Adam Woodburry
Joe VanSickle
Nate Gillard
CS-481A2 Ruiz

Research Biologist Project Report

Introduction

We selected our users from a population of research biologists who do laboratory work to accomplish their individual research goals. This user group includes a plant pathologist studying a virus that affects onions, a genomicist studying the role of DNA in evolution, and a genetic engineer developing genetic circuits for use in biofuel plants. Each user has a different set of tasks to complete and processes to follow to reach their research goals. However, they all perform one or more DNA, RNA, or PCR extraction technique. They tend to all follow a high level sequence which consists of: planning research and experiments, followed by data generation, followed by analysis, which is repeated until they are ready to consolidate data, and finally write their research papers.

We were able to identify some common recurring breakdowns in their work - specifically the ways in which they record and query data related to individual experiments, both during and after the experiment process. Each user has developed their own methods and tools for planning and analyzing their research tasks. Some of our users rely strictly on a paper notebook, which is portable and time efficient for planning and tracking progress, but leads to breakdowns when consulting records during analysis. Other users utilize a computerized notebook (Word document) to work around problems with searching records, but sacrifice mobility and a desirable level of recording ease.

We may consider developing something which preserve the efficiency of planning in a paper notebook, but improves the current breakdowns associated with searching in them during analysis work. The new system must address searching breakdowns without imposing a large time investment for the benefits during analysis.

User Description

While our users have different research goals, there are commonalities within the group. Each user has a master’s degree or Ph.D in a branch of biology, and conducts their research based on genetic data in addition to teaching undergraduate/graduate courses at the university. All of our users perform some bio-chemical extraction to obtain the genetic data (DNA or RNA) they will be working with. Our users are also all from a similar age group, late twenties to mid thirties, and so are assumed to have a similar level of proficiency in using computers.

Each user has a different research goal. User One is a plant pathologist studying the effects of iris yellow spot virus in onions and the role that pests play in the spread of the virus. User Two is an evolutionary genomicist who sequences, analyses, and compares genomes in order to answer questions about DNA’s role in evolutionary change. Our final user, User Three, is a genetic engineer who is developing a genetic on/off switch for controlling the expression of selected genetic traits, specifically to find alternative bio-fuel sources.

Task Description

User One

Planning

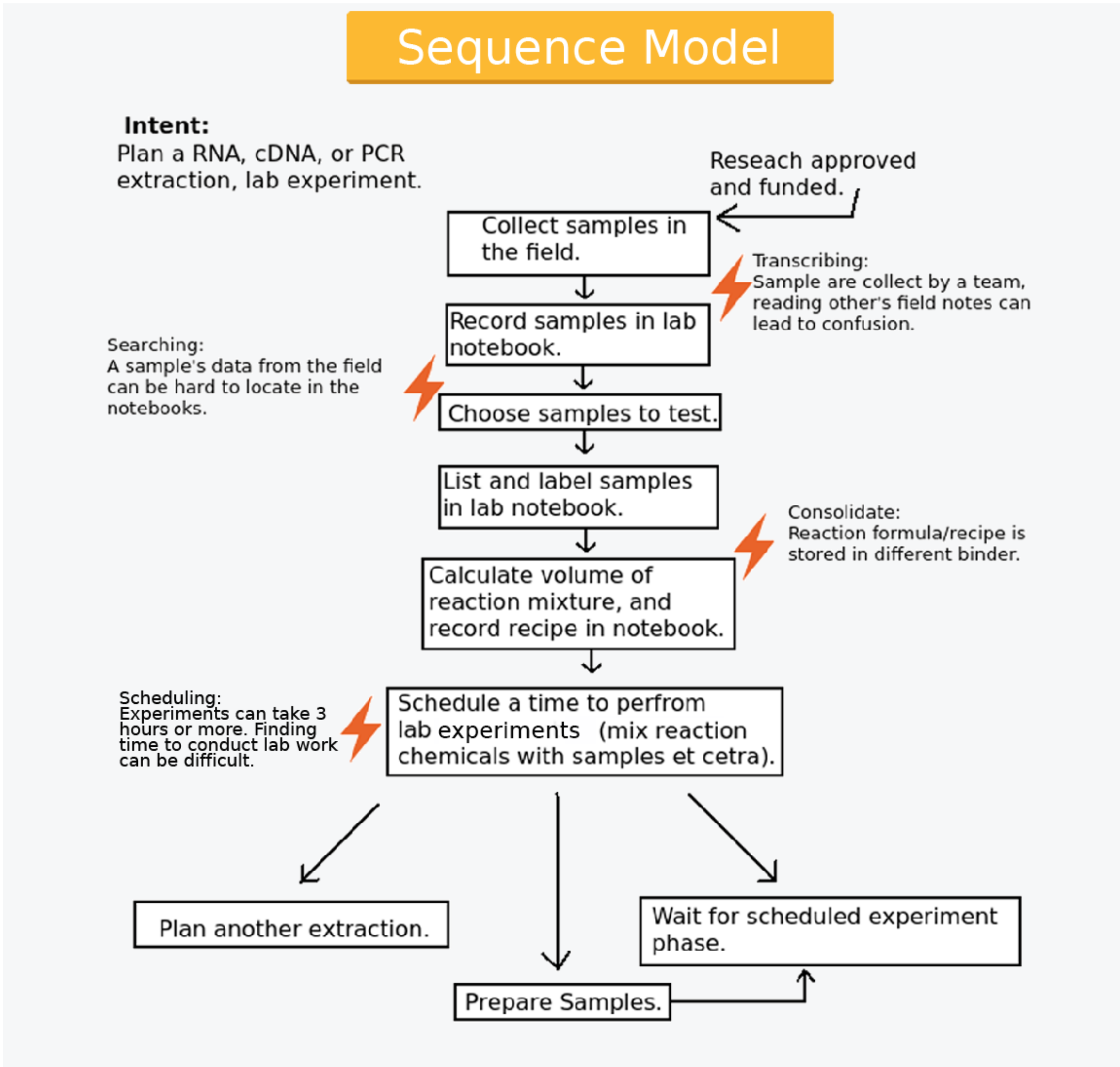
User One starts planning lab experiments by collecting samples of onion plants and pests (bugs) with a group of researchers in the field. Samples are shared between members of the field group and User One records identifiers and initial data for the samples in their lab notebook for referencing later. When User One is ready to plan an individual extraction experiment, they set up a page in the notebook to be used as a reference during the lab procedure and later as a complete record of the experiment. This typically involves identifying samples to be tested and locating them in the lab, looking up the reaction recipe in a separate binder, calculating the reaction formula volume for each ingredient, and recording the final recipe (see artifact 1). User One then plans a time when they will be available to perform the laboratory procedure.

Data Generation

Data generation involves first piping samples into test tubes with the reaction formula and then allowing the reaction to take place, often a computerized temperature control machine is programmed and samples are held in this machine until ready. Some samples are saved and stored for other extractions and depending on the experiment in progress the extracted materials are placed in an electrophoresis gel tray for more processing. If a electrophoresis was performed an image of the result is taken with special equipment, the image is saved on a computer and also printed out to be pasted into the experiment record in the user’s notebook.

Analysis

The user does some analysis to interpret the results and records the interpretation in the notebook. Later when the user is ready to begin consolidating experimental data the notebook record will be referenced. The user may also transcribe a set of experiment results into a spreadsheet.


Sequence Model
Intent:
Plan a RNA, cDNA, or PCR extraction, lab experiment.
Reseach approved and funded.
Collect samples in the field.
Transcribing:
Sample are collect by a team, reading other's field notes can lead to confusion.
Record samples in lab notebook.
Searching:
A sample's data from the field can be hard to locate in the notebooks.
Choose samples to test.
List and label samples in lab notebook.
Consolidate:
Reaction formula/recipe is stored in different binder.
Calculate volume of reaction mixture, and record recipe in notebook.
Scheduling:
Experiments can take 3 hours or more. Finding time to conduct lab work can be difficult.
Schedule a time to perfrom lab experiments (mix reaction chemicals with samples et cetra).
Plan another extraction.
Wait for scheduled experiment phase.
Prepare Samples.

Artifact Model

Intent:
Plan a RNA, cDNA, or PCR extraction, and create a guide for tracking progress through laboratory procedure.

Sample:
Identifies a specimen sample to be used in an extraction.

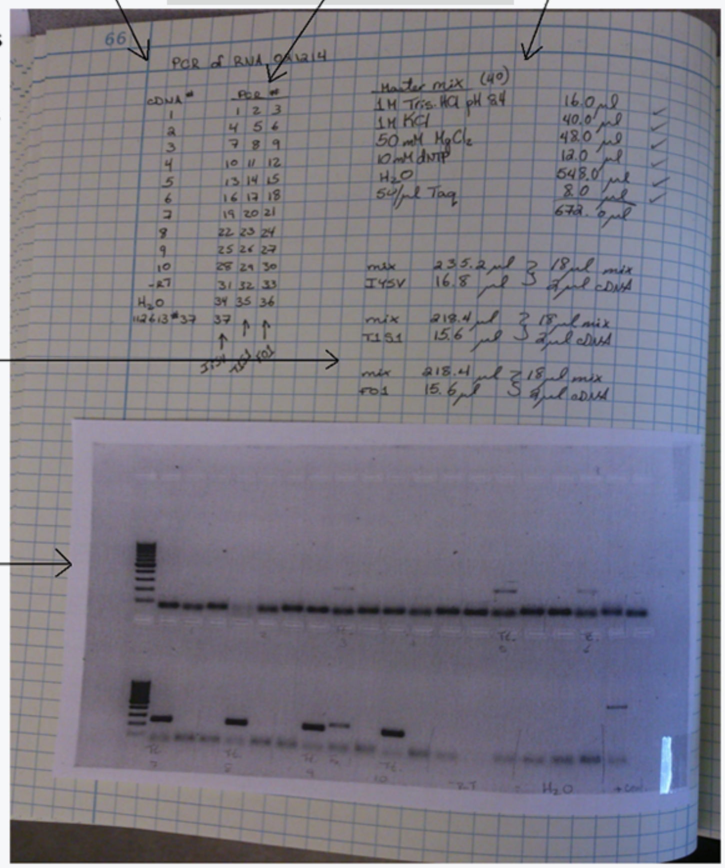
Test Tubes:
Identifies a test tube that will contain a sample and reaction mixture. Numbers listed match numbers on test tube in the lab.

Reaction Recipe:
Recipe for reaction mixture used to create bio-chemical reaction for extraction.

This is an example of Lab biologist's notebook and how it is used to plan an extraction. While components vary between extraction techniques our user used this general layout.

Directions and Notes:
Notes about how reaction mixture and sample are to be added to test tubes

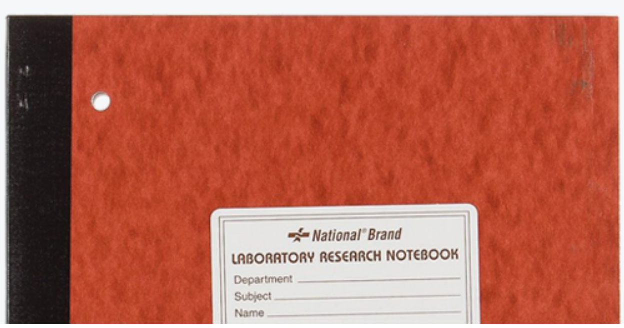
Results:
Image of results from lab work. In this case a gel-tray showing genetic information of samples.



Breakdown:
Recipes are stored in another binder.

-The notebook may be kept for the researcher's entire career.

-These experiment planning and layout forums live in a notebook.



Breakdown:
Searching for results from an experiment is not efficient. The lab notebook is often, not indexed.

- The notebooks are still needed. Results (and other notes) will be referred to when the researcher is ready to publish their findings

- These notebooks, once filled, are stored. Sometimes in a cluttered drawer.

Breakdown:
Since the notebooks are not indexed the researcher my need to search multiple notebooks relying only on his/her memory to find a record.

User Two

Planning

User Two starts planning by collecting relevant samples. Some samples are grown in the user's own greenhouse which the user maintains and often keeps notes on. The user records notes on paper in the greenhouse and later adds them to his electronic notebook (Word document). Samples used may come from other sources depending on the user's current research project, and the species to be studied. Once relevant samples have been selected, User Two continues his planning work by preparing tissue samples for DNA extraction and then uses an electronic lab notebook to record the samples used, calculated chemical formula, and steps taken. A DNA extraction technique is used to expose the DNA, and occasionally further biochemical reactions are used to prepare for sequencing.

Data Generation

Data generation is done by sending DNA samples to an offsite laboratory for sequencing. Sequencing involves breaking a long string of genetic data (a string of A,T,G,C) in to many small fragments and using sophisticated algorithms to compare fragments. The off site laboratory returns the sequenced data by postal delivery of a computer hard drive. Traditional postage delivery methods are preferred for this transfer, because the generated data set is too large to transfer efficiently over computer networks. Now the user can begin his analysis, which starts by assembling the sequences returned by the off site laboratory in to a genome and then annotating the genome to identify biological features of the organism studied. Annotation is done by comparison to published data that other biologist have generated and verified biochemically to identify sequences which code of genes or expressive traits. User Two uses computer software packages for assembly and annotation of a genome (Mitofiy and others). These programs run on powerful computer hardware and computation time depends on the size of the input genome. This user keeps a second electronic notebook which is used to record notes for the assembly and annotation software. This notebook contains reminders on how to run the programs, scripts for executing series of commands, and small programs.

Analysis

When the user has enough assembled and annotated genomes they will be ready to consolidate data and begin their research papers. User Two uses the data to answer question about the size of genomes, mutation rates, and other factors which may have an impact on evolutionary change.

User Three

Planning

User Three also starts by growing plant samples (sorghum) in a greenhouse and grow chambers (large temperature and light cycle controlled cabinets). This user takes some notes related to status of the plants, like when to water and the development stage of the plant, but for the most part does not record these in a notebook. Some notes are left on the grow chamber and some notes are written on individual plant tags (labels or tags stuck in the potting soil). When User Three is ready to plan a laboratory experiment they start by consulting databases and software tools which graphically organizes genetic data. Relevant genetic data is first looked up in an online database, then imported into software tools which annotate the genetic data and displays it graphically to make it easier for the researcher to identify critical sequences in the genetic information. The researcher uses these tools to find sequences which can be replaced by an engineered sequence for controlling gene expression. Once the user has located the desired genetic information he makes notes in his paper lab notebook identifying the genes and sequences to experiment on. The gene/sequence names are typically unpronounceable and difficult to remember, so our user refers to these notes often throughout research process.

Data Generation

In the data generation phase our user grows molds or bacteria to create the desired genetic sequence and uses a ligase enzyme to combine them with seeds or genetic information from plants grown in the greenhouses. A prototype plant is then grown from the genetically engineered seed and later used as a sample for DNA extraction. Another phase of planning is done with the same computer databases and software, this time to identify the desired configuration of the newly created gene/sequence. User three record the desired configuration in his notebook and plans a DNA extraction to test by electrophoresis whether the genetic sequence has been configured correctly and implanted into the plant.

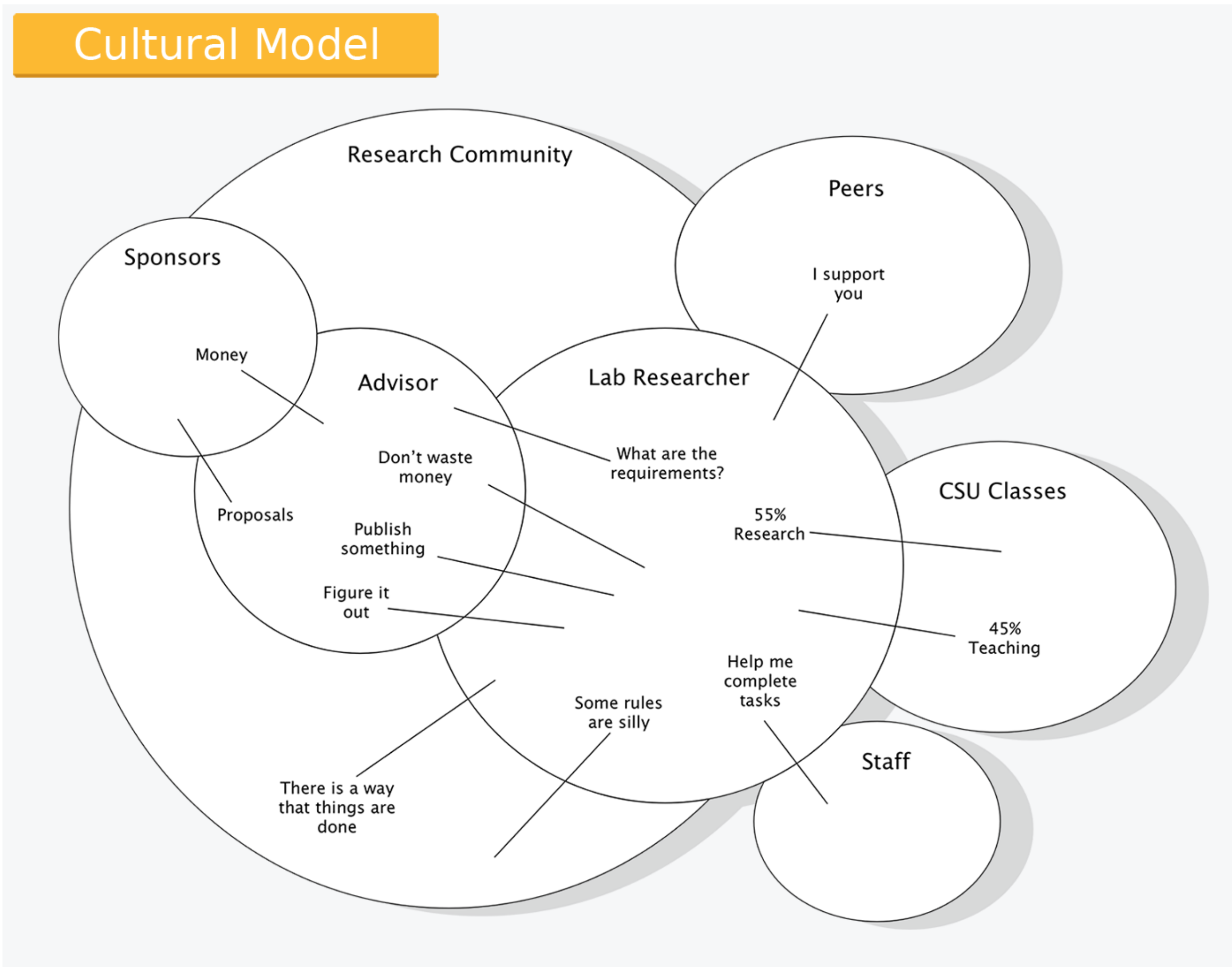
Analysis

User Three pastes an image of the electrophoresis in the record for an experiment. The user will collect and test several samples for his research and reference the paper lab notebook when it is time to write research papers.

Constraints

The most important constraint any new tool would have to honor is the scientific method itself. Scientific processes must be performed a certain way and those process cannot be modified by us. In this situation we rule out any tools or technology related to the Data Generation phase of our user's work. We are then left with the Planning and Analysis phases of work in which to focus our attention. The most apparent pressures on our users are Money, Time, and Form Factor.

Money is the most obvious of the three and it represents a rather large proportion of the total constraints. We would need to effectively demonstrate that the benefits of the new technology would justify the cost. If, for instance, our new tool was really great, but the same amount of money could be spent hiring a low cost intern who could do the same work, then our tool would not provide any value. Lets look at the notebook artifact as an example: any new tool or process would need to be more efficient than the notebook at performing the same task - enough so that the any added costs associated with the new tool would need to be recovered through more efficient use of our user's time. Form factor is a more tangible constraint in this situation, and although the pressure of this constraint may seem lessened, it is no less important to the overall goal of providing new tools. Using the notebook artifact again, we can see that our user's need to have the ability to perform the same tasks that the notebook currently does both in the field and in the laboratory.



Strengths and Limitations of Existing Practices

The Notebook

If we take a closer look at the notebook artifact we can begin to identify the significant breakdowns present in the Planning and Analysis phases of our user's work. It does have some strengths but it begins to fail very early in the first stages of the Planning process:

Strengths

1. The notebook is a very effective tool for quickly recording results of experiments.The notebook is a very effective tool for quickly recording results of experiments.
2. A user can carry it with them throughout the experimentation process and with only a brief glance, read and recall information relevant to the current experiment.
3. It is small enough to be carried with the user if they need to relocate across a room or even to another laboratory. User's can carry a notebook when doing fieldwork like

collecting samples.

Weaknesses

1. Information from past experiments that will be carried over to the current experiment must be copied by hand, requiring the user to flip through multiple pages while transcribing the information to the latest experiment's page. One user indicated that this process could influence the experiment itself - a recipe for a mixture might be copied completely along with the volume of each ingredient, meaning that the researcher would tend to use the same amount of mixture across multiple experiments rather than recalculating those values for a different amount.
2. Since the notebook is not indexed in a way that is accessible outside of itself, it can be a tedious process to find specific data or results. When asked during an interview to give an example of a certain positive result, the researcher had to spend multiple minutes flipping through notebooks in order to produce a single page with the relevant information. Not only is this process time consuming for the researcher who created these notebooks, it would take an extreme amount of time for anyone other than the researcher to find the desired information.
3. The information contained in the notebook cannot be relayed over a network without a tedious process of copying the data into a computing system - something which could take inordinate amounts of time, especially when combined with the previous breakdown in which the researcher first has to find the relevant data. This same breakdown can be observed when a user might want to make calculations using the data in the notebook.

Computerized System

In contrast to the notebook, a computerized system provides a powerful set of strengths and a new set of weaknesses.

Strengths

1. A researcher may initially spend more time learning an interface and it may take more time to input certain types of data but that time could be recovered through the ease with which data sets can be copied or even iterated upon during later reuse.
2. A database would allow quick access to specific information; searchable by date, result type, ingredients used in a mixture, sample type, etc. This process would be much faster than looking through notebooks, not only for the researcher but for anyone else as well.
3. It would be accessible over a network, which could alleviate the previous weakness that a computer cannot be carried easily through multiple locations (assuming we aren't already thinking of a computer in a tablet or laptop form factor).

Weaknesses

1. Entering data into a computer may not be as quick as writing it in a notebook because computing systems require data to be organized in a very specific way which give the user less freedom to deviate from the data entry process (A properly designed tool would need to allow the user to efficiently input many different types of data in differing formats).
2. The form factor of a computing system must also be taken into account - a powerful

desktop machine can not be taken out to do field work for example, nor carried from room to room.

3. Hard drives need to have redundant backups in case of failure, although we think this is hardly a weakness compared to a notebook, which could suffer the same kind of failure (fire or water damage for instance) and is much harder to “backup.”
4. Using a network to transfer data relies on the capacity of that specific network. For instance, User 3 has purchased dedicated hardware for storing his work which has a fast connection to national databases for sharing work with colleagues, but mentions that this dedicated hardware does not help in moving data between the user’s personal computers, since the data is too large to transmit along the institution’s internal network.

Next Steps

Many aspects of biologist’s work practices cannot be subjected to change. One area that does stand out however, is the planning phase of their work. All users that we interviewed use a notebook in planning, and then use it later for analysis as a reference. While a notebook is easy to use for recording, getting data out of a notebook may not be. One researcher uses theirs in the field and then, upon returning, re-records the data in a computer. This exemplifies the convenience of the notebook - it’s draw. Another user has 5+ notebooks (each with 156 pages) filled with notes, each takes one year to fill. It should seem clear that these breakdowns should be addressed in some form or another.