Handling PII's in research data management – Using IITA as case study

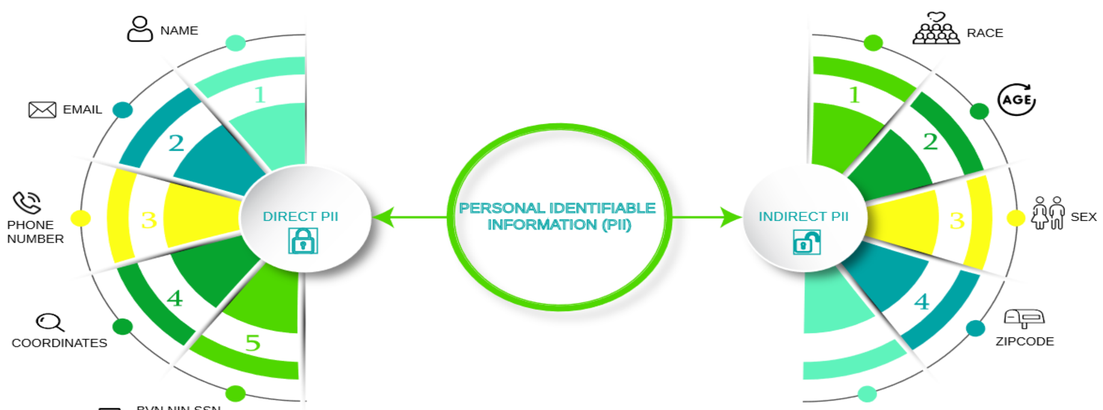
Adeoluwa, Olawamiwa[1], *orcid.org/0000-0003-3817-3807,* **Oluwasoga, Olayemi**[1], *orcid.org/0000-0003-0181-888*
[1]International Institute of Tropical Agriculture IITA, Data Management Unit, PMB 5320, Ibadan, Nigeria

Abstract
Sub-Saharan Africa is fast developing as a hub of research. There are many innovations, governance and sustainable development activities being carried out. These activities require that the critical role of data privacy, protection, and management in research becomes increasingly important. The International Institute of Tropical Agriculture, (IITA) is one of the leading institutions championing research in Sub-Saharan Africa with its headquarters in Nigeria, West Africa. With a focus on agriculture, the institute has different thematic areas of research ranging from (Plant health, Socio-economy, Food Nutrition, Natural Sciences, Genetics and biotechnology and Agribusiness). These research areas generate lots of robust data and some contain personally identifiable information (PII) which must be protected. In this poster, we attempt to show how IITA ensures that the collection, use, and dissemination of data adhere to ethical and legal standards. We highlight advanced techniques for securing research data, including data anonymization, encryption, and secure data-sharing protocols, which are essential for protecting sensitive information and maintaining public trust. This work aims to equip researchers, policymakers, and stakeholders with actionable insights and strategies to enhance data protection practices.

Keywords: sustainable development, data privacy, data management, data sharing protocols, GDPR, IRB.

Introduction
In an era where data is the lifeblood of research and innovation, the management of Personally Identifiable Information (PII) has become a critical concern. Ensuring the privacy and security of sensitive data is not only a legal obligation but also a moral imperative for researchers. This poster discusses the intricacies of handling PII within the framework of research data management, limiting the risk of re-identification.

Classification of PII



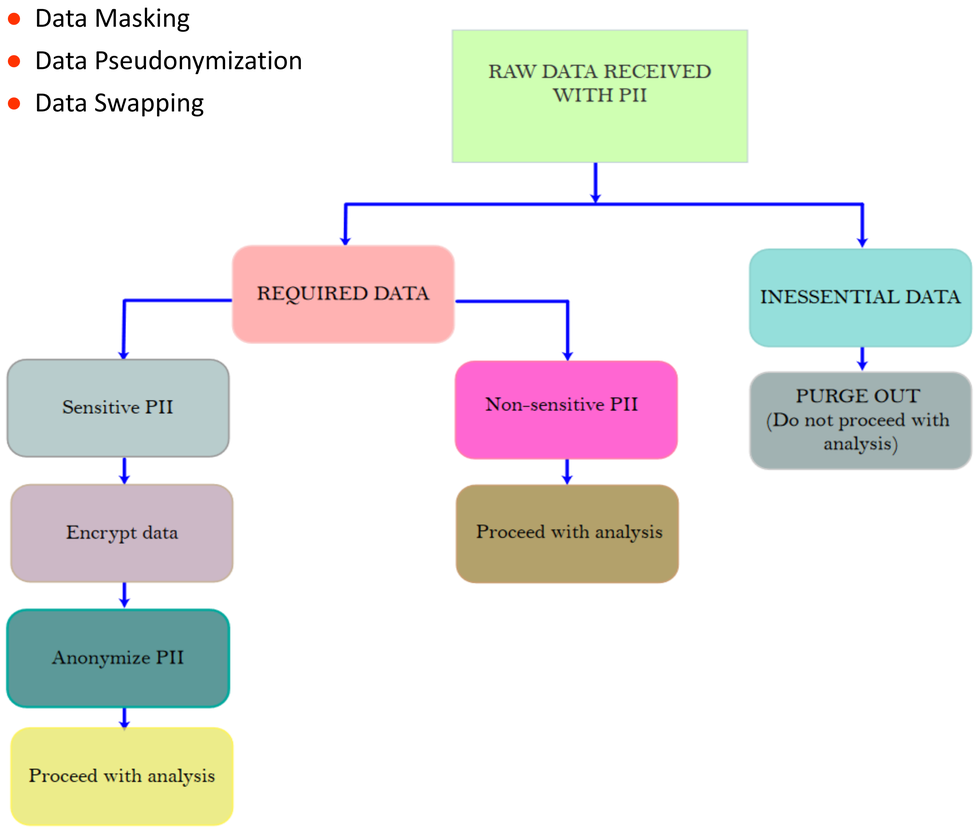
Objective
Importance of PII Management:
PII management plays a significant role in data management, particularly when it comes to handling data pertaining to specific individuals.

- Protecting Individual Privacy
- Ethical Responsibility
- Legal Compliance
- Avoid Data Breaching

Legal and Ethical Considerations in PII Management
Confidentiality is key, socioeconomic data often includes sensitive information about human respondents, Quasi identifiers, and more. Protecting this data is crucial to maintaining the privacy and trust of participants. These are treated by anonymizing such information and transforming it in such a way that it cannot easily be linked to a specific individual. Reducing k-anonymity to k-1.

A few ways of Anonymizing sensitive information:

- Data Masking
- Data Pseudonymization
- Data Swapping



IITA's Data Management Practices:
Below are few methods and protocols used by IITA to manage PII.

- Identify and classify PII in terms of sensitivity
- Determine how PII is stored, moved and shared if necessary
- Access to the sensitive information
- Encryption method/process
- Complete removal of columns containing sensitive information to avoid re-identification through joining of data from other sources.

Table 1. Original dataset. The attribute Name is an Identifier. Instead, age, Gender and Postcode are Quasi-Identifiers.

Name	Age	Gender	Postcode	Coordinate
Alice	24	F	80015	N00345`S35673
Musa	28	M	80019	N23348`S484677
Aderoju	42	F	85073	N54321`S12345
Nneka	49	M	85071	N60594`S20938

Table 2. Dataset K-anonymized. Considering the QI, the number of indistinguishable rows are two. So, the dataset is K-anonymized (K = 2)

Name	Age	Gender	Postcode	Coordinate
	20-35	F	800**	***
	20-35	M	800**	***
	36-50	F	850**	***
	36-50	M	850**	***

Effectiveness of these methods

- Identity of a respondent can no longer be identified either directly or indirectly.
- Reduces the risk of identifying a respondent
- Data can be deposited into an open repository for other research scientists or public.

Best Practices and Recommendation
These are a few strategies that research institutions can adopt to significantly enhance PII management, reduce the risk of breaches, and build trust with participants.

- Stay updated on privacy regulations and technologies: Adapt anonymization practices accordingly.
- Ensure research complies with relevant ethical guidelines.
- Develop Comprehensive PII Policies: Create clear guidelines for PII handling, storage, use, sharing, and disposal.
- Quality control and risk assessment to identify potential PII breaches and implement mitigation strategies.
- Data Pseudonymization/Anonymization: Consider transforming PII into non-identifiable data when feasible.
- Adopt secure Data Storage: Utilize secure data storage solutions, such as encrypted cloud storage or on-premises data centers.
- Regulatory Compliance: Ensure compliance with relevant data protection regulations (e.g., GDPR, CCPA).

Conclusion
Effective management of PII is essential for ensuring the integrity and trustworthiness of research data. Sensitive information must be anonymized in data because we cannot control how users will consume the data generated. The case study of IITA highlights both the challenges and successes in this domain, offering valuable insights and practical strategies for other institutions. By adopting these best practices, research organizations can enhance their data management frameworks, protect sensitive information, and comply with legal and ethical standards.

Acknowledgement
Special thanks to the International Institute of Tropical Agriculture (IITA), CGIAR for the opportunity given to attend this workshop. Also acknowledging the Data Management Team for their unwavering support all through.

References
- *Wilkinson, M. D., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship*. *Scientific Data, 3,* 160018. Available at: https://doi.org/10.1038/sdata.2016.18
- *Makulilo, A. B. (2016). African Data Privacy Laws*. *Springer International Publishing*. DOI: 10.1007/978-3-319-47317-8
- *Regulation (EU) 2016/679 of the European Parliament and of the Council* (2016). Available at: https://gdpr.eu