1. Open-source, cloud or billed-tools? Any constraints on this?
2. Download button in the interface to download what? Is it to download the initial pdfs or a report with the matching, etc.?
3. All input documents in PDFs format?
4. Top n of best matching document? How to order the output documents, how many?
5. In the search box, is it a tag or a random keyword that would be given? Or is it like in your website with filters based on hierarchy.
6. How do you currently classify documents based on a certain tag? What makes a document more relevant for a specific tag (business wise, as used today)?
7. We apparently have to generate the test-sets ourselves. How will you compare the different models based on different test sets?
8. How many observations for the test sets ?

Dans le règlement, ils mentionnent aussi qu’on sera dédommagé pour les deux demi-journées, comment ça fonctionne ?