MORINGA

Discover · Grow · Transform

What is Data Literacy?

Module 1

MORINGA
Discover · Grow · Transform

What is Data Literacy?

About this class

You will learn

Nice to meet you, data!

- What is data?
- Data vs opinions: is data neutral?
- Data and teaching

Where to find data and how to critically assess them

- Fact-checking and data verification: 3 questions to ask yourself when you meet data
- Where can I find you, data? Resources and techniques
- Data Portals

MORINGA
Discover · Grow · Transform

Competences

- **Understanding data**
  - Observing digital data
  - Reflecting on digital data
- **Using data**
  - Navigating digital data
  - Collecting digital data

Learning Outcomes

- Articulate information needs, to locate and retrieve digital data, information and content.
- Judge the relevance of the source and its content, thus being able to select reliable sources of information;
- Acquire knowledge on where and how to find data

MORINGA
Discover · Grow · Transform

1.1 Nice to meet you, data!

MORINGA
Discover · Grow · Transform

What is data?

A definition

Data is **information coded and represented in a specific form**. The coding can be done by a machine or by a human (or by a mix of both). Data is a **convention** through which the world and its phenomena are **measured** and **described**.

MORINGA
Discover · Grow · Transform

What is data?

A definition

A row in a spreadsheet contains data, but so does red pixel on your computer, or an image, or the message the cells of your skin transmit to your neurons on a sweaty August day. **Anything that gets measured or coded (through instruments or through your senses) is data**.

This is a basic definition, but it should be kept in mind throughout the course: **when we talk about data, we are not talking about the "Truth"**, but we are dealing with a **measurement or coding of the reality made by something or someone.**

MORINGA
Discover · Grow · Transform

What is data?

A definition

As a measurement or coding of the reality made by someone or something, any data comes with a certain level of **subjectivity**, the amount of which depends on the circumstances and methodology used to make the measurement or coding.

MORINGA
Discover · Grow · Transform

Data vs Opinions

“In God we trust, all others must bring data.”

“Without data, you're just another person
with an opinion.”

W. Edwards Deming. (engineer, professor, author, and management consultant)

What do you think about these two quotes?

Is there a clear-cut distinction between data and opinions?

Put another way: **Are data neutral?**

MORINGA
Discover · Grow · Transform

Data vs Opinions

"Data are always the product of unequal social relations—relations affected by centuries of history."

Quote from the book. "Data Feminism" - C. D'Ignazio, L.K. Klein

Data are not and cannot be neutral. While often presented as numbers, and derived from applied mathematics (such as statistics), data are not synonymous with numbers or with math.

What differentiates data from numbers is that numbers are mathematical abstractions, an idea. As such, they can be neutral. **But data, originating from the real world and real people, cannot**.

MORINGA
Discover · Grow · Transform

Data vs Opinions

Since data comes from the **real world** and are about **real people**, it is both useful and helpful to **better understand the reality around us**.
However, it's important to keep in mind that:

- **Data alone is not enough**
- **Data can lie**

Therefore, **fact-checking** and **data verification** are core activities when working with data.

MORINGA
Discover · Grow · Transform

Data and Teaching

In the next slides you will find two videos that will give you an overview of how **the effective use of data by a school community** can better inform the teaching practice, support students through their learning process and lead them to improved academic performance.

MORINGA

Data and Teaching

"Using Data to Support Teacher and Student Growth"

Source: Edutopia Youtube Channel

MORINGA
Discover · Grow · Transform

Data and Teaching

“Data: It's Just Part of Good Teaching”

Source: Data Quality Campaign Youtube Channel

MORINGA
Discover · Grow · Transform

1.2 Where to find data and how to critically assess them

MORINGA
Discover · Grow · Transform

Fact-checking and data verification

3 questions you should ask yourself when you meet data

Please note: you don't need to be a data scientist for that to happen! Reading the news is enough to bump into a bunch of data-driven information

- Who is the source?
- Is data presented in a correct and complete way?
- How has data been collected?

MORINGA
Discover · Grow · Transform

Fact-checking and data verification

Verifying our data source is key because sometimes it might be biased.

If so, it is to be considered unreliable because it pursues a self-serving interest.

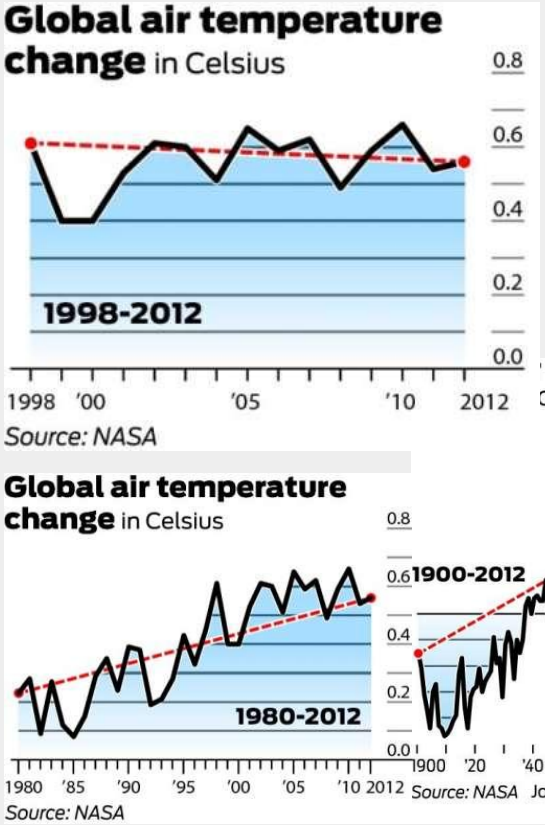
See the "data" being advertised by the big tobacco corporations in the post-war era.

MORINGA
Discover · Grow · Transform

Fact-checking and data verification

Is data presented in a **correct** and **complete** way?

Something to look out for is the tactic of **cherry-picking** the data, especially when working on political and controversial topics. As an example, many climate change deniers like to start they charts in 1998, an exceptionally hot year, to show that temperatures are actually getting cooler...

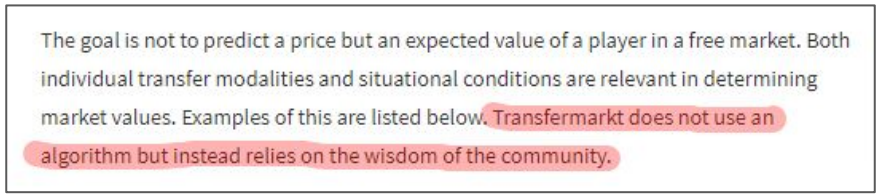


All three charts from "How climate deniers abuse statistics to mislead" by James Temple · SFGate

MORINGA
Discover · Grow · Transform

Fact-checking and data verification

How has data been collected?


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.

It is important to understand how data is collected or processed. For example: how does the website Trasfermarkt* collect the reported player market evaluations?

Looking for more information about the source, we discover that Trasfermarkt does not use an algorithm, but relies on the "wisdom of the community"...

* *Transfermarkt is a website that has footballing information, such as scores, results, statistics, transfer news, and fixtures.*

MORINGA
Discover · Grow · Transform

Fact-checking and data verification

Before investing time on working with a dataset, make sure you are dealing with a good and trustworthy one. Data is a source, and just like when dealing with any source, you first need to assess whether it can be trusted.

To start, as you would do for a person, a new-article, or a piece of research: **interview the data source.**

- Who collected the data
- Why did they collect it
- Why are they publishing it (and why now and not before/later)
- What time period does it cover (and why not a longer one? can you get data on the previous/following time period?)
- How they collected it (always read the methodology). Make sure you understand what is computed with each variable.

MORINGA
Discover · Grow · Transform

Fact-checking and data verification

Even on less controversial topics you should always think carefully about your data and whether it is fit for your scope.

For example, despite the popularity of using **social media data**, it might actually be inappropriate when you don't frame it correctly to account for its limitations. Survey results from trusted statistical institutes might be way more useful, as they should include methodological information in their metadata, should be designed to be inclusive and representative of different demographic groups; should include information on margins of error; etc.

MORINGA
Discover · Grow · Transform

Fact-checking and data verification

Social media data, on the other hand, despite the promise and expectations of big data, might in fact be less representative of the population. How many seniors use Twitters? Aren't city-dwellers, constantly covered by wi-fi, more likely to post than inhabitants of remote and rural regions?

This is to say that, **when verifying the sources of your data, you should do so not only in absolute terms (is this data from a trusted source) but also in relative terms (to what length can I trust this data to provide reliable answers to my specific question?)**

MORINGA
Discover · Grow · Transform

Fact-checking and data verification

Do numbers speak for themselves?

Put simply the answer is: no, they don't.

Knowing how to analyse numbers is as important as knowing how to interpret the results of the analysis.

What do those numbers mean?

Do I know enough the context of the phenomenon I am analysing?

In the following slide you will find a practical example of how numbers do not speak for themselves and of the importance of knowing the context of the analysis.

MORINGA
Discover · Grow · Transform

Fact-checking and data verification

Do numbers speak for themselves?

In 2014 Mona Chalabi wrote the article Mapping Kidnappings in Nigeria for FiveThirtyEight. This work was based on data from the Global Database of Events, Language and Tone (GDELT) referring to kidnappings in Nigeria since 1982 and it aimed at contextualising the kidnapping of nearly 300 schoolgirls by Boko Haram.

The article reported incredible numbers on the growth in the number of kidnappings in recent years and a map showing a very high concentration of those in a central region of Nigeria or, filtering for the most recent kidnappings, in Borno, where the kidnapping of the schoolgirls took place.

MORINGA
Discover · Grow · Transform

Fact-checking and data verification:

Do numbers speak for themselves?

What is wrong with these conclusions? The GDELT database is a secondary source, which in turn relies on media reports - newspaper mentions - to reconstruct the number of events to be reported in its database.

So, if according to GDELT there were 649 kidnappings in Nigeria in one month, the sense is that there were 649 newspaper stories about kidnappings in Nigeria.

The problem with the article is therefore not in the analysis of the numbers but in their interpretation: the data records media interest in kidnappings, not the kidnappings.

MORINGA
Discover · Grow · Transform

Where can I find you, data?

Data Portals

Data portals usually come in the form of directories hosting data (or links to the data), metadata and resources/tools about datasets contained.

Data will be licensed depending on the portal or publisher. If you wish, you can search specifically for Open Data Portals. (In the next module, we will talk about metadata and different licences)

In practice, data portals come in different shapes and sizes: they can be from international/national/local governments and administrations; from statistics institutes; from organizations covering specific topics (poverty, human rights, etc.); they can be data aggregators, etc.

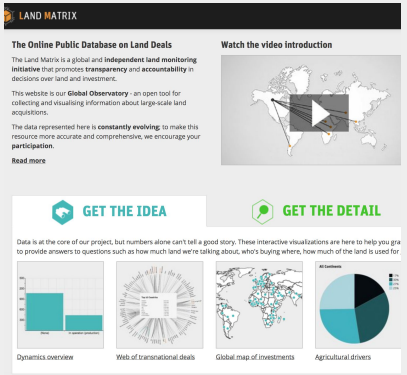
Many of them will be from public institutions, but they can also be from private businesses, NGOs, news companies, etc.

MORINGA
Discover · Grow · Transform

Where can I find you, data?

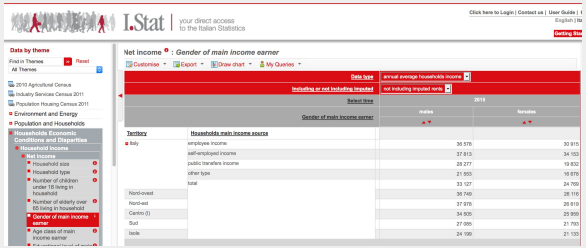
Data Portals

Examples of Data Portals



LandMatrix, data portal on land deals

City of Milan Open Data Portal



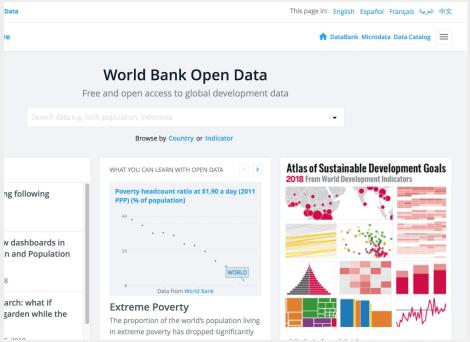
ISTAT.it, Data from the Italian Statistical Institute

MORINGA
Discover · Grow · Transform

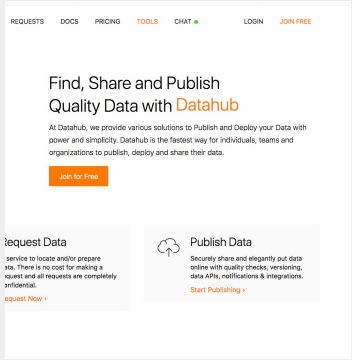
Where can I find you, data?

Data Portals

Examples of Data Portals



World Bank Open Data



DataHub

OECD Data

MORINGA
Discover · Grow · Transform

Where can I find you, data?

Open Access to Research Data

The term "*research data*" refers to quantitative and qualitative data produced, collected and/or analysed in the course of a research project.

More and more research programs encourage researchers to share not only their publications, but also their research data, in an open format.

There are several reasons behind this policy, like the fact that research is often publicly funded and so its output should be publicly available and the fact that sharing data behind research aids collaboration and facilitates progress and peer-review.

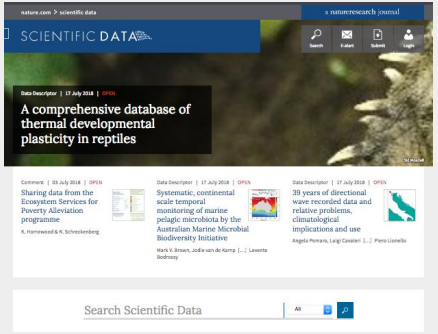
MORINGA
Discover · Grow · Transform

Where can I find you, data?

Open Access to Research Data

Examples of Open Research Data Sources

Scientific Data by Nature, a journal with Open Access research + mandates links to data repositories in each article



re3data, registry of 2000+ research data repositories

MORINGA
Learn · Grow · Transform

Where can I find you, data?

FOIA

(or other types of “Freedom of Information Access” legislations. For information on such laws in a specific country, you can browse resources at FOIA Advocates)

Many countries have laws regulating citizen’s access to information produced and held by public bodies.
The kind of laws are based on **the belief that such access, along with transparency of public institutions, is a universal civic right**.

In practice, this means any citizen can submit a requests for data, and this requests should be timely met unless there are serious obstacles to such release (like privacy concerns, national security issues, vague requests, excessively burdensome demands, etc.).

You can read more about how Freedom of information laws by country here.

MORINGA
Discover · Grow · Transform

Where can I find you, data?

FOIA

(or other types of “Freedom of Information Access” legislations. For information on such laws in a specific country, you can browse resources at FOIA Advocates)

While the process of getting data through FOIA requests can be daunting and entail long waiting times and lots of data wrangling, the results are often worth it: you are **actively contributing to your community** by opening up data that should have been public but was not accessible before.
Plus, **there can be really juicy stories behind never-published-before datasets.**

MORINGA
Discover · Grow · Transform

Where can I find you, data?

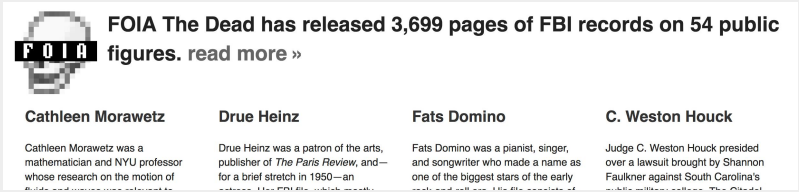
FOIA

(or other types of “Freedom of Information Access” legislations. For information on such laws in a specific country, you can browse resources at FOIA Advocates)

Examples of Stories Possible thanks to FOIA requests and similar

Vaccini, la mappa dell’Italia che rinuncia a proteggersi by Riccardo Saporiti for Wired Italy

In Cold War, U.S. Spy Agencies Used 1,000 Nazis, by Eric Lichtblau for The New York Times



FOIA The Dead by Freedom of the press Foundation

MORINGA
Discover Learn Transform

Where can I find you, data?

The Crowd

You can resort to “the crowd” for the data you need.
In practice, **this can mean several things**: from asking people to fill in surveys to collecting User Generated Content and social media data; from crowdmapping to using platforms enabling “human-in-the-loop” data sourcing and labelling.

When dealing with this type of data (as with any, actually), **always bare in mind the limitations and the biases weaved into your choice**.
Who are you asking to fill in the survey and where are you posting the request? Who is on this/that social media? Who will be inevitably excluded from my selected sample because of age, access to Internet, etc.?

MORINGA
Discover · Grow · Transform

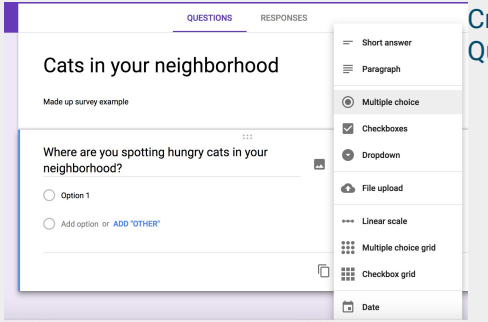
Where can I find you, data?

The Crowd

Active Collaboration

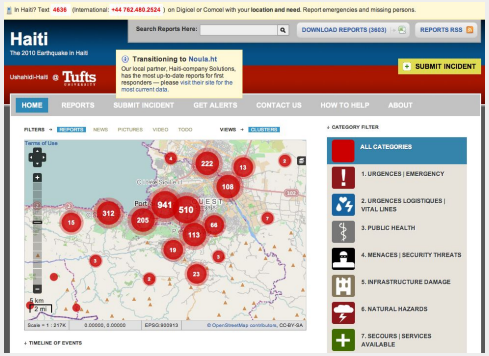
Asking users/audience to actively contribute to the production of the dataset needed

Examples of how to crowdsource data



Create a Google Form Questionnaire

Ushahidi can be used to help people collectively report and map on phenomena. Image from Ushahidi's blog post "Crisis Mapping Haiti: Some Final Reflections"



MORINGA
Discover · Grow · Transform

Some suggestions to start:

Data portals

- World Bank Open Data Portal
- United Nations Open Data
- OpenAfrica
- WikiData
- European Data Portal (data from EU countries) and EU Open Data Portal (data from EU institutions)
- OECD Data Portal
- DataHub.io
- NASA Open Data Portal
- Kaggle, training datasets for data science
- Data.World, a platform to upload, share and search for datasets + work collaboratively on projects
- OpenSpending, public spending and fiscal data
- NGOs, Universities and Research Groups sometimes also release data along with their reports and research. See the University of Maryland → Global Terrorism Database; SIPRI → data on Military Expenditure
- Google Public Data Explorer
- OpenCorporates, database on companies
- World Health Organization Data Repository
- Data portals f news organizations. (For example, the FiveThirtyEight Data Repo - the data-driven news company FiveThirtyEight shares the data behind its articles)