

Gestió de les dades de recerca: resultats de l'enquesta prospectiva a gener de 2016

(Doc.16/56) (B6SR\GDR\EnquestaPilot\1610 ResultatsEnquestaProspectivaH2020.doc ; GTSR 22.11.16)

El Grup de Treball de Suport a la Recerca del CSUC format per representats de la 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 i Universitat Jaume I va elaborar entre l'octubre i el desembre de 2015 una enquesta destinada als responsables dels diferents projectes de les Universitats de Catalunya que rebien finançament del programa Horitzó 2020 de la Comissió Europea, independentment de si formaven part o no de l'*Open Research Data Pilot*, amb la finalitat de conèixer les necessitats dels investigadors en matèria de gestió de dades de recerca.

1. Mostra

L'enquesta va ser enviada des de les diferents biblioteques i oficines de recerca de les universitats de Catalunya entre desembre de 2015 i gener de 2016 a una mostra de 164 projectes europeus de diferents àmbits de coneixement.

Aquest informe està fet sobre la base de les 73 respostes rebudes fins al 31 d'agost de 2016 i, tot i que el grau de resposta és desigual entre les diferents universitats (vegeu taula 1), suposa un percentatge de resposta del 45% .

Universitat	Nº enquestes enviades	Nº de respostes rebudes	Percentatge
UB	40	8	20%
UAB	24	14	58%
UPC	33	12	36%
UPF	25	9	36%
UdG	12	6	50%
UdL	8	7	88%
URV	13	11	85%
UOC	2	2	100%
UVic-UCC	2	0	0%
URL	5	4	80%
Total	164	73	45%

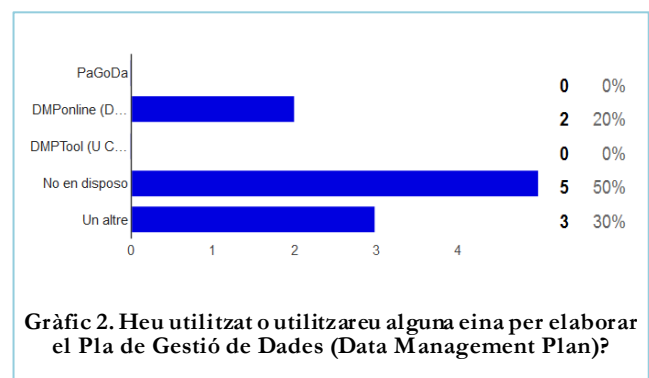
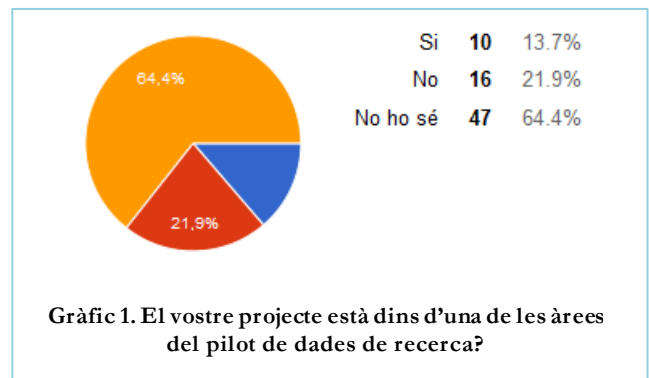
Taula 1. Nombre d'enquestes enviades i respostes rebudes per universitat

2. Anàlisi dels resultats

Sobre el projecte de recerca

Pel que fa a la pregunta inicial, més d'un 60% dels investigadors desconeixien si formaven part d'una de les àrees del pilot de dades (vegeu gràfic 1).

De les deu preguntes afirmatives, les àrees representades varen ser: “*Future and Emerging Technologies*”, “*Leadership in enabling and industrial technologies – Information and Communication Technologies*”, “*Nanotechnologies, Advanced Materials, Advanced Manufacturing and Processing, and Biotechnology*”, “*Societal Challenge: Climate Action, Environment, Resource Efficiency and Raw materials – except raw materials*”, “*Societal Challenge: Europe in a changing world – inclusive, innovative and reflective Societies*” i “*Science with and for Society*”. Cap d'aquests projectes van optar per excloure's del pilot i la meitat no disposaven d'una eina concreta per elaborar el Pla de Gestió de Dades (vegeu gràfic 2).



Pel que fa als 63 projectes que no formaven part d'una de les àrees del pilot o ho desconeixien, més del 73% no tenien previst elaborar un PGD (vegeu gràfic 3) tot i que més del 74% afirmaven produir dades de recerca.

Finalment, tan sols 16 projectes de les 73 respostes rebudes van afirmar que no produïen dades de recerca i van finalitzar l'enquesta en aquest punt.

La resta de projectes van continuar l'enquesta responnent qüestions sobre les dades de recerca, la difusió de les dades, la preservació de les dades i l'ús dels estàndards i la responsabilitat de les dades.

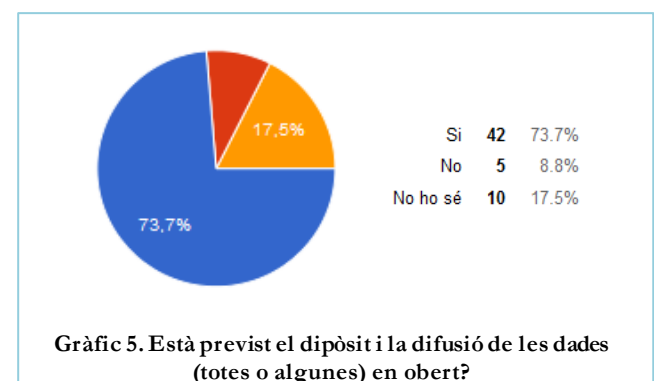
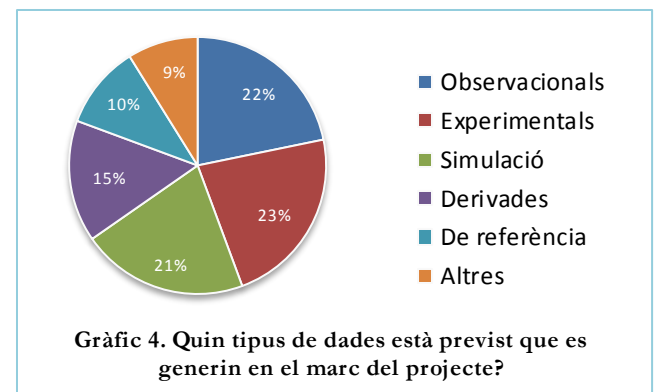
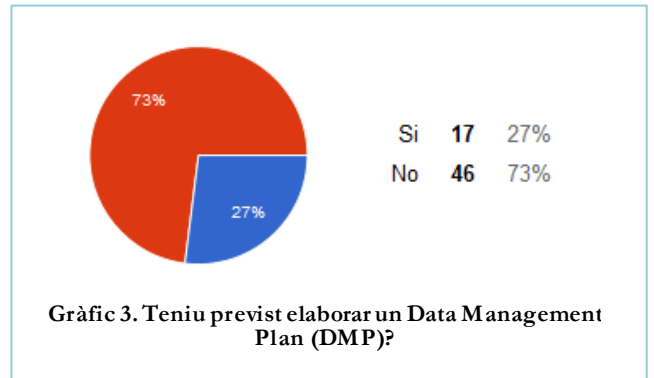
Sobre les dades de recerca

La tipologia de dades observacionals, experimentals i de simulació abasten més del 60% de dades que produeixen els grups d'investigació (vegeu gràfic 4).

Difusió de les dades

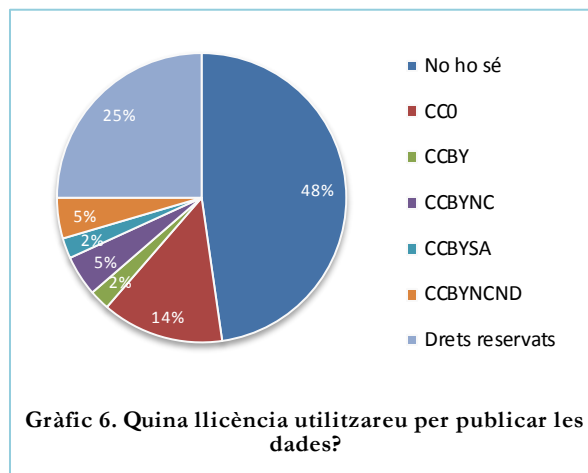
Respecte la difusió de les dades, més del 70% dels projectes tenien previst el dipòsit i la difusió en obert de totes o alguna de les dades de recerca (vegeu gràfic 5). La gran majoria preveïen publicar les dades associades a les publicacions (52%) i d'altres una vegada fossin tractades (26%) o s'hagués finalitzat el projecte (21%).

Pel que fa a la plataforma escollida a l'hora de publicar les dades és molt variada: principalment, l'opció escollida és el repositori institucional (34%) o el repositori d'un dels *partners* (29%) però també apareixen d'altres com les pròpies pàgines web dels projecte (20%),



els repositoris temàtics (9%) o Zenodo com a repositori avalat per la Comissió Europea (9%), entre d'altres.

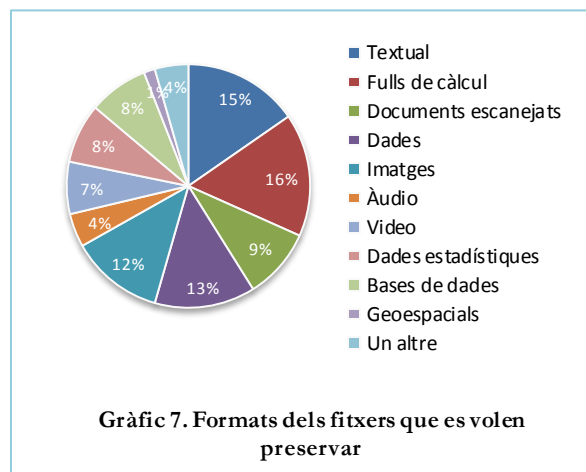
Les llicències utilitzades per publicar les dades tenen un alt grau de variabilitat (vegeu gràfic 6) però, sobretot, destaca el desconeixement dels investigadors respecte aquest tema (48%).



Preservació de les dades i ús d'estàndards

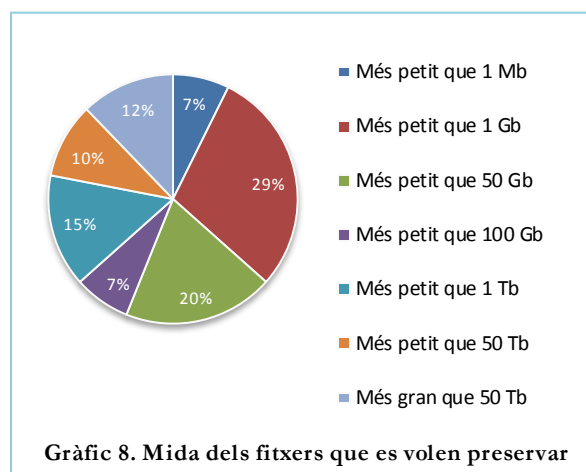
La naturalesa de les dades amb les que treballen els grups de recerca és variable. Pel que fa als formats, no en sobresurt cap respecte la resta (vegeu gràfic 7), tot i que els que apareixen en primeres posicions són els fulls de càlcul (16%), els textuals (15%) i les dades estructurades (13%).

Aquestes dades precisen d'un programari específic per ser consultades en el 40% dels casos i es deu al fet que alguns projectes han desenvolupat el seu propi programari o perquè en certes àrees destaquen programaris comercials.

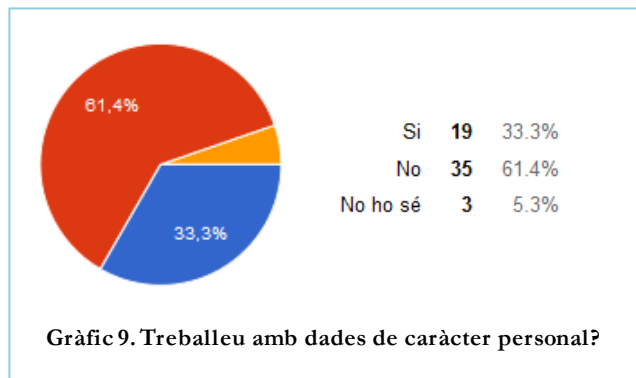


Pel que fa als volum d'emmagatzematge, un 25% dels grups desconeixen quina mida tindran els seus fitxers. El gràfic 8 mostra com el 36% de les respostes asseguren que el pes del fitxer serà relativament baix (inferior a 1Gb), i tan sols el 22% hauran de fer front a volums de dades superiors a 1 TB.

Respecte la preservació d'aquestes dades tampoc trobem un patró clar, tot i que el 44% afirmen que volen preservar-les per sempre. Tan sols 3 grups de recerca (que representen el 4%) asseguren que no els cal conservar les dades de recerca.

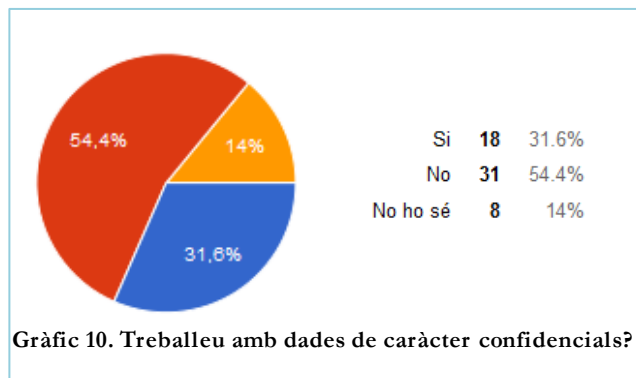


Quasi bé cap dels grups tenen contemplat el possible cost de preservació d'aquestes dades i, dels pocs resultats obtinguts, hi ha estimacions tan dispars que van des dels 500€ als 4.000€ (sense concretar si aquestes estimacions són anuals o per projectes).



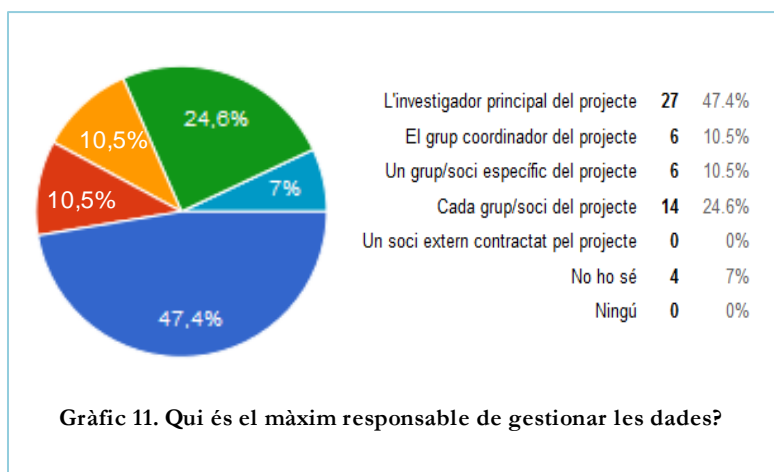
Dels 19 projectes que treballen amb dades de caràcter personal (vegeu gràfic 9), 11 preveuen publicar-les anonimitzades i només dos grups no les difondran fóra del seu grup de recerca de cap manera.

Només 18 projectes asseguren que treballen amb dades confidencials –secret empresarial, patents, etc.- (vegeu gràfic 10) i només faran accessibles aquestes dades dins el propi grup de recerca (21%) i en el marc del projecte (50%). Tan sols un 28% afirmen que seran publicades transcorregut un període temps.



Responsabilitats en la gestió de les dades

Quasi en el 50% dels casos, l'investigador principal del projecte és el màxim responsable de gestionar les dades (vegeu gràfic 11). De la resta de casos, cal destacar que el 24% de les vegades és cada grup/soci del projecte qui s'encarrega de gestionar les seves pròpies dades.



3. Conclusions

- Més del 60% dels investigadors enquestats desconeixen si el seu projecte pertany o no al pilot de dades de l'Horitzó 2020
- De les 73 respostes rebudes, 57 projectes produeixen dades de recerca (78%)
- Més del 70% dels projectes tenen previst dipositar i difondre les dades en obert (bàsicament als repositoris institucionals o en algun dels *partners*)
- Tan sols 2 projectes esmenten que no els hi cal preservar les dades
- Molta variabilitat de format i volum de dades
- 3 de cada 10 projectes treballen amb dades de caràcter personal o confidencials
- L'investigador principal és, en quasi el 50% dels projectes, el màxim responsable de la gestió de les dades

Gestió de les dades de recerca

Benvolguts investigadors,

En els darrers anys, les dades recollides, generades o utilitzades al llarg de les activitats de la recerca han rebut l'atenció de la comunitat científica i dels òrgans gestors de la recerca. La difusió d'aquestes dades és un dels elements constitutius del que s'anomena ciència oberta. Per aquesta raó, a Europa, dins del marc del programa Horizon 2020, s'ha engegat una acció pilot, adreçada específicament a algunes àrees del programa, on es demana tenir un pla de gestió de dades per als projectes atorgats.

Mentre que el valor de les dades de recerca augmenta, es constata la dificultat de la seva gestió. Arran d'aquesta constatació, les institucions de recerca més actives han començat a detectar quines necessitats tenen els investigadors i quins serveis els poden ser útils per tal de facilitar-los aquesta gestió. És per això que les biblioteques i les oficines de recerca de les universitats catalanes -que treballen coordinadament en el si de l'àrea de Biblioteques, Informació i Documentació del CSUC- hem iniciat una línia de treball en aquesta direcció.

La nostra intenció és posar en marxa un projecte pilot que, a partir de la informació provinent dels projectes de recerca finançats en el marc del programa Horizon 2020, analitzi les necessitats dels investigadors respecte de les dades de recerca i estableixi mètodes i processos estandarditzats per gestionar-les. Per recollir aquesta informació, us presentem aquesta enquesta que podreu respondre dedicant-hi entre cinc i deu minuts.

Gràcies per la vostra col·laboració,

*** Necessari**

Dades personals

Nom i cognoms *

Institució *

Marqueu només un oval.

- UB
- UAB
- UPC
- UPF
- UdG
- UdL
- URV
- UOC
- UVic-UCC
- URL

Sobre el projecte de recerca

Títol del projecte *

1. El vostre projecte està dins d'una de les àrees del pilot de Dades de Recerca? *

Marqueu només un oval.

- Si
- No [Passeu a la pregunta 9.](#)
- No ho sé [Passeu a la pregunta 9.](#)

Sobre el projecte de recerca

1.1 A quina? *

Marqueu només un oval.

- Future and Emerging Technologies
- Research infrastructures (including e-Infrastructures)
- Leadership in enabling and industrial technologies – Information and Communication Technologies
- Nanotechnologies, Advanced Materials, Advanced Manufacturing and Processing, and Biotechnology
- Societal Challenge: Food security, sustainable agriculture and forestry, marine and maritime and inland water research and the bioeconomy
- Societal Challenge: Climate Action, Environment, Resource Efficiency and Raw materials – except raw materials
- Societal Challenge: Europe in a changing world – inclusive, innovative and reflective Societies
- Science with and for Society
- Cross-cutting activities - focus areas – part Smart and Sustainable Cities.

1.2 Heu optat per excloure-us del pilot? *

Marqueu només un oval.

- Si
- No

(En cas afirmatiu) Per quina raó?

Marqueu només un oval.

- Els resultats seran explotats comercialment o de manera industrial
- Existeixen incompatibilitats per temes de confidencialitat en relació amb aspectes de seguretat
- Hem de protegir dades personals
- Participar-hi posa en perill l'objectiu principal del projecte
- No generarem ni recollirem dades
- Per altres raons legítimes
- No ho sé

1.3 Heu utilitzat o utilitzareu alguna eina per elaborar el Pla de Gestió de Dades (Data Management Plan)? *

Seleccioneu totes les opcions que corresponguin.

- PaGoDa
- DMPonline (DCC)

- DMPTool (U California)
- No en disposo
- Altres:

[Passeu a la pregunta 12.](#)

Sobre el projecte de recerca

1.4 Teniu previst elaborar un Data Management Plan (DMP)? *

Marqueu només un oval.

- Si
- No

1.5 Participeu de manera voluntària al pilot de dades? *

Marqueu només un oval.

- Si [Passeu a la pregunta 12.](#)
- No [Passeu a la pregunta 11.](#)

Sobre el projecte de recerca

1.6 El seu projecte produeix dades de recerca? *

Marqueu només un oval.

- Si
- No [Deixeu d'emplenar el formulari.](#)

Sobre les dades de recerca

2. Quin tipus de dades està previst que es generin en el marc del projecte? *

Seleccioneu totes les opcions que corresponguin.

- Observacionals: dades capturades en temps real. Per exemple, neuroimatges, dades de mostres, dades de sensors, dades d'enquesta
- Experimentals: dades capturades en equips de laboratori. Per exemple, seqüències de gens, cromatogrames, dades de camps magnètics
- Simulació: dades generades a partir de models de prova. Per exemple climatològiques, matemàtiques o models econòmics
- Derivades o compilades: dades reproduïble però de difícil reproducció. Per exemple, de text i mineria de dades, models 3D, bases de dades compilada
- De referència: conglomerat o conjunt de dades. Per exemple, bancs de dades de seqüències de gens, estructures químiques, o portals de dades espacials
- Altres:

Difusió de les dades

3. Està previst el dipòsit i la difusió de les dades (totes o algunes) en obert? *

Marqueu només un oval.

- Si
- No
- No ho sé

(En cas afirmatiu) Quan es preveu que es publiquin?

Seleccioneu totes les opcions que corresponguin.

- A mida que les anem obtenint
- Un cop les haguem tractat
- Publicarem les dades associades a les publicacions
- Al cap de sis mesos d'obtenir-les
- Al cap de dotze mesos d'obtenir-les
- Com a mínim fins que no hagi transcorregut un any i mig
- Quan hagi finalitzat el projecte
- Altres:

4. Utilitzareu alguna plataforma per dipositar i difondre dades? *

Marqueu només un oval.

- Si
- No
- No ho sé

(En cas afirmatiu) Indiqueu quina plataforma utilitzareu

Seleccioneu totes les opcions que corresponguin.

- Repositori institucional
- Repositori d'un dels partners
- Repositori temàtic
- Zenodo
- Figshare
- Dryad
- Altres:

5. Quina llicència utilitzareu per publicar les dades? *

Per tal de facilitar la reutilització de les dades s'aconsella indicar mitjançant una llicència els usos permesos a priori. En el cas que en el marc del vostre projecte no hàgiu establert cap llicència indiqueu-ho a la casella d'altres. Per a més informació: <http://www.dcc.ac.uk/resources/how-guides/license-research-data>

Seleccioneu totes les opcions que corresponguin.

- CC0 = Creative Commons Zero
- CCBY = Creative Commons Atribució
- CCBYSA = Creative Commons Atribució-Compartir Igual
- CCBYND = Creative Commons Atribució-Sense Obra Derivada
- CCBYNC = Creative Commons Atribució-No Comercial
- CCBYNCSA = Creative Commons Atribució-No Comercial-Compartir Igual
- CCBYNCND = Creative Commons Atribució-No Comercial-Sense Obra Derivada
- PDDL = Open Data Commons-Public Domain Dedication and License
- ODC-BY = Open Data Commons-Atribution License released
- ODC-BY = Open Data Commons-Atribution License released
- ODC-OBbL = Open Data Commons-Open Database License
- GPL = General Public License
- EUPL = European Union Public License
- No utilitzarem cap llicència, les publicarem amb tots els drets reservats
- Altres:

Preservació de les dades i ús d'estàndards

6. Feu una breu descripció dels procediments que preveieu per a la conservació de les dades més enllà de la vida del projecte, indicant: *

Preservació de les dades?

Seleccioneu totes les opcions que corresponguin.

- Volem preservar-les sempre
- Volem preservar-les més de 20 anys
- Volem preservar-les més de 10 anys
- Volem preservar-les més de 5 anys
- Volem preservar-les durant 1 any
- No ens cal preservar-les un cop finalitzat el projecte
- Altres:

Volum total de les dades que voldreu preservar? *

Seleccioneu totes les opcions que corresponguin.

- Més petit que 1 Mb
- Més petit que 1 Gb
- Més petit que 50 Gb
- Més petit que 100 Gb
- Més petit que 1 Tb
- Més petit que 50 Tb
- Més gran que 50 Tb
- Altres:

Podeu fer alguna estimació del cost de la preservació de les dades per any?

7. Quin és format de les dades amb les que treballeu: *

Seleccioneu totes les opcions que corresponguin.

- Textual (.doc, .docx, .log, .rtf, .txt...)
- Fulls de càlcul (.wks, .xls...)
- Documents escanejats (.pdf)
- Dades (.csv, .dat, .xml...)
- Imatges (.bmp, .gif, .jpg, .png, .psd...)
- Àudio (.mp3, .wav...)
- Video (.avi, .mov, .mp4...)
- Dades estadístiques (.sav, .sdq, .spv...)
- Bases de dades (.db, .mdb, .pdb, .sql...)
- Geoespacial (.kml, .gml, .wkt...)
- Altres:

8. Les dades amb què treballeu precisen d'un programari específic per ser consultades? *

Marqueu només un oval.

- Si
- No
- No ho sé

(En cas afirmatiu) De quin programari es tracta?

Preservació de les dades i ús d'estàndards

9. Treballeu amb dades de caràcter personal? *

(Recordeu que en cas que treballeu amb dades de caràcter personal ho heu de notificar a la vostra institució)

Marqueu només un oval.

- Si
- No
- No ho sé

(En cas afirmatiu) Què fareu amb aquest tipus de dades?

Marqueu només un oval.

- No les difondrem fora del nostre grup de recerca
- No les difondrem fora del nostre projecte
- Les publicarem anonimitzades
- Les publicarem agregades

10. Treballeu amb dades confidencials (secret empresarial, patents, etc.)? *

Marqueu només un oval.

- Si
- No
- No ho sé

(En cas afirmatiu) Què fareu amb aquest tipus de dades?

Marqueu només un oval.

- Només seran accessibles al nostre grup de recerca
- Només seran accessibles al nostre projecte
- Les publicarem transcorregut un període de temps

Responsabilitats en la gestió de les dades

11. Qui és el màxim responsable de gestionar les dades? *

Marqueu només un oval.

- L'investigador principal del projecte
- El grup coordinador del projecte
- Un grup/soci específic del projecte
- Cada grup/soci del projecte
- Un soci extern contractat pel projecte
- No ho sé
- Ningú

Tecnologia de

