

Organitzar, anomenar i versionar fitxers

Introducció

 [English version](#)

Quan es creen o es recopilen dades i es treballa amb molts fitxers aquests es poden desorganitzar fàcilment. Per estalviar temps i evitar errors, és important prendre decisions sobre com s'organitzaran i s'anomenaran els fitxers i quin control de versions s'aplicarà tant si es treballa de manera individual com en grup.

Disposar d'unes convencions per a l'estructura i l'organització dels fitxers i les dades de recerca en millorarà la seva gestió i farà que les dades de recerca siguin més consistents. Si a més, aquestes es documenten, altres investigadors podran entendre-les i reutilitzar-les a curt, mitjà i llarg termini.

En aquest document trobareu pautes per organitzar, anomenar i versionar fitxers. És una informació que moltes entitats de finançament de la recerca demanen incloure en el Pla de Gestió de Dades.

Organitzar fitxers

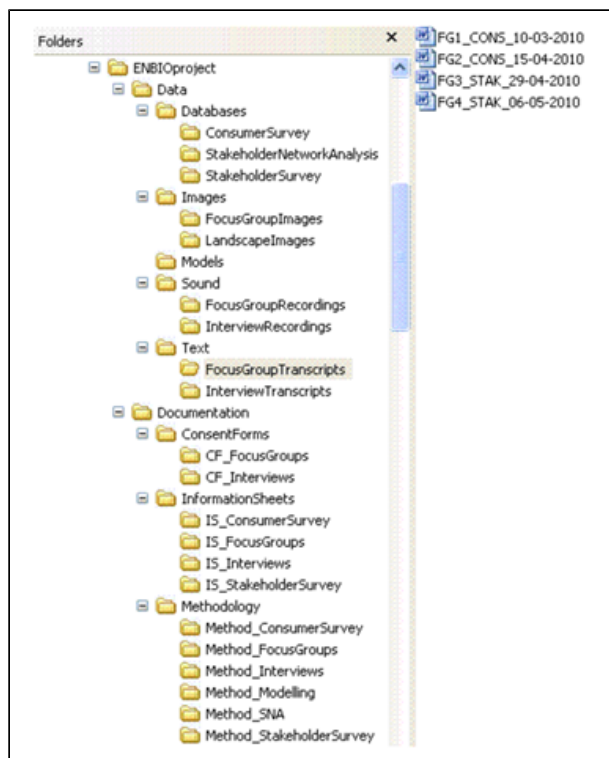
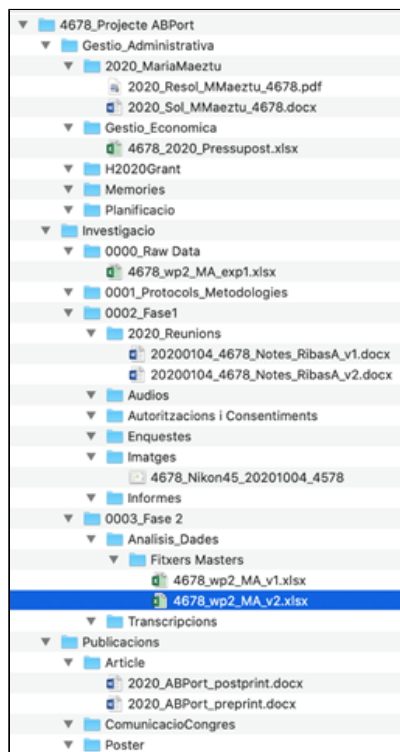
- Guardeu tots els documents d'un projecte en un mateix lloc, independentment del seu format i tipus, o com a mínim relacioneu-los.
- Trieu el lloc on gestionar i conservar els fitxers (serveis al núvol, servidors de la universitat, etc.) segons el tipus i l'ús que en fareu; no tots els llocs són adequats. Encara que treballeu des de diferents localitzacions i dispositius, establiu un únic lloc per arxivar-los tots i que aquest compleixi amb els requeriments de control, accés, seguretat i preservació necessaris.
- A l'inici de la recerca, definiu una estructura lògica de carpetes i subcarpetes per a que tots els membres del grup la utilitzin. No existeix un criteri únic d'organització de carpetes, dependrà del tipus, activitats i etapes del vostre projecte. Trieu el que millor s'adapti als vostres objectius i procediments de treball.

Es recomana una estructura jeràrquica de més general a més concreta, amb 3 o 4 sub-nivells ampliables si fos necessari. Per exemple:

- Podeu usar com a criteri les funcions per a les que es creen els fitxers: això és, un primer nivell per al nom del projecte i en un segon nivell, carpetes que agrupen els fitxers segons la seva finalitat d'ús (per a què s'ha creat el fitxer).
- Una altra opció seria un primer nivell per al nom de projecte i en un segon nivell, diferenciar entre dades i documentació, i en els nivells següents, per tipus i activitats a les que responen.

Completeu el directori de carpetes establint convencions per anomenar carpetes, subcarpetes i fitxers. Els noms de les carpetes han de reflectir el contingut, no la persona o equip que l'ha creat.

- Gestioneu acuradament les diferents versions dels fitxers, seguint les recomanacions del punt 4.
- Guardeu amb cura les dades originals (dades en brut o *raw data*) de la recerca:
 - Creeu una carpeta exclusiva configurada com a “només lectura” per a les dades en brut.
 - Utilitzeu còpies de les dades originals per treballar amb les dades i evitar perdre informació.
 - Donat que els fitxers es poden actualitzar contínuament i des de diferents dispositius, creeu una carpeta anomenada “ fitxers màsters” per guardar els fitxers definitius. Els fitxers amb els que trebal·leu han de ser tractats com a còpies temporals. Sincronitzeu-los o actualitzeu els fitxers màsters periòdicament o després de cada canvi.
 - Periòdicament feu còpies de seguretat de la carpeta amb els fitxers màsters, en diferents localitzacions.
- Assigneu metadades i elaboreu documentació explicativa sobre les dades, ja que resulta molt útil per poder entendre i interpretar-les sense haver d'obrir els fitxers. Fer-ho aporta valor afegit: en facilita la gestió, permet entendre el context en el que s'han generat les dades, trobar-les i reutilitzar-les. Existeixen tres tipus de metadades i documentació: administratives, tècniques i descriptives.
- Quan creeu un fitxer, utilitzeu formats estàndards per facilitar l'accés per part de tots els membres del projecte, la seva preservació i, posteriorment, la seva difusió i reutilització. Utilitzeu preferiblement formats oberts, no propietaris, compatibles amb diferents plataformes i àmpliament utilitzats. A la pàgina *Recomanacions de Formats* podeu ampliar aquesta informació.
- Controleu la seguretat de la informació i l'accés a les carpetes i fitxers al llarg de tot el projecte, ja que és essencial per complir amb la legislació vigent sobre protecció de dades i drets d'autor. Cal establir qui, quan i com pot accedir a cada carpeta i/o fitxer. Per això:
 - Acordeu en un document les condicions d'accés i ús per part de tots els membres del projecte.
 - Encripteu la carpeta, subcarpeta o fitxer, de manera que només els que coneguin el codi podran accedir-hi.
 - Establiu mesures de seguretat en el lloc on es guarden tots els fitxers: ús de programes antivirus i firewalls, xarxes segures, actualització contínua de programaris, còpies de seguretat periòdiques i contrasenyes d'accés.
- Establiu quins fitxers heu de conservar a llarg termini i durant quant de temps. Això dependrà de la finalitat del projecte i l'ús que es vulgui fer de les dades. Amb l'estructura de carpetes creada us serà més fàcil seleccionar els fitxers a conservar: la carpeta de “fitxers màsters” i la de “dades originals” les preservareu per sempre. En canvi, els documents que siguin esborranys, duplicats per a ús individual o fitxers temporals els podeu eliminar quan ja estiguin actualitzats o deixin de ser útils.



Font: elaboració pròpia

Font: <https://www.ukdataservice.ac.uk/manage-data/format/organising>

- Anomeneu els fitxers a partir d'elements com el nom o número del projecte, el nom del grup/equip de recerca, el tipus de mesura, el tema, la data de creació, el número de versió, etc.
- Els noms han de ser descriptius, clars i breus, màxim 25 caràcters. Han de mostrar la rellevància de cada document sense haver-los d'obrir.
- No utilitzeu paraules com: esborrany, presentació o full de càlcul.
- Utilitzeu noms i/o codis del més genèric al més específic per facilitar una ordenació senzilla dins les carpetes.
- Utilitzeu només caràcters dels conjunts A-Z, a-z, 0-9, guionet o guió baix. Eviteu l'ús de caràcters especials com ara &% \$ # : o bé)
- Si utilitzeu una numeració seqüencial, utilitzeu els zeros davant per garantir que els fitxers s'ordenin correctament: 0001, 0002 ... 1001, etc. i no: 1, 2 ... 1001, etc.
- Assegureu-vos que els noms de fitxers són independents de la seva ubicació per evitar problemes si es mouen els fitxers.
- En el cas de fotografies digitals i imatges escanejades, canvieu cada número de fitxer generat pel sistema a noms més significatius.
- Elaboreu un fitxer registre per explicar els elements codificats. Així tots els col·laboradors disposaran de la informació per establir els codis.

Exemples

- NTI_wp2_MA_exp1.csv: projecte, paquet de treball dins del projecte, tipus de mesurament, identificador de prova de la mesura
- MicroSensor_NTI023_20201031.xls: descripció de contingut, número del projecte, data seguint l'estàndard internacional
- NomExperiment_NomInstrument_DataCaptura_ImageID.tif

Versions d'un fitxer

Creeu una nova versió quan es produeixin canvis en l'estructura, el contingut o l'estat del recurs. En el cas de dades de recerca, creeu una nova versió d'un conjunt de dades quan processeu, corregiu o afegiu dades addicionals. El control de versions permet fer un seguiment dels canvis associats a dades dinàmiques.

La versió d'un fitxer es pot identificar mitjançant:

- La data registrada al nom del fitxer o dins del fitxer
- La numeració de versions al nom del fitxer (v1, v2, v3.3)
- La descripció de la versió al nom del fitxer o dins del fitxer (esborrany, final)
- L'historial de fitxers, la taula de control de versions o les notes incloses en un fitxer, on es registren versions, dates, autors i detalls dels canvis al fitxer

El control de versions també es pot fer mitjançant:

- L'ús de programaris de versions, per exemple GIT o Subversió (SVN)
- L'ús de serveis de compartició de fitxers com Dropbox, Google Docs o Amazon S3

Exemples

- 20200104_ProjecteA_Ex1Test2_RibasA_v1.xlsx
- 20200104_ProjecteA_ReunióNotes_RibasA_v2.1.docx

Fitxers duplicats

Quan es penja un fitxer, el programari de Dataverse realitza una verificació (o checksum) del nom i contingut del fitxer de manera que gestiona els fitxers duplicats de la següent manera:

- Els fitxers amb el mateix resultat de la verificació es poden incloure en un dataset, fins i tot si els fitxers es troben al mateix directori.
- Els fitxers amb el mateix nom es poden incloure en un dataset sempre que els fitxers estiguin en directoris diferents.
- Si un usuari carrega un fitxer a un directori on ja existeix un fitxer amb aquesta combinació de directori/nom de fitxer, Dataverse ajustarà la ruta i els noms del fitxer afegint "-1" o "-2" segons correspongui. Aquest canvi serà visible a la llista de fitxers que s'estan penjant.
- Si el directori o el nom d'un fitxer existent o recentment penjat s'edita de manera que es creï una combinació de directori/nom de fitxer que ja existeix, Dataverse mostrarà un error.
- Si un usuari intenta substituir un fitxer per un altre que tingui el mateix resultat en la verificació, es mostrarà un missatge d'error i el fitxer no es podrà substituir.
- Si un usuari intenta substituir un fitxer per un fitxer que tingui el mateix resultat en la verificació que un fitxer diferent del dataset, es mostrarà un avís.

Organitzeu els fitxers del projecte

Elimineu els fitxers obsolets periòdicament

Establiu el termini òptim de conservació dels fitxers

Feu còpies de seguretat

Controleu l'accés als fitxers i encripteu els que siguin necessaris

Bones pràctiques per a la vostra recerca

Trieu un únic lloc per guardar tots els fitxers



Definiu una estructura jeràrquica per crear carpetes i fitxers

Establiu convencions anomenar fitxers i carpetes

Controleu la versió dels fitxers

4678_NTI
Dades
NTI_wp2_MA_expl.xls
Publicacions
2020_NTI_Results_v1.doc
2020_NTI_Results_v2.doc

Per a més informació: <https://confluence.csuc.cat/x/wQReBQ>

  Grup de Treball de Suport a la Recerca del CSUC