
Gestió de dades en la recerca

Eduard Sala
22/2/2024

Doctorat
Departament de Geografia

UAB

Universitat Autònoma
de Barcelona

01

Temes

UAB

Universitat Autònoma
de Barcelona

Temes

- **Tipus de dades**
- **Fonts de les dades**
- **Ciència oberta**
- **Organització**
- **Eines de gestió**
- **Recursos en línia**
- **Seguretat**

02

Tipus de dades

UAB

Universitat Autònoma
de Barcelona

Tipus de dades

- Elegir format
- Usabilitat, accessibilitat, sostenibilitat
- Compatibilitat: readme.txt
- TIPUS: observacionals (en temps real com enquestes), experimentals (en laboratori, com seqüències de gens), simulació (a partir de models, com climatològics) i derivades (més difícil reproduir, models 3D)
- Tipus especials de dades: personals, sensibles (ètnia, política...), i confidencials (secrets comercials, industrials...).

Tipus de dades	Format recomanat per compartir, reutilitzar i preservar	Format habitual acceptat
Text	• PDF/A (.pdf) • ODT (.odt)	• Microsoft Word (.doc) • Office Open XML (.docx) • Rich Text File (.rtf) • PDF other than PDF/A (.pdf)
Text pla	• Unicode text (.txt)	• Non-Unicode text (.txt)
Llenguatges de marcatge	• XML (.xml) • HTML (.html) • Related files: .css, .xslt, .js, .es	• SGML (.sgml) • Markdown (.md)
Llenguatges de programació	• MATLAB • NetCDF • TextFabric	
Fulls de càlcul	• ODS (.ods) • CSV (.csv)	• Microsoft Excel (.xls) • Office Open XML Workbook (.xlsx) • PDF/A (.pdf)
Bases de dades	• SQL (.sql) • SIARD (.siard) • CSV (.csv)	• Microsoft Access (.mdb, .accdb) • dBase (.dbf) • HDF5 (.hdf5, .he5, .h5)
Dades estadístiques	• SPSS (.dat/.sps) • STATA (.dat/.DO) • R	• SPSS Portable (.por) • SPSS (.sav) • STATA (.dta) • SAS (.7dat; .sd2; .tpt)

03

Fonts de les dades

UAB

Universitat Autònoma
de Barcelona

Fonts de les dades

- CGPJ: [https://www.poderjudicial.es/cgpi/es/Poder Judicial](https://www.poderjudicial.es/cgpi/es/Poder_Judicial)
- INE: <https://www.ine.es/>
- Data Hippo: <https://datahippo.org/es/>
- Google: <https://datasetsearch.research.google.com/>
- Gobierno de España: <https://datos.gob.es/>
- Idescat: <https://www.idescat.cat/>
- Generalitat de Catalunya: https://governobert.gencat.cat/ca/dades_obertes/inici/
- Barcelona Open Data: <https://iniciativabarcelonaopendata.cat/ca/inici/>
- Open Data Ajuntament de Barcelona: <https://opendata-ajuntament.barcelona.cat/>

Fonts de les dades

- Format: [DOI Citation Formater](#), [Cornell University Citation Management](#), [Generador APA](#)
[Scribbr](#)
- Llicència: <https://ufal.github.io/public-license-selector/>
- TOT: ddd.uab.cat [Bones pràctiques en gestió de dades de recerca](#)

04

Ciència Oberta

UAB

Universitat Autònoma
de Barcelona

Ciència Oberta

“Ciència oberta descriu un moviment en curs sobre la manera com es realitza la recerca, com col·laboren els investigadors, es comparteix el coneixement i s'organitza la ciència. Afecta tot el cicle de la recerca i els seus grups d'interès, millora la ciència facilitant més **transparència, obertura, treball en xarxa i col·laboració.**”

Font: OpenAIRE <https://www.openaire.eu/what-is-open-science> 19/11/2021

Ciència Oberta: antecedents

- European Commission (2013). Digital science in Horizon 2020. March.
 - Document conceptual sobre el moment actual de la ciència; la idea: internet + tecnologies estan canviant radicalment la forma com es crea i difon el coneixement, això ha de sustentar-se en un entorn científic col·laboratiu i obert, i la nova manera de fer ciència permet que aquesta s'acosti a la societat.
- European Commission (2014). Public consultation: 'Science 2.0': Science in transition
 - Cal un nom global per a molts canvis, propostes considerades e-Science, open science, science2.0, digital science.
- European Comission (2016). Open innovation, open science, open to the world: A vision for Europe. Brussels: European Commission. Directorate-General for Research and Innovation.
 - Ciència oberta es relaciona amb innovació, benestar, nou model econòmic per Europa
- Recomendación de la UNESCO sobre la Ciencia Abierta (2021). UNESCO

Ciència Oberta: motivacions

- Científiques
 - La ciència millora quan és oberta més ràpida, reutilitzable
- Ètiques
 - Perquè el coneixement científic és un bé públic. El que està finançat públicament (de forma majoritària) ha de ser públic
- Econòmico-polítiques
 - Per al benestar futur cal un canvi de model econòmic. De societat industrial a societat informacional (basada en el coneixement)
- Per donar resposta als grans reptes socials del nostre temps
 - Més informació, implicació i participació ciutadanes, per transparència i control



Ciència Oberta: La investigació i les dades de recerca accessibles i obertes a tots els ciutadans

Open Science

Open Repositories
Repositoris oberts



Open Access
Accés obert

Accés sense barreres econòmiques, tecnològiques o jurídiques a les publicacions científiques

Open Access Journals
Revistes d'accés obert



Open Peer Review
Revisió per parells oberta



Open Science Evaluation
Avaluació de la ciència en obert

Avaluació oberta dels resultats d'investigació, ampliant la revisió tradicional amb la contribució de la comunitat

Open Metrics and Impact
Impacte i mètriques obertes



Open Reproducible Research
Investigació reproducible en obert
Accés lliure als elements experimentals per a la reproducció de la investigació



Open Research Data
Dades d'investigació obertes



Open Source in Open Science
Codi obert per a la ciència oberta

Open Data
Dades obertes
Dades disponibles en línia de forma gratuïta i que es poden fer servir, reutilitzar i distribuir



Open Big Data
Dades massives obertes



Open Government Data
Dades governamentals obertes



El personal de la teva Biblioteca et pot assessorar



Ciència Oberta: publicació

Daurada Revistes d'accés obert amb CC amb **cost pels autors**

Híbrida Revistes de subscripció amb **cost pels autors** si volen obrir els articles

Diamant Revistes d'accés obert amb CC sense cost pels autors

Ex.: revistes Servei de Publicacions de la UAB

Verda Repositoris institucionals o temàtics

Bronze Revistes amb accés gratuït sense cap CC. No és pròpiament accés obert

Plataformes de publicació Publicació de l'article en accés obert (sense passar per una revista) Ex.: Plataforma ORE (Open Research Europe)

Ciència Oberta: PDI UAB



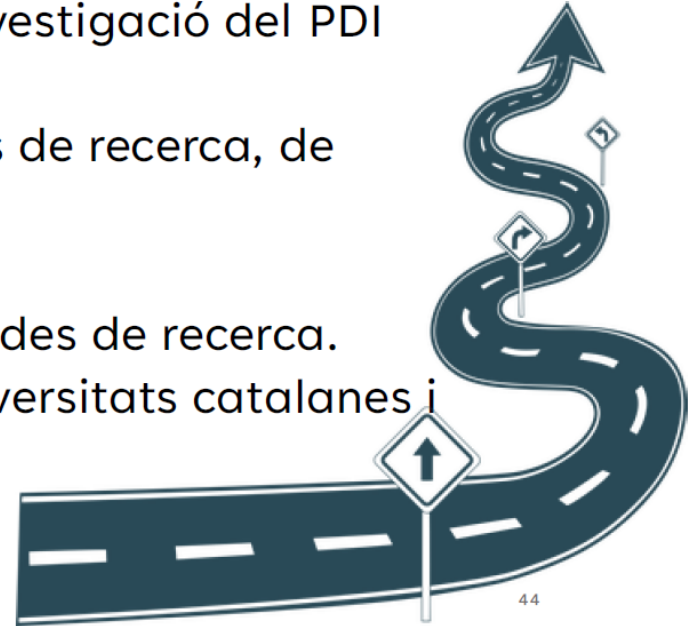
EGRETA : per referenciar els resultats de la investigació del PDI



DDD : per recollir i donar accés als documents de recerca, de docència i al fons patrimonial de la UAB.



CORA RDR : per recollir i donar accés a les dades de recerca. Repositori del CSUC que s'ofereix a les 34 universitats catalanes i centres CERCA.



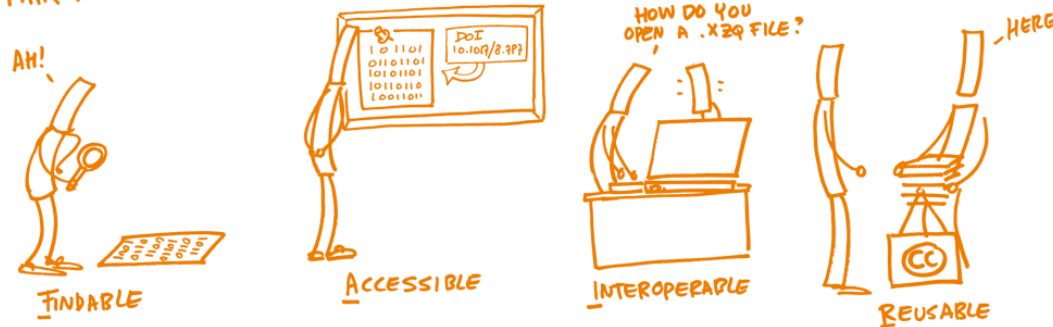
UAB

Universitat Autònoma
de Barcelona

Ciència Oberta: FAIR

- Findable (que es pugui trobar): DOI, repositoris amb estàndards internacionals.
- Accessible: gràcies als repositoris; les metadades sempre han de ser accessibles encara que les dades ja no estiguin disponibles.
- Interoperable: nomenclatures, codis controlats i segons estàndards.
- Reusable: bona descripció de les dades i amb les llicències corresponents.
- [RECOMANACIONS](#)

FAIR DATA PRINCIPLES



As Open as Possible, as Closed as Necessary



1. Findable

Identificador persistent



Identificació del arxius

Example:

Nom_<AAAAAMDD>_<versió núm.>

Prova_20231025_v01.csv

Metadades

 **DublinCore**



Etc.

Exemples:

CORA RDR: doi.org/10.34810/data674

Zenodo: doi.org/10.5281/zenodo.4571628

Paraules clau

Multidisciplinars



[Etc.](#)

Temàtics



[Etc.](#)

OAI-PMH

Open Archive Initiative-Protocol for Metadata Harvesting



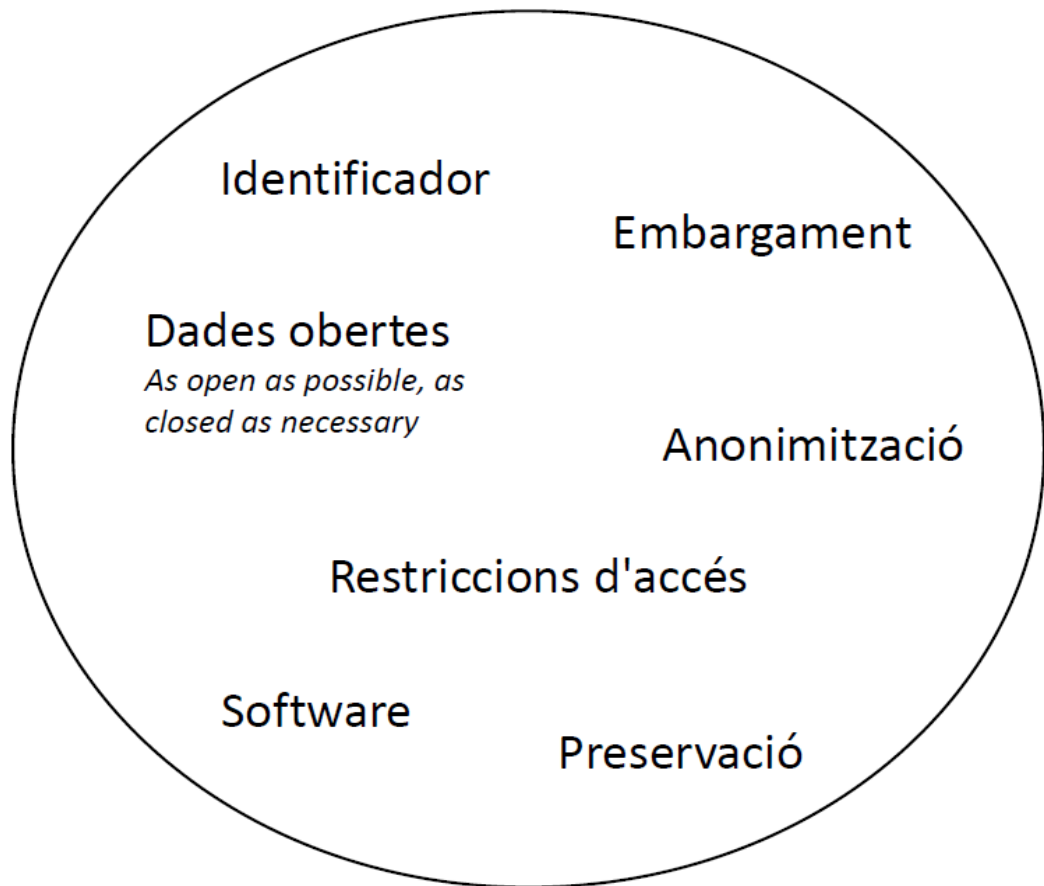


2. Accessible

Repositoris
de
confiança



Etc.





3. Interoperable

Quins
vocabularis, estàndards,
formats o metodologies
de
Dades i/o metadades
permetran
Interoperabilitat

ALGUNS EXEMPLES

OAI-PMH

Protocol d'Internet

.csv

.pdf

.txt

.jpg

Formats d'arxius

ASCII

Codi de caràcters

Estàndards de:

- Dates
- Temperatures
- Geolocalitzacions
- Etc.

MARC
Dublin Core

Estàndards de metadades



4. Reusable

Reutilització de les dades

[Readme.txt](#)

Llicències Creative Commons

(Cal comprovar requeriments de la convocatòria)



Ciència Oberta: DMP

Data Management Plan, actualitzable: [Exemple](#)

- ✓ Document formal que descriu què es farà amb les dades durant i després d'un projecte de recerca
- ✓ Ha de contenir informació sobre:
 - Quines dades s'obtidran per desenvolupar la recerca?
 - Com es recolliran i es processaran?
 - Quins estàndards i quina metodologia se seguiran?
 - Qui podrà accedir a les dades i com es difondran?
 - Com es preservaran quan la recerca finalitzi?
- ✓ Tot i tenir continguts similars, cal seguir els requeriments (o plantilles) de cada agència de finançament
- ✓ Un DMP és un document viu! El procés de recerca requereix que modifiqui la idea inicial i, per tant, cada canvi en la recerca, caldrà revisar i adaptar el DMP

DCC Checklist	DCC Guidance and questions to consider
Administrative Data	
ID	A pertinent ID as determined by the funder and/or institution.
Funder	State research funder if relevant
Grant Reference Number	Enter grant reference number if applicable [POST-AWARD DMPs ONLY]
Project Name	If applying for funding, state the name exactly as in the grant proposal.
Project Description	Questions to consider: - What is the nature of your research project? - What research questions are you addressing? - For what purpose are the data being collected or created? Guidance: Briefly summarise the type of study (or studies) to help others understand the purposes for which the data are being collected or created.
PI / Researcher	Name of Principal Investigator(s) or main researcher(s) on the project.
PI / Researcher ID	E.g ORCID http://orcid.org/
Project Data Contact	Name (if different to above), telephone and email contact details
Date of First Version	Date the first version of the DMP was completed
Date of Last Update	Date the DMP was last changed
Related Policies	Questions to consider: - Are there any existing procedures that you will base your approach on? - Does your department/group have data management guidelines? - Does your institution have a data protection or security policy that you will follow? - Does your institution have a Research Data Management (RDM) policy? - Does your funder have a Research Data Management policy? - Are there any formal standards that you will adopt? Guidance: List any other relevant funder, institutional, departmental or group policies on data management, data sharing and data security. Some of the information you give in the remainder of the DMP will be determined by the content of other policies. If so, point/link to them here.
Data Collection	
What data will you collect or create?	Questions to consider: - What type, format and volume of data? - Do your chosen formats and software enable sharing and long-term access to the data? - Are there any existing data that you can reuse?

Ciència Oberta

- 4 pilars de la ciència oberta: accés obert, avaluació de la ciència en obert, investigació reproduïble en obert, dades obertes.
- Començar a treballar les dades des del moment 0, establint el pla de gestió de dades en els primers 6 mesos pels HEU.
- El Open Research Europe és com un gran journal: plataforma on només hi poden publicar articles els qui tenen un HEU, avaluació en obert en dues, actua com una revista multidisciplinària
- Si son dades sensibles millor passar abans pel comitè d'ètica.
- Es demanarà en un futur que les tesis per compendi d'articles han de ser en accés obert.
- CORA cal posar el readme, dades, metadades i les llicències.
- Posar tot en un sol repositori, si està a més d'un es poden perdre citacions.
- A la web de biblioteques de la UAB, de Open Access, hi ha un apartat que pots buscar conjunts de dades, per veure si ja n'hi ha de similars.
- CURAR dades: Es compleixen els criteris requerits pel repositoris. Idealment es revisen els fitxers i les metadades; s'intenta entendre el contingut dels fitxers i s'executa el codi (si n'hi ha); es demana més informació si és necessària; s'afegeixen més metadades; es transformen els formats, si escau; i s'avalua la FAIRness.

Ciència Oberta: recursos

- Cursos: <https://data.europa.eu/en/academy>
- Portal UAB: <https://www.uab.cat/ca/ciencia-oberta>
- Publicar UAB: <https://www.uab.cat/web/comunicacio-de-la-recerca/Publicar-en-acces-obert-1345787284943.html>
- Producció en obert: <https://www.uab.cat/web/recerca/open-access-uab/produccio-cientifica-en-obert-1345691507345.html>
- Llicències UAB: <https://ddd.uab.cat/record/129205>
- Open Research Europe: <https://open-research-europe.ec.europa.eu/>
- European Open Science Cloud: <https://eosc-portal.eu/>
- Eina DMP: <https://dmp.csuc.cat>
- Respositori de dades: <https://dataverse.csuc.cat/>
- Gestio de DMP: <https://argos.openaire.eu/splash/>
- Zenodo: <https://zenodo.org/>



En defensa de los Repositorios de Acceso Abierto



Manifiesto y declaración de REBIUN en defensa de los repositorios de acceso abierto

<https://hdl.handle.net/20.500.11967/1347>

Utiliza el repositorio de tu institución para:



Depositar el preprint de tu artículo



Depositar tu artículo en acceso abierto si dispone de una licencia Creative Commons



Depositar tu versión revisada de autor (postprint) en acceso abierto o con embargo, de acuerdo con la política editorial

De esta forma garantizas:



El cumplimiento legal establecido tanto por la LCTI como por la LOSU

SEXENIOS

El cumplimiento de los requisitos de convocatorias de sexenios y acreditación



El cumplimiento de las convocatorias de proyectos de investigación



Mayor difusión de tus resultados de investigación



Una ciencia abierta y transparente

Desde REBIUN promovemos:

1 La inclusión de cláusulas de retención de derechos (retention rights) en las contrataciones de acuerdos transformativos con las editoriales.

2 Una mayor transparencia de las editoriales en la justificación de la distribución de costes y beneficios que afectan a la publicación en acceso abierto.

3 Nuevos desarrollos legislativos en España a favor de la ciencia abierta que faciliten el acceso abierto inmediato, sin periodos de embargo, a los resultados de investigación financiados con fondos públicos.

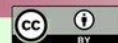


Si la editorial te ofrece mediante el pago de una tasa (por ejemplo el Article Development Charge (ADC)), el depósito y acceso abierto inmediato en un repositorio eliminando así el periodo de embargo que pueda afectar a tus versiones de autor, rechaza esta opción ya que en la práctica supone un doble pago innecesario a la editorial y un recorte en tus derechos de autor.

AT



El personal de tu Biblioteca te puede asesorar



05

Organització

UAB

Universitat Autònoma
de Barcelona

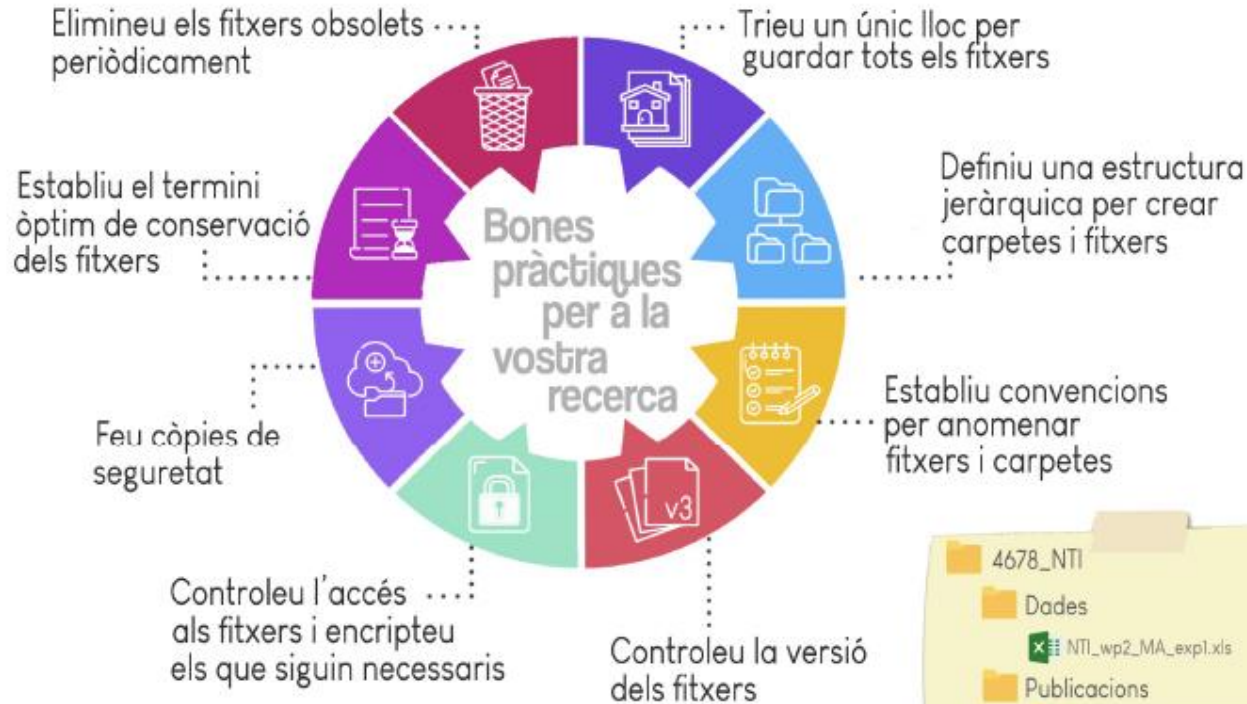
Organització

- Un mateix lloc
- Estructura
 - 3-4 subnivells
 - Noms de carpeta per contingut
 - Carpeta per dades en brut: només de lectura
 - **Carpeta per dades definitives:** actualitzable i més seguretat
- Treballar amb còpies de les originals
- Fitxers de metadades

Organització

- Noms
 - Màxims 25 caràcters
 - A-Z 0-9 _ sense caràcters especials ni espais
 - 000 al davant
 - Sense vinculació amb el lloc de desament
 - Dates inverses: 20220214
 - Versions: al nom o a l'historial
- [Endnote, Mendeley, Zotero](#)

Organitzeu els fitxers del projecte



Per a més informació: <http://hdl.handle.net/2072/377584>



Grup de treball de Suport a la Recerca del CSUC

06

Eines de gestió

UAB

Universitat Autònoma
de Barcelona

Eines de gestió

- EXCEL
 - Filtres
 - Ordenació
 - Dreceres
 - Taules dinàmiques
 - Gràfiques
 - Webs: [abreviatures](#) i [fórmules](#)
- Accés, SPSS: Valors perduts, dades, variables, anàlisis i sintaxis



07

Recursos en línia

UAB

Universitat Autònoma
de Barcelona

Recursos en línia

- Tableau: <https://public.tableau.com/s/>
- Open Access journals: <https://doaj.org/>
- RACO: <https://raco.cat/raco/>
- DDD UAB: <https://ddd.uab.cat/>
- Geografia:
 - Carto: <https://carto.com/>
 - Instamaps: <https://www.instamaps.cat/index.html#/>
 - Pintamaps: <https://betaserver.icgc.cat/pintamaps/#14/41.9715/2.8068>
 - Story Map: <https://storymap.knightlab.com/>
 - Trendsmap: <https://www.trendsmap.com/>
 - Google N-gram: <https://books.google.com/ngrams#>

08

Seguretat

UAB

Universitat Autònoma
de Barcelona

Seguretat

- Múltiples còpies
- Núvols vs. Privacitat
- Encriptament: 7Zip, Office, Bitwarden (o similar)
- [Virus Total](#)



UAB

Universitat Autònoma
de Barcelona