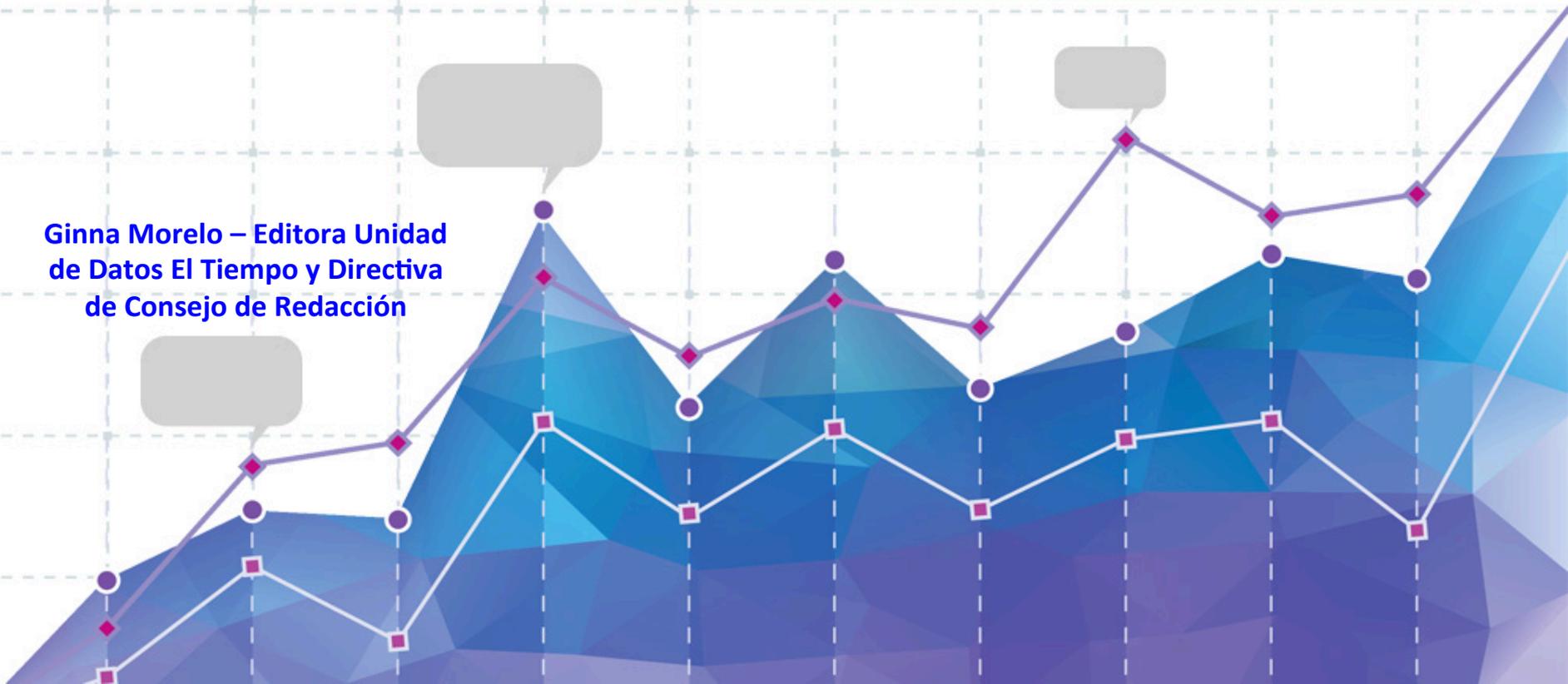


#istorias + datos

Nace un nuevo reportero

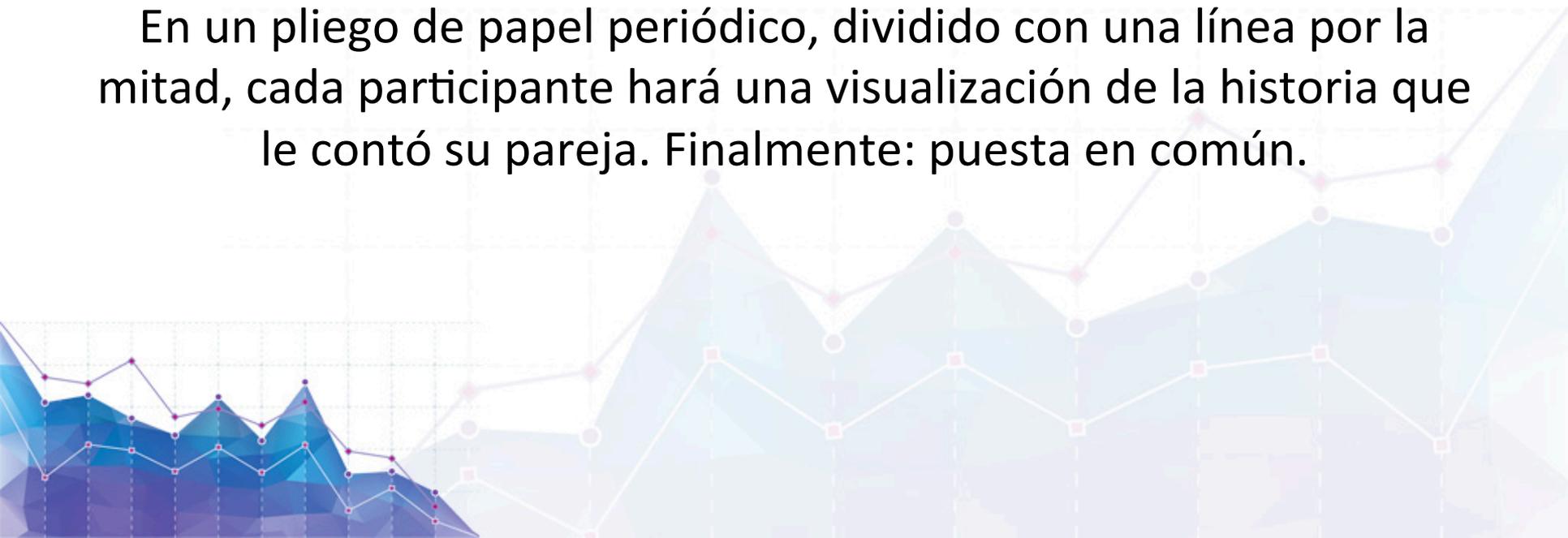
Ginna Morelo – Editora Unidad
de Datos El Tiempo y Directiva
de Consejo de Redacción



Día 1

Dinámica: Se conformarán parejas. Cada uno le contará al otro tres historias de su vida en donde el elemento común será un mismo número. El diálogo será a modo de entrevista.

En un pliego de papel periódico, dividido con una línea por la mitad, cada participante hará una visualización de la historia que le contó su pareja. Finalmente: puesta en común.



¡El detonante!

WIKILEAKS DICE QUE SERAN EXTERMINADOS ...

RAYMA



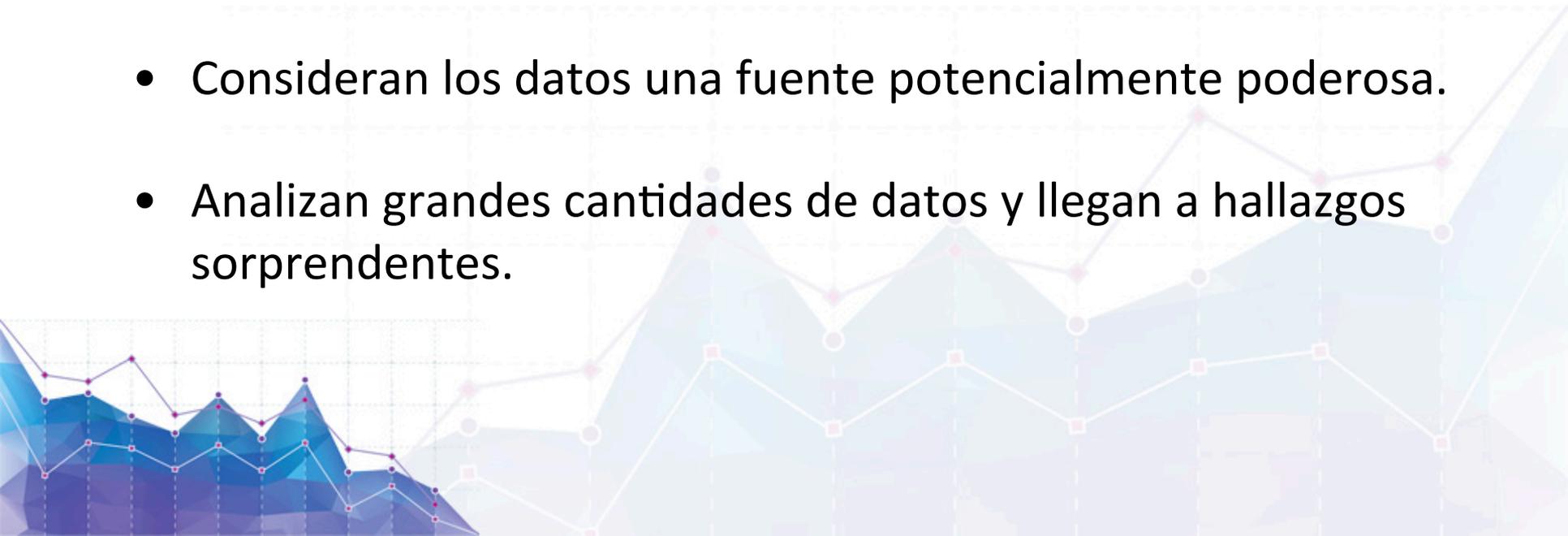
“Cuando la información era escasa, la mayor parte de nuestros esfuerzos estaban dedicados a buscarla y recogerla. Ahora que hay información abundante, el procesamiento es más importante”.

Philip Meyer, profesor emérito y profesor titular de la cátedra Knight de Periodismo de la Universidad de Carolina del Norte.

Los cambios

La reportería tradicional a la que nos hemos acostumbrado comienza a cambiar desde el momento en que los periodistas:

- Deciden comenzar a procesar grandes cantidades de datos.
- Consideran los datos una fuente potencialmente poderosa.
- Analizan grandes cantidades de datos y llegan a hallazgos sorprendentes.



El nacimiento

Sandra Crucianelli, especialista en periodismo de precisión, dice que el PD no es nuevo.

Comienza a conocerse con ese nombre a partir de hace un par de años.

En los 80 se postuló que el método científico podía aplicarse a la investigación periodística.

20 años antes Philip Meyer dijo que los métodos de la investigación social podían aplicarse al periodismo.

¿Qué es PD?

Es entonces el periodismo de datos, investigación en profundidad, la cual utiliza las herramientas ofrecidas por la tecnología para hacer un trabajo mucho más completo.

¿Qué incluye?

1. Métodos del periodismo en profundidad.
2. Precisión.
3. Análisis de los datos.

¿Qué hace al PD diferente?

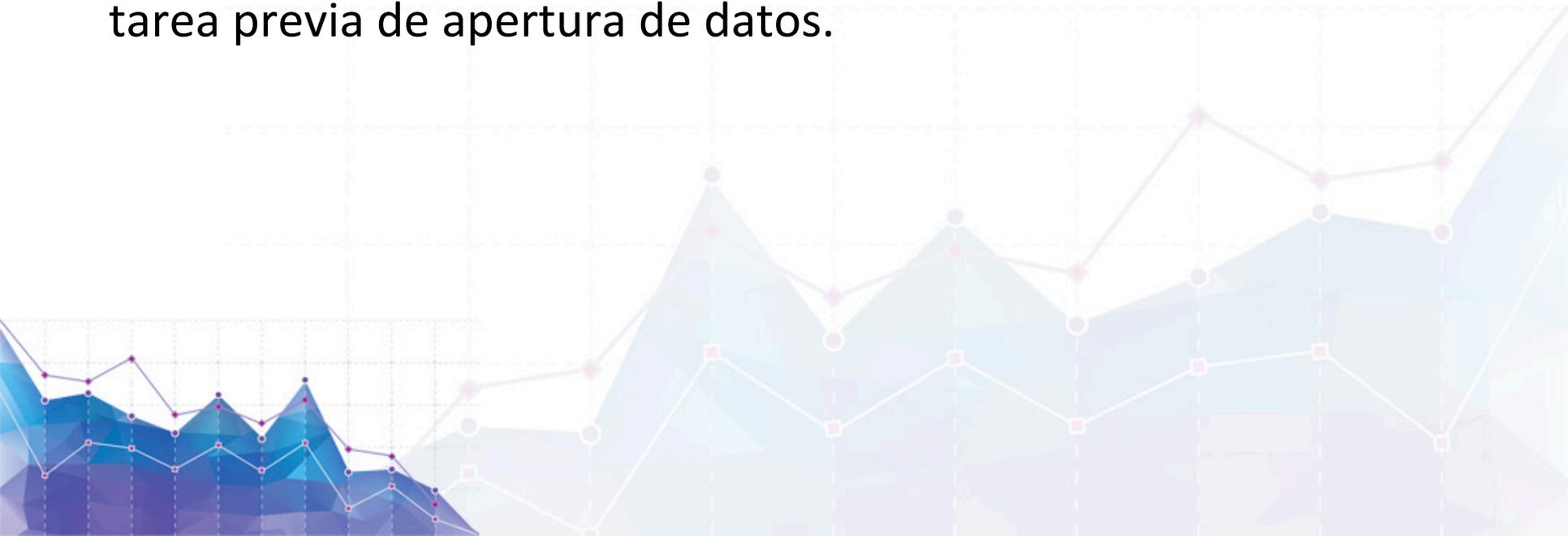
Que combina el tradicional “olfato para las noticias” y la capacidad de narrar una historia convincente, con la escala y alcance de la información digital disponible.

Acerca la información incomprensible que producen los gobiernos y las empresas que manejan nuestros intereses y recursos a los ciudadanos.

Plantea una nueva forma de narrar las historias a partir del periodismo para explicar.

¿Con qué trabaja el PD?

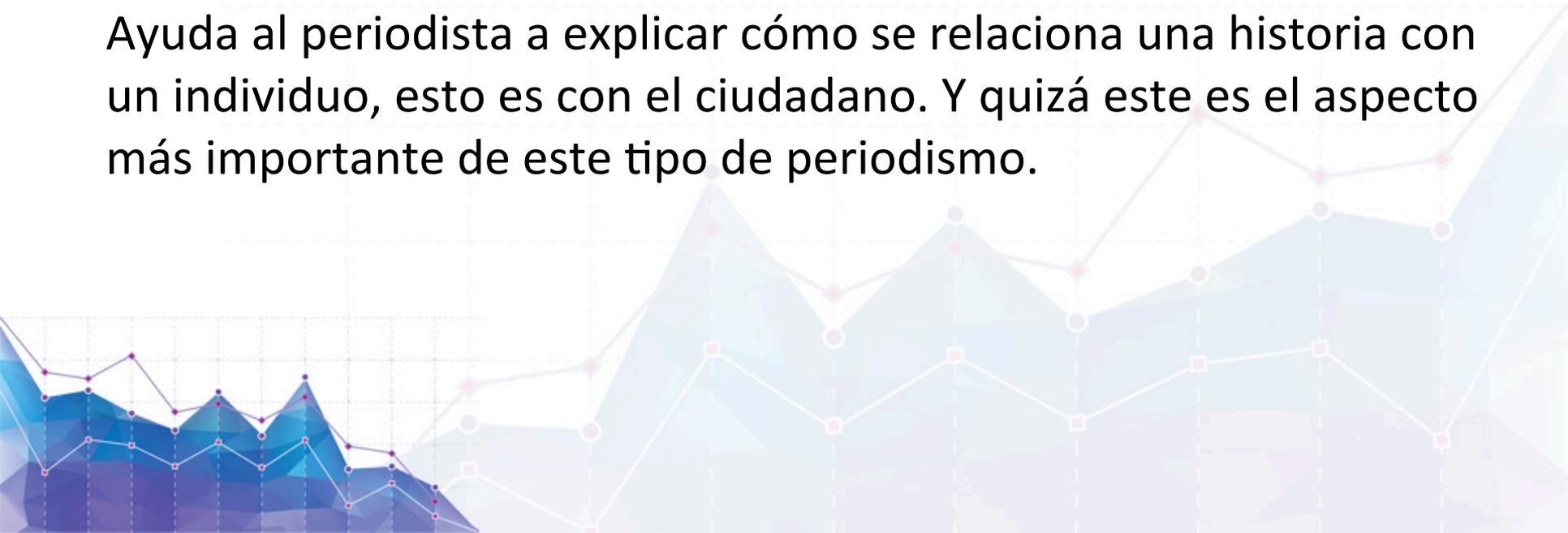
Con un volumen enorme de datos abiertos, muchos de los cuales están en formatos cerrados, por tanto, hay que hacer una tarea previa de apertura de datos.



¿Con qué fin se hace?

Ayuda al periodista a contar una historia convincente con una carga de datos alta.

Ayuda al periodista a explicar cómo se relaciona una historia con un individuo, esto es con el ciudadano. Y quizá este es el aspecto más importante de este tipo de periodismo.



¿Qué se requiere para hacer PD?

Armar un equipo apasionado. Un grupo de personas que de manera constante se esté formulando preguntas a partir de lo que no comprende la ciudadanía y frente a lo cual requiere respuestas.



3 aspectos a tener en cuenta

Si en el periodismo de investigación somos curiosos, en el de Datos auscultamos de más.

En periodismo de investigación consumimos enormes cantidades de documentos; en el de datos todo lo que vamos revisando lo vamos procesando administrándole una categoría de búsqueda que permita hacer usable y comprensible ese contenido.

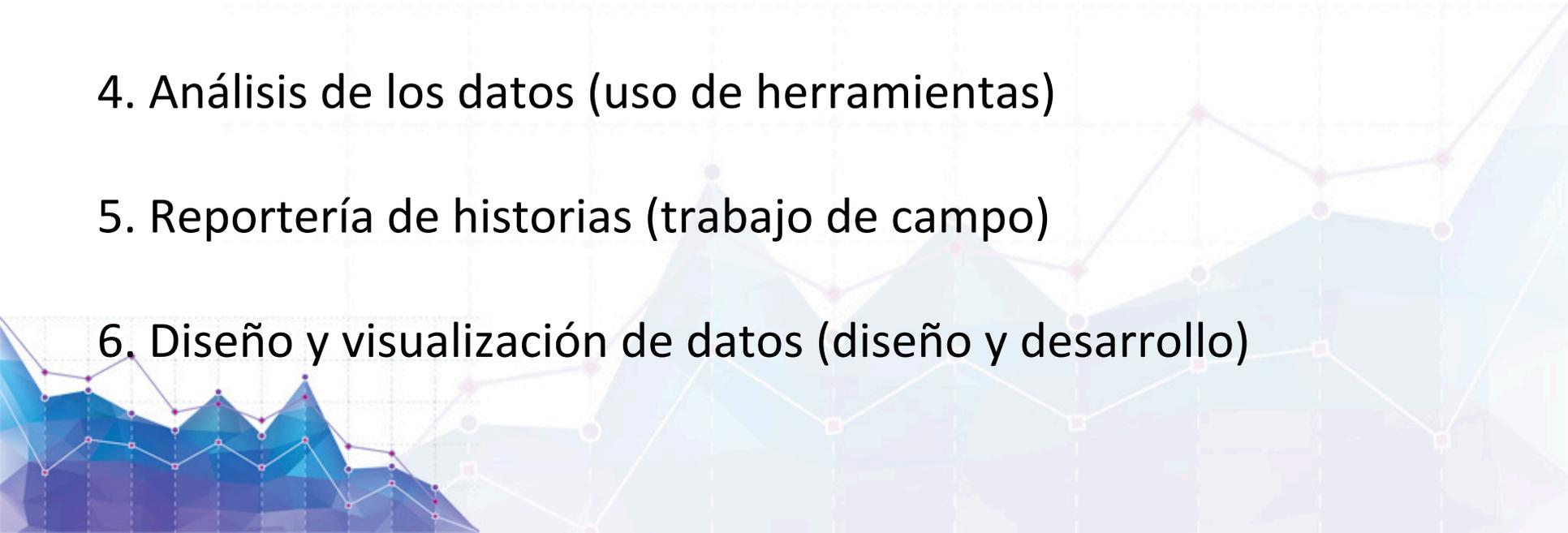
El periodismo de datos es el de la precisión, y cualquier error no solo tumba la historia, también su reputación.

Cuestión de equipo

- Periodistas.
- Ingenieros de sistemas.
- Diseñadores.
- Analistas de datos.

La riqueza de un buen equipo de periodismo de datos está en la suma de las disciplinas, que al final se complementan.

Proceso para hacer PD

1. Planeación desde la matriz (DataRakers)
 2. Minería de datos (reportería o búsqueda de información)
 3. Refinación de datos (limpieza de datos con herramientas)
 4. Análisis de los datos (uso de herramientas)
 5. Reportería de historias (trabajo de campo)
 6. Diseño y visualización de datos (diseño y desarrollo)
- 

CADA 4 MINUTOS UN PACIENTE MAL ATENDIDO PONE UNA TUTELA EN COLOMBIA

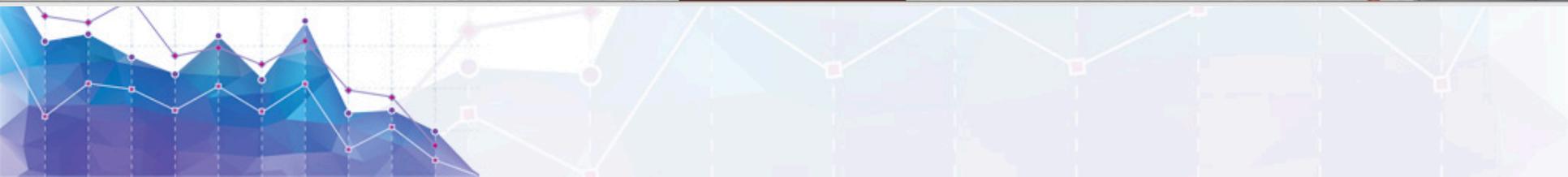
Servicios negados. Atención negligente. Pacientes que mueren en los pasillos esperando una autorización de su EPS. Hospitales atestados. En este país nos acostumbramos a padecer un sistema que funciona mal. Las cifras así lo evidencian: son cada vez más las quejas, los reclamos y las tutelas que buscan un servicio digno, que salve vidas. Mientras tanto, el Gobierno, al declarar la salud como un derecho fundamental, busca salidas a la crisis.

En 2013, los colombianos presentaron, según la Defensoría del Pueblo, 115.147 tutelas invocando el derecho a la salud para salvar su vida. La justicia sigue coadministrando la salud porque el sistema no es oportuno para salvar las vidas.

Si usted está enfermo en Colombia, lo más probable es que logre salvar su vida en juzgado y no en una entidad prestadora de los servicios de salud. Duro, pero así está sucediendo.

La salud es el segundo derecho más invocado por los colombianos, a juzgar por las tutelas que llegan a la Corte Constitucional, según la información estadística suministrada por la Defensoría. El primero es el derecho de petición y el tercero son los derechos económicos, sociales y culturales, relacionados con la dignidad humana.

[VER NOTA COMPLETA](#)



¿Para que?

Es una forma de potenciar el periodismo que nos conduce a un universo de verdades que se unifican a partir de las piezas fragmentadas que son los datos, los cuales van más allá de una simple declaración y mucho más lejos que la denuncia.



Matriz DATA RAKERS

MATRIZ

Data Rakers

Storytelling + Datos

Storytelling

¿Qué es?

Es una forma de contar historias, con las que se generan lazos emocionales con los usuarios por medio de diferentes formatos y dispositivos que le dan vida y relevancia a una historia.

¿Qué quiero que sientan los usuarios y que deben tener nuestras historias?



POSITIVAS

- Amor
- Añoranza
- Armonía
- Alegría
- Alivio
- Calma
- Cariño
- Entusiasmo
- Euforia
- Esperanza
- Excitación
- Expectativa
- Empatía
- Extasis
- Pasión
- Satisfacción
- Solidaridad
- Seguridad
- Sorpresa
- Serenidad
- Ternura

¿Qué quiero que sientan los usuarios y que deben tener nuestras historias?

NEGATIVAS

- Aburrimiento
- Angustia
- Apatía
- Ansiedad
- Asco
- Cólera
- Confusión
- Decepción
- Desamor
- Desconcierto
- Dolor
- Depresión
- Desconfianza
- Desesperación
- Desprecio
- Enfado
- Envidia
- Fastidio
- Frustración
- Fobia
- Humillación
- Indignación
- Inseguridad
- Ira
- Impotencia
- Inquietud
- Indiferencia
- Melancolía
- Pena
- Pánico
- Pereza
- Pesimismo
- Prepotencia
- Rabia
- Rechazo

¿Qué quiero que sientan los usuarios y que deben tener nuestras historias?



NEUTRAS

- Asombro
- Apego
- Curiosidad
- Compasión
- Intriga
- Nostalgia

Guía para investigar datos

Paso 1:

El recojo de los datos:

Es importante conocer el formato de los datos.

Extensión	Formato
Xlsx, xls	Excel
Csv	Texto separado por comas
Docx, doc	Word
Html	Página web
Pdf	Pdf
Pptx, ppt	PowerPoint
Mp3, mp4, wma, wav	Audio
Jpg, jpeg, png, tiff, gif, bmp	Imagen
Mp4, wmv, avi	Video
swf	Flash

Paso uno: El recojo de los datos

Convertir de CSV a Excel

	A	B	C	D	E	F	G	H	
1	Tipo de problema ambiental,	En el 2009,	En el 2010,	En el 2011,	En el 2012,	En el 2013,	Explicacion,	Fecha,	Links
2	Migraciones,	17833,	17192,	17746,	22070,	20938,	bla bla bla,	2/11/2016,	www.eltiempo.com
3	Sin alimentos,	14498,	12974,	15276,	21435,	22549,	bla bla bla,	3/11/2016,	www.eltiempo.com
4	Desnutricion,	59241,	60086,	66393,	80864,	91860,	bla bla bla,	4/11/2016,	www.eltiempo.com
5	Enfermos por vectores,	8405,	8098,	10248,	9228,	8933,	bla bla bla,	5/11/2016,	www.eltiempo.com
6	Afectados por inundaciones,	13147,	13344,	20432,	22651,	23200,	bla bla bla,	6/11/2016,	www.eltiempo.com

	A	B	C	D	E	F	G	H	I	J
1	Tipo de problema ambiental	En el 2009	En el 2010	En el 2011	En el 2012	En el 2013	Explicacion	Fecha	Links	
2	Migraciones	17833	17192	17746	22070	20938	bla bla bla	2/11/16	www.eltiempo.com	
3	Sin alimentos	14498	12974	15276	21435	22549	bla bla bla	3/11/16	www.eltiempo.com	
4	Desnutricion	59241	60086	66393	80864	91860	bla bla bla	4/11/16	www.eltiempo.com	
5	Enfermos por vectores	8405	8098	10248	9228	8933	bla bla bla	5/11/16	www.eltiempo.com	
6	Afectados por inundaciones	13147	13344	20432	22651	23200	bla bla bla	6/11/16	www.eltiempo.com	

Paso uno: El recojo de los datos

- ¿Hay bases de datos sobre el tema?
- ¿Cómo y para qué fueron hechas?
- ¿Están en una web oficial o debo hacer un pedido de acceso a información?
- ¿Si la base está en una web es descargable? ¿Cómo la puedo bajar?
- Si debo construir una base ¿qué variables debo incluir y qué podré demostrar?

Paso dos: depuración y contexto

Puede haber errores. Registros duplicados, casillas incompletas, palabras mal escritas. Hay que buscar herramientas para identificar estos problemas:

- ¿Uso excel?
- ¿Uso Refine?
- ¿En qué caso usar gestores de más capacidad como Mongo DB?

Paso dos: Depuración y contexto

- ¿Conozco todos los términos y variables que hay en las bases de?
- ¿Son los mismos que los usados en bases similares?
- ¿Los criterios apuntan al sentido de la pregunta que quiero responder o necesito entender cómo se construyeron?

Paso tres: Cruce y análisis

El valor de sus hallazgos depende de la calidad de las preguntas y de la combinación de dos o más registros para encontrar coincidencias reveladoras:

- ¿Tienen mis bases de datos un código común que me permita cruzarlas DNI, Código DANE, nombres completos (pero estos a veces generan problemas)

Paso tres: Cruce y análisis

- El cruce de las bases de datos muestra:
 1. Tendencias.
 2. Patrones
 3. Procesos evolutivos en un periodo determinado.

En todos pregunta: ¿EN QUÉ CONTEXTO?

Paso cuatro: Verificación

El periodista de investigación debe aplicar la metodología tradicional: ir a los lugares necesarios, entrevistar a las personas involucradas, revisar nuevos documentos para detectar debilidades y fortalezas de la base de datos:

- ¿Reflejan los datos la condición real de las personas?
- ¿Con qué experto puedo validar la metodología del cruce?
- Es posible que el hallazgo sea correcto, pero admita más de una interpretación?

Paso cinco: verificación

Hay que pensar desde el inicio en la forma más eficiente de presentar los hallazgos:

- Librerías **D3js.org** y repositorios software como **Gigithub** con ejemplos que uno puede adaptar a lo que necesita.

Pregúntese:

- ¿Conviene más una visualización o una aplicación?
- ¿Cómo debe ser la experiencia de usuario?
- ¿Qué es lo que la herramienta debe generar en el lector?
- ¿Qué elementos de mi visualización o aplicación la hacen necesaria para el usuario?

10 bases de datos para rastrear el dinero

ICIJ Offshore Leaks Database

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

Tiene datos de más de 100 mil compañías y fondos guardados en paraísos fiscales. Es parte de un archivo de 2,5 millones de documentos filtrados al Consorcio Internacional de Periodistas de Investigación (ICIJ).

10 bases de datos para rastrear el dinero

OpenCorporates

 <https://opencorporates.com/>

Contiene información sobre 80 millones de compañías y 90 millones de directores en más de 100 países. Se puede buscar por nombre de empresa, dirección y nombres de los directores.

10 bases de datos para rastrear el dinero

Open Spending

 <https://openspending.org/>

Sitio web que sigue el gasto público de los gobiernos en todo el mundo y lo presenta en distintas formas de visualización.

10 bases de datos para rastrear el dinero

Registro de propiedades de Miami

 [http:// www.miamidade.gov](http://www.miamidade.gov)

Base de datos libre del Condado de Miami Dade, que permite localizar propiedades en Miami, por nombre del dueño.

10 bases de datos para rastrear el dinero

Investigative Dashboard

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

Permite buscar información de los accionistas, directores y los reportes financieros de compañías alrededor del mundo. Hay enlaces a más de 450 bases de datos en línea de 120 países. Plataforma construida por el Organized Crime and Corruption Reporting Project (Occrp).

10 bases de datos para rastrear el dinero

Search Systems

 <http://publicrecords.searchsystems.net>

Portal especializado en la búsqueda de personas. Contiene más de 55 mil bases de datos organizadas por fecha de nacimiento, fecha de defunción, matrimonio, licencias, acciones, hipotecas, entre otras subdivisiones.

10 bases de datos para rastrear el dinero

Registro de compañías de Panamá

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

Sitio independiente que reordenó la data del registro público de Panamá para facilitar la investigación sobre más de 300 mil empresas inscritas en este país. Permite hacer búsquedas por nombre de personas.

10 bases de datos para rastrear el dinero

Personas de interés

 <https://www.personadeinteres.org>

Da acceso a archivos judiciales, registros de propiedades e informes de inteligencia sobre gente vinculada al crimen organizado, narcotráfico, corrupción, etc.

Tiene información sobre procesados por narcotráfico en Perú.

10 bases de datos para rastrear el dinero

FlightAware

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

Es una plataforma para rastrear los vuelos y su estado en todo el mundo. Ofrece una aplicación descargable en el celular. Permite buscar por el nombre de la línea aérea, el número de vuelo, la ruta y la matrícula de la aeronave.

10 bases de datos para rastrear el dinero

Marine traffic

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

Es una base de datos actualizada en tiempo real que permite rastrear la localización de cualquier tipo de barco o navío, además de las salidas, llegadas y recorridos.

Recursos digitales para cazar historias

ScraperWiki

<https://scraperwiki.com/>

Plataforma online que permite descargar información de la web y agruparla de forma ordenada en una base de datos excel, CSV u otro formato.



Move beyond digital paper

Improve use of data in your organisation by:

-  **1. Building skills**
ScraperWiki helps teams learn a bit of programming to fix real operational problems.
 - Spreadsheets into data using Python at the ONS
 - "Code in Browser" product which started the company
-  **2. Collaboration**
ScraperWiki is expert at data publishing, by caring about user needs first.
 - Sharing humanitarian data at the United Nations
 - Statistical summation for the Civil Service People Survey
-  **3. Artificial intelligence**
ScraperWiki uses complex algorithms to automate data in production.
 - Extract tables from PDFs with our product PDF Tables

Recursos digitales para cazar historias

DocumentCloud

<https://www.documentcloud.org>

Plataforma para gestionar documentos. Extrae el texto de una imagen mediante software. Permite resaltar datos, hacer anotaciones y organizarlas en links de fácil acceso. Ayuda a buscar por temas, embeber documentos y colocarlas en un catálogo público.

Overview

A tool for journalists, a document catalog, both.

Search

Our catalog has more than one million public documents.

Open Source

Our work is open source. Contribute your ideas.

Get Help

Learn about DocumentCloud's advanced features and API.

Who We Are

Meet the people making this happen.

The screenshot shows the DocumentCloud search interface. At the top, there are tabs for 'Search' and 'Help'. Below them is a search bar containing the text 'project: "Deepwater Horizon"'. Underneath the search bar are buttons for 'Open', 'Edit', 'Projects', 'Analyze', and 'Publish'. The main area displays a grid of document thumbnails with titles and small icons. The titles include: 'Advance Chapter on BP Well Blowout Investigation - 55 notes', 'Oil Spill Investigators' Letter to Commissioners', 'Report Shows Spill Poses Little Additional Risk', 'March 2010 BP Strategy Presentation', 'Wall Street on BP', and 'Deepwater Horizon Statement'.

DocumentCloud

Analyze, Annotate, Publish.
Turn documents into data.

The screenshot shows a document page with a table of financial data. The table is titled 'Type of Problem Change' and 'Positive Change'. The data is as follows:

Type of Problem Change	Positive Change
Forward Growth (a positive increase due to increased price and/or volume in a segment and/or activity, even which there is little to no discretion.)	8
Initiation (a positive increase due to discretionary program change versus changes.)	22
Internal Transfer (a positive transfer within the same department.)	11
Reduction (a positive decrease consistent with the company's stated strategy to reduce expenditures by \$100 million.)	(10)
Buyout (a positive decrease due to the ending of an entire activity.)	(10)
TOTAL	(97)

Recursos digitales para cazar historias

Visual Investigative Scenarios

<https://vis.occrp.org/>

Una herramienta de especial interés para los periodistas de investigación. Permite establecer relaciones entre personas y organizaciones y adjuntar documentos que prueben esa relación.



Recursos digitales para cazar historias

NEO4J

<https://neo4j.com/>

Permite identificar conexiones entre grandes cantidades de datos y mostrarlas en gráficos.



Experience Neo4j 3.0 -
the Next Major Evolution
in Graph Database
Technology



Búsqueda en la web

- Búsqueda avanzada



Búsqueda avanzada

Mostrar páginas que contengan...

todas estas palabras:

esta palabra o frase exactas:

cualquiera de estas palabras:

ninguna de estas palabras:

números del:

al

Para hacer esto en el cuadro de búsqueda

Ingresar las palabras importantes: `Terrier ratonero tricolor`.

Ingresar las palabras exactas entre comillas: `"Terrier ratonero"`.

Ingresar OR entre las palabras que desees:
En miniatura OR estándar.

Ingresar un signo menos justo delante de las palabras que no desees que aparezcan: `-Roedor, -"Jack Russell"`.

Ingresar dos puntos entre los números y agregar una unidad de medida:
`10..35 lb, $300..$500, 2010..2011`.

Luego restringe tus resultados por...

idioma:

Busca páginas en el idioma que seleccionas.

región:

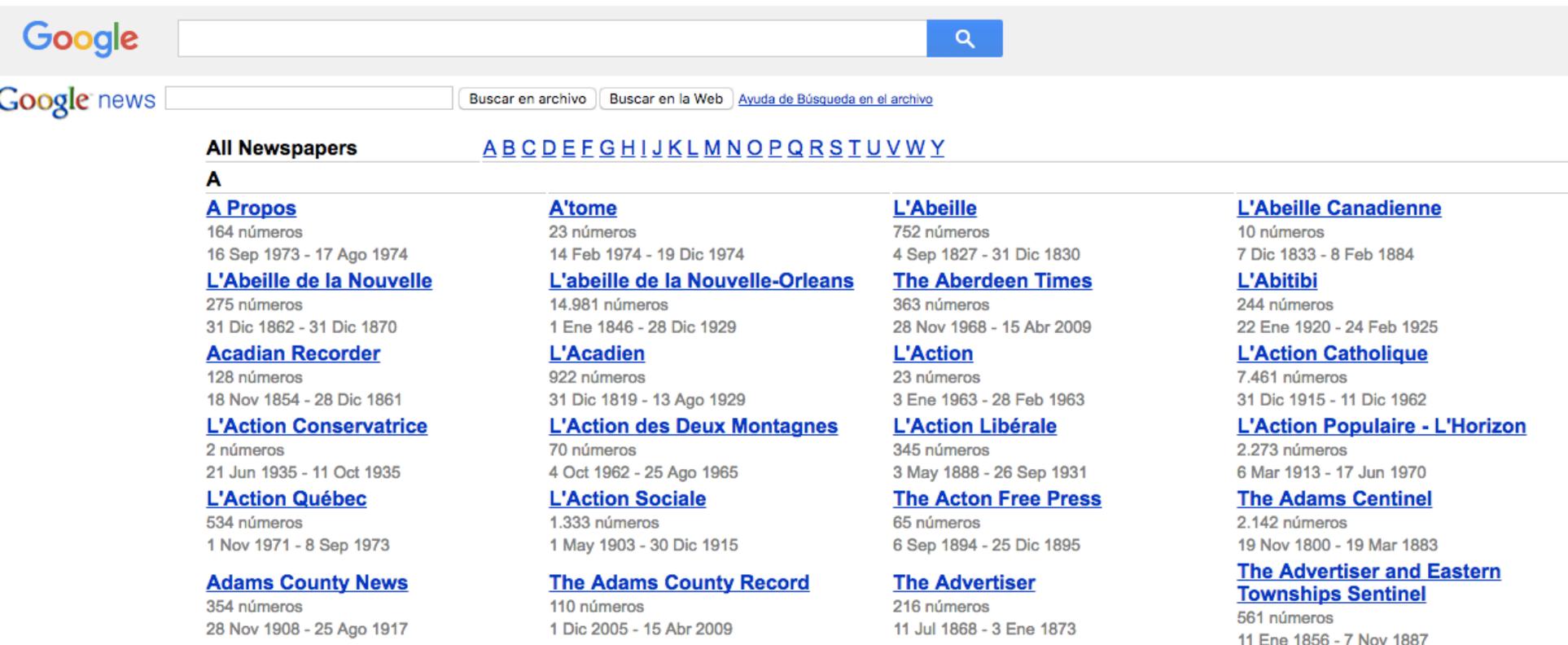
Busca páginas publicadas en una región determinada.

última actualización:

Busca páginas actualizadas en el transcurso del período que especificas.

Búsqueda de ediciones antiguas de periódicos en la web

- <https://news.google.com/newspapers?hl=es>



The screenshot shows the Google News interface. At the top is the Google logo and a search bar. Below it is the 'Google news' logo and a search bar with the text 'Buscar en archivo', 'Buscar en la Web', and 'Ayuda de Búsqueda en el archivo'. The main content area is titled 'All Newspapers' and features a navigation bar with letters A through Y. The 'A' section is expanded, displaying a grid of newspaper listings. Each listing includes the newspaper's name, the number of issues, and the date range.

All Newspapers			
A B C D E F G H I J K L M N O P Q R S T U V W Y			
A			
A Propos 164 números 16 Sep 1973 - 17 Ago 1974	A'tome 23 números 14 Feb 1974 - 19 Dic 1974	L'Abeille 752 números 4 Sep 1827 - 31 Dic 1830	L'Abeille Canadienne 10 números 7 Dic 1833 - 8 Feb 1884
L'Abeille de la Nouvelle 275 números 31 Dic 1862 - 31 Dic 1870	L'abeille de la Nouvelle-Orleans 14.981 números 1 Ene 1846 - 28 Dic 1929	The Aberdeen Times 363 números 28 Nov 1968 - 15 Abr 2009	L'Abitibi 244 números 22 Ene 1920 - 24 Feb 1925
Acadian Recorder 128 números 18 Nov 1854 - 28 Dic 1861	L'Acadien 922 números 31 Dic 1819 - 13 Ago 1929	L'Action 23 números 3 Ene 1963 - 28 Feb 1963	L'Action Catholique 7.461 números 31 Dic 1915 - 11 Dic 1962
L'Action Conservatrice 2 números 21 Jun 1935 - 11 Oct 1935	L'Action des Deux Montagnes 70 números 4 Oct 1962 - 25 Ago 1965	L'Action Libérale 345 números 3 May 1888 - 26 Sep 1931	L'Action Populaire - L'Horizon 2.273 números 6 Mar 1913 - 17 Jun 1970
L'Action Québec 534 números 1 Nov 1971 - 8 Sep 1973	L'Action Sociale 1.333 números 1 May 1903 - 30 Dic 1915	The Acton Free Press 65 números 6 Sep 1894 - 25 Dic 1895	The Adams Centinel 2.142 números 19 Nov 1800 - 19 Mar 1883
Adams County News 354 números 28 Nov 1908 - 25 Ago 1917	The Adams County Record 110 números 1 Dic 2005 - 15 Abr 2009	The Advertiser 216 números 11 Jul 1868 - 3 Ene 1873	The Advertiser and Eastern Townships Sentinel 561 números 11 Ene 1856 - 7 Nov 1887