2. OVERVIEW OF DATA COLLECTION

Data gathering with a Lyngngam community, Meghalaya, India.

Photo: D. Mijatović
2. OVERVIEW OF DATA COLLECTION

Agrobiodiversity research uses methods drawn from a range of disciplines (e.g. anthropology, ethnobotany, genetics, botany, biogeography, ecology). It requires approaches that integrate traditional and scientific knowledge and that can take account of different worldviews of diversity and the environment. Data collection procedures include commonly used methods such as household surveys and focus group discussions as well as specifically designed participatory methods such as the 'four cell analysis'. The methods presented here have been widely used to investigate the richness and distribution of species, varieties or breeds and their characteristics (traits), values and uses. Methods to study seed flows, landuse systems and the perceptions of the sources of resilience and ecosystem services are also described.

The choice of methods for data collection and analysis will depend on the specific questions that are being asked. For example, does the research focus on describing diversity (amount and distribution) or is it related to particular aspects of diversity management such as the revival of local seeds? Research questions can reflect the perspectives of specific disciplines (e.g. ethnobotany), or may be concerned with exploring various practical questions, such as how to conserve and increase diversity to improve productivity, resilience, livelihoods, nutrition and health.

The methods described in this Compendium provide information on:

- The amount and distribution of crop and livestock diversity at household and community level and the diversity of useful wild plants and pollinators
- Important characteristics (traits), management and uses of crops, crop varieties, livestock species and breeds and wild plants
- The ways in which seeds and planting materials are exchanged and affect diversity
- Changes in diversity over time
- Community perceptions of the landscape, and the importance of different land uses for the provision of ecosystem services and resilience
- Needs and opportunities for agrobiodiversity conservation and use.

2.1 DIVERSITY OF WHAT?

Agrobiodiversity encompasses both the differences among individual plants or animals, differences among crop varieties, between animal breeds or among wild plant populations, and the assortment of species, ecosystems and land uses. Most of the methods in this Compendium focus on assessing and describing variety, crop, breed or species diversity.

**Crop species** – plant species cultivated in agriculture or aquaculture. Crops and crop species are often but not always the same. For example, ‘wheat’ encompasses a number of species, including *Triticum aestivum* (bread wheat), *T. durum* (durum or pasta wheat) and *T. spelta* (spelt). In contrast, the species *Brassica oleracea* contains several crops, including kale, cabbage, cauliflower, broccoli and Brussels sprouts.

**Local or traditional varieties (landraces)** – dynamic populations of crops with certain characteristics selected by farmers. They have a distinct identity (phenotype) and are often genetically diverse and locally adapted. Modern varieties developed by plant breeding organizations are usually more uniform than traditional varieties.

**Domesticated animal species** – cattle, sheep, goat, pig, horse, donkey, buffalo, chicken and duck and some less-common species such as geese, llama, yak, camel and guinea pig.

**Local breeds** – groups within a domesticated animal species having common ancestors and identifiable external characteristics and appearance, homogeneous behaviour and/or other characteristics. Like local varieties, such breeds have evolved to suit local conditions.

**Wild plants** – wild species gathered for food, medicine, rituals, dyes, building material, etc.

**Pollinators** – animals, including insects, vertebrates and mammals, that pollinate plant species.
LANDSCAPE PERSPECTIVE

A landscape (or seascape) perspective allows a better understanding of the composition and patterns of agrobiodiversity, its management and uses at the community level.

Different crop varieties are cultivated on different types of soil, along elevation gradients and across different cultural groups. In addition to domestic crops and animals, local communities rely on wild species harvested along the continuum of land-use intensity in pastoral, rotational and other types of systems.

Although information about the crop, animal and wild plant diversity is often collected separately, different components of agrobiodiversity depend on each other and should be seen as part of a wider agroecological system managed by local communities. In such a system, practices for managing diversity, land and water are closely interrelated.

A landscape perspective allows a more complete understanding of the interactions between different components of diversity, e.g. the role of forests and sacred groves in providing food and medicine, maintaining pollinator populations and mitigating the effects of extreme weather events.

Figure 2.1 Community landscape map, Tshongogwe community, Zimbabwe. Source: Agrobiodiversity, Land and People Project, PAR and SAFIRE.
2.2 LOCAL KNOWLEDGE

Agrobiodiversity and its management are intrinsically linked to local knowledge and cultural practices. Aspects of local or traditional knowledge that are important for agrobiodiversity research include the following:

**Local classification systems (ethnotaxonomies)** – Local names and classification systems for crops, animals, forest or pasture flora, soil types and ecosystems reveal important information about diversity and reflect the interactions between people, plants, animals and the environment.

**Management practices and systems** – Agrobiodiversity is a result of distinct management systems in diverse environments. Practices including seed selection and exchange and the management of animals, soil, water sources, forest and other ecosystems all influence the evolution, richness and conservation of agrobiodiversity.

**Uses, values and beliefs** – Wild species, crops, varieties, domestic animals and breeds are associated with a diversity of cultural uses, values and practices. Specific varieties, breeds or species may have a special place within traditional worldviews (or cosmosvisions) or in local culture for their nutritional, culinary, medicinal or adaptive traits (e.g. adaptation to specific soil). Sacred groves and sacred woods have cultural and ecological importance.

---

**GENDER AND AGROBIODIVERSITY**

Agrobiodiversity knowledge and its acquisition are gender-differentiated. Knowledge arises out of experiences and daily acts and hence from gendered roles and responsibilities. Women and men have different roles in agricultural and pastoral production systems and consequently have different specialized knowledge about crops, animals, wild plants and the preparation of food, medicine and various crafts (e.g. weaving, natural dyes). Women have long been known for their specialized knowledge about seeds. The gender differences need to be taken into consideration to avoid gender-related bias in all stages of research.
LOCAL NAMES AND CLASSIFICATION SYSTEMS

Agrobiodiversity research requires a good understanding of local names and classification systems for crops, animals, wild plants, soil, seasons, pests and diseases and other features of diversity and the environment. Local names and classification systems are specific to cultures.

One of the simplest and most-effective ways to understand local categorization is to use a ‘freelisting’ method. For example, asking interviewees or focus group participants to list all vegetables, fruits or wild food plants can help to understand categorization from a cultural-domain perspective.

Once local and common names have been identified, the next step is to link them with scientific or Latin names, which consists of genus and species (e.g. *Malus domestica* for apple). The identification of species of crops, pollinators and wild plants often requires collaboration with botanists, entomologist and other experts. Specimens or photos can be used to consult the botanists. The photos need to capture specific parts or characteristics of importance for species or varietal identification. For example, for the identification of the species (and variety) of cereals such as wheat, rice and millets, the photos need to show the structure of the spike or panicle and the shape and colour of the seed.

In many cases, the correspondence between local name and scientific name is one to one: one local name corresponds with one scientific name.

However, for certain plants the correspondence may not be one to one, and this results in:

- Overdifferentiation (several local names refer to only one Latin name) or
- Underdifferentiation (one local name refers to several Latin names) (Table 2.1).

Table 2.1 Correspondence of local names for wild food plants with scientific names (example from the White Carpathians, Czech Republic).

<table>
<thead>
<tr>
<th>Correspondence Type</th>
<th>Folk Name</th>
<th>Scientific Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-to-one</td>
<td>Kokoška</td>
<td><em>Capsella bursa-pastoris</em></td>
</tr>
<tr>
<td>Overdifferentiation</td>
<td>Kašičky, Kozičky, Černý bez, Hural</td>
<td><em>Sambucus nigra</em></td>
</tr>
<tr>
<td>Underdifferentiation</td>
<td>Šťovík</td>
<td><em>Rumex acetosa</em></td>
</tr>
</tbody>
</table>

LOCAL VARIETY AND BREED NAMES AS MEASURES OF DIVERSITY

In many parts of the world, local varieties and breeds are recognized through local names. They may be named after places of origin, morphological characteristics, phenology or other specific traits. Names of varieties and breeds may change over time or vary from community to community or even from household to household. Individual farmers in a community may call the same variety or breed by different names or different breeds or varieties by the same name. This identity problem may increase when working with different communities where other factors such as differences in pronunciation may complicate identification further. Focus group discussions and four cell analysis are ways of coping with this problem. Further studies using field trials or even molecular genetic methods can shed additional light on the similarities and differences between varieties that farmers recognize.
2.3 DATA-GATHERING METHODS

Agrobiodiversity information is collected using a combination of quantitative (e.g. surveys) and qualitative (e.g. focus group discussions) methods. Field and participant observations, species inventories, field trials, nutritional composition analyses, pest and disease determination, remote sensing and molecular genetic studies are just some of the ways that can be used to obtain additional data. The Compendium describes some common data-gathering methods:

Household survey questionnaire is used to collect information from a sample of households in a community using a structured interview. Information collected includes land uses, management practices, characteristics of crops, varieties and breeds, seed sources and uses of wild plants. The household survey also provides information on demographics, socioeconomic status of households and other aspects to enable differentiation of the sample and analyses of changes over time. Further information is given in Section 5: Household surveys.

Focus group discussions (FGDs) are used to explore topics in more depth and from different perspectives within a community. FGDs are particularly useful to find out about diversity distribution, important characteristics, management practices, constraints and opportunities, and any other topic. FGDs are used to validate data from other sources and to reach a consensus at the community level, e.g. on variety identity and properties. Many methods in the Compendium draw on focus group methodology further described on page 15.

Key informant interviews are in-depth interviews with community members that have specialized knowledge about agrobiodiversity, e.g. medicinal plants, food or seed processing and beekeeping. These are conducted using semi-structured or structured interviews that consist of questions presented to all key informants in the same way. The Compendium gives examples of key informant interviews to collect information on animal diversity, wild plants and other aspects of diversity. Information obtained from key informants is complementary to information from household surveys and FGDs.
PARTICIPATORY TECHNIQUES

Household survey questionnaires, FGDs and key informant interviews are the main methods to collect information about agrobiodiversity. There are a number of techniques for systematic collection of agrobiodiversity data that can be applied, modified and combined in surveys, FGDs and key informant interviews. Some of these techniques and methods described in the compendium can be deployed to facilitate empowerment and decision-making in relation to agrobiodiversity and other resources. Participatory data collection techniques can contribute to processes of shared learning, enabling ownership and mobilization of knowledge to address issues faced by local communities (e.g. loss of diversity, climate change and malnutrition).

- Listing or freelistin involves creating lists of items with individuals or groups about species or other items (e.g. fruits, animals, wild plants, varieties of a crop). See Section 9: Uses of wild plants for further information on the freelistin process.

- Ranking, scoring or rating, pile sorting and similar techniques elicit attributes, similarities and relations among items within a domain (which have been identified through freelistin or some other method).

Two examples of ranking are given in Table 2.3 and Table 2.4. Another example is given in Figure 7.1, which shows the results of a scoring of rice varieties for different traits. Ranking, scoring, pile sorting and similar techniques can be applied and adapted in many different research contexts.

- Diagrams drawn by survey respondents, FGDs participants or key informants illustrate and explain processes, relationships and structures related to diversity, management practices or social institutions. An example is given in Figure 2.2, see Section 3: Transect walk for examples of transect diagrams.

- Mapping describes the location and distribution of resources, land uses and landscape features, their importance and changes over time. Landscape mapping is explained in Section 12, and mapping can be used to explore many other aspects of diversity and its management such as species distribution or migratory routes between dry- and wet-season pastures.

- Calendars and timelines show changes in uses, management and availability of diversity over time. While calendars show seasonal changes; timelines illustrate changes over a longer period of time, e.g. occurrence of droughts and floods, pest and disease outbreaks, introduction of commercial crop varieties or animal breeds. Examples of calendars are given in Section 4: Seasonal calendars; and an example of timeline is given in Section 13: Resilience assessment.
FOCUS GROUP DISCUSSIONS

Many methods described in this Compendium make use of FGD techniques to explore a specific topic with a group of participants. The information collected in FGDs draws from local knowledge and from experiences, beliefs, perceptions and attitudes of the participants. An FGD is a moderated discussion between participants, and not between the researcher and the participants. FGDs are not only for the researcher or facilitator to get information, but also provide a chance for the participants to exchange information among themselves.

FGDs can be organized around a set of open-ended questions on a specific topic, but other techniques such as scoring, ranking and diagramming can be used to obtain information. During the FGD, the information is recorded on, for example, a large sheet of white paper or on cards. The recorded information is not just for the researcher, but also for the participants.

FGDs require good planning and organization during research design, preparation and data collection. An FGD is conducted by a team consisting of a facilitator and one or more assistants, note-takers or rapporteurs. The facilitator manages the discussion, and needs to create a comfortable environment for all participants. The assistants’ role is to document the content of the discussion.

Designing FGDs
- Identify the main aim and the key research objectives
- Make a list of questions (schedule or script) as guidance for the FGD session
- Decide on the number of respondents (usually 4–15)
- Select the participants through purposive or convenience sampling
- Recruit the participants in advance
- Identify a venue for the discussion
- Prepare and organize material
- Organize refreshments for the participants
- Decide if to conduct a mixed or separate gender group according to the local socio-cultural context.

Conducting FGDs
- **Preparation** – Make sure the research team gets familiar with the script, committing the questions to memory as much as possible.
- **Pre-session** – Use the time before the FGD starts to become familiar with the group dynamics and make all participants comfortable.
- **Session** - Introduction – before proceeding with the questions and discussion:
  - The facilitator introduces the team and the topic and purpose of the FGD, and thanks the participants and organisers
  - The participants introduce themselves (one option is randomized self-introduction instead sequential introductions)
  - The facilitator initiates the discussion and proceeds with the script.

Facilitation
A successful FGD depends on a skilled facilitator to guide the group’s discussion. The facilitator needs to encourage discussion by creating a warm and comfortable environment. It is essential that the facilitator respects participants’ knowledge, experiences, opinions, perceptions and customs. Important facilitator skills include:
- Good speaking and listening skills
- Good observation of participants’ body language and group dynamics
- Some knowledge of the topic of discussion
- Flexibility to adapt to the flow of the discussion
- Ability to remain impartial and maintain verbal and non-verbal objectivity
- A sense of humour to keep the discussion relaxed and encourage sharing of information (Nyumba et al. 2018).
2.4 AGROBIODIVERSITY DATA

During or after interviews and FGDs, the information collected is organized and processed to create data tables that can later be analyzed. For example, Figure 2.1 shows a diagram of seed sources drawn by a farmer during a household survey. Such a diagram can be processed to create a table (Table 2.2). To encode data on seed sources identified by farmers, the code ‘1’ is assigned to those sources from which there is an arrow pointing to the farmer, and the code ‘2’ to those for which the farmer is the source.

Figure 2.1 Diagram of a farmer’s response to questions about seed sources. Source: Jarvis and Campilan (2006)

Table 2.2 Tabulated data from the farmer’s response to questions about seed sources from Figure 2.1.

<table>
<thead>
<tr>
<th>Respondent</th>
<th>Parents</th>
<th>Neighbour</th>
<th>Market</th>
<th>Relative</th>
<th>Extensionist</th>
<th>NGO</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

While some of the information collected will be quantitative (How many varieties? What is the size of the field?), much will be semi-quantitative or qualitative (Which variety is better? Why?).

Quantitative data (numbers, also called numerical) are observations that can be counted (discrete data, e.g. trees in a field) or measured (continuous data, e.g. area of land). As such, this type of data should always be associated with a unit of measurement (e.g. number of trees, hectare).

Qualitative data (text) describe characteristics or properties of a subject. Qualitative data are also called categorical as they express a categorical measurement not in terms of numbers, but in terms of words. Qualitative data can be extracted from questionnaires, interview transcripts, FGDs, diagrams and any other participatory data-gathering technique. For many analyses, qualitative data need to be quantified, which involves turning the words into numbers (coding) (e.g. fruit colour: orange = 1, red = 2, purple = 3). Coding requires construction of a category system that allows all of the data to be categorized systematically. After coding, the data can be organized, interpreted and analyzed for frequencies and relationships between variables, means and variance.

Qualitative data can be gathered using techniques that allow easy transformation into quantitative data (freelisting, ranking, rating, pile sorting). Examples of such transformations and their uses include the following:

**Characterization:**
- Binary – the informant in a survey or focus group is asked to say ‘yes’ or ‘no’ (e.g. Is this variety resistant to a disease?), or to choose between two possible values (e.g. Is this variety resistant or susceptible to a particular disease?)
- Categories – the informant is asked to choose or give a description, e.g. white, red or black for the colour of grains in rice.

**Comparative:**
- Rating – the informant is asked to rate an item on a numerical scale between two or more alternatives, e.g. yield: low, medium, high
- Ranking – the informant is asked to rank a list of items in order, for example, according to preference or importance. Tables 2.3 and 2.4 provide examples of the results of ranking for traits and functions of fruits.

**Belief statements:**
- The informant is asked to assess the truth of a statement against a predetermined scale, e.g. this variety is good for feeding to nursing mothers: true, intermediate, false.
**2.5 SAMPLING STRATEGIES**

Choosing the participants and sample size are two important first steps in any study. The choice of sampling approach is directly linked to the study objectives. The sample should be able to represent the population that is of interest to the study and be large enough to have sufficient statistical power to answer the research questions.

The sampling strategy should consider both the selection of communities and the selection of participants within communities and households for data collection. In some cases, it is desirable to target specific people ('knowledge holders'). If the focus of the study is a specific region or district, then communities should be selected to ensure a balanced reflection of the different social and environmental conditions in the region. Often, the focus of study is a specific village or set of villages that have been selected based on particular criteria, such as established trust and willingness to engage with the researchers. For impact assessments, the sample should include villages and households that are not participating in any specific interventions to serve as a 'control'.

The selection of participants should consider the variation of knowledge distribution among different age groups or social groups: certain knowledge can be held only by elders or specialist 'custodians of knowledge'. For example, knowledge about medicinal plants is commonly maintained by herbalists, traditional healers or shamans. It is important to keep in mind that knowledge is often gender-differentiated: women and men have different knowledge, preferences and concerns in relation to diversity. For example, in terms of preference for crop traits, traits important to women include qualities related to preparation and nutrition, while for men, important qualities are more likely to be related to productivity.

There are two main approaches to sampling: probability sampling and non-probability sampling. Probability sampling gives the best chance of obtaining a sample that is truly representative of a population. Non-probability sampling is used in specific cases, such as if the objective of the study is to document as much knowledge as possible in a short time or to document rapidly disappearing traditional knowledge. A summary of sampling strategies is provided below (based on Newing 2011).
PROBABILITY SAMPLING

Simple random sampling – Pick out individual cases (participants) from a sampling frame, using a random numbers table. For example, households can be randomly selected for inclusion in a survey by first preparing a list of households in the community, in consultation with local leaders, and then randomly selecting households from the list. In Microsoft Excel, the function =RANDBETWEEN(1,100) can be used to generate a random number for each household in the list, and then the households assigned the highest numbers can be selected for surveying.

Systematic sampling – Use a random numbers table to pick the first participant or household, and then select additional participants following a constant interval \( n = \frac{\text{total population (P)}}{\text{desired sample size (N)}} \). E.g. for a total population of 100 individuals or households, if the desired sample size is 20, the interval \( \frac{n}{=100/20=5} \). If the random number (first participant) is 7, then the second participant is 12 (7+5), the third participant is 17 (12+5)… until 20 participants or households are selected.

Cluster sampling – Divide the population into ‘clusters’ (often, geographical areas), take a sample of clusters, and then take a sample of cases from each selected cluster. This approach is particularly useful for a large, dispersed population. To achieve probability sampling, a sampling frame is needed for each cluster that is sampled.

Stratified random sampling – Stratifying the population before applying random sampling methods involves developing criteria for stratification (e.g. socioeconomic subgroups). Divide the population into ‘strata’ (groups of cases with certain characteristics, such as men and women, rich and poor, large and small landowners), and then take a random sample of cases from each stratum.

NON-PROBABILITY SAMPLING

Convenience sampling – Interview anyone that you can find who fits your broad criteria.

Targeted sampling – Seek out individuals who are most relevant to study. Targeted sampling is used in studies that focus on a particular group (e.g. particular ethnic group, pregnant mothers between the ages of 25-30) or people with specialized knowledge (e.g. traditional healers).

Purposive sampling – (also known as judgmental, selective, or subjective sampling) – Select a sample based on personal judgment of their suitability for the study.

Quota sampling – Define two or more subgroups (e.g. men and women) and set the proportion you want in each category (e.g. 50:50). Interview anyone you can find in each subgroup until you have reached the target sample size.

Snowball sampling – (also known as chain sampling, chain-referral sampling or referral sampling) and respondent-driven sampling – Seek out individuals who are most relevant to the study, interview them and ask if they know of others you could interview or who are linked to them in a specific way (e.g. in studying seed systems). Then interview those individuals suggested by already interviewed respondents.

Sample size is a key consideration in planning your research and different methods may require different sample sizes to give the amount of information needed. The sample size depends on the study design, and the methods for data collection and statistical analysis you are planning to use. In general, a sample size of at least 30 individuals is desirable for crop variety diversity information from a household survey. Economists, social scientists and others who want to have a more robust dataset tend to use larger sample sizes (50–100 households).

FURTHER INFORMATION / REFERENCES


PAR Climate Change Project (2010) FPIC – Agrobiodiversity and Climate Change project. http://agrobiodiversityplatform.org/climatechange/the-project/abd_and_cc_project_fpic/