Guidelines for Mapping the Preferences in Gender Sensitive Product Profiles to Crop Ontology and Creating a Consumer Segment Ontology

VERSION 1.0
FEBRUARY 2022

Elizabeth Arnaud, Sylviah Khamila, Charity Kibooga, Lilian Yiga Ndagire, Pricilla Marimo
Alliance Bioversity-CIAT

Contents

Acknowledgments

We are grateful to all researchers who answered the survey and accepted to be interviewed thus providing valuable contribution: Afola Agbona (IITA), Violet Akech (IITA), Vilma Hualla (CIP), Ivan Kabiita (NARO), Elizabeth Khakasa (NARO-NARL), Peter Kulakow (IITA), Godwill Makunde (CIP), Marie-Angélique Laporte (Alliance Bioversity-CIAT), Anne Rietveld (Alliance Bioversity-CIAT), Elisa Salas (CIP), Béla Teeken (IITA), Brigitte Uwimana (IITA). We thank Lora Forsythe, University of Greenwich, who led the development of the Template for the Product Profiles of Roots, Tubers and Bananas and provided us with her expert knowledge.

We acknowledge the important contribution of Amos Asiimwe, junior consultant at Alliance Bioversity-CIAT and Master's Student at Makerere University, Uganda, who initiated the extraction of processors' preferences for Roots, Tubers and Bananas food products and contributed to the preliminary discussions.

We also thank Anna Müller, Agricultural Socio-economist, Alliance Bioversity-CIAT, who provided feedback on the survey design and Gideon Kruseman, CIMMYT who provided literature for the definition of the consumer segments.

Citation

Arnaud E., Khamila S., Kibooga C., Ndagire L, Marimo P. - Guidelines for creating and using user-oriented ontology enabling a gender sensitive annotation of breeding product profiles. Alliance Bioversity-CIAT, February 2022.

Purpose of the Guidelines

The guidelines were developed to guide scientists, database managers and ontology curators to integrate in the ontology the proper social dimensions documenting trait preferences and use it to annotate the participatory trials or end-user surveys data stored in the databases or repositories. The development of breeding product profiles targeting specific market segments or consumer segments supposes that breeders can access interpretable information about the key preferences of the end users. The translation or interpretation of the collected preferences into traits and variables to make it interpretable and measurable by breeders is not always straightforward. A lot of contextual information needs to be included.

The Guidelines were produced for the Research Programme on Roots, tubers and Bananas (RTB) and the RTBFoods project with the support of the CGIAR Gender Platform. Provided examples are taken from the RTB Foods-Alliance Bioversity CIAT report entitled '*Gendered-food mapping on Matooke in Uganda: Understanding the Drivers of Trait Preferences and the 'Development of Multi-user RTB Product Profiles*' (Marimo P. et al, 2021) and the RTBFoods '*Gendered Food Product Profile Template*' (Forsythe et al, 2022),

This version will go through further revisions following additional feedback provided by experts.

Social Dimensions of a breeding strategy

The objective of most breeding programs in the public sector remains to address the food security issue with a social impact. Information on specific user groups along with their preferences on crop or food product qualities is required to develop product profiles that will adequately inform breeding decisions. A variety of studies have identified gendered trait preferences, but do not systematically analyse differences related to or interactions of gender with other social dimensions, household characteristics, and geographic factors (Teeken B. at al, 2021)

The RTB Breeding Community of Practice of the CGIAR Roots, Tubers and Bananas Research Programme (RTB) develops crop breeding product profiles that must integrate traits preferred by diverse social groups or market segments and include the gender dimension. It is not recommended to develop a separate product profile for gender which addresses specific traits unless if there is evidence that there is a niche product profile that is particularly relevant, and impactful for women, who may be involved across the value chain for such a product. Social scientists are exploring more nuanced examination of poverty level, food security and location and how these impact preferences. It is then crucial to relate the trait preference to individual, household- and farm-characteristics (Teeken, 2021).

Harmonizing Trait Naming and Description to Support Fieldbooks, Databases and Data Repositories

The Crop Ontology (www.cropontology.org) is a reference standard providing clear semantics (e.g., trait name, synonyms, definitions, …) for crop traits and variables recorded during experiments and surveys, supporting the harmonization and analysis of the evaluation data. The Crop Ontology is integrated in the RTB breedbases to annotate the trait data and create fieldbook templates. Most of the characteristics/traits in the current Crop Ontology are defined by the breeders. The current Guidelines do not have clear recommendations on how to incorporate gender, user-defined information, and other social aspects into the ontology and a breeding database. This is however needed to support the interoperability between product profiles and breeders' data.

In collaboration with RTB breeding community, the project 'Breeding Roots, Tubers and Banana products for end-user preferences (RTBFoods)', led by CIRAD, France, aims at providing data sets and information required by breeders to understand the preferences of end-users in each market segment for food product qualities and make them storable into the RTB breeding databases called Breedbase. Therefore, the connection of the breeding product profiles to the food product profiles per region is desirable. The Crop Ontology (CO, https://cropontology.org) compiles crop traits and food product quality traits, with methods of measurements or assessment and scales or unit. CO can support the harmonization of the assessments or measurements of products properties and support data interpretation across social groups and provide interoperability in the Breedbase.

Breeding and Food Product Profile Templates

RTB Breeding Product Profile

We took as example the Matooke Banana Product Profile developed by National Agricultural Research Organisation, (NARO), Uganda, in the context of the Research Programme on Roots, Tubers and Bananas (Table 1). Based on expert knowledge, the Matooke Banana breeding product profile was developed to include assessment or preferences for the agronomic traits and to be linked to the preferred qualities for Food product. It contains pre-harvest agronomic traits and stress traits. Hereunder, the traits in the profile were mapped to the Crop Ontology traits – see 'Crop Ontology Trait identifier' column. However, stress resistance traits are in fact related to several component traits and variables for observing presence/absence of the disease that can support conclusions about resistance (not listed in the table). The results need to be considered in the light of the environmental conditions.

	Banana PRODUCT PROFILE: Matooke				
Region/Market segment	Trait (economic, sustainability, livelihood) and value	Target trait level	Crop Ontology Trait ID	Market Priority	Selection Objective
	Highlands of East and Central Africa				
Fresh market and processing	Yield	30% greater than Mbwazirume variety across a range of soil and management conditions	CO_325:0000383 Trait Plant annual yield	1	Maximize

	Table quality (needs regional assessment)	A general acceptability score of at least 4 (on a hedonic scale of 1 to 6), using Mbwazirume as a check (acceptability is tested after cooking as taste, aroma, colour, texture/mouthfeel)	See the product Profile Template	1	Reach threshold
	Earliness: planting to harvest	300 to 390 days	CO_325:0000000 Trait Time from planting to shooting CO_325:0000003 Trait Time from shooting to harvest	2	Minimize
	Plant stature (girth at 1m/height ratio)	A ratio of at least 0.15	CO_325:0000012 Trait Plant circumference 100 cm from the collar	2	Maximize
	Plant height	Less than 350 cm	CO_325:0000009 Trait Plant height	2	Minimize
	Suckering behaviour	75% follower sucker growth at flowering, 3-4 suckers at flowering	CO_325:0000395 Trait Overall suckering quality	2	Optimize
	Resistance to black Sigatoka	INSL at flowering of 70% and above	Several traits in CO to be measured or estimated for the presence of Black Sigatoka symptoms	3	Reach threshold
	Resistance to weevils	40% resistance higher than that of the moderate resistant check (Kayinja)	Several traits in CO to be measured or estimated for the presence of weevils' damages	2	Maximize

Table 1: Excerpt of the Matooke Banana Product Profile developed by National Agricultural Research Organisation, (NARO), Uganda, Research Programme on Roots, Tubers and Bananas

RTB Gendered Food Product Profile Template

Within the RTBFoods project, in the work package called 'Surveys for trait preferences', a template was developed for describing a gendered Product Profile to support the recording of market segments' preferences on raw and cooked Products (Forsythe et al, 2022; Tables 2-4). The template records geography and scientists involved the of the product profile. Some elements were adapted from the template provided in the 'Demand-Led Breeding Product Profiles – A Practitioners' Guide: Creating product profile summaries', by the Demand-Led Breeding Community. The RTBFoods Template is divided into 3 sections: (1) Food Product Profile study details and context, (2) Preferred characteristics with Gender $^{+}$ assessment, (3) information about traits to avoid, producers' socio-economic characteristics and other study information.

3. Food Product Profile context
Type of processing for the product (household-local, processing centre-local, industrial etc.)
Alternative crop uses (fresh root sale, processed products)

Consumer segments for the product 2–3 sentences on the consumer segments (value chain actors) along the value chain that is relatively homogenous in their preferences, on which the completion of this document is based. These are the people that the product has been specifically designed to serve and may include one or more of the following: • Farmer • Transporter • Processor • Retailer • Consumer • Material producer • Seed distributors.
Market scale : some indication of the extent of demand for the product where it currently is and where it is expected to go.

Table 2: Excerpt of the Gendered Food Product Template, Section 1 Source: Forsythe et al, 2022

In Section 1 (Table 2), the social group surveyed is named '*Consumer segmen*t' which identifies food chain actors with a common set of preferences for a product which corresponds to producers, processors, traders and consumers (who often play one or more roles) preference for x product in x region(s).

A	B	C	D	E	F	G	H	I	J	K
Characteristic category	High quality characteristics	Indicator of characteristic	Driver(s)	Customer group(s)	Preference group(s)	Priority	Gender impact scores (G+ tools) (see legend)		Good, high quality varieties	Evidence
							Do no harm Score	Positive benefits		
1. Raw material characteristics (agronomic, post-harvest)										
2 Processing characteristics of raw material for the product quality during processing (technological, physicochemical)										
3 Characteristics of raw final product (to look at, touch, smell, taste, texture in mouth)										
4 Characteristics of cooked/ready to eat final product (to look at, touch, smell, taste, texture in mouth)										

Table 3: Section 2 - Gendered Food Product Template of RTBFoods. Source: Forsythe et al, 2022

Section 2 (Table 3) contains the information about preferences that is possible to map to Crop ontology and a consumer group ontology. It is accompanied by the explanation about the content of each column. Section 3 (Table 4) holds the information about the qualities that are disliked by the consumers and could be mapped to Crop Ontology.

Legend for the RTBFood Food Product Profile		
A	**Category**	Characteristics category as explained
B	**High quality characteristics**	Characteristics that give a good, high quality product
C	**Indicator of characteristic**	How respondents assess (evaluate, feel) the characteristic
D	**Driver** (adapted from Demand Led Breeding Template)	Reason why the characteristic is important. It is likely to be different for different actors and may include: • Productivity – food and feed; • Fodder/forage – biomass of crop; • Crop management and harvesting; • Durability and cost; • Raw material quality specification; • Processability; • Processing quality specification; • consumption quality specification; • Market value and price*; • Post-harvest storage; • Sales and profit • Scalability and cost; • Variety identification**
E	**Consumer Group (**adapted from Demand Led Breeding Template)	refers to the value chain actors who highly demand that specific characteristic.- These are the people that the variety has been specifically designed to serve and may include one or more of the following: • Farmer • Transporter • Processor • Retailer • Consumer • Material producer • Seed distributors
F	**Preference group** (adapted from Demand Led Breeding Template)	A preference group is a subset of customer groups that the characteristic is very important for – a deal breaker. This could be supported by qualitative evidence of its vital importance, such as labour reduction, or there may be a high citation and/or rank and/or high CATA scores. Women (W) –preferred by women; Men (M) – preferred by men; Youth (Y) –preference by men and women under the age of 30; W+M+Y (All) – for all users

H	**Gender impact: Do no harm** (Gender and Livelihoods assessment – adapted from G+ tools)	"Do no harm" analysis. An analysis of the possible harm that introducing a new trait might cause to women or any social category of customers identified for the analysis. *See G+ tools*
I	**Gender impact: Positive benefits** (Gender and Livelihoods assessment – adapted from G+ tools)	Positive benefit analysis. An analysis of the likelihood that a new trait will be beneficial to women and men or another social category of customer defined for the analysis. *See G+ tools*

Table 4: Column that can be mapped to an ontology. Excerpt of Section 2 of the Gendered Food Product Template provides the definitions of the information captured in the Template. Source: Forsythe et al, 2022

Providing Interpretable Descriptive information about Preferences

The RTBfoods Template includes complementary descriptive information called '**Indicator of Characteristic**' that aim at clarifying the meaning of the characteristics expressed by the informant and providing context. Examples are provided in the Table 5.

Characteristics	Indicator(s)
soft texture	on eating –feel in the mouth, smooth on fingers, easy to cut. *but what is the level of softness is desired? Perhaps physicochemical analyses in the lab can measure this
good smell	inhaling under the nose/ by smelling; smells like it's been cooked in banana leaves
yellow colour	visual assessment; yellow, golden yellow
holds/ sticks together when mashed (compact)	visual; compact
elastic/starchy	touch; feels elastic (*kunyururuka*)
homogenous texture	visual assessment, feel during eating, no particles (*obukote*), *kutakuterera*, no hard parts after mashing
smooth mouthfeel	smooth in the mouth during eating, smooth (*niguterera*)

Table 5: Excerpt of the table 'Indicators of high-quality characteristics of steamed-mashed Matooke' in the RTB Foods report on gendered food mapping in Uganda. Source: Marimo P. et al, 2020.

The quality of the provided '**Indicator of Characteristics'** varies a lot. For example, in the processors' preferences listed for matooke in Uganda, several preferences had no indicator which limits reuse by breeders or data scientists (Table 6).

Attractive on plate
Gives enough food during simmering
Soft skin when peeling
Easy to tie and remain in one place because fingers are big
Easy to be wrapped
Soft on mashing
Has not darkened
Does not stick to the leaves
Easy to mash (soft)
Shiny
Not watery
Not watery even after adding a lot of water
Remains clear *mubumba*

Table 6: Preferences recorded without indicators - **Source: Marimo P. et al, 2020.**

If most of the time the indicator brings clear information that can be interpreted, (Table 7a), it happens that the Indicator does not bring the expected clarity (Tables 7b).

Quality of the Attribute	Indicator
No hard particles	Visual assessment, feel during eating, no particles (obukote), kutakuterera, no hard parts after mashing

Table 7a: Example of an informative indicator.

Attribute (sensory trait	Quality of the Attribute	Indicator
Aroma	good smell	nose inhales steam of freshly served

Table 7b: Example of a descriptor with limited possible interpretation.

Therefore, it is **recommended** to record during surveys an indicator that:
1. Describes as precisely as possible how the informant is assessing the quality (e.g. bite the raw root, remove peel with the nail, smelling freshly cooked product, by hand, chewing, etc)
2. Provides, for each trait, a description using as much as possible terms/language relatable by breeders for instance, an end-user in a survey may mention '*Good branching'* as a preferred trait for desired cassava, to a breeder it may better be understood by '*Canopy or leaf area index*'.
3. Provides clear and relatable translations of responses from survey is vital. If possible, use thresholds to create clarity, for example provide colour scale to clarify what good or bad colour is
4. Include the name of a variety the person knows that holds the preferred qualities, so it provides a reference for the breeder or the food scientist.

Mapping users' preferences to traits interpretable by breeders and food scientists

What is a variable in Crop Ontology?

According to the Crop Ontology Guidelines v 2.1 (Pietragalla J et al, 2022. https://cgspace.cgiar.org/handle/10568/110906), the value of an observation or a measurement made by a breeder n a trait is associated with a variable. Therefore, the CO model provides a standard framework for the definition of breeders' variable that measures a trait such as Plant height (PH), Grain colour (GCOL) and Grain yield (GY).

The essence of the CO model is to decompose a variable recorded by the breeders into:

- A trait: “what is observed”
- A method: “how the observation is made”
- A scale: “how the observation is expressed”

The variable is the combination of 1 trait, 1 method and 1 scale. Such a definition can apply to quantitative data collected through surveys and participatory trials but requires complementary information to put the user preferences into context.

For the qualitative data and open questions, this variable definition does not directly apply. The variable can be created afterwards, once analytics are conducted to get the proportion of answers. RTBFoods had developed a template that supports the conversion of answers into scoring of traits per gender. However, the trait classes and trait names can be used to annotate surveys with the objective of supporting a trait data search in repositories or knowledge base.

To know more, you can consult the Crop Ontology Guidelines version 2.1. (https://cgspace.cgiar.org/handle/10568/110906)

Template for extracting the preferences before mapping to the Crop Ontology

A draft template was designed in Excel to extract from the RTBFoods reports on Gendered mapping Food study, the preferred and disliked qualities on as expressed by processors and consumers and map them to the traits in Crop Ontology (see tables 2a and 2b). The template includes the country, the region where the survey was conducted, the food product and its status (raw, ready to cook, ready to eat), the method used to collect data, the attribute that provide the trait sub-class (e.g., texture, odour). The purpose of this extraction was to verify whether the information was sufficient to map all to existing agronomic or sensory traits recorded in the Crop Ontology.

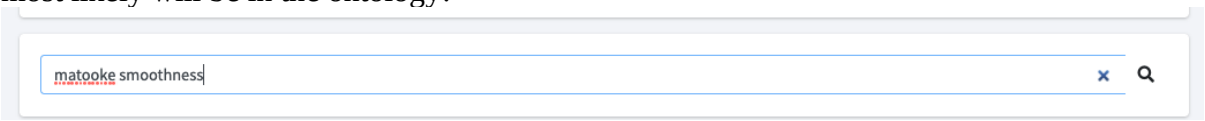
Country	District	Product	Stage of processing	User group	Method of data collect	Attribute (sensory trait class	Quality of the Attribute	Indicator	preferred material ?	Comparison with Sensory Traits in CO	Trait ID
Uganda	Mbarara and Nakasek	**Matooke**	Final ready to eat product	**processors**	Sensory evaluation by	Texture	Soft feel in mouth	Smooth in the mouth during eating, smooth		Matooke smoothness	CO_325:0002046

Table 2a: Example of extraction into the template of a preferred quality by processors for Matooke in Uganda. *Information source for extraction: Marimo at al. report, 2020.*

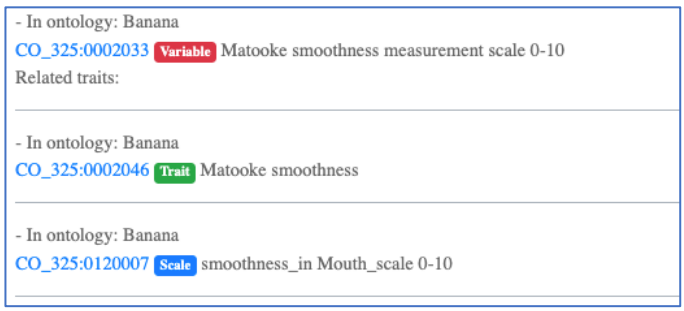
How to map a preferred or on-preferred quality to a Trait and a Variable in the Crop Ontology?

Ideally, a new sheet or accompanying file should be added to the Product Profile for the Variable & Trait Dictionary where concept identifiers of Crop ontology will be stored. The procedure to identify a trait and a variable to annotate the preference starts with a search in the Crop Ontology (cropontology.org):

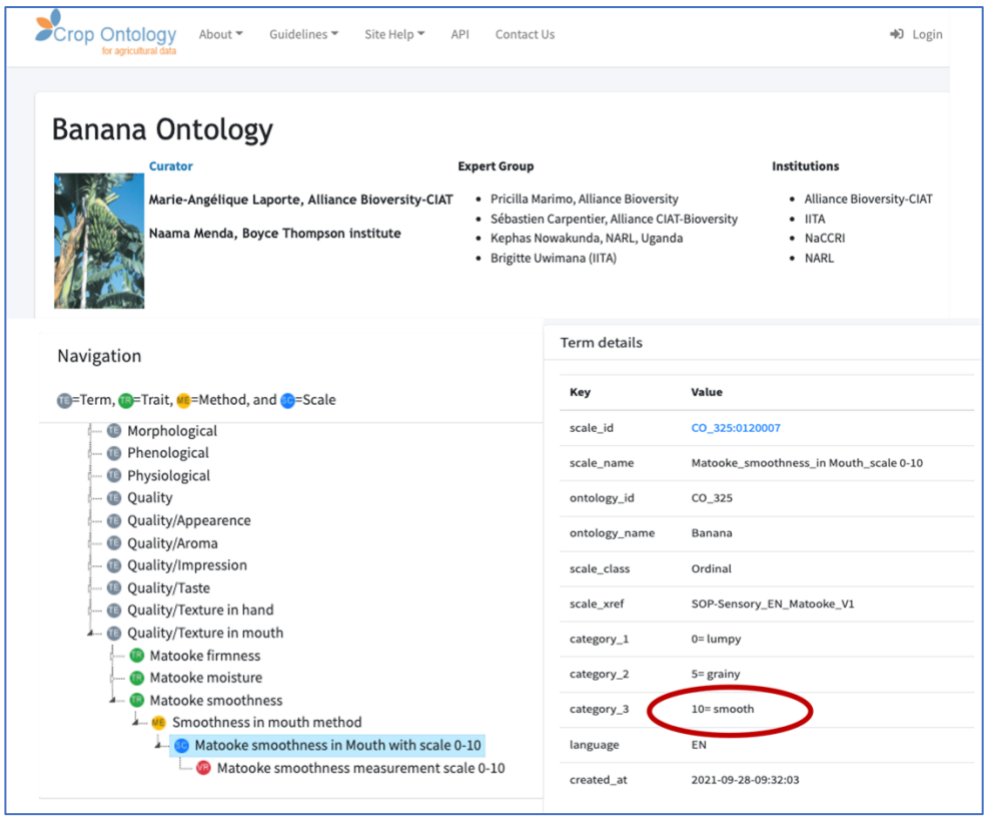
1) Search the term corresponding best to the quality (e.g. Matooke smoothness) that most likely will be in the ontology:



2) A trait and its variable will be displayed in the result list, each with its Crop Ontology concept identifier.



3) If you click on the variable (red tag) you will access it within the Banana ontology with its definition and complementary information. The variable was provided by the Lexicon of the Matooke Standard Operation Procedures for trained sensory panels (RTBFoods project Workpackage 2). It belongs to the class 'Quality/Texture in Mouth'.



In the scale, '*Smooth'* is coded 10 but corresponds to the way trained sensory panel will measure the trait. A new method should then be created for 'Smoothness in mouth_Assessement' with an hedonic scale if this was used. In the accompanying information, the method should describe the gender sensitive method used for scoring the trait.

The simplest is to select the Trait Identifier if the method is the one used in the survey, select the variable Identifier. Paste the Trait name and identifier in the Trait & Variable dictionary of your product profile.

Table 8 shows the proposed mapping result

Columns of Food Product ProfileTemplate				Crop Ontology		
		Definition	**Example**	**Mapping**	**Definition**	**Annotation**
A	**Category or family**	Class to which the quality belongs	Sensory quality	Trait Class	Agronomic, biotic stress, abiotic stress, quality, etc	Quality- Sensory in Mouth
B	**Characteristic**	Quality preferred or disliked as expressed by the informant	Matooke Smooth feel	Variable name and label	Variable name composed by Trait name, Method name and Scale ofr Unit, all abbreviated	Matooke_Texture homogenety_assessment by mouth_cat1-5 - CO_325:0002033
				Trait	Matooke Smoothness	CO_325:0002046
C	**Indicator of the Characteristic**	Complementary information indicating the way the informant estimates the quality	Soft feel in mouth during eating	Method	Method used to measure or assess the trait and provide a value	Texture in mouth
		n/a		Scale/unit	Categorical scale with scoring classes or unit	10: Smooth

Table 8: Example of mapping a preference recorded in the Food Product Profile Template to a Trait and variable in Crop Ontology.

If traits do not exist in the Crop Ontology, follow the guidelines provided by the web site regarding the term submission form: https://cropontology.org/page/Submit .

Adding the correspondence of preferences to multiple traits

It is crucial for all domains contributing trait information to the ontology (e.g. social scientists, breeders and food scientists) to reconcile or relate, as much as possible, across the domains, the traits, variables, and descriptions used. Having participatory discussions among the experts is important to validate terms used and create a common understanding of the meaning of the collected data.

Preferences mentioned by social groups may need to be broken down into measurable traits to make them interpretable by breeders. In some cases, more than one trait may constitute a preferred quality (called 'Composite Quality'). If this information was not captured in the 'Indicator' of the survey, then it is recommended to open a dialogue with the experts to identify which traits and variables compose the mentioned quality. For example, in user preferences survey for banana, end-users mentioned that they like to harvest or buy a '*Big bunch'*. This is typically a quality that does not mean much without a reference variety and/or indication of what the related traits that could be additionally assessed during the survey. Figure 1 proposes an example of '*Big Bunch* and its component traits.

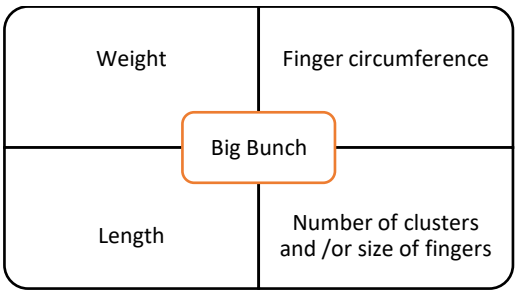


Figure 1: A banana composite trait

Majority of the experts consulted for the Guidelines development agree that there should be a way to link the user composite traits to their measurable component traits either in the ontology or the Breedbase. Currently, the Crop Ontology does not adequately capture the richness of all that information. However, for such a quality, it is possible to record in the Trait and variable dictionary accompanying the product profile the Crop ontology identifiers of the component traits. Example for 'Big Bunch':

CO_325:0000033 **Trait** Bunch weight
CO_325:0000478 **Trait** Number of hands in whole bunch
CO_325:0000324 **Trait** Average number of fingers in hand
CO_325:0000344 **Trait** Average finger circumference

In scenarios where it is difficult to dissociate composite traits, consultation can be made with Breeders, Crop specific ontology curators and expert groups, Food scientists, Social scientists, Data scientists. It is recommended to include in the survey data, all the subsets of the composite trait, and add average values, thresholds scored by the value chain actors interviewed in the survey.

An Ontology to Describe Market Segments

Providing Socio-demographic Context to the User's Preference Data

During participatory evaluations or surveys, a trait quality can be: (a) mentioned through surveys and free listing of preferences, or with comparative tests using hedonic scales or (b) measured by trained panels in participatory trials using defined attributes and categorical scales. Preferences can be collected about the agronomic and post-harvest traits on the plant,

fruit, root, etc and on sensory qualities on the raw or processed food product. Providing socio-demographic and economic information about the informant who belongs to the targeted market segment will contextualize the preferences (social group/market segment, gender, etc).

To harmonise breeder perspectives with characteristics reported by different value chain actors or social group, the development of inclusive market segment ontologies for Root, Tuber and Banana (RTB) crops was proposed. An ontology will provide a solid resource to support, the harmonization of the descriptions of the preferences, the socio-demographic context and its interpretation by breeders as well as supporting access in Breedbase. To include users' preference, it is necessary to integrate in Breedbase the informant characteristics beside the trait and variable name. As such, specific ontology is being developed to provide classes for the social groups along with their role in the food chain, and it will incorporate the user characteristics as recommended by domain experts.

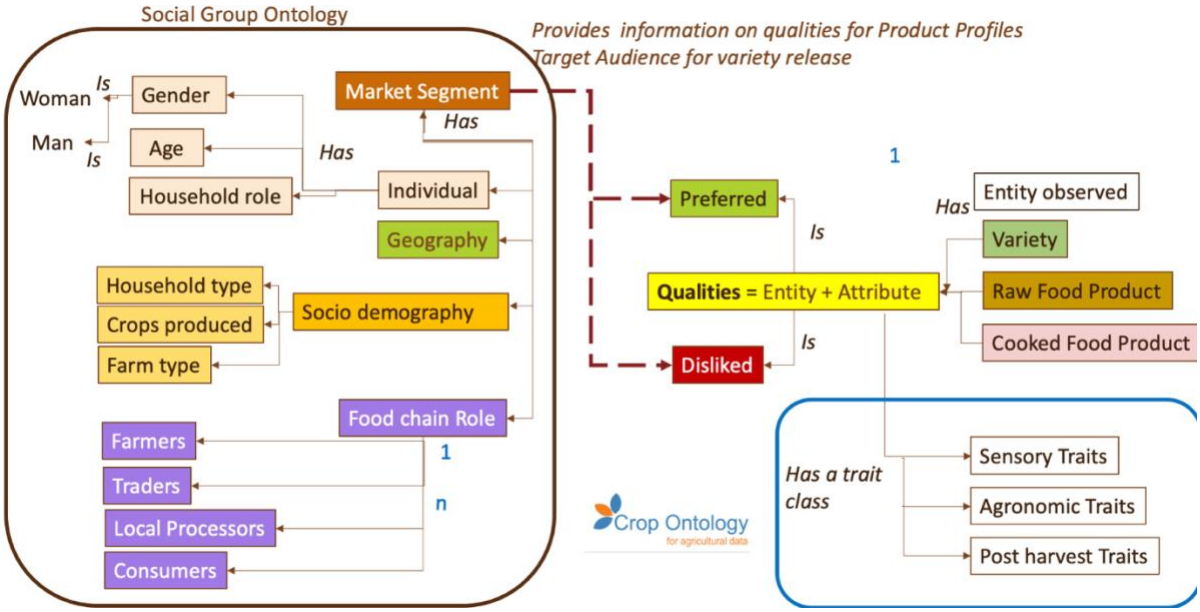


Figure 2: Elements for a user-oriented and gender sensitive ontology to document user preferences on product qualities.

It is necessary to develop an ontology for describing the person or market segment who provides the preferences. This ontology will support the annotation of the preferences with harmonized and well-defined market segment when it will be linked to the product profile (Figure 2). The ontology can also support the translation of farmers preferences into a breeder's measurable variable as far as this can be feasible as it is not always straight forward. Often, one preferred quality will correspond to *n* breeders' traits to measure: a 'banana big bunch' preferred quality can correspond to bunch weight, bunch diameter, peduncle length, number of hands, number of fruits, fruit length.

The Food Product Profile Template pf RTBFoods includes in Section1, one field for the description of the Consumers group surveyed and in Section 2, column E and F related to the consumers' group:

E	**Consumer Group** (adapted from Demand Led Breeding Template)	refers to the value chain actors who highly demand that specific characteristic.- These are the people that the variety has been specifically designed to serve and may include one or more of the following: • Farmer • Transporter • Processor • Retailer • Consumer • Material producer • Seed distributors

F	**Preference group** (adapted from Demand Led Breeding Template)	A preference group is a subset of customer groups that the characteristic is very important for – a deal breaker. This could be supported by qualitative evidence of its vital importance, such as labour reduction, or there may be a high citation and/or rank and/or high CATA scores. Women (W) –preferred by women; Men (M) – preferred by men; Youth (Y) –preference by men and women under the age of 30; W+M+Y (All) – for all users

To support the harmonization of the consumers' group naming and propose a generic definition that could guide scientists, we are developing a Market Segment ontology.

Recommendations of the Experts for the Ontology Content

To collect expert knowledge and their recommendations, an online survey was sent to RTB domain experts (Breeders, Social Scientists, Food Scientists, Data Scientists, Ontology curators) working on RTB crops. 12 responses were received to the survey and additional 8 interview were carried out were interviewed (see table in Annex). The experts who contributed were 5 breeders, 3 data scientists and ontology curators, 3 social scientists, 1 food scientist. Figure 3 provides the distribution of respondents per RTB crop.

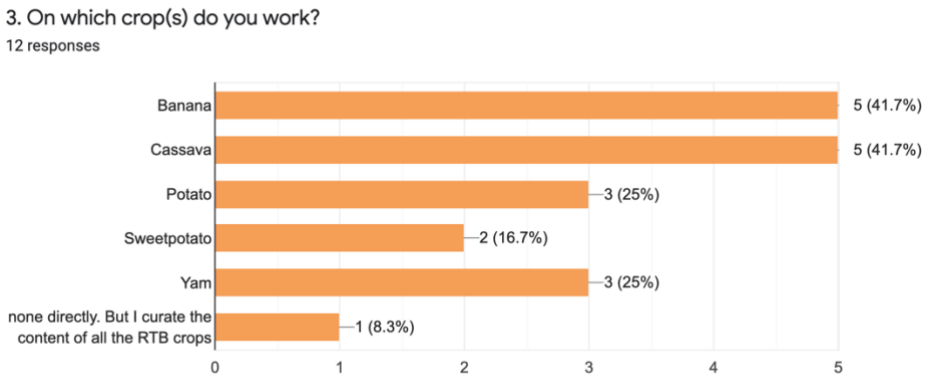


Figure 3: Distribution of the survey respondents per RTB crop

Survey results show that the top three types of information that must accompany a user preference for a product profile are:

- the gender of the informant
- a detailed and informative description of the quality preferred
- the comparative variety (check)

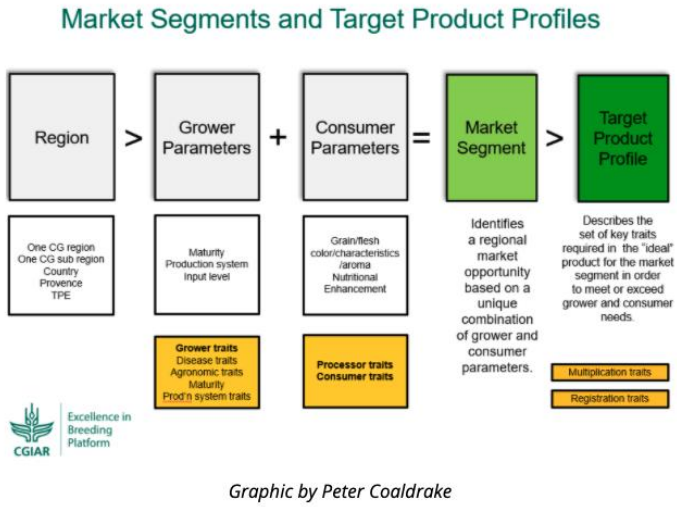
Additionally, indication of the social group, the geography of the informant is quite important.

Market Segment, Consumer Segment or Social Group? Definition and Characteristics

Definitions of social groups differ for breeders and social scientists and their research objectives. Social groups can be defined as social categories, market segments, value chain actors, food chain actors. Social scientists usually mention social groups defined by socio

demographic characteristics and recently by their roles in the food chain, which is not always linked to the food value chain that has market connotation. An ontology must be related to a precise science domain which in our case is crop breeding so we propose to use the 'market segment' concept. A market segment is **a group of people who possess one or more similar characteristics**.

For developing breeding product profiles and targeted breeding pipelines, CGIAR Excellence in Breeding (EiB) describes each targeted market segment by documenting-countries and agroecological zones, hectares grown, average yield and average selling price of the product. In addition, total population, rural population, and the population within the market segment footprint are determined, as well as the number of people in poverty and the number undernourished. The market segment is here defined with the perspective of identifying options for dissemination and adoption of new varieties bearing desirable 'added value' traits aside the mandatory agronomic traits – see figure 4.



Graphic by Peter Coaldrake

Figure 4: Market Segments and Target Product Profiles by P. Coaldrake, Excellence in Breeding (EiB)

Definitions provided by Excellence is Breeding:

Market Segment: Identifies a regional market opportunity based on a unique combination of grower and consumers' parameters

Target Product Profile Describes the set of key traits required in the 'ideal' product for the market segment to meet or exceed grower and consumer needs.

Commodity traders, processors, manufacturers who convert produce into food items and retailers, among others, are interposed between the producer and consumer. Scientists as breeders, plant biologists, nutritionists and chemists are part of the food value chain as they have made an immeasurable contribution to the development of agricultural production and food manufacture. (FAO, 1997)

A social group can be defined as any grouping within the social setup of a community based on one's area of interest such as farmers, seed producers, gender (male or female), age groups (youth, elderly), value chain actors and their roles in the value chain, researchers, wealth categories etc. Including value actors in the ontology cannot be bypassed because they are the people for whom the breeders are breeding. Each present preferences depending on the roles

they play. However, each group of players possesses social, economic and cultural characteristics that may also create differences in their preferences.

Consumer segment (value chain actors) for the product, as per The Food Product Profile Template of RTBFoods, is the segment along the value chain that is relatively homogenous in preferences, These are the people that the product will be specifically designed to serve and may include one or more of the following: • Farmer • Transporter • Processor • Retailer • Consumer • Material producer • Seed distributors

It should be noted that people do not exclusively belong to a particular social category/group in a community or a study site. As such, one must define their categorization within a given society to group them according to one's objectives The table 9 below highlights the responses from the surveyed experts on social group names that should be integrated in user-oriented crop ontologies in order of importance

1	Small holder Farmers
2	Market Traders
3	Local processors (Community- and family-based)
4	Farmer consumers
5	Urban consumers
6	Large-scale farmers
7	Processing companies
8	Small & medium entrepreneurs
9	Trading platforms
10	Peri-Urban consumers
11	Urban farmers

Table 9: Market segment names validated with the survey

Breeders suggested to integrate other social groups like youth, migrants/immigrants, disabled populations, food insecure populations, other vulnerable populations, depending on the region. Ethnicity was also mentioned as important as it is often connected to culinary traditions and specific ways of preparing food products, as well as cultural uses of the food products. The top priority socio-demographic characteristics defined by the survey respondents that should be used to define the informant and social groups are the following: gender, geography, age. It does not exclude to have additional characteristics in the ontology and in the database like: Main task(s) carried out within the value chain, socio-cultural background, decision making in stage, native to the region, years growing crops. For consumers, concepts describing the consumption patterns should be included.

An informant often belongs to several of groups mentioned above: Farmer-trader, Farmer-consumer, etc. This information will need to be collected. A well-designed ontology can integrate these one to multiple relationships.

Furthermore, survey respondents identified the important socio-demographic characteristics that should be used to define the informant and social groups as : gender, geography, age. It does not exclude to have additional characteristics in the ontology and in the database like: Main task(s) carried out within the value chain, socio-cultural background, decision making in

stage, native to the region, years growing crops. For consumers, concepts describing the consumption patterns should be included.

Inclusion of market segment into the Crop ontology will result in explosion of variables' number. It might be reasonable to limit the social groups to simple and more usable categories.

Definition of the consumers segment and roles

An ontology must include a valid definition of its concepts with the source reference to enable adequate reuse. The definition of the marker segment very context specific. However, a generic definition can be identify indicating that socio-demographic criteria will be specific to the location. If a 'farmer' has a definition, the 'smallholding size' depends of the region:

Example

Concept	Definition
Farmer	Those who owns, works on or operates a farm. https://www.vocabulary.com/dictionary/farmer
Smallholder	The definition of smallholders differs between countries and between agro-ecological zones. In favourable areas with high population densities they often cultivate less than one ha of land, whereas they may cultivate 10 ha or more in semi-arid areas, or manage 10 head of livestock - https://www.fao.org/3/y5784e/y5784e02.htm

Given that the Guidelines are for RTB crops, will be linked market segments to their specific production regions.

Integration of gender

A literature review by Marimo et al. (2020) indicated that there is less documented information on the gender differentiated trait preference evaluation for banana. However, farmers, irrespective of gender, reported similar characteristics related to production constraints, income generation and cultural uses of bananas as the main indicators for variety selection or preference. Among different value chain actors (farmers, producers, processors, and consumers), there was higher preference of traditional cultivars owing to their better consumption traits as compared to the new cultivars that had been bred to be higher yielding and resistant to pests. This points to the potential differences in the trait preferences among diverse groups in the banana value chain. These differences ought to be considered to increase the adoption of new varieties. Gender specific research and social group disaggregated data is, thus, important at the initial stages of breeding to ensure high adoption of new varieties.

The respondents emphasized the significance of using gender to define the demographic attributes of a social group. Gender roles may be specific to region, market segment or product as such generate variations in perspectives. Gender is one of the dimensions that must be used to define a social group as such, integrating it into the ontologies requires a thorough yet thought out process because it is important to know the source of information. In addition, either gender plays a different role which warrants the capturing of disaggregated perceptions. Gender disaggregated variable is best placed in the ontology than in the meta data for easy access by breeders. Integrating can be done by defining the gender groups or related terms in the ontology which can then be linked to the data. On the other hand, describing of gendered

roles for each segment of the data can be generalized and only specified in observations that require acknowledgment such as regional observations.

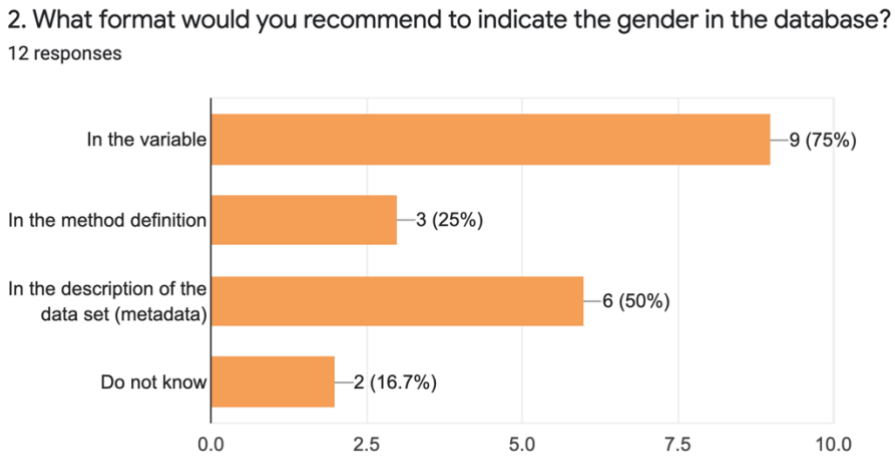


Figure 5: results **of the survey regarding the integration of the gender indicator in the ontology**

Gender roles are very sensitive to region and crop value chain, therefore, isolating traits by gender restricts the information to the regions from which the surveys were conducted.

If all survey respondents agree that a gender-sensitive variable should be linked to the role of the informant, opinions about how it should be integrated into the ontology and database vary. The ontologies on Banana, Potato and Sweet potato already integrate in the variable a gender disaggregation indicator. CIP has in fact adopted the format recommended by their Guidelines for potato PVS studies. Table 10 shows an example of Trait Dictionary where the variables directly integrate the gender indicator. Trait Dictionaries are uploaded in the Crop Ontology web site for public access. Figure 7 shows how the variables formatted in the Trait Dictionaries appear in the Crop Ontology web site.

There is currently no standard practice in Breedbase across RTB crops. For the Participatory Varietal Selection data, the farmer is recorded as a study to which socio-demographic information can be attached like gender. The answers collected with the fieldbook are therefore linked to the respondent gender. In ClimMob (https://climmob.net/), the CGIAR citizen science platform for on farm trials, the process is similar.

Variable name	Trait name	Trait class	Trait description	Entity	Attribute	Method name	Method class	Method description	Scale name	Scale class	Category 1	Category 2	Category 3
Overall plant appearance, as assessed by male respondents during preference scoring exercise	Overall plant appearance	Agronomic	Assessment of physical features of a plant, by individual male respondents, by giving it a score on the preference scale	Plant	Overall plant appearance	Preference scoring scale	Estimation	Individual male respondents give the plant one of three possible scoring cards: "Like", "Don't know", Don't like". Men and women use different colors of scoring cards.	Preference scale	Ordinal	Like	Don't know	Don't like
Overall plant appearance, as assessed by female respondents during preference scoring exercise	Overall plant appearance	Agronomic	Assessment of physical features of a plant, by individual female respondents, by giving it a score on the preference scale	Plant	Overall plant appearance	Preference scoring scale	Estimation	Individual female respondents give the plant one of three possible scoring cards: "Like", "Don't know", Don't like". Men and women use different colors of scoring cards.	Preference scale	Ordinal	Like	Don't know	Don't like
Overall plant appearance, as assessed by respondents during preference scoring exercise	Overall plant appearance	Agronomic	Assessment of physical features of a plant, by individual respondents, by giving it a score on the preference scale	Plant	Overall plant appearance	Preference scoring scale	Estimation	Individual respondents give the plant one of three possible scoring cards: "Like", "Don't know", Don't like". Men and women use different colors of scoring cards.	Preference scale	Ordinal	Like	Don't know	Don't like

Table 10: Variables disaggregated by gender and used for Participatory Varietal Selection (PVS) in the Banana Trait Dictionary developed by the Alliance-CIAT. Variable are integrated with the gender differentiation indicator.

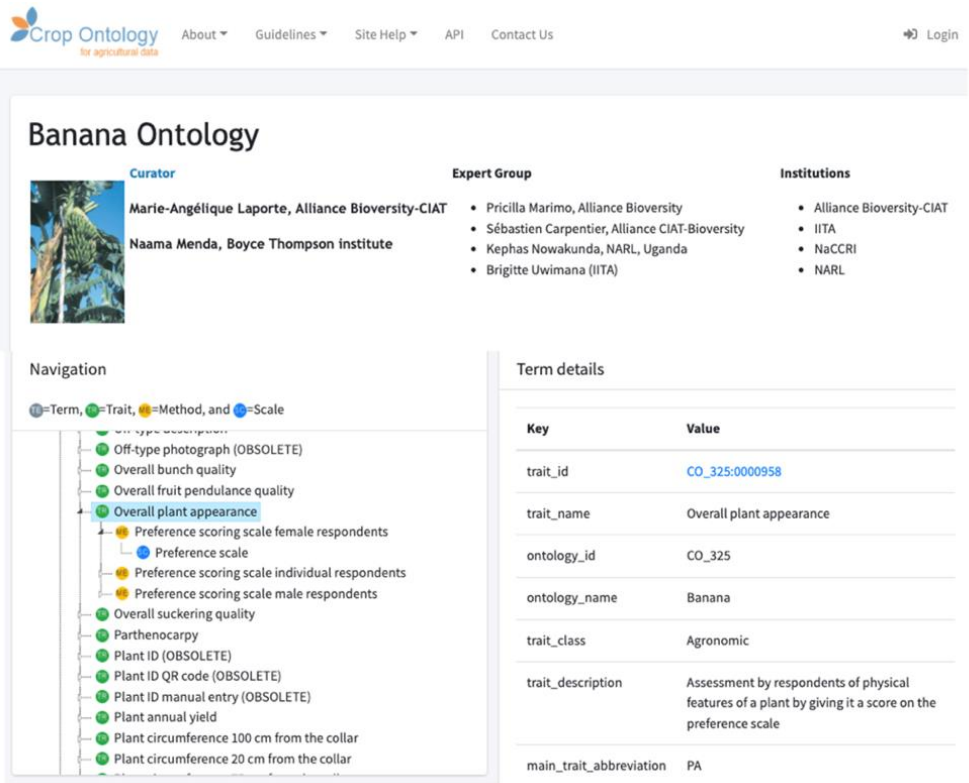


Figure 7: Format of the gender disaggregating variable in the Crop Ontology

The Cassavabase disaggregates information by user (i.e. Farmer, processor) but not by gender so the integration of the gender indicator in the ontology will facilitate identification of scoring per gender in the database. As little differences between gender preferences have been observed for cassava and plantain agronomic traits, no gender specific division is currently possible in the database. This will change with the post-harvest and quality traits assessments.

Integrating concepts of the Gender+ tools

The Food Product Template integrates elements of the Gender+ Template. Survey respondents who were aware of the Gender+ tools for developing gender sensitive product profile agreed that the user-oriented ontology should extract concepts from the G+ templates. This way, preference data collected with the G+ tools can be annotated with the ontology in the database.

The G+ approach for gender–responsive breeding, developed by CGIAR scientists since 2018 (https://www.cgiar.org/innovations/g-tools-for-gender-responsive-breeding/), offers an integrated, systematic and evidence-based protocol for breeding new crop varieties. The approach builds on work by the CGIAR Excellence in Breeding (EiB) Platform on the concept of a "product profile", which describes the traits that different actors want in a new variety, giving plant breeders a target. The G+ proposes 2 tools embedded into a Standard Operating Procedure

- G+ tool for gender-responsive customer profile that characterizes client groups targeted for new varieties, considering gender differences in knowledge, assets and decision-making which influence adoption. This makes it easier for breeders to develop the right product for the right customers.
- G+ product profile query tool (https://cgspace.cgiar.org/handle/10568/113189) guides collection of evidence to prioritize the traits in product profiles by examining both potential positive gender impacts of those traits, but also any negative impacts they might have. An example is provided in Table 11.

These two G+ tools enable breeding programs to meaningfully think through social inclusivity, and especially women's trait preferences and the special circumstances of different contexts, recognizing that one size does not fit all.

Respondents of the survey who know the G+ template recommended to integrate the concepts into the ontology.

	How to identify the trait	**The benefit of the trait to women**	**The negative effect of trait on women**
big-sized tubers	The tuber is bigger than a fist or an egg When you put the tuber in your palm you cannot close the palm 2 tubers should weigh a kilo 3 tubers can satisfy 5 people when cooked	Big sized tubers sell quickly in the market. They peel quickly and save labour A woman has control over what remains to be used for food. The big potatoes fill the basket quickly during harvest	'Some men will not agree with you on how to use the income. Yet you will have put in a lot of energy, and you feel demoralized.'
red skin and yellow flesh	By looking and by the pricking of a small piece. When growing, the flower is pink. The buds are also pink	It's mealy It has a longer shelf life It is what buyers want	
pest and disease resistance,	The tuber has no signs of diseases i.e. no black spots, no visible rotting	Disease resistance results in high yields thus fetching more income Labour for spraying is reduced. Disease resistant crops do not require a lot of spraying, therefore, saving on costs of agrochemicals and labour	Women do not control the income resulting from the higher yields

Table 11: Example of Sweetpotato trait qualities benefiting women- table provided by Netsayi N Mudege, CIP.

Mapping preferences to market segment

To develop a product profile, preferred traits are mapped to the market segments using statistical data analysis of the preference survey data. The RTBFoods Product Profile Template assists the scientists in this mapping. The consumer segment ontology will provide a consistent naming source of market segment dimensions to support the mapping in the database where the product profile will be stored

Additional metadata can be added in the breeding database that specifies the market segment, gender proportion, the target location and role in the value chain. The ontology will provide the controlled vocabulary with proper definitions.

When known, the economic importance of the most mentioned traits and variables for a crop or a food product could be indicated in the ontology if the information is available.

Mapping preferences to varieties

Currently, variables/traits in the ontology are not mapped to variety. In the ontology, traits are given a general description while in Breedbase, the information is always linked to a genotype/variety.

During participatory variety selection, evaluation is done on the variety in reference to cultivars that are well-known by users. This comparison contributes to better understanding of the preferred qualities as expressed by the end-users. For example, during surveys on banana hybrids in Uganda, respondents compared their preferences to Mbwazirume, their preferred cultivar. Data were captured indicated the trait quality in reference to this clone, e.g. 'yellow as Mbwazirume'. The trait measurements or assessments done on the check at the given location is captured in the data sets uploaded in breeding database to provide reference thresholds of the measurements or assessment made in a trial. It is not recommended to include the reference cultivar name directly in the scale of the trait in the ontology because (a) the ontology concept needs to be generic enough to be reused in other projects and (b) the phenotype of reference cultivar varies according to the environment of the growing location and the field management practices. It can be added as complementary information in the metadata of the trait as an information tightly linked to the location.

Quantitative scores, or ranks, or measures may be added as complementary information to indicate significance of a trait. Breeders indicated that adding economic weights to the traits will assist them to screen for the most important information. Breeders suggested that an index selection should be used to identify the important traits to allow release of the most optimal clone to the target environment. A Selection Index and calculated with mathematical tools and is used when several useful traits are selected simultaneously and is calculated upon trials data. In this method, the crop is scored for its merit in each of the traits included in selection. The formula to calculate a linear selection index (I) for a selection candidate is as follows: $\mathbf{I} = \sum \mathbf{i} = \mathbf{1}\ \mathbf{t}\ \mathbf{a}_{\mathbf{i}}\ \mathbf{g}_{\mathbf{i}}$, where **t** is the number of traits, and $\mathbf{a_i}$ and $\mathbf{g_i}$ are respectively the economic weight and genetic merit of the trait. To support full interpretation of the preferences, this information should be captured in the database, at the level of the product profile for a target market segment. As this value will change with the selected market segment it cannot be included in the ontology, or only as an informative element in the metadata.

Annex

Respondent to the interviews and online survey

Summary statistics for interviewed experts			
Breeders	Data scientist/curator	Social scientist	Food scientist
5	3	2	1

Summary statistics for experts who responded to the survey			
Breeders	Data scientist/curator	Social scientist	Food scientist
4	3	3	0

See the summary of the survey results in the Annex document.

Reference documents to read

De Haan, S., E. Salas, C. Fonseca, M. Gastelo, N. Amaya, C. Bastos, V. Hualla, and M. Bonierbale. 2019. Participatory Varietal Selection of Potato Using the Mother & Baby Trial Design: A Gender-responsive Trainer's Guide. Lima: Peru.International Potato Center. 81 pp

Demand-led Breeding. Demand-Led Breeding Product Profiles – A Practitioners' Guide: Creating product profile summaries. https://www.demandledbreeding.org/sites/g/files/zhg1501/f/2020/08/16/product_profile_guide_a4_professional_print.pdf

Lora FORSYTHE, Laurent ADINSI, Noel AKISSOE, Aurelie BECHOFF, Alexandre BOUNIOL, Pricilla MARIMO, Gerard NGOH, Benjamin OKOYE, Olamide OLAOSEBIKAN, Tessy MADU, Bela TEEKEN. (2022). RTBfoods Work Package 1 Gendered Food Product Profile. Chatham Maritime, United Kingdom: RTBfoods Calibration.

Forsythe L, Tufan H, Bouniol A, Kleih U, Fliedel G. An interdisciplinary and participatory methodology to improve user acceptability of root, tuber and banana varieties. *Int J Food Sci Technol*. 2021;56(3):1115-1123. doi:10.1111/ijfs.14680

Iragaba, P., Hamba, S., Nuwamanya, E., Kanaabi, M., Nanyonjo, R.A., Mpamire, D., Muhumuza, N., Khakasa, E., Tufan, H.A. and Kawuki, R.S. (2021), Identification of cassava quality attributes preferred by Ugandan users along the food chain. Int. J. Food Sci. Technol., 56: 1184-1192. https://doi.org/10.1111/ijfs.14878

Marimo Pricilla, Kankwasa Kenneth, Mpirirwe Innocent, Ndagire Lilian, (2021). Gendered Food mapping on Matooke in Uganda. Understanding the Drivers of Trait Preferences and the Development of Multi-user RTB Product Profiles, WP1. Kampala, Uganda: RTBfoods Field Scientific Report, 2021, 66 pp.

Marimo, P; Karamura, D; Tumuhimbise, R; Shimwela, MM; Van den Bergh, I; Batte, M; Massawe CRS; Okurut, AW; Mbongo, DB; Crichton, R, 2019. Post-harvest use of banana in Uganda and Tanzania: Product characteristics and cultivar preferences of male and female farmers. Lima (Peru). CGIAR Research Program on Roots, Tubers and Bananas (RTB). RTB Working Paper. No. 3. Available online at: www.rtb.cgiar.org

Pietragalla, J.; Valette, L.; Shrestha, R.; Laporte, M.-A.; Hazekamp, T.; Arnaud, E. (2022) Guidelines for creating crop-specific Ontology to annotate phenotypic data: version 2.1. Alliance Bioversity International and CIAT. 38 p.

Polar, V. and Ashby, J.A. (2021). Gender Report and Template Forms for G+ Tools. CGIAR Research Program on Roots, Tubers and Bananas, Report Template 2021-1. International Potato Center: Lima, Peru. www.rtb.cgiar.org/gbi.

Ndjouenkeu, R., Ngoualem Kegah, F., Teeken, B., Okoye, B., Madu, T., Olaosebikan, O.D., Chijioke, U., Bello, A., Oluwaseun Osunbade, A., Owoade, D., Takam-Tchuente, N.H., Biaton Njeufa, E., Nguiadem Chomdom, I.L., Forsythe, L., Maziya-Dixon, B. and Fliedel, G. (2021), From cassava to gari: mapping of quality characteristics and end-user preferences in Cameroon and Nigeria. Int. J. Food Sci. Technol., 56: 1223-1238. https://doi.org/10.1111/ijfs.14790

Teeken Béla, Garner Elisabeth, Agbona Afolabi, Balogun Ireti, Olaosebikan Olamide, Bello Abolore, Madu Tessy, Okoye Benjamin, Egesi Chiedozie, Kulakow Peter, Tufan Hale Ann. Beyond "Women's Traits": Exploring How Gender, Social Difference, and Household Characteristics Influence Trait Preferences' - Frontiers in Sustainable Food Systems, v 5., 2021, DOI=10.3389/fsufs.2021.740926
CIP Guidelines