SUBSECRETARÍA DE POSGRADO · SECRETARÍA DE ASUNTOS ACADÉMICOS
CURSO DE POSGRADO · JUNIO 2026

Armado y Uso de Base de Datos

para Proyectos de Investigación

Material complementario de lectura · Unidad 1

Dr. Javier Altamirano · Mg. Ana María Chalabe
Facultad de Humanidades y Ciencias Sociales – UNJu

Introducción

Este material de lectura está diseñado como texto complementario a las presentaciones del primer encuentro del curso. El objetivo es que puedan profundizar los conceptos trabajados durante la clase sincrónica, revisarlos a su ritmo y contar con una referencia escrita para las actividades prácticas.

El material cubre los temas centrales de la Unidad 1: qué es una base de datos, cuáles son sus características fundamentales, qué tipos existen, por qué son importantes en la investigación científica y cómo utilizar Excel como herramienta básica para su construcción y gestión. También se incluye una sección sobre principios de organización de datos y buenas prácticas en nomenclatura, y una guía sobre las funciones de Excel más utilizadas en proyectos de investigación.

UNIDAD 01

Introducción a las Bases de Datos

y su importancia en la investigación científica

1. ¿Qué es una base de datos?

En términos simples, una base de datos es una colección organizada de información estructurada, almacenada y accesible de manera sistemática. En el contexto de la investigación científica, podemos definirla como un conjunto de datos relacionados entre sí,

organizados bajo criterios lógicos que permiten registrarlos, gestionarlos y recuperarlos de forma eficiente.

Es importante distinguir una base de datos de una simple lista o de un archivo de texto. Lo que caracteriza a una base de datos es precisamente su estructura: cada dato ocupa un lugar definido, tiene un tipo establecido (numérico, texto, fecha, etc.) y se relaciona con otros datos de maneras predecibles y controladas. Esta estructura es lo que posibilita no solo almacenar información, sino también analizarla, filtrarla, cruzarla y comunicarla.

> *Una base de datos bien diseñada no es simplemente un depósito de datos: es una herramienta activa de investigación. La manera en que organizamos nuestros datos determina, en gran medida, las preguntas que podremos responder con ellos.*

1.1 Funciones principales de una base de datos

Toda base de datos cumple tres funciones esenciales que justifican su uso en proyectos de investigación:

Almacenar. La función más básica: guardar grandes volúmenes de datos en un lugar centralizado y organizado. A diferencia de las notas dispersas o los registros en papel, una base de datos permite concentrar toda la información de un proyecto en un único sistema coherente, facilitando su acceso y reduciendo el riesgo de pérdida.

Ordenar. Los datos no solo se guardan: se estructuran. Cada dato se ubica en un lugar lógico dentro de la estructura, lo que permite encontrar y relacionar información rápidamente. La posibilidad de ordenar, filtrar y agrupar datos es lo que transforma una colección de registros en una herramienta analítica.

Buscar y analizar. Una vez que los datos están organizados, es posible recuperar subconjuntos específicos y aplicar técnicas de análisis para extraer conclusiones. Esto incluye desde búsquedas simples ("todos los casos con diagnóstico X") hasta análisis estadísticos complejos (correlaciones, distribuciones, comparaciones entre grupos).

2. Características de una base de datos

No todos los sistemas de almacenamiento de datos son bases de datos en sentido estricto. Una verdadera base de datos se distingue por un conjunto de características que garantizan la calidad y la utilidad de la información que contiene:

Organización

Los datos se almacenan en una estructura definida: tablas con filas y columnas, donde cada columna tiene un nombre y un tipo de dato establecido. Esta organización no es estética sino funcional: permite que el sistema —y el investigador— sepa exactamente dónde buscar cada tipo de información. Una base de datos desorganizada no es una base de datos: es un repositorio de datos potencialmente incorrectos.

Redundancia mínima

En una base de datos bien diseñada, cada dato se registra una sola vez. Esto puede parecer un detalle técnico, pero tiene consecuencias profundas para la investigación: si el mismo dato aparece en múltiples lugares con valores diferentes, ¿cuál es el correcto? La redundancia genera inconsistencias que comprometen la validez de los resultados. El principio de "no te repitas" (conocido en inglés como DRY: Don't Repeat Yourself) es central en el diseño de bases de datos.

Consistencia

Los datos deben mantenerse coherentes en toda la base. Esto significa que las categorías, códigos y criterios de registro se aplican de la misma manera en todos los registros. Por ejemplo, si se registra el sexo biológico de los participantes, debe usarse siempre la misma nomenclatura: no mezclar "M" con "Masculino" o "masculino" en distintos registros. La inconsistencia en los valores es una de las fuentes más frecuentes de errores en el análisis de datos.

Acceso eficiente

Los datos pueden recuperarse de forma rápida usando filtros, consultas y búsquedas. En una base de datos bien estructurada, encontrar todos los registros que cumplen determinada condición (por ejemplo, "todos los participantes mayores de 60 años del departamento X") es una operación casi instantánea. Esto no solo ahorra tiempo: habilita tipos de análisis que serían imposibles con sistemas de registro manual.

Seguridad y control de acceso

Los datos se protegen mediante mecanismos que determinan quién puede ver, editar o eliminar registros. En investigación, esto es especialmente relevante cuando se trabaja con datos sensibles (datos de salud, datos de menores, datos personales). El control de acceso también protege contra modificaciones accidentales que podrían comprometer la integridad del dataset.

3. Tipos de bases de datos en investigación científica

Existen muchos tipos de bases de datos, diferenciados por su estructura, su modelo de datos y los sistemas que las gestionan. Para los proyectos de investigación, los tipos más relevantes son tres: las bases de datos tabulares (o planas), las relacionales y las longitudinales. Cada una responde a necesidades diferentes y tiene sus propias ventajas y limitaciones.

3.1 Base de datos tabular o plana

Es la forma más común y accesible de base de datos en investigación. Su estructura es simple: una tabla de dos dimensiones donde las filas representan observaciones o casos (personas, eventos, documentos, instituciones) y las columnas representan variables o atributos de esas observaciones. Cada celda contiene el valor de una variable para un caso específico.

Esta estructura se corresponde directamente con lo que los estadísticos llaman una "matriz de datos": filas × columnas. Es el formato nativo de herramientas como Excel, Google Sheets, SPSS o Stata. Es también el formato esperado por la mayoría de los procedimientos de análisis estadístico.

Ejemplo típico: una encuesta aplicada a 200 personas con 30 preguntas genera una base de datos tabular de 200 filas (una por encuestado) y 30 columnas (una por pregunta). Cada fila es un caso único; cada columna, una variable.

> *La base de datos tabular es el punto de partida de la mayoría de los proyectos de investigación. Dominar su diseño y gestión es la competencia más valiosa que puede desarrollar un investigador en formación.*

3.2 Base de datos relacional

Las bases de datos relacionales van un paso más allá: en lugar de una única tabla, utilizan múltiples tablas vinculadas entre sí mediante campos comunes llamados "claves". Este diseño permite evitar la redundancia y gestionar datos complejos con mayor coherencia y eficiencia.

Por ejemplo, en un estudio sobre escuelas y estudiantes, una base de datos relacional tendría una tabla de escuelas (con datos sobre cada institución) y una tabla de estudiantes (con datos sobre cada alumno), vinculadas por un identificador de escuela. Así, los datos de la institución no se repiten para cada estudiante: se almacenan una vez y se vinculan cuando es necesario.

Las bases de datos relacionales son gestionadas por sistemas especializados como Microsoft Access, MySQL, PostgreSQL o SQLite. Si bien su diseño es más complejo que el de las bases tabulares, ofrecen ventajas significativas en proyectos con datos interrelacionados, grandes volúmenes de información o múltiples usuarios.

3.3 Base de datos longitudinal

Una base de datos longitudinal registra las mismas unidades de análisis (personas, organizaciones, territorios) en múltiples momentos en el tiempo. A diferencia de las bases transversales (que capturan un único momento), las bases longitudinales permiten analizar cambios, trayectorias y procesos.

Este tipo de base de datos es fundamental en estudios de seguimiento, cohortes, paneles y evaluaciones de impacto. Su diseño presenta desafíos particulares: cómo identificar inequívocamente a cada unidad a lo largo del tiempo, cómo gestionar los datos faltantes en algunas mediciones, y cómo estructurar la información para facilitar el análisis de cambios.

Ejemplo: un estudio sobre trayectorias educativas que sigue a 500 estudiantes durante 5 años, registrando su situación académica una vez por año. El resultado es una base de datos donde cada estudiante aparece 5 veces (una por año de seguimiento), con variables que indican el momento de medición.

4. Importancia de las bases de datos en la investigación científica

El manejo adecuado de bases de datos no es un detalle técnico de la investigación: es una competencia metodológica central. La calidad de los datos determina, en última instancia, la calidad de las conclusiones. Un análisis sofisticado aplicado sobre datos mal organizados produce resultados inválidos. Inversamente, datos bien organizados pueden revelarse mucho más ricos de lo que parecían cuando se analizan con las herramientas adecuadas.

A continuación se detallan cinco dimensiones en las que el manejo de bases de datos impacta directamente en la calidad de la investigación:

4.1 Eficiencia en la gestión de datos

Una base de datos bien construida permite manejar volúmenes de información que serían imposibles de gestionar manualmente. En lugar de buscar registros en archivos físicos o documentos dispersos, el investigador puede filtrar, ordenar y recuperar información en

segundos. Esto no solo ahorra tiempo: libera capacidad cognitiva para concentrarse en la interpretación de los resultados en lugar de en la gestión de la información.

Además, muchas tareas repetitivas (verificar duplicados, identificar valores atípicos, calcular frecuencias) pueden automatizarse usando funciones y fórmulas, lo que reduce drásticamente el riesgo de errores humanos.

4.2 Mejora en la calidad de la investigación

Los datos bien organizados permiten realizar análisis más precisos y rigurosos. Cuando la información está estructurada de manera consistente, es posible detectar patrones, tendencias y relaciones que serían invisibles en datos dispersos o inconsistentes. Esto habilita la aplicación de técnicas estadísticas más sofisticadas y la formulación de hipótesis más precisas.

La calidad también se relaciona con la reproducibilidad: una base de datos bien documentada permite que otros investigadores —o el mismo investigador en el futuro— repliquen exactamente los pasos del análisis y verifiquen los resultados. Este principio es fundamental en la ciencia contemporánea.

4.3 Facilitación de la colaboración

En equipos de investigación, una base de datos centralizada y bien estructurada permite que múltiples personas trabajen con la misma información sin riesgo de crear versiones incompatibles. Cuando los datos están organizados con criterios claros y documentados (diccionario de datos, libro de códigos), cualquier miembro del equipo puede incorporarse al proyecto y comprender rápidamente la estructura de la información.

Esto también facilita la colaboración entre proyectos e instituciones: compartir una base de datos bien diseñada es mucho más eficiente que compartir notas o tablas ad hoc que solo el autor original puede interpretar.

4.4 Transparencia y reproducibilidad

Una de las demandas más importantes de la ciencia contemporánea es la transparencia en los procesos de investigación. Una base de datos bien documentada —con metadatos claros, diccionario de variables, registro de modificaciones y criterios de codificación explícitos— permite que cualquier interesado comprenda exactamente qué datos se usaron, cómo se recolectaron y cómo se procesaron.

Esta transparencia es la base de la reproducibilidad: la posibilidad de que otro investigador, con los mismos datos, llegue a los mismos resultados siguiendo los mismos procedimientos. La reproducibilidad es actualmente un criterio central en la evaluación de la calidad científica.

4.5 Preservación y accesibilidad de los datos

Los datos de investigación representan un activo valioso que puede tener utilidad mucho más allá del proyecto original. Una base de datos bien construida y documentada puede ser reutilizada en estudios secundarios, comparaciones históricas o síntesis meta-analíticas. Sin embargo, esto solo es posible si los datos están almacenados en formatos accesibles, con documentación suficiente para que puedan ser interpretados sin la mediación del investigador original.

Esto también implica mantener copias de respaldo (backups), definir políticas de acceso claras y contemplar el ciclo de vida de los datos desde el inicio del proyecto, no como un paso final.

5. Excel como herramienta para la gestión de datos de investigación

Microsoft Excel no es, en sentido estricto, un sistema gestor de bases de datos. No cuenta con los mecanismos formales de control de integridad, gestión de múltiples usuarios concurrentes o lenguaje de consultas (SQL) que caracterizan a los sistemas de bases de datos propiamente dichos. Sin embargo, para la gran mayoría de los proyectos de investigación, Excel es una herramienta completamente adecuada y tiene ventajas significativas: es accesible, flexible, ampliamente conocida y suficientemente poderosa para manejar bases de datos de tamaño mediano con análisis de complejidad moderada.

5.1 Capacidades y ventajas

Excel puede almacenar hasta 1.048.576 filas y 16.384 columnas por hoja de cálculo, lo que es más que suficiente para la mayoría de los proyectos de investigación. Sus funciones de análisis cubren estadística descriptiva, tablas dinámicas, gráficos, filtros avanzados, validación de datos y automatización mediante macros. La herramienta Power Query, integrada en las versiones modernas de Excel, permite importar, transformar y consolidar datos de múltiples fuentes con una interfaz visual sin necesidad de programar.

Entre sus principales ventajas para la investigación se destacan: la curva de aprendizaje relativamente baja, la compatibilidad con prácticamente cualquier software estadístico (R,

SPSS, Stata, Python), la facilidad para compartir archivos y la integración con el ecosistema Microsoft 365.

5.2 Limitaciones a tener en cuenta

Es igualmente importante conocer las limitaciones de Excel para usarlo con criterio:

- No está diseñado para el trabajo simultáneo de múltiples usuarios sobre el mismo archivo, lo que puede generar versiones conflictivas en equipos de investigación.
- No tiene mecanismos formales de integridad referencial, lo que significa que los errores de carga no se detectan automáticamente salvo que se configuren validaciones específicas.
- Es vulnerable a modificaciones accidentales: sin protección adecuada, es fácil borrar o alterar datos sin dejar registro del cambio.
- Los archivos de Excel pueden corromperse o perder datos si se cierra inesperadamente o se trabaja con versiones incompatibles.
- No es el formato más adecuado para datos sensibles sin medidas adicionales de seguridad.

5.3 Google Sheets como alternativa

Google Sheets es una alternativa gratuita y basada en la nube que comparte con Excel la mayoría de las funcionalidades básicas. Su principal ventaja es la colaboración en tiempo real: múltiples usuarios pueden editar el mismo documento simultáneamente, y los cambios se registran con historial de versiones. Esto la hace especialmente útil en equipos de investigación que trabajan de forma distribuida.

La siguiente tabla resume las principales diferencias entre ambas herramientas:

Característica	Excel	Google Sheets
Costo	De pago (Microsoft 365)	Gratuito
Instalación	Local (escritorio)	Basado en la nube
Capacidad máxima	1.048.576 filas × 16.384 col.	5.000.000 celdas aprox.
Colaboración	Limitada (requiere OneDrive)	En tiempo real, nativa
Funciones avanzadas	Más amplias (Power Query, VBA)	Limitadas pero suficientes
Uso en este curso	Herramienta principal	Alternativa válida

Para este curso utilizaremos Excel como herramienta principal, dado que cuenta con mayor potencia analítica y es el estándar en la mayoría de los entornos académicos e institucionales. Sin embargo, todas las actividades prácticas pueden realizarse también en Google Sheets, salvo aquellas específicamente relacionadas con Power Query.

6. Principios de organización de datos

Uno de los errores más frecuentes en la gestión de datos de investigación es confundir "tener los datos en una planilla" con "tener una base de datos bien organizada". Muchas planillas de Excel que circulan en entornos académicos e institucionales no son bases de datos en ningún sentido útil: son registros ad hoc, estructurados para ser leídos por humanos pero no para ser procesados por software.

El concepto de Tidy Data ("datos ordenados"), desarrollado por el estadístico Hadley Wickham, ofrece un conjunto de principios simples pero poderosos para organizar datos de manera que sean útiles tanto para el análisis automatizado como para la lectura humana. Estos principios son el fundamento de cualquier base de datos bien diseñada.

6.1 Filas = observaciones

Cada fila de la base de datos debe representar una única unidad de análisis: una persona, un caso, un evento, una localidad, un momento de medición. Esta unidad debe ser consistente a lo largo de toda la tabla.

El error más común en este punto es mezclar diferentes tipos de unidades en la misma tabla (por ejemplo, incluir en la misma planilla registros de personas y registros de instituciones), o usar filas para información que no corresponde a observaciones (filas de totales, filas de subtítulos, filas en blanco separadoras).

> *Regla práctica: si no puede responder en una sola oración "cada fila representa un/a ___", la estructura de su base de datos necesita revisión.*

6.2 Columnas = variables

Cada columna debe contener un único tipo de información: la edad de los participantes, su sexo, la fecha de registro, el valor de una medición, etc. Una columna no puede contener dos tipos de información distintos.

Los errores más frecuentes en este punto son: combinar en una sola celda información que debería estar en columnas separadas (por ejemplo, "nombre y apellido" en lugar de "nombre" y "apellido" en columnas separadas), y usar múltiples columnas para lo que en realidad es una sola variable categórica (por ejemplo, una columna por categoría en lugar de una única columna con la categoría como valor).

6.3 Celdas = valores únicos

Cada celda de la base de datos debe contener un único valor, del tipo correspondiente a la variable de esa columna. No deben incluirse fórmulas ni cálculos en las celdas de la tabla de datos (los cálculos se realizan fuera de la tabla, en otras columnas o en hojas separadas), ni celdas fusionadas, ni valores múltiples separados por comas dentro de una misma celda.

Este principio parece obvio, pero con frecuencia se viola en la práctica. Por ejemplo, registrar en la columna "diagnóstico" el valor "Hipertensión, Diabetes tipo 2" genera una celda con dos valores que el software interpretará como una sola cadena de texto, haciendo imposible filtrar o contar los casos de cada diagnóstico por separado.

7. Buenas prácticas en nomenclatura y estructura

La organización lógica de los datos (filas, columnas, celdas) es necesaria pero no suficiente. También es fundamental seguir convenciones de nomenclatura que hagan que la base de datos sea legible, procesable por software y comprensible para cualquier investigador que la use, incluido el propio autor tiempo después.

7.1 Nombres de variables

Los encabezados de las columnas (los nombres de las variables) deben seguir reglas específicas para ser compatibles con el mayor número posible de herramientas de análisis:

- Sin espacios: usar guion bajo (fecha_nacimiento) o camelCase (fechaNacimiento) como separador.
- Sin tildes, eñes ni caracteres especiales: no usar ñ, á, é, í, ó, ú, ü, ni ningún símbolo especial.
- Descriptivos pero cortos: el nombre debe indicar claramente qué mide la variable, pero sin ser tan largo que dificulte la lectura. Preferir sexo_biologico a sb o a sexo_biologico_del_participante.
- Consistentes: usar el mismo estilo en todos los nombres (todos en minúscula con guion bajo, o todos en camelCase, pero no mezclar).
- Únicos: no repetir el mismo nombre en dos columnas diferentes.

Ejemplos correctos: edad, sexo, fecha_nac, nivel_educ, ingreso_mensual, cod_localidad.

Ejemplos incorrectos: Edad del participante, sexo biológico, FECHA DE NAC., col. B, dato1.

7.2 Valores consistentes

La consistencia en los valores de las variables es tan importante como la consistencia en los nombres. Antes de comenzar a cargar datos, es necesario definir exactamente qué categorías

o valores se usarán para cada variable y documentarlos en el diccionario de datos. Una vez establecidos, esos valores deben usarse sin variación en todos los registros.

Por ejemplo, si se usa la categoría "Femenino" para una variable de sexo, no pueden aparecer en la misma columna variantes como "F", "femenino", "Fem.", "FEMENINO" o "Mujer". Aunque para un lector humano todas estas formas sean equivalentes, para el software son cadenas de texto distintas que se contarán como categorías diferentes.

7.3 Estructura del archivo

Más allá de las filas y columnas individuales, hay convenciones sobre la estructura general del archivo que facilitan enormemente el procesamiento de datos:

- Una sola fila de encabezado, en la primera fila: no usar dos filas de encabezado, no dejar la primera fila vacía, no usar encabezados fusionados.
- Sin filas ni columnas vacías dentro del rango de datos: las celdas vacías dentro de una tabla son ambiguas (¿valor faltante? ¿separador? ¿error?). Si un valor es faltante, debe indicarse explícitamente con NA o con el código establecido en el diccionario de datos.
- Sin filas de totales, subtotales o promedios intercaladas en la tabla de datos: esas operaciones se realizan fuera de la tabla o en una hoja separada.
- Sin formatos decorativos que codifiquen información: no usar el color de celda para indicar una categoría, porque esa información se pierde al exportar a otros formatos.

7.4 Identificación de datos faltantes

Los datos faltantes son inevitables en cualquier proyecto de investigación real. Lo importante es manejarlos de manera explícita y consistente. Algunas convenciones comunes:

- NA (Not Available): convención estándar en estadística, compatible con R, Python y la mayoría de los software de análisis.
- 999 o -9: convención histórica en ciencias sociales, usada en bases de datos grandes. Debe documentarse claramente en el diccionario de datos para no confundirse con valores reales.
- Celda vacía: técnicamente válido pero desaconsejado, ya que es ambiguo (puede confundirse con un error de carga).

Lo más importante es ser consistente: elegir una convención y aplicarla en toda la base de datos, documentándola en el diccionario de datos.

8. Funciones esenciales de Excel para la gestión de datos

Excel ofrece cientos de funciones, pero para la gestión y el análisis básico de bases de datos de investigación, un conjunto reducido de funciones cubre la mayoría de las necesidades. A

continuación se describen las cinco funciones más útiles en este contexto, con ejemplos aplicados a situaciones de investigación.

8.1 BUSCARV

BUSCARV (o VLOOKUP en inglés) busca un valor en la primera columna de un rango y devuelve el valor correspondiente de otra columna del mismo rango. Es la función más útil para cruzar información de dos tablas sin necesidad de hacerlo manualmente.

Sintaxis: =BUSCARV(valor_buscado; rango_de_búsqueda; número_de_columna; [exacto])

`=BUSCARV(A2,TablaRef,3,0)`	Busca el valor de la celda A2 en la primera columna de TablaRef y devuelve el valor de la tercera columna. El 0 indica coincidencia exacta.

Ejemplo de uso en investigación: tenemos una base de datos de participantes con su código de localidad, y otra tabla con los nombres y provincias correspondientes a cada código. BUSCARV permite agregar automáticamente el nombre de la localidad y la provincia a cada registro de participante, sin copiar y pegar manualmente.

Limitación importante: BUSCARV solo busca de izquierda a derecha. Si necesita buscar de derecha a izquierda, use la combinación INDICE+COINCIDIR, o la función BUSCARX disponible en versiones más recientes de Excel.

8.2 CONTAR.SI

CONTAR.SI cuenta el número de celdas en un rango que cumplen una condición específica. Es la función más directa para calcular frecuencias absolutas de variables categóricas.

Sintaxis: =CONTAR.SI(rango; criterio)

`=CONTAR.SI(B:B,"Femenino")`	Cuenta cuántas celdas en la columna B contienen exactamente el texto "Femenino".

Ejemplo de uso en investigación: en una base de datos de 500 participantes, necesitamos saber cuántos son de sexo femenino, cuántos tienen nivel educativo universitario, cuántos provienen de áreas rurales. CONTAR.SI permite calcular estas frecuencias en segundos.

Variante: CONTAR.SI.CONJUNTO (COUNTIFS) permite aplicar múltiples criterios simultáneamente: "¿cuántos participantes son mujeres Y tienen nivel universitario Y viven en el departamento X?"

8.3 SI

La función SI (IF) evalúa una condición lógica y devuelve un valor si se cumple y otro si no se cumple. Es fundamental para crear nuevas variables a partir de variables existentes, o para clasificar y etiquetar registros automáticamente.

Sintaxis: =SI(condición; valor_si_verdadero; valor_si_falso)

`=SI(C2>=18,"Adulto","Menor")`	Devuelve "Adulto" si el valor de C2 es mayor o igual a 18, y "Menor" en caso contrario.

Ejemplo de uso en investigación: tenemos la edad de los participantes y necesitamos crear una variable categórica que los clasifique en grupos etarios. O tenemos los ingresos y queremos crear una variable dicotómica que indique si el ingreso supera o no la línea de pobreza. La función SI permite hacer estas transformaciones de forma sistemática y replicable.

Las funciones SI pueden anidarse para crear clasificaciones con más de dos categorías: =SI(C2<18,"Menor",SI(C2<65,"Adulto","Adulto mayor")).

8.4 CONCATENAR / &

La función CONCATENAR (o el operador &) une el contenido de varias celdas en una sola cadena de texto. Es útil para crear identificadores únicos, unir campos que estaban separados, o generar etiquetas compuestas.

`=A2&"_"&B2`	Une el contenido de A2, el texto "_" y el contenido de B2. Ejemplo: "García" & "_" & "Juan" = "García_Juan".

Ejemplo de uso en investigación: en una base de datos donde las personas están identificadas por apellido y número de documento por separado, CONCATENAR permite crear un identificador único combinando ambos campos. También es útil para crear etiquetas descriptivas para los gráficos, o para unir variables de fecha que vienen en columnas separadas (día, mes, año).

8.5 Texto a columnas

Texto a columnas no es una función en el sentido estricto, sino una herramienta del menú Datos que divide el contenido de una celda en varias columnas usando un delimitador definido (coma, punto y coma, espacio, tabulación u otro carácter).

Es especialmente útil cuando se importan datos de fuentes externas (sistemas de información, exportaciones de formularios, archivos CSV) donde múltiples valores vienen concatenados en una sola celda.

Ejemplo de uso en investigación: una exportación de un sistema de salud incluye fecha y hora en una sola celda ("01/06/2024 14:30") pero necesitamos la fecha y la hora en columnas separadas. Texto a columnas permite hacer esta separación en segundos para toda la base.

Cómo acceder: pestaña Datos → grupo Herramientas de datos → Texto en columnas. El asistente guía paso a paso la configuración del delimitador y el formato de las columnas resultantes.

Referencias bibliográficas

Broman, K. W. y Woo, K. H. (2018). Data organization in spreadsheets. The American Statistician, 72(1), 2-10. https://doi.org/10.1080/00031305.2017.1375989

Marín-Arraiza, P., Puerta-Díaz, M. y Gregorio-Vidotti, S. (2019). Gestión de datos de investigación y bibliotecas: preservando los nuevos bienes científicos. Hipertext.net, (19), 13-31. https://doi.org/10.31009/hipertext.net.2019.i19.02

Melero, R. (2018). Recomendaciones para la gestión de datos de investigación dirigidas a investigadores. MareData: Red Española sobre Datos de Investigación en Abierto.

Valdés-Miranda, C. (2022). Excel 2022. Madrid: Anaya Multimedia.

Wickham, H. (2014). Tidy data. Journal of Statistical Software, 59(10), 1-23. https://doi.org/10.18637/jss.v059.i10