Module 4: Data Quality & Standardisation

Inclusion Data Quest

WE ARE OPEN

Module Objective

This presentation explores the critical role of data quality in enhancing gender and disability inclusion. We will assess key data quality dimensions. We will also discuss harmonisation challenges.

Agenda

1. Assessing Data Quality Dimensions
2. Challenges in Data Harmonisation
3. Case Study - Operationalizing KeSQAF for Gender & Disability Data
4. Improving Gender & Disability Data Availability
5. Addressing Data Quality in Marginalized Contexts
6. Key takeaways

1. Assessing Data Quality Dimensions

A ‘take on quality’ from KESQAF

- **Quality Assurance**: This is a planned and systematic pattern of all the actions necessary to provide adequate confidence that a product will conform to established requirements. It aims at achieving quality improvements on an ongoing basis.
- **Quality Assessment**: Part of quality assurance that focuses on an assessment of how well quality requirements (the stated needs or expectations) are fulfilled. It is a tool that identifies areas where quality improvements are most needed.

Consequently, instead of defining ‘Quality’ (or <u>WHAT</u>), Kenya Statistical Quality Assurance Framework (KESQAF, 2022) proposes standards and criteria for <u>HOW</u> both statistical data and methodologies used across the value chain.

This course adopts a similar approach.

Operationalising Data Quality and Availability - 1

- **Mode of Collection:** This has to do with how data is collected- is it digitised from the field? What tools are used? Among others.
- **Frequency of Collection:** The periodicity of collection and release should consider user requirements and timeliness as much as possible.
- **Granularity and Disaggregation:** The availability of disaggregated data – especially along critical variables of age, gender and location – are extremely valuable for public policy discourse, analysis, and planning.
- **Validation:** Validation provides an independent review of the data to increase the confidence of users and policymakers.
- **User Satisfaction**: Statistics generated must meet the needs of users; Priority needs should be identified and there must be ways to identify how to improve use cases which are addressed by data.
- **Accuracy:** Statistics must accurately and reliably portray realities on the ground.

The objective is to ascertain intentionality, ensure sustainability and scalability in strengthening the data ecosystems.

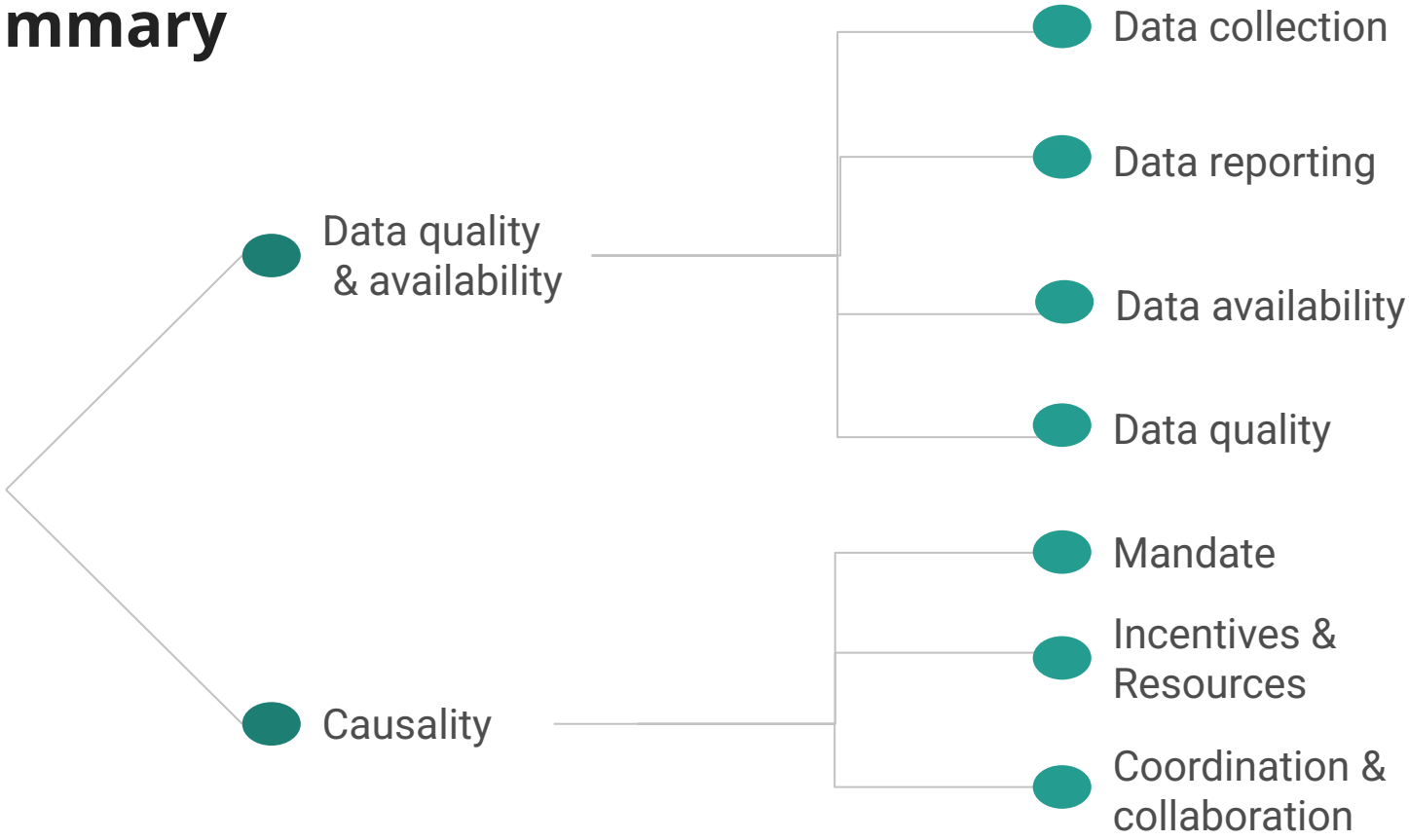
Operationalising Data Quality and Availability - 2

Causality focuses on the legislation and systems in place for data management across the entire sector. While the data quality and availability assessment will provide an indication of the robustness of data, it does not explain potential reasons for why this is the case and what should be done to improve data quality.

This is assessed through examining:

- **Mandates**: Data collection is legally mandated, with penalties for not reporting data
- **Resourcing**: What is the level of resourcing - both staff and material resources - to enable data quality and availability
- **Collaboration**: what opportunities for collaboration with entities outside of local and national government exist to strengthen data availability and quality

Summary


Data quality
& availability
Data collection
Data reporting
Data availability
Data quality
Causality
Mandate
Incentives &
Resources
Coordination &
collaboration

Let’s Pause

What is data quality?

Quality is a multidimensional concept, which means the degree to which a set of inherent characteristics fulfils requirements.

In other words, if quality is in reference to **statistical output**, then we are more interested in the rigor of the **statistical process that produced it**, the **institutional environment** surrounding the process, or the whole statistical system.

Additional tenets - Fundamental principles of Statistics

Principle 1: Impartiality, relevance, and equal access
Principle 2: Professional standards, scientific principles, and professional ethics
Principle 3: Transparency and accountability
Principle 4: Prevention of misuse
Principle 5: Sources of official statistics
Principle 6: Confidentiality
Principle 7: Legislation
Principle 8: National coordination
Principle 9: Use of international standards
Principle 10: International cooperation

United Nations Statistical Division. (2014). *Fundamental principles of official statistics*. https://unstats.un.org/unsd/dnss/hb/E-fundamental%20principles_A4-WEB.pdf

Let’s Pause

Insights on Data Quality and Availability from the UN Fundamental Principles of Official Statistics

- **Professional Standards Enhance Data Quality (Principle 2):** Adherence to scientific and ethical standards ensures data reliability and accuracy.
- **Transparency Builds Trust and Accessibility (Principle 3):** Publicly disclosing methods, sources, and procedures allows users to verify data integrity.
- **Appropriate Data Sources Ensure Relevance (Principle 5):** Using the best available sources (e.g., surveys, administrative records) ensures comprehensive and relevant data.
- **International Standards Promote Comparability (Principle 9):** Adopting internationally recognized methodologies and classifications improves data consistency across regions and timeframes.
- **Coordination Prevents Fragmentation (Principle 8):** Collaboration among statistical producers reduces duplication and ensures consistency.

Additional tenets - from Research methods

Research methods are guided by foundational principles that ensure **rigor**, **credibility**, and **ethical integrity**.

1. **Objectivity:** Avoid personal biases or preconceptions influencing the research process or outcomes.
2. **Reliability:** Consistency in results when the study is repeated under the same conditions.
3. **Validity:** The extent to which research measures what it claims to measure.
    a. **Internal Validity:** Refers to the extent to which a study accurately establishes a cause-effect relationship between variables *within the specific experimental conditions*, minimizing the influence of confounding factors or biases.
    b. **External Validity:** Refers to the degree to which the study's findings can be generalized or applied to *other populations, settings, or contexts* beyond the original research environment.
4. **Ethical Integrity:** Protecting participants' rights, privacy, and dignity especially if they are children or are drawn from marginalized groups
5. **Transparency:** Openly reporting methods, limitations, and potential conflicts of interest.

2. Challenges in Data Harmonisation

Components of a data ecosystem

The ecosystem encompasses multiple data communities; distinct types of data; institutions, laws and policy frameworks; technologies, platforms and tools; and the dynamic interactions among the actors within prevailing technological, infrastructural, legal, policy and other constraints

Challenges in harmonizing definitions within ecosystems

- **Divergent Definitions**: Addressing inconsistencies in classifying gender, disability, or ethnicity across datasets and within ecosystems.
- **Cultural Variability**: Lack of guidelines to reconcile differing cultural norms (e.g., varying perceptions of gender roles).
- **Technical Barriers**: Standardizing formats (e.g., JSON/XML schemas) to integrate multiple data sources and facilitate data interoperability.
- **Legal Frameworks**: Navigating conflicting data privacy laws that hinder cross-border harmonization

Power relations within the African data ecosystem

- The challenges of data collection, use and sharing in Africa are too often driven by non-African stakeholders.
- Those at the top of the iceberg wield more power and influence than those who remain ‘hidden’ – with individuals/citizens (especially indigenous and marginalized communities) at the very bottom

Let’s Pause

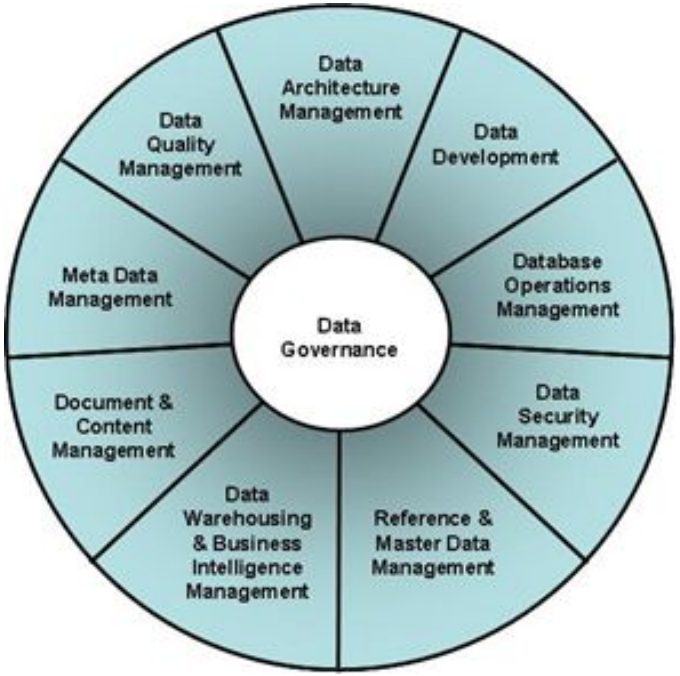
Addressing data harmonization

How can concepts in DATA GOVERNANCE assist to overcome challenges in data harmonization?

Data governance

It is is the exercise of authority, control and shared decision making (planning, monitoring and enforcement) over the management of data assets. Data governance is a core component of 9 disciplines

Specifically, data governance is the **approach that will shape the philosophy of data acquisition, management and archiving;** thus a governance program tries to bridge the Information Technology needs and those of various users in order to define data elements and the rules that govern the data across applications.



Why data governance is important

1. Whereas emerging technologies and expanded access and use of digital and mobile communications present new opportunities for Africa to harness new data sources for sustainable development, the increase in digitization brings new challenges. **These include trust in data, privacy protection and effective data governance.**
2. Needs of each of the stakeholders are different.
3. Information Technology is just an enabler; the interaction between **people, processes, technology and culture drive success of data governance**
4. Without identifying and addressing all these components, it is not possible to move from an ad-hoc state to a mature, focused process of data governance that is striving towards continuous improvement

Recap - Methodological soundness

1. Put in place the correct research design and analytical framework.
2. Select reliable and accurate data sources
3. Decide periodicity of data publication
4. Present data clearly in charts, maps and tables
5. Reports are published in hard and electronic copy
6. Metadata is provided alongside the published statistics and is availed
7. Mentioning focal persons on the report

Interoperability Requirements

1. **Data Portability**: It relates to the capability of moving data in and out of the data centre or cloud service environment.
2. **System/Application Portability**: It relates to the capability of moving specific application codes to or from the cloud service. To port code from one cloud service to another, the target application environment must be usable by the application being ported.
3. **Quality of Service (QoS):** Everything in cloud computing is delivered as a service, so it is essential to ensure the quality of cloud service for user satisfaction. The terms and conditions of contract between the Service Provider and its Client (in this case the county government) are defined in Service Level Agreements (SLA).
4. **Security:** Migrating an on-premise application to the cloud may present the county government with a number of security risks and threats like the protection of intellectual property, personally identifiable information that could fall into the wrong hands.

Interoperability guidelines

- Lack of data standards (and compliance to them) for effective interoperability and embedding of good meta data to support important trends is a key area of concern.

- ***Data interoperability*** *refers to the ability of diverse systems, organizations, or applications to seamlessly exchange, interpret, and use data through standardized formats, protocols, and structures. It ensures data consistency and compatibility across different platforms, enabling integrated analysis and decision-making without manual reformatting.*

3. Case Study - Operationalizing KeSQAF for Gender & Disability Data

Why KESQAF?

- It’s a tool that is used to continuously guide, monitor and assess the quality of statistics produced in the NSS.
- The framework also sets the base for the development of other tools such as the
  - Data Quality Assessment Frameworks,
  - Statistics Code of Practice
  - Frameworks for operationalizing the use of alternative sources of data for official reporting

Benefits of KESQAF

1. Address quality assurance in different circumstances and situations, thereby supporting the country in safeguarding the role of official statistics as a trusted source of information in a changing environment.
2. Provides context for quality concerns, activities and initiatives and explains the relationships between the various quality procedures and tools.
3. Formalizes both operational standards and criteria for evaluating the fitness of statistical data for their required purposes, as well as the methodologies used for data collection, processing, analysis and dissemination, and strategic and managerial issues of statistics.
4. Aims to improve consistency and efficiency in the NSS by reducing duplication of efforts; normalizing the use of international concepts, definitions, classifications, standards, sampling frames, and methodologies, where appropriate; and, creating datasets that are responsive to data sharing demands.
5. Aims at increasing effectiveness and efficiency in the production of statistics in the Country.
6. Forms a foundation for validating data from both primary and alternative sources.

Legislative framework around KESQAF

The development and implementation of KeSQAF is supported by various legal frameworks, strategies, policies, procedures and statistical quality frameworks. These include the:

1) Statistics Act, 2006;
2) Kenya Strategy for Development of Statistics (KSDS);
3) ISO 9001:2015 standards;
4) Existing Data Quality Assessment Frameworks (DQAFs);
5) United Nations National Quality Assurance Framework (UN-NQAF);
6) UN Principles Governing International Statistical Activities;
7) African Charter on Statistics;
8) International Statistical Institute's (ISI) Declaration on Professional Ethics;
9) Statistical Data and Metadata Exchange (SDMX);
10) European Statistics (Eurostat) Code of Practice (CoP).

Inclusive data in KeSQAF

Gender and Disability are only mentioned ONCE in KeSQAF - hence it is not clear from the framework how to integrate gender and disability data into this framework

The technical capacity building and harmonization of statistics within the NSS is achieved through technical working groups and committees, on different thematic/sectoral areas. The technical committees/groups are made up of statisticians and statistics focal persons from the various MDAs and meet on a quarterly basis to interrogate and harmonize data on the respective sectors/thematic areas. The established technical working groups/committees include Agriculture, Nutrition and Environment; Trade and trade Facilitation and Balance of payment; Gender Statistics; Health Statistics; Education Statistics; Governance Statistics; Tourism Statistics; Monetary Statistics; Disability Statistics; and Energy Statistics. The coordination also extends to the representation of the country in regional, continental and global statistical committees on various thematic areas.

Let’s Pause

Enhancing KeSQAF to meet the needs of Inclusive Data

How can KESQAF better address inclusive data?

Step 1 - Institutional Alignment and Governance

Establish Gender & Disability Focal Points:

- Create specialized roles within the KNBS Quality Working Group (QWG) and National Quality Working Group (NQWG) to oversee gender and disability data.
- Integrate gender and disability metrics into the **Kenya Strategy for Development of Statistics (KSDS)** to prioritize their inclusion in national statistical plans.

Stakeholder Collaboration

- Partner with civil society organizations (CSOs), such as gender rights groups and disability advocacy bodies, to co-design data collection tools and validate methodologies (as outlined in *Annex 4: CGD Validation*).

- Leverage the **Technical Working Groups** (e.g., Gender Statistics, Disability Statistics) to harmonize definitions and classification.

Step 2 - Methodological Enhancements

Data Collection Standards

- **Inclusive Sampling**: Ensure surveys and censuses include representative samples of women, marginalized genders, and persons with disabilities (PWDs). Address underrepresentation in hard-to-reach populations (*Principle 13: Managing Respondent Burden*).
- **Disaggregation**: Mandate gender and disability disaggregation across all statistical outputs (e.g., employment, education, health) to enhance *relevance* and *coherence* (Level D Principles).

Quality Validation

- Adapt the **Citizen Generated Data (CGD) Validation Criteria** (Annex 4) to assess gender and disability data. For example, measuring coverage gaps (e.g., rural vs. urban PWDs) and non-response biases

Ethical Safeguards

- Strengthen confidentiality protocols (*Principle 7: Statistical Confidentiality*) to protect sensitive data (e.g., gender-based violence, disability status).
- Train enumerators on ethical data collection practices, including informed consent and cultural sensitivity

Step 3 - Capacity Building & Resource Allocation

Staff Training

- Develop modules on gender-sensitive and disability-inclusive data practices for KNBS and NSS staff, focusing on:
- Avoiding biases in survey design (*Principle 10: Methodological Soundness*).

Technological Integration

- Invest in digital tools (e.g., mobile apps) to improve accessibility for PWD respondents (*Principle 17: Accessibility*).
- Create open-data platforms with gender- and disability-specific dashboards to enhance public access

Funding & Partnerships

- Mobilize resources through partnerships with donors (e.g., GIZ, Gates Foundation) to address gaps identified in the KeSQAF SWOT analysis (e.g., low staffing, financial constraints).
- Pilot innovative funding models, such as public-private partnerships, to sustain long-term data initiatives.

Step 4 - Monitoring & Continuous Improvement

Track progress using KeSQAF's quality dimensions:

- **Timeliness**: Ensure gender and disability data align with SDG reporting cycles.
- **Comparability**: Standardize indicators across counties for benchmarking (*Principle 18: Coherence*).

Feedback Loops

- Conduct regular **user satisfaction surveys** with marginalized groups to assess data utility and relevance.
- Publish annual quality reports highlighting gender and disability data improvements, aligned with *Principle 8: Commitment to Quality*.

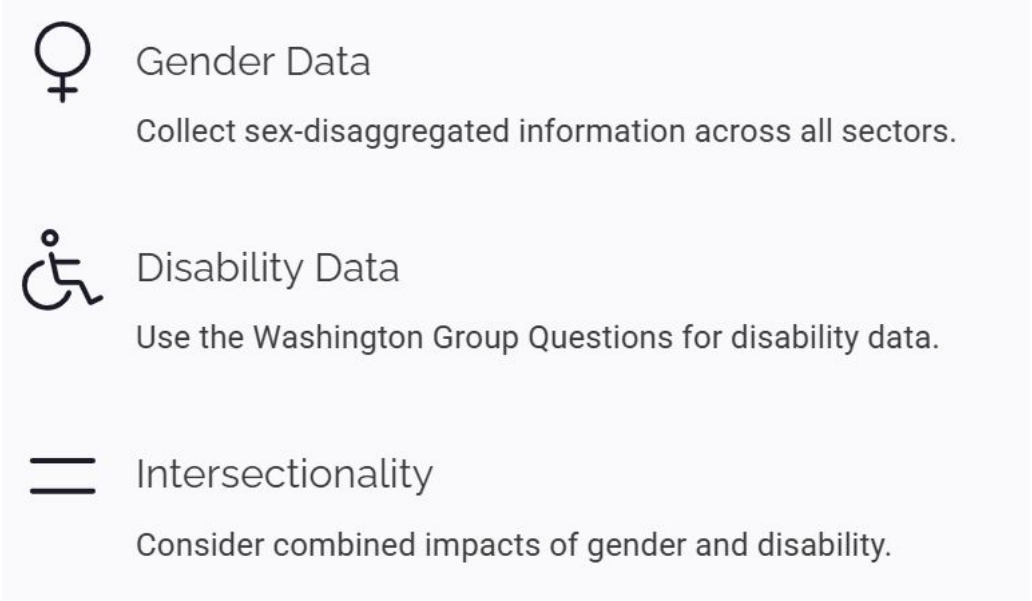
Let’s Pause

Key Considerations in addressing gender/disability data in KESQAF

- Revise KeSQAF annexes to include gender and disability-specific guidelines (e.g., expanding Annex 4).
- Intersectionality: Address overlapping vulnerabilities (e.g., women with disabilities) through multidimensional data frameworks.
- Legal Compliance: Align with Kenya’s Disability Act (2003) and Gender Policy 2021-2025 to enforce accountability.

4. Improving Gender & Disability Data Availability

Improving Gender & Disability Data Availability



- **Mandatory Disaggregation**: Enforce standards requiring gender and disability data splits in national surveys, censuses, and administrative records (e.g., non-binary gender options, disability categories aligned with the Washington Group questions).
- **Accessible Technology**: Deploy inclusive mobile/online tools (e.g., voice-to-text interfaces, sign language support) to reach underrepresented groups and reduce participation barriers.
- **Partnerships with Marginalized Communities**: Collaborate with NGOs, advocacy groups, and grassroots organizations to co-design surveys and access hard-to-reach populations (e.g., rural women, nomadic disabled individuals).

Improving Gender & Disability Data Availability

- **Capacity Building**: Train enumerators and analysts on intersectional data collection and ethical practices (e.g., cultural sensitivity, disability-inclusive methodologies).
- **Transparent Reporting**: Publish metadata and methodologies to clarify how gender and disability data are collected, ensuring accountability and enabling cross-validation.
- **Policy Alignment**: Link data collection to global frameworks (e.g., SDGs) to prioritize funding and institutionalize inclusive practices.
- **Ethical Validation**: Establish community review boards with gender and disability advocates to audit data quality and relevance.

Let’s Pause

What **systemic reforms** are necessary to institutionalize the consistent collection, accessibility, and utilization of intersectional gender and disability data across all levels of policy making and service delivery?

- This question compels stakeholders to address structural barriers (e.g., funding gaps, outdated methodologies), prioritize standardization (e.g., aligning with global frameworks like the Washington Group Short Set), and ensure marginalized voices shape data ecosystems to drive equitable outcomes.

5. Addressing Data Quality in Marginalized Contexts

1. Co-Design Data Tools with Marginalized Communities

How?

- Partner with local leaders, advocacy groups, and community members to co-create surveys, apps, and reporting systems that reflect local dialects, cultural norms, and accessibility needs (e.g., voice-based tools for low-literacy populations).

Why?

- Marginalized groups often distrust external data collectors due to historical exploitation. Co-design builds trust, ensures cultural relevance, and captures nuanced realities, leading to **higher accuracy** and **ethical representation** in datasets.

2. Invest in Intersectional Data Disaggregation

How?

- Mandate collection of granular demographic variables (e.g., disability status, ethnicity, gender identity) and analyze overlapping identities (e.g., rural women with disabilities).

Why?

- Broad categories (e.g., "Other") erase unique challenges. Disaggregation exposes hidden disparities, enabling **targeted interventions** and ensuring no group is invisibilized in policy decisions.

3. Implement Participatory Validation Processes

How?

- Establish community review boards to audit datasets for biases, missing populations, and misinterpretations. For example, Indigenous elders verifying land-use data.

Why?

- External analysts often misinterpret cultural context. Participatory validation corrects errors, strengthens **data integrity**, and ensures findings align with lived experiences.

4. Prioritize Accessible & Multimodal Data Collection

How?

- Use mobile tools with offline functionality, voice recordings, and visual aids (e.g., pictorial surveys) to reach populations excluded by text-heavy or internet-dependent methods.

Why?

- Marginalized communities often face barriers like illiteracy, disability, or lack of connectivity. Accessible methods improve **participation rates** and ensure data reflects the full spectrum of society.

5. Embed Ethical Safeguards Against Exploitation

How?

- Anonymize sensitive data, obtain informed consent in local languages, and share benefits (e.g., returning findings to communities as actionable insights).

Why?

- Marginalized groups are vulnerable to data misuse (e.g., stigmatization, surveillance). Ethical practices protect rights, foster trust, and ensure data collection **empowers rather than harms**.

6. Build Local Capacity for Sustainable Data Governance

How?

- Train community members as data stewards to manage collection, analysis, and dissemination, ensuring long-term ownership.

Why?

- External agencies often withdraw post-project, leaving gaps. Local capacity ensures **continuity**, reduces dependency, and aligns data practices with community priorities.

Key Takeaways from Module 4

- **Robust Quality Assurance Frameworks:** Use standards like KeSQAF to ensure data accuracy, relevance, and ethical integrity. This includes systematic validation, granular disaggregation and alignment with international principles.
- **Institutionalize Inclusive Data Practices:** Integrate gender and disability metrics into national frameworks (e.g., Kenya's KSDS) through specialized focal points, stakeholder collaboration (NGOs, advocacy groups), and ethical safeguards
- **Prioritize Intersectional Disaggregation:** Mandate granular data collection (e.g., "rural women with disabilities") to expose hidden disparities and enable targeted policies.
- **Invest in Capacity and Technology**: Train staff on inclusive methodologies and deploy digital tools (mobile apps, open-data platforms) to enhance accessibility and scalability.
- **Address Harmonization Challenges**: Overcome divergent definitions, cultural norms, and legal barriers by standardizing methodologies and leveraging interoperable formats (e.g., JSON/XML).