AFRICAN LONGITUDINAL DATA REPOSITORY [ALDR]

A New CEAR Project to be Launched Soon

<div align="center">**CONCEPT NOTE**</div>

1. Introduction

The Centre for Econometrics and Applied Research (CEAR) is dedicated to advancing rigorous, data-driven research addressing the most pressing socio-economic challenges and opportunities facing the African continent. A significant and persistent obstacle in this mission is the fragmented, inaccessible, and often inconsistent nature of critical long-term secondary data. Researchers, policymakers, and development partners frequently spend an inordinate amount of time sourcing, cleaning, and harmonizing data from disparate national and international sources rather than focusing on high-value analysis. This inefficiency stifles innovation, leads to duplicated efforts, and hampers evidence-based decision-making.

Recognizing this critical gap, CEAR proposes the development of the African Longitudinal Data Repository (ALDR). This will be a centralized data repository for macro analysis on Africa. This initiative aims to create a premier, one-stop digital platform that standardizes, aggregates, and provides seamless access to high-quality, long-term data on all African nations. The ALDR will be an indispensable public good for the global research community, fundamentally lowering the entry barrier for sophisticated econometric and applied research on Africa.

2. Problem Statement

The current landscape for accessing African data is characterized by several key challenges:

i. Data fragmentation: Vital data is scattered across numerous websites of national statistics offices, central banks, ministries, international organizations (such as the World Bank, IMF, AfDB, UN, UNCTAD, etc.), and individual research projects. There is no single, unified access point.

ii. Inconsistent formats and quality: Data is published in non-standardized formats (PDFs, inconsistent Excel sheets), with varying levels of documentation, and often with questionable reliability, making comparability across time and countries extremely difficult.

iii. Limited historical depth: While snapshots of recent data may be available, consistent long-term time series (20+ years) are hard to compile, limiting the ability to study trends, shocks, and long-run development processes.

iv.   High data transaction costs for researchers: The process of data collection is time-consuming, expensive, and requires significant technical skill for cleaning and harmonization, diverting resources away from core research activities.

v.   Underutilization of existing data: Valuable datasets, particularly from government sources, remain underused simply because they are not easily discoverable or accessible in a researcher-friendly format.

3. Project Vision and Objectives

Vision: To become the world's most trusted and comprehensive source of integrated long-term data on African economies and societies, empowering a new generation of research and fostering evidence-based policy across the continent.

Objectives:

The objectives of this project are to:

i.   systematically identify, acquire, and consolidate long-term time-series data from a wide range of credible sources, including national statistical agencies, international databases, academic institutions, and central banks.

ii.   clean, standardize, and document all datasets to ensure consistency, comparability across countries and time, and high quality. This includes converting units to common standards, addressing missing values transparently, and aligning geographic and administrative categorizations.

iii.   develop an intuitive, user-friendly, and powerful web-based platform that allows researchers to easily search, browse, visualize, and download data in multiple formats (CSV, STATA, R, Excel).

iv.   implement robust metadata standards for every dataset, providing clear information on sources, methodologies, definitions, and any transformations applied, ensuring full transparency and reproducibility.

v.   create a dynamic resource centre featuring research publications that have utilized the ALDR, working papers, and forums to encourage collaboration and knowledge sharing among Africa-focused researchers.

4. Scope of Data and Content

The ALDR will be built in phases, with Phase 1 focusing on core macroeconomic and development indicators. The scope will include, but not be limited to, the following domains for all 54 African countries, with historical data spanning as far back as reliably available:

S/N	Structure	Variables
1	Macroeconomic indicators	i. GDP (nominal, real, per capita); ii. inflation rates; iii. interest rates; iv. government revenue and expenditure; v. public debt; vi. trade balances; vii. exchange rates; viii. foreign direct investment.
2	Demographic and social data	i. population size and growth, ii. age structure, iii. fertility and mortality rates, iv. life expectancy, v. migration flows, vi. education enrolment and literacy rates, vii. health indicators (e.g., disease prevalence, healthcare access).
3	Sectoral data	i. agricultural production; ii. industrial output; iii. employment by sector; iv. mineral and resource production.
4	Financial development	i. Domestic credit to the private sector; ii. stock market capitalization; iii. bank branches per capita.
5	Infrastructure and development	i. access to electricity; ii. water and sanitation; iii. internet penetration; iv. mobile cellular subscriptions; v. road and rail density
6	Institutional and governance indicators	i. Data on governance; ii. political stability; iii. corruption perceptions and; iv. The rule of law is measured from recognized sources, such as the World Bank's Worldwide Governance Indicators (WGI) and the Mo Ibrahim Foundation.

Future phases could incorporate geospatial data, firm-level data, household survey data (e.g., LSMS, DHS), and climate/environmental data.

5. Technical Implementation Framework

The platform will be built on a modern, secure, and scalable technology stack using the following client-server architecture or web application stack:

i. Backend: A robust database management system (e.g., PostgreSQL) to store the harmonized data tables and metadata.

ii. Frontend: A responsive web application built with modern frameworks (e.g., React.js, Vue.js) to ensure an intuitive user experience across devices.

iii. Application Programming Interface (API): A comprehensive RESTful API will be a core component, allowing advanced users and other systems to programmatically access data, enabling integration with statistical software and custom applications.

iv. Data processing pipeline: Automated and semi-automated scripts (in Python/R) will be developed for the ETL (Extract, Transform, Load) process to ensure efficient and consistent data updates.

v. Hosting and Security: Secure cloud-based hosting (e.g., AWS, Azure) with regular backups, SSL encryption, and robust access controls to ensure data integrity and security.

6. Target Audience and Key Stakeholders

S/N	Users	Breakdown
1	Primary	i. Academic researchers, ii. graduate students, and iii. economists at universities and think tanks worldwide.
2	Secondary	i. Policy analysts and economists within African governments, regional bodies (AfDB, WAEMU, AU, AfCFTA), central banks, and multilateral institutions (World Bank, IMF, AfDB).
3	Other Beneficiaries	i. Journalists; ii. private sector analysts; iii. investors and iv. non-governmental organizations (NGOs) seeking reliable data for market analysis and project planning.
4	Key partners	i. National Statistics Offices (NBS, GSS, KNBS, SSA, CAPMAS, etc.#)[1];

---

[1] Nigeria Bureau of Statistics (NBS), Ghana Statistical Services (GSS), Kenya National Bureau of Statistics (KNBS), Statistics South Africa (SSA), Central Agency for Public Mobilisation and Statistics (CAPMAS)

		ii. African Development Bank (AfDB); iii. UN Economic Commission for Africa (UNECA); iv. World Bank; v. African Union; vi. other data-producing entities for collaboration and data-sharing agreements.

7. Project Implementation Plan (Phased Approach)

The ALDR project is expected to be nurtured and germinated, going through four (4) implementation phases, ranging from inception to sustainability and growth phase, as stated thus:

Implementation Phase	Title	Duration	Activities
I	Inception	1-3 months	i. Finalise concept; ii. secure initial funding; iii. assemble core project team; iv. establish technical architecture; v. initiate partnerships with key data providers.
	Development and Prototyping	4-9 months	i. Develop the core platform; ii. build the ETL pipeline for initial data domains; iii. populate the database with a pilot dataset (e.g., 20 key indicators for 10 countries); iv. beta version testing with a select group of researchers.
III	Launch and expansion	10-18	i. Public launch of the ALDR MVP (Minimum Viable Product); ii. continuous addition of new data domains and countries; iii. user feedback incorporation; iv. begin development of advanced features (API, advanced visualization tools).
IV	Sustainability and growth	Ongoing	i. Implement a long-term sustainability plan;

			ii. establish a user support system; iii. continuously update data; iv. explore funding models (grants, institutional subscriptions for advanced features, donor support)

8. Monitoring and Evaluation (M&E)

Success will be measured through both quantitative and qualitative metrics:

Usage Metrics: Number of registered users, monthly active users, dataset downloads, API calls.

Research Impact: Number of research papers, articles, and policy briefs that cite the ALDR as a primary data source.

User Feedback: Regular surveys and feedback mechanisms to assess user satisfaction and identify areas for improvement.

Data Coverage: Regular audits of data completeness, timeliness of updates, and breadth of countries and indicators covered.

9. Sustainability Plan

The long-term viability of the ALDR is critical. A multi-pronged sustainability strategy will be pursued:

i. Grant opportunity: The centre can pursue core funding from international development agencies and foundations aligned with data for development.

ii. Institutional partnerships: we can seek annual support from African regional institutions and central banks that are primary beneficiaries of the platform.

iii. Tiered access model: The CEAR can provide all core data free of charge to all users. Explore potential premium tiers for advanced features like personalized data alerts, extended API limits, and custom data harmonization services for institutional subscribers.

iv. Cost management: we can also utilize open-source technologies and cloud-based scaling to manage infrastructure costs efficiently.

10. Conclusion

The African Longitudinal Data Repository (ALDR) is a transformative initiative that directly aligns with CEAR's core mission. By solving a fundamental problem of data access and quality, the ALDR will catalyze higher-quality research, foster greater collaboration, and ultimately

contribute to more effective policies for Africa's development. This project represents a strategic investment in the continent's intellectual infrastructure, with the potential for a monumental return in the form of rigorous, evidence-based progress.