

Dossier

Transformació digital

La gestió dels formats digitals:
una habilitat clau





Jorge FRANGANILLO
 Facultat d'Informació i Mitjans
 Audiovisuals, Universitat de
 Barcelona
franganillo@ub.edu

Article rebut a l'octubre de 2022;
 revisat al gener de 2023

DOI
<https://doi.org/10.34810/itemn74id417347>

La gestió dels formats digitals: una habilitat clau

Resum: Els formats digitals codifiquen i estructuren en llenguatge binari informació de morfologia variada. Cada format, però, està dissenyat per a un propòsit específic i, per tant, és vital conèixer-ne les propietats tècniques. La tria d'un format inadequat o la gestió deficient de les metadades poden comportar pèrdues irreversibles d'informació. Sobre aquesta premissa, l'article presenta la gestió dels formats digitals com una competència essencial dels professionals de la informació. Aquesta habilitat implica, entre altres tasques, determinar quin format és adient per a cada escenari d'ús, i prendre decisions ben fonamentades quant a les metadades, la nomenclatura dels fitxers i l'estructura de les col·leccions digitals.

Paraules clau: formats digitals, gestió de la informació, metadades, preservació digital.

La gestión de los formatos digitales: una habilidad clave

Resumen: Los formatos digitales codifican y estructuran en lenguaje binario información de variada morfología. Pero cada formato está diseñado para un propósito específico, por lo que es vital conocer sus propiedades técnicas. La elección de un formato inadecuado o la gestión deficiente de los metadatos pueden conllevar pérdidas irreversibles de información. Sobre esta premisa, el artículo presenta la gestión de los formatos digitales como una competencia esencial de los profesionales de la información. Esta habilidad implica, entre otras tareas, determinar qué formato es idóneo para cada escenario de uso y tomar decisiones bien fundamentadas en cuanto a los metadatos, la nomenclatura de los ficheros y la estructura de las colecciones digitales.

Palabras clave: formatos digitales, gestión de la información, metadatos, preservación digital.

The management of digital formats: a key skill

Abstract: Digital formats codify and structure information of varied morphology in binary language. Each format, however, is designed for a specific purpose, so knowing their technical properties is vital. The choice of an inadequate format or a poor approach to metadata management can lead to the irreversible loss of information. Based on this premise, this paper presents file format management as a core competence of information professionals. This skill involves, among other tasks, determining which format is suitable for each usage scenario and making well-grounded decisions regarding metadata, file naming conventions and the structure of digital collections.

Keywords: file formats, information management, metadata, digital preservation.



Introducció

Tant en l'àmbit personal com en contextos professionals, tothom gestiona informació valuosa que ara està continguda, cada cop més sovint, en fitxers digitals: dades, text, imatges, so, vídeo i formes derivades o combinades, com ara animacions, pàgines web, models tridimensionals o contingut audiovisual amb subtítols. Cada dia és habitual que hom faci servir sèries de dades, documents de text, fotografies i altre material gràfic, enregistraments sonors, seqüències de vídeo, etc.

Certa part del temps quotidià la dediquem, aleshores, a generar, guardar, consultar, consumir, gestionar o compartir tota mena de fitxers: tiquets de compra, fitxers musicals, dades de contacte, mems, captures de documents, missatges d'àudio, peces audiovisuals... La llista és llarga i heterogènia, i evidencia que la informació digital nodreix nombroses activitats diàries i, per descomptat, el coneixement humà. Tots aquests recursos tenen un valor cridat a perdurar i constitueixen un material prou digne de conservació en previsió de necessitats futures.

Ara, per tal que aquesta informació estigui completa i sempre disponible, cal emmagatzemar-la en un format adient i els fitxers que en resulten s'han de regir per una gestió adequada. A l'hora de guardar un document, convé atorgar-li una nomenclatura apropiada i coherent, triar-ne el format idoni, incloure-hi prou metadades i evitar decisions que després puguin complicar-ne l'accés o malmetre'n el contingut de manera irreversible. La gran oferta de formats i la dificultat per trobar informació rigorosa sobre aquests formats, però, poden complicar-ne la gestió. I és precisament aquí on el professional de la informació i la documentació torna a tenir un paper decisiu.

Cal entendre que estem immersos en un procés, del tot global, de transformació digital. En aquest context, ens cal assumir que ara tenim una competència nova, encara que tal vegada

es deriva d'una competència que ja teníem: la gestió de col·leccions digitals, que en un sentit ampli abasta les habilitats per gestionar els actius digitals d'una organització. Ara ens cal guanyar, o entrenar, o millorar, les habilitats que són relatives al bon ús dels formats digitals.

Si hem d'entendre, ras i curt, que la transformació digital és l'aplicació de les tecnologies digitals en contextos on abans no hi eren, o no hi eren en la magnitud que ara es demana, ens serà fàcil entendre que aquí som essencials. Ho som perquè la gestió dels formats digitals és una competència específica que ens interessa desenvolupar amb prou solvència professional. Aquesta habilitat implica, entre altres tasques, la de determinar quin format és adient per a cada escenari d'ús i la de prendre decisions ben fonamentades quant a les metadades, la nomenclatura dels fitxers i l'estructura de les col·leccions digitals. Sobre aquesta base, a continuació s'ofereix una aproximació al coneixement i la gestió dels formats digitals.

1. Els formats, contenidors d'informació

Un format digital és una convenció que codifica i estructura en sistema binari informació de qualsevol tipus. Conté, preserva i descriu la informació, de la mateixa manera que un envàs embollica, conserva i identifica productes de consum. El disseny físic dels envasos és, per tant, una eloqüent analogia de l'estructura lògica dels fitxers.

Igual que els envasos es fabriquen de materials diversos per poder manipular, utilitzar, protegir i transportar el seu contingut, els formats digitals codifiquen de manera diferent els diversos tipus d'informació perquè aquesta es pugui tractar, transmetre i preservar correctament. Així mateix, considerant que els recipients no només protegeixen i transporten el contingut, sinó que també identifiquen el producte i en

controlen la cadena de traçabilitat, aquestes són igualment funcions dels formats digitals, que compten amb mecanismes per conèixer la procedència, les característiques i la ubicació d'un fitxer. L'etiqueta d'informació nutricional d'un aliment envasat és un símil dels atributs i les metadades que descriuen propietats d'un fitxer digital.

Així, doncs, els formats digitals tenen usos i propietats similars a les d'objectes ben quotidians. Sabut això, és fàcil comprendre per què la informació digital s'articula en un nombre cada cop més gran de formats i per què cadascun preveu unes funcionalitats concretes: igual que al món físic no hi ha un envàs universal, en l'àmbit digital tampoc no hi ha un format polivalent. Cada situació d'ús condiciona quins requisits ha de reunir un format perquè la informació s'hi pugui desar, transportar i utilitzar de manera adequada, fiable i segura.

Cada format està dissenyat amb unes propietats tècniques que el fan idoni per a certes situacions d'ús. Per tant, emmagatzemar informació digital implica envasar-la en un «recipient» que aguantí bé el pas del temps i que a més estigui ben «etiquetat», és a dir, dotat de metadades suficients que després permetin fer cerques ràpides i operacions massives de selecció i organització.

2. Criteris per a la tria de formats

L'elecció del format adient per a un determinat escenari d'ús és una decisió crucial que pot deixar seqüeles a curt i llarg termini si es pren sense un criteri assenyat. A la curta, una tria incorrecta, com ara l'ús d'un format comprimit amb pèrdua en un context que ho desaconse-

Un format digital és una convenció que codifica i estructura en sistema binari informació de qualsevol tipus.

llaria, pot destruir el contingut de forma immediata i irreversible. I a la llarga, optar per un format propietari, d'escassa acceptació entre la comunitat usuària o desproveït d'un esquema de metadades pot comprometre el futur accés a la informació o la seva possible reutilització. Atès l'impacte de la tria, cal valorar les opcions disponibles i l'apreciació ha de basar-se en criteris objectius. En aquesta consideració entren en joc dos tipus de criteris: els factors de sostenibilitat, i els de qualitat i funcionalitat.

2.1. Factors de sostenibilitat

En valorar la viabilitat de determinats formats per a la preservació a llarg termini convé preveure els set factors de sostenibilitat que, segons la Library of Congress¹ i The National Archives,² influeixen en la viabilitat i en el cost de preservar informació davant els canvis que experimenta l'entorn tecnològic. Aquests factors clau són els següents:

- **Divulgació.** Grau de disponibilitat d'especificacions completes i d'eines per a la validació de la integritat tècnica del format.
- **Adopció.** Mesura en què autors, distribuïdors i usuaris utilitzen el format, ja sigui com a format mestre, per a la distribució a usuaris finals o com a mitjà d'intercanvi entre sistemes.
- **Transparència.** Grau en què la representació digital és oberta a una anàlisi directa amb eines bàsiques, inclosa la llegibilitat humana amb un editor de text pla.

1. Library of Congress. *Sustainability of digital formats: planning for Library of Congress collections*. Washington, DC: Library of Congress, 2017. <https://loc.gov/preservation/digital/formats/intro> [Consulta: 05/10/2022].

2. Todd, Malcolm. *File formats preservation*. Londres: The National Archives, 2009. <https://dpconline.org/docs/technology-watch-reports/375-file-formats-for-preservation/file> [Consulta: 05/10/2022].

- **Autodocumentació.** Capacitat per a la inclusió de metadades descriptives, tècniques o administratives.
- **Dependències externes.** Mesura en què el format depèn d'un maquinari, un sistema operatiu o un programari específic per poder-lo utilitzar, i nivell de complexitat que implicarà afrontar aquestes dependències en el futur.
- **Impacte de les patents.** Grau en què l'accés al contingut d'un format podria ser inhibit per patents de programari.
- **Mecanismes tècnics de protecció.** Aplicació de mecanismes restrictius, com ara el xifratge mitjançant contrasenya o els mètodes anticòpia, que impedeixen la conservació fiable de continguts digitals.

2.2. Factors de qualitat i funcionalitat

En l'elecció d'un format sovint entren en joc principis addicionals: els factors de qualitat i funcionalitat. Aquests elements permeten valorar la capacitat d'un format per representar trets rellevants que es poden necessitar o esperar en una peça de contingut. L'expectativa s'estableix sobre el que s'anomena *representació normal*, un punt de referència sobre el comportament bàsic del contingut. Com que cada forma d'expressió —text, imatge, so— té una naturalesa diferent, els criteris varien segons la morfologia de la informació.

2.2.1 Factors de qualitat

Els factors de qualitat rellevants en l'elecció d'un format digital els determina la morfologia de la informació i, en conseqüència, s'observen en termes d'integritat (text), claredat (imatge fixa o en moviment) o fidelitat (so).

La **integritat** és la mesura en què el format permet representar contingut textual de manera rica i expressiva. Els documents de text transmeten intencions comunicatives per mit-

jà de paràgrafs, llistes, títols i indicadors d'èmfasi (negretes i cursives). Aquests són els trets bàsics que conformen la representació normal d'un document de text. Hi ha, però, trets d'especial rellevància que poden condicionar la tria del format, com ara la qualitat expressiva de la tipografia, la integritat de la maquetació, la navegació i els diagrames, i la capacitat per representar fórmules i il·lustracions.

La **claredat**, d'altra banda, és el grau de nitidesa amb què el format permet representar contingut gràfic o audiovisual. Normalment s'associa a la resolució, la densitat de píxels, la profunditat de color i la gamma de colors. Són preferibles, per tant, els formats capaços d'evitar o minimitzar la pèrdua de qualitat causada per la compressió de dades. En les imatges en moviment, a més, són rellevants trets propis del contingut audiovisual, com ara la velocitat de fotogrames i l'entrellaçat, que afecten la fluïdesa i la nitidesa del moviment.

La **fidelitat**, per últim, és la capacitat del format per reproduir un senyal sonor sense distorsions. S'associa, sobretot, a la freqüència de mostreig i a la quantificació. I en sentit negatiu, hi influeix la presència de soroll, en particular els artefactes audibles que resulten de la compressió amb pèrdua. Segons el context, pot ser preferible un format codificat amb pèrdues de qualitat, per alleujar-ne el pes, o bé un format sense pèrdua, per millorar la riquesa de l'escolta.

2.2.2 Factors de funcionalitat

Alguns formats ofereixen atributs excepcionals capaços de satisfer necessitats especials (capacitats addicionals, major expressivitat) en la manipulació del contingut.

En els formats de text poden ser útils capacitats especials com, per exemple, el control de canvis, la possibilitat d'afegir-hi anotacions o marcadors, o la capacitat de construir taules de contingut, índexs o bibliografies de forma automàtica.

La prevalença d'uns criteris o uns altres la determina el propòsit de la informació i la fase del cicle de vida en què es trobi aquesta.

En els formats gràfics són opcions rellevants la inclusió de mecanismes de gestió del color (perfils de color, espais de treball), la capacitat per aplicar nivells de transparència, ombres, filtres o altres efectes, i la possibilitat de generar animacions o imatges multipàgina, de representar estructures tridimensionals o d'emmagatzemar freqüències d'ona alienes a l'espectre visible per recolzar l'anàlisi científica.

En els formats d'àudio es preveuen factors com ara la representació multicanal (estèreo, quadri-fònic, envoltant), la distribució en temps real (*streaming*) o la capacitat per allotjar llistes de reproducció, text d'audiollibres o descripcions.

Els formats de vídeo poden oferir trets significatius, com poden ser la capacitat per a diverses pistes d'àudio i de subtítols, la distribució i reproducció en temps real, la capacitat per a la negociació de continguts (per exemple, el lliurament d'una versió en la llengua de l'espectador) o l'escalabilitat, és a dir, l'ajust automàtic del contingut a les limitacions de la xarxa i del dispositiu reproductor.

3. Equilibrar els factors

A la pràctica, la preferència per un format o un altre implica trobar un equilibri adequat entre tots els factors —sostenibilitat, qualitat i funcionalitat—, sovint en competència. La prevalença d'uns criteris o uns altres la determina el propòsit de la informació i la fase del cicle de vida en què es trobi aquesta.

En efecte, la informació digital travessa un cicle des que neix fins que es difon, s'explota, s'arxiva o s'elimina. Aquest procés tendeix a

motivar la conversió entre formats per satisfer necessitats específiques de cada fase, tot i que en algunes situacions es pot transitar entre elles tot mantenint el format d'origen. Per exemple, una instantània presa amb el telèfon mòbil es pot modificar amb una aplicació de retoc i després es pot difondre als mitjans socials sense transitar entre formats, però allò guanyat en temps de processament s'haurà perdut en qualitat gràfica.

El cicle de vida de la informació digital preveu diverses etapes, cadascuna de les quals condiciona l'elecció dels formats i la gestió que cal fer-ne:

- La **creació** o **adquisició** és l'acte de generar un document digital. En aquesta fase convé optar per formats capaços de satisfer les exigències dels posteriors processos d'edició i postproducció, i cal tenir en compte qüestions d'arxiu i preservació a llarg termini (en especial, la dotació de metadades) que determinaran les futures possibilitats de recuperar i reutilitzar la informació.
- La **producció** i **postproducció** abasta qualsevol procés d'elaboració i manipulació del contingut com a fase prèvia a la difusió final. El tipus de tasques involucrades depèn de la morfologia de la informació. En la producció editorial són habituals les feines de revisió i maquetació del contingut, mentre que l'edició i la postproducció audiovisuals comprenen processos propis del vídeo digital, com ara el muntatge, la subtitulació, els efectes especials o la correcció de color. L'elecció del format, en aquest context, implica valorar quins trets rellevants (factors de qualitat i funcionalitat) ofereix cada opció disponible.
- La **difusió**, sigui interna o externa, implica la distribució del document per al seu ús o consum. Per tant, interessa prioritzar formats que gaudeixin d'un elevat grau d'adopció, ja que això facilita que el contingut es pugui utilitzar de forma generalitzada, sense barreres de programari.



— Les accions d'**arxiu** impliquen assegurar la preservació del material, amb independència del context en què es va originar. En aquesta fase prevalen els formats oberts que no degradin el contingut, que tinguin capacitat per allotjar metadades internes i que estiguin lliures de dependències externes o de patents de programari. S'afavoreix així la futura localització i reutilització del contingut.

Amb tot, alguns formats utilitzats per a la difusió són propietaris o bé apliquen compressió amb pèrdua per garantir una transmissió eficient per les xarxes. El grau d'adopció i la bona proporció de compressió prevalen aleshores sobre altres factors. En altres casos, la capacitat del format per oferir alguna funcionalitat específica pot prevaldre sobre els factors de sostenibilitat.³ Les plataformes de vídeo sota demanda, per exemple, lliuren el contingut amb la possibilitat de canviar l'idioma de l'àudio o el dels subtítols.

4. Nomenclatura, estructura i metadades

En l'organització i explotació d'una col·lecció digital és clau desenvolupar una nomenclatura que apliqui bones pràctiques, ja establertes. En la denominació dels fitxers, convé:

- assignar noms breus, però descriptius;
- emprar com a elements separadors guionets, guions baixos i inicials majúscules, en lloc de punts i espais;
- evitar caràcters especials, com ara lletres amb accent;
- incloure un número de versió, si escau;
- ordenar els elements del nom del més general al més específic, per facilitar l'agrupació de documents afins;
- representar les dates amb el format ISO 8601 (AAAA-MM-DD).

3. Arms, Caroline; Fleischauer, Carl. "Factors to consider when choosing digital formats". A: Kresh, Diane (ed.). *The whole digital library handbook*. Chicago: American Library Association, 2007.

Sense el format que pertoqui, la informació es perd al bosc immens del fet digital, i encara que no sigui prou visible, hi és, hi ha d'ésser.

Ahora, convé que el criteri de nomenclatura sigui uniforme i que estigui descrit al pla de gestió de dades de la institució.⁴ Pel que fa a l'estructura de carpetes, convé que sigui poc profunda per estalviar temps de navegació, tant en l'emmagatzemament com en la recuperació de documents.⁵

Les metadades, per la seva banda, són declaracions estructurades que ajuden a identificar, localitzar i gestionar els fitxers d'una col·lecció digital. Per facilitar aquest triple propòsit interessa nodrir els documents amb metadades suficients. El valor d'un document augmenta com més metadades s'hi consignen, ja que faciliten operacions massives de cerca, selecció i organització.^{6,7}

5. Problemes i solucions

En l'actual context de transformació digital, a més de dominar elements que últimament han guanyat protagonisme, com ara la signatura electrònica i els certificats digitals, cal mantenir una actitud activa d'aprenentatge continu amb la qual es pugui anar més enllà del bon saber, prou necessari, de les noves eines i aplicacions. La signatura electrònica i el certificat digital substitueixen avui allò que semblava immortal, i que de fet és mil·lenari: la firma d'una persona en un document. Avui, en la nova onada de digitalització de processos en què estem, hi

ha mecanismes robustos que certifiquen l'autenticitat i la integritat d'un document. No n'hi ha prou, però. Cal continuar aprenent. Aquest aprenentatge continu implica conèixer, ara, quines perspectives, quines funcionalitats ofereix cada format digital, quin ús n'està fent la comunitat usuària i quins aspectes tècnics són rellevants en la gestió de col·leccions digitals.

Sense el format que pertoqui, la informació es perd al bosc immens del fet digital, i encara que no sigui prou visible, hi és, hi ha d'ésser. Des que els sistemes operatius amaguen per defecte l'extensió dels documents, s'ha diluït en part la noció de com n'és de crític «envasar» la informació en el «recipient» adequat per tal d'evitar pèrdues irreversibles. Tot i que el bosc és vast i força heterogeni, la resposta als reptes i interrogants que ens planteja aquest nou i complex escenari rau en la importància de no pretendre abastar molts formats, ni de fer tasques tedioses que impliquen més risc d'error, sinó de fer accions senzilles i pràctiques, fonamentades i coherents, amb el format digital més adequat. Aquestes accions en són garantia.

Un altre repte en la gestió de formats digitals és la recomanació de parar atenció en dos riscos: l'obsolescència i la proliferació. Sobre l'obsolescència s'ha de recordar que qualsevol format, obert o propietari, n'és vulnerable, ja sigui perquè els proveïdors de programari sovint recorren a l'obsolescència programada per estimular la venda de nous productes, o ja sigui perquè les comunitats de programari de codi obert poden deixar de promoure formats antics si la comunitat usuària ja no els demana. I sobre la proliferació de formats digitals s'ha de recordar que aquest fet, sobretot a les organitzacions, pot complicar la gestió documental

4. Briney, Kristin. *Data management for researchers: organize, maintain and share your data for research success*. Exeter: Pelagic, 2015.
5. Santaguada, Vincent. "Folder and file naming convention: 10 rules for best practice". *Exadox*. (2010), <https://exadox.com/en/articles/file-naming-convention-ten-rules-best-practice> [Consulta: 05/10/2022].
6. Lubas, Rebecca L.; Jackson, Amy S.; Schneider, Ingrid. *The Metadata Manual*. Oxford: Chandos, 2013.
7. Baca, Murtha (ed.). *Introduction to metadata*. 3^a ed. Los Angeles: Getty Research Institute, 2016. <https://getty.edu/publications/intrometadata> [Consulta: 05/10/2022].

perquè administrar molts formats, fins i tot en versions diverses, fa difícil identificar allò que està en risc, o esbrinar què és el millor per sortir-se'n.⁸

Ahora, cal tenir present que els sistemes digitals no són infal·libles. La informació digital roman exposada als errors de programari o de maquinari, a una possible alteració accidental, a un desastre natural o al mal ús derivat de la debilitat o la maldat humanes. Es pot millorar el control sobre les col·leccions si es prenen mesures adequades i proactives de preservació: cal crear, periòdicament, còpies de seguretat de les dades.

Per últim, en considerar que la transformació digital implica un repte que cal assumir, qui vulgui aprofundir en la correcta gestió dels actius digitals, en trobarà una bona aproximació al manual que l'autor d'aquestes pàgines tot just ha publicat sobre la gestió estratègica dels formats.⁹ Pioner en aquesta temàtica en castellà, el manual ofereix claus per determinar quin és el format digital més adequat per generar, manipular, distribuir o arxivar una peça d'informació segons l'ús al qual estigui destinada. Explica també les implicacions de convertir un fitxer a un altre format per adaptar-lo a un escenari diferent tot conservant-ne les propietats més rellevants. Perquè, al cap i a la fi, no n'hi ha prou de crear una bona peça d'informació i d'assegurar-li llarga vida, sinó que només es podrà dir que la feina ha estat ben feta quan aquesta informació arribi sana i estàlvia al seu destí.

6. Bibliografia

ARMS, Caroline; FLEISCHHAUER, Carl. "Factors to consider when choosing digital formats". A: Kresh, Diane (ed.). *The whole digital library handbook*. Chicago: American Library Association, 2007.

BACA, Murtha (ed.). *Introduction to metadata*. 3rd ed. Los Angeles: Getty Research Institute, 2016. <https://getty.edu/publications/intrometadata> [Consulta: 05/10/2022].

BRINEY, Kristin. *Data management for researchers: organize, maintain and share your data for research success*. Exeter: Pelagic, 2015.

Digital Preservation Coalition. "File formats and standards". A: *Digital preservation handbook*. Glasgow, York i Melbourne: Digital Preservation Coalition, 2021. <https://dpconline.org/handbook/technical-solutions-and-tools/file-formats-and-standards> [Consulta: 05/10/2022].

FRANGANILLO, Jorge. *Formatos digitales: propiedades técnicas y contextos de uso*. Barcelona: UOC, 2022.

Library of Congress. *Sustainability of digital formats: planning for Library of Congress collections*. Washington, DC: Library of Congress, 2017. <https://loc.gov/preservation/digital/formats/intro> [Consulta: 05/10/2022].

LUBAS, Rebecca L.; JACKSON, Amy S.; SCHNEIDER, Ingrid. *The Metadata Manual*. Oxford: Chandos, 2013.

SANTAGUIDA, Vincent. "Folder and file naming convention: 10 rules for best practice". *Exadox*. (2010), <https://exadox.com/en/articles/file-naming-convention-ten-rules-best-practice> [Consulta: 05/10/2022].

TODD, Malcolm. *File formats preservation*. Londres: The National Archives, 2009. <https://dpconline.org/docs/technology-watch-reports/375-file-formats-for-preservation/file> [Consulta: 05/10/2022]. ■

8. Digital Preservation Coalition. "File formats and standards". A: *Digital preservation handbook*. Glasgow, York i Melbourne: Digital Preservation Coalition, 2021. <https://dpconline.org/handbook/technical-solutions-and-tools/file-formats-and-standards> [Consulta: 05/10/2022].

9. Franganillo, Jorge. *Formatos digitales: propiedades técnicas y contextos de uso*. Barcelona: UOC, 2022.