

Integration and standardization of programme and finance data (2021–2025) to enable cost-effectiveness analysis

Lively Minds | Consultant: Data Engineer, Programme and Finance Data Integration | Duration: Short term consultancy: level of effort to be proposed.

1. Background and Rationale

Lively Minds is an NGO that helps LMIC governments deliver quality early childhood development at scale by activating the power of parents.

Across the four school years from 2021–2025 Lively Minds supported Government of Ghana to deliver a programme across 64 districts and 3,447 schools. The roll-out took place in staggered cohorts – where 6 new districts entered the programme each school term, and then received termly support from Lively Minds terms for 3 years.

Randomised evaluation has found that the programme in Ghana increased overall child cognition. Lively Minds would like to express those impact results as cost per unit of learning and be able to answer the question "what does an extra \$100 achieve?" for evidence-led funders.

The cost-effectiveness measurement should be based on the total number of children reached over the full 5 years and the total cost over these 5 years.

The impact and cost data needed for the cost-effectiveness analysis (CEA) exist but are spread across programme monitoring files and finance files that were never designed to be joined. The monitoring files also vary significantly across programme phase and changed slightly across the different school years.

2. Purpose and objectives

Overall purpose. To produce one validated, documented and reproducible dataset that integrates Lively Minds' Ghana programme data and finance data for four school years from 2021–2025, ready to be used directly in a cost-effectiveness analysis (CEA).

Specific objectives

1. Inventory every programme and finance file for 2021–2025 and document structure, coverage and gaps.
2. Agree a common set of identifiers (district, school, rollout cohort, term, programme phase) and apply them across all sources.
3. Combine all programme files and sheets into consistent school-by-term and district-by-term tables (containing KPI achievement); consolidate these across the four-school-year period from 2021–2025
4. Structure finance data into a district-by-term table that separates direct, shared and central costs, for each school year; consolidate these across the full four-school-year period from 2021–2025.
5. Join the programme and finance files into one district-by-term table and one school-by-term table, reconciled to LM's reported totals, with every assumption and unresolved issue logged.

4. Source data

The consultant will receive access to the following data. Lively Minds will provide a complete file list at contract signing.

Source	What it holds	Owner / contact	Format, location, volume	Known issues
Programme data files	Termly district reporting template – separate files in most years by programme phase (set-up, embed 1–3, embed 4, embed 5, sustain). Each file contains sheets with raw data, analysis based on the raw data, and some macro-enabled reports	MEL Manager	Excel files Folder structure for 2023-2025 follow similar structure: typically year, school term (1-3), then District/Excel templates. Folder structure for 2021-2022 varies more and is less complex. Approximately 6 Excel files per school term. Number of sheets per Excel varies by phase, with Embed having the most complex structure with approx. 26 sheets per file. These sheets include many raw data sheets pulled from ClickMedics and Cognito database, as well as sheets where school or district-level performance or reach is analyzed.	Excel template varies significantly with indicators and number of sheets depending on the phase of the district for the term. These indicators and KPIs also changed slightly over the years. The number of rows vary in each file based on the phase and type of dataset. However, the highest number of rows expected across the Excel sheets for one term is 3447 (the total number of schools for all phases put together). Names of columns are largely consistent for the same indicators, but there are some differences in indicators across phases and years. Files do not contain merged cells. Some may contain macros.
Finance data	Expenditure data disaggregated by year, term, Phase and district. Not disaggregated by the school, so school-level disbursements will be estimated	Finance Director	Excel ledgers from Xero accounting software but rebuild and analysed to add needed details	Accounting period differs from school term; currency conversion to GBP to be done using monthly Info Euro rates
Reference data	District and school lists, phase rollout schedule by district, enrolment targets	MEL Manager / Partnerships	Excel	Most files use district or school name, rather than a unique ID code for each. Unique IDs are generated by combining district, circuit and school names.
	Exchange rates	Finance Director	Online Monthly Info Euro rates	To be downloaded online

5. Target data structure

The target data structure for the four-school-year CEA data build will require the following five tables that share identifiers:

Table	Grain (one row per...)	Main contents
A. Reference tables	district; school; term	District ID and agreed district name; school ID and name variants; rollout cohort; term calendar; phase lookup (district × term).
B1. Programme, school level	school × term	All existing KPI and performance data, in long format with a variable dictionary. Also all reach data which includes, Parents, Children, Headteachers, and Teachers enrolled. This includes termly reach and cumulative reach over a period.
B2. Programme, district level	district × term	Roll-up of B1 plus any district-only sheets. New children and new parents enrolled; cumulative reach.
C. Finance	district × term × cost line	Reconciled district disbursements; direct LM implementation costs (teaching and learning materials, printing, training, travel, other); central and shared costs held in separate lines for later allocation; accounting period and attributed programme term.
D. CEA-ready merge	district × term school × term	B1 joined to C, and B2 joined to C, with currency and price-adjustment fields, completeness flags, and a log of exclusions.

Required identifiers on every table include: district_id, rollout_cohort, programme_term (SETUP, EMBED_1 to EMBED_5, SUSTAIN, SUSTAIN_11), plus year and school_term (1–3). Finance tables also carry accounting_period alongside the attributed programme term. School_id should additionally be on every school-level table.

6. Suggested Tasks and approach

The work is suggested to run in six stages, but candidates may propose an alternate approach or suggestions. Stage 2 ends with a sign-off by Lively Minds before the build starts.

Stage 1. Inception and file audit

- Catalogue every file and sheet: year, term, phase, level (district or school), variables, number of rows, completeness.
- Test a sample of files from each year for structural changes.
- Produce a gap and risk list and a confirmed workplan and effort estimate.

Stage 2. Data model and codebook (LM sign-off)

- Propose district and school master lists, with a mapping for spelling variants and any changes over time.
- Build the phase and cohort lookup.
- Map variables across years into one data dictionary, with definitions and units, noting where a KPI definition changed.
- Agree the target tables in section 5 with LM's data, finance and CEA leads.

Stage 3. Programme data pipeline

- Read all sheets by script (no manual copy and paste), standardise, append, and write to B1 and B2.
- Run checks: duplicate schools, orphan schools, impossible values, totals that differ between school and district sheets, unexplained jumps between terms.

Stage 4. Finance data pipeline

- Map the chart of accounts to the cost lines in section 5.

- Attribute each transaction to district, term and phase, and convert accounting periods to programme terms using a documented rule.
- Keep central and shared costs in separate lines with a proposed allocation key, but do not allocate unless LM instructs.
- Apply the agreed currency rule.

Stage 5. Merge, reconcile and validate

- Join B2 and C into table D. Flag districts and terms with spend but no activity, or activity but no spend.
- Reconcile totals to Lively Minds' reported programme figures and data shared from the accounting system, audited or management accounts.
- Walk Lively Minds staff through discrepancies and record decisions in an issues log.

Stage 6. Documentation and handover

- Deliver commented code, a README that lets another analyst rerun the pipeline, and the final codebook.
- Run a handover session for Lively Minds staff: MEL Manager, Finance Director, MEL Associate Director, and CEA consultant

7. Deliverables

#	Deliverable	Format	Payment
1	Inception report: file catalogue, risks, workplan	Document + spreadsheet	30%
2	Data model and codebook, including district, school, cohort, term and phase reference tables (LM sign-off)	Document + CSV	
3	Programme dataset (school × term; district × term)	CSV / Excel	70%
4	Finance dataset (district × term × cost line)	CSV / Excel	
5	Merged CEA-ready dataset (district × term)	CSV / Excel	
6	Reconciliation and data quality report with issues log	Document	
7	Reproducible code, README and handover session	Stata / R / Python, repository	

8. Consultant profile

Role: Data Engineer, Programme and Finance Data Integration

Essential Qualifications

- Degree in relevant field, i.e. computer science, information technology, economics, monitoring and evaluation, or other quantitative or analytical fields, or equivalent relevant experience
- 3-5 years' experience in data engineering, with at least 5 projects combining messy multi-file spreadsheet data into analysis-ready datasets using code (for example, R or Python).
- Experience with programme monitoring data in international development or education.
- Technical expertise in database design, data manipulation, and coding.

- Ability to read and structure financial data, including cost coding and period attribution.
- Record of reproducible, documented workflows (version control, readable code, data dictionaries).
- Fluent English. Strong written documentation.

Desirable Qualifications

- Prior work preparing data for cost-effectiveness or cost-per-outcome analysis.
- Experience in Ghana or West Africa, or with government education data.
- Familiarity with fuzzy matching of place and school names.

9. Data protection and security

- The consultant works only with data supplied by LM and stores it in LM SharePoint.
- No personal data leaves that environment. If any file contains child or parent names or other identifiers, LM will remove or pseudonymise them first, or the consultant will do so as the first processing step.
- The consultant signs LM's confidentiality agreement and follows LM's data protection policy and applicable law, including Ghana's Data Protection Act 2012 (Act 843).
- All data, code and documentation are LM's property and are returned or deleted at the end of the contract.

10. Management, reporting and LM responsibilities

- Reports to Associate Director MEL and works closely with Director Finance and MEL Manager, especially within the inception phase
- Weekly 30-minute check-in or written progress note.
- LM provides, within 2 working days of signing: all files, a named programme data counterpart, a named finance counterpart, and access to anyone who knows the history of the files.
- LM answers consultant queries within 3 working days. Delays beyond this may extend the timeline and are recorded in the issues log.
- Work location: Remote/Hybrid, ideally with possibility to work from office in Tamale, Ghana on select days. Time zone overlap needed: GMT

12. Contract, budget and payment

- Payment milestones: 30% on inception report and signed-off codebook (deliverables 1 and 2); 50% on merged dataset, final code and handover (deliverables 3-7).
- Travel and other costs: Reimbursed, if necessary, with prior approval.

13. Application and selection

Interested consultants should submit the following documents **by 27th October 2026**:

- A CV highlighting relevant qualifications and experience.
- A technical and financial proposal detailing relevant experience, intended approach and methodology, estimated level of effort/days required for the assignment and proposed fee per deliverable quoted in GHS or GBP.

- Two examples of similar assignments completed (anonymised if needed).

Please note: Due to the urgency of this assignment, applications will be reviewed on a rolling basis. Interested consultants are therefore encouraged to apply as early as possible, as the recruitment process may commence before the application deadline.

How to apply: To apply, please click on this link to submit your application [Apply here](#)

Short-listed candidates will be invited to interview and provided a sample of the data to refine their level of effort estimates for final consideration.

Criterion for candidate assessment	Weight
Relevant degree	15 %
Relevant data integration experience	30 %
Approach and understanding of the data problem	15 %
Finance and programme data skills	15 %
Price	25 %