

### **1) Protocol Title**

Play With Data for College-Bound High School Girls in West Philadelphia Promise Zone.

### **2) IRB Review History**

Not Applicable.

### **3) Objectives**

The purpose of this research project is to introduce participants to the open data movement and some basics of data driven science and investigate how this project impacts participants' approach to data science and career options.

Specific goals and objectives are as follows:

1. Cultivate interest in data driven STEM, and an understanding of the new, evolving scientific paradigm.
2. Develop a "Let data talk" methodology, educating participants, by having them understand data in their vicinity.
3. Contribute foundational research on strategies for data-related education for college-bound high school girls, particularly girls for underrepresented STEM population.

### **4) Background**

#### Introduction

The proposed CAMP Open Data program will provide college-bound high school girls with an opportunity to become data savvy citizens. The program will be led Metadata Research Center, College of Computing and Informatics at Drexel University in collaboration with the School of Education. The CAMP idea was inspired, in part, by Greenberg and Lin's CVDI project, and brings this into the RET environment.

CAMP Open Data will specifically focus on the (1.) open data movement, and (2.) basics of data- driven science. The majority of the work will be prepared by the RET participant, in collaboration with project Principle Investigator, Jane Greenberg (IS/CCI/Drexel),co PIs, Xia Lin (IS/CCI), and Penny Hammrich (School of Ed./Drexel). The camp program will be take place on Tuesdays and Thursdays from 5:00pm-7:00pm from January 10<sup>th</sup> -26<sup>th</sup>, 2017. The CAMP program will be hosted at the Dana and David Dornsife Center for Neighborhood, with support from Drexel University.

#### Rationale

Protocol Version Date: November 8, 2016

Over the last decade both federal and international funding agencies (e.g., National Science Foundation, National Institutes of Health, and UNESCO<sup>1</sup>) have mandated data sharing policies, thus contributing to and motivating the open data movement. Additionally, digital technologies have forced a paradigm shift in science. This shift is often explained as the *'Fourth Paradigm'*<sup>2</sup>, describing the change from empirical and theoretical based science to computational and data driven science. These changes are unprecedented, resulting in new requirements for anyone pursuing a STEM field. Students advancing in a STEM discipline, or simply seeking to become data savvy citizens and members of the work force, must have (1.) a comfort with digital data in their selected domain, and, (2.) at the very least, basic knowledge of computational and data driven approaches. These requirements make sense, and can impact successful pursuit of a STEM discipline or areas of business, health, and the social sciences. Despite knowledge of the importance of these requirements, there is clear indication that the necessary skills are not being sufficiently taught to college-bound high school girls, particularly girls in underrepresented STEM populations.

CAMP Open data will pilot an approach to address notes challenges through an informal learning opportunity. The program will recruit college-bound girls from the Philadelphia's promise zone, and introduce participants to the open data movement and some basics of data driven science.

**Specific goals and objectives are as follows:**

1. Does the camp open data experience inspire interest in STEM fields and potential careers?
2. Does the camp's informal learning approach increase interest in working with data?
3. Did the camp open data experience inspire independent learning with data.

## 5) Inclusion and Exclusion Criteria

Individuals taking part in this study are female secondary school students in West Philadelphia Promise Zone who are planning to attend college in the future.

- Adults unable to consent*
- Individuals who are not yet adults (infants, children, teenagers)*
- Pregnant women*
- Prisoners*
- Not Applicable*

## 6) Study Timelines

1. The duration of an individual subject's participation in this study is 2 hours a day for 6 days between January 10 and January 26.
2. Focus groups will occur at the end of the 6 days, and individual subjects in the focus group will participate in focus group discussions for approximately 30 to 40 minutes.
3. The duration anticipated to enroll all study subjects is approximately 3 weeks. This estimate excludes the recruitment process.
4. The estimated date to complete primary analysis is in February 2017.

## 7) Study Endpoints

The goal of this study is to describe the participants' experiences working with big data and if this experience changes their career goals. Participants will participate in surveys, focus groups, and journaling. These data points will be analyzed to see if the participants' experiences have any affect on their career goals.

## 8) Procedures or Methods Involved

The measurement instruments in this study include surveys, observations, participants' journal entries, and focus groups (see attached).

The methodology is a mixed methods where participants experiences working with big data changes will be examined to see how working with big data changes participants interest and career aspirations.

Responses will be analyzed by content analysis measuring patterns and trends on how students define the nature of science. Analysis will be conducted independently by both researchers and compared for inter-rater reliability. Pre-intervention response analysis findings will be compared to post-intervention response analysis findings to determine short-term impact (change over time).

## 9) Data Banking

Only the researchers will have access to the data. Data will be retained in a locked office for 7 years after the youngest participant turns 18. The data will also be stored in an encrypted password protected computer.

## 10) Data Management

Collected data include surveys, observation and focus groups records, and participants' journal entries. Individual subjects' identifiers will immediately be stripped after collection. Data will be stored on a researcher's' password-protected computer. Only the researcher will have

Protocol Version Date: November 8, 2016

access to the data. It will be the sole person responsible for transporting data from the interview site. As per Drexel University policy, all data will be retained according to the policy for three years after the study is closed for adults and seven years for the participants under the age of 18.

We do not consider the data stored to be of a sensitive nature. Participant names will be recorded as part of the informed consent process which will be given via the consent and assent forms. All data for this study will be collected confidentially. All data will be assigned to an ID number that is not associated with a student or any other identifying information for a student. The consent and assent forms will be kept in a separate locked file drawer from the data collected in the study. Because data from this study is collected confidentially, participant names will not appear in any report of the findings. All data will be securely stored in a locked file drawer for paper and on secured, encrypted servers for electronic data. Paper based information will be stored in the principal investigator's office and be destroyed five years after the study is completed.

The research team plans to disseminate the results from the project at related conferences as well as through related journals.

### **11) Provisions to Monitor the Data to Ensure the Safety of Subjects**

Only the research team will have access to the data obtained from the study participants during the time of the study. Data will be stored on secure, encrypted servers and will be accessible only by the research team. All subject's identifiers will be immediately removed after data collection, and students will be assigned a number and pseudonym.

### **12) Withdrawal of Subjects**

If any participants decide not to take part in the research or wish to stop at any time for any reason, it will not be held against them. The participant is expected to inform the researchers as soon as possible if they wish to withdraw or discontinue, and are made aware of this.

If any individual has decided to withdraw at any given time, the data gained up to that point might be included in the result.

### **13) Risks to Subjects**

There is a minimal risk to subjects in this study. However, participants may experience some challenges or discomfort when being introduced to unfamiliar softwares and applications and for the first time. The researchers will do all that they can to minimize participants' challenges or discomfort by giving them instructions step by step.

#### **14) Potential Benefits to Subjects**

We will be providing participating students with data-related education including introduction to Microsoft Excel and applications like tableau. Students will also be given instruction on Arduinos. This technology instruction as well as hands on activities will enhance participants' knowledge, competence, and understanding of data science. Participants will also gain knowledge of their vicinity because the data they will utilize is obtained through the community.

#### **15) Vulnerable Populations**

This research includes teenagers. These individuals will be protected from coercion through a parental consent form. Instructions, activities, and focus groups will be held at Dornsife Center for Neighborhood Partnerships, located within the teenagers' neighborhood community. The teenagers are familiar with the site as they have been recruited from the site. All individual subjects will be informed of their right to decline or discontinue the study at any time and for any reason.

#### **16) Multi-Site Research**

N/A. This research study is a single site study that will occur at the Dornsife Center in cooperation with Dornsife Center staff.

#### **17) Community-Based Participatory Research**

N/A. This is not a community-based participatory project.

#### **18) Sharing of Results with Subjects**

The investigators are willing to share research results with subjects/participants. Upon participants' request, investigators will email an electronic copy of the products (research articles) of the research project to participants.

#### **19) Setting**

The site that will be used for this project is the Dornsife Center for Neighborhood Partnerships. The research team will identify and recruit subjects from the Dornsife Center via flyers and meetings with families at Dornsife Center Events (regularly held Dornsife Center dinners). All research procedures will occur at the Dornsife Center, and the Dornsife center has agreed to take part in this research study.

#### **20) Resources Available**

- We will recruit anywhere from 15 to 20 students from the Dornsife Center. This number of participants should be sufficient to provide rich data to understand students learning through and with big data. This

Protocol Version Date: November 8, 2016

number will also allow us to match students with the amount of technology available at the Dornsife Center.

- PI Jane Greenberg in coordination with Co-PI Penny Hammerich will lead the design of the study, data collection, and analyses.
- Rooms within the Dornsife Center will be reserved to provide a space for the “Camp Open Data” (or Play with Data) to occur. These rooms will be reserved ahead of time, and all observations, camp activities, and focus groups will occur within the Dornsife Center.
- Medical and /or psychological resources are not required for this research.
- Each participant will be informed that he or she is participating in research intended to learn more about big data and how college bound High School girls respond and learn through big data.. They will also be informed that their participation will be limited to the duration of the camp and any subsequent focus group sessions.

### **21)Prior Approvals**

We have obtained prior permissions from the Research Offices of the School of Education, and the College of Computing & Informatics to proceed with the study described.

### **22)Recruitment Methods**

Students for this study will be recruited via two methods, the first being flyers that will be hung at the Dornsife Center. These flyers will include information about the When, Where, What, and Who of the study/camp. We will obtain permission from the Dornsife Center to hang and distribute flyers.

The second recruitment method will be in person talks about the program. The Dornsife Center holds frequent dinners for the community, and at several of these dinners the research team will explain the program and let members of the community know that we are actively recruiting participants for this camp/study. We will coordinate with the Dornsife Center to obtain permission to give these talks at the dinners.

### **23)Number of Subjects**

**We will recruit a maximum of 20 students from the Dornsife Center.** This number of participants should be sufficient to provide rich data to understand students learning through and with big data. This number will also allow us to match students with the amount of technology available at the Dornsife Center.

### **24)Confidentiality**

All data for this study will be collected confidentially. All data will be assigned to an ID number that is not associated with a student or any other identifying

Protocol Version Date: November 8, 2016

information for a student. All students will be assigned a pseudonym for the sake of confidentiality. All consent forms will be kept in a separate locked file drawer from the data collected in the study. Because data from this study is collected confidentially, students' names will not appear in any report of the findings. All data will be securely stored in a locked file drawer for paper and on secured servers for electronic data. Paper based information will be stored in the primary investigator's office, and be destroyed **seven years after the youngest participant turns 18** is completed.

**All electronic data from this study will be stored on investigators' secured encrypted password protected computers during the course of the research.**

According to Drexel University's policy, all forms of data from both studies will be retained in password-protected zip files on the investigator' secured computer for three years after the study closure. After seven years, the zip files will be deleted permanently.

## **25)Provisions to Protect the Privacy Interests of Subjects**

Participants' names will be recorded as part of the informed consent process which will be obtained via consent and assent forms. All data for this study will be collected confidentially. All data will be assigned to an ID number that is not associated with a student or any other identifying information for a student. The consent forms will be kept in a separate locked file drawer from the data collected in the study. Because data from this study is collected confidentially, students' names will not appear in any report of the findings. All data will be securely stored in a locked file drawer for paper and on secured servers for electronic data. Paper based information will be stored in the primary investigators office, and be destroyed three years after the study is completed.

All electronic data from this study will be stored on **investigators' secured encrypted password protected computers** during the course of the research.

According to Drexel University's policy, all forms of data from both studies will be retained in password-protected zip files on the investigator' **secured computer for seven years after the youngest participant turns 18.** After three years, the zip files will be deleted permanently.

## **26)Compensation for Research-Related Injury**

There will be no compensation for research-related injury. However, research-related injury is extremely unlikely to happen since the research will involve no more than minimal risk to subjects.

## **27)Economic Burden to Subjects**

There is no cost for participation for the participants.

## **28)Consent Process**

Participants who express an interest in participating in the program will be given the consent and assent forms and asked to return them to the Dornsife Center.

A written consent as well as assent will be acquired from each subject before subjects partake in the Play with Data Camp, and it will follow “SOP: Informed Consent Process for Research (HRP-090)”, as well as the “TEMPLATE OF ASSENT FORM”. The Dornsife Center will be provided consent and assent forms prior to the camp/study to interested participants, and these forms will also be given out during information sessions about the camp. Students will be required to hand in consent/assent forms prior to the start of camp, and all students under the age of 18 will be required to have a parent/guardian signature on the consent form.

### ***Subjects who are not yet adults (infants, children, teenagers)***

For subjects who are not yet adults (under the applicable law of the jurisdiction in which the research will be conducted), the subjects will need to provide a consent form signed by a parent or legal guardian. A consent script that is based on “TEMPLATE CONSENT DOCUMENT (HRP-502 Social/Behavioral/Educational)” will be used to obtain consent. An assent script based of the “TEMPLATE OF ASSENT FORM” will also be used to obtain assent from subjects who are not yet adults under the applicable law of the jurisdiction in which the research will be conducted). This process will take place prior to the start of camp.

### ***Non-English Speaking Subjects***

Does Not Apply.

### ***Cognitively Impaired Adults***

Does Not Apply.

### ***Adults Unable to Consent***

Does Not Apply.

## **29)Process to Document Consent in Writing**

A consent script that is based on “TEMPLATE CONSENT DOCUMENT (HRP-502 Social/Behavioral/Educational)” will be used to obtain consent. An assent script based of the “TEMPLATE OF ASSENT FORM” will also be used to obtain assent. A sample of the consent and assent forms



PROTOCOL TITLE: PLAYING WITH DATA

9

Protocol Version Date: November 8, 2016

will be provided with this application. The Dornsife Center staff will collect the consent and assent forms and the participants will have three weeks to decide.