

## **Data Collection White Paper EpiNu Project**

**11th December, 2023**

### **Background and Context**

Recent reports indicate that 40% of humanitarian funding is now allocated to climate-related issues, reflecting a global shift in attention towards climate change. This growing concern is particularly relevant for regions with limited resources, such as the Democratic Republic of Congo (DRC). Here, climate change-induced floods are posing a critical challenge, significantly affecting the nutrition of communities, especially pregnant women.

These floods disrupt crucial nutritional intake, increasing the risk of complications for pregnant women and compromising fetal development. Maternal exposure to environmental factors, combined with insufficient nutrition, poses lasting threats to offspring health, highlighting the urgent need to enhance birth anthropometry. Research in central South Africa reveals suboptimal dietary quality among pregnant women, indicating a lack of essential nutrients.

Addressing this nutritional deficit is crucial not only to reduce the risk of developmental delays but also due to broader global implications. The impact extends beyond regional boundaries, affecting maternal health and birth outcomes in low to middle-income countries. The magnitude of this issue is significant, with poor birth outcomes and maternal complications adding substantially to the healthcare burden in the DRC.

### **Project description**

To understand the daily nutritional intake in communities, the EpiNu current program asks women to capture images of their food using a shared smartphone – as an implementation of the use of ‘photovoice’ methodology. The objective of bringing women together now is to create a community space for these discussions and to build trust.

Currently, CEAFPD implements the guideline for image capture, only approximately 50% of the pictures taken by women are typically analyzable by the machine learning algorithm. The current manual filtering process by the NGO is not scalable. Hereby, we need to find a solution that will increase the success rate or percentage of pictures used.

As the program further progresses, it is envisaged that local NGOs may use the information gathered to guide how they may make small modifications to their nutritional intake to benefit their health and the health of their expected children. To augment guidance, the program is exploring the use of the food images to train an ‘edge’ machine learning algorithm informed by a ‘nutrition security database’ developed in tandem.

## **Low resource setting priorities (performance vs speed tradeoffs)**

There is a contrast difference of technology implementation priorities in low resource settings compared to developed countries. In low resource settings, human centered design should be the main focus, as a result, complex systems may not be favorable. For example, in terms of data collection, women with low digital literacy should be introduced to new phone technologies starting from handling until taking pictures with them. Hereby, there is a need for technology that will assist humans to achieve certain goals.

In this project, data collection are being done by women in the communities with the help of phones. A specific output of the picture should meet the requirement to be used as a training data in the machine learning algorithm. Terms such as angle, lighting quality, and position in the frame should be within selected range.

## **Solution Brainstorming**

To give a comprehensive solution that may help women to acquire pictures specific to the requirement, we should acquire some alternatives in collecting data so that more percentage of pictures are analyzed and reduce human resources to filter the pictures, which will be beneficial in terms of scaling the project.

### **1. Phone application**

Using a phone application presents an efficient and user-friendly solution for enhancing data collection. With a focus on empowering women, the application can guide users through the entire process, from handling phones to capturing pictures. The app ensures that the images meet specific requirements by providing real-time feedback on factors such as angle, lighting quality, and position within the frame. Its familiarity as a device and the potential for widespread use make it a high-impact, scalable option. Integration with existing systems may require some adjustments, but the overall ease of use and potential benefits make it a strong candidate.

### **2. Additional lighting**

Incorporating additional lighting aims to improve the quality of images captured during data collection. This solution involves the use of external light sources to enhance visibility and reduce shadows, especially in low-light environments. While relatively simple to set up, it may require adjustments to ensure optimal lighting conditions. However, its impact is moderate, and the scalability of this solution is limited to scenarios where improved lighting is crucial. Integration with existing systems may pose challenges, and the ease of use may vary depending on the context.

### **3. Tripod**

A tripod provides stability for capturing images during data collection. Its primary benefit lies in reducing motion blur and ensuring clear, focused photographs. Setting up a tripod is relatively simple, but it requires additional equipment. While offering stability, a tripod may limit mobility during data collection. Its impact is moderate, and scalability is restricted to scenarios where stability is a critical factor. Integrating a tripod into existing

mobile data collection processes may be challenging, and ease of use depends on the users' familiarity with such equipment.

#### 4. Photo box

A photo box creates a controlled environment for capturing images, minimizing external factors that could affect the quality of photographs. This solution involves setting up a confined space with consistent lighting conditions. The impact is high, particularly in scenarios where environmental factors need to be tightly controlled. However, its scalability is limited to specific use cases where a controlled environment is essential. Setting up and using a photo box may be less familiar to users, potentially impacting ease of use. Integration into existing data workflows may require additional steps and adjustments.

To analyze the best solution, we create a decision matrix that will help to determine the best technology. The terms that we consider in the the decision matrix are:

1. Impact: Assesses the potential positive influence on the data collection process.
2. Ease to Develop: Evaluates the complexity of developing and implementing the technology.
3. Ease to Use: Considers the simplicity and user-friendliness for women with low digital literacy.
4. Scalability: Examines the potential for the technology to be used on a larger scale.
5. Integration: Reflects the ease of integrating the technology with existing systems or processes.

Technology	Impact	Ease to Develop	Ease to Use	Scalability	Integration
Application	High - Empowers women	Moderate - Requires app development skills	High - Familiar for most users	High - Can be used by a large number of women	Moderate - May need integration with existing systems
Additional Lighting	Moderate - Improves image quality	Low - Simple to set up	Moderate - Requires adjustment	Moderate - Limited to specific scenarios	Low - May not integrate seamlessly with existing processes
Tripod	Moderate - Aids stability	Low - Simple hardware	Moderate - Requires setup	Low - Limited mobility	Low - May not integrate with mobile data collection
Photo box	High - Controlled environment	Moderate - Requires setup	Moderate - May be unfamiliar	Low - Limited to specific use cases	Low - May require additional steps in data

					workflow
--	--	--	--	--	----------

## Conclusion:

The phone application appears to be a strong candidate due to its high impact, ease of use, and scalability. Additional lighting, tripod, and photo box may have specific use cases but may lack in ease of use or scalability. The decision should consider the specific context and requirements of the data collection project.

## End users

Using an application may seem to give a personal recommendation and specific data sharing to each member of the community. However, to avoid the negative impact of personalized recommendation and privacy sharing, we suggest a recommendation based on the community.

## Features

### 1. Data generation assistance

The application provides support throughout the data generation process, offering assistance to users, particularly women, in capturing food pictures. This includes guidance on handling phones, taking pictures, and ensuring the collected data meet specific requirements.

### 2. Potential roles contributing to data cycle

#### 1. Image Labeling

The application allows users to label images based on the owner or the women who captured the pictures. This functionality enhances data governance by providing a means of categorization and identification, ensuring a more organized and meaningful dataset.

#### 2. Geographic Location

Utilizing the application enhances geolocation tagging, providing accurate information about the geographical location where each picture is taken. This feature contributes to a more comprehensive understanding of the spatial distribution of data, aiding in spatial analysis and contextualization.

#### 3. Specific Timing and Identification of Climate

The application includes features for capturing specific timing and identifying climate conditions when images are taken. This adds temporal context to the dataset, enabling users to correlate data with specific time frames and weather conditions. It enhances the richness of the dataset for various analyses.

### 3. In - built data governance features

This feature will help realize the data governance schema that we propose in the data governance white paper.

### 4. Future Potential of Recommendation

This data shouldn't just stop in recommendation of nutrition, furthermore, we could give recommendations on plants that may be suitable for a certain season to increase nutrient availability.

## **Role of NGOs**

We divide the NGO role based on the location and the participation role. Internal NGO are considered as those that have direct responsibility to train the women in using the application and assist data collection. External NGO are those using the application for data analysis and seeking funding for additional help.

1. Internal NGO

The role of an internal NGO is to introduce the use of this application to the women. Besides, they also have access to the data collected. For example, in this context CEAFPD is an internal NGO.

2. External NGO

The role of an external NGO is to use the data based on the agreement as described in the data governance white paper. External NGO may use the application to find data and use it within the agreement.

## **Prototyping**

To create a visualization of the suggested data collection method, we provide a prototype. To access the prototype, please open the link below:

1. Video of Prototype

Link: <https://www.youtube.com/watch?v=PdpMynISx0o>

2. Application Prototype

Link: <https://bit.ly/epinuptype>

## **Business Model**

We also consider several business models that will work to find funds for the application. We explore the following potential customers in different business canvases in this assignment. They are:

1. Government (Table 1)
2. Non-Government Organizations (Table 2)
3. Research Institutions (Table 3)

For each customer, we have tried to develop distinct value propositions, key partners, revenue models, and customer relationship strategies. For the most part, costs and key resources remain consistent across different customers.