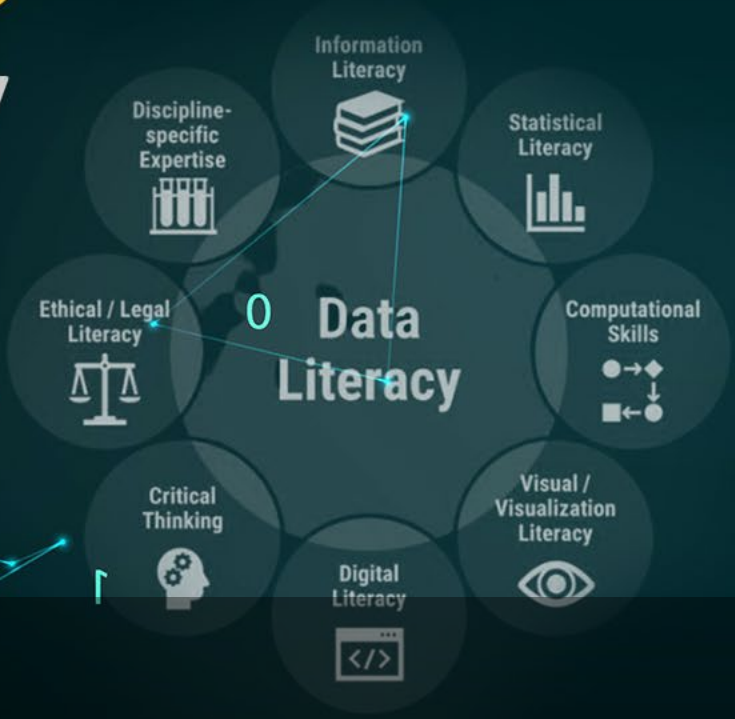
Alfabetización de Datos

Data Literacy

Organizado por la Dirección de Investigación – DI UTP



Mgter. Danny Murillo

Profesor / Investigador

Universidad Tecnológica de Panamá - CIDITIC

Objetivo

Desarrollar habilidades en la correcta utilización de los datos, abarcando fases como la normalización, manipulación y depuración de estos. Este aprendizaje permitirá realizar un adecuado análisis exploratorio de datos (AED), y posteriormente, transmitir de manera eficaz los resultados a través de visualizaciones comprensibles y dinámicas.

Contenido del curso

Tema 1: Introducción a los datos

Tema 2: Documentación de los datos

Tema 3: Manipulación y análisis de datos con Excel

Tema 4: Limpieza de datos con OpenRefine

Tema 5: Análisis Exploratorio de datos (AED)

Tema 6: Fundamentos de Visualización de datos

Tema 7: Gráficos en Datawrapper

Tema 8: Gráficos en Tableau

Metodología

En este curso virtual **sincrónico teórico-práctico,** exploraremos a fondo el tema a través de exposiciones dinámicas que se llevarán a cabo mediante recursos presentaciones en el entorno digital.

Haremos uso de **diversas herramientas** realizando prácticas en vivo, donde los participantes podrán aplicar directamente los conceptos teóricos adquiridos. Se fomentará la participación activa a través de sesiones de preguntas y respuestas, proporcionando a los participantes la oportunidad de aclarar dudas y profundizar en aspectos específicos del contenido.

La flexibilidad de este curso permitirá a los estudiantes avanzar a su propio ritmo, accediendo a los **recursos y actividades** proporcionados por el profesor, para optimizar su experiencia de aprendizaje.

¿Qué son los datos?

1. **Información** perteneciente a un hecho, utilizada como base para el razonamiento, la discusión o el cálculo.
2. Información en **digital** que puede ser trasmitida o procesada.
3. Información de entrada o salida de un proceso con información **relevante y/o redundante**.

¿Qué son los grandes datos?

Big data, se refiere a un gran volumen de datos que pueden extraer información. Utilizando grandes recursos computacionales pueden generar proyectos de aprendizaje y otras aplicaciones de análisis.

Características:

- Alto volumen: tamaño.
- Alta velocidad: ritmo rápido y continuo.
- **Alta variedad:** diferentes formatos, fuentes heterogéneas.

Fuentes de datos

Interno

- Recursos de Información
- RRHH
- Finanzas
- Cadena de suministros
- Registros de servicios

Externos

- API público
- Repositorios
- Fuentes de datos abiertos
- Empresas trasnacionales

¿Por qué es importante basarse en los datos?

- **Identificar Tendencias**

  informar prácticas eficaces, posibles soluciones.
- **Reducir el sesgo:**

  basarse en datos es más confiable que basarse en una percepción.
- **Desempeño de referencia:**

  Benchmarking.

Tipos de datos

Estructurado

y1	x1	x2	x3

Semiestructurado

```
<Books>
    <Book ISBN="0553212419">
        <title>Sherlock Holmes: Complete Novels...
        <author>Sir Arthur Conan Doyle</author>
    </Book>
    <Book ISBN="0743273567">
        <title>The Great Gatsby</title>
        <author>F. Scott Fitzgerald</author>
    </Book>
    <Book ISBN="0684826976">
        <title>Undaunted Courage</title>
        <author>Stephen E. Ambrose</author>
    </Book>
    <Book ISBN="0743203178">
        <title>Nothing Like It In the World</title>
        <author>Stephen E. Ambrose</author>
    </Book>
</Books>
```

Cuasi estructurado

17 de septiembre 02:33:08.536 [depuración] conexión_edge_process_relay_cell(): Ahora se ven 1802 celdas de retransmisión aquí (comando 2, flujo 5845).
17 de septiembre 02:33:08.536 [depuración] conexión_edge_process_relay_celda(): circ delivery_window ahora 933.

No estructurado

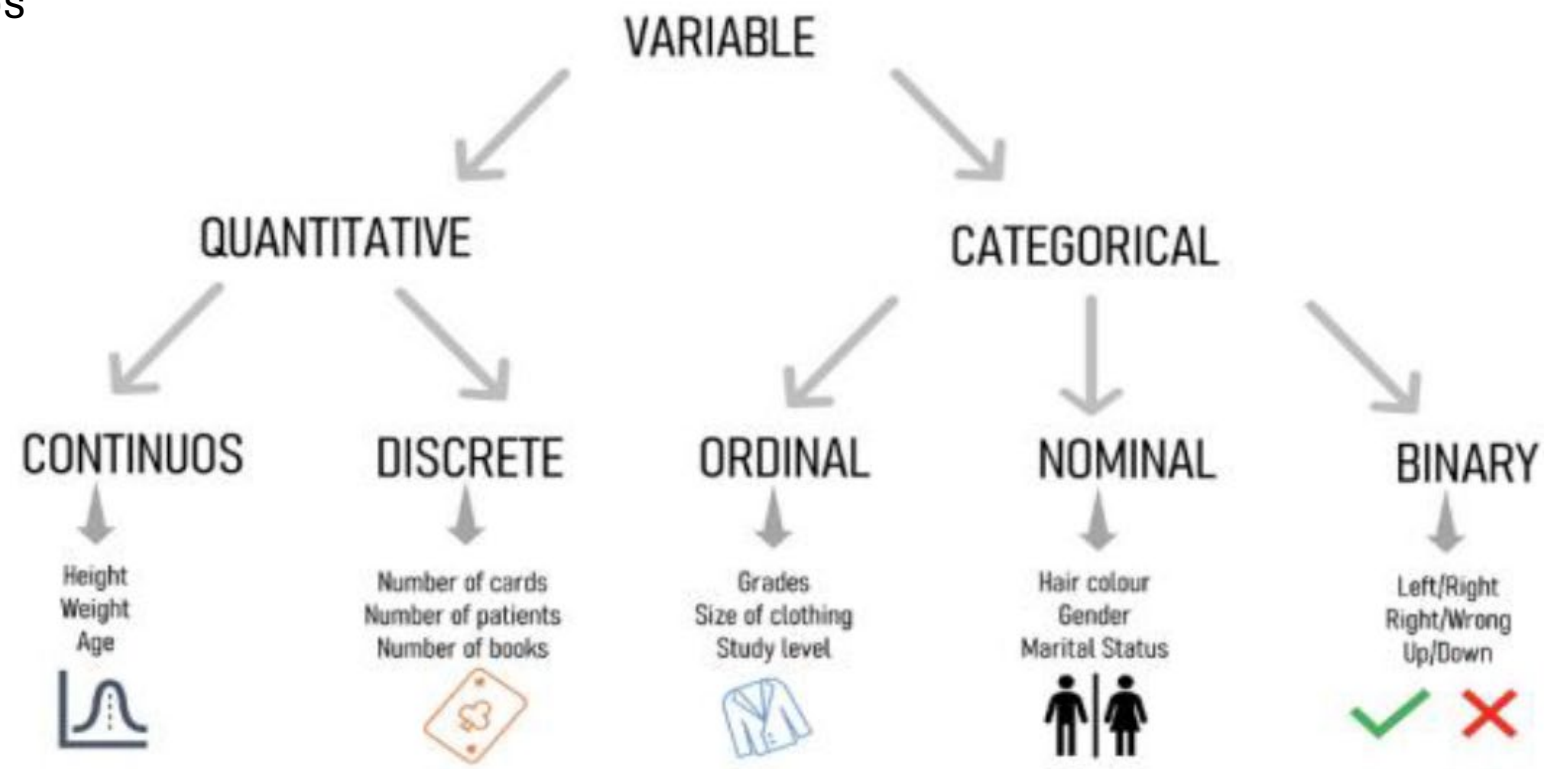
Tipos de datos

- Brutos
- Procesados
- Limpios
- Estadísticos
- Datos finales

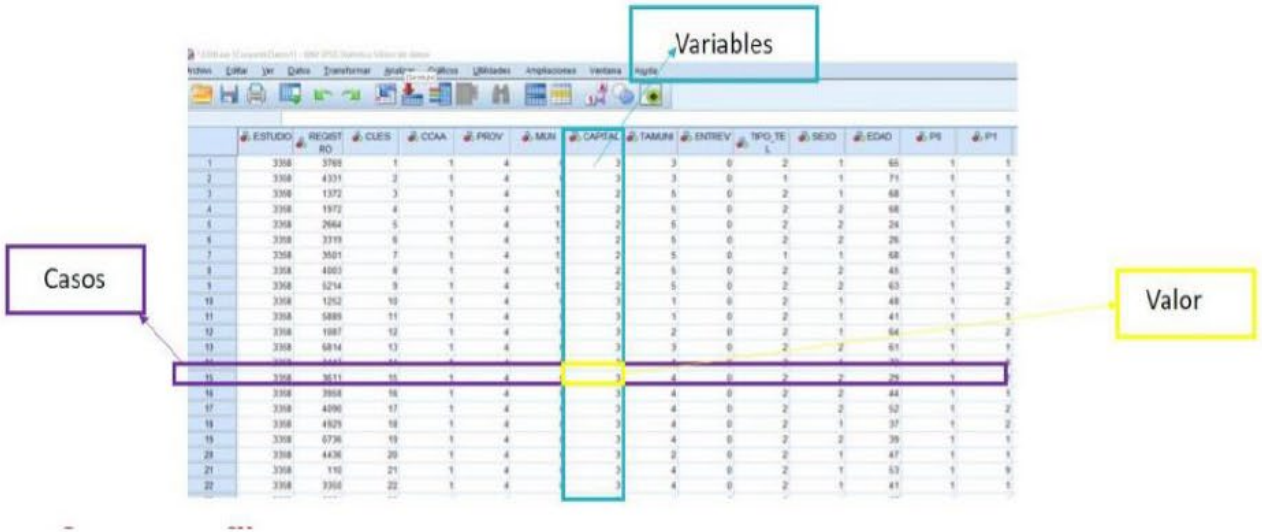
Tipos de datos / Tipos de Variables

Tipos de datos

- Cuantitativos
- Cualitativos



Estructura de datos | Datos tabulares



Caso	Sexo	Sentimiento
1	F	No me importa
2	M	Engañado
3	F	Me da risa

Estructura de datos | Datos tabulares


```
{
 "marcadores": [
  {
   "latitude": 40.416875,
   "longitude": -3.703308,
   "city": "Madrid",
   "description": "Puerta del Sol"
  },
  {
   "latitude": 40.417438,
   "longitude": -3.693363,
   "city": "Madrid",
   "description": "Paseo del Prado"
  },
  {
   "latitude": 40.407015,
   "longitude": -3.691163,
   "city": "Madrid",
   "description": "Estación de Atocha"
  }
 ]
}
```


marcadores			
latitude	**longitude**	**city**	**description**
40.416875	-3.703308	Madrid	Puerta del Sol
40.417438	-3.693363	Madrid	Paseo del Prado
40.407015	-3.691163	Madrid	Estación de Atocha

Estructura de datos | Formato CSV

Archivo en formato CSV :

```
marca,año,cilindros,consumo,potencia,aceleración
“chevrolet chevelle malibu”,70,8,18,130,12
“buick skylark 320”,70,8,15,165,11.5
“plymouth satellite”,70,8,18,150,11
“amc rebel sst”,70,8,16,150,12
“ford Torino”,70,8,17,140,10.5
```

- Tabla de datos:

marca	año	cilindros	consumo	potencia	aceleracion
chevrolet chevelle malibu	1970	8	18	130	12
buick skylark 320	1970	8	15	165	11,5
plymouth satellite	1970	8	18	150	11
amc rebel sst	1970	8	16	150	12
ford torino	1970	8	17	140	10,5

Estructura de datos | elección de Formatos

- Seleccionar **formatos abiertos**, no propietarios.
- Elegir **formatos comunes al campo disciplinar** al que se está trabajando: Para asegurar la interoperabilidad y la reutilización de los datos.
- Tener en cuenta el **tiempo en que se espera conservar los datos**
- Seleccione **formatos sin cifrar y sin compilar**

Yusnelkis Milanés Guisado Universidad Pablo de Olavide, Biblioteca/CRAi. España

Fuentes de datos

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Recolecta
RECOLECTOR DE CIENCIA ABIERTA
BUSCADOR DE CIENCIA ABIERTA

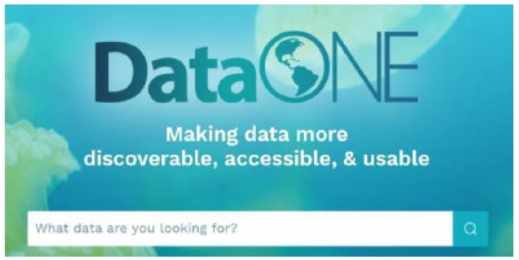
FAIRsharing.org
standards, databases, policies

DataCite
FIND, ACCESS, AND REUSE DATA

zenodo

UK • DATA ARCHIVE

Mendeley Data



LA Referencia
Red de repositorios de acceso abierto a la ciencia

dataMED
BETA version

¿Componentes de la cultura basada en datos?

La cultura basada en datos se puede separar en dos componentes:

Infraestructura de datos

- Acceso a los datos
- Almacenamiento de datos
- Recopilación de datos
- Procesamiento de datos

Alfabetización de datos

- Liderazgo de datos (necesidad de uso)
- Gobernanza de datos (estándares)
- Conocimiento de los datos
- Limpieza de datos
- Análisis de datos
- Visualización de datos

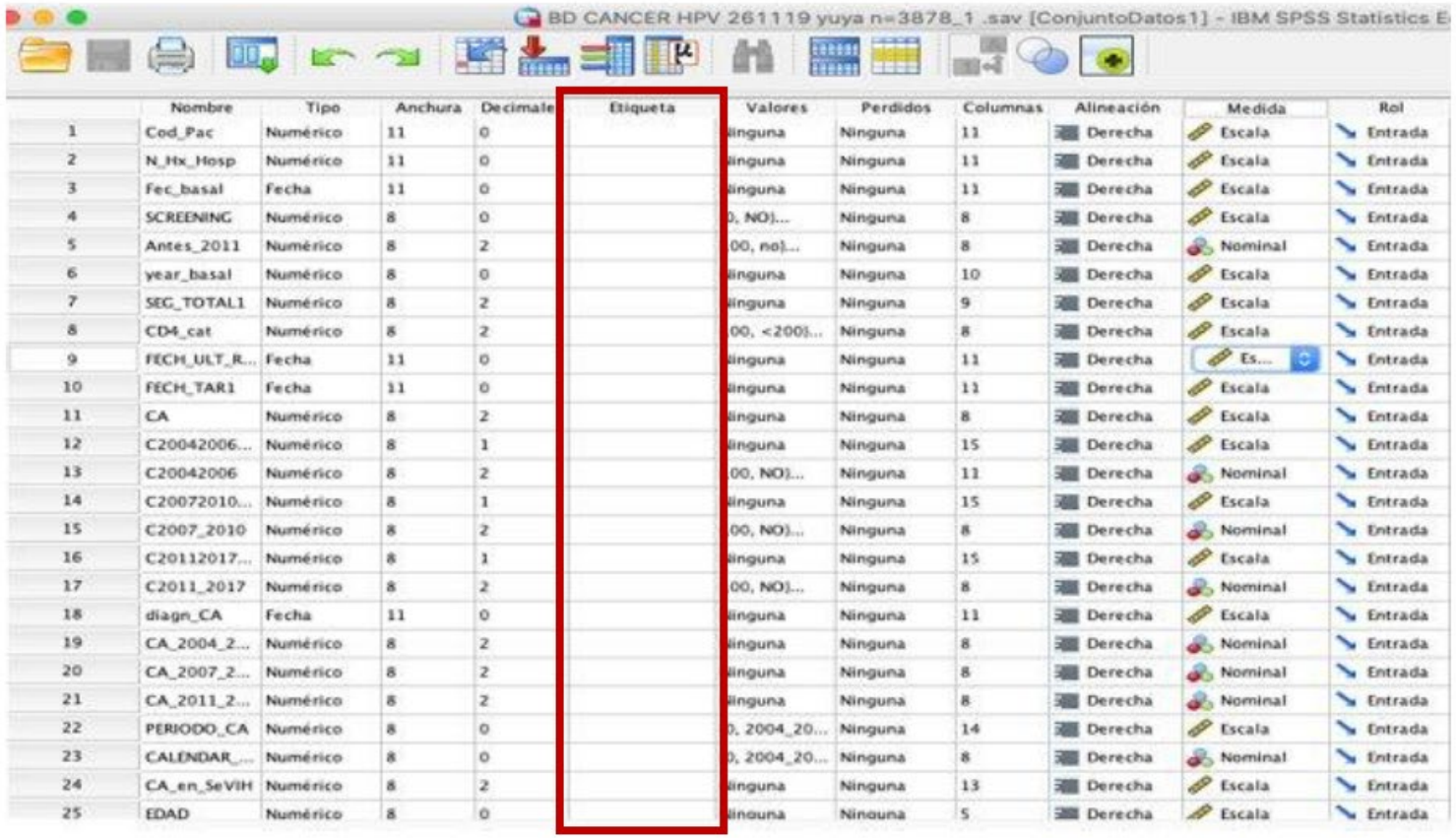
Errores en la Gestión de datos

- Si los datos que necesita todavía existen
- Si encontró los datos que necesita
- Si comprende los datos que encontró
- Si confía en los datos,
- Si puede utilizar los datos en los que confía
- Alguien hizo un buen trabajo en la gestión de datos.

Rex Sanders

Errores en la Gestión de datos

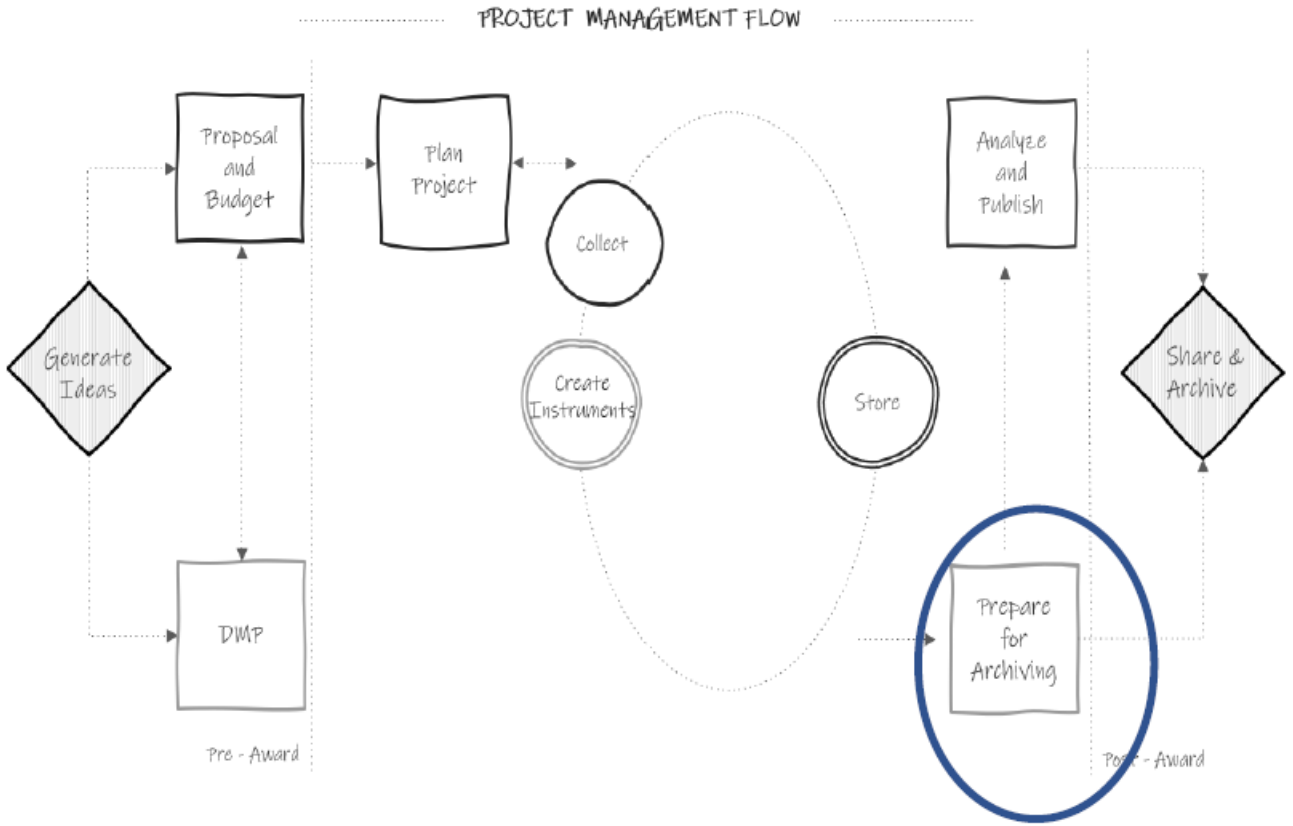
Ya trabajo con datos. ¿Para que la gestión?



	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Cod_Pac	Numérico	11	0		Ninguna	Ninguna	11	Derecha	Escala	Entrada
2	N_Hx_Hosp	Numérico	11	0		Ninguna	Ninguna	11	Derecha	Escala	Entrada
3	Fec_basal	Fecha	11	0		Ninguna	Ninguna	11	Derecha	Escala	Entrada
4	SCREENING	Numérico	8	0		0, NO}...	Ninguna	8	Derecha	Escala	Entrada
5	Antes_2011	Numérico	8	2		00, no}...	Ninguna	8	Derecha	Nominal	Entrada
6	year_basal	Numérico	8	0		Ninguna	Ninguna	10	Derecha	Escala	Entrada
7	SEG_TOTAL1	Numérico	8	2		Ninguna	Ninguna	9	Derecha	Escala	Entrada
8	CD4_cat	Numérico	8	2		00, <200}...	Ninguna	8	Derecha	Escala	Entrada
9	FECH_ULT_R...	Fecha	11	0		Ninguna	Ninguna	11	Derecha	Es...	Entrada
10	FECH_TAR1	Fecha	11	0		Ninguna	Ninguna	11	Derecha	Escala	Entrada
11	CA	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
12	C20042006...	Numérico	8	1		Ninguna	Ninguna	15	Derecha	Escala	Entrada
13	C20042006	Numérico	8	2		00, NO}...	Ninguna	11	Derecha	Nominal	Entrada
14	C20072010...	Numérico	8	1		Ninguna	Ninguna	15	Derecha	Escala	Entrada
15	C2007_2010	Numérico	8	2		00, NO}...	Ninguna	8	Derecha	Nominal	Entrada
16	C20112017...	Numérico	8	1		Ninguna	Ninguna	15	Derecha	Escala	Entrada
17	C2011_2017	Numérico	8	2		00, NO}...	Ninguna	8	Derecha	Nominal	Entrada
18	diagn_CA	Fecha	11	0		Ninguna	Ninguna	11	Derecha	Escala	Entrada
19	CA_2004_2...	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Nominal	Entrada
20	CA_2007_2...	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Nominal	Entrada
21	CA_2011_2...	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Nominal	Entrada
22	PERIODO_CA	Numérico	8	0		0, 2004_20...	Ninguna	14	Derecha	Escala	Entrada
23	CALENDAR_...	Numérico	8	0		0, 2004_20...	Ninguna	8	Derecha	Nominal	Entrada
24	CA_en_SeVIH	Numérico	8	2		Ninguna	Ninguna	13	Derecha	Escala	Entrada
25	EDAD	Numérico	8	0		Ninguna	Ninguna	5	Derecha	Escala	Entrada

Errores en la Gestión de datos

Error 1. Esperar hasta finalizar el proyecto para gestionar tus datos

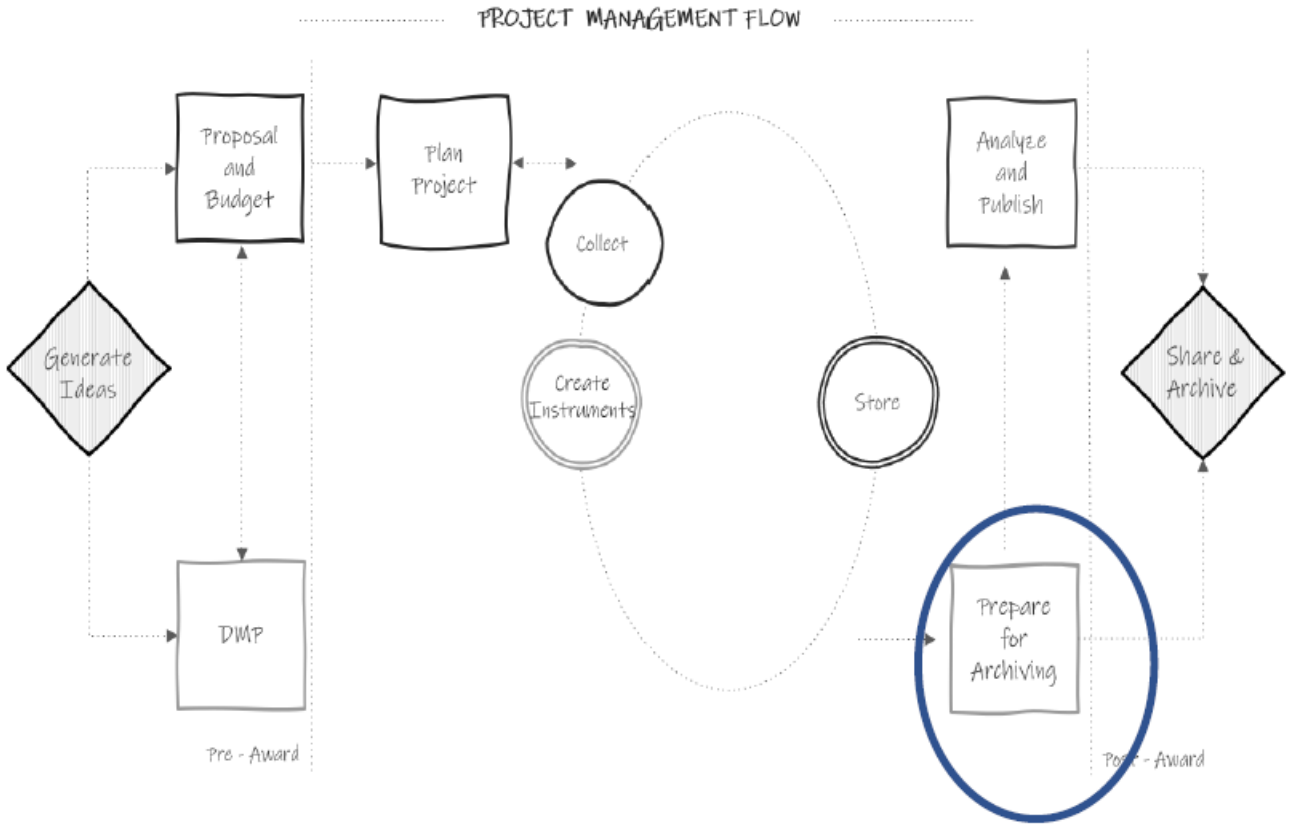


Milanés, Y. Errores vs buenas prácticas en Gestión de datos de investigación. https://zenodo.org/record/7540861#.Y8VL8BfMKUk

Errores en la Gestión de datos

Error 1. Esperar hasta finalizar el proyecto para gestionar tus datos

Solución: Un Plan de Gestión de Datos (PGD) debe ser tu aliado desde el inicio de la propuesta.



Milanés, Y. Errores vs buenas prácticas en Gestión de datos de investigación. https://zenodo.org/record/7540861#.Y8VL8BfMKUk

Errores en la Gestión de datos

Error 2. No usar guías de nomenclatura y organización de archivos y variables.

Misdatos.xls
Misdatosbuenos.xls
2001_data.xls
Version_buena.xls
Dataaltmetricsterminado.xls

Q1	q14_a	q14_b	Q15	Q16Open
1	1	m	10	3
2	5	f	11	2
4	13	f	8	1
5	22	m	15	4

Milanés, Y. Errores vs buenas prácticas en Gestión de datos de investigación. https://zenodo.org/record/7540861#.Y8VL8BfMKUk

Errores en la Gestión de datos

Error 2. No usar guías de nomenclatura y organización de archivos y variables.

Solución: Guía de estilos y nomenclaturas definidas.

Guías de estilo:

- ✓ Estructura de directorios
- ✓ Nomenclatura de ficheros (incluyendo versionado)
- ✓ Nomenclatura de variables
- ✓ Codificación de los valores de tus variables
- ✓ Codificación de los valores ausentes (Missing values)

Te permitirá:

- ✓ Mejorar la búsqueda
- ✓ Una fácil interpretación
- ✓ Mejorar la reproducibilidad
- ✓ Estandarizar

Milanés, Y. Errores vs buenas prácticas en Gestión de datos de investigación. https://zenodo.org/record/7540861#.Y8VL8BfMKUk

Errores en la Gestión de datos

Error 3. No documentar

- NO REPRODUCIBILIDAD
- REDUCE SEGURIDAD DE LOS DATOS
- BAJA CALIDAD EN LOS DATOS
- COSTES
- PÉRDIDA DE TIEMPO Y EFECTIVIDAD EN LA INVESTIGACIÓN

Milanés, Y. Errores vs buenas prácticas en Gestión de datos de investigación. https://zenodo.org/record/7540861#.Y8VL8BfMKUk

Errores en la Gestión de datos

Error 3. No documentar

Solución: Documentar todos los procesos realizados en los datos.

A nivel de proyecto: PROTOCOLO

- ✓ Reclutamiento
- ✓ Criterios de inclusión/exclusión
- ✓ Recolección de los datos / Procedimientos
- ✓ Procedimientos tratamiento de los datos
- ✓ Seguridad de los datos
- ✓ Control de la calidad
- ✓ Anonimización de los datos

Te permitirá:

- ✓ Reproducir en menos tiempo
- ✓ Calidad y fiabilidad de los datos y procedimientos
- ✓ Mejorar la reproducibilidad
- ✓ Estandarizar

Milanés, Y. Errores vs buenas prácticas en Gestión de datos de investigación. https://zenodo.org/record/7540861#.Y8VL8BfMKUk

Errores en la Gestión de datos

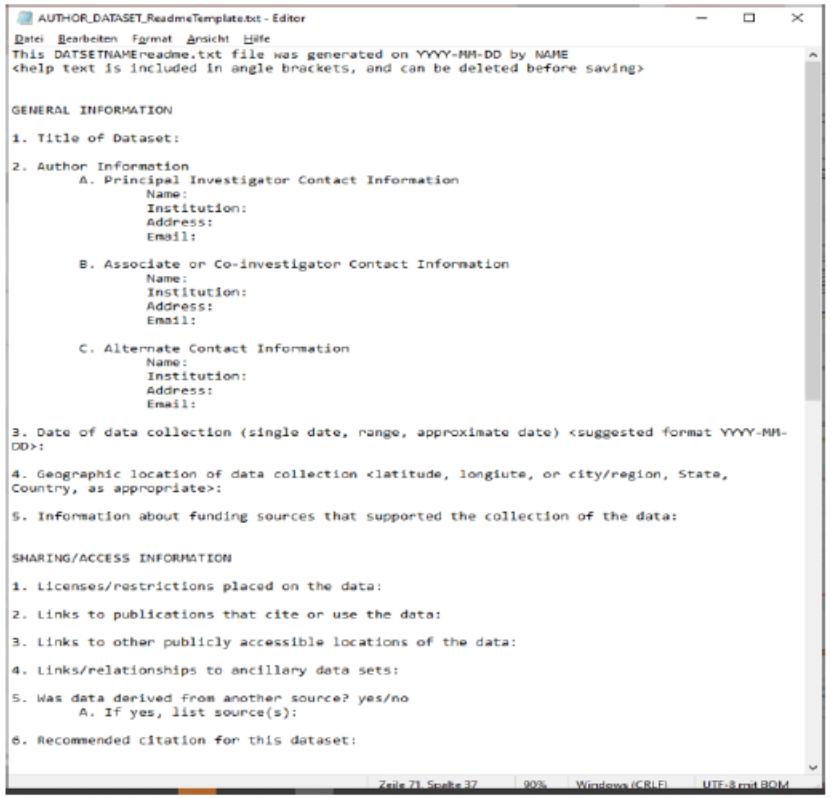
Error 3. No documentar

Solución: Documentar todos los procesos realizados en los datos.

Documéntalo TODO

¿Cómo documentarlo?

- ✓ Diccionario de datos
- ✓ **Fichero Readme.txt**
- ✓ Libro de código de variables
- ✓ Guías de usuario
- ✓ Software syntax
- ✓ Cuadernos de laboratorio



Milanés, Y. Errores vs buenas prácticas en Gestión de datos de investigación. https://zenodo.org/record/7540861#.Y8VL8BfMKUk

Errores en la Gestión de datos

Error 4. No crear un diccionario de datos antes de recoger los datos

Solución: Documentar todos los procesos realizados en los datos.

☹ NO TIENES LAS VARIABLES ESTANDARIZADAS

☹ REDUCE DRÁSTICAMENTE LA CALIDAD DE LOS DATOS

☹ COSTES

☹ PÉRDIDA DE TIEMPO Y EFECTIVIDAD EN LA INVESTIGACIÓN

Milanés, Y. Errores vs buenas prácticas en Gestión de datos de investigación. https://zenodo.org/record/7540861#.Y8VL8BfMKUk

Errores en la Gestión de datos

Error 4. No crear un diccionario de datos antes de recoger los datos

Solución: Ayuda a entender las variables. Tanto las RAW como las calculadas después.

Campos a tener en cuenta en un diccionario de datos

- ✓ Elemento identificador de la tabla (por sí tienes datos en varias tablas)
- ✓ Nombres de las variables
- ✓ Definición de cada variable (cómo se entiende en tu estudio)
- ✓ Tipo de datos
- ✓ Longitud del campo
- ✓ ¿Campo requerido? y/n
- ✓ ¿Valores nulos? Codificación

A1 Variables

	Variables	Código de la variable	Grupo de la variable	Definición conceptual	Tipo de variable	Naturaleza de la variable	Definición operativa
2	Género	Género	Demográficos	Identidad de género	Independiente	Cualitativa politómica	Mujer= 1; Hombre= 2, Otros= 3, Prefiero no decirlo= 4
3	Edad	Edad	Demográficos	Edad del paciente en el momento de realización del estudio.	Independiente	Cuantitativa continua	Edad (años)
4	Nivel de estudio	Estudios	Demográficos	Nivel máximo de estudios alcanzados	Independiente	Cualitativa politómica	Sin estudios= 0; Primaria= 1; Secundaria= 2; Bachillerato o Grado profesional= 3; Universitario= 4
5	Nivel socio-económico	SocioEcon	Demográficos	Nivel socio-económico (Renta Familiar)	Independiente	Cualitativa politómica	<12.000€=0; 12001-24000€=1; 24.001-36.000=2; 36.001-50.000=3; >50.000€=4
6	Pregunta MHA 1	MHA_1	Modificaciones en los hábitos alimentarios	1.- ¿Ha hecho usted un cambio restrictivo en su dieta debido a su intolerancia alimentaria?	Independiente	Cualitativa politómica	No he cambiado mi dieta nunca= 0; Lo hice después de hacerme la prueba=1; Lo hice incluso antes de hacerme la prueba= 2
7	Pregunta MHA 2	MHA_2	Modificaciones en los hábitos alimentarios	2.- ¿Ha hecho usted los cambios en la dieta siguiendo las pautas de algún médico o nutricionista?	Independiente	Cualitativa politómica	No he cambiado mi dieta nunca= 0; He hecho los cambios que me ha indicado mi médico= 1; He hecho los cambios que me ha indicado mi especialista= 2; He hecho lo cambios que me ha indicado mi nutricionista= 3
8	Pregunta MHA 3	MHA_3	Modificaciones en los hábitos alimentarios	3.- Si ha hecho usted algún cambio en la dieta, ¿lo mantiene en el tiempo?	Independiente	Cualitativa politómica	No he cambiado mi dieta nunca= 0; Con frecuencia consumo alimentos que me sientan mal= 1; Puntualmente consumo alimentos que me sientan mal= 2; Sigo mi dieta a diario y evito los alimentos que me sientan mal= 3
9	Pregunta MHA 4	MHA_4	Modificaciones en los hábitos alimentarios	4.- ¿Qué motivos principales cree que influyen en no poder cambiar su dieta para evitar alimentos no recomendados? (puede marcar varias respuestas)	Independiente	Cualitativa politómica	Los alimento recomendados son más caros= 1; El sabor de los alimentos adaptados es diferente al que estoy acostumbrado= 2; Si como fuera de casa tengo menos opciones para elegir un plato que me guste= 3; Es difícil de encontrar una oferta variada en mis tiendas habituales= 4; No considero que sea necesario cambiar m alimentación= 5
10	Cumplimiento de MHA	Cumplimiento	Modificaciones en los hábitos alimentarios	Adherencia al cambio dietético recomendado. Se calcula mediante la suma de las preguntas MHA_1 a MHA_3, se considera incumplidor si alguna de las	Independiente	Cualitativa dicotómica	Cumplidor= 0; Incumplidor= 1

Milanés, Y. Errores vs buenas prácticas en Gestión de datos de investigación. https://zenodo.org/record/7540861#.Y8VL8BfMKUk

Errores en la Gestión de datos

Error 5. Trabajar con el archivo original

Solución: crear una carpeta con una copia del Data Raw

Milanés, Y. Errores vs buenas prácticas en Gestión de datos de investigación. https://zenodo.org/record/7540861#.Y8VL8BfMKUk

Errores en la Gestión de datos

Error 6. No trabajar con datos tidy o " long

☹ SI LOS DATOS CRECEN HORIZONTALMENTE,

☹ REDUCE DRÁSTICAMENTE LA REPRODUCIBILIDAD DE LOS DATOS

☹ COSTES

☹ PÉRDIDA DE TIEMPO Y EFECTIVIDAD EN LA INVESTIGACIÓN

Corvallis_VegBiodiv_2006.csv

Site	Sp1	Sp2	Sp3	Sp4	Sp5
XYZ	12	4	8	16	3
PDQ	4	14	0	9	64

Corvallis_VegBiodiv_2007.csv

Site	Sp1	Sp2	Sp3	Sp4	Sp5	Sp6	Sp7
XYZ	21	0	0	45	0	12	6
PDQ	0	4	0	0	16	3	0

Milanés, Y. Errores vs buenas prácticas en Gestión de datos de investigación. https://zenodo.org/record/7540861#.Y8VL8BfMKUk

Errores en la Gestión de datos

Error 6. No trabajar con datos tidy o “ long

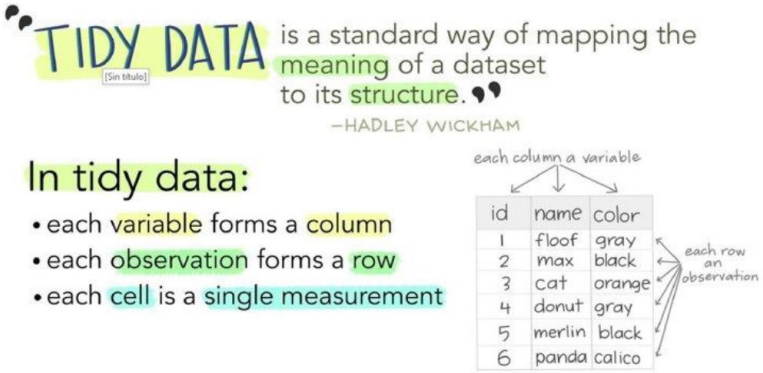
Solución: Trabajar con datos tidy long ”) o mantener ambas versiones

Corvallis_VegBiodiv_2006.csv

Site	Sp1	Sp2	Sp3	Sp4	Sp5
XYZ	12	4	8	16	3
PDQ	4	14	0	9	64

Corvallis_VegBiodiv_2006.csv

Site	SpName	Abundance
XYZ	Sp. 1	12
XYZ	Sp. 2	4
XYZ	Sp. 3	8
XYZ	Sp. 4	16
XYZ	Sp. 5	3
PDQ	Sp. 1	4
PDQ	Sp. 2	14
PDQ	Sp. 4	9
PDQ	Sp. 5	64

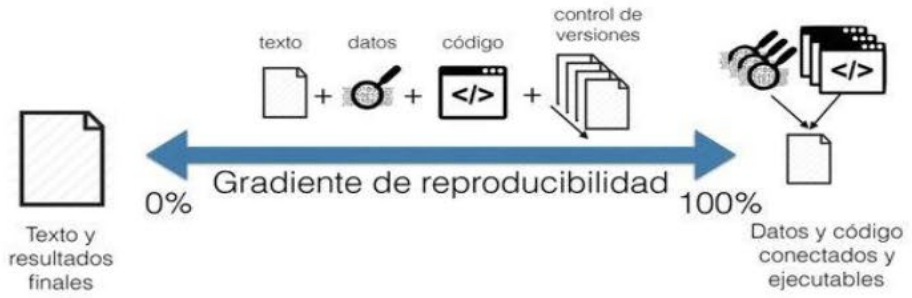


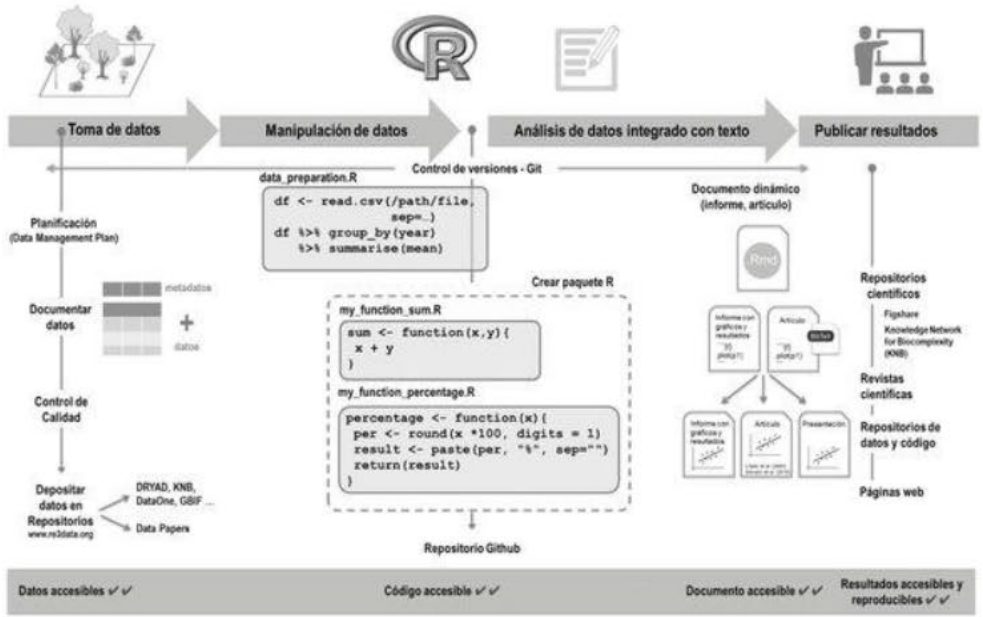
Milanés, Y. Errores vs buenas prácticas en Gestión de datos de investigación. https://zenodo.org/record/7540861#.Y8VL8BfMKUk

Errores en la Gestión de datos

Error 7. Evaluar la necesidad de datos de forma reproducible y/o automatizada desde el inicio.

Solución: Publicación + Código + datos ejecutables





Milanés, Y. Errores vs buenas prácticas en Gestión de datos de investigación. https://zenodo.org/record/7540861#.Y8VL8BfMKUk

Herramientas para la manipulación de datos

Alfabetización de datos

Data Literacy

La alfabetización de datos significa **habilidad para leer, analizar y argumentar datos** y comunicarse en base a ellos.

Utilizar datos en la **toma de decisiones y abrirlos al personal** en todos los niveles de la organización. La alfabetización de datos es la piedra angular del éxito empresarial, académico y científico.

Daniel Rosemberg, historiador (1700), los datos eran utilizados como **conceptos generales que servían para argumentar.**

Ventajas de la Alfabetización de datos

- Comunicarse en un lenguaje común de datos para entender mejor las conversaciones sobre ellos.
- Utilizar normativas para la correcta comunicación de los datos y correcta interpretación en su uso.
- Detectar problemas operativos inesperados e identificar las causas de fondo.
- Evitar que se tomen decisiones erróneas debido a una mala interpretación.

Importancia de la Alfabetización de datos

Los datos permiten a las organizaciones mejorar su negocio aumentando la precisión, la eficiencia y la capacidad de los empleados, **basando sus desiciones en datos**.
Los datos son el oro digital , pero no tienen valor, si no son fácilmente utilizables (FAIR).

Se estima que la falta de información cuesta 5 días de productividad por empleado.

Características de la Alfabetización de datos

Alfabetización de datos | Documentación de Datos

Nomenclatura de archivos:

Aspecto clave para administrar sus datos es nombrar y organizar sus archivos y carpetas asociadas de manera efectiva utilizando una buena nomenclatura.

1. Mantenga los nombres de los archivos breves y relevantes
2. No utilices caracteres especiales
3. No utilices puntos ni espacios
4. Formato de fecha (AAAA-MM-DD)
5. Identifique casos (entrevista, encuesta, etc)
6. Identifique proyecto (estructura de datos corta)

Nombre del archivo erróneos:

entrevista02.docx
Data_01.xlsx
Datos_marzo_2022.csv
Datos-limpios.xlsx
PRO1-encuesta-funciona.doc

https://mantra.ed.ac.uk/organisingdata/

Alfabetización de datos | Documentación de Datos

Nomenclatura de archivos:
Usando esta convención particular de nomenclatura de archivos, podemos averiguar fácilmente qué tipo de datos son y como están relacionados.

Nombre del archivo : **almetrics_en_R01_v3_ 20190731.csv**

Componentes del nombre del archivo: **almetrics en** R01 v3 **20190731**
almetric = Nombre del Proyecto
en = entrevista (tipo de datos)
R[n] = ID del investigador – investigador 01
V3 = versión de los datos
Fecha de entrevista en = AAAAMMDD – 20190731

**Puede encontrar en la estructura de carpeta , una carpeta llamada data con un archivo que en su nomenclatura indique RAW, son datos brutos u originales sin cambios.*

https://mantra.ed.ac.uk/organisingdata/

Alfabetización de datos | Documentación de Datos

Diccionario de datos:
Explican las variables utilizadas en un conjunto de datos con el fin de evitar malentendidos, también pueden entenderse como colecciones o descripciones de códigos, Algoritmos y cálculos utilizados en un proyecto.

Sheet_1

Show rows with cells including:

Variable	Variable name	Mesaurement unit	Allowed values	Description
Participant ID number	ID	Numeric	001-999	ID number assigned to participant in sequential order
Group number	GROUP	Numeric	1-30	Group assigned to participant based on ID number
Age in years	AGE	Numeric	18.0-65.0	Age of participant in years
Date of birth	DOB	mm/dd/yyyy	1-12/1-31/1951-1998	Participant's date of birth
Gender	SEX	Numeric	1 = male 2 = female	Participant's gender
Date of survey	SURVEY	mm/dd/yyyy	01/01/2015 – 01/01/2016	When the participant completed the survey
Self-reported consumer spending	SPEND	Numeric	0-100,000,000	Self-reported average yearly expenditure
Market sentiment	SENTIMENT	Numeric	1 = negative 2 = neutral 3 = positive	Sentiment towards US domestic economy
Actual GDP growth	GDP	Numeric	-5.0-5.0	Average US yearly GDP growth

Guide for data documentation | https://zenodo.org/records/1914401

Alfabetización de datos | Documentación de Datos

Versión de controles:
El control de versiones permite volver a una versión anterior de un archivo específico. Es posible utilizar un software automático (siempre preferible) o realizar un seguimiento manual de los archivos

Automático

Manual – definir nomenclatura de fecha en los nombres

Guide for data documentation | https://zenodo.org/records/1914401

Alfabetización de datos | Documentación de Datos

Readme-files:

Los archivos **Readme** son documentos de texto (en formato .txt) que proporcionan información sobre archivos de datos para garantizarse interpretan correctamente. Estos se vuelven especialmente importantes al compartir y publicar datos.

Puede contener información como:

- Autoría
- Título
- Descripción
- Metodología
- proyectos financiadores
- cobertura temporal y geográfica
- Derechos de uso y privacidad, etc.
- Listado de variables
- Procesos utilizados en la limpieza
- Herramientas

Ejemplo: https://cornell.app.box.com/v/ReadmeTemplate

Guide for data documentation | https://zenodo.org/records/1914401

Alfabetización de datos | Lectura de datos

Es recomendable seleccionar **formatos abiertos**, no propietarios como tomar en cuenta en que formato se trabaja en el campo según disciplina. Es necesario evaluar en que formato(s) se desea comunicar los datos finales.

Tipos de formatos más utilizados:

- XLS - formato propietario
- JSON – formato intercambio de datos (API)
- CSV – formato abierto

Otros

- XML
- RTF
- TXT
- Rdata (objeto de datos en R)

Alfabetización de datos | Lectura los datos

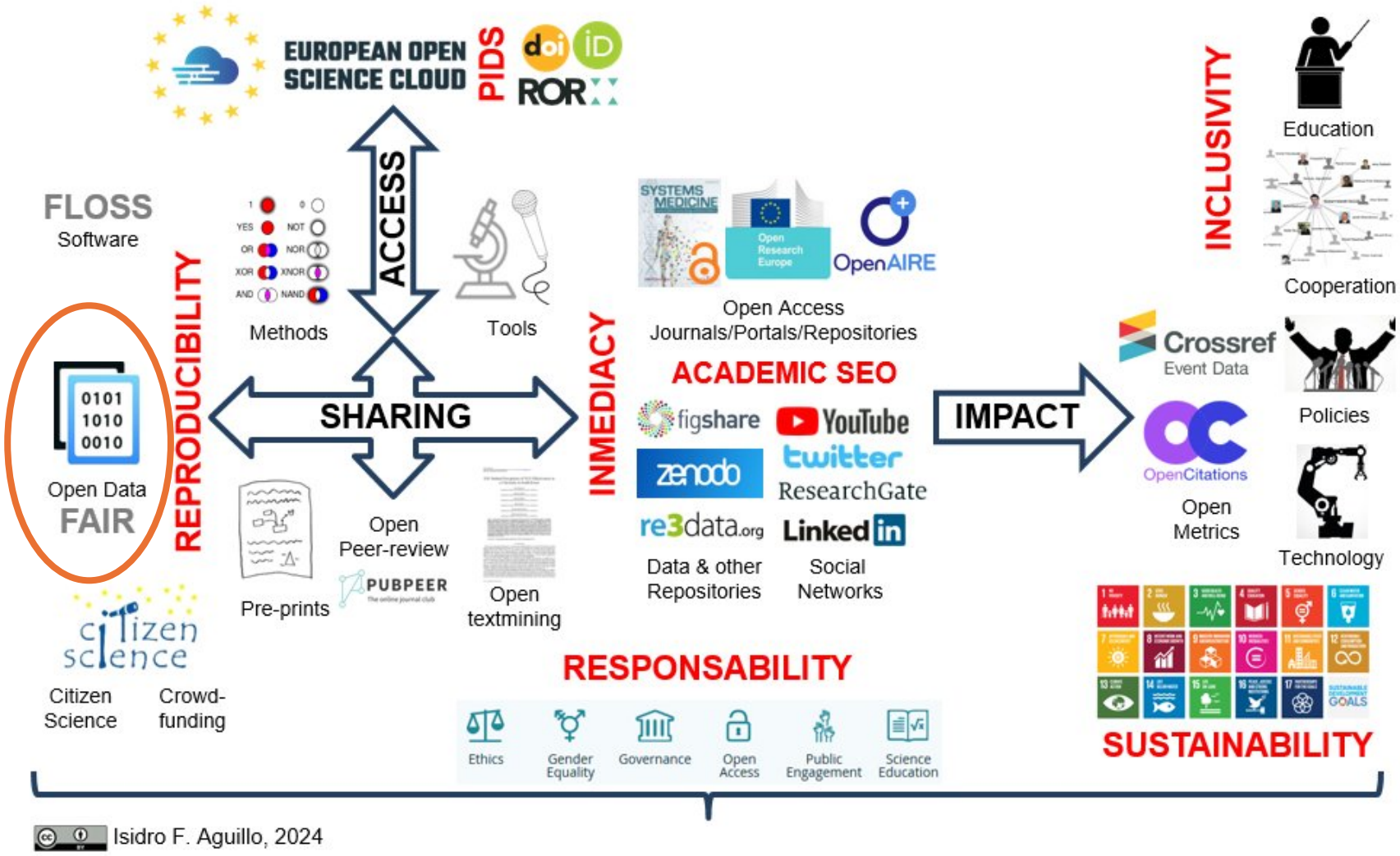
Tipos de datos más utilizados:

- XLS / XLSX - formato propietario
- JSON – formato intercambio de datos (API)
- CSV – formato abierto

Apertura de datos abiertos

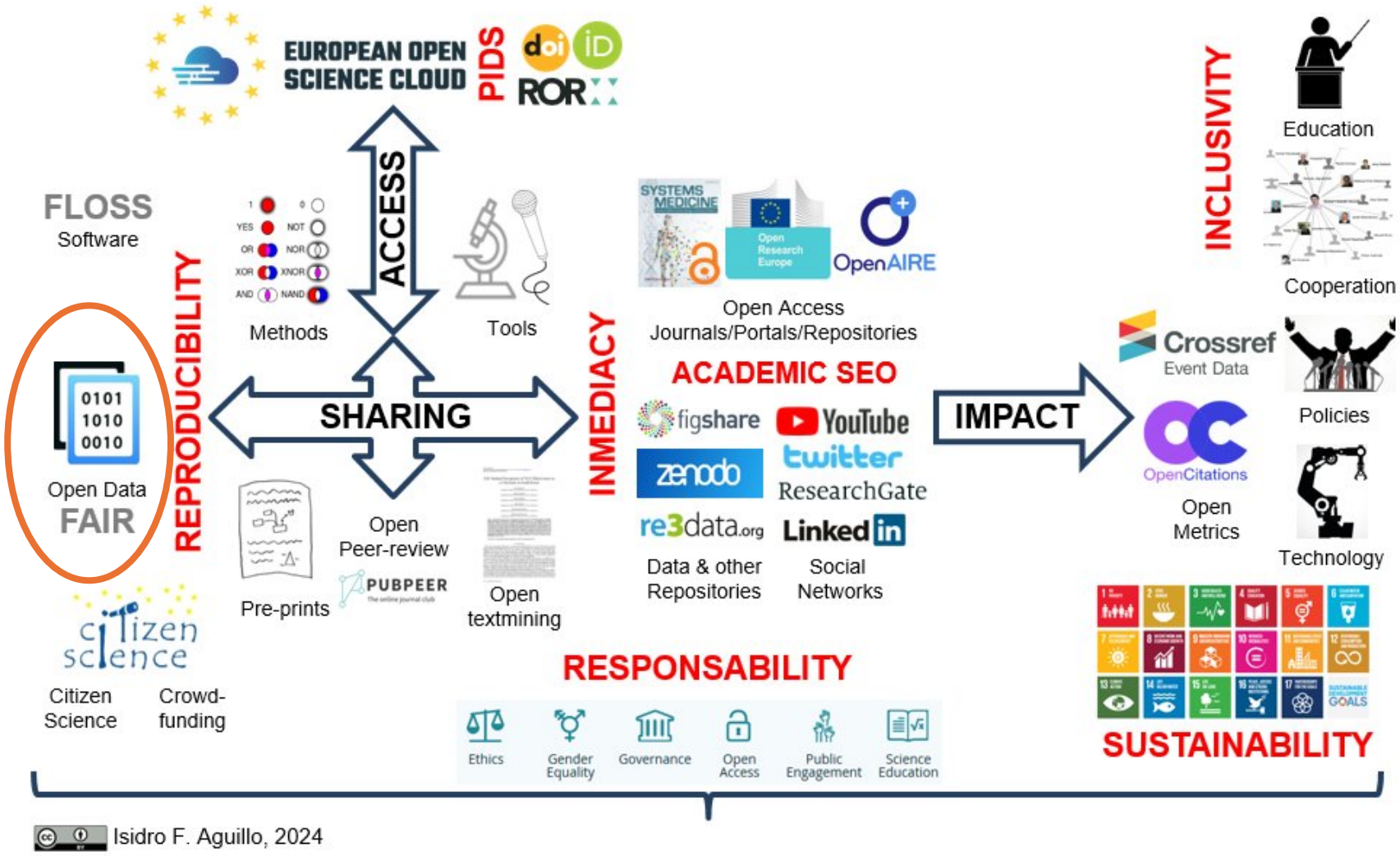
Alfabetización de datos | Lectura los datos

Ecosistema de Ciencia abierta| Datos



Alfabetización de datos | Lectura los datos

Ecosistema de Ciencia abierta| Datos



Alfabetización de datos

Estructura de datos | Formatos | Datos ordenados

Alfabetización de datos | Estructura de datos | XLSX

Ejemplo 1:
Datos del INEC

Abra el archivo **INEC-datos-morbilidad.xlsx**

¿Describa algunas características del conjunto de datos del INEC, filas , columnas, valores, apariencia?

MINISTERIO DE SALUD
DIRECCION DE PLANIFICACIÓN DE SALUD / DEPARTAMENTO DE REGISTROS Y ESTADÍSTICAS DE SALUD

MORBILIDAD POR CANDIDIASIS, ATENDIDAS EN LA CONSULTA EXTERNA DE LAS INSTALACIONES DEL MINISTERIO DE
AÑO 2020

Cód.	CAUSAS 1/	TOTAL			Menor de un año			1 a 4 años			5 a 9 años			10 a 14 años			15 a 19 años			20 a 24 años			25 a 34 años			35 a 49 años		
		T	H	M	T	H	M	T	H	M	T	H	M	T	H	M	T	H	M	T	H	M	T	H	M	T	H	M
		6660	**681**	**5979**	**316**	**153**	**163**	**487**	**204**	**283**	**181**	**69**	**112**	**127**	**27**	**100**	**596**	**22**	**574**	**1134**	**40**	**1094**	**1589**	**51**	**1538**	**1452**	**50**	**1402**
B37.3	Candidiasis de la vulva y de la vagina	4012	0	4012	10	0	10	17	0	17	31	0	31	62	0	62	386	0	386	825	0	825	1174	0	1174	1024	0	1024
B37.9	Candidiasis, no especificada	1535	198	1337	93	47	46	103	40	63	36	10	26	32	12	20	175	9	166	247	18	229	325	20	305	337	21	316
B37.0	Estomatitis candidiasica	800	386	414	169	87	82	322	145	177	105	54	51	20	9	11	24	10	14	26	16	10	35	21	14	39	20	19
B37.8	Candidiasis de otros sitios	148	59	89	42	18	24	38	15	23	6	3	3	5	3	2	3	3	0	6	2	4	16	4	12	14	3	11
B37.4	Candidiasis de otras localizaciones urogenitales	82	20	62	0	0	0	3	2	1	1	0	1	4	1	3	5	0	5	15	2	13	18	3	15	22	3	19
B37.2	Candidiasis de la piel y las uñas	52	11	41	1	0	1	1	0	1	2	2	0	3	2	1	1	0	1	6	1	5	16	3	13	11	2	9
B37.7	Sepsis debida a Candida	20	4	16	0	0	0	0	0	0	0	0	0	1	0	1	2	0	2	8	1	7	3	0	3	3	1	2
B37.1	Candidiasis pulmonar	7	3	4	1	1	0	3	2	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1
B37.5	Meningitis debida a Candida	3	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1
B37.6	Endocarditis debida a Candida	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
		397	**63**	**334**	**31**	**20**	**11**	**61**	**25**	**36**	**19**	**4**	**15**	**7**	**1**	**6**	**49**	**2**	**47**	**69**	**1**	**68**	**75**	**2**	**73**	**65**	**5**	**60**
B37.3	Candidiasis de la vulva y de la vagina	234	0	234	0	0	0	1	0	1	5	0	5	5	0	5	38	0	38	56	0	56	59	0	59	54	0	54
B37.0	Estomatitis candidiasica	93	39	54	22	13	9	45	16	29	11	4	7	2	1	1	3	0	3	2	1	1	5	1	4	3	3	0
B37.9	Candidiasis, no especificada	49	13	36	5	4	1	5	3	2	3	0	3	0	0	0	6	1	5	10	0	10	10	1	9	6	1	5
B37.8	Candidiasis de otros sitios	17	10	7	4	3	1	10	6	4	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0
B37.4	Candidiasis de otras localizaciones urogenitales	4	1	3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	1	2	1	1
		536	**31**	**505**	**10**	**6**	**4**	**29**	**12**	**17**	**9**	**2**	**7**	**2**	**0**	**2**	**39**	**0**	**39**	**102**	**1**	**101**	**142**	**3**	**139**	**135**	**5**	**130**
B37.3	Candidiasis de la vulva y de la vagina	410	0	410	0	0	0	0	0	0	2	0	2	1	0	1	34	0	34	81	0	81	124	0	124	111	0	111
B37.9	Candidiasis, no especificada	60	13	47	3	2	1	9	5	4	0	0	0	0	0	0	1	0	1	14	1	13	10	1	9	15	3	12
B37.0	Estomatitis candidiasica	43	14	29	7	4	3	17	6	11	6	1	5	1	0	1	3	0	3	2	0	2	1	1	0	3	1	2
B37.8	Candidiasis de otros sitios	7	2	5	0	0	0	3	1	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	3

Alfabetización de datos | Estructura de datos | XLSX

Ejemplo 1:
Datos del INEC

Títulos en las primeras filas.
Hoja de datos con celdas fusionadas, no hay un dato por cada columna.
No es una tabla tidy (ordenada).
Visualización para usuario, no para un análisis.

MINISTERIO DE SALUD
DIRECCION DE PLANIFICACIÓN DE SALUD / DEPARTAMENTO DE REGISTROS Y ESTADÍSTICAS DE SALUD

MORBILIDAD POR CANDIDIASIS, ATENDIDAS EN LA CONSULTA EXTERNA DE LAS INSTALACIONES DEL MINISTERIO DE
AÑO 2020

Cód.	CAUSAS 1/	TOTAL			Menor de un año			1 a 4 años			5 a 9 años			10 a 14 años			15 a 19 años			20 a 24 años			25 a 34 años			35 a 49 años		
		T	H	M	T	H	M	T	H	M	T	H	M	T	H	M	T	H	M	T	H	M	T	H	M	T	H	M
		6660	**681**	**5979**	**316**	**153**	**163**	**487**	**204**	**283**	**181**	**69**	**112**	**127**	**27**	**100**	**596**	**22**	**574**	**1134**	**40**	**1094**	**1589**	**51**	**1538**	**1452**	**50**	**1402**
B37.3	Candidiasis de la vulva y de la vagina	4012	0	4012	10	0	10	17	0	17	31	0	31	62	0	62	386	0	386	825	0	825	1174	0	1174	1024	0	1024
B37.9	Candidiasis, no especificada	1535	198	1337	93	47	46	103	40	63	36	10	26	32	12	20	175	9	166	247	18	229	325	20	305	337	21	316
B37.0	Estomatitis candidiasica	800	386	414	169	87	82	322	145	177	105	54	51	20	9	11	24	10	14	26	16	10	35	21	14	39	20	19
B37.8	Candidiasis de otros sitios	148	59	89	42	18	24	38	15	23	6	3	3	5	3	2	3	3	0	6	2	4	16	4	12	14	3	11
B37.4	Candidiasis de otras localizaciones urogenitales	82	20	62	0	0	0	3	2	1	1	0	1	4	1	3	5	0	5	15	2	13	18	3	15	22	3	19
B37.2	Candidiasis de la piel y las uñas	52	11	41	1	0	1	1	0	1	2	2	0	3	2	1	1	0	1	6	1	5	16	3	13	11	2	9
B37.7	Sepsis debida a Candida	20	4	16	0	0	0	0	0	0	0	0	0	1	0	1	2	0	2	8	1	7	3	0	3	3	1	2
B37.1	Candidiasis pulmonar	7	3	4	1	1	0	3	2	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1
B37.5	Meningitis debida a Candida	3	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1
B37.6	Endocarditis debida a Candida	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
		397	**63**	**334**	**31**	**20**	**11**	**61**	**25**	**36**	**19**	**4**	**15**	**7**	**1**	**6**	**49**	**2**	**47**	**69**	**1**	**68**	**75**	**2**	**73**	**65**	**5**	**60**
B37.3	Candidiasis de la vulva y de la vagina	234	0	234	0	0	0	1	0	1	5	0	5	5	0	5	38	0	38	56	0	56	59	0	59	54	0	54
B37.0	Estomatitis candidiasica	93	39	54	22	13	9	45	16	29	11	4	7	2	1	1	3	0	3	2	1	1	5	1	4	3	3	0
B37.9	Candidiasis, no especificada	49	13	36	5	4	1	5	3	2	3	0	3	0	0	0	6	1	5	10	0	10	10	1	9	6	1	5
B37.8	Candidiasis de otros sitios	17	10	7	4	3	1	10	6	4	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0
B37.4	Candidiasis de otras localizaciones urogenitales	4	1	3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	1	2	1	1
		536	**31**	**505**	**10**	**6**	**4**	**29**	**12**	**17**	**9**	**2**	**7**	**2**	**0**	**2**	**39**	**0**	**39**	**102**	**1**	**101**	**142**	**3**	**139**	**135**	**5**	**130**
B37.3	Candidiasis de la vulva y de la vagina	410	0	410	0	0	0	0	0	0	2	0	2	1	0	1	34	0	34	81	0	81	124	0	124	111	0	111
B37.9	Candidiasis, no especificada	60	13	47	3	2	1	9	5	4	0	0	0	0	0	0	1	0	1	14	1	13	10	1	9	15	3	12
B37.0	Estomatitis candidiasica	43	14	29	7	4	3	17	6	11	6	1	5	1	0	1	3	0	3	2	0	2	1	1	0	3	1	2
B37.8	Candidiasis de otros sitios	7	2	5	0	0	0	3	1	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	3

Alfabetización de datos | Leer los datos | XLSX

10 consejos útiles para estructurar sus datos en XLSX:

1. **No fusiones celdas**
2. **No mezcle datos y metadatos** (fecha de publicación, el nombre del autor).
3. **La primera fila de una hoja de datos debe contener encabezados de columna** .
4. **Las filas restantes deben contener datos**. No TOTAL o PROMEDIO.
5. **Los números en las celdas deben ser solo números** . No les pongas comas.
6. **Utilice identificadores estándar** : identifique países ISO 3166, ISO 8601 para fecha.
7. **No utilice colores u otras señales estilísticas para codificar información**.
8. **Deje la celda en blanco si un valor no está disponible** .
9. Si proporciona tablas dinámicas, asegúrese de que los **datos subyacentes también estén disponibles por separado**.
10. Si desea crear una **presentación de los datos amigable para las personas, hágalo en otra hoja**.

Alfabetización de datos | Leer datos | Datos ordenados

Ejemplo 2:
Abrir archivo **datos-practica-visualización.xlsx**
Vaya a la Hoja **country-data-wide.**
Describa la tabla

Ver documento de practica
practicas-datos-ordenados-Datawraper.doc

country	1952	1957	1962	1967	1972	1977	1982	1987	1992	
Albania		-9	-9	-9	-9	-9	-9	-9	-9	5
Argentina		-9	-1	-1	-9	-9	-9	-8	8	7
Armenia		-9	-7	-7	-7	-7	-7	-7	-7	7
Australia		10	10	10	10	10	10	10	10	10
Austria		10	10	10	10	10	10	10	10	10
Azerbaijan		-9	-7	-7	-7	-7	-7	-7	-7	1
Belarus		-9	-7	-7	-7	-7	-7	-7	-7	7
Belgium		10	10	10	10	10	10	10	10	10
Bhutan		-10	-10	-10	-10	-10	-10	-10	-10	-10
Bolivia		-4	-3	-3	-4	-7	-7	8	9	9
Brazil		5	5	5	-9	-9	-4	-3	7	8
Bulgaria		-7	-7	-7	-7	-7	-7	-7	-7	8
Canada		10	10	10	10	10	10	10	10	10
Chile		2	5	5	6	6	-7	-7	-6	8
China		-8	-8	-8	-9	-8	-7	-7	-7	-7
Colombia		-5	7	7	7	7	8	8	8	9
Costa Rica		10	10	10	10	10	10	10	10	10

Alfabetización de datos | Leer datos | Datos ordenados

Ejemplo 2:

Cinco problemas más comunes con conjuntos de **datos desordenados**:

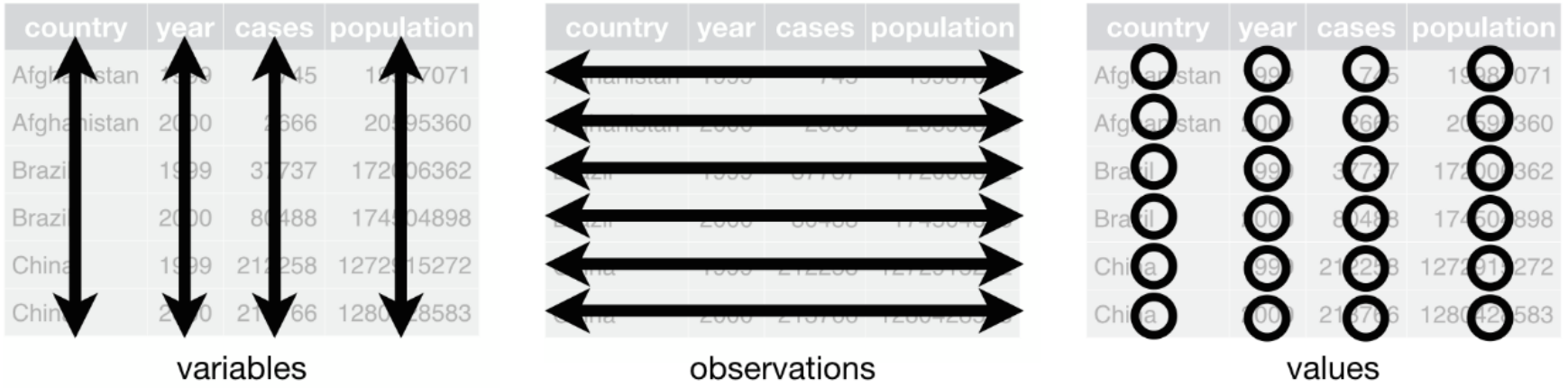
1. Los encabezados de columna son valores, no nombres de variables.
2. Se almacenan varias variables en una columna.
3. Las variables se almacenan tanto en filas como en columnas.
4. En la misma tabla se almacenan varios tipos de unidades de observación.
5. Una única unidad de observación se almacena en varias tablas.

country	1952	1957	1962	1967	1972	1977	1982	1987	1992	
Albania		-9	-9	-9	-9	-9	-9	-9	-9	5
Argentina		-9	-1	-1	-9	-9	-9	-8	8	7
Armenia		-9	-7	-7	-7	-7	-7	-7	-7	7
Australia		10	10	10	10	10	10	10	10	10
Austria		10	10	10	10	10	10	10	10	10
Azerbaijan		-9	-7	-7	-7	-7	-7	-7	-7	1
Belarus		-9	-7	-7	-7	-7	-7	-7	-7	7
Belgium		10	10	10	10	10	10	10	10	10
Bhutan		-10	-10	-10	-10	-10	-10	-10	-10	-10
Bolivia		-4	-3	-3	-4	-7	-7	8	9	9
Brazil		5	5	5	-9	-9	-4	-3	7	8
Bulgaria		-7	-7	-7	-7	-7	-7	-7	-7	8
Canada		10	10	10	10	10	10	10	10	10
Chile		2	5	5	6	6	-7	-7	-6	8
China		-8	-8	-8	-9	-8	-7	-7	-7	-7
Colombia		-5	7	7	7	7	8	8	8	9
Costa Rica		10	10	10	10	10	10	10	10	10

Alfabetización de datos | Leer datos | Datos ordenados

Tidy data (datos ordenados) : son simplemente datos organizados de una determinada manera donde el conjunto de datos deben cumplir con las siguientes tres reglas interconectadas:

1. Cada variable está en contenida en una columna.
2. Cada observación se encuentra en una fila.
3. Cada valor (observación/variable) corresponde a una celda.



Alfabetización de datos | Leer los datos | JSON

Ejemplo:

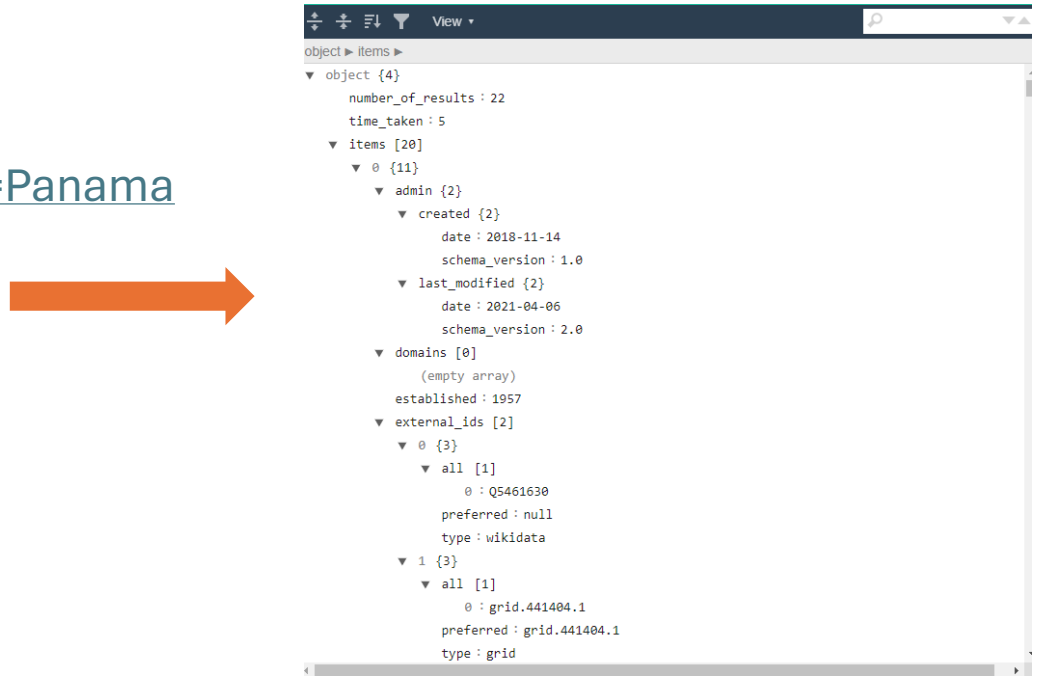
- Página de registros de Organizaciones de investigación https://ror.org/
- Explicación API ROR.org https://ror.org/tutorials/intro-ror-api/
- API con datos de organizaciones de Panamá https://api.dev.ror.org/v2/organizations?query=Panama

Alfabetización de datos | Leer los datos | JSON

JSON, o notación de objetos Javascript, es otro formato de archivo estándar abierto y legible por humanos que se utiliza ampliamente para intercambiar datos. Su extensión es .json.

Ejemplo:

- Página de registros de Organizaciones de investigación https://ror.org/
- Explicación API ROR.org https://ror.org/tutorials/intro-ror-api/
- API con datos de organizaciones de Panamá https://api.dev.ror.org/v2/organizations?query=Panama
- Visor Json con URL de API https://jsonformatter.org/json-viewer



Alfabetización de datos | Leer los datos | CSV

Es uno de los formatos de archivo más utilizados para intercambiar datos son los archivos CSV. CSV significa valores separados por comas y es un formato basado en texto, lo que significa que puedes abrirlo y editarlo con cualquier editor de texto. Los archivos CSV tienen una .csv extensión de archivo.

Los archivos CSV contienen datos en formato tabular, y la primera línea del archivo contiene los nombres de las columnas separados por comas.

Archivo Editar Ver

```
code,name,continent,region,surface_area,independence_year,population,life_expectancy,gnp,gnp_old,local_name,government_form,head_of_state,capital,code2
ABW,Aruba,North America,Caribbean,193,NuLL,103000,78.4,828,793,Aruba,Nonmetropolitan Territory of The Netherlands,Willem-Alexander,129,AW
AFG,Afghanistan,Asia,Southern and Central Asia,652090,1919,22720000,45.9,5976,NuLL,Afganistan/Afqanestan,Islamic Emirate,Mohammad Omar,1,AF
AGO,Angola,Africa,Central Africa,1246700,1975,12878000,38.3,6648,7984,Angola,Republic,Jose Eduardo dos Santos,56,AO
AIA,Anguilla,North America,Caribbean,96,NuLL,8000,76.1,63.2,NuLL,Anguilla,Dependent Territory of the uK,Elisabeth II,62,AI
ALB,Albania,Europe,Southern Europe,28748,1912,3401200,71.6,3205,2500,Shqiperia,Republic,Rexhep Mejdani,34,AL
AND,Andorra,Europe,Southern Europe,468,1278,78000,83.5,1630,NuLL,Andorra,Parliamentary Coprincipality,,55,AD
ANT,Netherlands Antilles,North America,Caribbean,800,NuLL,217000,74.7,1941,NuLL,Nederlandse Antillen,Nonmetropolitan Territory of The Netherlands,Willem-
Alexander,33,AN
ARE,united Arab Emirates,Asia,Middle East,83600,1971,2441000,74.1,37966,36846,Al-Imarat al- Arabiya al-Muttahida,Emirate Federation,Zayid bin Sultan al-Nahayan,65,AE
ARG,Argentina,South America,South America,2780400,1816,37032000,75.1,340238,323310,Argentina,Federal Republic,Fernando de la Rua,69,AR
ARM,Armenia,Asia,Middle East,29800,1991,3520000,66.4,1813,1627,Hajastan,Republic,Robert KotSarjan,126,AM
ASM,American Samoa,Oceania,Polynesia,199,NuLL,68000,75.1,334,NuLL,Amerika Samoa,uS Territory,George W. Bush,54,AS
ATA,Antarctica,Antarctica,Antarctica,13120000,NuLL,0,NuLL,0,NuLL,N/A,Co-administrated,,NuLL,AQ
ATF,French Southern territories,Antarctica,Antarctica,7780,NuLL,0,NuLL,0,NuLL,Terres australes francaises,Nonmetropolitan Territory of France,Jacques Chirac,NuLL,TF
```

Alfabetización de datos | Leer los datos | CSV

Es uno de los formatos de archivo más utilizados para intercambiar datos son los archivos CSV. CSV significa valores separados por comas y es un formato basado en texto, lo que significa que puedes abrirlo y editarlo con cualquier editor de texto. Los archivos CSV tienen una .csv extensión de archivo.

Los archivos CSV contienen datos en formato tabular, y la primera línea del archivo contiene los nombres de las columnas separados por comas.

Archivo Editar Ver

```
code,name,continent,region,surface_area,independence_year,population,life_expectancy,gnp,gnp_old,local_name,government_form,head_of_state,capital,code2
ABW,Aruba,North America,Caribbean,193,NuLL,103000,78.4,828,793,Aruba,Nonmetropolitan Territory of The Netherlands,Willem-Alexander,129,AW
AFG,Afghanistan,Asia,Southern and Central Asia,652090,1919,22720000,45.9,5976,NuLL,Afganistan/Afqanestan,Islamic Emirate,Mohammad Omar,1,AF
AGO,Angola,Africa,Central Africa,1246700,1975,12878000,38.3,6648,7984,Angola,Republic,Jose Eduardo dos Santos,56,AO
AIA,Anguilla,North America,Caribbean,96,NuLL,8000,76.1,63.2,NuLL,Anguilla,Dependent Territory of the uK,Elisabeth II,62,AI
ALB,Albania,Europe,Southern Europe,28748,1912,3401200,71.6,3205,2500,Shqiperia,Republic,Rexhep Mejdani,34,AL
AND,Andorra,Europe,Southern Europe,468,1278,78000,83.5,1630,NuLL,Andorra,Parliamentary Coprincipality,,55,AD
ANT,Netherlands Antilles,North America,Caribbean,800,NuLL,217000,74.7,1941,NuLL,Nederlandse Antillen,Nonmetropolitan Territory of The Netherlands,Willem-
Alexander,33,AN
ARE,united Arab Emirates,Asia,Middle East,83600,1971,2441000,74.1,37966,36846,Al-Imarat al- Arabiya al-Muttahida,Emirate Federation,Zayid bin Sultan al-Nahayan,65,AE
ARG,Argentina,South America,South America,2780400,1816,37032000,75.1,340238,323310,Argentina,Federal Republic,Fernando de la Rua,69,AR
ARM,Armenia,Asia,Middle East,29800,1991,3520000,66.4,1813,1627,Hajastan,Republic,Robert KotSarjan,126,AM
ASM,American Samoa,Oceania,Polynesia,199,NuLL,68000,75.1,334,NuLL,Amerika Samoa,uS Territory,George W. Bush,54,AS
ATA,Antarctica,Antarctica,Antarctica,13120000,NuLL,0,NuLL,0,NuLL,N/A,Co-administrated,,NuLL,AQ
ATF,French Southern territories,Antarctica,Antarctica,7780,NuLL,0,NuLL,0,NuLL,Terres australes francaises,Nonmetropolitan Territory of France,Jacques Chirac,NuLL,TF
```

Alfabetización de datos | Leer los datos | CSV

Ejemplo 3:
Abrir archivo **datos_practica-excel-CSV.xlsx**

Integrar archivos CSV
country-data-CODE-CSV.csv
country-data-continente-CSV

Ver documento de practica
practicas-datos-ordenados-Datawraper.doc

A	B	C	D	E	F
code	Num	name	continent	region	surface_area
ATA	1	Antarctica	Antarctica	Antarctica	13120000
ATF	2	French Southern territories		Antarctica	7780
BVT	3	Bouvet Island		Antarctica	59
HMD	4	Heard Island and McDonald Islands		Antarctica	359
SGS	5	South Georgia and the South Sandwich Islands		Antarctica	3903
AuS	6	Australia		AustraliaåandNewåZealand	7741220
AuS	6	Australia		AustraliaåandNewåZealand	7741220
CCK	7	Cocos (Keeling) Islands		AustraliaåandNewåZealand	14
CCK	7	Cocos (Keeling) Islands		AustraliaåandNewåZealand	14
CXR	8	Christmas Island		AustraliaåandNewåZealand	135
CXR	8	Christmas Island		AustraliaåandNewåZealand	135
NFK	9	Norfolk Island		AustraliaåandNewåZealand	36
NFK	9	Norfolk Island		AustraliaåandNewåZealand	36
NZL	10	New Zealand		AustraliaåandNewåZealand	270534
NZL	10	New Zealand		AustraliaåandNewåZealand	270534
EST	11	Estonia		BalticåCountries	45227
GBR	12	united Kingdom		BalticåCountries	242900
IRL	13	Ireland			70273
LTu	14	Lithuania		BalticåCountries	65301
LVA	15	Latvia		BalticåCountries	64589
ABW	16	Aruba	North America	Caribbean	193
AIA	17	Anguilla		Caribbean	96
ANT	18	Netherlands Antilles		Caribbean	800

Alfabetización de datos | Comunicación de datos

Al comunicar los datos existen diferentes fuentes en las que ellos es posible integrar los metadatos del conjunto de datos para publicación que permiten conocer de que se tratan los datos.

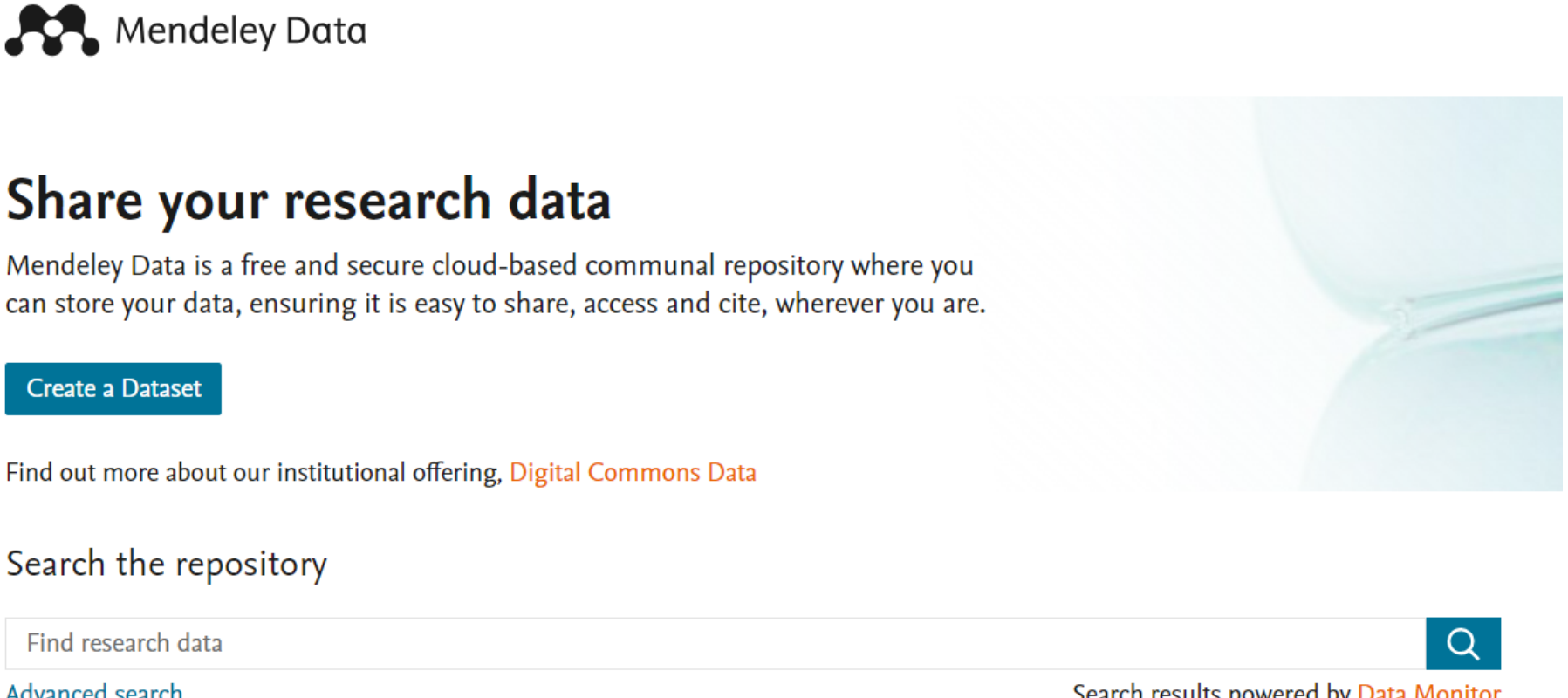
Algunos elementos al comunicar los datos:

- Metadatos (autor, afiliación, diccionario, fechas de creación, resumen de datos)
- Raw data (si es posible)
- Datos limpios
- Tidy data
- Archivo Reame.txt
- Identificador persistente DOI

Alfabetización de datos | Comunicación de datos

https://datasetsearch.research.google.com/

Alfabetización de datos | Comunicación de datos



https://data.mendeley.com/

Bibliografía

- https://www.innovaciondigital360.com/big-data/que-es-la-alfabetizacion-de-datos-data-literacy-y-por-que-tu-empresa-lo-necesita/
- https://runahr.com/mx/recursos/hr-management/que-es-la-alfabetizacion-de-datos-data-literacy/
- https://www.jumpingrivers.com/blog/best-practices-data-cleaning-r/
- https://www.lempert.com.ar/que-es-la-alfabetizacion-de-datos/
- https://chartio.com/learn/charts/how-to-choose-data-visualization/
- https://data.europa.eu/
- https://courses.tranzf.org/course/view.php?id=18
- https://chartio.com/learn/charts/how-to-choose-colors-data-visualization/
- https://www.youtube.com/watch?v=g1urAkB1ozs&t=8s
- https://www.altergeosistemas.com/blog/2014/01/12/normalizacion-datos-openrefine/
- https://openrefine.org/docs/manual/facets

MUCHAS GRACIAS

Twitter @dannymu
danny.murillo@utp.ac.pa
Orcid:0000-0003-0297-7213