

Recomanacions de formats

Introducció

 [English version](#)

En planificar un projecte de recerca, és important tenir en compte quin format de fitxer s'utilitzarà per emmagatzemar les dades. Els formats dels fitxers que utilitzeu tenen un impacte directe a l'hora de compartir les dades de recerca i en la seva reutilització en un futur.

Els formats que es presenten a la taula com a recomanats són formats de fitxer que oferiran les millors garanties a llarg termini en termes d'usabilitat, accessibilitat i sostenibilitat. Els formats acceptats són formats de fitxer que s'utilitzen àmpliament i que seran moderadament utilitzables i accessibles a llarg termini.

Els formats d'arxiu amb més probabilitat de ser accessibles tenen les característiques següents:

- No propietaris
- Estàndards oberts i documentats
- Utilitzats per la comunitat investigadora
- Representació estàndard (ASCII, Unicode)
- Sense xifrar
- Sense comprimir

Bones pràctiques

- Sempre que sigui possible, heu de desar les dades en un format obert i sostenible, no propietari (els programaris propietaris sovint permeten “desar com” un format obert sense dificultats). Si la conversió a un format de dades en obert genera una pèrdua de dades dels fitxers, considereu guardar-les tant en format propietari com en format obert. D'aquesta manera es podrà disposar d'una part de la informació.
- Quan sigui necessari guardar arxius en un format propietari, considereu incloure un arxiu *readme.txt* com a guia que documenti el nom i versió del programari utilitzat per generar l'arxiu, així com l'empresa que feia el programari. Pot ser de molta ajuda en cas de necessitar obrir aquests fitxers.
- Per evitar el risc d'obsolescència i garantir l'accessibilitat i la sostenibilitat dels fitxers es poden prendre diverses mesures. Una d'aquestes mesures és utilitzar formats de fitxer que tinguin una alta probabilitat de romandre utilitzables durant molts anys.

Tipus de dades

- Observacionals: dades capturades en temps real (p. ex., neuroimatges, dades de mostres, dades de sensors, dades d'enquesta)
- Experimentals: dades capturades en equips de laboratori (p. ex., seqüències de gens, cromatogrames, dades de camps magnètics)
- Simulació: dades generades a partir de models de prova (p. ex., climatològiques, matemàtiques o models econòmics)
- Derivades o compilades: dades reproduïble però de difícil reproducció (p. ex., de text i mineria de dades, models 3D, bases de dades compilada)

Formats recomanats i acceptats

Tipus de dades	Formats recomanats	Formats no recomanats però habitualment acceptats
Documents de text	• <i>PDF/A</i> (.pdf) • <i>ODT</i> (.odt)	• <i>Microsoft Word</i> (.doc) • <i>Office Open XML</i> (.docx) • <i>Rich Text File</i> (.rtf) • <i>PDF diferents a PDF/A</i> (.pdf)
Text pla	• <i>Unicode text</i> (.txt)	• <i>Non-Unicode text</i> (.txt)
Llenguatges de marcatge	• <i>XML</i> (.xml) • <i>HTML</i> (.html) • <i>YAML</i> (.yaml) • <i>JSON</i> (.json) • <i>ReStructuredText</i> (.rst) • Fitxers relacionats: .css, .xslt, .js, .es	• <i>SGML</i> (.sgml) • <i>Markdown</i> (.md)
Llenguatges de programació	• <i>NetCDF</i> • <i>TextFabric</i> • <i>R</i> (.r) • <i>Octave</i> (.mat)	• <i>MATLAB</i> (.mat)
Fulls de càlcul	• <i>ODS</i> (.ods) • <i>CSV</i> (.csv)	• <i>Microsoft Excel</i> (.xls) • <i>Office Open XML Workbook</i> (.xlsx) • <i>PDF/A</i> (.pdf)

Bases de dades	• <i>SQL</i> (.sql) • <i>SIARD</i> (.siard) • <i>CSV</i> (.csv) • <i>FITS</i> (.fits, .fit, .fts) • (<i>Apache</i>) <i>Parquet</i> (.parquet)	• <i>Microsoft Access</i> (.mdb, .accdb) • <i>dBase</i> (.dbf) • <i>HDF5</i> (.hdf5, .he5, .h5)
Dades estadístiques	• <i>SPSS</i> (.dat/.sps) • <i>STATA</i> (.dat/.DO) • <i>R</i> (.rdat/.rdara)	• <i>SPSS Portable</i> (.por) • <i>SPSS</i> (.sav) • <i>STATA</i> (.dta) • <i>SAS</i> (.7dat; .sd2; .tpt)
Imatges (mapa de bits)	• <i>JPEG</i> (.jpg, .jpeg) • <i>TIFF</i> (.tif, .tiff) • <i>PNG</i> (.png) • <i>JPEG 2000</i> (.jp2) • <i>DICOM</i> (.dcm)	
Imatges vectorials	• <i>SVG</i> (.svg)	• <i>Adobe Illustrator</i> (.ai) • <i>EPS</i> (.eps) • <i>WMF/EMF</i> (.wmf, .emf) • <i>CDR</i> (.cdr)
Àudio (contenidor i còdec)	• <i>BWF</i> (.bwf)	
Àudio (contenidor)	• <i>MXF</i> (.mxf) • <i>Matroska</i> (.mka)	
Àudio (còdec)	• <i>FLAC</i> (.flac) • <i>OPUS</i>	• <i>WAVE</i> (.wav) • <i>MP3</i> (.mp3) • <i>AAC</i> (.aac, .m4a) • <i>AIFF</i> (.aif, .aiff) • <i>OGG</i> (.ogg)
Vídeo (Contenidor)	• <i>MXF</i> (.mxf) • <i>Matroska</i> (.mkv)	• <i>MPEG-4</i> (.mp4, .m4a, .m4v) • <i>MPEG-2</i> (.mpg, .mpeg, .m2v, mpg2) • <i>AVI</i> (.avi) • <i>QuickTime</i> (.mov, .qt)

Video (còdec)	• <i>Theora</i> (.ogv)	
Disseny assistit per ordinador (CAD)	• <i>AutoCAD DXF version R12 (ASCII)</i> (.dxf) • <i>SVG</i> (.svg)	• <i>AutoCAD DXF diferent a R12 (ASCII)</i> (.dwg, .dxf) • <i>DWG</i> (.dwg) • <i>DGN</i> (.dgn)
Sistemes d'informació geogràfica (GIS)	• <i>GML</i> (.gml) • <i>MIF/MID</i> (.mif/.mid)	• <i>Esri Shapefiles</i> (.shp & related files) • <i>MapInfo</i> (.tab & related files) • <i>KML</i> (.kml) • <i>Esri Geodatabase</i> (.gdb) • <i>Project files/Workspaces</i> (.mxd, .wor, .qgs)
Imatges georeferenciades	• <i>GeoTIFF</i> (.tif, .tiff)	• <i>TIFF World File</i> (.tfw & .tif, possibly with additional files) • <i>JPEG World File</i> (.jgw & .jpg, possibly with additional files) • <i>ERDAS IMAGINE File Format</i> (.img)
GIS raster	• <i>ASCII GRID</i> (.asc, .txt)	• <i>Esri GRID</i> (.grd & related files) • <i>Surfer Grid</i> (.grd; .srf) • <i>ERDAS IMAGINE File Format</i> (.img)
3D	• <i>WaveFront Object</i> (.obj) • <i>Polygon file format</i> (.ply) • <i>X3D</i> (.x3d) • <i>COLLADA</i> (.dae)	• <i>Autodesk FBX</i> (.fbx) • <i>Blender</i> (.blend) • <i>3D PDF</i> (.pdf)
RDF	• <i>RDF/XML</i> (.rdf) • <i>Trig</i> (.trig) • <i>Turtle</i> (.ttl) • <i>NTriples</i> (.nt) • <i>JSON-LD</i> • <i>ATLAS.TI Copy bundle</i> • <i>NVivo Project file</i>	
Computer Assisted Qualitative Data Analysis (CAQDAS)	• <i>REFI-QDA</i> (Qualitative Data Analysis)	

Dades de proteòmica	• <i>MSF</i> (.msf)	
Estàndard de citometria de flux	• <i>FCS</i> (.fcs)	

Taula elaborada a partir de la taula de formats de *Data Archiving and Networked Services* (DANS)

Més informació de formats a *Recommended Formats Statement* de la Library of Congress