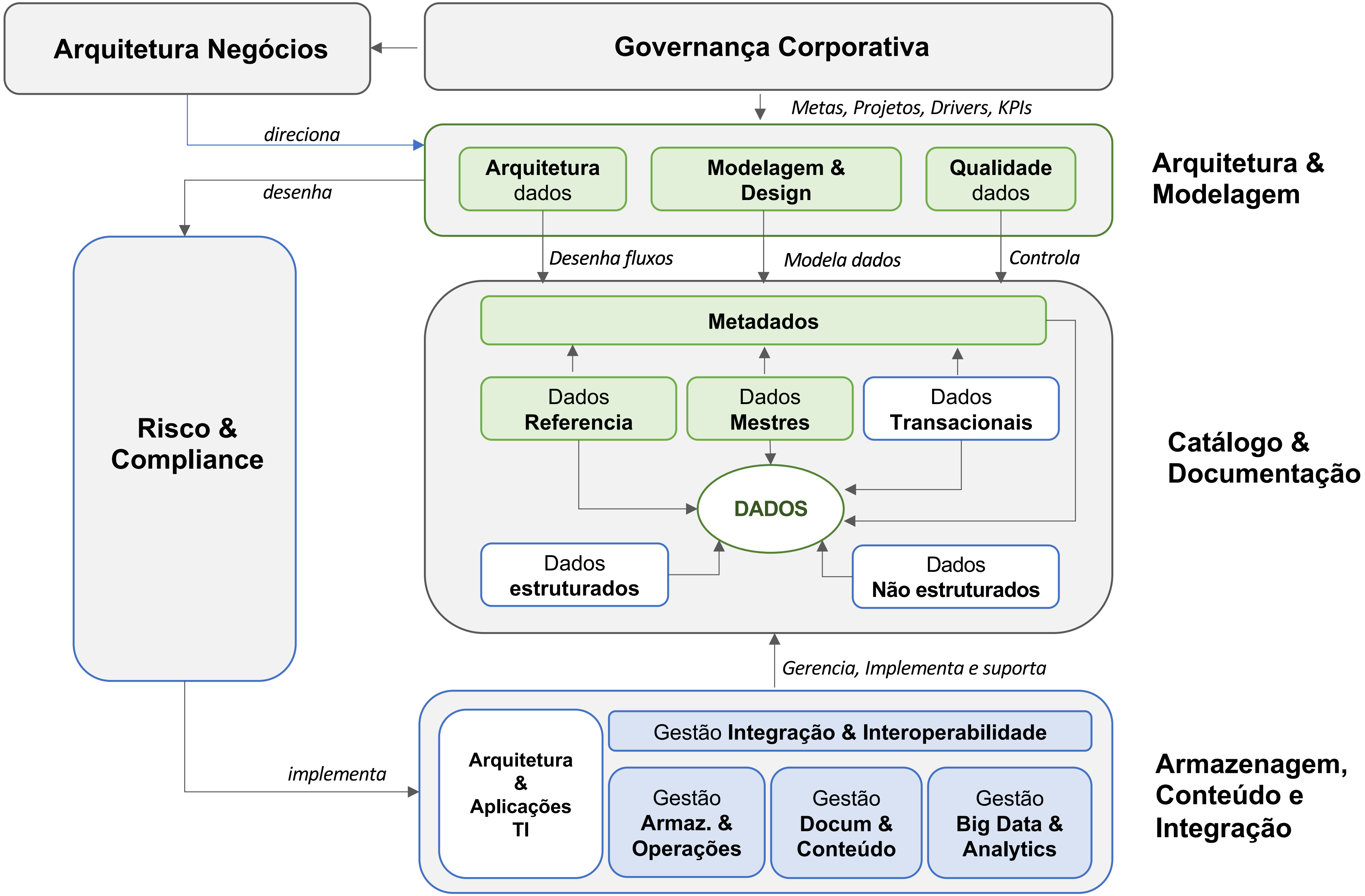
Resultados

- Definição dos processos de governança, definição dos papéis e custodiantes (Data Stewardship)
- Definição dos padrões, metodologias e ferramentas (MDM) de criação e manutenção do catálogo de dados a serem utilizadas no processoLinhagem de Dados (Data Lineage)

modulo 3

Gestão
Arquitetura & Modelagem

Gestão de Arquitetura & Modelagem



a) Gestão de Arquitetura & Modelagem de Dados

Objetivo

- Visa a melhoria dos processos de criação, evolução e gestão do ciclo de vida dos dados e informações estratégicas.
- Como são dados muito utilizados internamente em sistemas e relatórios, possuem grande impacto na performance e qualidade final nos clientes internos e externos

Plano de atuação

- Enterprise Data Model (EDM) – visão lógica dos dados relevantes ao negócio.
- Data Architecture Design- documentação e visualização da visão geral conceitual e funcional do ciclo de vida dos dados prioritários de negócios entre as principais entidades envolvidas (clientes, áreas e etc.).
- Data Plan - Elaborar e implementar o roadmap de iniciativa de melhoria da arquitetura atual (“as-is”) para a arquitetura futura desejada(“to-be”)

Resultados

- Definição dos processos de governança, definição dos papéis e custodiantes (Data Stewardship)
- Definição dos padrões, metodologias e ferramentas de criação, gestão e evolução da arquitetura de Dados, Modelagem e Linhagem de Dados (Data Lineage)

b) Gestão de Qualidade de Dados

Objetivo

- Visa a melhoria dos processos de gestão e manutenção da qualidade e alinhamento dos dados às necessidades de negócio, de acordo com os padrões estipulados entre áreas.
- Como são dados muito utilizados internamente em sistemas e relatórios, possuem grande impacto na performance e qualidade final nos clientes internos e externos

Plano de atuação

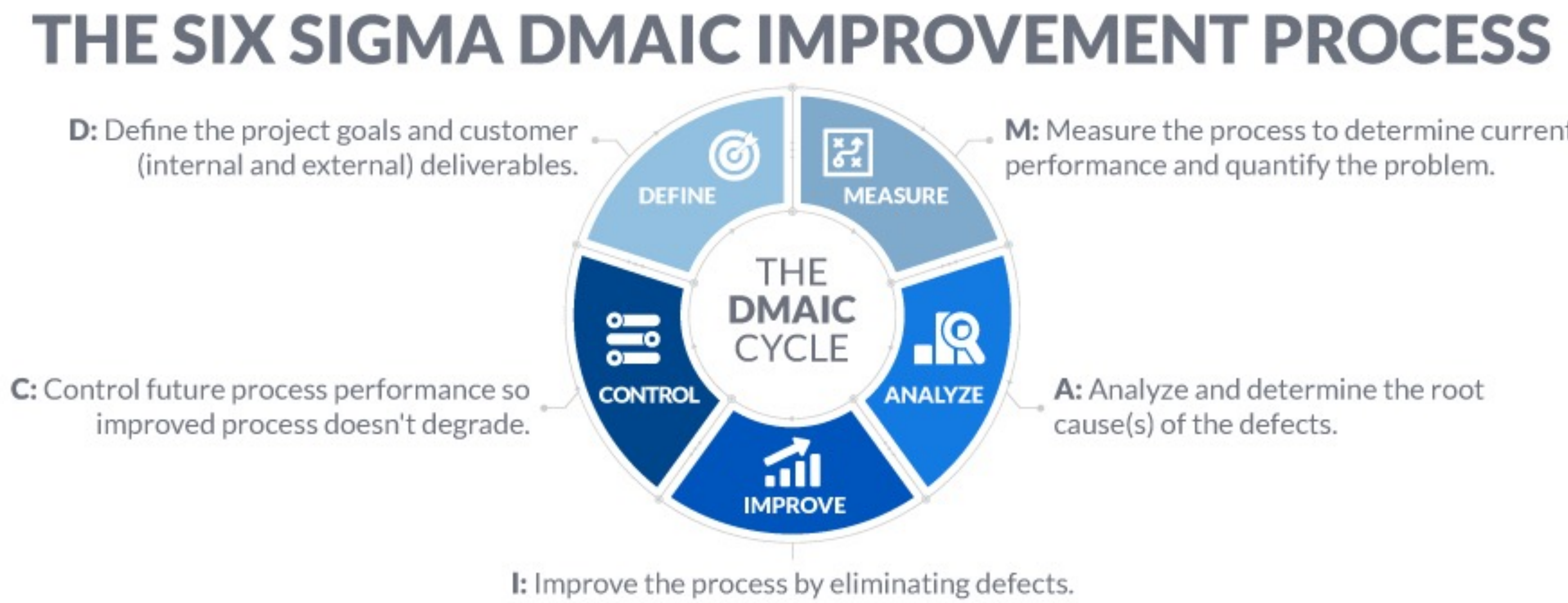
- Data Quality Assessment: efetuar levantamento interno dos pontos críticos de qualidade nas áreas envolvidas e expectativas de padrões esperados.
- Data Quality Plan: Definição das métricas de qualidade, priorização e planejamento das ações necessárias para melhoria da qualidade.
- Implementar iniciativas estruturadas de melhoria **contínua** da qualidade e monitorar resultados

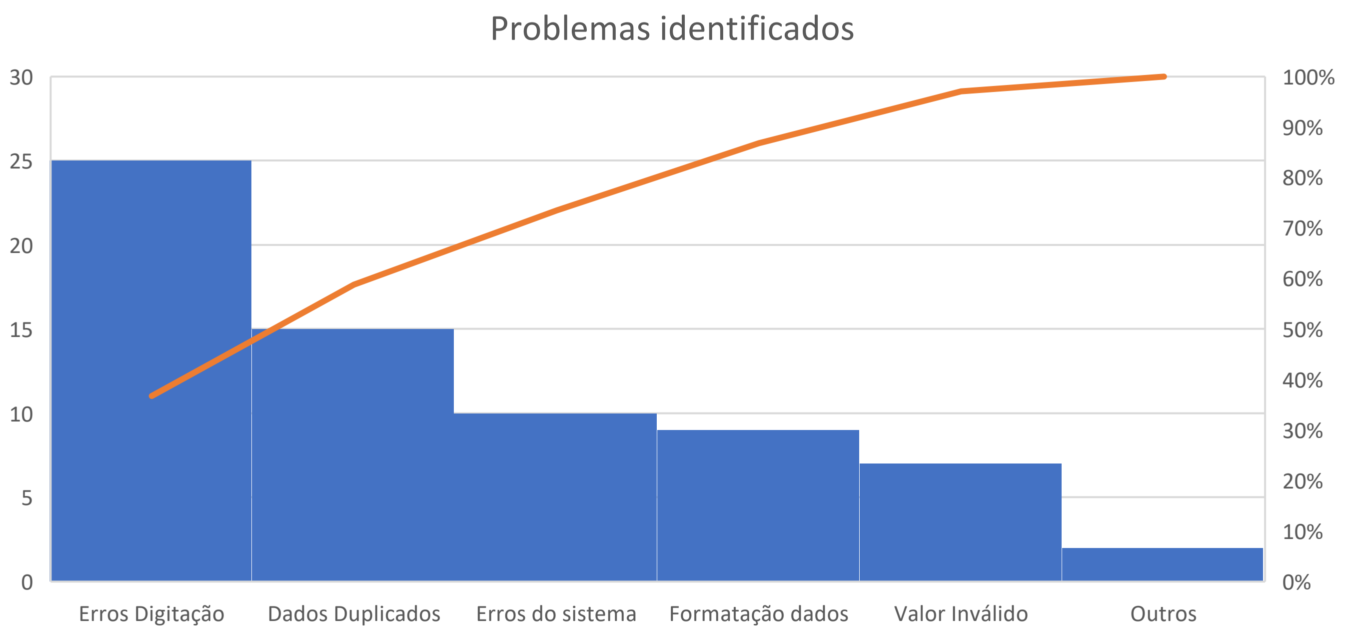
Resultados

- Definição dos processos de governança, definição dos papéis e custodiantes (Data Stewardship)
- Definição dos padrões, metodologias e ferramentas de criação, gestão e evolução da gestão da qualidade de dados

Data quality attributes

Attribute	What it means	Example of good practice	Example of bad practice	Metrics
Consistency	No matter where you look in the database, you won't find any contradictions in your data.	Your payment system shows that Jane Brown has made 5 purchases this month, and CRM system contains the same information.	Your payment system shows that Jane Brown has made 5 purchases this month, while CRM system shows she has made only 4.	The number of inconsistencies.
Accuracy	The information your data contains corresponds to reality.	Your customer's name is Jane Brown. And this is exactly how it's reflected in your CRM.	In your CRM, the customer's name is spelled Jane Brawn, though her actual name is Jane Brown.	The ratio of data to errors.
Completeness	All available elements of the data have found their way to the database.	You know that Jane Brown is born on 11/04/1975.	You have no idea how old Jane Brown is, as the date of birth cell is empty.	The number of missing values.
Auditability	Data is accessible and it's possible to trace introduced changes.	You can track down the changes made in Jane's data record. For example, on 12/5/2018, her phone number was changed.	It's impossible to trace down the changes in Jane's record.	% of cells where the metadata about introduced changes is not accessible.
Orderliness	The data entered has the required format and structure.	The entry for December 11, 2018 is in the format 12/11/2018.	The entry for December 11, 2018 is in the format 12/11/18, 12/11/2018 and even 11/12/18 (in your European stores).	The ratio of data of inappropriate format.
Uniqueness	A data record with specific details appears only once in the database.	You have only one record for Jane Brown, born on 11/04/1975, who lives in Seattle.	You have multiple duplicate records for Jane Brown.	The number of duplicates revealed.
Timeliness	Data represents reality within a reasonable period of time or in accordance with corporate standards.	On 02/15/2018, the customer informed you that her name is misspelled in the emails you send her. The customer's name was corrected the next day.	On 02/15/2018, the customer informed you that her name is misspelled in the emails you send her. Her name was corrected only in a month.	Number of records with delayed changes.





a) Medir e Identificar causas raiz

- "5 Why", "Causa & Efeito", etc

b) Priorizar causa raiz

- Pareto

c) Atacar o problema

- Implantar melhoria

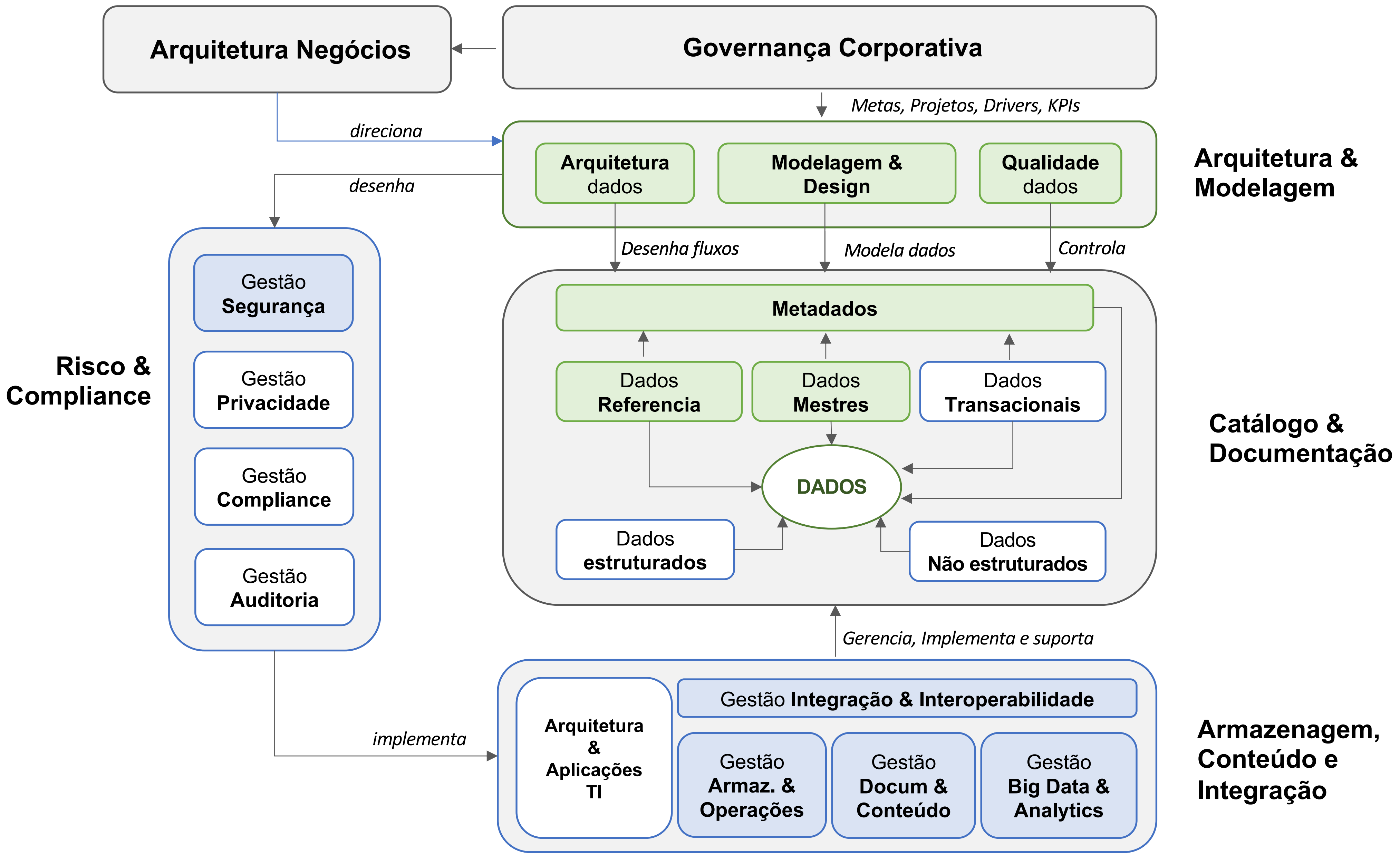
d) Medir o resultado

e) Atacar próxima causa...

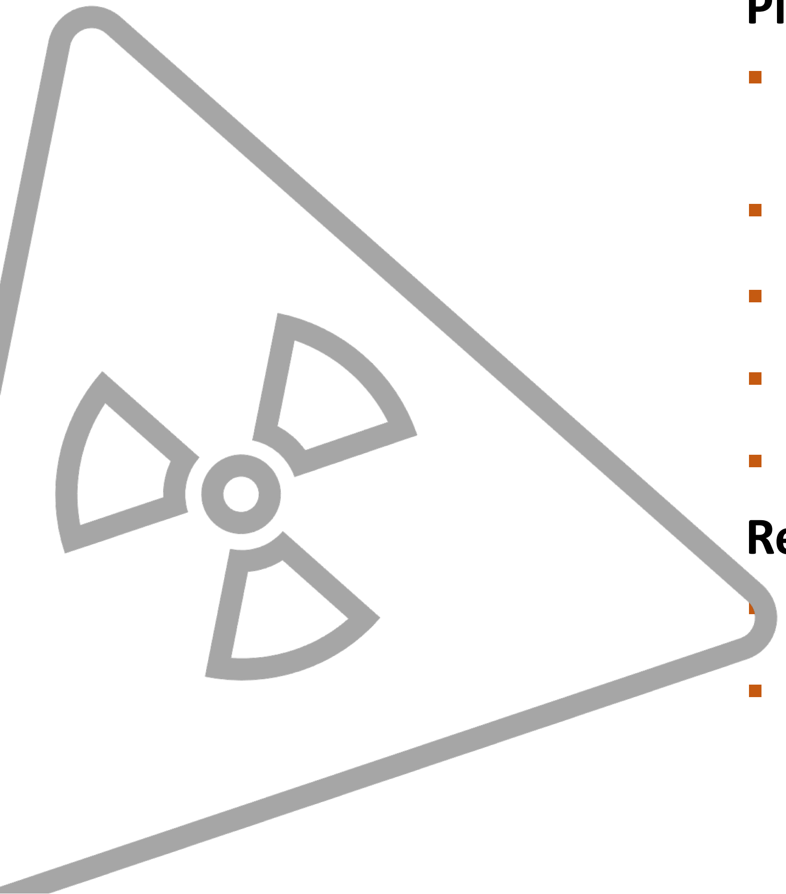
modulo 4

Gestão
Risco & Compliance

Gestão de Risco & Compliance



a) Gestão de Risco & Compliance



Objetivo

- Visa a melhoria dos processos dos processos de gestão de Risco dos dados corporativos.
- Esta gestão inclui as etapas de planejamento, desenvolvimento e execução de políticas e regulamentos definidos na Arquitetura de Dados (Data Architecture) definida nos módulos anteriores.

Plano de atuação

- Data Risk Requirements - definir, a partir da arquitetura de dados e modelagem e regulamentações, os requerimentos de gestão de risco. Caso aplicável adotar a regulamentação de segurança indicada.
- Data Risk Plan – plano de mapeamento, melhoria e monitoramento do risco
- Data Security Plan – plano de mapeamento, melhoria e monitoramento da segurança
- Data Privacy Plan – plano de mapeamento, melhoria e monitoramento de privacidade
- Data Audit Plan – plano de auditoria e performance

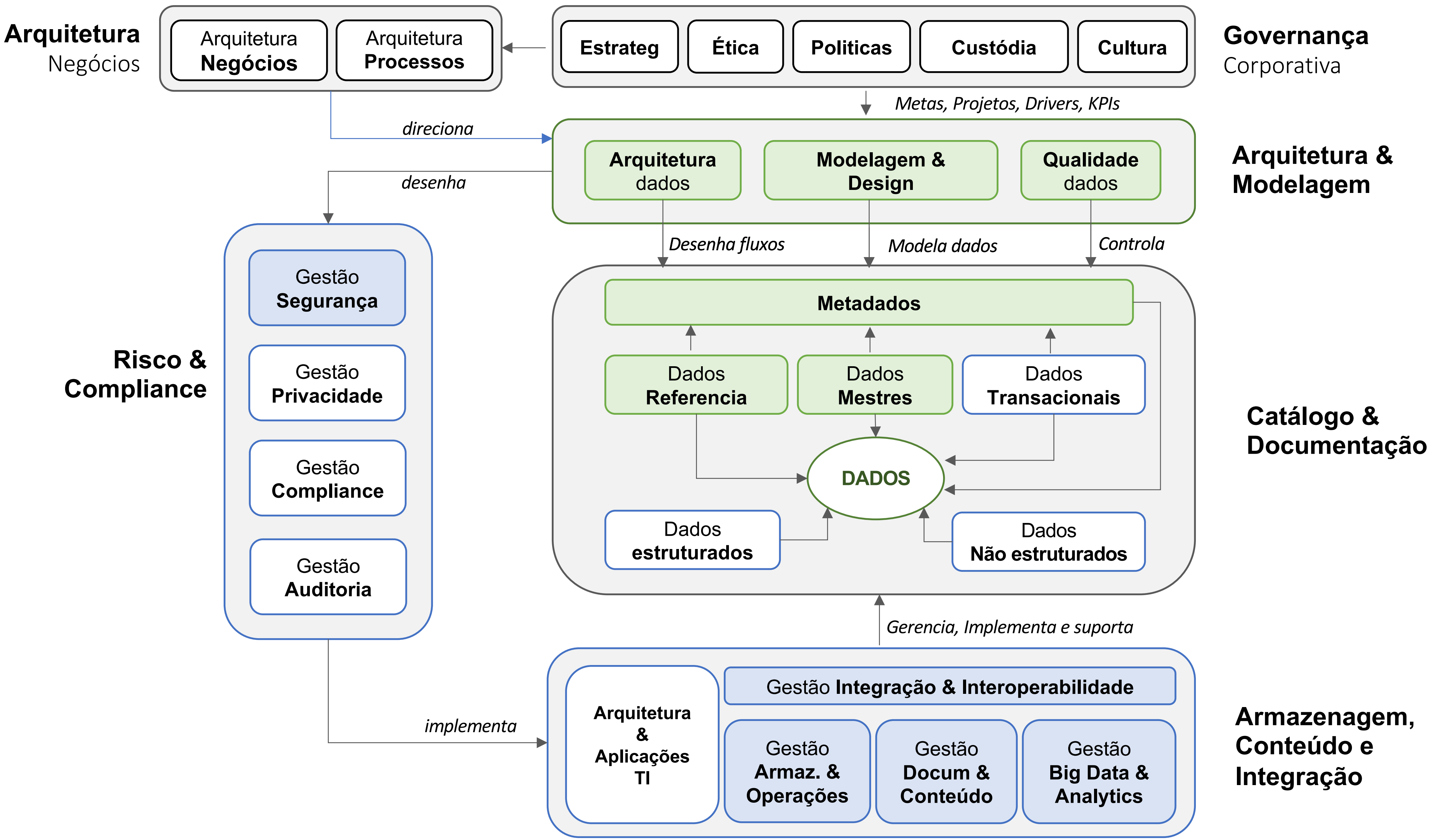
Resultados

- Definição dos processos de governança, definição dos papéis e custodiantes (Data Stewardship)
- Definição dos padrões, metodologias e ferramentas de criação, gestão do risco & compliance

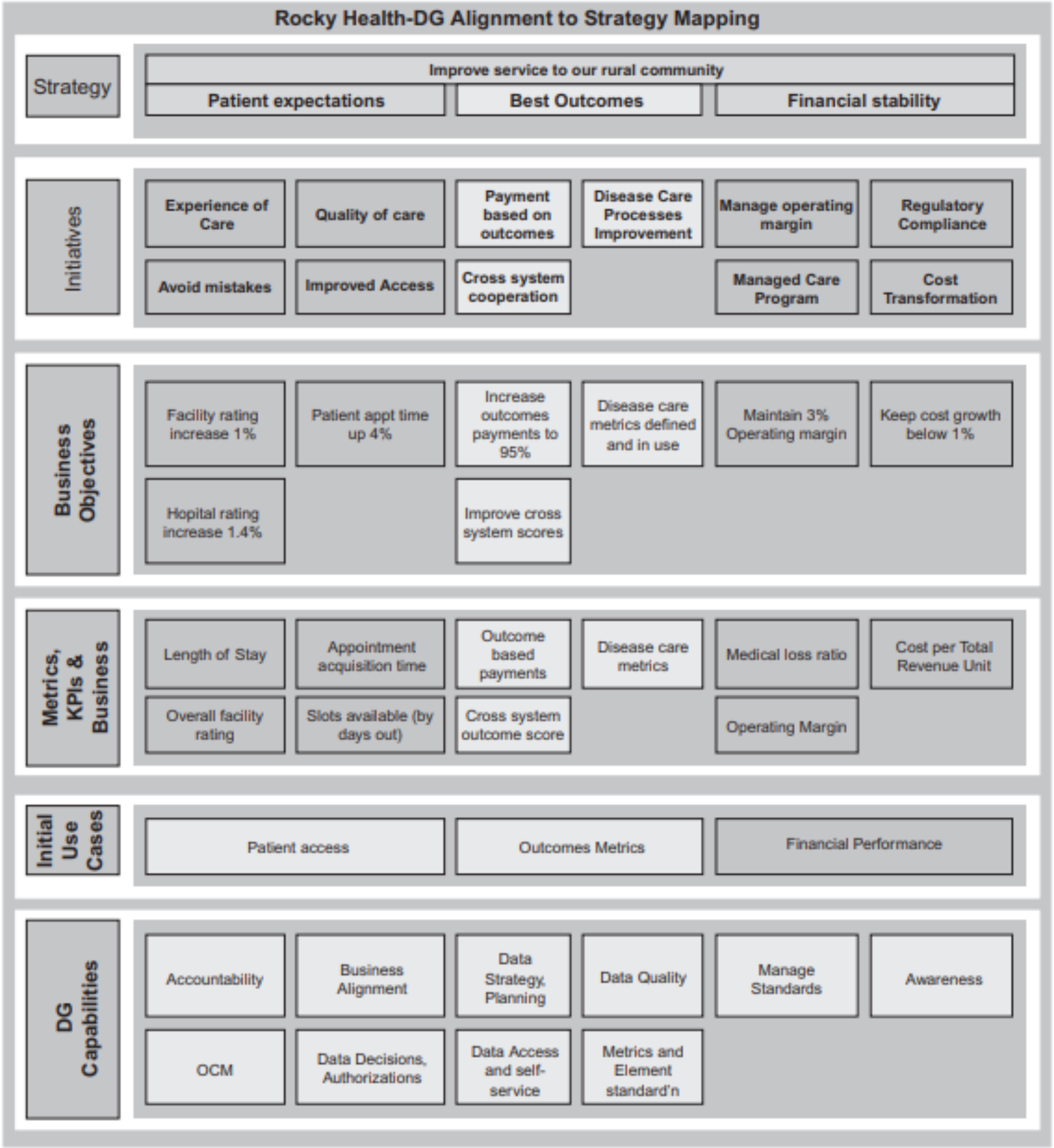
modulo 5

Gestão
governança & arquitetura de negócios

Governança & arquitetura de negócios



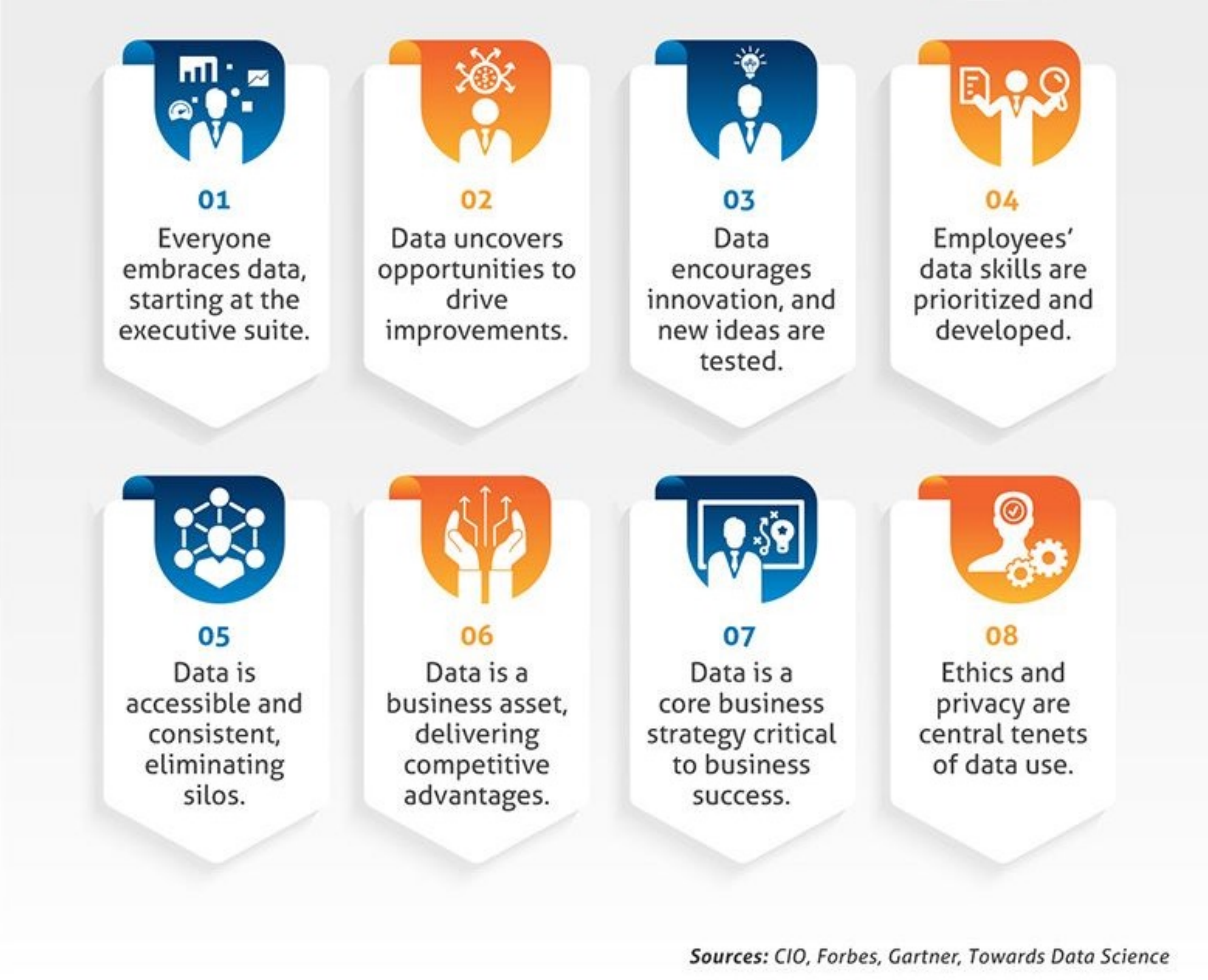
Data Strategy



Data Strategy

Driver	Goal	Documented Objectives	Initiatives	Business Data Mgmt Needs	Possible DG Capabilities
Growth in operating margins	Increase Revenue	Increase nonenergy product revenue 15%	Increase sales of existing products and services		
			Introduce programs to encourage efficient consumption of electricity		
	Reduce Costs	Reduce tool and material redundancy 25%	**Minimizing the tools & equipment needed to operate our business**	**Tool and equipment data management**	**Data lineage, data quality**
		Improve power output 5% in existing plants	Minimize production costs through improved plant capacity		
			Improve "capital efficiency" (getting the most out of every capital dollar spent)		
	Maximize return on assets (plant, people, processes)	Improve power output 5% in existing plants	Improve availability of information to facilitate efficient operations		
			Increase load factor on system		
			Effectively evaluate business opportunities in approval process		
		Reduce new transformer installation cycles by one week	**Attain more efficiency in asset set-up and management**	**Tool and equipment data management**	**Data standards**
	Position for Growth	Improve engineering project results-80% on time and budget	Plan, acquire, and position new and existing assets and resources (distribution, transmission, coal, wind, distributed generation, etc.)		
Effective Regulatory Position	Improved image with regulators	Reduce customer complaints 25%	Monitor customer privacy		
			Provide strong analytical and fact-based, well-documented positions		
			Establish strong support of and representation in community		
			Meet reliability targets		
			Meet compliance standards (e.g., reliability standards, call center responsiveness mandates, environmental mandates)		
			Analyze and communicate action steps necessary to minimize reported customer complaints		
Customer Satisfaction	Increase Value to Customer	Reduce new connection time to one week	**Improve responsiveness (cycle time, on time, information, etc.)**	**New service appointments, scheduling and asset availability data**	**Item, Inventory accuracy, data quality**
		Increase customer satisfaction on new service to 90%	**Deliver service to expectation levels**		
			Increase system reliability		
			Increase options to customers		
			Limit price increases		
			Increase value of offerings (maximize service received for each dollar spent by customer)		
			Improve Ability to Anticipate/React to System Swings		
			Ensure customer data is used appropriately and as specified		
			Reduce Interface Cycle Times		
Operational Excellence	Improve process efficiency (cost, cycle time) and effectiveness	Reduce cycle times and costs	Identify and define (map) processes to be able to effectively execute business strategies		
			Define process metrics with appropriate goals/targets and control limits		
			Provide tools/technology to effectively enable our processes		
Risk Management	Improve regulatory risk management skills	Reduce risk exposures by 8%	Ensure customer data privacy compliance		
			Meet compliance standards (e.g., SOX, Environmental, Employment, etc.)		

Data-driven culture



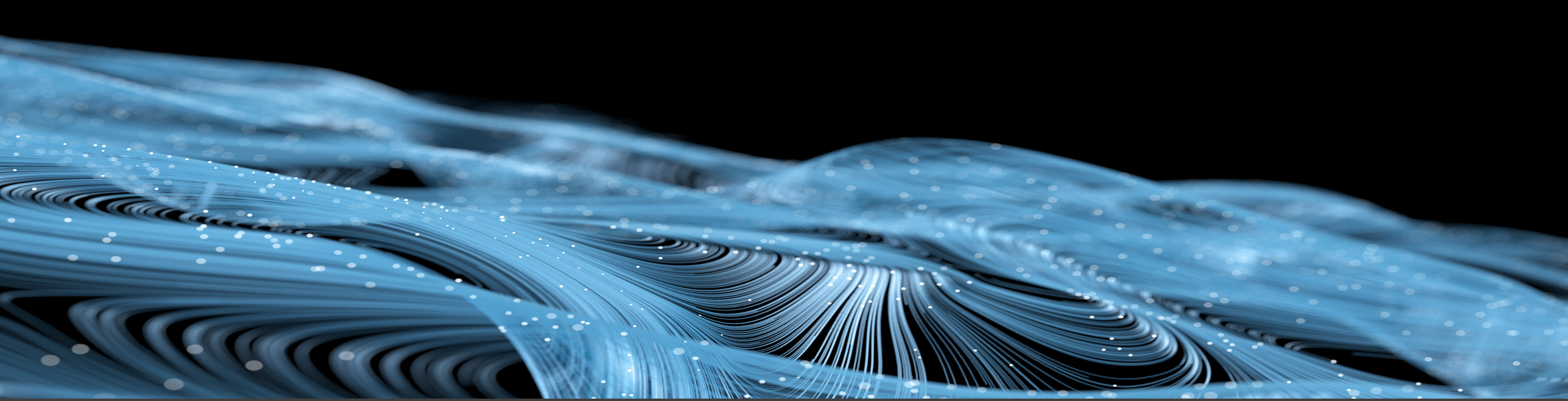
Data Literacy

Objetivo

- Evangelizar sobre o valor, relevância e importância estratégica dos dados
- Capacitar habilidades de leitura, analise e comunicação com dados
- Habilidades para gerar valor através dos dados e informações
- Entendimento da relação entre dados, transformação digital e inovação

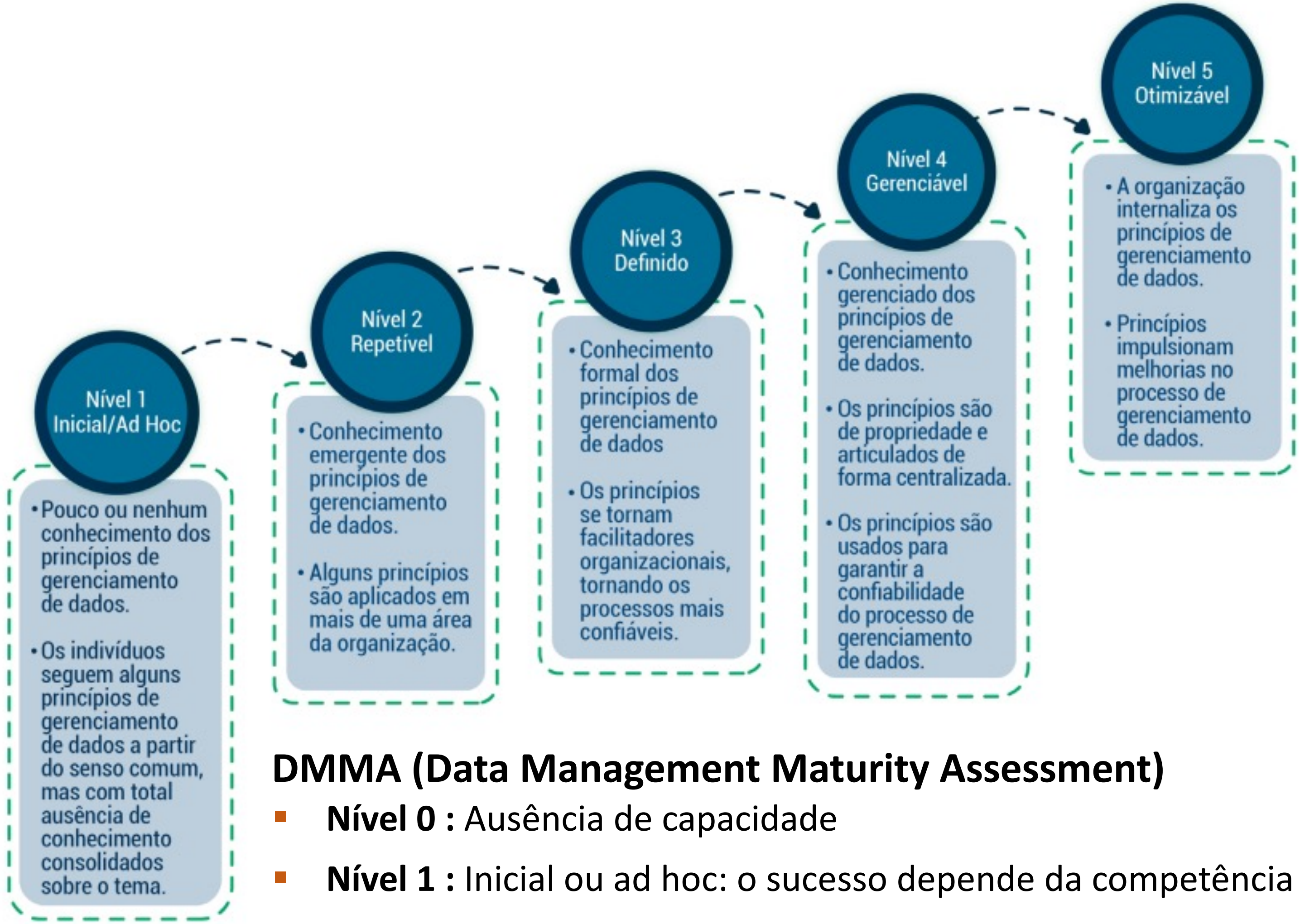
Plano de atuação

- Habilidades de processos de negócios para entender e planejar a criação de dados confiáveis.
- Habilidades de design para planejar sistemas nos quais os dados serão armazenados ou usados.
- Habilidades altamente técnicas para administrar hardware e criar software onde os dados são mantidos.
- Habilidades de análise de dados para entender questões e problemas descobertos nos dados.
- Habilidades analíticas para interpretar dados e aplicá-los a novos problemas.
- Habilidades linguísticas para trazer consenso às definições e modelos para que as pessoas possam entender dados.
- Pensamento estratégico para visualizar oportunidades de uso de dados para atender usuários/consumidores e atingir metas.



maturidade

2) Estágios



Objetivos

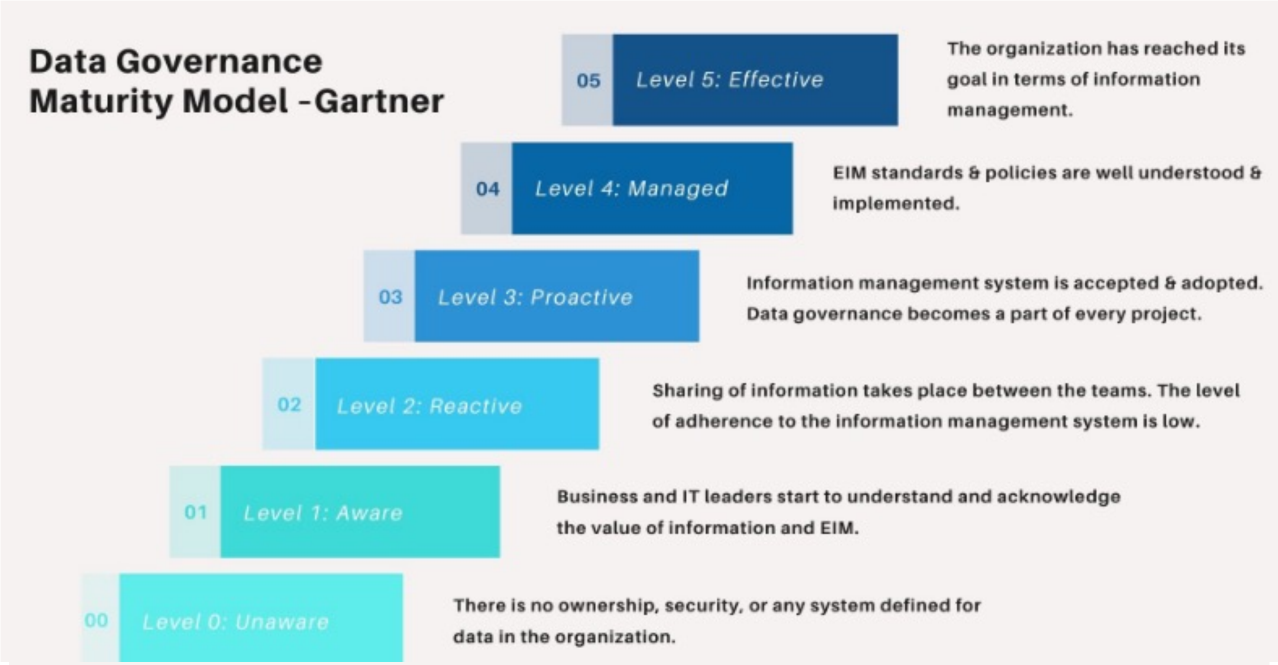
- Ajuda na compreensão de como é atualmente a sua gestão interna dos dados
- O modelo define um caminho para evoluir de acordo com as necessidades e estratégias
- Ajuda a medir a melhoria e comparar-se ao mercado (benchmarks) , concorrentes e parceiros

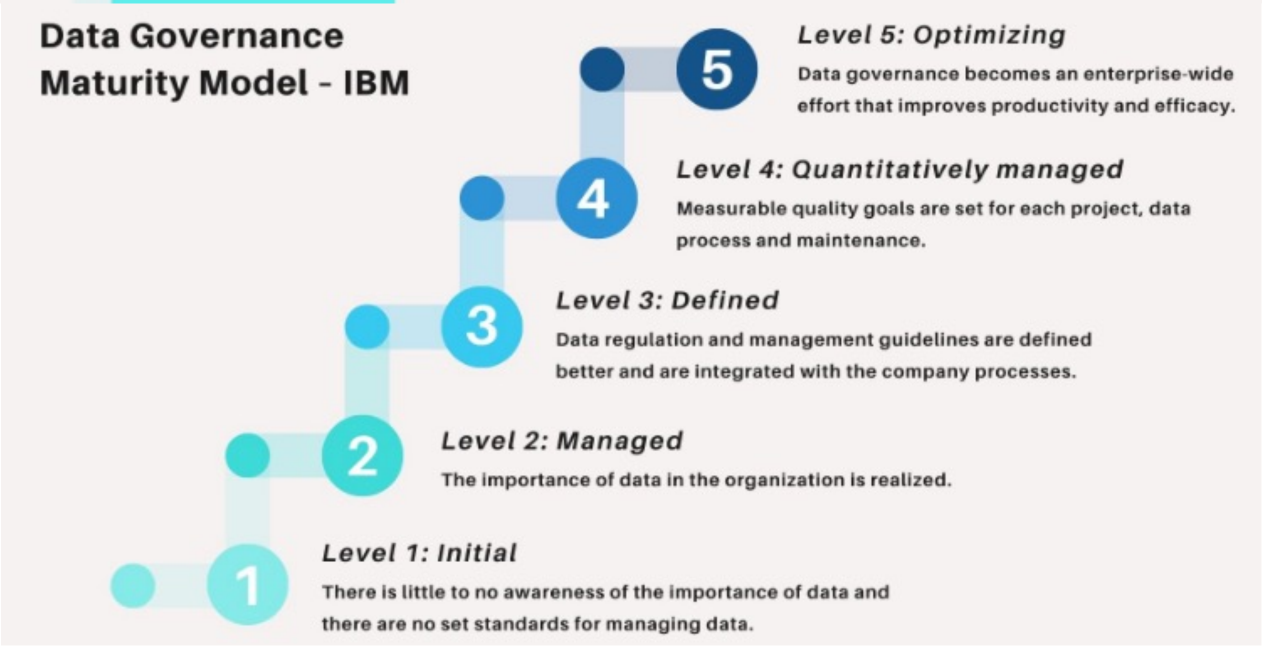
DMMA (Data Management Maturity Assessment)

- **Nível 0 :** Ausência de capacidade
- **Nível 1 :** Inicial ou ad hoc: o sucesso depende da competência dos indivíduos
- **Nível 2 :** Repetível: a disciplina mínima do processo está em vigor
- **Nível 3 :** Definido: os padrões são definidos e usados
- **Nível 4 :** Gerenciável: os processos são quantificados e controlados
- **Nível 5 :** Otimizável: as metas de melhoria de processo são quantificadas.

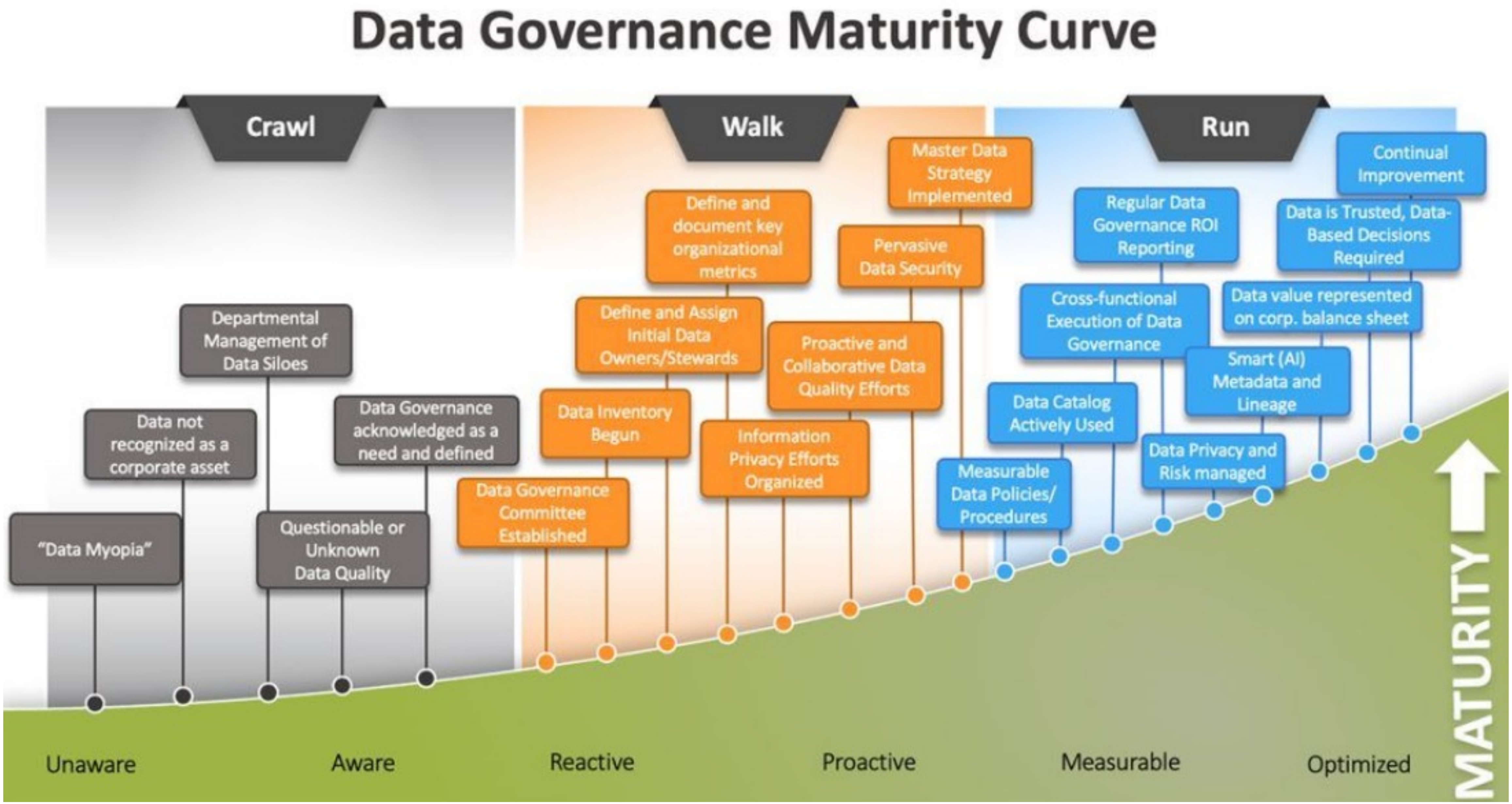
Outros Modelos

- Data Governance Maturity Model – Gartner
  - Level 0: Unaware
  - Level 1: Aware
  - Level 2: Reactive
  - Level 3: Proactive
  - Level 4: Managed
  - Level 5: Effective
- Data Governance Maturity Model – IBM
  - Level 1: Initial
  - Level 2: Managed
  - Level 3: Defined
  - Level 4: Quantitatively Managed
  - Level 5: Optimizing



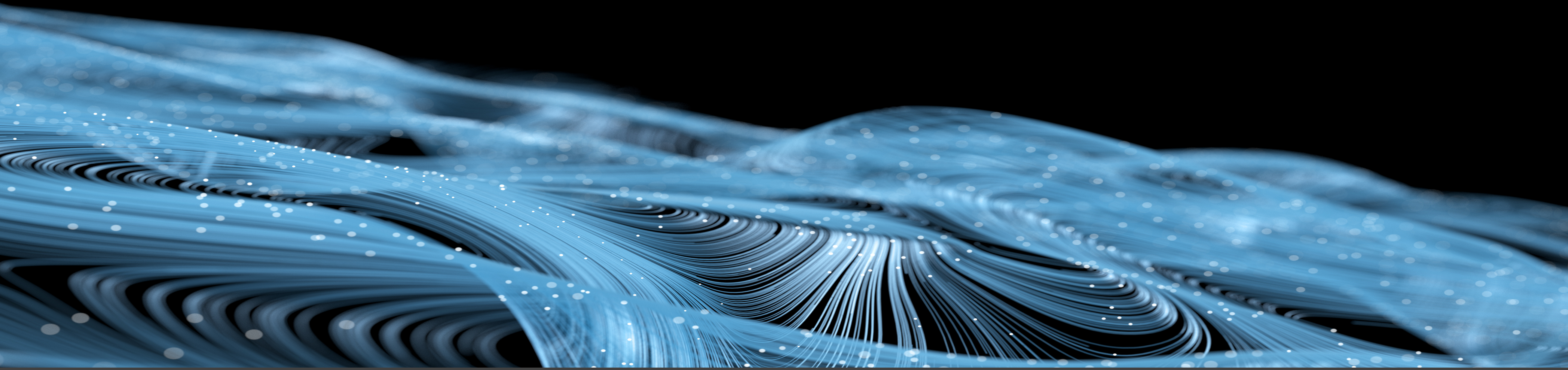


Visão por competências



Visão por processos

	Explorer	User	Leader	Innovator
Business Strategy	Data is used solely for reporting purposes	Data insights are used to inform business decisions	Competitive business strategy is built off of data	Data informs a continuous evolution of business strategy
Data	The organization solely uses its own internal data	The organization uses data providers to enrich & supplement its own data	Third party data is used as a differentiator	The organization is constantly looking to leverage new datasets from non-obvious sources
Culture	The use of data and analysis is left up to the individual	Data is a part of measuring results but not planning	Decision makers are enabled with the results of data analysis to maximize business outcomes	The organization has built true AI/ML algorithms that adapt and improve business objectives
Architecture	The business lacks a cohesive and coherent data architecture	Some architecture exists to automate and analyze data flow	The architecture enables all members of the organization to be data driven	The architecture is built for speed, market distribution and large volumes of data
Data Governance	Governance is largely manual and lacks consistency	Process is in place to protect data quality across the organization	There's universal confidence in the data and resulting insights	Data governance is integrated into all business processes
Procurement & Onboarding	No datasets have been onboarded to the organization	Individual teams are responsible for procuring and onboarding their own data	There's a streamlined process for data procurement, but onboarding isn't universal	The organization has a data procurement team that sources and onboards new data



Projeto tipico

1) Objetivos

Driver	Goal	Documented Objectives	Initiatives	Business Data Mgmt Needs	Possible DG Capabilities
Growth in operating margins	Increase Revenue	Increase nonenergy product revenue 15%	Increase sales of existing products and services		
			Introduce programs to encourage efficient consumption of electricity		
	Reduce Costs	Reduce tool and material redundancy 25%	**Minimizing the tools & equipment needed to operate our business**	**Tool and equipment data management**	**Data lineage, data quality**
		Improve power output 5% in existing plants	Minimize production costs through improved plant capacity		
			Improve "capital efficiency" (getting the most out of every capital dollar spent)		
	Maximize return on assets (plant, people, processes)	Improve power output 5% in existing plants	Improve availability of information to facilitate efficient operations		
			Increase load factor on system		
			Effectively evaluate business opportunities in approval process		
		Reduce new transformer installation cycles by one week	**Attain more efficiency in asset set-up and management**	**Tool and equipment data management**	**Data standards**
	Position for Growth	Improve engineering project results-80% on time and budget	Plan, acquire, and position new and existing assets and resources (distribution, transmission, coal, wind, distributed generation, etc.)		
Effective Regulatory Position	Improved image with regulators	Reduce customer complaints 25%	Monitor customer privacy		
			Provide strong analytical and fact-based, well-documented positions		
			Establish strong support of and representation in community		
			Meet reliability targets		
			Meet compliance standards (e.g., reliability standards, call center responsiveness mandates, environmental mandates)		
			Analyze and communicate action steps necessary to minimize reported customer complaints		
Customer Satisfaction	Increase Value to Customer	Reduce new connection time to one week	**Improve responsiveness (cycle time, on time, information, etc.)**	**New service appointments, scheduling and asset availability data**	**Item, Inventory accuracy, data quality**
		Increase customer satisfaction on new service to 90%	**Deliver service to expectation levels**		
			Increase system reliability		
			Increase options to customers		
			Limit price increases		
			Increase value of offerings (maximize service received for each dollar spent by customer)		
			Improve Ability to Anticipate/React to System Swings		
			Ensure customer data is used appropriately and as specified		
			Reduce Interface Cycle Times		
Operational Excellence	Improve process efficiency (cost, cycle time) and effectiveness	Reduce cycle times and costs	Identify and define (map) processes to be able to effectively execute business strategies		
			Define process metrics with appropriate goals/targets and control limits		
			Provide tools/technology to effectively enable our processes		
Risk Management	Improve regulatory risk management skills	Reduce risk exposures by 8%	Ensure customer data privacy compliance		
			Meet compliance standards (e.g., SOX, Environmental, Employment, etc.)		

Maturity Assessment

DATA GOVERNANCE MATURITY MODEL QUESTIONNAIRE

Data Access Management Questionnaire	Answer
Is a process in place for people to request data?	No
Are data access requests completed in a timely manner?	No
Is the data access request process efficient?	No
Are there designated data owners in your organization?	No
Are there designated data stewards in your organization?	No
Is data discoverable and searchable?	No
Is data universally accessable, not siloed, and easy for other teams to request access?	No
Is risk management and regulatory compliance important to every user in your organization?	No
Are there policies in place about who can access what data?	No
Is compliance and risk management an inegral part of your organization's data management?	No
Does your organization identify and communicate to the users of your data assets?	No
Do you have policies in place that address sharing PII outside your organization?	No
Is data encrypted at rest?	No
Does your organization have the ability to identify sensitive data in your ecosystem?	No
Does access to data change when a person is moved to another position in the company?	No
Is access regularly reviewed and adjusted for appropriateness?	No

ANSWER LEVELS

Level 1: No
Issues are dealt with as they appear.

Level 2: Beginning
The importance is becoming apparent and commonly accepted. Efforts are beginning as the organization determines what is needed.

Level 3: In Progress
Policies and documentation are being created to implement a solution. Actions are being taken such as appointing people to positions or installing a tool.

Level 4: Yes
The organization is now enforcing policies and procedures for data governance. There is an implemented solution being monitored for success.

Level 5: Absolutely
The implemented solution is working and only

Maturity Assessment

DATA GOVERNANCE MATURITY MODEL QUESTIONNAIRE

Data Quality Program Questionnaire	Answer
Are data quality standards defined across your organization?	No
Can users easily report a data quality issue in your organization?	No
Does data undergo the data quality improvement lifecycle process (define, collect, prioritize, analyze, improve, control)?	No
Is a prevention system in place for future data quality issues?	No
Are processes in place for performing root-cause-analysis to discover where data quality issues are occurring?	No
Are data quality rules made to fix previously found problems?	No
Do you profile and review data quality when creating Data Assets as part of the delivery process?	No
Is data classified and tagged for easy searchability?	No
Is data lineage tracked as data is moved and transformed?	No
Do you review data quality issues monthly to adress trends and global process improvements?	No
Does your organization proactively communicate to impacted users when quality issues arise?	No
Are ETL and Data Transformation errors logged as data quality issues?	No
Can you report all data quality issues as they apply to a specific asset in your organization?	No
Do you identify and track root cause, remediation, and long term solutions for your data quality issues?	No
Are there master data and reference master data policies in place so no duplicate data is made?	No
Overall Data Quality Program Level	

ANSWER LEVELS

Level 1: No
Issues are dealt with as they appear.

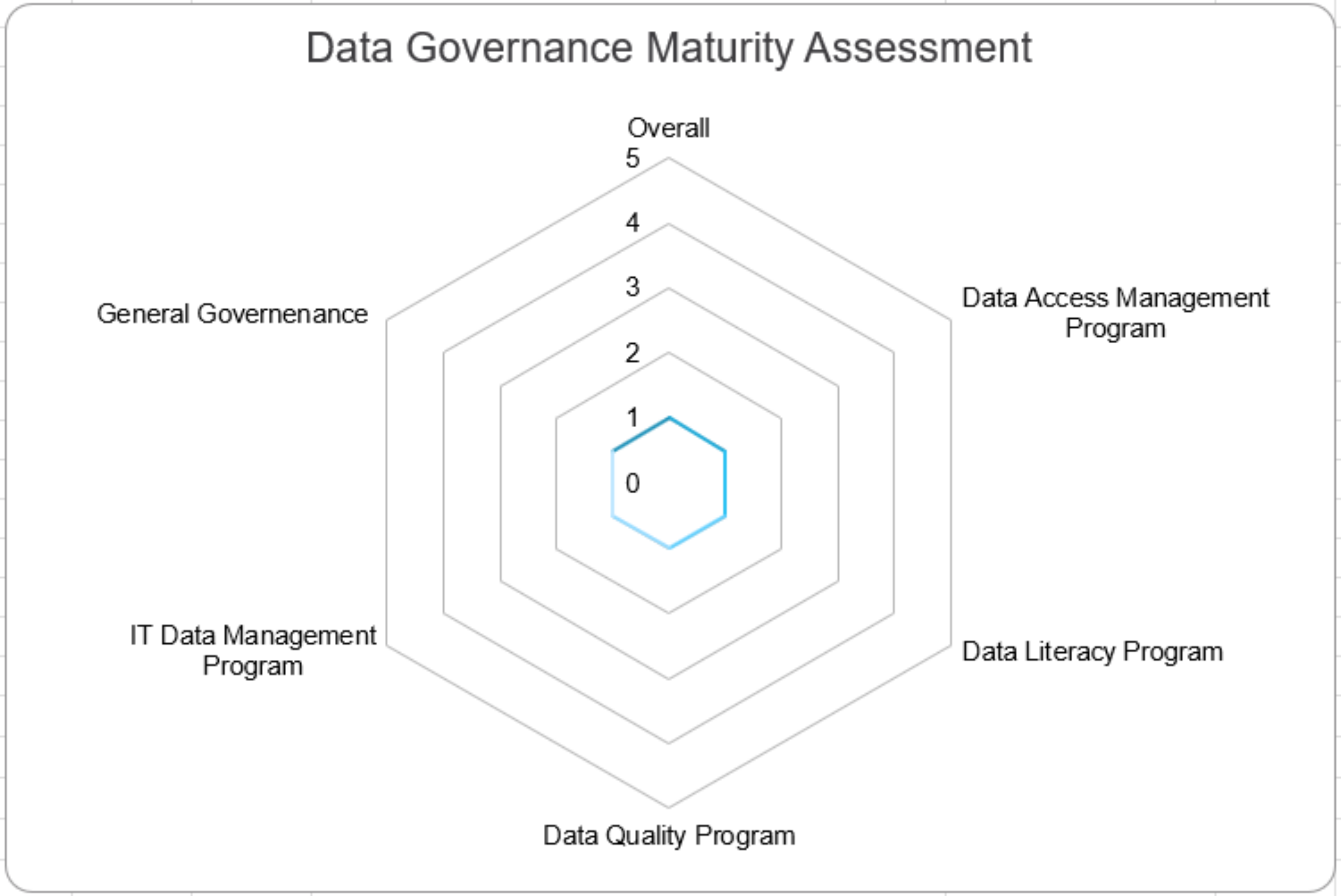
Level 2: Beginning
The importance is becoming apparent and commonly accepted. Efforts are beginning as the organization determines what is needed.

Level 3: In Progress
Policies and documentation are being created to implement a solution. Actions are being taken such as appointing people to positions or installing a tool.

Level 4: Yes
The organization is now enforcing policies and procedures for data governance. There is an implemented solution being monitored for success.

Level 5: Absolutely
The implemented solution is working and only

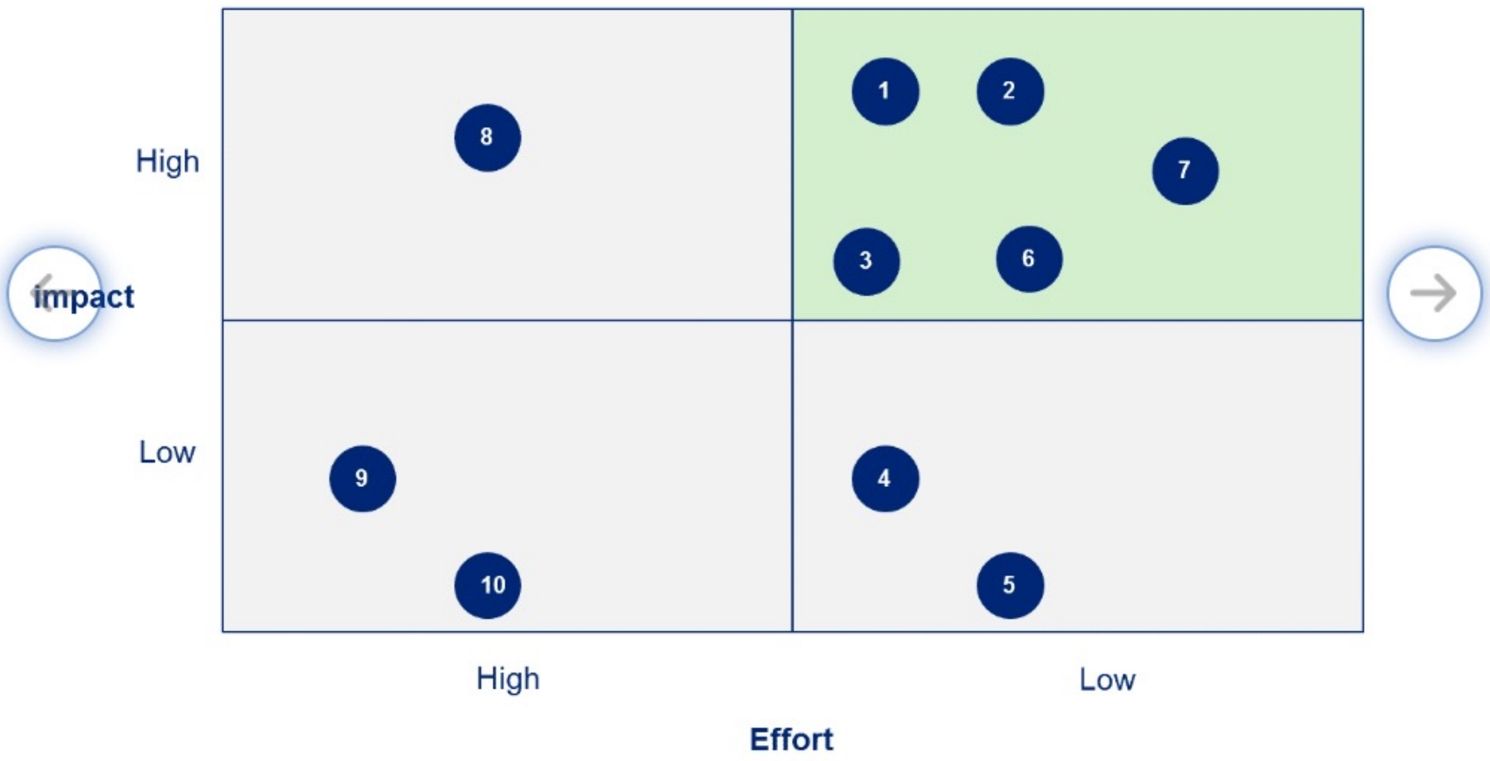
Maturity Assessment



2) Priorizar iniciativas

We recommend to prioritize the initiatives with high impact and low effort

Initiative Prioritization Matrix



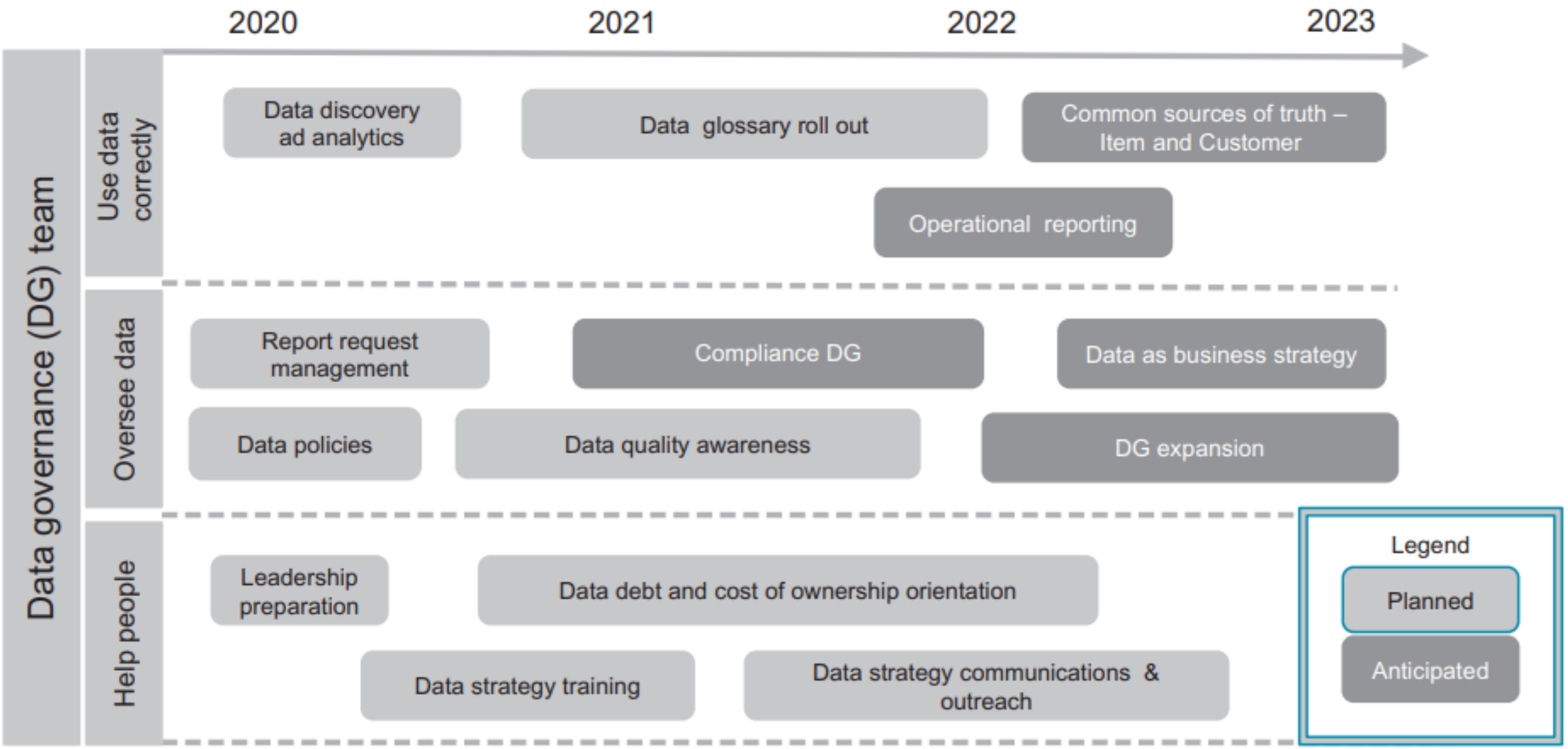
Impact	Effort
An initiative with a high impact would mean one of the followings: • The initiative needs to happen in order to achieve the desired future state • The initiative will significantly reduce our cost or increase our revenue	**The "Effort" criteria is assessed based on the followings:** • Ease of implementation • Time frame required • Resources required (Number of people, capital investment, etc.)

3) Planejar tarefas

1.Identify owners or custodians
2. Maturity Assessment
3.Managing roles and responsibilities
4.Data Access management
5.Business Glossary
6.Data Mapping & Classification
7.Data Quality Improvement Program
8.Data Literacy Program

1. Ownership and process defined and controlled
2. a clear process for resolving disputes
3. detailed documentation of business processes
4. regular data quality audits
5. a risk register that lists data-related business risks
6. data models for key business data domains; and
7. policies to limit access to critical and sensitive data
8. Audits process implemented
9. Metrics and analytics

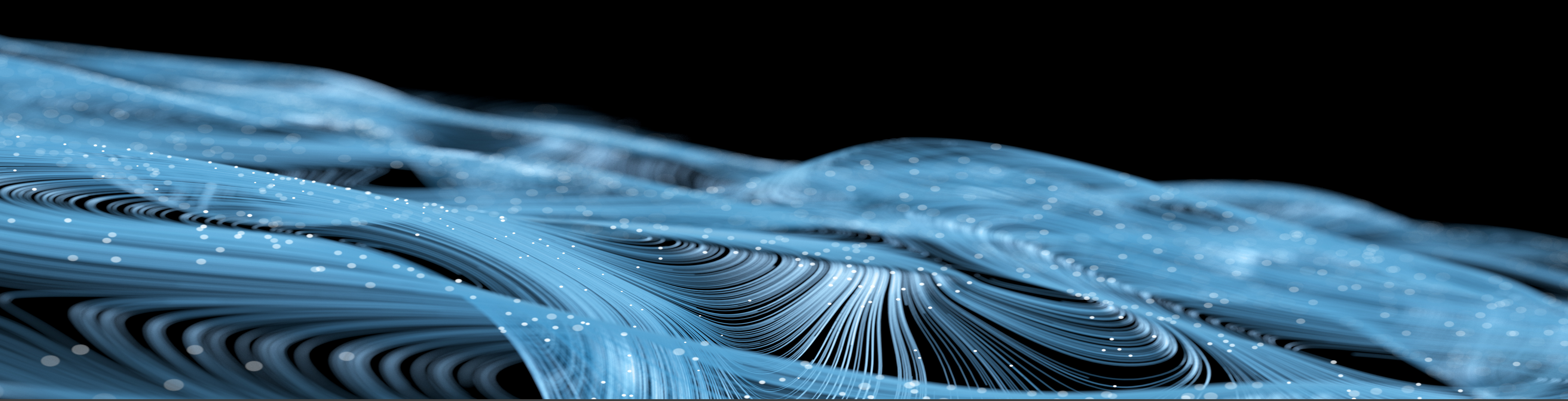
4) Fasear



2020
2021
2022
2023
Data governance (DG) team
Use data correctly
Data discovery ad analytics
Data glossary roll out
Common sources of truth – Item and Customer
Operational reporting
Oversee data
Report request management
Compliance DG
Data as business strategy
Data policies
Data quality awareness
DG expansion
Help people
Leadership preparation
Data debt and cost of ownership orientation
Data strategy training
Data strategy communications & outreach
Legend
Planned
Anticipated

Pontos importantes

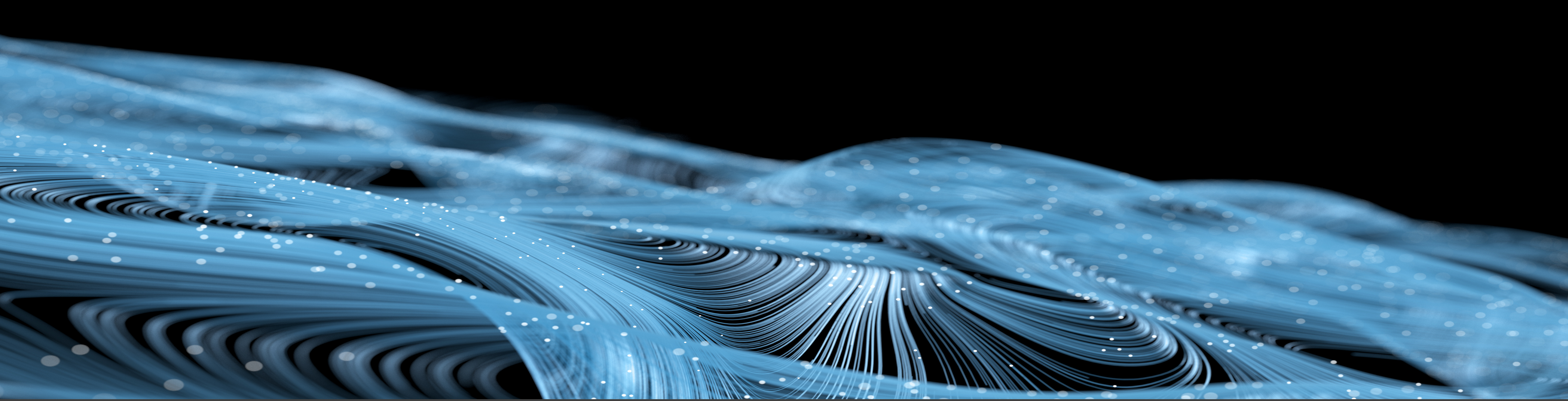
1. **Cenário atual** : motivo e razões (“as is” -> “to be”)
2. **Segmento**: qual vertical de atuação
3. **Regulamentação**: existe alguma implantada ou com prioridade (ex: LGPD)
4. **Maturidade atual**: Mapeamento da maturidade atual e a esperada.
    - Existe diagnóstico efetuado?
5. **Abrangência do escopo**: institucional, área ou projeto POV?
6. **Nível de engajamento da BU:** área de negócios/usuária
7. **Estrutura do projeto esperado** – expertise local, time local, recursos locais
8. **Subprojetos -** serão conduzidos internamentes ou por terceiros?
9. **Ferramentas atuais:** MDM, Cloud, Data Lake, Big Data, etc
10. **Métricas e expectativas de sucesso**: KPI (atual e esperada)
11. **Timeframe esperado**



Papéis & Responsabilidades

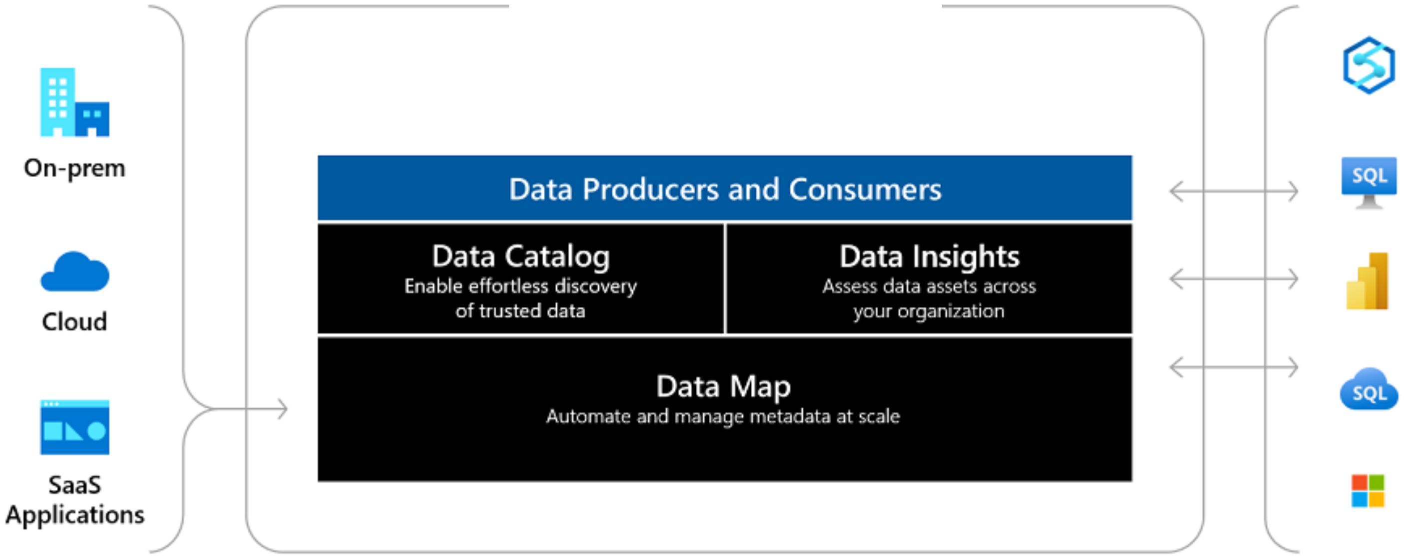
Papéis & Responsabilidades

Executive Sponsor	• coordinates data governance activities and programs • make decisions and take actions. Solve conflicts • Assign the Data Lead (ex: CDO/CIO/CFO, etc)	• works with the data governance team and taking responsibility for the implementation and ongoing data governance processes
Data governance lead	• responsible for implementing the data governance program vision, promoting the role of governance and enforcing policy, while following data governance best practices	• provide leadership, support, sponsorship, and understanding of data governance to other departments.
Data Owner	• A data owner is a person within your organization that has the authority to make decisions about business term definitions, data quality, accessibility and retention requirements as they tie to the business needs	• provides an in-depth knowledge and a sound business understanding of the data they own
Data Steward	• people with a working knowledge of the data and understand how it is used by the business on a day-to-day basis • sits under a data owner and is generally appointed by the data owner to work with them or act as their representative in data stewardship domain group meetings	• The data owner remains accountable, but they will delegate the day-to-day responsibility to a data steward. • Data stewards often tend to be the subject matter experts but are still reasonably senior because they must be trusted by their data owner
Data Custodian	• Responsible for maintaining data on your systems in accordance with the businesses requirements • Data custodians are typically part of IT departments. • They are usually divided in their areas of expertise: data modeling, data architecture, database administration, etc. • They are mainly responsible for maintaining, archiving, recovering, backing up data, preventing data loss/ corruption, etc	• Possesses the necessary technical knowledge, skill and experience • Follows good data management best practices • Is aware of regulations, policies, and standards governing the data they interact with
Data Stakeholder	• individual or group that could affect, or be affected by data governance decisions, processes, policies, standards, etc. • Examples : institutional researchers, data managers, data architects, and business intelligence staff	• Understands the importance of data and its impacts • Is collaborative and wants to be engaged • Is a champion in their area of expertise



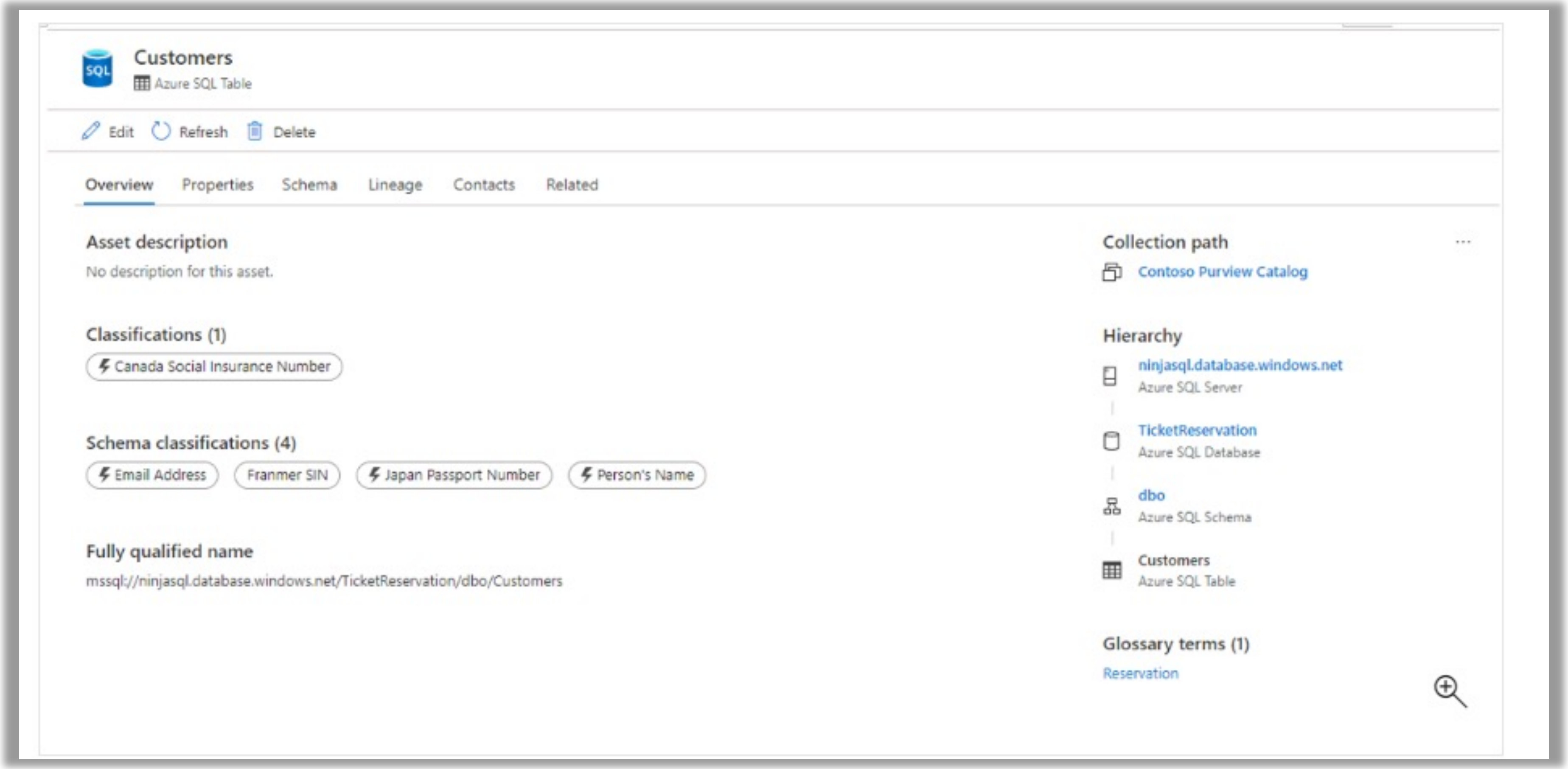
Ferramentas

Data Map & Catalog



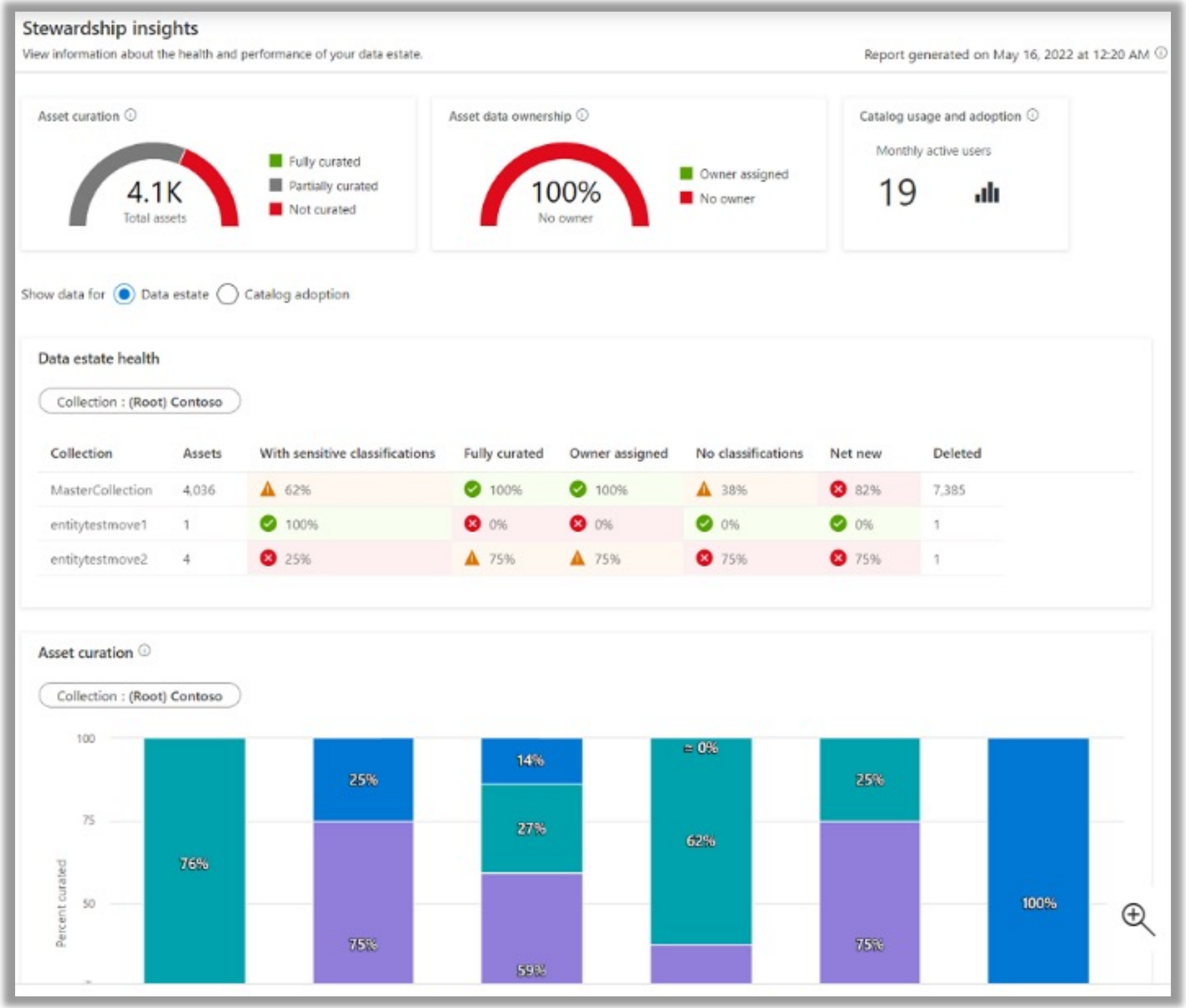
	Description
Data Map	Makes your data meaningful by graphing your data assets, and their relationships, across your data estate. The data map used to discover data and manage access to that data.
Data Catalog	Finds trusted data sources by browsing and searching your data assets. The data catalog aligns your assets with friendly business terms and data classification to identify data sources.
Data Estate Insights	Gives you an overview of your data estate to help you discover what kinds of data you have and where.

Data Map



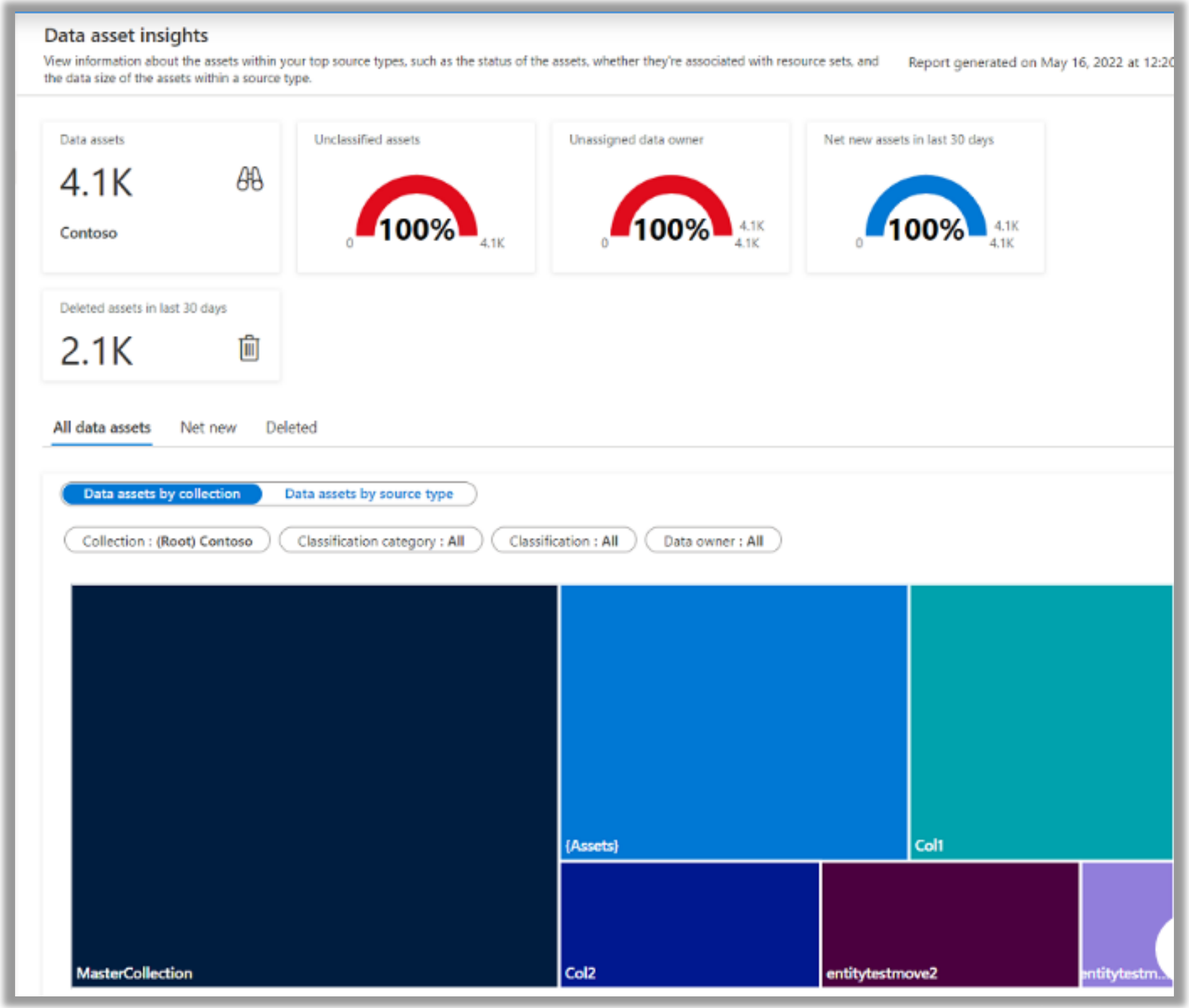
- Provides the foundation for data discovery and effective data governance.
- Service that captures metadata about enterprise data present in analytics and operation systems on-premises and cloud.
- Automatically kept up to date with built-in automated scanning and classification system.

Data Health



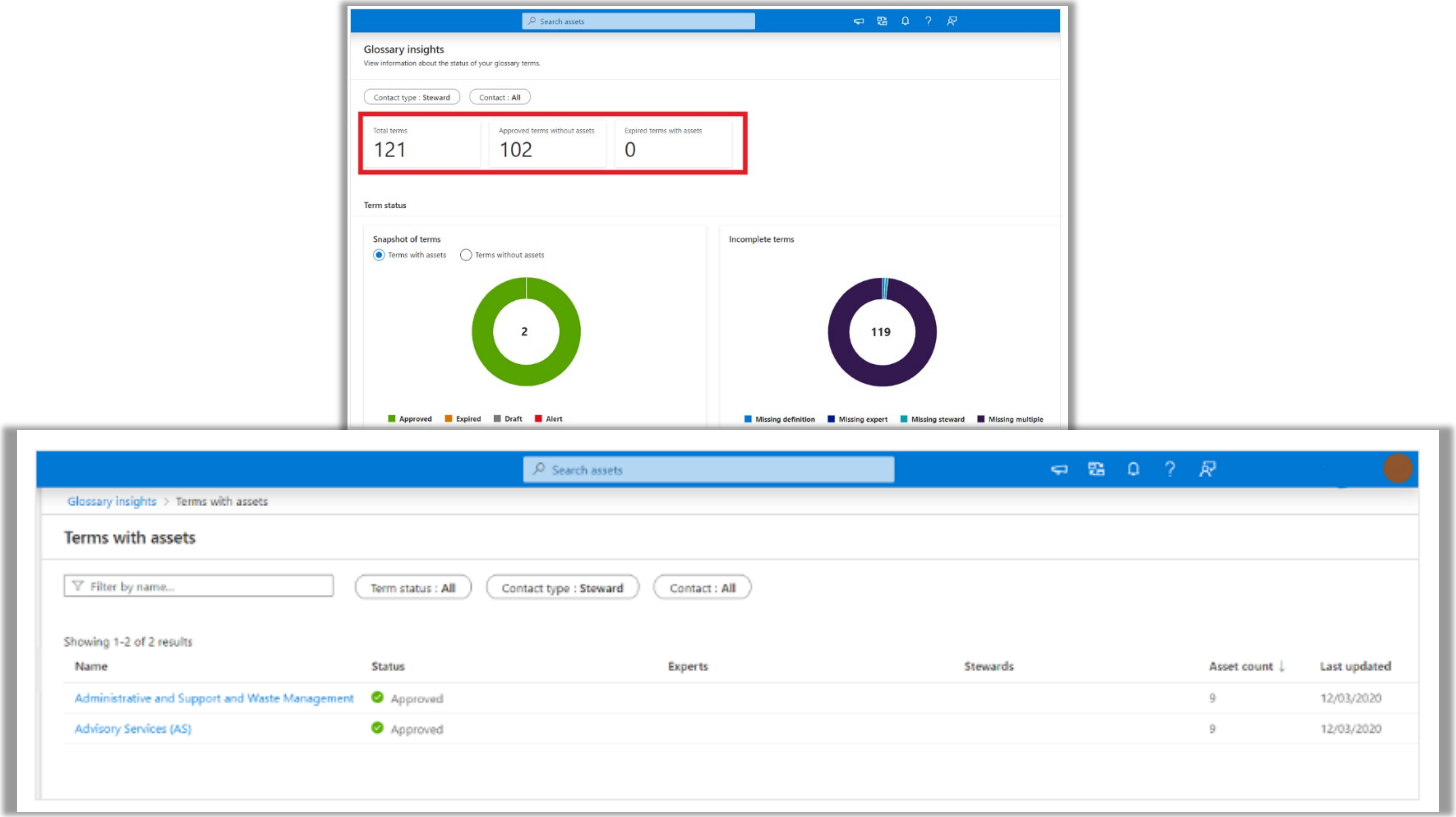
- The chief data officers (CDOs) and other governance stakeholders can get a **bird's eye view of their data estate** and can gain actionable insights into the governance gaps that can be resolved from the experience itself.

Data Inventory



- This area focuses on summarizing data estate inventory for data quality and management focused users, like data stewards and data curators.
- These dashboards provide key metrics and overviews to give users the ability to find and resolve gaps in their assets, all from within the data estate insights application.

Data Glossary



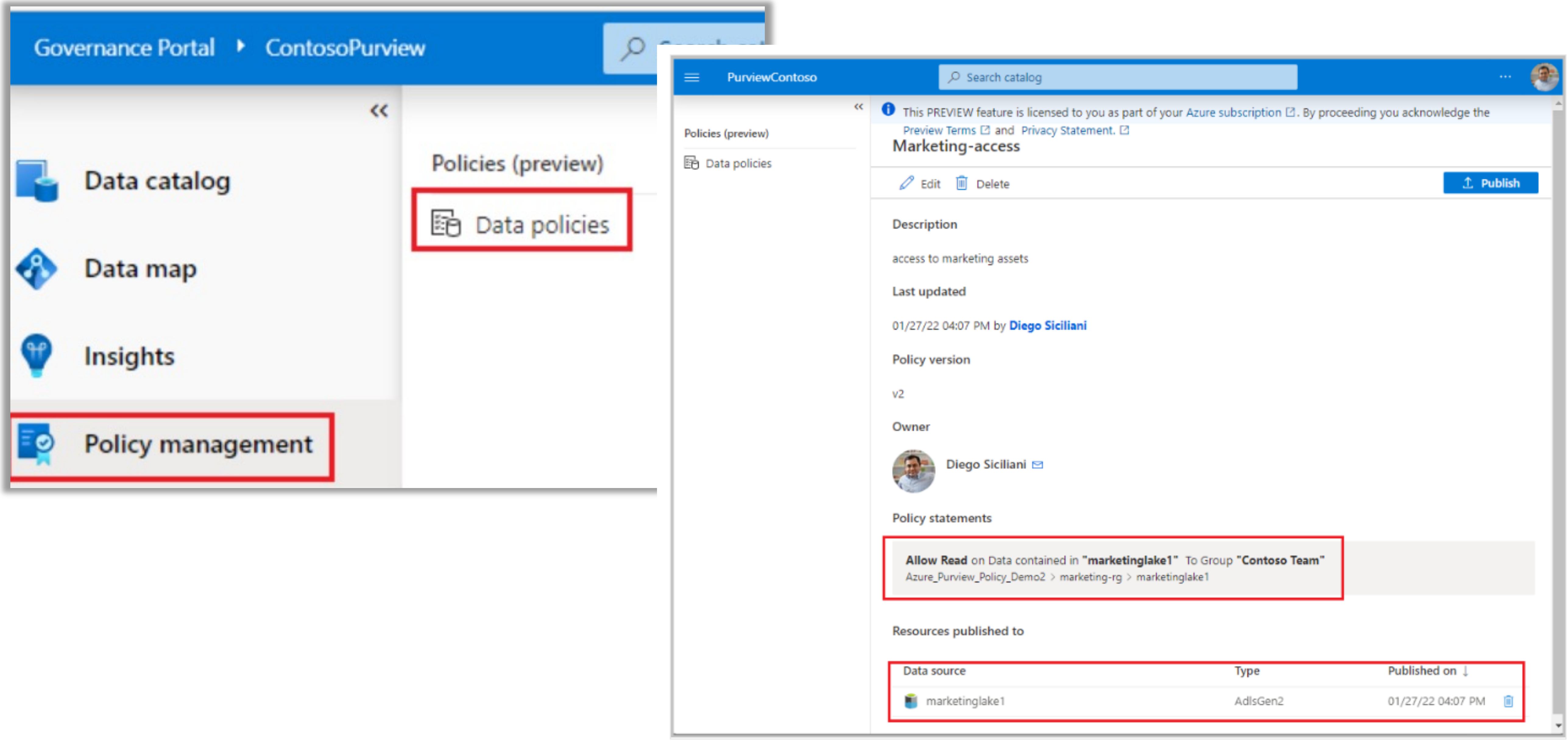
Data Catalog

- Business and technical **users can quickly and easily find relevant data** using a search experience with filters based on lenses such as glossary terms, classifications, sensitivity labels and more.

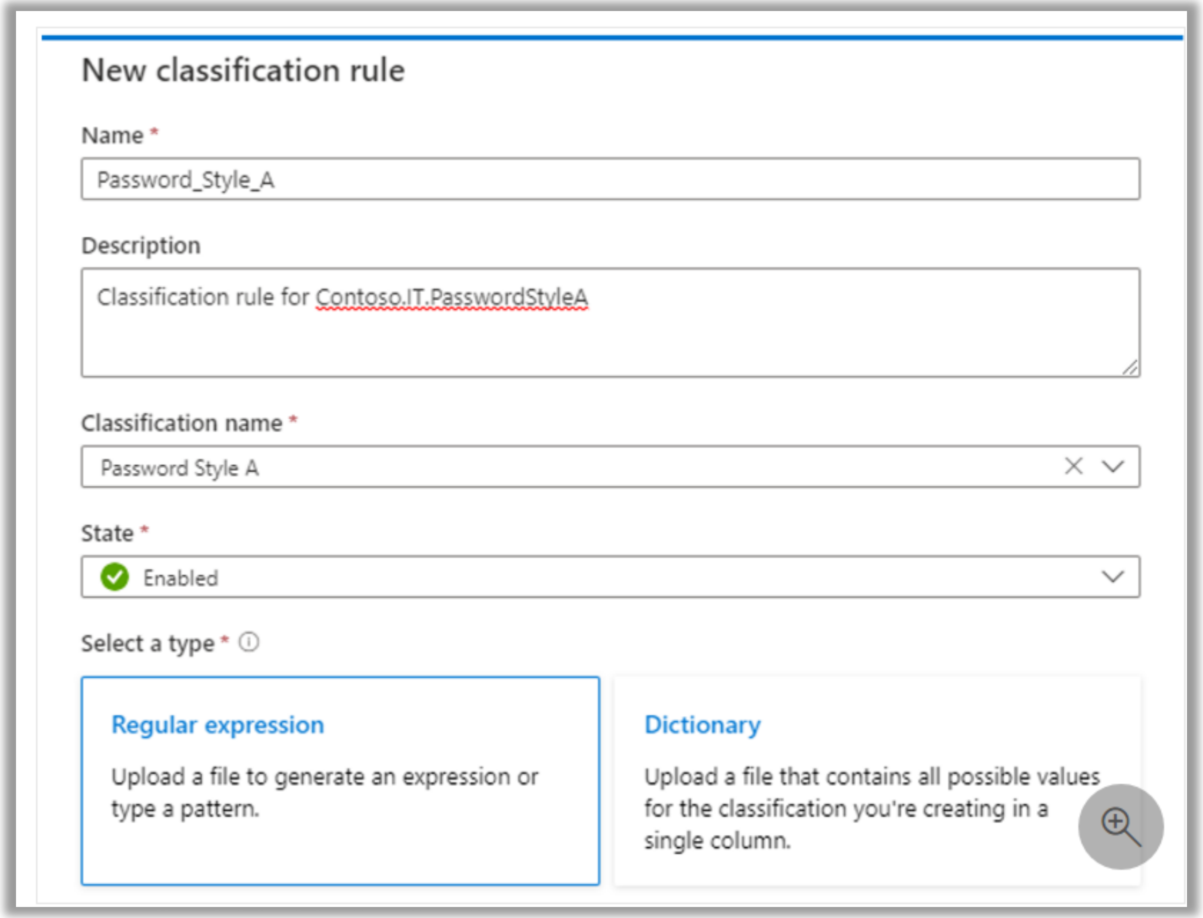
- For subject matter experts, data stewards and officers, Data Catalog provides **data curation features** such as business glossary management and the ability to automate tagging of data assets with glossary terms.

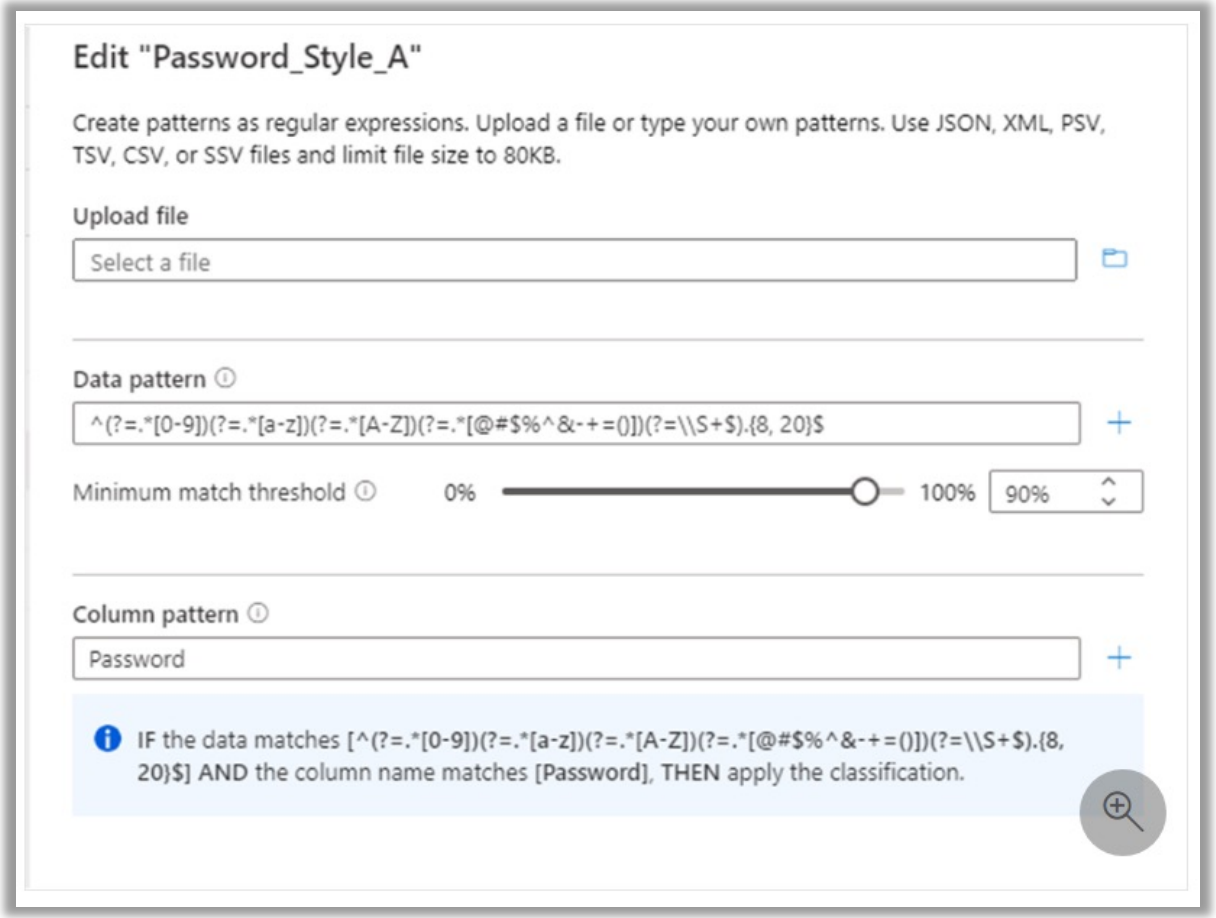
- Data consumers and producers can also visually **trace the lineage of data assets**: for example, starting from operational systems on-premises, through movement, transformation & enrichment with various data storage and processing systems in the cloud, to consumption in an analytics system.

Data policies

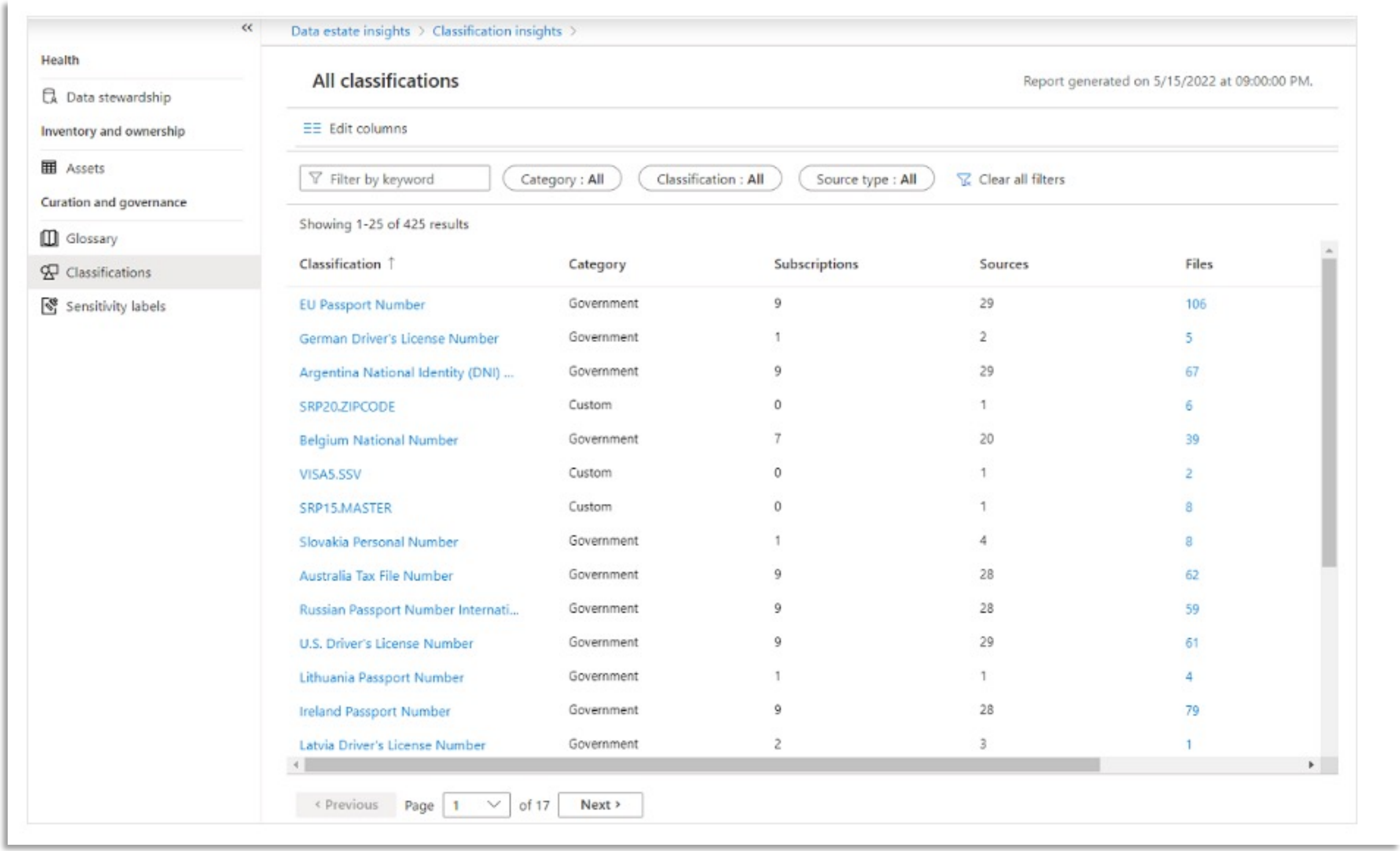


Classification rules

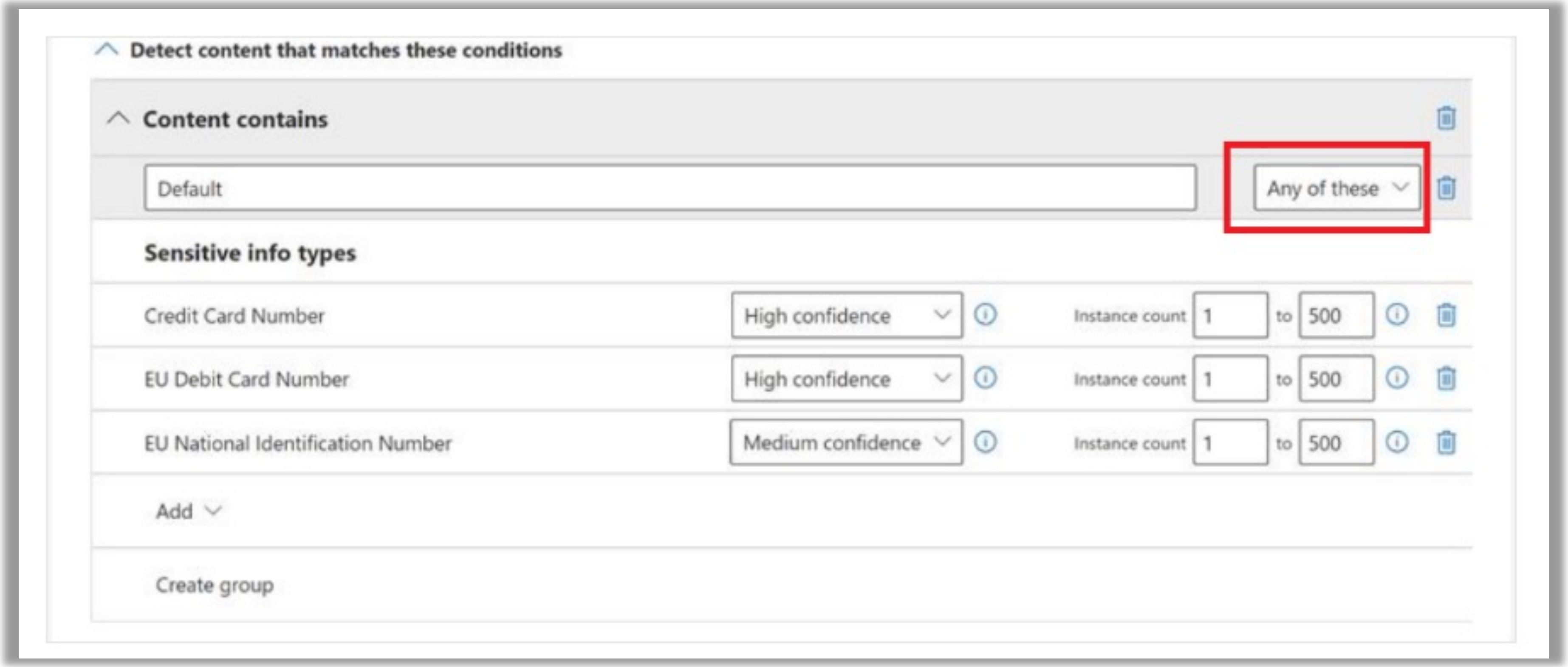




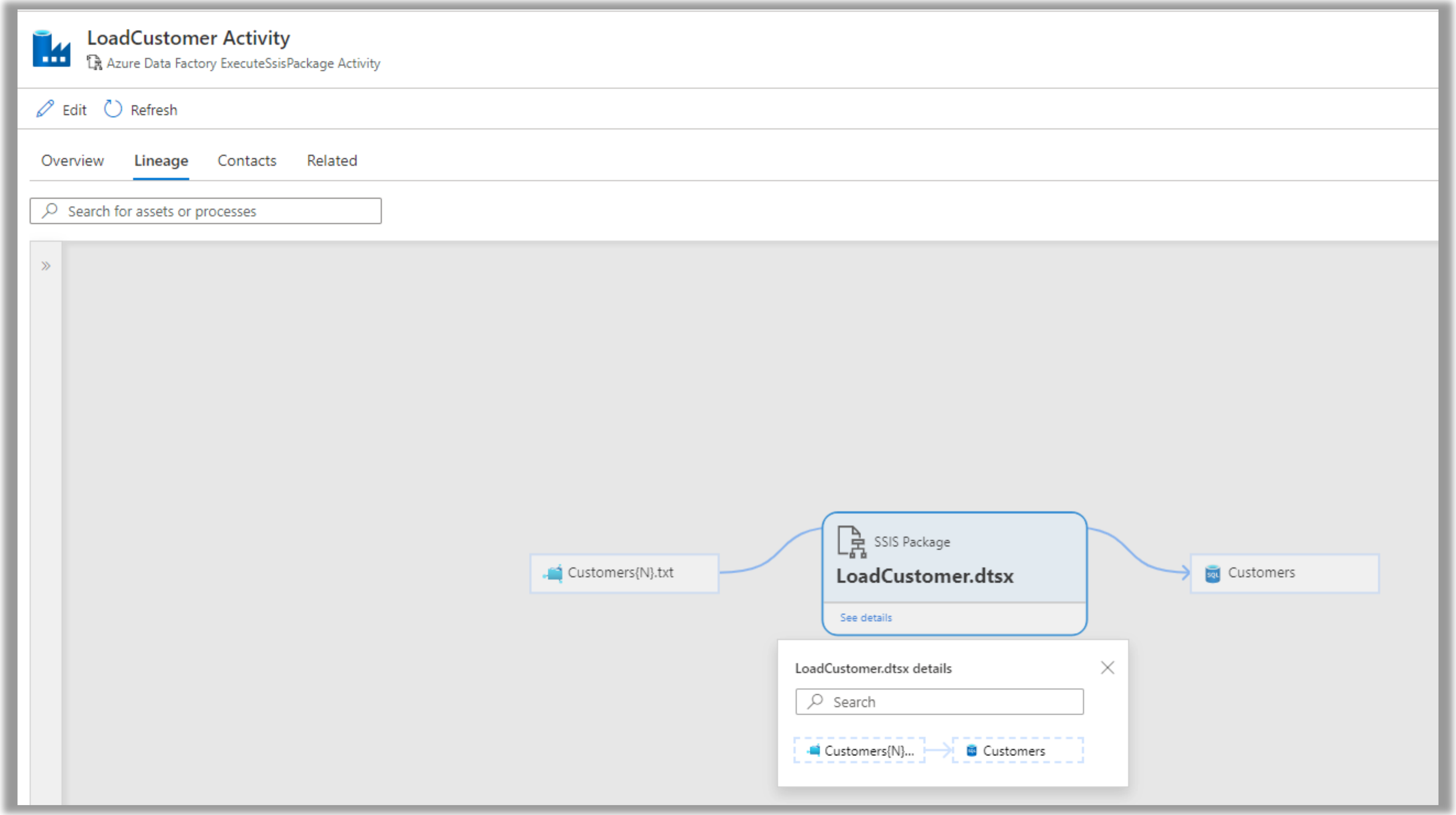
Classification



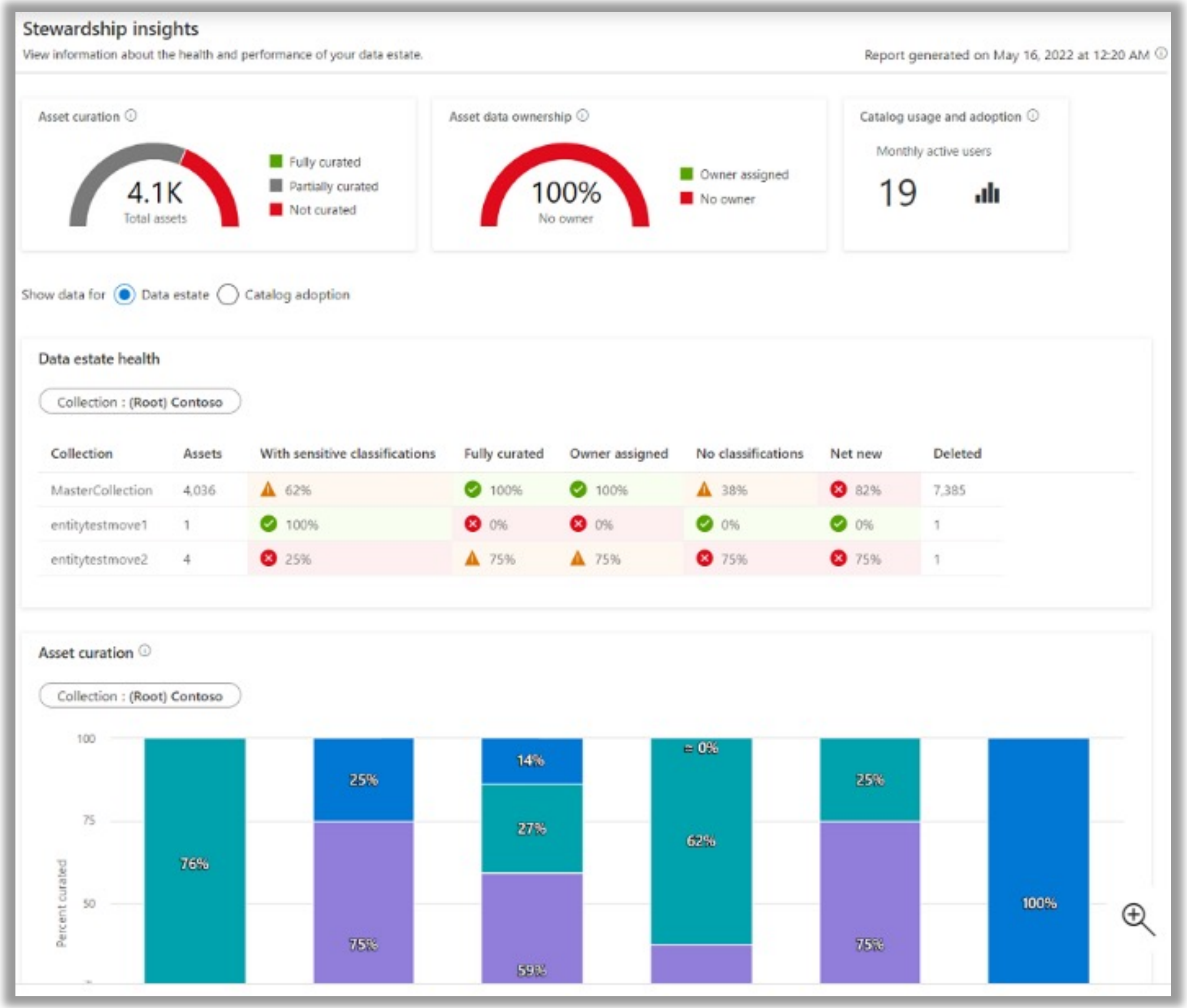
Data Labelling



Data Lineage

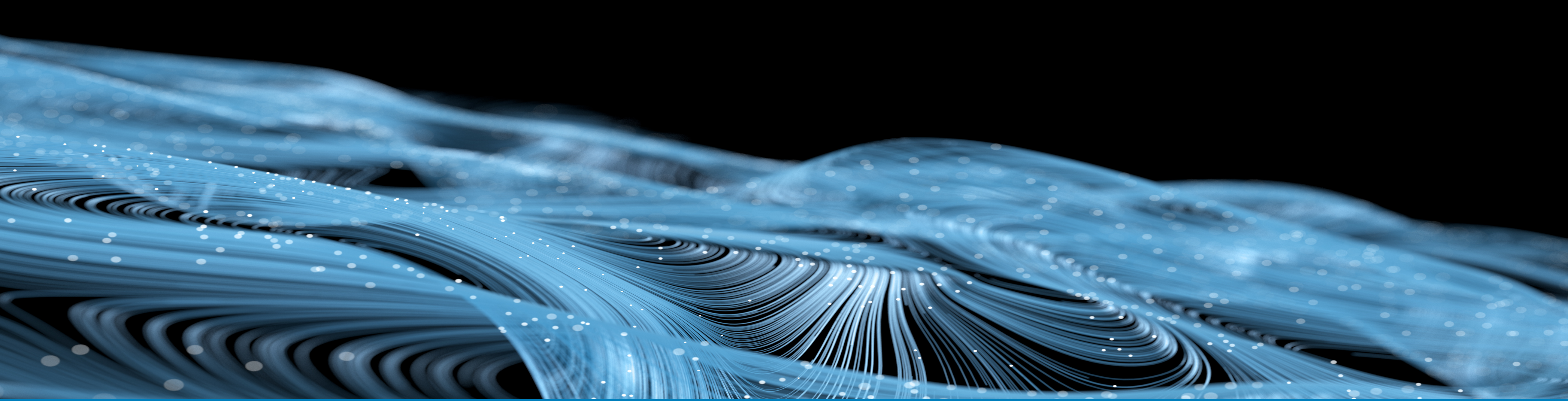


Data Stewardship report



Collection	Assets	With sensitive classifications	Fully curated	Owner assigned	No classifications	Net new	Deleted
MasterCollection	4,036	62%	100%	100%	38%	82%	7,385
entitytestmove1	1	100%	0%	0%	0%	0%	1
entitytestmove2	4	25%	75%	75%	75%	75%	1

- the KPIs tell what percentage of the data estate does not have assigned data owners and how many assets are fully, partially, and not “curated”.
- The KPIs also help with a high-level view of the catalog’s adoption, which gives an idea of whether there is significant return on the investment.



Obrigado