

# Workshop on Challenge Datasets ‘Juelich Problems’

November 8, 2018

## Participants:

Axer, Markus	Hoffmann, Maximilian	Polen, Tino
Bongen, Patrick	Ippoliti, Emiliano	Rank, Sven
Bruns, Benjamin	Jitsev, Jenia	Sachs, Christian Carsten
Denker, Michael	Kleinert, Felix	Scharr, Hanno
Dreyer, Jens	Fischbach, Andreas	Schiffer, Christian
Fritz, Jakob	Leufen, Lukas	Vereecken, Luc
Göbbert, Jens Henrik	Lobet, Guillaume	Wagenknecht, Dr.-Ing. Gudrun
Grün, Sonja	Nagel, Kerstin	Weis, Susanne
Günther, Gebhard	Neuner, Irene	Wenzel, Susanne
Hagemeier, Björn	Pieperhoff, Peter	

Slides are provided at:

[https://gitlab.version.fz-juelich.de/MLDL\\_FZJ/General\\_Wiki/wikis/Juelich-Problems](https://gitlab.version.fz-juelich.de/MLDL_FZJ/General_Wiki/wikis/Juelich-Problems)

## Context

- Publishing datasets is good scientific practice, but publishing datasets and adding an interesting challenge greatly increases visibility/citations
- Dataset challenges successfully published and promoted as “Juelich Problems” (working title) can support both the career development of individual researchers and the brand development of the Forschungszentrum
- Juelich problems should, in mid- or long term, have an HPC perspective, to link them to the strengths of the Forschungszentrum
- To engage in the community and build capacity in machine learning will help the researchers and groups to position themselves for HGF initiatives (and others) and thus to acquire funding

## Take Home Message

- Problem Statement Papers are a valid way of publishing problems rather than solutions
- FAIR+ principles should be addressed when publishing challenges / problems
- FAIR principles should be addressed when publishing data (for challenges or not)
- Promoting the problem increases chances that it is picked up by ML/CV researchers

## Discussion

- Several promising problems have been presented, which may be good candidates for future challenges (especially, the size of the datasets is small enough to be send to interested researchers).
- General positive feedback and first interests collected.
- ‘Juelich problems’ may be set up, in case a suitable set of challenges can be formulated. Critical mass not yet reached (3 concrete proposals discussed, more may come in the next days)
- Problems on big datasets (that cannot be send around) may need different approaches, as access to compute resources need to be managed properly. First steps and possible options were discussed e.g.
  - Set up a hierarchy of problems, where a small scale problem needs to be solved in a first challenge, and the winners of this challenge would be then invited to address the larger scale problem
  - Compute time would be applied for by the challenge proposer within the usual review schemes
  - Challenge winners would then get part of this compute time
  - When ‘real’ HPC is needed, rather than cloud computing, technical restrictions like available toolboxes may be formulated as part of the challenge, e.g. that DNNs need to be formulated in Keras or Pytorch or whatever toolbox performs best. Here the experience and groundwork of JSC people is indispensable to shape big data challenges properly.

## Next Steps

- Please show up, i.e., send us an Email, whether you are interested to publish your specific problem and according data.
- We will set up a meeting with these particular researchers in order to discuss details.
- We will set up a meeting with colleagues of JSC / HAF in order to discuss possible solutions for setting up a suitable platform for benchmarking (Contact: Björn Hagemeier)
  - ideally using Jupyter notebooks, as shown by Jens Henrik Göbbert
  - with cloud-based solution as minimal solution – sandboxing easily possible
  - ideally with perspective of using HPC
- Set up (web) platform to showcase Juelich problems