Enhancing Effective Use of Research Methods at the International Potato Centre (CIP): Case Study "Sweet Potato Action for Security and Health in Africa (SASHA) Project"

Januaris Nthenge Mbatha

A dissertation submitted in partial fulfillment for the degree of Master of Science in Research Methods in the Jomo Kenyatta University of Agriculture and Technology

2012

Declaration

This dissertation is my original work and has not been presented for a degree in any other university.

Signature…… Date……

Januaris Nthenge Mbatha

This dissertation has been submitted for examination with our approval as University supervisors.

1. Signature…… Date……..

Dr. Aggrey Bernard Nyende,
JKUAT, Kenya

2. Signature…….. Date……

Dr. Julius Sindi Kirimi
International Potato Centre (CIP)

Dedication

This dissertation is lovingly dedicated to my wife Nelima and our respective parents who have been our constant source of inspiration. They have given me the drive and discipline to tackle any task with enthusiasm and determination. Without their love and support this project would not have been made possible.

Acknowledgement

I am thankful to everyone who all supported me throughout my dissertation preparation. I am equally grateful to my supervisor Dr. A.B. Nyende who gave me professional support and guided me in different matters regarding this dissertation. He had been very kind and patient while suggesting me outlines of this dissertation and correcting my doubts. I thank him for his overall support.

I would like to thank my supervisor from CIP Dr. J.S. Kirimi who helped me a lot in gathering different information and guiding me from time to time in compiling this dissertation.

Last but not least, I would like to thank my colleagues in Research Methods class who gave me immense support through the moodle site, despite their busy schedules; they gave me different ideas in making this work unique.

Table of Contents

List of Tables

List of Figures

List of Appendices

Acronyms and Abbreviation

SSC	Statistical Service Centre
CIP	International Potato Centre
SASHA	Sweet potato Action for Security and Health in Africa
TNA	Training Needs Analysis
SUNY	State University of New York
CGIAR	Consultative Group on International Agricultural Research
JKUAT	Jomo Kenyatta University of Agriculture and Technology
M&E	Monitoring and Evaluation
SP	Sweet Potato
CSPRO	Census and Survey Processing System
SPHI	Sweet potato for Profit and Health Initiative

Abstract

The International Potato Center (CIP) was established in 1971 and seeks to reduce poverty and achieve food security on a sustained basis in developing countries through scientific research and related activities on potato, sweet potato, other root and tuber crops, and on the improved management of natural resources. The centre has been running a five year initiative, the Sweet potato Action for Security and Health in Africa, designed to improve food security and livelihood of poor families in sub-Saharan Africa by tapping the unexploited potential of sweet potato.

However, the program face several challenges in design, implementation, monitoring and evaluation of its projects which bordered on among others; lack of centralized project design and planning office which would be responsible for provision of guidelines and creation of a data base of projects progress, absence of data management support assistants to oversee the design of data collection instruments across the projects, capacity building of enumerators and data entry clerks, data collection, data organization, entry and analysis.

The objective of this study was to enhance effective use of research methods among researchers through training of researchers in data management practices, reinforce the data checking and organization before analysis and evaluation and reviewing of project proposals and literature on sweet potato seed systems.

Training followed the training cycle spanning training needs assessment whereby training activities, needs and goals were determined through administration of a structured questionnaire to nine data entry clerks and four researchers. The aim was to establish knowledge gaps among other areas of interest in use of CSPRO in data management. Training evaluation was done through administration of a structured questionnaire to establish the achievement of training objectives. Finally, an adoption survey was done as part of follow up to gauge the acceptance of the data management tool among researchers.

The role of data entry checks was demonstrated through analyzing the Tanzanian baseline survey data where two case scenarios were looked into. In one case data was entered into CSPRO without data entry checks and the second case the same data was entered with entry checks inbuilt before data entry. Two categories of

checks were used; checks before data entry which include range checks, automatic skips, numeric codes checks, missing and not applicable checks were used. The second category was checks after data entry; double data entry and exploratory data analysis which were used to compare the output from the two case scenarios.

A criterion to evaluate project proposals was developed and used to evaluate two proposals. It considered general components of the proposals as well as specific components, which were used to grade the proposals. Literature review was done on sweet potato seed systems. Three steps were adhered to in compilation of the sweet potato seed system paper; first, both published and unpublished material on sweet potato seed system were identified, second step entailed gathering the relevant information from the materials and finally writing up of the paper.

Results from the follow up study of the training showed that data from twelve projects were managed by researchers using CSPRO. The quality of the data output and limited data queries during analysis were some of the reasons behind majority recommendation of CSPRO software. However, the researchers identified some of the challenges to adoption of the software as resistance from scientists and lack of a functional data management unit among others.

Data entry checking was found to play a significant role in data analysis through increasing plausibility of the output and subsequent conclusions as evidenced by significant difference in output. At 5% both t test and chi square tests performed on two cases (with and without entry checks) gave a p value of <0.0001 which led to the conclusion that indeed there is a difference between output from checked and unchecked data.

The project proposals evaluated were ranked above 50% threshold on both the general and specific criterion with proposals reviewed scoring 100% on the specific criteria but <100% on the general criteria.

In order to enhance effectiveness of the research teams involved in these research projects it was therefore, recommended that CIP should design and promote training on data management, scientists to embrace data validation and organization and formulate a standardized criteria of evaluating and reviewing proposals and literature.

Chapter 1

1.0 Introduction

1.1Background information on host institution

The International Potato Center (CIP) was established in 1971 and seeks to reduce poverty and achieve food security on a sustained basis in developing countries through scientific research and related activities on potato, sweet potato, other root and tuber crops, and on the improved management of natural resources in the Andes and other mountain areas.

CIP headquarters are in La Molina, outside of Lima, Peru's capital, with regional offices in Indonesia, India, and Kenya. CIP is a member of the Alliance of the 15 centers of the Consultative Group on International Agricultural Research (CGIAR) and so receives its principal funding from the 58 governments, private foundations and international and regional organizations that constitute the CGIAR.

1.1.1 Core business

CIP works to rescue and conserve the biodiversity of the potato and other Andean tubers. CIP's gene bank holds almost 5000 varieties of potato, the largest collection in the world. The Center has distributed hundreds of thousands of samples of this germplasm to researchers worldwide. The organization studies the properties of the different varieties and uses them to breed new varieties of potato that can resist disease and grow well in developing countries to improve food supplies, earnings and health at the same time using integrated crop management to increase potato yields, as well as studying the role of potatoes in natural resource systems. Finally, CIP researches the links between agriculture and human health

The organization aims at contributing to reducing poverty and hunger; improving human health; developing resilient, sustainable rural and urban livelihood systems; and improving access to the benefits of new and appropriate knowledge and technologies through addressing these challenges by convening and conducting research and supporting partnerships on root and tuber crops and on natural

resources management in mountain systems and other less-favored areas where CIP can contribute to the achievement of healthy and sustainable human development (CIP 2010).

1.1.2CIP research

CIP research is geared towards reducing poverty and hunger while protecting the environment through sustainable use of potato, sweet potato and Andean roots and tubers. The center is dedicated to reducing poverty and hunger in developing countries, with research on potato, sweet potato and other roots and tubers, and management of natural resources in the Andes and other mountain areas.

1.1.3Natural resources management

Mountain ecosystems are found on every continent and sustain an estimated 10% of the world's population. In addition, billions of people living in the lowlands depend on these ecosystems for food and other resources such as water, raw materials and energy. These areas are also important sources of plant and animal diversity, both wild and domestic. In the past few decades, environmental changes and rapid increases in population densities in these mountain areas have increased problems for planning effective resource management strategies. Despite global recognition of the importance these areas, many mountain communities continue to live in poverty.

1.1.4 CIP programs

CIP’s program consists of six research divisions; impact enhancement, genetic resource conservation and utilization, germplasm enhancement and crop improvement, integrated crop management, production systems and environmental health and Agriculture. The organization works in partnership with three research institutions; CONDESAN, global mountain program and urban harvest Papa Andina all based in its headquarters in Lima, Peru, operating in coordination with four regional programs viz-a-viz; Latin America and the Pacific Sub-Saharan Africa South, West and Central Asia East, Southeast Asia and the Pacific.

1.2 SASHA projects in East Africa

The Sweet potato Action for Security and Health in Africa (SASHA) is a five-year project that serves as the foundation for the Sweet potato for Profit and Health

Initiative (SPHI) launched on 26 October 2010. SPHI aims to reduce child malnutrition and improve smallholder crop incomes in 10 million African families by 2020 through the effective production and expanded use of sweet potato. SASHA seeks to directly improve the food security of at least 155,000 SSA families by exploiting the untapped potential of sweet potato and to create the conditions for going-to scale (SASHA, 2010).

1.2.1 Projected impact of the project

The SASHA project seeks to directly improve the food security and livelihoods of at least 150,000 families in the Sub-Saharan Africa in five years and provide the evidence base for effective delivery systems to reach many more. Moreover, given widespread, informal farmer-to-farmer sharing of vines for planting, the number of direct plus indirect beneficiaries is likely to exceed 1 million families. As part of a broader, long-term, multi-donor Sweet potato for Profit and health Initiative, it is expected that the SASHA project will set the groundwork for improving the lives of 10 million Sub-Saharan households in 10 years.

It is a five-year initiative designed to improve the food security and livelihoods of poor families in Sub-Saharan Africa by exploiting the untapped potential of sweet potato. It will develop the essential capacities, products, and methods to reposition sweet potato in food economies of Sub-Saharan African countries to alleviate poverty and under nutrition.

1.2.2 Program components

1.2.2.1 Improved quality and range of available varieties

The focus of this component is on breeding a wide range of varieties with the combinations of traits suited to agro-ecological conditions and to consumer and producer demands. The aim is to create an integrated breeding system akin to the one that exists for cereal breeding, but focused on the producer and consumer preferences of resource-poor women and children.

1.2.2.2 Breeding weevil and resistant sweet potato varieties

This component draws on biotechnology to develop weevil-resistant sweet potato varieties for Sub-Saharan Africa. Swee potato weevils are the most important sweet

potato pest in the world – responsible for crop losses ranging from 60 to nearly 100% during pronounced drought. This situation may be critical during dry periods when sweet potato is sometimes the only food available. With climate change predictions of an expanding dry season in Sub-Saharan Africa, the urgency of developing resistance to weevils will likely intensify.

1.2.2.3 Developing sustainable sweet potato seed system

The access to and maintenance of quality planting material is a struggle for smallholder farmers. This component involves developing and testing strategies to ensure effective multiplication, dissemination, and exchange of disease-free vines from which new plants will be propagated. It involves strategies to more efficiently link farmers with public sector distribution programs and integrate those with for-profit nurseries. It will examine which strategies assure women the best access to vines and whether women are as successful as men at commercially-oriented vine production.

1.2.2.4 Proof of concept projects

This series of projects will examine broader institutional or market level issues affecting crop production, markets, potential market expansion (e.g., use of sweet potato as animal feed), and scalable approaches for improving nutrition with sweet potato. These projects will evaluate options that influence the capacity to scale up and achieve the outcomes on poverty and nutrition that are planned for the years following SASHA, in the longer, ten-year initiative.

1.2.2.5 Sweet potato support platforms and capacity strengthening

Three sub-regional support platforms, based in strong national research programs, will be established to provide the organizational and management structure for developing long-term breeding skills and capacity in Africa, for Africa. They will be located in each of three sub-regions: Ghana, for West Africa; Mozambique, for Southern Africa; and Uganda, for East and Central Africa.

1.3 Ongoing projects under SASHA

There are three projects which are already in progress under three different components in which research methods support is being given;

(i)Sweet potato value chain proof concept project in Rwanda. The project has got two objectives; to compare the farmer welfare outcomes from introduction of two models for producing sweet potato flour; and to test different models for inclusion of the poor and women in the development of new market chains for high value sweet potato products. These objectives are within the larger development question of how the poor, and particularly women, can capture the benefits of increased commercialization of staple crops such as sweet potato.

(ii)Proof-of-Concept Project Kenya agriculture-health linkages ; The project goal is to improve the health status of pregnant women and the nutritional status of children up to two years through an integrated orange fleshed sweet potato (OFSP) and health service delivery strategy through an existing health program (APHIA II) in selected districts of Western Kenya.

(iii)Going to scale with sweet potato vines distribution in Tanzania the major objectives of the study are; To provide farmers with quality seed of improved sweet potato varieties in a timely fashion; and To promote quality vine conservation among farmers , To stimulate increased demand for white and orange-fleshed sweet potato amongst rural and urban consumers, To assess the contribution of vine distribution and farmers' use of positive selection and vine conservation to raising productivity and improving food supply, To assess the cost effectiveness of the voucher based sweet potato seed system.

1.4 Statement of the problem

There are several challenges that face design, implementation and monitoring and evaluation of the SASHA projects which border but not limited to; lack of centralized project design and planning office which is responsible in provision of guidelines and creation of data base of projects progress, absence of data management support assistants to oversee the design of data collection instruments across the projects, capacity building/training of enumerators and data entry clerks, data collection, data organization and entry and analysis. Lack of research methods professionals under each scientist or attached to the projects is also a major challenge which continues to affect the timeliness of progress of projects as several consultative meetings are held to fine tune project designs and instruments.

The SASHA program has several projects which are complex both in design and implementation. Data sets generated are enormous and monitoring and evaluation instruments are equally complex. These factors compound the challenges of data management and analysis and overall project management at large. In the organization, there is absence of data management support officers relegating the role of data management, data analysis, review of project protocols and literature to research assistants who are discipline specific and only work under scientists who are engaged on specific projects. Capacity building is run and managed by scientists themselves. These challenges are cross cutting to all programs and units and monitoring and evaluation unit is not an exception consequently, affecting the efficiency and timeliness of the projects and surveys that precede the projects.

1.5 Justification

The challenges facing the organization in areas of data management, data analysis, review of project protocols and literature, planning and design of projects training of both enumerators on and data clerks, compromise on the timeliness, quality and the reliability of the research/ impact pathways of the program. To circumvent these problems there is a need to enhance effective use of research methods among scientists through research methods professional so that time spent on consultative meetings on design of project instruments is reduced, instances of messy data is reduced by both training of both data clerks and enumerators and finally to lift the burden of data management from specific scientists to a centralized research methods professional who would be providing backstopping services to researchers.

1.6 Objective

1.6.1 General objective

Enhance effective use of research methods among researchers in International Potato Centre (CIP)

1.6.2 Specific objectives

(i) Evaluate the role of training in adoption of data management practices among researchers

(ii)Assess the role of data entry checks and organization on data analysis for survey data

(iii) Evaluate and review project proposals and literature on sweet potato seed systems

Chapter 2

2.0 Literature Review

2.1 Data analysis

Data analysis refers to the systematic process of applying statistical or logical technique to describe, illustrate, condense, recap, and evaluate data (Sekaran, 2003). Various analytical procedures provide a way of drawing inductive inferences from data and distinguish the signal (phenomenon of interest) from noise (statistical fluctuations) present in data (Shamoo and Resnik, 2003). Thus in the most basic form data can be summarized as;

Data= pattern + residual (Statistical Service Centre, 2010)

The pattern part of the data is analyzed throughout the entire data collection phase (Sevenya and Robinson, 2004) and forms the basis of drawing conclusions. Pattern is the result of factors, such as the experimental treatments and other characteristics often determined by the layout. Identifying that part of the pattern that is due to the treatments is an important part of the analysis, because this relates directly to the objectives of the study. Residual on the other hand is the remaining unexplained variation. The residuals are evaluated to examine if they still contain some extra part of information that we could move into the pattern. It is also important to check the residuals in a model to see if they satisfy the assumptions of the analysis and if there are any oddities in the data (Statistical Service Centre, 2001).

Shephard, (2002) emphasizes the importance of accuracy in and appropriateness of analyses chosen for research findings. Integrity is of paramount concern because choice of improper statistical analysis distorts findings and gives flawed conclusions. Therefore, several issues come up for consideration when doing data analysis; one should have requisite skills for data analysis, follow acceptable norms for disciplines (Resnik, 2000), and provide accurate and honest analysis (North Illinois University, 2005).

2.2 Data management

To ensure current usability of data and long term preservation and access to data generated by research call for proper data management practices to be put in place. Various functions will be performed under data management in ongoing project at CIP (evaluation of sweet potato seed systems in East Africa) and subsequent documentation undertaken. The key areas that will be looked into include but not limited to data ownership, planning data flow, planning data collection, data entry and organization, statistical analysis, interpretation and write up, data storage and access, dissemination and feed back to data originators. The flow is outlined in figure 1.

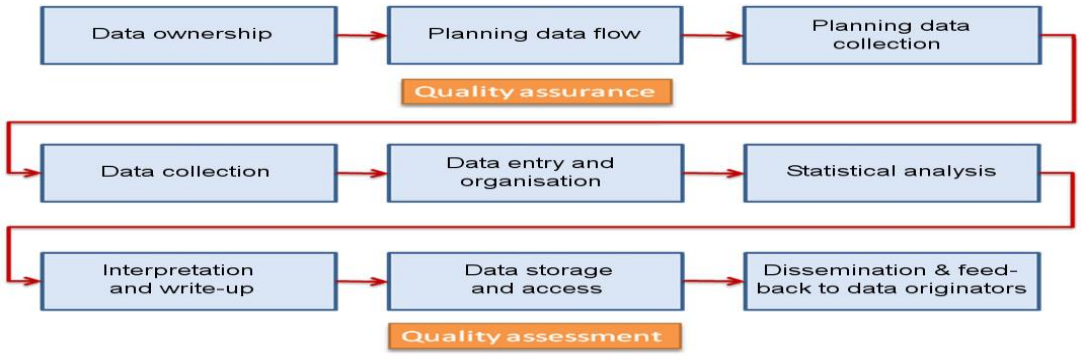


Figure 1: Research data cycle

Source: Statistical Service Centre, 2010

The data management component will be preceded by development of an elaborate data management plan which will consider the following components as suggested by structural reform group (2004);

i. Description of the project: e.g., purpose of the research, organizations and staff involved
ii. Description of the data to be collected: e.g., the nature and format of the data, how it will be collected, and overview of secondary data available on the topic
iii. Standards to be applied for formats, metadata, etc.
iv. Plans for short-term storage and data management: e.g., file formats, local storage and back up procedures, and security

v. Legal and ethical issues: e.g., intellectual property, confidentiality of study participants
vi. Access policies and provisions: i.e., how availability is made to others, any restrictions needed
vii. Provisions for long-term archiving and preservation: e.g., in a data archive
viii. Assigned data management responsibilities: i.e., which persons will actually be responsible for ensuring data management; how will compliance with this plan be monitored and ensured over time

2.3 Capacity building (Seminar or organize and offer a short training course)

Training refers to the acquisition of knowledge, skills and competencies as a result of the teaching of vocational or practical skills and knowledge that relate to specific useful competencies. Training done with key aim of upgrading, updating or simply to maintain skills in work environment is referred to as professional development (Cummings, 2005; Ghosh, 2000; Kubr, 1988). Training can be categorized as on the job training, which in this case the trainee continues to carry his/her usual duties using tools and equipment at the disposal or off job training where training is done away from work place.

Training to be effective should be preceded by training needs assessment (TNA). This helps in targeting the training in such that it is geared towards bridging some gap in an organization. Training is not a one stop gap measure and it follows a cycle (figure 2), which spans training needs analysis, design and development, implementation, evaluation and follow- up and report writing (JSI world Edu 2011).

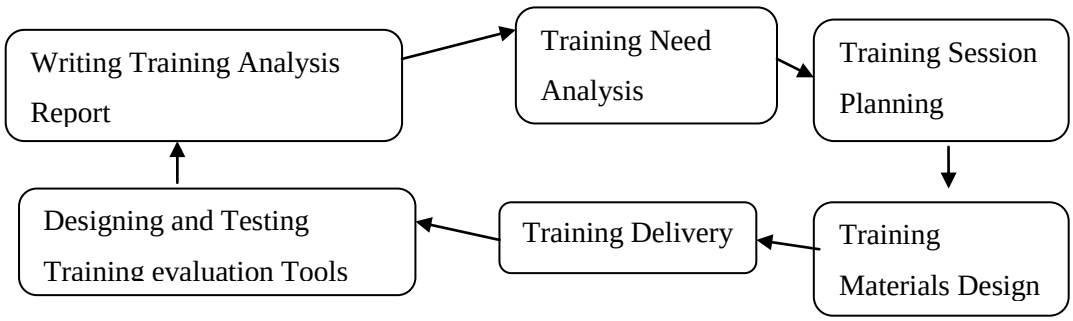


Figure 2: The Training cycle used

Source: Training cycle, guidelines and tool kit, (Angela, 2007)

In training needs analysis, trainers determine training/activity needs and goals. In design and development stage, trainers design a means to meet training activity/needs and goals. Implementation stage follows and here trainers carry out their training program or activity. In evaluation stage, assessment of the value of their training program is undertaken. Follow up stage of training the trainer ensures that the trainees successfully transfer the skills and knowledge, acquired during training activity to practice. Finally a report is compiled on the training analysis (JSI World Edu, 2011).

Through training needs analysis(TNA) the trainer should be able to identify an appropriate training methodology or a mix which is in tandem with the training objectives (i.e. if the training is intended to develop skills, change attitudes or even geared towards building knowledge). Available resources also work towards determining the appropriate training methodology (JSI world ED 2011). According to Info base Ltd, the methods of training are varied and can be divided into cognitive, behavioral methods and management development (Table 1). The cognitive category is more inclined towards theoretical training to the trainers. They provide the rules on how to do something, written or verbal information, demonstrates relationship among concepts.

The methods are associated with changes in knowledge and attitude by stimulating learning. On the other hand behavioral methods are geared towards giving practical training to the trainees and are best suited for skills development. A more forward looking method is management development and is more concerned with both the on the job training.

Table 1: Training methods

Cognitive Approach	Behavioral Approach	Management Development
Lectures	Games and simulations	Coaching
Demonstrations	Behavioral modeling	Mentoring
Discussion	Business games	Job rotation
Computer Based training	Case studies	Job instruction techniques (JIT)
	Equipment simulations	
	In basket techniques	
	Role plays	

Source: (Info base Ltd, 2007)

2.4 Written review of reports, proposals and protocols

This section will be targeted to look into previous research works done on sweet potato seed systems, the current status, methodological issues of previous studies, their findings, implications and the knowledge gaps on the same. The literature review on clonally propagated crops will be undertaken on research/projects done in East Africa and beyond on seed systems and will be centered on four areas as suggested by Tylor (2010).

Be organized around and related directly to the thesis or research question, synthesize results into a summary of what is and is not known, identify areas of controversy in the literature, formulate questions that need further research.

2.5 Participation in research planning and design

Research design is described as a structure that precedes data collection or analysis and it serves as a function of ensuring that the evidence obtained enables us to

answer the initial research questions (whom shall we study?, what shall we observe? When will the observation be made? How will the data be collected (Nachimias, 1996), describes the design as the blue print that enables the investigator to come up with the solutions to these problems and guides him/ her to the various stages of the research. It is the glue that holds the different elements or process of research process together (Troachim, 2006). It deals with logical problem not logistical problem (Yin, 1989). Several elements are tied together in logical manner; observation or measures, treatment or programs, groups, assignment to group and time (Troachim, 2006).

The purpose of the research plan is to describe what, why and how of the research. The what? Addresses the specific aim of the research, the why gives the background and significance of the proposed research. The how deals with the methodology and subsequent preliminary results (SUNY, 2011).

2.6 Conceptual framework

This study focused on enhancing effective use of research methods among researchers at the International Potato Centre (CIP), through capacity building in data management, assessment of the role of data entry checks and organization to data analysis and evaluation and review of project proposals and literature on sweet potato seed systems.

The conceptual framework is premised on the assumption that projects under SASHA program need support on research methods so that research teams can be able to manage data effectively, so that there is a match between research objectives to research methods, which eventually would lead to systematization of methodologies used in research in order to foster quality and efficiency of their research products.

Enhanced use of research methods by researchers is determined by acquisition of new knowledge and change of attitude through capacity building on data management on one side and on the other hand on programs' exposure variables;

capacity building on data management, data checking and organization and evaluation and review of project proposals and literature. These parameters have a bearing on the wider level institutional parameters; socioeconomic, psychographic and interpersonal characteristics.

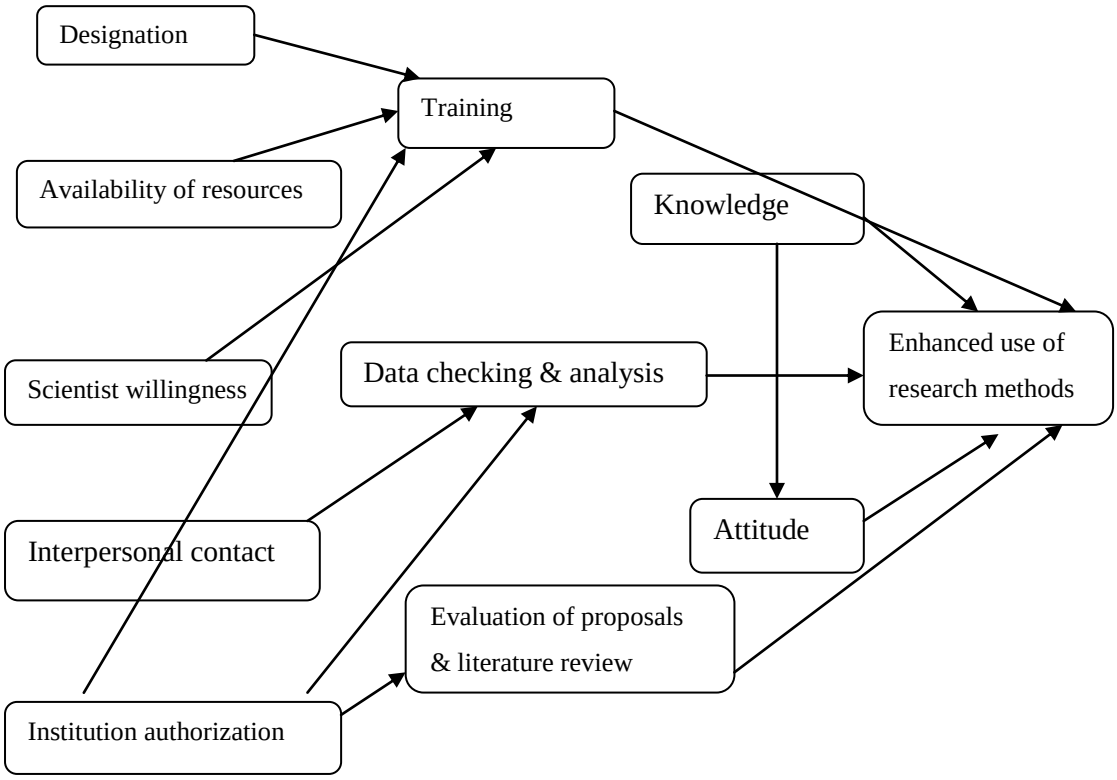


Figure 3: Conceptual framework for enhanced use of research methods skills

Chapter 3

3.0Project Approach

3.1 Evaluate the role of training in adoption of data management practices

Background: In CIP Projects were designed consultatively. Scientists providing input on project plans and designs, development of data collection instruments and design of impact pathways. Data collection was undertaken by enumerators under the supervision of the line researchers. Data entry and analysis was carried out through the use of excel or SPSS software. This resulted to various complexities ranging from lack of data controls and limited programmability. Further, the spreadsheet programs were found to be prone to errors, corruption and mismanagement of the data, making them suitable for small projects.

Owing to the limitations of the spreadsheet data management system (Excel and SPSS) experienced in handling survey data by data entry clerks and researchers, the M&E research team agreed to upgrade to a more powerful and programmable package that allowed controlled data entry/organization and data validation and checking before exporting to SPSS for analysis. CSPRO was considered the most powerful program for this purpose due to its ability to recreate data entry screens similar to questionnaires which makes it possible to capture all the information collected in the field. It also made it possible to program data validation checks (edit logics) and automatic skips in data entry. All these called for equipping of the data entry clerks with skills in CSPRO version 4 so that they could be able to manipulate and manage survey data effectively before analysis. The normal training cycle(figure 2) was adhered to which spanned training needs analysis (TNA), design and development, evaluation and follow up and preparation of a training report. A mixture of training methods was employed (Table 1)-discussions, demonstration and simulations for training.

Training Needs Analysis: Determination of training activities, needs and goals was done at this stage, through administration of a structured questionnaire(appendix 1)to data entry clerks and researchers with key aim of establishing their level of

experience in use of CSPRO, knowledge gaps and specific areas in use of the data management program.

Data was collected on several areas which ranged from education level, experience in data entry and data management, computer knowledge, previous experience with CSPro and inquiry to areas the data entry clerks wished to be introduced to in CSPro. Data collected was analyzed using SPSS version 18 and the findings synthesized into a report that profiled performance gaps, key priority areas and possible solutions to data management challenges.

Design and development stage: the means to meet training activity needs and goals was developed. This entailed development of training materials and manuals designed to cover both the theoretical and practical working of the program. The content of the training material was chosen on basis of three considerations guided by the training objectives identified in TNA. Information was prioritized on basis of 'Need to know' this tackled the critical content, 'good to know' formed the content that supported the training objectives and would be shared if time allowed and finally,' nice to know' content which was included for participants to go through after mastering the critical content.

Implementation of the training: training was rolled out to nine data entry clerks and four researchers handling baseline survey data. The training was structured to bridge the knowledge gaps identified in training needs analysis conducted to establish the level of competence in computer application and previous experience on use of CSPRO. The research team was exposed to CSPRO data management program on areas of creation of the data entry application, creation of the data entry forms, data checking and running double data entry. Through training needs analysis (TNA) and available resources, it was deemed fit to employ a mix of training methodologies, which were a mixture of both cognitive and behavioral methods (Table1). The cognitive method provided the theoretical basics on how to navigate in CSPRO program, written and verbal information was provided. These

methods were geared towards changing knowledge and attitude by stimulating learning- hence demonstrations and discussions were utilized.

On the other hand behavioral methods were geared towards giving practical training to the trainees and targeted skills development- case studies, and simulations came quite handy. More on the job training methods were also used i.e. coaching and mentoring.

Training evaluation: was done to establish the value of the designed training program to data management using CSPRO software for data entry and organization. Several areas of the training were evaluated; achievement of training objectives, length of training, material covered and the program support for the training. Focus group discussions, checklist of questions, questionnaire (appendix 2) and observation were used to assess the effectiveness of the training and a report compiled.

Follow up stage: This was undertaken to gauge whether the trainees successfully transferred the skills and knowledge acquired during training activity to practice. This was done through continuous monitoring of data capture process and performance of CSPRO functions as well as assessment of level of acceptance of the software among researchers. After the training the data entry clerks were tasked with creating the data dictionary, the data entry forms and perform other data management functions like data entry, data validation and checking for the Tanzanian sweet potato baseline survey 2010 (appendix 3).

Further follow up was done through checking on extend of adoption of CSPRO by researchers for their data management. An adoption questionnaire (appendix 4) was designed and administered four months after the training (June 2011) to assess the level of adoption of the CSPRO software among the data entry clerks and lead scientists involved in designing and executing baseline surveys among other studies in SASHA program. Nine data entry clerks and four researchers were engaged through a questionnaire to gauge the adoption of the software in data management

for their routine data entry and organization before analysis. Three key adoption metrics were used to study the adoption- usage, data quality and data management performance metric.

Data on the extent of use of CSPRO program in data management, effectiveness of the training in influencing acceptance of the data management practices and the overall bottlenecks in embracing of the package in data management was collected and analyzed using SPSS version 18.

3.2 Assess the role of data entry checks

The role of data entry checks and organization was determined using the data generated from Tanzania Seed sweet potato vine multipliers baseline survey Tanzania 2010.Thebaseline survey was carried out with key objectives as outlined below;

(1)To identify and characterize vine multipliers

(2) Understand the current sweet potato seed systems in Tanzania to help plan for effective projects implementation

(3) Understand the existing seed systems which will help in identification of solutions to the current problems

The data had already been collected using a structured questionnaire. Data collection quality checks (completeness, clarity and consistency of the questionnaires) had already been performed in the field. The next phase was to study the role of data checking and organization as a means to minimizing errors vis-a- vis; a background of absence of data entry checks and organization.

3.2.1. Structure of the data

The data collection questionnaire had the following characteristics which were recreated in CSPRO data entry application;

Data level: the data had only one level- household level, with 22 records which were defined and linked to the household information as shown in CSPRO record

figure4, which highlights sections of questionnaire and areas of inquiry to the household level information. Each square box with a + sign before indicates presence of list of questions available under that specific record. PG1-PG22 indicates the total pages of the questionnaire used for data collection.

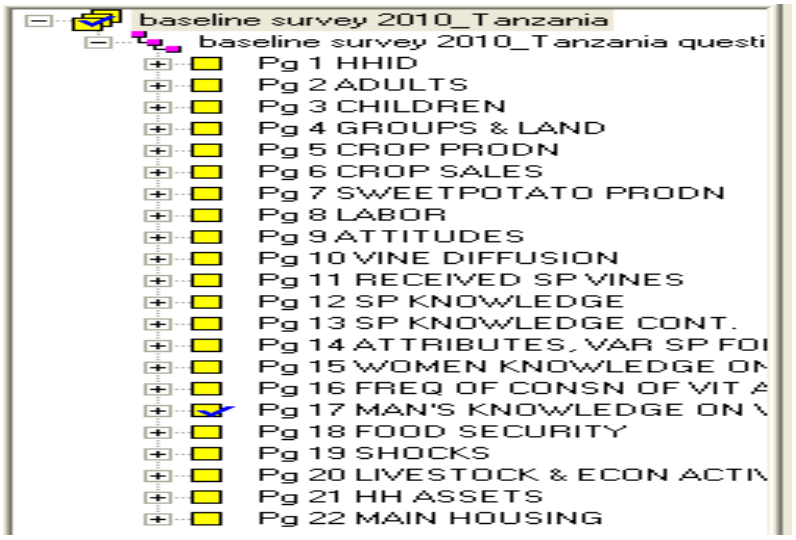


Figure 4: CSPro questionnaire records for the baseline survey from page 1 to page 22 of the questionnaire indicating subsequent records that information was collected on from the households sampled.

Items: these were the variables (923) and constituted responses to the questions in the questionnaire.

Data types: four data types were used to store the information viz; number, numeric codes, text (alpha) and date.

Number of cases: the data had 621 cases (respondents) who constituted the sample selected from the field in the baseline survey.

Levels of measurement: the data used the three levels of measurement; nominal, ordinal and scale.

Coding: the questionnaire used had codes for various variable responses and were adhered to when coding in CSPRO data entry screens (appendix 3). Consequently the same coding strategy was exported to SPSS.

3.2.2Data checking

To study the role of data checking two case scenarios were looked into whereby, one case data entry checks were created before data entry the latter case which had no entry checks. To study the effect of data checking, Record number 10- Vine Diffusion in (figure 4) was used as a case study and the data entered in CSPRO with and without checks prior to data entry or organization before being exported to SPSS for analysis.

3.2.2.1 Sample data Entry checks

Data entry forms were created in CSPRO to match the layout of the paper questionnaire. Data entry checks (edit logics) of different forms were created on variables which were deemed amenable to erroneous entries. Sample checks used are;

Range checks: These were checks defined to control lists of valid values for variables which needed to be checked against exaggerations like age, marital status and coded variables among others. Values outside the valid range were either disallowed or flagged as potential problems and investigated. An example is the variable year of birth- figure 5

```
PROC D4
POSTPROC

IF $<11 THEN
  $=2000+D4;
  ELSEIF $>10 & $<1900  THEN
   $=1900+D4;
  ENDIF;

IF $<1900 | $>2005 AND $<>8888 THEN
    ERRMSG("VALUE DOES NOT MAKE SENSE: PLEASE CHECK");
    REENTER;
    ENDIF;
```

Figure 5: Range checks for controlling against erroneous entries of variables that are interval in nature. D4 is the year of birth and it was to be recorded for every member of the household who was 5 years of age and above in years. The procedure was to be performed on D4 with the upper limit being equal or less than 2005 and

lower limit being 1900. Hence, any age outside 1900-2005 was returned as an error and an over rider specified to check it afresh.

Automatic skips in data entry: The survey questionnaires had instructions in some instances to skip questions that were not relevant or applicable to the respondent. This was intentioned to check false entries on some areas and save time at the end. For instance if in the section of details of household members with age <5 years the year of birth was >2005 and less than 2011 then that section was supposed to be skipped and fill the following section of demographic details of family members less < 5 years(figure 5).

```
POSTPROC
IF $>2005 & $<2011 THEN
    SKIP TO C6;
  ELSEIF  $>3 & $<11 THEN
     $=2000+C5;
     SKIP TO C6;
  ELSEIF $=99 THEN
    $=9999;
      SKIP TO C6;
  ELSEIF $=88 THEN
    $=8888;
    SKIP TO C6;
  ELSE
  ERRMSG ("NOT A LEGITIMATE YEAR:  PLEASE TRY AGAIN");
  REENTER;
  ENDIF;
```

Figure 6: Automatic skips checks to force skipping of questions not applicable to the respondents.

The age in years being considered to be expressed in months lie in the range 2004-2011 (limits excluded), forcing any entry in any other format to elicit an error message. If the member is a non resident then key in 999 and if the date of birth is unknown key in 888 and skip to C6 and fill age in months.

Labels for numeric codes checks: The numeric labels were defined in the data entry application and this command directed the entry clerks to open the drop down list of labels and key them appropriately (figure 7).

```
PROC BASELINE_SURVEY_2010_TANZANIA_FF


PROC A01
POSTPROC
 A01B=GETLABEL(A01,A01);
PROC A02
POSTPROC
 A02B=GETLABEL(A02,A02);
PROC A03
POSTPROC
 A03B=GETLABEL(A03,A03);
PROC A04
POSTPROC
 A04B=GETLABEL(A04,A04);
PROC A10
if A10=1 then skip to A11B endif;
PROC A12B
POSTPROC
A12A=GETLABEL(A12B,A12B);
```

Figure 7: Numeric codes checks used to prompt accurate identification of value labels.

Missing and not applicable values checks: The missing values were defined as 99 for continuous variables and 9 for discrete variables. The N/A values were checked with 88 in the CSPRO entry as shown figure 8.

```
PROC D5


PROC CHDNO
if CHDNO=0 then endgroup endif;
PROC C5

POSTPROC
IF $>2005 & $<2011 THEN
    SKIP TO C6;
  ELSEIF  $>3 & $<11 THEN
     $=2000+C5;
     SKIP TO C6;
  ELSEIF $=99 THEN
    $=9999;
     SKIP TO C6;
  ELSEIF $=88 THEN
    $=8888;
    SKIP TO C6;
```

Figure 8: Missing and not applicable checks; cases defined as 99 for continuous variables and 9 for discrete and 88 for not applicable cases.

3.2.2.2 Checks after data entry

Two checks were performed after data entry; Double Data Entry and Exploratory Data Analysis.

Double data entry: The data entry clerks exchanged their entered questionnaires and using separate copies of the data entry program in CSPRO the data was re-entered. The resulting files were compared and any discrepancies checked against the original questionnaires. Corrections were effected until the files were the same. This process worked well for the range variables like age among others which could not be easily checked by the edit logic.

Exploratory data analysis (EDA): At this point the Record No. 10- Vine Diffusion (figure 4) with entries as provided in a cross section of questionnaire in figure 9below was used to compare unchecked data and the one with checks. One record was created with checks and another without checks and the summary statistics compared.

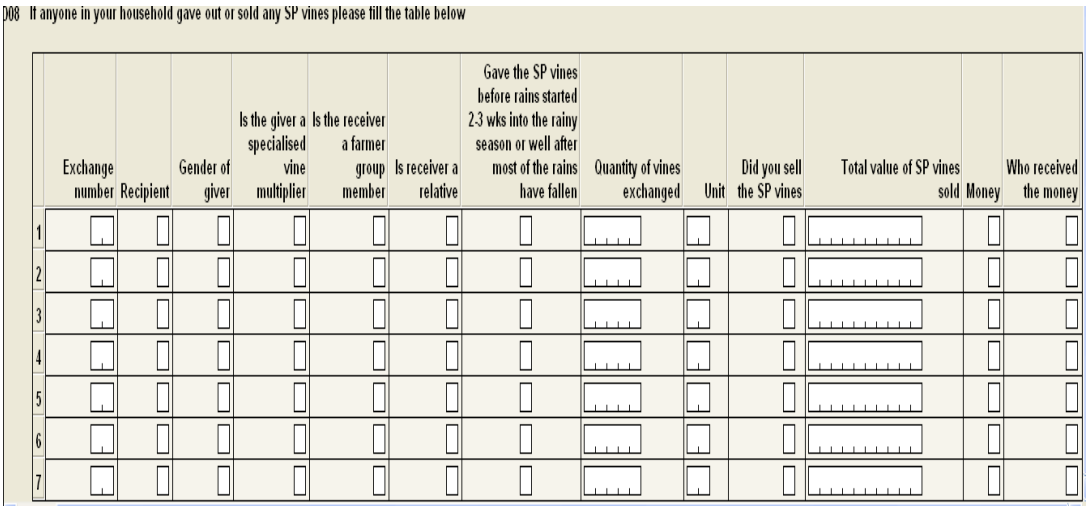


Figure 9: Section 08 of the questionnaire capturing vine diffusion information that was used to study the role of data entry checks.

This section of the questionnaire was geared towards documenting information on the available seed systems in the region. Questions were asked pertaining local transactions of vines to and from the household farms. For the categorical items (with numeric codes), frequency tables were used. Real numbers summary statistics such as means, range, inter-quartile range and graphical summaries were used. Box plots also came handy in comparing outliers in the data.

Due to enormous data generated in the baseline survey only record number 10 (figure 4) was considered for analysis purposes. A cross section of the questionnaire used (Figure 9) to capture the data was used and only four variable compared in the platform of with and without data entry checks. The data was analyzed using SPSS version 18 and significance tests performed on the two outputs to test whether the differences of the findings were significant.

Data analysis followed three major steps;

Data preparation- Cleaning and organizing the data for analysis- survey was first entered in CSPRO to avoid dealing with messy data as a result of direct entry to analytical software, data entry screens were developed in CSPRO, double data entry schemes developed with validation checks in built to the screens. Thereafter, it was imported to SPSS for analysis.

Descriptive statistics- The role of entry checks was assessed through running descriptive on a cross section of variables and ran a comparison analyses to check the divergences between with and without data checks scenario. The data was described through table of means, frequencies, box plots charts and graphs (chapter 4). After data entry and organization, data was analyzed to give the general spread and central tendency of the data.

Inferential statistics- Entailed testing hypothesis i.e. there is no difference between checked data entry and unchecked data through use of T test and chi square at 5 % significant level.

3.3 Evaluate and review project proposals and literature on sweet potato seed systems

Proposals and literature considered for evaluation and review were drawn from developing sustainable sweet potato seed system program component. Two proposals were reviewed and evaluated following a criteria inTable2, which had been developed in conjunction with the supervisors. Further, two data collection instruments to be used for data collection on the two projects were reviewed on the basis of the objectives of the study and recommendations made (appendix 7a & 7b).

Sections of the Project proposal considered in evaluating project proposals were divided into general criteria which accounted for 80% and specific criteria was awarded 20% of the rating.

Table 2: Sections of project proposal considered for evaluation

General criteria Technical Quality of the proposal	Points awarded (%)	Specific criteria Relevance to programs' objective	Points awarded (%)
Background / Introduction	10	Contribution to food security	10
Study Hypothesis /research questions	10	Contribution to community/ national development goals	10
Study Design / Protocol Development	10		
Rationale/Study Objective	10		
Preliminary Statistical Design	10		
Schema / Flowchart	10		
Communication Plan	10		
Dealing with Results	10		
Total	80		20

3.3.1 Summary of proposals reviewed and evaluated

Two project proposals that were reviewed and evaluated and subsequent summary and recommendations made. The first proposal was prepared by Dr. J.S Kirimi on development of high value sweet potato products value chain in Rwanda 2010/2011and the second one was prepared by Dr.MCEwan Margret.

Proposal one Title: Development of high value sweet potato products value chain in Rwanda

Author: Dr. Kirimi-International potato centre (CIP)

Overall objective: 1.To compare the farmer welfare outcomes from the introduction of two models for producing sweet potato flour, one where intermediate chips are produced by farmer groups and the other where the flour producer (and bakery) does all the processing,

2. To test different models for inclusion of the poor and women in the development of new market chains for high value sweet potato products. These objectives sit within the larger development question of how the poor, and particularly women, can capture the benefits of increased commercialization of staple crops such as sweet potato.

Specific objectives

1. Develop a sweet potato value chain based on contracting fresh roots from farmers which are then chipped, flash dried and milled into flour or either boiled and mashed at the SINA factory or other new processors.
2. Develop a sweet potato value chain based on farmer group formation, processing (chipping and drying, or boiling and mashing) by farmer groups in the production zone, and selling high quality semi-processed products to the SINA factory and other processors for their bakery line.
3. Support technology development and dissemination for high value white and orange- fleshed sweet potato product value chains, and evaluate change in consumer acceptance of sweet potato based products.
4. Compare and evaluate the relative efficiency of the two value chains and the potential to increase farmer incomes.

Hypothesis:

1. The development of value-chain for processed products with a private sector actor leads to improved returns to sweet potato cultivation for growers.

2. Farmer-based value-added pre-processing prior to delivery to factory results in greater participation of and revenues for women farmers than contract growing schemes managed by the private sector processor in which all value addition occurs at the factory.
3. Sweet potato will become high value crop among urban and up country consumers in target areas through effective marketing of an economically viable sweet potato processed

Research design & implementation plan: Project to be implemented in three districts- **Rulindo-**estimated population of 307,501 inhabitants **Muhanga-** population of 349,094 inhabitants and **Kamonyi-** population of 265,365 inhabitants.

Implementation: Key stakeholders (ISAR, CSR, SINA& local authorities) to raise awareness on the project objectives and agree on responsibilities.

Monitoring & evaluation: Designed baseline assessment will be undertaken in target districts. Questionnaires will be used and sample of 100 farmers from contract farmers, cooperative farmers and from farmers without access to a processing market
Methodology: Vision of success for each objective and activities defined and linkage between output and outcome delineated.

Recommendation/ conclusion: The specific objectives (outputs) are mismatched with the general objectives (intended outcome) of the project in a number of areas; the immediate output focuses more on the value chain formation and their subsequent efficiency and misses out on the welfare of the farmers- the poor and women.
To realign the outputs to the outcome it is suggested that the specific objectives be targeted more to capture the welfare aspects and emphasize the poor and women model in the project.

The design and implementation session though giving guidelines and flow of the project fall short in key areas; viz

(i)Definition of an elaborate sample selection in the three districts

(ii)Criteria of gender disaggregation and subsequent inclusion in the study

(iii)Decision criteria for the vision of success under the objectives

(iv)The data collection, analysis and reporting criteria are assumed in the project protocol

Proposal two title: Going to scale with sweet potato vines distribution in Tanzania

Author: Dr. MCEwan-International potato centre (CIP).

Overall objective: To improve the food security for subsistence farmers who rely on sweet potato as a staple and to enhance the incomes of a more commercially oriented group who produce cuttings or sell roots to generate income.

Specific objective:

1. To provide farmers with quality seed of improved sweet potato varieties in a timely fashion
2. To promote quality vine conservation among farmers
3. To stimulate increased demand for white and orange-fleshed sweet potato amongst rural and urban consumers
4. To assess the contribution of vine distribution and farmers' use of positive selection and vine conservation to raising productivity and improving food supply
5. To assess the cost effectiveness of the voucher based sweet potato seed system.

Research design and implementation: Project to be implemented in two regions; Mara and Mwanza with subsequent districts and divisions derived therein.

Implementation: The first activity in year 1 is to identify wet areas and sites with water access through GIS mapping and characterize existing informal vine distribution system

A formative research will be conducted in year one to comprehend the social, cultural and household barriers to the expanded production and consumption and diversified use of sweet potato to enable the design of an effective campaign to improve its image among urban consumers.

A subset of 30 farmers per site will be selected for continuous monitoring during 3 seasons. Mini-kits or simple field sampling kits will be used to gather information about virus incidence at the plot level with repeat visits

Hence a survey will be carried out to elicit farmers' willingness to pay for new varieties, early supplied vines and clean seed and the quantities they would demand under different price options.

An institutional analysis of the seed system will be carried out to identify the transaction costs associated with the system, these relate to search costs for information on the part of both multipliers and farmers and contract compliance.

Monitoring& evaluation: Field staff will collect data on vine production, distribution and use. The seed flows will also be mapped using GPS to track both formal and informal diffusion from different locations. Eventually results will be shared with stakeholders at SPHI meetings.

Methodology: Vision of success for each objective and activities defined and linkage between output and outcome delineated.

Recommendation/conclusion: The design and implementation section though giving guidelines and flow of the project fall short in key areas; viz

(i) Definition of an elaborate sample selection in the two regions

(ii) Criteria of gender disaggregation and subsequent inclusion in the study

(iii) Decision criteria for the vision of success under the objectives

(iv) The data collection, analysis and reporting criteria are assumed in the project protocol

3.3.2 Review of the data collection instruments

Two questionnaire instruments were reviewed for the above evaluated proposals. The key object of the review was to make them more understandable and free of bias as possible. Minor mistakes which could have altered the meaning and mistakes were highlighted. The review was guided by the study objectives and overall vision of success of the project.

Several factors were used to target questions in order to elicit answers deemed appropriate viz;

Clarity: Questions were subjected to this test in order to make them clear and unambiguous. The goal was to eliminate the chance that the question would mean different things to different people.

Phrasing: This was done to try as much as possible to make the questions mean the same thing to all respondents. For instance, the questionnaires developed for the value chain project in Rwanda had to be rephrased in a number of questions owing to use of the adjectives, verbs, and nouns inappropriately (appendix 7a &7b)

Prestige bias: Prestige bias is the tendency for respondents to answer in a way that make them feel better. People may not lie directly, but may try to put a better light on themselves. Questions which seemed to force the respondent depict themselves in an elitist manner were rephrased for example;

Q. Do you prefer to purchase processed or raw sweet potato? Was changed to; in which form do you purchase sweet potato? Processed ( ) Raw ( )

Leading questions: questions thought to imply certain responses were rephrased or changed all together. The ordering of questions was also factored and this was applied to increase the response rate by asking background information first before other questions.

3.3.3 Review of literature on sweet potato seed systems in East Africa

Seed systems have several purposes and effective seed systems provide the different categories of farmers with planting material i) in sufficient quantities ii) at the right time iii) of an appropriate physiological state, vigor and health, iv) of superior genotypes appropriate to the farmer's purposes and v) at an affordable price. In order to maintain superiority of genotypes and, in some cases, health, there needs to be capacity within seed systems for generation, dissemination and multiplication of new stock, new cultivars and/or pathogen-free material. Sweet potato is propagated through foliar cuttings: although it is possible to propagate it through roots, this is almost never done because the roots become malformed and also carry diseases on them. In Tanzania, Uganda and Rwanda, the seed system is almost entirely informal (Ndamagé, 1990; Bashaasha et al., 1995; Kapinga et al., 1995), with only occasional formal distributions of new varieties (Kapinga et al., 2000) and for disaster relief

(Kapinga et al., 2005). The lack of a major formal sector creates particular difficulties in development and dissemination of new improved cultivars and of disease-free stock, particularly freedom from asymptomatic viruses, in this clonally propagated crop.

The literature review followed three steps;

Step 1: identification of various published and unpublished material that were available on sweet potato seed systems in the region and gain access to them. This was majorly done through the sweet potato knowledge portal which is run and maintained by CIP data management unit. This portal allows scientists to share information and upload their work on sweet potato with seed system having a folder (figure10) within which are sub-folders (figure 11)where scientists or registered members upload articles and research of both published and unpublished work.

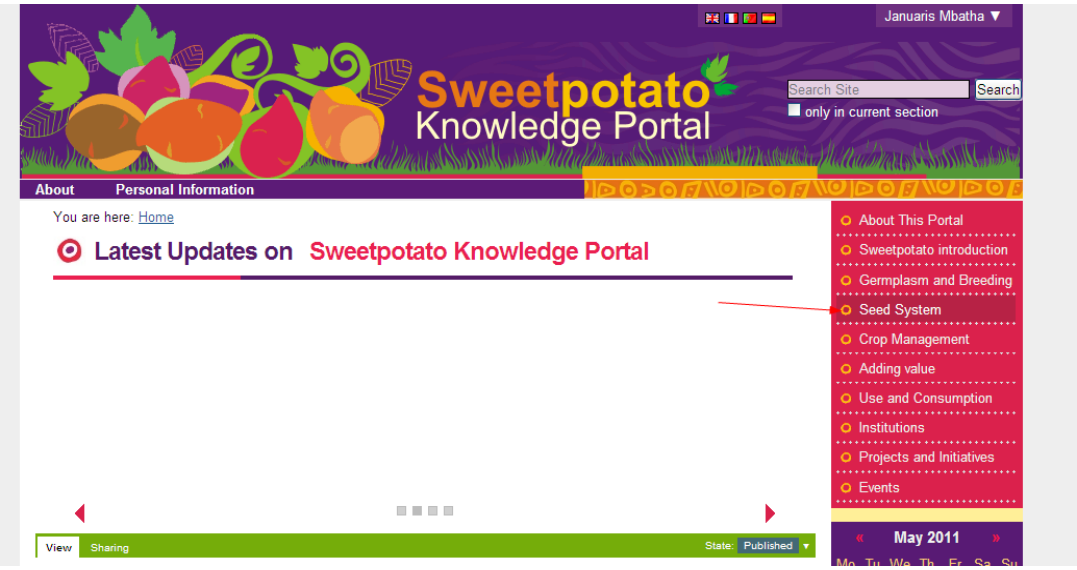


Figure 10: Sweet potato knowledge portal showing fields of uploading material on seed systems among others.

The subfolders as shown in figure 11 contained the bibliographic databases (the name of the author, the title of the article, source of publication, year, volume, and page numbers), as well as full articles and papers uploaded by registered members/scientists.

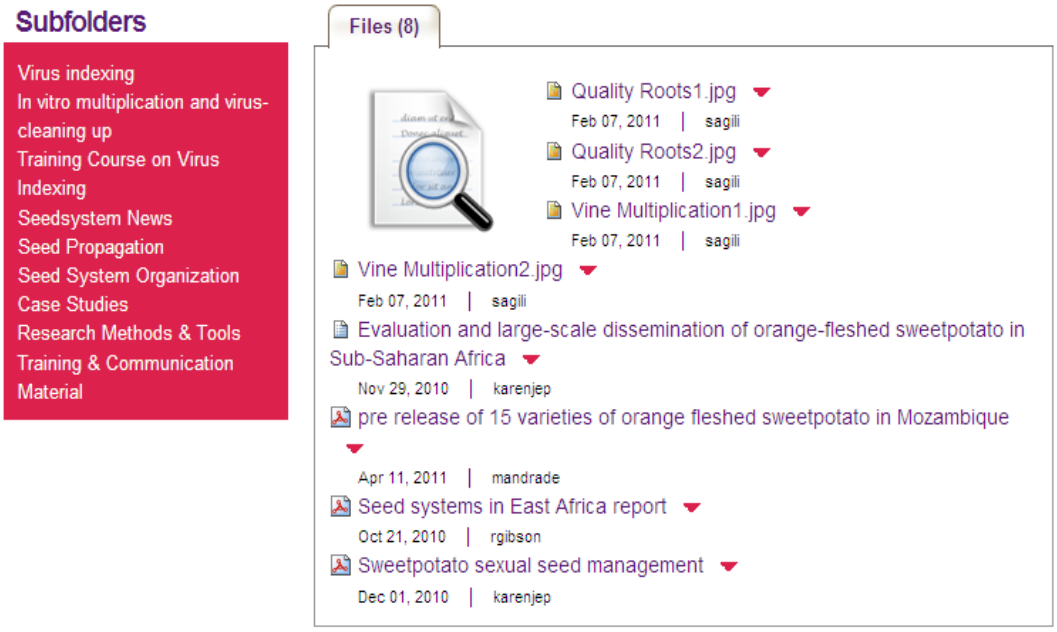


Figure11: Sample subfolders in sweet potato knowledge portal that contained specific articles on sweet potato seed system.

Step 2: Gathering of relevant information by going through the necessary materials. Over and above the knowledge portal other materials were accessed and relevant information derived in accordance to the objectives of the study. A total of 35 articles and the ultimate finding were recorded in a format as suggested by Sekaran (2003); name of the author, date of publication, title of the article, objectives of the study, summary of the methodology and main findings/conclusions.

Step 3: Writing up the literature review. All relevant information was put into logical order and structured in such a way to accentuate the problem of the study. The documents were referenced according to APA manual (2001) i.e. Name of the author (year). Title and the publisher. Finally, all articles considered relevant to the study were listed as references.

Chapter 4

4.0 Results

4.1.0 Role of training

Nine data entry clerks and four scientists were trained in use of CSPRO version 4 for survey data management. A TNA was conducted through the use of a structured questionnaire (appendix1) to establish the level of education, experience in data entry and data management, computer knowledge, previous experience with CSPro and inquiry to areas the data entry clerks would wish to be introduced to. Training evaluation and follow up were also conducted through administration of a questionnaire after training (appendix 2 & 4) to assess the role of training in adopting data management practices among researchers.

4.1.1 Training Needs Analysis (TNA)

Education Level: Table 3shows the education level of the trainees. Majority (44.5%) had bachelors level of education, followed by postgraduate level at 33.3% and diploma level of education at 22.2%.

Table 3: Education level of the trainees

Education level	frequency (n)	Percentage (%)
Bachelor degree	6	44.5
Diploma	3	22.2
Postgraduate degree	5	33.3
Total	14	100.0

Previous data management experience: All the participants had previous experience in one or more facet of data management cycle as shown in table 4. Majority of the trainees (57%) had previous experience in data collection, organization, cleaning, data entry and analysis with 14.5% having done data collection and analysis only, the rest (28.5%) having previous experience in data collection only.

Table 4: Data management experience

Data Management area	frequency(n)	Percentage (%)
Data collection	4	28.5
Data collection & analysis	2	14.5
Data collection, organization, cleaning, entry and analysis	8	57.0
Total	14	100.0

Trainees' computer knowledge on elected data management software's

Table 5 show the ratings on data management packages; (1) representing no skill or expertise, (2) Some skill or expertise, (3) Competent and (4) Denoting excellence that could be shared with group members during data entry exercise.

All the trainees were computer literate with divergent skills in data management packages with 14.3% of the respondents being competent in use of CSPRO application, 42.9% having excellence in masterly of the program which could be shared with group members during data entry exercise. More than 75% of the respondents had some skill or lack of it in use of SPSS and GENSTAT. The entire group was conversant with access, excel and word processor.

Table 5: Computing knowledge on selected computer software

Computer package	Rating Scale (%)				Total
	1	2	3	4	
Word processor				100	100.0
Excel		28.6	14.3	57.1	100.0
Access		42.8	28.6	28.6	100.0
SPSS	42.9	42.9	14.2		100.0
GENSTAT	57.1	28.6	14.3		100.0
CSPro	28.6	14.3	14.2	42.9	100.0

Key:

1.No skill or expertise 3. Competent 2. Some skill or expertise 4. Excellence that could be shared with group members during data entry exercise

Data entry experience using CSPRO

Majority (66.7%) of the trainees had previous experience in use of CSPRO program in performing various functions. Only 33.3% had no previous experience in use of the software.

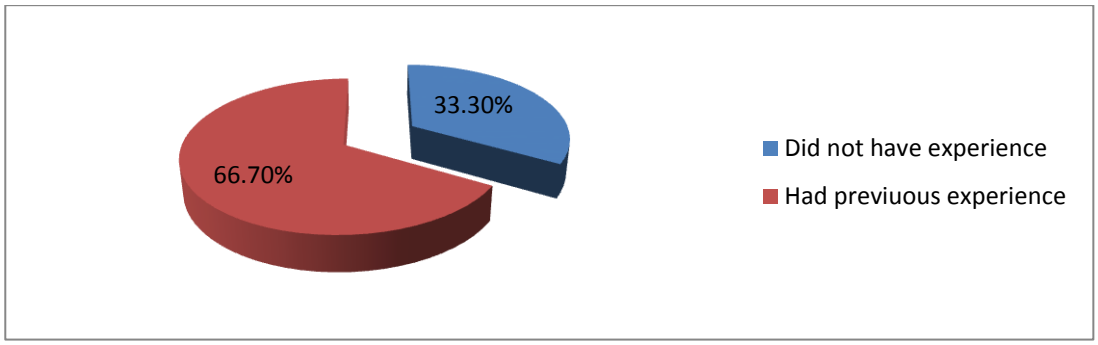


Figure12: Previous use of CSPRO showed majority had used the package to perform various tasks.

Previous tasks performed in CSPRO

Table 6show the ratings on level of competence of most critical CSPRO tasks; (1) representing no skill or expertise, (2) Some skill or expertise, (3) Competent and (4)Denoting excellence that could be shared with group members during data entry exercise.

More than half (57.1%) of the trainees were competent in launching CSPRO, 42.9% possessed excellent skills, which could be shared with other research group members in creation of data dictionary and data entry application respectively with 14.3% having excellent skills in adding edits to the data entry application. However, there were cases which had no skills or expertise in either of the functions which stood at 28.6% for starting CSPRO, creation of the data dictionary, creation of the data entry application respectively and 42.9% for adding edits.

Table 6: Rating of tasks performed previously in CSPRO

Task	Competence rating in %			
	1	2	3	4
Starting CSPro	28.6	14.3	57.1	
Creation of the data dictionary	28.6	14.3	14.3	42.9
Creation of the data entry form	28.6	14.3	14.3	42.9
Entering data	28.6	14.3	14.3	42.9
Adding edits to the data entry application	42.9	14.3	28.6	14.3

Key: 1.No skill or expertise
3. Competent
2. Some skill or expertise
4. Excellence that could be shared with group members during data entry exercise

Other areas that training was needed as per the respondents in CSPRO

The trainees expressed interest in learning various CSPRO functions with majority (42.9%)wishing to learn how to run batch application (data checking), 29% wanted to learn how to perform double data entry and rest (28%) preferring to learn how to merge files and export data from CSPRO to SPSS.

Table 7: Other CSPRO functions that trainees wanted exposed

CSPRO Function	Frequency (n)	Percentage (%)
Exporting data from CSPRO to SPSS	2	14.0
Merging files	2	14.0
Running the batch application/ edit logic	6	43.0
Running double entry	4	29.0
Total		100.0

4.1.2 Evaluation of the training program

An evaluation questionnaire (appendix 2) was administered to ascertain whether the training achieved its objectives- Start up CSpro program, Creation of the data dictionary, Creation of the data entry form, Entering data in CSPRO and Adding edits to data entry application.

Several areas were evaluated and summarized;

Evaluation of training activities: The participants rated achievement of programs objectives, organization of training and trainers knowledge were rated excellent. Matching of content and questions asked, trainers' ability to explain content clearly and achievement of personal expectations were rated as good and excellent by the trainees. Usefulness of visual aids and handouts was rated fair by 62.5% of the participants.

Table 8: Training evaluation

Criteria	Respondents rating (%)			
	Unsatisfactory	Fair	Good	Excellent
Achievement of program objectives				100.0
Achievement of personal expectations			50.0	50.0
Relevancy of content to my needs and interests	12.5		50.0	37.5
Organization of the program				100.0
Usefulness of visual aids and handouts		62.5	12.5	25.0
Trainer's knowledge				100.0
Match between content and my questions			62.5	37.5
Trainers ability to explain content clearly			13.0	87.0
Trainers ability to respond well to questions			37.5	62.5

Length of the training

Majority (75%) of the trainees were satisfied with the length of time spent on the training and checked it as just right. However, 12.5% said it was too long with same proportion (12.5%) saying it was too short.

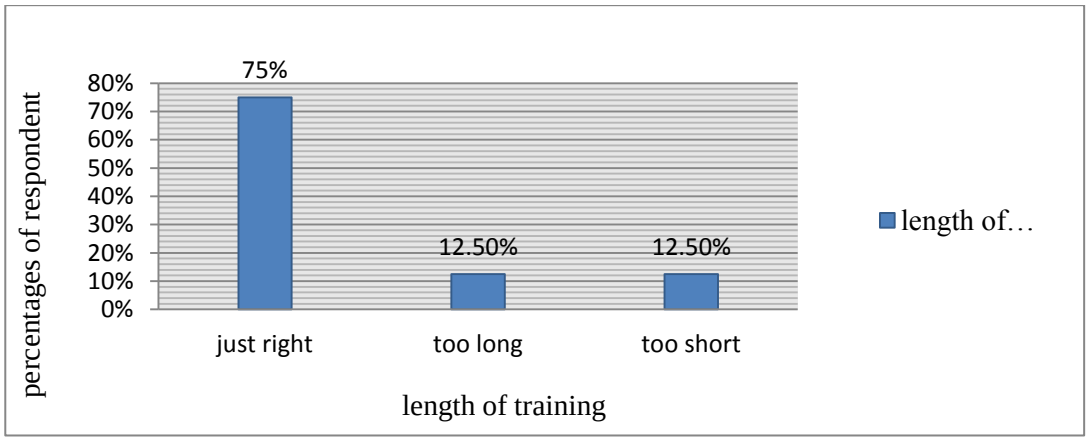


Figure13: Length of time spent on the CSPRO training program as rated by the trainees

Level of material covered

On the material covered 75% felt it was just right and they got what they wanted, with 12.5% checking it as just right and too high such that they felt overwhelmed respectively.

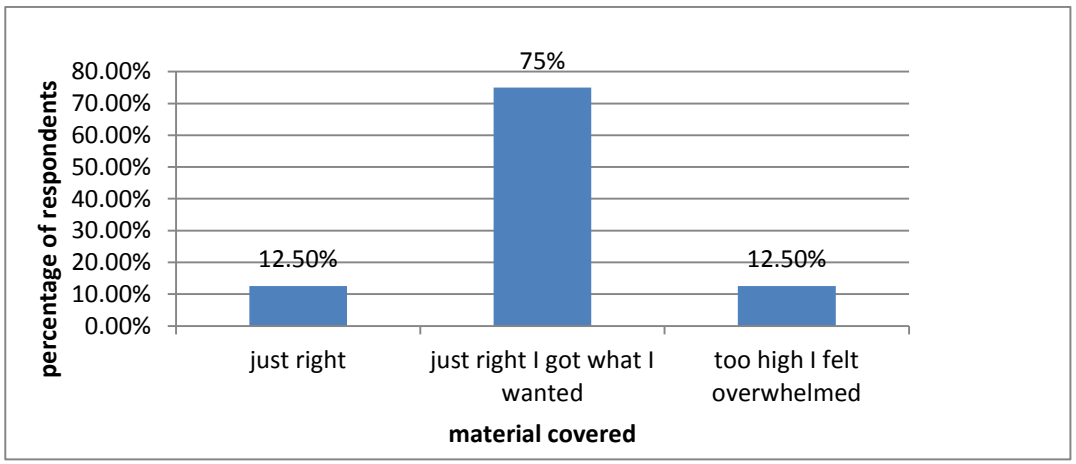


Figure 14: Level of material covered during the training

Areas which the trainees wanted more time to have been spent on

Majority (37.5%) of the trainees felt that more time should have been dedicated to discussing creation of data entry application in CSPRO (Table 4.7), with 25% saying that more time should have been spent in learning how to create data dictionary and the rest suggesting; creation of data entry logic, concatenating files and exporting to SPSS, and running tabulation application each at 12.5%.

Table 9: Areas of discussion where more time should have been spent

Areas	Frequency (n)	Percentage (%)
Adding edits to the data entry application	2	12.5
Concatenate files and exporting to SPSS or other software	2	12.5
Creation of data entry form	5	37.5
Creation of the data dictionary	3	25.0
Tabulation applications	2	12.5
Total	14	100.0

Additional topics for inclusion in future training

Table 10 gives the additional topics suggested for addition in future training with manipulation of data and exportation to analytical packages preferred by the majority (37.5%), running summary analysis in CSPRO (25%) with the rest 27.5% suggesting how to do double data entry, merging of files, and running batch application.

Table 10: Suggested topics for inclusion in future training

Topics	Frequency (n)	Percentage (%)
How to do double data entry	2	12.5
How to merge files created by different data entry clerks	2	12.5
Manipulation and export of data from CSPRO to other software	5	37.5
Run summary analysis in CSPRO	3	25.0
Running batch application	2	12.5
Total	14	100.0

Evaluation of the training support

The training support by the organization was rated as excellent in terms of the user friendliness of the computers provided by all the participants. Half (50%) felt that the organization of the training was excellent and 37.5% and 12.5% checking it as good and fair respectively. The administrative support was rated as good and fair by 87.5% and 12.5% respectively.

Table: 11: Training support

Training Aspect	Excellent (%)	Good (%)	Fair (%)	Unsatisfactory (%)	Total (%)
Versatility of computers	100.0				100.0
Organization of the training	50.0	37.5	12.5		100.0
Administrative support		87.5	12.5		100.0

Suggestions for improving the training

It was suggested that

- the lead scientist should take more active role in training arrangements
- data entry supervisor to attend the training
- the training to have a rapporteur
- visual aids to be used more in the training

4.1.3 Follow up on adoption of CSPRO in data management among researchers at CIP

Three adoption metrics were assessed (Usage, data quality and data management performance metric) to determine the level of adoption of use CSPRO amongst scientists over the last four months (February- June 2011) after the training. A structured questionnaire was administered (appendix 4) to nine data entry clerks and four researchers to gauge extend of adoption of data management practices (data checking and organization and data entry) using CSPRO before analysis. The findings are as documented below;

Gender of respondents

Females were majority at 75%while 25% were males with all the respondents drawn from both data entry clerks and researchers.

Table 12: Gender of the trainees

Gender	Frequency (n)	Percentage (%)
Female	10	75
Male	4	25
Total	14	100

Designation of the trainees

Majority were data clerks (66.7%) followed by researchers at 16.6%. the rest were gender expert and socio-economists at 8.3% respectively

Table 13: Designation of the respondents

Designation	Frequency(n)	Percentage (%)
Data entry clerk	9	66.7
Gender expert	1	8.3
Researcher	2	16.6
Socio-Economist	1	8.4
Total	14	100.0

.**Experience in use of selected data management computer programs**

The respondents were not new in use of data management software with all of them having used Excel, Access and CSPRO in the past (Table 14). Use of Epidata, Epi Info and Oracle was at a limited level (33.3%). SPSS which was being used for survey data analysis was at 58.3% with SAS and STATA being the least used at 8.3%.

Table 14: Experience in use of computer programs

Data management software	Previous use (%)	
	yes	No
Excel	100.0	
SPSS	58.3	41.7
Genstat	16.7	83.3
Access	100.0	
CSPRO	100.0	
Epidata	33.3	66.7
Epi Info	33.3	66.7
Oracle	33.3	66.7
Other (specify) SAS & STAT	8.3	91.7

CSPRO competence among researchers after the training

Table 15 show the ratings on level of competence on most of critical CSPRO tasks after the training; (1) representing no skill or expertise, (2) Some skill or expertise, (3) Competent and (4)Denoting excellence that could be shared with group members during data entry exercise.

The respondents were competent and majority (91.7%) had excellent skills in four out of five CSPRO functions. Adding edits was fairly ranked with 25% having some skills, 41.7% and 33.3% being competent and possessing excellent skills respectively.

Table: 15: CSPRO competence after the training

CSPRO Function	Competence rating (%)			
	1	2	3	4
Starting CSPRO			8.3	91.7
Creation of data dictionary			8.3	91.7
Creation of data entry forms			8.3	91.7
Data entry			8.3	91.7
Adding edits to the data entry application		25.0	41.7	33.3

Key:

1. No skill or expertise
2. Some skill or expertise
3. Competent
4. Excellence that could be shared with group members during data entry exercise

Usage adoption metric

Figure 15 show use of CSPRO by scientists after the training. Majority (83.3%) of the respondents used the data management program CSPRO in the last four months after the training for data entry and checking functions in their projects.

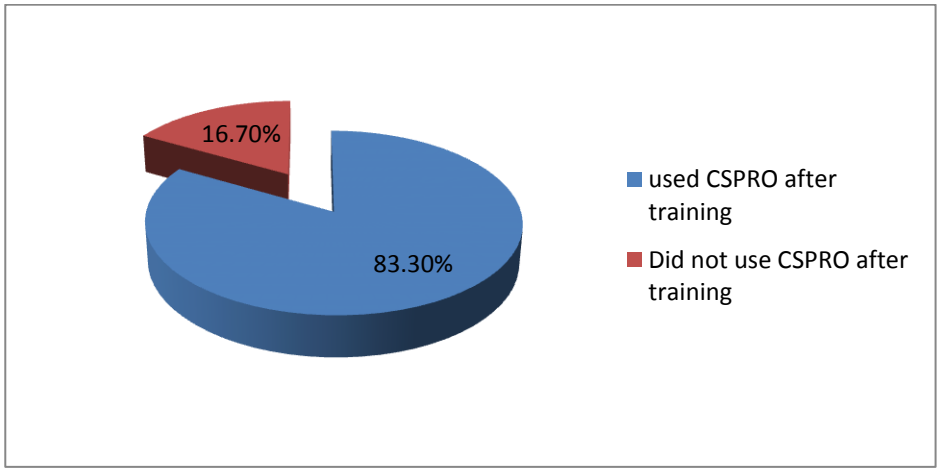


Figure15: CSPro use among researchers after the training

Table 16 show the number of projects management through CSPRO between February to June 2011.There was little use of other data management packages (excel) at 16.7% for data entry only. The rest used CSPRO with 41.7% of this group performing other functions beyond data entry.

Table: 16: Number of projects managed through CSPRO after the training

Number of projects	Frequency (n)	Percentage (%)
0	2	16.7
1	2	16.7
2	5	41.7
3	2	16.7
4	1	8.2
Total	12	100.0

Data Quality Metric

Data managed through CSPRO was reported as of very high quality by 58.3% of the respondents, 33.3% said it was excellent and only 8.3% rating it as good figure 16. This was made in reference to data queries raised during analysis and 66.7% said did not get queries at all and 33.3% saying they rarely got queries (figure 17).

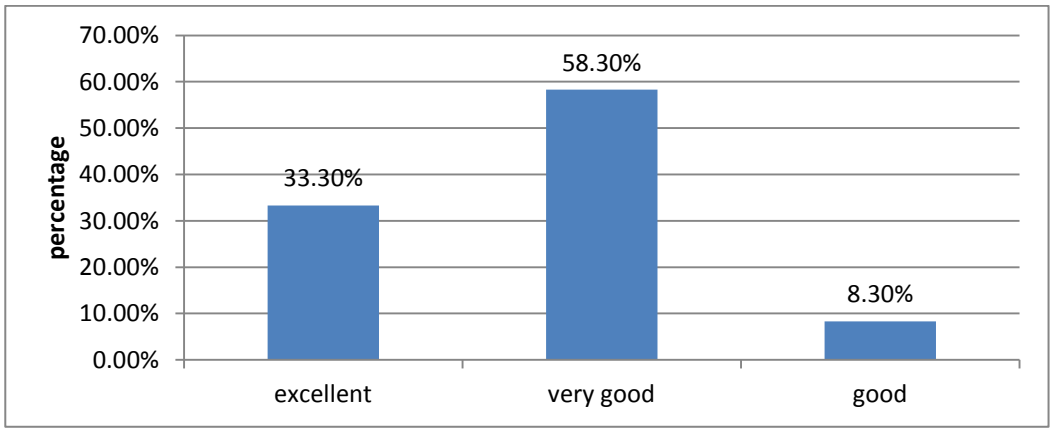


Figure 16: CSPro data quality: Majority (58.30%) of the respondents contented that the data output from CSPRO was of very good quality with 33.30% citing it as

excellent and minority (8.30%) checking it as good. It had also minimal queries as shown in figure 17

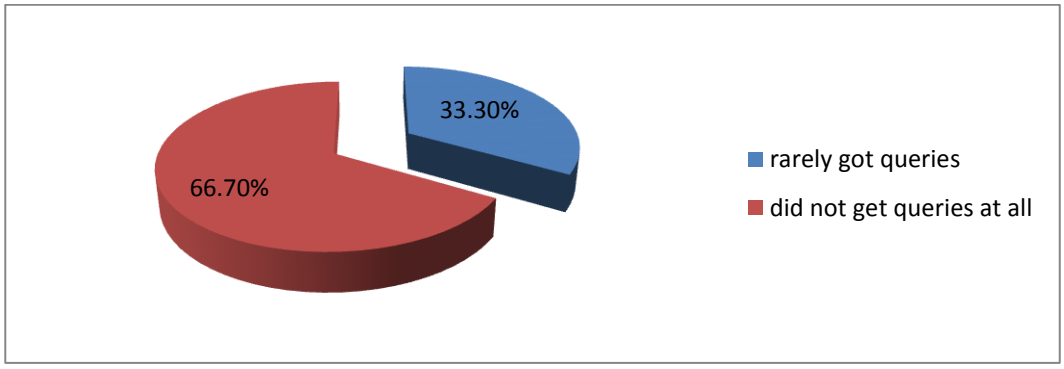


Figure 17: Frequency of CSPro data queries. Large proportion (66.7%) of the respondents reported that they did not get queries for clarification or verification during data analysis while 33.3% said the queries were rare.

Data management Performance Metric

This aspect checked at the efficiency of the program in promoting effectiveness of data management in the organization. Further the willingness to adopt the program across the scientific class in CIP.

All the respondents recommended CSPRO for data entry, checking and organization functions while ranking it superior to all other data management packages they were using before. Part of the reasons advanced for the recommendation; effective program for data checking (33%), highly flexible and interactive (40%) and 27% said it replicates the questionnaire instrument hence capturing all the variables.

Table17: Reasons for recommending CSPRO to other researchers

Reasons	Frequency(n)	Percentage (%)
Effective program for data checking	5	33
Highly flexible and interactive	6	40
Replicates the paper questionnaire	3	27
Total	14	100

Challenges encountered in using CSPRO for data management

Table17 summarizes some of the challenges that the researchers and data entry clerks encountered in using CSPRO in data management. Some solutions were also suggested. Key challenges bordered on resistance from scientists in adopting CSPRO, hiring data entry teams on temporally basis and lack of a functional data management unit. It was suggested that hiring of research methods professional, encouraging the scientists to take active role in training as well as creating a data management unit will go a long way in solving data management problems in the organizations.

Table18: Challenges faced by scientist in using CSPRO program

Challenges	Recommended solutions
Lack of contact person for consultation	Hire services of research methods professional
Resistance from lead scientists	Involve scientists in trainings
Cumbersome process of exporting data to analytical packages(SPSS)	Training
Hiring new data entry clerks on temporally basis	Create data management unit with permanent data entry clerks

4.2.0 Data entry checks and organization

To analyze the role of data entry checks a cross section of the questionnaire (figure 19) was used to compare the two scenarios i.e. With and without data entry checks inform of edit logic in CSPRO.

4.2.1.0 Entry checks used for record 10- vine diffusion section SD08 on seed systems survey

Two case scenarios were used to study the effect of data entry checks on analysis. First scenario was to enter data without entry checks and the second was to inbuilt checks before data entry, run the analysis and compare the output. Figure 18show a

defined entry check used to validate entry on a cross section of the questionnaire used (figure 19).

An example is provided from a questionnaire from vine diffusion section SD08 on the items to be entered. If the exchange number was not provided then all other entries would be null and void hence skipped. To control this entry an entry check was defined to restrict entry only on cases with an exchange number otherwise end the entry if exchange number missing (88).

```
POSTPROC
 IF $<>99 THEN
   X=$;       {X IS A PUBLIC VARIABLE WHOSE VALUE CAN CHANGE THAT IS DECLARED AT THE BEGINNING OF THE PROGRAM}
   SD01=D1(X);  {D1 IS THE NAME OF NEW ADULTS ON PAGE 2--WILL FIND MOTHER BASED ON ROW NUMBER, ASSUMING IDNOS ARE GIVEN IN ORDER}
   SKIP TO SD02;
  ELSEIF $=99 THEN
   SKIP TO SD02;
   ENDIF;
PROC SD08
If SD08=0 and SD07=1 then skip to SD10A(1) elseif SD08=0 and SD07=0 then skip to K01 endif;
PROC SD09A
if SD09A=88 then endgroup endif;
PROC SD10A
if SD10A=88 then endgroup endif;
PROC K01B
```

Figure 18: Entry check to control entry for SD08 (section to be filled by households who gave or sold SP vines only) vine diffusion Section

If exchange number was not applicable (N/A) was checked with 88 and other entries skipped.

The completed data form was exported to SPSS and comparisons made before and after analysis.

D08 If anyone in your household gave out or sold any SP vines please fill the table below

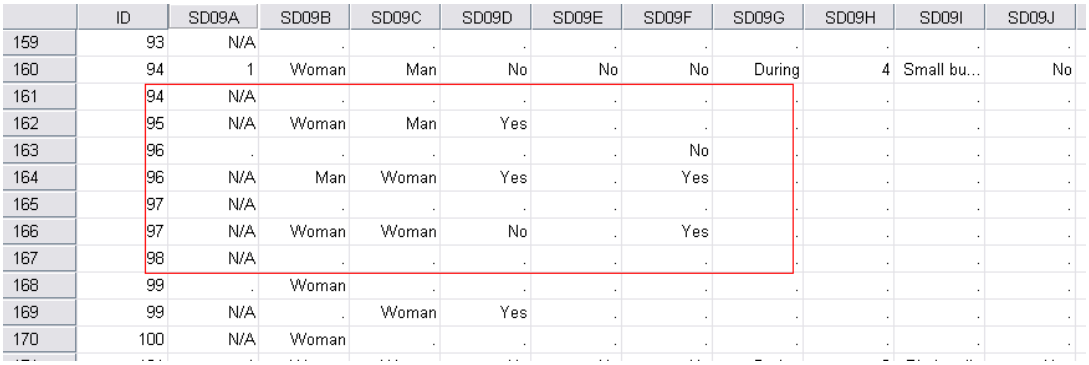
	Exchange number	Recipient	Gender of giver	Is the giver a specialised vine multiplier	Is the receiver a farmer group member	Is receiver a relative	Gave the SP vines before rains started 2-3 wks into the rainy season or well after most of the rains have fallen	Quantity of vines exchanged	Unit	Did you sell the SP vines	Total value of SP vines sold	Money	Who received the money
1													
2													
3													
4													
5													
6													
7													

Figure 19: Section of the questionnaire in CSPRO to be used to capture data for record 10 then export it to SPSS for analysis

4.2.2.0 Comparisons of data analyzed with and without data entry checks in SPSS Data View

Data view in SPSS without entry checks

SD094A (exchange number) is captured as not applicable but other fields (SD09B-recipient, SD09C-gender of giver, SD09D-specialised vine multiplier, SD09E-receiver a farmer group member?, SD09F- receiver relative?, SD09G- vines given before 2-3 weeks rain start?, SD09H- quantity of vines, SD09I-units, SD09J- did you sell vines) are completed creating confusion whether the exchange number was erroneously keyed in as N/A or the other fields were keyed in erroneously.



	ID	SD09A	SD09B	SD09C	SD09D	SD09E	SD09F	SD09G	SD09H	SD09I	SD09J
159	93	N/A	.	.	.	.	.	.	.	.	.
160	94	1	Woman	Man	No	No	No	During	4	Small bu...	No
161	94	N/A	.	.	.	.	.	.	.	.	.
162	95	N/A	Woman	Man	Yes	.	.	.	.	.	.
163	96	.	.	.	.	.	No	.	.	.	.
164	96	N/A	Man	Woman	Yes	.	Yes	.	.	.	.
165	97	N/A	.	.	.	.	.	.	.	.	.
166	97	N/A	Woman	Woman	No	.	Yes	.	.	.	.
167	98	N/A	.	.	.	.	.	.	.	.	.
168	99	.	Woman	.	.	.	.	.	.	.	.
169	99	N/A	.	Woman	Yes	.	.	.	.	.	.
170	100	N/A	Woman	.	.	.	.	.	.	.	.

Figure 20: Data view in SPSS without Checks entries were made despite the exchange number being N/A.

Data view in SPSS with data entry checks

Data view in SPSS with data entry checks the entries were barred and where SD09A is keyed in as N/A then subsequent entries were left blank (blacked out).

	ID	SD09A	SD09B	SD09C	SD09D	SD09E	SD09F	SD09G	SD09
160	94	1	Woman	Man	No	No	No	During	
161	94	N/A							
162	95	N/A							
163	96	.							
164	96	N/A							
165	97	N/A							
166	97	N/A							
167	98	N/A							
168	99	.							
169	99	N/A							
170	100	N/A							

Figure 21: Data view in SPSS with checks If SD09A (exchange number) is N/A then the subsequent entries are blank hence block lines

Variable view in SPSS without entry checks

Figure 22show variable view without entry checks for numerical, missing and level of measurement for the data as exported from CSPRO entries without entry checks. Several kinds of missing values were used some entered as none, 99, 1000. The level of measurement was given as scale for all variables making them amenable to quantitative manipulations.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
ID	Numeric	5	0	ID	None	None	7	Right	Scale
SD09A	Numeric	2	0	No.	{88, N/A}...	99	7	Right	Scale
SD09B	Numeric	1	0	Recipient	{0, Woman}...	1000	7	Right	Scale
SD09C	Numeric	1	0	Gender of giver	{0, Woman}...	None	7	Right	Scale
SD09D	Numeric	1	0	Is the giver a specialised vine multiplier	{0, No}...	9	7	Right	Scale
SD09E	Numeric	1	0	Is the receiver a farmer group member	{0, No}...	9	7	Right	Scale
SD09F	Numeric	1	0	Is receiver a relative	{0, No}...	9	7	Right	Scale
SD09G	Numeric	1	0	Gave the SP vines before rains started ...	{1, Before}...	999	7	Right	Scale
SD09H	Numeric	5	0	Quantity of vines exchanged	None	None	7	Right	Scale
SD09I	Numeric	2	0	Unit	{1, Kg}...	None	7	Right	Scale
SD09J	Numeric	1	0	Did you sell the SP vines	{0, No}...	9	7	Right	Scale
SD09K	Numeric	10	0	Total value of SP vines sold	None	None	12	Right	Scale
SD09L	Numeric	1	0	Money	{1, Tshs}...	9	7	Right	Scale
SD09M	Numeric	1	0	Who received the money	{0, Woman}...	9	7	Right	Scale

Figure 22: Variable definition without Checks

The data was wrongly coded. There was a mix up in variable definition of missing variables and measurement level. All variables were defined as scale therefore; all the variables were wrongly made amenable to quantitative processing.

Variable view in SPSS with entry checks

The entry checks used controlled entries in missing variable column and measurement column. Not applicable (N/A) variable were checked with 88, continuous missing variable checked with 99 and discrete missing variable checked with 9. Measurement levels were checked into either nominal, scale and ordinal as shown below

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
	ID	Numeric	5	0	ID	None	None	7	Right	Scale
	SD09A	Numeric	2	0	No.	{88, N/A}...	99	7	Right	Nominal
	SD09B	Numeric	1	0	Recipient	{0, Woman}...	None	7	Right	Nominal
	SD09C	Numeric	1	0	Gender of giver	{0, Woman}...	None	7	Right	Nominal
	SD09D	Numeric	1	0	Is the giver a specialised vine multiplier	{0, No}...	9	7	Right	Nominal
	SD09E	Numeric	1	0	Is the receiver a farmer group member	{0, No}...	9	7	Right	Nominal
	SD09F	Numeric	1	0	Is receiver a relative	{0, No}...	9	7	Right	Nominal
	SD09G	Numeric	1	0	Gave the SP vines before rains started 2-3 wk...	{1, Before}...	None	7	Right	Nominal
	SD09H	Numeric	5	0	Quantity of vines exchanged	None	None	7	Right	Nominal
	SD09I	Numeric	2	0	Unit	{1, Kg}...	None	7	Right	Nominal
	SD09J	Numeric	1	0	Did you sell the SP vines	{0, No}...	9	7	Right	Nominal
	SD09K	Numeric	10	0	Total value of SP vines sold	None	None	12	Right	Scale
	SD09L	Numeric	1	0	Money	{1, Tshs}...	9	7	Right	Nominal
	SD09M	Numeric	1	0	Who received the money	{0, Woman}...	9	7	Right	Nominal

Figure 23: Variable definition in SPSS with checks for missing values and level of measurements.

Comparison of sample exploratory data analysis- vine diffusion record No. 10

The two case scenarios were compared through exploratory data analysis on variables captured in SPSS data view on SD08.Quick descriptive were ran and comparison made through significance tests (T test and chi square) and subsequent conclusions from the two cases compared (Table 19).

Comparison of spread of SD09K (SP sales in TZshs of with and without entry checks using box plots

There are isolated cases of outliers with exceptional amounts corresponding to unlikely units of sale i.e. small bundle and big bundle prompting cross checking from raw data in both cases of with and without entry checks (figure 24 and 25).

Box plot 1: without entry checks. There was a tendency of the outliers to be concentrated on higher values in most of the cases forcing the spread from the median to either side of the quarters to be pulled towards smaller values.

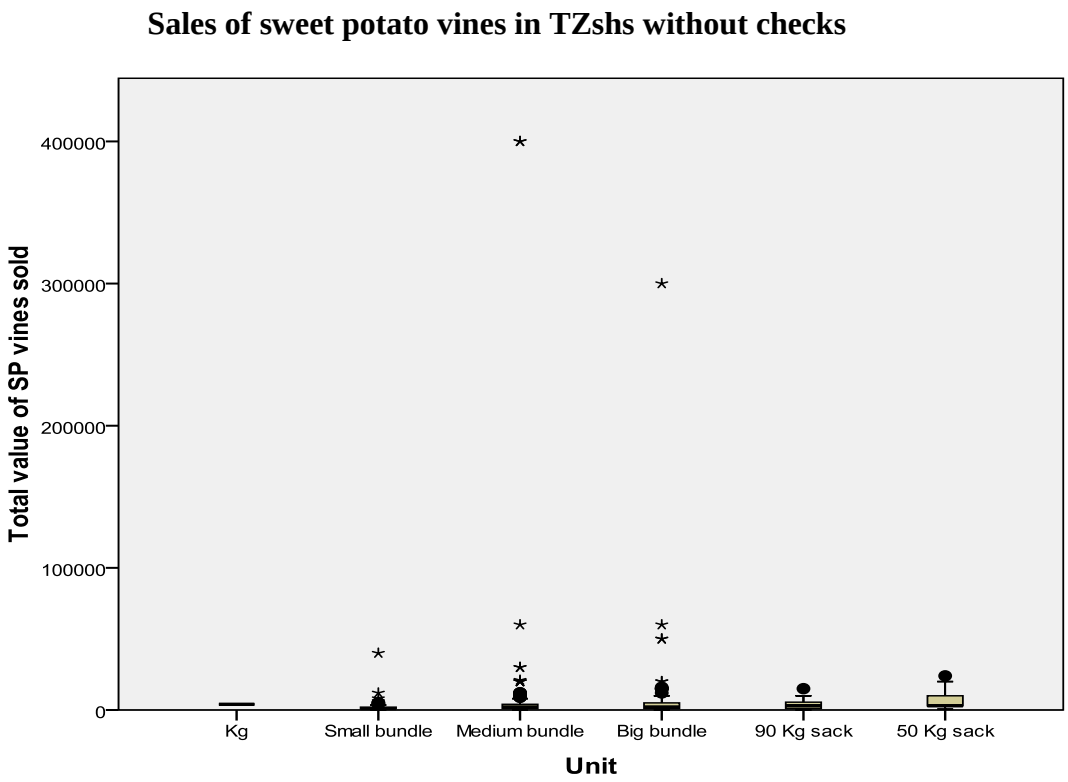


Figure 24: Box plot without checks showing spread of SP sales in TZshs (SD09K) by unit of Sale- without checks, outliers tending to higher values.

Box plot 2: Showing spread of SP sales in TZshs (SD09K) by unit of sale- with checks The outliers across all units appear random and within the range of between TZshs 5,000 to 60,000. The outliers are also consistent with price differentials within different regions. Greatest spread observed in 50 kg sack with second few outliers.

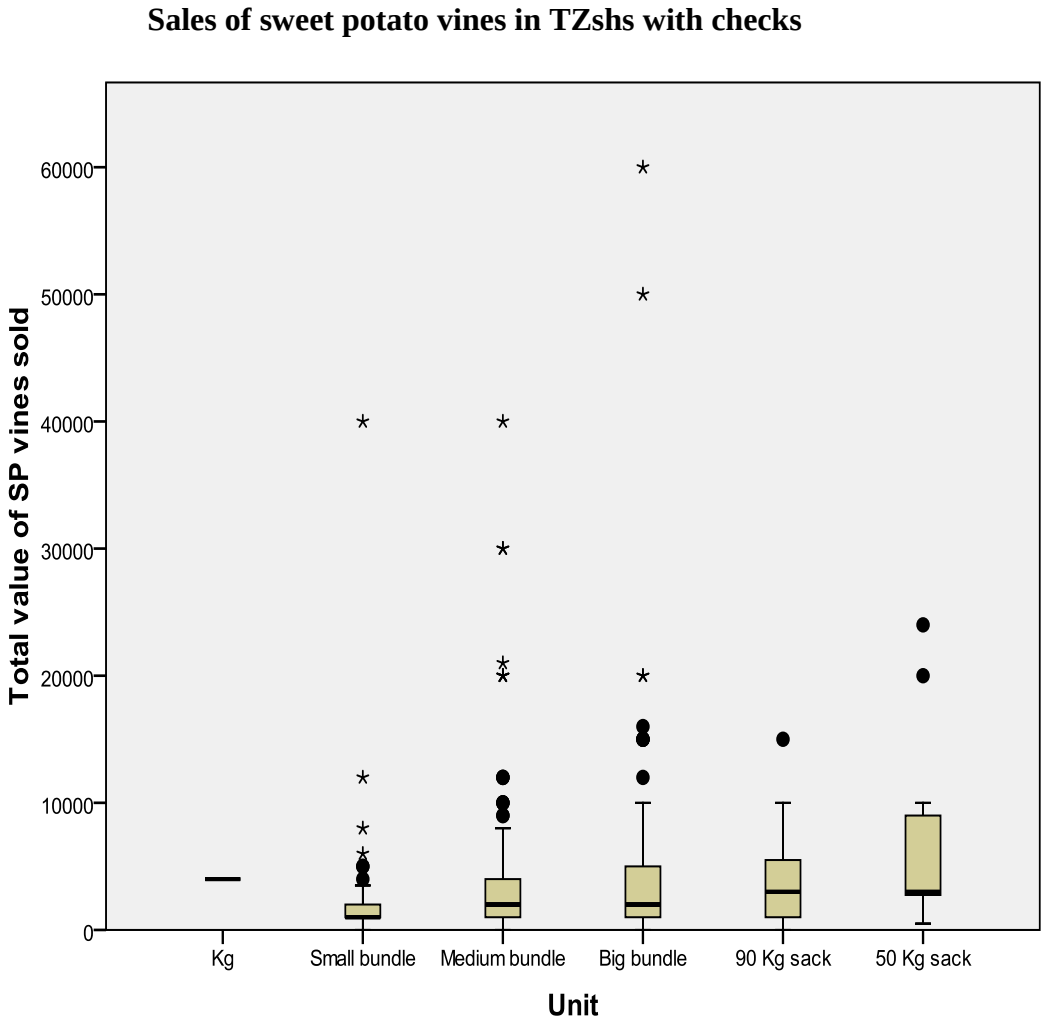


Figure 25: Box plot with checks showing spread of SP Sales in TZshs (SD09K) by unit of sale- without checks. Outliers more randomly distributed and the general spread of the SP sales is consistent with price differential as a result of differences attributable to region pricing regimes.

Comparison of the findings from the Output with and without data entry

Results from two different case scenarios differ markedly. Significance tests at 5% significance level for both chi square and t test were significant for all the four variables used for case study. The difference between the proportions of recipients of sweet potato from household was 1.35% for women, 1.2% men and 0.1% for NGO. A chi square test showed that the difference was significant at 5% significance level.

Difference between those who sold vines and those who did not with and without checks was 5%. A significance test using chi test at 5% significance test found the difference to be significant. The mean price for SP sale was also significantly different from a test performed at 5% significance level (Table 19).

Table 19: Comparison of the findings from the output with and without data entry checks

Variable	Conclusion with entry checks	Conclusion without entry checks	Remarks
Recipient of sweet potato vines from HH	93.8%were women, 5.8% men and 0.4% NGOs. Women represented the highest group sourcing for vines in the region	95.1% were women, men 4.6% NGOs 0.3%. Again women were highest recipient however; there is discrepancy in the actual percentages.	The proportions differ and reference to the raw data is recommended to verify the entries. A chi square significant difference at 5% significance level showed the difference was significant.
Did you sell the vines?	29.6% sold sweet potato vines while 70.4% did not sell the vines	24.1% sold the vines and 75.9% did not.	Both cases there is a difference of over 5% prompting a scrutiny. Significant chi square test at 5% showed the difference was significant
Units of issue/sale of vines	Medium bundle 46.1%, Big bundle 29.9%, Small bundle 18.3%, 90kg bag 1%, 50kg bag 3.2% and 1 Kg 0.1%	Medium bundle 46.1%, Big bundle 29.9%, Small bundle 18.3%, 90kg bag 1%, 50kg bag 3.2%, 1 Kg 0.1%, The figures were the same.	Same for both cases

Table 19 (continued): Comparison of the findings from the output with and without data entry checks

Mean price/range	Range=60,000, Mean price=TZshs 3,627.51, SD=TZshs 5,828	Range = TZshs400,000, Mean= TZshs 6,081.74, SD=TZshs 29,925.3	The difference between the two cases was significant at 5% T Test and reference to the raw data revealed erroneous entries shown as extreme outliers by **box plot 1**
Box plot on Vines sales **box plot 1 &2**	With an exception of 1kg unit all other units had divergent variations consistent with the unit in question.	Extreme variability was witnessed from the box plot and close check prompted.	Visual difference in variability between the two cases. An indication of varied outliers in the two cases. Checking of the raw data showed erroneous entries of Tzshs 400,000 and Tzshs of 300,000 instead of Tzshs 40,000 and Tzshs 30,000 respectively.

4.3.0 Evaluate and review reports, project proposals and literature on sweet potato seed systems

4.3.1 Evaluation of project proposals

Table 20 below show the evaluation results for the project proposal by Dr Kirimi J.S. On the general criteria the proposal was awarded 60% and 20% on the specific criteria which was the maximum score.

Table20: Evaluation of project proposal byDr. Kirimi:Development of high value sweet potato products value chain in Rwanda 2010/2011,

General criteria Technical Quality of proposal	Points awarded	Specific criteria: Relevance to programs' objective	Points awarded
Background / Introduction	10	Contribution to food security	10
Study Hypothesis /research questions	10	Contribution to community/ national development goals	10
Study Design / Protocol Development	10		
Rationale / Study Objective	10		
Preliminary Statistical Design	0		
Schema / Flowchart	10		
Communication Plan	10		
Dealing with Results	0		
Total	60		20

4.3.2 Evaluation of the project proposal by Dr.MCEwan

The proposal scored 40% on the general criteria and 20% on the specific criteria making a total score of 60%.

Table 21: Evaluation of project proposal by MCEwan: Going to scale with sweet potato vines distribution in Tanzania

General criteria Technical Quality of proposal	Points awarded	Specific criteria: Relevance to programs' objective	Points awarded
Background / Introduction	10	Contribution to food security	10
Study Hypothesis /research questions	0	Contribution to community/ national development goals	10
Study Design / Protocol Development	10		
Rationale / Study Objective	10		
Preliminary Statistical Design	0		
Schema / Flowchart	0		
Communication Plan	10		
Dealing with Results	0		
Total	40		20

4.3.3 Literature review on sweet potato seed system in East Africa

Identification of published and unpublished material on sweet potato seed systems

Both electronic sources, journals and other works done on sweet potato seed system were identified and listed in alphabetical order as per table below. The requisite literature source was derived from a total of 41 both published and unpublished documents. The most important ones were listed in the table 22. The literature was synthesized into a document titled Sweet potato Seed System in East Africa.

Table22: Articles considered relevant for literature citation in sweet potato seed system paper

Electronic Materials	Journals and other Materials
CGIAR 2005, Sweet potato, Research and Impacts: http//www.cgiar.org/impact/ research/sweet potato.html retrieved Dec 2010	Aritua, V., Alicia, T., Carey, E. E., & Gibson, R. W. 1998. Aspects of resistance to sweet potato virus disease in sweet potato. *Annals of Applied Biology* 132:387–398.
Farm concern (2010)- www.familyconcern.net accessed on 28 January 2011	Ateka, E. M., Njeru, R. W., Kibaru, A. G., Kimenju, J. W., Barg, E., Gibson, R. W., &Vetten, H. J. 2004. Identification and distribution of viruses infecting sweet potato in Kenya. *Annals of Applied Biology* 144:371–379.
	Bashaasha, B., Mwanga, R. O. M., Ocittip'Obwoya, P., &Ewell, P. T. 1995. *Sweet potato in the farming and food systems of Uganda: a farm survey report*.

Table22 (continued): Articles considered relevant for literature citation in sweet potato seed system paper

	Carey, E. E., Gichuki, S. T., Mwanga, R. O. M., Kasule, S., Fuentes, S., Macharia, C., & Gibson, R.W. (1998). Sweet potato viruses in Uganda and Kenya: results of a survey. *Proceedings of the Sixth Triennial Symposium of the International Society of Tropical Root Crops - Africa Branch (ISTRC-AB) on Root Crops and Poverty Alleviation*. 22-28 October 1995, Lilongwe, Malawi, pp457-461.
	Fuglie, K. O. 2007. Priorities for sweet potato research in developing countries: Results of a survey. *HortScience* 42:1200-1206.
	Fuglie, K. O., Zhang, L., Salazar, L., & Walker, T. 1999. *Economic impact of virus-free sweet potato planting material in Shandong Province, China*. International Potato Center, Lima, Peru, 27pp.
	Gibson, R. W., Mpembe, I., Alicai, T., Carey, E. E., Mwanga, R. O. M., Seal, S. E., &Vetten, H. J. 1998. Symptoms, aetiology and serological analysis of sweet potato virus disease in Uganda. *Plant Patholology* 47:95–102.
	Gibson, R. W., Mwanga, R. O. M., Kasule, S., Mpembe, I., & Carey, E. E. 1997. Apparent absence of viruses in most symptomless field-grown sweet potato in Uganda. *Annals of Applied Biology* 130:481–490.
	Gibson, R. W., Mwanga, R. O. M., Namanda, S. , Jeremiah, S. C., Barker, I. 2009. *Review of sweet potato seed systems in East and Southern Africa.* International Potato Center (CIP), Lima, Peru. Integrated Crop Management Working Paper 2009-1. 48 p.

Table22 (continued): Articles considered relevant for literature citation in sweet potato seed system paper

	Hall, A., Bockett, G., &Nahdy, S. 1998. Sweet potato postharvest systems in Uganda: Strategies, constraints and potentials. *Social Science Department Working Paper 1998-7. International Potato Center (CIP). Lima, Peru*. 99pp.
	Kapinga, R., Andrade, M.,Lemaga, B., Gani, A., Crissman, C. &Mwanga, R. 2005a. Role of orange-fleshed sweet potato in disaster mitigation: Experiences form East and Southern Africa. *African Crop Science Conference Proceedings (Edit: Tenywa, J. S., Adipala, E., Nampala, P., Tusiime, G., Okori, P., &Kyamuhangire, W.)* 7:1321–1329.
	Kapinga, R., Chirimi, B., Kiflewahid, B., Amour, R., & Carey, T. 2000. Rapid dissemination of improved sweet potato varieties through informal seed multiplication and distribution channels: experiences from the Lake Zone of Tanzania. *Proceedings of the 5th Triennial Conference of the African Potato Association, 29 May – 2 June 2000, Kampala, Uganda (Editors: Adipala, E., Nampala, P. &Osiru, M.).* pp91–98.

Table22 (continued): Articles considered relevant for literature citation in sweet potato seed system paper

	Kapinga, R. E., Jeremiah, S. C., Kileo, R, &Ewell, P. T. 1998. Sweet potato in Tanzanian farming and food systems. *Proceedings of the 6th Triennial Symposium of the International Society for Tropical Root Crops – Africa Branch, [Edit: Akoroda, M. O., &Ekanayake, I. J.], held in Lilongwe, Malawi, 22 – 28 October, 1995*.Pp528–535.
	Kapinga, R., Tumwegamire, S., Lemaga, B., Andrade, M., Mwanga, R., Mtunda, K., Ndolo, P., Nsumba, J., Agili, S., &Serwadda, B. 2005b. Development of farmer based seed systems for healthy planting material and increased sweet potato production in East and Southern Africa. *African Crop Science Conference Proceedings (Edit: Tenywa, J. S., Adipala, E., Nampala, P., Tusiime, G., Okori, P. &Kyamuhangire, W.)* 7:1169–1173.
	Mwanbene, R. O. F,Mwakyembe, C. M. A., & Mayora, C. M. 1994. Exploratory study of the farmers' view point on the production of sweet potato in the southern highlands of Tanzania. *Proceedings of the 5th Triennial Symposium of the International Society for Tropical Root Crops – Africa Branch*, [Edit: Akoroda, M. O.] held in *Kampala, Uganda, 22 – 28 November, 1992*.Pp218-224.
	Ndamagé, G. 1990. Case study "The Rwandan Promotion Programme for Sweet Potato". *Proceedings of the International Conference on: Roots, Tubers and Legumes Potential and limits for bridging nutritional gaps and food shortages in African countries. 2–6 October, 1989 Felestag, Germany*. Pp185–205.
	Tairo, F., Kullaya, A., &Valkonen, J. P. T. 2004. Incidence of viruses infecting sweet potato in Tanzania. *Plant Disease* 88:916

Additional material considered (appendix 6)

Chapter 5

5.0 Discussion

Training was undertaken to equip researchers with skills in data management by using CSPRO Version 4 for data entry, data checking and organization before exporting it to an analytical package (SPSS Version 18). The training followed the training cycle (figure2) whereby a TNA, training evaluation and follow up on adoption of the data management package were conducted to evaluate the knowledge gaps, effectiveness of the training and extent of adoption by the scientists of CSPRO for their data management processes.

Majority (77.8%) of the data entry clerks had education level of bachelors' degree and above with the least level of education being diploma. This can be attributed to the academic threshold set by the organization for employment or engagement. All the trainees had an experience in one or more facet of data management, it followed that they all understood largely the requirements of quality data in data management and less time would be spent on emphasizing the same principles. On consideration of the education level, experience in data management and good computer skills the training implementation was set to be elaborate and more problem based.

All the clerks had superior computing knowledge that could be shared with group members during data entry exercise. However, the analytical packages were not very popular with the data entry clerks. CSPRO was not an exception with 57% of the data clerks having competence in it and 28.6% and 14.4% having no skill and some skill respectively. Between these groups28.6% had no previous experience in use of CSPRO and it is on this basis that exposition to CSPRO in the following areas was deemed fit as per rating by the clerks:

- Starting CSPRO program
- Creation of the data dictionary
- Creation of the data entry form
- Entering data
- Adding edits to the data entry application

The above operations were to be done practically through the use of CSPRO version 4 practice guide "CSPRO getting *started version 4, by international program centre population division, US Census Bureau 2010".* Majority of data entry clerks having had prior experience in use of CSPRO for various application, they had transferable skills and the training was problem based and their input was instrumental in mentoring the new entrants. Close supervision and guidance was given where necessary.

Due to the timeframe before data entry, some areas that the data entry clerks suggested to be exposed to were limited to those areas which were critical and had immediate concern to the task at hand in addition to the five identified areas mentioned above. These were; merging files created by different clerks, running a double entry for verification and adding edits to the data entry application.

The course met its training objectives and areas noted for inclusion in future training signified different interest among the data entry clerks. This can be explained by the different levels of exposure by the participants and it is recommended that selection criteria for data entry clerks should be devised and a given threshold defined to avoid overemphasis on elementary tasks in training. However, it was apparent more emphasis should be given to creation of edit logic in data entry application over and above creation of the data dictionary and data entry forms and this was emphasized in data entry for the Tanzanian survey which was done practically.

According to Yang Z., et al (2004), usage metric is the first measurement of adoption and a basic indicator of success is the number of new data that had been managed, in this case, data entry, organization and checking performed through CSPRO program. This is evidenced by the paradigm shift in data management from the usual data entry into an analytical package and then proceeding with the data management process – analysis, reporting and dissemination to a more proactive process of data planning, data entry, organization and checking to minimize errors

prone to earlier on practice after the training. The usage metric is a clear pointer to willingness of the research team in making CSPRO an integral part of the data management process. Both the data quality and performance metric show concerted effort by researchers to continue promoting the program at a wider scale through recommendation to other scientists owing to its quality outputs and superiority in comparison to other packages known to the team.

The respondents through training had become competent and possessed excellent skills in CSPRO application, which could be shared amongst scientists at CIP in several areas; creation of data dictionary, data entry forms, running data entry and creating the edit logic
in data entry application. This formed an impressive cohort to promote effective data management practices in the entire organization. Training is in no doubt an influencing factor in adoption of data management practices through practical exposition of the underlying benefits of the practices, which promote the eventual quality of the final research output.

Data entry checking and organization play crucial role in the quality and validity of eventual output from analyses of the data (Emory, 1995; Nachimias, 1996; Sekaran, 2003). They considerably shorten the time spent on data verification, analysis and lend credibility to the conclusions and recommendations made from the data. Data entry in a platform of checked data entry (edit logic) was more efficient and effective in terms of the time spent for the actual data entry, data organization and crosschecking of the raw data compared to when edit checks were not inbuilt before data entry.

Fewer errors were experienced with checked data and instances of erroneous entries were minimal and this translated to less time spent in cross checking of the raw data to verify the entries during exploratory analysis.

Data verification i.e. double data entry for unchecked data was cumbersome and did not entirely guarantee accurate entries. Further, a detailed edit trail had to be kept to track the changes in the double data entry which were many and confusing to the data entry clerks. These challenges were solved through checking of the data entry process through creation of edit logic which worked effectively for the double data entry and all erroneous entries rectified within a framework of a second data entry.

Analysis output from the checked and unchecked data differed significantly. Therefore, data checking play a crucial role in ensuring quality output and credible conclusions drawn from the analysis output

There was no standardized format of evaluating project proposals and undertaking literature review. The followed format though developed consultatively with the head of M&E it needs refinement with input from scientists from all other component. However, when the two project proposals were subjected to the criteria, the only difference on the score was due to the general criteria. This criterion considered mainly the technical aspect of the proposal, which had more weight than the specific criterion.

The sweet potato knowledge portal formed the critical source of literature materials considered for review. However, other sources which constituted journals and internet sites were also used and documented for referencing.

Chapter 6

6.0 Conclusion and Recommendations

6.1 conclusions

- High adoption (83%) of CSPRO software for data management was achieved through training. A total of 12 projects were managed (creation of data entry application, data checking and organization) before analyzes by the researchers after the training. Before and after training scenario differed markedly in terms of competence rating on the use of the software. The rating changed from no skill at all and some skill to competent and possession of excellent skills (91.7%) after the training
- The process of data checking and organization minimizes the incidences of erroneous entries, thus giving more precise conclusions from the eventual analyzes output. This is deduced from the difference in the two case scenarios whereby in one case data was entered without validation and the analysis output compared with the output that was gotten from checked data. The difference was statistically significant; hence data checking and organization play a crucial role in the quality and validity of the eventual output from the analyses of the data
- Evaluation and review of project proposals and literature in an organization not only targets them more specifically to the objective of the study but also improves on the technical and relevance of the projects to the programs objectives'. Further a uniform format is promoted across the institutions approach

6.2 Recommendations

It is therefore recommended that;

- CIP should design and promote training on data management for efficient and effective delivery on project outputs.
- Scientists to embrace the data validation and organization before analyses in order to increase precision and validity of the conclusions from analysis output

- The institution to formulate a standardized criteria of evaluating and reviewing proposals and literature

References

Angela, C. (2010). The training cycle, guidelines and tool kit, available on http://www.childhope.org.uk /wcore/showdoc.asp

Bashaasha, B., Mwanga, R. O. M., Ocittip'Obwoya, P., &Ewell, P. T. 1995. Sweet potato farming in Uganda: a farm survey report. CIP Lima Peru

CGIAR, (2005), Sweet potato, research and impacts: available on http://www.cgiar.org/impact/research/sweetpotato.html

CIP, (2010). Sweet potato distribution; available on http://www.cippotato.org/

CGIAR, (2010). Research funding in CGIAR; available on http:www//cgiar.org/

Cummings, G.T., Worley, G.C.(2005). Organizational development and change. 8th edn. South-western publishing. Canada

DDI structural reform group (2004).Data management cycle.DDI Alliance; available on http://www.icpsr. umich.edu

Emory, C.W., Cooper, R.D. (1995). Business research methods. 5th edn. McGraw-Hill London.

Ghosh, B. (2000). Human resource development and management. Vikas publishing house PVT Ltd. New delhi.

Infobase Ltd. (2007). Training and development; available on http://www.traininganddevelopment naukrihub.com

JSI Word ED (2011).Planning and implementing, a community handbook for community training; available on http:www.traininganddevelopment.naukrihub.com

Kapinga, R. E., Ewell, P. T., Jeremiah, S. C., & Kileo, R. 1995. Sweet potato in Tanzanian farming and food systems: implications for research. International Potato Center (CIP)/ Ministry of Agriculture, Tanzania. 47pp.

Kapinga, R., Chirimi, B., Kiflewahid, B., Amour, R., & Carey, T. 2000. Rapid dissemination of improved sweet potato varieties through informal seed multiplication and distribution channels: experiences from the Lake Zone of Tanzania. Proceedings of the 5th Triennial Conference of the African

Potato Association, 29 May – 2 June 2000, Kampala, Uganda (Editors: Adipala, E., Nampala, P. &Osiru, M.). pp91–98.

Kapinga, R., Andrade, M., Lemaga, B., Gani, A., Crissman, C. &Mwanga, R. 2005a. Role of orange-fleshed sweet potato in disaster mitigation: Experiences form East and Southern Africa. African Crop Science Conference Proceedings (Edit: Tenywa, J. S., Adipala, E., Nampala, P., Tusiime, G., Okori, P., & Kyamuhangire, W.) 7:1321–1329.

Kapinga, R., Tumwegamire, S., Lemaga, B., Andrade, M., Mwanga, R., Mtunda, K., Ndolo, P., Nsumba, J., Agili, S., &Serwadda, B. 2005b. Development of farmer based seed systems for healthy planting material and increased sweet potato production in East and Southern Africa. African Crop Science Conference Proceedings (Edit: Tenywa,J.S., Adipala, E., Nampala, P., Tusiime, G., Okori, P. &Kyamuhangire, W.) 7:1169– 1173.

Kubr, M. (1998).Management consulting. 2nd edn. ILO Geneva

Nachmias,C,F, Nachamias D. (1996).Research Methods in the Social Sciences 5th Edition.St Martins Press Inc.

Ndamagé, G. 1990. Case study "The Rwandan Promotion Programme for Sweet Potato". Proceedings of the International Conference on: Roots, Tubers and Legumes Potential and limits for bridging nutritional gaps and food shortages in African countries. 2–6 October, 1989 Felestag, Germany. Pp185–205

North Illinois University, (2005). Data analysis; available on http//:www.niu.edu/rportal

Tylor, D, (2010). Writing literature review: a few tips of doing it; available on http//:www.writing.autoronto.ca

Resnik.D. (2000). Statistics, ethics, and research; an agenda for education and reforms. accountability in research. London press

Robinson,C.,(2000) Real world research (2nd Edition) Oxford Blackwel

SASHA, (2010), First annual technical report summary, available on http//:www.cippotato.org

Savenye, Wilhelmina,C., Robinson, R. (2004). Qualitative research issues and methods: An introduction for instructional technologists. Handbook of research on educational communications and technology. Lawrence Erlbaum publishing.

Sekaran, U. (2003). Research methods for business. 4th edition John Wiley & sons Inc.

Shamoo,A.E.,Resnik,B,R. (2003). Responsible Conduct of Research.Oxford University Press London.

Shamoo, A.E., Resnik, B.R. (2003). Responsible Conduct of Research. Oxford University Press.

Shephard, R.J. (2002). Ethics in Exercising Research. Journal of Sports Med, 32 (3)

SUNY, (2011). Research design; available on http//:www.suny.edu/govtrelations

Statistical service centre (2010). Data Flow: organizing action on Research Methods and Data Management. Research Methods Support for Collaborative Crop Research Program (CCRP) projects. Mc Knight Foundation.

Troachim W.M (2006). Data analysis knowledge base; available on http://www.socialresearchmethods. net/kb/index.php

Yang, Z, Cai, S, Zhou Z, Zhou N, (2004). Development and validation of an instrument to measure user perceived services quality of information presenting web portals. Journal of information and management, 42 (2005) 575-589

Yin, R.K.(1989). Case study research: Design and methods. Revised edition. London press

Appendices

Appendix 1: TNA Questionnaire

Needs Assessment Questionnaire CSPro

We would be grateful if you could complete this questionnaire to help us understand the
different skills, needs and interests within the CSPro orientation before formal data entry.

1. **Personal information**

First Name: ……..
Last Name: ……

Gender ( )

2. **Education level**

(a) Education level: Bachelor level ( ) postgraduate degree ( ) diploma ( )

3. **Experience:**

(i) What is your experience to date in data management?

.....

.....

.....

..... (ii) Which of these functions have you performed before in data management?

Data collection (X) Data organization (X) Data cleaning (X) Data entry ( ) Data analysis (X)

Other (specify) ……

.....

.....4. **Computer Knowledge**

(i) Please provide information below on your computer literacy. For each package rank your perceived level of skills/expertise:

1. Representing no skill or expertise

2. Some skill or expertise

3. Competent

4. Denoting excellence that could be shared with group members during data entry exercise

Computer package	**1**	**2**	**3**	**4**
Word processor				
excel				
access				
SPSS				
GENSTAT				
CSPro				

(ii) Have you ever used CSPro before? Yes ( ) No ( )

If yes in (ii) above Please provide the information requested below to help us identify the key areas on which to focus on during induction to CSPro. This information will be kept confidential. For each statement rank your perceived level of skill/ expertise:

1. Representing no skill or expertise

2. Some skill or expertise

3. Competent

4. Denoting excellence that could be shared with other group members during data entry

Activity in CSPro	**1**	**2**	**3**	**4**
Starting CSPro				
Creation of the data dictionary				
Creation of the data entry form				
Entering data				
Adding edits to the data entry application				

(ii) Which other area in CSPro package would you like exposed to? .....

Thank you for your time and cooperation. We look forward to meeting and working with you in the data entry exercise.

Appendix 2: CSPRO Training Evaluation Questionnaire

Training Session Title: **CSPro**Date: ……..
Location: Nairobi……..
Presenter: ……

Please evaluate each of the following aspects of the training program by circling a number on the scale below.

	excellent	good	fair	unsatisfactory	Not Applicable
Achievement of program objectives					
Achievement of personal expectrations					
Relevancy of content to my needs and interests					
Organization of the program					
Usefulness of visual aids and handouts					
Trainer's knowledge					
Match between content and my questions					
Trainers ability to explain content clearly					
Trainers ability to respond well to questions					

2. The length of the training was:

Too long ☐ Too short ☐ just right ☐

3. The level of material covered was:

Too high. I felt Overwhelmed. ☐ Just right. I got what needed ☐ beneath me ☐

I needed something more ☐

4. I wish we would have spent more time discussing

……..……
…
…….……
……

5. Next time, they should include additional topics such as

……..……
…
…….……
……

6. Please evaluate each of the following aspects of the training session by circling a number on the scale below:

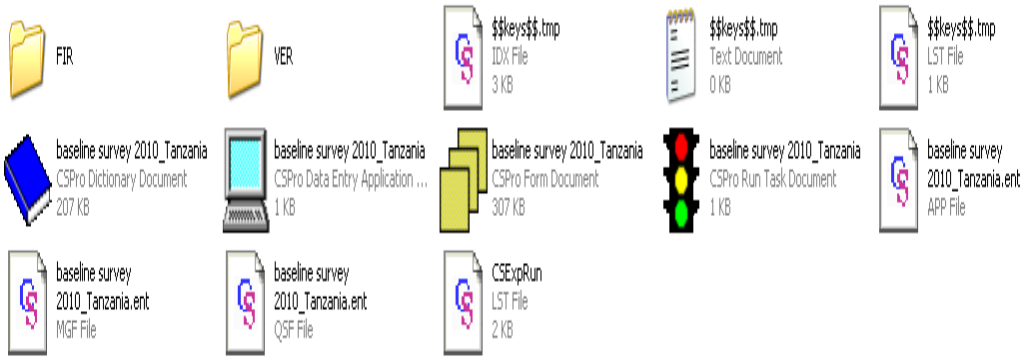
	excellent	good	fair	unsatisfactory	Not applicable
Versatility of computer					N/A
Organization of the training					N/A
Administrative support					N/A
					N/A

7. If you rated your experience as anything less than excellent, please tell us what we could
do to improve our efforts.

……..……
……
……..……
……

…….....……
…….....……
…….....……
…….....……
…….....……
…….....……
Thank you very much for your frank contributions. Your feedback will help us improve our efforts in order to serve you better.

Appendix 3: CSPRO Baseline line survey data files



Appendix 4: Adoption Questionnaire

Adoption of CSPRO in Data Management among data entry clerks and scientists at CIP

We would be grateful if you could complete this questionnaire to help us understand the effectiveness and bottlenecks in using CSPRO for data entry.

2. **Personal information**

First Name: …… Last Name: ……
Designation……..
Gender ( )

4. Education level

(b) Education level: Bachelor level ( ) postgraduate degree (X) diploma ( )

5. Experience in use of data management program:

Fill the following table on number of data management packages you have used before;

S/No	Data management package	Previous use	
		yes	No
1	Excel		
2	SPSS		
3	Genstat		
4	Access		
5	CSPRO		
6	Epidata		
7	Epi info		
8	Oracle		
9	Others (specify)		

6. Rate your CSPRO competence using the scales given in the following areas

This information will be kept confidential. For each statement rank your perceived level of skill/ expertise:

1. Representing no skill or expertise
2. Some skill or expertise
3. Competent
4. Denoting excellence that could be shared with other group members during data entry

Activity in CSPro	**1**	**2**	**3**	**4**
Starting CSPro				
Creation of the data dictionary				
Creation of the data entry form				
Entering data				
Adding edits to the data entry application				

7. Usage of CSPRO for Data Entry after the training

(a) Have used CSPRO for data entry after the training? Yes( ) No ( )
(b) If yes above how many projects have you created data entry application for in the last 4 months? .....
(c) If No which package did you use? .....
(d) Have you performed any other task using CSPRO? Yes ( ) No ( )
(e) If yes which one …….

8. Data Quality

(i) Use the following scale to rate the quality of data exported to SPSS from CSPRO
Excellent (1) Very Good (2) good ( ) poor quality (3)

(ii) How often did you get queries from data analyst for clarifications on the data entered in CSPRO?

Rarely (1) Not at all (2) very often (3)

9. Data management performance

(i) Would you recommend CSPRO for data management to other scientists in CIP? Yes ( ) No ( )

(ii) If yes above give the reasons……

How does CSPRO compare with other data management packages you have used before for data entry? Far much superior (1) same (2) inferior (3)

10. (i) Kindly list the challenges you encountered in using CSPRO for data management

……

……

……

……

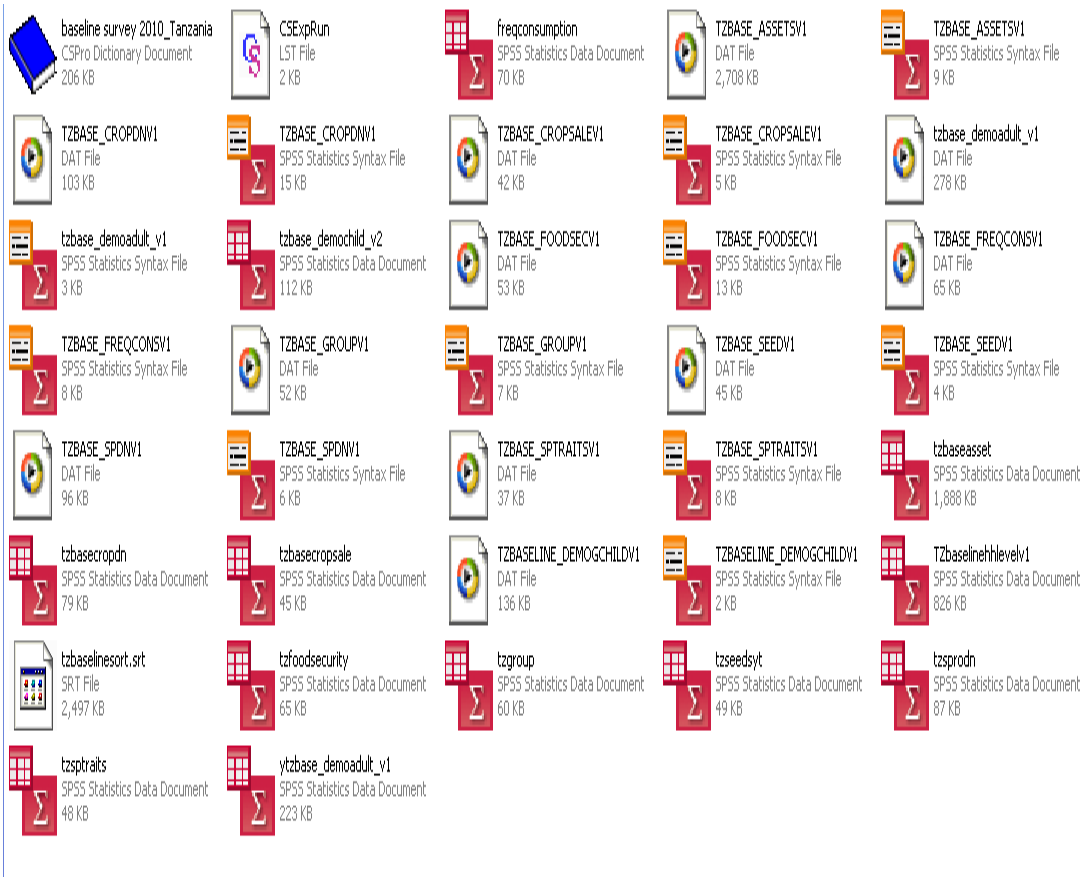
(iii) What are your suggested solutions for the above challenges

……

……

Thank you for your time and cooperation.

Appendix 5: SPSS data files from CSPRO application



Appendix 6: Index of Reviewed Papers