Research Article

Isaac Nyabisa Oteyo* and Mary Esther Muyoka Toili

Improving Specimen Labelling and Data Collection in Bio-science Research using Mobile and Web Applications

<https://doi.org/10.1515/comp-2020-0002>

Received Mar 23, 2019; accepted Nov 08, 2019

Abstract: Researchers in bio-sciences are increasingly harnessing technology to improve processes that were traditionally pegged on pen-and-paper and highly manual. The pen-and-paper approach is used mainly to record and capture data from experiment sites. This method is typically slow and prone to errors. Also, bio-science research activities are often undertaken in remote and distributed locations. Timeliness and quality of data collected are essential. The manual method is slow to collect quality data and relay it in a timely manner. Capturing data manually and relaying it in real time is a daunting task. The data collected has to be associated to respective specimens (objects or plants). In this paper, we seek to improve specimen labelling and data collection guided by the following questions; (1) How can data collection in bio-science research be improved? (2) How can specimen labelling be improved in bio-science research activities? We present *WebLog*, an application that we prototyped to aid researchers generate specimen labels and collect data from experiment sites. We use the application to convert the object (specimen) identifiers into quick response (QR) codes and use them to label the specimens. Once a specimen label is successfully scanned, the application automatically invokes the data entry form. The collected data is immediately sent to the server in electronic form for analysis.

Keywords: object tagging, bio-science, mobile phones, tissue culture, software engineering

1 Introduction

Advances in the agricultural industry have led to improvement in addressing food production even as demand increases. One of the ways to rapidly produce plants for food and feed is micropropagation (a plant tissue culture approach used for producing plantlets in mass). It is a specialised and technical means of propagating plants, requiring an aseptic growth of sections of a plant tissue or organ in a nutrient-defined culture medium within a controlled environment [1]. The success of micro-propagation has witnessed commercialisation of the process globally. Many biotechnology processes require the use of *in vitro* micropropagation for a multitude of uses including rapid plant multiplication, germplasm conservation, production of pathogen-free plants, genetic manipulations, and for secondary metabolite production for vaccines and pharmaceuticals [2]. As such, continuous research to develop protocols for the tissue culture of different plant species is rife and is fuelled by the specificity of the tissue culture protocols which tend to be species and variety specific.

Developing a tissue culture protocol for any plant requires a lot of experimental variations to decipher the ideal culture conditions for individual tissues used in the culture. Nutrients, plant growth regulators, explant type (starting tissue), type of medium, light intensity, temperature, pH, and humidity must all be considered in individual experiments. Several trials are setup and vast amounts of data must be visually and manually collected and recorded. Tracking the growing explant as it is multiplied several times from the mother plant requires precise and concise labelling skills which present a daunting task in a laboratory running different experiments. A commercial tissue culture laboratory producing millions of plantlets will spend immeasurable time collecting data and formulating labels for individual plants.

The conventional way for data collection has been the use of pen and paper [3]. It is a preferred method for its simplicity and minimal cost of initial implementation. However, this method is impractical for large studies associ-

*Corresponding Author: **Isaac Nyabisa Oteyo:** Software Languages Lab, Vrije Universiteit Brussel, Brussels, Belgium; Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya; Email: isaacoteyo@googlemail.com

Mary Esther Muyoka Toili: Plant Genetics Lab, Vrije Universiteit Brussel, Brussels, Belgium; Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

ated with several plant species and certainly for commercial tissue culture laboratories. It is not only time consuming, but the records are composed of volumes of paper requiring investment in storage space and security. The time and cost needed to transform such data into electronic format is considerable.

Advances in computing technologies provide new opportunities where researchers can use mobile devices to collect data and directly transform the data into electronic format and send it to servers for analysis. The receiver of the data can cross-check and validate the accuracy of the data as it arrives on the server. Electronic data collection tools have benefits in terms of improved survey (questionnaire) development and formatting, data quality, data collection, consistency in data collection, accuracy, data reliability, timeliness in data collection, and ease of enforcement and adherence to data collection protocols. With the widespread adoption of smartphones, this cost is drastically reducing. Smartphones are ubiquitous and can be optimally utilised at a minimal incentive to the owner to improve the data collection process. This requires applications to be developed to run on these devices to aid the data collection process.

Besides data, specimen identification is essential in research experiments. The identification is done by way of labelling using a unique identifier for each specimen. The data collected need to be directly linked to the associated specimen. Different alternative approaches have been proposed in the literature [4–6]. However, contextualising these methods still remains open for investigation.

In order to address the existing challenges, we seek to improve specimen labelling and data collection in bio-science research guided by the following research questions;

1. How can data collection in bio-science research be improved?
2. How can object (specimen) labelling be improved in bio-science research activities?

We address the aforementioned challenges by implementing a prototype mobile-web application to help bio-science researchers label their specimens and collect the associated data. Our application helps the researchers to convert object identifiers into QR codes for specimen labels. Each tag is linked to a data collection form which is automatically invoked on successful scan. In this paper, we identify challenges and benefits of using QR codes for labelling specimens and mobile-web applications for data collection. We present a bio-science use case and describe how the combination of QR codes and mobile-web applications can be used to improve object labelling and data

collection. We also provide the lessons learnt from the experience and directions for future research.

In the subsequent sections, the paper is organised as follows. We begin by describing our approach followed by the current scenario and associated challenges. We then give a description of our prototype application and its implementation. This is followed by the evaluation of the application in terms of performance, discussions on the evaluation and related work. And finally, we provide the summary, lessons learnt, the conclusion and directions for future work.

2 Approach

To address the challenges mentioned earlier, we followed a methodological approach to unravel the problems that we aimed to solve. We first started by developing a use case for our application. Our use case was motivated by experiences and interactions with bio-science researchers. Through the interactions, we were able to document the challenges faced by bio-science researchers that can be solved by application of technology. The interactions were made possible through a joint multidisciplinary research programme with multiple projects. At first, we observed the way they conducted their research which was mainly manual; specimen labelling was done manually and data collection was done on paper-based forms before being transferred to spreadsheets. From literature, the manual data collection is a classical approach even though it is prone to human errors, especially during data entry. From the interactions, we observed that improving data entry can considerably reduce the total time required to collect data from an experiment. We held focussed group discussions with the bio-science researchers to elicit more details on their research processes and activities to make our use case perceptible.

To move from manual to automated specimen label generation, the following requirements were found to be necessary.

1. Creating virtual experiments to simulate the actual physical experiment.
2. Each experiment is owned by a single user.
3. Specimens belong to an experiment.
4. An experiment can have zero or multiple specimens.
5. Each experiment has an identifier that should be translated into a QR code.
6. Data collection forms should be linked to the specimen identifier.

Figure 1: Current scenario for data collection

The requirements were then translated into the following modules.

1. User registration module to manage the application users.
2. Experiment management module.
3. Specimen management module.
4. QR code generation module.
5. QR code scanning module.
6. QR code printing module.
7. Data collection form creation module.
8. Internal notification module.

We also conducted literature surveys on the use of object labelling in other fields which we considered useful to bring into our use case as a way of improving the manual specimen labels. Our choice for QR codes was motivated by the associated cost and the ability to print and use them on any surface. We then implemented a prototype application that allows the user to convert object identifiers into QR codes and use the printed codes to label the specimens. To collect data, the user needs to scan the specimen label using the camera sensor available in most smartphones. Successfully scanning a specimen label automatically invokes a data collection form.

To implement the application, we settled on prototyping as the software engineering methodology. On technologies, React Native, ReactJS, Node.js, and MongoDB were our implementation technologies based on a number of reasons. First, ReactJS is component-based and allows the development of interactive user interfaces. Component-

based development allows the development of applications with the incremental introduction of new end-user requirements. React Native allows development of cross-platform applications from a single codebase. Node.js allows server-side execution of JavaScript without needing a web browser. MongoDB is a commonly used NoSQL database that stores data as JSON objects that are referred to as documents.

Our goal is to provide effective support for specimen labelling and data collection for researchers in bio-science. Using the scenario described in Section 3, we require;

1. Ability to generate specimen labels that are associated to unique specimen identifiers. The labels are printable with meta information to avoid information overload to the end user.
2. Ability to link the generated specimen labels to the data collection forms. The forms are loaded automatically upon successful scan of the specimen label by a camera sensor.

We describe the current scenario, the prototype application developed, and the services it provides in the subsequent sections.

3 Current scenario

Figure 1 shows the current data collection scenario that we observed while interacting with the researchers in bio-

Figure 2: Manually generated plant pot labels

Figure 3: Manual object labels in labs and growth chambers

Figure 4: Manual data collection

science. The research activities are conducted in an experiment site which can be inside a laboratory (closed environment) or open field (open environment). Closed environment experiments are done under controlled conditions. If we consider tissue culture research, initial experiments

begin in the closed environment where callus (a mass of rapidly proliferating cells) is induced from a carefully selected explant obtained from a mother plant with desirable qualities. After several sub-cultures (moving plant tissues into new media), new plantlets are generated from the initial callus. The callus is grown in media which is a combination of different chemical compounds considered suitable to provide the required ingredients to induce growth. The media is contained in pots. Every new pot at each point of sub-culture is labelled with short inscriptions for reference. The inscriptions include the variety of the plant and time related information like date of culture. After several sub-cultures, new plantlets are then generated from the initial callus. The new plantlets are also labeled with short inscriptions for reference. This process is repeated until well developed plantlets are acclimatised into actual soil in larger pots within a greenhouse. Each of the larger pots is labeled for identification (Figure 2 and Figure 3). The labels are plastic strips that are hand-written (Figure 2). The labelling is done manually and a manual record kept for all the labels. Data is manually collected on each plant label (Figure 4 and Figure 5) and later transferred to spreadsheets for processing (Figure 5). Data is periodically collected from the onset of the experiment until the plant is transferred to the open field.

Each experiment can have several to hundreds of plants that translate into several to hundreds of plant pots. Commercially, some techniques such as somatic embryogenesis result in hundreds of plantlets from a single callus culture. Some of the resulting plants perform better than others depending on a number of factors including variations in the media used to induce the callus growth, environmental conditions and the source of the explant (i.e., the mother plant). At this point, bio-science researchers find it necessary to trace the hierarchy of the plants to the original explant (and mother plant) and subsequently the media and environmental conditions that were used to grow the resultant plant. Manually tracing such a hierarchy is a challenging and difficult task.

This kind of data collection is highly human interactive and makes QR codes a possible candidate labelling mechanism to use, being applicable where human interaction is required. Once the object is scanned and verified, an electronic form is automatically invoked for data entry. The collected data is immediately transformed into an electronic format and sent to the server for processing. It is easier to work with the electronic format of the collected data in the later stages of data processing and analysis. It is on this basis that we focus our work on using QR codes to collect data from experimental setups in bio-science research. We use the QR codes in this context to label specimens for

Figure 5: Collected data transferred to spreadsheets for processing

Figure 6: Logical communication scheme

Figure 7: Frontend application architecture

identification. Once labeled it becomes easier to scan the specimen labels using a mobile phone with a camera sensor and have a form loaded for data collection. We use this reasoning to implement the *WebLog* application. The application was initially prototyped to help researchers collect basic physiological properties of plants. We describe the functionality and reasoning behind our prototype application in the subsequent sections.

4 Application architecture

In this section, we provide the description of *WebLog*, the application we prototyped to help researchers generate specimen labels and collect data using mobile phones. The application is composed of two frontends and one backend (server). One frontend runs on a mobile device as a mobile client while the other frontend runs on the browser as a web app. The connection between the frontend and the backend is made through the REST services which are exposed by the backend application.

4.1 Frontend architecture

The architecture of the application frontend is as shown in Figure 6 and Figure 7. Both the mobile client and web app are implemented in JavaScript using React Native and React respectively. By design, the web application is responsive i.e., automatically resizes to fit the device screen size. It is a design choice to have the two frontends to widen the scope and number of users for the application. Also, some services supported in the application are better managed from the web client as opposed to the mobile interface e.g., generating and managing QR codes that are used for labelling specimens. However, this requires maintaining two separate codebases. HTML5 and CSS are used for markup and styling respectively. Both frontend applications offer mechanisms for routing to different sections and views of the application. The views are represented as different application pages.

Figure 8: Application data model

4.2 Backend architecture

The backend application is implemented in JavaScript by using Node.js. Node.js is a run-time environment which executes JavaScript code without the need for a browser. It is a very popular variant of JavaScript for server-side programming because of its high performance attributed to its support for event-driven programming. Event-driven programming model uses callbacks to report when a task is complete which eliminates the need for threading.

5 Logical communication scheme

The logical communication scheme for the application is illustrated in Figure 6. The backend is linked to a MongoDB (NoSQL) database for storing data. It communicates to both the mobile and web client using RESTful APIs and HTTP requests. The choice of MongoDB was based on its popularity as a NoSQL database.

5.1 Application data model

The data model used by the application is illustrated in Figure 8. A user is a logical entity representing a project, re-

search group, lab, firm or any other entity that owns data from a study. Experiments are particular studies. Each experiment has one owner. A single experiment can have multiple specimens, but a specimen cannot belong to more than one experiment. Each specimen has a unique identifier associated with it at the time of creation. Data is collected on a per specimen basis. The collection can occur multiple times. The data model enforces the following policies:

1. A user can have zero or more experiments (studies).
2. An experiment can have zero or more specimens.
3. Each specimen has a unique identifier associated with it.
4. A specimen can have zero or more datasets.
5. Each dataset is a collection of values and a timestamp at the time of collection or modification.

5.2 Internal notification

Internally, the application runs a grow-shrink counter to keep track of each dataset recorded into the application. The counter grows on the addition of new records and shrinks on the reduction of records (Figure 9). The counter value is displayed visibly for the user to see the number of datasets per specimen.

Figure 9: Dataset counter

6 Application description and service specification

In this section, we present the application overview together with the specification of the implemented services. For each service, we present its functional description, integration with external services and how critical cases are handled. We also describe the key aspects of the implementation of each service.

6.1 Application overview

The application presents two perspectives; one for logged in users and one for visitors. A landing page is presented to a user visiting the application and is not logged in. The page allows the user to navigate to the registration or login view. The registration view allows a new user to create an account. Once the account is created, the user can log in and proceed with other services presented by the application.

The application has a frontend and backend. The frontend is divided into two; one that runs on a mobile device as a mobile app and another that runs on a browser as a web app. This design approach is targeted to reach out to more users and also provide an interface for generating and managing specimen tags. The mobile frontend is implemented using React Native while the web frontend is implemented using React. Both React Native and React are variants of JavaScript. The application has one backend implemented using Node.js. Data is stored in a MongoDB. The application runs several services accessible to the end user.

6.2 Identity management service

To gain access to the services provided by the application, a user is required to have an active account. From the application home page, a user with an account already can proceed to log into the application or navigate to the registration page to create a new account. User registration is a

public service that allows users to sign up into the application. The registration workflow is represented in Figure 10. The registration view of the application provides a new user with an interface for creating a new account. The user must fill in and submit a form from the registration page. The data filled in the registration form include email address, first name, last name, and password. For an account to be created, the data filled in the form is required to pass the validation criteria. And for safety reasons, the user must confirm the password provided in an extra password field. Validations happen at both the client and server side. If the client-side validation fails, the wrong fields are highlighted for the user to correct. If the server-side validation fails, the server gives a response on where the problem is likely to have occurred. The user registration service is a RESTful service; the new account details are sent in the HTTP POST method body and after validation, the server responds with a representative message.

By default, a new user is expected to provide an email address, first name, last name, and a password. The email address and password are used as login credentials. The email address serves as a username to uniquely identify a registered user. A user successfully logs into the application by providing the correct email address and password. On successful login, the user is redirected to the application dashboard. The dashboard acts a launch pad from where a user can access different services in the application. Also, from the dashboard, the user can create a profile by providing additional information besides those already provided during registration.

6.3 Experiment management service

This service allows users to create their own experiments in the application. The service can be accessed by anyone who has an active account. The experiment is the main entity of the application. Most of all other services are anchored on the experiment entity. Creating an experiment requires the experiment name and a description of what the experiment is about. Validations are done to ascertain these fields have been provided by the user. Logically, an experiment should not override another one and can contain zero or more specimens. At the initial state, an experiment contains zero specimens.

6.4 Specimen management service

This service allows users to create specimens within an experiment. To avoid conflict during data collection, each

Figure 10: User registration workflow

Figure 11: Generating QR codes for specimens

specimen belongs to a single experiment. A single specimen cannot belong to multiple experiments. To create a specimen, the parent experiment has to be selected first followed by the number of specimens to be created. Each specimen is assigned a unique identifier. This unique identifier is encoded into a QR code and stored in a database. The corresponding QR code image is displayed for the user to print and use to physically label the specimen. The specimen identifier is linked to the data collection form for each experiment. The workflow for generating the QR codes for each specimen is as shown in Figure 11. The QR codes are generated at once for the number of specimens specified at creation time. The length of time required depends on the number of specimens. It increases with a higher number of specimens to be created. This is limiting especially in terms of computational resources e.g., processor speed and memory.

6.5 Data management service

The collected data is stored in a database. The user can get basic analytics in terms of counts, maximum, minimum, ranges and averages. However, if the elaborate analysis is needed, then a third-party application is required. The application allows the user to export data into a CSV file that can be fed into third-party tools for elaborate processing and analysis. Meta information is automatically collected to help filter the data based on the collection time and the owner.

6.6 Notification service

This is a service through which the application can notify users of the occurrence of various events generated by the other application services. The current implementation of the application utilises a counter to indicate to the user any

Figure 12: Creating a dynamic data collection form

Figure 13: Dynamic form fields

change in the number of datasets per specimen (Figure 9). Adding a new dataset increments the counter by one and its effect is reactively displayed on the user interface. Similarly, deleting a dataset decrements the counter by one. The user can use this counter to monitor the current number of datasets collected per specimen. This number is useful in determining whether the sample threshold for data analysis has been achieved or not.

6.7 Analytics and visualisation service

This service allows users to view data generated in a more user-friendly manner in the form of charts and aggregate values. The service can be accessed through the application analytics view and the reports view. The reports view shows the visualisations while the analytics view shows the aggregate values. The analytics view is limited to basic statistical features. Advanced statistics are left for future work as at the time of writing this paper.

6.8 Dynamic data collection forms

Initially, the data collection form was implemented statically. This proved difficult to handle different user requirements. Each user collects different kinds of data from an

experiment. Also, at the specimen level, different datasets are collected. Having a static form, therefore, proved difficult to handle each user's data collection requirements. To address this challenge and widen the user scope, we proposed to have the user create the data collection form at the point of adding specimens to an experiment. To create the form, the user only needs to specify the form fields. To demonstrate how this works, we take a simple hypothetical example of a researcher collecting data for specimens in a growth chamber and wishes to collect details on callus weight, colour of callus, number of shoots, and the number of roots observed in an experiment. The required specimen details potentially show the form fields. All we need is then to drag, drop, and provide the required validations for these fields. The form fields can be found on the ToolBox section labeled (B) in Figure 13. We can preview the form before submitting it for generation (Section (A) on Figure 13). Submission prompts the server to generate a schema in the backend to hold the data collected from the form fields.

The workflow for generating the data collection form is as illustrated in Figure 12. The detailed illustration of the implementation for creating a new form is shown in Figure 14. The user is provided with a drag and drop feature for specifying new form fields (Figure 13). Validations for the form fields are required to be explicitly specified by

Figure 14: Specifying new form fields

Table 1: Loading time from different locations

Location	Load time	First byte	Start render
London, UK	3.825s	0.681s	3.966s
New York, USA	3.894s	0.903s	4.100s
Sao Paulo, Brazil	5.472s	0.784s	5.533s
Rose Hill, Mauritius	3.489s	1.205s	3.200s
Sydney, Australia	6.637s	1.514s	6.700s
Istanbul, Turkey	3.396s	0.735s	3.500s

the user, otherwise, generic form validations are assumed (Section (C) on Figure 13). Once the fields are validated, the form is dispatched and the corresponding schema created for storing the data received from the form.

7 Performance analysis

This section presents the performance analysis of the web application. The analysis measures the simulated load time from different locations. Load time is an important aspect for measuring the performance of web applications [7]. This influences the user experience on the application. As depicted in the literature, a user leaves the application if the application load time is over 2 – 2.5 s. The specification of the server and distance between the testing location and the application hosting server highly influence the performance of the application.

The application was deployed on a Linux virtual server. The server had 8 CPU cores with 12GB memory and 500GB of hard disk space. Ubuntu Desktop 16.04.6 LTS was used as the host operating system. These features were selected to ensure no constraint on computational resources during the running of the application on the server-side. The performance tests were done with the help of Web-

Figure 15: Distribution of HTTP requests

Figure 16: Bytes dimension

PageTest¹, which is an online tool that allows complex load testing of web applications from different locations around the world. The test locations were chosen based on the existing literature [7]. The test results are presented on Table 1. The load time column represents the total time spent by the browser to load and render the page. The load time is not satisfactory because of the distance between the simulated test location and the application host. Also, the operating system deployed supports normal desktop applications. The load time is longer on first access but reduces on subsequent accesses. Given that the first Byte is received after 0.7 – 1.5s, the application still offers good performance.

The distribution of the number of HTTP requests on file types is as shown in Figure 15. Most files fetched are JavaScript (54.5%) followed by CSS (27.3%). Regarding the dimension (Figure 16) of the fetched files, the order is maintained with JavaScript at 85.7% followed by CSS at 14.1%.

¹ <https://www.webpagetest.org>

Figure 17: Comparison of manual and automated label generation

On label generation, we performed a comparison between the manual and automated method. We sampled 21 different users who manually generated specimen labels. We recorded the time it took to generate labels for three different specimen samples per user. We also generated labels for the samples using our prototype application and recorded the time. In order to compare the two approaches, we plotted the average time it took the sampled users and our application to generate the labels. The comparison is as shown on Figure 17. From Figure 17, generating specimen labels manually takes considerable time compared to the automated method.

8 Discussion

The choice of mobile device-based data collection using QR codes was motivated by the high adoption rates for mobile devices in developing countries [8]. QR codes are used to verify the identity of a specimen [9] and are less costly compared to other specimen labelling techniques. They provide researchers with an alternative to manually generated labels. The end user has access to the electronically collected data, can generate preliminary analysis reports, and start observing trends that may require human intervention. Access to mobile internet connectivity is still costly in the developing countries, though projections are that these costs will reduce drastically in the future. This paves way and gives the motivation for designing and developing technologies that rely on mobile internet connec-

tivity. Also, reducing extra steps required in paper-based data collection approaches serves as a good incentive to explore specimen labelling techniques. This approach adds to the already existing electronic data collection methods. The use of QR codes makes data collection easier compared to manually generated labels.

QR codes have their own weaknesses which make them not ideal for specimen labelling considerations. Though they have been used widely, their black and white appearance is not visually pleasing [10]. They have low human visualisation effect [11]. Solving these challenges has led to the development of different variants of the QR codes like the colour and logo QR codes. For binary data, QR codes have a maximum data storage of 2,953 bytes. For mobile applications, QR codes are generated between versions 1 and 10 for accurate data decoding with a maximum data storage of 119 and 271 bytes for 7% and 30% error correction recovery respectively [12]. However, the wide spread usage of mobile devices and the development of applications for generating and reading these codes, makes them potentially feasible for use in specimen labelling and identification. Mobile applications greatly improve the data collection process. Digital specimen labels permits linking to respective data collection forms which are invoked when the labels are scanned successfully. Developing requisite software applications requires experienced programmers who are often expensive to engage.

On improving the data collection process, *WebLog* provides custom user defined forms. The forms are designed on creation of experiment specimens. Once generated, the data collection form can be used by multiple enumerators. For large sample sizes, this reduces the time it takes to collect and aggregate research data. Once keyed in the data is converted into electronic format that makes it is to store and share among different users and also between different software applications. Also, using mobile phones, makes it more flexible to collect data and remotely send it to the server. For the experiment manager, it becomes easy to follow and regulate the data collection process for experiments with large sample sizes. While the data comes to a central server, field enumerators can be deployed in remote and distributed locations. Preliminary analysis can already provide some insights to the experiment manager on the quality of the data being collected. Such insights take longer to get with manual methods.

On improving specimen tagging and labelling, *WebLog* automatically generates each specimen identifiers. The identifiers are then converted into QR codes. Generating these identifiers and associated labels manually takes considerable amount of time (Figure 17). The identifier is automatically linked to the associated data collection form.

Table 2: Comparison of different labelling techniques

Technique	Data capacity	Durability to damage	Cost	Strengths
1D barcodes	Limited	Less durable; low error correction margin	Cheap	Cheap to generate; for communicating small amounts of data
Active RFID	Large	Less durable; requires power source	Expensive	Requires internal power source; good for tracking items
Passive RFID	Large	Durable in harsh conditions	Expensive	Does not require battery; good for tracking items
QR codes	Large	Reading possible with up to 30% damage; error correction margin ranges from 7-30%	Cheap	Easy to use; cheap to generate

The data collection form can readily be invoked when the specimen label is successfully scanned. The scanning (reading) of the specimen labels is done using the camera sensors that are readily available in most smartphones. The specimen labels can be reprinted if damaged. Table 2 shows the comparison of the different labelling and tagging techniques. Each of the techniques is well suited for the application scenario in context. For example, RFID tags are most suited for tracking large quantities of items, especially on transit. 1D barcodes have limited data capacity compared to RFID tags and QR codes.

We can summarise the challenges and benefits of using QR codes to label objects as follows. First, there are indirect costs associated with printing. Printers, special paper, and ink are requisite before the QR codes can be used in production environments. Depending on the volume of labels required, the costs can increase and decrease appropriately. Once printed, the codes can be damaged if subjected to harsh and extreme environmental conditions. Despite these challenges, the associated benefits of QR codes are more relevant and make them ideal for use in production environments. They are free to use i.e., there are no license costs associated to the use of these codes in production environments. There are numerous software libraries that are freely available for use to generate these codes. The codes are precise and help in reducing information overload to the end user. The end user needs to scan the code and the encoded information is decoded and displayed for viewing. In some cases, the codes can be used to reduce memorising long web addresses or even long specimen labels. QR codes are fairly easy to integrate with other technologies even though some technical skills are necessary.

For mobile applications, the associated challenges and benefits include, but are not limited to the following. Mobile phones are expensive, especially in developing countries. Despite the cost, the rate of adoption of smartphones is high. This makes them readily available and the camera sensors that are embedded in them are useful in reading QR codes. The phones use cellular networks that make sending data to remote servers easier. Combining QR codes and mobile applications provides a good mixture to label specimens and also collect data on each object.

In summary, to improve object labelling and data collection, we combine QR codes and data collection forms in our prototype application. The application integrates QR code generation and reading and links each code to the associated data collection form. The form is automatically invoked once the code is scanned successfully.

8.1 Application areas for QR codes

Different kinds of QR codes are described in the literature [13–15]. The application areas for QR codes transcend beyond their original application in motor vehicle manufacturing. The application areas range from their originally intended tracking of parts in motor vehicle manufacturing to mobile content distribution in electronic learning environments. These are not limited to but include shopping, marketing and advertisement to reinforce brand presence [16–18], transport and logistics, content accessibility in teaching (education) [4, 18, 19], architecture, tourism, augmented reality, mobile content distribution, gaming authentication and security sensitive operations [20, 21], software development for device pairing, in research as a means for linking physical spaces with digital information [20], and mobile payments as user interface enablers to auto-fill forms for mobile payment applications [22]. They have also been used in educational libraries to label reading resources [23, 24]. Some academic journals add a QR code on the cover of the journal, or on the title page or at the end of each article. This helps readers to go directly to the associated (linked) material e.g., a website, an address or other contact information, a video, an animation, or other forms of web content.

Other useful applications of QR codes include user authentication into web platforms and Wi-Fi networks. To log into a web platform, the code is presented on the login interface. A registered user is required to scan the code with a verified mobile device. User authentication is performed by the mobile device which contacts a remote server [25]. In this case, the mobile device acts as a helper device for authenticating the user. Google tested such a login method in January 2012 [26]. Similarly, to access a Wi-Fi network, the access credentials are encrypted into the QR code. The user scans the code and joins the network without manu-

ally entering the access credentials [27]. This technique is only valid for static SSID passwords.

Quick response codes are a quick and easy medium to take a user from the physical world to the virtual world [16]. Even though these codes are widely used in different areas, still there is need to explore more possible application areas and document the experiences to improve on the adoption process. This is particularly useful in new and promising areas that have not been explored before.

9 Related work

Since our work cuts across two fields of specimen labelling and data collection, we present the related work in two-fold. First, we present related work on the use and application of QR codes (Section 9.1). And secondly, we present related work on data collection (Section 9.2).

9.1 QR codes

Different studies have presented varied contextual scenarios for the use and application of QR codes. The codes have been used in object searching and recognition [28]. In governance, the codes have been used to increase the scope of digital services [29]. The codes have also been used in product packaging and conducting market campaign promotions. Product stores are using these codes to make transaction payment easier [10, 30]. In information security, QR codes have been used to develop schemes for sharing secret information among participants [31, 32]. The codes have also been used to hide information and protect confidential data [33]. Only authorised users can extract the hidden confidential data. The codes have been applied in photon counting encryption [34]. The counting is done via an optical security method for object authentication.

More applications of QR codes in security are seen in enhancing the security of optical mark recognition systems used in assessing examinations [35]. The details of the examinee and number of filled questions are embedded in an encrypted QR code which is then printed on the optical mark recognition sheet. Also, in fighting counterfeits, holographic structures have been combined with QR codes to form aesthetic holographic markers that conceal secret messages about products [36]. In information holding and sharing, the codes have been used in hospitals to share vast amounts of information [37]. The codes are placed at strategic points where they can be scanned easily with smartphones or tablets. The codes have also been

used in channel capacity analysis [12]. The study done by Melgar *et al.* [12] uses mutual information theory to map code symbols into the output sequence of each channel. The codes have also been used in exchanging data and text files through mobile devices in the absence of existing wireless (Wi-Fi and cellular networks) technologies in personal area networks [38]. In this application, colour QR codes send data by manipulating colours into multiplexed QR codes. Colour QR codes are able to hold and transmit more data at a higher throughput compared to black and white QR codes. The QR codes have also been used in augmented reality (AR) for tagging virtual objects [39]. With the help of a marker, a camera can efficiently recognise the tag of the virtual object while capturing the real scene image.

Object labelling has been widely embraced in sales and marketing [16, 30, 40]. However, it is yet to find more applications in research environments e.g., collecting data from experiment setups.

9.2 Data collection applications

Several tools and applications that support data collection exist in literature. The applications are focused on collecting data for baseline surveys, impact evaluation and informing the policy-making process. The applications are either proprietary or open source. The commonly used open source tools for data collection include OpenXdata, Open Data Kit (ODK), GeoDK and CommCare.

OpenXdata is an open source data collection tool that was initially designed to run on Java (J2ME) enabled phones. The tool has three main blocks i.e., "design", "collect" and "manage". End users can design electronic forms for data collection on mobile phones or web browsers. The collected data is uploaded directly to an electronic database [42, 43]. Though it functions on low-cost phones, the phones must be Java enabled [41].

ODK is a tool for data collection, aggregation and visualisation [44, 45]. It supports standard questions, voice, image, and video recordings. The tool runs on Android phones. It was designed to be modular and interoperable. Though, popular in resource-constrained environments, it is SMS based. The cost of sending and receiving SMS is considerable and can be high in developing regions given a large number of messages to be sent and received.

GeoODK² is a variant of ODK that supports collection, analysis, visualisation, and manipulation of geo-

² <http://geoodk.com>

referenced data. It supports mobile and web application. The mobile application was derived from ODK with added features for online and offline mapping and developer options for deploying surveys with the app.

CommCare runs on Java-enabled phones. The application was designed for data collection in resource-constrained environments [46, 47]. Though not well documented in the literature, there are more tools for data collection. These are mostly proprietary and include EpiCollect, Formhub³, KoboToolBox⁴, Device Magic⁵, GoFormz⁶, Fulcrum⁷, Formitize⁸, iFormBuilder⁹, Push Forms¹⁰, Nexticity¹¹, Magpi¹², TrackVia¹³, and Sensr.

EpiCollect provides a web application for the generation of data collection forms [48, 49]. Data can be collected using multiple Android devices or iPhone devices, synchronised and viewed centrally on the project website or directly on the phones. Sensr provides non-technical users with visual drag-and-drop programming environment to create data collection forms [50]. However, it supports a limited set of data types.

Though these tools exist for use, they need contextual configuration to address specific needs. They can only handle discrete data that makes them acceptable for surveys. To the best of our knowledge, none of these tools links object labels to data collection forms. None of the existing tools incorporate QR codes as a way of identifying and labelling objects. Also, most of the tools are proprietary and have an implication on the budgetary constraints already experienced in research projects. These constraints force researchers to build their own tools which takes considerable time.

10 Summary and lessons learnt

In this paper, we present *WebLog*, a prototype application to help in bio-science data collection. The application generates QR codes to replace manual labels for use in labelling specimens.

Generating specimen labels this way saves on time and effort required for the manual labels. The QR codes are scanned using a mobile phone to invoke the data collection form. The data is converted into electronic format and relayed to the server immediately. This provides a better alternative compared to paper-based data collection methods and makes the process easy. The approach is less prone to errors and saves on the overall time spent in data collection and management. This further reduces the cost of collecting data from large sample sizes in terms of labor and materials.

Though this approach simplifies specimen labelling (identification) and data collection, it requires high technical expertise to develop the requisite software applications. Building software applications for novices is a daunting task. Engaging experienced programmers is expensive. Novice users need to be empowered to build their own applications. Bio-science research is often done in remote and distributed environments that require Internet connectivity to relay collected data to the servers. Also, the number of specimen samples used in the experiments is small compared to commercial applications where the impact of manually generating specimen labels is apparent.

Bio-science research involves processes that have to be done manually and some can be automated. Also, a key aspect in the integration of technology to make research processes more sustainable is to involve different actors from multiple disciplines to tap into the available knowledge pool, reconcile problems and solution options. The solutions developed need to be applied and validated contextually. We suggest multidisciplinary, community-based, interactive, or participatory research approaches as appropriate means to meet the requirements for contextual solutions, advancing, and application of technology.

11 Conclusion and future work

In this paper, we have presented a prototype application to help researchers label specimens using QR codes and collect data from those specimens in a more automated way. We have proposed and implemented the application to link specimen labels to data collection forms that run on a mobile phone and web interface. The application has a frontend and backend; the two communicate via RESTful services.

We have also described the application data model and services. We have highlighted the operation of each service and integration to external services. In our future work, we will refine the application and implement it using

³ <https://formhub.org>

⁴ <https://www.kobotoolbox.org>

⁵ <https://www.devicemagic.com>

⁶ <https://www.goformz.com>

⁷ <https://www.fulcrumapp.com>

⁸ <http://www.formitize.com>

⁹ <https://www.zerionsoftware.com/iforbuilder>

¹⁰ <http://www.getpushforms.com>

¹¹ <https://nexticity.com>

¹² <https://home.magpi.com>

¹³ <https://www.trackvia.com>

flow-based programming and perform a comparative analysis. We will also include 'in-application' advanced data analysis features.

Acknowledgement: We thank the anonymous referees for their insightful comments. We also thank the Kenya Education Network (KENET) for providing the hosting services for the prototype application. Any opinion, findings, and conclusions or recommendations expressed in this paper do not necessarily reflect the views of the affiliate institutions.

References

- [1] Kumar N., In vitro Plant Propagation: A Review, *Journal of Forest and Environmental Science*, 2011, 27(2), 61–72
- [2] Espinosa-Leal C. A., Puente-Garza C. A., García-Lara S., In vitro plant tissue culture: means for production of biological active compounds, *Planta*, 2018, 248(1), 1–18, doi:10.1007/s00425-018-2910-1
- [3] Jayaraman P. P., Yavari A., Georgakopoulos D., Morshed A., Zaslavsky A., Internet of things platform for smart farming: Experiences and lessons learnt, *Sensors*, 2016, 16(11), 1–17
- [4] Duarte M., Baptista A., Pinto G., Exploring the use of Quick Response Codes: accessing videos in the laboratory, *TEEM Conference 2015*, 2015, 99–103, doi:10.1145/2808580.2808596
- [5] Georgakopoulos D., Jayaraman P. P., Internet of things: from internet scale sensing to smart services, *Computing*, 2016, 98(10), 1041–1058, doi:10.1007/s00607-016-0510-0
- [6] Rouillard J., Contextual QR codes, *Proceedings of the 3rd International Multi-Conference on Computing in the Global Information Technology (ICCGI 2008) in Conjunction with CompP2P 2008: The 1st Int. Workshop on Computational P2P Networks: Theory and Practice (Athens, Greece)*, 2008, 50–55, doi:10.1109/ICCGI.2008.25
- [7] Colezea M., Musat G., Pop F., Negru C., Dumitrascu A., Mocanu M., CLUEFARM: Integrated web-service platform for smart farms, *Computers and Electronics in Agriculture*, 2018, 154, 134–154, doi:10.1016/j.compag.2018.08.015
- [8] Ganesan M., Prashant S., Jhunjhunwala A., A Review on Challenges in Implementing Mobile Phone Based Data Collection in Developing Countries, *Journal of Health Informatics in Developing Countries*, 2015, 6(1), 366–374, doi:10.1523/ENEURO.0029-14.2015
- [9] Nandhini S., Performance Evaluation of Embedded Color QR Codes on Logos, *2017 Third International Conference On Science Technology Engineering and Management (Chennai, India)*, 2017, 1009–1014
- [10] Wang Y., Sun C., Kuan P., Lu C., Wang H., Secured Graphic QR Code with Infrared Watermark, *2018 IEEE International Conference on Applied System Invention (Chiba)*, 2018, 690–693, doi:10.1109/ICASI.2018.8394351
- [11] Lay K., Chen T., Visual QR Codes with Lossless Picture Embedding, *2018 3rd International Conference on Intelligent Green Building and Smart Grid (Yi-Lan, Taiwan)*, 2018, 1–4, doi:10.1109/IGBSG.2018.8393546
- [12] Melgar M. E. V., Santander L. M., Channel Capacity Analysis of 2D Barcodes : QR Code and CQR Code-5, *2016 IEEE Colombian Conference on Communications and Computing (Cartagena, Colombia)*, 2016, 1–5, doi:10.1109/ColComCon.2016.7516376
- [13] Chu H., Chang C., Lee R., Mitra N. J., Halftone QR codes, *ACM Transactions on Graphics*, 2013, 32(6), 1–8, doi:10.1145/2508363.2508408
- [14] Nurwono K. A. H., Kosala R., Color Quick Response Code for Mobile Content Distribution, *MoMM2009*, 2009, 267–271
- [15] Pearson J., Robinson S., Jones M., Nanavati A., Rajput N., ACQR: Acoustic Quick Response Codes for Content Sharing on Low End Phones with No Internet Connectivity, *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services - MobileHCI '13 (Munich, Germany)*, 2013, 308–317, doi:10.1145/2493190.2493195
- [16] Asare I. T., Asare D., The Effective Use of Quick Response (QR) Code as a Marketing Tool, *International Journal of Education and Social Science*, 2015, 2(12), 67–73
- [17] Cata T., Patel P. S., Sakaguchi T., QR Code: A New Opportunity for Effective Mobile Marketing, *Journal of Mobile Technologies, Knowledge and Society*, 2013, 2013(2013), 1–7, doi:10.5171/2013
- [18] Law R., Interactive lectures using quick response codes, *Proceedings of the 18th ACM conference on Innovation and technology in computer science education - ITiCSE '13*, 2013, 317, doi:10.1145/2462476.2465576
- [19] Durak G., Ozkeskin E. E., Ataizi M., QR Codes in Education and Communication, *Turkish Online Journal of Distance Education*, 2016, 17(2), 42–58
- [20] Lerner A., Saxena A., Ouimet K., Turley B., Vance A., Kohno T., Roesner F., Analyzing the Use of Quick Response Codes in the Wild, *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services - MobiSys '15 (Florence, Italy)*, 2015, 359–374, doi:10.1145/2742647.2742650
- [21] Narayanan A. S., QR Codes and Security Solutions, *International Journal of Computer Science and Telecommunications*, 2012, 3(7), 69–72
- [22] De P., Dey K., Mankar V., Mukherjee S., An Assessment of QR Code as a User Interface Enabler for Mobile Payment Apps on Smartphones, *7th International Conference on HCI (Guwahati)*, 2015, 81–84, doi:10.1145/2835966.2835977
- [23] Barker K. R., Attridge E., Bennett J., Hiserman T., Horne A. S., Moody D., Ramsey E. C., Son I. K., Vaughn P., The implementation of embedded quick response codes into library resources to improve service delivery, *Journal of Medical Library Association*, 2012, 100(1), 68–71
- [24] Coleman J., QR Codes: What Are They and Why Should You Care?, *Kansas Library Association College and University Libraries Section Proceedings*, 2011, 1(1), 16–23
- [25] Adhatrao K., Gaykar A., Jha R., Honrao V., A Secure Method for Signing in Using Quick Response Codes with Mobile Authentication, *International Journal of Student Research in Technology and Management*, 2013, 1(1), 1–11
- [26] Naraine R., Google testing login authentication via QR codes, 2012, <https://www.zdnet.com/article/google-testing-login-authentication-via-qr-codes>, Retrieved[22-03-2019]
- [27] Bhattacharjee M., Share your Wi-Fi SSID & Password using a QR Code, 2015, <http://www.technostarry.com/share-wi-fi-details-using-qr-code>, Retrieved[22-03-2019]

- [28] Xue Y., Tian G., Li R., Jiang H., A new object search and recognition method based on artificial object mark in complex indoor environment, 2010 8th World Congress on Intelligent Control and Automation (Jinan, China), 2010, 6648–6653, doi:10.1109/WCICA.2010.5554161
- [29] Lorenzi D., Shafiq B., Vaidya J., Nabi G., Chun S., Atluri V., Using QR codes for enhancing the scope of digital government services, Proceedings of the 13th Annual International Conference on Digital Government Research - dg.o '12 (College Park, MD, USA), 2012, 21–29, doi:10.1145/2307729.2307734
- [30] Liu Y., Yang B., Yang J., Bar Code Recognition in Complex Scenes by Camera Phones, Fourth International Conference on Natural Computation (Jinan, China), 2008, 462–466, doi:10.1109/ICNC.2008.288
- [31] Tan L., Liu K., Yan X., Wan S., Chen J., Chang C., Visual Secret Sharing Scheme for Color QR Code, 2018 IEEE 3rd International Conference on Image, Vision and Computing (Chongqing, China), 2018, 961–965, doi:10.1109/ICIVC.2018.8492724
- [32] Wan S., Lu Y., Yan X., Liu L., Visual Secret Sharing Scheme With (k, n) Threshold Based on QR Codes, 2016 12th International Conference on Mobile Ad-Hoc and Sensor Networks (Hefei, China), 2016, 374–379, doi:10.1109/MSN.2016.068
- [33] Lin P., Chen Y., QR code steganography with secret payload enhancement, 2016 IEEE International Conference on Multimedia & Expo Workshops (Seattle, WA, USA), 2016, 1–5, doi:10.1109/ICMEW.2016.7574744
- [34] Markman A., Javidi B., Tehranipoor M., Photon-Counting Security Tagging and Verification Using Optically Encoded QR Codes, IEEE Photonics Journal, 2014, 6(1), 1–9, doi:10.1109/JPHOT.2013.2294625
- [35] Tiwari S., Sahu S., A Novel Approach for the Detection of OMR Sheet Tampering Using Encrypted QR Code, 2014 IEEE International Conference on Computational Intelligence and Computing Research (Coimbatore, India), 2014, 1–5, doi:10.1109/ICCIC.2014.7238430
- [36] Włodarczyk K. L., Ardrón M., Weston N. J., Hand D. P., Holographic watermarks and steganographic markings for combating the counterfeiting practices of high-value metal products, Journal of Materials Processing Tech., 2019, 264, 328–335, doi:10.1016/j.jmatprotec.2018.09.020
- [37] Uzun V., QR-code based Hospital Systems for Healthcare in Turkey, 2016 IEEE 40th Annual Computer Software and Applications Conference (Atlanta, GA, USA), 2016, 71–76, doi:10.1109/COMPSAC.2016.173
- [38] Toh S. R., Goh W., Yeo C. K., Data Exchange via Multiplexed Color QR Codes on Mobile Devices, 2016 Wireless Telecommunications Symposium (London, UK), 2016, 1–6, doi:10.1109/WTS.2016.7482035
- [39] Lin P., Teng C., Chen Y., Diverse Augmented Reality Exhibitions for Differential Users Based upon Private Quick Response Code, 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (Hong Kong, China), 2015, 1121–1125, doi:10.1109/APSIPA.2015.7415445
- [40] Wittman T., Lost in the supermarket: Decoding blurry barcodes, SIAM News, 2004, 37(7), 1–2
- [41] Mancini F., Gejibo S., Mughal K. A., Valvik R. A. B., Klungsøyr J., Secure mobile data collection systems for low-budget settings, Proceedings - 2012 7th International Conference on Availability, Reliability and Security (ARES 2012), 2012, 196–205, doi:10.1109/ARES.2012.23
- [42] Mancini F., Mughal K. A., Gejibo S. H., Klungsøyr J., Adding security to mobile data collection, 2011 IEEE 13th International Conference on e-Health Networking, Applications and Services, 2011, 86–89, doi:10.1109/HEALTH.2011.6026793
- [43] Shrestha P. L., Usability and desirability of OpenXdata – an electronic data capture tool for data collection in health research, MSc thesis, University of Bergen, Norway, 2014
- [44] Brunette W., Sundt M., Dell N., Chaudhri R., Breit N., Borriello G., Open data kit 2.0: Expanding and Refining Information Services for Developing Regions, Proceedings of the 14th Workshop on Mobile Computing Systems and Applications (Jekyll Island, Georgia, USA), 2013, 10:1–10:6, doi:10.1145/2444776.2444790
- [45] Hartung C., Anokwa Y., Brunette W., Lerer A., Tseng C., Borriello G., Open Data Kit: Tools to Build Information Services for Developing Regions, Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development (London, United Kingdom), 2010, 18:1–18:12
- [46] Dayalu R., Wacksman J., Chen M., Loudon M., Lesh N., Adoption and Continued use of Mobile Technology: An Analysis of CommCare Data, Procedia Engineering, 2015, 107, 247–254, doi:10.1016/j.proeng.2015.06.079
- [47] Medhi I., Jain M., Tewari A., Bhavsar M., Matheke-Fischer M., Cutrell E., Combating rural child malnutrition through inexpensive mobile phones, Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design (Copenhagen, Denmark), 2012, 635–644, doi:10.1145/2399016.2399113
- [48] Aanensen D. M., Huntley D. M., Feil E. J., Al-Owaini F., Spratt B. G., EpiCollect: Linking smartphones to web applications for epidemiology, ecology and community data collection, PLoS ONE, 2009, 4(9), 1–7, doi:10.1371/journal.pone.0006968
- [49] Aanensen D. M., Huntley D. M., Menegazzo M., Powell C. I., Spratt B. G., EpiCollect+: linking smartphones to web applications for complex data collection projects, F1000Research, 2014, 3, 1–15, doi:10.12688/f1000research.4702.1
- [50] Kim S., Mankoff J., Paulos Eric., Sensr: Evaluating a Flexible Framework for Authoring Mobile Data-Collection Tools for Citizen Science, Proceedings of the 2013 conference on Computer supported cooperative work, 2013, 2(18), 1453–1462, doi:10.1145/2441776.2441940