

NOVA

IMS

Information
Management
School

Utilização da Ciência de Dados na Auditoria dos Fundos Europeus

PEDRO SIMÕES COELHO | PSC@NOVAIMS.UNL.PT



Dados, o petróleo do Séc. XXI

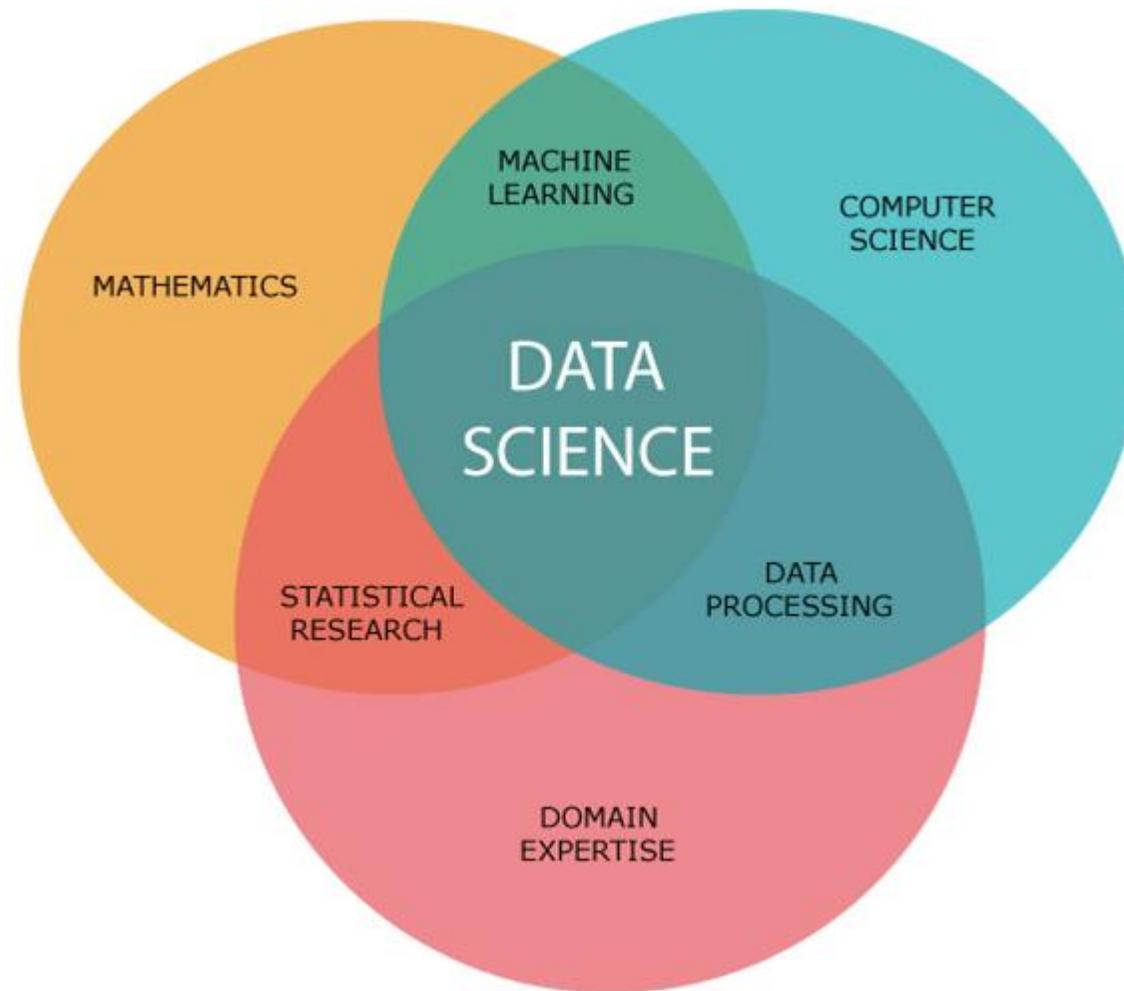
An iceberg floating in the ocean. The tip of the iceberg is above the water surface, while the much larger, jagged base is submerged. A horizontal line represents the water surface. The sky is blue with some clouds, and the water is a deep blue. Two white arrows point from the text to the respective parts of the iceberg.

← O que sabemos...

← O resto...

BIG DATA

Ciência de dados



Ciência de dados

Ciência de dados é uma área multi-disciplinar que utiliza métodos científicos, processos, algoritmos e sistemas para extrair conhecimento a partir de dados estruturados e não estruturados.

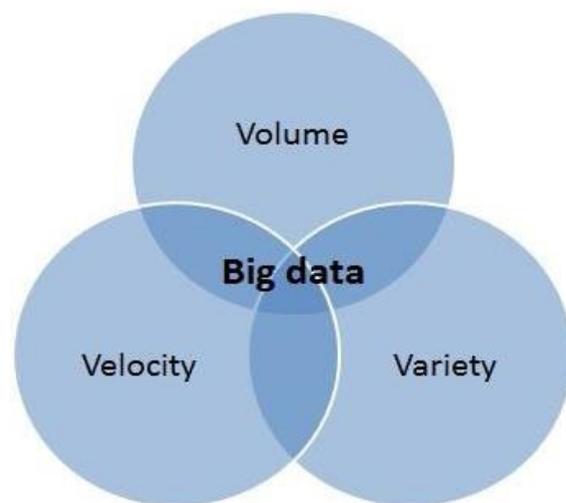
Ciência de dados é um “conceito que unifica estatística, análise de dados, aprendizagem automática e outros métodos relacionados” para “compreender e analisar fenômenos atuais” com dados. Emprega técnicas e teorias de várias áreas dentro do contexto da matemática, estatística, ciências da computação e ciências da informação.

[Wikipedia](#)

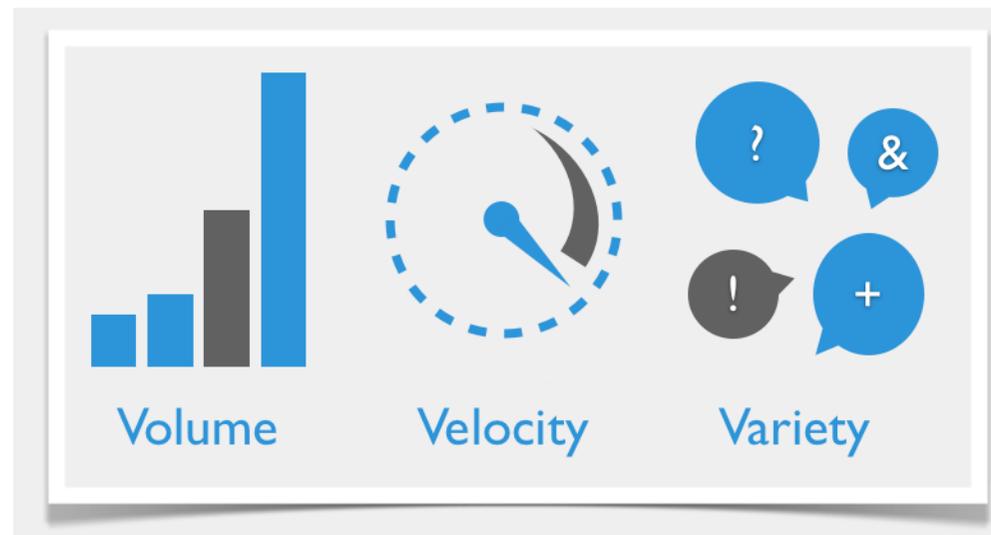
Ciência de dados

O que é *Big data*?

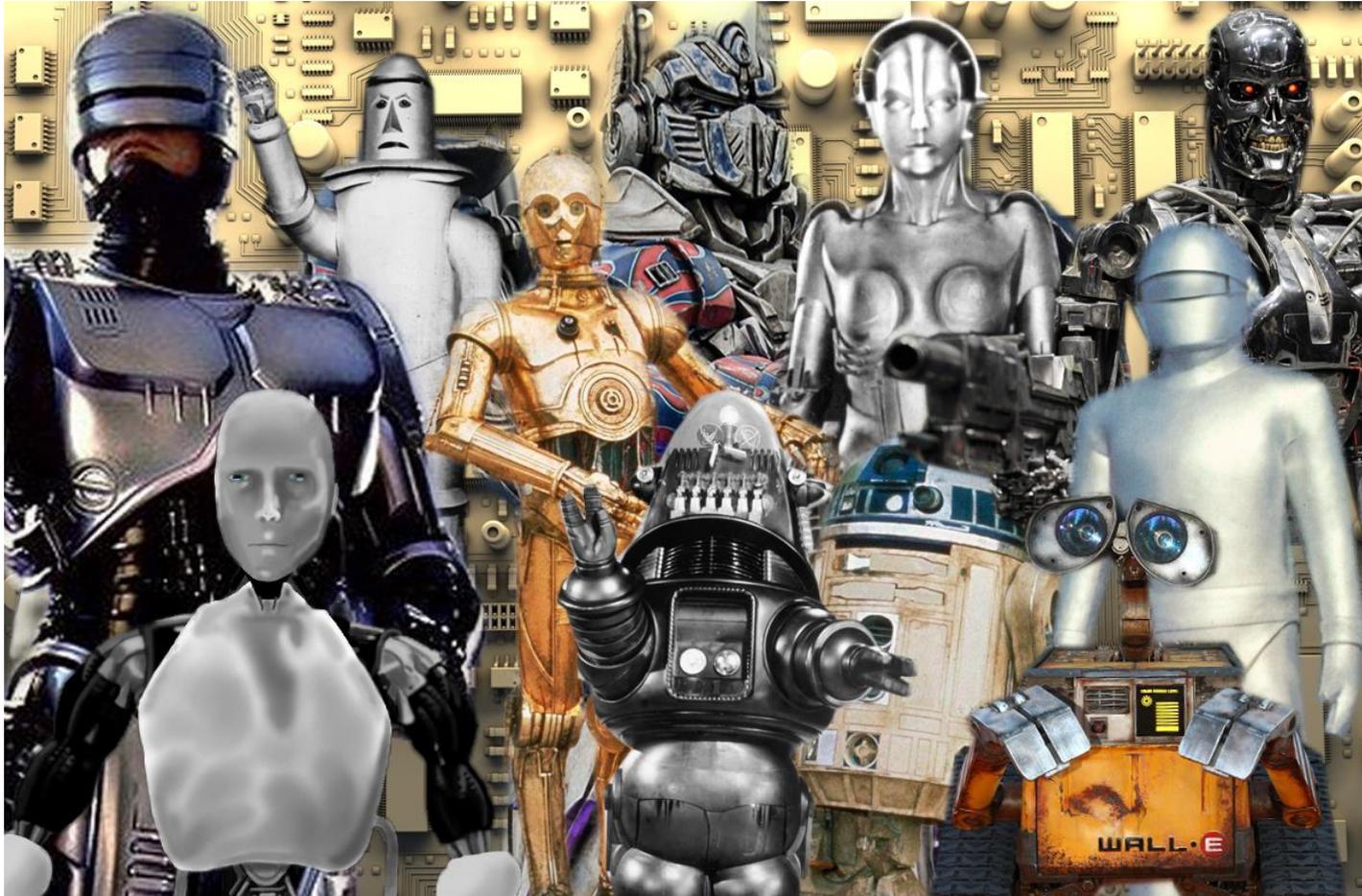
- Grandes volumes de dados complexos, disponíveis a grande velocidade e com significativa heterogeneidade em termos de estrutura, fontes e formatos.



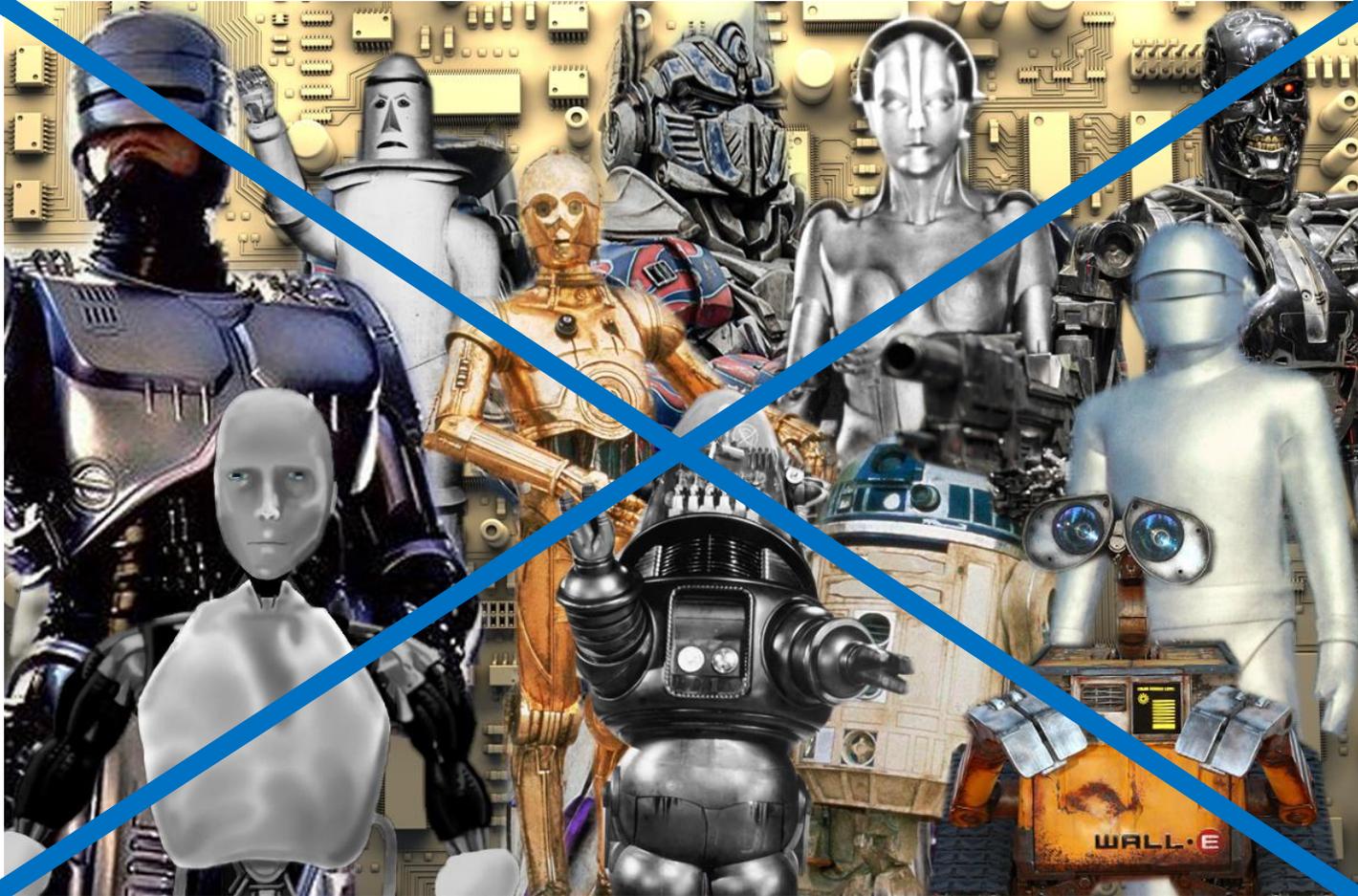
Circa 2010-12: the 3 Vs of Big Data



O que é a Inteligência Artificial?



O que é a Inteligência Artificial?



Não!!

Inteligência Artificial

- Inteligência Artificial é a programação de computadores para realizar tarefas que normalmente requerem inteligência humana.
- Isto inclui a capacidade para perceber e monitorizar informação visual/espacial e auditiva, fazer previsões, interagir com humanos e máquinas, e **continuamente aprender e melhorar**.

Artificial Intelligence for Citizen Services and Government - Ash Center

https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

► Humanos

- Ouvem e falam
- Escrevem e leem texto numa linguagem
- Veem com os olhos e processam o que veem
- Reconhecem imagens à sua volta
- Compreendem o seu ambiente e movem-se no mesmo
- Veem e entendem padrões
- Usam o seu cérebro composto por neurónios
- Lembram-se do passado

► IA

- Reconhecimento da voz
- Processamento de Linguagem Natural (NLP)
- Visão Computacional
- Processamento de imagem
- Robótica
- Reconhecimento de padrões
- Redes Neuronais (NN) e Aprendizagem Profunda
- Redes Neuronais Recorrentes (RNN)

O histórico da ideia – alguns projetos com a administração pública

- **DS4AA** - Understanding the drivers of academic achievement: Evidence for Portugal's high school system
- **MapIntel** – Plataforma de análise visual interativa para inteligência competitiva
- **AICEP** - Criação da Infraestrutura de Processamento de Dados para a Unidade de Inteligência
- **Forester and Ipsters** – Detecção Remota
- **BINDER** - Improving Bio-INspired DEep Learning for Radiomics
- **GADGET** - Online Gambling Addiction Detection
- **AICE** - Data Science and Over-Indebtedness: Use of Artificial Intelligence Algorithms in Credit Consumption and Indebtedness Conciliation in Portugal
- **IPQ Metadata** – IA para a definição de uma estratégia de predição de instrumentos de medição
- **Juris App Impacto legislativo** – Sistema de IA para avaliação de impacto legislativo
- **IVDP Data+** - A utopia da previsão na IA
- **DGT i Territorio** - Modelação Inteligente do Acesso Territorial a Serviços

O histórico da ideia – Juris App Impacto Legislativo

Sistema de IA para avaliação de impacto legislativo

- O exercício de **avaliação de impacto legislativo** em desenvolvimento em Portugal, tem por objetivo fundamental a estimação dos encargos suportados por cidadãos e empresas que possam resultar de iniciativa legislativa do Governo.
- Até recentemente era realizado por julgamento humano e extremamente intensivo em capital humano.

O histórico da ideia – Juris App Impacto Legislativo

Sistema de IA para avaliação de impacto legislativo

Objetivos

Implementar instrumentos que apoiem o legislador no exercício de **deteção das obrigações impostas num diploma legal**, permitindo a **estimação de encargos e a deteção de medidas a introduzir ou a alterar.**

Melhorar o processo da estimativa de custos administrativos para uma legislação concreta, utilizando os dados estatísticos recolhidos para obrigações padronizadas e ajustando os valores a novas obrigações

Introduzir algoritmos de Aprendizagem Automática e de Inteligência Artificial para **previsão dos impactos legislativos.**

Identificando:

- Obrigações de informação associadas
- Atividades administrativas
- Custos administrativos gerados

O histórico da ideia – Modelos na produção de estatísticas oficiais



Exemplos:

- Pereira, L.; Mendes, J. Coelho, P.S. (2013) **Model-Based Estimation Of Unemployment Rates In Small Areas Of Portugal**. *Communications in Statistics – Theory and Methods*. Vol. 42, pp. 1-18.
- Pereira, L.N. e Coelho, P.S. (2013). **Estimation of House Prices with Small Sample Sizes: the case of Portugal**. *The Annals of Regional Science*. Vol. 50, pp. 603-621. DOI 10.1007/s00168-012-0507-3.
- Coelho, P.S. (2000). **Statistical Gap Allocation in Area Estimation-An application to an Agricultural Survey**. *RevStat*, Vol 3, 31-57.

A ideia

- Desenvolvimento de novas metodologias baseadas em ciência de dados, incluindo inteligência artificial, para **estimar/prever a ocorrência e o montante de erros e seus impactos** nos **programas operacionais** suportados por fundos europeus em Portugal.
- O objetivo é o de poder utilizar estas **metodologias inovadoras em substituição ou complemento à tradicional auditoria de operações** exclusivamente baseada em auditoria por amostragem.
- Estas metodologias inovadoras poderão ser usadas para **reduzir drasticamente o esforço humano envolvido em auditoria por amostragem, aumentando a fiabilidade** das conclusões.

O contexto

- Em Portugal são realizadas anualmente milhares de atividades de **auditoria a operações financiadas**, que requerem durações, recursos humanos e custos muito significativos.
- Os **erros estimados ao nível dos programas têm implicações diversas** que podem envolver correções financeiras significativas (incluindo correções de erros extrapolados), a aplicação de taxas forfetárias de correção por parte da comissão europeia e/ou a necessidade de desenvolver planos de melhoria.
- Torna-se da maior importância:
 - **aumentar a fiabilidade das estimativas produzidas, simplificando simultaneamente o esforço;**
 - **reduzir situações em que os resultados da auditoria baseada em amostragem são inconclusivos**, conduzindo à necessidade de realização de trabalho de auditoria adicional.

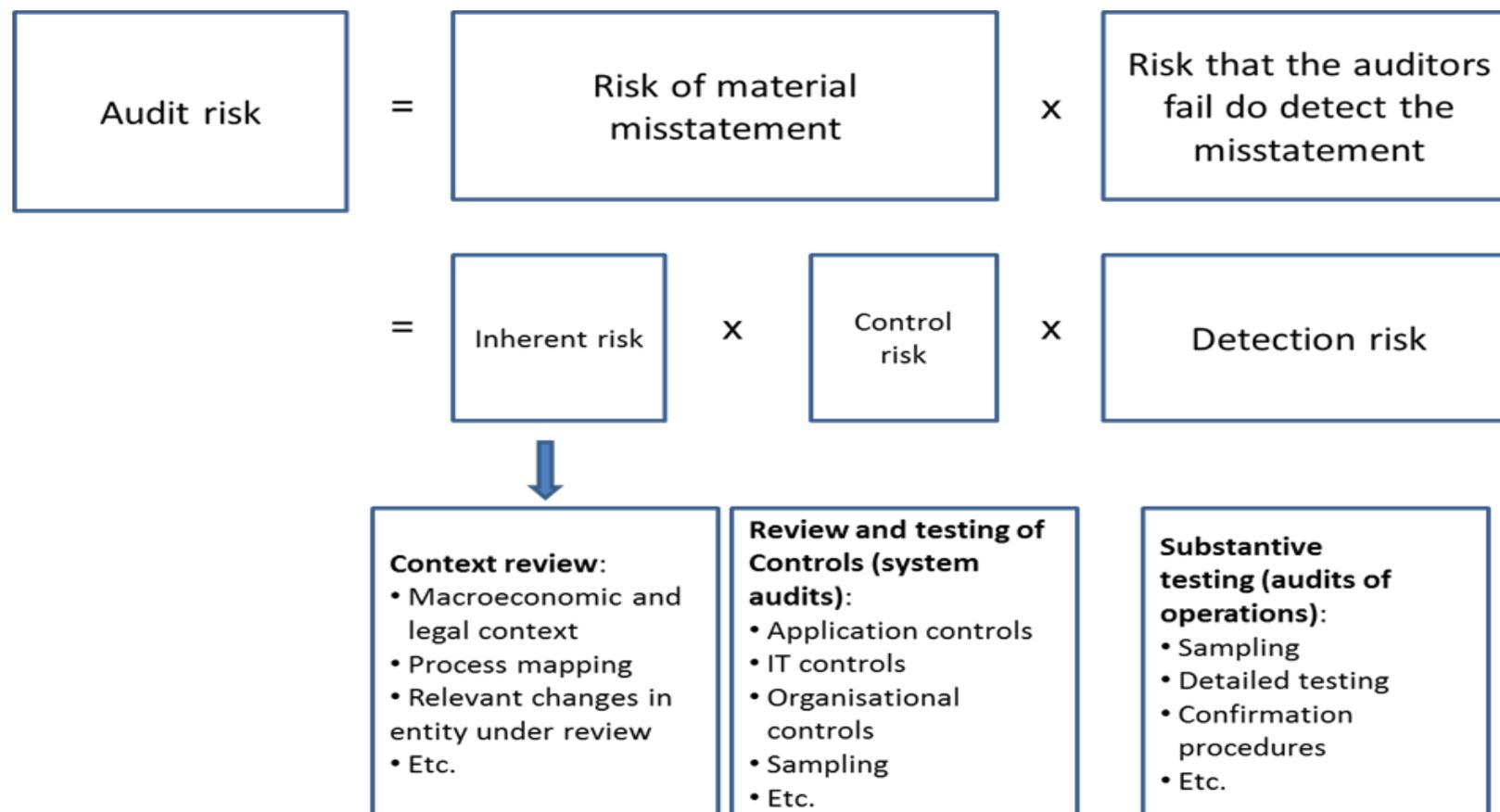
O problema

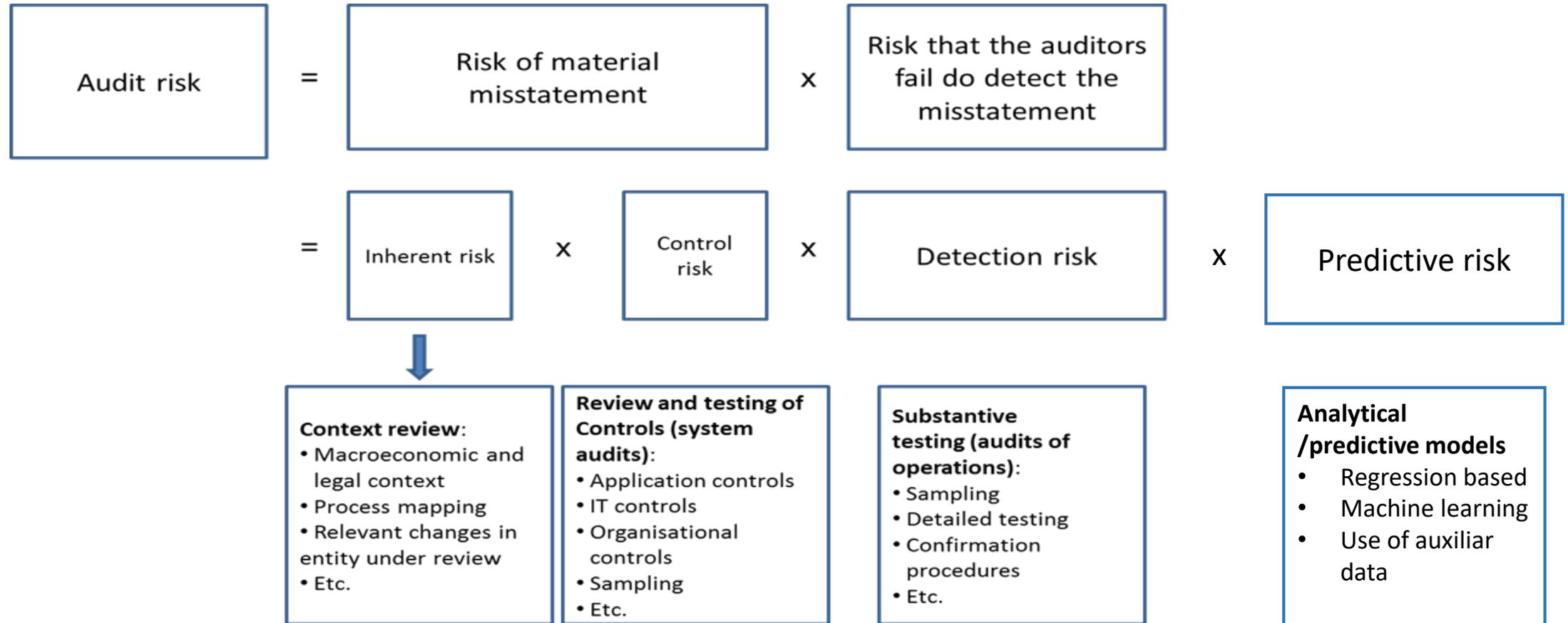
- Os problemas que se pretende resolver são de 3 naturezas:
 - A extraordinária **carga de trabalho** humana que atualmente está associada à auditoria dos fundos europeus gera custos e encargos muito significativos para a administração pública;
 - O facto de este ónus ser maioritariamente sentido pelos **beneficiários auditados**, por vezes sujeitos a auditorias múltiplas, muitas vezes morosas e perturbadoras das suas atividades organizacionais;
 - O facto de frequentemente a **garantia** pretendida pelas actividades de auditoria tradicionais **não poder ser alcançada com um esforço razoável**, conduzindo por vezes a **resultados inconclusivos** das actividades de auditoria.

Objetivos

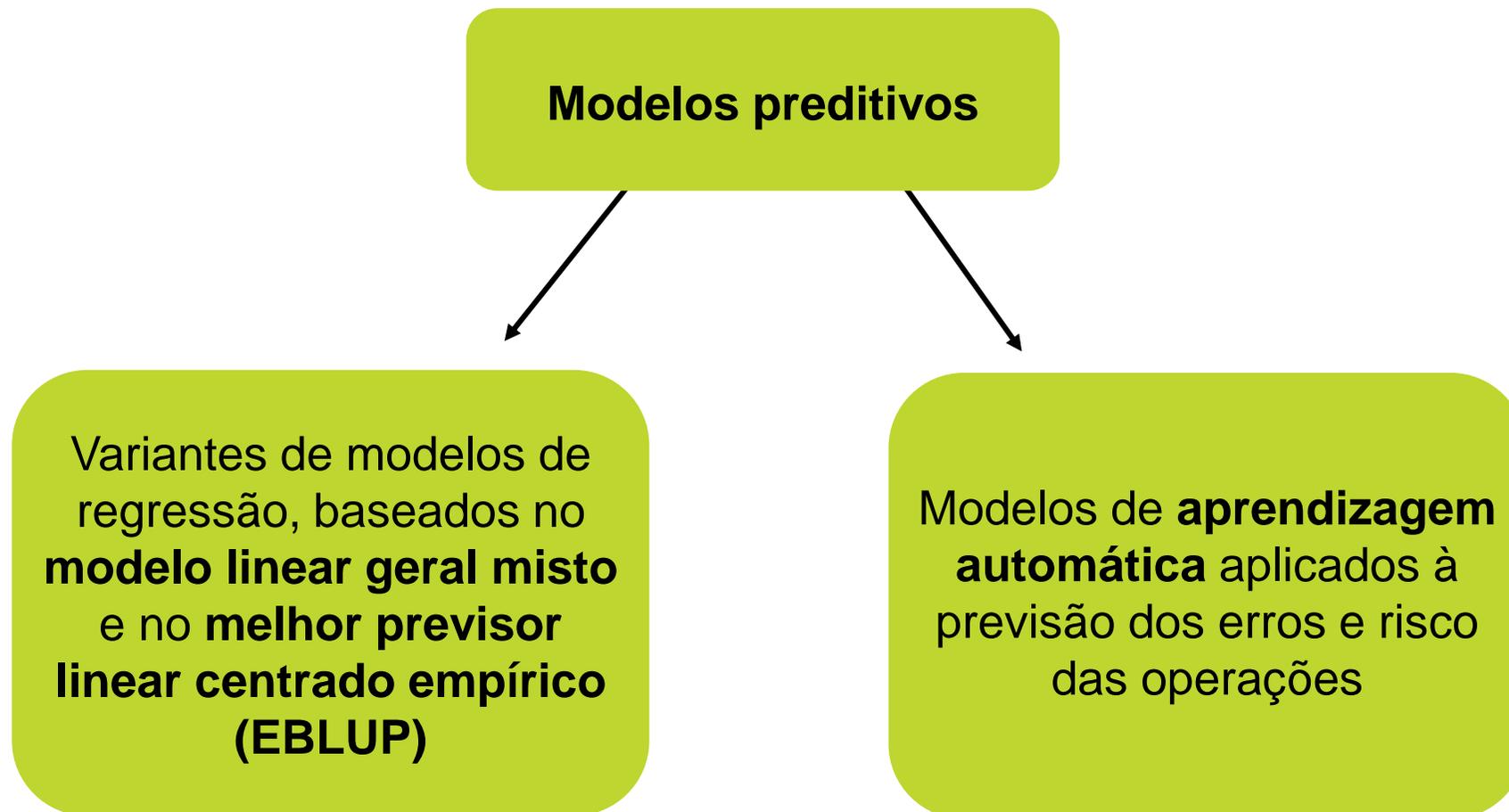
- Desenvolver metodologias inovadoras em **substituição ou complemento à tradicional auditoria** de operações exclusivamente baseada em auditoria por amostragem.
- Garantir a **fiabilidade das estimativas produzidas**, simplificando simultaneamente o esforço, nomeadamente em recursos humanos, colocado na produção das mesmas.
- **Evitar a ocorrência de resultados da auditoria baseada em amostragem inconclusivos**, que conduzam à necessidade de realização de trabalho de auditoria adicional.
- **Aumentar a proporcionalidade das auditorias**, reduzindo o ônus para os beneficiários gerado por auditorias que podem ser parcialmente substituídas por atividades analíticas com menor impacto sobre os beneficiários.
- Induzir uma **melhor exploração de dados já disponíveis na Administração Pública**.
- Garantir a aplicação de metodologias suportadas pelo poder computacional que possam **eliminar ou reduzir significativamente o esforço** atualmente realizado por recursos humanos.
- Contribuir para uma **melhoria contínua dos sistemas de gestão e controlo** e da eficiência da utilização da despesa pública.

Desenvolvendo modelos analíticos preditivos, (baseados simultaneamente em métodos analíticos estatísticos e computacionais) aplicados à previsão dos erros e risco das operações.





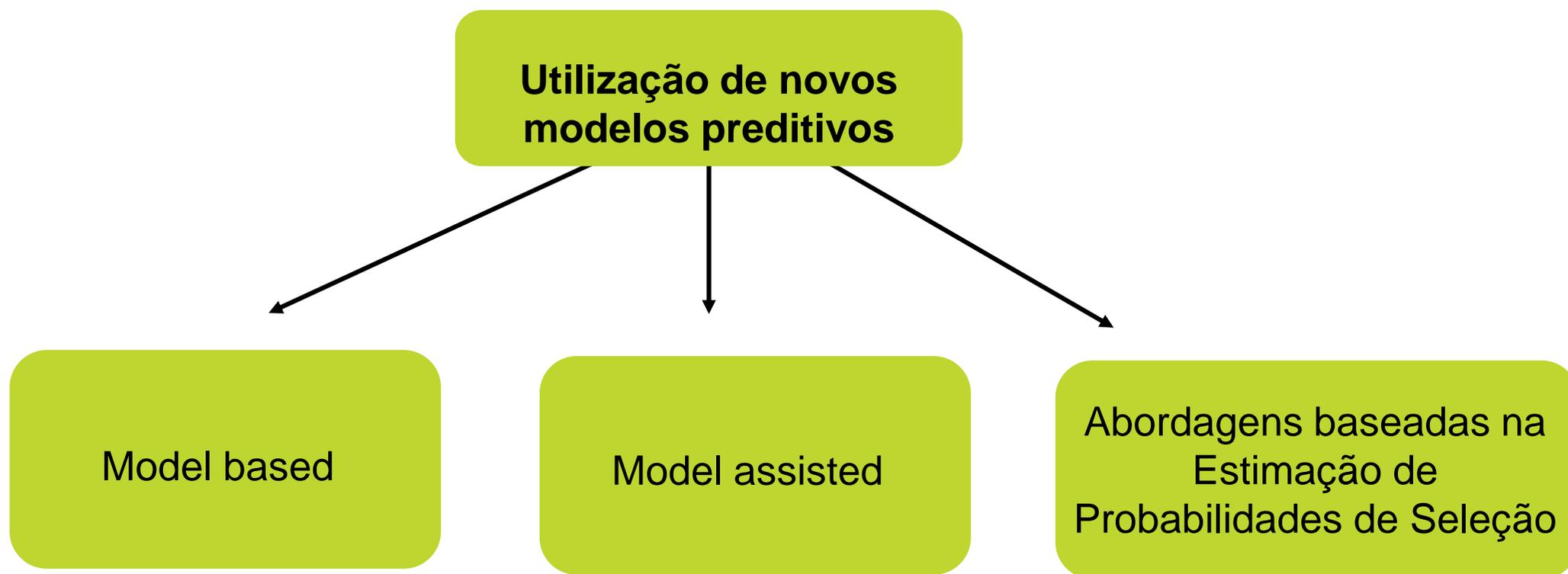
Linhas de Investigação



Linhas de Investigação

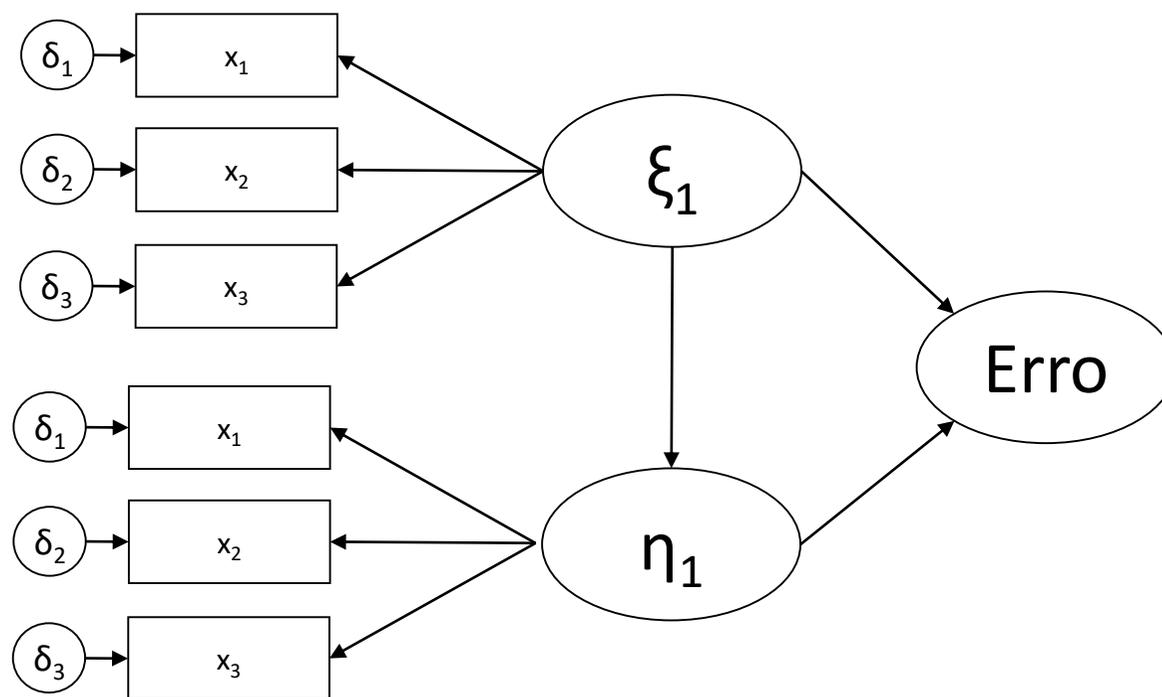
- Os modelos preditivos que se pretende vir a desenvolver passam por duas linhas de investigação complementares:
 - **Variantes de modelos de regressão**, baseados no modelo linear geral misto e no melhor previsor linear centrado empírico (EBLUP), tirando partido de trabalho realizado pela mesma equipa de investigação na estimação em domínios
 - **Modelos de aprendizagem automática** aplicados à previsão dos erros e risco das operações, tirando partido do trabalho que está a ser desenvolvido pela mesma equipa de investigação no âmbito de avaliação de impacto da legislação portuguesa.

Linhas de Investigação



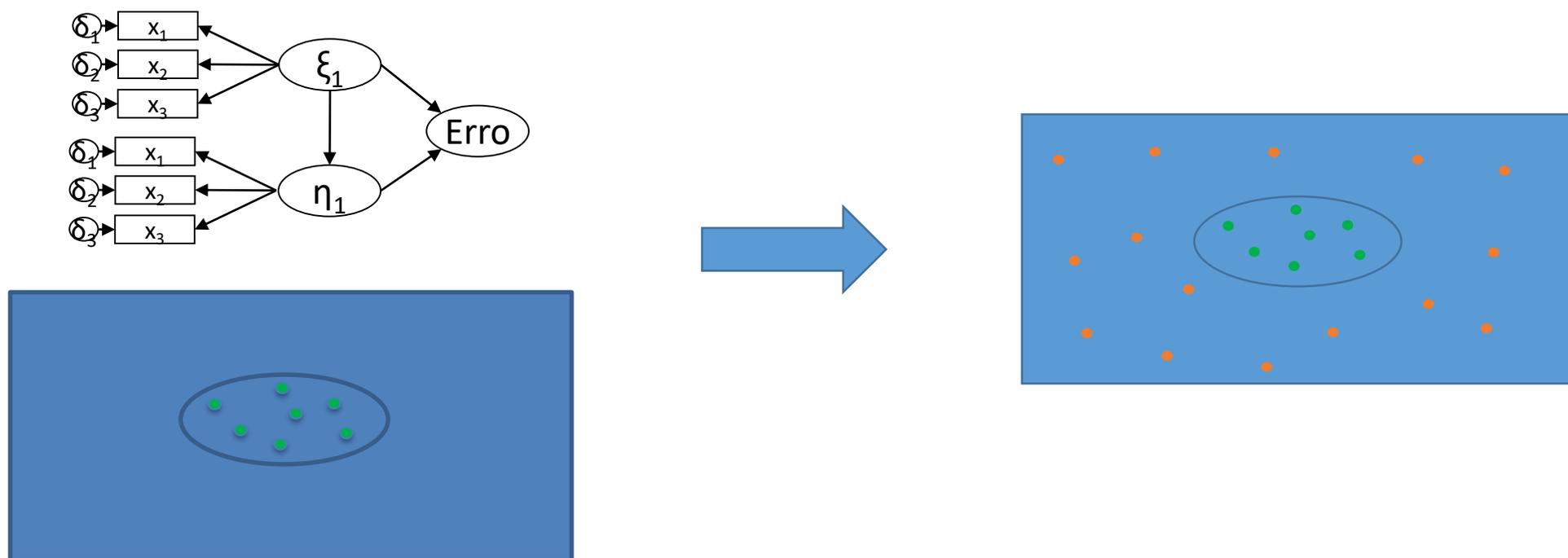
Linhas de Investigação

- **Model based:** em que os modelos serão usados para prever os montantes dos erros, o risco de erro material e a precisão da previsão; nesta abordagem os modelos podem ser usados em substituição dos métodos de amostragem em situação em que a sua precisão garante resultados conclusivos (i.e. garantia de existência ou inexistência de erro material)



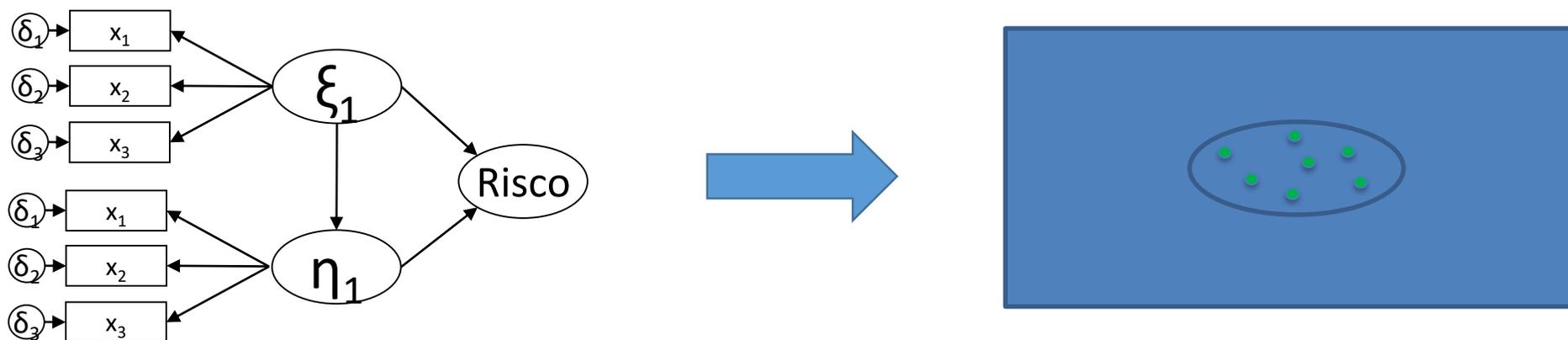
Linhas de Investigação

- **Model Assisted:** em que os modelos serão usados como parte integrante do processo de extrapolação resultante da amostragem, com o objetivo de através da explicação de parte dos erros, reduzir as dimensões amostrais (e conseqüente esforço de auditoria) e simultaneamente robustecer a qualidade da estimação, garantindo melhores níveis de precisão.



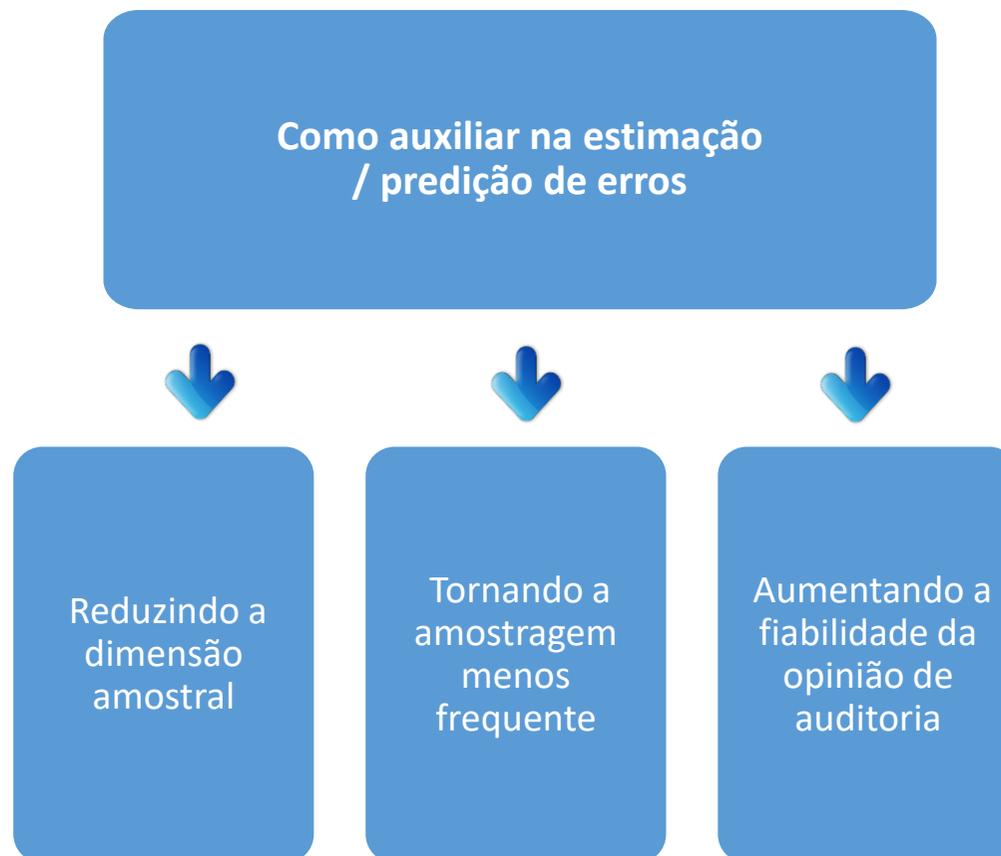
Linhas de Investigação

- **Abordagens baseadas na Estimação de Probabilidades de Seleção:** não obstante os progressos feitos nos últimos anos com o desenvolvimento do novo método das unidades monetárias, as probabilidades de seleção continuam a ser exclusivamente baseadas no montante de despesa, havendo um enorme potencial de as adequar aos riscos que venham a ser previstos pelos modelos.



Limitações / Benefícios

Mesmo que os modelos não possam substituir totalmente os testes substantivos, poderão pelo menos ser usados:



Desafios



Necessidade de adaptar os métodos de Amostragem.



Novas formas de estimação algorítmica e sua combinação com os resultados dos testes substantivos.



Necessidade de produção de medidas de precisão que permitam manter o controlo do risco de auditoria.



Restrições legais e regulamentares.

Três requisitos para a aplicação útil da ciência de dados nas organizações:

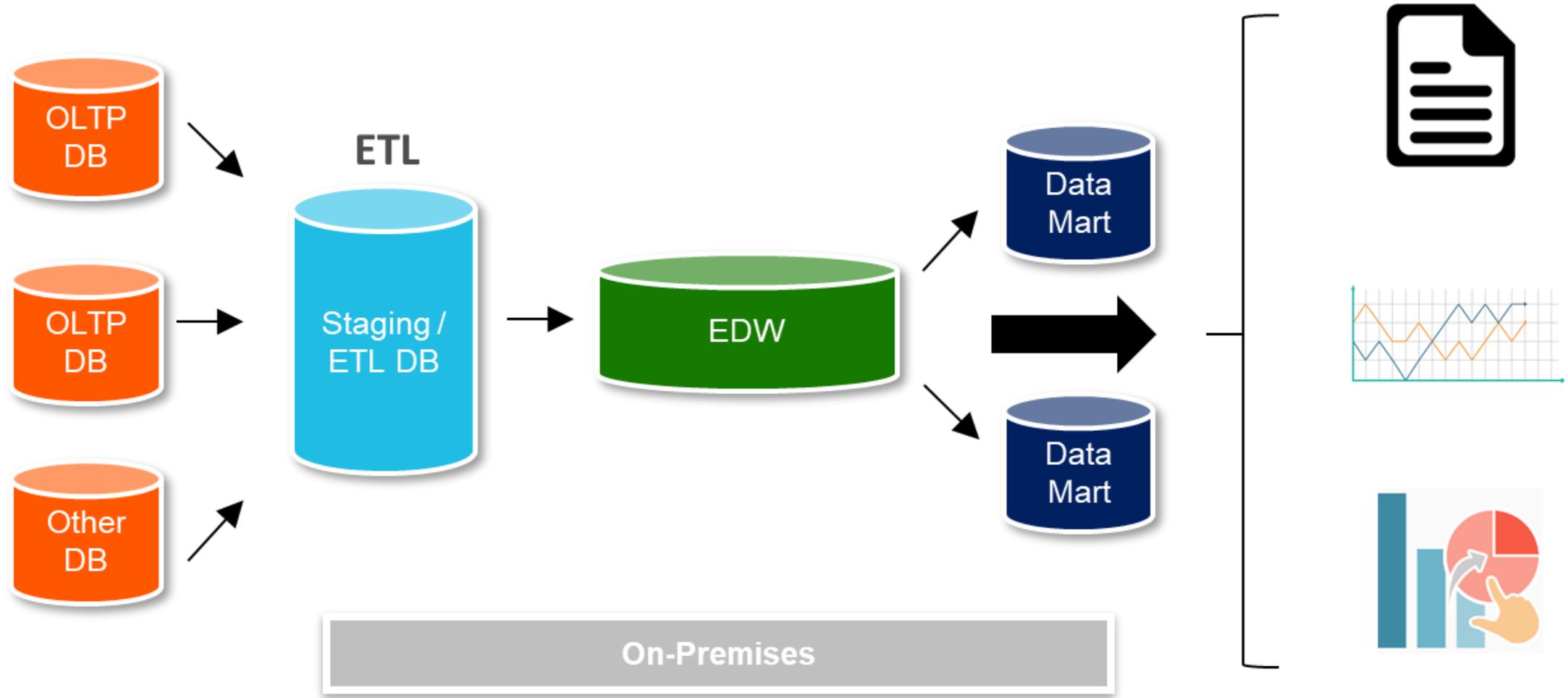
- Adequada gestão de dados
- Disponibilidade de recursos humanos
- Aquisição de conhecimento

Gestão de Dados

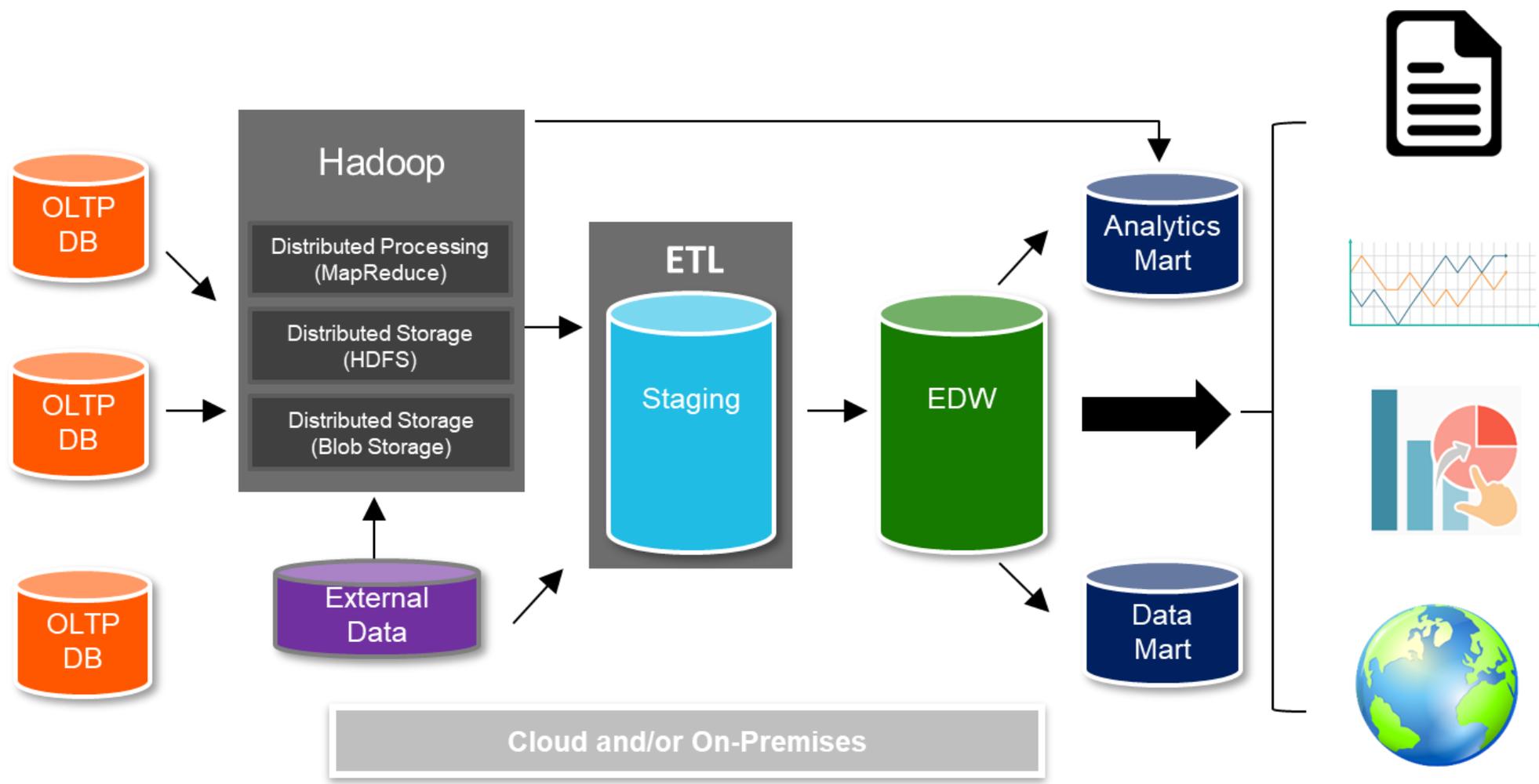


- Relevância
- Acesso
- Quantidade
- **Qualidade**
- **Governança**
 - ▶ Segurança
 - ▶ Privacidade

Arquitetura de Armazenamento de Dados clássica

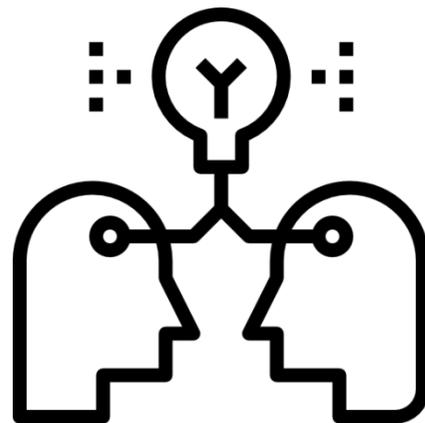


Arquitetura de Armazenamento de Dados hoje



Aquisição de reconhecimento

- Um dos principais desafios é a **transferência de conhecimento**
- O auditor de hoje é também um analista de dados
- São necessárias novas formações híbridas que desafiam as fronteiras dos conhecimentos tradicionais



The Jobs Landscape in 2022



Formação Académica

A formação académica deve **combinar novos saberes académicos** em temas como:



Saberes tradicionais
relacionados auditoria

O que implica

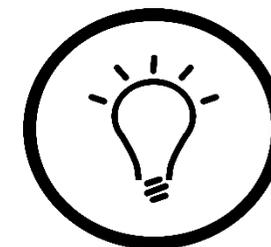
Trabalho de **parceria** entre a **universidade** e a **administração pública** no desenho dos cursos e também a construção de corpo docente misto.

A formação académica deve igualmente assumir novas **formas diferentes** dos tradicionais cursos conferentes de grau. Com **durações e formatos diversos**.

Formação Académica

A **NOVA IMS** lançou nos últimos anos um **conjunto de pós graduações/mestrados** que **respondem a estes desafios**:

- **Gestão e controlo orçamental (IGF);**
- Gestão de Informação e Business Intelligence na Saúde (**ACSS**);
- Gestão de Informações e Segurança (**SIRP e IDN**);
- Sistemas Estatísticos (**Banco de Portugal**); spin off para o **Sistema Estatístico Europeu e Sistema Europeu de Bancos Centrais** (utilização de ensino à distância);
- Métodos Analíticos para as Finanças Públicas (**DGO**);
- Cidades Inteligentes (**ANMP, CML, CMC**).



NOVA

IMS

Information
Management
School

Utilização da Ciência de Dados na Auditoria dos Fundos Europeus

PEDRO SIMÕES COELHO | PSC@NOVAIMS.UNL.PT