



OjoPúblico

# Data journalism and the need of an Open Government

A Latin American journalism experience

# Data journalism

It's a new trend of  
journalistic investigation  
based on Big Data.

# Data journalism

## INVESTIGATIONS WITH PUBLIC DATA

90 countries in the world – including 14 in Latin America and the Caribbean - have Freedom Of Information laws, according to the Regional Alliance for Free Expression and Information. This presents an advantage for investigative journalism: it is now possible to track how corporate power influences state decisions and the impact this has on people's lives.

## INVESTIGATIONS WITH LEAKED DATA

Few processes show the peculiarities of this time as well as data leaks. Even with new forms of communicating with sources using specialised software, we face the challenge of understanding large amounts of information that could bury any enthusiasm if we didn't have the help of programmers. These are some examples.

## INVESTIGATIONS WITH BUILT DATABASES

In recent years, big cases started in investigation projects with data bases build by the journalists themselves. Here we put together seventeen experiences that can inspire the development of new topics.

# Open Government: Perú

- In 2003, Perú approved the Act of Access to Public Information.
  - 1. All information held by the State is presumed to be public, except for cases expressly provided for in article 15 of this Law.
  - 2. The State adopts basic measures that guarantee and promote the transparency in the performance of Public Administration entities.
  - 3. The State has the obligation to deliver the information demanded by the people in application of the principle of publicity.



## Los dueños del agua en el Perú

Por OjoPúblico - [Investigación](#)

Ojo-Publico.com documenta las historias de la desigual lucha de poder por el agua en Perú. Una serie de investigaciones, un documental y un mapa de datos identifican a los privilegiados dueños del agua, en base al análisis de casi medio millón de derechos de explotación del agua otorgados en medio siglo. El especial está disponible en tres idiomas: quechua, inglés y castellano.

28 Agosto, 2018

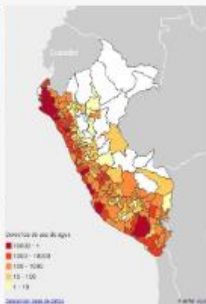


## Los despojados

Por Nelly Lana Amancio - [Investigación](#)

Un viaje a las comunidades indígenas de Aparímac, Cusco y Arequipa, ubicadas en el corredor minero sur, para narrar su lucha por acceder al recurso hídrico más valioso del país.

28 Agosto, 2018



## Data revela quiénes concentran los derechos de agua

Por OjoPúblico - [News Apps](#)

Más de medio millón de permisos para explotación de agua a nivel nacional.

28 Agosto, 2018



## El poder empresarial detrás del agua

Por Elizabeth Salazar Vega - [Investigación](#)

Agroexportadoras, mineras y embotelladoras lideran el ranking de las privilegiadas compañías que acceden a este recurso.

28 Agosto, 2018



## El precio del agua: Los pagos irregulares de la minería

Por Leslie Moreno Custodio - [Investigación](#)

Las Bambas y Volcan y la omisión de pago de un millón de soles.

28 Agosto, 2018



## Jueces del círculo de Hinostroza y Ríos favorecieron a empresario acusado de pesca ilegal

Por Luis Enrique Pérez y Milagros Salazar - [Investigación](#)

"Rey de la pesca negra" operó sin

26 Agosto, 2018



## Jueces retiran confianza a su representante en el CEPJ tras audio con Hinostroza

Por OjoPúblico - [Actualidad](#)

La Sala Plena de la Corte Suprema decidirá esta semana su suerte.

27 Agosto, 2018



## Walter Ríos se reunía con empresario denunciado antes de juicios por terrenos en el Callao

Por Eduardo Alonso García y Luis Enrique Pérez - [Investigación](#)

Policía registró encuentros.

26 Agosto, 2018



## Pedro Chávary defendió como abogado al vendedor de armas peruano James Stone



## Fiscal Gálvez fue confidente de operador de "Cuellos Blancos del Puerto"



## Congresistas piden a ministra de Cultura evaluar denuncia contra Reinaldo Narango



## Gustavo Yepes: "La corrupción es una institución y tiene reglas de juego"



## OCMA suspende a expresidente de corte de Lima Sur por Audios de la Corrupción



## Jueza de caso Sánchez Paredes: "No sé qué interés tendrá Hinostroza"



## Procuraduría señala a fiscal Gálvez como parte de "Los Cuellos Blancos del Puerto"



## Corte Suprema declara en emergencia al Consejo Ejecutivo del Poder Judicial



## Jueza del Consejo Ejecutivo del Poder Judicial renuncia tras audios con Hinostroza

# OjoPúblico

- OjoPúblico is a lab of innovation for journalism focused in a mix of deep reporting, the developing of news apps, and the exploration of the new digital storytelling.
- The strength of our methodological proposal is based on journalism and technology.

## GUIDE TO INVESTIGATE WITH DATA

Paul Bradshaw, author of the blog *Online Journalism*, argues that the journalistic work with databases comprises five stages: gathering, debugging, analysis, verification and presentation of findings. We can take that sequence to propose the following exercise when starting a project.

### The collection of data



You must know the file formats that contain the information and the tools you need to collect them. You can get masses of data using a script, a simple program that enables automated downloading of information. This process is known as scraping.

Are there any databases on the subject? How and why were they created? Are they on an official websites or should I make an FOI request?

If the database is on the web, is it downloadable or do I need to scrape it?

What is the best format (Excel, CSV, json) to request a copy of these databases in? If the information is in PDF or JPG, how can I change its format to an Excel file?

If I build a new database: what variables should I include and what can I prove?

### Debugging and context



There are several types of errors you might find: duplicated records, incomplete boxes, misspelled words, etc. You will need tools to identify and solve these problems. This is called 'cleaning the data'.

Is the database complete? How many lines of information does it have? Can I clean it using Excel or Open Refine? When do I need to do it by hand? When do I have to use data managers with more capacity, such as MongoDB?

Do I know and understand all of the terms, variables and acronyms that appear in the database? Are they the same ones used in similar databases? Is the criteria aligned with the meaning of the question I want to answer or do I need to see the same data in reverse?

### Crossbreeding and analysis



In this phase, the value of the findings depends on the quality of the questions asked and the combination of two or more records to find revealing matches.

Do my databases have a concept or code in common that will allow me to cross the datasets: ID, Tax Code, full name?

Does the database crossing show trends, patterns or evolutionary processes in a given period? In what context?

Or, on the contrary, does it reveal atypical behaviours? In what context?

### Verification



Investigative journalists must apply the traditional methodology: go to the necessary places, interview the people involved, and review new documents to detect weaknesses and strengths of the database.

Does the data represent the actual condition of the people? What has changed in the life of the alluded people: his health, economic stability, legal status or relations?

Does it affect the meaning of the findings? Does it confirm its relevance? Does it stress it or make it relative?

Is there an expert that can validate the methodology of the crossing? Is it possible to have correct findings that offer more than one interpretation?

### Presentation



To present your findings in the most efficient way, you have to think about this from the beginning. There are libraries like d3js.org and software repositories like Github with examples that you can adapt to what you need.

What is more convenient: visualisation or an app? Which one will make the best contribution to the meaning of the story?

What should the user experience be like? What emotions should the chart or tool induce in the reader? Which elements of my app or visualisation are necessary for the user?

Is the tool responsive? Will it look good on mobile phones and tablets? Can it be shared? Can it be embedded?

# What have we done?

- -Corruption / *Panama Papers, Lava Jato*
- -Human rights / *La justicia de los pobres, Memoria*
- -Organized crime / *Oro sucio, Memoria Robada*
- -Environment/ *Amarakaeri*
- -Corporate power/ *Cuidados Intensivos, The Big Pharma Project*



## We work in different formats:

- -Longform (investigative reporting, non fiction stories)
- -Digital tools (search engines, news apps )
- -Visualizations (maps, nodes)
- -New digital storytelling (Web domentary, comic)

# Longform



# Digital tools

OjoPúblico ¿Cómo se hizo? | Reportajes




## SUPREMA FORTUNA

Una radiografía del patrimonio y la carrera de los jueces en el Perú

Busca el perfil de tu juez por nombre, cargo o distrito **BUSCAR**

OjoPúblico **THE BIG PHARMA PROJECT** 521 f t SUSCRIBETE



HERCEPTIN SOVALDI REYATAZ ISENTRESS  
BARACLUDE HUMALOG MIX EVOLTRA ENBREL  
ATRIPLA AVASTIN GLIVEC KALETRA

OjoPúblico ¿Qué es Cuidados Intensivos? Investigación Entrevistas Videos DataViz Denuncia

## cuidados INTENSIVOS

¿Qué sabes de tu médico, clínica y aseguradora?

Buscar médicos, establecimientos, seguros, EPS. (Campo obligatorio) **Buscar**

Establecimientos  Médicos  Aseguradoras, EPS y programas de salud

32 clínicas acumulan S/ 10 millones en multas en Indecopi por malas prácticas.  
61.372 médicos prestan servicios en el Perú. En 10 años dos fueron expulsados de su gremio.  
56% de servicios privados de salud no ha sido supervisado por Susalud a la fecha.

OjoPúblico **MEMORIA ROBADA**

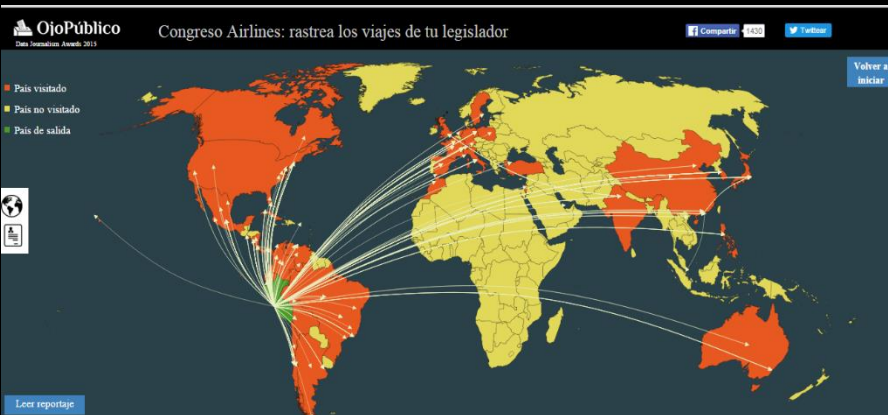
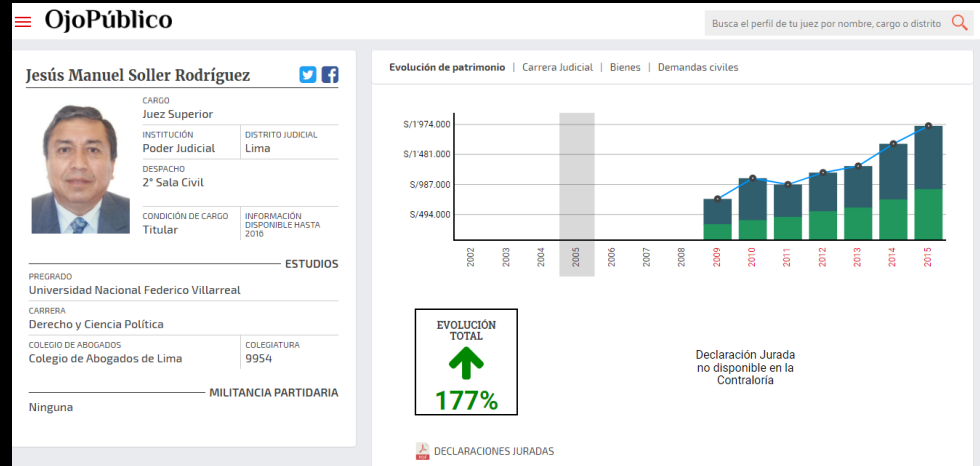
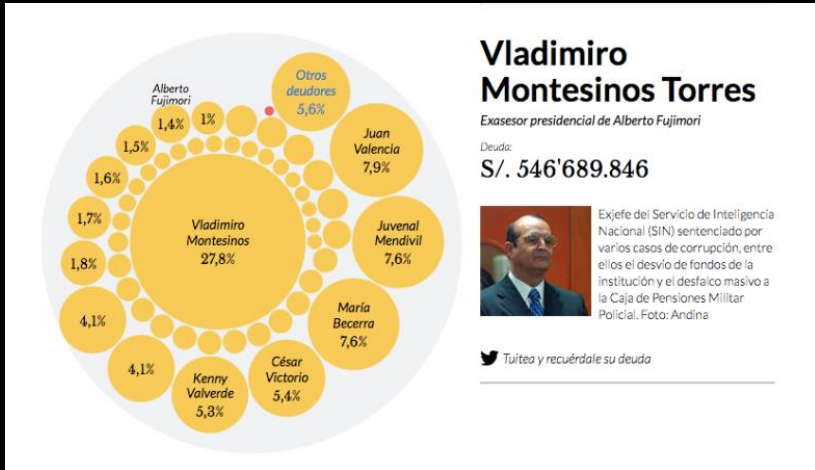
ROBOS SUBASTAS REGISTRO REPATRIACIONES



Descubre las 39 mil obras de arte robadas de 132 países y reportadas a Interpol en los últimos 100 años.



# Visualizations



# New digital storytelling



# How do we work?

The best OjoPúblico's experiences  
could be summarized  
in 3 points

# 1: the 4 dimensions of data

- In contrast to the old way, we design our investigations. Since the starting point we have a pretty clear idea of what we need to find, what kind of results we expect, what content we're going to develop.
- The new journalist is an architect of a digital reality that has 4 components:
  - Unlimited dimension of data
  - Global scope
  - A complex narrative structure
  - Digital security

# Digital security

Experiences like Panama Papers and Paradise Papers have confirmed this factor not just a possibility, but as a need.

- Encrypted email
- VeraCrypt



**CAJA DE SEGURIDAD**

*[Herramientas para evitar que te espíen en Internet]*

**Tor Project**  
<https://www.torproject.org/>  
Software gratuito para comunicaciones seguras. Oculta la dirección de IP de los dispositivos utilizados. Permite navegar sin ser detectado ni dejar rastro de sitios visitados o la ubicación geográfica del usuario.

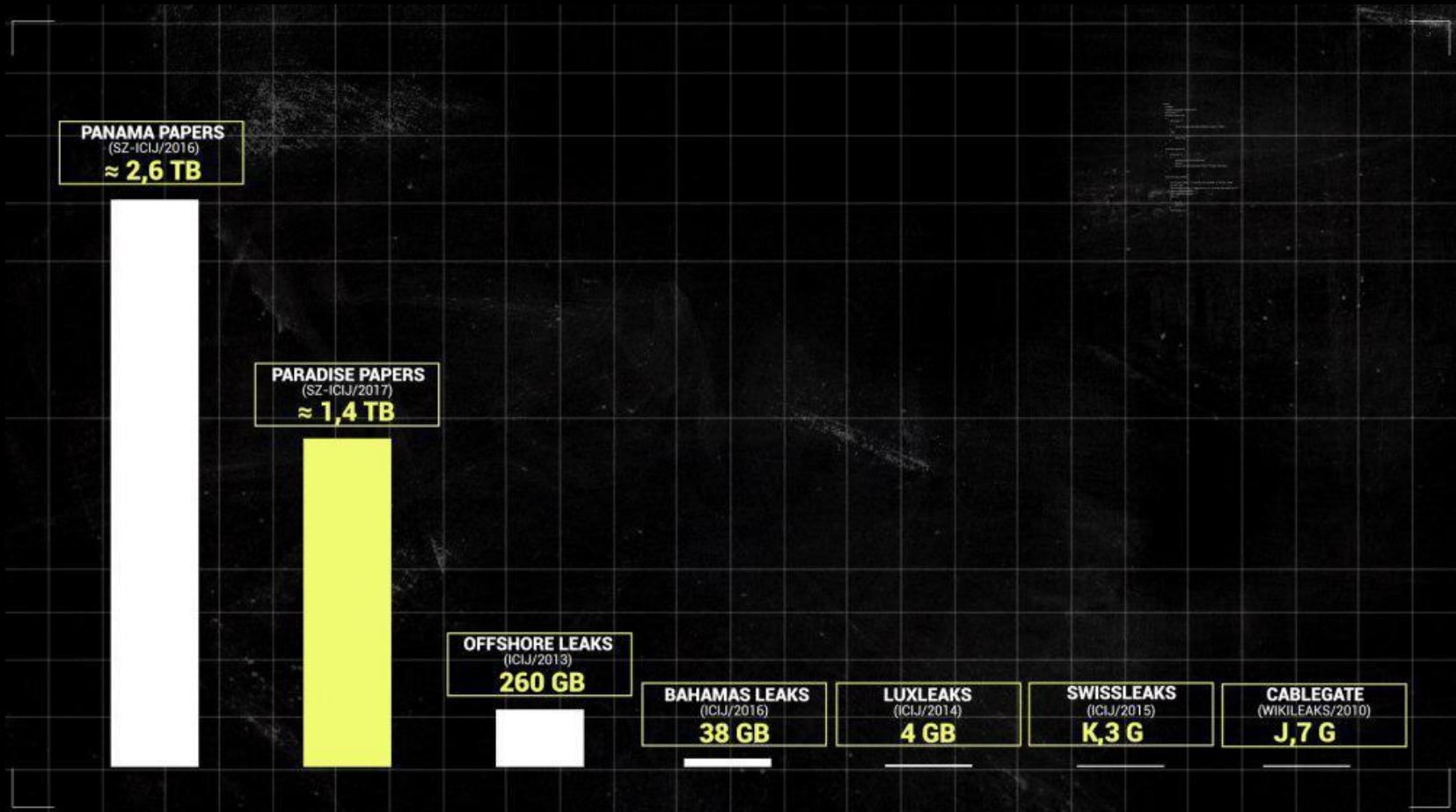
**Cryptocat**  
<https://crypto.cat/>  
Servicio de mensajería instantánea privada en la web que puede ser usado desde cualquier navegador y en celulares.

**Mozilla Thunderbird**  
<https://www.mozilla.org/en-US/thunderbird/>  
Programa de correo libre y de código seguro para recibir, enviar y almacenar mensajes electrónicos. Se puede gestionar varias cuentas de correo con un solo programa.

**Enigmail**  
[www.enigmail.net](http://www.enigmail.net)  
Complemento del Thunderbird que permite enviar correos protegidos con claves cifradas. El usuario mantiene su clave. Para usar Enigmail, se debe instalar también GNU Privacy Guard (GnuPG).



# Unlimited dimension



El Confidencial / ICIJ

Paradise Papers: 13.4 millón documents

# Global scope

## 10 DATABASES TO TRACK THE MONEY AND OTHER LEADS

This is a suggestive selection collected by Ojo Público from those attending the Global Investigative Journalism Conference 2015.

### ICIJ Offshore Leaks Database

<http://offshoreleaks.icij.org/search>

It has data on more than 100,000 companies and funds held in tax havens. It is part of an archive of 2.5 million documents leaked to the International Consortium of Investigative Journalists (ICIJ).

### OpenCorporates

<https://opencorporates.com/>

It contains information on 80 million companies and 90 million directors in more than 100 countries. You can search by company name, address and name of director.

### Persons of interest

<https://www.personadeinteres.org>

It gives access to court records, property records and intelligence reports on people linked to organised crime, drug trafficking, corruption, etc. It has information on trials for drug trafficking in Peru.

### Open Spending

<https://openspending.org/>

Website that follows the public spending of governments around the world and presents it in various forms of visualisation.

### Registering property in Miami

<http://www.miamidade.gov>

Miami Dade County's free data, which can locate properties in Miami, on behalf of the owner.

### Investigative Dashboard

<https://www.investigativedashboard.org/>

Allows you to search shareholders, directors and financial reports of companies around the world. There are links to more than 450 online databases in 120 countries. Platform built by the Organized Crime and Corruption Reporting Partnership (OCCRP).

### Search Systems

<http://publicrecords.searchsystems.net>

Specialised portal. It contains more than 55,000 databases by date of birth, date of death, marriage, licenses granted, stocks, mortgages, among other subdivisions.

### Registration of companies Panama

<http://ohuiginn.net/panama/>

Independent site that reordered the public record information from the Panama leak to facilitate the research of more than 300,000 companies registered in the country. It allows searches by people's names.

### FlightAware

<https://es.flightaware.com/>

A worldwide platform, including a mobile app, to track airline flights and their status. It allows you to search by the name of the airline, flight number, route and aircraft registration.

### Marine traffic

<http://www.marinetraffic.com/>

It is a database updated in real-time that can track the location of any boat or ship, in addition to departures, arrivals and routes.

# A complex narrative structure



## HOW TO TRACK CRIMES IN A DATABASE

Twenty research projects that changed  
the way journalism is done

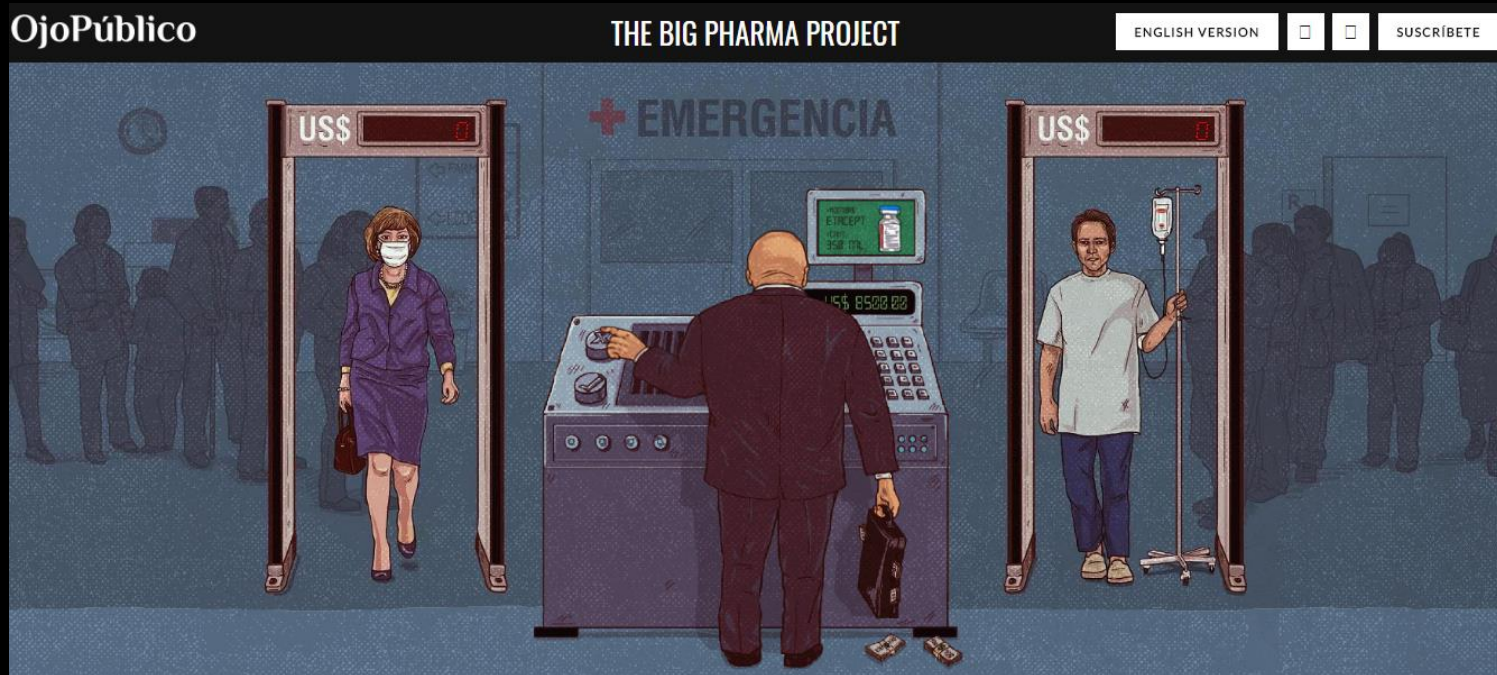
## 2: the need of alliances in the digital era

The Stolen Memory Project was our first transnational investigation. It was a collaborative effort of Latin American journalists in 5 countries. OjoPúblico applied the methodology learned in the Panama Papers investigation: a collaborative project is a relationship in which all partners win.



## 2: the need of alliances in the digital era

The Big Pharma Project was one of our recent projects that replied that model with successful results: we offer our knowledge and developing, and we receive data and reporting from places we couldn't reach in other way.





# 3: the challenge of experimentation

The road to a culture of innovation is full of experiments.

1. When you have limitations in the access to public data, you can build new probatory schemes with data from private resources.
2. When you have a small, local team, you can expand the scope of your work with a trasnational collaborative project.
2. When you need a new way to tell a story, you can explore the new formats of the digital storytelling.
2. There's no limitation for experimenting.

# Data + journalism: fact-checking



LATEST  
NEWS

FACT-  
CHECKING

STRUCTURED  
JOURNALISM

ABOUT  
THE LAB

Duke  
Reporters' **LAB**

## FACT-CHECKING NEWS

FACT-CHECKING NEWS

Global fact-checking up 50% in past year

February 16, 2016



[SEE ALL FACT-CHECKING NEWS](#) ▶

## GLOBAL FACT-CHECKING SITES

The Reporters' Lab maintains a database of global fact-checking sites. You can use the map to explore sites around the world or use the menu below.

PROVIDER LIST



OjoPúblico

**Thank you**

**David Hidalgo**

NEWS DIRECTOR

Twt: @Club\_insolito

dhidalgo@ojo-publico.com