# Notes on the PePPEr project:

**The platform will be called PePPEr:** A Permeability , Porosity, and Pycnometry volcanic Eruption sample database

Django will be the ideal hosting platform as it is Python-controlled, which will be much easier for post-project editing/accessing for myself. It also means the platform is hosted as a webspace, which makes it very user friendly to access. Check the “Possible design” slides to help with the platform framework.

General users - most likely an earth scientist looking to download/plot/upload and compare data of a similar nature. Most of these will be volcanologists who collect similar data in their research. However, it will be completely open source, so anyone can use it if interested. Might be of use for student research or class projects.

Any user can download the data, but only the authors and programme developers (yourself, Samuel, Mathieu) can edit and maintain the main framework and structure. Best approach: using Django for development and Github for locked file storage and repository (author and developer access)

Currently there are 24 categories and over 6100 data entries. Data categories are unlikely to change, but that option needs to be available in the future

I have provided the main database as both csv and txt. .csv will be easier to work with for updating (if possible).

Calculations can be discussed at a later stage. Currently all of the data in the database has no equations. Only the additional tools in the platform may be added calculations. No need for calculus or partial differentials, but may need the ability to use matrices. For now, we will just add in the “polydispersivity” calculation (see later). I have included an xlsx file of an example calculation for this. Example of Polydispersivity tool given in polydis.\_calc.xlsx file. Currently I have only included the Polydispersivity tool in the design framework. I have provided the spreadsheet where the calculation of the S value is made based on 2 sets of input data.

I need to think a little more on the other calculation tools, as it might be a little more complicated, but it is not essential to the framework. It is an addition tool as a bonus.

If using Django, the web platform will likely need a URL domain and github repository. Feel free to set up the Github (if you are familiar). URL suggestions would be https://www.pepper.database.com/ (or something therefore). Let me know what the best hosting would be for a django site. (I am happy to purchase a domain name, so the site is ad-free).

At the end of the project, a handover meeting where I go through the Python code and terminal will be very useful so there is a full handover of the data platform, and also transfer of access to the files and Github (if used).

A downloadable empty data template where users can input data that gets sent to the platform hosts to review and then update the main database - I have built this into the proposed platform, design slides. We can discuss what way is best, should users download this as a csv/xlsx file they have to fill in and then re-upload, or can it all be done within the web platform itself?

The uploaded/submitted new data files would come to me/Mathieu in some way so that we can verify the data and then manually add the data back into the main database that feeds the website itself. We will do any data quality checks, but let us think about error messages e.g. some cells will be number values only.

**Provided full database as of 30/05/23. 6160 data entries and 24 data categories**

* A: Publication – the scientific publication from which the data is found
* B: Volcano – name of the volcano the sample was collected from
* C: Eruption – the year, age, unit/phase of the eruption sample comes from
* D: Data DOI – Full DOI of the publication
* E: Chemistry – Broad classification of the whole rock chemistry (TAS classification)
* F: Bulk SiO2 (wt %) – the mass weight % of rock that is silica (SiO2)
* G: Bulk Na2O+K2O (wt%) – the mass weight % of rock that is alkali (Na2O+K2O)
* H: Glass SiO2 (wt %) – the mass weight % of glass that is silica (SiO2)
* I: Glass Na2O+K2O (wt%) – the mass weight % of glass that is alkali (Na2O+K2O)
* J: Chemistry DOI – Full DOI of the publication with chemistry
* K: Rock/experiment type – Brief one/two word description of the rock texture or style of experiment
* L: Subaerial/submarine – was the sample from an eruption on land or in the water
* M: Eff/exp – was the eruption the sample was from explosive or “effusive” 🡨 just lava flows
* N: Sample no. – the name of the number as quoted directly from the paper
* O: Bulk porosity (%) – The % of the whole sample that is void space/bubbles
* P: Connected porosity (%) – the proportion of bulk porosity that is connected. Measured by pycnometry.
* Q: Connectivity – =P/O (value between 0 and 1) – already calculated.
* R: Permeability k1 – the Darcian permeability of the rock samples, measured by a permeameter
* S: Permeability k2 – the inertial permeability of the rock samples, measured by a permeameter
* T: Vesicle number density – how many volcanic gas bubbles are trapped in a given rock volume
* U: S (polydispersivity) – a measure of the breadth/skewness of the bubble size distribution (value of 0 to 1)
* V: Total crystallinity (%) – total volume % of rock that is crystals (all sizes)
* W: Phenocrystallinity (%) – total volume % of rock that is larger crystals (phenocrysts)
* X: Microcrystallinity (%) – total volume % of rock that is smaller crystals (microlites)

Specific questions from you:

1. Could you provide more details about the visual appearance and functionalities of the data hosting platform? If possible, it would be helpful to see examples of similar platforms to better grasp the concept.
   1. Given within the previous email and also in the “possible design slides”. Let me know if you want any more examples for similar sites.
2. I would like to understand the specific features and functionalities that should be supported in our platform. Based on what I know so far, there is a download option and the ability to categorize data based on certain criteria. Could you provide more specific requirements? For example, how many different categorizations should we consider? Are there any additional features we should incorporate?
   1. The options I have given in the possible slide design, outline what data categories can be selected, and what data categories plots can be made from. Let me know if you want more clarification on this.
3. I also learned that we need to incorporate several specific calculations as functions, which will overlay as models on the graphs. I'm curious about the implementation of this feature and how we should interact with these functions in the platform. Specifically, what operations should be performed using these functions, and what results are we expecting?
   1. We will focus on the polydispersity tool for now (in the slide sign). Use the spreadsheet as the way to calculate S for the two datasets you would add in. I will figure out the set of equations for the remaining calculations. Other calculations tool need to be discussed with my project partner.
4. I noticed the need for a downloadable empty data template, where users can input data and submit it for review and database updates. Could you provide some guidance on the format and structure of this input? It would be helpful to have a template as a reference to understand the required functionalities for user input and the expected format of the data to ensure data accuracy.
   1. One of the tabs in FULL\_database provides an empty template for this type of data entry (with an example). Ideally the user will download this template, fill in the data and then submitted back into the system. This would then be emailed to me/Mathieu to review and then manually add into the main database. Unless you can add in a data entry function into the platform itself to prevent the need to download an fill in the template externally.

Breakdown for Polydispersity (S) calculations:

* User inputs dataset A[1:n] and B[1:n].
* Calculate ΣB[1:n] = K
* R1 = A[n]\*B[n] R2 = (A[n])^2\*B[n] R3 = (A[n])^3\*B[n]
* ΣR1 = Σ(A[1:n]\*B[1:n]) ΣR1 = Σ((A[1:n]^2)\*B[1:n]) ΣR3 = Σ((A[1:n]^3)\*B[1:n])
* X = ΣR1/K Y = ΣR2/K Z = ΣR3/K
* **S** value = (X\*Y)/Z