

How to open data?

Camila Salazar
@EscuelaDeDatos
@milamila07

Agenda

1. Data architecture basics
2. Getting, creating or extracting data
3. Data cleaning and formatting
4. Data interoperability

Data basics

Definition

Open Data is information which can be freely **used, reused, and redistributed**, by **anyone, anywhere, for any purpose**

See: opendatahandbook.org

Definition

Legally open: available under an open (data) license that permits anyone freely to access, reuse and redistribute

Technically open: available in machine readable format and bulk form.

See: opendatahandbook.org

Open data principles

Completeness: Datasets released by the government should be as complete as possible, reflecting the entirety of what is recorded about a particular subject.

Primacy: be primary source data.

Easy access: Datasets should be as accessible as possible, without barriers.

See: [Sunlight Foundation](#)

Open data principles

Machine readable: information should be stored in widely-used file formats that easily lend themselves to machine processing

Timeliness: Open data is only valuable if it's still relevant.

Permanence: It should be available online in archives in perpetuity

Open data formats

CSV

School name,School type,Address3,Postcode

Alfriston School,"Infant And Junior School, 5-11",Polegate,BN26 5XB

Bexhill High School,Academies,Bexhill-on-sea,TN39 4BY

Bourne Primary School,"Infant And Junior School, 5-11",Eastbourne,BN22 8BE

Brede Primary School,"Infant And Junior School, 5-11",Rye,TN31 6DG

Castledown Primary School,"Infant And Junior School, 5-11",Hastings,TN34 3QT

Chailey School,"Comprehensive All-Through, 11-16",Lewes,BN8 4PU

Chyngton School,"Infant And Junior School, 5-11",Seaford,BN25 3ST

Claverham Community College,"Comprehensive All-Through, 11-16",Battle,TN33 0HT

College Central,Pupil Referral Unit,Eastbourne,BN22 9RB

Oakwood Academy,Academies,Eastbourne,BN22 0SS

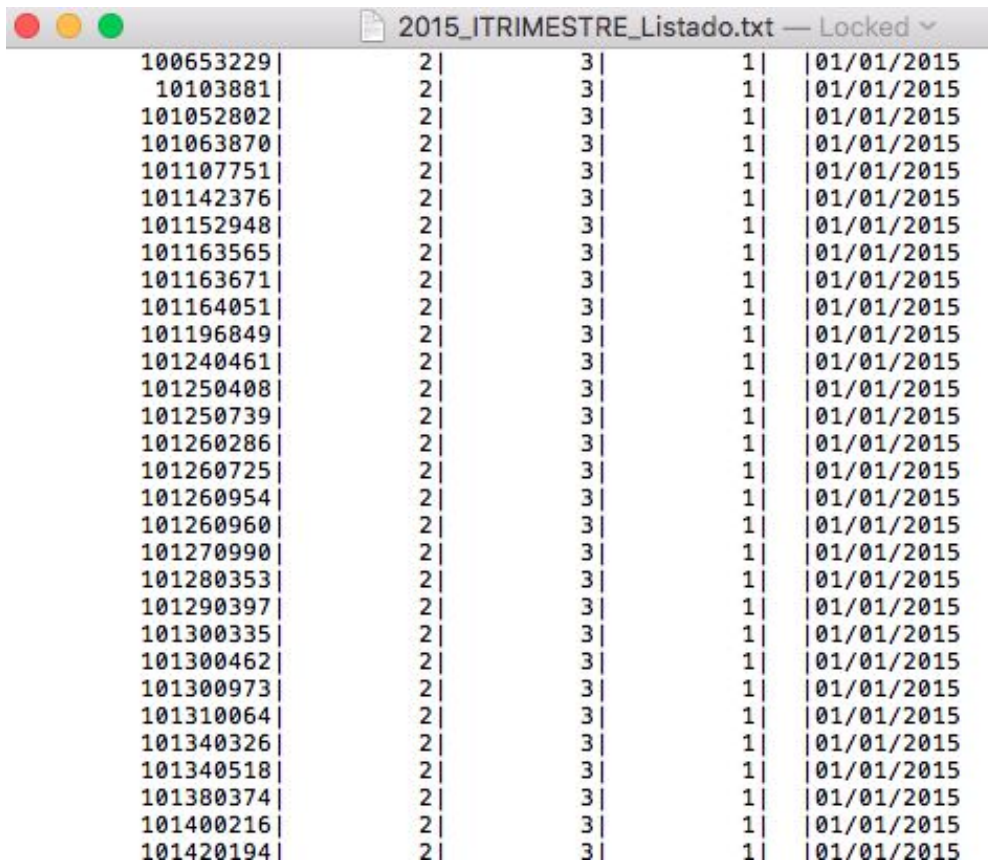
Ocklynge Junior School,"Junior School, 7/8-11",Eastbourne,BN20 8XN

Spreadsheets



Some files might be proprietary

Text files (txt)



A screenshot of a text editor window titled "2015_ITRIMESTRE_Listado.txt" with a "Locked" status. The window displays a list of 30 rows of data, each containing a 10-digit ID, three single-digit numbers, and a date in YYYY-MM-DD format. The IDs range from 100653229 to 101420194. The three single-digit numbers are consistently 2, 3, and 1. The date for all entries is 01/01/2015.

100653229	2	3	1	01/01/2015
10103881	2	3	1	01/01/2015
101052802	2	3	1	01/01/2015
101063870	2	3	1	01/01/2015
101107751	2	3	1	01/01/2015
101142376	2	3	1	01/01/2015
101152948	2	3	1	01/01/2015
101163565	2	3	1	01/01/2015
101163671	2	3	1	01/01/2015
101164051	2	3	1	01/01/2015
101196849	2	3	1	01/01/2015
101240461	2	3	1	01/01/2015
101250408	2	3	1	01/01/2015
101250739	2	3	1	01/01/2015
101260286	2	3	1	01/01/2015
101260725	2	3	1	01/01/2015
101260954	2	3	1	01/01/2015
101260960	2	3	1	01/01/2015
101270990	2	3	1	01/01/2015
101280353	2	3	1	01/01/2015
101290397	2	3	1	01/01/2015
101300335	2	3	1	01/01/2015
101300462	2	3	1	01/01/2015
101300973	2	3	1	01/01/2015
101310064	2	3	1	01/01/2015
101340326	2	3	1	01/01/2015
101340518	2	3	1	01/01/2015
101380374	2	3	1	01/01/2015
101400216	2	3	1	01/01/2015
101420194	2	3	1	01/01/2015

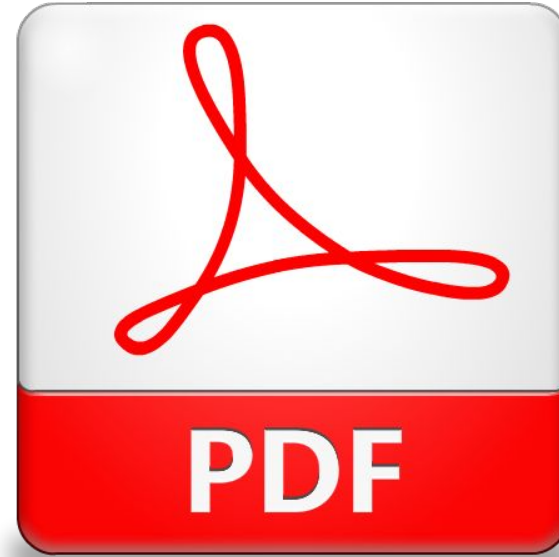
Json

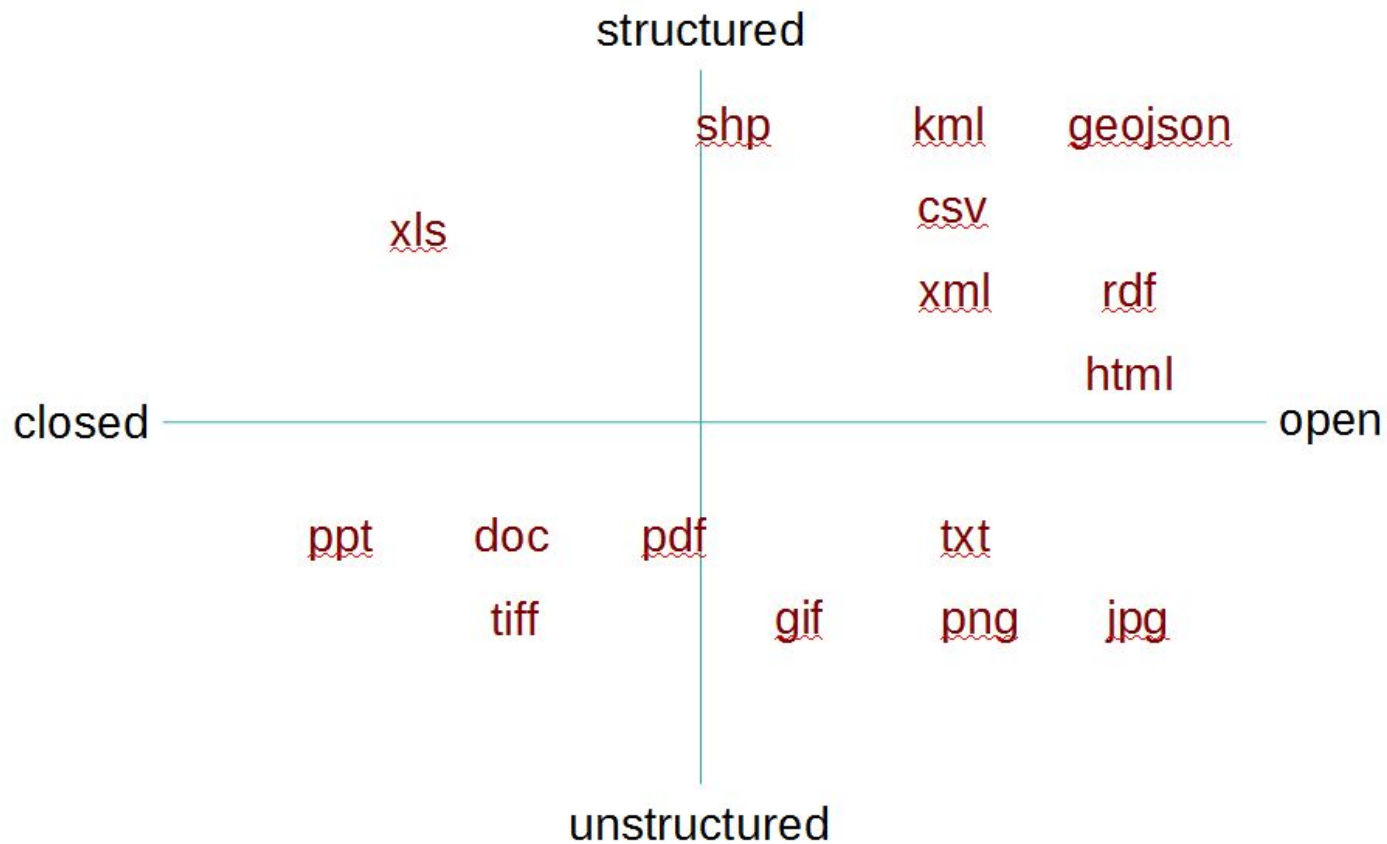
Useful format for programmers

{ JSON }
JavaScript Object Notation

```
[
  {
    "School name": "Chyngton School",
    "School type": "Infant And Junior School, 5-11",
    "Facilities": [
      {
        "name": "First grade classroom"
        "capacity": 30
      },
      {
        "name": "Auditorium"
        "capacity": 100
      },
      {
        "name": "Science Lab"
        "capacity": 40
      }
    ]
    "Address3": "Seaford",
    "Postcode": "BN25 3ST"
  },
  {
    "School name": "Claverham Community College",
    "School type": "Comprehensive All-Through, 11-16",
    "Address3": "Battle",
    "Postcode": "TN33 0HT"
  },
],
```

Avoid PDFs





How does it look like?

DATOS ABIERTOS DEL PODER JUDICIAL DE COSTA RICA

Distribución en CSV

Título Distribución CSV de las Estadísticas policiales de Costa Rica

Descripción Distribución CSV de las Estadísticas policiales de Costa Rica

Tipo de medio text/csv

Licencia [CC BY 4.0](#)

Fecha de publicación 2017-03-04

Última modificación 2018-06-01

Direcciones de descarga

- 2015 https://pjcrdatosabiertos.blob.core.windows.net/datosabiertos/PJCROD_POLICIALES_V1/PJCROD_POLICIALES_V1-2015.csv
- 2016 https://pjcrdatosabiertos.blob.core.windows.net/datosabiertos/PJCROD_POLICIALES_V1/PJCROD_POLICIALES_V1-2016.csv
- 2017 https://pjcrdatosabiertos.blob.core.windows.net/datosabiertos/PJCROD_POLICIALES_V1/PJCROD_POLICIALES_V1-2017.csv
- 2018 https://pjcrdatosabiertos.blob.core.windows.net/datosabiertos/PJCROD_POLICIALES_V1/PJCROD_POLICIALES_V1-2018.csv

	A	B	C	D	E	F	G	H	I	J	K	L
1	Delito	SubDelito	Fecha	Victima	SubVictim	Edad	Genero	Nacionali	Provincia	Canton	Distrito	
2	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	MOTOCIC	Mayor de	HOMBRE	COSTA RIC	ALAJUELA	ALAJUELA	SAN ISIDRO	
3	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	AUTOMO	Mayor de	HOMBRE	COSTA RIC	HEREDIA	SAN RAFA	SAN JOSECITO	
4	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	AUTOMO	Desconoc	HOMBRE	NICARAGU	HEREDIA	SAN RAFA	SAN JOSECITO	
5	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	MOTOCIC	Mayor de	HOMBRE	COSTA RIC	LIMON	LIMON	LIMON	
6	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	MOTOCIC	Mayor de	HOMBRE	NICARAGU	LIMON	GUACIMC	POCORA	
7	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	MOTOCIC	Mayor de	HOMBRE	COSTA RIC	LIMON	GUACIMC	RIO JIMENEZ	
8	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	MOTOCIC	Mayor de	HOMBRE	COSTA RIC	LIMON	LIMON	LIMON	
9	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	MOTOCIC	Mayor de	HOMBRE	COSTA RIC	PUNTARE	PARRITA	PARRITA	
10	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	AUTOMO	Mayor de	HOMBRE	COSTA RIC	SAN JOSE	VASQUEZ	SAN ISIDRO	
11	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	MOTOCIC	Mayor de	HOMBRE	COSTA RIC	GUANACA	NICOYA	NICOYA	
12	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	MOTOCIC	Mayor de	HOMBRE	COSTA RIC	ALAJUELA	UPALA	UPALA	
13	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	AUTOMO	Mayor de	HOMBRE	COSTA RIC	CARTAGO	PARAISO	PARAISO	
14	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	MOTOCIC	Menor de	HOMBRE	COSTA RIC	SAN JOSE	MORAVIA	SAN VICENTE	
15	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	CARGA LI	Mayor de	HOMBRE	COSTA RIC	SAN JOSE	TIBAS	SAN JUAN	
16	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	RURAL 4X	Mayor de	MUJER	COSTA RIC	CARTAGO	TURRIALB	TURRIALBA	
17	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	AUTOMO	Mayor de	MUJER	COSTA RIC	ALAJUELA	ALAJUELA	ALAJUELA	
18	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	CARGA LI	Mayor de	HOMBRE	COSTA RIC	ALAJUELA	ALAJUELA	SAN JOSE	
19	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	AUTOMO	Mayor de	MUJER	COSTA RIC	CARTAGO	LA UNION	SAN RAFAEL	
20	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	MOTOCIC	Mayor de	HOMBRE	COSTA RIC	ALAJUELA	ALAJUELA	GUACIMA	
21	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	MOTOCIC	Mayor de	HOMBRE	COSTA RIC	SAN JOSE	DESAMPA	GRAVILIAS	
22	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	MICROBU	Mayor de	HOMBRE	COSTA RIC	SAN JOSE	SAN JOSE	HATILLO	
23	ROBO DE	DESCUIDO	2018-01-0	VEHICULO	CARGA LI	Mayor de	HOMBRE	COSTA RIC	SAN JOSE	SAN JOSE	PAVAS	

Variables

Observations

Closed data (PDF)

SAN JOSE SAN JOSE

UNIDAD SOCIAL CRISTIANA
ALCALDE

105870719 JORGE EDUARDO SANCHEZ SIBAJA
VICEALCALDESA PRIMERA
105610212 SUSY BIBIANA MORENO AMADOR
VICEALCALDE SEGUNDO
105870165 EDGAR ANTONIO GARITA SANCHO

SOLIDARIDAD
ALCALDE

301590969 ROBERTO ARTURO BARRANTES GUERRERO
VICEALCALDESA PRIMERA
111550236 DYANA IBARRA JIMENEZ
VICEALCALDE SEGUNDO
301740491 GUILLERMO SANABRIA RAMIREZ

LIBERACION NACIONAL
ALCALDE

900130151 GUIDO ORLANDO GRANADOS RAMIREZ
VICEALCALDESA PRIMERA
105240209 MARCIA VALLADARES BERMUDEZ
c.c. MARZHIA VALLADARES BERMUDEZ
VICEALCALDE SEGUNDO
900450133 GUILLERMO ARAYA RODRIGUEZ
c.c. WILLIAM ARAYA RODRIGUEZ

Open structured data (csv)

	A	B	C	D	E
1	Provincia	Cantn	id ca	Ndecdu	Nombredelcandidato
2	San José	San José	101	112680372	Morua Calvo Laura Melissa
3	San José	San José	101	601270337	Castillo Villanueva Fernando Alexis
4	San José	San José	101	900130151	Guido Granados Ramirez
5	San José	San José	101	105250577	Ortiz Barrionuevo Gustavo Adolfo
6	San José	San José	101	104760724	Araya Monge Johnny Francisco
7	San José	San José	101	105870719	Jorge Eduardo Sánchez Sibaja
8	San José	San José	101	104200371	Israel Guillén González
9	San José	San José	101	301590969	Roberto Barrantes Guerrero
10	San José	San José	101	112240208	Daguer Hernández Vazquez

DECIDA POR SU CANTÓN

Este 7 de febrero usted elegirá al alcalde de su cantón, quien debe velar por las obras y servicios como el de recolección de basura. Se postulan 605 candidatos. La Nación examinó su historia judicial y sanciones administrativas. Su situación como patronos con el Seguro Social. También se analizó su experiencia política y comunal. En esta aplicación además se enterará del desempeño de su cantón en las 10 áreas que más impactan su vida. Esperamos que esta información acompañe su decisión de voto.

[Ver metodología >](#)

VEA LOS ALCALDES ELECTOS

SELECCIONE SU PROVINCIA

Information
Knowledge

Get the data

Where's your data?

Make an inventory:

- What information do you have
- Where
- In which formats

Indicadores generales de la condición de actividad
II TRIMESTRE 2013 – II TRIMESTRE 2015

Indicadores generales	Trimestre								
	II 2013	III 2013	IV 2013	I 2014	II 2014	III 2014	IV 2014	I 2015	II 2015
	Abr - Jun	Jul - Set	Oct - Dic	Ene - Mar	Abr - Jun	Jul - Set	Oct - Dic	Ene - Mar	Abr - Jun
Población total	4 703 581	4 717 478	4 733 053	4 747 489	4 763 662	4 778 864	4 792 617	4 807 263	4 823 600
Población de 15 años y más	3 582 545	3 597 351	3 613 321	3 628 363	3 645 426	3 661 332	3 676 598	3 692 461	3 709 802
Fuerza de trabajo	2 245 133	2 226 050	2 277 577	2 310 113	2 252 474	2 294 204	2 279 775	2 282 882	2 305 296
Ocupado	2 008 405	2 026 738	2 088 282	2 084 210	2 048 011	2 065 801	2 059 600	2 051 208	2 087 363
Desempleada	236 728	199 312	189 295	225 903	204 463	228 403	220 175	231 674	217 933
Fuera de la fuerza de trabajo	1 337 412	1 371 301	1 335 744	1 318 250	1 392 952	1 367 128	1 396 823	1 409 579	1 404 506
Porcentajes									
Tasas de la población de 15 años o más									
Tasa neta de participación	62,7	61,9	63,0	63,7	61,8	62,7	62,0	61,8	62,1
Tasa de ocupación	56,1	56,3	57,8	57,4	56,2	56,4	56,0	55,6	56,3
Tasa de no participación	37,3	38,1	37,0	36,3	38,2	37,3	38,0	38,2	37,9
Tasas de la fuerza de trabajo									
Tasa de desempleo	10,5	9,0	8,3	9,8	9,1	10,0	9,7	10,1	9,5
Tasa de presión general	19,6	17,5	16,7	18,8	18,6	19,7	19,3	19,1	18,4
Porcentaje de la población ocupada									
Porcentaje de ocupados con seguro por trabajo	68,4	66,8	66,7	67,2	68,7	68,6	67,3	67,7	68,3
Porcentaje de ocupados con subempleo	12,5	13,9	12,2	11,9	11,8	13,4	14,2	12,9	13,5

Fuente: INEC, Encuesta Continua de Empleo (ECE), 2015.

CARGO	PERIODO DE PAGO												AGUINALDO *	SALARIO BRUTO PROMEDIO MENSUAL y AGUINALDO	
	2013-12	2014-1	2014-2	2014-3	2014-4	2014-5	2014-6	2014-7	2014-8	2014-9	2014-10	2014-11			2014-12
DECANO(A)	4.391.670,1	4.522.520,1	4.522.520,1	4.522.520,1	4.522.520,1	4.547.823,1	4.596.167,9	4.684.831,5	4.684.831,5	4.684.831,5	4.729.163,2	4.730.864,3	4.729.163,2	4.595.021,9	4.620.982,9
DECANO(A)	2.339.011,9	2.408.582,2	2.408.582,2	3.488.845,4	2.766.669,9	2.776.671,9	2.418.583,3	3.623.195,4	3.108.568,4	3.108.568,4	3.137.983,4	3.137.983,4	3.137.983,4	2.893.770,5	2.995.229,8
DECANO(A)	2.810.000,5	2.893.700,5	2.893.700,5	2.893.700,5	2.893.700,5	2.893.700,5	2.552.119,5	2.283.461,9	2.280.259,6	2.314.248,5	2.336.098,6	2.336.098,6	2.336.098,6	2.615.065,8	2.578.611,8
DECANO(A)	2.376.561,4	2.447.258,3	2.447.258,3	2.447.258,3	2.447.258,3	2.447.258,3	2.447.258,3	2.494.389,6	2.494.389,6	2.494.389,6	2.553.640,1	2.562.561,4	2.562.561,4	2.471.623,4	2.485.931,1
DECANO(A)	2.595.385,0	2.672.646,6	2.672.646,6	2.672.646,6	2.721.191,8	2.724.659,3	2.724.659,3	2.777.176,9	2.777.176,9	2.777.176,9	2.803.435,9	2.803.435,9	2.803.435,9	2.726.853,1	2.742.857,0
DECANO(A)	1.751.514,0	1.803.159,5	1.821.547,8	1.842.563,1	1.842.563,1	3.240.311,9	2.552.614,9	2.601.597,7	2.601.597,7	2.601.597,7	2.624.826,7	2.708.610,9	2.692.998,7	2.332.708,7	2.405.130,6
DECANO(A)	2.083.032,0	2.145.523,0	2.145.523,0	2.145.523,0	2.145.523,0	2.145.523,0	2.145.523,0	2.187.183,6	2.229.896,3	5.035.780,0	2.836.963,5	2.836.963,5	2.836.963,5	2.506.913,0	2.564.907,8
DECANO(A)	4.053.180,7	4.173.876,1	4.173.876,1	4.178.415,1	4.243.521,1	4.243.521,1	4.243.521,1	4.325.337,0	4.325.337,0	4.325.337,0	4.366.244,9	4.366.244,9	4.366.244,9	4.251.534,3	4.275.616,2
DECANO(A)	2.411.288,7	2.486.011,9	2.486.011,9	2.486.011,9	2.486.011,9	2.486.011,9	2.486.011,9	2.532.648,0	2.532.648,0	2.532.648,0	2.555.966,0	2.598.068,5	2.599.520,3	2.506.611,5	2.521.090,9
DIR ADMINISTRATIVO EN SERV SECRET	1.953.595,0	2.012.441,3	2.012.441,3	2.012.441,3	2.012.441,3	2.012.441,3	2.044.829,5	2.086.729,7	2.086.729,7	2.086.729,7	2.106.522,9	2.106.522,9	2.106.522,9	2.044.488,8	2.056.252,5
DIR ASESOR EN DES HUMANO	2.826.283,1	2.837.739,9	2.837.740,0	2.837.740,0	2.837.740,0	2.837.740,0	2.837.740,0	2.890.878,5	2.890.878,5	2.890.878,5	2.917.447,6	2.917.447,6	2.929.330,9	2.863.354,4	2.871.281,2
DIR ASESOR EN DES TECNOLOGICO	1.755.174,1	1.810.373,1	1.810.373,1	1.810.373,1	1.810.373,1	1.810.373,1	1.810.373,1	1.843.780,8	1.843.780,8	1.843.780,8	1.871.575,7	1.908.017,8	1.908.017,8	1.827.362,3	1.839.119,5
DIR ASESOR EN DES TECNOLOGICO	2.450.435,0	2.437.967,9	2.884.721,6	2.764.860,9	2.776.095,3	2.767.669,5	2.767.669,5	2.560.242,9	2.534.273,3	2.534.273,3	2.557.927,7	2.557.927,7	2.557.927,7	2.632.838,7	2.641.107,2
DIR ASESOR EN GEST JURIDICA	2.401.115,7	2.473.546,5	2.473.546,5	2.473.546,5	2.473.546,5	2.473.546,5	2.484.323,7	2.568.388,7	2.568.388,7	2.568.388,7	2.592.715,7	2.592.715,7	2.592.715,7	2.612.980,8	2.526.719,4
DIR ASESOR EN PLAN ESP	2.183.081,8	1.549.820,5	3.315.088,4	2.532.336,0	2.532.336,0	2.532.336,0	2.532.336,0	2.331.062,4	2.268.492,0	2.268.492,0	3.481.833,8	2.605.848,0	2.847.425,9	2.511.088,6	2.562.191,9
DIR ASESOR EN PLANEAMIENTO	2.217.460,3	2.286.607,3	2.286.607,3	2.286.607,3	2.286.607,3	2.286.607,3	2.286.607,3	2.376.292,4	2.376.292,4	2.376.292,4	2.398.041,0	2.398.041,0	2.398.041,0	2.321.838,6	2.335.729,4
DIR ASESOR EN PROVEEDURIA	2.493.240,2	2.571.534,8	2.571.534,8	2.571.534,8	2.571.534,8	2.571.534,8	2.571.534,8	2.619.068,0	2.637.901,9	2.666.152,7	2.690.367,6	2.690.367,6	2.690.367,6	2.602.192,2	2.617.355,9
DIR ASESOR EN SERV ADMIS	1.751.510,4	1.804.135,2	1.804.135,2	1.804.135,2	1.804.135,2	1.836.466,7	1.850.323,1	1.886.197,2	1.886.197,2	1.886.197,2	1.904.134,2	1.904.134,2	1.904.134,2	1.843.475,1	1.855.215,3
DIR ASESOR EN SERV ADMIS	2.501.189,1	2.579.960,9	2.579.960,9	2.612.292,4	2.626.148,7	2.626.148,7	2.626.148,7	2.674.578,8	2.674.578,8	2.674.578,8	2.710.028,1	2.701.602,3	2.701.602,3	2.632.268,0	2.647.684,4
DIR ASESOR EN SIST DE INF DOC	2.661.814,9	2.774.080,3	2.774.080,3	2.774.080,3	2.774.080,3	2.774.080,3	2.774.080,3	2.827.891,4	2.827.891,4	2.827.891,4	2.854.796,9	2.854.796,9	2.869.056,8	2.791.630,4	2.807.572,0
DIR ASESOR FINANCIERO CONTABLE	2.233.970,1	2.303.771,4	2.320.707,0	2.349.959,3	2.349.959,3	2.364.002,3	2.352.767,9	2.396.489,4	2.396.489,4	2.396.489,4	2.418.350,1	2.418.350,1	2.452.053,3	2.358.442,1	2.375.217,8
DIR ASESOR GENERAL EN DES TECNOLOGICO	2.749.084,4	2.739.154,8	2.744.772,0	2.741.963,4	2.741.963,4	2.741.963,4	2.741.963,4	2.906.840,4	2.906.840,4	2.906.840,4	2.933.106,9	2.981.907,6	2.985.393,3	2.819.700,0	2.837.877,6
DIR EJECUTIVO EN SERV ADMIS	2.150.917,5	2.219.101,5	2.224.718,8	2.221.910,1	2.221.910,1	2.221.910,1	2.221.910,1	2.262.490,8	2.262.490,8	2.263.787,9	2.322.064,7	2.322.064,7	2.338.916,3	2.242.939,7	2.257.401,2
DIR EJECUTIVO EN SERV ADMIS	1.695.947,9	1.749.369,9	1.749.369,9	1.749.369,9	1.749.369,9	1.772.607,9	1.812.492,0	2.279.500,9	2.279.500,9	2.259.275,5	2.279.936,4	2.279.936,4	2.009.313,1	1.927.212,2	1.951.317,2
DIR EJECUTIVO EN SERV ADMIS	1.863.615,7	1.920.398,5	1.920.398,5	1.926.760,5	1.958.570,3	1.958.570,3	1.958.570,3	1.996.000,9	1.996.000,9	1.996.000,9	2.014.716,2	2.014.716,2	2.014.716,2	1.960.359,9	1.971.983,0
DIR EJECUTIVO EN SERV ADMIS	1.654.555,9	1.707.133,7	1.707.133,7	1.707.133,7	1.707.133,7	1.738.264,1	1.738.264,1	1.778.708,8	1.778.708,8	1.778.708,8	1.793.112,8	1.793.112,8	1.793.112,8	1.723.664,3	1.734.322,6
DIR EJECUTIVO EN SERV ADMIS	1.506.229,2	1.551.575,0	1.551.575,0	1.551.575,0	1.551.575,0	1.551.575,0	1.551.575,0	1.594.564,6	1.620.506,6	1.620.506,6	1.635.886,4	1.635.886,4	1.635.886,4	1.576.919,1	1.586.892,7
DIR EJECUTIVO EN SERV ADMIS	1.851.770,5	1.910.344,2	1.910.344,2	1.944.698,8	1.970.984,8	1.954.133,2	1.954.133,2	1.989.896,1	1.989.896,1	1.989.896,1	2.007.777,5	2.007.777,5	2.007.777,5	1.956.804,3	1.968.804,8
DIR EJECUTIVO EN SERV ADMIS	1.556.520,0	1.603.215,6	1.603.215,6	1.603.215,6	1.603.215,6	1.603.215,6	1.671.961,9	1.673.259,0	1.673.259,0	1.673.259,0	1.689.194,7	1.689.194,7	1.689.194,7	1.631.056,9	1.641.262,6
DIR EJECUTIVO EN SERV ADMIS	1.819.923,0	1.874.918,0	1.886.152,6	1.880.535,3	1.880.535,3	1.880.535,3	1.916.668,8	1.916.668,8	1.916.668,8	1.916.668,8	1.934.735,5	1.956.996,2	1.974.019,1	1.895.406,1	1.907.259,6
DIR EJECUTIVO EN SERV PUB E IMP	1.278.570,0	1.352.554,1	1.355.098,9	1.355.098,9	1.355.098,9	1.355.098,9	1.355.098,9	1.381.411,5	1.381.411,5	1.381.411,5	1.394.567,7	1.394.567,7	1.394.567,7	1.361.665,7	1.370.588,6
DIR ESPECIALISTA CONT INT Y FISC	2.495.488,1	2.606.217,9	2.624.450,5	2.618.833,3	2.618.833,3	2.618.833,3	2.618.833,3	2.668.975,6	2.668.975,6	2.668.975,6	2.694.046,5	2.694.046,5	2.710.898,1	2.633.042,4	2.649.612,4
DIR ESPECIALISTA CONT INT Y FISC	2.104.923,6	2.196.502,6	2.200.827,7	2.200.827,7	2.354.100,8	2.660.647,0	2.660.647,0	2.712.092,0	2.712.092,0	2.712.092,0	2.737.814,3	2.532.755,7	2.264.602,2	2.482.110,2	2.494.393,1

PDF

- Transform it into open formats (Tabula)



- Search for the primary source (it may be in a machine readable format)

Houston Museum of Natural History

1935	Accession Catalogue		
Oct 29 L. S. Turley Houston	\$ 2574		
Small paper impressions on diatomite. Muddoon, Tex.	\$ 12817		
Oct 29 Herbert Lodwin 1122 Holman	\$ 2575		
Three floats for fishing nets, Japan	Q 12818		
Newspaper of Aug. 17, 1925, printed by mimeograph at Shanghai, China. The paper carries story of Rogers' Post death of Aug. 15.	Q 12817		
Nov 3 Mrs. N. Currier Fairbanks, Tex.	\$ 2576		
Small lot of sandy fossils from Grayson Co. Kentucky	\$ 12820	RECORDED 20-14-35	
Nov 15 City of Houston	\$ 2577		
Melanterite (Cinnabar) (Coppens)	M 12821		
22, 50 & 7420. 3701 (2.101)			
Nov 20 Mrs. E. O. Raller Beaumont, Texas	\$ 2578		
Whip made from horse hair, Tex.	Q 12822		
Leather cushion cover, Kentucky	Q 12823		
Red cover from Monkey Cabin, Tex.	Q 12824		
Nov 21 Mrs. M. J. Bennett 2142 Bolivar Rd.	\$ 2579		
Surgicome forceps for blood letting	Q 12825		
Cup for taking blood	Q 12826		
2 - Drinking glasses (moulded)	Q 12827		
with Pinkney Buck emblem	Q 12827		

Houston Museum of Natural History

1935	Accession Catalogue		
Nov. 21 Mrs. C. J. Hutchinson, Houston	\$ 2600		
1 shot bow to Seal claim	M 12828		
Dec 3 Nagel, Elber. Insufed Cigarettes to Casanova Bay on 1927.	\$ 2601		
Box of lawn fish.	\$ 12829		
(No record found at this time of this item being previously recorded. (Hamm))			
Dec 3 Mrs. Home Nagel Houston	\$ 2602		
11 pcs of Coral representation of type found near the Texas- Gulf shore.	\$ 12830		
Dec 16 Julia A. Tamm Houston	\$ 2603		
Repetori silk used by Prof. Roman (teacher of Mr. Tamm) on his work in France L. 17 years count.	Q 12831		
Field and laboratory notes on paleontology by Prof. Roman	E 12832		
Dec 19 Harry Frank, Bellville, Texas	\$ 2604		
Reeler horn for pistols, W.D.A.	Q 12833		
Quartz tile	Q 12834		
Block with handle from cane sword (No history known of the above.)	Q 12835		
Dec 27 Mrs. Thomas Ascardi 541 Hamersburg St.	\$ 2605		
1 Sandalwood Toyon Japan	Q 12836		

Contrato Numero No. MH-2013-000040-00 Número de contrato

Contratación Directa N° 2013CD-000429-13802 Procedimiento de contratación

"Arrendamiento de los pisos 7 y 8 del Condominio Vertical Comercial CCI Torre C"
Entre el Ministerio de Hacienda
y la empresa CCI Torre C, S.A.

Entre nosotros, el Ministerio de Hacienda, representado en este acto por el señor Edgar Ayales Esna, mayor, casado una vez, economista, portador de la cédula de identidad número uno – trescientos ochenta y tres – trescientos diecisiete, vecino de San José, en mi condición de Ministro de Hacienda, según Acuerdo de Nombramiento Número 553-P del veinticinco de abril del año dos mil doce, publicado en el Alcance Digital No. 61 de La Gaceta No. 89 del 9 de mayo de 2012, con facultades suficientes para representar al Ministerio de Hacienda cédula jurídica número dos–cien–cero cuatro dos cero cero cinco, en adelante "El Ministerio", y la empresa CCI Torre C, S.A., con cédula jurídica número tres – uno cero uno – cinco tres nueve uno nueve cero, inscrita en el Registro Público, Sección Mercantil, bajo la cédula jurídica número tres – ciento uno – quinientos treinta y nueve mil ciento noventa, representada en este acto por la señora Diana Escobar Pardo, cédula de identidad número ocho – cero setena y cuatro – trescientos sesenta y cinco, mayor de edad, divorciada en segundas nupcias, ejecutiva de negocios, vecina de Escazú, en calidad de Apoderada Generalísima sin límite de suma, de dicha sociedad de conformidad con la inscripción número quinientos cincuenta y uno – dos mil trece, del veintitrés de agosto

Empresa
Cédula
jurídica

conven
049-20
regirá ;

El edificio está ubicado en un bien inmueble inscrito en el Registro Nacional de la Propiedad, bajo el Sistema de Folio Real matrícula número 1-610088-000, localizado en la provincia de San José (01), Cantón San José (01), Distrito Hospital (03); situado en el costado noreste de la intersección de la avenida 8 con la calle 28, en Barrio Don Bosco.

Ubicación



Institucion contratante	Tipo de procedimier	Unidad de compra	No de procedimiento	Objeto contrac	Adjudicado en cc	Cedula adju	Adjudicatario
MINISTERIO DE HACIENDA	Contratacion Directa	GESTION DE INGRESOS INTERNI	2013CD-000086-13402	ARRENDAMIENTO	957,871,232.00	3101024626	INMOBILIARIA MIRA SOI
MINISTERIO DE HACIENDA	Contratacion Directa	GESTION DE INGRESOS INTERNI	2013CD-000677-13402	ARRENDAMIENTO	1,275,000.00	3101168676	PROYECTO CINCO SOCIE
MINISTERIO DE HACIENDA	Contratacion Directa	ACTIVIDADES CENTRALES	2013CD-000715-13200	ARRENDAMIENTO	64,738,560.00	3101008940	COMPANIA LA LLACUNA
MINISTERIO DE HACIENDA	Contratacion Directa	REGULACION Y REGIS. CONTABI	2013CD-000368-13605	ALQUILER DEL ED	53,435,980.00	3101469973	COPPERMINE INC SOCIE
MINISTERIO DE HACIENDA	Contratacion Directa	GESTION ADUANERA	2013CD-000714-13403	ALQUILER DE APA	2,703,846.42	3101129353	HOTELERA QUINTO CEN'
MINISTERIO DE HACIENDA	Contratacion Directa	CENTRO DE FORMACION E INVE	2013CD-000429-13802	ARRENDAMIENTO	47,601,538.92	3101539190	C C I TORRE C SOCIEDAD
MINISTERIO DE HACIENDA	Contratacion Directa	GESTION DE INGRESOS INTERNI	2012CD-000471-13402	ARRENDAMIENTO	163,484,992.00	106400079	QUIROS JIMENEZ RODRI
MINISTERIO DE HACIENDA	Contratacion Directa	REGULAC.Y COORDIN.DE LA COI	2012CD-000146-13603	“ARREND	9,488,966.74	3101081697	REFRIMPORT SOCIEDAD
MINISTERIO DE HACIENDA	Contratacion Directa	REGULACION Y GESTION DEL EM	2012CD-000548-13606	ARRENDAMIENTO	9,000,000.00	3101021994	INVERSIONES AGRARIAS
MINISTERIO DE HACIENDA	Contratacion Directa	GESTION DE INGRESOS INTERNI	2010CD-000282-13402	ARRENDAMIENTO	50,773,800.00	3101018688	PROCAMO SOCIEDAD AI
MINISTERIO DE HACIENDA	Contratacion Directa	GESTION DE INGRESOS INTERNI	2010CD-003998-13402	ARRENDAMIENTO	46,347,567.78	3101241601	COMERCIAL SOL Y LLUVI
MINISTERIO DE HACIENDA	Contratacion Directa	GESTION DE INGRESOS INTERNI	2010CD-003427-13402	ALQUILER EDIFIC	42,500,000.00	3101292582	PENARROYA SOCIEDAD /
MINISTERIO DE HACIENDA	Contratacion Directa	GESTION ADUANERA	2010CD-004225-13403	ALQUILER EDIFIC	17,822,146.00	3101477544	HELLINIKH COMPANY SC
MINISTERIO DE HACIENDA	Contratacion Directa	GESTION ADUANERA	2010CD-005928-13403	ALQUILER DE EDI	3,756,767.25	3101144610	ADMINISTRADORA FIDU
MINISTERIO DE HACIENDA	Contratacion Directa	CENTRO DE FORMACION E INVE	2010CD-003192-13802	ALQUILER DE CIN	11,497,999.21	3101027962	HOGARES DE COSTA RIC.
MINISTERIO DE HACIENDA	Contratacion Directa	DIREC.Y COORD.DEL PROC.PRES	2010CD-005778-13602	ALQUILER DE EDI	45,179,495.00	3101590656	INVERSIONES INMOBILI/

<http://database.ameliarueda.com/alquileres/>

Information on paper

- Identify the variables and formats
- Digitalize the information
- Ask for help: civil society, journalists

Cons: labor intensive, time consuming, subject to human error.

Information in other formats

- Start collecting the data in open and structured formats

Clean the data

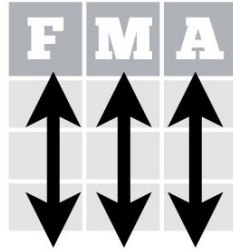
Cleaning data

Processing a dataset to make it easier to consume. This may involve fixing inconsistencies and errors, removing non-machine-readable elements such as formatting, using standard labels for row and column headings, ensuring that numbers, dates, and other quantities are represented appropriately, conversion to a suitable file format, reconciliation of labels with another dataset being used

See: opendatahandbook.org

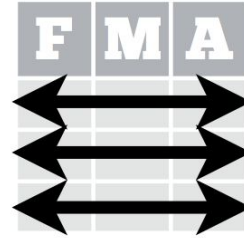
Tidy data

In a tidy data set:



Each **variable** is saved in its own **column**

&



Each **observation** is saved in its own **row**

See: [Tidy data by Hadley Wickham](#)

	A	B	C	D	E	F	G	H	I
1		Landgrabber	e	C t	Hectares	nductive oil, potatoes	ted investment	Status of deal	Summary
2		Al Qudra	E	r e a r	31000			Done	<p>Dhabi in 2005. In February 2008, the company told Le Monde that it had acquired concessions covering 31,000 ha of agricultural lands in Algeria where it intends to produce potatoes, olives and dairy. It also said that it was planning to set up a joint venture with Moroccan investors to produce olive oil on 14,000 ha in Morocco. In September 2009, The National reported that Al Qudra was considering business land in Dakinjetta, Syria. Vietnam: the Angolan government had approved GAMCE's proposed project to construct a rice mill in Langa and establish a 1,500-ha pilot rice farm in</p>
3		neering Co. Ltd	a	r u	1500		US\$77 million	Done	
4		ENI	y	E n e r g y	12000	Oil palm		In process	<p>The project is a joint venture between Sonargol, an Angolan state-owned company responsible for the exploration, production and transportation of hydrocarbons in Angola, and ENI, one of the world's largest oil companies. In December 2011, the two companies signed an agreement for the execution of a pilot project in food and biodiesel production.</p>
5		AfriAagro	al	' J e c o m m	5000	palm	\$30-35 million	Done	<p>company acquired a 50-year lease on 5,000 ha of land near the city of Ambur in Baniu Province, Mozambique. It intends to expand to</p>
6		Eurico Ferreira	al		30000	Sugar cane	\$200 million	Done	<p>© 2012. All rights reserved. This report contains confidential information. Please do not disseminate this report without the prior written consent of the author. For more information, please contact the author at info@schoolofdata.com.</p>

description	description2	location	properties	price_euros	reviews
	Cozy little cabin	Cahuita	Tipo de habitac	30 \$	4.7 $\hat{\mu}$ $\hat{\sigma}$ • 40 evaluaciones
Cuisine Miguel...	OCEANFRONT	Samara	Tipo de habitac	28 \$	4.9 $\hat{\mu}$ $\hat{\sigma}$ • 20 evaluaciones
Mountain cottage	This is truly a ma	Turrialba	Tipo de habitac	35 \$	5.0 $\hat{\mu}$ $\hat{\sigma}$ • 3 evaluaciones
This small unique	little house is in	Monteverde	Tipo de habitac	24 \$	5.0 $\hat{\mu}$ $\hat{\sigma}$ • 16 evaluaciones
We have direct b	Our houses are s	Puerto Viejo de T	Tipo de habitac	60 \$	4.8 $\hat{\mu}$ $\hat{\sigma}$ • 58 evaluaciones
	The 3500 square	Puntarenas	Tipo de habitac	153 \$	4.8 $\hat{\mu}$ $\hat{\sigma}$ • 34 evaluaciones
This beautiful ho	This house is a 2	Nosara	Tipo de habitac	144 \$	4.8 $\hat{\mu}$ $\hat{\sigma}$ • 31 evaluaciones
Beautifully desig	Casa Troya refle	Puerto Jimenez	Tipo de habitac	87 \$	5.0 $\hat{\mu}$ $\hat{\sigma}$ • 14 evaluaciones
Colores de Pacif	The location is g	Playa Flamingo	Tipo de habitac	125 \$	5.0 $\hat{\mu}$ $\hat{\sigma}$ • 22 evaluaciones
One bed in a dor	Surf with local le	Jaco	Tipo de habitac	13 \$	Habitacion compartida $\hat{\mu}$ $\hat{\sigma}$
King size bed be	King size bed be	tamarindo	Tipo de habitac	93 \$	4.9 $\hat{\mu}$ $\hat{\sigma}$ • 46 evaluaciones
Very lightful, spa	Is very centric!!!	San Jose	Tipo de habitac	13 \$	1 evaluacion
Hostal del Mar e	Hostal del Mar e	Golfito	Tipo de habitac	35 \$	4.3 $\hat{\mu}$ $\hat{\sigma}$ • 4 evaluaciones
	Beautiful new ho	Puerto Viejo	Tipo de habitac	89 \$	4.7 $\hat{\mu}$ $\hat{\sigma}$ • 54 evaluaciones
Stunning NEW lu	Playa Penca is o	Potrero	Tipo de habitac	80 \$	4.3 $\hat{\mu}$ $\hat{\sigma}$ • 4 evaluaciones
Ubicado en el m	Ademas de disfr	Manuel Antonio	Tipo de habitac	25 \$	4.6 $\hat{\mu}$ $\hat{\sigma}$ • 62 evaluaciones

	A	B
1	Columna1	
2	SAN JOSE SAN JOSE	
3	UNIDAD SOCIAL CRISTIANA	
4	ALCALDE	
5	105870719 JORGE EDUARDO SANCHEZ SIBAJA	
6	VICEALCALDESA PRIMERA	
7	105610212 SUSY BIBIANA MORENO AMADOR	
8	VICEALCALDE SEGUNDO	
9	105870165 EDGAR ANTONIO GARITA SANCHO	
10	SOLIDARIDAD	
11	ALCALDE	
12	301590969 ROBERTO ARTURO BARRANTES GUERRERO	
13	VICEALCALDESA PRIMERA	
14	111550236 DYANA IBARRA JIMENEZ	
15	VICEALCALDE SEGUNDO	
16	301740491 GUILLERMO SANABRIA RAMIREZ	
17	LIBERACION NACIONAL	
18	ALCALDE	
19	900130151 GUIDO ORLANDO GRANADOS RAMIREZ	
20	VICEALCALDESA PRIMERA	
21	105240209 MARCIA VALLADARES BERMUDEZ	
22	VICEALCALDE SEGUNDO	
23	900450133 GUILLERMO ARAYA RODRIGUEZ	
24	FRENTE AMPLIO	

Clean data

- Variables in the correct format
- Tidy data
- No merged cells
- Single value per cell
- No strange characters
- Without errors

Common errors

Wrong format: text stored as number, wrong dates

09/12/2014

12/09/2014

Each column should have only ONE data type

Multiple values

	A	B	C	D	E
1	PROGRAMADO N° VOOS		ATRASADOS (%)	ATRASADOS N°	CANCELADOS (%)
2	TOTAIS	1534	44 (2.9 %)	3 (0.2 %)	85 (5.5 %)
3	Altamira-PA	7	0 (0 %)	0 (0 %)	0 (0 %)
4	Aracaju-SE	11	2 (18.2 %)	0 (0 %)	1 (9.1 %)
5	Belém-PA	39	0 (0 %)	0 (0 %)	0 (0 %)
6	Belo Horizonte -	8	0 (0 %)	0 (0 %)	0 (0 %)
7	Belo Horizonte -	104	5 (4.8 %)	0 (0 %)	10 (9.6 %)
8	Boa Vista-RR	4	0 (0 %)	0 (0 %)	0 (0 %)
9	Brasília-DF	129	5 (3.9 %)	0 (0 %)	10 (7.8 %)
10	Campina Grande	1	0 (0 %)	0 (0 %)	0 (0 %)

Absolute numbers?
Percentage?

11730 101 3 PROFESIONAL SERVICIO CIVIL 2 a 667,450 Cls c/u

11731 101 3 PROFESIONAL SERVICIO CIVIL 3 a 726,100 Cls c/u

Different variables in a single cell

Multiple values

country	year	column	cases
AD	2000	m014	0
AD	2000	m1524	0
AD	2000	m2534	1
AD	2000	m3544	0
AD	2000	m4554	0
AD	2000	m5564	0
AD	2000	m65	0
AE	2000	m014	2
AE	2000	m1524	4
AE	2000	m2534	4
AE	2000	m3544	6
AE	2000	m4554	5
AE	2000	m5564	12
AE	2000	m65	10
AE	2000	f014	3

(a) Molten data

country	year	sex	age	cases
AD	2000	m	0-14	0
AD	2000	m	15-24	0
AD	2000	m	25-34	1
AD	2000	m	35-44	0
AD	2000	m	45-54	0
AD	2000	m	55-64	0
AD	2000	m	65+	0
AE	2000	m	0-14	2
AE	2000	m	15-24	4
AE	2000	m	25-34	4
AE	2000	m	35-44	6
AE	2000	m	45-54	5
AE	2000	m	55-64	12
AE	2000	m	65+	10
AE	2000	f	0-14	3

(b) Tidy data

Typing errors: the same entity should be typed precisely. Acronyms vs full names.

Name
Ana Rojas
Ana Rojas
Ana Rojas

This should be a variable called income

Values as variables

religion	<\$10k	\$10–20k	\$20–30k	\$30–40k	\$40–50k	\$50–75k
Agnostic	27	34	60	81	76	137
Atheist	12	27	37	52	35	70
Buddhist	27	21	30	34	33	58
Catholic	418	617	732	670	638	1116
Don't know/refused	15	14	15	11	10	35
Evangelical Prot	575	869	1064	982	881	1486
Hindu	1	9	7	9	11	34
Historically Black Prot	228	244	236	238	197	223
Jehovah's Witness	20	27	24	24	21	30
Jewish	19	19	25	25	30	95

Table 4: The first ten rows of data on income and religion from the Pew Forum. Three columns, \$75–100k, \$100–150k and >150k, have been omitted.

Values as variables

Tidy format

religion	income	freq
Agnostic	<\$10k	27
Agnostic	\$10–20k	34
Agnostic	\$20–30k	60
Agnostic	\$30–40k	81
Agnostic	\$40–50k	76
Agnostic	\$50–75k	137
Agnostic	\$75–100k	122
Agnostic	\$100–150k	109
Agnostic	>150k	84
Agnostic	Don't know/refused	96

Table 6: The first ten rows of the tidied Pew survey dataset on income and religion. The column has been renamed to `income`, and `value` to `freq`.

Check:

[The Quartz guide to bad data](#)
[Tidy data by Hadley Wickham](#)

Metadata

Information about a dataset such as its title and description, method of collection, author or publisher, area and time period covered, licence, date and frequency of release, variables. It is essential to publish data with adequate metadata to aid both discoverability and usability of the data.

Level of education
1
5
3
1
4
2

???

Level of education: *The individual's last year of schooling*

1= elementary

2= primary education

3=lower secondary

4=upper secondary

5=bachelor degree or higher



Metadatos estructurales

Campo	Nombre	Descripción	Tipo de dato	Es llave	Es requerido
Delito	Delito	Descripción del delito	Texto	No	Sí
SubDelito	Subdelito	Descripción del Subdelito	Texto	No	Sí
Fecha	Fecha	Fecha del delito	Fecha	No	Sí
Victima	Victima	Descripción de la victima	Texto	No	Sí
SubVictima	Subvictima	Descripción de la Subvictima	Texto	No	Sí
Edad	Edad	Descripción de la edad	Texto	No	Sí
Genero	Genero	Descripción del genero de la victima	Texto	No	Sí
Nacionalidad	Nacionalidad	Descripción de la nacionalidad de la victima	Texto	No	Sí
Provincia	Provincia	Descripción de la provincia	Texto	No	Sí
Canton	Canton	Descripción del cantón	Texto	No	Sí
Distrito	Distrito	Descripción del distrito	Texto	No	Sí

Sexo (A5S)

Archivo: I Trimestre 2011



PREGUNTA TEXTUAL

¿Es hombre o mujer?

FUENTE DE INFORMACIÓN

Persona de 15 años o más, que tenga un amplio conocimiento de las características sociodemográficas y económicas de los miembros del hogar (puede ser auto informante, otro informante del hogar o un informante fuera del hogar).

CATEGORÍAS

Valor	Categoría	Casos	Weighted	
1	Hombre	13054	2309320	 50.6%
2	Mujer	13505	2258144	 49.4%

Precaución: estas estadísticas reflejan el número de casos encontrados en el archivo de datos. No deben ser interpretadas como un resumen estadístico del universo de interés.

Interoperability

What?

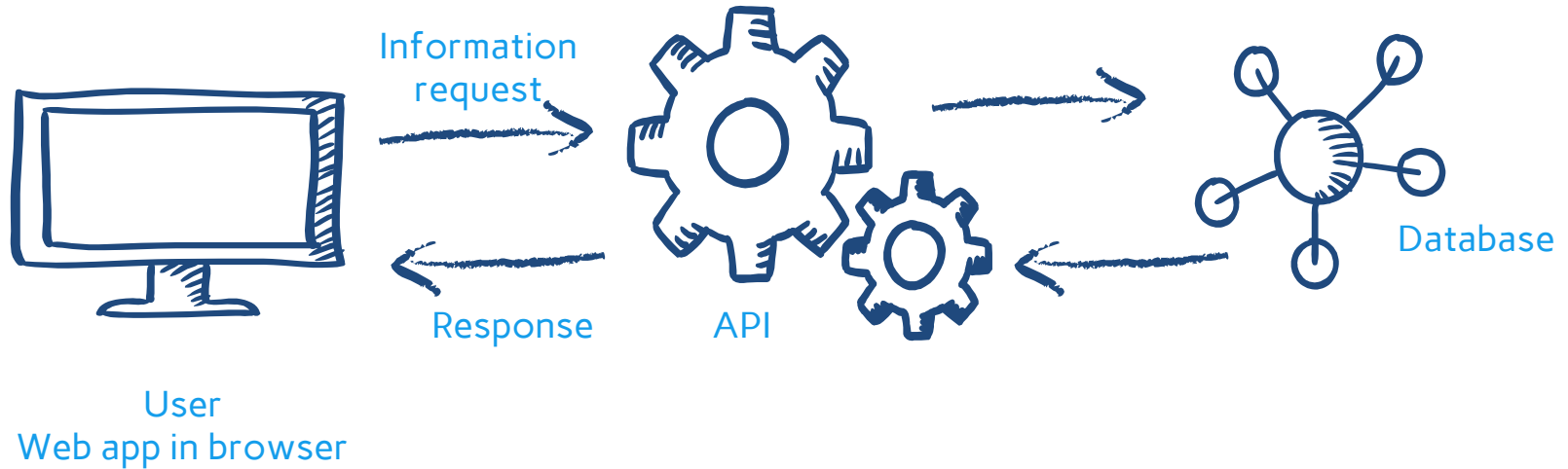
Interoperability denotes the ability of diverse systems and organizations to work together (inter-operate). In this case, it is the ability to interoperate - or intermix - different datasets.

See: opendatahandbook.org

API

Application Programming Interface. For data, this is usually a way provided by the data publisher for programs or apps to read data directly over the web.

API



Why?

- Saves time for users and public servants. They can easily access and analyse different datasets.
- The fact that the information is well documented and structured, allows to work with big databases and vast amounts of information.
- Increases the consumption of information within institutions.

Final remarks

Open Data

Don't be afraid to start an open data process.

Opening data has a large potential economic value, increases transparency, creates business opportunities, benefits consumers.

Thanks!