

Anexo D2. Tipología de datos de la investigación

¿Qué datos va a generar su proyecto?

Los datos de investigación se generan por diferentes motivos a través de diferentes procesos y pueden ser:

- Observacionales: datos capturados en tiempo real (neuroimágenes, datos de muestras, datos de sensores, datos de encuestas, etc.).
- Experimentales: datos capturados en equipos de laboratorio (secuencias de genes, cromatogramas, datos de campos magnéticos, etc.).
- Simulación: datos generados a partir de modelos de prueba (climatológicas, matemáticas o modelos económicos, etc.).
- Derivadas o compiladas: datos reproducibles, pero de difícil reproducción (texto y minería de datos, modelos 3D, base de datos compilada, etc.).
- De referencia: conglomerado o conjunto de datos (bancos de datos de secuencias de genes, estructuras químicas, portales de datos espaciales, etc.).
- Otros

¿Cómo se verán sus datos?

TIPO DE DATO	DESCRIPCIÓN - FORMATO ACEPTADOS
Datos cuantitativos (tabulados, variables, matriz de datos, entre otros) y bases de datos	- Formato estadísticos (.por / .sav / .dta / .sas7bdat) - Texto estructurado (.xml / .ddi) - Bases de datos (.mdb / .accdb / .dbf / .xml / .json / .ris / .bib) - Valores separados por comas (.csv) - Valores separados por tabulaciones (.tsv) - Texto delimitado (.txt) - Hojas de cálculo (.xls / .xlsx / .ods) - SQLite (.sqlite / .db) - SQL Dump (.sql)
Datos geoespaciales o vectoriales	- ESRI Shapefile (.shp / .shx / .dbf / .prj / .sbx / .sbn) - ESRI geodatabase (.mdb / .gdb) - TIFF geo-referenciado (.tif / .tiff) - Datos CAD (.dwg) - KML (.kml) - MapInfo (.mif) - Adobe Illustrator (.ai) - Datos CAD (.dxf / .svg) - JSON geoespacial (.geojson) - Keyhole Markup Language (Google Earth) (.kml, .kmz) - Geography Markup Language (.gml) - GeoTIFF (.tif) - Formato de intercambio GPS (.gpx)



Datos cualitativos	- Extensible Markup Language (.xml) - Formato enriquecido (.rtf) - Datos de texto sin formato (.txt / UTF-8 –Unicode-) - HyperText Markup Language (.html) - Editor de texto (.doc / .docx) - Otros formatos (Atlas.ti / NVivo / Latex)
Datos de imagen	- Imagen de alta calidad, sin pérdida (.tif / .tiff) - Imagen comprimida sin pérdida (.png) - Imagen comprimida con pérdida (.jpeg / .jpg) - Norm RAW (.raw) - Photoshop (.psd) - imagen animada o estática (.gif) - Bitmap (.bmp) - Gráficos vectoriales (.svg) - PostScript encapsulado (.eps)
Datos de audio digital	- Audio sin pérdida (.flac) - Audio sin compresión (.wav) - Audio comprimido (.mp3) - Audio inter change file format (.aif)
Datos vídeo digital	- MPEG-4 (.mp4) - Video Matroska (.mkv) - JPEG 2000 (.mj2) - AVI (.avi) - MXF FLAC (.mxf) - Formato de Apple para video de alta calidad (.mov)
Documentación	- Texto plano (.txt) - Formato de texto enriquecido (.rtf) - HTML (.htm / .html) - Procesadores de texto (.doc / .docx / .odt) - LaTeX (.tex) - PDF (.pdf / .pdf/a)
Código y scripts	- Python (.py) - R (.r / .rda / .rds / .RData) - Jupyter Notebook (.ipynb) - MATLAB (.m / .mat) - Lenguajes de programación diversos (.java / .cpp / .c / .js) - Scripts de automatización (.sh, .bat) - Jupyter Notebook (.ipynb) - Markdown (.md, .markdown) - Otros (.html / .css / .yaml / .yml)
Modelos, simulaciones y resultados computacionales	- Modelos biológicos o químicos (.model / .sim / .sbml) - Modelos 3D (impresión o simulación) (.stl / .obj) - Datos climáticos y de simulación (.netcdf / .h5 / .grib)



Datos científicos específicos	- Secuencias genéticas (.fasta / .fastq) - Datos de espectrometría de masas (.mzML / .mzXML) - NetCDF (datos climáticos y ambientales) (.cdf / .nc) - Hierarchical Data Format (.hdf5 / .h5) - Imágenes neurológicas NIfTI (.nii / .nii.gz)
Archivos comprimidos y empaquetados	- Archivos comprimidos (.zip / .tar / .gz / .7z) - BagIt (.bag)

Fuente: elaboración propia con base en Couto (2016)