



VII Jornadas Técnicas AEFP de Valoración

3 y 4 de octubre de 2024

Valorando datos y series datificadas

Pilar Campos Martínez

Jordi Serra Serra

To see collections as data begin with reframing all digital objects as data.

Padilla, Thomas (2017). *On a Collections as Data Imperative*.

The background features a stylized illustration of a globe with blue and yellow waves. A large, semi-transparent blue 'V' shape is centered over the globe. Below the globe, there are several sheets of paper or data tables, some of which are partially visible and appear to be floating or falling. The overall color palette is dominated by blue and yellow/gold tones.

Dato, dataset y serie datificada

Del documento a los datos...

¿un paso más en la transformación digital?

Conceptos

¿De qué estamos hablando cuando hablamos de datos?

Aclaración: estas son solo algunas de las definiciones que existen y son las que se usarán en esta presentación.

Conceptos

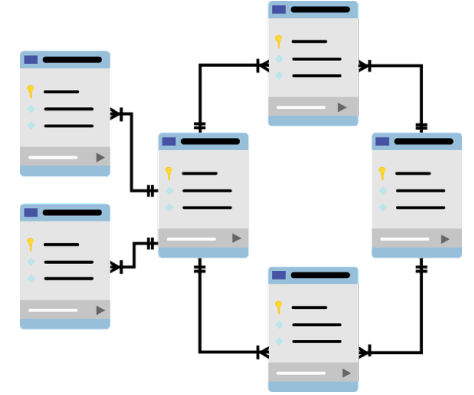
Dato (inglés *data item*): Ítem que se considera indivisible de un contexto concreto. Es la unidad de datos para la cual la definición, identificación, representación y valores permitidos se especifican mediante un conjunto de atributos (metadatos). (ISO/IEC, 1994)

Ejemplo: 04-10-2024

Datos (inglés *data*): Representación interpretable de información de manera adecuada para la comunicación, la interpretación y el procesamiento. (ISO/IEC, 1993).

Ejemplo: Fecha de la VII Jornadas Técnicas de AEFP: 04-10-2024

Conceptos



Base de datos (inglés *database*): 1. Colección de datos con una estructura para incorporar, almacenar, proveer o pedir datos para múltiples usuarios. 2. Colección de datos interrelacionados organizados según un esquema de bases de datos para dar servicio a una o más aplicaciones (Butterfield et al., 2016)

Para definir, manipular y acceder a los datos de las bases de datos, así como para mantener su integridad o seguridad se hace mediante un **Sistema de Gestión de Base de Datos (SGBD)** que actúa como “interfaz” entre el usuario y la base de datos.

Aclaración: en esa presentación y para simplificar lo trataremos como un conjunto bajo el nombre de ‘bases de datos’ a no ser que se indique lo contrario.

Conceptos

Conjunto de datos (inglés *dataset*): Conjunto de datos estructurados creados para un propósito específico. Los conjuntos de datos se pueden almacenar, administrar y publicar en una variedad de formatos y tecnologías (The National Archives, 2011)

Conceptos

Documento (inglés *record*): Se entiende por documento toda expresión en lenguaje oral, escrito, de imagen o sonidos, natural o codificado, recogida en cualquier tipo de soporte material, y cualquier otra expresión gráfica que constituya un testimonio de las funciones y actividades sociales del hombre y de los grupos humanos, con exclusión de las obras de investigación y de creación (Llei 9/1993 del Patrimoni cultural català).

¿Conjunto de datos Vs Documento?

Conceptos

Documento (inglés *record*): Se entiende por documento toda expresión en lenguaje oral, escrito, de imagen o sonidos, natural o codificado, recogida en cualquier tipo de soporte material, y cualquier otra expresión gráfica que constituya un testimonio de las funciones y actividades sociales del hombre y de los grupos humanos, con exclusión de las obras de investigación y de creación (Llei 9/1993 del Patrimoni cultural català).

¿Conjunto de datos Vs Documento?

“Data record”: una colección de datos relacionados tratados como unidad, como los campos de una fila de una tabla de bases de datos. (SAA, 2005-2022)

¿Un dataset es documento, expediente, colección, serie...?

Conceptos

Pongamos un ejemplo...

Ejemplo 1: Libro actas de calificación

- Listas de alumnos + notas
- Firmadas por el profesor
- Encuadernadas anualmente
- 1 hoja por asignatura / profesor
- 1 libro = 1 expediente

Libro en papel

59

UNIVERSIDAD
DE
BARCELONA.

Facultad de *Jurisprudencia* Año *5^o*

Barcelona *6* de *Junio* de 1856

En el día de hoy, ante los Sres. Jueces notados al margen, se han celebrado los exámenes ordinarios de los alumnos que á continuación se espresan.

SEÑORES JUECES.	NOMBRES Y APELLIDOS.	CALIFICACIONES.
<i>co. J. Raguil</i>	<i>D. Manuel Lloix y Kneller</i>	<i>hechano</i>
<i>co. J. Raguil</i>	<i>D. Federico Lloix y Kneller</i>	<i>hechano</i>
<i>co. J. Raguil</i>	<i>D. Franco Reina y Alina</i>	<i>Bueno</i>
	<i>D. Romulo Moragas</i>	<i>hechano</i>
	<i>D. Felipe Botana y de Saut</i>	<i>Bueno</i>
	<i>D. Jose de S. y de Badal</i>	<i>Bueno</i>
	<i>D. Miguel Salas</i>	<i>hechano</i>
	<i>D. Jose de S. y de Saut</i>	<i>hechano</i>
	<i>D. Jose Bahiguer</i>	<i>Bueno</i>
	<i>D. Leopoldo Andaguer</i>	<i>hechano</i>
	<i>D. Similiano Agui y Saut</i>	<i>Bueno</i>
	<i>D. Pablo Riquelme y Saut</i>	<i>Bueno</i>

D. J. Raguil

J. Raguil

J. Raguil

Ejemplo 1: Libro actas de calificación

- Listas de alumnos + notas
- Firmadas por el profesor
- Encuadernadas anualmente
- 1 hoja por asignatura / profesor
- 1 libro = 1 expediente

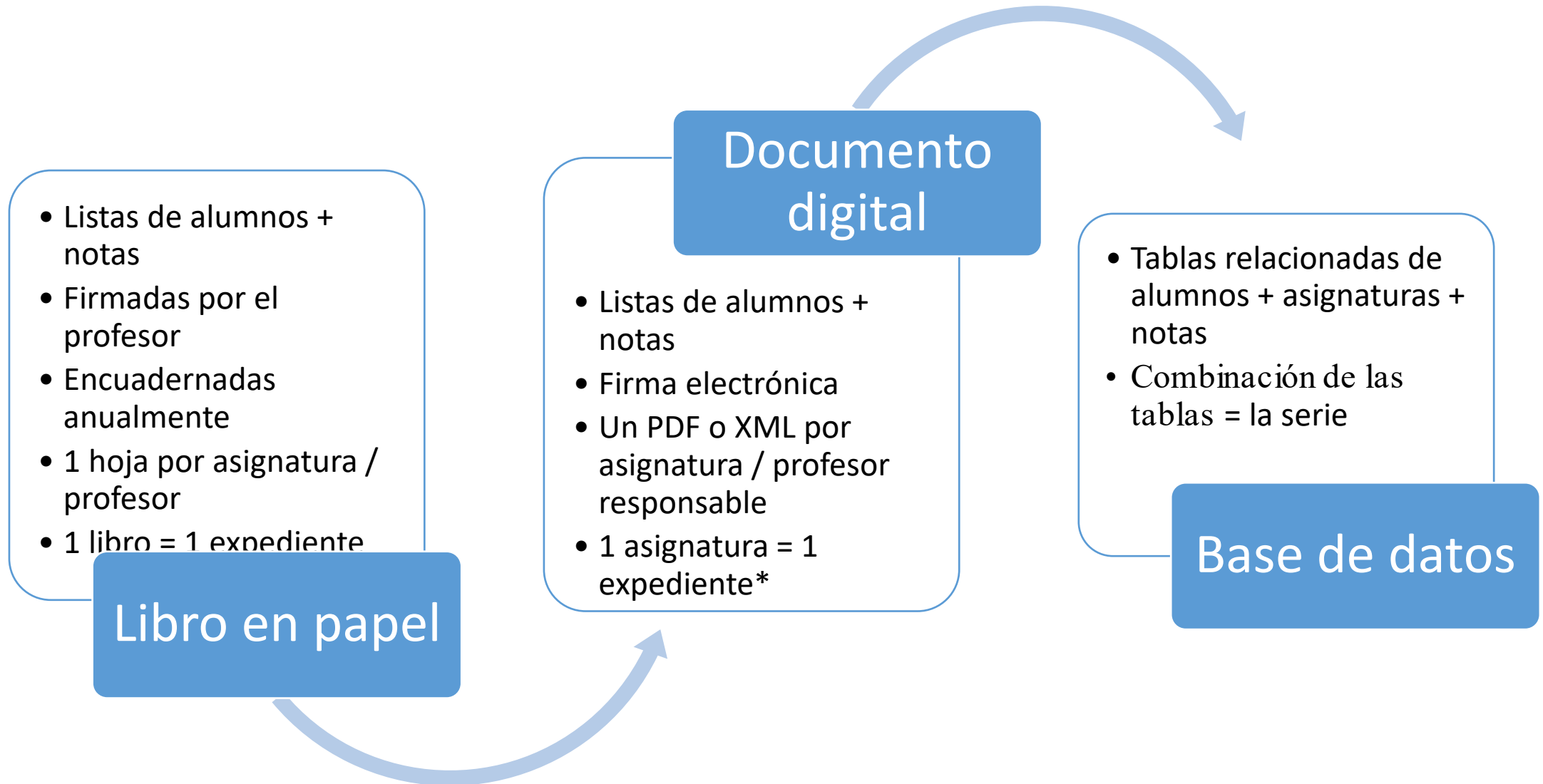
Libro en papel

Documento digital

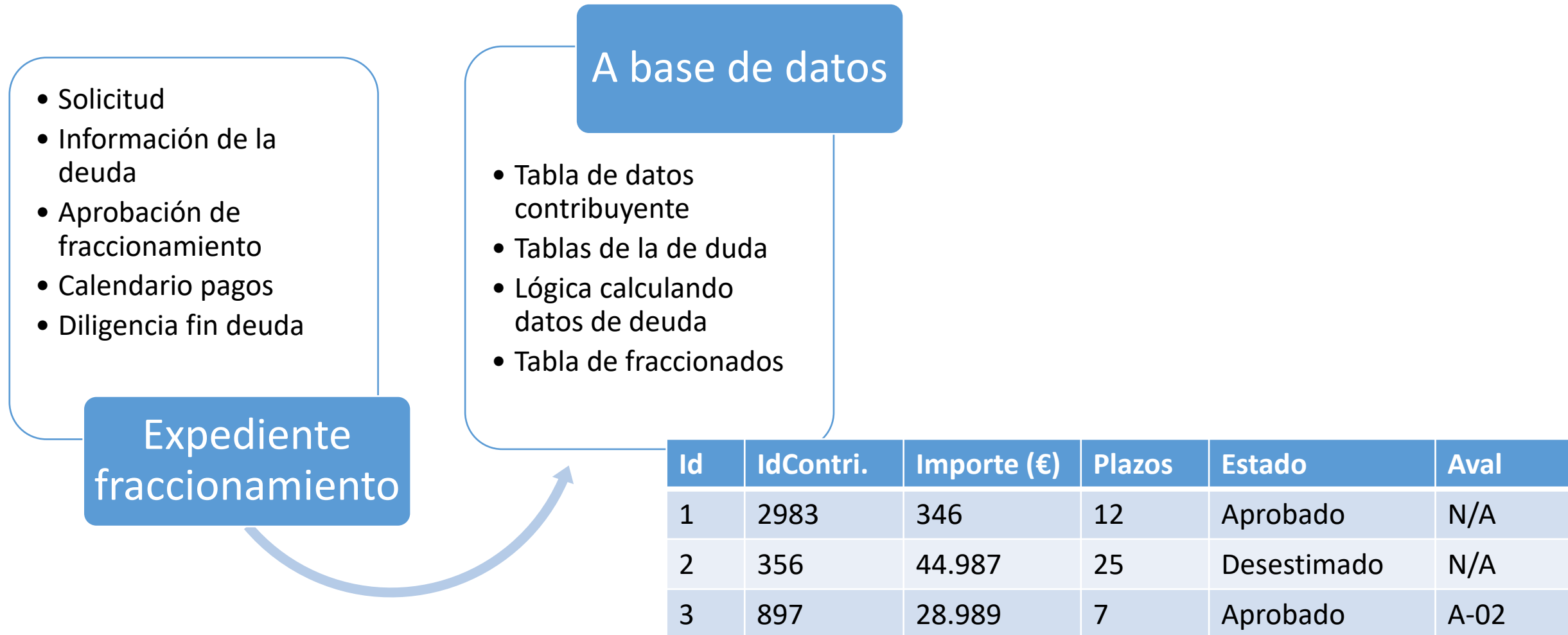
- Listas de alumnos + notas
- Firma electrónica
- Un PDF o XML por asignatura / profesor responsable
- 1 asignatura = 1 expediente*

```
<bc_assignatura id="36005/">Gestio Documental a les Organitzacions</bc_assignatura>
<bc_curs id="2010">Curs 2010/2011</bc_curs>
<bc_timat id="0">Oficial </bc_timat>
<bc_quadrimestre id="2">2</bc_quadrimestre>
<bc_periode id="1">1</bc_periode>
<bc_ensenyament id="*****">Tots els ensenyaments</bc_ensenyament>
<bc_grup id="M1">M1</bc_grup>
<bc_incompatibles id="*">Tots</bc_incompatibles>
<bc_lliure_eleccio id="*">Tots</bc_lliure_eleccio>
<bc_annex_acta id="00">Acta Original</bc_annex_acta>
<bc_convocatoria id="1">1</bc_convocatoria>
<bc_tipus_qual id="N">NUMERICA</bc_tipus_qual>
<bc_acta_matricula id="1">DRET A DOCENCIA I EXAMEN</bc_acta_matricula>
<bc_numero_alumnes>20</bc_numero_alumnes>
<bc_ensenyament_gestio id="1">G1030</bc_ensenyament_gestio>
<bc_linia00>G10300000002010-20110540200 Segon 0</bc_linia00>
<bc_assig id="360057">Gestió Documental a</bc_assig>
<bc_uga id="0540200">FAC.BIBLIOT.I DOCUM.</bc_uga>
</bloc_condicions>
<!-- Bloc: ENSENYAMENTS_ACTA -->
<bc_ensenyaments>
  <bc_ensenyament id="G1030">Informació i Documen</bc_ensenyament>
</bc_ensenyaments>
<!-- Bloc: SIGNATURA_ACTA -->
<bc_signatures>
  <bs_centre id="0540200">FACULTAT DE BIBLIOTECONOMIA I DOCUMENTACIO</bs_centre>
  <bs_departament id="0540200">FACULTAT DE BIBLIOTECONOMIA I DOCUMENTACIO</bs_departament>
  <bs_professor id="          , JORDI      </bs_professor>
</bc_signatures>
<!-- Bloc: ALUMNES_ACTA -->
<bc_alumnes_acta>
  <bloc_alumne id="          ">
    <ba_numOrdre>00001</ba_numOrdre>
    <ba_niub>14      </ba_niub>
```

Ejemplo 1: Libro actas de calificación

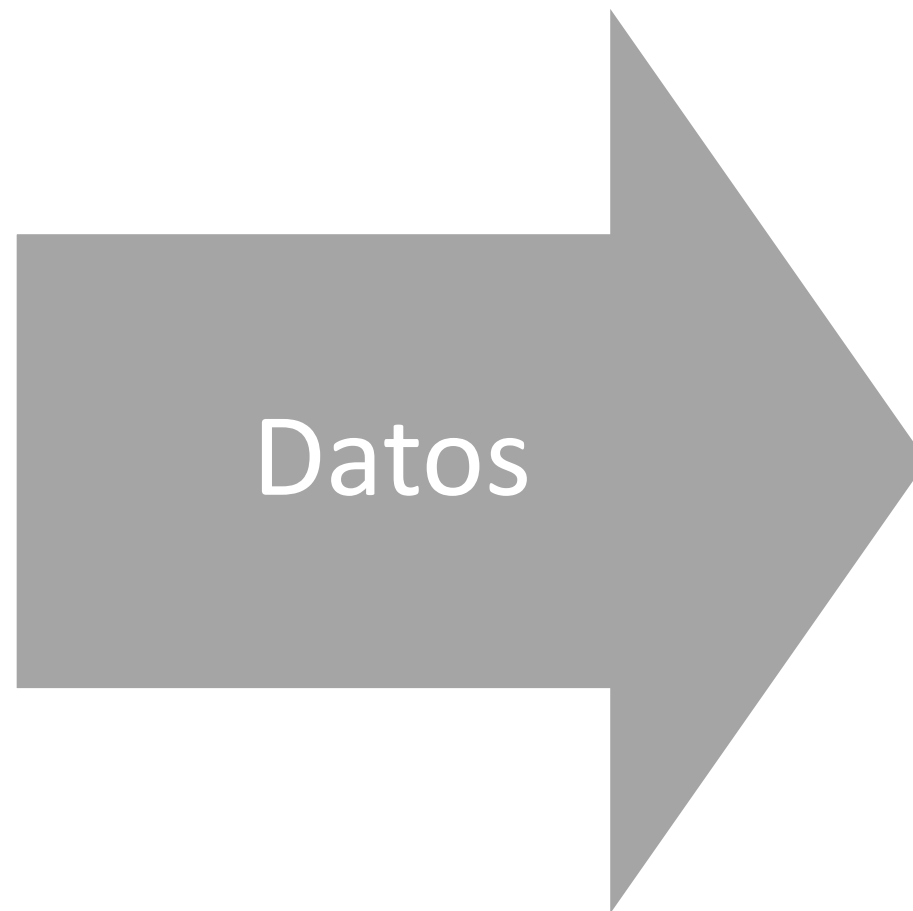


Ejemplo 2: Pasando de 1 expediente a base de datos



Del documento a los datos...

¿lo mismo o lo contrario?



ABSTRACCIÓN



MANIFESTACIÓN

Gestión física o implementada (instanciada)

Documentos o evidencias (dato instanciado)

Datos (estructuras puramente tecnológicas)

Gestión (onto)lógica

Datos modelados (dato elemental)

Documentos o evidencias (dato instanciado)

Documento

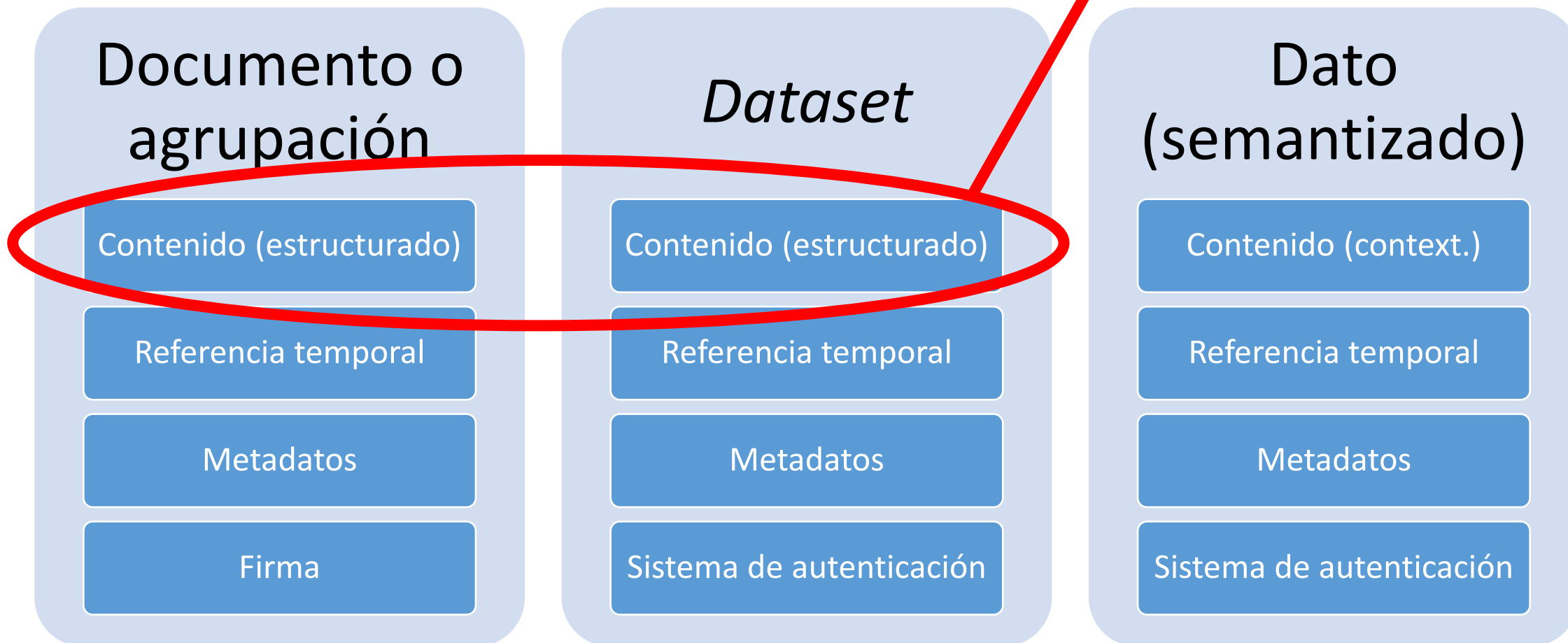
- Instancia
- Original y copias
- Vinculado a un contexto
- Utilizable (contexto pertinente)
- Estructurado preferentemente de acuerdo con una lógica funcional

Dato

- Concepto
- Original ubicuo (multiplicable)
- Suma contextos con el uso
- Reutilizable (contextos distintos)
- Estructurado de acuerdo con distintas lógicas (funcional, tecnológica, ontológica...)

El impreciso concepto de “*dataset*”

Significación



Significado
("archivístico")
del dato

Dominio

Subdominio

Término

Persona

Trabajador/a

Nombre

Dato lógico
(elemental)

Manifestación
tecnológica (física)
del dato

SI08.006 AMProc.	Fiabilidad media
SI02.012 CATPROC	Fiabilidad alta
SI05.007 CCP	Fiabilidad alta
SI08.007 NForm.	Fiabilidad alta

Dato
instanciado
(físico)

Extracción e ingesta 003	05/05/2024 – 23:02 h.	Vigente
Extracción e ingesta 002	02/05/2024 – 23:15 h.	Obsoleto
Extracción e ingesta 001	26/04/2024 – 23:00 h.	Historificado

Dato utilizable

Significado
("archivístico")
del dato

Dominio	Subdominio	Término
Persona	Trabajador/a	Nombre

Dato lógico
(elemental)

Matriz ORG003-001	Análisis y mejora de procesos	Columna 15	Alfanumérico	Fiabilidad media
Matriz ORG004-001	Normalización de formularios	Columna 8	Alfanumérico	Fiabilidad alta

Dato
instanciado
(físico o lógico?)

Manifestación
tecnológica (física)
del dato

SI08.006 AMProc.	Fiabilidad media
SI02.012 CATPROC	Fiabilidad alta
SI05.007 CCP	Fiabilidad alta
SI08.007 NForm.	Fiabilidad alta

Dato
instanciado
(físico)

Extracción e ingesta 003	05/05/2024 – 23:02 h.	Vigente
Extracción e ingesta 002	02/05/2024 – 23:15 h.	Obsoleto
Extracción e ingesta 001	26/04/2024 – 23:00 h.	Historificado

Dato utilizable

Significado
("archivístico")
del dato

Dominio	Subdominio	Término
Persona	Trabajador/a	Nombre

Dato lógico
(elemental)

Matriz ORG003-001 Análisis y mejora de procesos	Columna 15	Alfanumérico	Fiabilidad media
Matriz ORG004-001 Normalización de formularios	Columna 8	Alfanumérico	Fiabilidad alta

Dato
instanciado
(com contexto)

Manifestación
tecnológica (física)
del dato

SI08.006 AMProc.	Fiabilidad media
SI02.012 CATPROC	Fiabilidad alta
SI05.007 CCP	Fiabilidad alta
SI08.007 NForm.	Fiabilidad alta

Dato
instanciado
(físico)

Extracción e ingesta 003	05/05/2024 – 23:02 h.	Vigente
Extracción e ingesta 002	02/05/2024 – 23:15 h.	Obsoleto
Extracción e ingesta 001	26/04/2024 – 23:00 h.	Historificado

Dato utilizable



La valoración de datos: particularidades y nuevos retos

Funciones de archivo con datos

Funciones archivísticas

Captura/Transferencia

Clasificación y descripción

Preservación

Acceso

Valoración/Eliminación

Funciones de archivo con datos

Funciones archivísticas

Captura/Transferencia

Clasificación y descripción

Preservación

Acceso

Valoración/Eliminación

Retos

Hemos basado la evaluación en la serie... ¿datificada?

Funciones de archivo con datos

Funciones archivísticas

Captura/Transferencia

Clasificación y descripción

Preservación

Acceso

Valoración/Eliminación

Retos

Hemos basado la evaluación en la serie... ¿datificada?

Qué normas aplicar

Funciones de archivo con datos

Funciones archivísticas

Captura/Transferencia

Clasificación y descripción

Preservación

Acceso

Valoración/Eliminación

Retos

Hemos basado la evaluación en la serie... ¿datificada?

Qué normas aplicar

Nuevos usos

Funciones de archivo con datos

Funciones archivísticas

Captura/Transferencia

Clasificación y descripción

Preservación

Acceso

Valoración/Eliminación

Retos

Hemos basado la evaluación en la serie... ¿datificada?

Qué normas aplicar

Nuevos usos

Afectaciones a la integridad

Funciones de archivo con datos

Funciones archivísticas

Captura/Transferencia

Clasificación y descripción

Preservación

Acceso

Valoración/Eliminación

Retos

Hemos basado la evaluación en la serie... ¿datificada?

Qué normas aplicar

Nuevos usos

Afectaciones a la integridad

Saber interpretar el contexto

Quien tiene un dato tiene un tesoro

Problema

Real Decreto 203/2021, de 30 de marzo, por el que se aprueba el Reglamento de actuación y funcionamiento del sector público por medios electrónicos.

Art. 54.3 "Conservación de los documentos electrónicos"

(...)

Los plazos de conservación de esta información [incluida en base de datos] están sujetos a los mismos plazos establecidos para los correspondientes documentos electrónicos.

Quien tiene un dato tiene un tesoro

Problema

Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

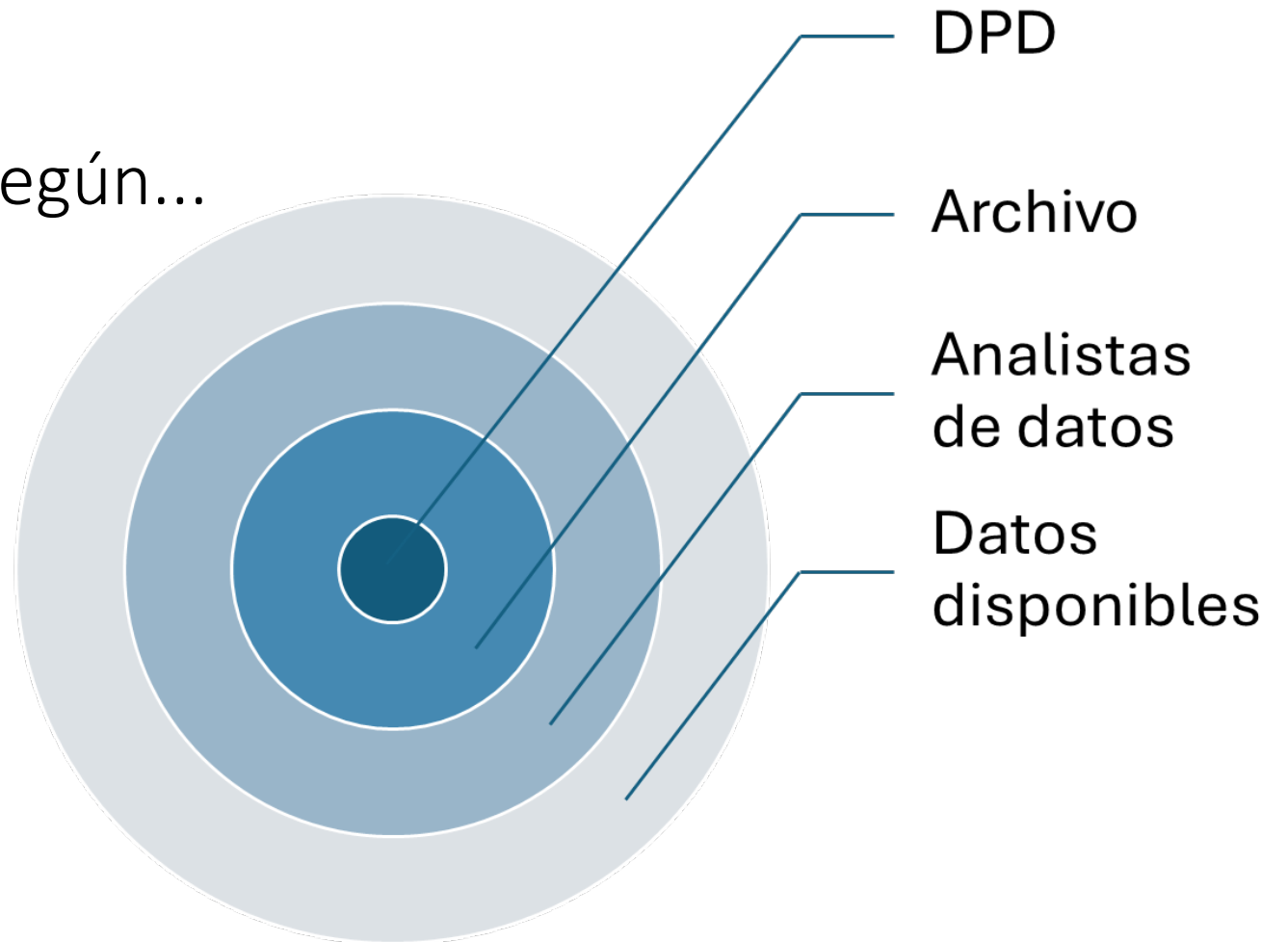
Artículo 26. Tratamiento de datos con fines de archivo en interés público por parte de las Administraciones Públicas.

Será lícito el tratamiento por las Administraciones Públicas de datos con fines de archivo en interés público...

Quien tiene un dato tiene un tesoro

Problema

Qué datos podemos conservar según...



El ciclo de vida de los datos

- El ciclo CRUD y su significado
- Vigencia vs. Volatilidad
- Los valores primarios y secundarios, y el encaje de la reutilización
- Formas de historificación y aplicación de las resoluciones de valoración (anonimización, etc.)

Los datos a conservar

- Auténticos (ciertos)
- Desagregados
- Únicos (pseudo-unicidad)
- Temporalizados
- Contextualizados (significado)

Desafíos

- Determinación del objeto real de la valoración: *dataset (raw data)*, sistema de información, matriz de datos de proceso?
- Determinación del criterio básico de conservación o eliminación:
 - Valor histórico tradicional
 - Potencial de reutilización (perfilado, actos automatizados, entrenamiento de IA...)
 - Potencial de riesgos (protección de datos, sostenibilidad, derecho al olvido...)
- Apertura de nuevas perspectivas a la valoración
 - Racionalización mediante la valoración por niveles
 - Creación de estructuras ad hoc para la preservación

The background features a stylized illustration of a globe with blue and yellow waves. Overlaid on the globe are several documents or forms, some of which are partially obscured by a semi-transparent white banner. The overall color palette is dominated by shades of blue and yellow.

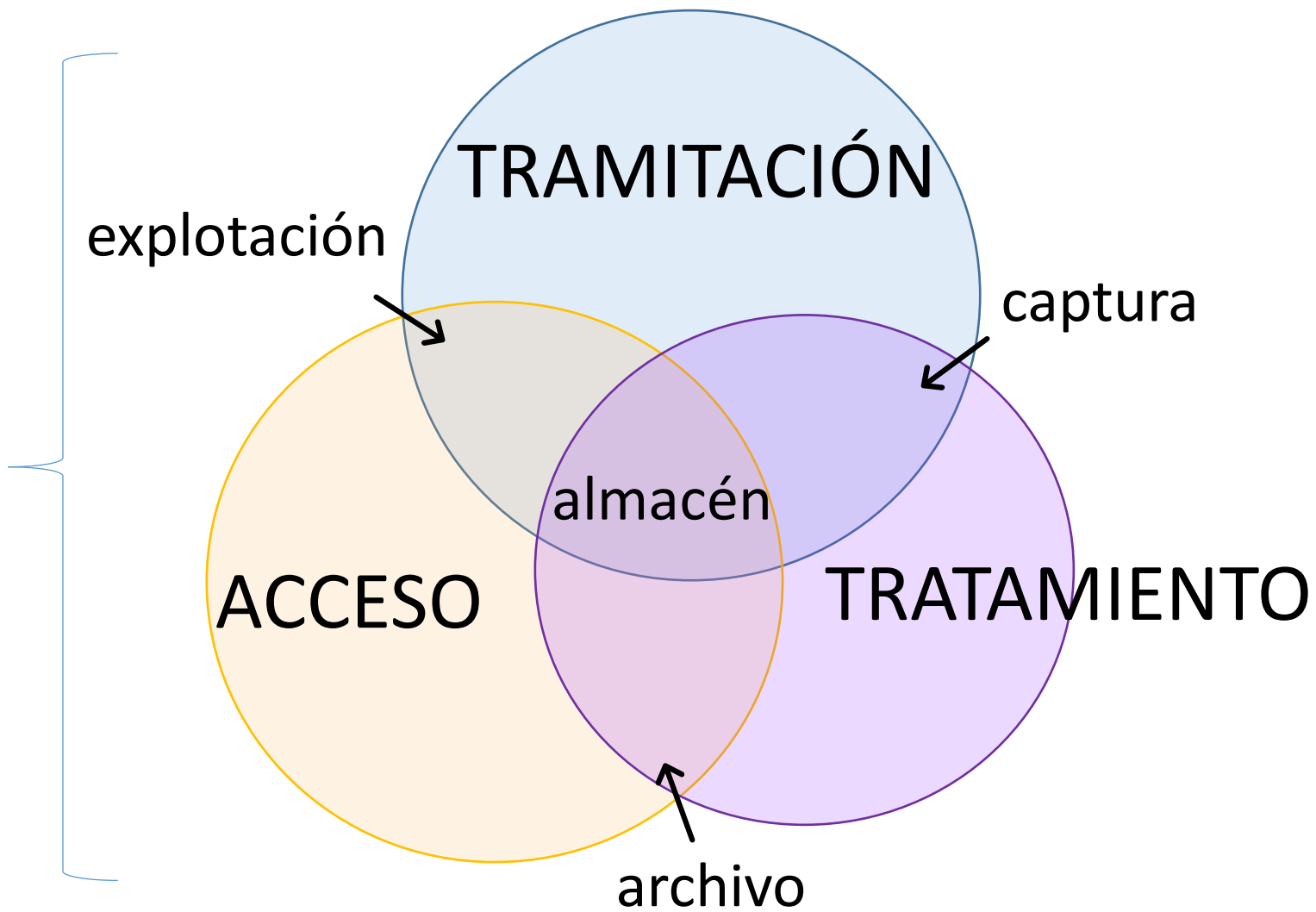
El papel del archivero en el gobierno del dato

Conjunto de datos en los archivos

Gestión de datos
Vs
Gestión Documental

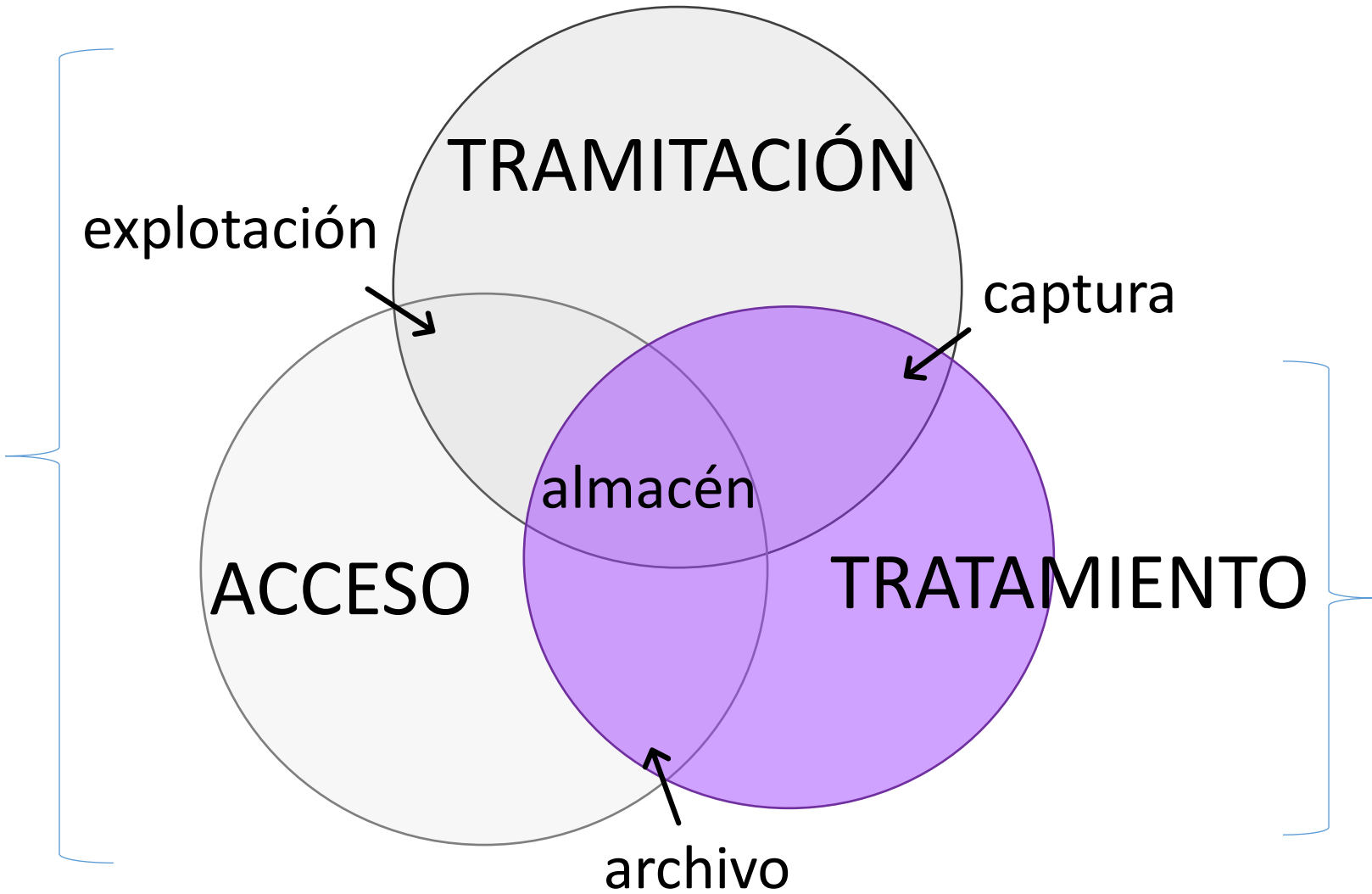
Data

management



Data

management

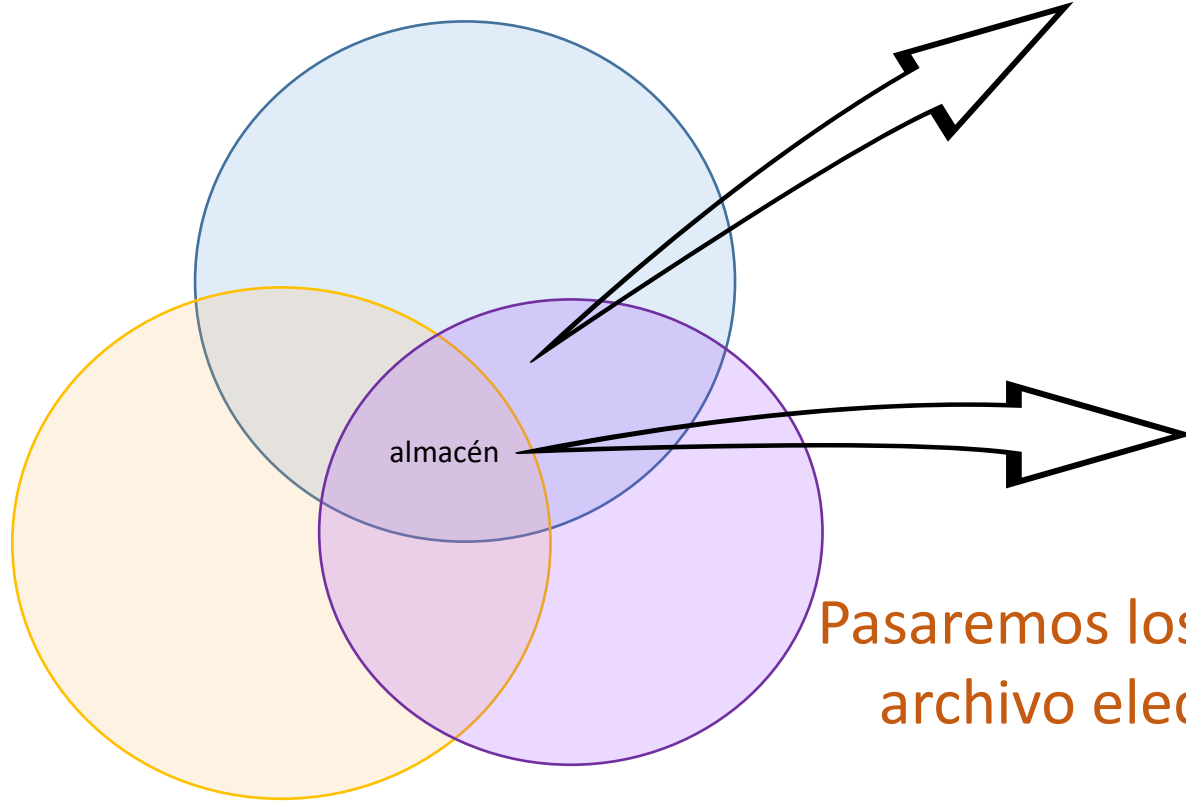


management

Records

Identificación y análisis de los datos
Intervención en el diseño
Valoración de los datos

Hoy, si necesitamos acotar...

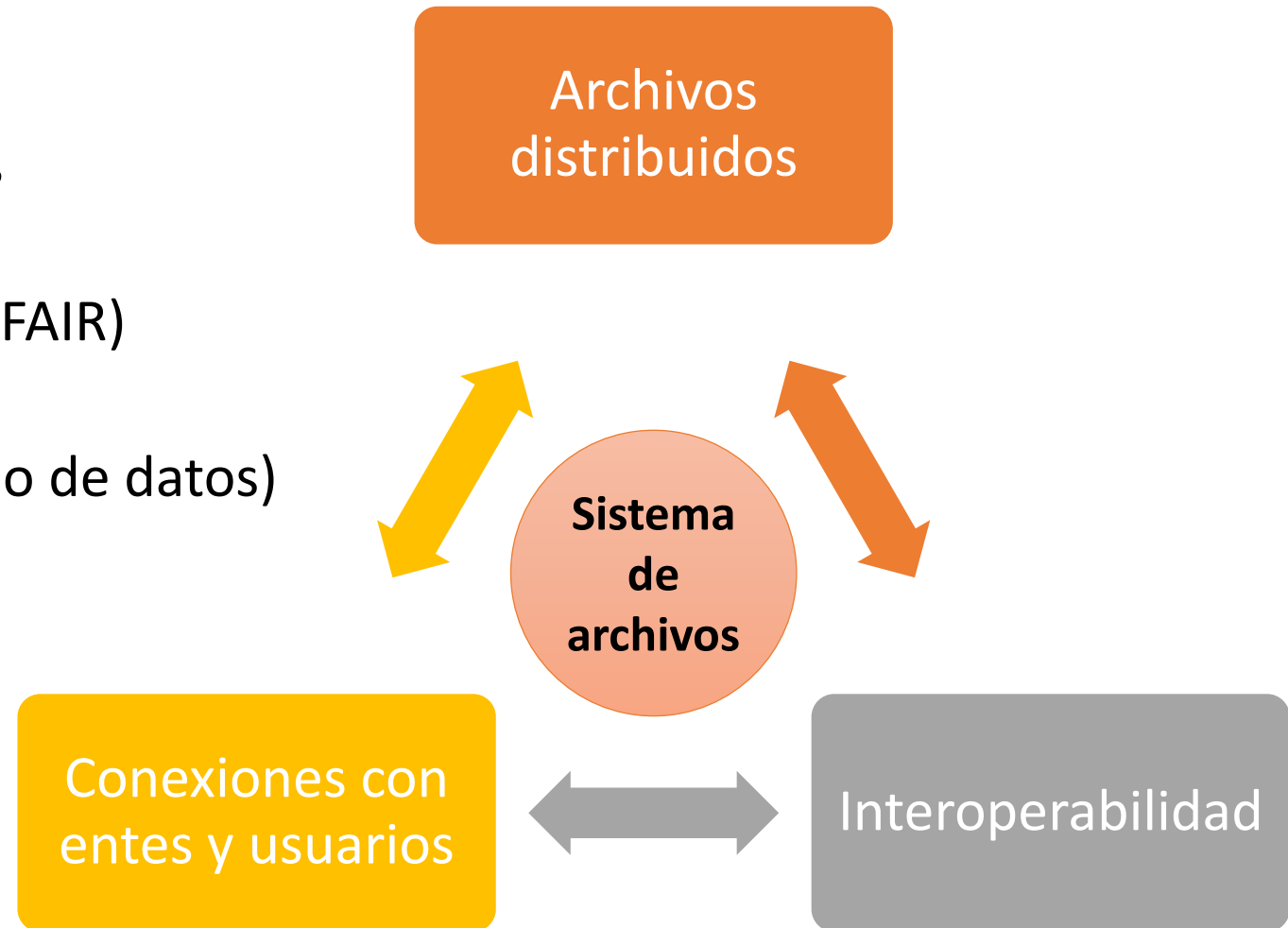


Pasaremos los datos a nuestro
archivo electrónico actual

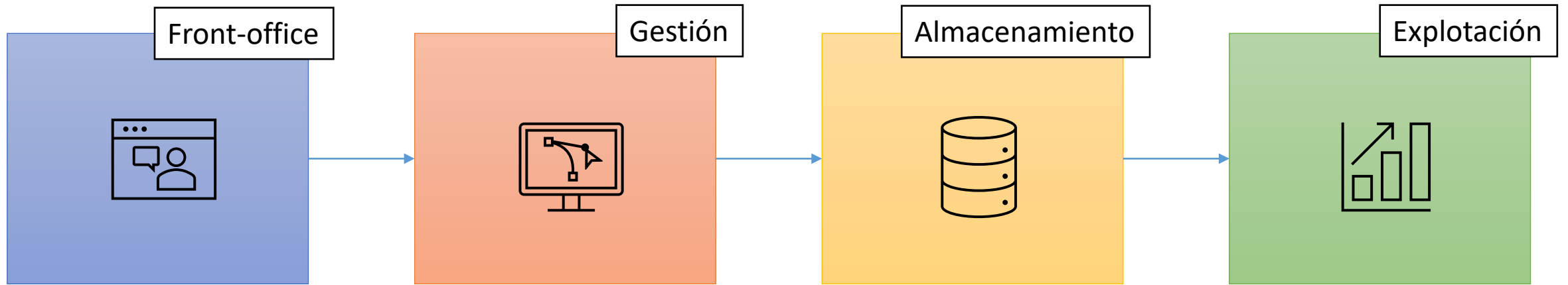


¿Llegamos tarde? ¿Existen ya archivos de datos?

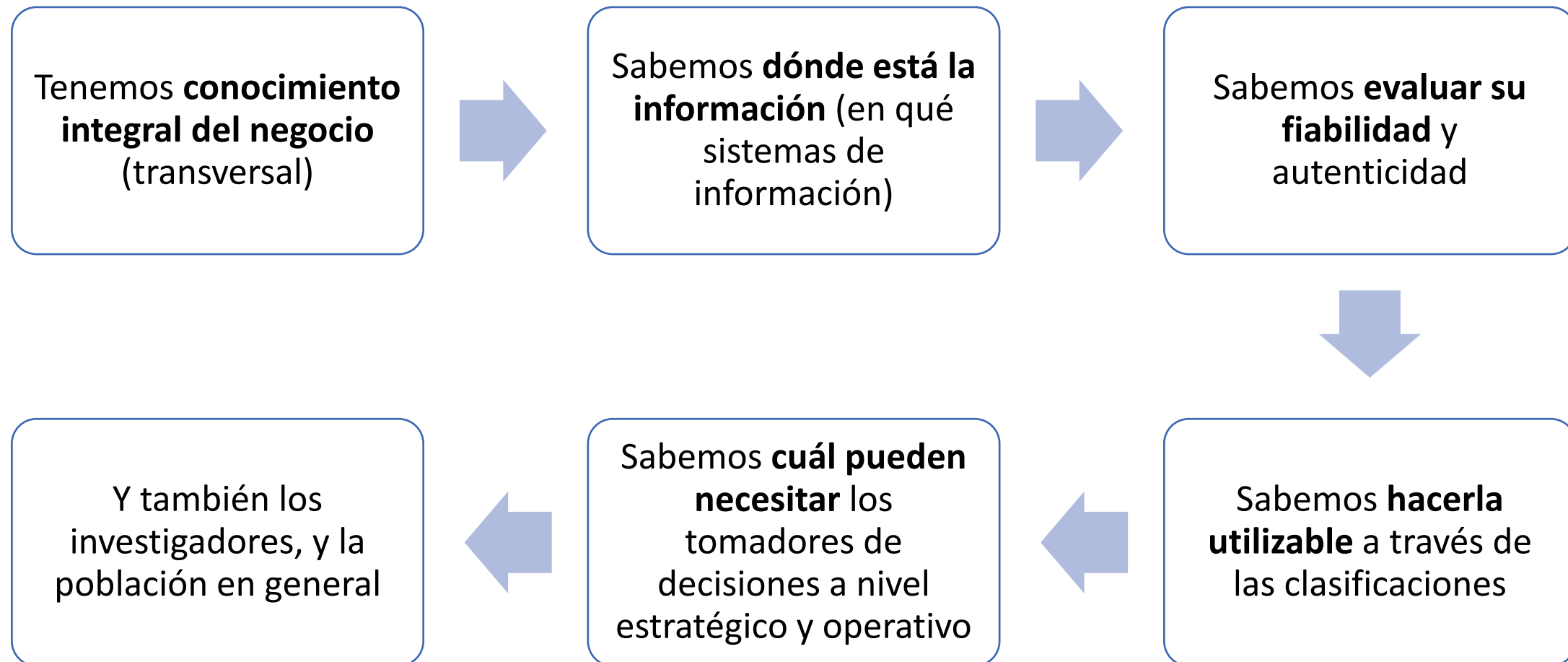
- Servicios oficiales de estadística
- Portales públicos de datos abiertos
- Centros y grupos de investigación (FAIR)
- Servicios privados (mercado privado de datos)



Pero no es solo un tema de archivo.
La clave está en la generación y gestión del dato



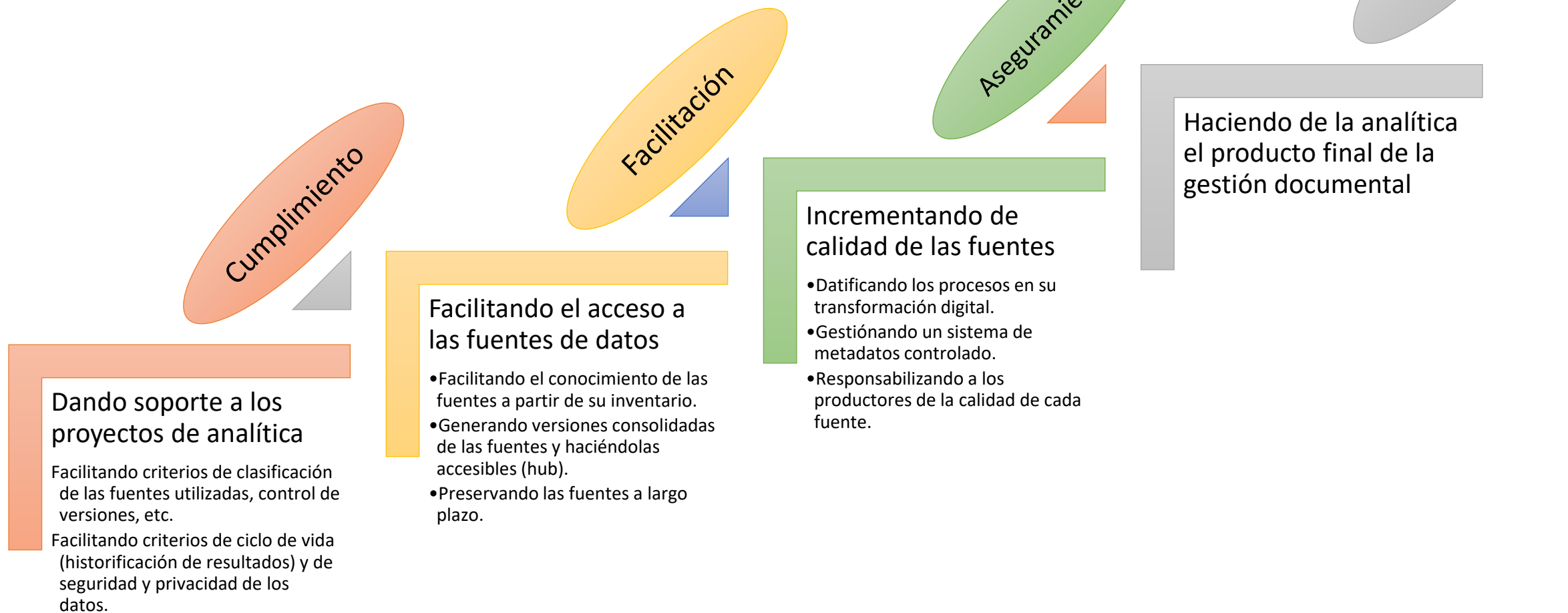
El archivero y la gestión del dato



¿Por dónde podemos empezar?

- La principal aportación, y más sencilla, es facilitar los orígenes para la analítica de datos: **identificar e inventariar las fuentes originales.**
- El trabajo de la gestión documental (identificación y valoración de series) identifica de forma natural las aplicaciones y bases de datos donde existen datos originales, incluso aquellos que tienen base ofimática.

¿Y por dónde podemos seguir?



Disponibilidad

Fiabilidad y consistencia

El lugar del archivo está en la gobernanza



Archivo como **servicio**
(generación de productos...)



Archivo como **garante**
(acceso, linaje,
preservación...)



Archivo como **facilitador**
(catálogo de datos,
sistemas...)

Dudas, preguntas, comentarios...

A stylized illustration in shades of blue and yellow. It features large, rolling waves that appear to be composed of or carrying numerous sheets of paper or documents. The overall effect is one of dynamic movement and information flow.

VII Jornadas Técnicas AEFP de Valoración

3 y 4 de octubre de 2024

Muchas
gracias

Pilar Campos Martínez
Jordi Serra Serra