

TUTORIAL

DATOS ABIERTOS

Apertura y uso de datos presupuestarios

Identificar la fuente y formato

Identificar la fuente y el formato de los datos

Una vez priorizados los datos abiertos a publicar, de acuerdo con los usuarios potenciales y sus demandas de información, resulta vital ubicar la procedencia y el formato de la información.

¿Por qué es importante la procedencia y el formato de los datos? Si bien ahora la tendencia es generar sistemas que puedan ser abiertos de forma predeterminada, la realidad actual es que la mayoría de las veces **la información con la que contamos no fue creada pensando en la publicación y reuso, sino que es el producto de las gestiones cotidianas del gobierno**, por lo que trabajar en su apertura implicará análisis, procesamiento y creación de nuevos procesos para garantizar la sostenibilidad.

Por ello, vale la pena responder a las siguientes preguntas:

- ¿Los datos están en formato **impreso únicamente**?
- ¿Los datos a publicar se encuentran en el **mismo lugar**?
- ¿Cuántas personas son responsables de **generar, concentrar y/o publicar** estos datos?
- ¿En qué **formato** se encuentran: PDF, Word, Excel, CSV u otro?
- ¿Estos datos provienen de un **sistema informático**?
- De ser así, ¿de qué manera se puede **acceder** a ellos?
- ¿Están todos en el mismo sistema o se requiere **unir datos de varias herramientas tecnológicas**?



Considerar estas preguntas ayuda a establecer tiempos para la apertura de los datos; identificar las áreas involucradas en el proceso; analizar el estado de la calidad de la información; así como definir los retos y oportunidades del proceso de apertura.

TIP

Una recomendación es realizar un calendario operativo en conjunto con las áreas involucradas en el proceso de apertura de datos presupuestarios, para definir fechas y acciones hacia la publicación de la información.

Algunas consideraciones relevantes para identificar la fuente y el formato de los datos que se abrirán son:

1. Actores involucrados

La información presupuestaria puede estar dispersa entre áreas e incluso entre dependencias gubernamentales. Es importante tomar en cuenta que entre más dispersa se encuentre la información y más personas sean responsables de la misma, el proceso de apertura de datos dependerá en mayor medida de la construcción de consensos.

En el mismo sentido, un conjunto de datos que está en un solo lugar con menor número de responsables y del cual existe interés por utilizar, se convierte en la punta de lanza del proceso de apertura.

Así mismo, es significativo considerar que las áreas responsables de la información pueden no estar familiarizadas con los conceptos de apertura de datos, mientras que las personas encargadas de la apertura pueden no estar muy familiarizadas con los conjuntos de datos que se buscan publicar, por lo que es indispensable iniciar un proceso de aprendizaje de dos vías continuo, ya que de esto dependerá que la base de datos resultante y su diccionario de datos sean utilizables y de calidad.

2. Formato de los datos

Para que la información sea aprovechada efectivamente, es necesario considerar qué tan maleables o rígidos son los datos, si están ordenados para su procesamiento o si se pueden retomar completos o por partes. Esto dependerá, en primera instancia, de si se encuentran no estructurados o estructurados, y a partir de ahí determinar sus posibilidades de uso.


Los **datos no estructurados** se refieren a información desorganizada, es decir, se trata de un conglomerado de información de la que no pueden extraerse atributos específicos de manera ágil y/o automatizada.

Estos datos pueden ser públicos o no, pero se caracterizan por la forma en que están almacenados, por ejemplo: en formato impreso para el caso de información que se obtiene de manera directa de personas beneficiarias de un servicio, o producto de interacciones oficiales entre servidores públicos; en imagen digital (PDF, JPG, PNG) para información digitalizada a partir de impresos o fotografías; en archivos de procesadores de texto o presentaciones (DOCX, PPTX) para reportes o informes, por ejemplo, de la situación económica trimestral; entre otros.

Si interactuamos con este tipo de formatos de manera cotidiana sabemos que no es viable obtener agregados de información de manera directa, dado que su procesamiento es más complicado, por lo que será necesario transformar estos datos a estructurados mediante su digitalización.

En contraste, los **datos estructurados** son aquellos que son legibles por máquina (vía software libre o de paga) ya que cuentan con un arreglo u ordenación fácilmente identificable. Por ejemplo, si la estructura de la información es de filas y columnas, los formatos de hoja de cálculo o despliegue lineal permiten su ágil procesamiento (XLSX, CSV, etc.); mientras que para estructuras de árbol, los formatos que permiten jerarquización de información son los más comunes (JSON o XML).

Si se identifica que los datos a publicar ya se encuentran en formato estructurado, puede agilizarse el proceso de apertura.



SHCP

1.2 Producción y empleo

1.2.1 Producción

Durante el segundo trimestre de 2017, el PIB registró un crecimiento anual de 1.8 por ciento real. Al excluir el efecto estacional, el PIB creció a una tasa de 3.0 por ciento anual, y de 0.6 por ciento en términos trimestrales. A su interior, se observaron los siguientes resultados:

- La producción agropecuaria se incrementó a una tasa anual de 0.7 por ciento. Cifras ajustadas por estacionalidad indican que este sector tuvo una reducción trimestral de 1.9 por ciento.
- La actividad industrial tuvo una disminución anual de 1.1 por ciento. Dicho resultado fue consecuencia, principalmente, de la menor producción de la minería petrolera y las obras de ingeniería civil. Por su parte, la producción de manufacturas registró un aumento anual de 2.0 por ciento, debido a la expansión de los subsectores de equipo de transporte, equipo de computación, alimentos y fabricación de maquinaria y equipo, entre otros. Al eliminar el efecto estacional, la producción industrial no presentó cambio alguno respecto al trimestre previo.
- El sector servicios registró un crecimiento anual de 3.2 por ciento como resultado del desempeño de las actividades de servicios financieros, comercio, información en medios masivos, inmobiliarias y de alquiler, y transportes, correos y almacenamiento, principalmente. Cifras ajustadas por estacionalidad indican que los servicios aumentaron a una tasa trimestral de 0.8 por ciento.

PRODUCTO INTERNO BRUTO, 2015-2017 * / (Variación % anual)

	Enero-junio			2015			2016			2017		
	2015	2016	2017	Q	III	IV	I	II	III	IV	I	II
Total	2.7	2.4	2.3	2.5	2.8	2.5	2.2	2.6	2.0	2.3	2.8	1.8
Agropecuaria	2.7	1.8	1.1	0.5	0.5	0.4	-0.4	1.9	5.5	5.1	6.4	0.7
Industrial	1.3	0.4	-0.3	0.7	1.2	0.1	0.0	0.8	-0.9	-0.1	0.5	-1.1
Minería	-5.0	-4.0	-10.1	-4.1	-4.3	-4.1	-3.2	-4.8	-8.2	-8.5	-11.6	-8.5
Construcción	2.5	3.3	-1.2	0.2	1.9	2.1	1.0	5.7	3.9	2.8	-0.5	-1.9
Comercio	3.9	2.2	-0.1	2.9	3.4	-0.7	1.3	3.0	0.0	3.0	1.5	1.5
Manufacturas	2.8	1.1	1.6	2.7	2.6	2.0	0.7	1.5	1.1	1.9	5.2	2.0
Servicios	3.3	3.4	1.5	3.4	3.7	3.8	3.4	3.9	3.4	3.4	3.7	3.2
Comercio	4.9	3.0	3.0	4.4	5.1	4.3	3.7	2.3	1.3	2.9	3.4	2.4
Transportes	4.0	3.1	1.8	3.9	5.3	3.9	2.9	3.3	2.3	2.7	4.1	3.6
Info. en medios masivos	3.1	9.1	4.2	2.5	8.6	16.1	8.9	9.2	13.4	8.8	5.3	7.0
Financieros y de seguros	2.6	7.8	9.7	3.1	5.9	6.1	8.2	7.5	7.9	7.2	9.5	9.9
Inmobiliarias y del alquiler	2.7	2.0	2.3	3.6	2.3	2.2	2.1	1.8	1.8	1.8	2.4	1.8
Renta	2.4	2.7	2.6	2.6	1.5	1.8	2.2	3.1	3.6	3.3	3.0	2.2

* Fuente: SHCP, datos MEXI.



	M	N	O	P	Q	R	S
ID_RECURSO	RECURSO	CLAVE_PROG	PROGRAMA	ID_RAMO	RAMO	INSTITUCION	
2	Aportaciones I013		FONE Servici		33	Aportaciones Secretaría de	
2	Aportaciones I012		FAFEF		33	Aportaciones SECRETARÍA	
3	Convenios K031		Proyectos de		9	Comunicación CARRETERAS	
2	Aportaciones I002		FASSA		33	Aportaciones SECRETARÍA	
1	Subsidios U005		Seguro Popul		12	Salud SERVICIOS D	
3	Convenios K040		Proyectos de		9	Comunicación Secretaría de	
1	Subsidios U006		Subsidios par		11	Educación Pó UNIVERSIDA	
1	Subsidios U006		Subsidios par		11	Educación Pó Universidad	
1	Subsidios U006		Subsidios par		11	Educación Pó UNIVERSIDA	
1	Subsidios U128		Proyectos de		23	Provisiones S Comisión Est.	
1	Subsidios U128		Proyectos de		23	Provisiones S Comisión Est.	
1	Subsidios U092		Fortalecimier		23	Provisiones S MUNICIPIO E	
3	Convenios U006		Subsidios par		11	Educación Pó UIED	
1	Subsidios U130		Fortalecimier		23	Provisiones S Secretaría de	
3	Convenios K031		Proyectos de		9	Comunicación CARRETERAS	
3	Convenios K041		Sistema de Tr		9	Comunicación Secretaría de	
1	Subsidios U005		Seguro Popul		12	Salud SECRETARÍA	
1	Subsidios U005		Seguro Popul		12	Salud SECRETARÍA	
1	Subsidios U022		Programas Ri		23	Provisiones S SECRETARÍA	
3	Convenios U006		Subsidios par		11	Educación Pó UNIVERSIDA	
2	Aportaciones I014		FONE Otros c		33	Aportaciones Secretaría de	
1	Subsidios U006		Subsidios par		11	Educación Pó UNIVERSIDA	
3	Convenios U012		Fortalecimier		12	Salud SECOPE	
1	Subsidios U057		Fondo Metro		23	Provisiones S Secretaría de	
1	Subsidios U087		Fondo de Caj		23	Provisiones S Secretaría de	
1	Subsidios U005		Seguro Popul		12	Salud INSTITUTO D	
1	Subsidios U006		Subsidios par		11	Educación Pó Universidad	
1	Subsidios U130		Fortalecimier		23	Provisiones S Tribunal Sup	
3	Convenios K031		Proyectos de		9	Comunicación CARRETERAS	

Ejemplo de datos estructurados vs no estructurados. Del lado izquierdo, datos del Producto Interno Bruto que son públicos pero de difícil procesamiento; del lado derecho, datos de las transferencias de recursos federales a gobiernos sub-nacionales en formato estructurado en filas y columnas (.XLSX) para ágil procesamiento.

3. Vinculación entre sistemas informáticos.

Cuando la información ya se encuentra estructurada en sistemas informáticos, es recomendable proceder a diagnosticar su calidad y conocer sus contenidos; diseñar una propuesta para la salida pública de la información (en caso de que el formato y estructura no sean usables) y realizar pruebas para finalmente publicarla.

En esta situación, es conveniente aspirar a los mayores niveles de apertura desde el comienzo del proceso de apertura para tener ventaja en su publicación. Para conocer más sobre los niveles de apertura de datos, **revisa el video y documento de consulta** “¿Qué son los datos abiertos?”.

Con estas consideraciones sobre la identificación de la fuente y el formato de los datos **el siguiente paso será su depuración y limpieza.**

Revisa el siguiente documento de consulta para conocer este proceso.



Referencias

Kyocera. Diferencia entre datos estructurados y no estructurados. Disponible en:
<https://smarterworkspaces.kyocera.es/blog/diferencia-datos-estructurados-no-estructurados/>

Escuela de Datos. Conoce el formato de tus datos. Disponible en:
<https://es.schoolofdata.org/2013/10/31/conoce-el-formato-de-tus-datos/>

