Oluwafolajinmi Olugbodi

NHGRI/ NIH HuBMAP Presentation

Kim’s Collection – A Meta-Data Collection App

July 29, 2021

Good afternoon everyone, thank you for coming. Today, I will be discussing a novel redesign to Metadata Collection.

To begin, we first have to understand, what is metadata, why is it important to collect data, and how is this important to genomics or HuBMAP as a whole.

Let’s start with collection, data provenance. Data provenance is where a piece of data comes from, historical ledger of data. Why is this relevant to genomics? Genomics data is collected after a complex set of protocols all of which can affect the quality and authenticity of the data collected. Metadata describes, explains, and makes it easier to manage information and informative resources.

What kinds of metadata are we collecting?:

RNA sequence Samples 🡪 like the **sampleID**, **harvestBy**- describes researcher who collected the sample from the source material be it culture dish, tissue; **Species**- constitutes the species from which the source material is derived, **Strain**- is the strain of source animal

So in the culmination of data, there are problems that arise

Inputting data can be rather labor-intensive, especially because it happens after the experiments are already done, making the task of recalling prerecorded data time consuming. Researchers sometimes dread this obligation as it can feel an inconvenience to the basis of their work. We, the Kim Lab seek to remove this feeling by proposing a solution where experimenters, lab technicians and the like can capture metadata conveniently while performing experiments

Which leads me to my objective – capturing metadata in order to establish data provenance for HuBMAP before, during, and after a procedure using a user-friendly and complicated input interface

In analyzing the previous method of data encapsulation, I saw that in capturing a Sample of data, that is what we are essentially trying to do, collect attributes of Samples. In collecting Samples, many of the sample attributes or fields were analogous. The similar Samples could be categorized by groupings of similar properties and features. Groups then could share common features, and they too could be categorized into Batches. This creates a hierarchical scheme shown on-screen

A Batch could have a given set of attributes common to all groups, all samples of a group then could share similar fields---allowing a top-down approach to meta-data collection which would make inputting data much simpler.

To further explain, I have captured some images of my hierarchical approach entry method.

Imagine only having to log onto your device during a lab procedure of 100 samples, creating groups of samples and batches of groups. In the batches, you add a field attribute that instantaneously assign the attribute of 100 samples. Or in groups, you add a field that assigns 45 of the 100 samples. This significantly reduces the task of individually filtering through and entering in the data manually. It also reduces the risk of inputting errors.

As you see in the addition Tab View, we have 3 banks, the Batch Bank, the Group Bank, and the Sample Bank. One is able to create a Batch, then for the given batch, as many groups as they desire, and for each grouping, a given number of samples.

To continue, the Data Entry Tab View. Each level on the approach will have it’s on level of attributes which will affect the scope of the given level. Meaning, each level will have a Species that can be edited within the selected vision parameters

The same for the Strain. These attributes also have built in logic, as shown with the Strain picker, depending on the selection, different options are presented.

Now, to achieve these views and to implement this logic, I did have to learn to use SwiftUI and Xcode which is a programming language and a simulation environment respectively that allow you to create apps, making the Meta-data collection mobile. In addition, I also became familiar with iOS Core Data, an object Graph and Persistence Framework….essentially it allowed me to persist and save the data created upon entry locally to the device.

With knowledge of Core Data, and the help of God and the Kim Lab, and family, I was able to make a functioning design that I plan to persist in the future.

I would also like to acknowledge NHGRI for its sponsorship, the NIH, Dr. Kim, Dr. Fisher and Dr. Nordgren, as well as the Kim Lab as a whole for their suppose and guidance!