



## Deliverable 7.2

### Analysis report on the use of data and open science results

#### Version 1.4

**Due date:** 31st December 2021

**Actual submission date:** 30th December 2021

**Project start date:** January 1st, 2021 - Duration: 36 months

**Work Package concerned:** WP7

**Concerned work package leader:** Marco Boscolo

**Task leader:** formicablu

**Authors:** Marco Boscolo, Giulia Bonelli, Elisabetta Tola

**Contributors:** all partners

Dissemination level:

- PU: X
- CO:
- CL:



## REVISION HISTORY

Revision	date	Contributor	Description
v1.0	07.2021-09.2021	Marco Boscolo, Giulia Bonelli, Elisabetta Tola, formicablu	Selection of sample to analyse - discussion on the structure of the analysis
v1.1	01.12.21	formicablu	Internal revision of the first draft
v1.2	22.12.21	ENJOI consortium	Sharing of the draft for comments
v1.3	29.12.21	formicablu	Revision of comments and final formatting
v1.4	30.12.21	formicablu	The final format and upload

## QUALITY ASSURANCE

To ensure the quality and correctness of this deliverable, we arranged an internal review and validation process. The deliverable was drafted by the work package leader (Marco Boscolo, formicablu). All partners contributed and reviewed the overall draft. Before submission, the formicablu team made a final review and validation.

### STATEMENT OF ORIGINALITY AND DISCLAIMER

This deliverable contains original, unpublished work except where clearly indicated otherwise. It builds upon the experience of the team and related work published on this topic. Acknowledgement of previously published material and others' work has been made through appropriate citation, quotation, or both.

The views and opinions expressed in this publication are the authors' sole responsibility and do not necessarily reflect the views of the European Commission and the Research Executive Agency (REA), that are not responsible for any use that may be made of the information here contained.



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication  
 This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)





## TABLE OF CONTENTS

1.	Summary	3
2.	Project overview	6
3.	What is Data Journalism	7
3.1	A cross-border perspective on data journalism	10
4.	Methodology	12
5.	Analysis of data journalism case studies	14
6.	Analysis of data journalism case studies on Covid-19	39
7.	Conclusions	48
8.	Further references	52



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)

# 1. SUMMARY

As the ENJOI project was moving its first steps, at the beginning of 2021, the whole world was entering the second year of the Covid-19 pandemic. We all were waiting for solutions, mitigations, and other measures to counteract the deadly SARS-CoV-2 virus that has brought unprecedented distress to all countries. All around the world, media outlets were packed with infographics and data visualisations. **Many more journalists than ever before were reporting on the pandemic data.** At the same time, more and more people in general, both with and without any real expertise, were commenting on topics like how to "flatten the curve", debating on the quality of the data collected and disseminated by the health authorities and so on.

This experience we collectively had during the pandemic is the most striking reminder of **the importance of addressing the quality and the excellence of using data in journalism.** Data journalism arose to public attention at the end of the 2000s, and it seemed a promise of objectivity and fact-baseness in the field of journalism. Nevertheless, it started as a small niche: few journalists had the skills and knowledge necessary to use data into journalistic practice fully. At the same time, data has become more and more a precious asset. In 2014 an article in the American edition of the Wired Magazine titled: "Data is the new oil of the digital economy."<sup>1</sup>.

Since then, the use of data as a bearer of objective value has been questioned and criticised.<sup>2</sup> What is undeniable is that **never before in this era, such an amount of data has been within reach of journalists' hands.** With the data came an array of skills, crafts and - above all - accessible digital tools for their manipulation. As a result, data journalism became a more and more widespread practice in a growing number of media outlets in different countries.

The ever-increasing use and manipulation of data in the journalistic practice went parallel with the availability of more and more data produced by scientific research. Cultural pushes like those coming from movements such as Open

---

<sup>1</sup> J. Toonders, 2014. [Data is the New Oil of the Digital Economy](#), Wired.com

<sup>2</sup> J. Gray, L. Bounegru *The Data Journalism Handbook*, Amsterdam University Press, 2021, p. 12-13.



Science or the Public Library Of Science (PLOS), to name a few, contributed to the availability of scientific data to the general public and journalists. However, as the Covid-19 coverage has demonstrated, the **availability of data and the possibility to use them to find stories to fill science pages in daily newspapers or slots in radio and TV programs do not necessarily bring the ability to use them properly**. The risk of misunderstanding, be it on purpose or by chance, is part of the more significant problem of misinformation.

The deluge of data we are exposed to due to the pandemic is arguably the most significant one during the entire history of journalism. And, we are now more aware than ever that numbers, data, and graphs are often at the basis of public decision-making. That is why we should not take data lightly without exerting classical journalistic critical thinking.

Alberto Cairo, a visual data designer and lecturer at the School of Communication of the University of Miami (Florida), wrote an article in 2014 titled *Data journalism needs to up its own standards*. Cairo's piece analyses the usage of data made by American outlets FiveThirtyFive and Vox. At that time, critics started to question the method, and Cairo stated that they (and the data journalist community at large) "need to treat their data with more scientific rigour."<sup>3</sup>

This Analysis Report on Data Journalism from the ENJOI project does not aim to tackle such a broad problem. Here we will focus on analysing some examples of data journalism that, for different reasons, can be helpful to **understand how the use of data can specifically contribute to excellent journalism**. We selected pieces with diverse formats coming from various languages, even though this report does not aim to cover the entire production of data journalism worldwide. The report focuses on 15 examples that the community of data journalists itself regarded as outstanding ones. **Special attention has been given to some aspects of the use of the data in the construction of the stories**: where do the data come from; whether they are from openly available sources; how clear the source of the data is made to the reader; how transparent is the manipulation of the data performed by the journalists; the method used to get from the data to

---

<sup>3</sup>A. Cairo, [Data journalism needs to up its own standards](https://niemanlab.org/2014/07/09/data-journalism-needs-to-up-its-own-standards/), Niemanlab.org, 9th July 2014.

the story. We have also been particularly careful to evaluate the efficacy in visualisation and storytelling of the data in each story.

The result is a repertoire of practices provided here as input data for further reflection and research within the ENJOI project. These insights might also prove helpful for researchers and specialists outside the consortium as part of a more extensive analysis of what is working and what is not in today's journalism. To this repertoire, we added a special section on data stories related to the Covid-19 pandemic because "as the COVID-19 pandemic continues to roll on around the world, numbers, graphs and rankings have become widely shared, thematised, politicised and depoliticised."<sup>4</sup>.

This Analysis Report on Data Journalism follows a Digital Engagement Focus Report realised in the summer of 2021 within the ENJOI project (WP7, deliverable 7.1). We tracked 15 practical experiences testing **engagement** as a key innovation asset in science journalism and communication in Europe in this previous work.

Overall, within WP7, we explore the existing media landscape in science journalism and communication through the prism of **engagement, data, general innovation, and solutions**.

In collaboration with the other WPs, our long-term goal is to embed these concepts in science journalism and communication and create a solid theoretical framework around the future **ENJOI Observatory for an Outstanding Open Science Communication**.

---

<sup>4</sup>J. Gray, L. Bounegru *Data Journalism in Question* in J. Gray, L. Bounegru (eds) *The Data Journalism Handbook. Towards a Critical Data Practice* Amsterdam University Press 2021.



## 2. PROJECT OVERVIEW

ENJOI (ENgagement and JOurnalism Innovation for Outstanding Open Science Communication) will explore and test **engagement** as a key asset of innovation in science communication distributed via media platforms, with a strong focus on journalism.

Through a combination of methodologies and in collaboration with producers, target users and stakeholders of science communication, ENJOI will co-create and select a set of **standards, principles and indicators** (SPIs) condensed to a **Manifesto for an Outstanding Open Science Communication**. ENJOI will deploy a series of actions via **Engagement Workshops, Labs, field and participatory research, evaluation and testing** phases.

It will also build an **Observatory** as its landmark product to make all results and outputs available to foster capacity building and collaboration of all actors in the field. ENJOI will work in four countries: Belgium, Italy, Portugal and Spain, taking into account different cultural contexts.

ENJOI's ultimate goal is to **improve science communication** by making it more consistently reliable, truthful, open and engaging. Contextually, ENJOI will contribute to the **active development of critical thinking, digital awareness and media literacy** of all actors involved in the process.



### 3. WHAT IS DATA JOURNALISM

Before diving into the analysis of the selected examples of data-driven journalism, a brief introduction to its history and practice is provided.

The practice of data-driven journalism as we know it today emerged in the 2000s, primarily in the United Kingdom and in the United States. The decade saw several factors that helped this new form of journalism move its first steps in mainstream journalism. First, there was an unprecedented surge in the availability of data through the Internet: vast amounts of data, both publicly produced (for example, from public institutions, such as National Institutes of Statistics), became accessible to every citizen accessing to the Web. Second, an international movement for "data liberation" emerged. It was not an organic phenomenon but rather the development of a civic sensibility from several different perspectives, such as universities, fostering the idea that data produced through public money must be publicly available and easily accessible to all. This motion took different forms: a growing open data culture within the public sector in several countries in Europe and beyond and the implementation and enforcement of specific laws on transparency. Third, the field of journalism started a transformation process that has not yet reached stability but implicated more and more digital products and fewer and fewer printed outlets. That coupled with the fourth factor, i. e. the rise in the adoption of digital devices to access information from the general public. Lastly, but not less important, is the higher availability of digital tools for analysis and visualisation of data that could be used not only by specialists but by journalists too. All these factors applied in general to the digital transformation of the journalistic sector but opened a specific space to data journalism.

The first example in a mainstream newspaper was the foundation of the Guardian Datablog at the end of 2010, thanks to Simon Rogers. He was a staff journalist of the British newspaper while it was trying to expand its international reach through innovation and experimentation. The first breakthrough was the coverage of the





so-called England Riot in 2011. This investigation based on public data (poverty indexes, police public records) and data scraped from the Internet, especially from Twitter, is a landmark in the recent history of data journalism. It showed the practice's potential, showing how the UK Government provided a wrong analysis of the phenomenon and how data could help find clues to breaking stories.

More and more mainstream newspapers adopted, to some degree, the data journalism approach. In the beginning, only small units were dedicated to it. Local and smaller media outlets followed soon, as is the case of the Sarasota Herald-Tribune in Florida. Staff writer Paige St. Johns conducted a comprehensive investigation on house insurances in Florida after the new law that made it mandatory to ensure your own house against hurricanes. The investigation published in 2011 led to the first Pulitzer Prize awarded to a data-driven inquiry.

The use of data in journalism became more and more frequent both in the US and internationally. More examples of uses of data journalism have arrived in the last years from different organisations, both in print and online. Dedicated units have been founded in newsrooms in different countries, and more and more small organisations were born. Especially the small newborn organisations dedicated to journalism saw in the digital revolution described above the chance to compete in the journalistic market against more prominent competitors.

Today, data journalism is not a fringe practice for small units or more experimental groups. The Guardian Datablog does not exist anymore, but because of the expertise it gathered, it has become an integral part of the newsroom<sup>5</sup>. The Interactive team of the New York Times underwent a similar fate. The same is happening in the major newsrooms worldwide: the digital (and data) revolution brought a more interconnected organisation among the different components of the newsroom. The digital transformation blurred the boundaries between the traditional sectors of the newspaper and their relative desks.

---

<sup>5</sup> Pamela Duncan. [Number you can tell stories with: a decade of Guardian data journalism](#). 13th September 2021.



In Europe, the situation is very diverse. While data journalism has become an integral part of many small and big newsrooms and is not even highlighted as a separate section anymore in most of the resulting outputs, there are countries where it is still difficult to see it integrated into the daily journalistic practice. In Italy, most of the data journalists, with very few exceptions, are not integrated into newsrooms and work as freelancers. The number of resources needed to invest in data collection, scraping, wrangling or even simply in designing original data visualizations is hardly available to these journalists. The result is that very often data analysis is limited to small datasets, with visualization produced through available tools and without any specific personalization and original design. On the contrary, Spain has been ready to integrate data journalism in many local and national newsrooms from the beginning of this practice. In recent years, a number of cross-border networks have also started joining forces to develop big data investigations: it is the case of Investigate Europe, of the OCCRP, of European Investigative Collaborations. There is a growing scene of collaborative data journalism efforts gathering annually at the European Investigative Journalism Conference and Data Harvest organised and promoted by Arena for Journalism in Europe in Mechelen, Belgium (see next section). In other words, there is a growing and vibrant scene of investigative journalism in Europe that is increasingly implementing data analysis in its practices. But the situation in each country remains highly diversified and therefore the impact of data journalism on local and national readers is also very heterogeneous. While data are needed more and more to manage analyses of the complex challenges our societies face, the ability of journalists, on one side, and of readers on the others to read, understand and use the data are very very different and not always properly supported even at the institutional level.



### 3.1 A cross-border perspective on data journalism

The Covid-19 pandemic highlighted the importance of addressing the quality and excellence of the use of data in journalism. Working with data was frequently also translated into the pressing need for data journalists to transcend national borders.

The ENJOI team interviewed Brigitte Alfter, a senior German-Danish journalist, director of the [Arena for Journalism in Europe](#), the leading European network supporting cross-border collaborative journalism at the EU level, to explore the cross-border perspective within data journalism.

Having practised journalism on local, national and European levels, Alfter realised the need for cross-border collaborative journalism structures. Since 2008, she has developed European support and infrastructures for cross-border journalism and advises on all aspects of collaborative journalism.

She has authored the Cross-border Collaborative Journalism Handbook<sup>6</sup>, first published in 2015, an inspiring guide to transnational reporting with a strong focus on cross-border data journalism.

According to Alfter, collaborating across the globe enhances the opportunities to produce impactful and in-depth reporting. Science journalism and data journalism can hugely benefit from this cross-border collaborative approach for two main reasons:

- When it comes to public data (such as the ones broadly spread and used during the pandemic), the reliability of information highly depends on transparency. However, depending on the specific situations in the different countries, data transparency was often incomplete or missing. Cross-border journalistic projects can be the key to filling this transparency gap and guaranteeing more open and accurate information.

*Example:* [Farmsubsidy](#), co-founded by Brigitte Alfter in 2005 and still active. It is a follow-the-money journalism and data project. For the first time in the history of the EU, it brought to light beneficiaries of the annual billion Euro subsidies to the farming sector (for a detailed analysis of

---

<sup>6</sup> Brigitte Alfter. 2019. [Cross-Border Collaborative Journalism. A Step-By-Step Guide](#). Routledge.



Farmsubsidy, see Section 5 of this document).

- Networks of journalists are more powerful than single journalists. This can be reflected in the quality of cross-border data journalism projects and the creation of "virtual newsrooms" across geographical borders.

*Example:* [DataHarvest](#), annual European Investigative Journalism Conference organised by the Arena for Journalism in Europe. It facilitates open networks for information sharing to help journalists collaborate across borders. It also promotes cross border collaborative journalism through training, publishing articles, speaking, developing educational materials and advising on cross-border collaborative journalism education. DataHarvest also led to successful investigations such as the cross-border data journalism project [Cities for Rent: Investigating Corporate Landlords Across Europe](#) (for a detailed analysis, see Section 5 of this document).

Other outcomes of the interview with Brigitte Alfter are available in D7.1, with a specific focus on engagement as a key ingredient for successful cross-border investigative projects. Here, we aimed to embed the cross-border vision in defining excellent data journalism, as the sections below illustrate.



## 4. METHODOLOGY

The ENJOI project bases its activities on a **participatory methodology**, involving a variety of science information producers, users and stakeholders to assess high quality in science communication and journalism.

All the partners are setting up a **multi-step iterative approach** to research, co-create and produce SPIs together with recommendations and guidelines that will be condensed into the ENJOI Manifesto.

As a first step, this participatory and iterative approach was implemented within the ENJOI consortium itself to perform the first research task and deliverable 7.1. A similar approach was used for task 7.2 on **data journalism** (see section 3 of this document).

Based on the preliminary work conducted through a series of dedicated consortium meetings, allowing each partner to contribute to the discussion, the WP7 selected a “shortlist” of the 15 examples of data journalism coming from different countries and a variety of contexts. In order to better represent the state of the art of global data journalism offer, we selected the pieces according to a mix of criteria. On the one hand, we looked at international awards and data international data journalism conferences, using them as a valuable benchmark to identify relevant examples, especially coming from countries usually less considered. On the other hand, we also relied on the in-the-field experience of the WP7 team's members. Finally, the selection was also guided by the will to represent various ways in which data journalism is used and performed. The combination of these criteria has also been harmonised with the selection of interviewees for the D7.1 deliverable.

The selection of the examples took into account the **six main relevant macro-areas** sketched by the Digital Engagement Focus Report produced by this research team. The macro-areas are:

- Cross-border journalism.
- Investigative and slow journalism.
- Fact-checking.
- Local journalism.



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)

- Social media.
- Newsletters.

Except for 'social media', all other categories can be found in the selected stories.

More factors were taken into account to fully understand the potential and the variety of applications of data-driven journalism. In particular, we looked at how data has been obtained: freely available from public sources, open access data (both from institutions and/or scientific literature), obtained and liberated via FOIA, collected in person by the authors (for example, thanks to sensors), obtained via user-centred approaches and Open Source INTelligence (OSINT) methodologies.

On the other hand, we looked at the topics selected. The corpus of stories covers some of the most critical topics in international journalism at the moment: environment and pollution, climate change adaptation and mitigation, and health-related subjects, with a unique selection of five examples of data-driven journalism covering the Covid-19 pandemic.

All of this is taken into account as a step in developing the ENJOI project and its primary goals. Nevertheless, we also believe that it should and could be a valuable contribution to the construction of the **ENJOI Observatory** for an Outstanding Open Science Communication. The idea is to embed the concepts and the best practices emerging from this analysis of data journalism within the project itself, building and expanding the **ENJOI community**.



## 5. ANALYSIS OF DATA JOURNALISM CASE STUDIES

This section of the document comprises two different sets of data journalism examples analysed. The first part presents 15 examples of data journalism stories in different languages published in recent years in several countries. In this part, we mainly concentrate on the use of data made by the authors, whether it is coming from publicly available sources, open science databases or have been collected from the journalists themselves (for example, thanks to FOIA or similar laws requests).

The second part is a collection of 5 data journalism stories that focus on pandemic-related topics. Data visibility during the Covid-19 pandemic has increased: more data, visualisations, infographics, and so on have been published by mainstream media. Depending on the trends and the events, (i.e. the daily or the weekly number of Covid-19 cases), the data stories or visualisations stand on the front page. They are well visible for the reader with a primary role in the news hierarchy.

Table 1: List of 15 examples analysed

	Title	Media (Country)	Public data	Notes
1	Paradise Lost. Hawaii's Disappearing Beaches	ProPublica/Star Advertiser	Yes, partially	Use of FOIA
2	The Story Behind a Line	Independent project	No	Collected by the authors
3	L'Italia delle slot	GEDI Group (Italy)	No	Use of FOIA
4	El estronismo climático que no nos deja respirar	El Surtidor (Paraguay)	Yes	
5	Dirty Subsidies	Investigative Europe and various media outlets in the EU	Yes	Public data available in public documents gathered and organised for reuse



6	Os Devolvidos	Expresso (Portugal)	Yes	Data coming from NGOs filling a gap in data gathering from EU and the Member States
7	La grande onda	Independent (International)	Yes	Gathering data from different public and open science sources
8	Pazienti dimenticati	Independent (International)	No	Use of FOIA
9	Mar Menor: historia profunda de un desastre	Datadista (Spain)	Yes	Gathering data from different public and open science sources
10	Red Alert. Slowing Down CO2	Straits Times (Singapore)	Yes	Gamification/interactivity
11	Farm subsidies	Cross-Border, Independent (International)	Yes	Public data available in public documents gathered and organised in a reusable format
12	Who Gets to Breathe Clean Air in New Delhi?	NYT (USA)	Yes, partially	Part of the data collected by the authors via sensors
13	How Big Tech got so big: Hundreds of acquisitions	Washington Post (USA)	Yes, partially	Use of public documents and research study necessarily coupled with data coming from private analysts and research companies
14	Cities for Rent	Cross-Border, EU	Yes, but not publicly accessible	The data is collected at a local level and integrated into a data catalogue
15	Liquido e incerto	Folha de São Paulo (Brasil)	Yes, partially	

## 5.1 Paradise Lost. Hawaii's Disappearing Beaches

**Authors:** Ash Ngu, Sophie Cocke

**Keywords:** Environment, climate change, sea-level rise, glocal journalism

**Media Outlet:** ProPublica and Star Advertiser



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication  
This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)



**Country and language:** USA, English

**Date of publication:** December 29, 2020 + subsequent updates

**Link:** <https://projects.propublica.org/hawaii-beach-loss/>

**Short description:**

Sea level rises are affecting the coastal profile of many places worldwide, and the Hawaii archipelago is no exception. In particular, iconic Hawaii's seaside sidewalks are highly relevant for real estate and touristic purposes: they define the seaside panorama. They are a crucial element for the value of the houses and mansions that overlook the ocean. In many places on the islands, private citizens erected seaside walls to protect the estates from the rising water. The problem is that they are just protecting the houses and not the beaches, which are publicly owned. During the last century, three of the archipelago's major islands, Oahu, Maui and Kauai, had lost a quarter of their beaches. The seawalls constructed to protect estates will worsen the situation, resulting in quicker shores disappearing. The radical transformation of the seashores of Hawaii's tourist destinations affects not only the economy but also the environment at a deeper level, causing the loss of many coastal species. Moreover, with the pollution caused by the waves hitting them, the new walls are causing a severe problem to coral reefs around the islands. The scrolly telling story is enriched by videos and images that show the difference between today and half a century ago when the shores were more extensive and in higher numbers. The story's centre is the visualisation of the data on an interactive and searchable map.

**Source of data:**

The publication clearly states the origin of each data utilised to source the story. A dedicated section at the bottom of the main story explains that the data for seawall erections have been collected through a request to the public offices supervising the construction process.

The description allows the reader to understand the period the data refers to (2000-2020). The data was, at least in part, handed over in a paper format, and for some specific locations, there is a clear statement by the authors that it was



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)



impossible to link the data and the address.

All of the data was not publicly available, and for this reason, the authors of the stories had to file a formal request, probably thanks to the Freedom Of Information Act (FOIA), but the description in the stories does not specify this aspect.

### **Transparency of data usage and method:**

All the collected data have been made available, cleared and refined, through a direct link to the database.

The main story also contains a search engine that allows users to browse the data by address or locality. The data can also be browsed via the interactive map embedded in the main story.

A dedicated methodological paragraph briefly explains how the data has been treated after collection and how the final database has been constructed.

---

## **5.2 The Story Behind a Line**

**Authors:** Federica Fragapane with Alex Piacentini

**Keyword:** Migration, refugee, asylum seekers

**Media Outlet:** Independent project

**Country and language:** Italy, English

**Date of publication:**

**Link:** <http://www.storiesbehindaline.com/>

### **Short description:**

The Story Behind a Line is a data journalism project that tells the story of six asylum seekers that arrived in Italy in 2016. There is little text to read, and the reader is catapulted into each story through the path (for this reason, the "lines") each asylum seeker travelled from the point of departure until he/she/they reach their destination. For each path, the interactive visualisation allows the reader to know the kind of transportation utilised in each stage of the journey (train, bus, car,



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)



truck, boat and on foot), the length of every stop in cities of other places (in days) and the distance travelled (in km). The six stories are meagre in detail, symbolising the little knowledge available about each asylum seeker's individual story. Nonetheless, they possess the necessary data to understand the difficulties encountered during the journey. This fact alone, was the implicit intent of the project, should be enough to have a different perspective on the asylum seeker and migrant waves Italy and Europe are experiencing.

#### **Source of data:**

The data comes from individual interviews Fragapane conducted with the six asylum seekers in 2016. During the interviews, Fragapane asked them to help her fill up these lines with data and information. At some points, the interviewees freely gave more information on single facts, and these notes are embedded in the visualisation tool.

#### **Transparency of data usage and method:**

The data source is clearly stated in the methodology text at the bottom of the project homepage. Some data are missing because, as Fragapane wrote, "either they were not able to recall the information or I was so moved that I simply forgot to ask it. I think that imperfections in these projects are inevitable, and I think that's right". Clearly, the lack of information in this project can speak volumes of the tragedy behind the data and the "lines".

---

### **5.3 L'Italia delle slot**

**Authors:** Dataninja, Effecinque, Visual Lab and the journalists of 13 local newspapers of the GEDI Group

**Keywords:** Health, Compulsive gambling, FOIA, glocal journalism

**Media Outlet:** GEDI Group - Quotidiani locali

**Country and language:** Italy, Italian



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)



**Date of publication:** December 2017 and following updates

**Link:** <https://lab.gedidigital.it/finegil/2017/italia-delle-slot/>

**Short description:**

*L'Italia delle slot* (literally "the Italy of slot machines") is a data enquiry on the habits of playing with slot machines in Italy. For a long time, local newspapers have covered the proliferation of slot machines in local shops around Italy and the consequent proliferation of compulsive gambling. However, none of the efforts at the local level could provide enough national attention that could grasp the attention of the general public and policymakers. That is why 13 local newspapers owned by GEDI Group, the same publisher of the national newspaper *La Repubblica*, sparked the idea to work together. The Visual Lab, the internal visual and interactive division that worked for every newspaper of the Group, put together a working team that brought together journalists from several newsrooms and some freelance journalists and experts in managing data. The aim was to build a web app that could work as a searchable database for each municipality in Italy and allow the readers and the journalists to compare territories.

To request the data, Raffaele Mastrodonato, an external journalist and consultant for the project, used the Freedom Of Information Act (FOIA) that was introduced in Italy on the 23rd of December 2016. The data requests went to the owner of the data, Agenzia delle Dogane e dei Monopoli. The result was a massive delivery of more than 10 thousand pages filled with data in pdf format.

After actually translating the pdfs into a proper database, the team went on in producing an internal report. The idea was that the local journalists did not necessarily have the skills to manipulate the data, so the small team composed of external experts produced a series of ideas for stories based on the data they extracted from the pdfs. The Visual Lab realised a web app for the national online newspaper website, and the local journalists were left free of researching stories based on the internal report. All the first wave of data stories were published simultaneously on the 13 newspapers and their respective websites. This strategy



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)



produced maximum impact, also resulting in a press conference held at the Italian Parliament on the 15th of December 2017<sup>7</sup>.

#### **Source of data:**

The data was obtained through a series of FOIA requests to the Agenzia delle Dogane e dei Monopoli, the actual owner of the data on slot machines. The data requested comprised the geographical localisation of the slot machines, all of which must be authorised by the same Agenzia and must respect specific rules, for example, in terms of distance from school buildings. The data also contained the total money spent on every slot machine. The team could then estimate how much was spent for each municipality in the Country.

#### **Transparency of data usage and method:**

The source of the data is part of the credits of the web application, but there is no explicit explanation about the work done on the data. This story is contained in a post on Medium written by Marianna Bruschi<sup>8</sup>, the then head of the Visual Lab. In this methodological piece, Bruschi explains how the team has worked together and assigned the single tasks to the Visual Lab, the external consultants and journalists, and the local newspapers' journalists. She also explains in 7 points the method utilised to extract data from the pdfs, explicitly citing the software used.

---

## **5.4 El estronismo climático que no nos deja respirar**

**Authors:** Maximiliano Manzoni, Lis García, Guillermo Achucarro

**Keywords:** Climate change, Politics, History

**Media Outlet:** El Surtidor

**Country and language:** Paraguay, Spanish

**Date of publication:** 3rd February 2021

---

<sup>7</sup> [Data Journalism, l'Italia delle slot: al via il nuovo lavoro del Visual Lab e del gruppo Gedi.](#)

<sup>8</sup> Marianna Bruschi. [L'Italia delle slot. Ovvero: di Foia, pdf, database e inchieste collettive.](#)



**Link:** <https://elsurti.com/futuros/culpables/especial/estronismo-climatico/>

**Short description:**

*El estronismo climático que no nos deja respirar* is a special data investigation in the form of scrolly tell. The word "estronismo" refers to Alfredo Stroessner, the Paraguayan army officer that took the power of the country with a coup in 1954 and served as President of Paraguay from 15 August 1954 to 3 February 1989. El "estronismo climático" is the concept elaborated by the three authors of the investigation after they analysed the relationship between dictatorial policies and the validity of the production model that led the country to be one of the countries with the highest Greenhouse Gas (GHG) emissions per inhabitant in the region. The investigation points out that the economic policies of *estronismo* shaped the deepening of Paraguay's extensive land estate structure, placing Paraguay as a supplier of raw materials for foreign capital, particularly Brazilian. "We cannot talk about the increase of Paraguay's emissions without considering the state terrorism since 1954 – disappearances, kidnappings, torture and murders – as a method to silence those who opposed the model that fuels these broadcasts", conclude the authors.

So the story investigates the relationship between policy choices and the consequences of climate change in today's Paraguay using an interactive temporal line, where the rise of the GHG is directly related to facts from Paraguayan history.

**Source of data:**

The story relies only on publicly available data from the publicly funded Potsdam Institute for Climate Impact Research (PIK). The authors write that they decided to use the greenhouse gas inventory developed by the PIK because it provides information prior to 1990, crossing official information provided by the States with estimates and mathematical climate models that allow tracing greenhouse gases since 1850.

**Transparency of data usage and method:**

The PIK inventory is also characterised by using official data provided by the States



- including Paraguay - to the United Nations Framework Convention on Climate Change (UNFCCC). It includes additional sources from other related institutions such as The Carbon Dioxide Information Analysis Center (CDIAC), Missions Database for Global Atmospheric Research (EDGAR) and FAO. This information is clearly stated in the FAQ section at the end of the scrolly tell. Parallel to the FAQ section about the building of the investigation, there is a glossary where the authors explain some of the technical terms. There is also an email address for further readers' enquiries. The data used for the story is freely downloadable from the page.

---

## 5.5 Dirty Subsidies

**Authors:** Investigate Europe

**Keywords:** fossil fuels, subsidies, green deal, oil and gas

**Media Outlet:** Investigative Europe and various media outlets in the EU

**Country and language:** English, French, German, Italian

**Date of publication:** 2nd July 2020

**Link:** [www.investigate-europe.eu/en/2020/fossil-subsidies/](http://www.investigate-europe.eu/en/2020/fossil-subsidies/)

### Short description:

For certain aspects of the work done by the international team of investigative journalists behind the collective name "Investigate Europe", their *Dirty Subsidies* is traditional investigative journalism meeting data harvesting. Published in mid-2020, the bulk of stories looked at how much public money is spent subsidising the fossil fuel industry in the European Union (and some other European countries). This investment still happens despite the public announcements of decarbonising the European society and progressively stepping away from the extraction and use of fossil fuels under the Green Deal framework. The team, composed of journalists from 11 different European countries, scanned public data and documents to measure the extent of the "dirty subsidies" at the European and national levels. The



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)

result is a simple map representing access to a collection of stories published both on the team's websites and on several European media outlets. In a nutshell, the EU plus Norway, Switzerland and the United Kingdom subsidise the fossil industry with 137 € billions annually. The reader can access the general data for each country and the related stories on the map. As a result, the map is a powerful journalistic and communicative tool that embodies another frequent outcome of data journalism: being a tool for citizens, activists, and stakeholders.

### **Source of data:**

The data sources page<sup>9</sup> is a database itself, comprising all the different documents (with the relative link when available) that have been used. The data refer to the period 2016-2019 and have also resulted from a discussion with the German Institute for Applied Ecology for validation.

### **Transparency of data usage and method:**

The data sources page has a dedicated column to details, where the reader can easily understand the specific issues that a single data source could provide. The decision to count direct subsidies and tax reliefs and tax exemptions, investments in infrastructure for the fossil sector and the calculations on the free allocations of emission rights is clearly stated in the general description of the data and detailed in the "Category" column.

---

## **5.6 Os Devolvidos**

**Authors:** Ana França, Sofia Miguel Rosa, Bülent Kiliç (AFP), Tiago Pereira Santos, Maria Romero

**Keywords:** Migrants

**Media Outlet:** Expresso.pt

**Country and language:** Portugal, Portuguese

---

<sup>9</sup> [Data sources for subsidies investigation.](#)





**Date of publication:** 2020

**Link:** <https://multimedia.expresso.pt/migrantesdevolvidos2020/>

**Short description:**

*Os Devolvidos* means "the pushed back" in English and refers to the persons that try to enter the European Union asking for asylum and are pushed back through police actions. The right to apply for asylum for people fleeing persecution, war or the effects of climate change is theoretically granted by international law. However, the European Union says it has no tool to sanction the member states that do not respect such a right. The main story of *Os Devolvidos* is a visualisation of the 150 cases of pushbacks that happened in the European Union between January and September 2020. The primary publication is a long scroll-text. Every single story is placed on the European map. The piece reads as a 9-month long journey along the European borders and through police physical and psychological violence. The team decided to explore some of these single stories from the main story in deeper detail. These stories have been published in the print and online edition of Expresso.

**Source of data:**

The authors declare that the exact number of persons involved is not essential. Their story has more of symbolic meaning: should it be just one unjustified rejection, the argument of an unwelcoming European Union would stand. Nevertheless, the story underlines the lack of official statistics on this matter at the European level. So the story is based on the database compiled by Border Violence Monitoring Network (BVMN)<sup>10</sup>, a non-profit organisation that monitors the violence at the European borders. The organisation started to gather information on the subject in 2017, and since then, its database has reached 850 pushback requests, involving more than 11 thousand persons. Their main activity in this field is assembling data on pushbacks via collecting direct testimonies thanks to the work of several organisations on the field and partly rely on a sort of Open Science Intelligence (OSINT) practice scanning institutions, organisations and so on for

---

<sup>10</sup> [Border violence.](#)



information. This practice, powered by the journalistic narration, exposes the European Union and the single-member state's capacity in measuring the phenomenon.

**Transparency of data usage and method:**

The database is built with a standardised methodology by BVMN. Journalists do not declare any wrangling of the data.

---

## 5.7 La grande onda

**Authors:** Davide Mancino

**Keywords:** Covid-19, pandemic, economics, social crisis

**Media Outlet:** Independent

**Country and language:** Italy, Italian and English

**Date of publication:** 2021

**Link:** [www.grandeonda.it/en](http://www.grandeonda.it/en)

**Short description:**

*La grande onda* ('The Great Wave') is a chronicle and a data-based reconstruction of the 2020 Covid-19 pandemic in Italy, the first country that registered cases and deaths in Europe at the end of February 2020. The five-chapter scrolly tell long-form starts from the first cases registered in Northern Italy in Val Seriana, a busy and densely populated area near Bergamo, Lombardy. Local administrators and journalists (see the Covid-19 related section of this report) started to grasp the dimension of the event only when they looked at the excess mortality, i. e. comparing the number of deceased people in a particular month in 2020 with the average for the previous years. From this scenario, set thanks to infographics and data visualisation, La Grande Onda tackles the first year of the pandemic in Italy from different angles: the long term effects on health in the survivors, the mental health consequences, the subsequent economic and social crisis, the scientific



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)



analysis of the effects of the pandemic for different countries and what science learnt on how the SARS-CoV-2 spread from person to person. These themes are treated in a data-based manner, sourcing a great variety of data from mostly open access sources. La Grande Onda shows the potential of gathering vast quantities of information from readily available sources using a method similar to OSINT that can support journalism in making sense of an event of such magnitude.

#### **Source of data:**

The data come primarily from open access sources: the National and European statistics institutes (ISTAT in Italy, the Office for National Statistics in the United Kingdom, Eurostat); national institutions (Italy National Bank); international organisations (OECD, World Bank, UNWTO); and scientific literature. These sources have been coupled with a vast number of research interviews made by the author to frame the stories that emerged from the data correctly.

#### **Transparency of data usage and method:**

The data sources are indicated for each chapter in the final methodological note. All the quotations are retrieved from published articles and are directly linked in the text. La Grande Onda has been financed by a grant from the Scuola Internazionale Superiore di Studi Avanzati (SISSA) with a project prompted by the writer Paolo Giordano and this is declared openly in the text.

---

## **5.8 Pazienti dimenticati**

**Authors:** Riccardo Saporiti

**Keywords:** health, Covid-19, FOIA, health care system

**Media Outlet:** Independent (website and Infodata, Il Sole 24 Ore)

**Country and language:** Italy, International both in Italian and English

**Date of publication:** 24th June 2021

**Link:** <https://forgottenpatients.com/>



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication  
This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)



**Short description:**

*Pazienti dimenticati* ('Forgotten patients' in English) looked at the effects of Covid-19 pandemic on the access to treatments for Italian patients: "two out of three surgeries postponed, more than one exam or outpatient visit out of three referred, just one in ten cancer screening performed during the first lockdown", as the lead of the story says. The piece looked at different hospital activities: surgery, exams and visits, Intensive Care Units (ICUs), oncology, and impact on healthcare personnel. The result is a thorough investigation of hospital treatments' delays due to the pandemic emergency.

The author writes on the first page of the investigation's website: "The ministerial choice was to postpone elective non-oncological hospitalisations with priority classes B and C (i.e. to be carried out within 60 and 180 days respectively) and elective class D hospitalisations (without a defined maximum waiting period and linked to clinical cases that do not cause pain, dysfunction or disability) and to postpone deferred (30 days) and programmable (90/120 days) outpatient services". The consequences of these delays are still all to be assessed. However, a series of scientific studies are beginning to help the scientific community understand the burden of such a situation. The first results indicate that these delays tend to lead to a higher percentage of deaths and a higher amount of adverse effects on the patients' health.

**Source of data:**

The infographics in *Pazienti dimenticati* allows the reader to get information on the postponement of health performances at a single hospital level. This data is not publicly available, and that is why Saporiti submitted more than 200 requests for access to the public records (FOIA - Freedom Of Information Act) to as many NHS local offices (Aziende sanitarie locali, Asl) and hospitals (Aziende ospedaliere, AO) in the country. The requests were to provide data for the period between 1st March and 30th April 2020. "Of the 200 local health districts and hospitals surveyed, 57 ignored our request, 21 rejected it. The other 122 sent the requested data, although not always in a complete form". These are the bases for Saporiti's analysis.



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)

### **Transparency of data usage and method:**

The data collected by the author are redistributed under a Creative Commons Attribution-NonCommercial 4.0 International Public License. The data is freely downloadable from the first page of the website. The investigation also comprises a glossary to help the readers with healthcare-related jargon. So this is a clear case of data liberation via a journalistic investigation that took advantage of the Italian version of the FOIA.

*Pazienti dimenticati* has been financed by a grant from the Scuola Internazionale Superiore di Studi Avanzati (SISSA) with a project prompted by the writer Paolo Giordano, and this is declared openly in the text.

---

## **5.9 Mar Menor: historia profunda de un desastre**

**Authors:** Antonio Delgado, Ana Tudela

**Keywords:** Environment, Urban Development

**Media Outlet:** Datadista

**Country and language:** Spain, Spanish and English

**Date of publication:** October 2019

**Link:** <https://especiales.datadista.com/medioambiente/desastre-mar-menor/eng/>

### **Short description:**

*Mar Menor: historia profunda de un desastre* is an example of the application of data journalism to a local environmental story. Datadista is a Spanish independent team of journalists and designers working on data stories and innovative formats both in Spanish and English. In this story from 2019, they investigate the pollution and environmental alterations of the Cartagena lagoon in the last decades. The story develops through a dedicated area of their organisation's website in the form of a scrolly tell with data visualisations and videos embedded in. The journalistic investigation is two-folded. On the one hand, it explains how and why the lagoon was not protected against excessive urban development, much higher than what is



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)

allowed by the regulations that, on paper, protect it. On the other hand, it uses data from different sources to show the consequences of this situation on ecosystems and the natural environment. The lagoon has received less and less fresh water from rivers due to irrigation intensification, and the urban development along the coastline has resulted in an increase in the pollutants that end up in the lagoon itself.

#### **Source of data:**

The data used in the investigation are both coming from public institutions' databases, such as the regional water management body, and scientific literature and technical documents. Data has been collected from several sources to give the reader an organic vision of the different causes that contributed to the Mar Menor environmental disaster. The data not only is at the basis of the charts and data visualisations but also concurred to the illustrations of the geography and geology of the area that is rendered in animations along some of the chapters of the long-form.

#### **Transparency of data usage and method:**

Each source of data is credited in the visualisation, while there is not a clear indication of the method used to re-elaborate it in some cases.

---

## **5.10 Red Alert. Slowing Down CO2**

**Authors:** Alyssa Mungcal, Anton Dzeviatau, Rebecca Pazos, Zachary Tia

**Keywords:** Environment, Climate Change

**Media Outlet:** Straits Times

**Country and language:** Singapore, English

**Date of publication:** 18th October 2021

**Link:**



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)



<https://www.straitstimes.com/multimedia/graphics/2021/10/singapore-carbon-emissions-climate-change/index.html>

### **Short description:**

The data story from the Singapore daily newspaper is based on a timeline of the CO<sub>2</sub> in the world from 1850 until today. While the reader scrolls down to read the contents that appear in a series of small boxes, a red smoke progressively fills the back of the screen, symbolising the carbon dioxide emissions. The red smoke is measured in two ways. The first is the total amount of emissions in time, the other is the percentage of the carbon budget, where 100% represents the limit the world is trying not to reach. As the "time passes" scrolling down, more information is plotted on the red smoke, showing the reader how much carbon dioxide has been emitted globally since Singapore's independence in 1965: 56% of the global budget. The second part of the piece is based on interactive elements where the reader can insert, for example, the year he/she is born and get on-screen the amount of budget emitted during his/her lifetime. Then the reader can play with some parameters involving personal and public choices to see how these can modify the projection of the future emissions. This is an example of simple gamification involving data in a journalistic context. The reader is no longer completely passive but can actively interact with the data of the piece. This helps to cut the distance between global information and personal everyday life, resulting in an engaging data story.

### **Source of data:**

The data used in the piece comes from a variety of sources, mainly research and scientific sources freely reachable online. The first cited is Our World in Data, a website that publishes data from scientific research on "the world's largest problems", as they claim on their website<sup>11</sup>. Another source is Global Carbon Project, a "Global Research Project of Future Earth and a research partner of the World Climate Research Programme" that gathers scientific data on emissions<sup>12</sup>. Other sources include the publications of the Intergovernmental Panel on Climate

---

<sup>11</sup> [Our world in data.](#)

<sup>12</sup> [The global carbon project.](#)



Change (IPCC) and Singapore's National Climate Change Secretariat (NCCS).

### **Transparency of data usage and method:**

The data sources are clearly indicated at the bottom of the piece's page, even if a direct link to the original sources could not be found. Apparently, there has been no particular wrangling of the original data, with the exception of the necessary steps for visualisation and the building of the gamification part.

---

## **5.11 Farm Subsidies**

**Authors:** Danish Nils Mulvad, British Jack Thurston and Danish-German Brigitte Alfter (founders)

**Keywords:** Agriculture, Subsidies, EU policies, Cross country

**Media Outlet:** Independent

**Country and language:** International, English

**Date of publication:** from 2005 continuously update

**Link:** <https://farmsubsidy.org/>

### **Short description:**

The project started with the purpose of facilitating access to farm subsidies in the European Union. "The means have been to obtain and provide data about farm subsidy beneficiaries and make them accessible to the public and to further journalism in the field."<sup>13</sup>.

The result of this ongoing project is a publicly accessible database where the reader can find information about the farm subsidies at the European level but can also access information on each country down to the level of each subsidy recipient. The information is visualised in a series of simple tables but represents the basis for a series of journalistic pieces developed by the community at large that orbits around the project.

---

<sup>13</sup> [Farm subsidy](#).





Farm subsidies represent both an example of cross country collaboration of journalists, but also a civic act of data liberation. Not that the original data is hidden behind closed doors - it is all public data - but it is difficult to find, especially in just one simple and accessible place, as farmsubsidies.org is. Moreover, according to the founders of the projects, there is too little coverage of the topic in all European press. Hence the project puts itself as a point of reference for all journalists and citizens.

#### **Source of data:**

The original sources of the data are the official European documents on the agriculture policies in the Union<sup>14</sup>. The data is also downloadable in raw format.

#### **Transparency of data usage and method:**

The project clearly declares the supporters: DICAR (2005-2007), EU Transparency (2007-2011) and Journalismfund.eu (2011-2017) in collaboration with the Open Knowledge Foundation (2013-current).

The data collected from a series of documents, often in the PDF format, and a process assimilable to the gathering of information through OSINT techniques and is only transformed in a working format for the publication on the website.

---

## **5.12 Who Gets to Breathe Clean Air in New Delhi?**

**Authors:** Jin Wu, Derek Watkins, Josh Williams, Shalini Venugopal Bhagat, Hari Kumar and Jeffrey Gettleman (story), Karan Deep Singh and Omar Adam Khan (cinematography).

**Keywords:** Environment, pollution, sensor data

**Media Outlet:** New York Times

**Country and language:** USA, English

**Date of publication:** 17th December 2020

---

<sup>14</sup> <https://ec.europa.eu/info/food-farming-fisheries> and [http://europa.eu/pol/agr/index\\_en.htm](http://europa.eu/pol/agr/index_en.htm)



**Link:**

<https://www.nytimes.com/interactive/2020/12/17/world/asia/india-pollution-inequality.html>

**Short description:**

This is a story about inequality and people's exposure to air pollutants in New Delhi, India, especially PM2.5 (but even the composition of the pollution is discussed). This inquiry is based both on public records of air pollution in the city and on data on the quality of the air collected by the NYT team, thanks to a series of sensors. The story follows two kids, Aamaya and Monu, during their day. The two kids come from two different areas of the city and from two families with different economic backgrounds. Collecting data on the air they breathe during a single day and taking into consideration the possibility of the two families having access to air cleaning technology in their homes. The same happens in the two different schools they go to. Different schools mean different chances for having air cleaning technology installed. The overall story shows that even in a city with a chronic air pollution issue, the largest burden of the pollution lies on the people with less economic power. Aamaya and Monu are used as real and effective symbols of the problem in Delhi and in the Country, showing the probable effects of exposure to air pollution on the lifespan of the kids.

**Source of data:**

The public data comes from two AirBeams. They are commercial portable devices designed to measure air pollution that the NYT teams modified in order to guarantee a longer battery autonomy and the connection to a Raspberry Pi computer for data collection. This data is laid on the background of a series of public records on the air pollution in India and in the city.

Additional sources for the story is a series of experts credited as consulted for the piece: Joshua Apte (assistant professor of environmental health sciences and environmental engineering at the University of California, Berkeley), Kalpana Balakrishnan (professor of environmental health engineering at the Sri Ramachandra Institute of Higher Education and Research), Anumita Roy



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)

(Chowdhury, executive director of research and advocacy at the Centre for Science and Environment), Douglas Dockery (former professor of environmental epidemiology at Harvard University), Randeep Guleria (director of the All India Institute of Medical Sciences), Arden Pope (professor of economics at Brigham Young University), Harshal Ramesh Salve (assistant professor at the Centre for Community Medicine, All India Institute of Medical Sciences), Anant Sudarshan, (executive director for South Asia at the Energy Policy Institute, University of Chicago).

This is an example of sensor data journalism, i. e. the use of data-collecting devices for analysing specific situations where there is no specific data available. It is a resource-consuming and complex way of data collecting that is emerging in the later years. This opens a new field for journalistic investigations that is still relatively rarely explored.

**Transparency of data usage and method:**

The process of data collection and data manipulation is clearly described in the notes at the bottom of the piece (*How we collected the data* and *How we processed the data* section). The authors describe the process of selection of the chosen sensors and their modification, citing and linking the scientific literature that guided their choices. There is also a clear discussion of the process of estimation of the effects of air pollution on the kids' lifespan (*How we estimated how pollution might affect the kids' lifespans*). This last section also discusses the little scientific literature on the topic and indicates this field as a sector that needs scientific investigation and public decision-makers attention.

---

## 5.13 How Big Tech got so big: Hundreds of acquisitions

**Authors:** Chris Alcantara, Kevin Schaul, Gerrit De Vynck and Reed Albergotti

**Keywords:** Tech

**Media Outlet:** Washington Post



**Country and language:** USA, English

**Date of publication:** 21st April 2021

**Link:**

<https://www.washingtonpost.com/technology/interactive/2021/amazon-apple-face-book-google-acquisitions/>

**Short description:**

*How Big Tech got so big* by the interactive team at the Washington Post is a classic installation in the "follow-the-money" genre in journalism. It looks at how the so-called tech giants (Apple, Amazon, Google and Facebook) got so big and predominant in their respective markets. The piece is both an explanatory piece and a data search piece. Acquisitions are the key element represented in the infographics, and they are the main explanation for the actual size of the companies today. The story comprises a general timeline of acquisitions for all companies analysed and then a dedicated timeline for each of them, where data is accompanied by a series of facts concerning two different areas: acquisitions in the core business and outside of it. The result is an exposure of the enterprise size of the four giants, but also and primarily a map of the sectors the general public is not aware of the direct interests of these companies. This highlights a series of concerns on the role of tech companies in today's public life.

**Source of data:**

Under the US Law, companies are not obligated to disclose details on acquisitions, so the authors relied on a series of data coming from analysts, research companies and government agencies: Refinitiv, a financial and economic analyst from the London Stock Exchange; the American Economics Liberties Project, an American non-profit organisation that advocates for corporate accountability legislation and the enforcement of antitrust regulations; a report by the US House (*Investigation of the Competition in Digital Markets*).

To these sources, the authors also added studies from two universities: Yale University and the University of Utah.



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)



### **Transparency of data usage and method:**

The sources of the data are clearly exposed in the "about this story" section at the end of the piece. All the sources are directly linked in the text. In terms of clarity on the fact that, to a certain extent, the story's conclusions are estimates (see above, in the "Source of data" section), there is a disclaimer ("Because companies aren't required to disclose all mergers and acquisitions to the public, the list is likely to be incomplete. Some acquisition dates are approximate") and the methodology of data collection and analysis is sufficiently explained directly into the main body of the story.

---

## **5.14 Cities for Rent**

**Authors:** Cities for Rent crossborder journalists' network

**Keywords:** Landlords, Subsidies, EU policies, Cross country

**Media Outlet:** Independent and a number of local/national media

**Country and language:** International, English

**Date of publication:** 2021

**Link:** <https://cities4rent.journalismarena.media/>

### **Short description:**

*Cities for Rent: Investigating Corporate Landlords Across Europe* is a cross-border collaborative investigation on what the authors define as "corporate landlords". Corporate landlords are companies that buy, sell and own many houses. In the last few years, some of these companies have become the biggest landlords in cities and countries across Europe. For example, it was claimed that Blackstone, an American investment fund, was the single biggest private landlord in Europe. The problem is that there isn't publicly available data to support or deny those claims and to allow for a properly informed public debate about corporate landlords. *Cities for Rent's* goal is to try and find the relevant data about corporate landlords in Europe.



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)



**Source of data:**

The data is collected at municipal levels through a thorough research process carried out by the group of journalists collaborating with the project. For getting each data set, the struggle and the obstacle could be different, but the overall characteristic is the lack of information even at the level of municipal offices.

**Transparency of data usage and method:**

All the results are published on the project website in many European languages, and the data is searchable through a dedicated page (data catalogue). Overall, the investigation shed light on a real economic and social problem, on which it was really hard to collect data. Results of the project showed that many European municipalities were trying to regulate housing markets without being able to know who were the biggest landlords in their cities and how many houses they owned. Overall, the main result of the investigation was the methodology the team developed and the way journalists collaborated across European borders to build up from scratch some of those missing databases.

---

## 5.15 Liquido e incerto

**Authors:** Marcelo Leite, Lalo De Almeida, Eduardo Geraque, Fernando Canzian, RAfael Garcia, Dimmi Amora

**Keywords:** Environment, Water resources

**Media Outlet:** Folha de São Paulo

**Country and language:** Brasil, Portuguese

**Date of publication:** September 2015

**Link:** <http://arte.folha.uol.com.br/ambiente/2014/09/15/crise-da-agua/>

**Short description:**

*Liquido e incerto* ('Liquid and uncertain' in English) is an investigation carried out by six staff reporters on the water crisis in Brazil. They investigated three extreme situations – drought in São Paulo and in the semi-arid region of the Northeast and



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)



the floods on the Madeira River – that highlighted the country's unpreparedness to face the emergencies to come in relation to water. The interactive data-driven story is divided into four chapters, and each of them comprises video footage, data visualisations and reporting from the field.

**Source of data:**

The data comes from publicly available resources on the Brazilian water resources and the emergencies related to water management. But it also comes from international private data consultants, such as Webster Pacific, a company that provides data for analysis on a variety of topics, including water management around the world.

**Transparency of data usage and method:**

No dedicated statement on the data manipulation methodology.



## 6. ANALYSIS OF DATA JOURNALISM EXAMPLES ON COVID-19

“Data does not just provide neutral and straightforward representations of the world, but is rather entangled with politics and culture, money and power. Institutions and infrastructures underpinning the production of data—from surveys to statistics, climate science to social media platforms—have been called into question”. These are the words Jonathan Gray and Liliana Bounegru, the two curators of the influential *Data Journalism Handbook*<sup>15</sup>, wrote in their preface to the volume. They were writing at the peak of the first year of the pandemic, and their writing reflects a common feeling both for the general public and for journalists, especially data journalists: “The pandemic has broadened awareness of not just using but also critically reporting on numbers and data.”<sup>16</sup>.

We spent hours upon hours during the lockdown and the period after it talking about “flattening the curve”, criticising the way data on deaths and ICUs occupation were presented both by the media and the institutions. The magnitude of the pandemic, and its consequences on the public discourse, has pushed some journalists to consider it a sort of tipping point in the communication of data and scientific information in general. Never before the fact-checking process has taken such a central role in scientific journalism<sup>17</sup>.

At the same time, the complex and varied nature of the public discourse on COVID-19 and the science related to it could have worsened the polarisation around science, as proposed by Agnes Arnold-Forster<sup>18</sup>.

All of the above exposed “the fragility and provisionality of such data has been widely reported on, with concerns around the under-reporting, non-reporting and classification of cases, as well as growing awareness of the societal and political implications of different kinds of data from sources—from hospital

---

<sup>15</sup> J. Gray, L. Bounegru *Data Journalism in Question* in J. Gray, L. Bounegru (eds) *The Data Journalism Handbook. Towards a Critical Data Practice* cit. p. 12.

<sup>16</sup> id. p. 13.

<sup>17</sup> Michelle Z. Donahue. [COVID-19 impact on science journalism will have lasting effects](#) SciDev.net. 27th August 2021

<sup>18</sup> Agnes Arnold-Foster. 2021. [Polarisation, incivility, and scientific debate during covid-19](#) BMJ; 374





figures to research estimates to self-reporting to transactional data from tracing apps”<sup>19</sup>.

As the Italian writer and journalist Paolo Giordano, ideator and funder of an Italian grant dedicated to data journalism and pandemic in collaboration with the International School of Advanced Study in Trieste (Italy), put it: "numbers have become the protagonists of our life. Our actions, our perception of risk, the decisions of governments have been based on numbers and percentages. Data journalism is a [form of] journalism that selects and analyses data to get new information and evidence. For this reason, it looks to us the most suitable means of investigation to read the recent past and our present". That's why, in this report, there is a specific section dedicated to the analysis of five prominent examples of data journalism in relation to the global pandemic.

All the projects the WP7 team selected started during the pandemic as tentative answers to the insecurities, critics, doubts that emerged when talking about data and Covid-19. This is not a definitive list - and it cannot otherwise be with just five examples - but an effort to show different solutions to different needs. The five examples that we regard as outstanding are coming from different countries, and they were produced starting with different resources and objectives. All together, they show the potential in working with data with different approaches, trying to contribute to the clarity, reliability and validity of the information on the pandemic.

Table 1: List of 5 examples of data usage in covering Covid-19 related topics analysed in this report

	Title	Media (Country)	Public data	Notes
1	Covid-19 Data Dispatch	Independent (USA)	Yes	Gather data resources from USA public agencies and put them at disposal for reuse
2	The Covid Tracking Project	The Atlantic (USA)	Yes	Gather and verify data from US public agencies. Reused for journalism and scientific research

<sup>19</sup> J. Gray, L. Bounegru cit. p. 13.



3	No appointments for mental health patients during the COVID-19 pandemic	Civio/EDJN (Spain/International)	Yes	
4	Quanto é que a covid-19 já custou em contratos com o Estado?	Público (Portugal)	Yes	Collected in a searchable dashboard
5	Coronavirus, the real death toll: 4.500 victims in one month in the province of Bergamo	Eco di Bergamo (Italy)	Yes, partially	From local to international level

## 6.1 Covid-19 Data Dispatch

**Authors:** Betsy Ladyzhets

**Keywords:** Newsletter, Resources repository

**Media Outlet:** Independent

**Country and language:** USA, English

**Date of publication:** ongoing

**Link:** <https://covidatadispatch.com/>

### Short description:

The *Covid-19 Data Dispatch* started in March 2020 as a personal project by data journalist Betsy Ladyzhets during the first wave of the pandemic in the United States. The idea was to provide a digest on the Covid-19 data in the Country through a free weekly newsletter. The basic idea was to provide in a single and easily readable outlet the most relevant data on the pandemic "for journalists, science communicators, educators, and many other concerned citizens working to keep their communities safe", as the project claims. Among the data deluge that hit mainstream media, Ladyzhets - who specialises in covering science and health - worried that the reliability of the information was at risk because of contradictory data interpretations, an unclear time reference for the information provided, a difficulty to get to the original source of the information and the data itself. The project also evolved into a website with a repository of the newsletter and a variety



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)

of resources. The latter is mainly in the form of reusable materials both for communicators, community members and journalists. The project is funded via direct donations, but the newsletter and the website are completely free. The subscribers will get privileged access to extra material via a dedicated Slack channel.

**Source of data:**

The data used in the contents of the project comes from a list of more than 200 (and counting) sources from public agencies.

**Transparency of data usage and method:**

The sources all are listed in a spreadsheet file that is not only accessible but also freely downloadable and copyable. The data sources are organised by category (ranging from "Census" to "Vaccine", from "Vulnerable communities" to "testing"). The project in its entirety also became a tool for information on the activities of the American public agencies in relation to the Covid-19 pandemic, as in the traditional journalistic approach of monitoring the work of the government and its branches.

---

## 6.2 The Covid Tracking Project

**Authors:** Robinson Meyers, Alexis Madrigal, Erin Kissane and the Covid Tracking Project Team

**Keywords:** Tracking, Resources repository, Scientific Research

**Media Outlet:** The Atlantic

**Country and language:** United States, English

**Date of publication:** 7th March 2020 - 7th March 2021

**Link:** <https://covidtracking.com/>

**Short description:**

As the outbreak of the Covid-19 pandemic hit the USA, staff writer at The Atlantic



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)



Robinson Meyers and contributor Alexis Madrigal felt the need for orientation on the data produced by National Agencies. Joined by independent journalist Erin Kissane, they launched a call for volunteers to start a project of monitoring the Covid-19-related data produced by all 50 States and all the Territories, trying to provide in a single place the most updated information on the pandemic in the United States. The Project ran for an entire year, updated daily and was able to attract a wealth of data resources that contributed to more than 7.000 stories on different media, including *The New York Times*, *The Washington Post*, *CNN*, *Vox* and *ProPublica*.

The Project was coordinated by the Atlantic but was possible thanks to a community of hundreds of volunteers from every corner of the United States. The result is an amount of data that highlights not only the spread of the pandemic, the vaccination progression and the effects on public health in general, but also helps to understand more sensitive topics, like the relation between race and Covid-19 and understanding the impact of the pandemic on nursing homes and long-term care facilities.

### **Source of data:**

The data comes from a daily and weekly update made by the volunteers. Since the beginning of the project, for Meyers it was clear that in order to understand the test rate in the Country, journalists (and decision-makers) couldn't rely on the dashboards provided by the single States or the Agencies: States frequently changed how, what, and where they reported data and often published data in different ways, rendering comparisons across states exceptionally challenging. That's why he and the Project felt the need for constant monitoring of the data published by the public offices. The effort resulted in more than 20.000 hours of manual data entry that could lead to some mistakes but limited the errors due to erroneous interpretation of the data published.

For these reasons, the data from the Covid Tracking Project has become the basis not only for journalistic investigations but also for more than 1.000 scientific papers, including works published by *The New England Journal of Medicine*, *Nature*, and *JAMA*, and 11 letters by federal lawmakers demanding answers on the



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)



pandemic response from government leaders and commercial labs.

### **Transparency of data usage and method:**

The Projects had a comprehensive section on the methods utilised to get to the published data on the website<sup>20</sup> and it frequently updated the methods of the different sub-projects run by part of the team. Everything is published on the website and reusable solely for journalistic, research, academic, medical, healthcare, and/or non-commercial use. The materials produced by the project are protected by copyright and intellectual property and are owned by The Atlantic Monthly Group LLC, the Atlantic publishing company.

---

## **6.3 No appointments for mental health patients during the COVID-19 pandemic**

**Authors:** Ángela Bernardo, María Álvarez Del Vayo, Carmen Torrecillas with contributions by Monica Georgescu and Olalla Tuñas

**Keywords:** Mental Health

**Media Outlet:** Civio.es and European Data Journalism Network (EDJN.eu)

**Country and language:** Spain/EU, Spanish and English

**Date of publication:** 3rd December 2020

**Link:**

<https://civio.es/medicamentalia/2020/12/03/mental-health-coronavirus-covid-19/>  
<https://www.europeandatajournalism.eu/esl/Noticias/Noticias-de-datos/Sin-cita-con-el-psiquiatra-por-el-coronavirus>

### **Short description:**

This story on the lack of mental health treatment during the first phases of the pandemic is the first part of a larger investigation on mental health in the European Union coordinated by the Spanish independent media Civio.es and published by the

---

<sup>20</sup> [The Covid tracking project.](#)



European Data Journalism Network, a platform for data-driven news on European affairs in up to 12 languages supported by funding from the European Commission (grant agreement LC-01591825). This story highlights the power of data, especially public data, coupled with research on the field to identify less known stories and effects of a phenomenon like the pandemic on civil society. The story shows how the block of all treatments considered non-vital and non-urgent by the European hospital due to the response to the emergency had severe effects on the health of specific groups of patients. The piece not only looked at the lack of in-hospital treatment for mental health during the lockdown period but also at the different effects that the obligation of staying at home had on different persons: for example, people affected by autism spectrum disorders did experience intense stress due to the loss of their daily routines and the inability to go to specialised rehabilitation centres.

**Source of data:**

Data comes from a variety of public sources, such as the World Health Organization, Eurostat and a survey by the European Psychiatric Association

**Transparency of data usage and method:**

All the sources are listed in the “Method” section at the end of the story and all the data collected are linked and downloadable.

---

## 6.4 Quanto é que a covid-19 já custou em contratos com o Estado?

**Authors:** Rui Barros, Claudia Carvalho Silva, José Volta e Pinto, Gabriel Sousa, David Mano

**Keywords:** public contracts, expenditure, Covid19

**Media Outlet:** Público

**Country and language:** Portugal, Portuguese

**Date of publication:** 2020 (last update: 7th November 2020)



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)



**Link:** <https://www.publico.pt/interactivo/gastos-covid-19#/>

### **Short description:**

The Público journalists have produced a dashboard where all the contracts signed by public administrations in Portugal are tracked. Since the beginning of the emergency, government, municipalities and public institutions have signed 16,996 contracts with private companies in response to the pandemic. The dashboard is intended to highlight how much of this money has already been spent and to which goal they were allocated in the first place. In total, the Portuguese public expenditure has reached 49.240.203 euros. In more than 96% of the cases, they are spent on contracts with more than 4.500 private companies. The contracts are also divided into categories. For example, 4768 contracts regard personal protective equipment for more than 200 million euros. The data can be searched by category (contract, product, company, institution), allowing for detailed research.

### **Source of data:**

The data - explains the methodology at the bottom of the dashboard - were obtained through the contracts available on the public open data portal<sup>21</sup> where contracts signed under Decree-Law no. 10-A/2020 have been published and classified by public bodies as being related to covid- 19 and on the BASE portal<sup>22</sup> (the Portuguese public database on contracts signed by the public administration) with keywords related to the pandemic. Where possible, the authors kept only materials and services with a direct link to the pandemic.

### **Transparency of data usage and method:**

From the methodological note at the bottom of the page there is a link to an explanation of the process of data collection and analysis. All the databases consulted are listed and explained, and the authors explain in detail how they proceeded in cases where the correct significance of the data itself wasn't clearly addressable. Moreover, a dedicated email address is available for the readers that

---

<sup>21</sup> <https://dados.gov.pt/pt/>

<sup>22</sup> <https://www.base.gov.pt/base4>



should find errors or want to contribute with their knowledge and expertise.

---

## 6.5 Coronavirus, the real death toll: 4.500 victims in one month in the province of Bergamo

**Authors:** Isaia Invernizzi

**Keywords:** Local, death rate, lack of data

**Media Outlet:** Eco di Bergamo

**Country and language:** Italy, Italian and English

**Date of publication:** April 2020

**Links:**

[https://www.ecodibergamo.it/stories/premium/Cronaca/coronavirus-il-numero-real-e-dei-decessiin-bergamasca-4500-in-un-mese\\_1347415\\_11/](https://www.ecodibergamo.it/stories/premium/Cronaca/coronavirus-il-numero-real-e-dei-decessiin-bergamasca-4500-in-un-mese_1347415_11/)

[https://www.ecodibergamo.it/stories/bergamo-citta/coronavirus-the-real-death-tool-4500-victims-in-one-month-in-the-province-of\\_1347414\\_11/](https://www.ecodibergamo.it/stories/bergamo-citta/coronavirus-the-real-death-tool-4500-victims-in-one-month-in-the-province-of_1347414_11/)

### Short description:

This investigation is the result of the analysis carried out by the local Italian newspaper *Eco di Bergamo* and the research and data analysis agency [InTwig](#), using the data provided by local municipalities. The starting point was a lack of data about the real Covid-19 death rate in Bergamo, one of the Italian cities most hit by the Coronavirus outbreak at the beginning of 2020. In particular, official figures didn't say that in March 2020 more than 5.400 people had died in Bergamo province, 4.500 of which were due to coronavirus: six times more than the previous year.

The "official" certified deaths caused by Covid-19 in the local hospitals were "only" 2.060, but there were around 2500 people who generically died for "interstitial pneumonia". The investigation was launched to fill this data gap.

### Source of data:



ENJOI - ENgagement and JOurnalism Innovation for Outstanding Open Science Communication

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n°101006407  
[www.enjoiscicomm.eu](http://www.enjoiscicomm.eu)



In order to verify the increase of total mortality compared to the previous years and collect the missing data, Isaia Invernizzi launched a survey among all the 243 municipalities of Bergamo province: 91 administrations answered the call, representing 607.000 people, which is more than 50% of the total population. The official data enabled L'Eco di Bergamo to make an accurate estimate of deaths and infections over every area of the province. Results demonstrated that 4.500 of the overall deaths in March 2020 in Bergamo were due to coronavirus: more than double of the official sources. This investigation is a great example of a data journalism project that started at the local level and gained international visibility.



## 7. CONCLUSIONS

The Covid-19 pandemic has shown as never before how much data is of paramount importance in public communication. Governments and public institutions all around the world faced an unprecedented scenario both in terms of emergency and communication. But the deluge of data on various aspects of the pandemic showed that using data in communication is not enough: there is a clear need for critical reporting on data and numbers.

This is where journalism, and data-driven journalism in particular, can and must have a more prominent role: asking for the availability of more and more data, interrogating data for uncovering hidden stories and issues, opening the public discussion from different angles and perspectives.

The repertoire of data journalism examples analysed in this report is not a definitive list of works that should be taken as a benchmark. It is too short a list of case studies and too limited in terms of the diversity of media taken into consideration. It was not our intention to provide an extensive review on the state-of-the-art of contemporary data journalism: for that, more comprehensive texts have been published in recent years, and many curations are available online.

Our main goal here was to show a selected number of outstanding examples of different practices related to data journalism that underline the potential contribution of this specific journalistic practice to the public discussion around scientific topics. The selection itself is not done on a scientific and rigorous basis, but rather more drawing from the experience of the authors, all practising data journalists for many years, who use these and many other case studies during their activity as data journalism lecturers as well as in public speaking for other journalists, communicators, local policymakers, students and teachers. In other words, these case studies well represent, in their diversity, a number of issues we should take into consideration when evaluating the opportunities and the challenges of data journalism as a practice that can contribute to outstanding open science communication and journalism.



The 15 selected examples provide hints on how to:

- uncover a story using publicly available data and combining them with journalistic methodology;
- use open-access data and publicly available scientific data to report on a science-related story;
- make public information that is scattered across several official documents easily available and accessible to citizens as well as to journalists via a dedicated repository;
- used OSINT-like techniques to gather information on topics not covered by official data collections;
- use FOIA and FOIA-like tools to get access to public information and data that contain stories of public interest;
- use sensors, with the validation of scientific literature, to collect data in cases where no data is available;
- collaborate from different countries to uncover topics on a transnational level.

All of this is even more evident when we focus on the selection of examples of works on Covid-19 because of the magnitude of the crisis. But the same size of the situation, where public institutions were not completely able to manage the communication and thoroughly inform the citizens, underlined the need for more and more attention on the data and the increasing need for data that our communal living demands. Data that has to be critically analysed, fully interrogated and validated, need to be explained and contextualised by the work of more and more data journalists.

Our main goal is to highlight how this type of journalism can demonstrate where public institutions and governments are not completely fulfilling the right of citizens to have access to information and data. Data journalism can be a powerful watchdog on the public availability, transparency and access to data that are crucial in taking decisions and describing local and global trends.

Moreover, it can provide a useful framework to look at a complex situation from different perspectives, to avoid falling always into the confirmation bias that is



one of the main mechanisms behind the current crisis of misinformation and infodemics. Data are not objective per se, they do not tell a story, they help contextualize a story and help highlight the voids, the gaps, the need for further analyses and for further and better data collections.

This is a pre-condition to have a very open and public discussion on measures that need to be taken to control a public emergency or to embrace practices of prevention of risks or, even, to design future strategies of sustainable development. Be it a pandemic, a climate crisis, an environmental problem, a technological development. Data can be of help but in order to achieve this goal, they need to be understood and become a much more common ingredient of public conversation on any socially and ethically relevant issue.



## 8. FURTHER REFERENCES

Batsell, J. 2015. [Engaged Journalism: Connecting with Digitally Empowered News Audiences](#), Columbia University Press.

Cairo, A. 2019. [How Charts Lie: Getting Smarter about Visual Information](#). W. W. Norton & Company.

Davies, S. R., Franks, S., Roche, J., Schmidt, A. L., Wells, R. and Zollo, F. 2021. [The landscape of European science communication](#), Journal of Science Communication, Volume 20, Issue 3.

EBU Media Intelligence Service. 2021. [Trust in media report](#)

Kleis Nielsen, R. 2015. [Local Journalism. The decline of newspapers and the rise of digital media](#), I.B.Tauris & Co. Ltd in association with the Reuters Institute for the Study of Journalism, University of Oxford.

Lakomý, M., Hlavová, R., and Machackova, H. 2021. [Open Science and the Science-Society Relationship](#), Society, Volume 56.

Newman, N., Fletcher, R., Schulz, A., Andi, S., and Nielsen, R. 2020. [Reuters Institute Digital News Report](#), Reuters Institute for the Study of Journalism, University of Oxford.

Tola, E. 2019. [Delving into the state of data journalism](#). Journalism Innovation, the Medium journal of the Craig Newmark Graduate School of Journalism, CUNY. (Translated and published in Italian as [Ripensare il data journalism](#), June 21st 2019, European Journalism Observatory).

Tola, E. 2020. [Making sense of science stats](#). In: The KSJ Science editing Handbook. Knight Science Journalism Fellowship at MIT.

