

Capacity building towards digitization of national vegetable databases to address regional and national priorities in food and nutritional security in Eastern Africa.

Tsvetelina Stoilova, Maarten van Zonneveld

National genebanks in East and Southern Africa conserve various food plants, but lack online databases documenting their collections. Access to seed, passport and characterization data is essential for effective germplasm conservation, and for the use of vegetable diversity for variety development. Geographical information of the location of the accessions is crucial to guide collecting missions and select germplasm based on agro-climatic adaptation. Germplasm databases group accessions according to taxonomy, and geographic and thematic data information, and must be updated once additional data on the entries become available. To establish and maintain national germplasm databases, curators need to have database management skills. The main objective of the project is to train germplasm curators from Burundi, Tanzania and Zambia on digitizing information for national germplasm databases with a focus on indigenous vegetables to improve data availability and sharing.

The biodiversity information project aims to increase availability of national biodiversity data through GBIF.org.

Through the project implementation five - days training workshop on main principles of data management: data digitizing, cleaning, editing and publishing was conducted at World Vegetable Center, Eastern and Southern Africa based in Arusha, Tanzania from 28 May to 1 June.

Project activities completed:

Data mobilization workshop

Programme for BID-AF2017-0310-SMA project 28 May - 1st June 2018

Day	Activity	Responsibility	Time
Monday Morning May 28th	Registration		8.30
	Welcome address from Regional Director		9.00
	INTRODUCTION	Trainers	
	Presentations from the participants		
	WorldVeg presentation on the project activities and achievements	Tsvetelina	
Monday Afternoon May 28th	DATA mobilization and standardizing Presentations , Data mobilization and digitization	Tsvetelina	
	Open Refine		
	Exercises, Use case 1		
Tuesday morning May 29th	DATA FORMATTING		8.30
	Decision making – traditional vegetables, cultivation and consumption. Presentation and followed by discussion	Innocent Ritte	8.30
	Introducing to Darwin Core - presentation	Tsvetelina	10.30
Tuesday afternoon May 29th	Exercises using Use Case 1 HS, excel sheet	Trainers	
	Presenting results	Trainers	
Wednesday morning May 30th	GIS INTRODUCTION	Maarten	8.30
	Presentation Maarten		
	Exercises GIS manual		
Wednesday afternoon May 30th	DATA QUALITY AND CLEANING	Maarten	
	Presentation Maarten		
	Exercises on GIS programme and presenting their results	Trainers	
Thursday morning May 31st	PUBLISHING	Tsvetelina	8.30
	Presentation on using IPT to publish	Trainers	
	Exercises on publishing data using test mode of IPT	Trainers	
Thursday afternoon May 31st	GIS AND IUCN CONSERVATION STATUS	Maarten/ Lourance	
	Presentation Maarten/ Lourance		
	Exercises GIS manual		
Friday morning June 1st	PREPARATION OWN ANALYSIS		8.30
Friday morning- after tea	Presentations from the participants	Trainers	
	Presentations on publishing from each of the participant using their own data		
	Feedback discussion		
Friday afternoon	Visit to NPGRC and ECHO East Africa		14.00



Group Photo, BID workshop 28 May – 1st June, 2018

The workshop was held in Arusha, Tanzania at World Vegetable Center, Eastern and Southern Africa. We received total of 16 participants from seven organizations (Table 1). One of the participant was representing local government and he presented the status of crop cultivation at the district level and new trends for increasing traditional vegetables growing to improve food security of the populations.

Table 1. List of participants

	Name	Affiliation	Contacts
1	Dr. Barnabas Kapange	SPGRC, Zambia	bkapange@spgrc.org.zm
2	Dr. J. Claude Bigirimana	ISABU, Burundi	clautbigg@yahoo.fr
3	Dr. Pavithra Sharan	NMAIST, Tanzania	pavisst@gmail.com
4	Dr. Mussa Ally	NMAIST, Tanzania	mussa.ally@nm-aist.ac.tz
5	William Crispo Hamisy	NPGRC, Tanzania	whamisy@hotmail.com
6	Odilia Shirima	NPGRC, Tanzania	oshirimas@yahoo.co.uk lourance65@yahoo.com
7	Laurence Mapunda	NPGRC, Tanzania	
8	M.S. Kabululu	NPGRC, Tanzania	raphlulu@yahoo.co.uk
9	Fatma Hussein	HORTI Tengeru, Tanzania	fkiruwa@gmail.com
10	Emmanuel Laswai	HORTI Tengeru, Tanzania	emmanuel.28lz2001@yahoo.com
11	Sophia Kasubi	ECHO East Africa, Tanzania	skasubi@echonet.org
12	Martha Munisi	WorldVeg, Tanzania	martha.munisi@worldveg.org
13	Salome Mushi	WorldVeg, Tanzania	salome.mushi@worldveg.org

14	Mary Matovolwa	WorldVeg, Tanzania	mary.matovolwa@worldveg.org	
15	Omary Mbwambo	WorldVeg, Tanzania	omary.mbwambo@worldveg.org	
16	Innocent Ritte	Meru District Council, Tanzania	innocent.ritte61@gmail.com	
17	Maarten van Zonneveld	WorldVeg, Taiwan	maarten.vanzonneveld@worldveg.org	facilitator
18	Tsvetelina Stoilova	WorldVeg, Tanzania	tsvetelina.stoilova@worldveg.org	facilitator

Maarten van Zonneveld provided information on digitization and cleaning data section and georeferencing and presenting data in GIS. The participants were attracted by the mapping section using DIVA GIS programme. Some of the participants were able to generate a map for crop distribution using their data. The other topics of the curriculum like, data mobilization and digitization, introduction to Open Refine and Darwin Core and Basic concept of publishing data were facilitated by Tsvetelina Stoilova.

Most of the participants heard for the first time about GBIF and got information on publishing and using data from the gbif.org website.

The first day three of the participants, Dr. Kapange (SPGRC), Dr. Bigirimana (ISABU) and Dr. Hamisy (NPGRC) presented their genebank activities and databases. Ms. Sophia Kasubi from ECHO East Africa attended this workshop as they maintain data and collection of local plants and play significant role for seed distribution to the farmers.

The presentations given on **Day 1** were **on main objectives of our project, basic principles of publishing data and data mobilization** (Presentations 1, 2 attached). The participants received information about GBIF and importance of publishing biodiversity data. The explanations were given on different datasets, occurrence, checklist and sampling event. It was very well explained why the data have to be published and become publicly available for use from different researchers and Institutions. The availability of data will allow to do a more accurate assessment of species distribution and threats of biodiversity.

The next presentations was focused on data **digitization and planning**. The digitization helps to preserve our data. Digitized format is better for use and increase visibility of the Institutions/ organizations holding the data. The information explained the importance of each step: planning for mobilization, clustering the components and make a good project plan. For all activities we have to select the right stakeholder's - resources and to use them to achieve our goal. The equipment which will need for project implementation has to be bought before project start, and train people how to use it. Good planning includes timeframe for the project activities, in which phase of the project life they need to be done: before, during or after the project implementation.

The same day the session for **Open Refine** programme was introduce to the participants. The programme was installed to each computer and participants were trained how to use it and supervised during the exercises. For this purpose we used the BID CE Workshop, Session 10. Hard and soft copies were shared with each participant.

Day 2 started with presentation of Mr. Innocent Ritte, from Meru District Council. He gave information on crops cultivation at district level and role of traditional vegetables for food security of the populations. Mr. Ritte explained that the Government still does not focus on traditional vegetables as

complimentary nutrient food. The main constraint is seeds, which are not available in the market and it is difficult to find them at the right time before rainy season. He explained that the government still does not have policy for the traditional crops, but farmers are interested to grow them as the market demand is increasing during the last decade. It is necessary to continue working on awareness creation on traditional vegetables and importance for consumption and as a source for income. The farmers have to be aware about the importance of conserving local plants and their seeds to ensure any problem can be solved in the future.

The next session was an introduction to **Darwin Core Standard**. The participants understood why it is important to use the standards instead everyone uses different sheets and ways of recording. Biodiversity information standards (TWDG) realized in the 80s the need for taxonomic information to have standards that are similar worldwide.

Using Darwin Core Standard, the data can be easily published. Darwin Core Standard provides standardized descriptors for sharing biodiversity data. It plays a fundamental role in the sharing, use and reuse of open-access biodiversity data. The Darwin Core Standard is the most common standard used.

The participants were able to do exercises on the Excel spreadsheet which was distributed to them. They were asked to fill it with their own data. At the end of this session every participant had an opportunity to present his exercised work and discussed the terms which were not enough clear to them. All participants took part of the discussion. Many participants were concerned about the missing information in their data.

Day 3 consisted of a one-day GIS training including data import, data cleaning, connecting climate data to passport data, georeferencing, and the development of species richness maps. The program DIVA-GIS was installed on each of the computers and participants received a manual as well. DIVA-GIS offers several analyses to show geographic patterns of diversity of crops. The sources of the geographical coordinates were also given to the participants. Some of them were able to generate maps on some crop distribution using their data.

The participants in the afternoon were involved in exercises to learn how to use the program and to make species richness maps.

On **Day 4** the participants learned how to publish data on GBIF using the Integrated Publishing Toolkit (IPT). For the purpose of exercising IPT test mode was used. All the participants had to organize their own data and try to publish them.

Day 5 was the last day in which every participant presented his/her results and published their data in the IPT test mode. Almost all the participants had successfully published their data and we agreed on their real publishing process to start immediately after the workshop.

Last 15 minutes were given to the participants to share their opinion for the workshop. According to their feedback the most interesting part was GIS program and presenting data for crop diversity using this program. All participants appreciated the course on data mobilization which will improve their knowledge for collecting and publishing data. Some of them shared their impression of this information about data which they did not consider so important before this workshop. The

information enabled them to share and make visible their data. Network building was also mentioned as part of the workshop.

In the afternoon the participants visited the National Plant Genetic Resources Centre (NPGRC) and ECHO East Africa.

A story from the workshop was published on WorldVeg website, with link:

<https://avrdc.org/germplasm-curators-dig-into-data/>

The materials from the workshop are attached.

Photos are attached.