

Unearthing Silent Data: Back to Basics in Process Design

Jae Sook Cheong

Africa Multiple Cluster of Excellence
University of Bayreuth

Junseong Bang

Management Strategy Headquarters
Ymatics Corp.

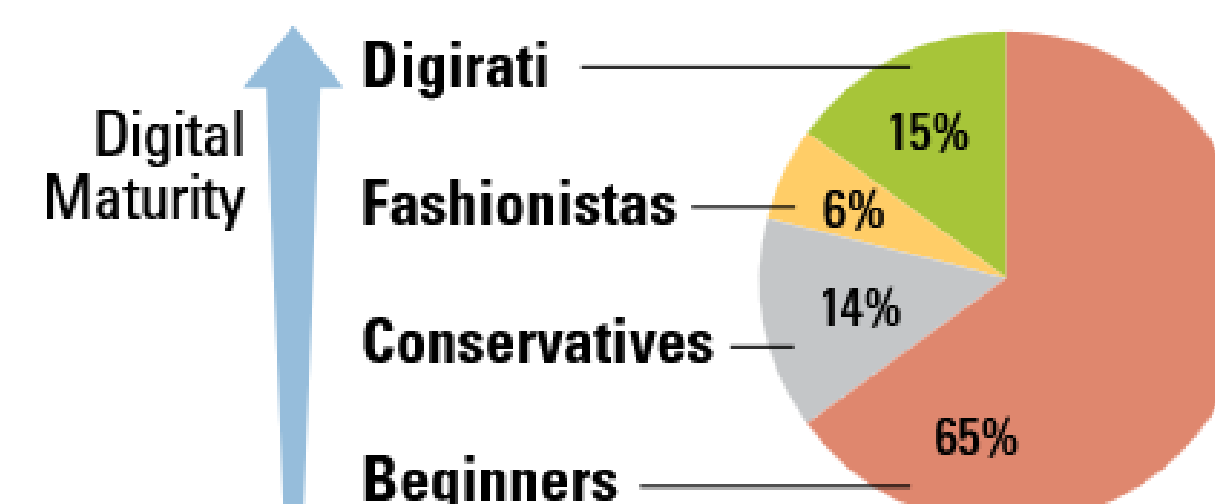
Abstract

We live in technology-driven societies, especially shaped by Artificial Intelligence (AI). The performance of AI models depends heavily on data quality—particularly the issue of **data bias**. Among the many sources of bias, one critical yet often overlooked type is **silent data** and **dark data**. This often stems from organizations in earlier stages of digital maturity, which may encounter difficulties in fully leveraging their data for operational and strategic purposes. These difficulties begin at the foundational level—data collection and data quality maintenance—hindering their ability to achieve strategic goals.

In this work, we proposed comprehensive protocols for the management of the research projects and related administrative operations with a focus on data collection into two databases and on enhancing data quality. We also developed a strategy of maintaining data quality with limited resources, addressing the challenges posed by diverse and scattered data sources.

The ratio of companies in digital maturity levels.

M. Fitzgerald, N. Kruschwitz, D. Bonnet, and M. Welch. "Embracing digital technology: A new strategic imperative." MIT sloan management review 55, no. 2 (2014): 1.



Digitalization for the Project Objectives

The project's objectives relevant to digital systems are:

- Expanding and reconfiguring African studies. Especially by broadening the range of disciplines involved and developing cross-fertilising methods.
- Encouraging inter- and transdisciplinary collaborations through a research data management system.
- Data ownership. The Cluster assists partner institutions in establishing independent research database systems and promotes their data ownership.

Our Approach

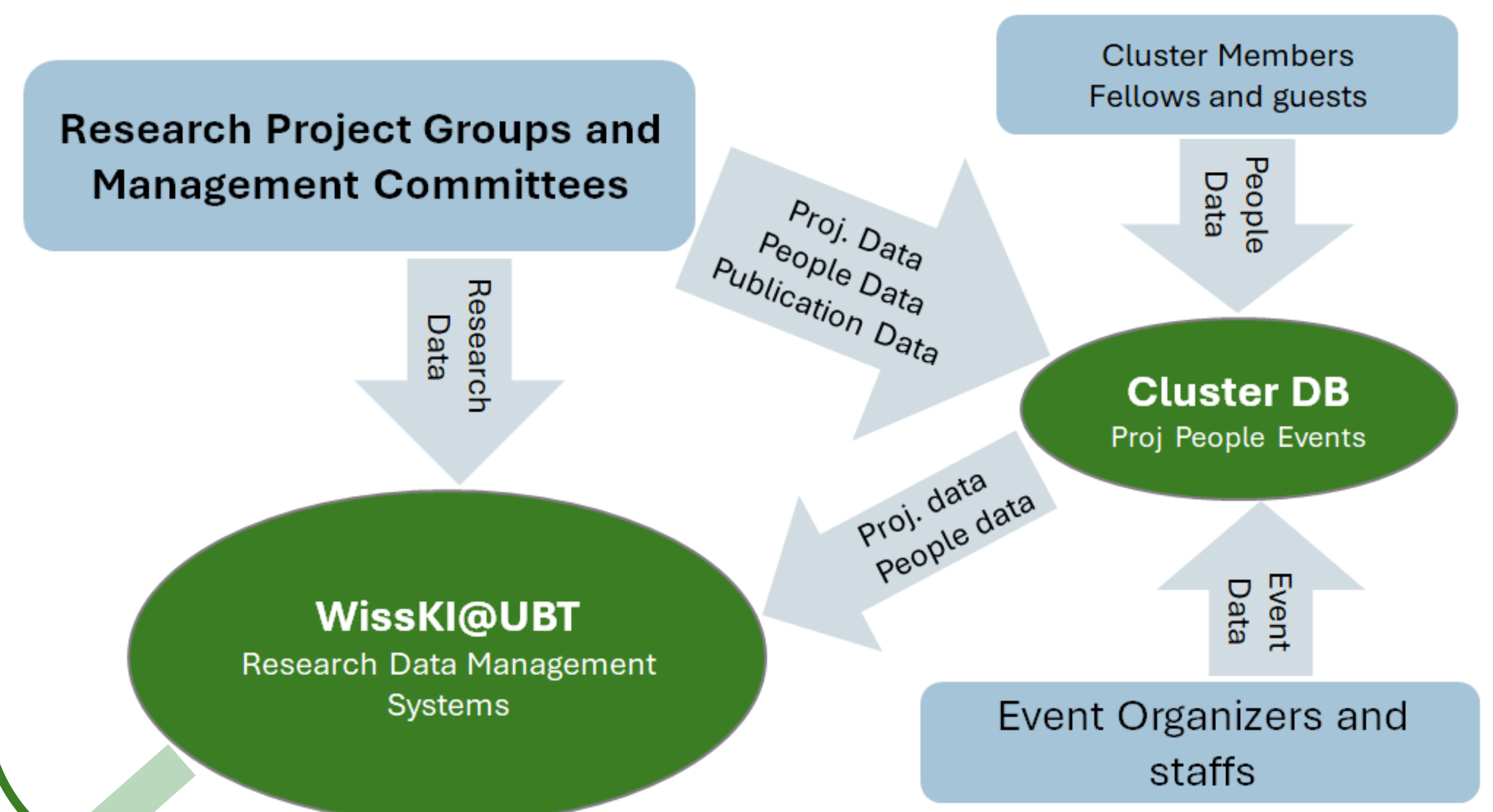
Criteria:

Data Reliability, Data Consistency across DBs, and Process Simplicity.

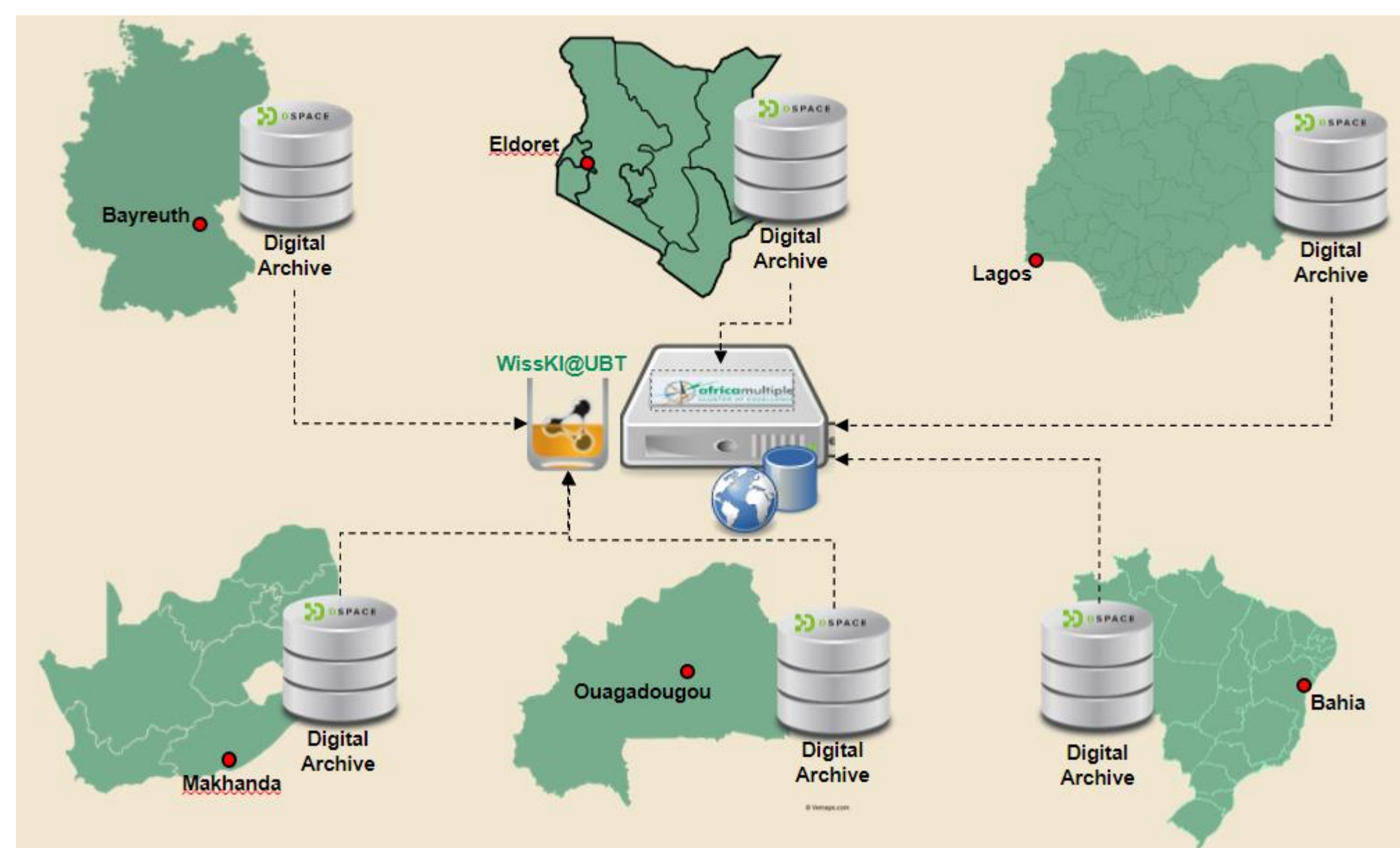
Steps for robust data collection and data quality assurance based on these criteria:

- Design databases with clearly defined purposes.
- Identify all required data collection procedures.
- Develop a stable data collection process for each data category.
- Categorize data into Focus Groups to efficiently maintain data quality with limited resources.

Overview of Databases and dataflow



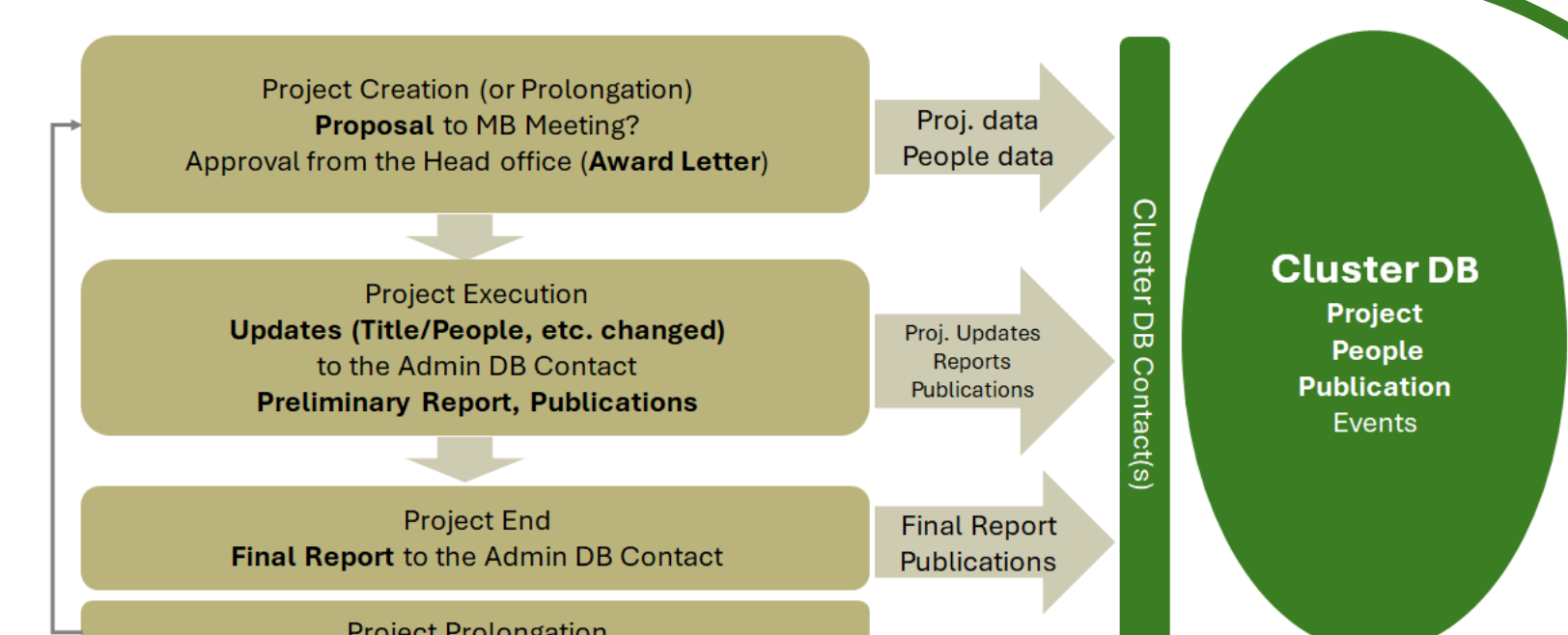
AM Research Data Management Environment



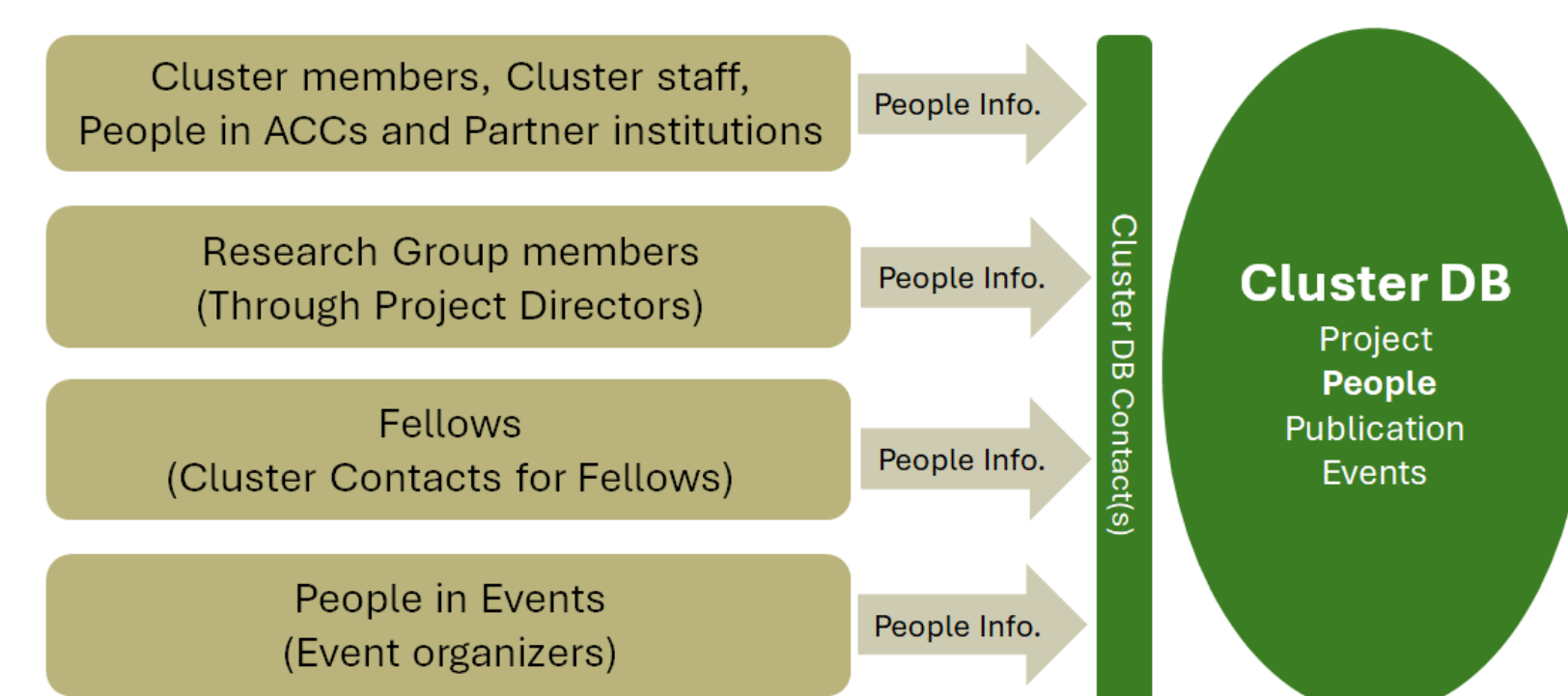
Consultation 1, 2 → Technical Consultation 1, 2
→ Data Intake (Metadata + Research Data) into WissKI@UBT

Data Collection Processes

For Project Data

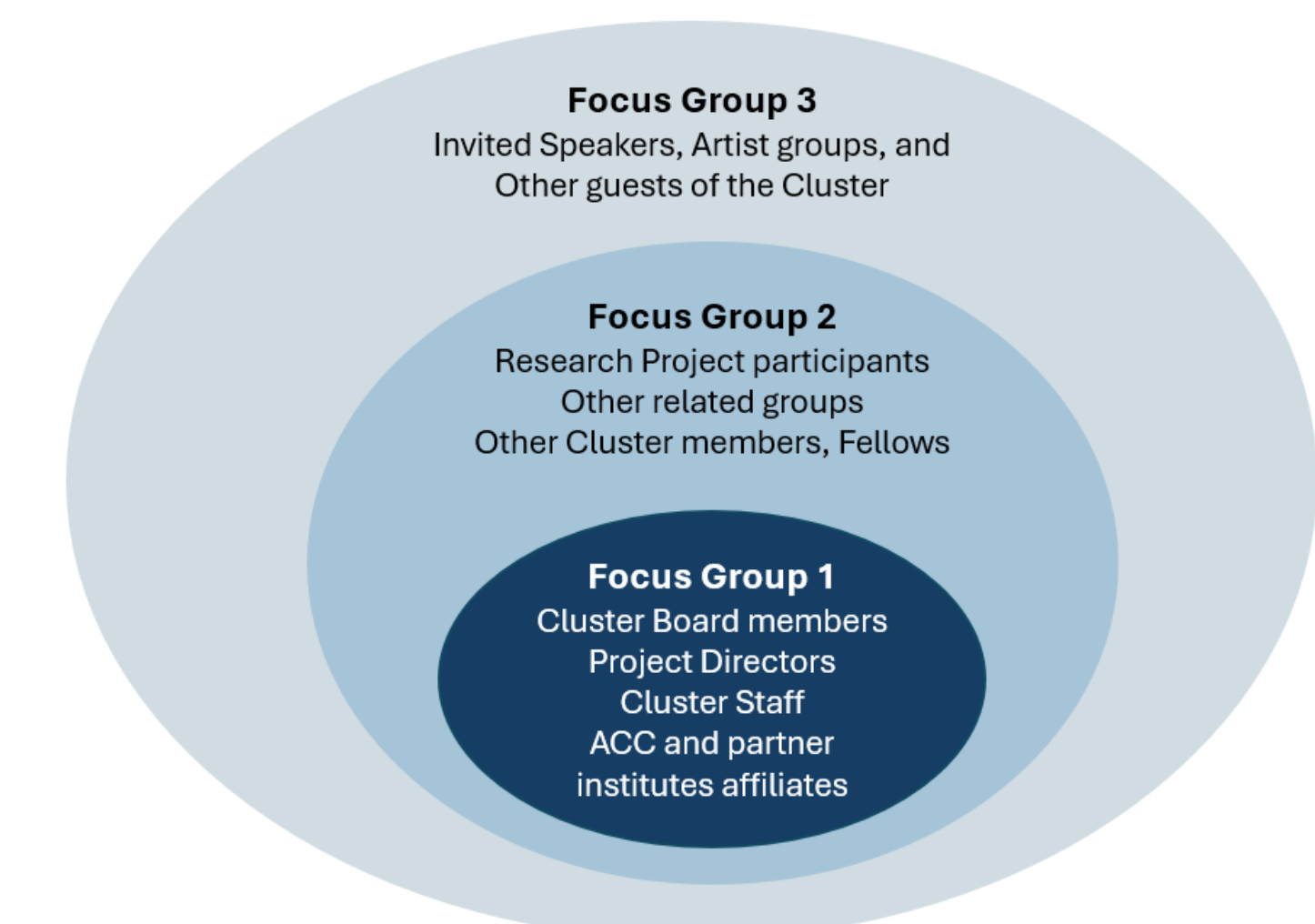


For People Data



Data Quality Maintenance

- The data collected from diverse and scattered sources
- Maintenance with limited resources



Information Sharing (extracted from 12 interviews)

- Centralized announcements of events (lectures, exhibitions, social activities) replacing email broadcasts.
- Accessible archives for magazines, newsletters, and other communications.
- Comprehensive administrative manuals facilitating task execution and staff onboarding.
- Secure, role-specific information portals for different task groups.
- Interactive platforms for feedback, questions, and improvement suggestions.

For effective communication protocols

- Standardized and consistent terminology for administrative operations and research project management;
- Clearly defined data directories with corresponding access groups
- Organizational directory explaining the groups and organizations related to the Cluster

Conclusion

We proposed processes with optimal dataflow around two databases in the Cluster for data collection, and for improving data quality. A strategy to effectively maintain the quality of the data collected from diverse and dispersed sources with limited resources.

Future work include enhancing the user experience for research data exploration in WissKI@UBT, promoting the utilization of the databases, and maintaining the associated work processes.