



DIGITAL DATA LITERACY FOR EDUCATION

O que é a literacia de dados?

Módulo 1

Funded by the
Erasmus+ Programme
of the European Union



O que é a literacia de dados?

Sobre esta aula

O que irá aprender

Prazer em conhecer-vos, dados!

- O que são dados?
- Dados vs. opiniões: Os dados são neutros?
- Os dados e o ensino.

Onde podemos encontrar dados e como os podemos avaliar criticamente?

- Verificação de factos e verificação de dados: Três perguntas a fazer ao encontrar dados;
- Onde vos posso encontrar, dados? Recursos e técnicas;
- Portais de dados.

Competências

- **Compreender dados**
 - Observar dados digitais;
 - Refletir sobre dados digitais.
- **Utilizar dados**
 - Navegar através de dados digitais;
 - Recolher dados digitais.

Resultados de aprendizagem

- Articular necessidades de informação, localizar e extrair dados, informação e conteúdo digitais;
- Avaliar a relevância da fonte e do respetivo conteúdo, sendo capaz de selecionar fontes de informação fidedignas;
- Adquirir conhecimento sobre como e onde encontrar dados.



1.1 Prazer em conhecer-vos, dados!

O que são dados?

Definição

Dados são informação codificada e representada de uma forma específica. A codificação pode ser feita por uma máquina ou por um ser humano (ou por uma combinação de ambos). São uma **convenção** através da qual o mundo e respetivos fenómenos são **medidos e descritos**.

O que são dados?

Definição

Uma linha numa folha de cálculo contém dados, tal como um pixel vermelho no seu computador, uma imagem ou a mensagem que as células da sua pele transmitem aos neurónios num dia quente de agosto. **Qualquer coisa que seja medida ou codificada (por meio de instrumentos ou dos seus sentidos) pode ser considerada como um dado.**

Esta é uma definição básica, mas que deve ser tida em mente ao longo do curso: **quando falamos de dados, não estamos a falar de «verdade»**; estamos a lidar com uma **medida ou codificação da realidade, feita por algo ou alguém.**

O que são dados?

Definição

Enquanto medida ou codificação da realidade, feita por algo ou alguém, qualquer dado inclui um certo nível de **subjetividade**, o qual depende das circunstâncias e da metodologia utilizada para criar essa medida ou codificação.

Dados vs. opiniões

«Acredito em Deus, todas as outras pessoas devem apresentar dados.»

«Sem dados, é só mais uma pessoa com uma opinião.»

W. Edwards Deming (engenheiro, professor, autor e consultor de gestão)

O que pensa de ambas as citações?

Existe uma distinção clara entre dados e opiniões?

Por outras palavras: **Os dados são neutros?**

Dados vs. opiniões

«Os dados são sempre um produto de relações sociais desiguais — relações afetadas por séculos de história.»

*Citação do livro «Data Feminism» de
C. D'Ignazio, L.K. Klein*

Os dados não são nem podem ser neutros.

Embora apresentados como números e derivados da matemática aplicada (por exemplo, da estatística), os dados não são sinónimos de números ou matemática.

O que diferencia dados de números é o facto de que os números são abstrações matemáticas, uma ideia. Como tal, podem ser neutros. **No entanto, este não é o caso dos dados, os quais são originados no mundo real e por pessoas reais.**

Dados vs. opiniões

Uma vez que os dados são originados no **mundo real** e referem-se **pessoas reais**, é útil **compreender melhor a realidade à nossa volta**.

No entanto, é importante ter em conta que:

- **os dados por si só não são suficientes,**
- **os dados podem mentir.**

Por isso, a **verificação de factos** e a **verificação de dados** são atividades fundamentais ao trabalhar com dados.

Os dados e o ensino

Nos diapositivos seguintes, irá encontrar dois vídeos que lhe permitirão ter uma ideia geral sobre a forma como **a utilização de dados eficaz por parte de uma comunidade escolar** pode informar melhor a prática letiva, apoiar a comunidade estudantil no processo de aprendizagem e contribuir para um melhor desempenho académico da mesma.

Os dados e o ensino

«Utilizar dados para apoiar o desenvolvimento de docentes e estudantes.»



Fonte: [Canal do YouTube da Edutopia](#)

Os dados e o ensino

«Dados: São apenas uma parte de um ensino de qualidade.»



Fonte: [Canal do YouTube da Data Quality Campaign](#)



1.2 Onde podemos encontrar dados e como os podemos avaliar criticamente?

Verificação de factos e verificação de dados

Três perguntas a fazer ao encontrar dados

*Nota: Não precisa de ser data scientist para o fazer!
Ler as notícias é suficiente para encontrar imensa
informação baseada em dados.*

- Qual é a fonte?
- Os dados estão apresentados de forma correta e completa?
- Como foram recolhidos os dados?

Verificação de factos e verificação de dados



É essencial verificar a fonte dos dados porque, por vezes, esta poderá não ser imparcial.

Nesse caso, a mesma deve ser considerada inexata porque serve um interesse pessoal.

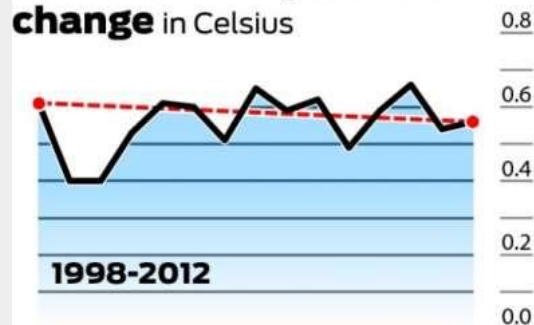
Veja os «dados» publicitados pelas grandes empresas de tabaco no pós-guerra.

Verificação de factos e verificação de dados

Os dados estão apresentados de forma **correta e completa**?

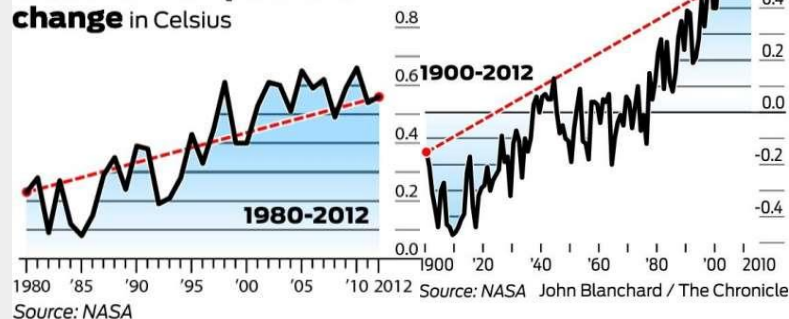
Algo que devemos observar é a tática de **supressão de evidências** relacionadas com os dados, ou *cherry picking* em inglês, especialmente ao trabalhar com tópicos políticos e controversos. Por exemplo, muitos negacionistas das alterações climáticas gostam de iniciar os gráficos aos quais recorrem no ano de 1998, um ano excepcionalmente quente, para mostrarem que as temperaturas estão, na verdade, a diminuir.

Global air temperature change in Celsius



Source: NASA

Global air temperature change in Celsius



Os três gráficos foram retirados de «[How climate deniers abuse statistics to mislead](#)» de James Temple, SFGate

Verificação de factos e verificação de dados

Como foram recolhidos os dados?

The goal is not to predict a price but an expected value of a player in a free market. Both individual transfer modalities and situational conditions are relevant in determining market values. Examples of this are listed below. **Transfermarkt does not use an algorithm but instead relies on the wisdom of the community.**

É importante compreender a forma como os dados são recolhidos ou processados. Por exemplo, de que forma é que o site Transfermarkt* recolhe as avaliações do mercado de jogadores comunicadas?

Se procurarmos mais informações sobre a fonte, descobrimos que o Transfermarkt não utiliza qualquer algoritmo, dependendo da «sabedoria da comunidade».

* [Transfermarkt](#) é um site com informações sobre o futebol, tais como classificações, resultados, estatísticas, notícias sobre transferência e jogos.

Verificação de factos e verificação de dados

Antes de dedicar tempo a trabalhar num conjunto de dados, certifique-se de que está a lidar com um conjunto de dados bom e fidedigno. Os dados são uma fonte e, tal como quando lida com qualquer fonte, precisa primeiro de avaliar se pode confiar na mesma.

Para começar, tal como o faria com pessoa, um novo artigo ou um artigo de investigação:

pesquise a fonte de dados:

- Quem recolheu os dados?
- Porque foi feita essa recolha?
- Porque estão a ser publicados os dados (e porquê agora e não antes/depois)?
- Que período abrangem? (E porque não um período mais longo? É possível obter dados sobre o período anterior/seguinte?)
- Como foram recolhidos? (Leia sempre a metodologia.) Garanta que compreende o que é calculado com cada variável.

Verificação de factos e validação de dados

Mesmo em tópicos menos controversos, deve sempre refletir cuidadosamente sobre os dados e verificar se são adequados ao assunto em questão.

Por exemplo, apesar de ser uma prática popular utilizar **dados de redes sociais**, na verdade, pode ser inapropriado utilizá-los se não os enquadrar corretamente para ter em consideração as limitações dos mesmos. Os resultados de inquéritos de institutos de estatística fidedignos podem ser muito mais úteis, uma vez que deverão incluir informações sobre a metodologia utilizada nos respetivos metadados, ser projetados para serem inclusivos e representativos de vários grupos demográficos e incluir informação sobre margens de erro, etc.

Verificação de factos e verificação de dados

Por outro lado, apesar da promessa e das expectativas dos macrodados, os dados de redes sociais podem, na verdade, ser menos representativos da população. Quantas pessoas seniores utilizam o Twitter? As pessoas que vivem na cidade, que têm constante acesso ao Wi-Fi, não têm uma maior probabilidade de fazer publicações do que as pessoas que vivem em regiões remotas e rurais?

Com isto queremos dizer que **ao verificar as fontes dos dados que encontrar, deve fazê-lo não só em termos absolutos (estes dados provêm de uma fonte fidedigna?), mas também em termos relativos (em que medida posso confiar nestes dados para oferecer respostas fiáveis às minhas perguntas específicas?).**

Verificação de factos e verificação de dados

Os números falam por si
mesmos?

De um modo simples, a resposta é: não.

Saber como analisar números é tão importante quanto saber como interpretar os resultados da análise.

O que significam esses números?

Conheço suficientemente bem o contexto do fenómeno que estou a analisar?

No diapositivo seguinte, irá encontrar um exemplo prático da forma como os números não falam por si mesmos e da importância de conhecer o contexto da análise.

Verificação de factos e verificação de dados

Os números falam por si
mesmos?

Em 2014, Mona Chalabi escreveu o artigo [Mapping Kidnappings in Nigeria](#) para a FiveThirtyEight. Este trabalho baseou-se nos dados que a Global Database of Events, Language and Tone (GDELT) tem recolhido sobre os raptos na Nigéria desde 1982 e tinha por objetivo contextualizar o rapto de cerca de 300 meninas em idade escolar pelo Boko Haram.

O artigo apresentava valores incríveis sobre o aumento do número de raptos nos últimos anos e um mapa que mostrava uma elevada concentração de raptos numa região central da Nigéria ou que filtrava os raptos mais recentes em Borno, onde ocorreu o rapto das meninas.

Verificação de factos e verificação de dados:

Os números falam por si
mesmos?

O que não está certo nestas conclusões? A base de dados da GDELT é uma fonte secundária que, por sua vez, depende de relatórios da comunicação social — menções em jornais — para reconstituir o número de casos que irá comunicar na sua base de dados.

Assim, se de acordo com a GDELT houver 649 raptos na Nigéria no espaço de um mês, então, foram escritas 649 notícias nos jornais sobre os raptos na Nigéria.

Deste modo, o problema com o artigo não é a análise dos números, mas sim a representação dos mesmos: os dados refletem o interesse da comunicação social nos raptos e não os raptos em si.

Onde vos posso encontrar, dados?

Portais de dados

Os portais de dados, normalmente, estão disponíveis na forma de diretórios que alojam dados (ou ligações para os dados), metadados e recursos/ferramentas sobre os conjuntos de dados contidos.

Os dados serão licenciados consoante o portal ou a entidade publicadora. Se pretender, pode pesquisar especificamente por portais de dados abertos. (No módulo seguinte, iremos falar sobre metadados e as várias licenças existentes.)

Na prática, os portais de dados podem ter várias formas e dimensões: podem pertencer a governos e administrações internacionais/nacionais/locais, a institutos de estatística, a organizações que trabalhem em domínios específicos (pobreza, direitos humanos, etc.), podem ser agregadores de dados, etc.

Muitos deles pertencerão a instituições públicas, mas também poderão pertencer a empresas privadas, ONG, agências de notícias, etc.

Onde vos posso encontrar, dados?

Portais de dados

Exemplos de portais de dados

[LandMatrix](#), portal de dados de negócios sobre propriedades

[Portal de dados abertos da cidade de Milão](#)

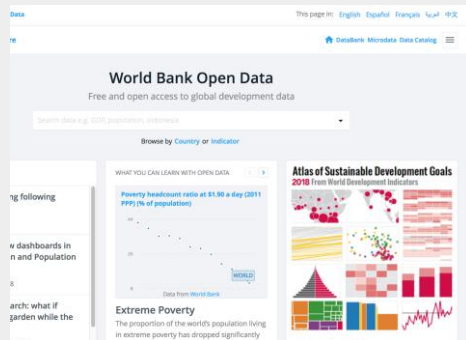
Settore	Stipendio medio lordo annuo	2015		2016	
		maschi	femmine	maschi	femmine
Totale					
• Alta					
• Impiegati	employee income	36 676	36 614	36 676	36 614
• Imprenditori	self-employed income	37 613	36 133	37 613	36 133
• Imprenditori	public business income	23 877	16 633	23 877	16 633
• Imprenditori	other type	21 682	16 676	21 682	16 676
• Imprenditori	tax	33 127	26 749	33 127	26 749
• Medio-bassa					
• Imprenditori	employee income	26 749	26 749	26 749	26 749
• Imprenditori	self-employed income	27 876	26 813	27 876	26 813
• Imprenditori	public business income	26 655	26 655	26 655	26 655
• Imprenditori	tax	27 085	27 085	27 085	27 085
• Bassa					
• Imprenditori	employee income	24 198	24 198	24 198	24 198

[ISTAT.it](#), dados do Instituto de Estatística Italiano

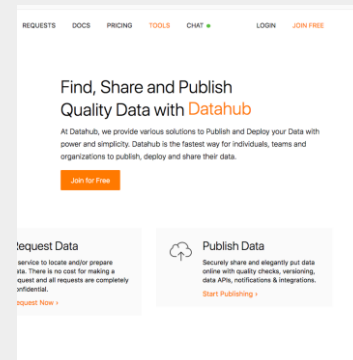
Onde vos posso encontrar, dados?

Portais de dados

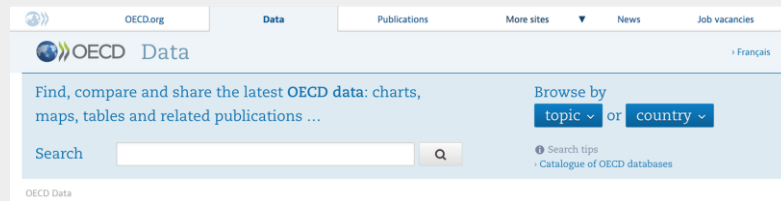
Exemplos de portais de dados



[World Bank Open Data](#)



[DataHub](#)



[OECD Data](#)

Onde vos posso encontrar, dados?

Acesso aberto a dados de investigação

O termo «*dados de investigação*» refere-se aos dados quantitativos e qualitativos produzidos, recolhidos e/ou analisados no decorrer de um projeto de investigação.

Cada vez mais programas de investigação encorajam investigadores e investigadoras a partilharem não só as suas publicações, como também os dados de investigação em formato aberto.

Existem várias razões por trás desta política, tais como o facto de que a investigação conta, frequentemente, com financiamento público e, por isso, os resultados devem estar publicamente disponíveis e o facto de que a partilha dos dados da investigação contribui para a colaboração e facilita o progresso e a revisão entre pares.

Onde vos posso encontrar, dados?

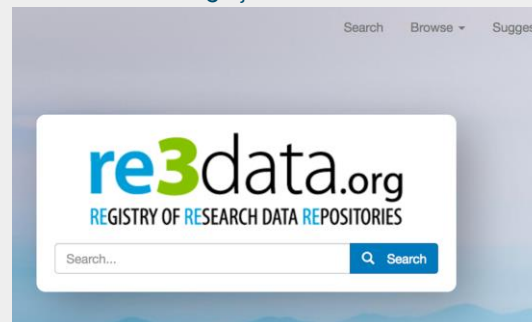
Acesso aberto a dados de investigação

Exemplos de fontes de investigação abertas

[Scientific Data](#) do Nature, um jornal com fontes de investigação de acesso aberto e ligações para repositórios de dados em cada artigo.



[re3data](#), registo de mais de 2000 repositórios de dados de investigação.



Onde vos posso encontrar, dados?

FOIA

(ou outros tipos de leis de liberdade de acesso à informação. Para conhecer mais informações sobre este tipo de leis num país específico, pode realizar uma pesquisa em [FOIA Advocates](#).)

Muitos países possuem leis que regulam o acesso de cidadãos e cidadãs à informação produzida e detida por organismos públicos. Este tipo de leis baseia-se na crença de que **tal acesso, juntamente com a transparência das instituições públicas, é um direito cívico universal.**

Na prática, isto significa que qualquer cidadão ou cidadã pode submeter pedidos de dados e que estes pedidos devem ser satisfeitos atempadamente, a menos que haja obstáculos consideráveis à divulgação dos mesmos (tais como preocupações de privacidade, questões de segurança nacional, pedidos vagos, pedidos excessivamente difíceis de satisfazer, etc.).

Pode saber mais sobre as leis relativas à liberdade de informação de cada país [aqui](#).

Onde vos posso encontrar, dados?

FOIA

(ou outros tipos de leis de liberdade de acesso à informação. Para conhecer mais informações sobre este tipo de leis num país específico, pode realizar uma pesquisa em [FOIA Advocates](#)).

Embora o processo de obtenção de dados através de pedidos FOIA possa ser intimidante e implicar uma longa espera e longas trocas de informação, os resultados, normalmente, valem a pena: **estará a contribuir ativamente para a sua comunidade** ao abrir o acesso a dados que deveriam ser públicos, mas que ainda não tinham sido tornados acessíveis.

Adicionalmente, **poderá descobrir histórias interessantes sobre conjuntos de dados até então não publicadas.**

Onde vos posso encontrar, dados?

FOIA

(ou outros tipos de leis de liberdade de acesso à informação. Para conhecer mais informações sobre este tipo de leis num país específico, pode realizar uma pesquisa em [FOIA Advocates](#)).

Exemplos de histórias conhecidas graças a pedidos FOIA e semelhantes



Wired Italy | Sezioni | Wired Next Fest | Gallery | Wired Next

TOPIC: ESTATE TRAILER MATERACIA FACEBOOK GOVERNO MIGRANTI GOOGLE SERIE TV...

HOME ATTUALITÀ **POLITICA**

Vaccini, come il Foia ha permesso di costruire la mappa delle coperture

È stato possibile ottenere i dati sui vaccini grazie al Freedom of information act. Ecco come è andata l'esperienza di Wired con il Foia

[Vaccini, la mappa dell'Italia che rinuncia a proteggersi](#) de Riccardo Saporiti para a Wired Italy



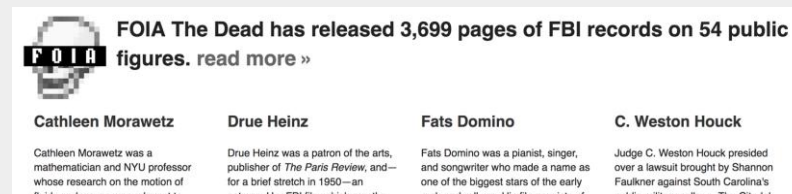
In Cold War, U.S. Spy Agencies Used 1,000 Nazis

By Eric Lichtblau

Oct. 26, 2014

WASHINGTON — In the decades after World War II, the C.I.A. and other United States agencies employed at least a thousand Nazis as Cold War spies and informants and, as recently as the 1990s, concealed the government's ties to some still living in America, newly disclosed records and interviews show.

[In Cold War, U.S. Spy Agencies Used 1,000 Nazis](#) de Eric Lichtblau para o The New York Times



FOIA The Dead has released 3,699 pages of FBI records on 54 public figures. read more >>

Cathleen Morawetz	Drue Heinz	Fats Domino	C. Weston Houck
Cathleen Morawetz was a mathematician and NYU professor whose research on the motion of...	Drue Heinz was a patron of the arts, publisher of <i>The Paris Review</i> , and— for a brief stretch in 1950—an...	Fats Domino was a pianist, singer, and songwriter who made a name as one of the biggest stars of the early...	Judge C. Weston Houck presided over a lawsuit brought by Shannon Faulkner against South Carolina's public military college, The Gradu...

[FOIA The Dead](#) da Freedom of the press Foundation

Onde vos posso encontrar, dados?

O público

Também pode recorrer «ao público» para obter os dados que precisa.

Na prática, **isto pode significar várias coisas**: desde pedir às pessoas que respondam a inquéritos para recolher conteúdo gerado pelo utilizador e dados de redes sociais, a recorrer ao *crowdmapping* para utilizar plataformas que permitam a pesquisa e etiquetagem de dados «human-in-the-loop».

Ao lidar com este tipo de dados (tal como com qualquer outro, na verdade), **tenha sempre em consideração a limitação e a parcialidade das suas escolhas**.

A quem está a pedir para responder ao inquérito e onde está a publicar esse mesmo pedido? Que pessoas estão presentes em determinada rede social? Que pessoas serão inevitavelmente excluídas da amostra selecionada por razões de idade, acesso à *Internet*, etc.?

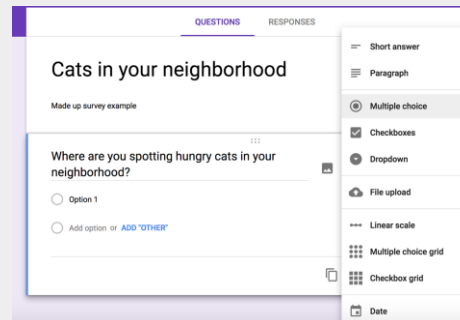
Onde vos posso encontrar, dados?

O público

Colaboração ativa

Pedir aos utilizadores/público que contribua ativamente para a produção do conjunto de dados necessário.

Exemplos de como obter dados por *crowdsourcing*



Criar um questionário no [Google Forms](#)

Ushahidi pode ser utilizado para ajudar as pessoas a criarem coletivamente relatórios e mapas sobre fenómenos. Imagens da publicação de [blogue do Ushahidi «Crisis Mapping Haiti: Some Final Reflections»](#)



Algumas sugestões para começar:

Portais de dados

- [Portal de dados abertos World Bank](#)
- [Dados abertos das Nações Unidas](#)
- [OpenAfrica](#)
- [WikiData](#)
- [Portal de Dados Europeu](#) (dados de países da UE) e [Portal de Dados Abertos da UE](#) (dados de instituições da UE)
- [OECD Data Portugal](#)
- [DataHub.io](#)
- [Portal de dados abertos da NASA](#)
- [Kaggle](#), conjuntos de dados de formação sobre *data science*
- [Data.World](#), uma plataforma para carregar, partilhar e pesquisar conjuntos de dados e trabalhar colaborativamente em projetos
- [OpenSpending](#), dados fiscais e de despesa pública
- As ONG, as universidades e os grupos de investigação, por vezes, também divulgam dados juntamente com relatórios e investigações publicadas. Veja a Universidade de Maryland → [Global Terrorism Database](#); SIPRI → [Data on Military Expenditure](#)
- [Explorador de dados públicos da Google](#)
- [OpenCorporates](#), base de dados sobre empresas
- [Repositório de dados da Organização Mundial da Saúde](#)
- Portais de dados de agências de notícias. (Por exemplo, o [repositório de dados da FiveThirtyEight](#) – A agência de notícias baseada em dados FiveThirtyEight partilha os dados que estão por trás dos seus artigos)