



Centro Singular de Investigación
en Tecnoloxías da
Información

Título

Guía para a publicación de Datasets na Web
do CiTIUS

citi.usc.es

Ónde podo publicar un Dataset?

No seguinte [enlace](#) podes publicar os teus Datasets. A información que se debe publicar para cada Dataset dependerá do obxectivo do mesmo. Por exemplo, no caso dun Dataset de clasificación podería ser:

- Descrición do Dataset, indicando nome, número de instancias, entradas, clases, etc.
- Referencia a algunha publicación a cal se asocie o Dataset.
- Licenza de dereitos de uso do Dataset.
- Arquivos de descarga.

A seguinte imaxe mostra como aparece un dataset na web do CiTIUS:

Fish Ovary

The four tar.gz files contain four data sets. Each tar.gz file includes several files: see the readme.txt file inside the tar.gz file for further information about each one. The data are in the files 'XX_R.dat' (e.g. oocytes_merluccius_states_2f_R.dat). These data are texture features of oocytes extracted from histological images of fish ovary of species Merluccius and Trisopterus. There are 4 data sets (each one corresponds to a different combination of texture features, as described in the reference below):

Set	Instances	Inputs	Classes
Merluccius States	1022	25	3
Merluccius Nucleus	1022	41	2
Trisopterus States	912	25	3
Trisopterus Nucleus	912	25	2

The values of each input are standardized with zero mean and standard deviation one. The classes in the Merluccius and Trisopterus state data sets are ordinal, i.e., there is an order from class 0 to class 1 and from class 1 to class 2, which the three classes are states of development of oocytes. Therefore, these two data sets are ordinal classification problems.

Reference

```
{
  @article{ruffinolo,
    author={E. González-Rufino, P. Carrilón, E. Cernadas, M. Fernández-Delgado, R. Domínguez-Petit},
    journal={Pattern Recognition},
    pages = {2391-2407},
    title = "[Exhaustive comparison of colour texture features and classification methods to discriminate cells categories in histological images of fish ovary]",
    volume = {46},
    year = {2013}
  }
```

License



You can use this dataset on your publication as long as you include a citation to the reference on this page. When including a link to this dataset, please use this page instead of linking the file directly.





Download

- [oocytes_merluccius_states_2f.tar.gz](#)
- [oocytes_merluccius_nucleus_4d.tar.gz](#)
- [oocytes_trisopterus_states_5b.tar.gz](#)
- [oocytes_trisopterus_nucleus_2f.tar.gz](#)






1. Presentación dun Dataset na web do CiTIUS


Qué son as licencias Creative Commons?

As licencias [creative commons](#) son un conxunto de modelos de licenza desenvolto pola organización Creative Commons. Este tipo de licencias permiten asociar unha serie de dereitos as persoas que queiran utilizar a obra, como son:

	Recoñecemento: En calquera explotación da obra autorizada pola licenza fará falta recoñecer a autoría.
	Non Comercial: A explotación da obra queda limitada a usos non comerciais.
	Sen obras derivadas: A autorización para explotar a obra non inclúe a transformación para crear unha obra derivada.
	Compartir Igual : A explotación autorizada inclúe a creación de obras derivadas sempre que manteñan a mesma licenza ao ser divulgadas.

Pódese elixir unha combinación de calquera destes dereitos, así se xeran os seguintes modelos de licencias:

	Recoñecemento: Permítese o uso da obra con calquera fin, incluída a explotación comercial, e a creación de obras derivadas sen ningún tipo de restrición.
	Recoñecemento – NonComercial: Permítese a xeración de obras derivadas sempre que non se faga un uso comercial. Tampouco se pode utilizar a obra orixinal con finalidades comerciais.
	Recoñecemento – NonComercial – CompartirIgual: Non se permite un uso comercial da obra orixinal nin de posibles obras derivadas, a distribución das cales debe facerse cunha licenza igual á que regula a obra orixinal.
	Recoñecemento – NonComercial – SenObraDerivada: Non se permite un uso comercial da obra orixinal nin a xeración de obras derivadas.
	Recoñecemento – CompartirIgual: Se permite o uso comercial da obra e das posibles obras derivadas, a distribución das cales debe facerse con unha licenza igual á que regula a obra orixinal.







	<p>Recoñecemento – SenObraDerivada: Se permite o uso comercial da obra pero non á xeración de obras derivadas.</p>
---	--

Cómo asocio unha licenza a un dataset da web co CiTIUS?

Asociar unha licenza a un Dataset da Wiki é sinxelo. No apartado *Licencia*, que aparece na imaxe 1, e suficiente asociar o logo da licenza que se vai a utilizar así como un pequeno texto indicando as condicións de uso. Por exemplo, no caso anterior:

*You can use this dataset on your publication as long as you include a citation to the reference on this page.
When including a link to this dataset, please use this page instead of linking the file directly*

Podes utilizar *emoticonos* para incluír, dunha maneira sinxela, os logotipos asociados a licenza. Por exemplo:

	<p>:cc-by:</p>
	<p>:-cc-by-nc:</p>
	<p>:cc-by-nc-sa:</p>
	<p>:cc-by-nc-nd:</p>
	<p>:cc-by-sa:</p>
	<p>:cc-by-nd:</p>