Overview of task areas

The openTECR & de.NBI hackathon focuses on building a database infrastructure based on free components only. This will most likely be GitHub and Zenodo. We have a specific interest in distributing information about specific physical values: the apparent equilibrium constants. But first, a bit about the background of the story.

Maybe like you, we have spent time searching for apparent equilibrium constants of enzyme-catalyzed reactions (or somewhat similar thermodynamic data which is somehow relevant for biology) on the internet and in good old paper journal articles. While it's often great to read scientific prose, we were actually looking for "the one" database that would serve us that information in condensed form. Because we didn't find one, we decided to build it.

We want to save time when looking for previous work on our enzymatic reaction of interest (or related ones), and when constraining our models. We want a place to submit our newly acquired data to. We want to correct the errors we found in the database, so that others don't have to repeat the error checking.

We're creating an open database that simplifies access to and sharing of energy data for biochemical reactions in living organisms, built by and for modellers and experimentalists alike.

You can find some (limited) information about the current state of our community here: https://github.com/opentecr/opentecr

Within that setting, below you will find some aspects about the work described in more detail. Those are:

- Curation (new data)
- Curation (quality control / integration)
- Backend development
- Website development
- Advancing the science
- Improving computational tools

See also our General information -- openTECR & the de.NBI hackathon document and be sure to take the survey: openTECR hackathon survey.

Below you find more info on each task area.

Curation (new data)

You join the project because you want to bring new data in. You know this data is out there and just waits to be discovered by you.

openTECR wouldn't be the best open database, if it didn't contain the latest (and most) data on the whole web.

We are confronted with a flood of scientific papers, of which only a tiny fraction is relevant for us. There is no standard journal for the information that interests us. We have to rely on text search, on keywords and other hints, to guide us to the publications of interest. To help us navigate this, we can prioritize ("triage") publications based on educated guesses and our experience.

We extract the information from the identified publications in a database-ready format.

Connected tasks:
1
2
→ see Task list in detail

Curation (quality control / integration)

You find errors pretty annoying; and it's pretty obvious that there are quite some in the openTECR database! You're ready to dive into what we have, and make sure it comes out cleaner than before - you're our super turbo washing machine!

openTECR wouldn't be the best open database, if it wasn't quality-checked many times.

The openTECR database grew organically from a pen'n'paper collection, index cards, to printed review publications, to an online database, to a csv file. It's pretty certain that some mistakes happened along the way. Some of these mistakes can be identified by comparison of different versions of the "same" data. Others by chemical common sense, and by scripting some checks. Others require re-search to understand what others already did and how this aligns with the current state of data.

We flag problematic entries and propose / implement solutions to resolve mistakes.

Connected tasks:
3
4

5
7
8
→ see Task list in detail

Backend development

You are interested in the “database” aspect behind all of this. Computers are good friends of yours, and because they are pretty dumb, you help them achieve what we want them to be good for. You’re excited to get the system running and connect the dots!

openTECR wouldn’t be the best open database, if it wasn’t running on free infrastructure and wouldn’t link with other databases.

The data behind openTECR involves many different fields and entities, like chemicals, enzymes, metabolic reactions, literature references, and physicochemical properties. There are already specialized databases out there, from which it makes sense to integrate / pull their information, to avoid reinventing the wheel. It is very open, however, how to implement the corresponding pipeline exactly. GitHub actions have the advantage of being available without cost for us, and can achieve a lot. openTECR stores all data in json files, and is to be released in different shipping formats for each release.

We develop code to go automated from input files to static database website, while keeping true to the 15 FAIR guiding principles.

Connected tasks:
6
16
18
19
→ see Task list in detail

Website development

You know the project would benefit from some proper communication… like, seriously, where can you learn right at all what openTECR is about?! Clearly there was no super-power communicator like you around – until now!

openTECR wouldn’t provide the best community experience in the world, if it hadn’t a website that gets straight to the point, and its communication wouldn’t flow like water.

We are a bunch of scientists and tend to be not the strongest visual and/or textual communicators. There is a GitHub repository with some information about openTECR, and a mailing list. Some communication materials were developed for presentations, conference posters, and for the OLS program. But there is still no website which communicates crisply what openTECR is, what it's about, and how to get involved.

We create a web presence for openTECR.

Connected tasks:
9
10
11
17
20
→ see Task list in detail

Advancing the science

You are here for the science. You are building upon the data of openTECR and want to see where that takes you and humanity.

openTECR wouldn't be a VIP (very important project), if the science which it enables wouldn't be the coolest science.

In the advancement of science, data resources like openTECR are the first stepping stone (with publication being even a step before that). Tools like eQuilibrator, COBRA, dGPredictor, and possibly other packages build upon those data aggregations. And then comes science again – a lot is still to be done and the next idea is probably already in your head. openTECR is intended to be a safe community, where you can share research ideas and realize them together with friends, not competitors. Further experiments to add more apparent equilibrium constants would be great, but are probably impractical in a hackathon, so we will most likely focus on advances which can be realized through computations.

We work together to create scientific insight while standing on the shoulder of giants.

Connected tasks:
13
→ see Task list in detail

Improving computational tools

You contribute to computational tools that build upon openTECR data. You want to take your time to improve your tool, and would be happy to have some company on that journey.

openTECR wouldn’t be a great community, if we wouldn’t meet people interested in the same things here.

Data resources like openTECR only thrive when they are used. The data behind openTECR is already used in eQuilibrator, COBRA, dGPredictor, and other packages. These tools were often developed by researchers, not professional software engineers. They do exactly what they were intended to do – if you know how to make them run and use them. There is always something which can be improved, documented better, … you name it.

We make the experience with tools based on openTECR data more smoothly.

Connected tasks:
12
14
15
→ see Task list in detail