ENERGY ACCESS EXPLORER

Data standards/practices for Energy Access Explorer

Contents

1. Introduction 2

1.2 Scope of the Framework 2

2. Principles and Values 3

2.1 Open Data and Interoperability 3

2.2 Quality, Consistency, and Transparency 3

2.3 Equity and Inclusion in Data Representation 3

2.4 Respect for Data Ownership and Licensing 3

2.5 Scalability and Adaptability 3

3. Overview of the EAE Data Lifecycle 4

3.1 Data Collection: 4

3.1.1 Spatial Data Quality and Evaluation 5

3.2 Data pre-processing (transformation): 5

3.3 Integration of data into EAE 8

3.3.1 Categories/ Data Schemas 8

3.3.2 Dataset 12

3.3.3 Metadata 13

3.3.4 Data access 14

3.3.5 Data visualization 14

3.3.6 Geographies 14

3.4 Monitoring and Maintenance 15

3.5 Archiving and Deprecation 15

4. Summary 15

Appendix 1: Data monitoring report format 18

ENERGY ACCESS EXPLORER

1. Introduction

As a Digital Public Good (DPG), the Energy Access Explorer (EAE) is designed to be open, adaptable, and reusable across geographies and institutions. Its value lies not only in the data it hosts, but also in its potential to be deployed, extended, or forked (cloning) by governments, civil society organizations, and researchers to support inclusive and integrated energy planning. To enable such decentralized adoption[i], whether through sandboxes, localized instances, or integrated decision platforms, it is essential to establish a clear set of data standards and practices.

To support this decentralized and scalable model, it is essential to establish robust data standards and Quality Assurance / Quality Control (QA/QC) practices that ensure consistency, reliability, and interoperability of the data hosted on the platform. A clearly defined framework for standardizing data naming, metadata documentation, symbology, data quality validation, and normalization will streamline data ingestion and reduce friction for contributors and adopters alike. By aligning with international best practices like ISO 19115 (metadata), ISO 19157 (data quality), CEOS QA4EO (data quality) and OGC standards (spatial interoperability), these practices help reduce onboarding friction, improve cross-region comparability, and enhance user trust. Ultimately, a robust data standards framework empowers EAE to function not just as a tool, but as a foundational infrastructure that others can confidently build upon in the pursuit of sustainable energy access.

Developing data standards and practices is a foundational step toward ensuring that EAE remains a trusted, scalable, and impactful tool for inclusive energy planning.

1.2 Scope of the Framework

This framework applies to all data layers and datasets used in or contributed to EAE, including:

- **Raster data** (e.g., solar irradiance maps, population density grids, terrain elevation)
- **Vector data** (e.g., mini-grid locations, administrative boundaries, road networks)
- **Tabular data** (e.g., energy demand by region, household survey indicators)
- **Derived indicators** (e.g., energy access gaps, least-cost electrification potential)

It encompasses both **global datasets** (e.g., satellite-derived data, global basemaps) and **local datasets** (e.g., utility asset maps, subnational demographics) and includes both **open-access** and **restricted or sensitive datasets**, where appropriate safeguards and attribution protocols are needed. The framework is intended for use by all contributors, curators, administrators, and adopters of the platform to ensure that the data within EAE remains accurate, transparent, and fit for decision-making.

2. Principles and Values

EAE operates at the intersection of open data, spatial intelligence, integrated and inclusive energy planning. As a Digital Public Good (DPG), EAE is guided by a set of core principles and values that underpin its approach to data governance and quality assurance. These principles ensure that the data powering the platform remains equitable, transparent, accurate and actionable - regardless of geography or user.

2.1 Open Data and Interoperability

EAE prioritizes the use and promotion of open data wherever possible. All datasets integrated into the platform are expected to be accompanied by clear licensing, metadata, and documentation to enable free access and responsible reuse with attribution to the authors and tool. In parallel, the framework encourages alignment with internationally recognized data standards such as ISO, OGC, and FAIR (Findable, Accessible, Interoperable, Reusable) principles to ensure that data can seamlessly integrate across systems and scales.

2.2 Quality, Consistency, and Transparency

Data used in EAE must meet clear quality thresholds in terms of completeness, accuracy, logical consistency, and traceability (section 3.1). All datasets should be accompanied by standardized metadata and, where relevant, data quality statements following ISO 19157 or QA4EO guidelines. Transparency in data processing methods, transformations, and limitations is critical for enabling informed decision-making and maintaining user trust.

2.3 Equity and Inclusion in Data Representation

The platform is intended to support inclusive planning that accounts for marginalized or underserved populations. Therefore, contributors and curators are encouraged to ensure that data represents all population groups equitably, avoids bias (e.g., urban/rural oversampling), and reflects local realities. Where feasible, local knowledge and ground truthing should complement global or modelled or secondary datasets.

2.4 Respect for Data Ownership and Licensing

EAE respects intellectual property, community data rights, and licensing conditions. Contributors must ensure that all data submitted to the platform is accompanied by appropriate licensing (e.g., CC BY , ODbL), attribution information, and usage restrictions or sensitivity flags, if required. Sensitive or proprietary data should be handled with appropriate access controls while still contributing to the platform's overall utility. The access controls can be discussed with Product Owner or Data QA/QC lead.

2.5 Scalability and Adaptability

To support the platform's DPG nature, the governance model is designed to be scalable across geographies and adaptable to local institutional contexts. Standards and protocols are modular and

can be extended or localized without compromising core integrity. This ensures that data governance practices remain relevant as EAE expands to new regions or evolves in functionality.

Table 1: Core Value/principles for EAE Data Governance

Principle	Description	Alignment/Reference
Open Data and Interoperability	Promote the use of open, machine-readable, and standards-aligned datasets with clear licensing and metadata.	FAIR Principles, OGC Standards, ISO/TC 211
Quality, Consistency and Transparency	Ensure data completeness, accuracy, logical consistency, and clear documentation of sources and processing methods.	ISO 19115-1:2014 - Metadata[ii], ISO 19157:2013 – Data Quality
Equity and Inclusion in Data	Prioritize data that reflects the needs and realities of all communities, including underserved and marginalized populations.	SDG 7, DPG Standards
Respect for Data Ownership & Licensing	Uphold intellectual property rights, community data sovereignty, and appropriate licensing or usage restrictions.	Creative Commons, Open Data Commons

3. Overview of the EAE Data Lifecycle

To ensure the reliability, transparency, and long-term utility of the Energy Access Explorer (EAE), a structured approach to managing data throughout its lifecycle is essential. The data lifecycle defines how data enters, flows through, and eventually exits the EAE platform. It outlines who is responsible (table: 3) at each stage, what quality controls must be applied, and how updates, archiving, and deprecation should be handled. The lifecycle framework ensures that whether data is contributed by a local government partner, an NGO, or a GIS/energy researcher, they adhere to a shared set of principles and practices.

The lifecycle includes five main stages:

3.1 Data Collection:

This step includes following the data tracking sheet and ensuring we are checking that the most recent data is being collected. While collecting the data we need to ensure that we are ensuring the guidelines for the geography of interest, and we know the nuances of the region.

We need to check for the following:

☐ Is the data relevant? Add the reason on data tracking sheet
☐ Data license including permissions and usage restrictions
☐ Initial metadata availability (description, license, date, why is the data required)

3.1.1 Spatial Data Quality and Evaluation

Spatial data quality is usually described through a set of principles (or dimensions) that help determine whether a dataset is *fit for its intended purpose*. For integration of data into EAE, we need to check for the following:

☐ Positional Accuracy: Does the dataset's location match its true placement on the ground? (e.g., a ±10 m shift is acceptable for schools or healthcare facilities).
☐ Attribute Accuracy: Are the attributes recorded correctly and without mislabelling? (e.g., a 33 kV line is listed as 33 kV, not 11 kV).
☐ Completeness: Is the dataset complete[iii], and what percentage of expected records or coverage is included? Is this information included in the metadata?
☐ Temporality: Is the dataset up to date with the most recent version available?
☐ Traceability back to standards – Can the Earth Observation (EO) or derived dataset be traced back to internationally accepted standards (e.g., Quality Assurance Framework for Earth Observation (QA4EO) developed by Committee on Earth Observation Satellites (CEOS))? ☐ Uncertainty Quantification: Has uncertainty been quantified for the EO data, and is this information included in the metadata?

3.2 Data pre-processing (transformation):

EAE hosts data from diierent sources which might come in diierent formats, scales, and units making it diiicult to compare them raw or directly. By applying methods like normalization, downscaling, and resampling, raw data is converted into consistent, comparable, and policyrelevant layers that can support reliable decision-making in tools like EAE.

Please note: Here we are not talking about the Extract, Transform, Load (ETL) process in EAE which is handled by Paver. Paver is the tool developed within EAE which allows users to preprocess the data in the format required for visualization and analysis in EAE Few questions for the geography administrators to review:

- Is the data in same resolution as your EAE geography? If not, does it need resampling?
- Does the data require normalization?
- Are there gaps in your data? Can they be addressed through interpolation?

Please see the checklist below for the data preprocessing:

- **Normalization (type of rescaling)**
  ☐ Convert all units to a common standard (e.g., W/m²/year, people/km²)

☐ Apply scaling (min–max or z-score) if required for comparability. For example: flood vulnerability index in Assam was normalized so that the values can be included while calculating the demand index.

☐ Reclassify categorical values into harmonized classes. For example: Landuse Landcover across all the geographies should have same number of classes. It includes 8 classes (or 9 if permanent snow cover): water bodies, forest, agricultural land, grassland, shrubland, bare land, built-up area, wetlands. ☐ Document the normalization process for metadata.

- **Resampling**

☐ Select resampling method: Are you moving from finer resolution to coarse resolution or vice versa. If you are aggregating, please see checklist A and for downscaling, please see checklist B.

Checklist A: Data Aggregation

☐ **Choose Aggregation Method by Data Type.** This will depend on whether the data is categorical[iv] or continuous[v].

For categorical data:

☐ Majority rule → assign most common class in the desired raster resolution.

☐ Mode → same as majority, but tracks ties[vi].

☐ Weighted dominance → prioritize critical classes (e.g., "urban" overrides "forest").

☐Check for minority-class loss (e.g., small wetlands or waterbodies disappearing).

For continuous data:

☐ Mean (average value, e.g., solar radiation).

☐ Sum (for additive quantities, e.g. energy demand, population count (not population density)).

☐ Median (robust to outliers, e.g., income).

☐ Max/Min (for extremes, e.g., max travel time).

☐**Do you need to handle the boundary?**

☐Identify partial pixels at edges of the grid or study area.

☐Decide rule: clip, weight (fractional area), or include/exclude.

☐Define no-data handling strategy (ignore, assign, or flag).

☐Apply rule consistently across the dataset.

☐ **Data Validation**

☐ Check mass balance (total before = total after, for additive variables).

☐ Compare aggregated results to oiicial totals (e.g., census).

☐ Validate class proportions (categorical) vs original dataset.

☐ **Documentation for the metadata**

☐Record original resolution, target resolution, aggregation method.

☐Report validation results (e.g., mass balance preserved, accuracy %).

☐Note uncertainty and any lost detail for cautions while using the data.

☐Document aggregation and validation choices in metadata.

Checklist B: downscaling

ENERGY ACCESS EXPLORER

☐**Choose Resampling / Downscaling Method** ○

Simple resampling:

- Nearest neighbour (categorical, e.g., land cover).
- Bilinear or cubic interpolation (continuous, e.g., temperature). ○

Statistical / ML downscaling:

- Regression against high-resolution predictors (elevation, land cover).
- Kriging / Co-kriging with auxiliary data.
- Machine learning (Random Forest, Gradient Boosting, Neural Networks).
- Bias correction (e.g., quantile mapping for climate data).

☐**Boundary Handling** ○ Focus only on the

area of interest

☐ Clip outputs strictly to study area.
☐ Do not extrapolate outside valid extent of input dataset.
☐ If predictors extend beyond input, mask to input coverage.
☐ For admin boundaries, aggregate only within oiicial units (avoid overlaps).

○ Does your data have No-Data Cells?

☐ Identify no-data regions in coarse dataset (cloud gaps, missing tiles).
☐ Decide: Mask, Fill (if small gaps), or Hybrid (fill <10%, mask larger).
☐ Ensure predictors with missing values do not drive spurious outputs.
☐ Flag filled areas in metadata as uncertain under 'Data Cautions'.

☐ **Check for boundaries**

☐ Use buier zones in geostatistical or ML methods to avoid border artifacts.
☐ If tiling, apply overlap margins and crop back to avoid seams.
☐ Validate downscaled values near borders against input aggregation.

☐ **Data Validation**

☐ Check mass balance: aggregated downscaled results = coarse input totals. ☐ Compare edge areas with independent reference data (if available).
☐ Report RMSE, Mean Absolute Error (continuous), or Kappa/accuracy (categorical).
☐ Produce uncertainty layers (kriging variance, prediction intervals).

☐**Documentation**

☐Record original resolution, target resolution, processing steps, parameters and software used.

☐Report validation results as well as any deviations.
☐Note uncertainty and any lost detail for cautions while using the data.
☐Document lineage, uncertainty and validation choices in metadata.

3.3 Integration of data into EAE

To successfully integrate the data into EAE, we will follow the checklist for geography, categories, dataset, data access, metadata and data visualization.

3.3.1 Categories/ Data Schemas

In EAE, a category is a label that groups spatial data into classes (for example, healthcare facilities or global horizontal irradiance), making it easier to analyze and visualize. More formally, a category is a discrete label that allows spatial data to inherit defined properties, such as qualitative attributes or quantitative thresholds, thereby structuring the data into meaningful groups for systematic analysis and visualization.

☐ **Check if the category already exists**, before creating a new one. If it exists, simply use it. Please use 'Category Overrides' while adding new data, instead of creating a new category when similar category exists. User can override the styling template of the category for their dataset.

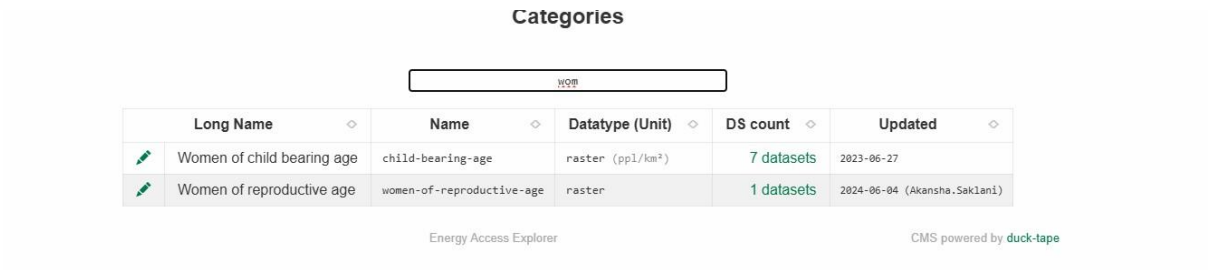


Figure 1: Both the categories are being used for similar data. It's better to have only 1 in this case.

Creating a new category

☐ Define the long name. Name should be intuitive, for example: Healthcare Facilities or Global Horizontal Irradiance. Avoid jargons like GHI_5.5, use intuitive and clear naming.

☐ Define a short name. When adding data, the tool searches using the short name, so it must be clear and closely aligned with the long name. Avoid jargons and abbreviations if possible.

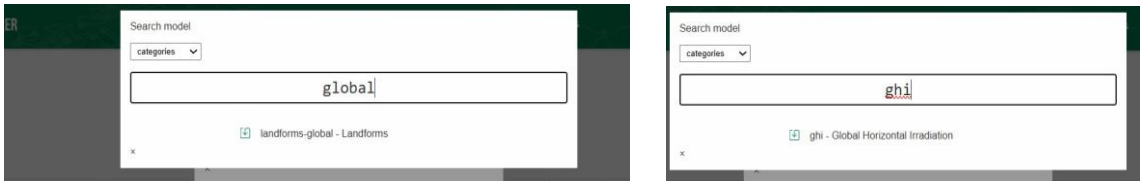


Figure 2: Screenshot shows how categories are added to the data. For example, the long name is Global Horizontal Irradiance, but the short name is GHI. The tool searches using the short name (GHI), not the long name.

ENERGY ACCESS EXPLORER

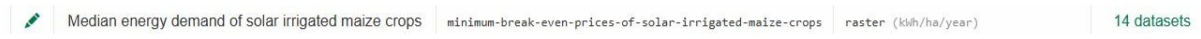


Figure 3: Long name and short name are not matching. It needs to be intuitive for the users, since short name is used to add new data

☐ Define the unit. Please include the units, unless the category is unitless.

☐ Keep circle as 'Everywhere'. This will reduce the duplication if users are using one category across geographies.

☐ Mention and link mutant clearly while naming the category. This will help understand the linkages between two categories.



Figure 4: mutant should be clearly mentioned so that the link between categories is understandable

☐ Add clear and concise descriptions of the category. This description is also visible in the front end below the data toggle button. Avoid adding very long description. Users can also override description for each geography.



Figure 5: Screenshot showing description of the category. If the description is very long, it might take more space.

☐ Colours and colour stops should be intuitive and easy to interpret. For example: 'Continuous energy supply' should not be shown in 'red colour'.

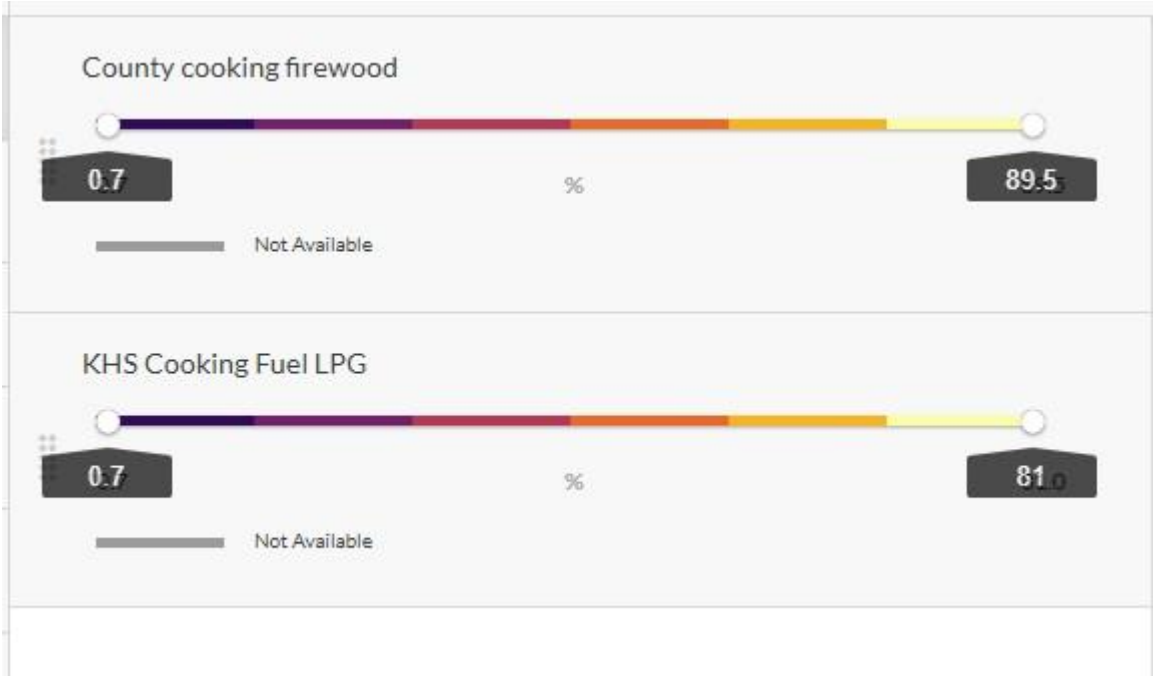


Figure 6: Dataset on LPG cooking and firewood cooking utilizes same colour palette, which can be changed to show user where the problem is

Please see the table 2 below which aligns with the cartographic conventions.

Table 2: Data type and colouring style

Data Type	Includes	Recommended Style	Example (Conventional Colours)
Point	Location of schools, cold storage,	Use simple and intuitive colours. Each facility colour should be unique, so the	Powerplant – black circle, industries - brown
	industries, minigrids, substations	user can diierentiate them visually.	
Line	Transmission lines, rivers or road network	Use contrasting line colours with suitable thickness. Follow conventions.	Rivers: Blue colour
Polygon	Administrative boundary, census or timeline data at admin level, Protected areas, Settlements	Intuitive colouring for the users to understand if the data is negative or positive.	For ex: Institutional delivery follows Red → Green colour, while Maternal death follows Green →Red colour. This allows users to understand that high isn't always good.

ENERGY ACCESS EXPLORER

Raster: Discrete (Categorical)	Landuse Landcover (LULC), soil type	Follow cartographic norms. Use distinct, easily distinguishable colours (no gradients).	For ex: LULC with 9 categories will have following colours. Water = Blue (#1A5BAB); Forest = Dark Green (#358221); Agricultural land = Yellow(#FFDB5C); Grassland = light green(#a8eba2); Shrubland = brownish green(#6a755e); barren land = brown(#3d221c); snow = white(#F2FAFF); built-up area = red(#e01f19); wetlands = blue-green/teal(#008080)
Raster: Continuous	Elevation, Global Horizontal Irradiance, Wind Speed	Follow cartographic norms. Use sequential or diverging gradients that reflect natural progressions.	GHI = Blue-> Green-> Light Yellow → Orange → Dark Red;

☐Have clarity about the indices. This will define how a category contributes to MCA and the rules it follows like does the data need to be included (inclusion buier). It is to understand if the data will contribut to energy supply or demand.

☐Directors to check if a category is linked to mutant, if yes → add the linkage, if no→ if it doesn't have data, delete the category.

	District Hospital Geolocation	health-dhstat	points	0 datasets	2022-05-13 (Santiago.Sinclair-Lecaros)
	Energy Demand	energy-demand	raster (MW)	0 datasets	2025-07-30 (Akansha.Saklani)
	Energy Meters	energy-meters	points (distance in km)	0 datasets	2025-07-28 (Akansha.Saklani)
	Extreme Heat Days (>35C), 2040-2070 Projections	extreme-heat-days-35c-2040-2070-projections	mutant	0 datasets	2023-11-15 (Jake.Stockman)
	Extreme Heat Days (>35C), 2040-2070, RCP 2.6	extreme-heat-days-35c-2040-2070-rcp-26	raster	0 datasets	2023-11-16 (Jake.Stockman)
	Extreme Heat Days (>35C), 2040-2070, RCP 4.5	extreme-heat-days-35c-2040-2070-rcp-45	raster	0 datasets	2023-11-16 (Jake.Stockman)
	Extreme Heat Days (>35C), 2040-2070, RCP 8.5	extreme-heat-days-35c-2040-2070-rcp-85	raster	0 datasets	2023-11-16 (Jake.Stockman)
	Farmland Crops	farmland-crops	raster	0 datasets	2025-04-23 (frederic.famba)
	Grid Connected Solar Plant	grid-connected-solar-plant	polygons-timeline (ksp)	0 datasets	2024-03-01 (shikha.anand)
	Groundwater	groundwater		0 datasets	2022-02-07
	HREA Consumption	hrea-consumption	raster	0 datasets	2023-10-25 (Jake.Stockman)
	HREA Electrification Status	hrea-electrification-status	raster	0 datasets	2023-10-25 (Jake.Stockman)
	HREA Reliability	hrea-reliability	raster	0 datasets	2023-10-25 (Jake.Stockman)
	Health Facilities	health-facilities	points	0 datasets	2023-04-25 (abdul.khalid)
	Heat Index	heat-index	raster	0 datasets	2025-06-06 (shikha.anand)
	Land Cover	land-cover	raster	0 datasets	2024-05-07 (Santiago.Sinclair-Lecaros)
	Land Use Land Cover	land-use-land-cover	raster	0 datasets	2023-05-17 (abdul.khalid)
	Land Use Land Cover 2	land-use-land-cover-2	raster	0 datasets	2024-02-03 (abdul.khalid)
	Land Use and Land Cover	lulc	raster	0 datasets	2022-05-13 (Santiago.Sinclair-Lecaros)
	Landforms	landforms-global	raster	0 datasets	2024-01-24 (Jake.Stockman)
	Landslide Susceptibility	landslide-susceptibility	raster-valued	0 datasets	2024-09-03 (abdul.khalid)
	Least Cost Electrification Scenarios	least-cost-electrification-scenarios	mutant (Technology)	0 datasets	2025-07-31 (alemayehu.agizew)
	Length of Growing Period (Days, Total)	length-of-growing-period-days-total	raster	0 datasets	2023-11-16 (Jake.Stockman)
	Livelihood Zones	livelihood-zones	polygons	0 datasets	2024-05-22 (alemayehu.agizew)

Figure 7: Categories where there is no data; Directors to decide whether to keep them or delete

3.3.2 Dataset

☐Is the name intuitive for the user to understand? Avoid names which doesn't communicate the meaning of the data.

	Name
	all-chc-4dist
	all-chc-6dist
	all-chc-solar-jh
	all-e-j
	avg-energy-supply-sod

Figure 8: Name of diZerent dataset available in the CMS. It's very diZicult to understand what does the data mean or why its added for a particular geography.

☐Is the Long name enough for the user to understand? Please note: Long names are visible in the front end.

☐Are the attributes mapped for the pop-up in the front end?

☐ Is properties_search (only enable for vector data) enabled for the users to search for the data?

☐ Are the colours intuitive?

☐ Are the colours following the EAE guidelines/norms?

☐Is the metadata complete? please see 'Metadata' for details on how to ensure accuracy and completeness.

3.3.3 Metadata

Metadata is "data about data". It describes the content, source, and context of a dataset. In EAE, metadata is essential for clarity, interoperability, trust, and consistent use across geographies. ***Please note: No dataset should be added to production (public) or protection without complete metadata.*** It is acceptable for a dataset to remain 'private' without metadata, but in protected environment many partners use password-protected versions of their data.

☐ Is 'description' inherited from 'category' enough for the data? In case of edits, please override it. This should explain 'why this dataset is being used' ☐ Is the description included in the data?

☐ Are the dataset methodology and processing by data provider covered? ☐ Is normalization or resampling methodology covered for the dataset? Please see the documentation on these steps.

☐ Ensure suggested citations include the citations provided by the data provider and 'Accessed Through Energy Access Explorer, [date] www.energyaccessexplorer.org.' ☐ Are the cautions updated and relevant to the dataset?

☐Have you included cautions provided by the data provider?

☐Have you included accuracy and cautions after data preprocessing?

☐Check if the dataset has disaggregated categories.

☐Add a clear caution message to prevent double counting. For example, "*Gender and age categories are disaggregates of total population. Do not combine total population density with gender-based layers in the same analysis, as this will lead to duplicate counts*"

☐Test in summary reports to confirm no duplication occurs.

☐Have you included the spatial resolution of the data?

☐In case of publicly available data, have you provided source link to download the data?

☐In case of publicly available data, have you provided link for the users to learn more about the data?

☐Have you included data license?

☐Have you included source of the data?

☐Have you included content date?

3.3.4 Data access

☐ Does the data follow any other dataset?

☐ Check if the dataset follows governance rules: Who can read, edit, or publish it? Pay special attention if the data is under an NDA or MoU with limited permissions ☐Confirm dataset complies with governance rules.

- ☐Identify ownership clearly.
- ☐Define who can read, edit, and publish.
- ☐Flag datasets under NDA or MoU.
- ☐Review governance status regularly.

☐Conduct a quarterly data access review.

- ☐Are the permissions up to date?
- ☐Confirm that the data is not published accidentally.

3.3.5 Data visualization

Data visualization is to recheck that the data in frontend is following the EAE standards and guidelines.

☐Ensure title, legend, and unit are clear.

☐Colours follow intuitive/cartographic standards.

☐Set legend ranges close to the dataset's min–max values so users can clearly see diierences (e.g., GHI 1350–1950, not 500–2500)

☐Choose intervals that are intuitive and distributed (e.g., equal intervals, Jenks interval or quantiles), so diierences across classes are easy to interpret.

☐Symbology is standardized across geographies.

☐Maps are readable with the diierent base layers.

☐Check pop-up function is working

☐Check you can search the data, if the feature is activated in the CMS.

3.3.6 Geographies

EAE has multiple geographies (administrative units). This checklist is to ensure that the geographies follow common standards, so the governance is manageable.

☐Ensure oiicial name has been given. If there are multiple instances of the geography, add 'training', 'test' with the name so that admins can diierentiate.

☐Ensure the circles are mentioned correctly.

☐For directors: Test circle access is correct. For instance, users can only access the geography they have the right to.

☐Ensure the instance is deployed correctly. For ex: Nagaland-Training should be accessible only in 'training environment' and not in 'protected environment'. ☐For directors: Remove instances which are not in use.

3.4 Monitoring and Maintenance

Monitoring and maintenance ensure that datasets, metadata, and visualizations in EAE remain accurate, up-to-date, and compliant with governance rules. Regular checks, at a fixed interval, help sustain trust, usability, and consistency across geographies.

This exercise should be integrated into working group meetings. At the agreed frequency, working groups should dedicate time to reviewing dataset updates, verifying alignment with identified needs, and ensuring compliance with licensing requirements.

Datasets are updated as per schedule.
☐ Metadata is revised as the data is updated.
☐ Access permissions reviewed quarterly.
☐ Data loading & visualization tested.
☐Monitor data usage (via hotjar or google analytics).
☐ User feedback reviewed.
☐ QA check performed after updates.
☐ Old datasets archived properly.
☐ NDA/MoU restrictions verified.
☐ Half-yearly monitoring report (Appendix 1) logged.

3.5 Archiving and Deprecation

Archiving and deprecation are the processes of marking outdated datasets as obsolete, moving them out of active use while keeping them stored for reference. In EAE, this is needed to avoid duplication, prevent misuse of old data, and ensure users always work with the most current and reliable datasets.

☐ Has the dataset been identified as outdated or superseded? For example: population density of 2015 can be superseded by population density of 2020 or 2025 (if available). ☐ Has the dataset and its metadata been updated to include a reference to the newer version?
☐ Has the dataset been moved to OneDrive archive repository (not deleted)?
☐ Review archives half-yearly to see if any dataset can be permanently deleted.
☐ In case of timeline dataset, check for how many years you want data to be available? For example: Is data from 2012 still valid in 2025
☐ Maintain archived datasets (or methodology) for reproducibility

4. Summary

To summarize, this document outlines the diierent stages of the data lifecycle and provides a checklist for each step. Table 3: includes the data life cycle and along with the roles and responsibility. By following it, you can ensure the following:

☐ Is the data of good quality (complete, accurate, reliable)?

☐ Is the data represented correctly? For ex: Water is shown as blue and not green; if the district boundaries are changing, we are not splitting the data by ourselves. Like year 2020 there are 27 districts, we can't disaggregate the data to 33 districts if we don't have granular data. It's simple: 5 can be 0+5, 5+0, 1+4, 4+1, 2+3, 3+2.

☐ Are names, data types, and colours consistent across geographies?

☐ Are attribute fields using unified definitions?

☐ Attribute Integrity: Are attribute values valid, non-duplicated, and correctly linked?

☐ Are units standardized and documented?

☐ Semantic Consistency: Are terms, categories, and meanings consistent across datasets and geographies?

☐ Is metadata complete and clear?

☐ Are data governance rules followed (access rights, NDA/MoU compliance)?

Table 3: Data lifecycle summary with roles and responsibilities

Stage	Key checks	Outputs	Frequency	Reviewer (Data QA/QC + product owner)	Geography Admin**	EAE team
Data collection	Verify source, license, quality and relevance	Raw dataset + any notes	On data acquisition	Spot check if needed	Responsible	Informed/Consulted
Data Transformation	Cleaning, resampling, normalization , reprojection, derivation of new indicators. Ensure transformations are documented.	Preprocessed dataset + processing log.	On ingestion	Random spot-checks	Responsible	Informed/Consulted

Metadata Creation	Title, short name, keywords, CRS, temporal coverage, owner, license.	Metadata record	On upload	Consulted/random review	responsible	Informed/consulted
Data Governance	Access rights (read/write/publish), NDA/MoU compliance, audit log.	Access log + permissions with the support of Working groups.	Halfyearly	Consulted/ random 4–5 datasets and reviewed by working group	responsible	Informed, by product owner and data QA/QC lead
Data Visualization	Intuitive colours, symbology, legends, intervals, accessibility.	Publishe d layer	On publish	Spot checks	Responsible	Informed
Monitoring & Maintenance	QA check, metadata update, permissions review, monitoring lo	Half yearly review report	Halfyearly	Spot checks	responsible	Informed, by product owner and data QA/QC lead
Archiving & Deprecation	Mark deprecated, archive dataset, link new version, add caution note, retention policy.	Archived dataset + caution message.	As needed + half yearly review	Spot checks	responsible	Informed, by product owner and data QA/QC lead

** Geography administrators must engage and work with the EAE working group across all data lifecycle stages.

ENERGY ACCESS EXPLORER

Appendix 1: Data monitoring report format

It is not feasible to review every dataset individually. Instead, a designated person should randomly review 4–5 datasets across diierent categories and geographies, while the geography administrator remains responsible for ensuring that all standards are upheld

Field	Details
Name and long name	
Geography	
Category (including mutant layer)	
Data type (Raster/vector)	
Source	
Last updated on	
Next scheduled update	
Metadata verified (Y/N)	
QA check completed (Y/N)	
Access permissions reviewed (Y/N)	
Governance restrictions reviewed (Y/N)	
Notes/ Issues found	
Reviewed by	
Date of Review	

[i] Decentralized adoption of tools is the organic, permissionless integration of technologies driven by their value proposition without relying on a central governing authority. [ii] Will be replaced by ISO/AWI 19115-1, which is under development [iii] If the dataset is not complete – please add it as a caution in metadata

[iv] Categorical data refers to data that represent discrete classes or categories, rather than continuous numerical values. These categories are usually qualitative in nature and are often represented in thematic maps. For example: land use land cover or soil types.

[v] Continuous data represents values that change smoothly over space and can take on an infinite range of values within a given interval. For ex: GHI, temperature. [vi] Here track ties mean, that we have 50% urban and 50% forest. While majority will give value to the first occurrence, mode will make it ambiguous or highlight the user to choose i.e. breaking the tie.