

MEGSI (SIOS e ESI), MSI, MiEGSI, Erasmus, ...

Apresentação e Visualização de Dados (Inf.)

Jorge Oliveira e Sá

jos@dsi.uminho.pt

2022/23

Dados

- ‘[...] data are becoming increasingly valued and relied upon, as they come to play an ever more important role in decision-making and knowledge about the world’ Kennedy, Hill, Aiello & Allen (2016b, p. 715) <https://doi.org/10.1080/1369118X.2016.1153126>
- Os dados são gerados, recolhidos, armazenados, transformados e tornam-se mais acessível do que nunca. A recolha de dados ocorre em muitos domínios, muitas vezes por lei, incluindo comércio, educação, saúde, transporte e vida social.
- Esses dados oferecem insights sobre padrões sociais que de outra forma ficaria invisível e despercebido.

Dados

- Datafication (Datafy) – surgiu nas organizações. Através de dados extraem informação e conhecimento para executar processos críticos. Uma organização também usa dados para tomar decisões, definir estratégias e objetivos chave.
- O passo seguinte é tornar esses dados acessíveis a todos
- Iniciativas de Dados Abertos contribuíram para acelerar a disseminação de visualização de dados
- Outro acelerador é a tecnologia, pois surgem novas ferramentas e técnicas para recolher, filtrar, analisar e visualizar dados

Dados

- Em contexto científico, os dados são geralmente entendidos como resultado da geração, coleta, observação ou registro de objetos, eventos ou processos adequados para servir algum propósito analítico
- Em contexto de DataVis, dados podem ser qualquer coisa que possa ser sujeita a uma categorização, abstração e translação para uma representação gráfica. Por exemplo: pessoas, lugares, documentos, relações, salários, ...

Dados

- Os dados podem ser qualitativos ou quantitativos
- Qualitativos - são valorizados pela singularidade de cada unidade individual, seja um poema, uma frase ou uma entrevista
- Quantitativos - são valorizados por características compartilhadas por todas ou muitas unidades num *dataset*. São suas características compartilhadas que os tornam objetos para contar ou medir e, portanto, para representação numérica e processamento estatístico.

Tipos de dados (origens dos dados)

- Origens dos dados:
 - IoT
 - streaming data (os dados necessitam de ser tratados em tempo-real)
 - Cloud
 - armazenamento de dados na cloud pode ser uma solução para disponibilizar dados
 - Múltiplas fontes
 - existência de várias plataformas de software
 - essas fontes de dados estão espalhadas em diferentes lugares (distribuídas)
 - os donos dessas fontes de dados são distintos

Tipos de dados

- numéricos – só contém números (valor das vendas, percentagens de margens, ...)
- textos – alfanumérico (nomes, matrícula de viaturas, ...)
- compostos – combinação de vários dados (JSON, XML, YAML, RDF, RSS, OWL, ...)
- documentos – texto não estruturado (words, PDFs, simples txts, ...)
- imagens – ferramenta de diagnóstico na saúde, reconhecimento facial, ...
- grafos – dados estão relacionados (semântica – ligações, entre eles, por exemplo artigos científicos – autores)
- áudios – sinais que podem ser recolhidos (ruído, batimentos cardíacos, ...)
- conversas – registos de voz
- vídeos – registos audiovisuais (procedimentos de produção, instruções operacionais, ...)

Tipos de dados (desafios)

- Cada vez mais a necessidade de utilizar volumosos dados
 - isto continua em crescimento
- Fontes (origens) variadas
- Privacidade dos dados
- Os dados são mais do que números e alfanuméricos e cada vez mais é necessário visualizar a variedade de dados existentes

Tipos de dados (problemas)

- A recolha de dados de várias fontes contém muitos problemas:
 - incorreções
 - dados em falta
 - má codificação
 - incongruências
 - incompletos
- A principal razão para isto acontecer é devido à recolha de dados ser:
 - observacional, e
 - não experimental

Tipos de dados (problemas)

- Isto também é provocado porque os dados estão distribuídos
 - provoca que um determinado dado pode ter diferentes nomes e codificações
 - levanta problemas no momento da integração desses dados
- Dados em falta, má codificação e incongruências
 - podem ser parcialmente corrigidos de uma forma automática com ferramentas adequadas

Tipos de dados

- Dados complexos
 - os dados não são totalmente compreendidos
 - impossibilita aplicar análises hipotético-dedutiva aos dados
 - dificuldade em perceber como e porquê os dados são recolhidos (diferentes fontes)
- Os desafios podem ser sintetizados em:
 - qualidade dos dados
 - privacidade dos dados

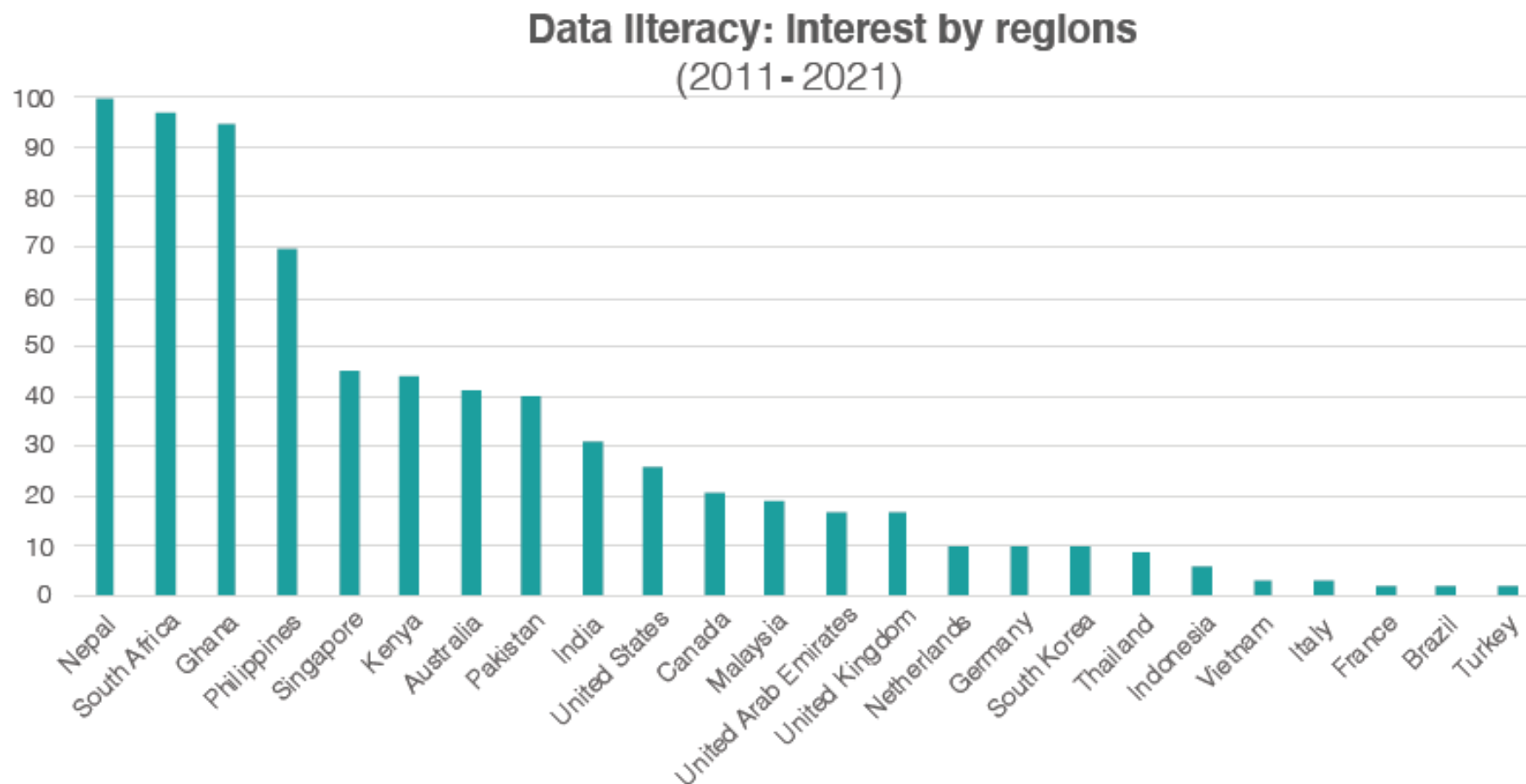
Tipos de dados (literacia)

- O que significa ser (literalmente) alfabetizado (literate)?
 - o significado original da alfabetização está ligado a letras- um sistema de convencional de símbolos que representam palavras.
 - a maioria dos sistemas de escrita são alfabéticos no sentido de que estes símbolos não representam as palavras diretamente, mas os sons das palavras
 - aprender a ler e escrever não é fácil, e muitas línguas tornam a aprendizagem de leitura ainda mais desafiador porque os seus sistemas ortográficos contêm inconsistências

Tipos de dados (tipos de literacia)

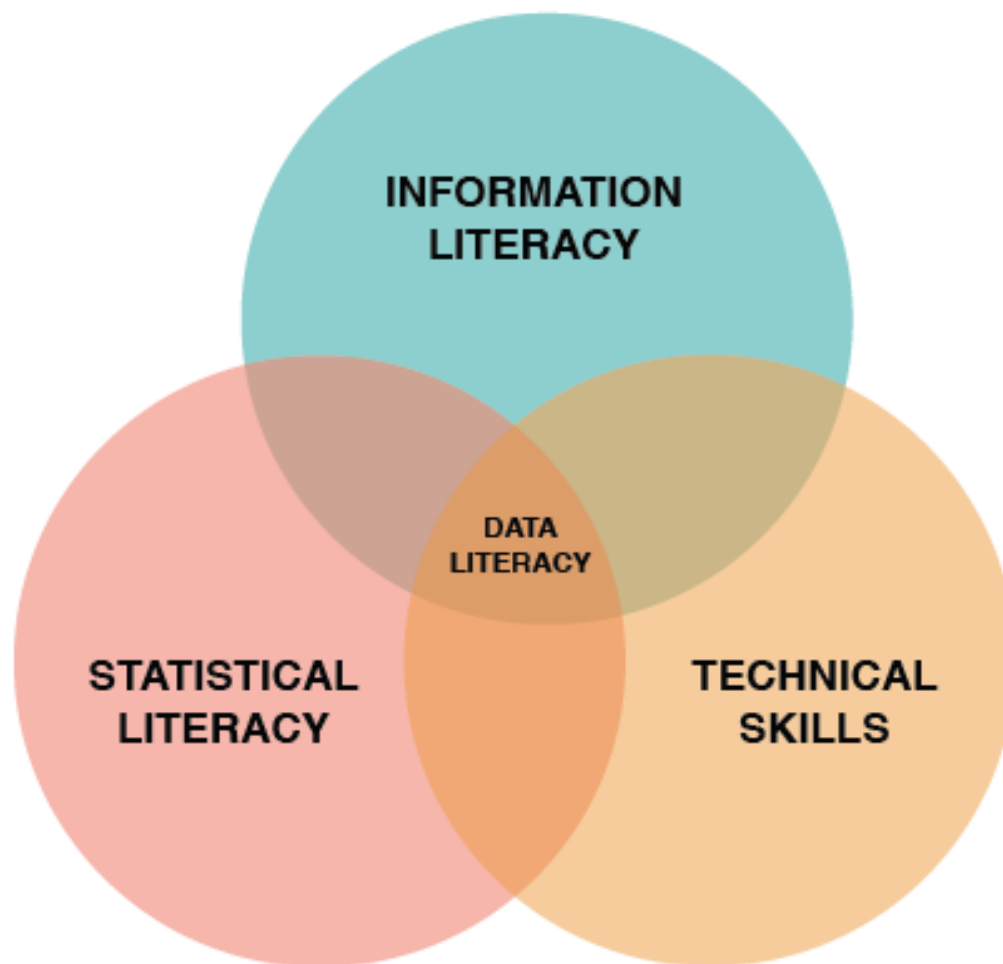
- Visual literacia – este termo foi usado em 1950 pelo fotógrafo Henry Smith, e devido ao interesse em uso de fotografias na educação como meio de comunicação. A literacia visual é abordada na psicologia Gestalt e inclui a gramática de elementos visuais (pontos, linhas, formas, direções, tons, cores, textura, escalas, movimento).
- Media literacia - capacita as pessoas a serem pensadoras críticas e produtoras criativas de uma gama cada vez mais ampla de mensagens usando imagem, linguagem e som. As tecnologias de comunicação transformam a sociedade, elas impactam a nossa compreensão de nós mesmos, nossas comunidades e as nossas diversidades culturais, tornando a media literacia uma capacidade essencial para a vida no século XXI.
- Informação literacia – capacita uma pessoa a ser capaz de reconhecer quando a informação é necessária e tem a capacidade de a localizar, avaliar e usar efetivamente a informação necessária. Em última análise, as pessoas alfabetizadas em informação são aquelas que aprenderam a como aprender.

Tipos de dados (data literacy)



Source: [Google Trends](#), as of March 14, 2021. "Interest" is captured by Google searches for the term "data literacy". Searches pertain to the period 2011-2021. The vertical axis represents search interest relative to the highest point on the chart, and doesn't convey absolute search volume. For more information, please refer to the Google Trends FAQ

Tipos de dados (data literacy)



Source: What is data literacy?" graphic reproduced from [UN Data Revolution website](#)

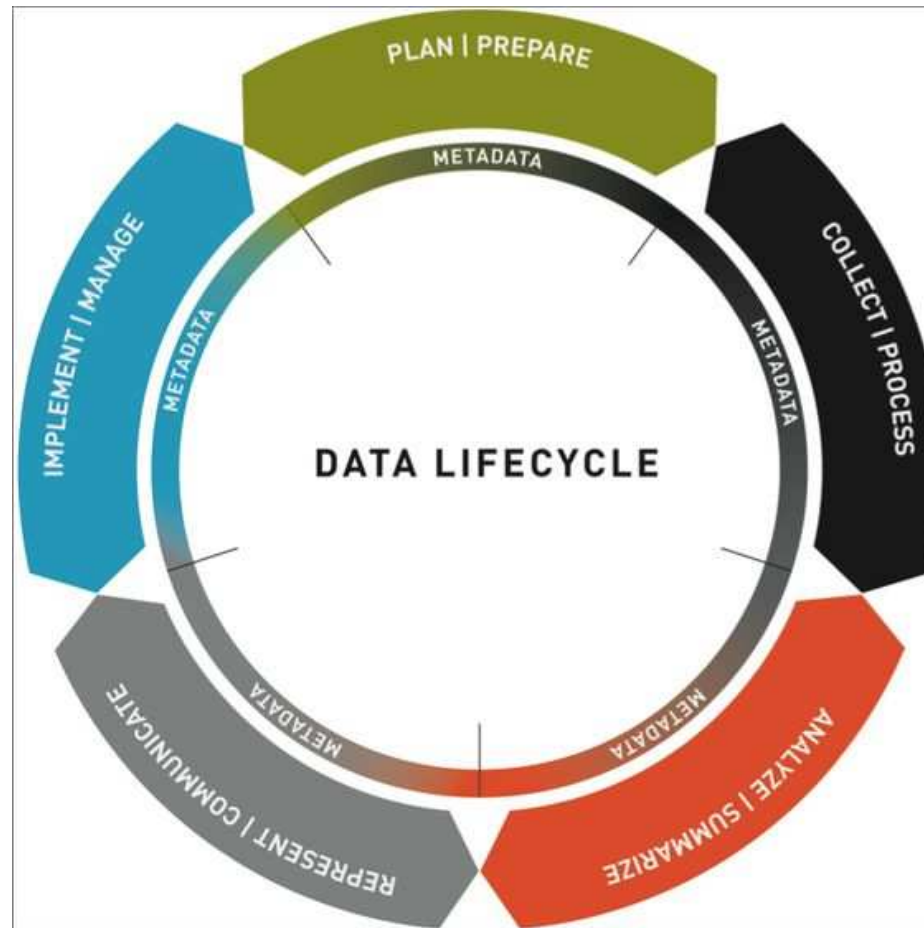
Tipos de dados

- Não existe consenso na definição de data literacy. Podem ser identificadas três visões:
 - use-centric phenomenon, “the ability of non-specialists to make use of data” (Frank et al. 2016)
 - “collect and manage” dados (Risdale et al, 2015)
 - “select and clean” dados (Wolff et al, 2016)

Tipos de dados (data life-cycle)

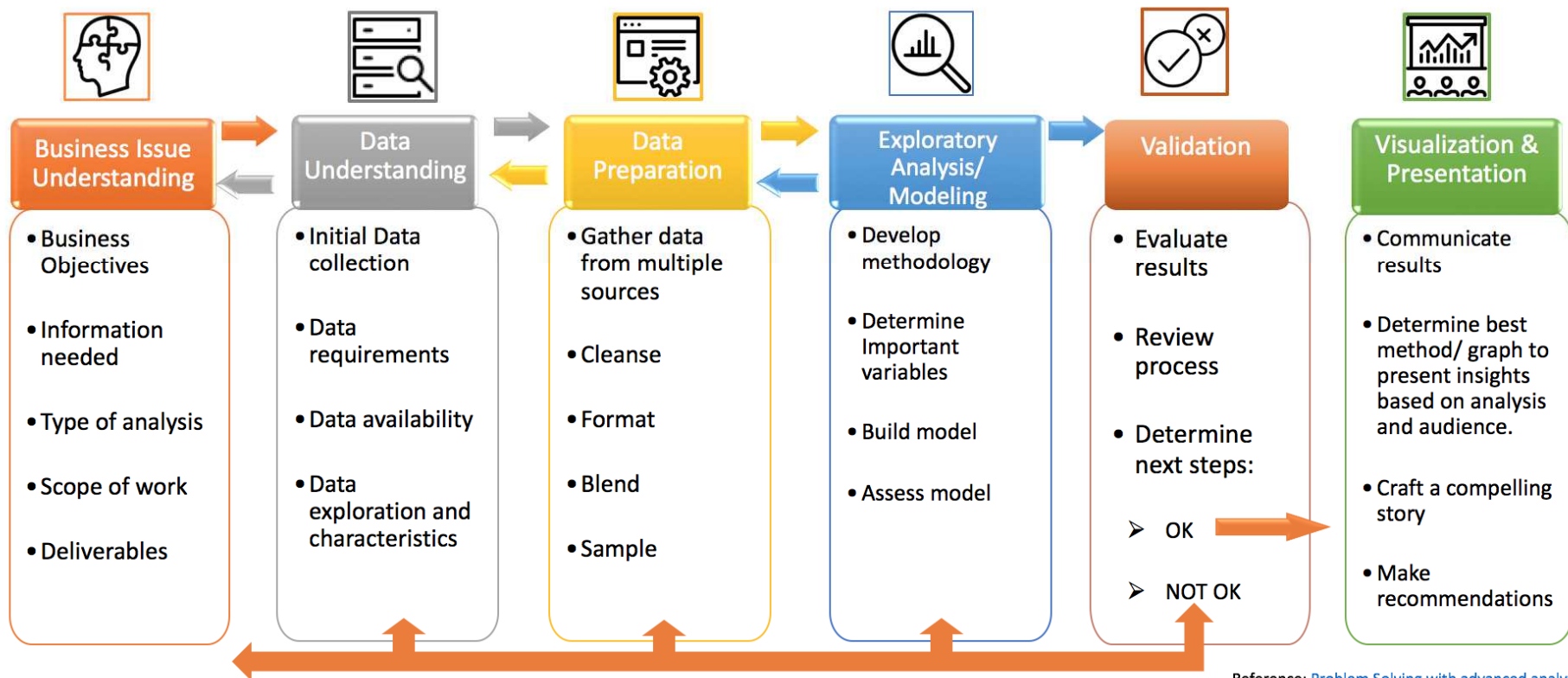
- Os estádios do ciclo de vida dos dados podem ser vistos como:
 - contexto: inclui uma consciência geral dos dados e de seu ambiente. Isso envolve uma compreensão do contexto subjacente dos dados, suas origens, propósito e sistemas associados.
 - planeamento: inclui a identificação das necessidades de dados com base no cenário do mundo real e investigação crítica, desenvolvimento de uma estratégia ou planos de dados, considerando aspetos de *data governance*
 - produção: inclui aspetos relacionados com a aquisição de dados, *design*, fontes, recolha, processamento, gestão e disseminação
 - avaliação: inclui o aspeto de "trabalhar com dados", como interpretação, aplicação, análise e visualização dos dados
 - utilização: inclui comunicação de dados, crítica, argumento e defesa relacionada à tomada de decisão baseada em dados. Isso também pode incorporar o compartilhamento ético de dados e práticas de reutilização

Tipos de dados (data life-cycle)



Tipos de dados (data life-cycle)

Life Cycle of a data analysis project based on
CRISP- DM methodology



Reference: https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining