PHARM

PREVENTING HATE AGAINST REFUGEES AND MIGRANTS

Handbook for creating stories with data

INTRODUCTION, CASES AND TOOLS FOR DATA JOURNALISM

November 2021

This project is funded by the European Union's Rights, Equality and Citizenship Programme (2014-2020). REC-RRAC-RACI-AG-2019. Grant Agreement number 875217

http://pharmproject.usal.es

This report constitutes Deliverable 24, for Work Package 4 of the PHARM project.

November 2021

© 2021 – PHARM, Preventing Hate Against Refugees and Migrants, – GA number 875217.

General contact: pharm.project.eu@gmail.com

For more information type the e-mail address of the corresponding author

Information may be quoted provided the source is stated accurately and clearly.

This publication is also available via https://pharmproject.usal.es/

This project is funded by the European Union's Rights, Equality and Citizenship Programme (2014-2020). REC-RRAC-RACI-AG-2019. Grant Agreement number 875217

The content of this report represents the views of the author only and is his/her sole responsibility. The European Commission does not accept any responsibility for use that may be made of the information it

PHARM
PREVENTING HATE AGAINST REFUGEES AND MIGRANTS

VERSION CONTROL SHEET

Deliverable number + title	D24. Handbook for creating stories with data.
Prepared by	PHARM Team
Work package number	4
Work package leader	Sergio Splendore
Dissemination level (PU, CO)	PU
Delivery date	Noviembre 2021
Submission date to EC	Noviembre 2021
Main authors	Datasketch
Reviewers	Carlos Arcila Calderón, Martín Oller, Patricia Sánchez Holgado, David Blanco Herrero

Handbook

CREATING STORIES WITH DATA

INTRODUCTION, CASES AND TOOLS FOR DATA JOURNALISM

VNiVERSiDAD D SALAMANCA
CAMPUS DE EXCELENCIA INTERNACIONAL

ARISTOTLE UNIVERSITY OF THESSALONIKI

UNIVERSITÀ DEGLI STUDI DI MILANO

Funded by the European Union's Rights, Equality and Citizenship Programme (2014-2020)

PHARM
PREVENTING HATE AGAINST REFUGEES AND MIGRANTS

This project is funded by the European Union's Rights, Equality and Citizenship Programme (2014-2020). REC-RRAC-RACI-AG-2019. Grant Agreement number 875217.The content of this report represents the views of the author only and is his/her sole responsibility. The European Commission does not accept any responsibility for use that may be made of the information it contains.

PHARM

PREVENTING HATE AGAINST REFUGEES AND MIGRANTS

HANDBOOK FOR CREATING STORIES WITH DATA.

INTRODUCTION, CASES AND TOOLS FOR DATA JOURNALISM

November 2021

Elaborated by Datasketch https://datasketch.co for the PHARM project https://pharmproject.usal.es/

All contents of this manual are licensed under a Creative Commons Attribution-Noncommercial, Share Alike license. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

VNiVERSiDAD D SALAMANCA
CAMPUS DE EXCELENCIA INTERNACIONAL

ARISTOTLE UNIVERSITY OF THESSALONIKI

UNIVERSITÀ DEGLI STUDI DI MILANO

Funded by the European Union's Rights, Equality and Citizenship Programme (2014-2020)

This project is funded by the European Union's Rights, Equality and Citizenship Programme (2014-2020). REC-RRAC-RACI-AG-2019. Grant Agreement number 875217.The content of this report represents the views of the author only and is his/her sole responsibility. The European Commission does not accept any responsibility for use that may be made of the information it contains.

Introduction to data journalism.

If we want to consider how to tell stories with data, we should approach the discipline of data journalism and its importance in communicating information. Data journalism uses data and statistics to share relevant information in a specific context, using visualizations and graphic tools.

One of the reasons why data journalism has gained so much relevance in recent years is because, through the Internet, content users have increased significantly. In addition, online communication has opened up the possibility of creating novel and engaging content.

Data journalists constantly combine data analysis with their ability to narrate and tell stories. That allows them to offer a rigorous analysis of the information and a refined study that makes the data understandable and selects the most relevant.

In order to practice data journalism, it is essential to have a series of tools to tell stories more efficiently. Some of them are technologies for analyzing large and small databases, generating calculations from the information, and making data visualizations that accompany the story. In addition, data journalists should be knowledgeable about:

- Accessibility and information requesting: this includes the ability to access public databases or databases of any kind.
- Data cleaning: in many cases, databases are not tidy and clean and should be prepared in the best way.
- Verification of information: databases do not have the absolute truth, and the journalist's rigor requires contrasting this data with other sources of information.
- Visualization and communication with data: determine how the information will be attractive, precise, and clear. It requires the use of design and visualization tools.

There are different ways to reflect data-based projects from a journalistic point of view, with products that can complement each other such as

- Data-driven journalistic articles focused on very particular topics from a much broader and general issue. They can be used, for example, to publicize a city's crime rate or the schools with the best performance in international education rankings.

- Interactive visualizations focus on simple or more complex graphics, which allow understanding databases based on cross-references of information that usually complement a news item or a story.
- Interactive applications and visualization dashboards. They include interactive visualizations where the user can choose what type of information to graph and cross-reference, according to a series of predetermined change controls.
- Downloadable datasets allow the article journalistic product user or reader to reuse the data free of charge and without access barriers.

To learn more about the uses of data journalism globally, we suggested reviewing digital portals of media such as *The New York Times*, *The Guardian*, *Financial Times*, and *El País*.

Creating stories with data

Data can be a complex subject to understand. We can think of them as information records, such as a person's age or favorite color. However, they mean nothing without providing context, analysis, or telling a story with them (Nussbaumer Knaflic, C., 2015)[1]. We find that databases are created and made available to anyone. Still, no one uses them, investigates, or takes advantage of them to reach conclusions beyond the numbers (Datos abiertos, Gobierno de España, 2019)[2].

One way to tell stories from the information in these databases is by cross-referencing data using technologies and providing narratives that allow for understanding. Generally, creating this type of story or communication includes constant collaborative work between people with different disciplinary approaches, including journalists, citizens, designers, and developers, who play important roles from their expertise (Morales Agiss, Oliver, 2021)[3].

We have historically seen how stories are told with data from an academic point of view. Proof of aforementioned is the wide range of related subjects within the curricula of the University of Salamanca, in subjects such as "Data analysis in education" in the Degree of Pedagogy, "Economic analysis of data" in the Degree of Economics or the Master's Degree in Advanced Multivariate Data Analysis.

This perspective focuses on end-users with expertise in the topics. Still, it shows a challenge to be understood by more general audiences who want to approach information

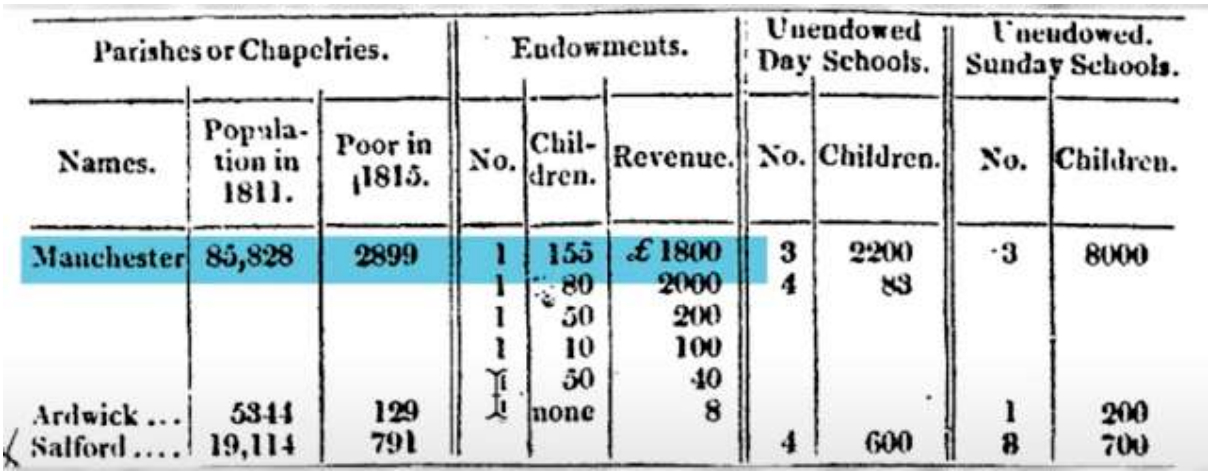
[1] Nussbaumer Knaflic, C. (2015). Storytelling with data (C. N. Knaflic, Ed.). John Wiley & Sons.

[2] Datos.gob.es (2019). 3 retos de la reutilización de la información en el sector público y cómo se pueden superar. Available in : https://datos.gob.es/es/noticia/3-retos-de-la-reutilizacion-de-la-informacion-en-el-sector-publico-y-como-se-pueden-superar

[3] Morales Agiss, Oliver (February 25, 2021). *Multidisciplinary universes. Telling stories with data*. Datasketch. https://www.datasketch.co/blog/data-culture/multidisciplinary-universes-telling-stories-with-data-by-oliver-morales-agiss/

simply and attractively. In this sense, we will address the data stories in this course that have an artistic and visual component, which invites anyone to get to know the project developed.

This module approaches the discipline of data journalism, which could be defined as those actions of data collection and analysis that seek to communicate information in the most understandable way and impact citizens through data visualizations and human narratives[4]. To begin with an example, the British newspaper The Guardian published in 1821[5] a table that counted the number of children going to school in Manchester, the cost of these students, and compared the rates of child poverty and schooling. It is considered the first piece of data journalism in the media and opened the debate on access to education in the country.



Parishes or Chapelries.			Endowments.			Unendowed Day Schools.		Unendowed. Sunday Schools.	
Names.	Population in 1811.	Poor in 1815.	No.	Children.	Revenue.	No.	Children.	No.	Children.
Manchester	85,828	2899	1	155	£1800	3	2200	3	8000
			1	80	2000	4	83		
			1	50	200				
			1	10	100				
			1	50	40				
Ardwick ...	5344	129		none	8			1	200
Salford	19,114	791				4	600	8	700

Source: https://image.guardian.co.uk/sys-files/Guardian/documents/2011/09/25/ManchesterGuardianback.pdf

Information flow of a data-driven project

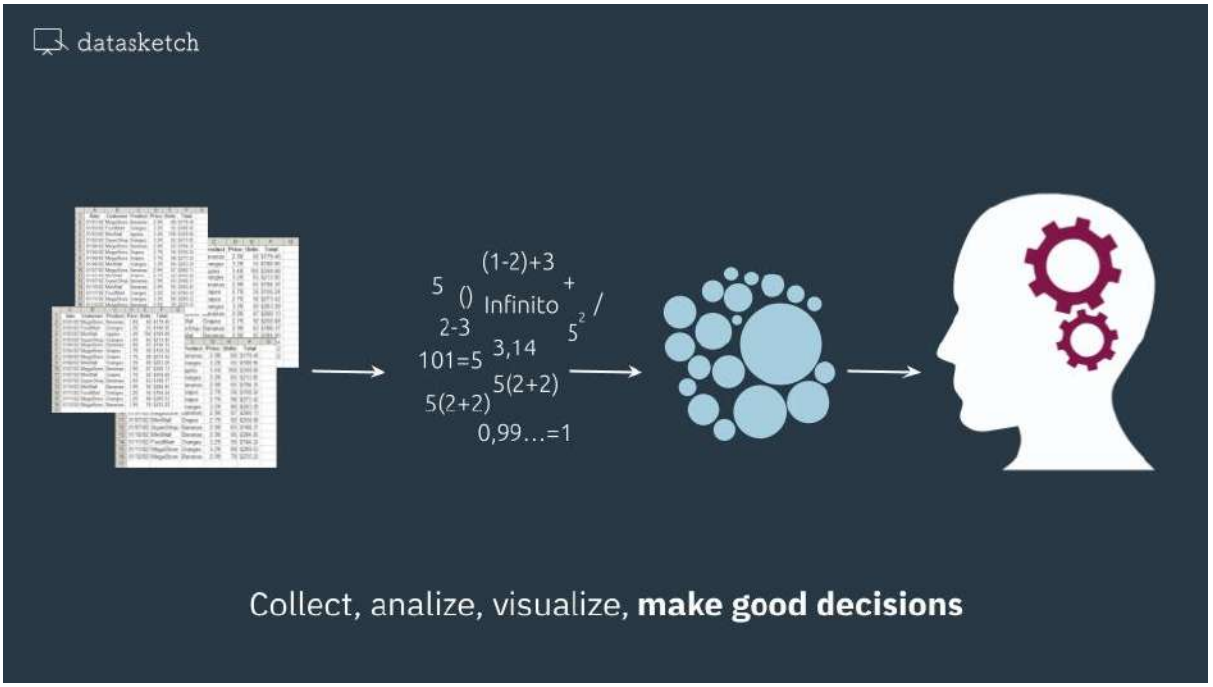
A data project usually has a flow of information that follows a particular process. We show here a five-step summary of how it works. However, sometimes this flow is not so linear or fast because, for example, the data may not be easy to collect, or you do not know the best visualizations that should be used to communicate the information assertively.

- Start with an idea or proposal on the topic to be addressed in the project. Questions should be raised about the interest of the topic, its potential impact, and the objectives of managing the information.
- Next is the collection of data. In this step, questions should be asked, and solutions sought to issues such as: where is the information? How do we access it? In what formats is it?

[4] Alberto Cairo (2020), an expert in data journalism, states in an interview that "Data never gives a complete view of a news story. We need the human side.

[5] The first Guardian data journalism: May 5, 1821 | News | theguardian.com

- Follow with data analysis. Here is where you question the data and begin to identify the best ways to communicate the information.
- It continues with data visualization. Here, you choose the best ways to graph the data accompanying a product, note or article, so that the recipients can make good decisions.



When carrying out a project of these characteristics, many barriers and challenges can be encountered, such as access to data or the need for data that do not exist, are not collected, or are only partially available.

An example of this is often data on violence against LGBTI+ people because, although there is progress (ILDA, 2021)[6], there is no widely accepted methodology for its collection, if it is done at all. Also, various sources of information can be found with different data or highly complex information. Data can be found in "good formats" but have poor quality, or it can be relevant and exciting data but open in "bad formats" (Tim Berners-Lee, 2009)[7].

Therefore, when telling a story, we must take multiple factors into account. We can tell stories aimed at large and small audiences. That is reflected in the selection of the graphic components that will accompany our narratives. Sometimes, what is behind the numbers is much more important than the numbers themselves. For example, behind databases on evictions or terrorism, some people have been victims or who have been evicted. In these cases, a component of humanization is necessary for which a record becomes part of the story.

[6] https://idatosabiertos.org/wp-content/uploads/2021/08/9agostovisibilidadLGBTI.pdf
[7] https://5stardata.info/es/

Proposed idea or topic to be addressed.

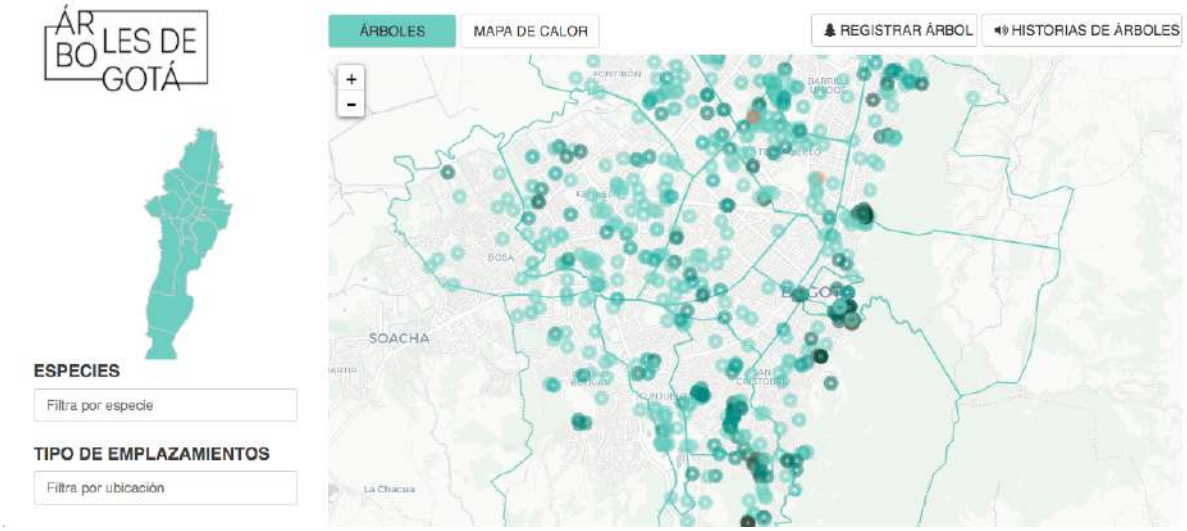
There are many ways to extract information from data sources, so it is essential to list issues of interest. To do this, we can identify the problems and the first data related to them. From there, generate a first quick analysis based on the initial impressions of the data.

Once we have defined the problem, we should develop an extensive list of questions with possible cross-referencing of variables and other valuable sources of information, even if we do not have access to them at the moment. It is essential to recognize that there may be patterns and trends in the data sets relevant to our clients.

In this first stage, the information analysis should be considered an "interview" of the data. It includes knowing the context of the issue, the information available, the entity or organization that collects it, how it is stored, and what information is needed to complete the analysis.

Let's look at an example. If we have a database on the trees in a city, we can ask ourselves how many trees there are in total or specific sectors. In this case, the answers can probably be found in the same database. We could also ask how many trees per inhabitant or fewer trees are in the city's poorest neighborhoods. This information should be sought using other sources of information: we will need to find the number of inhabitants of the city and the socioeconomic classification of each neighborhood.

It is much easier to determine the type of information to search for, collect, filter, and cross-check with these initial questions. In turn, it will help us identify if there are patterns in the data and focus the research.



Source: Datasketch. Trees of Bogota.

We suggest creating a document containing all the project information, defining dates and people in charge of each task to carry out these actions. These may include requesting information, other inquiries, cleaning and organizing databases, writing texts, data visualizations or graphics, web design, web development, launching, and promoting the project.

Collecting information

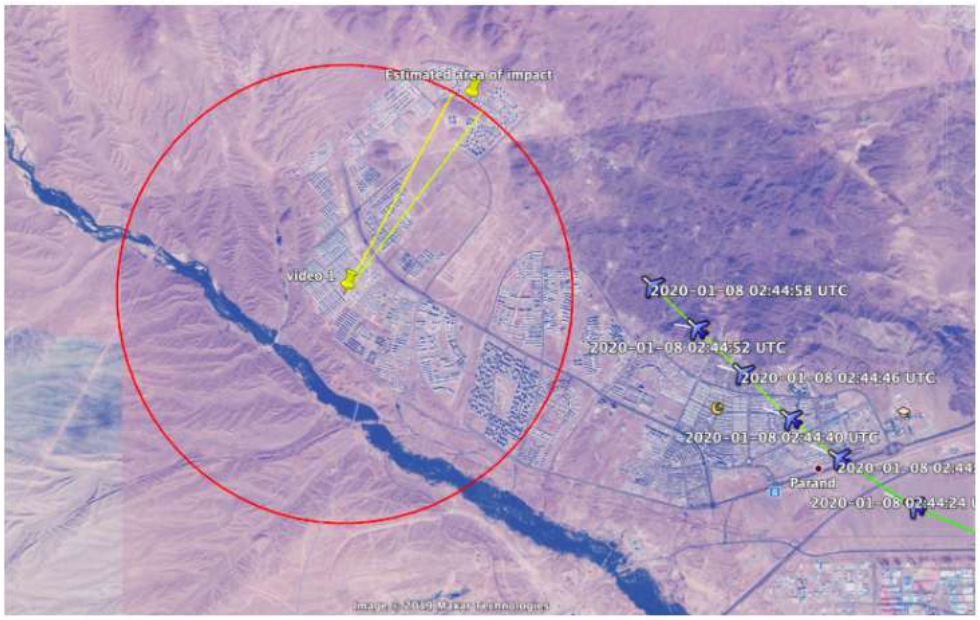
The collection of data can be done from different sources. One of the most common is using public data that government agencies compile and make available to citizens in compliance with their fundamental right to access public information. Some countries have legislation on the obligation to publish information of interest to citizens[8] and regulate requests for access to information[9].

Before resorting to requests for information to public bodies, it is advisable to carry out an initial investigation and documentation since it is likely that the required data is open or can be located on the Internet. We should use the citizen petition as a last resort, and, in any case, it should be formal, precise, and specific about the information requested and the formats required. It is also recommended to mention the relevant access to information law and request data limited to one particular period.

If the information cannot be obtained through this request, there are other ways to collect it. It can come to us through information leaks (as in the case of the Pandora Papers), which are a helpful tool but may have specific interests behind them. You can also initiate legal actions in national or international jurisdiction to make the information public, do your information gathering, or, in most cases, simply work with the available information.

The use of public data has generated trends in information gathering and communication. One of them is Open Source Intelligence (OSINT), which uses open sources to reveal information in real-time. In early 2020, Bellingcat, a British media outlet, used this information-gathering mechanism to tell the story about the authenticity of a plane crash in Iran, using satellite imagery, video, and metadata to verify the information.

[8] BOE-A-2013-12887 Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno.
[9] BOE-A-2001-21090 Ley Orgánica 4/2001, de 12 de noviembre, reguladora del Derecho de Petición.



Source: Bellingcat. Video Apparently Showing Flight PS752 Missile Strike Geolocated to Iranian Suburb.

Researching with existing data based on new approaches that generate new knowledge is essential for information work. In Colombia, the organization Datasketch (2020) developed a consolidated database on murdered social leaders from information available from various official, journalistic and international organization sources. That, in turn, allowed for its reuse and a unique feature to narrate the information discovered from the data analysis. Storytelling with data involves collaboration between different organizations or individuals who contribute from their own experience.

When public and open sources of information are not available, an alternative is to create your databases from other information that is available or can be accessed. That can be done from news generated by other media or from information provided by primary information sources, for example, from interviews, surveys, and forms.

In addition, when it is necessary to go beyond a single source of information, other sources can complement what is provided in the story. For example, if you are writing a story about the arrival of ships carrying food or commodities in a country, you can use official maritime import databases.

Suppose you do not have complete confidence in the official information. In that case, it is essential to resort to other non-official sources, such as sources from third sector organizations, the private sector, or international agencies.

Data analysis and organization

The information collected has to be structured in a relevant way. It is essential to have an extensive list of questions that we want to solve from the available databases and existing sources of information. That will make it possible to determine the best ways to organize the information and, if necessary, to obtain more.

Data structuring and subsequent analysis must be considered in what formats the data sources store and open the data. It can be in formats such as excel or CSV, which will facilitate the journalist's work, but it can also be offered as an image or pdf, which will require a different treatment. Those above will make it possible to extract and convert the information into a suitable and user-friendly format for the technological tools that will help analyze the data, which must be clean and orderly.

Once the information is in the correct format, it must be sorted. In this step, it is essential to identify the variables that make up the information.

Let's look at an example: Suppose we have downloaded a database on exports, and this is the information found in it.

Product	**Banana**	**Coffee**	**Rice**	**Apple**
Tons	120	200	82	90
Destination	USA	UK	France	Belgium

Three variables are located here: product, tons and destination. We know they are because they are the ones that give meaning to the information. In this example, the variables are easy to locate, but they are not organized correctly. It is at this point that we will have to perform a data cleaning (tidy data). It will be done through a logical structure in which each attribute (variable) in the database must be a column and each observation (a record) must be a row.

Thus, the table in the example should be organized as follows.

Product	**Tons**	**Destination**
Banana	120	USA
Coffee	200	UK

Rice	82	France
Apple	90	Belgium

The categories are located at the beginning of the column, are not modified, and serve as a guide to find a specific type of information. It will allow generating filters, analyzing them better, and visualizing them correctly. If we had to generate it from scratch instead of working with an already created database, we had to develop it from scratch. We should also follow this logic, which applies to both simple and complex cases.

After organizing the information, the values should be standardized. The data collected in each of the table rows should have the same format as the rest. For example, if you have a variable with the gender of people, you must decide how the information will be presented.

You cannot put in one row the letter F, in another the word "Male," and in another "Female." Another case is when dealing with weight units. We cannot use "Kg" and "kilograms." We shall choose only one option. Although they mean the same thing, for a computer, they are different variables.

Identifying problems with the available data is one of the fundamental tasks when processing information. It includes incomplete and missing information in the analysis, erroneous records of the information (which may be human typing errors), and data that may be inconsistent with the reality of each subject. These problems are fundamentally solved when the context of the data and its structuring is understood. Some considerations that should be taken into account are:

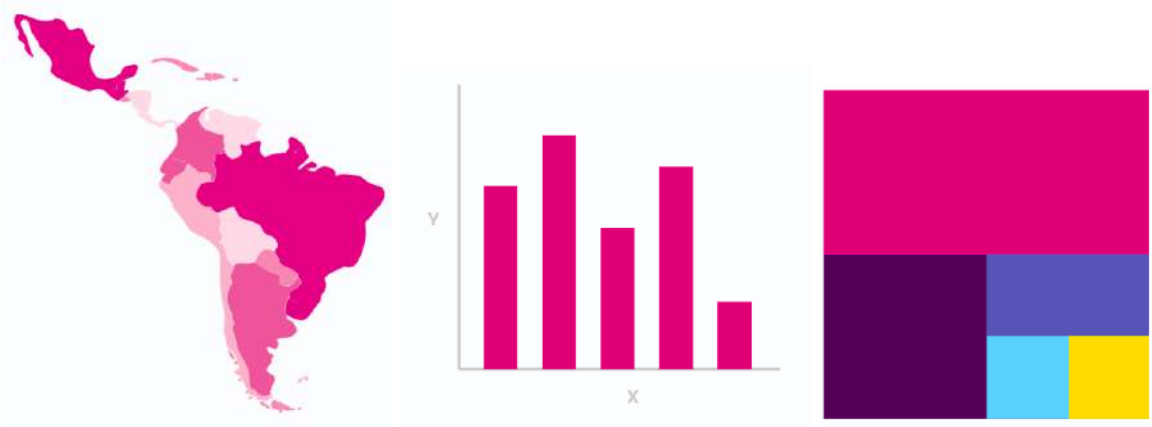
- The column headings must occupy a single row.
- Each variable must be in a single column.
- Each column must have the same data type.
- Each data record is a row.
- There should be no hidden variables in the cells.

Once the data are clean and sorted, the most appropriate analysis methodology should be selected. This step will depend on the questions to be answered and the type of research desired.

Data visualization

After the information structuring and analysis, it is necessary to clearly define what is to be communicated with the information and the visual guides or data visualizations that will accompany the story from the analysis results.

Data visualizations are encodings of information designed to be understood by the human eye. For example, if you have a database of populations by cities in a country. In that case, it is possible to graph the information in different ways, communicating the concentration of people by geographic location or showing in detail the number of people with a bar chart. It is also possible to compare the density of people in cities with a tree diagram (treemap).



That is useful when providing sizes, shapes, inclinations, distances, colors, textures, and movements, transforming the information into visual forms to accompany our story. Then, it is necessary to explore the types of existing visualizations and determine the advantages and disadvantages at the moment of telling the information.

There are different ways of communicating, depending on the audience. Some will be simpler, and others more complex. If the message must reach a broad and not very specialized audience, we can offer simple graphics.

On the contrary, the graphics should be more explanatory and focused on particular issues when the audiences are more specific.

Defining the type of information to be communicated is fundamental. The same data can be shown in different ways and to other people, reaching a similar communication but with another degree of depth. When we talk about migration, we should not offer anyone who works in a refugee reception center the same information. Their knowledge base on the subject is different, and their understanding of the issue may be further.

When thinking about how to display data, do not forget everyday references that do not necessarily count as data visualizations, such as the subway routes of a city.

The Black Cultural Archives organization developed a subway map of black history in London, based on black people's contribution to the history of the town and the United Kingdom. It is information generated with data created from scratch and was communicated with widely known graphic cues.

Data visualizations can also be thought of from more analogical points of view. This is the case of these T-shirts with different designs based on data covering fundamental issues of reality, such as gender, environment, human rights, etc. Here, the aim is to communicate complex information focused on significant audiences.

How to select visualization types

There are multiple ways to visualize data, more than 200 types. In this great universe of options, we can find numerous possibilities that can be difficult to navigate. Fortunately, different tools allow you to know the different types of visualization and what each is for. Here we recommend some tools. All of them will enable you to search and find visualization types and know some examples of their use cases.

- Datasketch - Types of visualization: It allows you to explore different forms of visualization https://www.datasketch.co/es/tipos-visualizacion/ and their respective variations, with specific examples https://www.datasketch.co/es/pseudoviz/.
- Dataviz Project is a design studio tool that helps you visualize and learn about applications of different types of graphics https://datavizproject.com/.
- Data Viz Catalogue allows knowing about the use of various graphics, describing their visual and functional components https://datavizcatalogue.com/ES.

It is crucial to consider the objective when choosing the most appropriate visualization and having the correct data to make the graphic.

Some online charting tools

To make data visualizations, we recommend some tools:

- Infogram: allows you to create simple charts with a good design component. It may have some limitations when loading data files with multiple variables or a larger

volume. It works very well for infographic exercises. Available at https://infogram.com/.

- Datawrapper: is a fairly intuitive and easy-to-use interactive graphing tool. It has a very flexible interface for loading and customizing graphics. It is widely used by different media around the world and has many interesting educational resources. Available at https://datawrapper.de.

- Tableau: is a very versatile tool for creating dashboards or interactive control panels. For this reason, it has a very high learning curve and it is necessary to invest in learning several data visualization concepts to use it well, for example mapping different types of variables, coordinates, colors and other visual cues. Available at http://tableau.com.

- Datasketch: is a tool with multiple applications that help the different processes of access, analysis and visualization of information. Unlike other platforms, Datasketch can also explore open data from different entities uploaded by other users. It also has visualization applications, tools for cleaning and cross-referencing databases, text analysis, among others. Available at http://datasketch.co[10].

Graphic concept

In addition to data visualizations, data-driven stories have the potential to convey the best way to communicate the story visually in a way that brings the proper meaning and is eye-catching to the audience.

The first step in thinking through the graphic concept of the story is to get inspiration. Creativity is, in essence, the ability to make novel connections between concepts or ideas. That can come in many forms: from contrasting colors seen while walking through a park to a striking image on the Internet. Each of these ways of relating to the world implies a possible way of determining the graphic concept.

The next step is to define the visual concept itself. For this, it is a good idea to collect and organize the ideas that arise from the inspiration process, using digital tools that allow you to group images according to some concept or do it analogously in inspiration boards, better known as mood boards. These images or ideas must not be limited to being referents of our theme but are somehow related. It will allow us to expand the possibilities of our graphic concept.

[10] Existen múltiples herramientas para facilitar el trabajo con datos, Datasketch cuenta con una colección por diferentes casos de uso como el análisis geográfico, herramientas de web scraping para estructurar información no estructurada, herramientas para procesamiento de textos, entre otras.

Next, we must create the color palettes that can be used for the visualizations. For this, it is vital to take into account three aspects. The first is to choose the leading colors according to the theme of the story. Usually, when talking about a particular topic, the mind relates it to specific colors. That is the case of medical or hospital data, where blue and green predominate as representative colors. Choosing the wrong color could confuse our audience about the subject matter.

Second, create color schemes according to the data to be communicated. There are two types of color scales: categorical and continuous. Absolute palettes are used to denote names of things: months, car brands, names of countries, names of political parties, etc. Continuous scales help display numerical data, such as a country's population or the percentage of Republicans or Democrats in a given area of the United States.

Finally, consider the accessibility of the color palettes chosen, as our story should be equally exciting and understandable to people with different forms of color blindness or limited vision. Thus, it is essential to achieve an adequate contrast between the different colors and the background.

There are tools such as WAVE Web Accessibility Evaluation Tool, Vischeck, or the W3C's guidelines, which will help identify the accessibility level of your images and projects. In turn, there are others in which it is possible to learn how to create image color palettes and generate your palettes, such as Paletton, Khroma, or Image Palette.

Writing and editing the story.

To tell the story and provide it with a humanistic approach, we can turn to experts or to people directly related to the story being told. Those people who "live" the data records.

A legal validation of the information and assertions made in the report should also be carried out. If the story describes a conflict, whether legal, social, economic, political, or environmental, the counterpart, who is the actor generating the conflict or problem, should be consulted.

Always check that all the information contained in the story is accurate. It is essential to edit the texts and make spelling and writing revisions and final adjustments to the web design to match the texts and visualizations.

For text editing, we should answer the following questions:

- Was it verified that all information stated in the document is verifiable?

- Was the spelling and wording of each sentence of the text checked? Is the title of the story related to what is to be communicated and the text's content? Does the title invite you to read it?
- Is the introduction striking and in line with the remaining content of the text? Does it contain the necessary elements for the reader to learn about the central aspect of the story?
- Does the text have adequate subheadings to provide the reader with a fair outline of the content?
- If there are references in the text to other texts, interviews, or other sources of information, are they well-cited, and are references to those who carried out these projects correct?

- Are the databases used public and in appropriate formats for downloading and reuse?

In addition, it is essential to add the collaborators of the project (sponsors, illustrations, design, editing, among others).

It is relevant to add the methodology for data collection, analysis, and creating the visualizations. The step should also include the databases used in the project: remember that raw data is always more relevant for those who want to reuse them or tell new stories with the data. If you tell stories, you should also open up the information and allow readers to make their inquiries.

Publish the story

The next step is to compile all parts of the project into a single website through web development to build and realize what has already been planned. It consists of creating the websites with which the texts and the project's design will be accessible. It also makes the functionalities through an interface for the use of the users. There are many ways to implement a website. It can be through programming languages or programs for easy publishing on the web, such as Github Pages or WordPress.

Tips and tricks for information visualization

1. **Choose the right graphs and check that they really work to communicate your message.** Choosing the most appropriate graph to represent the available data can be complex, as it implies knowing the different types of graphs available and being aware of the data we have, but it is essential if we want readers to understand the information we want to convey without any misunderstanding. An outstanding

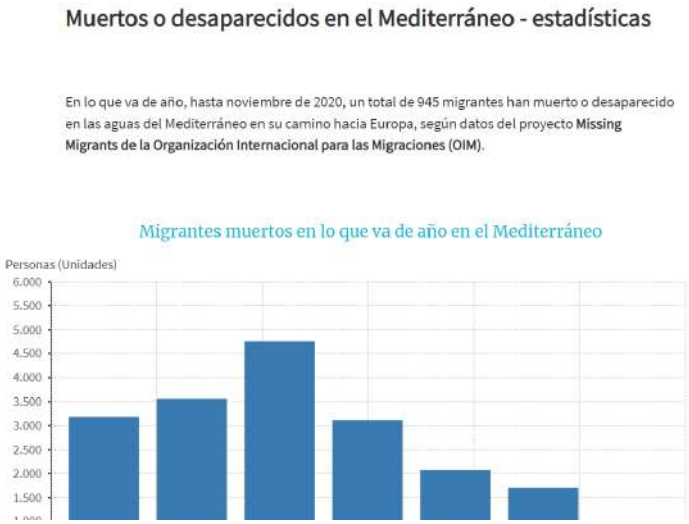
example of how to successfully use data and visualizations with a narrative line is "Venezuela on the run", a multimedia report by the Data Unit of the newspaper *El Tiempo* and the digital media Efecto Cocuyo, which won the Gabo Award in the Coverage category in 2018.

The use of data in the platform *A Welcoming Europe? Asylum applications in the EU* is also interesting. There, it provides information on the number of refugees, asylum seekers and the population of countries, as well as the countries with the most asylum applications and the evolution of these applications.

2. **Explain the scale you are going to use and use it correctly.** It is necessary to use visual aids to illustrate the news, but when the design and numbers do not match you can give the wrong message and lose credibility in the media and in the data. Alberto Cairo (2019) points out that the central element of most graphics is their visual coding. This depends, among other things, on the length, height, area and angle of the graphic and the position of the points in it. It also depends on the title, subtitle, labels and textual notes, and mainly on the color. If you want to explore this further, we recommend reading this guide to expressive graphic design by Nam Wook Kim, assistant professor of computer science at Boston College, and this article by Sarah Leo, visual data journalist at The Economist.

3. **Always provide full context.** Data needs context in order to be interpreted and communicated in the right way. Graphs require an appropriate title, a caption to interpret what is in the image and/or a text, or verbal communication, to explain them. The reader must know all the information that influences the data in order to have an accurate idea of what they are reading and not a part of it. This visualization from *EpData* on migrants who have died or disappeared in the Mediterranean during their journey to Europe is a good example of a graphic with a title and description

appropriate to the data being presented.



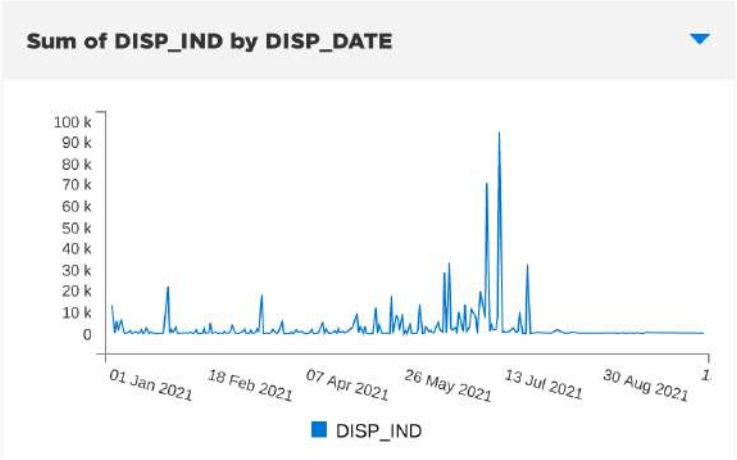
4. **Graphics and data should have a logical relation.** Making information about the pandemic appealing and presenting it in an understandable way has been a challenge faced by media and governments almost everywhere. Great work has been done, but there have also been many mistakes. One of the most repeated failures we have observed is in the design of the visualizations.

   This graph was broadcast by Telemadrid in October 2020. In this case, the data shown are the number of deaths in Madrid corresponding to July, August, September and part of October. The presenter of the program explained that the number of deaths had not stopped growing and that the trend in October was the same but, when we look at the graph, we find that this statement is not true. The accumulated number of deaths in September was 1,136 and in October 1,056, but the graph shows a higher bar in the second case, giving rise to an erroneous interpretation, like that of the presenter.

On this occasion, Telemadrid apologized, attributing the failure to human error, and made the necessary adjustments to correct the display.

5. Explain well why they exist and where the missing data are. Sometimes, the data we have does not have the values in all the variables we are considering and that causes uncategorized data to exist. This presents a challenge when it comes to displaying it graphically, which we must consider and explain to the reader.

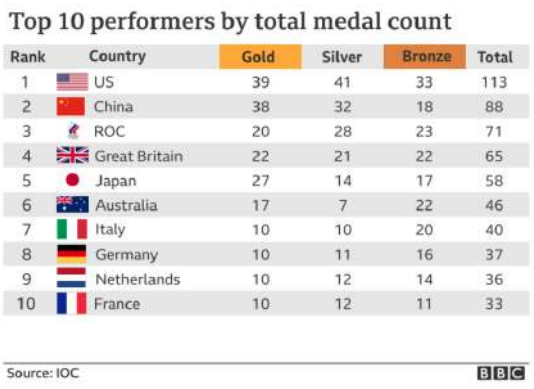
Humanitarian Data Exchange collects information on conflict-induced population displacement in Afghanistan. During the Taliban takeover of the country, they stopped publishing data and, although the page had an updated date, the data has not been updated, as can be seen below.



On the page they show a legend referring to this lack of update and indicate that it will be done as soon as they have new data, which is a good practice.

6. Respect your reader. This advice should not be necessary. Showing information at convenience, in order to prevent the recipient from perceiving a negative message, is manipulation. We cannot take readers for fools. Sooner or later they will realize it, you will lose your credibility and you will be exposed.

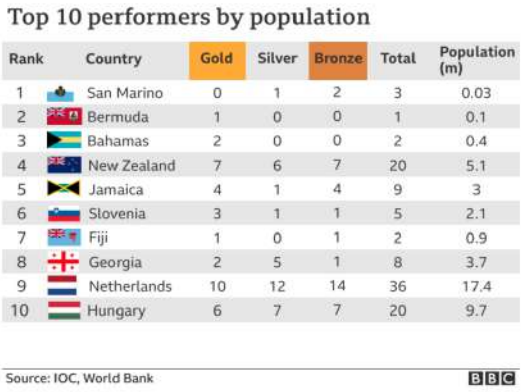
In the article "Olympic medals: An alternative table -with US 15th", published in August in BBC Canada, Robin Levinson-King proposed ranking the countries participating in the Tokyo 2020 Olympic Games by other variables in addition to the number of medals -gold, silver and bronze- won during the different competitions. He proposed the use of variables such as population or per capita income and GDP (gross domestic product) of the countries to show that any country outside the medal-winning triad (United States, China and Russia) could reach the podium if the variables that favor them were chosen.



Top 10 performers by total medal count

Rank	Country	Gold	Silver	Bronze	Total
1	US	39	41	33	113
2	China	38	32	18	88
3	ROC	20	28	23	71
4	Great Britain	22	21	22	65
5	Japan	27	14	17	58
6	Australia	17	7	22	46
7	Italy	10	10	20	40
8	Germany	10	11	16	37
9	Netherlands	10	12	14	36
10	France	10	12	11	33

Source: IOC BBC



Top 10 performers by population

Rank	Country	Gold	Silver	Bronze	Total	Population (m)
1	San Marino	0	1	2	3	0.03
2	Bermuda	1	0	0	1	0.1
3	Bahamas	2	0	0	2	0.4
4	New Zealand	7	6	7	20	5.1
5	Jamaica	4	1	4	9	3
6	Slovenia	3	1	1	5	2.1
7	Fiji	1	0	1	2	0.9
8	Georgia	2	5	1	8	3.7
9	Netherlands	10	12	14	36	17.4
10	Hungary	6	7	7	20	9.7

Source: IOC, World Bank BBC

Thus, building on Levinson-King's paper, Kaiser Fung published "Demo of how data can tell any story you want, ode to Tokyo" in which he shows how, when you have enough dimensions of a data set, you can give the maximum range to whatever you want. Fung, in his challenge to make any country participating in the Olympics number one at something, has shown that you can always find a way to be the best at something, it's just a matter of method in the selection of variables.

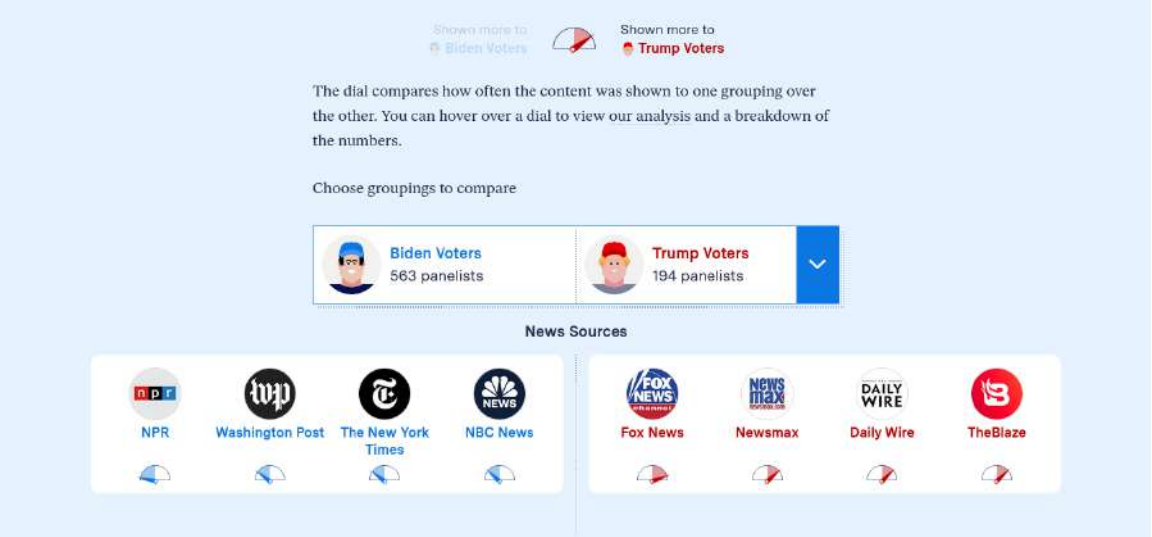
7. The graphs must be based on real data. This point is closely related to the previous one. Using graphs when they add value and serve to explain the data is fine, but using false visualizations, without data to support them because they "give credibility" to the messages we want to convey, is dangerous.

In September 2021, Abraham Weintraub, a well-known Brazilian politician and Bolsonaro's former Minister of Education, published a graph to illustrate the evolution or increase in the number of arbitrary arrests and violations of processes against conservatives.

The graph does not look good in general, does not provide sources of information, has no title and the axes show some illegible texts, with no legend to explain them. All these are red flags for a cautious reader, like the Twitter user who identified the falsity of this image. The fact is that if we look up the definition of graph in the Portuguese version of Wikipedia we find a graph used to describe the general concept of graph, which is strikingly similar to the one used by Weintraub. The difference between the two is the cut axes and the image quality, which in the politician's example makes the data unreadable.

8. Be careful with research methodologies and do not massify messages that are not relevant. One of the objectives of scientific research is that the methodologies applied always give the same results. When researching and communicating the results of a research, even a small one, the limitations of the data should be considered and reported, the methodology used should be checked and the conclusions should be really relevant.

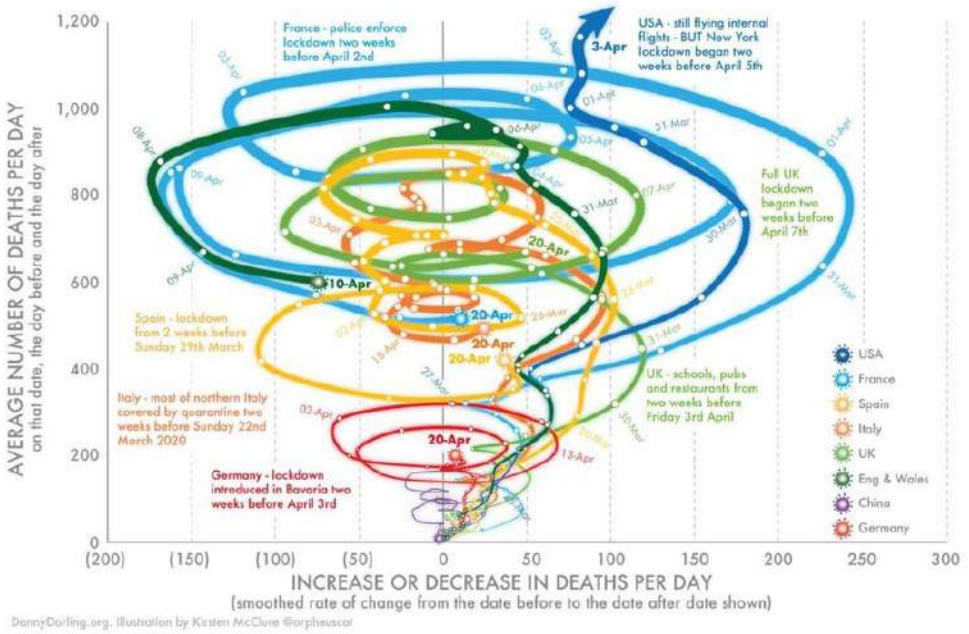
The Markup did a special report on the differences in the information Americans find on their Facebook feeds. There, they visualized how the algorithm of this social network showed people news from various media, depending on their political affiliation, gender and generational interval. Although it is a small visualization, the organization explained in detail how the information was collected, the methodology used to review the feeds and why this data is important.



This image shows the visualization created by The Markup. As you can see, it is small and simple but, before presenting it, a link is included where you can access the explanation mentioned above.

9. Do not overload charts with information. Graphs are a useful tool for communicating complex information. That does not mean that the visualization should be too. It is not advisable to oversimplify, but to clearly explain the message we hope to convey, so a good design is essential. It is better to present three good graphics than one bad one.

The image you will see below has been widely commented on in networks and some classified it among [the worst visualizations of 2020](), because it is so complicated that the design does not allow to identify the data.



It shows the number of deaths by Covid-19 per day and also the rate of change in that number for seven countries. When a line goes to the right, deaths per day increase. When it goes to the left, they decrease. The loops indicate a change in the trend. In its favor we can say that the idea is not bad and the accompanying text is quite explanatory, but the visualization itself does not meet the communicational objective. In this case, several images could have been used, perhaps one for each country, to collect the information separately and avoid overload.

10. Before publishing, communicating or replicating, you should review the information. Reviewing the information before publishing or disseminating it is a fundamental step. We recommend that you check that you have complied with the advice given in this section, that the mathematics add up and, above all, that you are not misinforming with your data.

Verifying information

Finally, although it is an essential step throughout the process of creating a news or data-driven story, the veracity of the data and its sources must be verified. If you start with erroneous information, the entire investigation will be compromised. Therefore, it is necessary to know tools and techniques to verify third-party information.

First of all, the source of information must be checked to ensure that it is clear and reliable. The references we use must offer the sources of information they use, the information must be contrasted by different sources and the data must be up to date.

It may be useful to find the origin of the images to be used as references. If the image illustrating a news item is intended to be used as a reference, you should check its origin. Google offers a service called "Reverse Image Search," which takes an image and searches for similar images to find where a photo was first published. There is also the Pinterest browser extension, which can help you with this task.

The steps to verify the data are multiple- Some of the ones to consider are:

- **Select a statement from the public domain.** It can be a figure or data that impacts these issues of collective interest. In case the original source deletes the content from the Internet, a copy of the object to be verified should be archived.
- **Consult the source of the information.** It is essential to contact the author of a statement, or the source of a piece of information that has been selected for verification. The statement may have been misquoted or quoted out of context. If the original source cannot be found or is unwilling to respond, or if it is the case of information found on a web page, you can visit WHOIS Search, Domain Name,

Who.is to check who has registration for that domain and contact the administrator directly.

- **Check the information and data in the sentence against official and reliable sources.** After verifying the source, you should contrast what he/she said with official and reliable sources, such as governmental and state entities that produce periodic reports. From here you can get data or figures to contrast.
- **Consult alternative and expert sources.** In addition to official entities, you can check the data with other reliable sources, such as respected international organizations like the UN, respected academic think tanks in your field, or a recognized individual expert in the field.
- **Put the verification in context.** In many cases, the figures or data that we verify are out of context, which can lead to confusion or misinformation.
- **Reach a verdict on the veracity of the information.** This is the step that requires the most care and must be based on all the previous steps.

Bibliography

Cairo, A (2013). *The Functional Art: An Introduction to Information Graphics and Visualization*. New Riders.

Chaparro, M. A. (2013). La evolución del periodismo de precisión: el blog de The Guardian sobre periodismo de datos. En: Universidad Internacional de la Rioja. *I Congreso Internacional de Comunicación y Sociedad Digital*. UNIR. http://reunir.unir.net/handle/123456789/1736

Crucianelli, S. (2013). ¿Qué es el periodismo de datos?. *Cuadernos de periodistas*. https://www.cuadernosdeperiodistas.com/que-es-el-periodismo-de-datos/

DataCamp. (9 de noviembre de 2017). *Principles of tidy data* [Vídeo]. https://www.youtube.com/watch?v=oQuupzfX9OQ

Datasketch. (2021). *Líderes sociales asesinados en Colombia desde enero de 2016 a junio de 2020*. Recuperado de https://www.datasketch.co/datasketch/lideres-sociales-asesinados-en-colombia-desde-enero-de-2016-a-junio-de-2020

Díez Oronoz, I. (2016). *Storytelling: transformar datos en historias* [Tesis de Máster, Universidad del País Vasco]. https://addi.ehu.es/bitstream/handle/10810/19297/TFM-IbaiDiezOronoz.pdf?sequence=1&isAllowed=y

Escuela de Periodismo UAM-El País. (2019). E*l Periodismo de datos en 10 vídeos*. Recuperado el 10 de noviembre de 2021, de https://escuela.elpais.com/el-periodismo-de-datos-en-10-lecciones/

Knight Center. (30 de septiembre de 2019). *Ejemplos del tipo de historias basadas en datos que verás en este MOOC* [Vídeo]. https://www.youtube.com/watch?v=E0NDykwGO8I

Reuser's Information Services (2021). *The Open Source Intelligence resource discovery toolkit*. Recuperado el 10 de noviembre de 2021, de https://rr.reuser.biz/

Rogers, S. (26 de septiembre de 2011). The first Guardian data journalism: May 5, 1821. *The Guardian*. Recuperado de http://www.theguardian.com/news/datablog/2011/sep/26/data-journalism-guardian

Simplilearn. (4 de diciembre de 2018). *Data Science In 5 Minutes | Data Science For Beginners | What Is Data Science? | Simplilearn* [Vídeo]. https://www.youtube.com/watch?v=X3paOmcrTjQ

Toulouse Lautrec. (10 de octubre de 2019). *Periodismo y Marketing de Contenidos* [Vídeo]. https://www.youtube.com/watch?v=4xQHsmiJvdQ

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59(10), 1-23. https://www.jstatsoft.org/article/view/v059i10

PHARM

PREVENTING HATE AGAINST REFUGEES AND MIGRANTS

VNiVERSiDAD D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

ARISTOTLE UNIVERSITY OF THESSALONIKI

UNIVERSITÀ DEGLI STUDI DI MILANO

Funded by the European Union's Rights, Equality and Citizenship Programme (2014-2020)

This project is funded by the European Union's Rights, Equality and Citizenship Programme (2014-2020). REC-RRAC-RACI-AG-2019. Grant Agreement number 875217.The content of this report represents the views of the author only and is his/her sole responsibility. The European Commission does not accept any responsibility for use that may be made of the information it contains.

PHARM

PREVENTING HATE AGAINST REFUGEES AND MIGRANTS

ABOUT PHARM (2020-2022)

Preventing Hate against Refugees and Migrants (PHARM)

Migration to Europe has grown in the last years in scale and complexity. The so called 'refugee crisis' and the migratory pressure is particularly acute in southern EU countries as the main entrance to the EU.
The main goal of PHARM project is to monitor and model hate speech against refugees and migrants in Greece, Italy and Spain in order to predict and combat hate crime and also counter its effects using cutting-edge techniques, such as data journalism and narrative persuasion. The activities distributed in 5 coordinated work packages include:
(i) Implementation of a conceptual and methodological common framework for large-scale analysis and detection of hate speech; (ii) Implementation and evaluation of machine learning approaches to model and predict hate crimes against refugees and migrants based on hate speech features; (iii) Survey journalists to understand how they inform and raise awareness about hate speech and how they can help building and disseminating counter-narratives based in data-driven news pieces; (iv) Creation, evaluation and dissemination of counter-narrative fictional stories adapted to different characteristics of citizens using large-scale narrative persuasion.

COORDINATOR

University of Salamanca, Faculty of Social Sciences (ES)

PARTNERS

University of Milan, (IT)
Aristotle University of Thessaloniki (GR)

VISIT: http://pharmproject.usal.es
CONTACT US: pharm.project.eu@gmail.com
FOLLOW US:

@Pharm_project

Pharm_project

Pharm Project

Pharm Project