



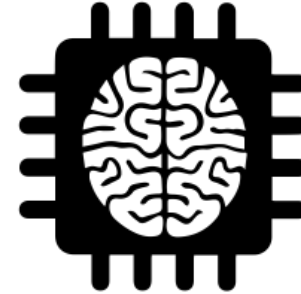
# COLECTA DE DATOS

STEVEN SOTELO



# Introducción

# Minería de datos



Colectar datos

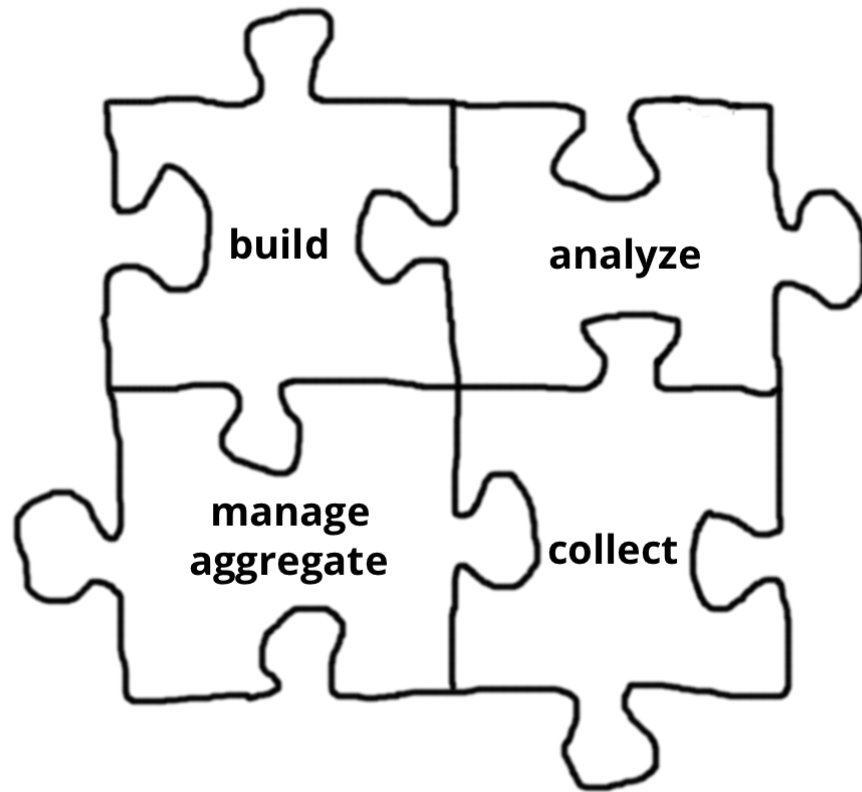
Almacenar  
datos

Limpiar datos

Analizar datos

Mostrar  
resultados

# OPENROSA



- OpenRosa es el nombre de un ecosistema de herramientas para el análisis de datos, las cuales son diseñadas bajo estándares, protocolos y código abierto, para facilitar la comunicación entre sí. Algunas herramientas conocidas dentro de este ecosistema son Open Data Kit (ODK), KoBoToolbox y Ona.



# ODK


Open Data Kit

# Qué es ODK?

Open Data Kit es un conjunto de herramientas de código abierto para la colecta y administración de datos



# Herramientas de ODK

- **ODK Collect:** Aplicación móvil para la colecta de datos
  - **ODK Aggregate:** Servidor web para el almacenamiento y análisis de datos
  - **ODK Central:** Servidor con RESTful API
  - **ODK Build:** Aplicación de escritorio para el diseño de formularios
  - **ODK XLSForm:** Diseño de formularios a través de excel
  - **ODK Briefcase:** Aplicación de escritorio para importar y exportar datos
- 



Herramientas para crear  
formularios



# ODK Build

ODK Build es una aplicación para el diseño de formularios que sigue los estándares de Open Rosa. Los formularios debe ser exportados a XML para poder ser interpretados por ODK

- Ventajas:
  - Herramienta drag and drop
  - Integración con otras herramientas
  - Exportación para diferentes formatos

Cuenta con dos versiones:

- Web (online): <https://build.opendatakit.org/>
- Escritorio (offline):  
<https://github.com/opendatakit/build/releases>






[no caption text yet]  
untitled1

✕



[no caption text yet]  
untitled2

✕




[no caption text yet]  
untitled3

✕



[no caption text yet]  
untitled4

✕




Device ID  
untitled5

✕



[no caption text yet]  
untitled6

✕



[no caption text yet]  
untitled7

✕

## Properties

Data Name

Label

Hint

Default Value

☐ Read Only☐ Required

Advanced &lt;

## Information: Time

Collects just a time of day. If you wish to collect a time on a certain date, use Date/Time instead.

An appropriate time widget will be presented on mobile devices.

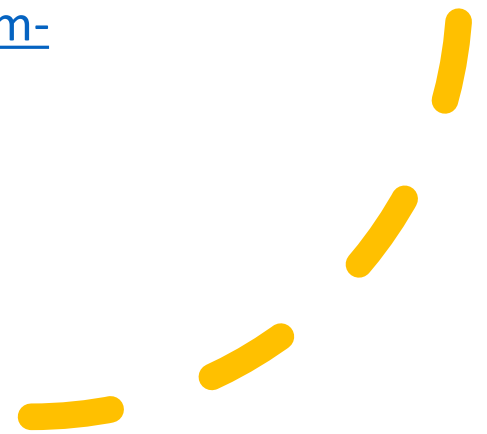
# XLSForm

XLSForm es un formulario estándar creado para simplificar la construcción de formularios en Excel. Los formularios debe ser exportados a XML para poder ser interpretados por ODK

- Ventajas:
  - Diseño a través de Microsoft Excel
  - Programación del formulario

Para exportar a XLM se requiere usar la herramienta XLSForm. Cuenta con dos versiones:

- Web (online): <https://opendatakit.org/xlsform/>
- Escritorio (offline):  
<https://github.com/opendatakit/xlsform-offline/releases/tag/v1.7.0>



# XLSForm – Estructura de archivo

Un XLSForm tiene la siguiente estructura:

- survey: Hoja que contiene la definición de preguntas del formulario
- settings: Contiene los parámetros globales del formulario
- choices: Esta hoja tiene los listados de opciones para las preguntas de selección



aeaps\_production\_event.xlsx - Excel

Sotelo, Humberto (CIAT)

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW TEAM

Clipboard Font Alignment Number Styles Cells Editing

A1 type

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	type	name	label::Español (es)	hint::Español (es)	required	default	constraint	constraint	relevant	calculator	choice_filt	appearance	
2	begin group	technical_assistant	Asistente técnico										
3	select_one choices_assistant_partnership	assistant_partnership	Asociación		yes								
4	select_one choices_kind_document	assistant_kind_document	Tipo de documento		yes	1							
5	text	assistant_document_number	Número de identificación		yes		(if(\$assist	El número de identificación debe ser numérico y no tener espacios					
6	calculate	assistant_found								(if(pulldata('people','name', 'document_key', :			
7	begin group	assistant_found_group	Asistente técnico en base de datos							(\$assistant_found)=true')			
8	text	assistant_found_name	Nombre		yes		(regex(., '[	El nombre debe conti		pulldata('people','name', 'document_key', \$a			
9	text	assistant_found_last_name	Apellidos		yes		(regex(., '[	El apellido debe conti		pulldata('people','last_name', 'document_key'			
10	text	assistant_found_cellphone	Celular		yes		(regex(., '[	El número de celular		pulldata('people','cellphone', 'document_key'			
11	select_one choices_sex	assistant_found_sex	Sexo		yes					pulldata('people','sex', 'document_key', \$assi			
12	select_one choices_country	assistant_found_country	País		yes					pulldata('people','country', 'document_key', \$			
13	select_one choices_adm1	assistant_found_adm1	Departamento o región		yes					country = \${assistant_found_count			
14	select_one choices_adm2	assistant_found_adm2	Distrito		yes					country = \${assistant_found_count			
15	text	assistant_found_address	Dirección		yes					pulldata('people','address', 'document_key', \$			
16	text	assistant_found_email	Email		yes		(regex(., '[	El email se encuentra		pulldata('people','email', 'document_key', \$a			
17	end group												
18	begin group	assistant_group	Asistente técnico							(\$assistant_found)=false')			
19	text	assistant_name	Nombre		yes		(regex(., '[	El nombre debe contener solamente texto y no tener espacios en bl					
20	text	assistant_last_name	Apellidos		yes		(regex(., '[	El apellido debe contener solamente texto y no tener espacios en bl					
21	text	assistant_cellphone	Celular		yes		(regex(., '[	El número de celular se encuentra mal diligenciado. Ejemplo: 555 5					
22	select_one choices_sex	assistant_sex	Sexo		yes								
23	select_one choices_country	assistant_country	País		yes	pe							
24	select_one choices_adm1	assistant_adm1	Departamento o región		yes	1					country = \${assistant_country}		
25	select_one choices_adm2	assistant_adm2	Distrito		yes	1					country = \${assistant_country} and		
26	text	assistant_address	Dirección		yes								
27	text	assistant_email	Email		yes		(regex(., '[	El email se encuentra mal diligenciado					
28	end group												
29	end group												
30													
31	begin group	farmer	Agricultor										
32	select_one choices_kind_document	farmer_kind_document	Tipo de documento		yes								

survey choices warnings settings

READY 85%

	A	B	C	D	E	F
1	form_title	form_id	public_key	submission_url	instance_name	version
2	AEPS - Evento productivo	aeps_production_event			concat(concat(concat('aeps-',\${farmer_document_number}),'-'),\${plot_name})	20180829

# XLSForm – Estructura de archivo (settings)

- form\_title: Título de formulario (nombre que se verá en los listados)
- form\_id: Identificador del formulario
- public\_key: Llave criptográfica para cifrar formularios
- submission\_url: Url para sincronizar datos
- instance\_name: Nombre que debe llevar cada formulario diligenciado
- version: Version del formulario

# Formularios XML

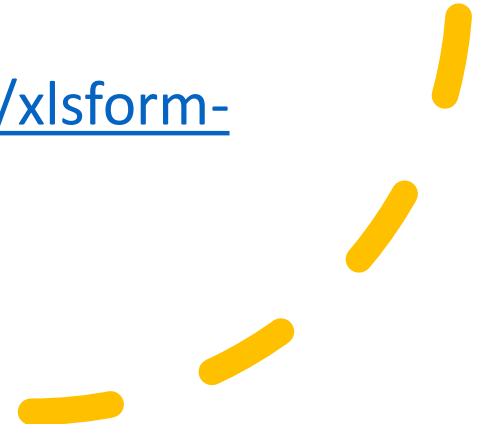
Los formularios que son representados por las aplicaciones encargadas de coleccionar los datos requieren que los formularios estén en formato **XML**



## Transformación de archivos

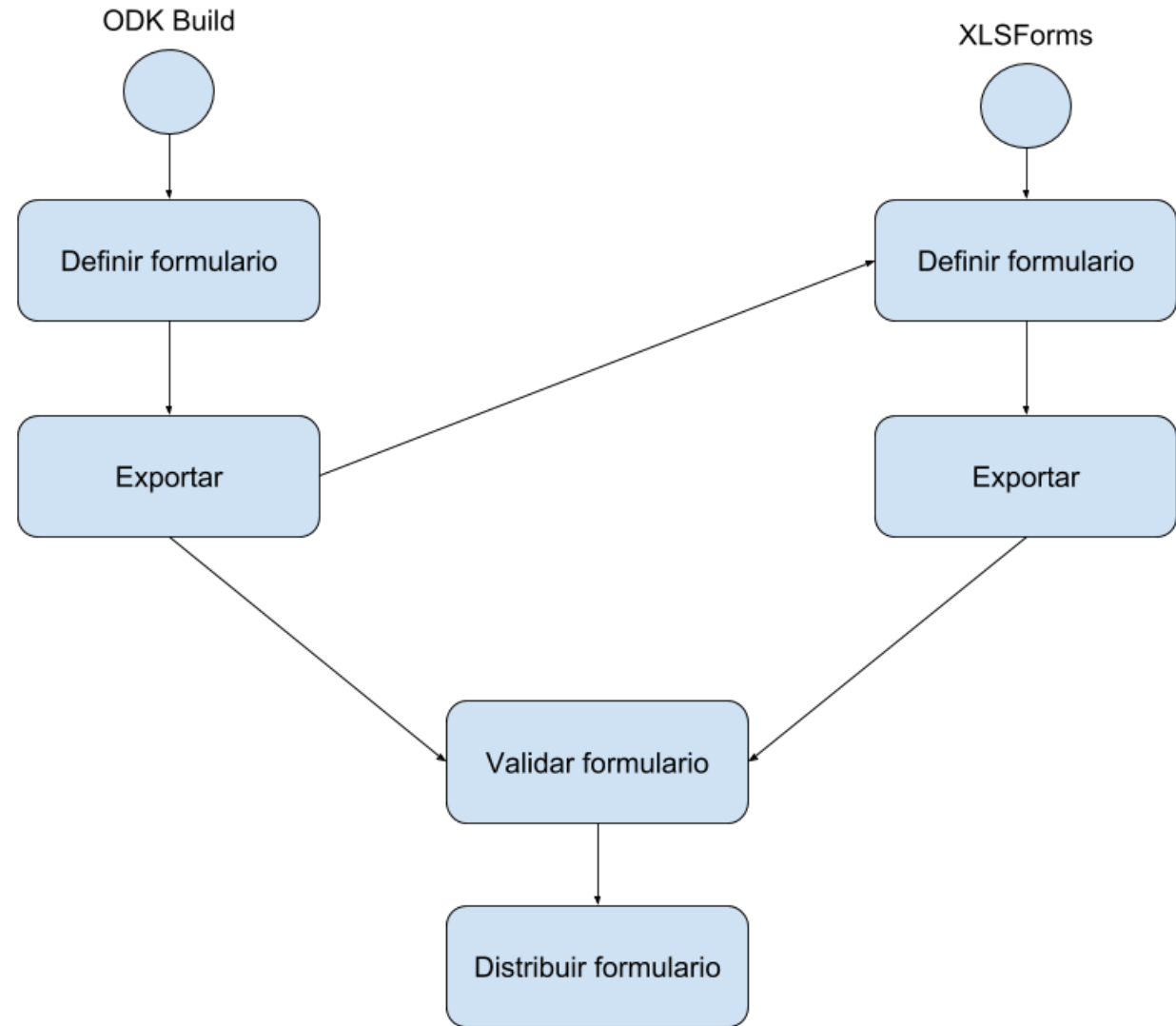
Dependiendo de la herramienta utilizada para definir los formularios puede transformar estos con otros programas.

- ODK Build permite exportar directamente a XML
- Los formularios creados con XLSForms pueden transformarse desde:
  - Online: <https://opendatakit.org/xlsform/>
  - Offline: <https://github.com/opendatakit/xlsform-offline/releases>





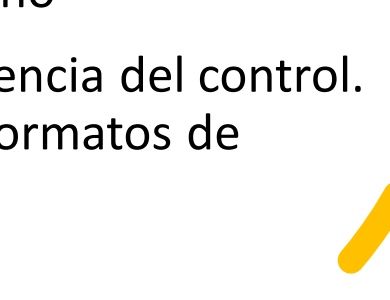
# Ciclo de creación de formularios





Preguntas

# Propiedades que posee una pregunta en ODK

- **Type:** Indica cual es el tipo del campo. Son definidos por un estándar
  - **Name:** Nombre del campo. Este es un nombre en format máquina (alfabeto inglés, en minúsculas y sin símbolos especiales)
  - **Label:** Descripción de la pregunta. Es un texto que le sirva a la persona que diligencia el formulario a comprender el texto. Se puede tener por cada idioma
  - **Hint:** Texto de ayuda o indicaciones de como diligenciar el campo. Se puede tener por cada idioma
  - **Default:** Valor por defecto del campo
  - **Require:** Indica si el campo es obligatorio o no
  - **Appearance:** Establece como será la apariencia del control. Combinado con type, se puede establecer formatos de colecta
  - **Constraint:** Reglas de validación del campo
- 

# Tipos de preguntas

Texto	Númericas	Fecha y hora
Selección	Ranking	Ubicación
Media	Archivos	Código de barras
Rangos	Notas	Campos ocultos
Campos calculados		

# Texto, numéricos, fecha y tiempo

Son tipos de preguntas básicos en ODK.

- Texto: Es el tipo de pregunta básica y permite ingresar caracteres alfanuméricos. Todos los controles heredan de este objeto.
  - type: *text*
- Numéricos: Este tipo de pregunta permite el ingreso únicamente de valores numéricos. Los tipos disponibles son: Enteros y Decimales
  - type: [integer, decimal]
- Fecha y tiempo: Este tipo de pregunta permite el ingreso de fechas, horas y una combinación entre ambos. Los tipos disponibles son:
  - type: [date, time, dateTime]
  - appearance: month-year, year

# Selección

Este tipo de pregunta permite a los usuarios seleccionar valores una lista predeterminada de valores. Este tipo de campo ofrece las opciones de selección única o varios valores

- Selección única
  - `type: select_one {list_name}`
  - `appearance: minimal, autoadvance, autocomplete`
- Selección múltiple
  - `type: select_multiple {list_name}`
  - `appearance: none`

Los campos de tipo selección requieren de una lista de valores predeterminados. Estas listas pueden ser reutilizadas en diferentes campos del mismo formulario. Cada elemento de la lista tiene las siguientes propiedades:

- `list_name`: Nombre de la lista a la que pertenece el elemento
- `name`: Código o nombre máquina del elemento (no visible)
- `label`: Descripción o etiqueta que el usuario observa en el formulario (visible)

# Selección (Listas)

	A	B	C	D
1	list name	name	label::Español (es)	country
120	choices_country	co	Colombia	
121	choices_country	pe	Perú	
122	choices_adm1	1	Amazonas	co
123	choices_adm1	2	Antioquia	co
124	choices_adm1	3	Arauca	co
125	choices_adm1	4	Atlántico	co
126	choices_adm1	5	Bolívar	co
127	choices_adm1	6	Boyacá	co
128	choices_adm1	7	Caldas	co

Los campos de tipo selección pueden establecer jerarquías entre sí. Estas establecen que valores van a ser visualizados a medida que se avanza cuando se diligencia el formulario. Para establecer jerarquías se requiere que en la definición del formulario indicar en el atributo “choice\_filter” cual será el campo y el valor a filtrar. Por consiguiente la sección de elementos debe estar igualmente diligenciado.

# Ubicación

Este tipo de pregunta permite a los usuarios seleccionar obtener valores sobre ubicación geográfica. Esto puede ser la ubicación actual, una ubicación seleccionada en un mapa o una ruta.

- type: [geopoint, geoshape]
- appearance: [map, placement-map]

Si se requiere se puede realizar una traza de un recorrido mediante los campos de ubicación. Este permite que se realice manual o en un intervalo de tiempo predeterminado.

- type: geotrace

## AEPS - Ejemplo

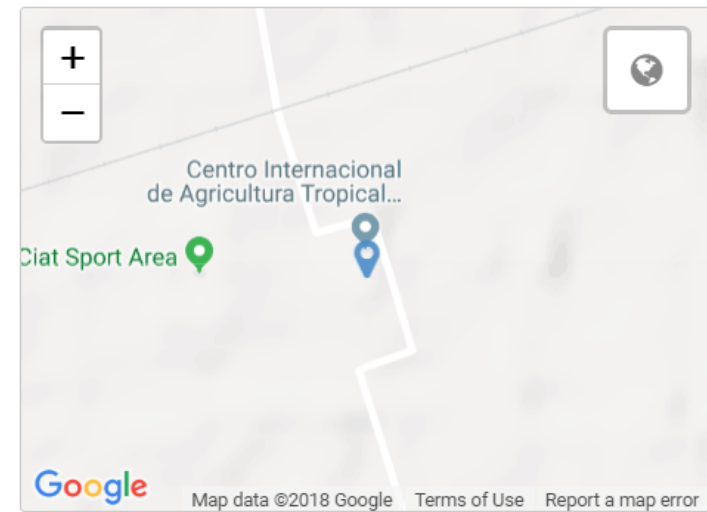
### Ubicación actual

latitude (x.y °)

longitude (x.y °)

altitude (m)

accuracy (m)





# Archivos y media

Este tipo de pregunta permite a los usuarios seleccionar coleccionar información a través de archivos media, tales como imágenes, videos y audios.

- type: [image, video, audio, file]
- appearance: [none, annotate, new, new-front, draw]

# Campos ocultos

Este tipo de campos obtienen información indirecta de los usuarios que diligencian los formularios. Estos campos no son visibles en la captura de información, pero son almacenados dentro de cada encuesta. La información que se puede obtener mediante estos campos son:

- Fecha de inicio de la encuesta (start)
- Fecha de finalización de la encuesta (end)
- Fecha actual (today)
- Identificador del dispositivo (deviceid)
- Identificador del usuario (subscriberid)
- Código de la tarjeta sim (simserial)
- Nombre del usuario (username)
- Número de teléfono (phonenumber)

# Campos calculados

Este tipo de campo permiten evaluar complejas expresiones, para almacenar los valores y usarlos después. Para evaluar las expresiones, se requiere que en la definición del formulario se adicione un atributo adicional llamado “calculation”, ya que será este atributo donde se evalúan las expresiones

- type: calculate

# Expresiones

Las expresiones, a veces llamadas también fórmulas, son evaluadas dinámicamente a medida que un formulario está siendo diligenciado. Una expresión puede contener variables, operadores, funciones y las combinaciones entre ellas. Estas son usadas en diferentes contextos de un formulario, que incluyen validación de campos, cálculos de variables, reglas de negocio, condiciones para despliegue de información, etc. Veamos como son las expresiones:

- Expresiones aritméticas:
  - **`${factura} * 0.16`**
- Expresiones lógicas:
  - **`${edad} >= 18`**
- Expresiones de funciones:
  - **`concat(${nombre}, ' ', ${apellidos})`**



# Operadores

Tipo	Operador	Explicación	Ejemplo
Aritméticos	+	Suma	$\${total} + \${propina}$
Aritméticos	-	Resta	$\${precio} - \${descuento}$
Aritméticos	*	Multiplicación	$\${factura} * \${impuesto}$
Aritméticos	div	División	$\${monto} \text{ div } \${cantidad}$
Aritméticos	mod	Módulo	$\${numero} \text{ mod } 2$
Lógico	=	Igual a (sirve para números y cadenas)	$\${afiliado} = \text{'si'}$
Lógico	!=	Diferente (sirve para números y cadenas)	$\${afiliado} != \text{'si'}$
Lógico	>	Mayor	$\${edad} > 17$
Lógico	>=	Mayor o igual	$\${edad} >= 18$
Lógico	<	Menor	$\${edad} < 64$
Lógico	<=	Menor o igual	$\${edad} <= 63$
Booleano	and	Retorna verdadero si ambas premisas son verdaderas	
Booleano	or	Retorna verdadera si alguna de las premisas es verdadera	

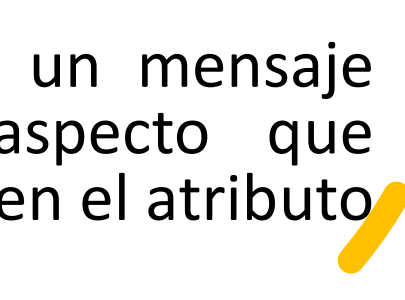
# Validaciones

ODK permite realizar validaciones al contenido de los formularios diligenciados. Esto nos permite asegurar la calidad de los datos colectados. Algunas de las validaciones que podemos realizar son:

- Campos obligatorios
- Campos dependientes
- Validaciones numéricas
- Validación de expresiones regulares
- Validación de fechas
- Entre otras



# Validaciones

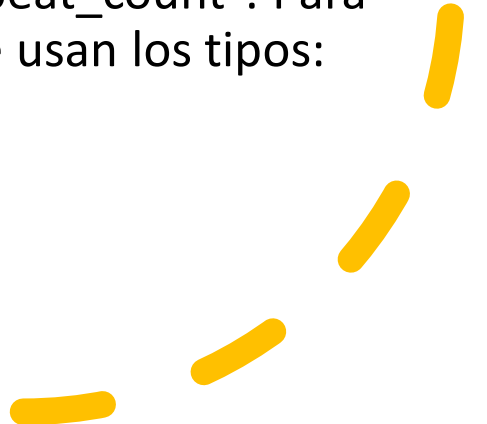
- La validación de campos permite evaluar el contenido de una respuesta luego de que el campo fue diligenciado. Las validaciones de campos deben ser lógicas, ya que para que un campo con validación sea aceptado, al evaluar la expresión debe ser verdadera.
  - Para validar el contenido de un campo se debe crear una expresión que evalúa el contenido. Esta se crea en el atributo “constraint”.
  - Adicionalmente puede establecer un mensaje personalizado para indicar el aspecto que generó el error, el mensaje debe ir en el atributo “constraint\_message”
- 

# Grupos y repeticiones

Los grupos permiten enmarcar un conjunto de preguntas. Las agrupaciones dentro de sus muchas ventajas permiten realizar validaciones a un conjunto de variables o mostrar todos los campos en una sección. El usuario cuando esté diligenciando el grupo podrá adicionar cuantos quiera del mismo conjunto. Para definir un grupo se usan los tipos:

- `begin_group` (indica donde inicia)
- `end_group` (indica donde termina).

Un usuario puede repetir un grupo de preguntas. Se puede establecer una cantidad fija de repeticiones o dejar abierta la posibilidad. Para definir la cantidad de repeticiones se debe establecer un número entero en el atributo “`repeat_count`”. Para definir un grupo de preguntas que se repiten, se usan los tipos:

- `begin_repeat` (indica donde inicia),
  - `end_repeat` (indica donde termina)
- 



# Precargar datos en formularios

Los datos de precarga se realizan cuando uno quiere referirse a datos preexistentes en un formulario de encuesta. Puede hacer referencia a los datos en su formulario de encuesta (de la encuesta que ahora está creando), a partir de datos preexistentes en un formulario de encuesta específico o cualquier otra fuente. Por ejemplo, si tiene datos preexistentes de una encuesta de hogares y desea recopilar datos de seguimiento sobre los ocupantes del hogar. Puede confiar en que se refiera a los datos de la encuesta de hogares en su formulario de encuesta. Para hacer referencia a datos preexistentes en un formulario de encuesta:

Cargue uno o más archivos .csv como archivos de soporte cuando cargue la definición de su formulario (de la misma manera que carga los archivos de soporte de medios como se explica en la sección Medios). La primera fila de cada archivo .csv debe ser un encabezado que incluya:

- nombres únicos para cada columna
- filas posteriores que deben contener los datos en sí

Tomado de: <http://xlsform.org/en/#pre-loading-csv-data>



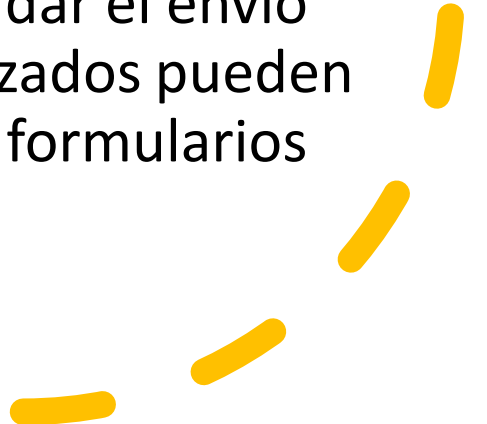


Colecta de datos

# ODK Collect

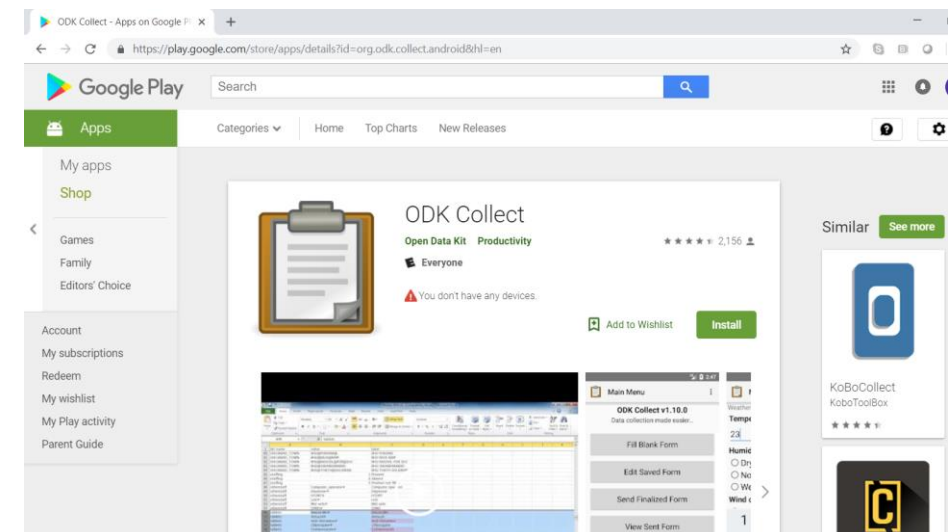
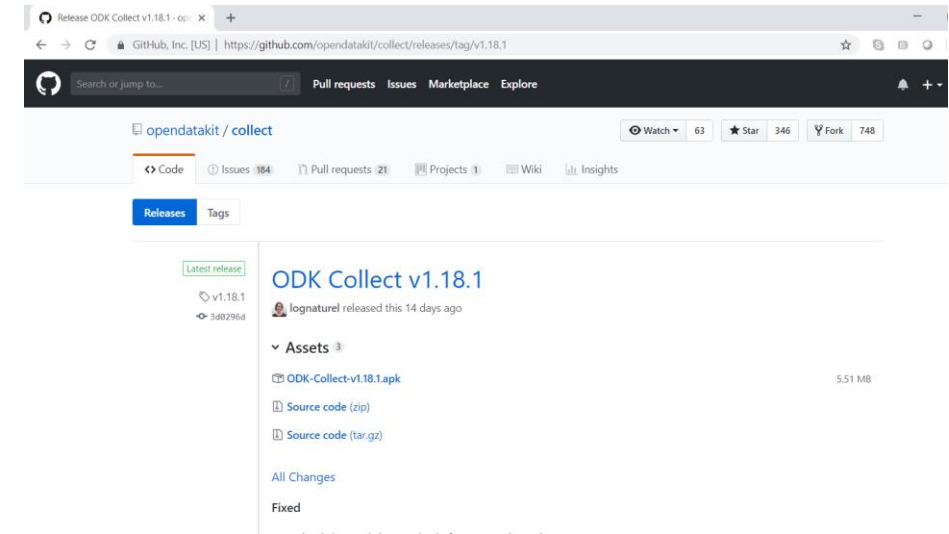
ODK Collect es una aplicación de código abierto para Android que reemplaza los formularios en papel utilizados en la recopilación de datos basados en encuestas. Admite una amplia gama de tipos de preguntas y respuestas, y está diseñado para funcionar bien sin conectividad de red.

ODK Collect presenta los formularios en una secuencia de solicitudes de entrada que aplican la lógica de formularios, las restricciones de entrada y la repetición de subestructuras. Los usuarios trabajan a través de las indicaciones y pueden guardar el envío en cualquier momento. Los envíos finalizados pueden enviarse (y se pueden descargar nuevos formularios desde) un servidor.



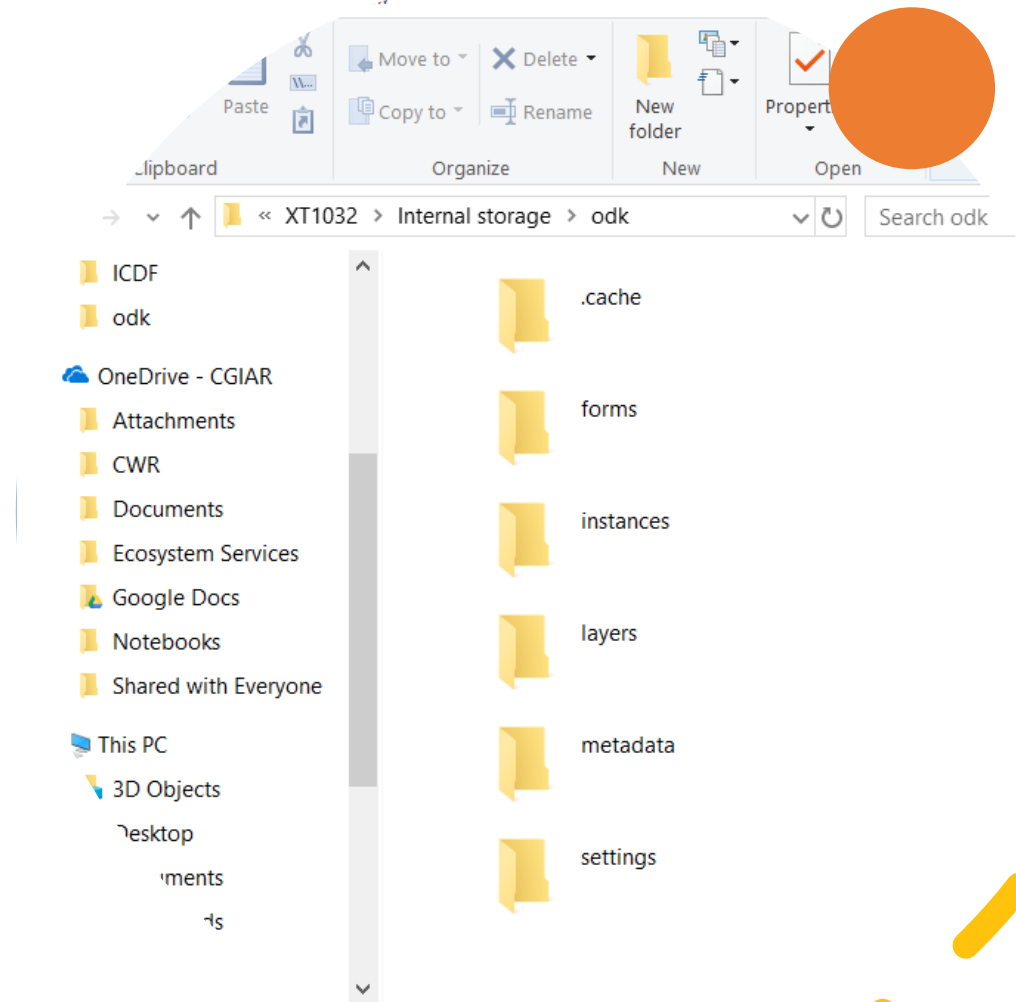
# ODK Collect

- Manual
  - Descarga desde github:  
<https://github.com/opendatakit/collect/releases/latest>
  - Instrucciones:  
<https://docs.opendatakit.org/collect-install/>
- Automática
  - Play store:  
<https://play.google.com/store/apps/details?id=org.odk.collect.android&hl=en>



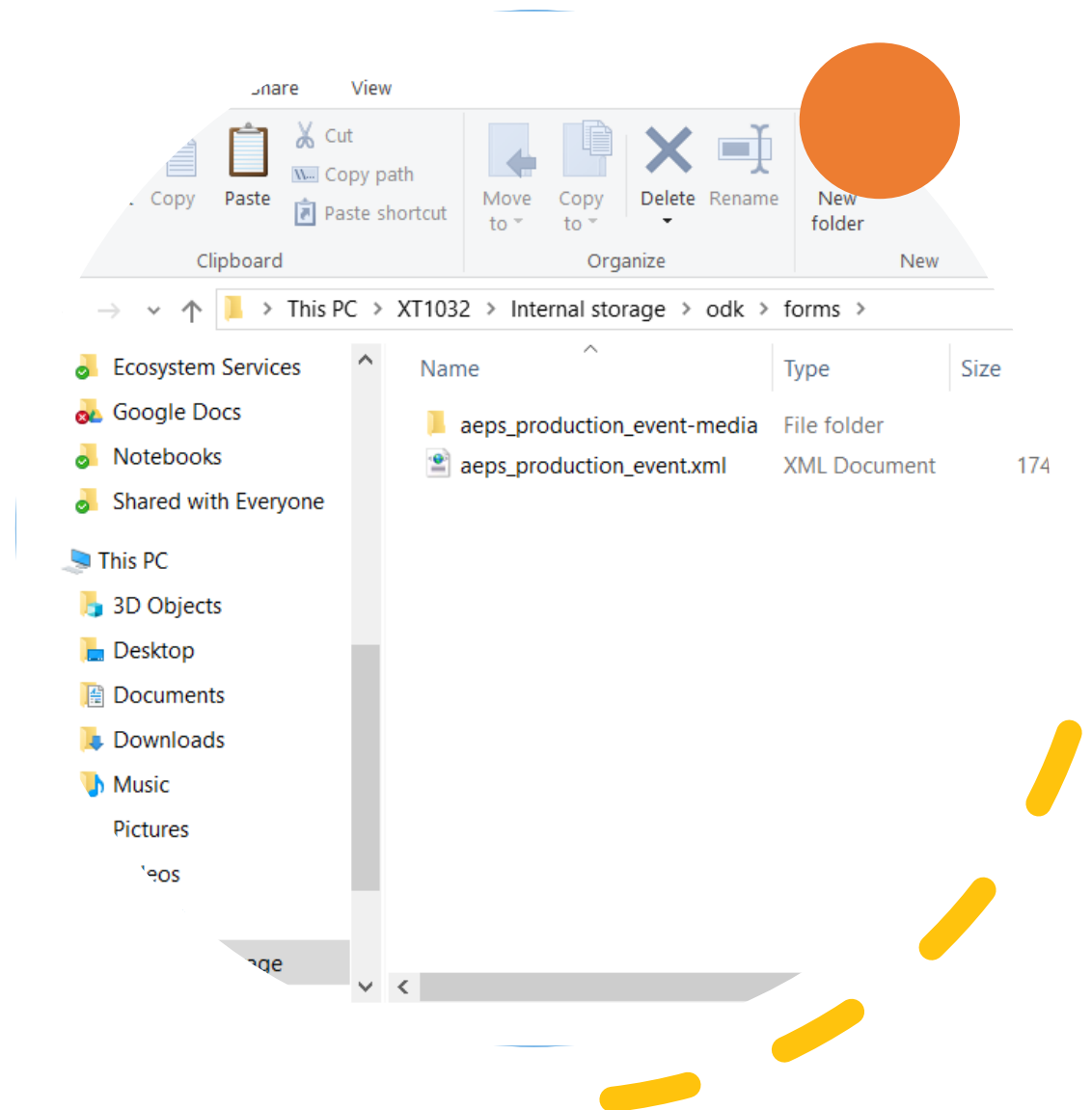
# ODK Collect - Estructura de carpetas

- .cache: Folder para archivos temporales
- forms: Esta carpeta almacena todos los formularios y archivos asociados a estos
- instances: Almacena las respuestas de los formularios
- layers: Este folder tiene las capas geográficas para usarlas en modo offline
- metadata: Este folder tiene metadatos de los demás
- settings: Se almacena un archivo con las configuraciones globales y un archivo con la imagen de un código QR que contiene las mismas configuraciones



# ODK Collect - Carpeta forms

- Este folder contiene todo los formularios disponibles para ser diligenciados.
- Cada formulario puede o no contener archivos multimedia
- Los folder multimedia deben tener el mismo nombre del formulario con un extension “-media”





Sincronización

# Sincronización de datos

ODK cuenta con varios sistemas para sincronizar los datos entre los dispositivos móviles y un servidor:

- Google Drive: [https://github.com/CIAT-DAPA/aeps\\_platform\\_forms](https://github.com/CIAT-DAPA/aeps_platform_forms)
- ODK Aggregate: <https://docs.opendatakit.org/aggregate-intro/>
- ODK Central: <https://docs.opendatakit.org/central-intro/>





# Información complementaria

- <https://docs.opendatakit.org/>
- <http://xlsform.org/en/>
- <https://opendatakit.org/software/>
- [https://github.com/CIAT-DAPA/aeps\\_platform\\_forms](https://github.com/CIAT-DAPA/aeps_platform_forms)



Gracias por su atención