



EGI-ACE Open Call no.1

Checkpoint meeting with Shepherds

Protein pKa and isoelectric point calculations

Jérôme PANSANEL (CNRS)

Dissemination level:

Disclosing Party:

Recipient Party:



EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.

Background about the scientific use case



The protein pKa and pKi calculation service

- An easy-to-use service that allows fast and accurate pKa and isoelectric point calculations
- Based on a Python software for flexible Poisson-Boltzmann pKa calculations
- Include a complete pKa database (pKPDB) of all proteins in the Protein Data Bank
- Currently hosted by the Portuguese National Distributed Computing Infrastructure (INCD)
- Mostly used by regional users
- <https://www.pypka.org>

Ambition, Impact, Challenge(s)



This service is of high interest for people working in protein science, and could benefit to a larger audience. Two objectives have been defined so far:

- Open the service widely at the international scale, so it can benefit a variety of scientific researchers world-wide, with an emphasis on structural biologists and bioinformaticians.
- Build a large dataset of pKa values and isoelectric points, which would be pivotal to train machine-learning algorithms.

Participants



Name	Organisation	Role
Miguel MACHUQUEIRO	Uni. Lisbon	Customer
Pedro REIS	Uni. Lisbon	Customer
Alexandre BONVIN	WeNMR	WeNMR Manager
Antonio ROSATO	WeNMR	WeNMR Manager
Mario DAVID	LIP	Resource Provider
Carlos FERNANDEZ	CESGA	Resource Provider
Jérôme PANSANEL	IN2P3-IRES	Sheperd and Resource provider

Integration Support



- To make the service usable by a larger community, they need to access a bigger amount of CPUs that they are already using. A convenient way would be to use a distributed SLURM cluster distributed over the resource providers. It could be done with the Elastic Cloud Computing Cluster (EC3) service. Another option would be to use DIRAC if the first option is not feasible.
- The customer has already an ongoing SLA with EGI (Bioisi) : <https://documents.egi.eu/public/ShowDocument?docid=2876>
The customer asks us if this SLA could be updated to include this new use case, rather than to create a new SLA.

Capacity Requirements



To develop further the PypKa service, the customer is asking for the following resources:

- 200 cores to be shared between the server, the compute nodes and the database
- 256 GB of RAM
- 5 TB of disk storage for the the database
- External IP addresses should be available for the public services.

The database should be hosted by a single server with:

- 16 cores
- 64 GB of RAM
- 1 TB SSD for the cache
- Additional 2 TB of storage (could increase up to 4 TB in 2024)

- **Current status**
 - Starting
- **When a new release of the service will be available for first tests**
 - The regional service is already on-line: <https://pypka.org/> ; expected to be available for test on EGI Infra: August
- **When the setup will be ready for production use**
 - October
- **Plans to release it in EOSC?**
 - Yes, it is planned. The WeNMR team may provide some support.
- **Plans to scale-up the setup in stages?**
 - The database does not need 5 TB at start (only 2TB)
- **Dissemination activities**
 - To be defined



Thank you!

Contact: egi-ace-po@mailman.egi.eu
Website: www.egi.eu/projects/egi-ace



[EGI Foundation](#)



[@EGI_eInfra](#)



EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.