

# Explorar o potencial da Inteligência Artificial no contexto do arquivo

Luis Borges Gouveia

lmbg@ufp.edu.pt, Universidade Fernando Pessoa – Portugal

V Colóquio REVIS-ARQ

Grupo de Pesquisa Registros Visuais e Sonoros: Arquivo e Memória

24 de Novembro de 2021

**V COLÓQUIO**  
REVIS-ARQ



Dia 23, 24 e 25 de novembro/2021

---

Transmissão ao vivo pelo canal Revis-Arq no YouTube  
<https://youtube.com/c/REVISARQGrupodepesquisaREVISARQ>

# Explorar o potencial da Inteligência Artificial no contexto do arquivo

Luis Borges Gouveia

*A crescente sofisticação do uso do digital e o seu impacto tem proporcionado novos desafios e oportunidades. Os arquivos, bem como a necessidade de preservação da informação e da memória, também são confrontados com estes novos desafios do digital.*

*A inteligência artificial (IA), entendida aqui como a capacidade de simular os processos da inteligência humana por meios computacionais, tira partido do digital e da crescente digitalização da atividade humana. Tal adiciona um conjunto de desafios significativos também para o arquivo. Existe assim um potencial para a IA no contexto do arquivos que importa analisar.*

**V Colóquio REVIS-ARQ**, 24 de Novembro de 2021

# O contexto atual

Complexo, volátil e em transformação constante

# O mundo mudou...

- Cada vez mais **digital**:
  - **Próximo**  
(o **espaço físico** é menos importante e menos distante)
  - **Rápido**  
(o **tempo humano** está a ser gradualmente substituído pelo **tempo das máquinas**)

# O mundo mudou...

- Cada vez mais **complexo**:

- **Próximo**

A sustentabilidade ambiental, económica e social do mundo, afeta a nossa região, que afeta o nosso país, que afeta a nossa cidade e, por sua vez, implica com a nossa comunidade e conosco

(sociedade, comunidade, individuo) **para** (global, regional, nacional, local e pessoal)

- **Desordenado**

As instituições que governam tem de se adaptar a um ecossistema de partilha de poder em rede, substituindo hierarquias por redes de negociação – governança

(do comando e controlo **para** a **partilha** e **colaboração**)

# O mundo mudou...

- Cada vez mais **alternativo**

- **Próximo**

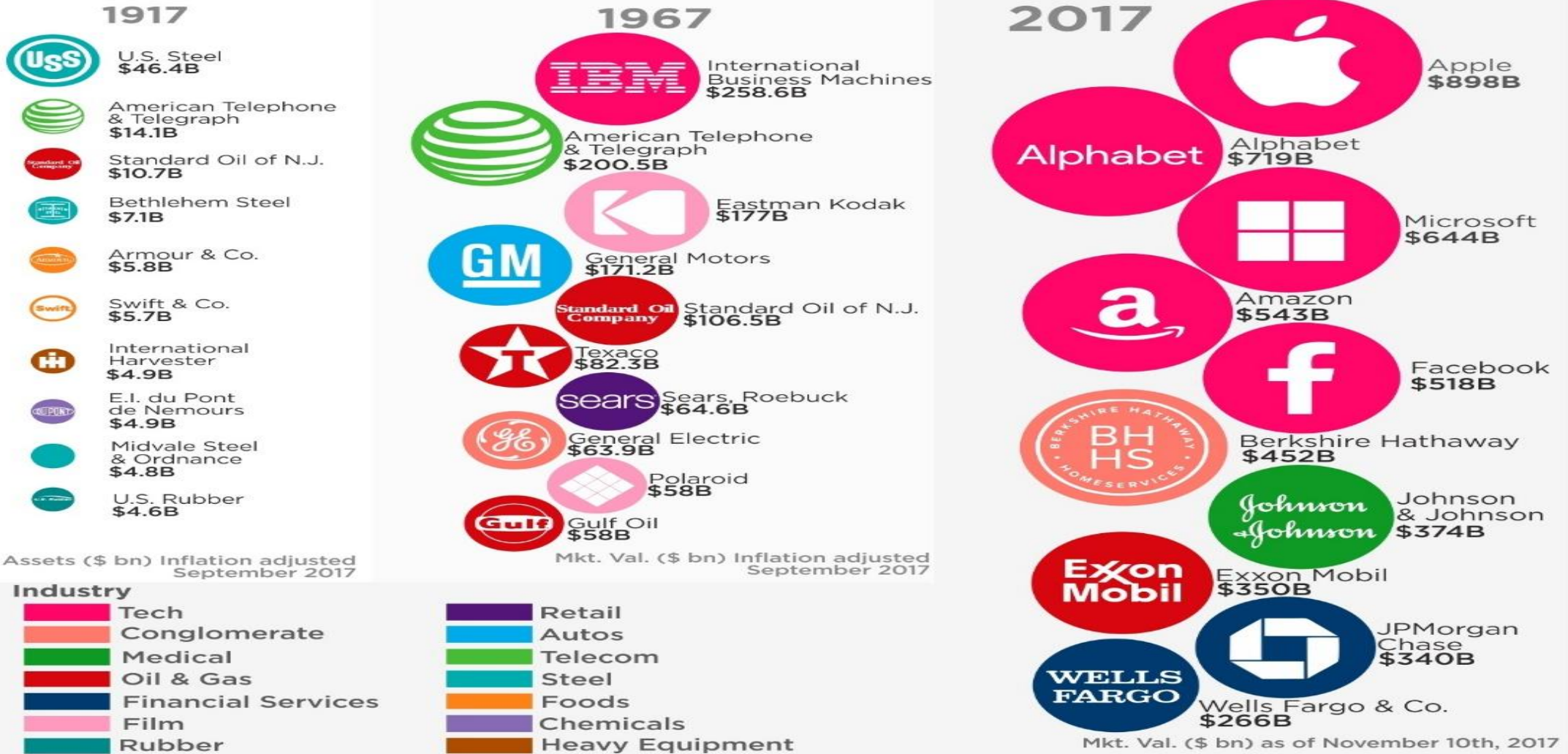
- A forma como trabalhamos, nos divertimos e aprendemos está a ser transformada pela coexistência de meios e **plataformas digitais** (do processo para o algoritmo)

- **Ordenado**

- Os dados assumem importância crescente e o ser humano convive no seu dia-a-dia cada vez mais com máquinas que realizam atividades físicas, mas também cognitivas, numa associação crescente do humano e artificial, desenvolvendo novas parcerias (do ter para o **ser**, incorporando e sendo incorporado por indivíduos e organizações)

# Cada vez mais digital

## 100 Years of America's Top 10 Companies



Source and Article:  
<https://howmuch.net/articles/100-years-of-americas-top-10-companies>  
<https://forbes.com>

# Algumas notas sobre esta (nova) fonte de poder (digital)

- 50% das grandes marcas atuais são da área da tecnologia (economia de base digital)
- a diversidade dos setores representados foi baixando: de 7, em 1917 e 1967 para 5, em 2017
- O setor mais representativo: em 1917 (30% **aço**, 1º, 4º e 9º) para 1967 (30% **petróleo**, 5º, 6º e 10º) e em 2017 (50% **tecnologia**, com os cinco primeiros lugares)
- O **digital** representa uma concentração de poder nunca antes vista, em influência global, número de utilizadores e complexidade do seu ecossistema, fugindo à regulação e ao poder dos Estados

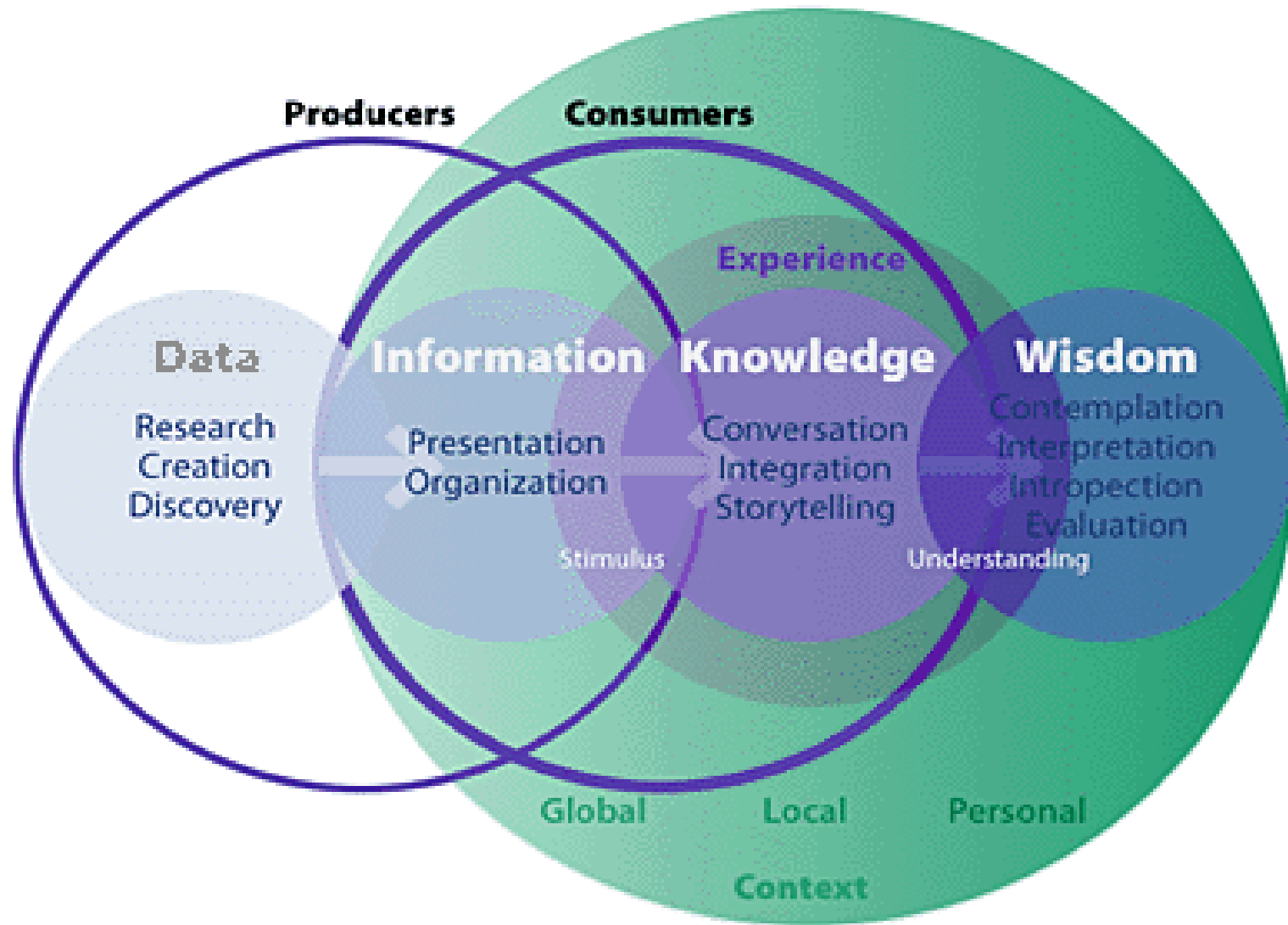
*Em conjunto, estas  
mudanças impactam  
profundamente a  
memória*

*Uma das tecnologias da  
memória é o arquivo*

# Digital, Dados e Informação

Manter o equilíbrio entre o digital e a memória

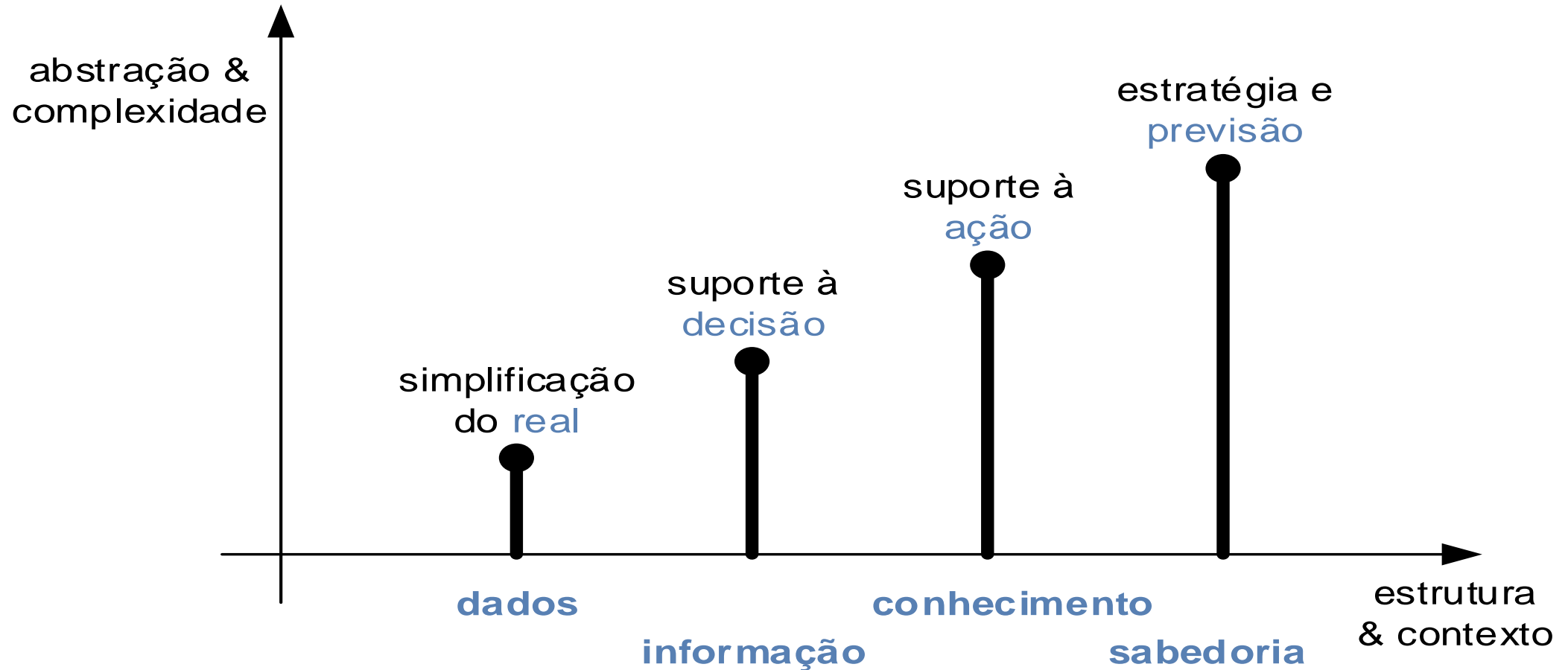
- Os **dados** são o produto da pesquisa, criação e descoberta
- É o material que se obtém ou é criado para suporte à **mediação**
- Muito do que é experimentado não é informação, mas sim dados, logo o seu valor é menor, pois não constitui **comunicação**
- A maior parte das **TIC** é para lidar com o armazenamento, o processamento e a transmissão de dados e não de informação
- Os dados são apenas úteis para os produtores, que são aqueles com capacidade de impor estruturas com significado aos dados – **metadados**
- Quando é criada uma **experiência**, deve ser evitada a partilha de muitos dados, sem fornecer um **contexto**



Information Anxiety, Richard Saul Wurman

# A informação já não é o que era...

(e nem toda é igual)

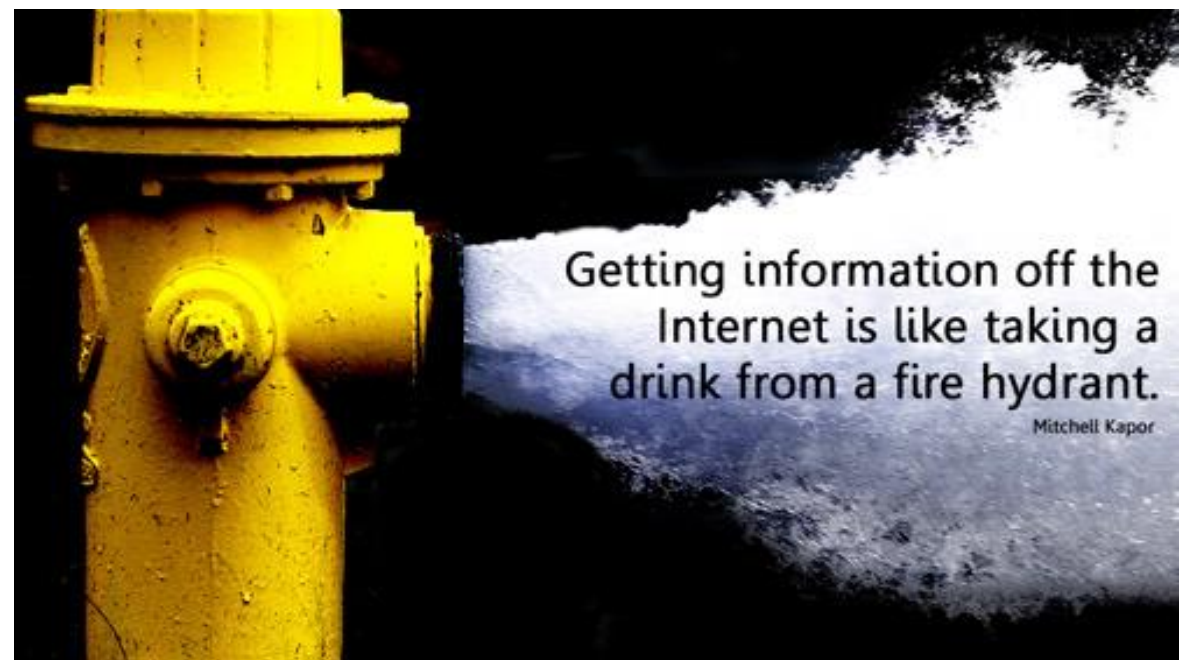


# No digital

- Crescente **mediação** de computadores e redes no relacionamento humano
- **Desmaterialização** de actividades e processos associados
- **Transformação** da actividade humana
  - O tempo com diferentes ciclos
  - Virtualização e transformação do conceito de tempo e espaço (exemplo: o sítio na Web...)
  - (implica) espaço (físico) com diferentes significados

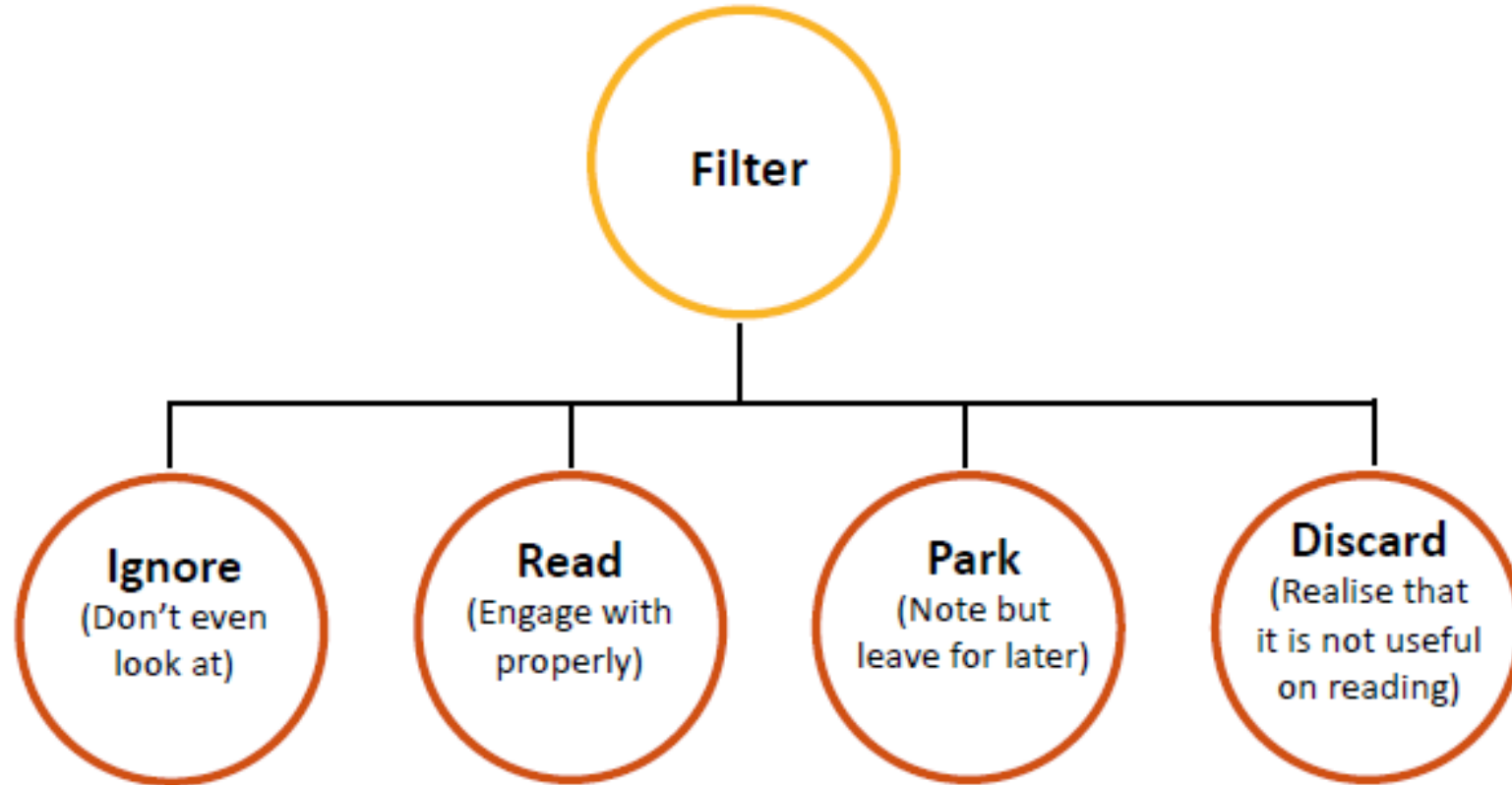
# E o digital?

- Amplificou a **quantidade** de informação
- Tornou mais **complexo** o seu **uso**
- Força a **alteração** de hábitos
- Força a perda da **escala humana**
- Força a **perda do controle** dos fluxos
- Aumenta a **dependência** de terceiros



<http://www.examiner.com/article/warning-information-overload>

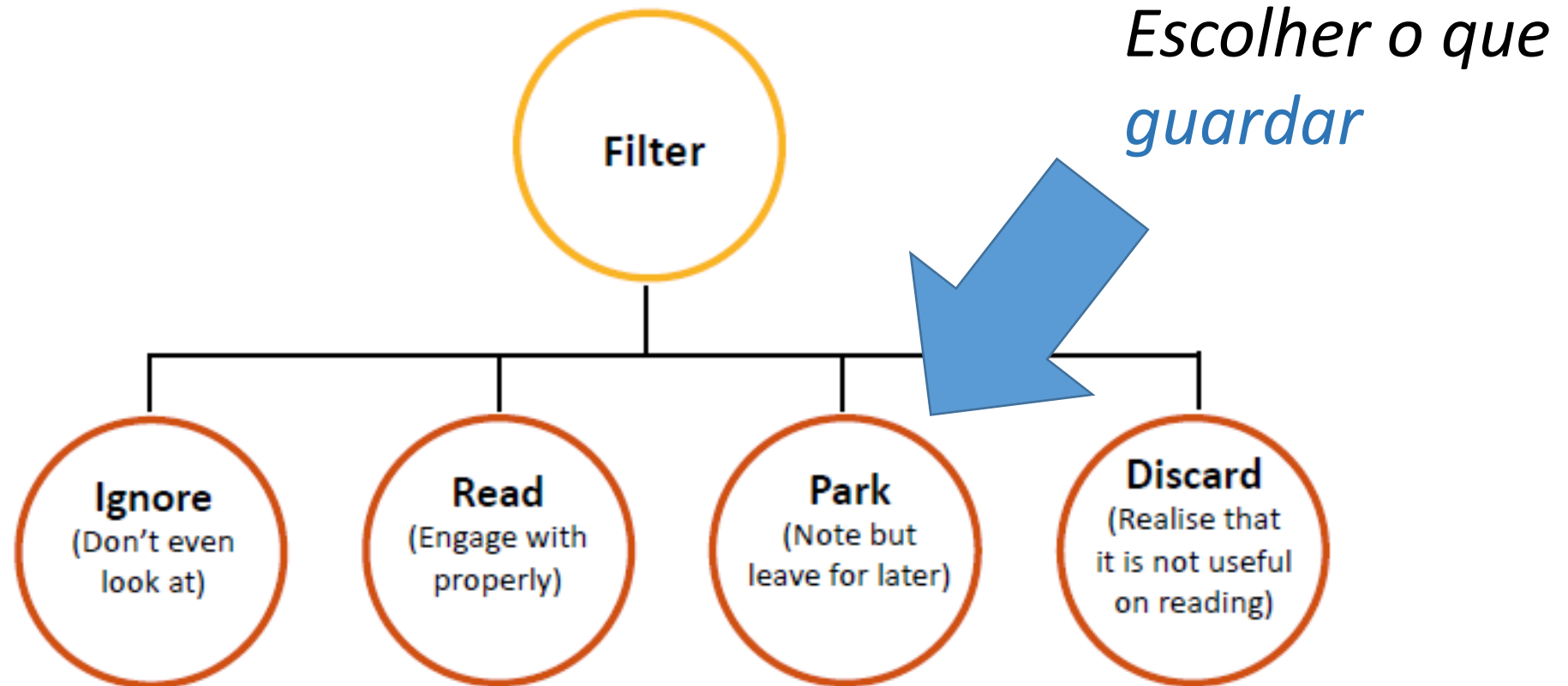
# Repensar a nossa relação com a informação (de recolectores para **curadores**)



<http://srobalino.wordpress.com/2011/07/04/supporting-researchers-with-social-tools/>

# Nem tudo é informação...

(de recolectores para **curadores**)



<http://srobalino.wordpress.com/2011/07/04/supporting-researchers-with-social-tools/>

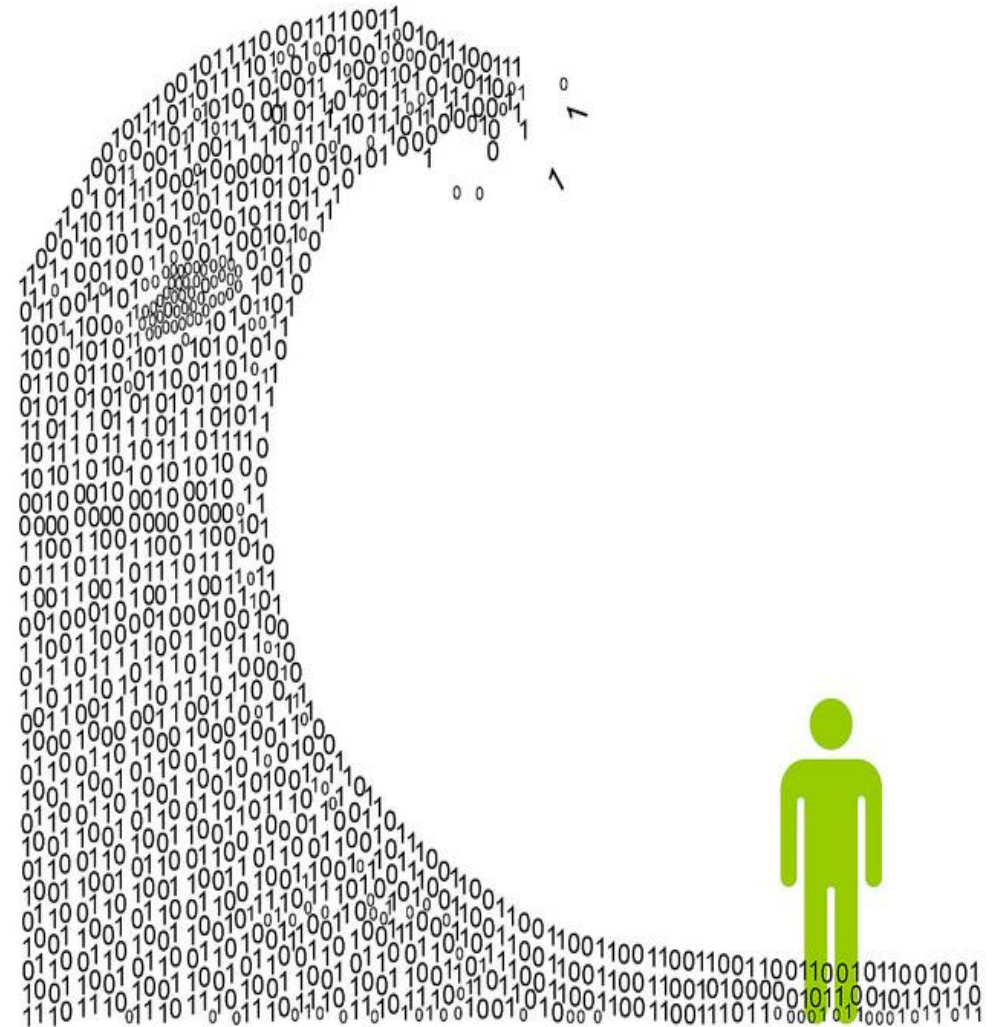
# No contexto profissional (muito papel/digital)



<http://athandz.blogspot.pt/2012/08/information-overload-and-mobile.html>

# A esperança do digital (mudar processos e tornar esforço colaborativo)

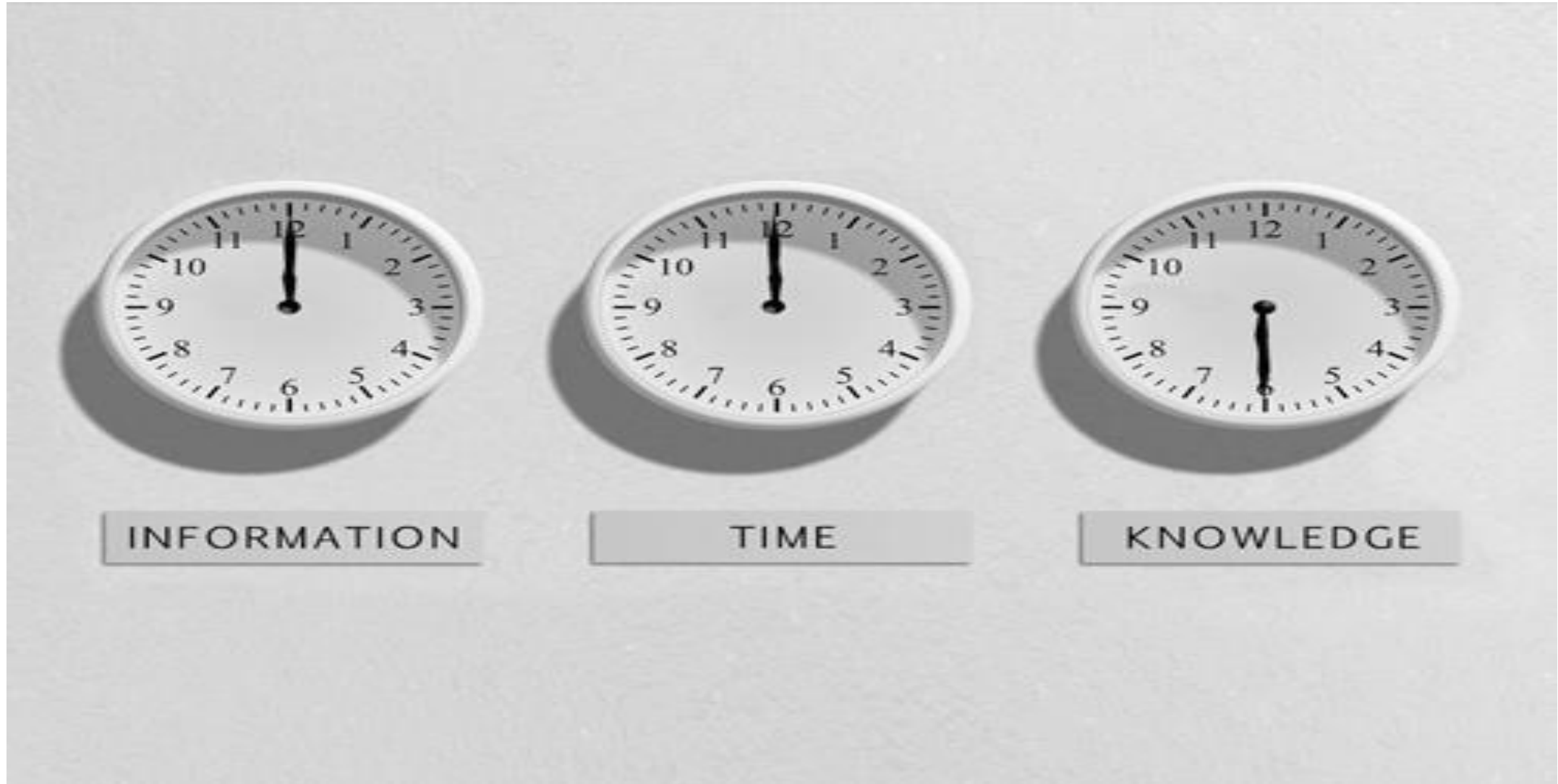
- **Integrar e sincronizar** informação de múltiplas origens
- Cruzar e confirmar informação de forma mais **automática**
- Interagir com outros **minimizando** informação e tornando esta mais fácil de entender
- **Permitir e praticar o descarte** de informação
- **Reutilizar e simplificar** (em especial recorrendo a estruturas existentes e evitando novas ou alterações destas)
- **Filtrar e colaborar** com terceiros para facilitar a obtenção, confirmação e entendimento da informação



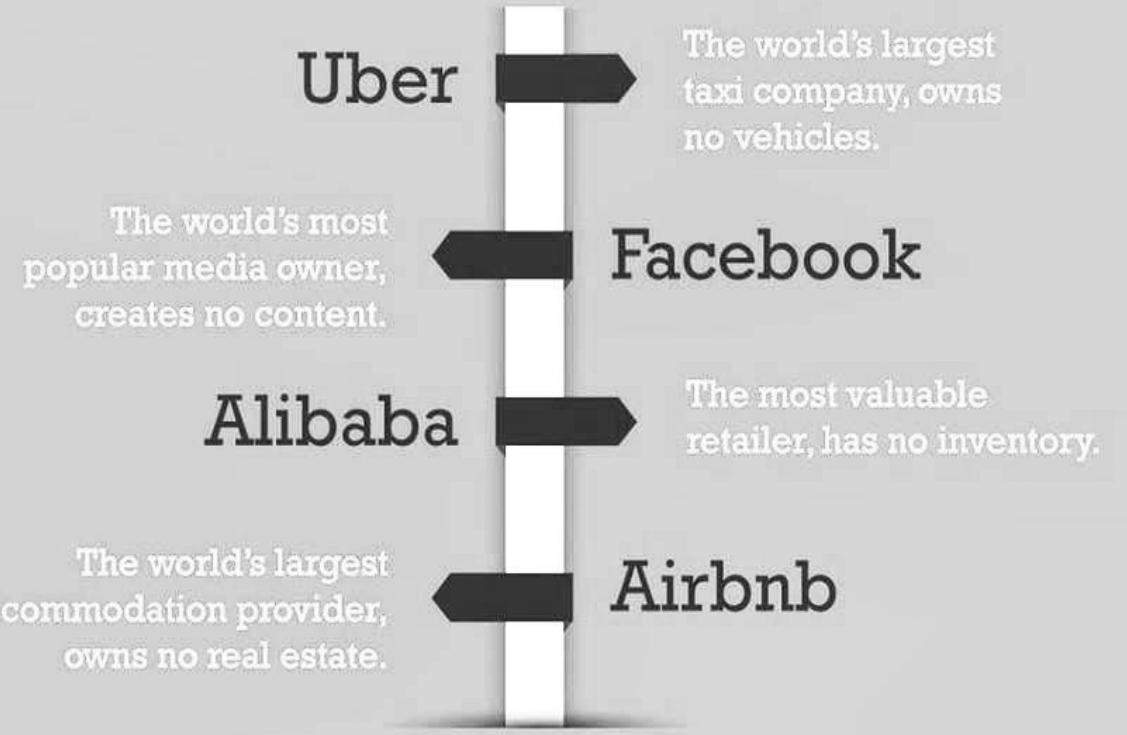
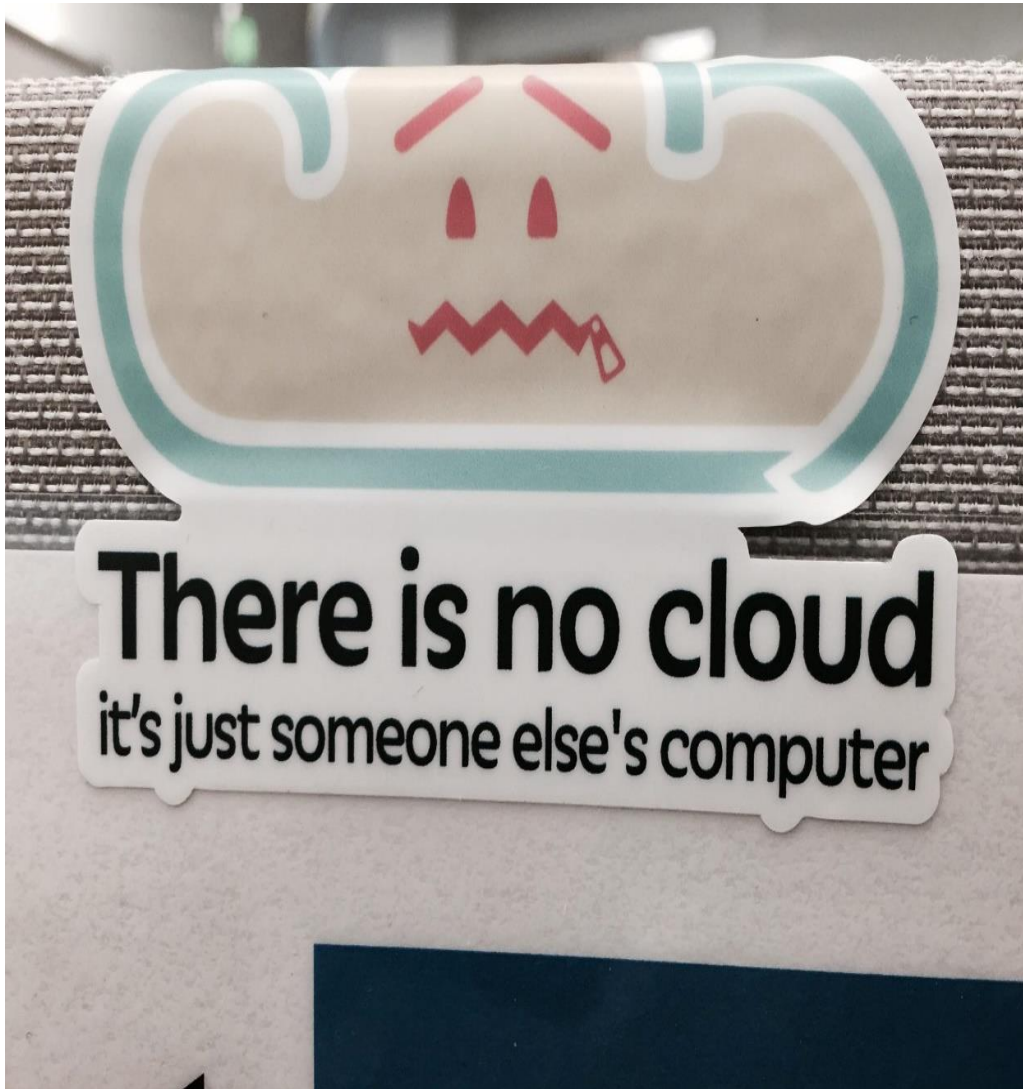
# Da tecnologia ao arquivo

O potencial da IA para o arquivo

O nosso tempo (há já alguns anos...)



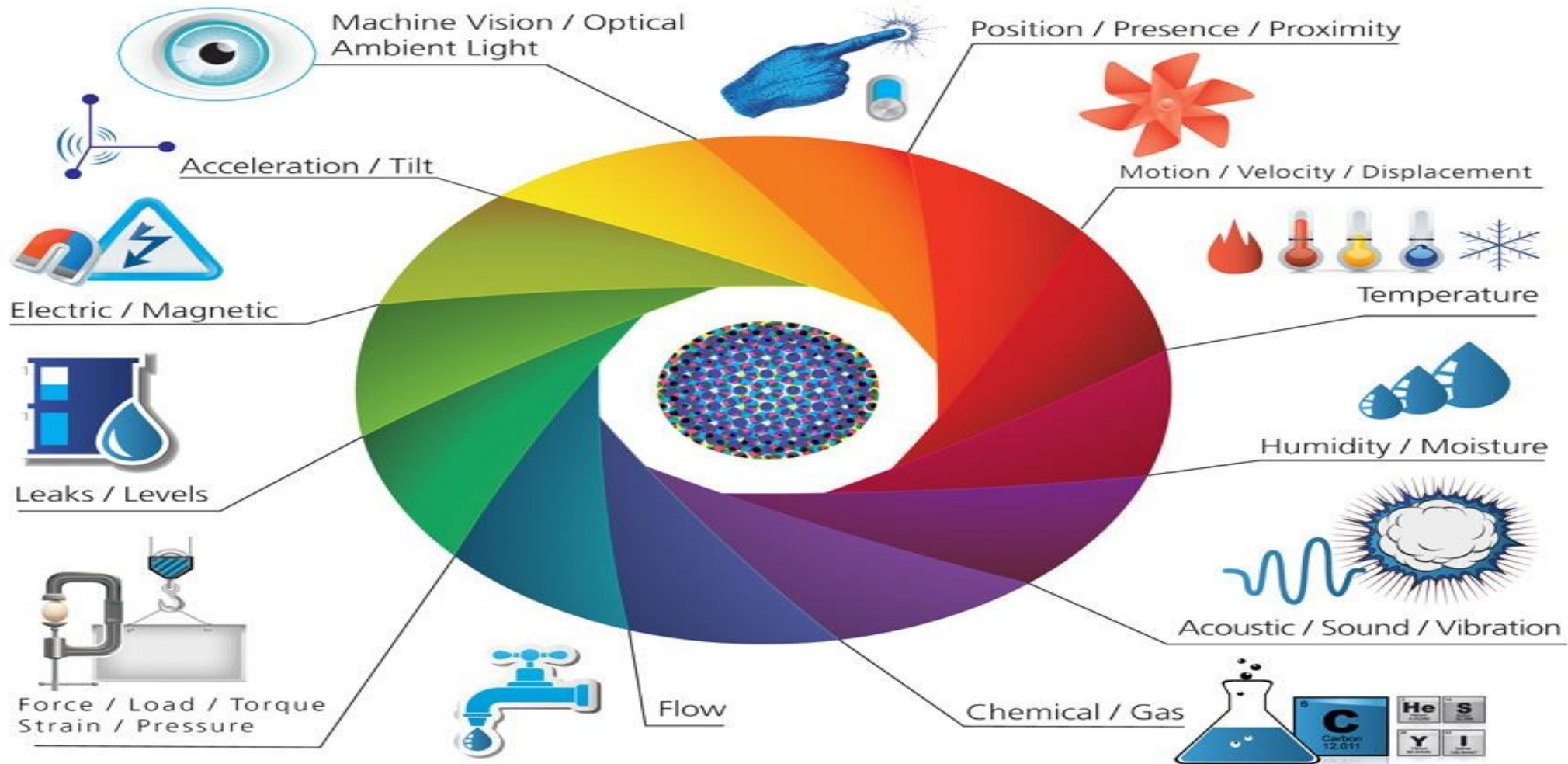
# A realidade *cloud*



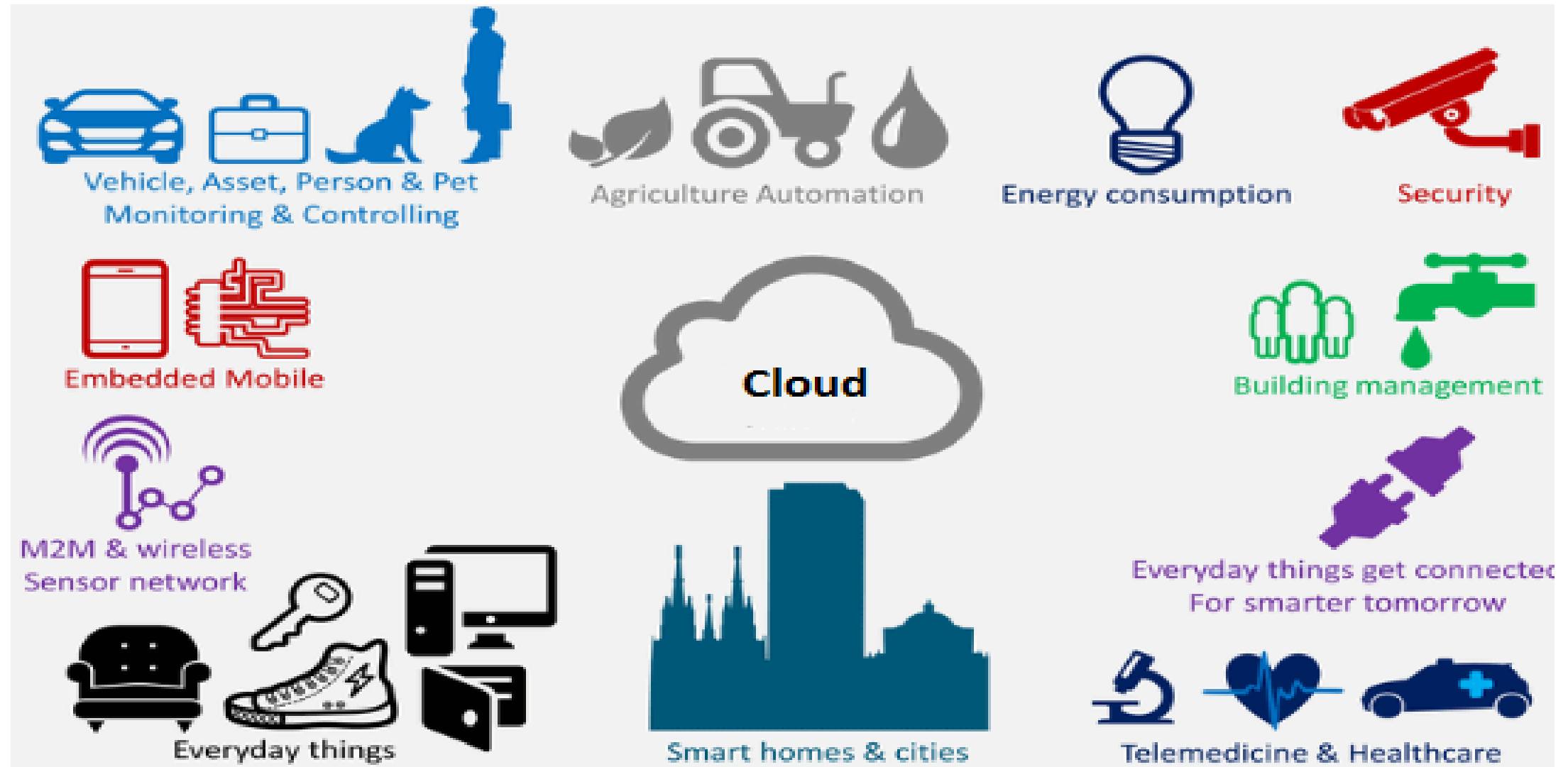
Something interesting is happening.

TOM GOODWIN

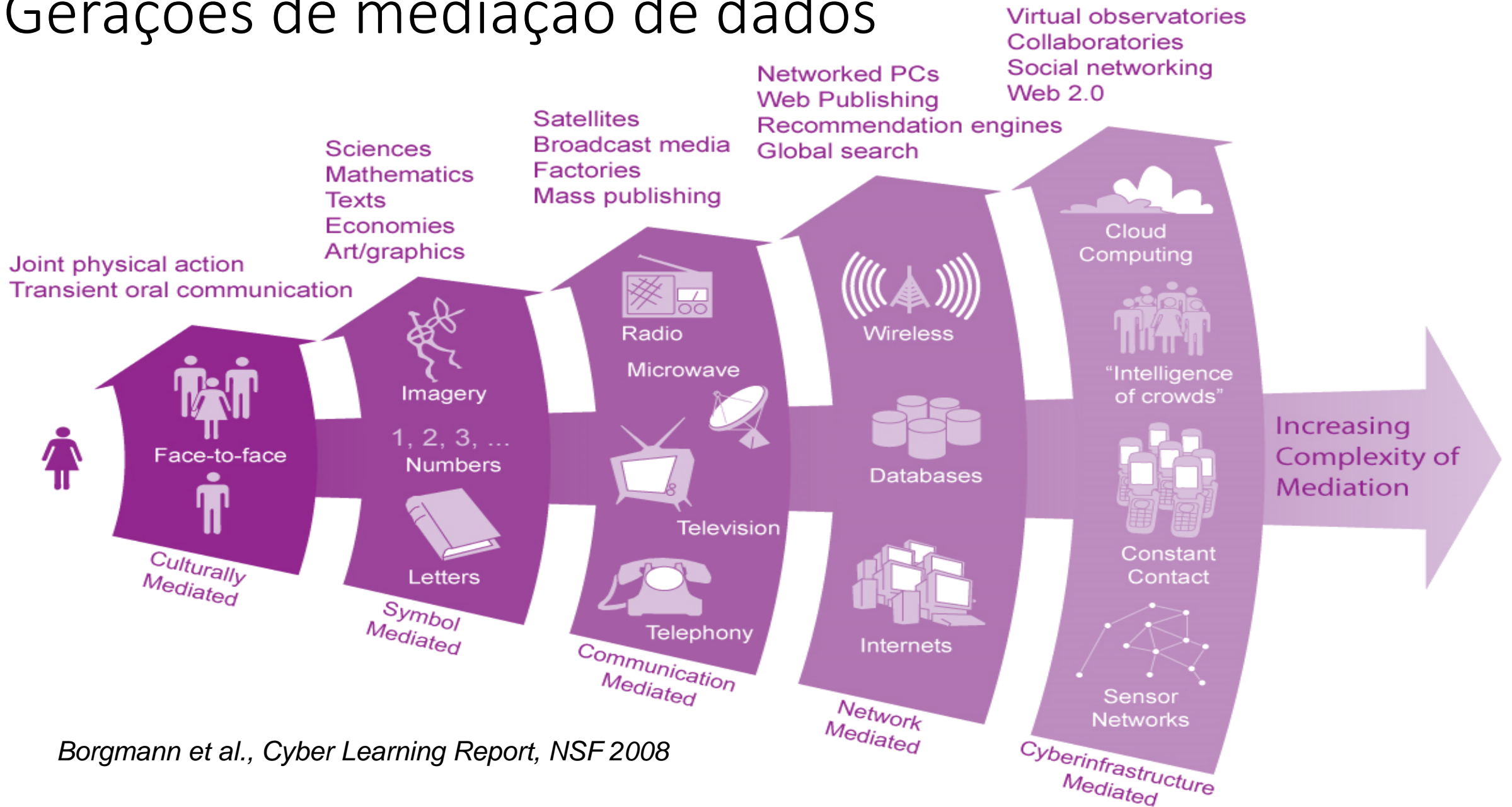
# A ubiquidade dos sensores e atuadores



# IoT (Internet das Coisas)



# Gerações de mediação de dados



Borgmann et al., *Cyber Learning Report*, NSF 2008

# A dimensão do *Big Data*

## Every 60 seconds



**98,000+** tweets



**695,000** status updates



**11 million** instant messages



**698,445** Google searches



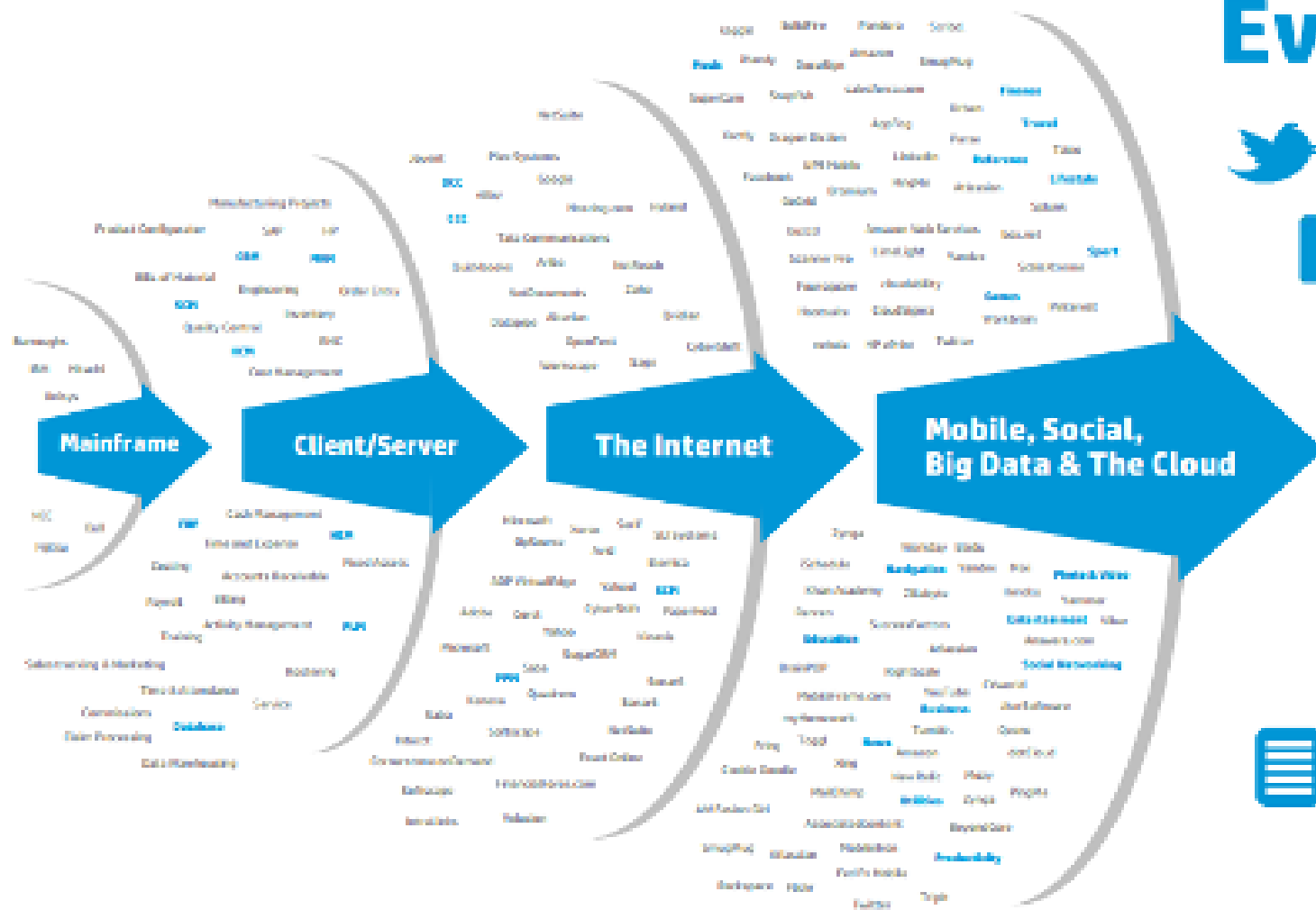
**168 million+** emails sent



**1,820TB** of data created



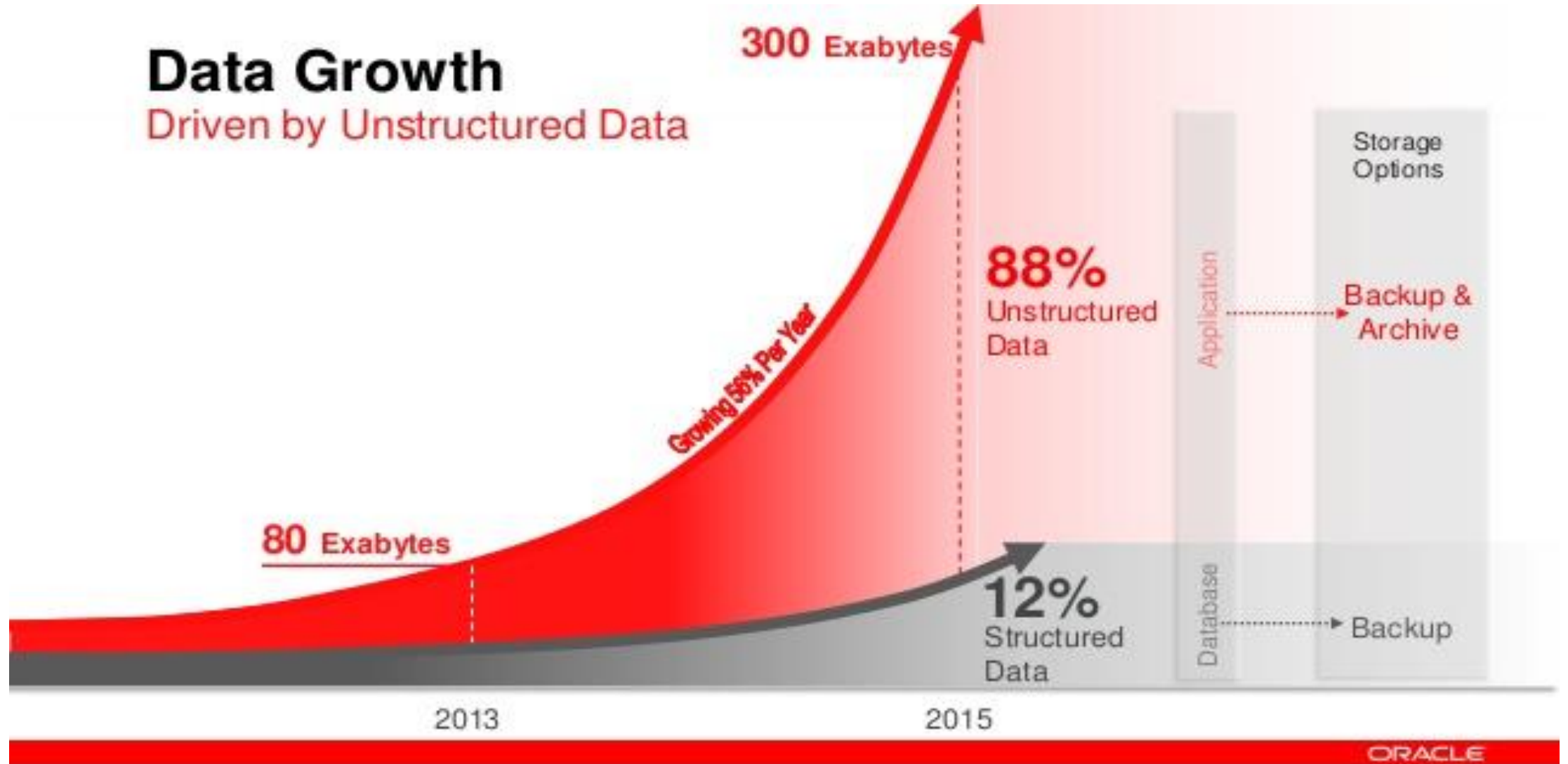
**217** new mobile web users



# Importância do *Big Data*

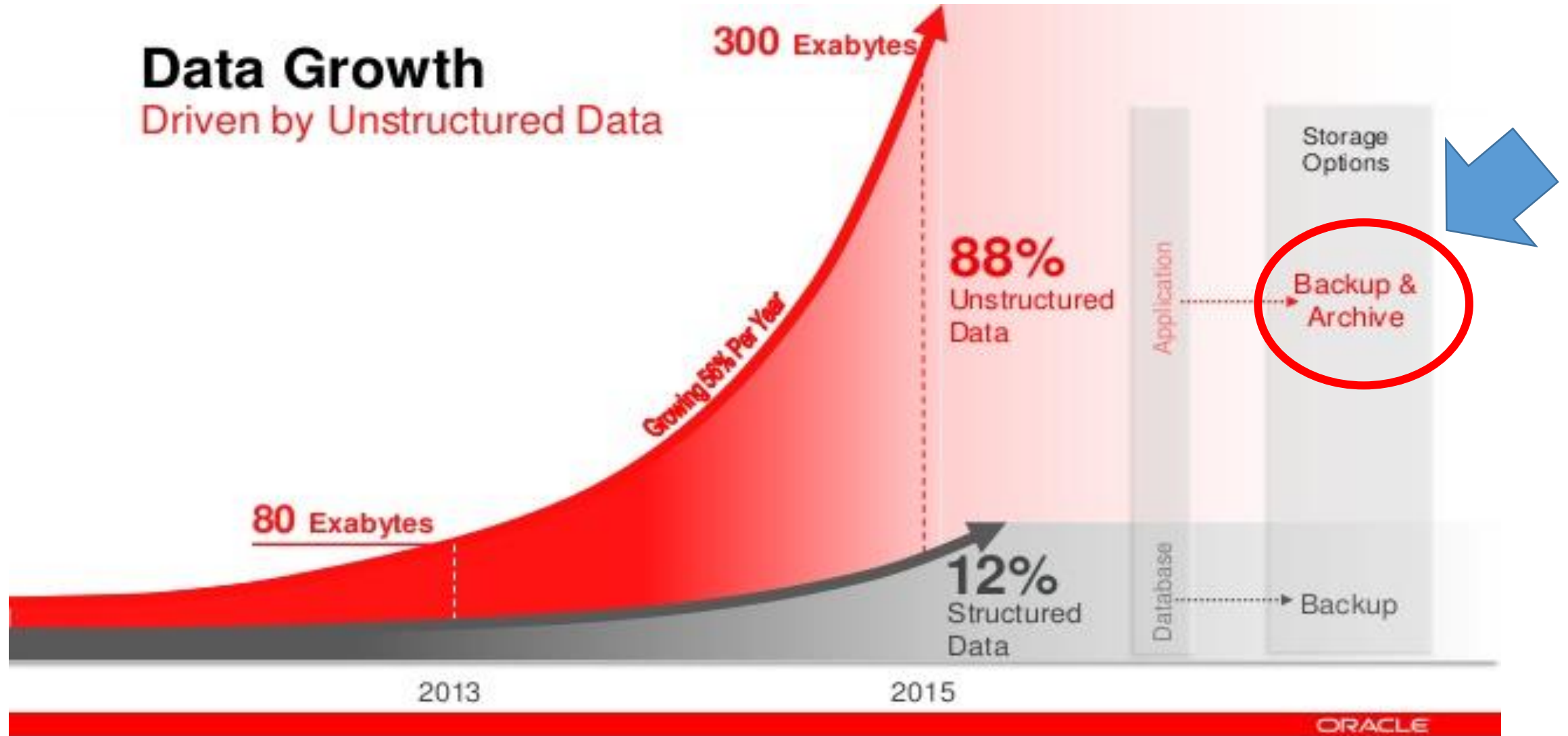
- Os dados podem ser analisados no seu **formato e organização nativa**, seja ela estruturada, não estruturada ou *streaming* (fluxo constante de dados)
- Dados podem ser captados em **tempo real**
- Os dados podem ser **transformados em conhecimento**
- ***Datafication***
  - Tudo se pode transformar em dados que podem ser analisados
  - **O utilizador como produto** (e produtor de dados)
- Os dados adquirem um potencial enorme de realizar **valor**
  - Os dados como o **novo capital**

# Crescimento exponencial



\* Source: ESG Digital Archive Market Forecast

# O papel do arquivo



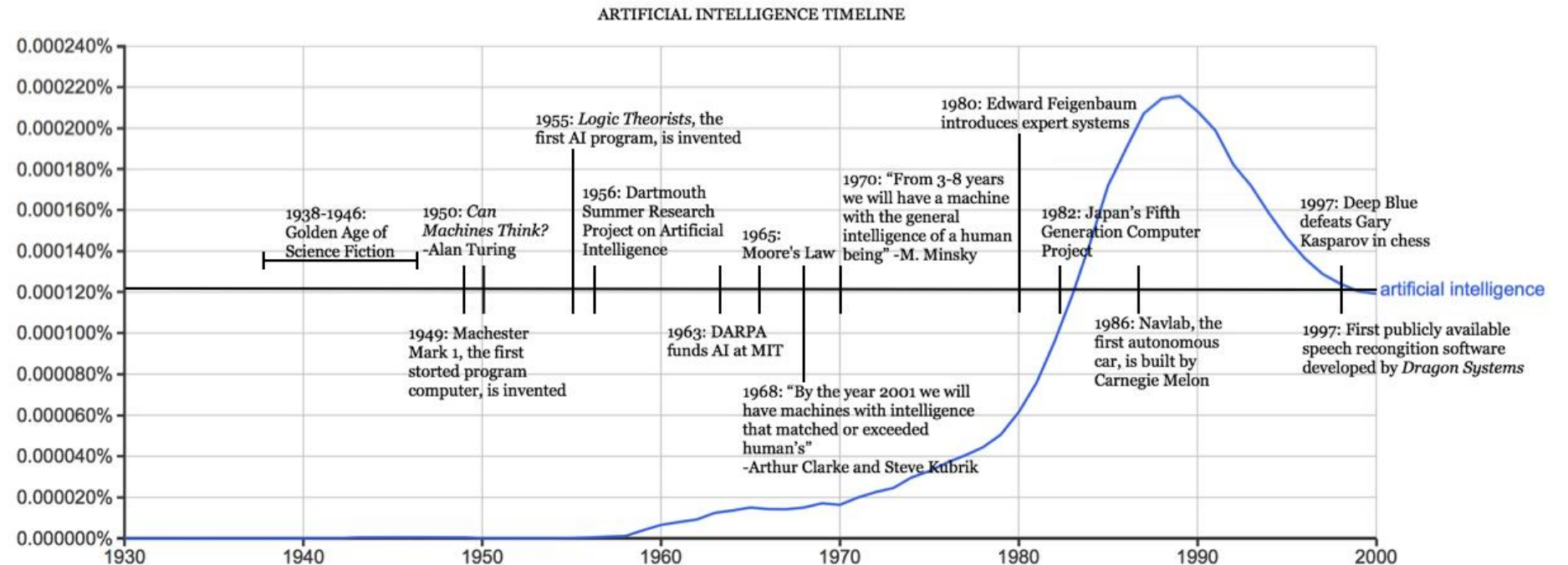
\* Source: ESG Digital Archive Market Forecast

# Inteligência

- pode ser definida de muitas formas: **capacidade** para lidar com lógica, entendimento, auto consciência, aprendizagem, conhecimento emocional, raciocínio, planejamento, criatividade e resolução de problemas
- capacidade de **análise de informação** e adaptação, percebendo a informação disponível de forma a formar conhecimento que proporcione ação, de modo a permitir comportamentos adaptativos ao ambiente circundante ou a um contexto



**Alan Turing**  
(1912, 1954)



**A brief History of Artificial Intelligence**

<https://www.youtube.com/watch?v=056v4OxKwII>

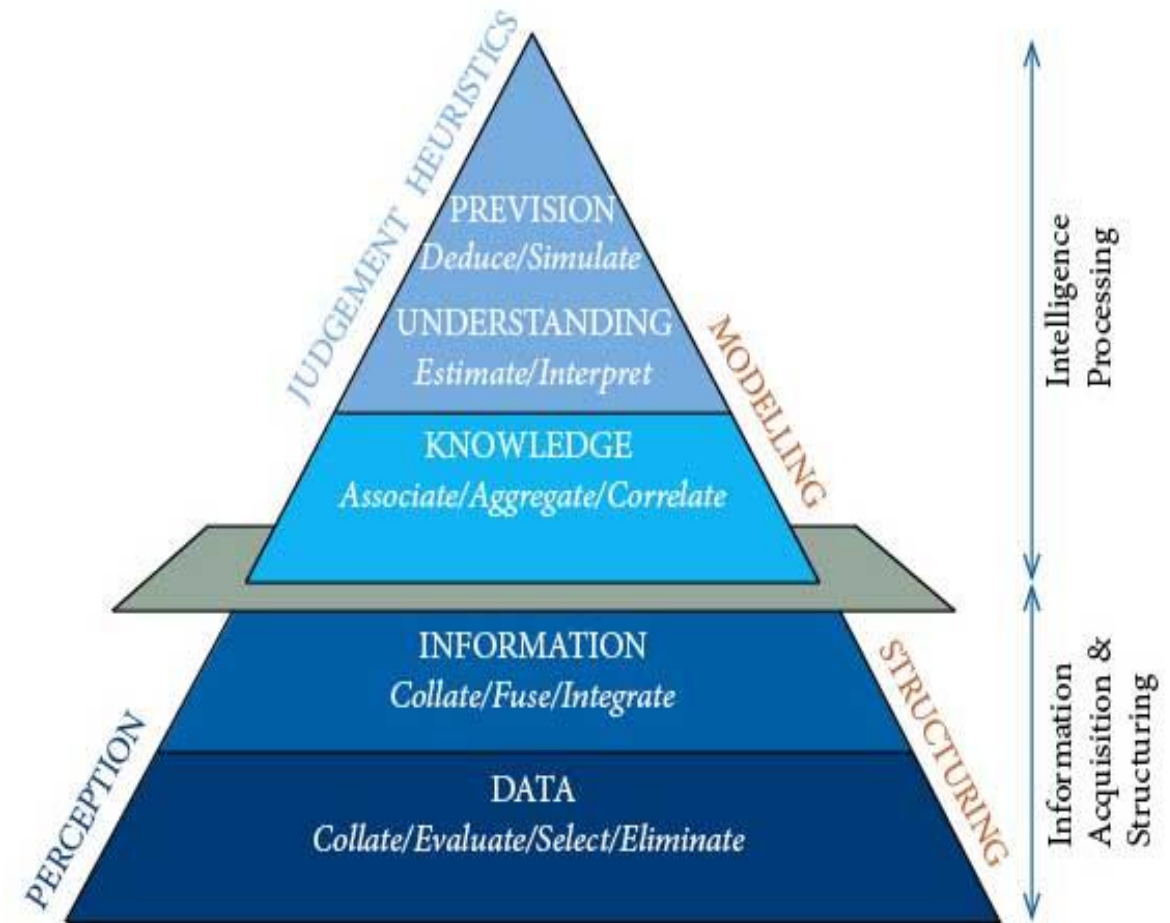
**Can machine think?**

<http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>

# Inteligência Artificial (IA)

## *Artificial Intelligence (AI)*

- Stuart J. Russell e Peter Norvig definem inteligência artificial como um *conjunto de teorias e técnicas usadas para criar máquinas capazes de simular a inteligência humana*
- Definição mais abrangente: área da ciência dos computadores que estuda a *criação de máquinas inteligentes que trabalham e reagem como os seres humanos, aprendendo, planejando, classificando, resolvendo problemas e reconhecendo dados, informação e conhecimento com o objetivo de criação de aplicações autónomas ou de suporte à atividade humana*



# Olá Futuro

*Como ser Humano na era dos algoritmos*

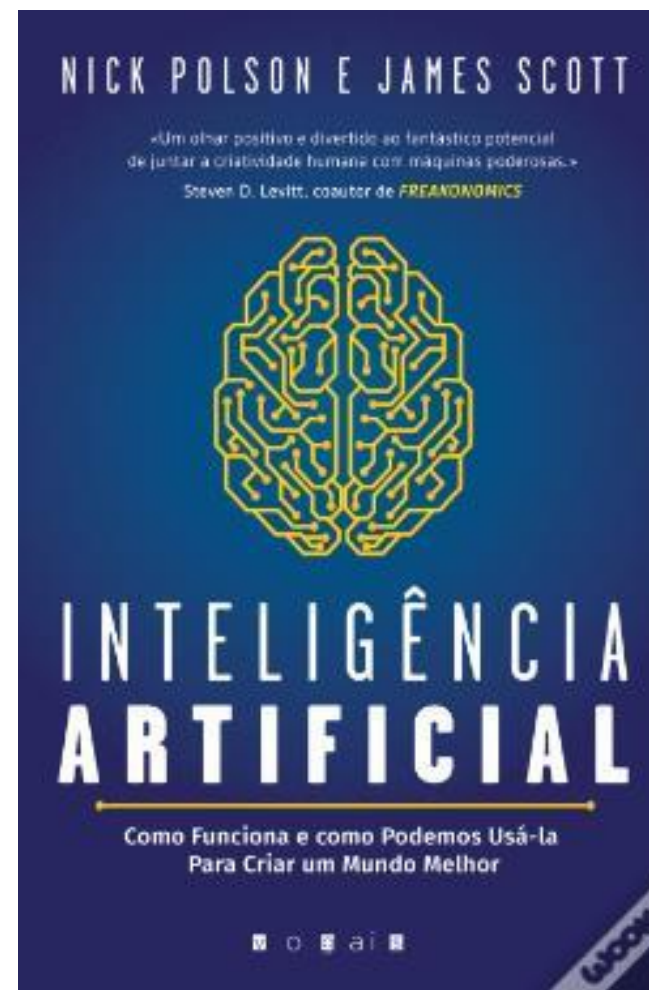
Hannah Fry



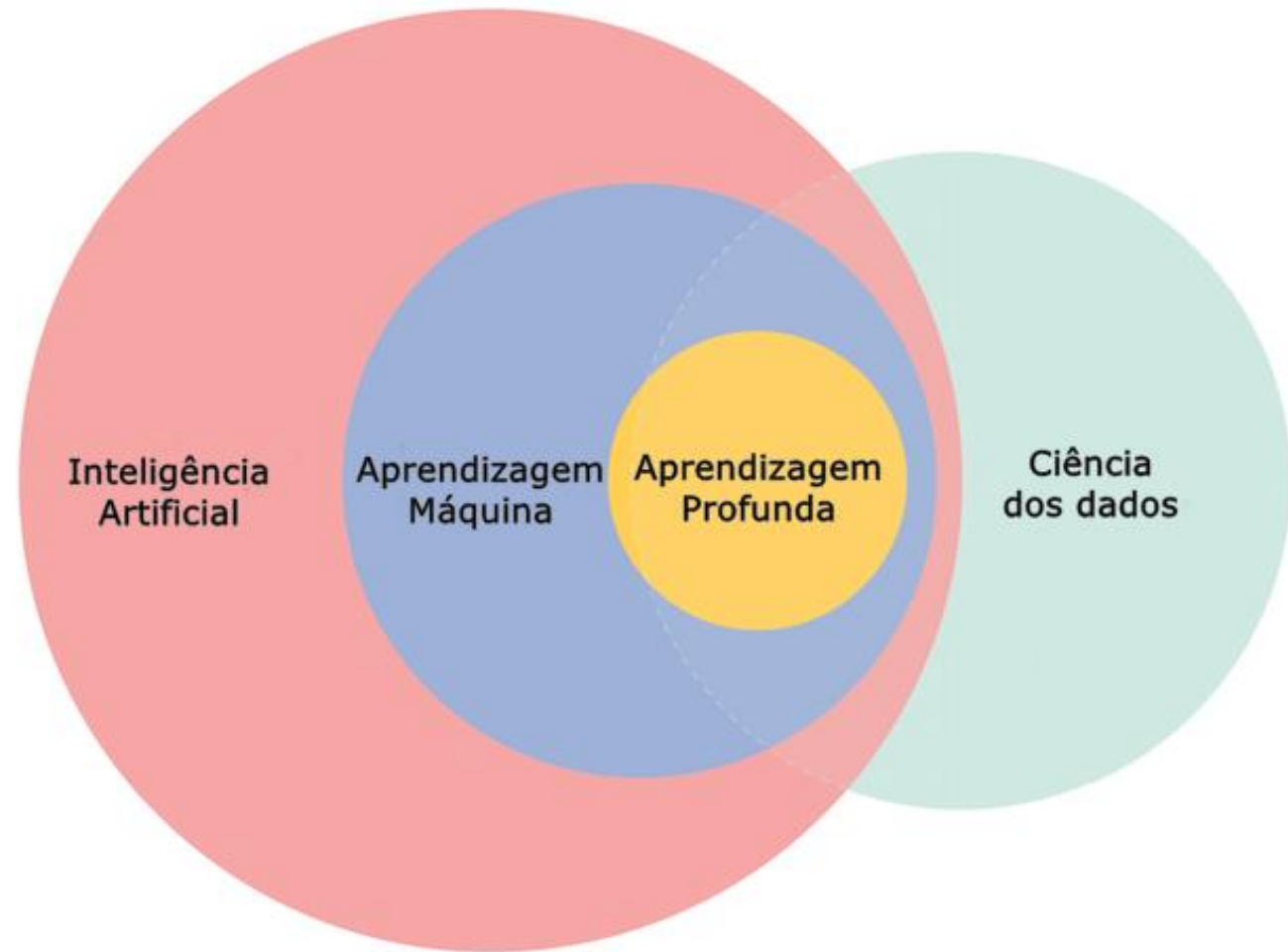
# Inteligência Artificial

*Como funciona e como podemos usá-la para criar um mundo melhor*

Nick Polson e James Scott



- **Machine learning** (DL): a utilização e desenvolvimento de sistemas informáticos capazes de aprender e adaptar-se sem seguir instruções explícitas, utilizando algoritmos e modelos estatísticos para analisar e extrair inferências a partir de padrões nos dados
- **Deep learning** (DL): aprendizagem profunda; um tipo de aprendizagem máquina baseada em redes neuronais em que são utilizadas múltiplas camadas de processamento para extrair progressivamente características de nível superior dos dados
- **Diferença entre ML e DL:** a aprendizagem máquina utiliza algoritmos para analisar os dados, aprender com esses dados, e tomar decisões informadas com base no que aprendeu. A aprendizagem profunda organiza algoritmos em camadas para criar uma rede neural artificial que pode aprender e tomar decisões inteligentes por si só, de modo autónomo
- **Data science** (DS): a ciência dos dados é o campo de estudo que combina conhecimento de um domínio específico, com conhecimentos de programação, de matemática e estatística para extrair percepções significativas dos dados, integrando também técnicas associadas com ML e DL



## Archives and AI

An Overview of Current Debates and Future Perspectives

Giovanni Colavizza

University of Amsterdam, Amsterdam, The Netherlands, [g.colavizza@uva.nl](mailto:g.colavizza@uva.nl)

Tobias Blanke

University of Amsterdam, Amsterdam, The Netherlands, [t.blanke@uva.nl](mailto:t.blanke@uva.nl)

Charles Jeurgens

University of Amsterdam, Amsterdam, The Netherlands, [K.J.P.F.M.Jeurgens@uva.nl](mailto:K.J.P.F.M.Jeurgens@uva.nl)

Julia Noordegraaf

University of Amsterdam, Amsterdam, The Netherlands, [J.J.Noordegraaf@uva.nl](mailto:J.J.Noordegraaf@uva.nl)

### ABSTRACT

The digital transformation is turning archives, both old and new, into data. As a consequence, automation in the form of artificial intelligence techniques is increasingly applied both to scale traditional recordkeeping activities, and to experiment with novel ways to capture, organise and access records. We survey recent developments at the intersection of Artificial Intelligence and archival thinking and practice. Our overview of this growing body of literature is organised through the lenses of the Records Continuum model. We find four broad themes in the literature on archives and artificial intelligence: theoretical and professional considerations, the automation of recordkeeping processes, organising and accessing archives, and novel forms of digital archives. We conclude by underlining emerging trends and directions for future work, which include the application of recordkeeping principles to the very data and processes which power modern artificial intelligence, and a more structural – yet critically-aware – integration of artificial intelligence into archival systems and practice.

### KEYWORDS

Archives, Recordkeeping, Machine Learning, Artificial Intelligence

<https://arxiv.org/ftp/arxiv/papers/2105/2105.01117.pdf>

LISE JAILLANT (ED.)

# ARCHIVES, ACCESS AND ARTIFICIAL INTELLIGENCE

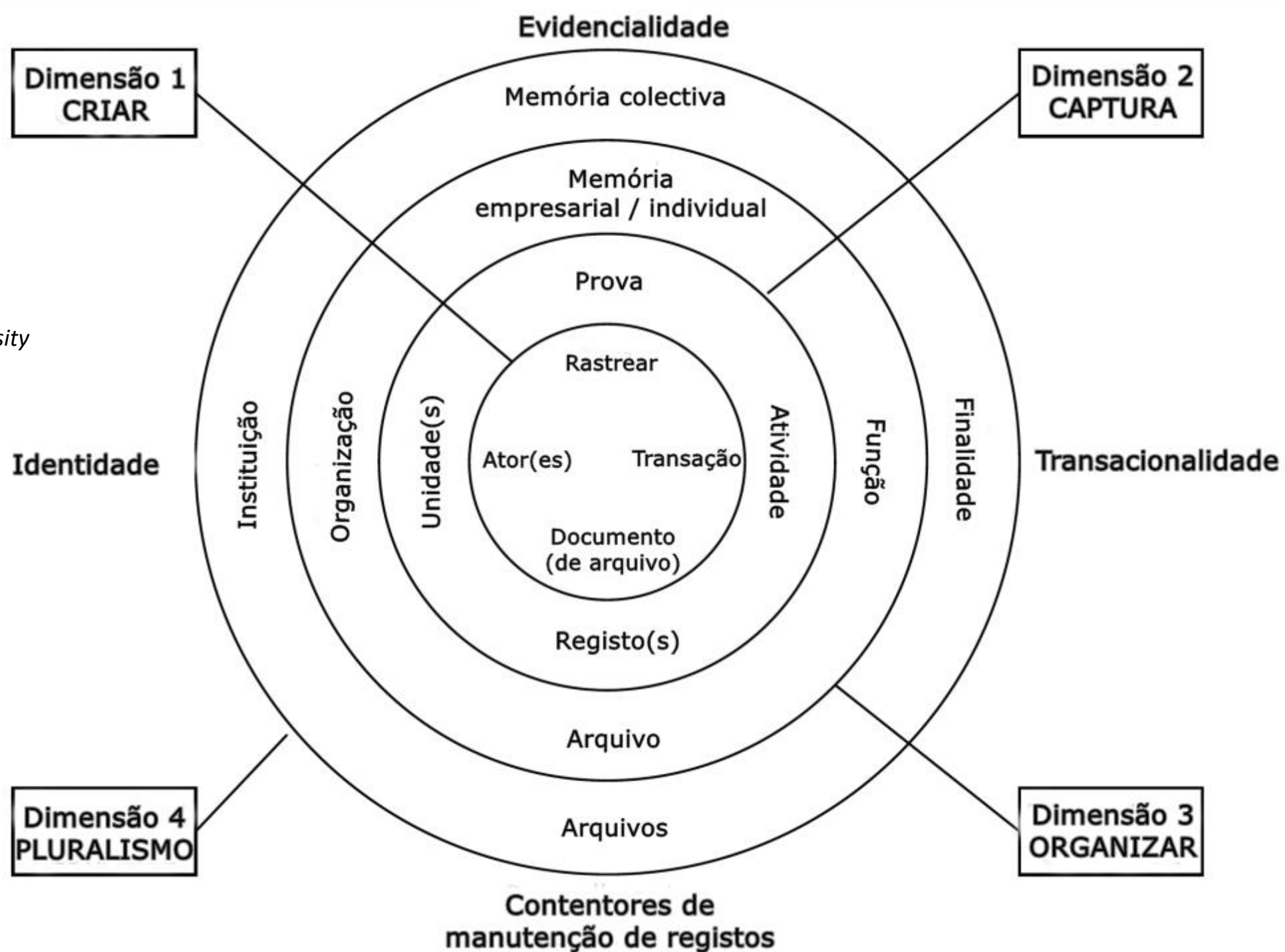
WORKING WITH BORN-DIGITAL AND  
DIGITISED ARCHIVAL COLLECTIONS

DIGITAL HUMANITIES RESEARCH  
BIELEFELD UNIVERSITY PRESS

<https://www.amazon.com.br/Archives-Access-Artificial-Intelligence-Born-Digital/dp/3837655849>

# The Records Continuum Model

Upward, F.; Reed, B.; Oliver, G. & Evans, J. (2018). *Recordkeeping Informatics for a Networked Age*. Clayton, Victoria: Monash University Publishing.



# Algumas notas finais (I)

- **Considerações teóricas e profissionais:** os métodos tradicionais de avaliação tornaram-se obsoletos. Não só são necessárias novas ferramentas de mediação homem máquina, como também novos métodos na utilização e análise de arquivos como o uso e exploração de (ciência dos) dados. Requer novos grupos profissionais, novas competências e conhecimento
- **Automatizar processos e decisões de manutenção de registos:** o potencial de tecnologias de IA promete automatizar partes do fluxo de trabalho dos arquivos. As expectativas relacionadas com este potencial estão predominantemente situadas na Captura e Diversidade dos conteúdos associados. A curadoria manual, de alta qualidade e colaborativa de dados continua a ser uma abordagem bem sucedida em muitas situações
  - **Avaliação:** os profissionais de manutenção de registos são confrontados com volumes cada vez maiores de registos não estruturados e não categorizados. A IA tem potencial para ajudar a gerir volumes tão grandes e a literatura reporta inúmeras experiências e casos de sucesso
  - **Tratamento de informação sensível:** as instituições devem evitar que a incapacidade de identificar informação sensível torne virtualmente impossível conceder liberdade de acesso e pedidos de informação ou de permitir a investigação arquivística
  - **Metadados:** questões de descrição e extração de metadados para auxiliar os criadores de registos do fornecimento de metadados (classificação), aumentando a qualidade dos metadados existentes (catálogos)

# Algumas notas finais (II)

- **Organização e acesso a arquivos:** A quantidade crescente de materiais de arquivo digitalizados e de base digital exige a automatização e mesmo a reinvenção da organização e do acesso aos registos. A abordagem tradicional pode agora ser complementada por outros meios, centrados principalmente no conteúdo dos registos. Esta mudança não só permite aos utilizadores dar sentido a materiais de arquivo em escalas e complexidade maiores, como o fazer de formas inovadoras
  - Extração e indexação automática de conteúdos
  - Registos de arquivo para leitura/aquisição de modo remoto
  - Pesquisa e recuperação
- **Novas formas de arquivos digitais:** facilitar novos tipos de recolha de dados, como nos sistemas integrados, sistemas inteligentes, com recurso à Internet das Coisas (IoT). A informação resultante é capturada e organizada, de acordo com a conformidade a leis, regulamentos e expectativas relativamente à privacidade e outros valores públicos. Os princípios de manutenção de registos geralmente aceites para arquivo (responsabilidade, transparência, integridade, proteção, conformidade, disponibilidade, retenção e disposição) fornecem um referencial para a organização destas novas formas de arquivo digital

# Luis Borges Gouveia

Professor Catedrático  
Universidade Fernando Pessoa.

Coordenador do Grupo de I&D Informação,  
Comunicação e Cultura Digital, CITCEM  
Universidade do Porto.

Possui a Agregação em Engenharia e Gestão  
Industrial (Universidade de Aveiro),  
Doutoramento em Ciências da Computação  
(Universidade de Lancaster, UK) e é Mestre em  
Engenharia Eletrônica e de Computadores pela  
Universidade do Porto.

Os seus interesses estão relacionados com o  
digital e como as tecnologias podem melhorar  
a vida das pessoas.

Página pessoal: <http://homepage.ufp.pt/lmbg/>

