
BONES PRÀCTIQUES EN GESTIÓ DE DADES DE RECERCA

Recomanacions de formats

Aquest document ha estat elaborat pel Grup de Treball de Suport a la Recerca del CSUC que està format per representants de les universitats següents:

Universitat de Barcelona
Universitat Autònoma de Barcelona
Universitat Politècnica de Catalunya
Universitat Pompeu Fabra
Universitat de Girona
Universitat de Lleida
Universitat Rovira i Virgili
Universitat Oberta de Catalunya
Universitat de Vic-Universitat Central de Catalunya
Universitat Ramon Llull
Universitat de les Illes Balears
Universitat Jaume I

Octubre 2020

Sumari

1. Introducció	3
1.1 Bones pràctiques.....	3
2. Tipus de dades	3
3. Taula de formats	4

Control de versions

Versió	Data	Autors	Notes
1.0	13/10/2020	UdL, UOC, URL	Versió presentada al GTSR

Llicència



Aquest document està subjecte a la llicència de [Reconeixement](#) de Creative Commons (CC-BY 4.0)

1. Introducció

Al planificar un projecte de recerca, és important tenir en compte quins 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

1.1 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.

2. 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)

3. Taula de formats

Tipus de dades	Format recomanat per compartir, reutilitzar i preservar	Format habitual acceptat
Text	<ul style="list-style-type: none"> • PDF/A (.pdf) • ODT (.odt) 	<ul style="list-style-type: none"> • Microsoft Word (.doc) • Office Open XML (.docx) • Rich Text File (.rtf) • PDF other than PDF/A (.pdf)
Text pla	<ul style="list-style-type: none"> • Unicode text (.txt) 	<ul style="list-style-type: none"> • Non-Unicode text (.txt)
Llenguatges de marcatge	<ul style="list-style-type: none"> • XML (.xml) • HTML (.html) • Related files: .css, .xslt, .js, .es 	<ul style="list-style-type: none"> • SGML (.sgml) • Markdown (.md)
Llenguatges de programació	<ul style="list-style-type: none"> • MATLAB • NetCDF • TextFabric 	
Fulls de càlcul	<ul style="list-style-type: none"> • ODS (.ods) • CSV (.csv) 	<ul style="list-style-type: none"> • Microsoft Excel (.xls) • Office Open XML Workbook (.xlsx) • PDF/A (.pdf)
Bases de dades	<ul style="list-style-type: none"> • SQL (.sql) • SIARD (.siard) • CSV (.csv) 	<ul style="list-style-type: none"> • Microsoft Access (.mdb, .accdb) • dBase (.dbf) • HDF5 (.hdf5, .he5, .h5)
Dades estadístiques	<ul style="list-style-type: none"> • SPSS (.dat/.sps) • STATA (.dat/.DO) • R 	<ul style="list-style-type: none"> • SPSS Portable (.por) • SPSS (.sav) • STATA (.dta) • SAS (.7dat; .sd2; .tpt)
Imatge	<ul style="list-style-type: none"> • JPEG (.jpg, .jpeg) • TIFF (.tif, .tiff) • PNG (.png) • JPEG 2000 (.jp2) • DICOM (.dcm) 	
Imatges (Vectors)	<ul style="list-style-type: none"> • SVG (.svg) 	<ul style="list-style-type: none"> • Adobe Illustrator (.ai) • EPS (.eps) • WMF/EMF (.wmf, .emf) • CDR (.cdr)
Àudio	<ul style="list-style-type: none"> • BWF (.bwf) • MXF (.mxf) • Matroska (.mka) • FLAC (.flac) • OPUS 	<ul style="list-style-type: none"> • WAVE (.wav) • MP3 (.mp3) • AAC (.aac, .m4a) • AIFF (.aif, .aiff) • OGG (.ogg)

Vídeo	<ul style="list-style-type: none"> • MXF (.mxf) • Matroska (.mkv) 	<ul style="list-style-type: none"> • MPEG-4 (.mp4, .m4a, .m4v) • MPEG-2 (.mpg, .mpeg, .m2v, mpg2) • AVI (.avi) • QuickTime (.mov, .qt)
Bases de dades	<ul style="list-style-type: none"> • SQL (.sql) • SIARD (.siard) • CSV (.csv) 	<ul style="list-style-type: none"> • Microsoft Access (.mdb, .accdb) • dBase (.dbf) • HDF5 (.hdf5, .he5, .h5)
Disseny assistit per ordinador (CAD)	<ul style="list-style-type: none"> • AutoCAD DXF version R12 (ASCII) (.dxf) • SVG (.svg) 	<ul style="list-style-type: none"> • AutoCAD other versions than R12 (ASCII) (.dwg, .dxf) • DWG (.dwg) • DGN (.dgn)
Informació geogràfica	<ul style="list-style-type: none"> • GML (.gml) • MIF/MID (.mif/.mid) 	<ul style="list-style-type: none"> • Esri Shapefiles (.shp & related files) • MapInfo (.tab & related files) • KML (.kml) • Esri Geodatabase (.gdb) • Project files/Workspaces (.mxd, .wor, .qgs)
Imatges georeferenciades	<ul style="list-style-type: none"> • GeoTIFF (.tif, .tiff) 	<ul style="list-style-type: none"> • TIFF World File (.tfw & .tif, possibly with additional files) • JPEG World File (.jgw & .jpg, possibly with additional files) • ERDAS IMAGINE File Format (.img)
Raster GIS	<ul style="list-style-type: none"> • Raster GIS 	<ul style="list-style-type: none"> • Esri GRID (.grd & related files) • Surfer Grid (.grd; .srf) • ERDAS IMAGINE File Format (.img)
3D	<ul style="list-style-type: none"> • WaveFront Object (.obj) • Polygon file format (.ply) • X3D (.x3d) • COLLADA (.dae) 	<ul style="list-style-type: none"> • Autodesk FBX (.fbx) • Blender (.blend) • 3D PDF (.pdf)
RDF	<ul style="list-style-type: none"> • RDF/XML (.rdf) • Trig (.trig) • Turtle (.ttl) • NTriples (.nt) • JSON-LD 	
Computer Assisted Qualitative Data Analysis (CAQDAS)	<ul style="list-style-type: none"> • REFI-QDA (Qualitative Data Analysis) 	<ul style="list-style-type: none"> • ATLAS.TI Copy bundle • NVivo Project file

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