

## The quality of shared datasets in agriculture: a further step in the study of shared data quality

Petrosyan, Luiza (1); Sixto-Costoya, Andrea (1,2); Valderrama-Zurián, Juan Carlos (2,3); Peset, Fernanda (1); Ferrer-Sapena, Antonia (1); Aleixandre-Benavent, Rafael (2,4)

1. Institute of Pure and Applied Mathematics - IUMPA (Polytechnic University of Valencia - UPV). Spain; 2. Social and Health Research Unit. UISYS Group; 3. Department of History of Science and Information Science. School of Medicine. University of Valencia. University of Valencia. Spain; 4. Ingenio (CSIC-UPV)

### INTRODUCTION.

Seeking the best strategies to improve Agricultural research is a need recognized by all international organizations, including the UN Sustainable Development Goals. Our objective is to find methodologies to determine compliance with the FAIR principles in this field through two resources: Data Citation Index (DCI) and F-UJI tool, based on 16 metrics and developed by the FAIRsFAIR team.

### METHODOLOGY.

Web of Science (WoS) was used to retrieve all records of journals in Agriculture-related categories in the Q1 of the Journal Citation Reports - 2020 (n=175).

67,763 records were retrieved, of which 201 contained references to datasets. These 201 records were searched in the DCI in WoS, resulting in 116 datasets.

27 duplicate records were deleted, leaving a total of 87 datasets. Information collected: title of the dataset, repository and unique identifier.

Datasets were assessed through F-UJI. The results are divided into three FAIR levels:  
1) advanced  
2) moderated  
3) initial

### RESULTS.

Of the 87 datasets, 60 (those with a DOI) could be analysed. The repositories used were Dryad (n=30) and Zenodo (n=30).

Of those that could not be analysed, 13 belonged to the European Nucleotide Archive repository, 11 to Gene Expression Omnibus, and 3 to MetaboLights.

For each principle of advanced category datasets, the best results with 80% were findable and accessible, meanwhile the percentage of interoperable and reusable principles is only 1.7%

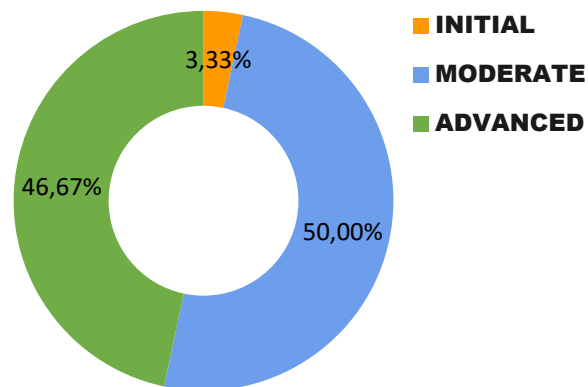


Figure 1: Results obtained in F-UJI by level

### CONCLUSIONS.

The first two principles (findable and accessible) are being fulfilled in a very complete way. However, reusability and interoperability are still far from being developed, which shows us that there is still work to do in this area to reach the FAIRness of datasets.

**Acknowledgement:** Grant “Stable methodologies to evaluate and measure quality, interoperability, blockchain and reuse of open data in the agricultural field” funded by MCIN/ AEI /10.13039/501100011033. LP is beneficiary of a predoctoral grant (PRE2020-092585) and ASC is beneficiary of a postdoctoral grant (MS21-020).