OPEN DATA TUTORIAL

Opening and promoting use of budget data

Debugging and cleaning

SHCP
SECRETARÍA DE HACIENDA Y CRÉDITO PÚBLICO

TRANSPARENCIA PRESUPUESTARIA
Observatorio del gasto

GIFT
GLOBAL INITIATIVE FOR FISCAL TRANSPARENCY

Start debugging and cleaning

When governments carry out their daily activities, they generate data, whether it is captured for processing by machines (in a system or structured as CSV files, for example) or published in formats that make it difficult to reuse them (unstructured, for example, in PDF documents).

Although the **data** is **structured** (those with an easily identifiable order such as rows and columns) in a data set, this does not imply that it is ready to be published or used, since there is a possibility that it contains errors. Therefore, it is necessary to propose a **debugging and cleaning strategy**, either manual or automatic.

Data cleaning is an essential step, since a data set with erroneous or duplicated information, empty fields and spelling mistakes, makes it impossible to correctly process, use and reuse the information.

Debugging and cleaning data can be a tedious and complicated task; however, by carrying out this process, an optimum level of presentation of information is guaranteed, which not only ensures the integrity, timeliness and readability of the data, but also its quality.

During the process of cleaning and debugging budget data, it is common to encounter certain obstacles.

In order to solve them and to obtain data as clean as possible from its origin, it is important to identify the good practices that guarantee the quality of the information and make sure that these good practices are shared with those responsible for generating it.

Below, you will find a check-list that shows some of the most common mistakes faced by those responsible for opening budget data:

TIP
As a good practice, those responsible for publishing the information should be motivated to modify or add the cleaning tools/practices that they consider pertinent.

1. Use of special characters

It is common to find data with characters such as currency signs ($), percentages (%), at (@), and many others. These special characters hinder the correct reading, and the automatic interpretation of information by data processors.

For example, when capturing data or receiving information showing the physical progress of a project, there is a chance to find the percentage sign (%) in the data source. In this case, it is suggested to select the column with this information and adjust its format to "Number" so that 100 and not 100% appears.

How does the public know that the data in this column is expressed as a percentage if it has already been deleted? That's important information and there are at least two simple ways to know it:

1.1 Incorporate a methodological note or specification in the data dictionary, indicating that a variable is expressed as a percentage.

1.2 Add a column named "Unit of Measure" in which the text "Percentage" is incorporated.

Another very common example is to find some currency sign if the captured data refers to a monetary amount. Again, it is suggested to adjust the format to "Number" so that the sign no longer appears in the corresponding cell.

Making these adjustments, as well as verifying the removal of any special character spotted improves the data processing software's ability to correctly read the information.

	A	B	C	D	E	F	G	H	I
1	IDENTIFICAD	NOMBRE	FECHA_INI_(	FECHA_FIN_(	FECHA_INI_F	FECHA_FIN_I	AVANCE_FISICO	MONTO_TOTAL_IN	FASE
2	'0209J3A0001	Construcción	01/01/2001	01/12/2015	01/01/2001	01/12/2015	90%	$196,831,070.00	Concluido
3	'0306G1C001	Bienes Inmu	01/12/2002	01/11/2014	01/12/2002	01/12/2002	90%	$7,053,728.00	Concluido
4	'0306G1C001	Bienes Inmu	01/05/2002	01/05/2014	01/05/2002	01/12/2002	90%	$5,236,295.00	Concluido
5	'0309J3A0002	Proteccion d	01/01/2001	01/12/2017	01/01/2001	01/12/2017	100%	$394,941,593.00	Concluido
6	'0409J3A0004	Oficinas adm	01/01/2005	01/12/2014	01/01/2005	01/12/2012	100%	$202,448,403.00	Concluido
7	'0410K2N000	Construcción	01/08/2005	01/12/2014	01/08/2005	01/12/2014	90%	$204,821,522.00	Concluido
8	'0509632002(	Pueblo Hidal	01/08/2005	01/12/2015	01/08/2005	01/12/2005	100%	$70,742,185.00	Concluido
9	'0509J3A0004	REUBICACIÓI	01/02/2006	01/12/2015	01/01/2011	01/12/2013	90%	$92,720,542.00	Concluido
10	'0510K8V000	Adquisición	01/07/2006	01/12/2014	01/07/2006	01/12/2009	90%	$26,667,664.00	Concluido
11	'0610K2N000	Construcción	01/08/2006	01/12/2014	01/08/2006	01/12/2012	90%	$82,167,033.00	Concluido
12	'0809J3A0002	Construccior	01/01/2009	01/12/2015	01/01/2009	01/12/2015	95%	$113,349,489.00	Concluido
13	'0809J3B0001	Desarrollo d	01/01/2008	01/12/2015	01/01/2008	01/12/2015	95%	$3,078,141,596.00	Concluido
14	'0810K2N000	Optimizació	01/09/2008	01/12/2015	01/09/2008	01/12/2015	95%	$105,560,129.00	Concluido
15	'0812514000	Construcción	01/06/2008	01/12/2015	01/10/2008	01/12/2008	90%	$1,518,676,724.00	Concluido
16	'0812NBG00(	Construcción	01/01/2008	01/12/2014	01/01/2008	01/12/2015	97%	$382,656,604.00	Concluido
17	'0909J2U0007	DESARROLLC	01/01/2010	01/04/2015	01/01/2010	01/04/2015	97%	$25,907,499.00	Concluido
18	'1008I000002	Dragado de r	01/09/2010	01/04/2016	01/09/2010	01/12/2012	97%	$165,213,751.00	Concluido
19	'1008I000003	Dragado de r	01/09/2010	01/04/2016	01/09/2010	01/12/2012	97%	$169,940,128.00	Concluido
20	'1009625000	Antiguo Gen	01/06/2010	01/12/2017	01/06/2010	01/12/2010	100%	$131,460,229.00	Concluido
21	'1009635002	Camino a Lu	01/07/2010	01/12/2014	01/07/2010	01/12/2010	100%	$61,418,778.00	Concluido
22	'1009635002	Puente Gran	01/07/2010	01/12/2017	01/07/2010	01/12/2010	100%	$16,672,839.00	Concluido
23	'1009635002	Jiquipilco el	01/07/2010	01/12/2015	01/12/2010	01/12/2010	100%	$36,706,878.00	Concluido
24	'1010K2N000	Reparacion g	01/01/2011	01/12/2015	01/01/2011	01/12/2013	100%	$45,003,136.00	Concluido
25	'1010K2N001	Optimizacior	01/01/2011	01/12/2014	01/07/2011	01/12/2012	100%	$23,682,767.00	Concluido
26	'1011MAR00(	Programa de	01/03/2011	01/09/2014	01/09/2012	01/12/2014	100%	$63,510,322.00	Concluido
27	'1105613000	Programa de	01/01/2012	01/12/2014	01/07/2011	01/12/2011	100%	$39,181,420.00	Concluido
28	'1109510001(	Proyecto de	01/01/2012	01/12/2015	01/07/2012	01/12/2012	100%	$213,354,992.00	Concluido

Example on the use of special characters. In this template, columns G and H use the percentage sign (%) and the sign of currency ($) in specific fields, which makes them difficult to process and read.

	A	B	C	D	E	F	G	H	I	J
1	institucion	unidad_resp	sistema_regi	descripciÃ³n	campos_contenidos					
2	Agencia Mexica	DirecciÃ³n G	Registro Nac	Registro de l	RegiÃ³n, contraparte, tÃ-tulo del proyecto, sector, vertiente, Ã¡mbito					
3	ComisiÃ³n Fede	CoordinaciÃ³	Anteproyect	Registro de l	Cambio en la regulaciÃ³n; RegulaciÃ³n; Fecha del cambio; InstituciÃ³n					
4	ComisiÃ³n Fede	CoordinaciÃ³	Registro Fed	Inventario el	Homoclave; Siglas; Dependencia; Unidad Administrativa; Nombre del trÃ¡mite					
5	Instituto Nacio	ComisiÃ³n de	Registro Fed	-	-Entidad					
6	Instituto Nacio	ComisiÃ³n de	Organizacior	Procedimier	OSC; Expediente administrativo; Fecha y Acuerdo de la SesiÃ³n; Fecha de Resc					
7	Instituto Nacio	Subsistema I	Directorio Es	Se ofrecen lo	Datos de identificaciÃ³n; ubicaciÃ³n; actividad econÃ³mica; y tamaÃ±o de los r					
8	Secretariado Ej	ProcuraduriÃ	Incidencia de	Presuntos de	Campo,DescripciÃ³n ANO,AÃ±o de registro de las Averiguaciones Previas y/o c					
9	ProcuraduriÃ-a	ProcuraduriÃ	EstadÃ-stica	Control de P	Por tipo de consignaciÃ³n (con y sin detenido); Por Unidad; Por delito; Por aÃ±					
10	Servicio de Adr	Jefatura de S	Directorio de	Registro de	RFC; RazÃ³n social; Tipo donataria; entidad federativa; Donativo efectivo nacio					
11	SecretariÃ-a de	SubsecretariÃ	Sistema de i	Integra un re	RazÃ³n social, estado, municipio, domicilio, colonia, cÃ³digo postal, telÃ©fon					
12	SecretariÃ-a de	DirecciÃ³n G	Listado del R	Conjunto de	NÃºmero de expediente; RazÃ³n social; PaÃ-s de origen; Sector; Estado; Munic					
13	SecretariÃ-a de	CoordinaciÃ³	Plataforma N	Se integra la	tipo de predio, nÃºmero_cta, predio, delimitaciones					
14	SecretariÃ-a de	DirecciÃ³n G	Diario Oficia	Normas Ofic	descripciÃ³n, alta, fecha, autores					
15	SecretariÃ-a de	Unidad de Pe	Inventario de	-Normas	Tipo de norma; Homoclave; Nombre de la norma; Materia; InstituciÃ³n; Tipo d					
16	SecretariÃ-a de	Unidad de Pe	Sistema de I	Registro de l	Ramo, Sector , DescripciÃ³n, Estatus, Resultado					
17	SecretariÃ-a de	Unidad de Pe	CompraNet	Datos, bajo e	Gobierno; Siglas; Dependencia; Clave de Unidad Compradora; Nombre de la U					
18	SecretariÃ-a de	DirecciÃ³n G	Directorio de	Datos a 2015	Infractor; NÃºmero de expediente; Fecha de notificaciÃ³n de resoluciÃ³n; Pub					
19	SecretariÃ-a de	Unidad de Co	Programa An	Programa An	Ejercicio; NÃºmero; Ente pÃºblico; Tipo de Ente pÃºblico; Proyectos financiado					
20	SecretariÃ-a de	DirecciÃ³n G	Lista de firm	Listado de fi	Ejercicio; Mes; Corte; No.; Firma; Socio_Director; DirecciÃ³n, TelÃ©fono					
21	SecretariÃ-a de	DirecciÃ³n G	Resultados d	Datos de la d	DesignaciÃ³n; No.; Ente pÃºblico; Firma de auditores externos; Ã³rgano Intern					

Example of how special characters are displayed when processed by an XLSX or CSV file.

2. Misspellings and homonyms that create or disappear geographic spaces

Data sets must correctly express the data to which they refer, for this, it is necessary to establish language criteria that ensure effective communication, otherwise, it could lead to misleading interpretations of the information.

Ortographic errors are frequent, such as the inclusion of a misplaced accent or the omission of a letter. However, it is also true that they can completely alter the meaning of the information.

For example, *Mexico* is not the same as *Mexiko*; the first case (Mexico) refers to a country that has over 120 million inhabitants, while the second one (Mexiko), refers to a "place" that cannot be located on a map.

The recommendation is to use and maintain the names provided by the institution in charge of the statistical or geographic information in each country or to use an international standard as a reference, which may provide a guideline on how the information is written appropriately.

	A	B	C	D	E	F	G	H	I	J	K	L
1	FECHA_EVENTO	TIPO_EVEI	TIPO_INFRAESTRU	ID_INFRAESTRUC	DESC_INFRAESTRUCTURA	DIRECCION_I	ID_ENTIDAD_	ENTIDAD_FEDE	ID_MUNICI	MUNICIPIO	LATITUD	LONGITUD
2	2017-09-19	Sismo	Hospitalario Terc	36A101112153	HG CMN La Raza	Calzada Valle	15	México	2	Azcapotzalcc	19.4860	-99.1457
3	2017-09-19	Sismo	Hospitalario Terc	36A102142153	HGO 3 CMN La Raza	Calle Seris es	15	México	2	Azcapotzalcc	19.4660	-99.1820
4	2017-09-19	Sismo	Hospitalario Terc	36A1041C2153	HES CMN La Raza	Calle Seris y	15	México	2	Azcapotzalcc	19.4665	-99.1458
5	2017-09-19	Sismo	Apoyo Hospitalar	36A131322153	C. Externa CMN La Raza	Calle Seris es	15	México	2	Azcapotzalcc	19.4665	-99.1458
6	2017-09-19	Sismo	HospitalarioTerce	37B504132153	HONCO CMN Siglo XXI O	Avenida Cua	15	México	15	Cuauhtémoc	19.4084	-99.1544
7	2017-09-19	Sismo	Hospitalario Terc	37B503122153	HP CMN Siglo XXI Pediati	Avenida Cua	15	México	15	Cuauhtémoc	19.4087	-99.1544
8	2017-09-19	Sismo	Hospitalario Terc	37B5091C2153	HES CMN Siglo XXI Espec	Avenida Cua	15	México	15	Cuauhtémoc	19.4093	-99.1545
9	2017-09-19	Sismo	Hospitalario Terc	37B5071B2153	HC CMN Siglo XXI Cardio	Avenida Cua	15	México	15	Cuauhtémoc	19.4089	-99.1545
10	2017-09-19	Sismo	Hospitalario Segu	362001062151	HGR 25 I. Zaragoza	Avenida Igna	15	México	7	Iztapalapa	19.3882	-99.0398
11	2017-09-19	Sismo	Hospitalario Segu	181103022151	HGR MF 7 Cuautla	Calle Tulipan	15	México	6	Cuautla	18.8111	-98.9497
12	2017-09-19	Sismo	Hospitalario Segu	181501022151	HGZ 5 Zacatepec	Avenida Láza	15	Méjico	31	Zacatepec	18.6520	-99.1980
13	2017-09-19	Sismo	Hospitalario Segu	180112072151	HGRMF1 Cuernavaca	Avenida Plar	15	México	7	Cuernavaca	18.9217	-99.2058
14	2017-09-19	Sismo	Hospitalario Segu	224190012151	HGZ 5 Metepec	Km. 4.5 Carre	15	México	19	Atlixco	18.9305	-98.4713
15	2017-09-19	Sismo	Hospitalario Segu	38A520012151	HGZ 32 Villacoapa	Calzada del H	15	México	12	Tlalpan	19.3072	-99.1316
16	2017-09-19	Sismo	Hospitalario Prim	385348252110	UMF 21 Francisco del Pas	Francisco del	15	Méjico	17	Venustiano C	19.4177	-99.1133
17	2017-09-19	Sismo	Hospitalario Segu	220120062151	HGR 36 San Alejandro	10 Poniente	15	México	114	Puebla	19.0582	-98.2164
18	2017-09-19	Sismo	Hospitalario Prim	221504252110	UMF 32	Independen	15	México	51	Chietla	18.4842	-98.6973
19	2017-09-19	Sismo	Administrativa	95218	Tokio 80	Tokio 80, Col	15	México	15	Cuauhtémoc	19.4232774	-99.1725336
20	2017-09-19	Sismo	Administrativa	912343	Toledo 21	Toledo 21, C	15	México	15	Cuauhtémoc	19.4231342	-99.1723549
21	2017-09-19	Sismo	Administrativa	9166715	Villalongin 117	Manuel villal	15	México	15	Cuauhtémoc	19.4340043	-99.1654824
22	2017-09-19	Sismo	Administrativa	Arrendado	Revolución OIC	Avenida Rev	15	México	10	Álvaro Obreg	19.350171	-99.1902677
23	2017-09-19	Sismo	Administrativa	Arrendado	Hamburgo 18	Hamburgo 18	15	México	15	Cuauhtémoc	19.428864	-99.1592234
24	2017-09-19	Sismo	Administrativa	Arrendado	Tiburcio Montiel	Tiburcio Mor	15	México	16	Miguel Hidal	19.4146704	-99.1851953
25	2017-09-19	Sismo	Administrativa	911275	Durango 289	Durango 289,	15	México	15	Cuauhtémoc	19.4190328	-99.1714798
26	2017-09-19	Sismo	Administrativa	912352	Durango 291	Durango 291,	15	México	15	Cuauhtémoc	19.4190166	-99.1715328
27	2017-09-19	Sismo	Administrativa	911284	Durango 323	Durango 323,	15	México	15	Cuauhtémoc	19.4187264	-99.1729352
28	2017-09-19	Sismo	Administrativa	9550001	Hamburgo 289	Hamburgo 28	15	México	15	Cuauhtémoc	19.4226741	-99.1723415
29	2017-09-19	Sismo	Administrativa	912263	Reforma 476	Reforma 476	15	México	15	Cuauhtémoc	19.4235294	-99.1729305
30	2017-09-19	Sismo	Administrativa	912334	Sevilla 33	Sevilla 33, Cc	15	México	15	Cuauhtémoc	19.42304	-99.1705658
31	2017-09-19	Sismo	Administrativa	9166528	Tláloc 90	Tlalóc 90, Col	15	México	16	Miguel Hidal	19.4386457	-99.17071
32	2017-09-19	Sismo	Administrativa	95343	Tokio 104	Tokio 104, Cc	15	México	15	Cuauhtémoc	19.4229612	-99.173409
33	2017-09-19	Sismo	Administrativa	912307	Tokio 92	Tokio 92, Col	15	México	15	Cuauhtémoc	19.4230729	-99.1731676
34	2017-09-19	Sismo	Administrativa	9166475	Toledo 10	Toledo 10, C	15	México	15	Cuauhtémoc	19.4237878	-99.1725051

Example of the alteration of geographic spaces. In column G, misspellings when writing Mejico refer to places that cannot be referenced on a map.

3. Creation of new categories and data based on personal or abitrary criteria

To preserve the quality of the information, is important to keep the data as it was provided by its source; otherwise, when modifications are made, it will be difficult to review and analyze them, especially when working with large data sets.

For example, having a category already defined for a variable: Not available and N/A (meaning not applicable), it is only required for the field to be filled with these two values. By adding a category that was not originally considered as "Not applicable" instead of N/A, it becomes difficult to query the data since the information is altered.

Categories allow grouping data according to a shared characteristic or property. Therefore, it is worth respecting the assigned catalogue and not establishing new identification values. In this sense, cleanliness implies that previously established attributes are preserved.

	A	B	C	D	E	F	G
1	STATE	EJCUC_EJID	EJCUC_PARC	EJCUC_CAMI	EJCUC_CAM(	REGISTRY	
2	Aguascalient	393	614	934	547	N/A	
3	Baja Californ	371	224	305	598	Not applicable	
4	Baja Californ	96	176	458	1403	N/A	
5	Campeche	1052	543	686	1427	N/A	
6	Chiapas	1409	1120	1422	887	N/A	
7	Chihuahua	826	555	350	1222	N/A	
8	Ciudad de M	1013	199	1355	64	N/A	
9	Coahuila	1378	1200	690	54	Not applicable	
10	Colima	1268	668	361	1470	Not applicable	
11	Durango	544	1013	961	904	Not applicable	
12	Guanajuato	1420	335	1095	718	Not applicable	
13	Guerrero	641	617	650	584	Not applicable	
14	Hidalgo	558	550	192	367	Not applicable	
15	Jalisco	125	373	381	54	Not available	
16	México	191	1457	254	608	Not available	
17	Michoacán	220	194	761	204	Not available	
18	Morelos	683	896	722	623	Not available	
19	Nayarit	732	628	85	129	Not available	
20	Nuevo León	1314	1111	971	919	N/A	
21	Oaxaca	603	10	481	837	N/A	
22	Puebla	478	384	1079	1012	Not applicable	
23	Querétaro	1040	191	1100	179	Not applicable	
24	Quintana Ro	618	1219	1192	303	Not available	
25	San Luis Pot	581	676	949	491	Not available	
26	Sinaloa	935	847	1076	704	Not available	
27	Sonora	990	716	608	683	N/A	
28	Tabasco	768	808	410	189	N/A	
29	Tamaulipas	1343	746	293	31	N/A	
30	Tlaxcala	61	223	222	1463	N/A	
31	Veracruz	1464	723	529	941	N/A	

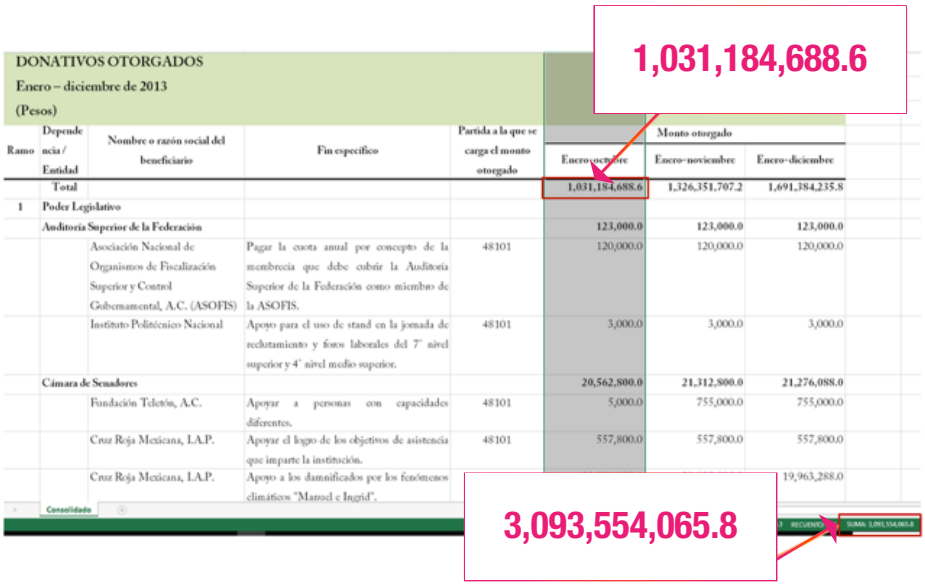
	A	B
1		
2		
3	**Etiquetas de fila**	**Cuenta de REGISTRY**
4	N/A	15
5	Not applicable	9
6	Not available	8
7	**Total general**	**32**
8		
9		

Example of modifications in terms of defined categories. In this case, the corresponding catalog in column F (Registration) requires that only “N/A” or “Not available” should be written; nevertheless, the modification of the categories when adding “Not applicable” alters the query of the data. When performing a dynamic table, we see that writing erroneous categories causes different results and makes it difficult to group data.

4. Additional calculations in tabular formats that result in deceptive sums of information

Using subtotals is a common practice in reports for executive reading. However, when working with large data sets meant to be used for specific purposes, the inclusion of columns or cells in which calculations have been made can lead to double counting of the data.

Therefore, it is recommended not to group the information or to make any other calculation inside the data set that may be confused with the data itself.



DONATIVOS OTORGADOS
Enero – diciembre de 2013
(Pesos)

Ramo	Dependencia / Entidad	Nombre o razón social del beneficiario	Fin específico	Partida a la que se carga el monto otorgado	Monto otorgado		
					Enero-octubre	Enero-noviembre	Enero-diciembre
	Total				1,031,184,688.6	1,326,351,707.2	1,691,384,235.8
1	Poder Legislativo						
	Auditoría Superior de la Federación				123,000.0	123,000.0	123,000.0
		Asociación Nacional de Organismos de Fiscalización Superior y Control Gubernamental, A.C. (ASOFIS)	Pagar la cuota anual por concepto de la membrecía que debe cubrir la Auditoría Superior de la Federación como miembro de la ASOFIS.	48101	120,000.0	120,000.0	120,000.0
		Instituto Politécnico Nacional	Apoyo para el uso de stand en la jornada de reclutamiento y ferias laborales del 7° nivel superior y 4° nivel medio superior.	48101	3,000.0	3,000.0	3,000.0
	Cámara de Senadores				20,562,800.0	21,312,800.0	21,276,088.0
		Fundación Teletón, A.C.	Apoyar a personas con capacidades diferentes.	48101	5,000.0	755,000.0	755,000.0
		Cruz Roja Mexicana, I.A.P.	Apoyar el logro de los objetivos de asistencia que imparte la institución.	48101	557,800.0	557,800.0	557,800.0
		Cruz Roja Mexicana, I.A.P.	Apoyo a los damnificados por los fenómenos climáticos "Manuel e Ingrid".				19,963,288.0

Example of how additional calculations hinder the interpretation of information. There is a subtotal that is already accounted for in the lower part, leading to an apparent duplication of the original amount.

	A	B	C	D	E	F	G	H	I	J	K	L
1	CICLO	DESC_UR	DESC_FF	ID_EN	ENTIDAD_	ID_CLAVE_CAR	MONTO_AP	MONTO_MOI	MONTO_DEVE	MONTO_PAC	MONTO_	MONTO_EJERCICIO
2	2017	Juntas Dis	Recursos fiscales	4	Campeche	00000000000	10080.00	6678.00	6678.00	6678.00	0.00	6678.00
3	2017	Juntas Dis	Recursos fiscales	5	Coahuila	00000000000	25200.00	25200.00	25200.00	25200.00	0.00	25200.00
4	2017	Juntas Dis	Recursos fiscales	6	Colima	00000000000	10080.00	10080.00	10080.00	10080.00	0.00	10080.00
5	2017	Juntas Dis	Recursos fiscales	7	Chiapas	00000000000	60480.00	62160.00	62160.00	62160.00	0.00	62160.00
6	2017	Juntas Dis	Recursos fiscales	8	Chihuahua	00000000000	45360.00	43456.00	43456.00	43456.00	0.00	43456.00
7	2017	Juntas Dis	Recursos fiscales	9	Ciudad de	00000000000	136080.00	125424.00	125424.00	125424.00	0.00	125424.00
8	2017	Juntas Dis	Recursos fiscales	1	Aguascalie	00000000000	88210.00	0.00	0.00	0.00	0.00	0.00
9	2017	Juntas Dis	Recursos fiscales	10	Durango	00000000000	110124.00	0.00	0.00	0.00	0.00	0.00
10	2017	Juntas Dis	Recursos fiscales	11	Guanajuat	00000000000	385434.00	0.00	0.00	0.00	0.00	0.00
11	2017	Juntas Dis	Recursos fiscales	12	Guerrero	00000000000	247872.00	0.00	0.00	0.00	0.00	0.00
12	2017	Juntas Dis	Recursos fiscales	13	Hidalgo	00000000000	189487.00	0.00	0.00	0.00	0.00	0.00
13	2017	Juntas Dis	Recursos fiscales	14	Jalisco	00000000000	515456.00	0.00	0.00	0.00	0.00	0.00
14	2017	Juntas Dis	Recursos fiscales	15	Estado de	00000000000	1114460.00	0.00	0.00	0.00	0.00	0.00
15	2017	Juntas Dis	Recursos fiscales	16	Michoacár	00000000000	332382.00	0.00	0.00	0.00	0.00	0.00
16	2017	Juntas Dis	Recursos fiscales	14	Jalisco	00000000000	206681.00	203918.40	203918.40	203918.40	0.00	203918.40
17	2017	Juntas Dis	Recursos fiscales	15	Estado de	00000000000	434963.00	435869.45	435869.45	435869.45	0.00	435869.45
18	2017	Juntas Dis	Recursos fiscales	16	Michoacár	00000000000	137080.00	130952.23	130952.23	130952.23	0.00	130952.23
19	2017	Juntas Dis	Recursos fiscales	17	Morelos	00000000000	57161.00	59072.40	59072.40	59072.40	0.00	59072.40
20	2017	Juntas Dis	Recursos fiscales	18	Nayarit	00000000000	34327.00	29171.73	29171.73	29171.73	0.00	29171.73
21	2017	Juntas Dis	Recursos fiscales	19	Nuevo Le	00000000000	131959.00	123853.15	123853.15	123853.15	0.00	123853.15
22	2017	Juntas Dis	Recursos fiscales	20	Oaxaca	00000000000	125587.00	115866.53	115866.53	115866.53	0.00	115866.53
23	2017	Juntas Dis	Recursos fiscales	21	Puebla	00000000000	182824.00	169812.06	169812.06	169812.06	0.00	169812.06
24	2017	Juntas Dis	Recursos fiscales	22	Querétaro	00000000000	45744.00	51273.12	51273.12	51273.12	0.00	51273.12
25	2017	Juntas Dis	Recursos fiscales	23	Quintana I	00000000000	34269.00	39837.42	39837.42	39837.42	0.00	39837.42
26	2017	Juntas Dis	Recursos fiscales	24	San Luis P	00000000000	78940.00	76848.03	76848.03	76848.03	0.00	76848.03
27	Total de Ejercido											1709472.52

Subtotal example. There is a figure already accounted for in the last row, which makes interpretation difficult.

5. Numbers are not text, that is, the format of presentation of the data must correspond to the type of information to which it refers

During the process of cleaning, it is important to validate that the format of the cell corresponds to its content.

For example, the fields that express monetary quantities or magnitudes, must remain in numeric format of integer type; the category fields must have a text format and the date and time fields must be encoded with the ISO-8601 standard (for more information on the standardization of data, you can review the document "Verify national and/or international standards").

When this attribute is not met, machines can misinterpret the data and lead us away from the desired results; for example, it might be that a spreadsheet processor does not perform the sum of numbers stored in text format, assign encoding of exponential mathematical operation to text identifiers that include numbers and are not classified as text, etc.

	A	B	C	D	E	F	G
1	Nombre del	Nombre del	Tipo de Recu	Convocatoria	Rubro	Fecha de Inicio	Fecha de Finalizacio
2	BIOINSUMOS	LOPEZ AGUIL	RECURSO PR	SECRETARIA DEL MEDIO AMBIENTE	DESARROLLO SECTORIAL	01/10/2012	31/12/2021
3	OPERACION	VILLARREAL	RECURSO PR	MARIMEX DEL PACIFICO, S.A. DE C.	APOYO EMPRESAS	02/10/2012	01/01/2022
4	INNOVACIO	MAZON SUA	RECURSO PR	PRODUCTORA DE ESPECIES ACUATI	APOYO EMPRESAS	03/10/2012	02/01/2022
5	EVALUACION	VILLARREAL	RECURSO PR	FUNDACIONES INTERNACIONALES	APOYO EMPRESAS	04/10/2012	03/01/2022
6	MEJORAMIE	PEREZ ENRIC	RECURSO PR	BLUE GENETICS MEXICO, S.A. DE C.	APOYO EMPRESAS	05/10/2012	04/01/2022
7	SERVICIOS D	LOPEZ AGUIL	RECURSO PR	CH2M HILL INC.	APOYO EMPRESAS	06/10/2012	05/01/2022
8	ANALISIS DE	LOPEZ TAPIA	RECURSO PR	PATRONATO DEL BOSQUE Y ZOOLC	DESARROLLO REGIONAL	07/10/2012	06/01/2022
9	INNOVACIO	MAZON SUA	RECURSO PR	PROPIOS	APOYO EMPRESAS	08/10/2012	07/01/2022
10	DESARROLLC	MAGALLON I	RECURSO PR	PROMOTORA INDUSTRIAL ACUASIS	APOYO EMPRESAS	09 de octubre de 2012	08/01/2022
11	PROYECTO C	BURROLA SA	RECURSO PR	VITALFOODS, S DE R.L. DE C.V.	APOYO EMPRESAS	10 de octubre de 2012	09/01/2022
12	PLANTA PILC	BELTRAN MC	RECURSO PR	INGENIERIA RECONSTRUCTIVA, S.A	APOYO EMPRESAS	11 de octubre de 2012	10/01/2022
13	SERVICIOS D	CORONADO	RECURSO PR	PROPIOS	APOYO EMPRESAS	12/10/2012	11/01/2022
14	COMPARATI	VILLARREAL	RECURSO PR	SKRETTING AQUACULTURE RESEAR	APOYO EMPRESAS	13/10/2012	12/01/2022
15	CARAVANAS	DIAZ CASTRC	RECURSO PR	Gobierno de BCS	DESARROLLO REGIONAL	14/10/2012	13 de enero de 2022
16	ESTUDIOS AI	SALINAS ZAV	RECURSO PR	CFE	APOYO EMPRESAS	15/10/2012	14 de enero de 2022
17	ENVIROMEN	SALINAS ZAV	RECURSO PR	BAJA CALIFORNIA SUR GAS TRANSI	APOYO EMPRESAS	16/10/2012	15 de enero de 2022
18	CARACTERIZ	MALDONADO	RECURSO PR	PROPIOS	APOYO EMPRESAS	17/10/2012	16/01/2022
19	SISTEMA SUF	QUIROZ GUZ	RECURSO PR	GRANJAS MARINAS DE SINALOA, S	APOYO EMPRESAS	18/10/2012	17/01/2022
20	BIOTECNOLC	RODRIGUEZ I	RECURSO PR	PROPIOS	GENERACION DE CONOC	19/10/2012	18/01/2022
21	SERVICIOS D	CRUZ HERNA	RECURSO PR	PROPIOS	APOYO EMPRESAS	20/10/2012	19/01/2022
22	EVALUACION	BURROLA SA	RECURSO PR	PROPIOS	APOYO EMPRESAS	21/10/2012	20/01/2022
23	ESTUDIO OCI	BURROLA SA	RECURSO PR	PROPIOS	APOYO EMPRESAS	22/10/2012	21/01/2022
24	ANALISIS A M	CORONADO	RECURSO PR	PROPIOS	APOYO EMPRESAS	23/10/2012	22/01/2022
25	CREACION D	CHAVEZ VILL	RECURSO PR	PROPIOS	DESARROLLO REGIONAL	24/10/2012	23/01/2022
26	OBTENCION	HERNANDEZ	RECURSO PR	PROPIOS	APOYO EMPRESAS	25/10/2012	24/01/2022
27	ESTUDIOS Y A	VALENZUELA	RECURSO PR	PROPIOS	APOYO EMPRESAS	26/10/2012	25/01/2022
28	ESTUDIOS AI	SALINAS ZAV	RECURSO PR	FUNDACIONES INTERNACIONALES	APOYO EMPRESAS	27/10/2012	26/01/2022
29	LOS ARRECIF	CORDOVA M	RECURSO PR	UNIVERSIDAD AUTONOMA DE B.C.	GENERACION DE CONOC	28/10/2012	27/01/2022
30	MUESTREO Y	SALINAS ZAV	RECURSO PR	SEASE	EMPRESAS APOYADAS	29/10/2012	28/01/2022
31	INNOVACIO	MAZON SUA	RECURSO PR	.	EMPRESAS APOYADAS	30/10/2012	29/01/2022

Example of misrepresentation of the information for the date format according to ISO-8601.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Contrato	Anio	Servicio	Unidad	Entidad	Empresa	Fechainicio	Fechatermin	Porcentaje d	Porcentaje d	Avance Fina	Avance Finan	Porcentaje d	Estatus
2	CGA-SOC-C-(	2006	Unidad de M	Caborca	Sonora	DARVIO CON	18/07/2006	24/11/2006	100	100	3777736.99	3.78E+06	100	Concluidas
3	CGA-SOC-C-(	2007	Hospital Alta	Leon	Guanajuato	Servicios de	18/06/2007	16/12/2007	100	100	5546281.08	5.55E+06	99.9999998	Concluidas
4	CGA-SOC-C-(	2007	Hospital Ger	Gral. Jose M	Distrito Fede	Ingenieros y	20/06/2007	31/12/2007	100	100	4822214.42	4.82E+06	99.999443	Concluidas
5	CGA-SOC-C-(	2007	Hospital Ger	Dr. Dario Fe	Distrito Fede	Electromeca	05/07/2007	22/12/2007	100	100	3281048.89	3.31E+06	99.236986	Concluidas
6	CGA-SOC-C-(	2007	Clinica de Es	Complejo Zo	Distrito Fede	McBuild de K	05/07/2007	24/12/2007	100	100	12092823.5	1.21E+07	100	Concluidas
7	CGA-SOC-C-(	2007	Hospital Ger	Dr. Santiago	Durango	Coorporativc	09/07/2007	11/08/2008	100	100	9341097.47	9.34E+06	100	Concluidas
8	CGA-SOC-C-(	2007	Clinica Hospi	Rio Bravo	Tamaulipas	Darvio Const	09/07/2007	21/12/2007	100	100	7105557.85	7.73E+06	91.8800001	Concluidas
9	CGA-SOC-C-(	2007	Hospital Alta	Primero de	Distrito Fede	Constructora	10/07/2007	21/12/2007	100	100	4768405.85	5.37E+06	88.815501	Concluidas
10	CGA-SOC-C-(	2007	Clinica Hospi	Cd. Cancun	Quintana Ro	McBuild de K	16/07/2007	28/12/2007	100	100	4371015.3	4.39E+06	99.5432647	Concluidas
11	CGA-SOC-C-(	2007	Clinica Hospi	Nueva Rosita	Coahuila	JAE Grupo Cc	16/07/2007	10/05/2008	100	100	23932234.4	2.39E+07	100	Concluidas
12	CGA-SOC-C-(	2007	Hospital Ger	Tepic Nay.	Nayarit	Gritzel Const	16/07/2007	24/12/2007	100	100	2168577.37	2.17E+06	100	Concluidas
13	CGA-SOC-C-(	2007	Clinica Hospi	Matehuala	San Luis Pote	Grupo Integr	16/07/2007	30/12/2007	100	100	9204075.62	9.20E+06	100	Concluidas
14	CGA-SOC-C-(	2007	Clinica Hospi	Nogales	Sonora	Hector Const	18/07/2007	31/12/2007	100	100	4858825.41	4.86E+06	100	Concluidas
15	CGA-SOC-C-(	2007	Hospital Ger	Acapulco	Guerrero	JOCO Ingeni	13/07/2007	31/12/2007	100	100	3712591.9	3.71E+06	100	Concluidas
16	CGA-SOC-C-(	2007	Clinica de Es	Guadalajara	Jalisco	Medkitec S..	30/07/2007	03/12/2009	100	100	65226766.5	6.67E+07	97.7520263	Concluidas
17	CGA-SOC-C-(	2007	Hospital Ger	Acapulco	Guerrero	Grupo Triton	01/08/2007	23/10/2008	100	100	45642504.4	4.57E+07	99.8558132	Concluidas
18	CGA-SOC-C-(	2007	Clinica de M	Milpa Alta	Distrito Fede	Ingenieros y	20/08/2007	26/12/2007	100	100	1978013.89	1.98E+06	99.8485624	Concluidas
19	CGA-SOC-C-(	2007	Unidad de M	Tierra Blanca	Veracruz	Inmobiliaria	20/08/2007	26/12/2007	100	100	5706079.76	5.92E+06	96.3289864	Concluidas
20	CGA-SOC-C-(	2007	Clinica Hospi	Rio Bravo	Tamaulipas	Guteza Cons	20/08/2007	19/07/2008	100	100	10390510.9	1.04E+07	99.9939796	Concluidas
21	CGA-SOC-C-(	2007	Estancia Ter	E.T.T.E Miste	Distrito Fede	Axtral S.A. d	05/11/2007	05/07/2008	100	100	5630760.36	6.16E+06	91.363185	Concluidas
22	CGA-SOC-C-(	2007	Clinica de M	Cd. Lerdo	Durango	Aplicaciones	08/10/2007	14/12/2008	100	100	25873583.7	2.59E+07	100	Concluidas
23	CGA-SOC-C-(	2007	Clinica Hospi	Teziutlan	Puebla	Grupo Const	15/10/2007	11/02/2008	100	100	2925074.78	3.37E+06	86.6909306	Concluidas
24	CGA-SOC-C-(	2007	Clinica de M	Mixquiahuala	Hidalgo	C.V.M. Const	15/10/2007	28/11/2008	100	100	12433738.4	1.24E+07	100	Concluidas
25	CGA-SOC-C-(	2007	Unidad de M	Cosamaloapa	Veracruz	Del Valle Co	15/10/2007	11/05/2008	100	100	7032184.53	7.41E+06	94.9100288	Concluidas
26	CGA-SOC-C-(	2007	Clinica de M	Torreon	Coahuila	Medkitec S..	05/11/2007	31/03/2009	100	100	32922067.3	3.45E+07	95.4579492	Concluidas
27	CGA-SOC-C-(	2007	Clinica de Es	Churubusco	Distrito Fede	Constructora	05/11/2007	24/12/2007	100	100	1242686.53	1.26E+06	98.6040297	Concluidas
28	CGA-SOC-C-(	2007	Unidad de M	San Jose Del	Baja Californ	Maquinaria y	03/12/2007	29/06/2008	100	100	8393604.07	8.39E+06	99.9999993	Concluidas
29	CGA-SOC-C-(	2007	Clinica Hospi	Dr. Patricio T	Campeche	Grupo Const	27/11/2007	27/12/2007	100	100	1255202.1	1.26E+06	99.9999984	Concluidas
30	CGA-SOC-C-(	2007	Clinica Hospi	Piedras Negr	Coahuila	GR Edificacic	17/12/2007	14/05/2008	100	100	3637490.08	3.64E+06	100	Concluidas
31	CGA-SOC-C-(	2007	Clinica Hospi	Celaya	Guanajuato	Grupo Const	13/12/2007	09/07/2008	100	100	3222850.22	3.22E+06	100	Concluidas

Example of wrong format for the presentation of numbers; this type of format does not allow the captured data to be added.

6. Empty columns that presuppose information gaps

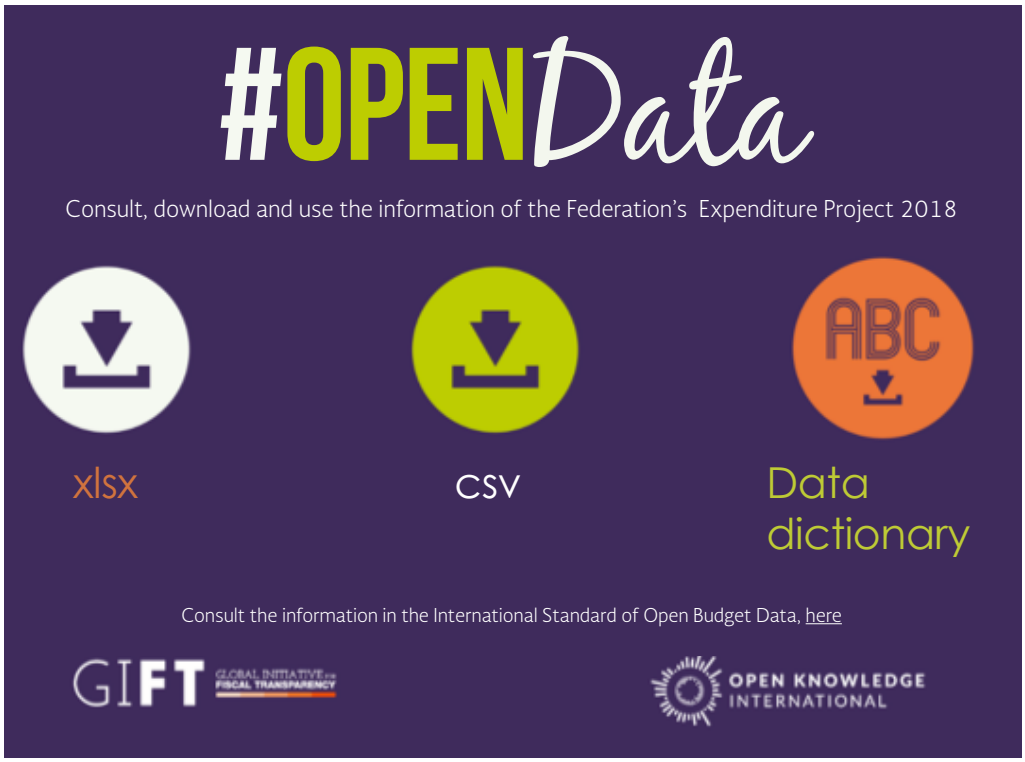
During the data cleaning process, it is important to eliminate the blank fields, especially when they are complete rows and columns, for they surely correspond to errors in the design of the data structure.

	A	B	C	D	E	F	G	H
1	ID	No.	Solicitante	Tipo.de.Perr	Destino	Autopista	Cadenamiento	Espacio
2	94	94	Inmobiliaria	Señal Inform	Señalamient	Cuernavaca-	101+390, 102+430, 103+330 y 103+860	
3	282	282	Adatiz S.A. d	Señal Inform	Señalamient	Chamapa-Le	14+200	
4	335	335	Parques Indu	Señal Inform	Señalamient	México-Puel	112+950 y 115+950	
5	374	374	CODECAM C	Señal Inform	Señalamient	Guadalajara-	15+383 y 45+447	
6	376	376	Concesionar	Señal Inform	Señalamient	México-Puel	31+100, 32+000, 32+400	
7	377	377	Concesionar	Señal Inform	Señalamient	Chamapa-Le	4+675, 6+140 y 4+725	
8	417	417	Concesionar	Señal Inform	Señalamient	México-Pach	24+420,25+600 y 27+600	
9	418	418	Concesionar	Señal Inform	Señalamient	México-Que	33+700, 55+100, 56+200 y salida a hue	
10	477	477	Centro Cultu	Señal Inform	Señalamient	Chamapa-Le	23+398	

Example of empty columns presentation (see column H).

7. Omission of a data dictionary, or the exclusion of the additional guide that helps the user understand the information provided

Every data set must be accompanied by a file that helps the user to identify the meaning of each field, as well as the additional information required to interpret it (references to legislation, other related data, methodology or calculus memory, etc.).



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	CICLO	ID_RAMO	DESC_RAMO	ID_UR	DESC_UR	GPO_FUNCIO	DESC_GPO_FUN	ID_FUNCION	DESC_FUNCION	ID_SUBFUNC	DESC_SUBFUNCION	ID_AI	DESC_AI	ID_MODALID	DESC_MODALIDAD	ID_PP	DESC_PP	ID_CAPITULO
2	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	K	Proyectos de inversión	27	Mantenimiento de infraestructura	6000
3	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
4	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
5	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
6	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
7	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
8	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
9	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
10	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
11	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
12	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
13	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
14	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
15	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
16	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
17	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
18	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
19	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
20	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
21	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
22	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
23	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
24	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
25	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
26	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
27	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
28	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000
29	2018	1	Poder Legislativ	100	H. Cámara de Dip	1	Gobierno	1	Legislación	1	Legislación	4	Llevar a cabo el proceso Le	R	Específicos	1	Actividades derivadas del trabajo legisla	1000

	A	B	C
1	**Variable**	**Descripción Variable**	**Etiqueta de Datos Abiertos**
2	Actividad Institucional	Clave que identifica las acciones sustantivas o de apoyo que realizan los ejecutores de gasto con el fin de dar cumplimiento a los objetivos y metas contenidos en los programas, de conformidad con las atribuciones que les señala su respectiva ley orgánica o el ordenamiento jurídico que les es aplicable (3 dígitos). Esta información está disponible de 2008 a 2018.	ID_AI
3	Adefas	Los adeudos de ejercicios fiscales anteriores (ADEFAS) son los gastos devengados y no pagados al último día de cada ejercicio fiscal, derivados del ejercicio del Presupuesto de Egresos, incluido el gasto devengado de las dependencias cuya cuenta por liquidar correspondiente está pendiente de presentarse a la Tesorería, así como las cuentas por liquidar presentadas a ésta que quedaron pendientes de pago. Esta información está disponible de 2014 a 2016.	MONTO_ADEFAS
4	Aprobado	Son las asignaciones presupuestarias anuales comprendidas en el Presupuesto de Egresos a nivel de clave presupuestaria en el caso de los ramos autónomos, administrativos y generales, y a nivel de los rubros de gasto que aparecen en las carátulas de flujo de efectivo para las entidades. Esta información está disponible de 2008 a 2018.	MONTO_APROBADO
5	Capítulo	Corresponde a la clave para el mayor nivel de agregación del Clasificador por Objeto del Gasto, que sirve para identificar conjuntos homogéneos y ordenados de bienes y servicios requeridos por las Entidades y Dependencias de todos los niveles de gobierno. En la clave del Clasificador por Objeto del Gasto, puede identificarse con el primer dígito de izquierda a derecha. Esta información está disponible de 2008 a 2018.	ID_CAPITULO
6	Ciclo	Determina el ciclo presupuestario de referencia. Esta información está disponible de 2008 a 2018.	CICLO

Example of a data dictionary, which must serve as a companion to a data set. Columns A, L and R are presented in the data set and are described in the dictionary.

A good practice to promote data cleaning is to indicate who is/are responsible for generating and publishing the information.

Thus, the public will be able to identify who is to be consulted about the data shown should any questions arise, since the area responsible for publishing the information is not necessarily responsible for generating it.

This document is a checklist to allow those responsible for generating and publishing information to solve common mistakes.

However, its purpose is merely illustrative, which is why this manual of good practices has also recommendations on how to deal with the most common errors in data processing, and while not an exhaustive list either, it provides a practical input for a deeper review.

TIP

This checklist can incorporate more practices since the data cleaning process is constant and generates specific experiences for those responsible for publishing budget data. Taking these points and the manual as examples, allows us to build the check list that best suits the needs of the information in practice.

Remember: data cleaning is an iterative process, not static.

After reviewing some strategies for the debugging and cleaning of information, be sure to check the next document, which explains the importance of thinking about the standards ruling information openness, so the data can be compared with those in other circumstances.

SHCP
SECRETARÍA DE HACIENDA Y CRÉDITO PÚBLICO

TRANSPARENCIA PRESUPUESTARIA
Observatorio del gasto

GIFT
GLOBAL INITIATIVE FOR FISCAL TRANSPARENCY