# Overview

## Our Client

Our client for this project is Mark Pajak, who is Head of Digital Development for Bristol Museums Galleries and Archives. Mark and his team manage technology online and on-site and his interests involve digitising museum collections. Bristol Museums Galleries and Archives do important work to conserve the artifacts and history of not only Bristol but the world. Along with biological specimens, the museum has texts regarding wildlife observations. Bridging the gap between physical and technological curation is especially relevant today so our client has requested a system that aids in the transcription of handwritten notebooks. It's important that museums don't lose important information stored in an analogue form.

## The Application Domain

The main sample of notebooks we aim to help transcribe are biological recordings of butterflies, which contain important data on species-some of which are now extinct. If this data can be easily digitised, it can be aggregated online for use by scientists/historians. Mark hopes that the data extracted from these notebooks could be added to the National Biodiversity Network Atlas. Bristol Museum has transcribers to work on this project of digitising the information and our system aims to act as a unified tool for them. There is physical access to the notebooks but the museum can photograph them, allowing transcribers to work remotely. Currently, to transcribe notebooks, they simply read through the handwriting and type up a transcription in a text editor.

## The Problem

Transcribing handwritten notes is a task with difficulty that is subject to varying aspects of the handwriting and notetaking. Currently, this takes a long time as a transcriber must learn the handwriting of each individual collector and manually transcribe the text into a digital format. Making this worse, there is not one single system that the transcribers can use for creating, storing, managing and exporting transcriptions. The most useful piece of software available would most likely be a standard text editor coupled with a cloud storage system, but this is less than elegant considering that there isn't support to be able to view an image while transcribing it. If the handwriting is uniform and therefore not challenging to read, there isn't a system to automatically digitise it and our extension will be to attempt to speed up transcription this way.

## Our Vision

We will create a piece of software to speed up the process of transcribing notebooks in Bristol Museum. The system will be incorporated into an open-source web application written mainly in Java and using the Spring Boot framework. In the backend, the website will be hosted on an Amazon Web Services instance. The reason we have opted to implement our application as a web-app is to allow for the museum transcribers to freely access the tool anywhere and to avoid installation troubles. Our main implementation goal is to create a tool that allows the user to select an image of a notebook page on one half of the screen and transcribe the text into an area on the other half. Along with this we will create a backend system that allows for organised storage of materials and transcriptions. Finally, we are aiming to incorporate machine learning to automatically transcribe the text in an attempt to reduce the manual work for transcribers.