



# Ciencia Abierta y Principios FAIR: Buenas Prácticas en Publicaciones, Datos y Software

**Authors**

Iker Martín-Álvarez

Contribución al progreso del conocimiento en el área	10%
Impacto científico	60%
Impacto social	10%
Contribución a la ciencia abierta	10%
Aportación preferente	10%
<b>Posibles circunstancias reductoras de la calificación</b>	
Reiteración de publicaciones	Exclusión de aportaciones
Conflicto de interés editorial	Exclusión de aportaciones
Malas prácticas de publicación	Exclusión de aportaciones
Procesos de revisión no contrastables	-25% a -100%
Autoría insuficientemente justificada	-15%



Publicar en abierto, artículos, datasets y software

Molts organismes finançadors inclouen com un dels seus requisits la publicació en obert de les dades d'investigació:

- **Horizon Europe.** Obligacions per als beneficiaris:
  - Gestionar les dades d'investigació de forma responsable i que les dades compleixen amb els **principis FAIR** (Findable, Accessible, Interoperable, Reusable).
  - Elaborar un Pla de gestió de dades (Data Management Plan)
  - Dipositar les dades en un repositori de confiança
  - Assegurar l'accés obert a les dades via el repositori amb una llicència CC-BY o CC-0 o equivalent.
  - Les metadades de les dades d'investigació han de ser obertes i amb una llicència CC-0 o equivalent.
- **Plan Estatal de Investigación Científica y Técnica y de Innovación (PEICTI) 2024-2027.** Entre els principis bàsics està:
  - Accés obert als resultats de les activitats subvencionades, tant les publicacions científiques com les dades d'investigació
  - Les dades d'investigació han de complir els principis FAIR, i sempre que siga possible, s'han de difondre en obert.
- **H2020.** A partir de 2017, tots els projectes beneficiaris d'aquest programa tenen l'obligació de:
  - Elaborar un Pla de gestió de dades (Data Management Plan)
  - Dipositar les dades, respectant sempre la normativa relativa a les dades de caràcter personal i altres possibles excepcions, en un repositori d'accés obert.
  - Les dades d'investigació han de complir els principis FAIR.

# Índice:

1. Objetivo
2. Publicación en abierto
3. Secciones extra artículos
4. Principios FAIR
5. Datasets
6. Software
7. Planes de Gestión de Datos
8. Conclusiones

## Se valora positivamente:

- Artículo final en abierto
- Preprint en abierto
- Postprint en abierto

Articles most recently published online for this journal.



Open Access | Research article | First published Jul 13, 2025


[Characterization of quantized inference with transformer encoders on low power CPUs](#)

Héctor Martínez , Sandra Catalán , Adrián Castelló , Enrique S Quintana-Ortí 



Restricted access | Research article | First published Jun 16, 2025

[Simulation-based machine learning for real-time assessment of side-branch hemodynamics in coronary bifurcation lesions](#)

Arash Ghorbannia, Cyrus Tanade, Ayman Yousef, Nusrat Sadia Khan, Madhurima Vardhan  [...]

[View all](#) 

## Se valora positivamente:

- Artículo final en abierto
- Preprint en abierto
- Postprint en abierto



A screenshot of an arXiv paper page. The header includes the Cornell University logo and the text 'We gratefully acknowledge support from the Simons Foundation, member institutions, and all contributors.' with a 'Donate' button. The navigation bar shows 'arXiv &gt; cs &gt; arXiv:2403.17107' and search options. The main content area is titled 'Computer Science &gt; Distributed, Parallel, and Cluster Computing' and features the paper title 'Design Principles of Dynamic Resource Management for High-Performance Parallel Programming Models' by Dominik Huber, Martin Schreiber, Martin Schulz, Howard Pritchard, and Daniel Holmes. The abstract discusses Dynamic Resource Management (DRM) in HPC. The page also includes a 'Submission history' section and a 'References &amp; Citations' sidebar with links to 'Access Paper' (PDF, HTML, TeX, etc.) and 'Current browse context: cs.DC'.

Se valora positivamente:

- Artículo final en abierto
- Preprint en abierto
- Postprint en abierto

Es la versión aceptada del artículo, pero sin maquetar por la revista.

Subir a repositorio abierto (El de la UJI por ejemplo)

**REVISAR LICENCIA**



ACTES DE LA CONFERÈNCIA

**Malleability Implementation in a MPI Iterative Method**

Martin-Alvarez, Iker ; Aliaga, Jose I. ; Castillo, Maria Isabel ; Mayo, Rafael ; Iserte, Sergio

IEEE

Proceedings / IEEE International Conference on Cluster Computing, 2021, p.801-802

[HTML](#) >

[Disponible en línea](#) >

AMUNT

ENVIAR A

VEURE EN LÍNIA

INFORMACIÓ DE...

ENLLAÇOS

ETIQUETES

EXPLORAR

Enviar a

CORREU ELECTRÒNIC

MENDELEY

EXPORTACIÓ BIBTEX

EXPORTACIÓ RIS (MENDELEY, ZOTERO)

ENDNOTE

ENLLAÇ PERMANENT

CITA

Veure en línia

# Índice:

1. Objetivo
2. Publicación en abierto
3. **Secciones extra artículos**
4. Principios FAIR
5. Datasets
6. Software
7. Planes de Gestión de Datos
8. Conclusiones

Researcher I. Martín-Álvarez was supported by the pre-doctoral fellowship ACIF/2021/260 from Valencian Region Government and European Social Funds.

### Data availability

The source codes for DMR (<https://gitlab.bsc.es/siserte/dmr>), MaM ([https://lorca.act.uji.es/gitlab/martini/malleability\\_benchmark/-/tree/PPAM24-JournalSpecialIssue](https://lorca.act.uji.es/gitlab/martini/malleability_benchmark/-/tree/PPAM24-JournalSpecialIssue)) and MPDATA (<https://gitlab.bsc.es/siserte/mpdata-dmr>) are publicly available. All data used to create the tables and figures in the Experimental Results section are available at (39).



- [16] A. B. Yoo, M. A. Jette, Utility for Resource Management (JSSPP), 2003, pp. 44–60.
- [17] P. K. Smolarkiewicz, Multitransport algorithm: an overview, *Journal of Computational Physics* 50 (10) (1982) 109–117.
- [18] K. Rojek, R. Wyrzykowski, Stencil-based 3D MPDATA: Simulations on GPU Clusters, *Journal of Supercomputing* (2017) 664–675.
- [19] K. Rojek, R. Wyrzykowski, Stencil-based 3D MPDATA: Computation: Practice and Performance, *Journal of Supercomputing* (2017) 664–675.
- [20] J. I. Aliaga, M. Castillo, S. Aliaga, Survey on Malleability Solutions, *Applied Science* (2021) 1–12.
- [21] A. Tarraf, M. Schreiber, A. C.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data availability

Data will be made available on request.

### Acknowledgments

The research presented in this work is based upon work by



Researcher I. Martín-Álvarez was supported by the predoctoral fellowship ACIF/2021/260 from Valencian Region Government and European Social Funds.

### Data availability

The source codes for DMR (<https://gitlab.bsc.es/siserte/dmr>), MaM ([https://lorca.act.uji.es/gitlab/martini/malleability\\_benchmark/-/tree/PPAM24-JournalSpecialIssue](https://lorca.act.uji.es/gitlab/martini/malleability_benchmark/-/tree/PPAM24-JournalSpecialIssue)) and MPDATA (<https://gitlab.bsc.es/siserte/mpdata-dmr>) are publicly available. All data used to create the tables and figures in the Experimental Results section are available at (39).

- [16] A. B. Yoo, M. A. Jette, Utility for Resource Management (JSSPP), 2003, pp. 44–60.
- [17] P. K. Smolarkiewicz, Mult transport algorithm: an over Methods in Fluids 50 (10) (
- [18] K. Rojek, R. Wyrzykowski, Simulations on GPU Cluste (2017) 664–675.
- [19] K. Rojek, R. Wyrzykowski, Stencil-based 3D MPDATA Computation: Practice and I
- [20] J. I. Aliaga, M. Castillo, S Survey on Malleability Soli Computing, Applied Scienc
- [21] A. Tarraf, M. Schreiber, A. C

The screenshot shows the GitLab interface for the 'Proteo' repository. At the top, there's a dropdown menu set to 'master' and the repository name 'malleability\_benchmark'. Below that, a notification banner indicates 'Minor changes to README' by 'iker\_martin' 8 months ago. The main part of the screenshot is a table with two columns: 'Name' and 'Last commit'. The table lists several files and folders, each with a corresponding commit message.

Name	Last commit
Analysis	New version of Proteo
Codes	New version of Proteo
Exec	New version of Proteo
Results	New version of Proteo
.gitignore	New version of Proteo
Manual.pdf	New version of Proteo
README.md	Minor changes to README

## Dataset (2025) for article "Resource Optimization with MPI Process Malleability for Dynamic Workloads in HPC Clusters"

Martin Álvarez, Iker (Contact person)<sup>1</sup> ; Iserte, Sergio (Contact person)<sup>2</sup>

Show affiliations

### Contributors

#### Contact persons:

Aliaga Estellés, José Ignacio ; Castillo Catalán, María Isabel ; Rojek, Krzysztof<sup>1</sup>

Folwarska, Weronika ; Peña Monferrer, Antonio J.<sup>2</sup>

Show affiliations

# Índice:

1. Objetivo
2. Publicación en abierto
3. Secciones extra artículos
4. Principios FAIR
5. Datasets
6. Software
7. Planes de Gestión de Datos
8. Conclusiones

FAIR:

- 4 principios.
- Solo dos requieren trabajo personal

Se encarga el repositorio (DOI por ejemplo)

Trabajo principalmente del investigador

## 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.

Licencias:

- Datos → CC-BY
- Software → GNU GPL
- <https://choosealicense.com/>



## 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.

Utilizar vocabulario aceptado por la comunidad para describir datos.

Incluir metadatos que describan los datos para que puedan ser comprendidos por 3<sup>o</sup>.

## 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.

# Índice:

1. Objetivo
2. Publicación en abierto
3. Secciones extra artículos
4. Principios FAIR
5. **Datasets**
6. Software
7. Planes de Gestión de Datos
8. Conclusiones

Subir a Zenodo los datos usados para cada artículo.

Tiene control de versiones, es gratuito y permite tener un DOI. (Referencia en artículo)

Subir descripción de datos, datos y códigos usados en el análisis (<https://www.kaggle.com/>)

The screenshot shows the Zenodo interface for a dataset. At the top, there's a blue header with the Zenodo logo, a search bar, and navigation links for 'Communities' and 'My dashboard'. A notification banner indicates a 'Planned intervention' on Tuesday, July 22nd, 05:00 UTC. The dataset is published on February 5, 2025, and is version 1.0.0. It has buttons for 'Dataset', 'Open', 'Edit', 'New version', and 'Share'. The dataset title is 'Dataset (2025) for article "Resource Optimization with MPI Process Malleability for Dynamic Workloads in HPC Clusters"'. The authors listed are Martín Alvarez, Iker (Contact person) and Iserte, Sergio (Contact person). There are 26 views and 8 downloads. The description states the dataset is organized into three stages: raw data, preprocessed data, and processed data. It also provides a 'Dataset Structure' section and a 'Cite all versions?' section with a DOI: 10.5281/zenodo.14812022.

<https://zenodo.org/records/14812022>

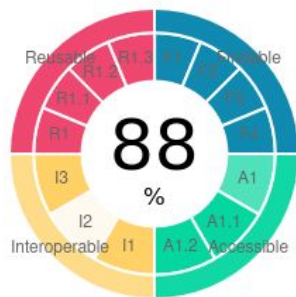
<https://www.kaggle.com/datasets/uciml/student-alcohol-consumption>

Comprobar que sigue FAIR en <https://www.f-uji.net/index.php>

Requiere que los datos ya estén en un repositorio público como Zenodo.

Asegurar que el dataset tiene un nivel adecuado.

### Summary:



	Score earned:		Fair level:
<b>Findable:</b>	7 of 7		<b>advanced</b>
<b>Accessible:</b>	6 of 7		<b>moderate</b>
<b>Interoperable:</b>	4 of 6		<b>moderate</b>
<b>Reusable:</b>	6 of 6		<b>moderate</b>

# Índice:

1. Objetivo
2. Publicación en abierto
3. Secciones extra artículos
4. Principios FAIR
5. Datasets
6. **Software**
7. Planes de Gestión de Datos
8. Conclusiones

Dos opciones en el artículo:

- Enlace a GitHub.
- Enlace a Zenodo.

Utilizar una rama o tag para marcar el código para ser reutilizado.

Crear un README con las especificaciones de la máquina, software extra, pasos de instalación y como ejecutar

Ejemplo:

[https://lorca.act.uji.es/gitlab/martini/malleability\\_benchmark/-/tree/JournalSupercomputing23/24?ref\\_type=heads](https://lorca.act.uji.es/gitlab/martini/malleability_benchmark/-/tree/JournalSupercomputing23/24?ref_type=heads)

### Installation

---

#### Prerequisites

Before installing, ensure you have the following prerequisites:

- MPI (MPICH) installed on your system. This code has been tested with MPICH versions 3.4.1 and 4.0.3
- Slurm is installed on your system. This code has been tested with slurm-wlm 19.05.5.

#### Steps

1. Clone the repository to your local machine:

```
$ git clone http://lorca.act.uji.es/gitlab/martini/malleability_benchmark.git
$ cd malleability_benchmark
$ git checkout PPAM24-JournalSpecialIssue
```



Los artifacts son para demostrar el funcionamiento de un código.  
Se adjuntan junto al artículo como apéndice.



ANACONDA®



docker®

Normalmente basados en un Docker que tras iniciarlo se tiene un entorno preparado para probarlo.

# Índice:

1. Objetivo
2. Publicación en abierto
3. Secciones extra artículos
4. Principios FAIR
5. Datasets
6. Software
7. Planes de Gestión de Datos
8. Conclusiones



## 10 pasos para elaborar un Plan de Gestión de Datos

Un Plan de Gestión de Datos (PGD) o Data Management Plan (DMP) es un **documento formal, que debe presentarse al inicio de la investigación, en el que se describe qué**

**vas a hacer con tus datos durante y después de finalizar tu investigación** y que puede modificarse si se producen cambios en el proceso de la misma.

### ¿Por qué?

Es una **buena práctica**, es un **elemento clave de Open Science** y es **obligatorio** en los nuevos proyectos H2020.

### Herramientas gratuitas para elaborar un PGD



FGDSonline  
(Consortio Madrid)  
<http://dmp.consortio-madrid.es/>



DMPonline (Digital Question  
Centre, UK)  
<https://dmponline.dcs.ac.uk/>



01

Revisa los **requerimientos** de la entidad financiadora (H2020).



02

Identifica los **datos**: tipología, procedencia, volumen, formatos y ficheros.



03

Define cómo se **organizarán y gestionarán los datos**: nombre de los ficheros, control de versiones, software necesario...



04

Explica cómo se **documentarán los datos**: identifica la información a procesar, consulta si hay estándares o esquemas de metadatos, identifica herramientas que permitan gestionarlos.



05

Describe los procesos que aseguran una **buena calidad de los datos**.



06

Prepara una **estrategia de almacenamiento** (durante el proceso) y de preservación de datos (repositorio).



07

Define las **políticas de datos del proyecto**: cuestiones sobre propiedad intelectual y cómo se tratarán los datos sensibles y personales.



08

Describe cómo se **difundirán los datos**: donde, cuáles, cuándo se van a difundir. Si publicarás los datos en un repositorio, como información suplementaria del artículo o como un "data paper".



09

Asigna **roles y responsabilidades** para las personas y organizaciones participantes en el proyecto.



10

Prepara un **presupuesto realista**: la gestión de datos cuesta tiempo y dinero en términos de software, hardware, servicios y personal.

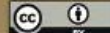


Universidades Españolas Red de Bibliotecas REBIUN

Adaptado de McInerney, W.C. Ten simple rules for creating a good data management plan. PLOS Comp Biol. 2015; 11(10)



El personal de tu Biblioteca te puede asesorar



El PGD es obligatorio para nuevos proyectos.

Se puede usar [ARGOS](#) para preparar el documento.

Hay que depositar el documento en un repositorio público.

# Cómo gestionar los datos a lo largo de proyecto

Planificar	<ul style="list-style-type: none"><li>• Crear PGD</li></ul>
Organizar	<ul style="list-style-type: none"><li>• Establecer estructura de carpetas y nomenclatura ficheros</li></ul>
Describir	<ul style="list-style-type: none"><li>• Usar metadatos para registros y contenido</li><li>• Crear Fichero Readme.txt</li></ul>
Procesar y analizar	<ul style="list-style-type: none"><li>• Estructura, limpieza, validación, revisión..</li></ul>
Publicar y compartir	<ul style="list-style-type: none"><li>• Introducir en repositorio</li><li>• Asignar licencias</li></ul>
Preservar	<ul style="list-style-type: none"><li>• Establecer período de conservación de los datos</li><li>• Medidas para conservar</li></ul>

10

# Índice:

1. Objetivo
2. Publicación en abierto
3. A añadir en artículos
4. Principios FAIR
5. Datasets
6. Software
7. Planes de Gestión de Datos
8. Conclusiones

**Publicar datasets y códigos, citándolos en el artículo**

**Hacer uso de Zenodo y F-UJI para seguir principios FAIR en datasets o PGD**

**Asegurar que todo lo publicado tiene una licencia adecuada**



# Ciencia Abierta y Principios FAIR: Buenas Prácticas en Publicaciones, Datos y Software

## Authors

Iker Martín-Álvarez ([martini@uji.es](mailto:martini@uji.es)),