

OjoPúblico

Narrar el conflicto desde el ánalisis de datos

Nelly Luna Amancio

Fundadora y editora general en Ojo-Publico.com

@nellylun / @Ojo_Publico / @OjoLab

Huancayo, agosto 2017



OjoPúblico

- Ojo-Público es una organización periodística sin fines de lucro constituida en Lima el 2014 y es en este momento el medio digital peruano dedicado al periodismo de investigación de mayor alcance en el país y el extranjero.
- Medio digital y laboratorio de innovación para investigar el poder. Miembro del Global International Journalism Network (GIJN) y socio del Consorcio Internacional de Periodistas de Investigación (ICIJ).
- Equipo de periodistas y programadores: 12 integrantes. Ganador del Data Journalism Awards 2015 / Premio Nacional de Derechos Humanos en Perú 2015 / Ganador del Tercer Premio #Colpin2016
- Promovemos la formación de periodistas a través del programa de entrenamiento @OjoLab.

OjoPúblico



Consideraciones para la cobertura de conflictos socioambientales





Consideraciones para el oficio

- No somos activistas.
- Transversalidad del tema: economía, política, comercio exterior, derecho
- Las fuentes: conocer la estructura del Estado e identificar la importancia de la fuente científica.
- Mirada intercultural: Entender al otro
- Los riesgos en el tratamiento: prejuicios, paternalismo
- Riesgos en el reporte: violencia, corrupción



¿Por qué trabajamos con bases de datos?

- Para profundizar el contexto y desnudar un sistema.
- Para dismitificar el conflicto
- Para ir más allá de la cobertura diaria y del reporte cotidiano de los hechos.
- Para aproximarnos al origen y causas de un hecho.
- Para investigar los patrones y las prácticas sistemáticas de los grupos de poder.

OjoPúblico

**¿Qué es el análisis de
datos aplicado al
periodismo?**





¿De qué hablamos cuando
hablamos de periodismo de datos?

El dato no es **El** dato: son los
millones de datos estructurados

OjoPúblico

The scale of the leak

Volume of data compared to previous leaks

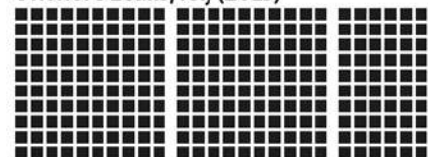
1,7 GB

Cablegate/Wikileaks (2010)



260 GB

Offshore Leaks/ICIJ (2013)



4 GB

Luxemburg Leaks/ICIJ (2014)



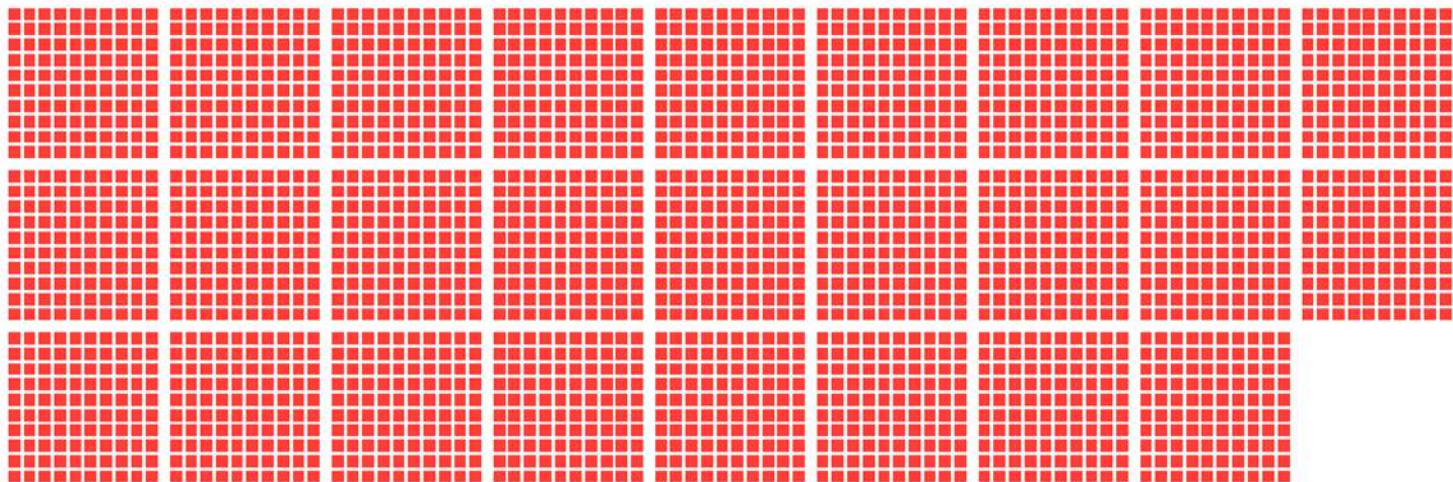
3,3 GB

Swiss Leaks/ICIJ (2015)



≈ 2,6 TB

Panama Papers/ICIJ (2016)



OjoPúblico

El objetivo del análisis de datos en el periodismo de investigación no es la visualización:

Método + Noticia + Narrativa

Revelar

OjoPúblico

**Las fuentes humanas son
irremplazables e imprescindibles.**

¿Cómo elaborar un proyecto de investigación basado en análisis de datos?

[Adiós al lobo solitario]

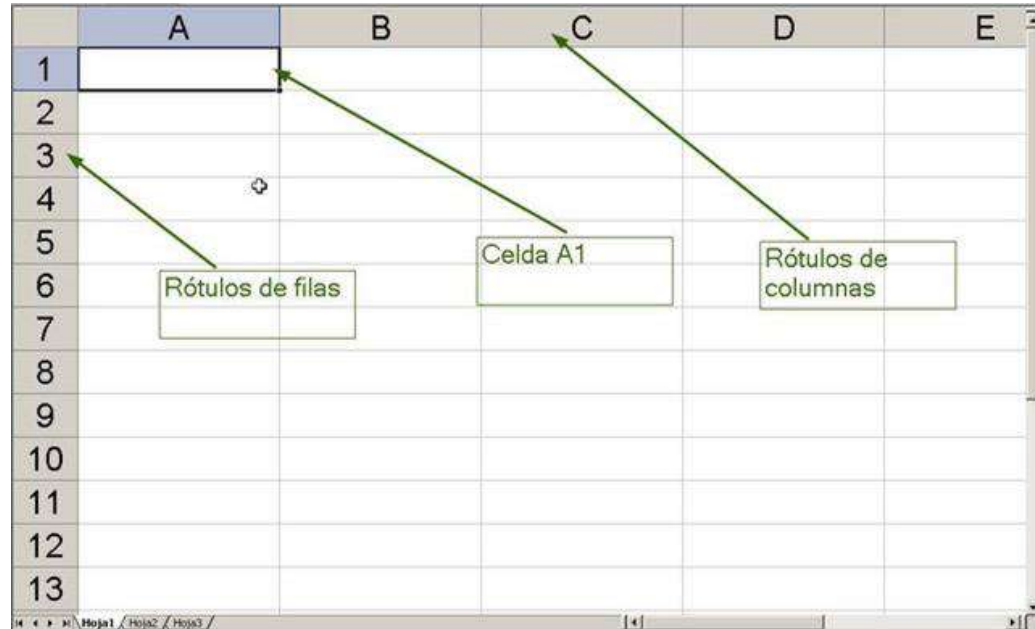


¿Qué es una base de datos?

Una base de datos es un conjunto de información organizada de forma que un programa de un computador pueda seleccionar rápidamente los fragmentos de datos que se necesiten.

Una hoja de cálculo

Documento que nos permite analizar y sistematizar datos (numéricos y alfanuméricos) ordenados en forma de tablas bidimensionales: filas y columnas). La celda es la unidad básica de información en la hoja de cálculo.



Los formatos de los datos

Hojas de cálculo:

.xls (Los archivos XLS, Xml, XLSX las genera microsoft Excel)

.ods (archivos Libre Office, paquete de software libre y de código abierto)

.csv (comma-separated values. Formato abierto)

.json acrónimo de JavaScript Object Notation, es un formato de texto ligero para el intercambio de datos.

Consulta amigable MEF

Tor Browser		Graficar		Año 2017 Actividades/Proyectos									
¿Quién gasta?		¿En qué se gasta?		¿Con qué se financian los gastos?		¿Cómo se estructura el gasto?		¿Dónde se gasta?		¿Cuándo se hizo el gasto?			
Ejecutora		Proyecto/Proyecto	Función	Fuente	Rubro	Genérica		Departamento		Trimestre	Mes		
TOTAL						142,471,518,545	169,693,361,544	133,690,793,675	109,551,593,570	82,190,330,268	75,993,763,527	73,463,903,982	44.8
Nivel de Gobierno R. GOBIERNOS REGIONALES						22,218,924,639	30,309,814,794	22,428,300,131	18,606,726,357	14,555,773,720	14,069,515,719	13,562,543,118	46.4
Sector 99: GOBIERNOS REGIONALES						22,218,924,639	30,307,912,189	22,428,412,025	18,605,942,825	14,554,990,189	14,068,742,987	13,561,776,390	46.4
Pilego 450: GOBIERNO REGIONAL DEL DEPARTAMENTO DE JUNIN						1,179,576,382	1,754,369,926	1,229,212,420	1,089,547,634	876,322,780	856,876,154	815,347,823	48.8
Categoria Presupuestal						PIA	PIM	Certificación	Compromiso Anual	Atención de Compromiso Mensual	Devengado	Girado	Avance %
Buscar Item						Por código Por descripción		¿Cómo buscar?					
0001: PROGRAMA ARTICULADO NUTRICIONAL						49,356,464	57,593,953	37,731,498	31,064,480	27,652,612	27,070,487	26,144,559	47.0
0002: SALUD MATERNO NEONATAL						42,243,044	124,862,353	106,685,373	86,608,385	59,938,133	59,102,380	58,130,879	47.3
0016: TBC-VIH/SIDA						8,897,096	14,618,657	8,132,098	6,858,227	6,140,135	6,050,215	5,864,195	41.4
0017: ENFERMEDADES METAXENICAS Y ZOOZOSIS						9,919,105	11,621,207	8,529,874	6,001,357	5,780,628	5,634,445	5,499,081	48.5
0018: ENFERMEDADES NO TRANSMISIBLES						7,419,550	11,284,757	6,853,388	5,692,591	5,281,278	5,184,137	4,994,367	45.9
0024: PREVENCIÓN Y CONTROL DEL CÁNCER						5,115,166	66,624,912	62,352,190	59,422,032	46,100,420	45,873,671	45,806,984	68.9
0042: APROVECHAMIENTO DE LOS RECURSOS HIDRICOS PARA USO AGRARIO						0	15,861	0	0	0	0	0	0.0
0046: ACCESO Y USO DE LA ELECTRIFICACION RURAL						0	7,996	0	0	0	0	0	0.0
0047: ACCESO Y USO ADECUADO DE LOS SERVICIOS PUBLICOS DE TELECOMUNICACIONES E INFORMACION ASOCIADOS						388,599	388,569	285,953	285,953	183,301	183,301	182,973	47.2
0051: PREVENCIÓN Y TRATAMIENTO DEL CONSUMO DE DROGAS						1,889,832	1,905,837	1,618,831	1,603,300	982,675	964,096	889,173	50.6
0057: CONSERVACION DE LA DIVERSIDAD BIOLÓGICA Y APROVECHAMIENTO SOSTENIBLE DE LOS RECURSOS NATURALES EN ÁREA NATURAL PROTEGIDA						986,566	1,162,366	1,092,131	156,126	87,376	62,053	53,253	5.3
0068: REDUCCION DE VULNERABILIDAD Y ATENCION DE EMERGENCIAS POR DESASTRES						4,644,042	8,851,749	5,947,652	3,925,886	3,228,492	2,952,212	2,585,935	33.4
0072: PROGRAMA DE DESARROLLO ALTERNATIVO INTEGRAL Y SOSTENIBLE - PIRDAIS						0	448,422	319,893	301,809	300,309	139,352	139,352	31.1
0082: PROGRAMA NACIONAL DE SANEAMIENTO URBANO						12,211,688	45,855,703	20,055,160	19,412,976	7,340,303	6,460,772	2,359,058	14.1
0083: PROGRAMA NACIONAL DE SANEAMIENTO RURAL						3,062,538	5,301,033	4,903,082	4,649,821	3,396,540	3,241,502	3,231,322	61.1
0090: LOGROS DE APRENDIZAJE DE ESTUDIANTES DE LA EDUCACION BASICA REGULAR						497,074,626	589,070,376	381,953,064	341,382,466	330,996,093	329,936,960	322,240,613	56.0
0091: INCREMENTO EN EL ACCESO DE LA POBLACION DE 3 A 16 AÑOS A LOS SERVICIOS EDUCATIVOS PUBLICOS DE LA EDUCACION BASICA REGULAR						625,940	15,837,813	12,016,147	6,816,465	5,990,670	5,732,047	5,456,520	36.2
0104: REDUCCION DE LA MORTALIDAD POR EMERGENCIAS Y URGENCIAS MEDICAS						2,016,026	2,075,673	1,467,795	1,314,924	1,072,166	1,041,637	1,025,726	50.2
0106: INCLUSION DE NIÑOS, NIÑAS Y JOVENES CON DISCAPACIDAD EN LA EDUCACION BASICA Y TECNICO PRODUCTIVA						4,013,332	4,322,067	2,433,564	2,101,868	2,076,470	2,068,251	2,007,309	47.8
0107: MEJORA DE LA FORMACION EN CARRERAS DOCENTES EN INSTITUTOS DE EDUCACION SUPERIOR NO UNIVERSITARIA						4,713,422	5,369,598	2,768,821	2,760,517	2,760,517	2,760,517	2,677,611	51.4

MEF: formato hoja de cálculo

Consulta Amigable Consulta de Ejecución del Gasto

Fecha de la Consulta: 26-julio-2017

Año de Ejecución: 2017

Incluye: Actividades y Proyectos

TOTAL	142471518545	169693391544	133690793675	109551593570	82190330268	75993763527
Nivel de Gobierno R: GOBIERNOS REGIONALES	22218924639	30309814794	22428300131	18606726357	14555773720	14069515719
Sector 99: GOBIERNOS REGIONALES	22218924639	30307912189	22426412025	18605942825	14554990189	14068742987
Pliego 450: GOBIERNO REGIONAL DEL DEPARTAMENTO DE JUNIN	1179576582	1754369928	1229212420	1089547634	878322780	856878154
Categoría Presupuestal	PIA	PIM	Certificación	Compromiso A	Ejecución Atención de C	Devengado
0001: PROGRAMA ARTICULADO <u>NUTRICIONAL</u>	49356464	57593953	37731498	31064480	27652612	27070487
0002: SALUD MATERNO NEONATAL	42243044	124882353	106685373	86608385	59938133	59102380
0016: <u>TBC-VIH/SIDA</u>	8867096	14618657	8132098	6858227	6140135	6050215
0017: ENFERMEDADES <u>METAXENICAS Y ZOONOSIS</u>	9919105	11621207	8529874	6001357	5780628	5634445
0018: ENFERMEDADES NO TRANSMISIBLES	7419560	11284757	6853388	5692591	5281278	5184137
0024: <u>PREVENCION Y CONTROL DEL CANCER</u>	5115166	66624912	62352190	59422032	46100420	45873971
0042: APROVECHAMIENTO DE LOS RECURSOS <u>HIDRICOS PARA USO</u>	0	15861	0	0	0	0
0046: <u>ACCESO Y USO DE LA ELECTRIFICACION RURAL</u>	0	7996	0	0	0	0
0047: <u>ACCESO Y USO ADECUADO DE LOS SERVICIOS PUBLICOS DE</u>	388569	388569	285953	285953	183301	183301
0051: <u>PREVENCION Y TRATAMIENTO DEL CONSUMO DE DROGAS</u>	1889832	1905837	1618831	1603300	982675	964096
0057: <u>CONSERVACION DE LA DIVERSIDAD BIOLÓGICA Y APROVECHAM</u>	986566	1162366	1092131	156126	87376	62053
0068: <u>REDUCCION DE VULNERABILIDAD Y ATENCION DE EMERGE</u>	4644042	8851749	5947652	3925886	3228492	2952212
0072: <u>PROGRAMA DE DESARROLLO ALTERNATIVO INTEGRAL Y SO</u>	0	448422	319893	301809	300309	139352
0082: <u>PROGRAMA NACIONAL DE SANEAMIENTO URBANO</u>	12211688	45855703	20055160	19412976	7340303	6460772
0083: <u>PROGRAMA NACIONAL DE SANEAMIENTO RURAL</u>	2062508	5201023	4000000	4000000	3200000	3200000

OjoPúblico

1. El dato disparador: una fuente, un documento, la recomendación de un experto. Los datos y los temas de inspiración están en todos lados.

OBSERVACIÓN.

2. Formulación del proyecto de investigación. Hipótesis, objetivos, costos, y tiempo. ¿Vamos detrás de un caso o de un sistema? Aquí se define el método. Si vamos detrás de un patrón necesitaremos un dataset.

3. Identificar la bases de datos que voy a necesitar. ¿Dónde se encuentran? Grado de dificultad, acceso y construcción. Localizar LA fuente.

OjoPúblico

4. Conformar un equipo de trabajo. Periodistas, estadísticos, programadores, economistas, ingenieros, ilustradores.

5. Construir y analizar la base de datos. Información pública, extraoficial, filtrada, construida, descargada.

6. Escribir reportes permanentes de los hallazgos. No esperar al final, la cantidad de información es abrumadora y puede desviarte del camino.

7. Prototipar la presentación de la investigación: Buscadores, visualizaciones, gestores de contenido, mapas, plataformas. NewsApps

OjoPúblico

8. El impacto. Analizar los hallazgos y entender el impacto que podría tener. Establecer alianzas con otros medios, nacionales o extranjeros.

9. Conoce a tu audiencia. Identificar el mejor timing. Entiende los tiempos y hábitos del usuario/lector. No escribes para ti ni para otros periodistas.

10. La noticia como experiencia. Con las NewsApps el lector navega y elige qué contar, dónde empezar y qué compartir. Una app no pierde vigencia.

¿Cómo construir y acorralar una base de datos?

[La persistencia hace al
periodista]



La importancia del método

1. La construcción de la base de datos
2. Estandarización y contexto
3. Cruce, análisis
4. Verificación
5. Producción

OjoPúblico

1. La construcción de los datos

- Bases de datos construidas (a través de pedidos de acceso a la información)
- Bases de datos filtradas (Panamá Papers)
- Bases de datos obtenidas extraoficialmente (El Club de los deudores)
- Bases de datos extraídas (Deudores de la corrupción)

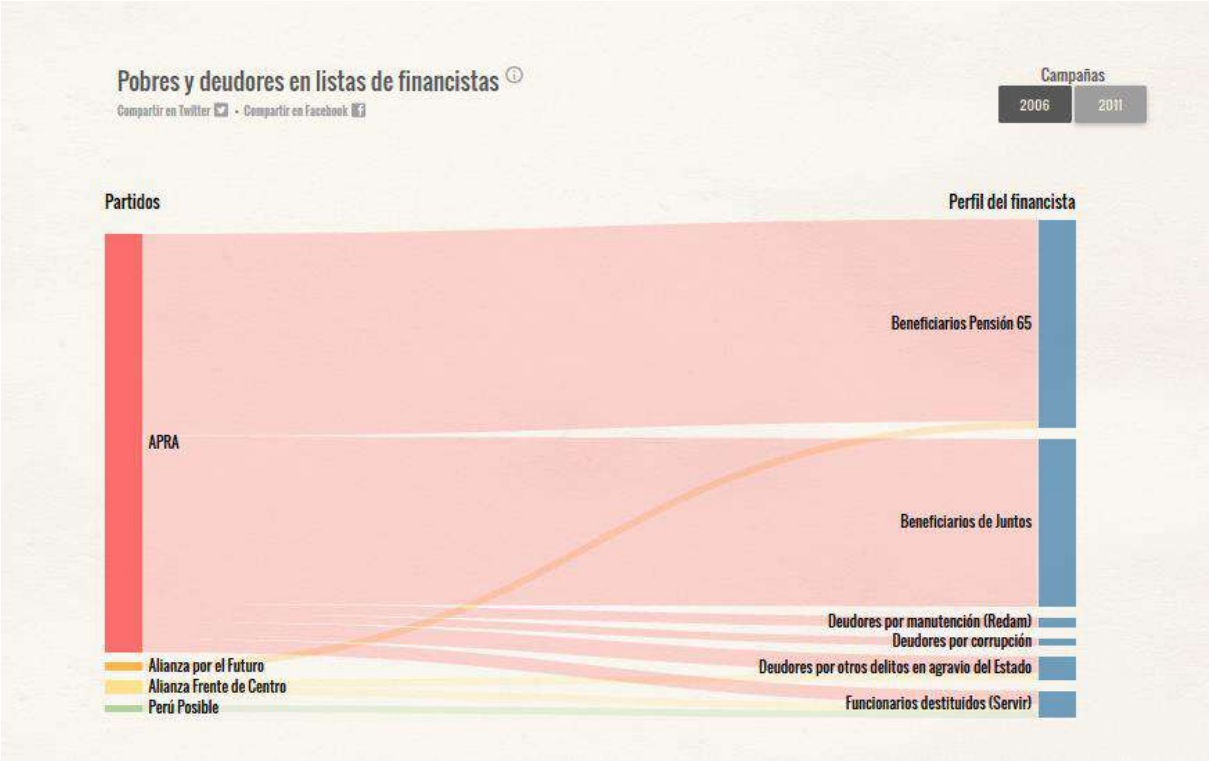
OjoPúblico

2. Estandarización, limpieza y contexto

The image shows a screenshot of a Microsoft Excel spreadsheet. The spreadsheet contains a table with the following columns: tipoOrganizacion, partido, ruc, tipo, año, periodo, fecha, procesoEli, apellidoPa, apellidoMa, nombres, tipoDoc, Num Doc, and Tipo Ap. The data rows show various records, including actions and alliances, with names and dates. Blue arrows point to specific cells: one points to the 'num Doc' column for the first row, another points to the 'num Doc' column for the row with 'MACHICA MAMANI BELISSA ZEPEDA', and a third points to the 'num Doc' column for the row with 'ROCHA VDA DE JAPLANCA MARCELA MARIANA'. A fourth arrow points to the 'procesoEli' column for the row with 'ELERA GARCIA WILMAR ALVARO'.

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
	tipoOrganizacion	partido	ruc	tipo	año	periodo	fecha	procesoEli	apellidoPa	apellidoMa	nombres	tipoDoc	Num Doc	Tipo Ap	
2	1	ACCION PO	20509374881	aporCampE	2011		3/13/2010	EG2011	FINQUINO	MERA	JORGE RAP	DNI	5254458	EFFECTI	
3	1	ACCION PO	20509374881	aporCampE	2011		3/24/12/2010	EG2011	INGA	VASQUEZ	LEONARDO	DNI	5602293	EFFECTI	
4	1	ACCION PO	20509374881	aporCampE	2011		3/24/12/2010	EG2011	MACHICA	MAMANI	BELISSA ZEP	DNI	499701	EFFECTI	
5	1	ACCION PO	20509374881	aporCampE	2011		3/24/12/2010	EG2011	MORALES	MANSILLA	PEDRO AN	DNI	19862158	EFFECTI	
5	1	ACCION PO	20509374881	aporCampE	2011		3/27/12/2010	EG2011	ALEGRIA	VARONA	GONZALO P	DNI	6513752	EFFECTI	
7	1	ACCION PO	20509374881	aporCampE	2011		3/27/12/2010	EG2011	AVILES	YEPEZ	MARCIAL E	DNI	21860659	EFFECTI	
3	1	ACCION PO	20509374881	aporCampE	2011		3/27/12/2010	EG2011	CORREA	VELASQUE	CARMELA P	DNI	20117146	EFFECTI	
3	1	ACCION PO	20509374881	aporCampE	2011		3/27/12/2010	EG2011	CUBAS	CARRANZA	JANET ISAB	DNI	16409684	EFFECTI	
0	1	ACCION PO	20509374881	aporCampE	2011		3/27/12/2010	EG2011	GUEVARA	AMASIFUEN	MESIAS AN	DNI	9871134	EFFECTI	
1	1	ACCION PO	20509374881	aporCampE	2011		3/27/12/2010	EG2011	RIOS	MARQUEZA	MARIA JULI	DNI	509665	EFFECTI	
2	1	ACCION PO	20509374881	aporCampE	2011		3/28/12/2010	EG2011	ROCHA	VDA DE JAP	BLANCA MA	DNI	10558625	EFFECTI	
4	1	ALIANZA PA	20510688661	aporCampE	2006		3/06/02/2006	EG2006	TOSO	DE VERA	RENEE SU	DNI	16447992	EFFECTI	
5	1	ALIANZA PA	20510688661	aporCampE	2006		3/08/02/2006	EG2006	SEGURA	DIAZ	GUILLERMO	DNI	16545163	EFFECTI	
7	1	ALIANZA PA	20510688661	aporCampE	2006		3/09/02/2006	EG2006	APORTE EN	PROCESO DE	IDENTIFI	OTROS		0	EFFECTI
9	1	ALIANZA PA	20510688661	aporCampE	2006		3/10/01/2006	EG2006	AGUILAR	DE CRUZ	LUISA TER	DNI	2608286	EFFECTI	
0	1	ALIANZA PA	20510688661	aporCampE	2006		3/10/01/2006	EG2006	ELERA	GARCIA	WILMAR AL	DNI	2644140	EFFECTI	
1	1	ALIANZA PA	20510688661	aporCampE	2006		3/10/01/2006	EG2006	NUÑEZ	CAMPOS	JOSE LUIS	DNI	16504031	EFFECTI	
3	1	ALIANZA PA	20510688661	aporCampE	2006		3/11/01/2006	EG2006	AZABACHE	CASTRO	CARLOS AL	DNI	18040789	EFFECTI	
4	1	ALIANZA PA	20510688661	aporCampE	2006		3/11/01/2006	EG2006	GALLARDO	GARCIA	EDUARDO	DNI	17909952	EFFECTI	
5	1	ALIANZA PA	20510688661	aporCampE	2006		3/11/01/2006	EG2006	RAMOS	BARON	WALTER AR	DNI	17837486	EFFECTI	
6	1	ALIANZA PA	20510688661	aporCampE	2006		3/11/01/2006	EG2006	SANTISTEB	CHAVEZ	VICTOR RA	DNI	17917388	EFFECTI	
9	1	ALIANZA PA	20510688661	aporCampE	2006		3/14/12/2005	EG2006	AZABACHE	CASTRO	CARLOS AL	DNI	18040789	EFFECTI	
0	1	ALIANZA PA	20510688661	aporCampE	2006		3/14/12/2005	EG2006	GALLARDO	GARCIA	EDUARDO	DNI	17909952	EFFECTI	
1	1	ALIANZA PA	20510688661	aporCampE	2006		3/14/12/2005	EG2006	RAMOS	BARON	WALTER AR	DNI	17837486	EFFECTI	
2	1	ALIANZA PA	20510688661	aporCampE	2006		3/14/12/2005	EG2006	SANTISTEB	CHAVEZ	VICTOR RA	DNI	17917388	EFFECTI	
3	1	ALIANZA PA	20510688661	aporCampE	2006		3/15/02/2006	EG2006	VALENZUE	CUELLAR	JULIA	DNI	7277158	EFFECTI	
6	1	ALIANZA PA	20510688661	aporCampE	2006		3/23/02/2006	EG2006	AZABACHE	CASTRO	CARLOS AL	DNI	16040789	EFFECTI	

3. Cruce y análisis



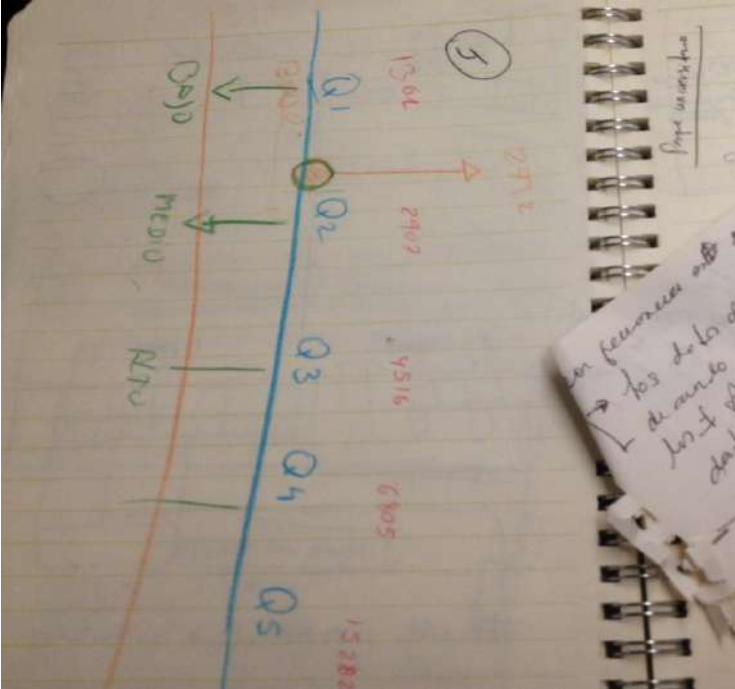
4. Verificación y contrastación

Ensúciate los zapatos



OjoPúblico

5. Producción



Bases de datos públicas



opencorporates

The largest open database of companies in the world

All jurisdictions ▾



Companies Officers

[Browse all jurisdictions](#)

Winner Open Data Business Award 2015

Geocatmin

The screenshot displays the Geocatmin web application interface. The main map shows the border region between Ecuador and Peru, with major cities like Quito, Guayaquil, Iquitos, and Lima labeled. The interface includes a top navigation bar with social media links (Facebook, YouTube) and utility buttons like 'Introducir dirección', 'Malla', 'Lista de Capas', and 'Mapa base'. A 'Lista de Capas' (Layers List) panel is open on the right, showing a list of geological layers with visibility and style controls. A 'Buscar DM' (Search DM) panel is open in the bottom left, featuring input fields for 'Código', 'Nombre', and 'Titular', along with date pickers for 'Desde' and 'Hasta', and 'Buscar' and 'Limpiar' buttons. The bottom of the page shows coordinate information and data sources.

geocatmin.ingemmet.gob.pe/geocatmin/

GEOCATMIN
Sistema Geológico Catastral Minero

Me gusta 5373

Introducir dirección

Malla

Lista de Capas

Mapa base

Lista de Capas

Visibilidad de capa

- Geología 50,000
- Geología 100,000
- Geología Regional
- Geología Integrada
- Geología Nacional I
- Metalogenético
- Certificación Ambie
- Geomorfología
- Hidrogeología
- Peligros Geológico:
- Neotectónica
- Vulcanología
- Geoquímica
- Geofísica
- Mapa Geológico Bo
- Carta Nacional
- Geografía
- Catálogo de Imágen

Buscar DM

Código:

Nombre:

Titular:

Fecha de Formulación

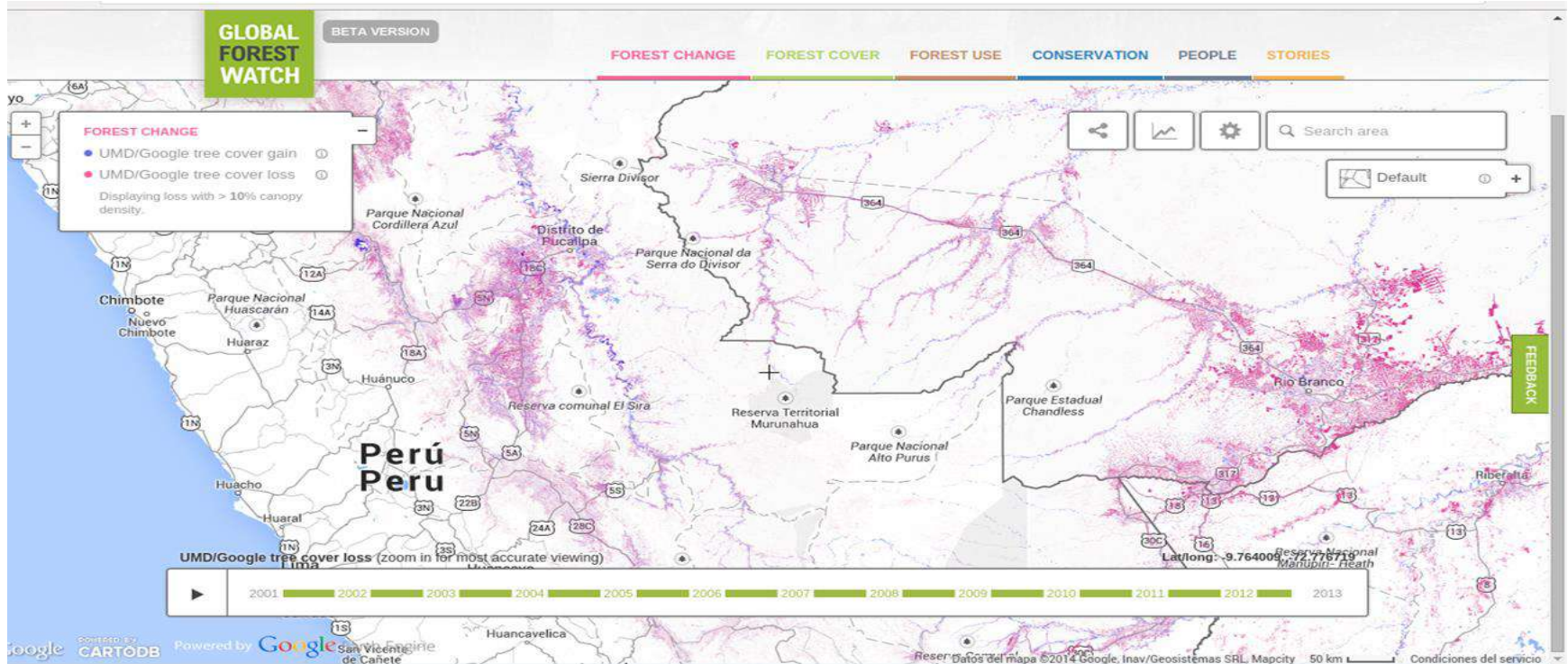
Desde: Hasta:

Buscar Limpiar

S 1984 : Esc: 9,244,648 300 millas
D X: -68°46'45.09" Y: -6°35'17.03" UTM => Este: 524409.13 Norte: 9271783.59 Zona: 19

Sources: Esri, DeLorme, USGS, NPS | Sources: Esri, USGS, NOAA

Global Forest Watch



Offshore Leaks

ICIJ

Offshore Leaks Database

Search by country

People, companies and addresses connected to offshore entities

Search by jurisdiction

Offshore companies and trusts

All countries



SEARCH

To search, enter a term in the box. Leave it blank and select a country or jurisdiction to see a full list of names.

Popular searches today: [Russia](#) · [Japan](#) · [Tony lee honkin](#) · [Eli klein holdings ltd.](#) · [Emma watson](#)



[You can download this database](#)



Got a tip? [Contact us](#) or [leak to us securely](#)

Banco de datos - BM



Search data e.g. GDP, population, Indonesia

[DataBank](#) [Microdata](#) [Data Catalog](#)



Data Portals & Tools

One of the most effective ways to display development indicators is through graphs and charts. A visual display of data makes comparisons easier and promotes a better understanding of trends. They provide data dashboards on various topics as well as access to all the underlying data through our latest data visualization and sharing application, [DataBank](#). Below, you will find list of many available portals:

- [Atlas of Social Protection - Indicators of Resilience and Equity \(ASPIRE\)](#)
- [Climate Change Knowledge Portal \(CCKP\)](#)
- [Country Policy and Institutional Assessment \(CPIA\) for Sub-Saharan Africa](#)
- [Debt Portal](#)
- [EdStats](#)
- [Financial Inclusion Data](#)
- [G20 Basic Set of Financial Inclusion Indicators](#)
- [Gender Data Portal](#)

Products

World Development Indicators

International Debt Statistics

Other Books & Reports

Country & Lending Groups

Data Portals & Tools

Mobile Apps

Third Party Apps

[Help / Feedback](#)

Enlaces de interés

- Datamart Minería: http://intranet2.minem.gob.pe/datamart/dac_publico/
- Ciencia: <http://alicia.concytec.gob.pe/vufind/>
- Estadísticas económicas:
<https://estadisticas.bcrp.gob.pe/estadisticas/series/index>
- Censos: <http://censos.inei.gob.pe/Cenagro/redatam/>
- Mercado de valores:
http://www.smv.gob.pe/Frm_LisConstitucion.aspx?data=5957F36437C9754AFE9203064F36D817B0E06A47DA