

Gestión de datos de investigación y Planes de Gestión de Datos para documentalistas

5-8 de junio de 2023

Horario: 12-14/ 11.30-14/ 11.30-14/ 11.30-14:30/

Tutoría 13 de junio a las 12 h

Remedios Melero. IATA-CSIC, Valencia, España

Tema 1. Acceso abierto y ciencia abierta (5 de junio, 12:00-14:00)

- El acceso abierto a las publicaciones y a los datos de investigación
- El ciclo de vida de los datos
- Políticas sobre los datos de investigación

Tema 2. Fundamentos de la gestión de datos de investigación (5 de junio, 12:00-14:00)

- Los datos de investigación: Definiciones y tipología
- Formatos
- Datos abiertos. Datos FAIR

Tema 3. Como citar y enriquecer los datos. Aspectos legales y éticos (6 junio, 11:30-14:00)

- Licencias para compartir datos
- Cómo citar datos. Cómo enriquecer los datos
- Normativa GDPR. Anonimización de datos

Tema 4. Depósito y localización de datasets (7 junio, 11:30-14:00)

- Depósito de datasets
- Repositorios de datos
- Localización de datos

Tema 5. El plan de gestión de datos (PGD) (8 junio, 11:30-14:30)

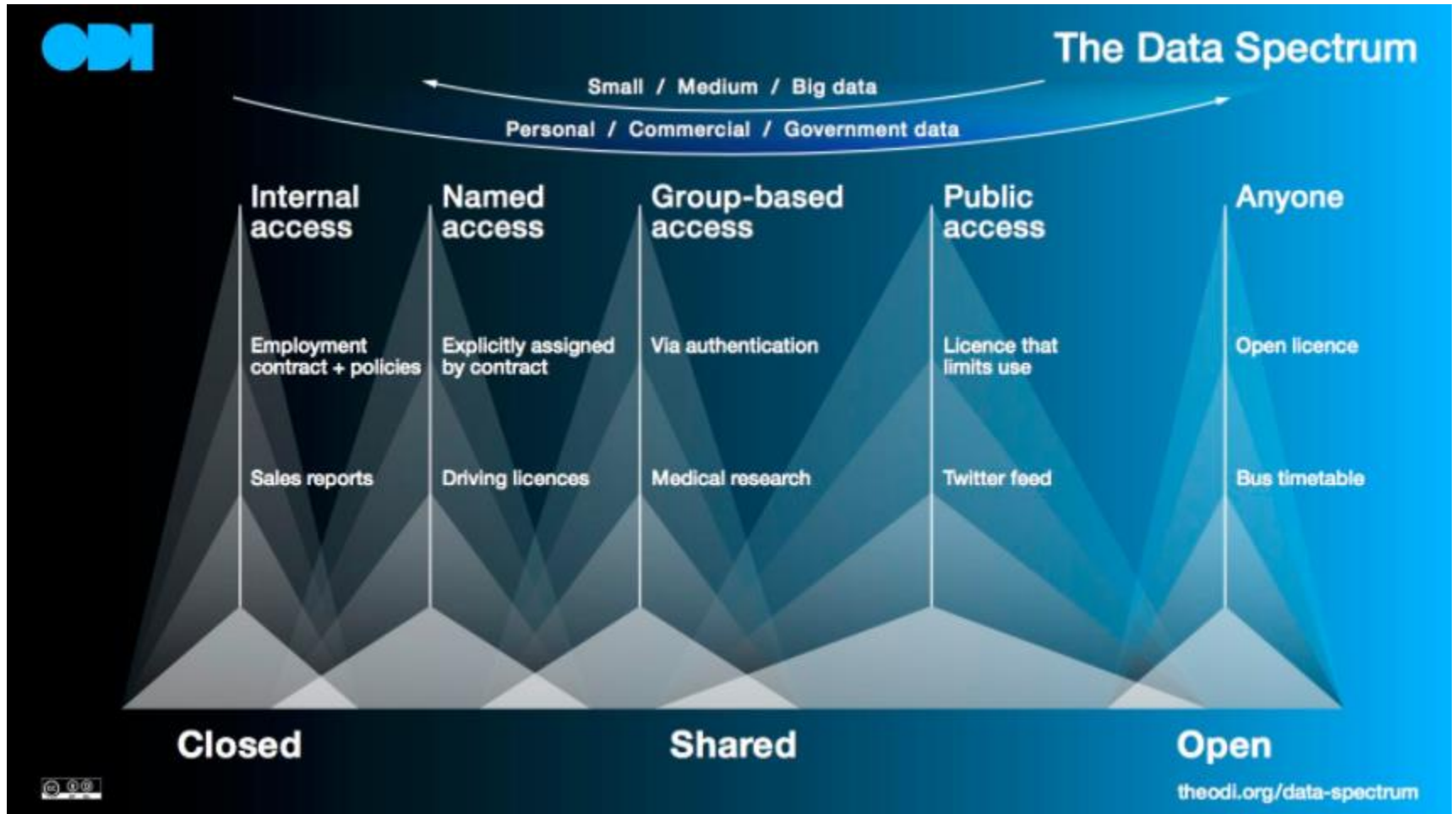
- Qué es un PGD
- Recursos y herramientas para generar PGD

Tema 2. Fundamentos de la gestión de datos de investigación (5 de junio, 12:00-14:00)

- Los datos de investigación: Definiciones y tipología
- Formatos
- Datos abiertos. Datos FAIR

Tipología de datos y formatos de datos en función de su naturaleza.

Datos FAIR



Datos en función de su formato





Recommended formats

SHARE

Data lifecycle

Plan to share

Legal and ethical

Rights

Document your data

● **Format your data**

File formats

– Recommended formats

Organising

Quality

Versioning

Transcription

This table contains guidance on file formats recommended and accepted by the UK Data Service for data sharing, reuse and preservation.

You may need to convert your data files to a preservation file format.

We welcome [queries](#) from researchers about appropriate file formats for working and preservation, particularly early in the research process.

If you are unsure of the suitability of your file formats for the data you want to deposit with the UK Data Service, please [get in touch](#).

Type of data	Recommended formats	Acceptable formats
Tabular data with extensive metadata variable labels, code labels and	SPSS portable format (.por) delimited text and command ("setup") file (SPSS, Stata, SAS, etc.)	proprietary formats of statistical packages: SPSS (.sav), Stata (.dta), MS Access (.mdb/.accdb)

Tipo de datos	Formatos Recomendados	Formatos aceptables
Imágenes	TIFF 6.0 uncompressed (.tif)	JPEG (.jpeg, .jpg, .jp2) if original created in this format GIF (.gif) TIFF other versions (.tif, .tiff) RAW image format (.raw) Photoshop files (.psd) BMP (.bmp) PNG (.png) Adobe Portable Document Format (PDF/A, PDF) (.pdf)
Audio	Free Lossless Audio Codec (FLAC) (.flac)	MPEG-1 Audio Layer 3 (.mp3) if original created in this format Audio Interchange File Format (.aif) Waveform Audio Format (.wav)
Video	MPEG-4 (.mp4) OGG video (.ogv, .ogg) motion JPEG 2000 (.mj2)	AVCHD video (.avchd)
Scripts	Rich Text Format (.rtf) PDF/UA, PDF/A or PDF (.pdf) XHTML or HTML (.xhtml, .htm) OpenDocument Text (.odt)	plain text (.txt) widely-used formats: MS Word (.doc/.docx), MS Excel (.xls/.xlsx) XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHTML 1.0

Tipo de datos	Formatos recomendados	Formatos aceptables
Datos en tablas con nombres de variables, códigos.. (extensive metadata)	SPSS portable format (.por) delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) structured text or mark-up file of metadata information, e.g. DDI XML file	proprietary formats of statistical packages: SPSS (.sav), Stata (.dta), MS Access (.mdb/.accdb)
Datos en tablas con metadatos mínimos (encabezamientos de columnas/nombre de variables)	comma-separated values (.csv) tab-delimited file (.tab) delimited text with SQL data definition statements	delimited text (.txt) with characters not present in data used as delimiters widely-used formats: MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf), OpenDocument Spreadsheet (.ods)
Geoespaciales (Vectorial y raster)	ESRI Shapefile (.shp, .shx, .dbf, .prj, .sbx, .sbn optional) geo-referenced TIFF (.tif, .tfw) CAD data (.dwg) tabular GIS attribute data Geography Markup Language (.gml)	ESRI Geodatabase format (.mdb) MapInfo Interchange Format (.mif) for vector data Keyhole Mark-up Language (.kml) Adobe Illustrator (.ai), CAD data (.dxf or .svg) binary formats of GIS and CAD packages
Textuales	Rich Text Format (.rtf) plain text, ASCII (.txt) eXtensible Mark-up Language (.xml) text according to an appropriate Document Type Definition (DTD) or schema	Hypertext Mark-up Language (.html) widely-used formats: MS Word (.doc/.docx) some software-specific formats: NUD*IST, NVivo and ATLAS.ti

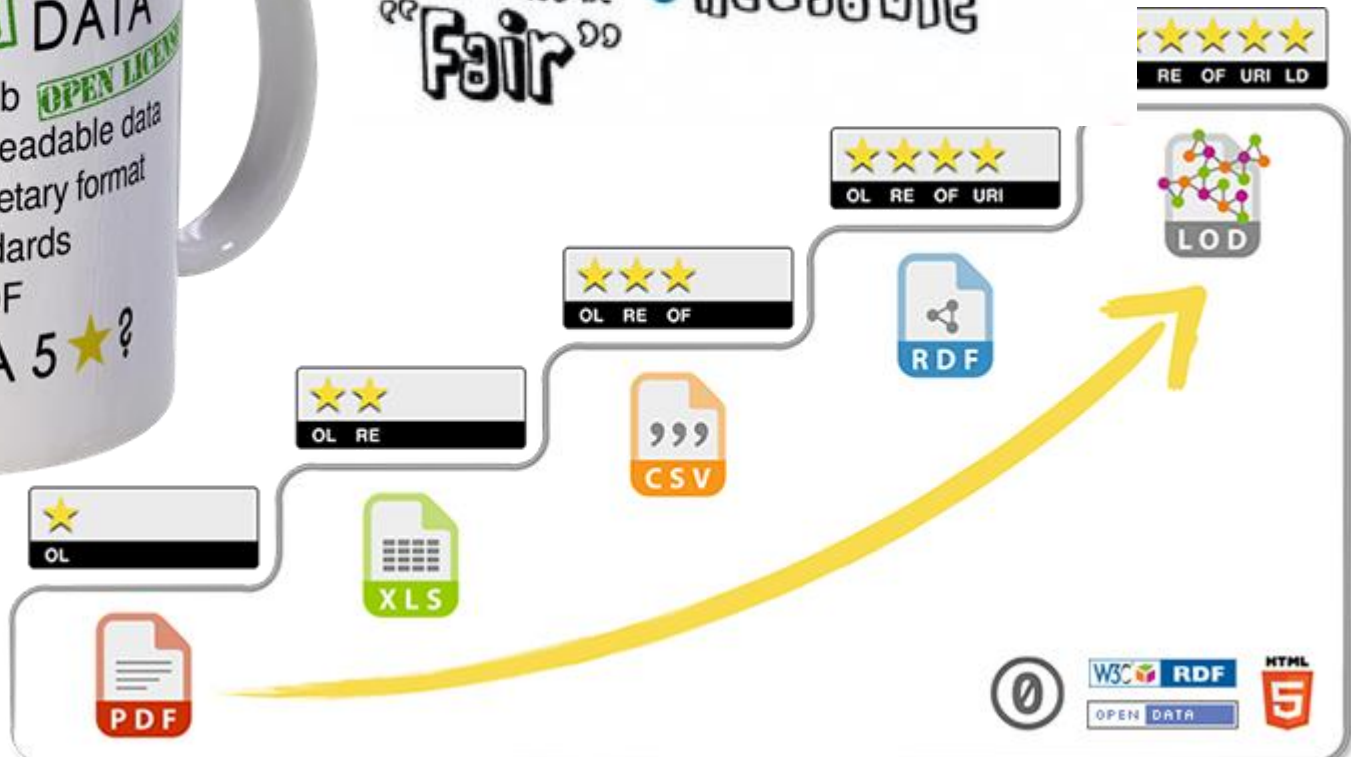


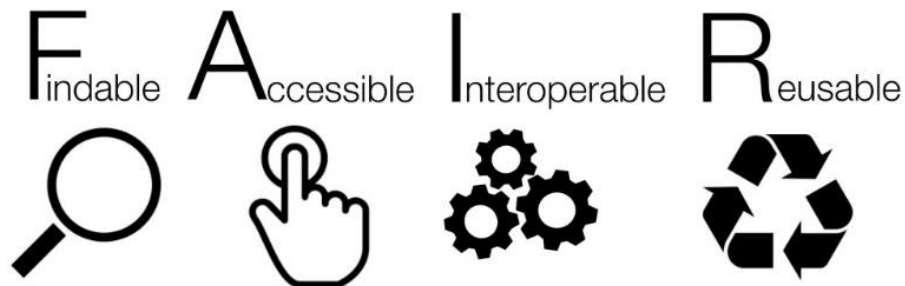
Open data must be accessible, useable, assessable and intelligible (definición de *Science as an Open Enterprise*, 2012)



Open data
is about
MORE
THAN
DISCLOSURE
it must be
Fair

- Findable
- Accessible
- Interoperable
- Reusable





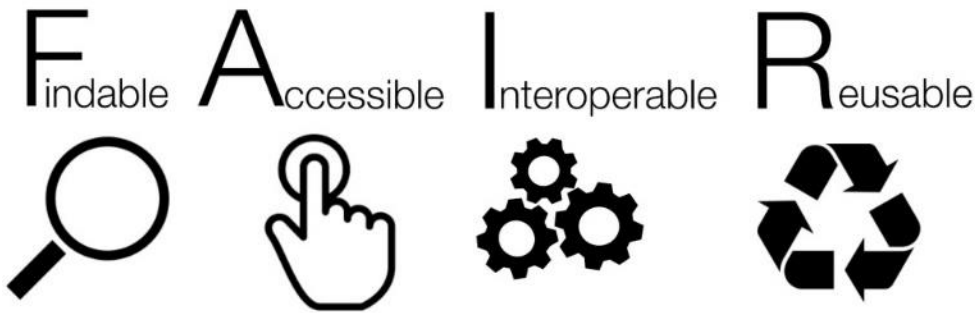
Principios FAIR (<https://www.force11.org/group/fairgroup/fairprinciples>)

Para ser localizable (**Findable**)

- F1. A los (meta)datos se les debe asignar un identificador único y persistente.
- F2. Los datos se describen con metadatos enriquecidos.
- F3. (meta)datos se registran o indexan en un sistema que disponga de motor de búsqueda.
- F4. Los metadatos contienen uno para el identificador del dataset.

Para ser accesible (**Accesible**):

- A1. Los (meta)datos se pueden recuperar mediante su identificador utilizando un protocolo de comunicación estandarizado.
 - A1.1 El protocolo es abierto, gratuito y universalmente aplicable.
 - A1.2 El protocolo permite un procedimiento de autenticación y autorización, cuando sea necesario.
- A2. Los metadatos son accesibles, incluso cuando los datos ya no están disponibles.



Principios FAIR (<https://www.force11.org/group/fairgroup/fairprinciples>)

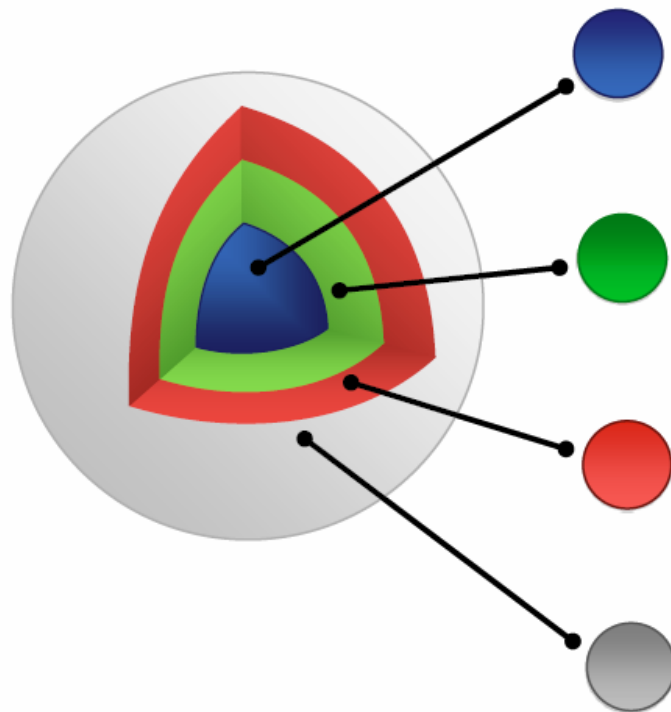
Para ser interoperable (Interoperable):

- I1. Los (meta)datos utilizan un lenguaje formal, accesible, compartido y ampliamente aplicable para la representación del conocimiento.
- I2. Los (meta)datos usan vocabularios que siguen principios FAIR.
- I3. Los (meta)datos incluyen referencias calificadas a otros (meta)datos.

Para ser reutilizable (Reusable)

- R1. Los (meta)datos tienen disponibles atributos precisos y pertinentes.
 - R1.1. Los (meta)datos se publican con una licencia de reutilización clara y accesible.
 - R1.2. Los (meta)datos están asociados con su procedencia.
 - R1.3. Los (meta)datos cumplen con los estándares propios de una comunidad de un área de conocimiento determinada.

Turning FAIR data into reality. Interim report from the European Commission Expert Group on FAIR data



DATA

The core bits

At its most basic level, data is a bitstream or binary sequence. For data to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by Persistent Identifiers (PIDs), metadata and code. These layers of meaning enrich the data and enable reuse.

IDENTIFIERS

Persistent and unique (PIDs)

Data should be assigned a unique and persistent identifier such as a DOI or URN. This enables stable links to the object and supports citation and reuse to be tracked. Identifiers should also be applied to other related concepts such as the data authors (ORCID), projects (RAIDs), funders and associated research resources (RRIDs).

STANDARDS & CODE

Open, documented formats

Data should be represented in common and ideally open file formats. This enables others to reuse the data as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve. Data also need to be accompanied by the code use to process and analyse the data.

METADATA

Contextual documentation

In order for data to be assessable and reusable, it should be accompanied by sufficient metadata and documentation. Basic metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when and by whom the data were created. To enable the broadest reuse, data should be accompanied by a 'plurality of relevant attributes' and a clear and accessible data usage license.

Model for FAIR data objects

