

# Programa de formació sobre gestió de dades CSUC-CERCA

31 de maig de 2021

## Pautes per establir un servei de gestió de dades de recerca

Ignasi Labastida i Juan  
CRAI Universitat de Barcelona



Consorti de  
Serveis Universitaris  
de Catalunya



# Programa

El context

La planificació

L'obtenció i el processament

La publicació i la preservació

Els costos de la gestió

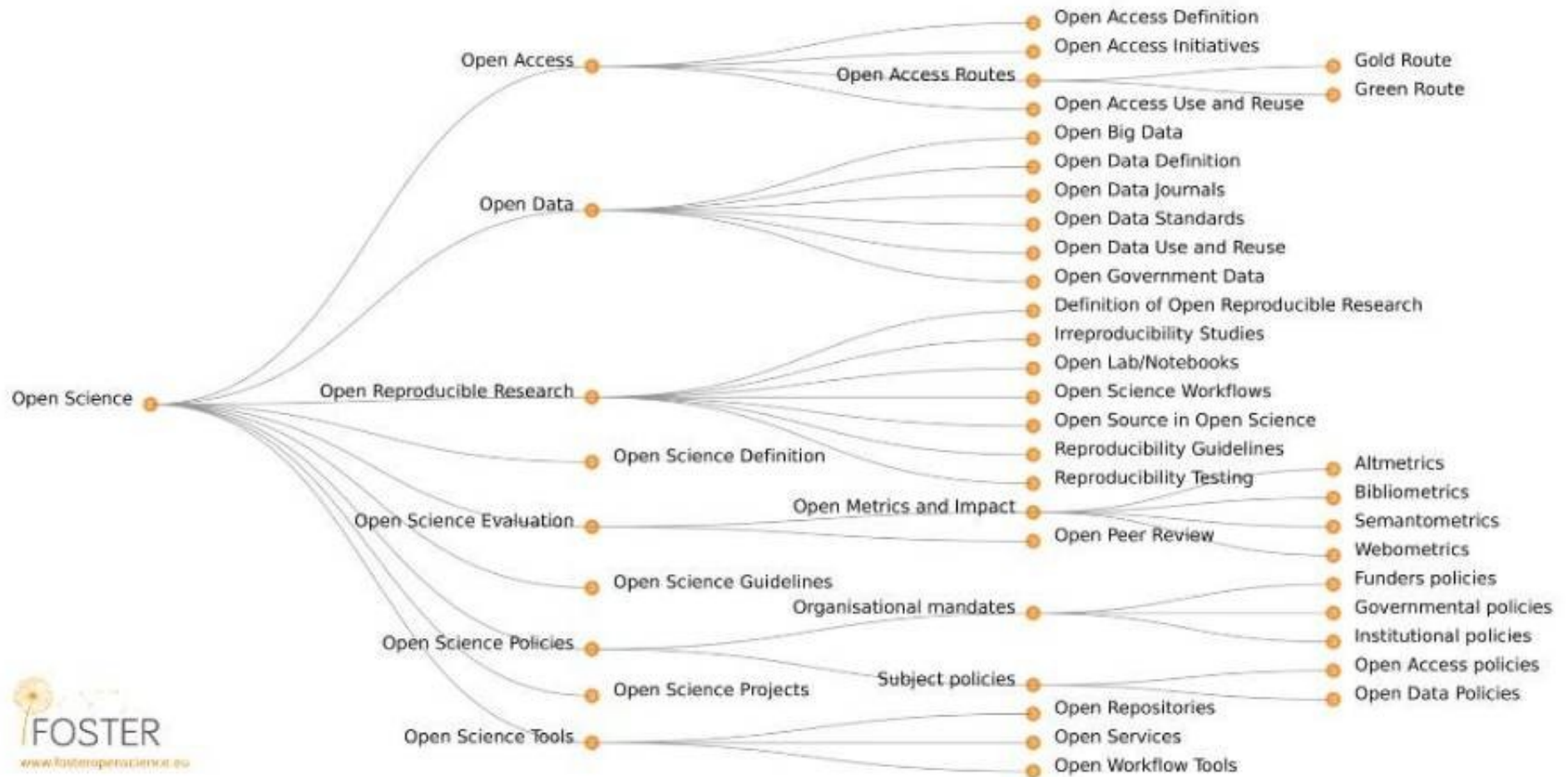
La política institucional

Recomanacions finals

# El context

# La Ciència Oberta

## Open Science Taxonomy



# El perquè de la Ciència Oberta

Open Science aims at transforming science through ICT tools, networks and media, to make research more open, global, collaborative, creative and closer to society.

<https://ec.europa.eu/digital-agenda/en/open-science>

Eficiència, qualitat, integritat, reproduïbilitat, equitat, transparència, participació pública...

# Les dades en el centre de la ciència oberta



This Call for Action is the main result of the Amsterdam conference on 'Open Science-From Vision to Action' hosted by the Netherlands' EU presidency on 4 and 5 April 2016...

A multi-actor approach was formulated to reach two important pan-European goals for 2020:

1. Full open access for all scientific publications

This requires leadership and can be accelerated through new publishing models and compliance with standards set.

**2. A fundamentally new approach towards optimal reuse of research data**

**Data sharing and stewardship is the default approach for all publicly funded research. This requires definitions, standards and infrastructures.**

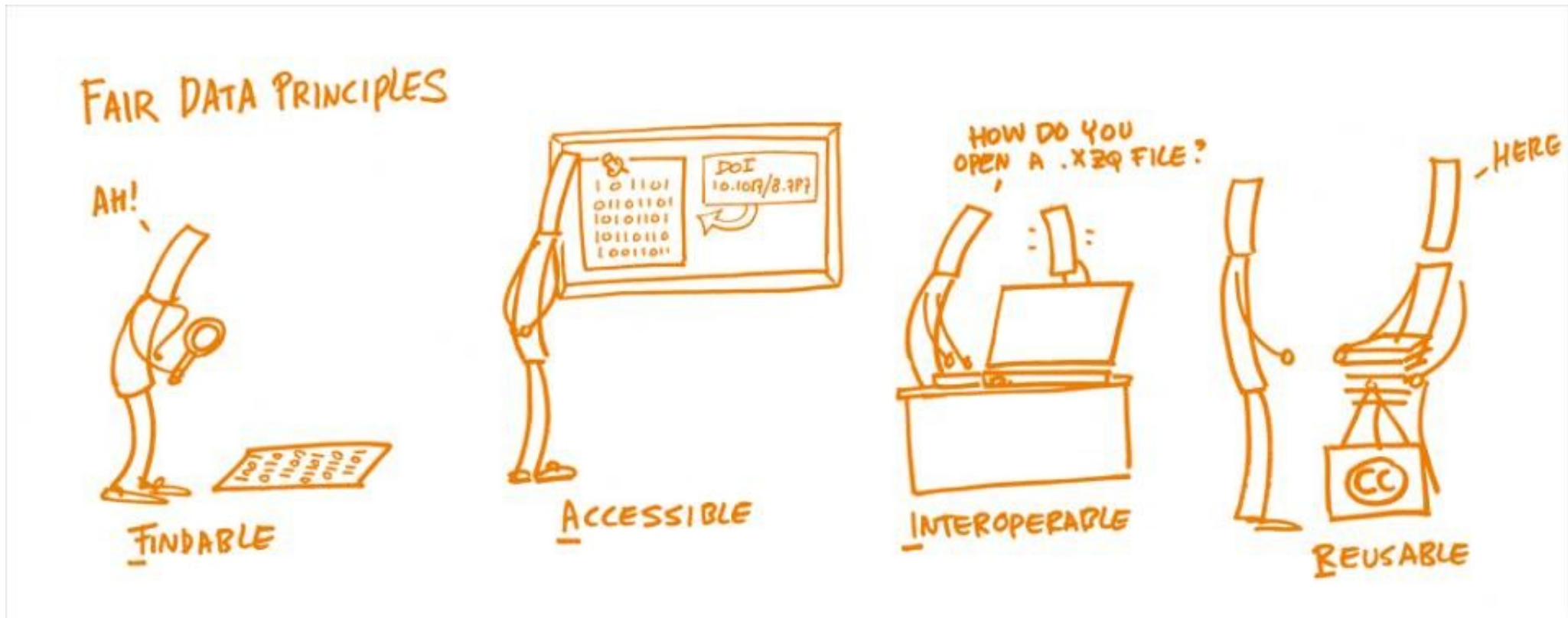
# Quines dades?

Data that are used as primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity, and that are used as evidence in the research process and/or are commonly accepted in the research community as necessary to validate research findings and results. All other digital and non-digital content have the potential of becoming research data.

Research data may be experimental data, observational data, operational data, third party data, public sector data, monitoring data, processed data, or repurposed data.

<https://casrai.org/term/research-data/>

# Objectiu: dades FAIR



<https://book.fosteropenscience.eu/>



# Objectiu: dades FAIR

## FAIR Principles

## Compliance



### Findability

Resource and its metadata are easy to find by both, humans and computer systems. Basic machine readable descriptive metadata allows the discovery of interesting data sets and services.

- ✓ F1. Resource is uploaded to a public repository.
- ✓ F2. Metadata are assigned a globally unique and persistent identifier.



### Accessibility

Resource and metadata are stored for the long term such that they can be easily accessed and downloaded or locally used by humans and ideally also machines using standard communication protocols.

- ✓ A1. Resource is accessible for download or manipulation by humans and is ideally also machine readable.
- ✓ A2. Publications and data repositories have contingency plans to assure that metadata remain accessible, even when the resource or the repository are no longer available.



### Interoperability

Metadata should be ready to be exchanged, interpreted and combined in a (semi)automated way with other data sets by humans as well as computer systems.

- ✓ I1. Resource is uploaded to a repository that is interoperable with other platforms.
- ✓ I2. Repository meta- data schema maps to or implements the CG Core metadata schema.
- ✓ I3. Metadata use standard vocabularies and/or ontologies.



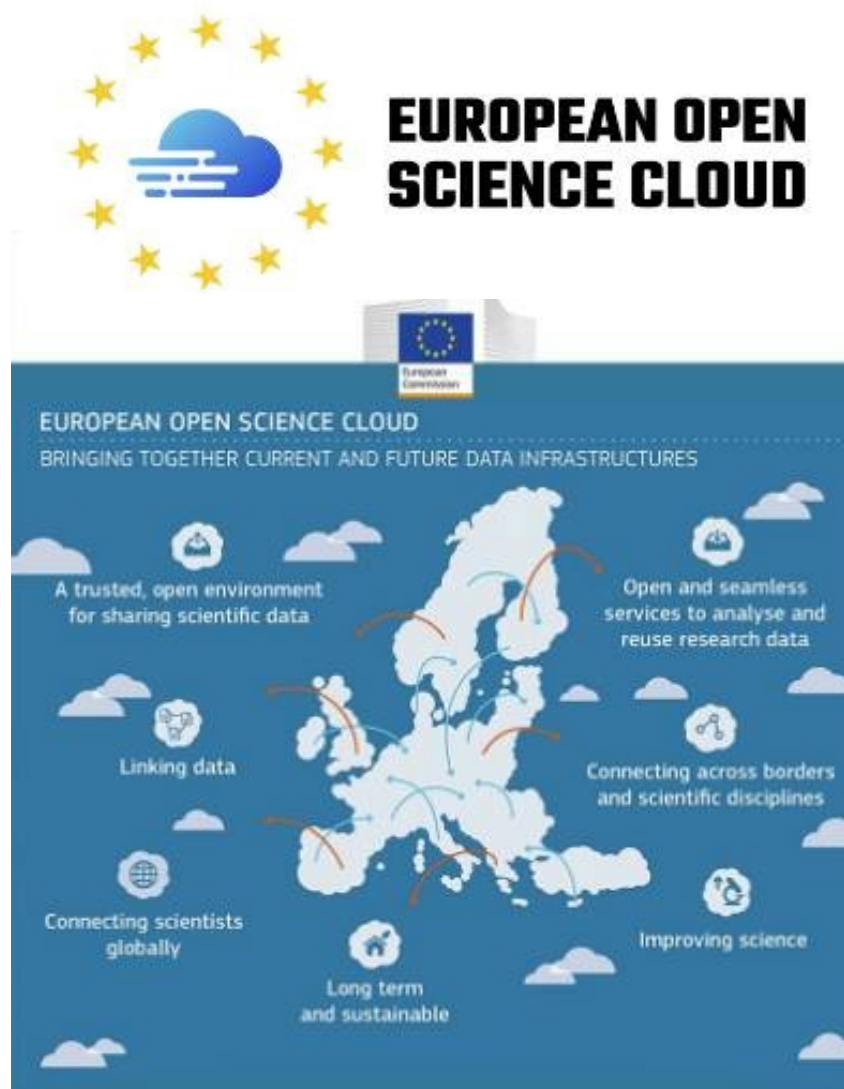
### Reusability

Data and metadata are sufficiently well-described to allow data to be reused in future research, allowing for integration with other compatible data sources. Proper citation must be facilitated, and the conditions under which the data can be used should be clear to machines and humans.

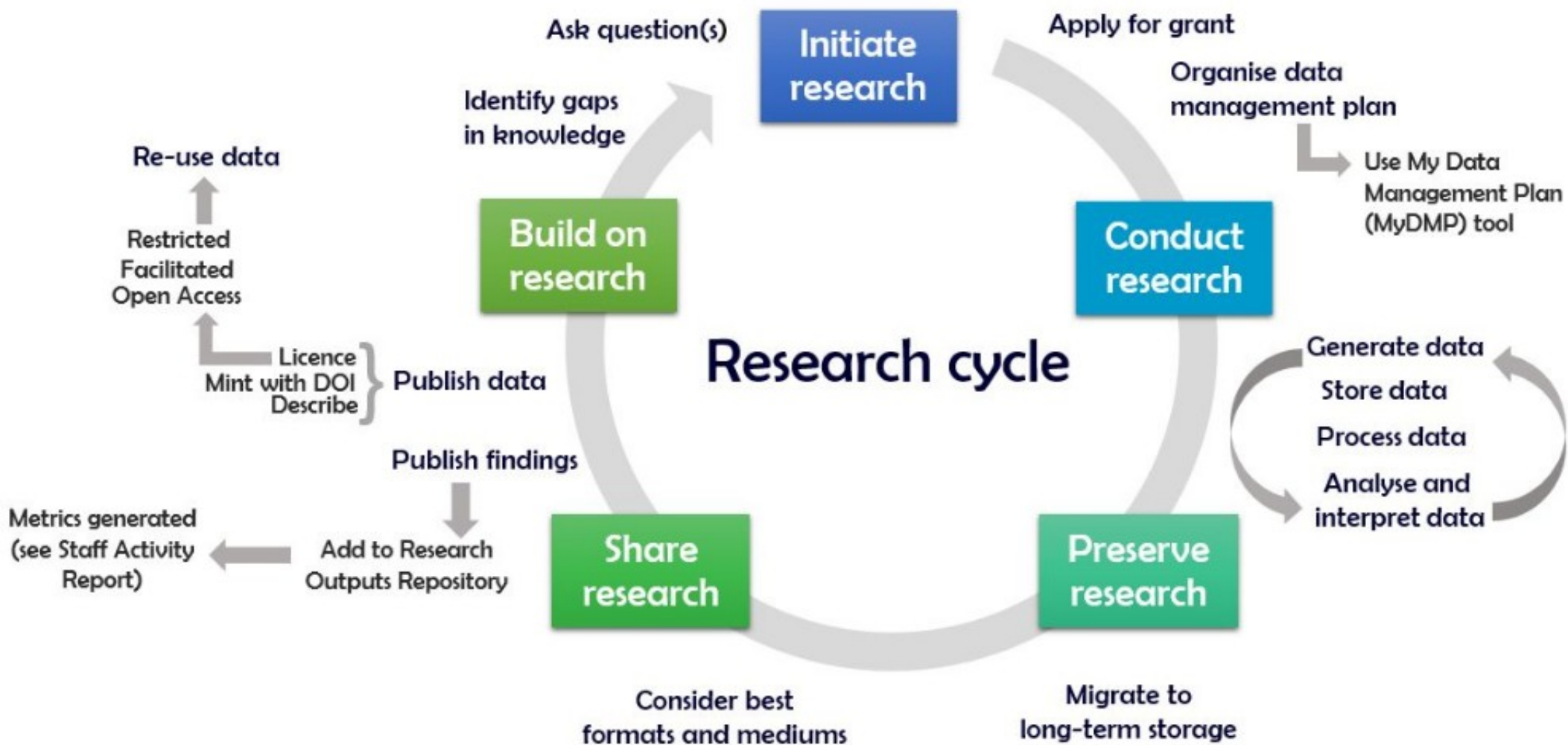
- ✓ R1. Metadata are released with a clear and accessible usage license.
- ✓ R2. Metadata about data and datasets are richly described with a plurality of accurate and relevant attributes.

<https://ccafs.cgiar.org/open-access-and-fair-principles>

# Objectiu: construir l'EOSC



# Actuar en tot el cicle de la recerca



<http://guides.library.unisa.edu.au/ResearchDataManagement>

# La planificación

# Aspectes a tenir en compte

Per començar a planificar la gestió de les dades haurem de tenir en compte temes com:

- Les lleis o normatives aplicables, tant externes com internes
- Les restriccions aplicables a projectes, convenis, col·laboracions
- Els possibles aspectes ètics de la investigació a desenvolupar
- La normativa de propietat intel·lectual del centre

# Aspectes a tenir en compte

Però també hem de preveure com:

- Conscienciem als investigadors
- Analitzem el nivell de preparació del centre per gestionar dades
- Formalitzem les pràctiques de gestió de dades
- Definim indicadors de la política o estratègia establerta

# Primeres passes

Per començar podem identificar

- els grups d'interès i els «campions» de la gestió de dades
- les dades amb les quals es treballa
- les pràctiques actuals en la gestió de dades
- els plans de gestió de dades utilitzats
- les mancances existents per una gestió «ideal»

I llavors establir un «protocol»

# Els protocols de gestió de dades

Science Europe proposava l'any 2018 elaborar protocols de gestió per disciplines

Són marcs de referència que permeten elaborar individualment els plans de gestió de dades, que inclouen els detalls

Es pot definir un protocol per a un centre de recerca

Cal tenir en compte les lleis i normatives aplicables, les infraestructures i els serveis dels quals es disposa, els estàndards de la disciplina, com es pot aconseguir que les dades siguin FAIR, el suport a la gestió i es pot definir una plantilla model per fer un pla de gestió

[https://www.scienceeurope.org/media/nsxdyvqn/se\\_guidance\\_document\\_rdmeps.pdf](https://www.scienceeurope.org/media/nsxdyvqn/se_guidance_document_rdmeps.pdf)



# Elements d'un protocol de gestió de dades

- Lleis i normatives aplicables
- Responsabilitats de la gestió i l'accés a les dades
- Unitats i serveis de suport a la gestió de les dades
- Infraestructures disponibles per treballar amb dades
- Infraestructures disponibles per publicar i preservar dades
- Estàndards i formats acceptats per la comunitat
- Especificitats per a dades personals, confidencials, sensibles
- Aspectes ètics a tenir en compte

# Els plans de gestió de dades

Plans generalment demanats per institucions o finançadors

Responen a preguntes com:

Quines dades s'obtindran o es reutilitzaran i en quins formats

Com es processaran, s'emmagatzemaran i preservaran les dades

Quan es podrà accedir a les dades i qui ho podrà fer

Com es podran reutilitzar

Qui té la responsabilitat de la gestió i l'accés a les dades

S'han tingut en compte els possibles aspectes ètics relacionats amb les dades

# Els plans de gestió de dades

Hi ha diverses plantilles disponibles:

- Horizon 2020
- European Reserach Council
- Fundació la Caixa

Hi ha diverses eines per elaborar-los

- Eina DMP <https://dmp.csuc.cat>
- Argos <https://argos.openaire.eu/splash/>
- Data Steward Wizard <https://ds-wizard.org>

# La plantilla de l'H2020

1. Data Summary

2. FAIR data

2.1. Making data findable, including provisions for metadata

2.2. Making data openly accessible

2.3. Making data interoperable

2.4. Increase data re-use (through clarifying licences)

3. Allocation of resources

4. Data security

5. Ethical aspects

6. Other issues

[http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan\\_en.docx](http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx)

# La gestió de la propietat intel·lectual

Moltes dades no són protegibles per la propietat intel·lectual.

Però la seva compilació o inclusió en una base de dades, sí.

Tipus de dades protegibles: imatges, textos. Han de ser «creacions originals»

La creació d'una base de dades concedeix al fabricant uns drets d'extracció i reutilització dels continguts

Una base de dades pot ser original pel que fa a la presentació i disposició de les dades

# La gestió de la propietat intel·lectual

Qui ostenta els possibles drets de propietat intel·lectual?

Quan un investigador o altre tipus de personal assalariat de la universitat, dins del marc de les activitats de la institució, creï un programa d'ordinador, la titularitat dels drets d'exploració corresponents, correspondran, exclusivament, a la universitat, a no ser que hagi pacte en contrari. En el mateix sentit, les bases de dades que per la seva selecció o disposició dels seus continguts constitueixin creacions intel·lectuals, sens perjudici, en el seu cas, dels drets que poguessin subsistir sobre els esmentats continguts.

En tots els convenis, contractes de recerca i altres tipus de col·laboració realitzats amb entitats públiques o privades i que participi el personal descrit a l'apartat anterior d'aquest precepte, es determinarà a qui correspon la titularitat dels drets de propietat intel·lectual que puguin sorgir com a conseqüència dels mateixos, així com la distribució o repartiment dels possibles beneficis.

Normativa universitat catalana

# La gestió de la propietat intel·lectual

Qui ostenta els possibles drets de propietat intel·lectual?

Art 53

Las disposiciones de este Capítulo son de aplicación a los Organismos públicos de investigación, las universidades públicas,... y otros centros de investigación dependientes de la AGE

Estas disposiciones son aplicables a los resultados de las actividades de investigación, desarrollo e innovación realizadas como consecuencia del desempeño de las funciones que les son propias, por el personal investigador de las entidades mencionadas en el apartado anterior, cualquiera que sea la naturaleza de la relación jurídica por la que estén vinculados a ellas

Art 54

Los derechos de explotación relativos a la propiedad intelectual corresponderán a las entidades en que el autor haya desarrollado una relación de servicios, en los términos y con el alcance previsto en la legislación sobre propiedad intelectual.

<https://www.boe.es/eli/es/l/2011/03/04/2/con>

# Els aspectes ètics

Quan es fa recerca amb éssers humans cal tenir en compte dos aspectes:

- Validació ètica de la investigació

Normalment cal la validació d'un comitè ètic

- Investigació amb dades personals

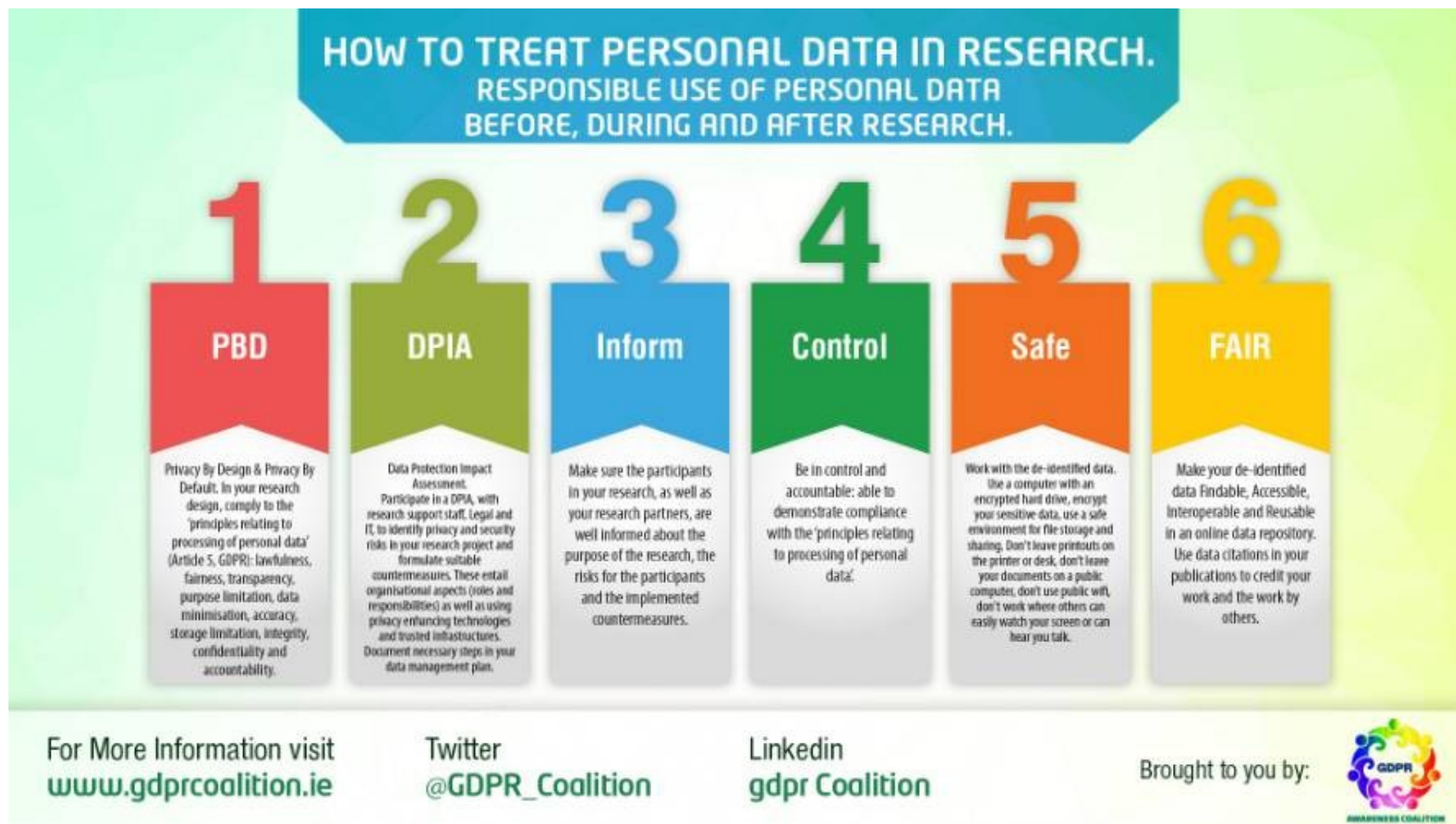
Caldrà tenir en compte els drets de les persones, minimitzar la recollida de dades, aplicar mesures de seguretat, aplicar tècniques d'anonimització o pseudonimització, ...

Consultar amb el delegat de protecció de dades del centre

Fer una avaluació d'impacte relativa a les dades personals



# Els aspectes ètics



# L'obtenció i el processament

# L'obtenció de dades

## Les dades externes

1. Tenim permís per utilitzar-les?
2. Tenim permís per publicar-les?

Un parell d'exemples:

ISSP Research Group (2016). International Social Survey Programme: Citizenship II - ISSP 2014. GESIS Data Archive, Cologne. ZA6670 Data file Version 2.0.0, <https://doi.org/10.4232/1.12590>

Kasybekova, K.; Kessell, S.; Mateeva, K.; McCormack, M.; Nogoibaeva, E.; Toktogazieva, S., 2018, "Fair Trials in the Kyrgyz Republic - Perceptions among Justice Officials and the Public", <https://doi.org/10.7910/DVN/JENNW0>, Harvard Dataverse, V1

# L'obtenció de dades

ISSP Research Group (2016). International Social Survey Programme: Citizenship II - ISSP 2014. GESIS Data Archive, Cologne. ZA6670 Data file Version 2.0.0, <https://doi.org/10.4232/1.12590>

All downloads from this catalogue are free of charge. Please respect our Terms of use.

Unless explicitly indicated differently, GESIS makes data and documents available only for scientific analysis carried out in a project limited to a specific time period in academic research and teaching.

The user is obligated to ensure that the data provided made available, specifically including all backup copies, modified copies, and extract files are deleted from all data processing media and storage devices once the right to process and use the data has ceased.

# L'obtenció de dades

Kasybekova, K.; Kessell, S.; Mateeva, K.; McCormack, M.; Nogoibaeva, E.; Toktogazieva, S., 2018, "Fair Trials in the Kyrgyz Republic - Perceptions among Justice Officials and the Public", <https://doi.org/10.7910/DVN/JENNW0>, Harvard Dataverse, V1

Our Community Norms as well as good scientific practices expect that proper credit is given via citation. Please use the data citation above, generated by the Dataverse.

CC0 - "Public Domain Dedication"

# Els formats de les dades

No és fàcil establir uns formats específics de dades perquè cada disciplina té els seus estàndards i de vegades els formats poden venir fixats per determinats aparells o programari utilitzats

Tanmateix es pot establir un llistat de formats recomanats i suportats que facilitaran la reutilització i la preservació a llarg termini

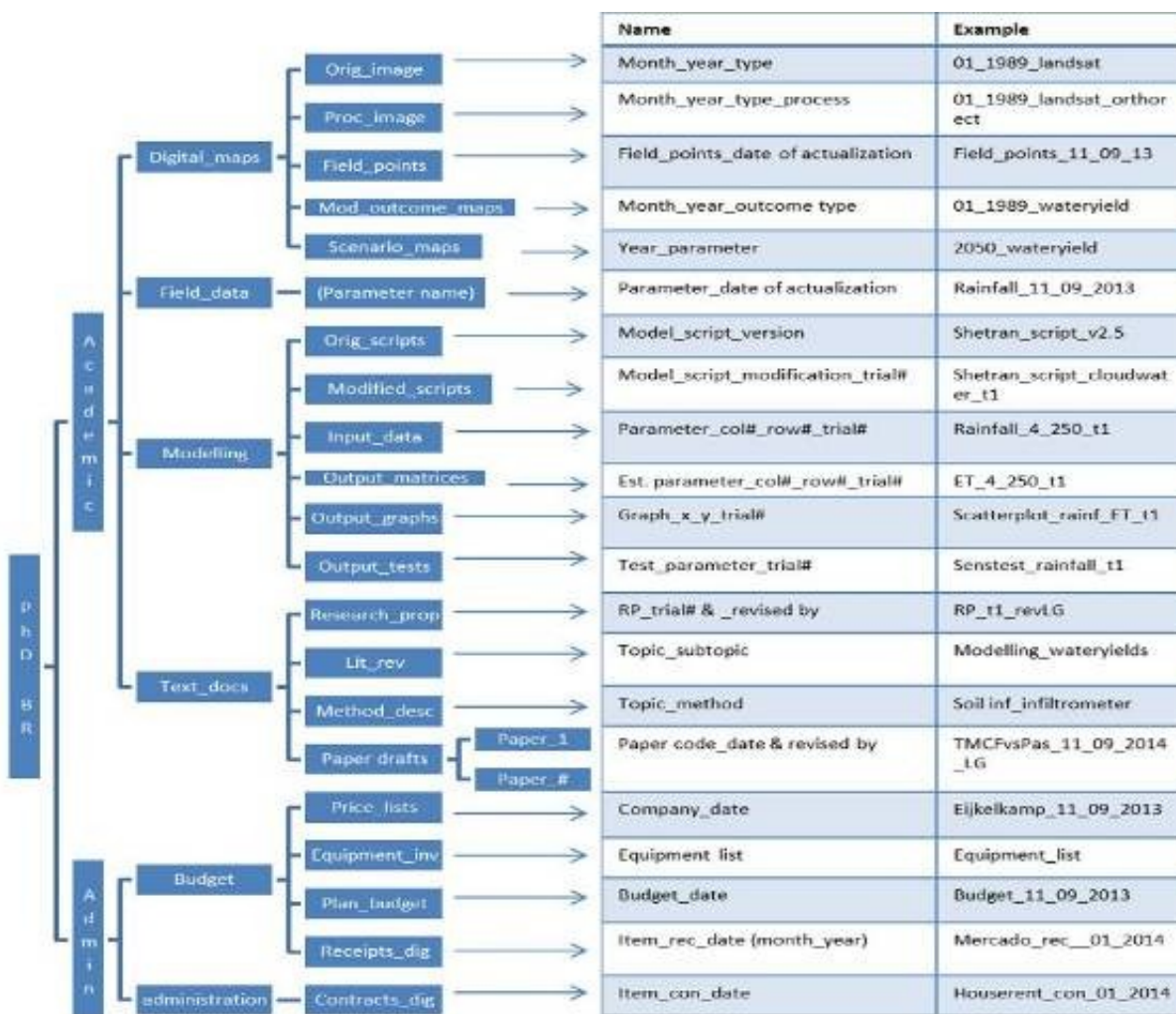
Un centre hauria d'identificar quins formats utilitzen generalment els investigadors, veure com són d'estàndard i elaborar un llistat de formats recomanats. Es poden fer servir llistats existents i adaptar-los a les característiques del centre

# Els formats de les dades

Type	Preferred format(s)	Non-preferred format(s)
Text documents	<ul style="list-style-type: none"><li>PDF/A (.pdf)</li><li>ODT (.odt)</li></ul>	<ul style="list-style-type: none"><li>Microsoft Word (.doc)</li><li>Office Open XML (.docx)</li><li>Rich Text File (.rtf)</li><li>PDF other than PDF/A (.pdf)</li></ul>
Plain text	<ul style="list-style-type: none"><li>Unicode text (.txt)</li></ul>	<ul style="list-style-type: none"><li>Non-Unicode text (.txt)</li></ul>
Markup language	<ul style="list-style-type: none"><li>XML (.xml)</li><li>HTML (.html)</li><li>Related files: .css, .xslt, .js, .es</li></ul>	<ul style="list-style-type: none"><li>SGML (.sgml)</li><li>Markdown (.md)</li></ul>
Programming languages	<ul style="list-style-type: none"><li>MATLAB</li><li>NetCDF</li><li>TextFabric</li></ul>	
Spreadsheets	<ul style="list-style-type: none"><li>ODS (.ods)</li><li>CSV (.csv)</li></ul>	<ul style="list-style-type: none"><li>Microsoft Excel (.xls)</li><li>Office Open XML Workbook (.xlsx)</li><li>PDF/A (.pdf)</li></ul>
Databases	<ul style="list-style-type: none"><li>SQL (.sql)</li><li>SIARD (.siard)</li><li>CSV (.csv)</li></ul>	<ul style="list-style-type: none"><li>Microsoft Access (.mdb, .accdb)</li><li>dBase (.dbf)</li><li>HDFS (.hdfs, .he5, .h5)</li></ul>
Statistical data	<ul style="list-style-type: none"><li>SPSS (.dat/.spss)</li><li>STATA (.dta/.DO)</li><li>R</li></ul>	<ul style="list-style-type: none"><li>SPSS Portable (.por)</li><li>SPSS (.sav)</li><li>STATA (.dta)</li><li>SAS (.7dat, .sd2, .rpt)</li></ul>
Raster images	<ul style="list-style-type: none"><li>JPEG (.jpg, .jpeg)</li><li>TIFF (.tif, .tiff)</li><li>PNG (.png)</li><li>JPEG 2000 (.jp2)</li><li>DICOM (.dcm)</li></ul>	
Vector images	<ul style="list-style-type: none"><li>SVG (.svg)</li></ul>	<ul style="list-style-type: none"><li>Adobe Illustrator (.ai)</li><li>EPS (.eps)</li><li>WMF/EMF (.wmf, .emf)</li><li>CDR (.cdr)</li></ul>
Audio	<ul style="list-style-type: none"><li>BWF (.bwf)</li><li>MXF (.mxf)</li><li>Matroska (.mka)</li><li>FLAC (.flac)</li><li>OPUS</li></ul>	<ul style="list-style-type: none"><li>WAVE (.wav)</li><li>MP3 (.mp3)</li><li>AAC (.aac, .m4a)</li><li>AIFF (.aif, .aiff)</li><li>OGG (.ogg)</li></ul>
Video	<ul style="list-style-type: none"><li>MXF (.mxf)</li><li>Matroska (.mkv)</li></ul>	<ul style="list-style-type: none"><li>MPEG-4 (.mp4, .m4a, .m4v)</li><li>MPEG-2 (.mpg, .mpeg, .m2v, .mpeg2)</li><li>AVI (.avi)</li><li>QuickTime (.mov, .qt)</li></ul>
Computer Aided Design (CAD)	<ul style="list-style-type: none"><li>AutoCAD DXF version R12 (ASCII) (.dxf)</li><li>SVG (.svg)</li></ul>	<ul style="list-style-type: none"><li>AutoCAD other versions than R12 (ASCII) (.dwg, .dxf)</li><li>DWG (.dwg)</li><li>DGN (.dgn)</li></ul>

<https://dans.knaw.nl/en/about/services/easy/information-about-depositing-data/before-depositing/file-formats>

# L'organització de les dades

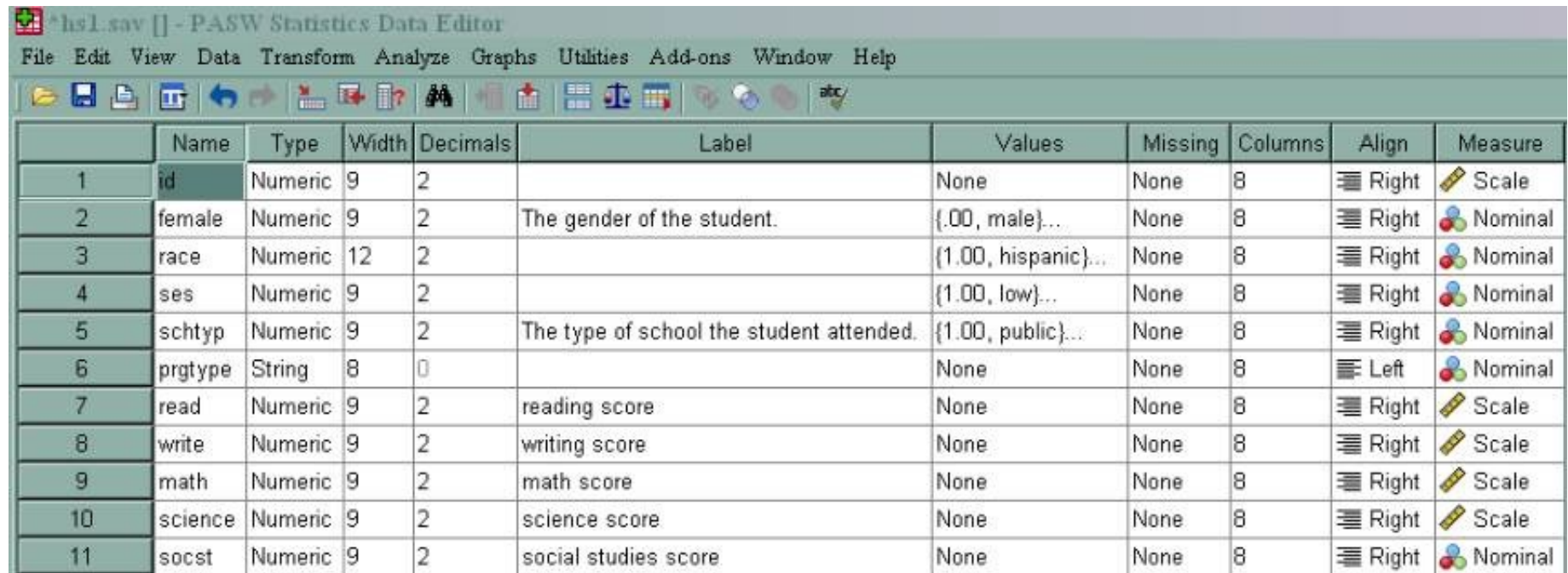


<http://www.wageningenur.nl/en/Expertise-Services/Data-Management-Support-Hub/Browse-by-Subject/Organising-files-and-folders.htm>



# La documentació de les dades

Rarament les dades es poden entendre sense una bona documentació



The screenshot shows the PASW Statistics Data Editor interface. The main window displays a table with 11 variables. The table has columns for Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, and Measure. The variables are: id (Numeric, 9 width, 2 decimals, Scale measure), female (Numeric, 9 width, 2 decimals, Nominal measure, label: 'The gender of the student.'), race (Numeric, 12 width, 2 decimals, Nominal measure, label: '{1.00, hispanic}...'), ses (Numeric, 9 width, 2 decimals, Nominal measure, label: '{1.00, low}...'), schtyp (Numeric, 9 width, 2 decimals, Nominal measure, label: 'The type of school the student attended.', values: '{1.00, public}...'), prgtype (String, 8 width, 0 decimals, Nominal measure), read (Numeric, 9 width, 2 decimals, Scale measure, label: 'reading score'), write (Numeric, 9 width, 2 decimals, Scale measure, label: 'writing score'), math (Numeric, 9 width, 2 decimals, Scale measure, label: 'math score'), science (Numeric, 9 width, 2 decimals, Scale measure, label: 'science score'), and socst (Numeric, 9 width, 2 decimals, Nominal measure, label: 'social studies score').

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	id	Numeric	9	2		None	None	8	Right	Scale
2	female	Numeric	9	2	The gender of the student.	{.00, male}...	None	8	Right	Nominal
3	race	Numeric	12	2		{1.00, hispanic}...	None	8	Right	Nominal
4	ses	Numeric	9	2		{1.00, low}...	None	8	Right	Nominal
5	schtyp	Numeric	9	2	The type of school the student attended.	{1.00, public}...	None	8	Right	Nominal
6	prgtype	String	8	0		None	None	8	Left	Nominal
7	read	Numeric	9	2	reading score	None	None	8	Right	Scale
8	write	Numeric	9	2	writing score	None	None	8	Right	Scale
9	math	Numeric	9	2	math score	None	None	8	Right	Scale
10	science	Numeric	9	2	science score	None	None	8	Right	Scale
11	socst	Numeric	9	2	social studies score	None	None	8	Right	Nominal

# L'emmagatzematge

Quan es treballa amb dades cal decidir on s'emmagatzemaran per tenir-les segures i controlar-ne l'accés en funció de la mida i tipologia de les dades, qui ha d'accedir-hi, el grau de seguretat i control en l'accés...

- ordinadors, portàtils, discs, usb,...
- serveis externs al núvol
- espais institucionals compartits
- infraestructura institucional específica

# L'emmagatzematge

## Data Storage Finder

1. Do you want to share your data with colleagues during your research? ⓘ

No

Yes, but only within the university

Yes, with persons outside the university

2. Is the total amount of your data exceeding 1 TB? ⓘ

Yes  No

3. What is the classification of your data regarding INTEGRITY? ⓘ

Low  Basic  Sensitive

Critical

4. What is the classification of your data regarding CONFIDENTIALITY? ⓘ

Low  Basic  Sensitive

Critical

5. How much does your research depend on continuous availability? ⓘ

Low  Basic  Sensitive

Critical

6. Do you want to be able to access previous versions of the data yourself? ⓘ

Yes  No

7. Do you have special performance needs? ⓘ

I am likely to have more than 1,000 files in a single directory, collected within two years.

My data interactions demand high transaction or transfer rates.

No, performance is not an issue

8. How do you want to access your data?

From a browser, anywhere (https)

Windows network drive (CIFS)

Linux NFS  Webdav

This Data Storage Finder is a guide through the range of IT solutions for storing and managing data. Depending on your demands, the possible solutions are shown. Click on a solution to see more details.

For more information please contact the ICT Servicedesk or Research Data Management Support [ES](#)

<b>BASIC STORAGE</b> ITS Basic storage is a cheap solution for storing a lot of data.	<b>BETA FILE SYSTEM (BFS)</b> Beta File System (BFS) is the Science Faculty storage solution for storing large amounts of data.
<b>BOXCRYPTOR</b> Encryption software.	<b>MANAGED PC/LAPTOP LOCAL DRIVE</b> A device managed by ITS.
<b>MOBILE STORAGE (ENCRYPTED)</b> Encrypted files stored on a portable device.	<b>O: NETWORK DRIVE (SHARED)</b> Fileshare available on Windows for all employees and students.
<b>ONEDRIVE FOR BUSINESS</b> Location in Office 365 to store and share personal business files.	<b>PUBLIC CLOUD (E.G. DROPBOX)</b> Any free public cloud storage.
<b>SURFDRIVE</b> A cloud service for sharing documents.	<b>TEAMS</b> Work together and store and share files.
<b>TEAMSITE</b> An Office 365 SharePoint site to store and share files.	<b>TEAMSITE SECURE</b> An extra secure version of an Office 365 teamsite.
<b>U: NETWORK DRIVE (PERSONAL)</b> Fileshare available on Windows for all employees and students.	<b>UNMANAGED PC/LAPTOP LOCAL DRIVE</b> Any device not managed by ITS.
<b>USB STORAGE</b> USB based storage on an external device such as a stick of drive.	<b>YODA</b> A research data management service that enables Utrecht University researchers and their partners to deposit, publish and preserve research data.

<https://tools.uu.nl/storagefinder/>

# Els nous rols

Els investigadors necessiten suport de personal en la gestió de dades

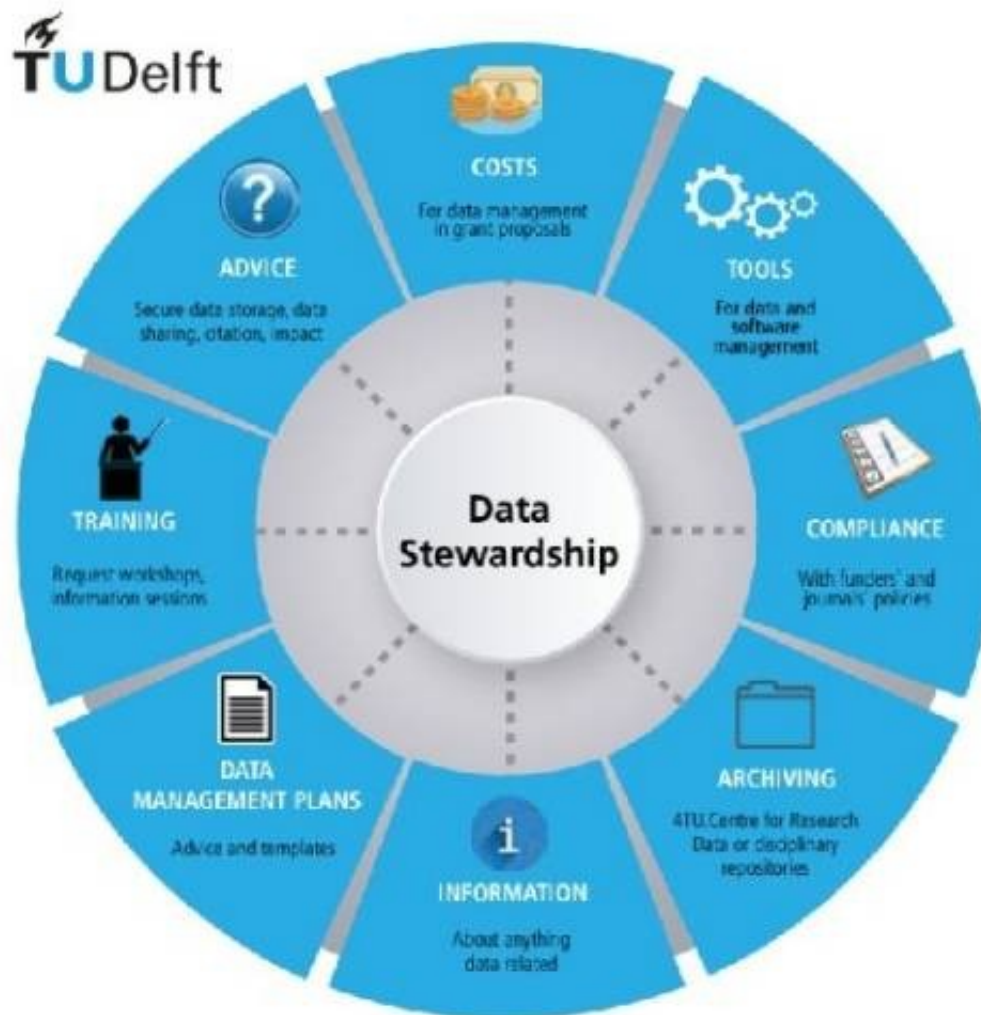
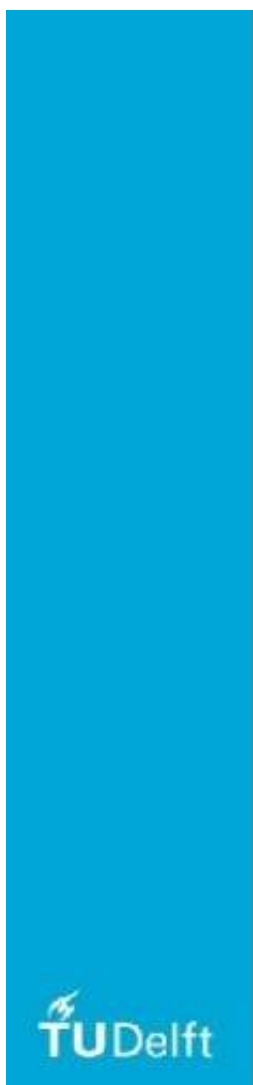
Apareixen nous rols com el curador de dades, el científic de dades, el bibliotecari de dades, l'enginyer de dades, ...

Possiblement el més estès és el «data steward»

Què fa un data steward?

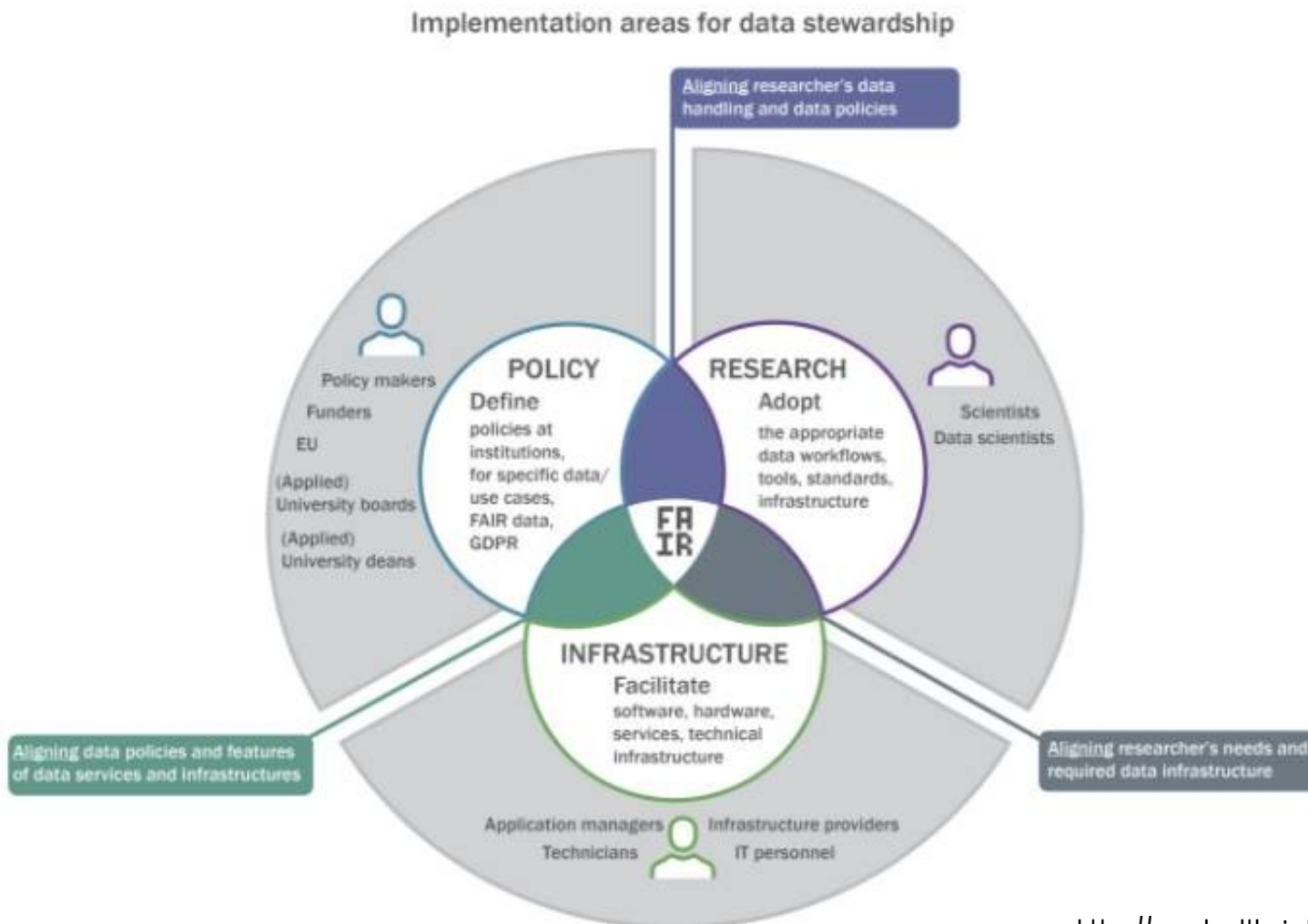
- Connecta investigadors amb eines, polítiques, informació
- Ofereix consell i formació
- Intervé en el disseny dels serveis

# Els nous rols



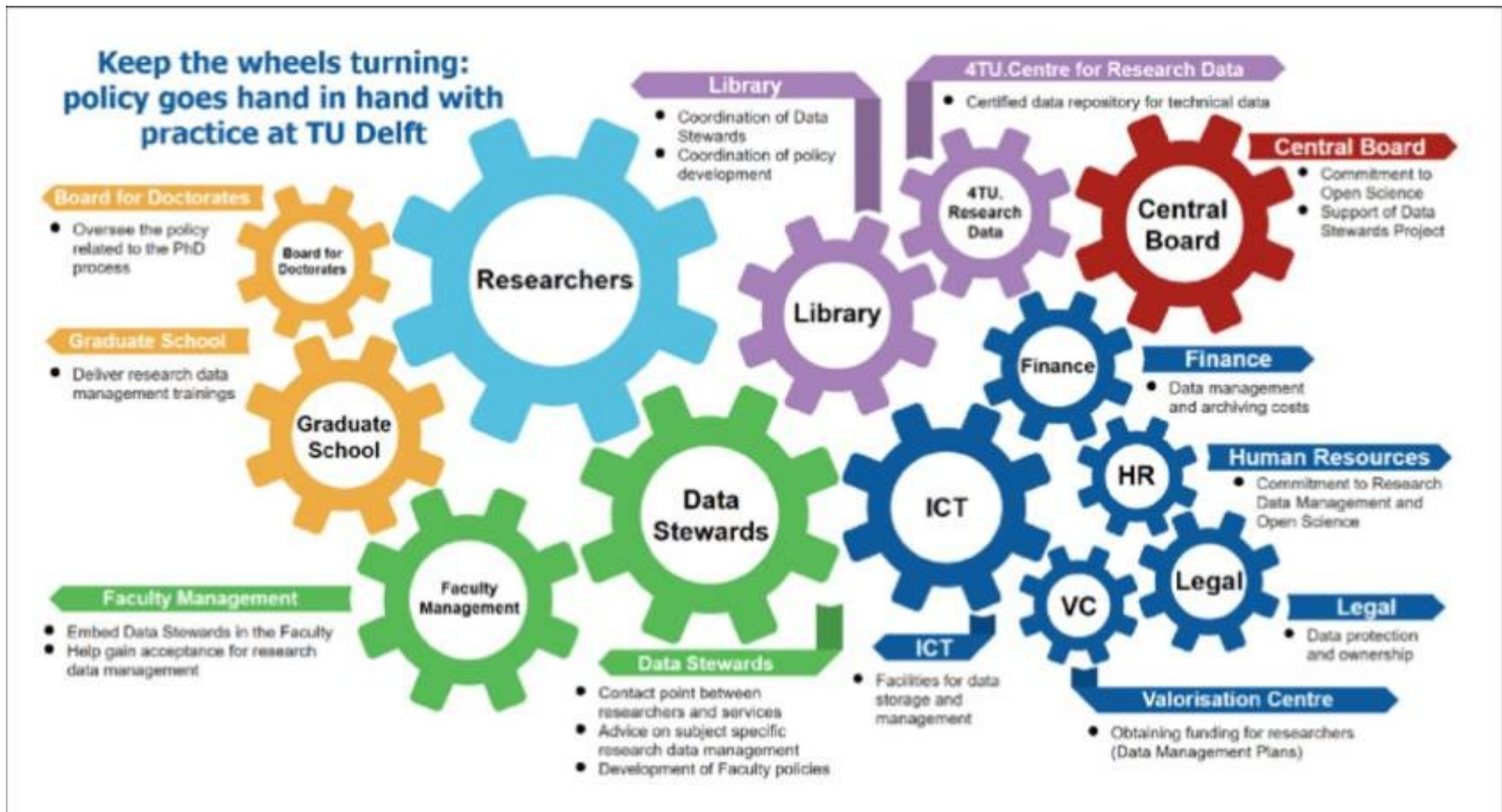
<https://slidetodoc.com/data-stewardship-at-tu-delft-the-netherlands-esther/>

# Els nous rols



<https://www.health-ri.nl/definitions-data-stewardship>

# Un sistema amb molts engranatges



Policy Needs to Go Hand in Hand with Practice: The Learning and Listening Approach to Data Management  
Data Science Journal, 18(1), p.45. DOI: <http://doi.org/10.5334/dsj-2019-045>

# La publicació i la preservació



# La publicació de les dades

## La tria del repositori

1. Tenim un repositori temàtic?
2. Tenim un repositori propi?
  - De la institució
  - D'una institució participada
  - Consorciat
3. Quines opcions multidisciplinàries tenim a fora?

# La publicació de les dades

## Repositoris temàtics

Si tenim un repositori temàtic consolidat en el nostre àmbit d'investigació aquesta seria la primera opció

Directori de repositoris: <http://www.re3data.org/>

En un centre de recerca hauríem d'identificar possibles repositoris temàtics i oferir-los com a opció als investigadors

# La publicació de les dades

## Repositoris propis

Si volem tenir un repositori propi cal pensar en una eina pensada per gestionar dades, que permeti aplicar els principis FAIR

### FAIR x FAIR

Requeriments factibles, assolibles i  
implementables per a un repositori de  
dades de recerca FAIR

Consorci de Serveis Universitaris de Catalunya (CSUC)

Maig de 2018

# La publicació de les dades

## Repositoris multidisciplinaris externs

Hi ha diverses opcions externes per publicar dades

En molts casos s'ofereixen opcions gratuïtes amb limitacions de mida o nombre de conjunts de dades







Hi ha opcions de pagament que amplien les opcions gratuïtes i que poden esdevenir institucionals

Seria recomanable tenir un catàleg institucional de dades sigui quina sigui la seva ubicació

# La publicació de les dades



<https://ddd.uab.cat/record/150829>

						
<b>1. Àmbit temàtic i geogràfic</b>	Multidisciplinar Internacional	Multidisciplinar Internacional	Multidisciplinar Internacional	Amb èmfasi en ciència i medicina Internacional	Multidisciplinar Internacional	Multidisciplinar Internacional
<b>2. Tipologia de dades (programari, imatges, raw data, etc.)</b>	<u>Tot tipus de formats</u> i també documents publicats: publicacions, literatura grisa, imatges i audiovisuals, text pla, etc.	<u>Tot tipus de formats</u>	<u>Tot tipus de formats:</u> xifres, conjunts de dades, imatges i audiovisuals, literatura grisa, conjunts d'arxius, etc.	<u>Tot tipus de formats</u> (amb exclusió de software) i amb preferència per formats de codi obert	<u>Tot tipus de formats</u> amb funcions específiques per alguns (Stata, SPSS, R, Excel, CSV, etc.)	<u>Tot tipus de formats:</u> text sense format, imatges i audiovisuals, PDF, HTML, etc.
<b>3. Mida aproximada dels fitxers a incloure</b>	Fins a 50 GB Per a mides superiors es pot <u>contactar</u> amb ells	Fins a 2 GB. Per a mides superiors, <u>contactar</u> amb ells	Fins a 5 GB	20 GB per un conjunt de dades ( <u>cost</u> addicional per sobre de 20 GB)	Fins a 2 GB. Per fitxers més grans <u>contactar</u> amb ells	10 GB
<b>4. Dades obertes, embargades, restringides o tancades</b>	Obertes Tancades Restringides Embargades ( <u>Publica</u> )	Obertes Restringides Embargades. La descripció del conjunt de dades és pública.	Obertes Tancades (per defecte fins que no es publiquen) Restringides Embargades	Obertes <u>Embargades</u> (per defecte fins el moment de la publicació)	Obertes Restringides La descripció (títol, etc.) d'un conjunt de dades és pública però els arxius poden ser públics o restringits. Sense informació sobre embargaments	Obertes Tancades (per defecte fins que no es publiquen) Restringides Embargades
<b>5. Llicències amb les que es volen difondre les dades</b>	Permet varietat de llicències encara que recomanen llicències obertes. La descripció de les dades està sota <u>CC0</u>	Permet varietat de llicències per a dades i per a software	- <u>CC0</u> per a dades ( <u>metadata &amp; datasets</u> ) - <u>CC-BY</u> per a no dades (audiovisuals, papers, etc.) - <u>MIT</u> per a codi i software	<u>CC0</u> ( <u>EAC0</u> )	Per defecte <u>CC0</u> però es pot canviar	Permet varietat de llicències: CC-BY 4.0, CC-BY-NC, CC0 1.0, MIT, Apache 2.0, BSD 3 clause, BSD 2 clause, gplv3
<b>6. Identificador permanent</b>	DOI	Handle	DOI	DOI	DOI	DOI
<b>7. Condicions per depositar o retirar contingut</b>	Es permet <u>retirar</u> les dades justificant-ne el perquè. El DOI i la URL quedaran retinguts.	Per retirar contingut <u>contactar</u> amb <u>EUDAT</u>	Sense informació	Cessió de dades <u>irrevocable</u>	Sense informació	Es permet <u>retirar</u> les dades abans de publicar-les. Una vegada publicades cal <u>contactar</u> amb Mendeley.

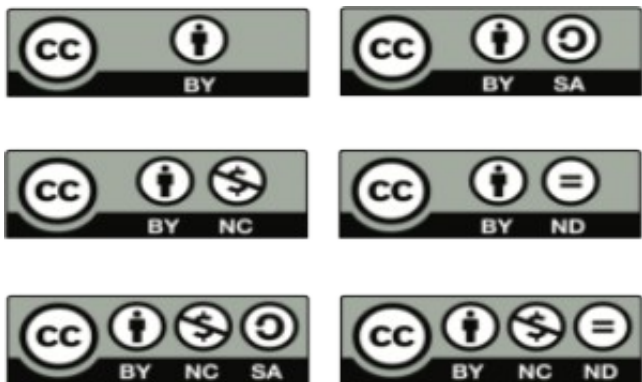
# L'obertura de les dades

Quan es publiquen els dades cal indicar quin és el seu grau d'obertura, és a dir quins usos estan permesos per defecte i per quins cal demanar un permís exprés

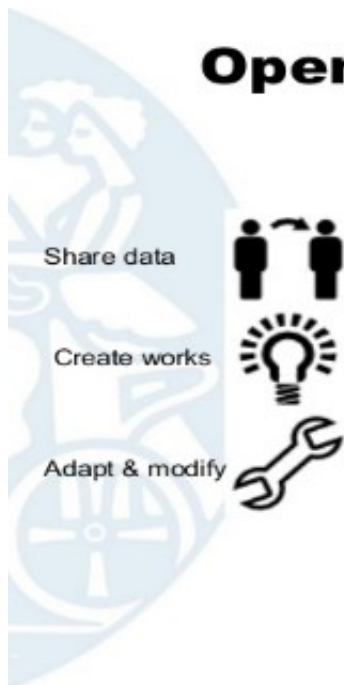
No dir res equival a només permetre els usos previstos per defecte a la llei vigent aplicable

El titular dels possibles drets de propietat intel·lectual és el responsable de decidir com s'obren les dades

# Llicències per obrir les dades



## Open Data Commons



### ODC Open Database License (ODbL)

<http://opendatacommons.org/licenses/odbl/>



Attribution, Share alike & keep open

### ODC Attribution

<http://opendatacommons.org/licenses/by/>



Attribution

### ODC Public Domain Dedication

<http://opendatacommons.org/licenses/pddl/summary/>

None

# El tancament de les dades

Hi ha molts casos on cal que l'accés a les dades no estigui obert a tothom sinó que cal que estigui restringit. Parlarem de dades «tancades», però poden ser FAIR.

Podem obrir les metadades i informar sobre les condicions d'accés i utilització

Podem dissenyar acords d'accés i ús estàndard anomenats DUA (data user agreement)

Podem establir processos d'anonimització, agregació o privacitat diferencial per donar un determinat grau d'accés a les dades



# El tancament de les dades

Tag Type	Description	Security Features	Access Credentials
Blue	Public	Clear storage, Clear transmit	Open
Green	Controlled public	Clear storage, Clear transmit	Email- or OAuth Verified Registration
Yellow	Accountable	Clear storage, Encrypted transmit	Password, Registered, Approval, Click-through DUA
Orange	More accountable	Encrypted storage, Encrypted transmit	Password, Registered, Approval, Signed DUA
Red	Fully accountable	Encrypted storage, Encrypted transmit	Two-factor authentication, Approval, Signed DUA
Crimson	Maximally restricted	Multi-encrypted storage, Encrypted transmit	Two-factor authentication, Approval, Signed DUA

<http://datatags.org/>

# El tancament de les dades

## **Article 1 Contract Data Basis**

(1) Data Provider shall provide Data Recipient with the following statistical individual data free of charge: **XXXXXXX**

- hereinafter referred to as the data basis -

The designation "data basis" also embraces parts of the data basis and duplicates of the data basis.

(2) Data Recipient is only permitted to use and to process the data basis within the framework of the scientific project with the title:

**XXXXXXXXXXXXXXXXXXXXX.**

Passing on of the data basis to a third party is not permitted. Processing or use for other purposes, particularly in the context of reports for private clients, requires a separate agreement.

## **Article 2 Entitlements and Liability**

(1) If Data Recipient proves that the data basis provided is significantly different to that agreed to, Data Recipient shall be entitled to subsequent improvement within a week of the first use of the data basis. Farther-reaching claims of whatever sort are excluded.

(2) Data Recipient shall be liable vis-à-vis NIFU for all damage resulting from handling of the data basis that is not in keeping with the agreement reached and shall to this extent exempt Data Provider of all liability claims of third parties. Cases of liability shall justify termination for good cause.

## **Article 3 Prohibition of Deanonymization, Secrecy**

(1) Data Recipient must refrain from all action that is intended to, or suitable for deanonymization of the anonymized individual data contained in the data basis.

(2) Results in publications and presentations must be summarized such that no inference to individuals can be made. Neither the data basis nor any parts of it may be published.

## **Article 4 Persons with Rights of Use**

(1) The data basis may only be used by persons named in this contract, persons working for Data Recipient who are entrusted with the processing of the research project indicated in article 1, subsection 2. These are the following persons:

- **XXX**
- **XXX**

[http://risis.eu/wp-content/uploads/2016/04/MORE\\_Data\\_Use\\_Agreement.pdf](http://risis.eu/wp-content/uploads/2016/04/MORE_Data_Use_Agreement.pdf)

# La preservació de les dades

Un cop finalitzat un projecte o una investigació s'haurà de decidir quines dades es preserven. Podem tenir dades publicades i dades tancades

Hi ha codis de conducta que indiquen que les dades cal guardar-les almenys 10 anys, i hi ha obligacions legals de guardar algunes dades fins a 20 anys. Mentre que les dades personals s'haurien d'esborrar tan aviat com fos possible

Cal seleccionar les dades a preservar i les que es poden eliminar

Es pot elaborar una llista de situacions:

<https://stanford.app.box.com/s/3lcv1cj0zynmqy3z5s41k4pe08u51pxj>

# Els costos

# Els costos de la gestió de dades

La gestió de dades de recerca implica uns costos que es poden dividir en:

Obtenció de les dades

Documentació de les dades

Emmagatzematge de les dades

Accés i seguretat de les dades

Preservació de les dades

Disponibilitat i reutilització de les dades

Costos generals

<https://www.uu.nl/en/research/research-data-management/guides/costs-of-data-management>

# Els costos de la gestió de dades

## Obtenció de les dades

Costos relacionats amb obtenció de dades externes

Costos relacionats amb l'organització i el format de les dades

Costos relacionats amb transcripcions

Costos per obtenir autoritzacions

## Documentació de les dades

Costos relacionats amb la descripció de les dades i la metodologia emprada, creació de metadades,

# Els costos de la gestió de dades

## Emmagatzematge de les dades

Costos relacionats amb l'espai utilitzat, les còpies de seguretat

## Accés i seguretat de les dades

Costos relacionats amb accessos de personal extern

## Preservació de les dades

Costos relacionats amb la conversió a formats oberts i estàndards

Costos d'emmagatzematge a llarg termini

# Els costos de la gestió de dades

## Disponibilitat i reutilització de les dades

Costos relacionats amb la preparació de les dades per temes legals

Costos de digitalització

Costos de publicació en un repositori

## Costos generals

Personal del servei

Formació



# La política institucional

# Un possible model

## Basat en el projecte LEARN

Preàmbul amb la justificació

Jurisdicció aplicable

Drets de propietat intel·lectual

Gestió de les dades

Responsabilitats, drets i deures

Validesa

Annex amb definicions

Política <https://doi.org/10.14324/000.learn.26>

Pautes <https://doi.org/10.14324/000.learn.27>

# Indicadors per fer-ne seguiment



## List of KPIs for implementation

Number	KPI	LEARN Toolkit Theme	Measurement	Expected Value	Rationale for measure
I1	Monitoring of the institutional policy	Policy and Leadership, Tool Development	Monitoring activity exists	YES	Any research performing institution with a RDM policy in place should have a way to monitor it
I1.1	Review of the policy	Policy and Leadership, Tool Development	Number of reviews in a year	1	A yearly review is advised as best practice to monitor the policy
I1.2	Updates of the policy	Policy and Leadership, Tool Development	Number of updates since its enforcement	< 1/year	An update of the policy means that some of the initial expectations are not fulfilled. It is advisable to make as many updates as needed, but several updates in a year would mean that the policy is not well designed
I2	Disciplines engaged in RDM within the research institution	Advocacy; Subject approaches	Percentage of disciplines engaged in RDM activities	90%	Any research performing institution expects to engage all its researchers in a policy. Therefore, is advisable to monitor the engagement of all disciplines, even if some have difficulties in starting work on RDM
I3	Datasets stored in the policy-designated infrastructure(s)	Research Data Infrastructure	Number of datasets stored	Increasing year after year	If the institution establishes an infrastructure to store datasets, it must monitor the growth of the infrastructure and measure the number of datasets available
I4	Datasets published on the policy-designated platform(s) with a clear statement on terms of use	Open Data; Research Data Infrastructure; Tool development	Number of datasets published with a clear statement (I4.1) / Number of all datasets published (I4.2)	100%	Although open data could be a final goal, many datasets cannot be shared openly due to some restrictions such as confidentiality, security, privacy. Nevertheless, it is required that at least metadata is publicly available and there is an indication about the degree of openness of each dataset



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

[http://learn-rdm.eu/wp-content/uploads/KPIs\\_MD.pdf](http://learn-rdm.eu/wp-content/uploads/KPIs_MD.pdf)

# Recomanacions finals

# Recomanacions per a la gestió de dades

Identificar la situació actual

Definir rols, serveis i infraestructures

Avaluar els costos de la gestió

Establir una política o estratègia de gestió per al centre

Establir un protocol de gestió de dades per als investigadors

Conscienciar als investigadors i oferir formació

# Gràcies

Ignasi Labastida i Juan  
CRAI Universitat de Barcelona

ilabastida@ub.edu  
@ignasi

# Estem preparats?

## Enquesta LEARN

Realitzada Juny 2016-17

<https://zenodo.org/record/810493>

<https://forms.gle/y4QivyNZarDiPrFy5>

