

XV Workshop REBIUN. Castellón, 29-30 septiembre 2016

# Sobreviviendo a los datos

## La Biblioteca de la UPF y la gestión de los datos de investigación

30 de septiembre de 2016



Universitat  
Pompeu Fabra  
Barcelona



crue

Universidades  
Españolas

Red de Bibliotecas  
REBIUN

# **Sobreviviendo a los datos**

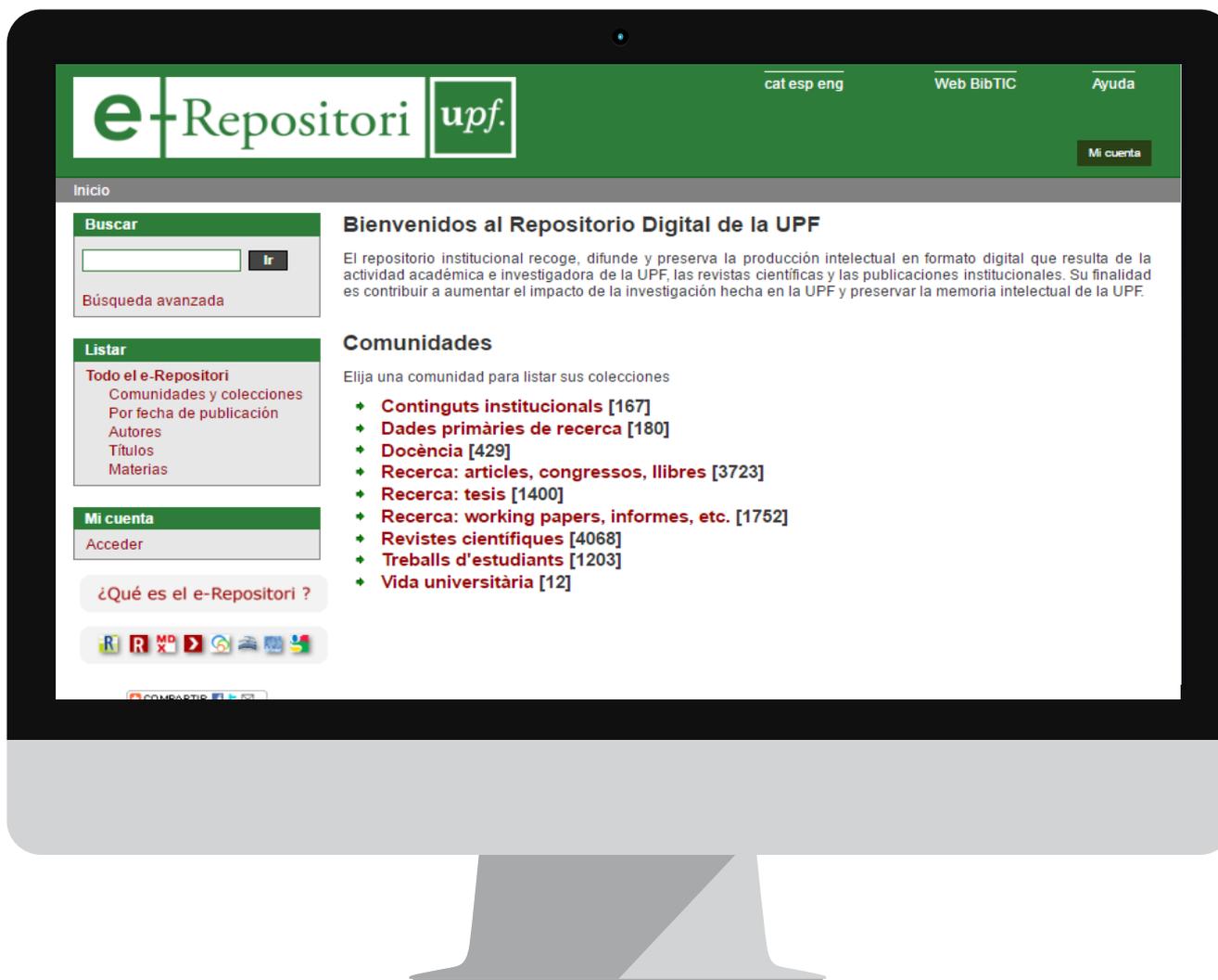


# **Sobreviviendo a los datos**





# 1. Contexto





# 1. Contexto



## e-Repository

### Acceso abierto

98 % documentos

### Producción UPF

No incluye los fondos patrimoniales

### Archivo delegado

Introducción de documentos por bibliotecarios

### Dspace versión 3.1

Migrando a Dspace 5.5

### Datos de investigación

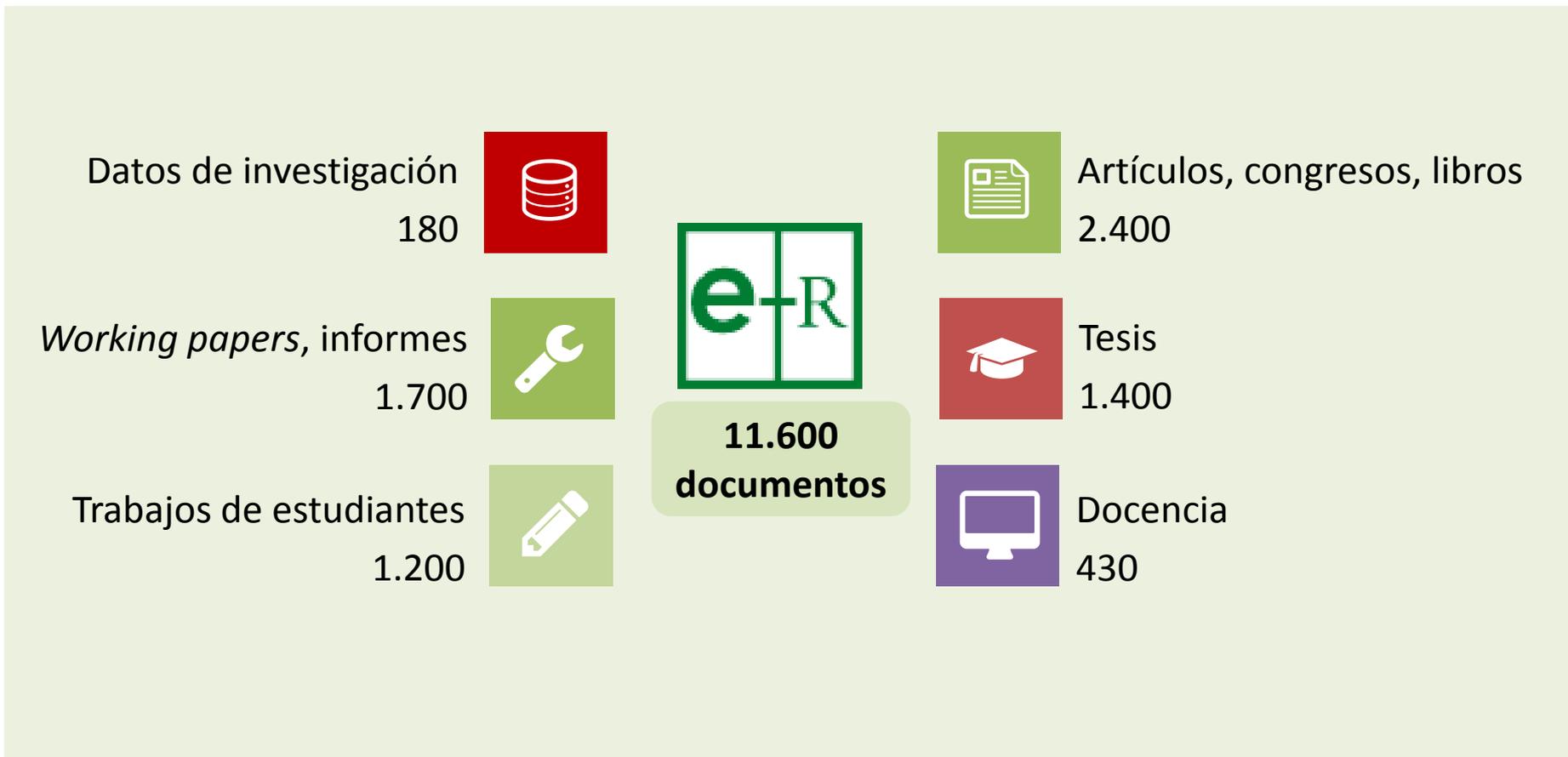
Desde 2011

### Política de acceso abierto

2011 (Recomendación)



# 1. Contexto



# 1. Contexto Aspectos...

## Económicos

- Coste del almacenamiento
- Preservación
- Recursos humanos



## Tecnológicos

- Múltiples formatos
- Diferentes necesidades
- Diversidad de repositorios

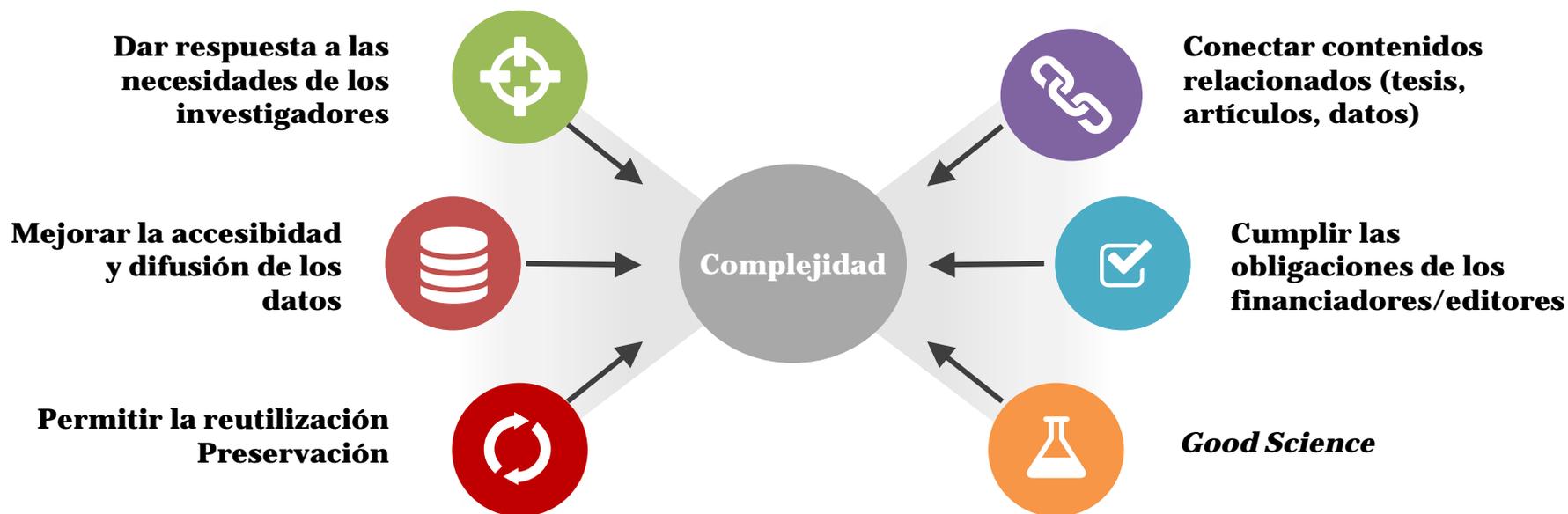


## Legales

- LPI
- LOPD
- Formación / Información



# 1. Contexto ¿Porqué?





# 1. Contexto Colaboración

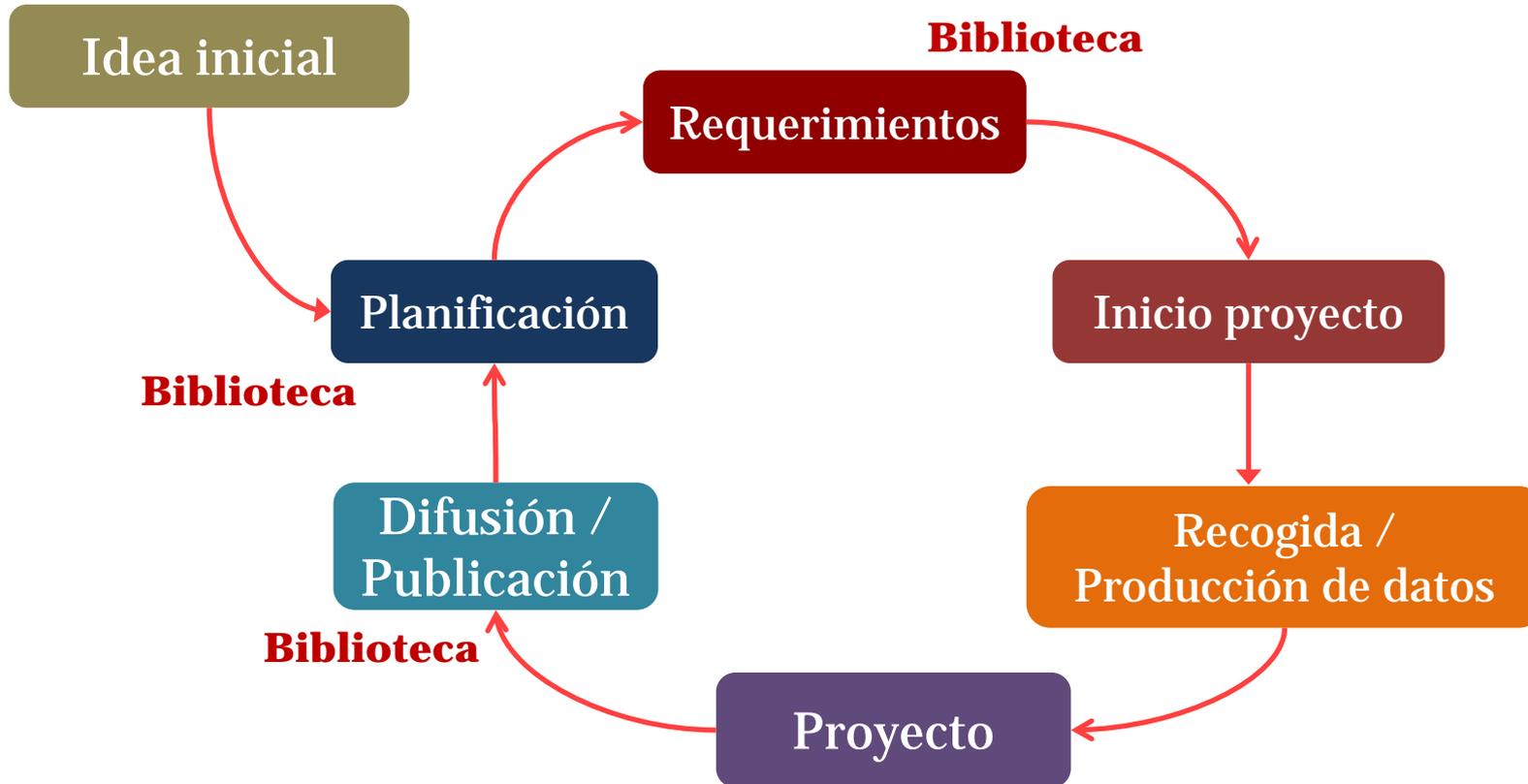


# **Sobreviviendo a los datos**



## 2. Carta de servicios

### Ciclo de la investigación y Biblioteca



## 2. Carta de servicios



- **Formación y asesoramiento**
- **Páginas web informativas**
- **Plan de Gestión de Datos**
- **Publicación de *datasets***
- **Colaboración con proyectos**
- **Resolución de dudas y consultas**

## Datos de investigación: Apoyo en la planificación y difusión

Search this Guide  Carcar

### Apoyo en la planificación y difusión

### Publicar datos en el e-Repository

#### ¿Por qué publicar los datos en abierto?

Publicar los datos de investigación en abierto permite **garantizar el acceso**, de forma libre y universal, a estos datos, así como una correcta preservación, explotación, reproducción, difusión, visibilidad e impacto de los mismos. También permite cumplir con los requisitos de los organismos financiadores.

#### Datos de investigación en el Programa H2020



Los proyectos del Programa Horizon 2020 que formen parte del Piloto de Datos de Investigación en Abierto deben desarrollar un Plan de Gestión de Datos que debe ser entregado durante los seis primeros meses del proyecto y deben depositar los datos en abierto en un repositorio. Más información en:

- [Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020](#)
- [Guidelines on Data Management in Horizon 2020](#)

#### Datos de carácter personal

Consultad la legislación española y europea y el procedimiento para el tratamiento de datos de carácter personal en los proyectos de investigación, desarrollo e innovación. La información UPF sobre protección de datos de carácter personal y las medidas de seguridad, están disponibles en [Campus Global](#) > [La Universidad](#) > [Datos de carácter personal](#)

#### Plan de Gestión de Datos



Data Management Plan template (english)  
Plantilla de la UPF (word) para realizar un Plan de Gestión de Datos siguiendo los requisitos del Programa H2020.

Otras herramientas para crear y gestionar el Plan de Gestión de Datos:



DMP Online  
Digital Curation Centre (UK). El registro es necesario. Permite editar y exportar los planes colaborativamente y modificarlos cuando sea necesario. Tiene una plantilla para los proyectos H2020 de la Comisión Europea.



PGDOnline  
Traducción y adaptación al castellano del DMPonline. Consorcio Madrofo (Madrid).



DMP Tool  
Universidad de California. El registro es necesario. Permite compartir, editar y exportar el Plan de Gestión de Datos cuando sea necesario. No dispone de plantilla para proyectos del Programa H2020.

#### Publicar datos en abierto

Para publicar vuestros datos tenéis diferentes opciones:

1. Comprobar si disponéis de un repositorio temático adecuado. Consultad [re3data.org](http://re3data.org)
2. Publicar datos en el e-Repository de la UPF
3. Otros repositorios multidisciplinares (Dryad, Figshare, Zenodo, GitHub, Mendeley Data, etc).

Si no publicáis en el Repositorio Digital de la UPF, informadnos para enlazar los datos con las publicaciones relacionadas.

#### Como citar los datos

Datacite.org: Autor(es) (Año). Título. Editor. Identificador (DOI, Handle)

Ejemplo: Irino, T. Tada, R. (2009). Chemical and mineral compositions of sediments from ODP Site 127-797. Geological Institute, University of Tokyo. <http://dx.doi.org/10.1594/PANGAEA.726855>

Force11: Autor(es), Año, Título del dataset, Repositorio, Versión, Identificador (DOI, Handle)

Ejemplo: Chan, AH, Anderson, K. 2015. X-Ray Diffraction data for RT52A from E. Coli. PDB C0de SC25, SBGrid Data Bank, v1. <http://dx.doi.org/10.15785/5BGRD/1588>

Vancouver: Autor(es) Título [splogia], Nombre de la institución / Repositorio donde se deposita. Localización; Año de publicación [fecha de acceso]. Disponible en: Identificador (DOI, Handle)

Ejemplo: ComPlusic. Beijing Opera Percussion Pattern Dataset [dataset]. Repositorio Digital de la UPF: Barcelona; 2016 [citado 23 abril 2016]. Disponible en: <http://hdl.handle.net/10230/25720>

#### Versiones

Català English

#### Consulta también

- [Publicar en acceso abierto](#)
- [Proyectos de investigación](#)
- [Publicar en revistas científicas](#)

#### Contacto / Ayuda

Contactad con la Biblioteca por correo electrónico en la dirección [repositor@upf.edu](mailto:repositor@upf.edu) o por teléfono al 93 542 17 27.

<http://guesbibtic.upf.edu/data/es/datos>



## Plan de Gestión de Datos:

- Creación de una plantilla siguiendo requerimientos Proyecto H2020
- Asesoramiento en la redacción y planificación
- Uso del DMPonline CSUC (próximamente)

### Data management Plan - UPF Template

This template has been designed specifically for researchers of the Universitat Pompeu Fabra. The Data Management Plan consists of five sections, as determined by the [H2020 guidelines](#). Each section is accompanied with a Checklist, providing the most important questions to be answered in your Data Management Plan. Many questions include example options for answers – use or delete as appropriate, or replace with your own response.

The DMP should address the 5 points below on a dataset by dataset basis.

#### 1. Data set reference and name

Unique name identifying the dataset

Examples:

- The ProjectX dataset identification follows the naming: Data\_<WPno>\_<serial number of dataset>\_<dataset title>. Example: Data\_WP2\_1\_User generated content.
- Once the data set is created and stored in the UPF Repository, a persistent and unique identifier (handle) will be generated automatically.

#### 2. Data set description

Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.

##### 2.1 Data collection/generation

Checklist:

- Are there any existing data or methods that you can reuse?
- Do you need to pay to reuse existing data? Are there any restrictions on the reuse of third-party data?
- Can the data that you create - which may be derived from third-party data - be shared?
- How does your data complement and integrate with existing data?

Examples:

- The research objectives require qualitative data that are not available for other sources. Some data exist that can be used to situate the findings of the proposal research and which will supplement data collect as part of the proposed research. Some quantitative data are available, but they are insufficiently detailed. In their current form, they would not permit as full a comparison across the cases as is desirable.
- Data are collected through online questionnaires, structured observations and prototyping.
- Existing sources of data: group previous research, academic collaborators, publicly available database/archive in x repository/archive/web.
- Data collected through questionnaire survey among Project XXX partners.
- Data collected through an online form that will become publicly available and distributed through the partners' networks.
- Data under copyright: photographs / film extracts / musical compositions

## Datos de investigación: Publicar datos en el e-Repositori

Escriu les paraules que vols cercar

Apoyo en la planificación y difusión

**Publicar datos en el e-Repositori**

### e-Repositori



El e-Repositori es el repositorio institucional de la Universitat Pompeu Fabra que tiene por objetivo recoger, difundir y preservar la producción intelectual hecha en la universidad, incluyendo tanto las publicaciones como los datos de investigación.

### Depositar los datos en el e-Repositori

- Si prevés que se pueda hacer una transferencia de resultados, consulta con la [Unidad de Innovación](#).
- Si quieres difundir tus datos de investigación, puedes hacerlo a través de el e-Repositori con las siguientes condiciones:

- 1/ Como mínimo uno de los autores tiene que pertenecer a la comunidad universitaria de la UPF.
- 2/ Los datos no son de consulta restringida y no contienen ninguna información de carácter confidencial, ni protegida legalmente, ya sea por cuestiones de privacidad, de propiedad intelectual, industrial u otros.
- 3/ Los datos se depositarán en acceso abierto. Se puede optar por embargar el acceso por un período máximo de 2 años.
- 4/ Cada uno de los ficheros debe ser más pequeño de 2 GB.
- 5/ Los datos de investigación son finales o bien asociados a una publicación.

Para depositar los datos, rellena el formulario y envíanos los archivos a [repositori@upf.edu](mailto:repositori@upf.edu).

Si tienes dudas en relación a estas cuestiones, contacta con nosotros: [repositori@upf.edu](mailto:repositori@upf.edu).

### Formulario para el depósito de los datos en el e-Repositori

[Formulario](#)





**e-Repository upf**

**Entrada de dades de recerca a l'e-Repository**

Indicar que veu a registrar un nou registre de dades de recerca a l'e-Repository de l'UPF. El registre de dades de recerca de l'UPF està gestionat pel Servei de Recerca i Innovació Tecnològica (SRIT) de l'UPF. El registre de dades de recerca de l'UPF està gestionat pel Servei de Recerca i Innovació Tecnològica (SRIT) de l'UPF. El registre de dades de recerca de l'UPF està gestionat pel Servei de Recerca i Innovació Tecnològica (SRIT) de l'UPF.

**Autor/s \***  
Autor individual o grup de recerca

**Títol \***  
Títol descriptiu de la naturalesa i tema del fitxer de dades, ex: General Household Survey, 2001-2002.

**Nom del fitxer/s \***  
Sense espai, guioi baix per a separar.

**Resum**  
Propòsit i descripció general del fitxer de dades (màx. 300 paraules).

**Cobertura**  
Proporcionar informació sobre cobertura geogràfica (país, regió, ciutat), temporal (dates), població (característiques del grup estudiant) o altres.

**Enllaç a pàgines web**  
Enllaç a pàgines web amb informació rellevant sobre el fitxer de dades.

**Paraules clau \***  
Paraules clau descriptives del contingut

**Programari**  
Indicar el programari utilitzat per a generar el fitxer de dades.

**Descripció: data de creació \***

**Descripció: format i mida dels fitxers \***

**Descripció: llengua \***

anglès  
 castellà  
 català  
 altres

**Descripció: tipus \***

Dataset  
 Image  
 Audiovisual  
 Programari  
 So  
 Text  
 Altres

**Finançament**

No s'indica  
 European Commission H2020  
 European Commission FP7  
 Altres

**Publicacions relacionades**

**Accés \***

Accés obert (metadades i fitxer visible)  
 Accés amb embargament (metadades visibles, fitxer visible en X mesos (indicar els mesos a l'opció Altres))  
 Altres

**Llicència / Copyright \***

Creative Commons CC BY  
<https://creativecommons.org/licenses/by/4.0/>  
 Creative Commons CC0  
<https://creativecommons.org/publicdomain/zero/1.0/>  
 Una altra opció

**Transferència de fitxers \***

Correu electrònic: [repositori@upf.edu](mailto:repositori@upf.edu)  
 Google Drive (compartir amb: [repositori@upf.edu](mailto:repositori@upf.edu))  
 Una altra opció

**Altres comentaris**

**Condicions \***

Accepto les condicions

Vull rebre una còpia de les meves respostes.

**e-Repository upf**

**Entrada de dades de recerca a l'e-Repository**

Declaro que estic legitimat/da per enviar aquest formulari i acceptar les condicions de dipòsit i difusió a l'e-Repository de les dades que s'hi descriuen (perquè sóc titular dels drets sobre les dades o bé perquè disposo de la competència o dels permisos per fer-ho)

Declaro que el dipòsit i difusió de les dades no vulnera drets de tercers (ni de propietat intel·lectual, ni industrial, ni de privacitat, ni de cap mena) de manera que exoneri la UPF de qualsevol obligació o responsabilitat davant qualsevol acció legal que es pugui suscitar derivada del contingut, dipòsit o difusió de les dades.

En enviar aquest formulari, es registra el nom d'usuari (**marina.losada@upf.edu**). Aquest usuari no és vostre? [Tanqueu la sessió](#)

**\* Necessari**

**Autor/s \***  
 Autor/s individual o grup de recerca

La vostra resposta

**Títol \***  
 Títol descriptiu de la naturalesa i tema del fitxer de dades, ex: General Household Survey, 2001-2002.

La vostra resposta

**Nom del fitxer/s \***  
 Sense espai, guioi baix per a separar.

La vostra resposta

**Resum**  
 Propòsit i descripció general del fitxer de dades (màx. 300 paraules).

La vostra resposta

**Cobertura**  
 Proporcionar informació sobre cobertura geogràfica (país, regió, ciutat), temporal (dates), població (característiques del grup estudiant) o altres.

La vostra resposta

**Enllaç a pàgines web**  
 Enllaç a pàgines web amb informació rellevant sobre el fitxer de dades.

La vostra resposta

**Paraules clau \***  
 Paraules clau descriptives del contingut

La vostra resposta



## Depositar los datos en el e-Repositori

- Si prevés que se pueda hacer una transferencia de resultados, consulta con la [Unidad de Innovación](#).
- Si quieres difundir tus datos de investigación, puedes hacerlo a través de el [e-Repositori](#) con las siguientes condiciones:

1/ Como mínimo uno de los autores tiene que pertenecer a la comunidad universitaria de la U

Datos primarios o de investigación \* (*sólo si se han recopilado*)

Autorizo la introducción de los datos primarios que he recopilado para elaborar la tesis en el *e-Repositori*, (<http://repositori.upf.edu>) con una licencia Creative Commons CC0 Dominio Público, y me comprometo a suministrar los ficheros correspondientes. Estos datos no vulneran derechos de terceros y por lo tanto asumo cualquier responsabilidad que se pueda derivar de los mismos y de su publicación.

Y para que quede constancia, firmo este documento en \_\_\_\_\_ (lugar),  
el día \_\_\_\_\_

## Entrada de datos de investigación en el e-Repositori

Declaro que estoy legitimado/a para enviar este formulario y aceptar las condiciones de depósito y difusión en el *e-Repositori* de los datos que en él se describen (o bien soy titular de los derechos sobre los datos o dispongo de la competencia o de los permisos para hacerlo).

Declaro que el depósito y difusión de los datos no vulnera derechos de terceros (ni de propiedad intelectual, ni industrial, ni de privacidad, ni de ningún otro tipo) de manera que eximo a la UPF de cualquier obligación o responsabilidad ante cualquier acción legal que se pueda derivar del contenido, depósito o difusión de los datos.

Al enviar este formulario, se registra el nombre de usuario ([nombre.apellido@upf.edu](mailto:nombre.apellido@upf.edu)).

¿Este usuario no es el vuestro? [Cerrar la sesión](#) |



## 2. Carta de servicios

### Colaboración en proyectos de investigación

Departamento de TIC



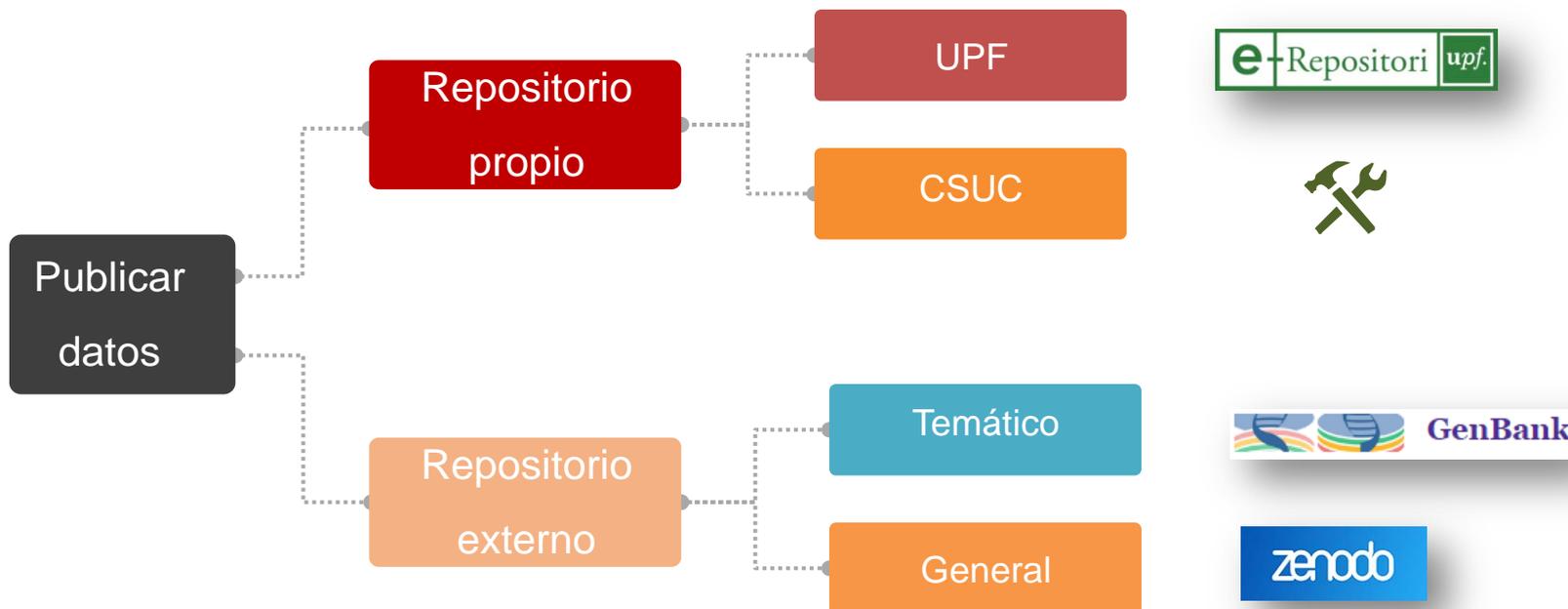
- Mejorar el  para hacer más visible la relación de las publicaciones (artículos) con los datos de investigación (*datasets*)
- Plantilla para elaborar un Plan de Gestión de Datos
- Mejorar los flujos de trabajo: un único punto de entrada de datos
- Analizar infraestructura para la fase de trabajo con los datos → Figshare

# **Sobreviviendo a los datos**





# 3. Infraestructura Opciones





## 3. Infraestructura Recomendaciones UPF

### Publicar datos en abierto

Para publicar vuestros datos tenéis diferentes opciones:

1. Comprobar si disponéis de un repositorio temático adecuado. Consultad [re3data.org](http://re3data.org)
2. Publicar datos en el e-Repositori de la UPF
3. Otros repositorios multidisciplinares (Dryad, Figshare, Zenodo, GitHub, Mendeley Data, etc).

Si no publicáis en el Repositorio Digital de la UPF, informadnos para enlazar los datos con las publicaciones relacionadas.

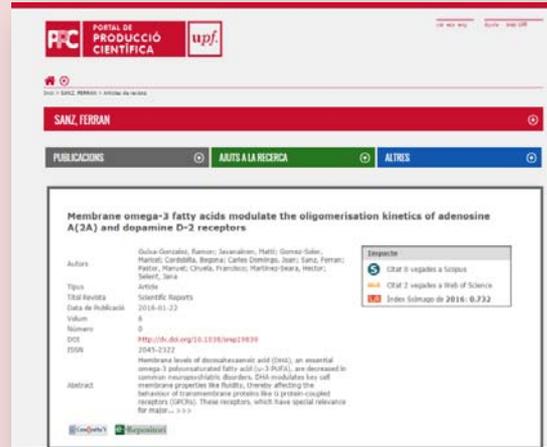
<http://guiesbibtic.upf.edu/data/es/datos>

# 3. Infraestructura

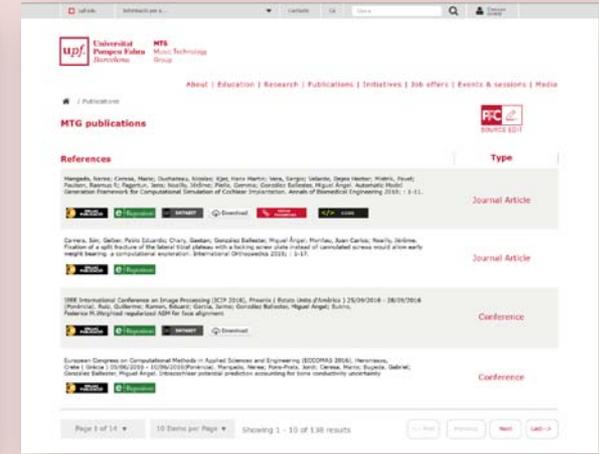
## Retos

1. Un único punto de entrada de datos y múltiples *outputs*
2. Relación publicación-*dataset*





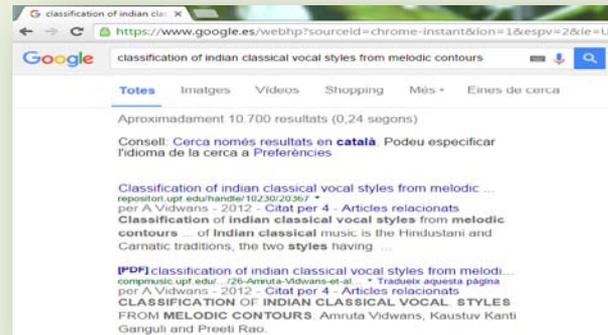
CRIS UPF



Páginas web



e-Repositori



Google



## MTG publications



SOURCE EDIT

### References

### Type

Mangado, Nerea; Ceresa, Mario; Duchateau, Nicolas; Paulsen, Rasmus R; Fagertun, Jens; Noailly, Jérôme; [Generation Framework for Computational Simulation](#)



ector; Mistrik, Pavel; [Automatic Model](#)  
ngineering 2016; : 1-11.

Journal Article

Carrera, Ión; Gelber, Pablo Eduardo; Chary, Gaetan; Gonzalez Ballester, Miguel Ángel; Morrao, Juan Carlos; Noailly, Jérôme. [Fixation of a split fracture of the lateral tibial plateau with a locking screw plate instead of cannulated screws would allow early weight bearing: a computational exploration](#). International Orthopaedics 2016; : 1-17.



Journal Article

[IEEE International Conference on Image Processing \(ICIP 2016\)](#), Phoenix ( Estats Units d'Amèrica ) 25/09/2016 - 28/09/2016 (Ponència). Ruiz, Guillermo; Ramon, Eduard; García, Jaime; González Ballester, Miguel Ángel; Sukno, Federico M. Weighted regularized ASM for face alignment



Conference

[European Congress on Computational Methods in Applied Sciences and Engineering \(ECCOMAS 2016\)](#), Heronissos, Crete ( Grècia ) 05/06/2016 - 10/06/2016(Ponència). Mangado, Nerea; Pons-Prats, Jordi; Ceresa, Mario; Bugeda, Gabriel; González Ballester, Miguel Ángel. Intracochlear potential prediction accounting for bone conductivity uncertainty



Conference

# Estimating the Rank Correlation between the Test Collection Results and the True System Performance



**Publication Type** Conference Paper  
**Year of Publication** In Press  
**Conference Name** International ACM SIGIR Conference on Research  
**Authors** Urbano J

**Impacte**

- Citat 0 vegades a Scopus
- Wos Citat 2 vegades a Web of Science
- Index Scimago de 2016: 0.732

- ENLACE PUBLICACIÓ
- Repositori
- DATASET
- Other resources
- CODE

## Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam venenatis nisi facilisis nibh cursus vehicula. Aenean pharetra sapien mi, sed eleifend erat vestibulum quis. Ut consequat nunc eget enim cursus, at scelerisque velit facilisis. Donec fringilla est ligula, at vulputate mauris bibendum vel. Sed euismod ipsum sapien, vitae elementum enim sodales et. Curabitur ante tortor, hendrerit in sagittis id, pulvinar porttitor metus. Curabitur nunc tellus, pellentesque sed odio ut, aliquet gravida urna.

## Supplementary Materials

### Run the code

All the code is pure python and uses only numpy, scipy, matplotlib. The code was run with ipython.

```
$ ipython --version
3.2.1
```

We use anaconda to install python, numpy, matplotlib, etc.



```
$ ipcluster start -n x
```

where `x` is the number of engines you want to use. You can change the number of loops directly in the script line 42. Then, run the command

```
$ ipython figure_MSE_sim.py
```

This will run the long simulation needed. The result will be stored in the folder `sim_data` and the name of the file will contain the date and time.

Copy the date and time in the file `figure_MSE_plot.py` line 61-64. Then run

```
$ ipython figure_MSE_plot.py
```

**Buscar**

Buscar  
 Esta colección

Búsqueda avanzada

**Listar**

**Todo el e-Repository**

- Comunidades y colecciones
- Por fecha de publicación
- Autores
- Títulos

- Autores
- Títulos
- Materias

**Mi cuenta**

Acceder

¿Qué es el e-Repository ?



COMPARTIR

**Estadísticas**

Ver estadísticas del ítem

## Beijing Opera Percussion Pattern Dataset

Para citar o enlazar este documento: <http://hdl.handle.net/10230/25720>



Autor: **CompMusic**

Fecha: 2016

### Resumen:

Beijing Opera Percussion Pattern (BOPP) dataset is a collection of 133 audio percussion patterns covering five pattern classes. The dataset includes the audio and syllable level transcriptions for the patterns (non-time aligned). It is useful for percussion transcription and classification tasks. The patterns have been extracted from audio recordings of arias and labeled by a musicologist.

**Publicación relacionada:** <http://hdl.handle.net/10230/25678>; <http://hdl.handle.net/10230/25677>

### Consultad el texto completo



Nombre: BOPP\_Annotations.zip  
Tamaño: 27.77Kb  
Formato: Desconocido

[Ver/Abrir](#)



Nombre: Beijing Opera Percussion Pattern ...  
Tamaño: 18.98Kb  
Formato: Microsoft Word 2007

[Ver/Abrir](#)

[Mostrar el registro completo del ítem](#)

Este documento está disponible bajo una licencia Creative Commons:



**Registro *dataset***

**Buscar**

Buscar  
 Esta colección

[Búsqueda avanzada](#)

- Listar**
- Todo el e-Repository**
- Comunidades y colecciones
  - Por fecha de publicación
  - Autores
  - Títulos
  - Materias
- Esta colección**
- Por fecha de publicación
  - Autores
  - Títulos
  - Materias

**Mi cuenta**

[Acceder](#)

¿Qué es el e-Repository ?

COMPARTIR

**Estadísticas**

[Ver estadísticas del ítem](#)

## A study of instrument-wise onset detection in Beijing Opera percussion ensembles

Para citar o enlazar este documento: <http://hdl.handle.net/10230/25678>



**Autor:** Tian, Mi ; Srinivasamurthy, Ajay ; Sandler, Mark ; Serra, Xavier

**Fecha:** 2014

**Resumen:**

Note onset detection and instrument recognition are two of the most investigated tasks in Music Information Retrieval (MIR). Various detection methods have been proposed in previous research for western music, with less focus on other music cultures of the world. In this paper, we focus on onset detection for percussion instruments in Beijing Opera, a major genre of Chinese traditional music. A

**Datos relacionados:** <http://hdl.handle.net/10230/25720>

DOI: <http://dx.doi.org/10.1109/ICASSP.2014.6853981>

Consultad el texto completo

	<b>Nombre:</b> Tian_ICASSP2014.pdf	<a href="#">Ver/Abrir</a>
	<b>Tamaño:</b> 522.8Kb	
	<b>Formato:</b> PDF	

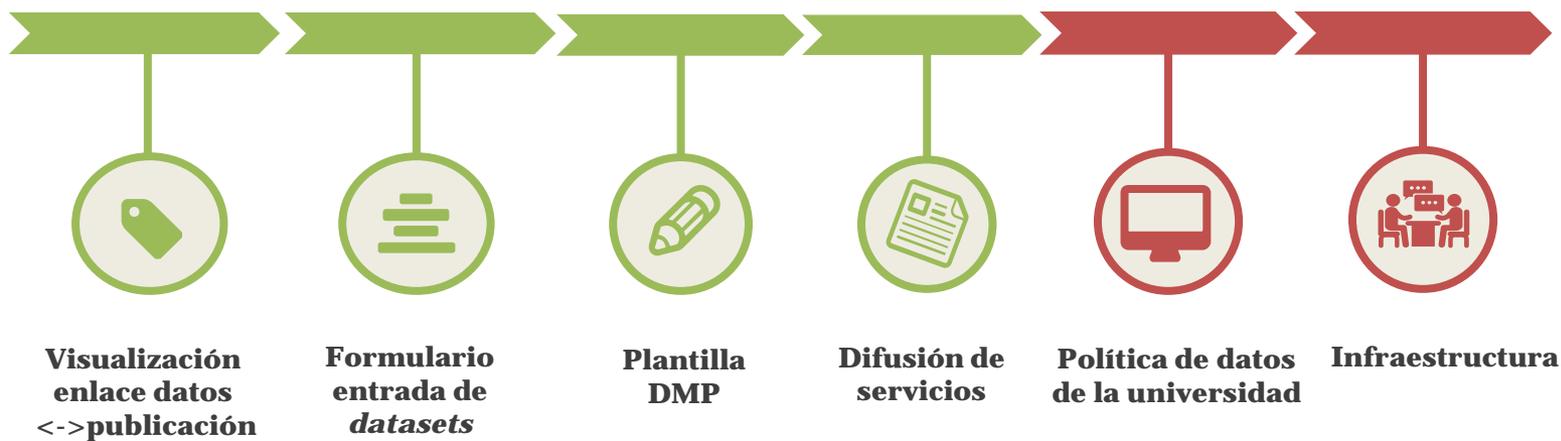
[Mostrar el registro completo del ítem](#)

**Registro publicación**

# **Sobreviviendo a los datos**



## 4. Líneas de trabajo





Formación y benchmarking



Cooperación



Beta permanente



Rol activo



**¡ Sobreviviendo a los datos !**





¡Muchas gracias!

Anna Casaldàliga

[anna.casaldaliga@upf.edu](mailto:anna.casaldaliga@upf.edu)

Marina Losada

[marina.losada@upf.edu](mailto:marina.losada@upf.edu)

Natàlia Plancheria

[natalia.plancheria@upf.edu](mailto:natalia.plancheria@upf.edu)



Biblioteca de la UPF

<http://www.upf.edu/bibtic>



**Universitat  
Pompeu Fabra**  
*Barcelona*