Chapter 1. How it started

First of all, you should know initial purpose of the project.

1. To help Harvard to analyze data.
2. To help KTU students with their career: give experience in development, be an author of scientific papers.

Let’s start with first one. Scientists from Harvard are researching causes of eyes diseases, genetic causes. They are geneticist, and the problem is that a lot of tools for analysis exist only as API or as GitHub repository. They don’t know how to use it and they need help of IT specialists. Of course, we have to understand what they are searching for to help them.

> So, I suggest you to watch 1.5 hour lecture provided by Harvard. Ask current project manager to get access to it.

I will make just a note in terms describing goal.

1. Human has DNA.
2. Sometimes mutations happen in DNA.
3. Those mutations cause diseases.
4. We have to find which mutation caused disease.

Scientists across the world were taking DNA from people and analyzing mutations. Information about this mutations is saved in databases such like LOVD, Clinvar and GnomAd. Additionally, there were tools developed to find what mutation is good or bad: CADD, SpliceAI, REVEL. From IT prospective goal is

1. Download data from database;
2. Use tools to get score if mutation good or bad;
3. Give to Harvard and be happy.

In reality things are a bit more complicated, but generally it is more than correct. Details I will describe in next chapters.

> You will have to communicate with Harvard a lot, for this I created small glossary explaining basic stuff. Find it here Glossary or in pages of Confluence.

Human DNA has different genes. For analysis Harvard chose gene EYS. It was proved that eyes diseases are most likely to be caused with mutation in this gene. Understanding that Harvard is not the only institution that struggles with analysis of data shifted our focus to more general solution. So, we decided to implement solution for any gene.

> Unfortunately, at that time we implemented such option only for LOVD.