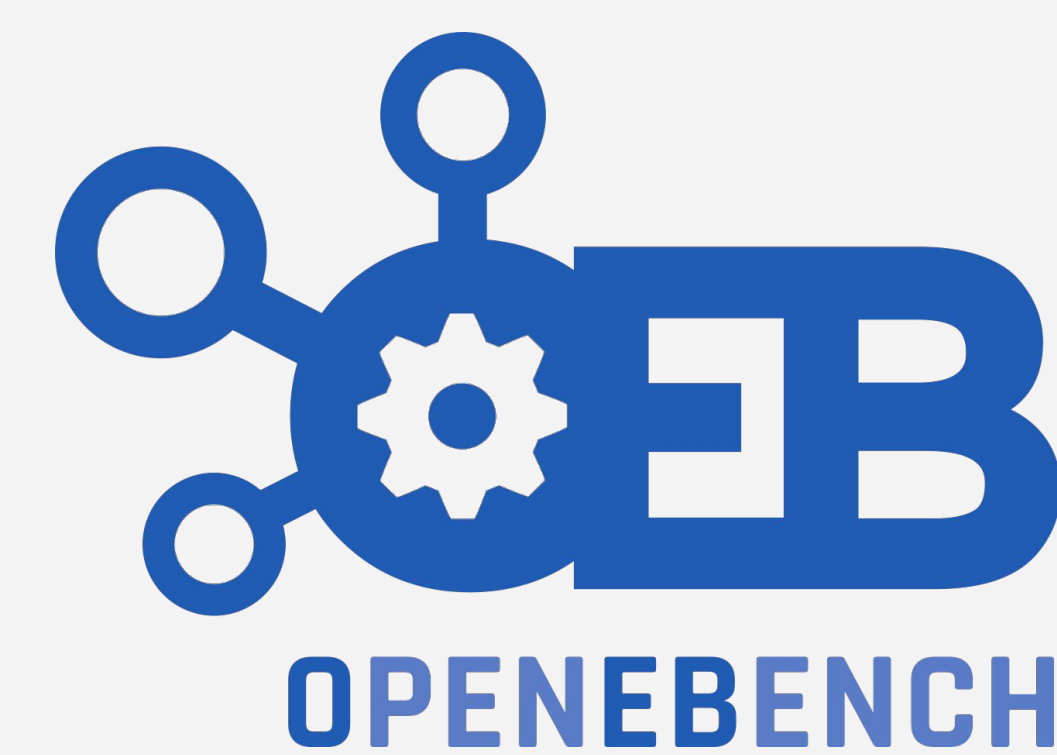


# OpenEBench 2024 update

## Capturing the new perspectives and current needs of communities and scientific projects for scientific benchmarking



Anna Redondo Guitarte <sup>1,2</sup>, Laura Portell-Silva <sup>1,2</sup>, Carles Hernandez-Ferrer <sup>1,2</sup>, Laia Codó <sup>1,2</sup>, José M. Fernández <sup>1,2</sup>, Dmitry Repchevsky <sup>1,2</sup>, The OpenEBench Team, Josep Lluís Gelpí <sup>1,2,3</sup>, and Salvador Capella-Gutierrez <sup>1,2</sup>

<sup>1</sup> Barcelona Supercomputing Center (BSC); <sup>2</sup> Spanish National Bioinformatics Institute (INB/ELIXIR-ES); <sup>3</sup> Dept Biochemistry and Molecular Biomedicine, University of Barcelona (UB)

**OpenEBench** is the ELIXIR open-data collaborative platform to support community-driven scientific benchmarking activities. As part of the **ELIXIR Tools Platform**, OpenEBench is dedicated to foster scientific benchmarking and technical monitoring practices of research software in Life Sciences. The open nature of the platform facilitates its use and adoption by different communities within ELIXIR and beyond.



**OpenEBench has captured an overall diversity of scientific benchmarking needs since its creation. This diversity is also present among the new communities.**

- 1** The **OpenEBench** platform hosts 12 active communities onboard in May 2024. It is expected to reach around 20 in the near future, driven by the engagement with different European projects like **DataTools4Heart**, **EuCanImage** or **EUCAIM**.

Previous communities reached level 2 of engagement, while upcoming engagements will extend to levels 2 and 3, including the evaluation of AI models.



### When to engage to each level

#### Level 3 (reproducible event)

- Non disclosure of the data for evaluation.
- Need for additional metrics, e.g. technical performance.

#### Level 2 (reproducible metrics)

- Evaluate participants contributions within the platform.
- Facilitate continuous evaluation even when the benchmarking event has finalized.

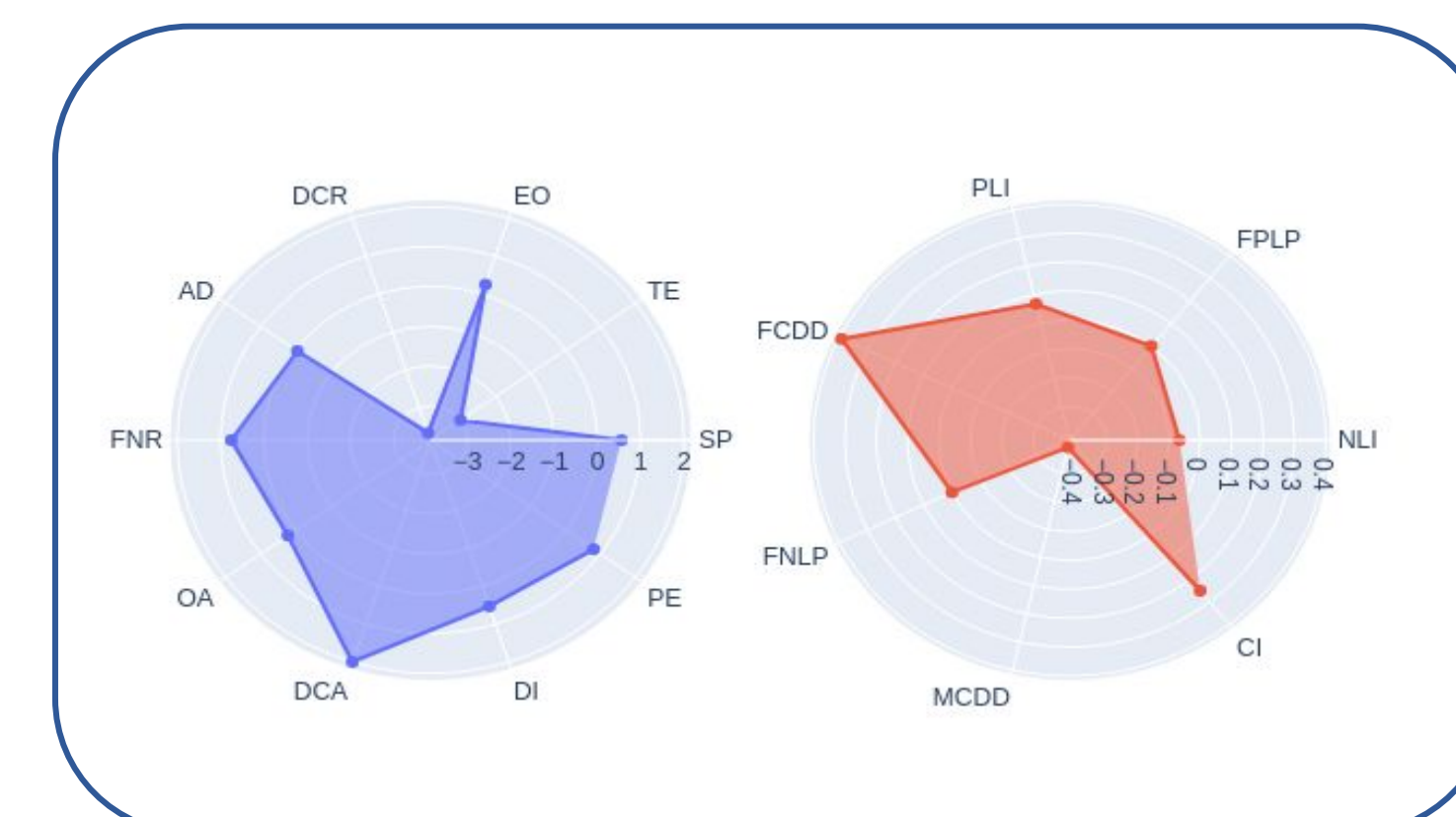
#### Level 1 (initial)

- Facilitate publishing benchmarking results without any additional requirement.
- Long-term storage for results following the FAIR data principles.

- 2** New scientific communities are focusing on the rigorous evaluation of **AI models**, with particular emphasis on ensuring these systems' **trustworthiness** and **ethical integrity**.

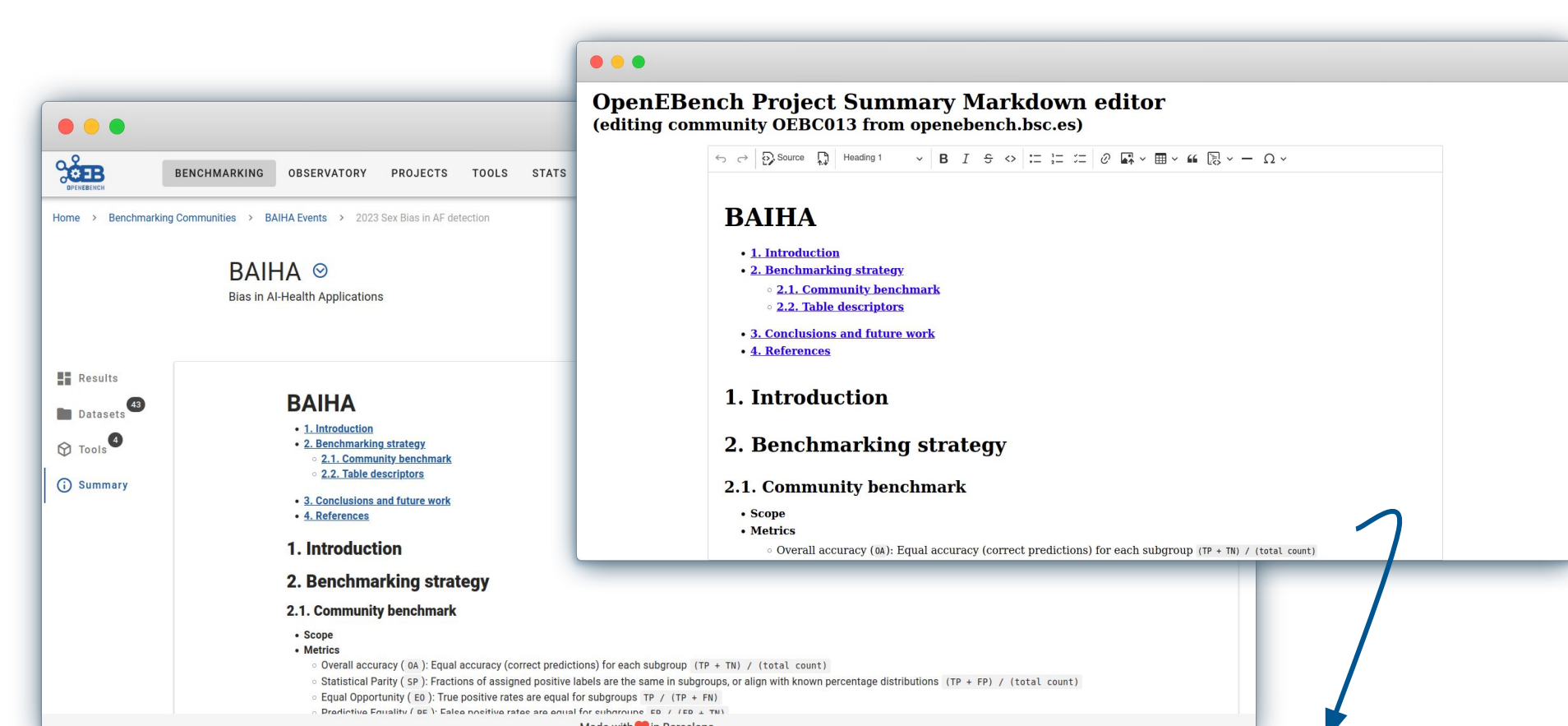
OpenEBench has responded to the influx of new participants by enhancing its back-end capabilities to accommodate these emerging challenges. New visualizations will also be introduced in the near future.

### Multidimensional metrics for biases in AI models



Assessment of sex and gender biases metrics (e.g. Treatment Equality, Female Positive Label Percentage, Negative Label Imbalance) for a particular AI dataset, as evaluated by the BAHIA community.

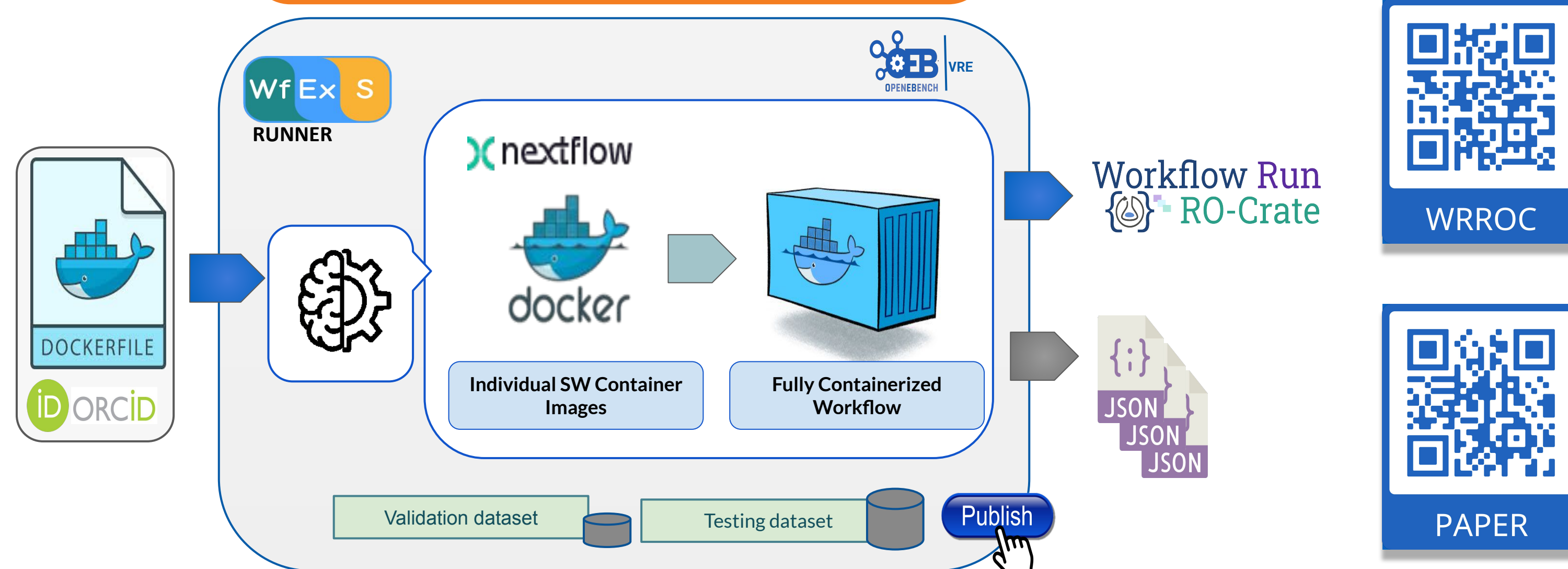
- 3** **Project Spaces** provides a flexible web space for projects and **communities** to document their efforts, provide guidelines and **share relevant information and benchmarking data** with anyone interested in engaging with them.



[inab/oeb-project-markdown-editor](https://github.com/inab/oeb-project-markdown-editor)

- 4** Progressive adoption of **RO-Crate** to facilitate results **reproducibility** and **replicability**, e.g., the same evaluation workflow can be used across different events within the same community with different reference datasets.

### Reproducible workflow orchestration



## Contact

Anna Redondo Guitarte  
[ana.redondo@bsc.es](mailto:ana.redondo@bsc.es)

The OpenEBench Team  
<https://openebench.bsc.es/team>

**Barcelona Supercomputing Center (BSC)**  
Plaça Eusebi Güell, 1-3, 08034-Barcelona, Spain

