Development Engineering 6 (2021) 100062

Contents lists available at ScienceDirect

Development Engineering

journal homepage: www.elsevier.com/locate/deveng

Design of remote data collection devices for social impact indicators of products in developing countries☆

Bryan J. Stringham, Christopher A. Mattson *

Department of Mechanical Engineering, Brigham Young University, Provo, UT, 84602, USA

ARTICLE INFO

Keywords:
Engineering for global development
Design for the developing world
Social impact
Sensor systems
Internet of things
Development canvas

ABSTRACT

Social impact indicators provide one effective way to measure the social impacts of products in developing countries and ensure that engineering design is producing positive impacts on individuals. Remote data collection devices enable the use of sensors to collect user data required to calculate social impact indicators remotely, continuously, and potentially less expensively than other methods that require direct interaction with users. However, many key decisions and questions must be considered during the development and use of such devices to avoid risk of failure. To provide a systematic way for researchers and engineers to consider critical device development questions, the parts of device development and use can be decomposed into Data Identification, Device Design, and Device Deployment. This paper discusses the key decisions within each part of development along with critical questions, common options, and considerations that should be addressed during each part of device development, thus increasing the likelihood of success. A sensor development canvas outlining the key decisions is also provided as a design tool to easily identify deficiencies in the device during development. Considering these critical questions while developing and deploying data collection devices can help researchers and engineers successfully collect social impact indicator data that may be used to ensure engineered products are producing desired positive impacts on individuals.

1. Introduction

Engineered products have great potential to improve quality of life for individuals that comprise the *Base of the Pyramid* (BOP) – the approximately 4 billion people in the world that live on less than $8 per day (United Nations Development Programme Growing Inclusive Markets Initiative, 2008; Wood and Mattson, 2016; Jagtap, 2019). However, not all engineering design efforts result in improvement of the lives of the BOP and can unintentionally harm individuals instead (Wood and Mattson, 2016).

Measuring or otherwise assessing the social impacts of an engineered product, or the effects that product has on the daily quality of life of an individual, is essential to verifying that engineering design is producing positive effects and avoiding negative ones (Burdge, 2004; Mattson and Wood, 2014; George and Shams, 2007). Many different approaches exist for evaluating the social impacts of a product. Encouragingly, many of these approaches acknowledge the need for data to evaluate the social impact instead of assuming impacts (United Way of America, 1996; Kellogg, 2006; Clark and Anderson, 2004; Stevenson et al., 2018; Stevenson et al., 2020; Hutchins et al., 2009).

Social impact indicators, which combine user data in a meaningful way to indicate the social impact of a product, are one useful way to quantify and track the social impact of a product over time (Stevenson et al., 2020). These social impact indicators can be identified through considering the eleven social impact categories identified by Rainock et al., including health and safety, education, paid work, conflict and crime, family, gender, human rights, stratification, population change, social networks and communication, and cultural identity and heritage, and then identifying measurable metrics relative to the impact categories of interest (Stevenson et al., 2020; Rainock et al., 2018). These social impact indicators also have potential to be used alone in some cases or combine with other indicators to represent high-level objectives, such as the United Nations' Sustainable Development Goals (United Nations Statistics Division, 2020; Johnson et al., 2021). This paper does not attempt to map sensor data all the way to Sustainable Development Goals. Instead, it focuses only on helping designers map

☆ An earlier version of this paper was published in conference proceedings by ASME. ASME has granted permission for the authors to use any part of the original version in the revised paper published here.

* Corresponding author.
E-mail addresses: bryan.stringham@byu.edu (B.J. Stringham), mattson@byu.edu (C.A. Mattson).

https://doi.org/10.1016/j.deveng.2021.100062
Received 29 January 2021; Received in revised form 7 May 2021; Accepted 10 May 2021
Available online 25 May 2021

sensor data to social impact indicators.

Historically, the user data required to calculate social impact indicators of products in developing countries has primarily been obtained through methods that require direct interaction with or observation of users (Wood and Mattson, 2019; He et al., 2014). This data, referred to as *direct data* in this paper, is typically rich in information but can usually only be collected at relatively low frequency intervals because of the high amount of human facilitation to obtain and because researchers and engineers are often geographically removed from the developing country of interest (Stringham et al., 2020).

Another potential way to obtain data used to calculate and monitor social impact indicators related to the use of products over time is through using electronic data collection devices. These devices are installed onto products or used by individuals and utilize sensors to translate physical phenomena into meaningful user data without requiring manual collection via human interaction with users, which data is referred to as *indirect data* in this paper. One advantage of these devices is their ability to collect data that would not be plausible to continuously collect through manual data collection. This continuously collected data can either be stored onboard the device for manual retrieval after a period of time or transmitted wirelessly as in the case of *Internet of Things* (IoT) sensor devices (Pessôa and Becker, 2020). When this data is transmitted wirelessly, these devices also enable autonomous and remote monitoring of social impact indicator and other data. In some cases, this sensor data can be used to calculate social impact indicators directly. In other cases, a simultaneously collected set of direct data (collected manually) and indirect data (collected via the sensor) can constitute a *training dataset* used to create a correlation model that predicts direct data given the sensor data, thus enabling the continuous and often remote prediction of rich direct data at the often lower cost of sensor data (Stringham et al., 2020).

The use of remote data collection devices also has the potential to reduce the cost of data collection by providing lower data collection costs than manual collection or increasing the quality of the data obtained (Stringham et al., 2020). These devices can also reduce costs by providing data that improves a product, service, or intervention delivery and thereby enable a proactive and preventative response to a crisis as opposed to a more costly, reactive one. For example, a widespread network of these devices monitoring water hand pumps in East Africa can ensure that water hand pumps are functional in case of a drought and prevent a much more costly emergency water delivery service (Thomas and Brown, 2021). However, the per unit deployed cost of these devices must incorporate the amortized engineering and development cost along with bill of materials and deployment costs for an accurate cost comparison.

Additionally, the costs of developing and using these electronic devices has reduced, and development of these devices has become more accessible in recent years. Individuals with little experience developing electronic sensor devices can now develop them more easily and inexpensively than ever before. Much of this lowering cost and easier accessibility has been aided by the rapid growth of IoT devices as reflected by the growing number of IoT devices in the world from 15 billion in 2015 to an estimated 75 billion in 2025 (IHS, 2015). Also, the growing ubiquity of cellular networks and Wi-Fi connectivity in developing countries has made potential applicability of these devices to social impact and other data collection in developing countries more viable.

A small but growing number of researchers have already used sensor devices to monitor the usage or social impact of products in developing countries. Existing products that have previously been remotely monitored using sensors include water filters (Thomas et al., 2013a), improved cookstoves (Thomas et al., 2013a; Wilson et al., 2015; Ventrella et al., 2020), latrines (Andres et al., 2018; Sinha et al., 2016), solar panels (Collings and Munyehirwe, 2016), and water hand pumps (Andres et al., 2018; Nagel et al., 2015).

Ottosson et al. lists many additional existing products for which the social impact could be monitored (Ottosson et al., 2019). These devices also have high potential for assisting in the improvement of quality of life when the products monitored are critical to health and of which there is a historically high rate of failure as in the case of water hand pumps (Ottosson et al., 2021). Furthermore, these data collection devices also have potential to provide data useful for improving water, sanitation, and energy service delivery applications as these have special design considerations (Sharpe et al., 2019). In addition to providing research value on a small scale, these types of remote data collection devices also have significant potential to provide product usage and operational insights for wide-scale monitoring as well as facilitate pay-for-performance contracting as shown by Thomas et al. (2020). However, due to the financial constraints typically present in the developing world, there are only limited applications in which wide scale deployment of these sensor devices is feasible. Thorough financial analysis is of utmost importance to ensure feasibility in developing world applications that require high volume manufacturing and deployment.

When developing these sensor devices for use in the developing world, there are unique challenges that must be overcome to ensure that reliable, accurate, and affordable devices can be deployed for collecting social impact or other data. The “Principles for Digital Development” framework outlines nine principles that should be reviewed and considered by any practitioner desiring to improve their likelihood of success in using digital tools to collect and use data in developing world applications (Wagaman, 2016). However, despite this and other current resources available for developing electronic and/or IoT sensor devices, there is little in the literature that discusses principles of how to effectively design these systems for use in measuring the social impacts of products in developing countries. Thomas et al. propose the use and benefits of one specific commercial hardware platform for remotely collecting data (Thomas et al., 2013b). Kipf et al. propose a platform for managing and using backend data from remote sensor systems (Kipf et al., 2015). Stringham et al. propose a framework that includes some considerations that are important to sensor device design for social impact measurement (Stringham et al., 2020). However, designing the mechanical, electrical, and many other aspects of sensor devices for monitoring the social impact and other usage data of products in developing countries has its own challenges unique from the design of other types of systems that are not addressed in the literature.

Principles of effective design of such systems will become increasingly applicable as more researchers, NGOs, businesses, and others seek to use them to measure the social impacts or collect critical and actionable data related to the use of products in developing countries. The purpose of this paper is to provide principles and guidelines that should be considered for successful sensor device development and use for the collection of user data for social impact indicator calculation and other applications that have potential to improve lives in developing countries.

These principles have been identified through the design, development, testing, manufacturing, deployment, and use of many remote data collection devices by the authors and other members of the Design Exploration Research Group at Brigham Young University. These devices include environmental sensor devices (Brazil and Utah), a two-part Bluetooth-connected classroom usage and environmental monitor device (Cambodia), human-powered water borehole drill monitoring device (Utah), water hand pump usage monitor and failure detection devices (Uganda), and other sensor devices. Nearly all of these devices connect to the internet via cellular or Wi-Fi networks. Altogether, more than 80 of these devices have been manufactured and deployed since 2017 and are primarily in the developmental part of pilot testing. Many other prototypes have been manufactured and tested during the development of these devices. While we acknowledge that many others have deployed greater quantities of devices, we believe the principles identified through our experience developing these devices for a wide variety of applications will be useful to others looking to do the same.

The remainder of this paper is organized as follows: Section 2 discusses the approach taken to identify and organize the key decisions and critical questions that should be considered when developing a remote data collection device; Section 3 provides an overview for the canvas that can be used to visualize and track device design progress; and Sections 4, 5, and 6 discuss the key decisions, questions, options, and considerations that should be examined during the Data Identification, Device Design, and Device Deployment parts of device development, respectively.

2. Device Development Overview

The successful development and use of devices for collecting data used to calculate social impact indicators can be difficult to achieve, especially for those without experience developing or using such devices. Many factors can cause the project to result in failure if not considered properly. However, by decomposing the process of device development and use into parts, it becomes clear what key decisions need to be made throughout the process.

One way to distinguish the parts of device development and use for these specific devices includes Part 1: Data Identification, Part 2: Device Design, and Part 3: Device Deployment. Key decisions that need to be made in the Data Identification part relate to the social impact indicators that will be calculated, the physical phenomena that represent the indicators and can be measured by sensors, and the data correlation that will be used to correlate sensor data with the user data required to calculate the indicators. Key decisions that must be made in the Device Design part relate to the device sensor data, data retrieval, device computer, power supply, device housing, and non-sensor inputs and outputs. Key decisions to be made in the Device Deployment part relate to data utilization, training data collection, testing, manufacturing, installation, maintenance and operations, ethics and regulations, and design and deployment strategy. As with any product development process, decisions in one part can affect another part, and therefore iteration is expected.

For each key decision in the process of development and use, there are a number of critical questions that need to be answered to help ensure the best decision is made for collecting and using social impact indicator data. Sections 4 through 6 include many questions that are critical to answer for each key decision in each part of device development and use. Included with the questions are common options and examples that could be possible answers to the questions as well as considerations and guidelines for answering that question. While these options, considerations, and guidelines are by no means exhaustive, acknowledging and answering them through the process will assist practitioners interested in collecting and using sensor device data. Extensive analysis of which option to pursue for a given question is not included as this will vary greatly by application and would be prohibitive to include here.

Accompanying the critical questions for each factor is a canvas, which can be used as a design tool throughout the development process. The Business Model Canvas (Osterwalder and Pigneur, 2010), E-Spot Canvas (Mehta and Mehta, 2011), and Design for Developing World Canvas (Wood and Mattson, 2016) provide the precedents upon which this canvas is based. A canvas was chosen instead of a simple list of questions because a list does not capture interactions between the key decisions. A canvas was also chosen instead of a flow chart because a flow chart implies a strict order in which the key decisions should be made. While the presented canvas will generally be filled out in a clockwise order starting from the top left (Data Identification to Device Design to Device Deployment), it could be filled out in any order that is most beneficial or seems most appropriate to the person or team using it.

A canvas provides a tool that can be used repeatedly during the development process as key decisions are made and the design progresses. During design reviews, the canvas can be useful to guide discussion and help identify what is currently known and unknown about the design. Relative to each box on the canvas, design teams can ask "Have we made a choice relative to each key decision?" If a decision has not been made, teams can ask "What are the requirements that drive the decision?" or "What is needed to be able to make a decision?" However, it is not critical that all of the answers to the questions within each box be answered in one session with the canvas, and it is recommended that sessions reviewing the canvas last less than 1 hour. It is typical that in the first session with the canvas, only part of the canvas will be answered using knowledge available initially and assignments will be made to various team members to further research the remaining parts. In subsequent meetings, the team will progressively and iteratively complete the canvas. Hence, the canvas includes a place to record revision, date, and current state of development. The ultimate goal of the canvas is to facilitate decision making and ensure that critical questions are answered and key decisions are made deliberately.

3. The Social Impact Sensor Canvas Overview

Fig. 1 shows a scaled down version of the canvas for illustration purposes. A more practically useful 11 in x 17 in format is available under the Resources tab and Design Resources option at gdi.byu.edu.

The heading of the canvas allows tracking of the date and revisions of the canvas and should be updated as new versions are created. The "Product" field can be used to write the name of the product for which the social impact will be measured. The "Current State" field can be used to describe the current state of development of the remote data collection device. Relative to the canvas and this paper, *device* refers to the data collection sensor device that is being developed while *product* refers to the product of which the social impact is being measured.

Each part (section) below discusses the key decisions (subsections) and questions (sub-subsections) that should be considered when developing a system for collecting data used to calculate social impact indicators. Each part and key decision includes a description below the heading for clarity.

4. Part: Data Identification

Data Identification is the part of the process in which the practitioner identifies what social impact indicators will best indicate the social impact of the product over time. The calculation of social impact indicators can be made possible through the collection of user data based on sensor data that correlates sensor output with physical phenomena related to how a product is used.

Often the social impact indicators will be the starting point and will already be known if this canvas is used in conjunction with another approach such as Stringham et al.'s framework for combining direct and indirect data for social impact measurement (Stringham et al., 2020). Alternatively, the physical phenomena that will be measured or is available, such as the movement of a water pump handle, may be used as the starting point from which possible social impact indicators could be identified.

4.1. Social Impact Indicators

The social impact indicators written in this box of the canvas during device development can come from the eleven social impact categories derived by Rainock et al. (2018). The process for determining which indicators should be collected is discussed extensively by Stevenson et al. (2020), but the process generally involves identifying the social impact categories that are relevant to the product and its application, identifying the data that could be collected, and selecting meaningful indicators that represent an outcome of interest. The following considerations will help lead to a careful decision of which indicators and data should be collected.

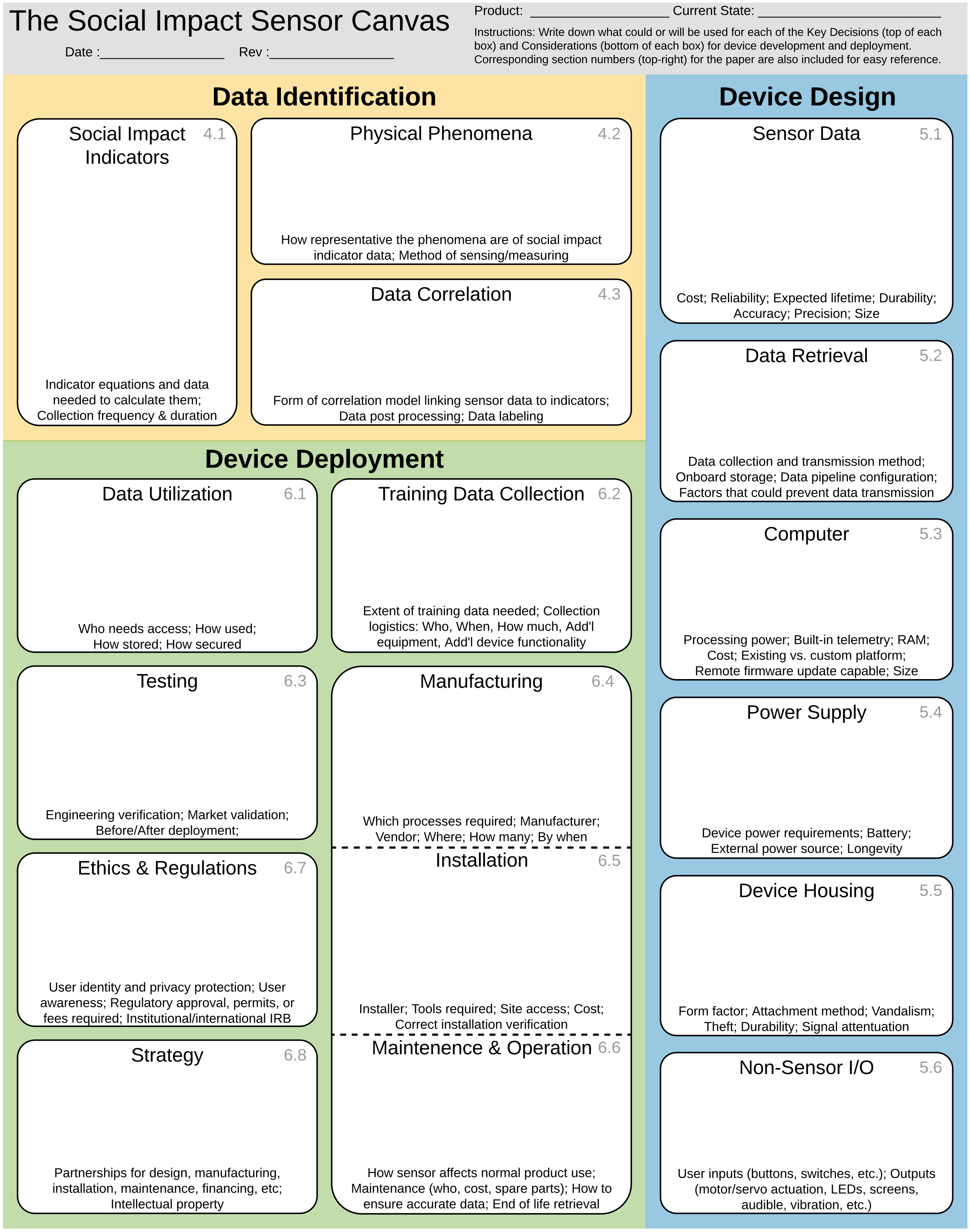

Fig. 1. The Social Impact Sensor Canvas for guiding the development of sensor systems used to measure the social impacts of products in developing countries.

4.1.1. What user data is needed for calculation of the social impact indicators?

4.1.1.1. Common options/examples. The user data required for calculation of the social impact indicators will be identified from determination of the social impact indicator equations. For example, data needed to calculate social impact indicators related to water hand pumps could include number of hours using a hand pump and number of strokes of a hand pump by user type (man, woman, or child) (Stringham et al., 2020).

4.1.1.2. Selection considerations. The user data used to calculate the social impact indicators provides the basis for the entire data collection process; as such, care should be exercised when selecting the data that will be collected, and the motivation for collecting each data source should be clear. For each relevant social impact category, one or more social impact indicators should be identified and selected. The indicators chosen and data collected may be refined over time.

A sufficient number of social impact indicators should be selected so as to meet the desired objectives for the sensor device application but not so many indicators that the data is prohibitively expensive to obtain. When possible, select a physical phenomenon along with a sensor data source (discussed in the following sections) that can be used to calculate multiple indicators to obtain the most data possible with the least complex device. For example, one previous study used the single sensor data source of water pump handle angle over time to calculate ten social impact indicators in five different social impact categories (Stringham et al., 2020).

4.1.2. How frequently does the data need to be obtained?

4.1.2.1. Common options/examples. Continuously to annually.

4.1.2.2. Selection considerations. The main consideration here is how frequently the data will be needed in order to be useful for the specific application. Some situations, such as when individual safety is involved, may require near real-time data collection in order to be able to inform time-critical decisions. Other applications such as for product or program evaluations may not require more than quarterly, semiannual, or annual collection of data stored on-board the device. The trade-off of increasing data collection frequency is often a higher cost of data collection due to increased power and data transmission needs or more frequent manual retrieval of data.

Generally, collect the least amount of data needed to calculate the social impact indicators or perform the needed functionality. If near real-time data is needed, transmit or obtain data daily when possible to have only a small delay between when data is collected and when the data can be used.

4.1.3. How long will the data need to be collected for?

4.1.3.1. Common options/examples. Several weeks to many years.

4.1.3.2. Selection considerations. One of the main advantages of using a data collection device to collect data is the lower cost of long term, high frequency data collection. If there is little need to identify trends in data over time, the cost of developing and deploying a long term data collection system may not be justified. The main trade-off of collecting data for an extended period of time is the cost of recurring data collection. This potentially includes data transmission and network connection fees in addition to data storage and/or the cost of paying personnel to collect or manage the data.

Collect data for a minimum of one to six months to make the effort of creating the setup and data collection process worthwhile. If the data can be collected over a few weeks or less, it may be more cost and time effective to instead collect the data manually.

4.2. Physical Phenomena

Physical phenomena refers to the physical phenomena that will be measured by the sensors to collect the user data used to calculate the social impact indicators. The selected phenomena should be written in this box of the canvas.

4.2.1. What physical phenomena are representative of the data needed?

4.2.1.1. Common options/examples. The physical phenomena could include any number of sensor-measurable physical phenomena that could be indicative of the user data needed to calculate the social impact indicators. Some examples include movement, acceleration, rotation, force, water flow, and temperature.

4.2.1.2. Selection considerations. In selecting the physical phenomena that will be measured to collect the data used to calculate the social impact indicators, effort should be made to avoid as many confounding factors as possible. Ideally, the physical phenomena will individually and completely capture the data of interest.

If the physical phenomenon that can be measured is identified before selection of social impact indicators, this could also dictate the selection of social impact indicators.

4.3. Data Correlation

Data correlation refers to the data correlation that will be used to translate sensor data into user data subsequently used to calculate the social impact indicators.

4.3.1. What type of correlation would most effectively capture the relationship between the sensor data and user data required for calculating social impact indicators?

4.3.1.1. Common options/examples. A wide variety of predictive modeling approaches could be used to correlate sensor data with the user data needed to calculate the social impact indicators. While new modeling approaches will inevitably be developed, some current approaches include linear regression (Ramsey and Schafer, 2012), logistic regression (Ramsey and Schafer, 2012), random forests (Breiman, 2001), decision trees (Quinlan, 1986), deep learning models (such as neural networks) (Goodfellow et al., 2016), and multivariate adaptive regression splines (Friedman, 1991).

4.3.1.2. Selection considerations. When deciding which correlation modeling approach to use, some factors to consider include the type of data that will be collected, the researchers individual experience with the chosen modeling approach, the amount of data that will be available or required to create the correlation, the mathematical attributes of the data being measured, and the performance or model accuracy required.

Begin by considering the most straightforward modeling approach available based on the experience of the researcher. If the performance of the most straightforward model is acceptable, there is no need to explore other approaches. However, if higher model accuracy is needed, other modeling approaches can be explored according to the time and resources available by the researchers to create such models.

4.3.2. What data post-processing and labeling will be needed to prepare data for creation of and use with the correlation model?

4.3.2.1. Common options/examples. Potential data preprocessing techniques that may be beneficial prior to using the sensor data to create a correlation model may include normalizing all data channels to a common scale (typically 1), shifting the data to have a mean of 0, data filtering and smoothing, and adjusting for sensor drift or hysteresis.

The approach that will be used to label the data from the sensor, or manually assign a class to the various classes of sensor data, should also be considered prior to data collection. Possible options include real-time class labeling through observation, recording video for later data labeling through observation, or surveying users for later data labeling.

4.3.2.2. Selection considerations. The data post-processing that is required is largely a function of the model in which the data will be used so the identification of what post-processing is required should be done in tandem with the creation of the correlation model. If class labeling is to be done through observation, video recording is typically preferred over real-time labeling to enable fewer labeling errors as long as the privacy of individuals is not compromised. In terms of resource management, it is beneficial to consider the amount of data labeling that will need to be completed to identify the amount of resources that will be required for data preparation prior to the actual data collection.

To minimize time and effort required to post-process the data, it is recommended to begin with the least amount of data post-processing possible. If the minimally processed data results in a sufficiently accurate model, no additional processing is necessary. If the minimally processed data results in an insufficiently accurate model or if the most accurate model possible is needed, continue to perform post-processing techniques.

5. Part: Device Design

The key decisions within the Device Design part are the subsystems that comprise a typical data collection device, which can be identified using subsystem decomposition.

5.1. Sensor Data

The sensor data key decision incorporates the sensor(s) that will be used to measure the physical phenomena along with any intermediary data processing or modeling to translate raw sensor data into a physically meaningful data source. In the example of a water handpump, this could entail both the selection of an inertial measurement unit (IMU) sensor along with data processing required to translate IMU sensor data into pump handle angle over time.

5.1.1. What sensor(s) should be used to measure the physical phenomena?

5.1.1.1. Common options/examples. Accelerometers, gyroscopes, temperature sensors, humidity sensors, hall effect sensors, ultrasonic sensors, strain gauges, force transducers, passive infrared sensors, optical sensors, GPS sensors, potentiometers, light sensors, and level sensors include several common sensor options available at the time of writing that could be used to measure the physical phenomena of interest.

5.1.1.2. Selection considerations. The cost, reliability, expected lifetime, durability, and repeatability are several factors that should be considered when selecting a sensor to measure the physical phenomena of interest. If a low quantity of devices are needed, the cost of the sensor may be a less critical consideration than when a large quantity of devices is needed.

The least expensive sensors should be chosen that still meet the reliability and performance requirements. If possible, choose sensors for which software libraries have been developed for whichever computer (typically a microcontroller) is being used to reduce required effort and potential bugs related to communicating with and obtaining readings from the sensor.

5.1.2. What modelling or data processing technique(s) will be used to translate raw sensor data into a physically meaningful data source?

5.1.2.1. Common options/examples. Due to the large number of possibilities here based on the various available techniques applicable to the hundreds or thousands of different types of sensors, the practitioner is advised to research the specific techniques for the sensor and application of interest.

5.1.2.2. Selection considerations. The translation of raw sensor data into a physically meaningful data source will often occur onboard the device's computer; therefore, it is critical to ensure that whatever technique is used is within the processing capabilities of the computer.

5.2. Data Retrieval

The data retrieval subsystem includes all necessary radios, telemetry, on-board storage, and other components required to transmit and store data in order to move the data from the device to the practitioner. The team's choices relative to this should be written in this part of the canvas.

5.2.1. What data collection method should be used?

5.2.1.1. Common options/examples. Data can be collected through either having individuals retrieve removable media containing stored data or through data transmission via one of the many radio technologies and transmission protocols. Several common data transmission technologies at the time of writing include satellite, cellular, LPWAN, Wi-Fi, Zigbee, Bluetooth, and near-field communication (NFC).

5.2.1.2. Selection considerations. When determining whether to store the data onboard the device and collect it manually or transmit the data through one of the radio technologies, it is important to consider the hardware cost, difficulty of data transmission development, recurring cost of transmission, power consumption, availability by area, range of transmission, reliability, and possibility and cost of manual data collection. If data transmission is to be used, it is often optimal to use the option that has the lowest cost subject to the power availability, transmission capability, and other functionality constraints. Regarding currently available technologies, cellular data transmission has quickly grown to be one of the most promising transmission methods due to the near ubiquity of cellular networks even in remote areas of developing countries and the current ease of development and relatively low cost of cellular-enabled development boards. Many companies provide cellular-based microcontrollers or shields, which make cellular device development much more simple and straightforward now than in the past. However, large amounts of data will be restrictive due to the higher cost of transmission. Wi-Fi is often the ideal method of data transmission for situations in which large amounts of data are required to be transmitted due to lower or absent restrictions on data amounts. Also, the hardware costs of Wi-Fi based microcontrollers are less expensive than cellular or satellite options. LPWAN (Low power wide area network) is desirable for situations with minimal cellular coverage but require remote or long distance transmission with low data rates. These require the use of a cellular or Wi-Fi gateway to provide connectivity to the internet. Bluetooth Low Energy, which has superseded the original Bluetooth, typically requires lower energy than any of the previous technologies but has much lower transmission distances and also requires a gateway to connect to the internet or must be collected by hired individuals on a periodic basis. Satellite data transmission should typically only be used when cellular, Wi-Fi, or other radio options are out of range due to the higher cost and fewer development resources available than other approaches.

If a large amount of data is needed for something such as video data or deep learning training data, it may be cost prohibitive to transmit this information and may require the use of on-board storage that is manually collected at periodic intervals with data either transferred electronically or physically shipped.

Where available, it is recommended to use Wi-Fi as the transmission method due to the relatively low cost of the hardware and ability to inexpensively transmit large amounts of data. However, Wi-Fi is often not available in remote, developing world settings. When Wi-Fi is unavailable, cellular will typically be the most promising option, especially since cellular networks are nearly ubiquitous even in remote areas of developing countries. There are also several companies that currently provide microcontroller development boards with full cellular service for data costs as low as $1–3 per month.

5.2.2. How should the data pipeline be configured?

5.2.2.1. Common options/examples. The data pipeline is the means by which the data will go from the deployed device to a useful, accessible form for the researcher and other necessary parties. Pre-built dashboard service providers offer application programming interfaces (APIs) that publish data directly to their servers and easily allow the setup of an online dashboard for visualizing and storing the data. Some currently available IoT dashboard providers include providers such as Adafruit IO, AskSensors, Ubidots, Thingspeak, and others. An example basic option involves a process as simple as using a data publish that triggers a Webhook that is captured by the web service IFTTT (If This Then That) and stores data in a Google Sheet. An additional option is building a fully custom data pipeline using proprietary web servers, databases, and a custom dashboard.

5.2.2.2. Selection considerations. The primary drivers for data pipeline decisions are cost (both upfront and recurring), functionality/customization, who needs to view the data, and reliability.

If the design team has little web development experience, a pre-built dashboard service provider is an ideal way to develop a proof of concept for data retrieval, storage, and visualization. These pre-built dashboard service providers may also be a viable long term solution if the quantity and time period of use is low. However, their lower upfront cost typically comes at the cost of a higher recurring cost. Another consideration is that 3rd party services may limit the frequency or amount of received data, only store it for a limited amount of time, or be less reliable in capturing and storing the published data than custom pipelines.

Custom data pipelines enable any desired functionality and customization but can have significant upfront development costs and require a longer development timeline. Thomas et al. provides one data pipeline framework for effectively collecting and processing data from developing world sensor applications (Thomas et al., 2013b).

For prototyping and initial use, it is usually sufficient to use free data pipeline and dashboard options such as those mentioned above. To avoid many debugging issues and minimize initial costs, it is recommended to use an existing dashboard provider. However, when scaling, it is recommended to pay the price to develop a custom dashboard and data pipeline to minimize recurring costs and provide the specific functionality needed for the specific application.

5.2.3. What factors could prevent effective data collection and transmission?

5.2.3.1. Common options/examples. Signal attenuation, unable to purchase data for transmission network connections (i.e. SIM card for cellular network), vandalism, battery energy depletion, and power disconnection.

5.2.3.2. Selection considerations. When cellular data transmission is chosen, use mobile virtual network operators (MVNOs) that have systems of automatic billing available instead of requiring manual "recharging" of cellular data to simplify paying for cellular data and avoid disruptions in cellular service. Most companies that provide cellular microcontroller boards also provide cellular service. Guidelines for the other considerations are included in their more relevant sections below.

5.2.4. How much data needs to be collected?

5.2.4.1. Common options/examples. Data requirements include few kilobytes per day for basic usage-based sensors to many kilobytes or megabytes per day for constant data collection using basic sensors to gigabytes per day for video.

5.2.4.2. Selection considerations. It is important to consider the amount of data required to identify a statistically or practically meaningful trend or result and weigh that against the cost.

When using the data in a deep learning application, the amount of data transmitted may be large. When using any type of video data, the amount of data will likely be large. For simple sensors, the amount of data may be very small.

If using Wi-Fi, transmission of large amounts of data is typically acceptable. However, if cellular is used, it is recommended to compress data before transmission and to transmit less than a few megabytes of data per month according to current cellular costs to minimize data costs. If video or large amounts of data are to be collected, it is likely to be more cost effective to perform manual data collection by partnering with someone in the country in which data will be collected.

5.3. Computer

The computer used to provide the critical functionality of controlling the overall device, sensors, data transmission, and data storage will typically be a microcontroller, although a single board computer, microcomputer, personal computer, or any other type of computer could be used. Microcontrollers will receive the most attention here since they typically best meet the size, processing, and functionality requirements for these applications.

5.3.1. What computer should be used?

5.3.1.1. Common options/examples. While the technology and preferable or available options will change over time, common current options at the time writing for bespoke device applications include cellular, LPWAN, Wi-Fi, or Bluetooth-based development microcontroller boards (including Arduino, ESP32-based boards, Particle products, Pycom products, Raspberry Pi products, Adafruit/SparkFun development boards). For large scale applications, the primary option is a custom microcontroller system.

5.3.1.2. Selection considerations. It is necessary to consider power availability, data transmission and storage needs, data processing needs (processing and RAM), programming language, built-in telemetry, and ability to remotely update firmware during installation or operation when selecting a microcontroller or other computer. Numerous data sheets, tutorials, and blogs outline the benefits of each type of computer or microcontroller. Custom microcontroller systems enable lower manufacturing cost at high volume but come at the cost of much higher development cost.

Initially, development boards should be used as opposed to custom boards to reduce potential issues caused by custom microcontroller development. Upon need to massively scale the project, it may become necessary to develop a custom microcontroller.

5.4. Power Supply

The power supply subsystem includes all necessary components and factors that affect how the device is powered.

5.4.1. What are the power requirements?

5.4.1.1. Common options/examples. Current and voltage required by computer, data transmission unit, sensors, and all other components and peripherals.

5.4.1.2. Selection considerations. The power requirements could set the power that needs to be obtained, or the power that is available could set the max allowable power and energy consumption by the device and thereby dictate the device and data transmission type and amount that is possible.

Use lower power data transmission methods and lower power sensors and microcontrollers when possible. If possible, use grid power combined with battery backup to ensure reliable transmission of data.

5.4.2. What are the battery needs?

5.4.2.1. Common options/examples. Rechargeable or not and battery chemistry type. Current common battery chemistries available at the time of writing include alkaline, lead acid, nickel metal hydride (NiMh), lithium polymer (Li–Po), lithium ion (Li-Ion), lithium iron phosphate (LiFePO4), lithium thionyl chloride (LiSOCl2).

5.4.2.2. Selection considerations. Energy storage capacity, rate of charge and discharge capacity, intermittency of battery recharging, functional temperature range, cost, cycles to failure, and environmental impact are all factors to be considered when selecting a battery.

In most developing world situations when data is critical, a backup battery is needed even for devices powered by grid power supply due to intermittency of grid power. The specified storage capacity of the battery should be based on how unreliable grid power is.

Another easily overlooked factor is the max allowable capacity and type restrictions for the expected transportation and shipping of the device. For example, if lithium batteries are to be used, air transportation and shipping regulations should be checked to ensure compliance and avoid confiscation during shipping and customs processing.

Regarding currently available batteries, Li–Po batteries provide a well performing, rechargeable, readily available, and energy dense option. The primary restriction in using them is the operating temperature range as most Li–Po batteries can only be used between 0 and 10 and 45 degrees C if the battery will be intermittently charged during use (Ascent International Group Co. Ltd, 2020; Particle Industries and Inc, 2020). For extreme temperatures of −60 to 85 degrees C, no promising rechargeable option currently exists, but lithium thionyl chloride batteries can provide a non-rechargeable option (Jauch Quartz GmbH, 2020).

5.4.3. What are the external power supply needs?

5.4.3.1. Common options/examples. Grid-connected, solar photovoltaic, wind, hydro, and motion energy harvesting.

5.4.3.2. Selection considerations. The lowest cost and typically most reliable method of externally powering the device is grid-connected power. However, if availability is limited, other options may be more appropriate. When selecting an external power source, the likelihood of tampering or theft of the device is a major consideration that should be made. If motion energy harvesting is used, longevity and durability is likely the most important design consideration.

Generally, grid power is recommended when available due to its low cost. However, grid power is often unavailable in remote monitoring situations. Photovoltaics is the next easiest power supply option to implement and could be used when available with panels sized based on power needs.

5.5. Device Housing

The device housing subsystem includes all relevant aspects of the design to ensure the device is protected and secured.

5.5.1. What is the required form factor of the device housing?

5.5.1.1. Common options/examples. Off-the-shelf housing vs. custom housing design that is 3D printed, injection molded, or machined.

5.5.1.2. Selection considerations. The method of integration of the device with the product, space constraints, attachment method, and environmental protection are examples of factors that will affect whether an off-the-shelf housing could be used or whether a custom housing is needed. If a custom housing is used, extensive testing will likely be needed to ensure that adequate environmental protection is provided.

It is recommended to use off-the-shelf housing for the low cost and higher finish quality than 3D printed designs. If the form factor required is not conducive to an off-the-shelf housing, it is recommended to use 3D printed housing for low cost as long as waterproofing is not needed. If waterproofing is needed, it can be very difficult to achieve complete waterproofing using 3D printed housings. However, as 3D printing quality increases and cost decreases, it is possible that 3D printed housing quality will approach injection molded quality.

5.5.2. What environmental factors must the housing protect from?

5.5.2.1. Common options/examples. Dust, temperature, and moisture are likely to be the most common factors that must be considered for the developing world. Water, vibration, impact, and other factors may also be encountered in certain situations.

5.5.2.2. Selection considerations. It may be tempting to design the device to protect from every possible environmental factor. However, to reduce cost, only the factors that may be encountered based on the desired system lifetime should be designed for to keep cost as low as possible.

As mentioned, off-the-shelf housings are recommended where possible because they typically provide the best environmental protection and can be purchased based on designated ingress protection (IP) or NEMA ratings for whatever level of environmental protection is needed. For dust protection only, 3D printed housings with O-ring cord stock lining the lid may provide sufficient protection.

5.5.3. How will the housing or installation location affect any data transmission required?

5.5.3.1. Common options/examples. The use of a metal housing or electromagnetic interference caused by motors or other components in close proximity could negatively affect data transmission reliability.

5.5.3.2. Selection considerations. The primary consideration in potential data transmission issues is the required signal reliability. If successful reception of all data is critical for the device to be useful, all potential signal attenuation should be avoided. Cost and device security are two potential trade-offs that could come with designing for maximum signal reliability.

When possible, use plastic over metal housings to minimize signal attenuation. If a metal housing must be used, explore the possibility of using an external antenna.

5.5.4. How will the housing be secured to the structure of the product?

5.5.4.1. Common options/examples. Permanent (welding, gluing, etc.) vs. removable (screwing, bolting, snap-fitting, double-sided taping, etc.), integrated with product vs. add-on.

5.5.4.2. Selection considerations. As customary in any design process, the advantages and disadvantages of each potential method should be considered.

Whichever method that can be completed using the tools available at the time of installation should be used. For example, even in a case where welding is the ideal attachment method, screwing or gluing may be the better option if it would be prohibitively difficult to provide a welder at the location of device installation.

Obtain feedback from locals who will be using the product to ensure that the method of securing the device does not adversely affect the use of the product.

5.5.5. How will the housing be secured to prevent vandalism or theft?

5.5.5.1. Common options/examples. Security by obscurity, locked housings, tamper proof screws, permanent attachment methods such as welding or high strength adhesives on fasteners, or metal housings.

5.5.5.2. Selection considerations. Security by obscurity, or securing the device so it does not appear in plain sight or valuable, is one promising option for helping prevent vandalism or theft. Other measures of installation that are permanent or more difficult to remove can provide additional protection of the device if security by obscurity is not possible.

Generally, make the device as discreet and unattractive as possible to help avoid the potential for theft and vandalism. Additionally, using high strength attachment methods and security bolts or screws with unique heads may help prevent vandalism.

5.5.6. What endurance or fatigue issues need to be considered?

5.5.6.1. Common options/examples. Cyclically moving parts, wear situations, and devices exposed to the elements or UV degradation.

5.5.6.2. Selection considerations. When designing a device to measure the social impact of a product with moving or exposed parts, long term fatigue and endurance challenges should be addressed.

When possible, avoid integrating the sensor device with the product using methods subject to fatigue, for example, by avoiding designs that would require cyclic bending of wires or other components. When unavoidable, extensive testing should be performed to avoid premature failure.

5.5.7. What maintenance or data collection access will the device need?

5.5.7.1. Common options/examples. Removable media storage (i.e. microSD card) access, battery replacement, entire unit replacement, component lubrication, or sacrificial part replacement if moving parts or corrosion are involved.

5.5.7.2. Selection considerations. If data will be stored on removable media such as a microSD card that will be retrieved periodically, the housing should be designed such that card access is easily accessible. In applications where it is not possible to constantly power the device, battery replacement may be required and the housing should accommodate straightforward battery replacement without needing to completely remove or uninstall the device. If moving parts are integrated into the sensing function, the device should be designed such that lubrication or sacrificial/consumable parts can be replaced.

5.6. Non-Sensor I/O

Non-sensor I/O refers to the non-sensor input and output components needed for proper functionality and installation of the device.

5.6.1. What user inputs must the device accept?

5.6.1.1. Common options/examples. Buttons and switches.

5.6.1.2. Selection considerations. The inclusion of a power button or switch is typically needed to provide an externally accessible and easy way to power on the device. Other buttons or switches may also be needed for configuration in the field during installation or use.

5.6.2. What outputs must the device create?

5.6.2.1. Common options/examples. Motor or servo actuation, LEDs, screens, audible, and vibration.

5.6.2.2. Selection considerations. Outputs for indicating and ensuring correct setup and installation of the device may help prevent other issues in the future. Any other movement or actuation that the device should provide should also be considered.

At a minimum, incorporate an externally visible indicator LED to identify whether the device is able to connect to transmission networks and function properly.

6. Part: Device Deployment

Device Deployment includes all of the key decisions that must be considered when deploying the device into the field to collect the data.

6.1. Data Utilization

Data utilization refers to how the data will be used after being transmitted or retrieved from the device.

6.1.1. Who will need access to the data?

6.1.1.1. Common options/examples. Researchers, organizations, individuals, maintenance workers, and pay-for-service contractors, including whether these parties access the data through open versus closed access.

6.1.1.2. Selection considerations. The first consideration relates to the ease of data use, which may be greater with a user-friendly dashboard and lower with a database. If only researchers or data analysts will be using the data, the cost of developing a user-friendly dashboard may not be justified.

The second consideration relates to whether the data will be made open access for use by the public or closed access and only available to specified stakeholders. If the data is made open access, there is potential for use by a larger community to work toward solving additional challenges, growing businesses, and increasing government accountability (Wagaman, 2016). However, there is added potential for personal privacy and security to be compromised. The protection of individual privacy and security is of utmost importance as discussed in Section 6.7. If the data is closed for access only by specified stakeholders, the data has less potential for use, but data security is easier to ensure.

6.1.2. How will the data be used?

6.1.2.1. Common options/examples. For observing general trends, directly inputting data into a correlation or predictive model, or failure notification.

6.1.2.2. Selection considerations. The required use of the data will directly affect the means of retrieving, storing, displaying, and using the data that should be developed.

If the data is only needed for observing general trends, a basic dashboard should be developed. If the data is used only for failure notification, a text message based approach may be sufficient without need for a dashboard. If the data will be used within a correlation or predictive model, a dashboard may not be needed if the predicted values are all that is needed, but a dashboard may be helpful for visualizing the modeling results.

6.1.3. How will the data be stored after retrieved from the device?

6.1.3.1. Common options/examples. Web-based (cloud-based) server (local server or 3rd party), local computer.

6.1.3.2. Selection considerations. Cost and who will need access to the data are the primary drivers for this decision. If multiple users in different locations need access to the data, a web-based server may be the best option, although this may come at a higher recurring cost than local, non-web based storage.

Web-based servers are recommended for most applications due to their minimal or non-existent upfront costs, reasonable recurring costs, and great flexibility in scaling or adjusting.

6.2. Training Data Collection

Training data collection is the process of simultaneously collecting direct and sensor data for use in creating or training a correlation model. It is an optional factor that will only need to be considered when correlating sensor data with some higher level type of data via a correlation model (Stringham et al., 2020).

6.2.1. What additional equipment is required for training data collection?

6.2.1.1. Common options/examples. Cameras or other custom devices connected to the sensor device wirelessly by Bluetooth, for example.

6.2.1.2. Selection considerations. In the process of collecting training data, cameras may be required for recording training data that is subsequently labeled. Custom devices may need to be connected to the primary sensor device through Bluetooth, for example, to increment counters or record a start/stop time on the data collection device for straightforward use in later labeling of training data.

If possible, based on local privacy laws and IRB approval, it is recommended to use video to record data needed for labeling sensor data and training the model. This allows greater accuracy in most cases than real-time data labeling. If real-time labeling is required, the process should be as automated as possible such as through pressing buttons on a remotely connected custom device, smart phone, or computer, for example, to result in highest possible accuracy.

6.2.2. What additional or different device functionality does the training data collection device need from the deployed device?

6.2.2.1. Common options/examples. Same as or different than deployed device; on-board storage may be required even if deployed device does not need it.

6.2.2.2. Selection considerations. The device used during collection of training data may need a larger battery capacity or on-board storage for storing large amounts of training data.

It is recommended to include onboard storage for the Training Data Collection device to enable data recording.

6.3. Testing

Although intermediate testing should be done regularly through the device design process for each of the subsystems, testing warrants inclusion as its own box in the canvas due to the critical nature of testing to successful long term use and additional possible causes of system failure in a developing world environment.

6.3.1. What pre- and post-deployment testing is needed to help ensure successful long term use?

6.3.1.1. Common options/examples. Examples include fatigue, durability, data transmission, installation, environmental and long-term testing of the sensor device.

6.3.1.2. Selection considerations. The testing required for this device is similar to that of any product except some of the primary challenges faced by this device may be different than other typical products. For example, long term testing may be more critical due to the importance of correct long term operation to the success of the entire data collection process. Also, environmental or signal conditions can be difficult to predict and may require greater effort to simulate, so the installed conditions and time period of use should be simulated and tested. Some testing may and should only be completed in the country of use.

It is recommended to complete a thorough failure modes and effects analysis (FMEA) and formal testing plan for the device so that potential risks of failure may be mitigated as much as possible (Mattson and Sorensen, 2020). As much testing as possible may be completed in the home location of the practitioners, but extensive testing of both the hardware and data pipeline in the country of use is still highly recommended.

6.4. Manufacturing

Manufacturing refers to how the devices will be manufactured at low as well as high volumes.

6.4.1. How will the devices be manufactured and assembled?

6.4.1.1. Common options/examples. In-house, partner/contractor, or hybrid.

6.4.1.2. Selection considerations. Quantity, cost, and supply chain include several of the factors that should be considered when deciding how the devices will be manufactured and assembled.

For low quantity, in-house or a hybrid approach (partner manufactured and in-house assembled) could be the best, whereas contract or partner manufacturing and assembly is typically the best choice for large quantities of devices.

6.4.2. How many devices will need to be manufactured?

6.4.2.1. Common options/examples. One to thousands.

6.4.2.2. Selection considerations. The quantity required depends on the scale of impact that is desired to be measured and the stage of the device development.

Initially, a small quantity of devices could be deployed to prove the concept and perform validation testing. This could be followed by the refinement of the device and use of a large quantity of devices.

6.4.3. When will device manufacturing need to be completed?

6.4.3.1. Common options/examples. Few weeks to many years.

6.4.3.2. Selection considerations. The manufacturing time of proof-of-concept devices can be as short as several weeks if off-the-shelf microcontrollers and component breakout boards are used. Manufacturing time will often extend to two or more months for large quantities or for custom PCBs, components, and housings due to time required to establish supply chains, build tooling, and setup manufacturing lines.

It is recommended to allow as long of a lead time for manufacturing as possible while still meeting desired deadlines because longer lead times typically mean lower manufacturing costs.

6.5. Installation

Installation refers to how the devices will be designed for proper installation and how they will be installed during deployment. The correct deployment and installation by field staff and partners in remote situations is of particular significance.

6.5.1. How will the devices be installed?

6.5.1.1. Common options/examples.
Self, field staff, partner organization, or individual including residents local to where device will be used.

6.5.1.2. Selection considerations.
Catastrophic failure can occur due to lack of training in the correct installation of the device, even when a device has been otherwise well-designed. As such, training materials in the form of manuals and/or videos should be used to ensure correct installation of the device. Language, cultural, or experience differences should be considered when deciding how the devices will be installed. If a partner with a language, cultural, or experience barrier is used, visual instructions may be better than written instructions only.

Additionally, the device should be designed such that it is difficult to install incorrectly. Feedback from the field staff or others that will be installing the device should be sought during the design process to ensure that issues will not arise during the installation process.

It is recommended that the practitioners install or be present at the installation of the first devices when possible to ensure that installation is performed correctly. However, for long term and scaled deployment, it is recommended to partner with a local individual or organization to reduce installation costs due to travel.

6.5.2. How will correct device installation be recognized?

6.5.2.1. Common options/examples.
Visual or audible indicator(s) on the device for the installer; backend or dashboard-based analysis for end user of data.

6.5.2.2. Selection considerations.
A procedure of determining that data transmission and collection is correct should be setup for the practitioner to know that the device was installed correctly to ensure the data is useful and can be trusted.

If a partner individual or organization will be installing the device(s), a visual or audible indicator should be used to aid the installer to know the installation was performed correctly. A diagnostic test routine such as moving the product in a specific way for which there is a known correct response on the sensor device should also be used to ensure correct installation.

6.5.3. How will transmission of accurate and representative data be ensured?

6.5.3.1. Common options/examples.
Through periodic inspection of the device, periodic inspection of the data, or anomaly detection in data processing.

6.5.3.2. Selection considerations.
Ensuring correct installation is a major aspect of being able to trust the data that has been transmitted. However, the device is usually not completely immune to tampering, so it is beneficial to have a method of determining that the device is continuing to transmit useful and representative data.

The data being transmitted should be inspected periodically to ensure there are not data artifacts such as drift, hysteresis, or other data anomalies not representative of the actual behavior of the system. If the integrity of the data is sensitive to device positioning, the installed device should be inspected periodically when possible or after detecting data anomalies to ensure it has not been tampered with. For large scale use, autonomous methods of anomaly detection should be established.

6.6. Maintenance & Operation

Maintenance and operation refers to any aspects of device functionality and reliability post-installation.

6.6.1. What maintenance will the device require?

See Section 5.5.7.

6.6.2. What training will be needed by those maintaining or operating the device?

6.6.2.1. Common options/examples.
On-board storage retrieval, battery replacement, device or component replacement training, device operation.

6.6.2.2. Selection considerations.
Catastrophic failure can occur due to lack of training in the correct maintenance and operation of the device, even when a device has been otherwise well-designed. As such, training materials in the form of manuals and/or videos should be used to ensure correct installation of the device. Language, cultural, or experience differences should be considered when deciding how the devices will be maintained and operated. If a partner with language, cultural, or experience barrier is used, visual instructions may be better than written instructions only.

Additionally, the device should be designed such that it is difficult to maintain or operate incorrectly. Feedback from the field staff or others that will be maintaining and operating the device should be sought during the design process to ensure that issues will not arise during the maintenance or operation processes.

6.6.3. How can the device be installed to minimize or eliminate affecting the typical use of the product during and after installation?

6.6.3.1. Common options/examples.
Device hidden versus not hidden, integrated within product vs. added on externally.

6.6.3.2. Selection considerations.
If the product whose impact will be measured by the device has not been fully designed, it is possible and perhaps desirable to integrate the data collection device into the design of the product. Regardless, the ways in which the device affects normal product use or perception of the product should be minimized. This is especially the case for products that affect the health and safety of individuals. Feedback from locals using the product should be sought during the design process to ensure its functionality is acceptable.

When possible, data collection devices should be hidden and integrated within the product to prevent tampering and promote normal use of the product.

6.6.4. What hazards could result from the installation and use of the sensor device?

6.6.4.1. Common options/examples.
Shock hazard, potential battery fire, or contamination.

6.6.4.2. Selection considerations.
Care should be taken to ensure that no additional risks are posed by implementing the data collection device.

Testing of the device in extreme use conditions should be performed under controlled conditions to ensure no catastrophic or hazardous outcomes can result from the device's use.

6.6.5. How will the device be retrieved when data collection is complete?

6.6.5.1. Common options/examples.
Partner individual/organization or self.

6.6.5.2. Selection considerations.
An important step at the end of the data collection process is to retrieve and dispose of the product. This step should not be overlooked so as to reduce pollution and minimize negative environmental impact.

Using a partner organization local to the device's location will often result in lower retrieval costs.

6.7. Ethics and Regulations

Ethics and regulations include any ethical considerations that should be made when collecting data from users in addition to privacy concerns, data security issues, and any regulations or laws that could govern the collection of the data. As with all canvas boxes, the team should write the answers to the following questions directly on the canvas.

6.7.1. How will user identity and privacy rights be protected?

6.7.1.1. Common options/examples. Not collecting personally identifiable information; securing the stored data through encryption, security keys, or password protection; removing personally identifiable information before publicly sharing; or using layers of access between public and private data (Wagaman, 2016).

6.7.1.2. Selection considerations. It is important to consider whether or not the user should be aware that their data is being collected. Check local laws and do not collect personally identifiable information unless that data is secured using best practices and IRB approval is obtained.

For situations in which data will be shared publicly, personally identifiable information should be removed, but it is still possible to re-identify individuals in some cases. The utilization of layers of access between public, non-sensitive data and private, sensitive data can help protect individuals privacy and data security (Wagaman, 2016).

6.7.2. How will relevant regulatory body or individual permissions be secured and fees be paid?

6.7.2.1. Common options/examples. Privacy and data transmission laws, waivers, and consent forms.

6.7.2.2. Selection considerations. The collection and transmission of product usage data may not be legal in all countries and applications.

Permissions from all relevant regulating bodies such as governments (local or national), village leaders, community groups, consortia, families, and individuals should be obtained when necessary.

6.7.3. What IRB requirements are present and how will institutional and international IRB secured, if necessary?

6.7.3.1. Common options/examples. International IRB approval can usually be obtained at the institutional level.

6.7.3.2. Selection considerations. IRB approval may be required for legal and ethical collection of data.

If needed, IRB approval through the researcher's institution should be obtained. International IRB approval is similar to national IRB approval with potentially additional considerations.

6.8. Strategy

Strategy refers to partnerships or collaborations that could aid the practitioner in the design or deployment process of collecting needed data as well as intellectual property considerations relevant to developing and deploying these devices.

6.8.1. What partnerships could be formed for financial support, installation, maintenance, and data collection as needed?

6.8.1.1. Common options/examples. Individuals identified through personal connections, academia, local governments, religious institutions, NGOs, non-profits, businesses, or other organizations.

6.8.1.2. Selection considerations. In resource constrained settings such as research, it can be beneficial to identify individuals or organizations that share a common interest in the research questions that will be answered by the device. These organizations may be willing to provide financial or other material support for deploying the device. They may also be able to provide local connections for installation and device integration assistance. Partner individuals or organizations could be US-based, international, or local to the country in which the devices will be deployed. Partners may also be helpful in navigating regulations in the country where the device will be deployed.

Partners are highly recommended when possible. Potential partners should be vetted for trustworthiness. If partners are being hired, the costs and expectations of required work of both parties should be made clear and verified in writing at the beginning of the partnership to avoid potential issues caused by differences in expectations.

6.8.2. What approach should be used relative to the protection and distribution of intellectual property created during the hardware and software development of the device and data collection process?

6.8.2.1. Common options/examples. Proprietary approach in which technology is protected for exclusive use by specific stakeholders; open source approach in which technology developed is admitted into the public domain for use and development by others; or a mixture of both approaches that varies by subsystem of the sensor device and data collection process.

6.8.2.2. Selection considerations. A proprietary approach can incentivize a more cost effective production and delivery than an open source approach. Primarily, a proprietary approach to these applications in developing countries can come through patents, trade secrets, and copyrights. Since patents are territorial and subject only to the laws of the country in which they are filed, patent protection must be obtained from each territory in which patent protection is desired (International Trade Administration, 2018). Before pursuing a patent, the ability and cost to litigate the defense of the patent should be weighed against costs of obtaining the patent. Trade secrets in which the design of the hardware or software is confidentially protected and never disclosed publicly can nevertheless still provide sufficient protection in some instances.

An open source approach can enable greater collaboration within the development community and prevent the duplication of work. It also has the potential to distribute workload between a broad community of invested individuals and enable more work to be performed than would be possible by a smaller group of individuals within an organization. However, open source should not be confused with free. When deciding whether to take an open source approach, the costs of deploying the code, maintaining the code base and hosting environment, supporting continued community engagement, and any other long term costs should be considered (Wagaman, 2016).

A mixed approach includes a mixture of benefits of both the proprietary and open source approaches and may be appropriate for some applications. This could include, for example, open sourcing the data collected or the code used in data processing while keeping the hardware used to collect the devices for proprietary use.

7. Concluding Remarks

One primary way to ensure that engineered products are producing positive effects on individuals in developing countries is through social impact measurement of those products. The use of social impact indicators provides one effective way to combine relevant user data in meaningful way to measure a product's social impacts (Stevenson et al., 2020; Stringham et al., 2020). Electronic sensor devices provide one potential way to remotely, continuously, and inexpensively collect user data that can be used to calculate social impact indicators for engineered products in developing countries. However, many key decisions must be considered to effectively design, test, manufacture, and deploy these remote data collection devices.

Decomposing the device development process into parts of Data Identification, Device Design, and Device Deployment provides a useful and structured way by which to consider the many key decisions that must be considered. Further decomposing the Data Identification part into the key decisions of Social Impact Indicators, Physical Phenomena, and Data Correlation; the Device Design part into key decisions or subsystems of Sensor Data, Computer, Data Retrieval, Power Supply, Device Housing, and Non-Sensor I/O; and the Device Deployment part into the key decisions of Data Utilization, Training Data Collection, Testing, Manufacturing, Installation, Maintenance and Operation, Ethics and Regulations, and Strategy further provides structure by which essential aspects of the development and use can be considered. This paper provides critical questions, common options, and selection considerations and guidelines for each of the key decisions. A canvas is also provided as a tool to help researchers track their answers to these questions and visualize areas of potential concern throughout the development and use of these devices. By systematically considering all of these questions, researchers can more effectively develop and deploy remote data collection devices for social impact measurement of engineered products in developing countries.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to recognize the National Science Foundation for providing the Graduate Research Fellowship Grant 1247046 and Grants CMMI–1632740 and CMMI–1761505 that funded this research. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Andres, L., Boateng, K., Borja-Vega, C., Thomas, E., 2018. A review of in-situ and remote sensing technologies to monitor water and sanitation interventions. Water 10 (6), 756.

Ascent International Group Co. Ltd, 2020. Ascent International Group Co. Ltd. Component Specification Page. Available at: http://aliumbatteries.com/products/secondary-batteries/li-polymer-battery/li-po-low-high/(2020/07/16.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Burdge, R.J., 2004. A Community Guide to Social Impact Assessment, vol. 2. Social Ecology Press.

Clark, H., Anderson, A.A., 2004. Theories of change and logic models: telling them apart. In: American Evaluation Association Conference.

Collings, S., Munyehirwe, A., 2016. Pay-as-you-go solar pv in Rwanda: evidence of benefits to users and issues of affordability". *Field Actions Science Reports*. J. Field Actions 94–103 (Special Issue 15).

Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 1–67.

George, C., Shams, A., 2007. The challenge of including customer satisfaction into the assessment criteria of overseas service-learning projects. Int. J. Serv. Learn. Eng. 2 (2), 64–75 sep.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT press.

He, L., Wang, M., Chen, W., Conzelmann, G., 2014. Incorporating social impact on new product adoption in choice modeling: a case study in green vehicles. Transport. Res. Transport Environ. 32 (oct), 421–434.

Hutchins, M.J., Gierke, J.S., Sutherland, J.W., 2009. Decision making for social sustainability: a life-cycle assessment approach. In: International Symposium on Technology and Society, Proceedings. IEEE, pp. 1–5.

IHS, 2015. Internet of Things (IoT) Connected Devices Installed Base Worldwide from 2015 to 2025 (In Billions).

International Trade Administration, 2018. Is My U.S. Patent Good in Other Countries?.

Jagtap, S., 2019. Design and poverty: a review of contexts, roles of poor people, and methods. Res. Eng. Des. 30 (1), 41–62.

Jauch Quartz GmbH, 2020. Jauch Quartz Gmbh Component Specification Page. Available at: https://www.jauch.com/blog/en/advantages-and-special-features-of-lithium-thionyl-chloride-batteries/(2020/07/16.

Johnson, G.E., Fisher, M.J., Salmon, J.L., Mattson, C.A., 2021. Product development using perceived correlations between the united Nations sustainable development goals and social impact categories. In: Proceedings of the ASME IDETC/CIE 2021, Aug, pp. 1–8.

Others Kellogg, W.K., 2006. WK Kellogg Foundation Logic Model Development Guide. WK Kellogg Foundation.

Kipf, A., Brunette, W., Freitas, J., Podolsky, M., Rosa, J., Sundt, M., Wilson, D., Borriello, G., Brewer, E., Thomas, E., 2015. A proposed integrated data collection, analysis and sharing platform for impact evaluation. Dev. Eng. 1.

Mattson, C.A., Sorensen, C.D., 2020. Product Development: Principles and Tools for Creating Desirable and Transferable Designs. Springer.

Mattson, C.A., Wood, A.E., 2014. Nine principles for design for the developing world as derived from the engineering literature. J. Mech. Des. 136 (12), 121403.

Mehta, C., Mehta, K., 2011. A design space and business strategy exploration tool for infrastructure-based ventures in developing communities. Int. J. Serv. Learn. Eng. Humanit. Eng. Soc. Entrepren. 6 (2), 30–57.

Nagel, C., Beach, J., Iribagiza, C., Thomas, E., 2015. "Evaluating cellular instrumentation on rural handpumps to improve service delivery – a longitudinal study in rural Rwanda". Environ. Sci. Technol. 49.

Osterwalder, A., Pigneur, Y., 2010. Business Model Generation: a Handbook for Visionaries, Game Changers, and Challengers. John Wiley & Sons.

Ottosson, H.J., Mattson, C.A., Dahlin, E.C., 2019. Analysis of Perceived Social Impacts of Existing Products, with Implications for New Product Development. Manuscript in review.

Ottosson, H.J., Mattson, C.A., Johnson, O.K., Naylor, T.A., 2021. Nitrile cup seal robustness in the India mark II/III hand pump system. Dev. Eng.

Particle Industries, Inc, 2020. Particle Industries, Inc. Component Specification Sheet. Available at: https://docs.particle.io/assets/pdfs/new-certs/battery/zn-103450-iec621331.pdf (2020/07/16.

Pessôa, M.V.P., Becker, J.M.J., 2020. Smart design engineering: a literature review of the impact of the 4th industrial revolution on product design and development. Res. Eng. Des. 1–21.

Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1 (1), 81–106.

Rainock, M., Everett, D., Pack, A., Dahlin, E.C., Mattson, C.A., 2018. The social impacts of products: a review. Impact Assess. Proj. Apprais. 36 (3), 230–241.

Ramsey, F., Schafer, D., 2012. The Statistical Sleuth: A Course in Methods of Data Analysis. Cengage Learning.

Sharpe, T., Muragijimana, C., Thomas, E., 2019. Product design supporting improved water, sanitation, and energy services delivery in low-income settings. Sustainability 11 (23).

Sinha, A., Nagel, C.L., Thomas, E., Schmidt, W.P., Torondel, B., Boisson, S., Clasen, T.F., 2016. Assessing latrine use in rural India: a cross-sectional study comparing reported use and passive latrine use monitors. Am. J. Trop. Med. Hyg. 95 (3), 720–727.

Stevenson, P.D., Mattson, C.A., Bryden, K.M., MacCarty, N.A., 2018. Toward a universal social impact metric for engineered products that alleviate poverty. J. Mech. Des. 140 (4), 41404 feb.

Stevenson, P.D., Mattson, C.A., Dahlin, E.C., 2020. A method for creating product social impact models of engineered products. J. Mech. Des. 142 (4).

Stringham, B.J., Smith, D.O., Mattson, C.A., Dahlin, E.C., 2020. Combining direct and indirect user data for calculating social impact indicators of products in developing countries. J. Mech. Des.

Thomas, E., Brown, J., 2021. Using feedback to improve accountability in global environmental health and engineering. Environ. Sci. Technol. 55 (1), 90–99. PMID: 33305578.

Thomas, E.A., Barstow, C.K., Rosa, G., Majorin, F., Clasen, T., 2013a. Use of remotely reporting electronic sensors for assessing use of water filters and cookstoves in Rwanda. Environ. Sci. Technol. 47 (23), 13602–13610 dec.

Thomas, E.A., Zumr, Z., Graf, J., Wick, C., McCellan, J., Imam, Z., Barstow, C., Spiller, K., Fleming, M., 2013b. Remotely accessible instrumented monitoring of global development programs: technology development and validation. Sustainability 5 (8), 3288–3301.

Thomas, E.A., Kathuni, S., Wilson, D., Muragijimana, C., Sharpe, T., Kaberia, D., Macharia, D., Kebede, A., Birhane, P., 2020. The drought resilience impact platform (drip): improving water security through actionable water management insights. Front. Clim. 2, 6.

United Nations Development Programme Growing Inclusive Markets Initiative, 2008. Creating Value for All: Strategies for Doing Business with the Poor. United Nations Development Programme.

United Nations Statistics Division, 2020. E-handbook on Sustainable Development Goals Indicators. Tech. rep.. United Nations, New York City.

United Way of America, 1996. Measuring Program Outcomes: a Practical Approach. Alexandria, VA.

Ventrella, J., Lefebvre, O., MacCarty, N., 2020. Techno-economic comparison of the fuel sensor and kitchen performance test to quantify household fuel consumption with multiple cookstoves and fuels. Dev. Eng. 100047.

Wagaman, A., 2016. From Principle to Practice: Implementing the Principles for Digital Development. Tech. rep.. The Principles for Digital Development Working Group, Washington, DC.

Wilson, D.L., Adam, M.I., Abbas, O., Coyle, J., Kirk, A., Rosa, J., Gadgil, A.J., 2015. Comparing cookstove usage measured with sensors versus cell phone-based surveys in darfur, Sudan. In: Technologies for Development. Springer, pp. 211–221.
Wood, A.E., Mattson, C.A., 2016. Design for the developing world: common pitfalls and how to avoid them. J. Mech. Des. 138 (3), 31101.
Wood, A.E., Mattson, C.A., 2019. Quantifying the effects of various factors on the utility of design ethnography in the developing world. Res. Eng. Des. 1–22 jan.