



FAIR² CERTIFIED™
Explore interactive data here:
<https://doi.org/10.71728/senscience.k2f7-p5v9>

OPEN ACCESS

EDITED BY

Vikas K. Sangal,
Malaviya National Institute of Technology,
Jaipur, India

REVIEWED BY

Freddy Zambrano Gavilanes,
Technical University of Manabi, Ecuador
Ayodeji Olawore,
Kwara State University, Nigeria

*CORRESPONDENCE

Kok Sin Woon,
✉ koksinwoon@hkust-gz.edu.cn

RECEIVED 03 October 2025

REVISED 23 December 2025

ACCEPTED 24 December 2025

PUBLISHED 27 January 2026

CITATION

Hoy ZX, Woon KS, Chin WC, Fan YV and Yoo SJ
(2026) Global waste sector dataset
(1990–2050): scenario-based projections of
generation, emissions, and
socioeconomic drivers.
Front. Environ. Sci. 13:1717992.
doi: 10.3389/fenvs.2025.1717992

COPYRIGHT

© 2026 Hoy, Woon, Chin, Fan and Yoo. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the copyright
owner(s) are credited and that the original
publication in this journal is cited, in accordance
with accepted academic practice. No use,
distribution or reproduction is permitted which
does not comply with these terms.

Global waste sector dataset (1990–2050): scenario-based projections of generation, emissions, and socioeconomic drivers

Zheng Xuan Hoy^{1,2}, Kok Sin Woon^{3,4*}, Wen Cheong Chin⁵,
Yee Van Fan⁶ and Seung Jick Yoo⁷

¹Energy Studies Institute, National University of Singapore, Singapore, Singapore, ²Institute for Global Environmental Strategies, Hayama, Japan, ³Carbon Neutrality and Climate Change Thrust, Society Hub, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China, ⁴Department of Energy and Chemical Engineering, Xiamen University Malaysia, Bandar Sunsuria, Malaysia, ⁵Department of Mathematics, Xiamen University Malaysia, Bandar Sunsuria, Malaysia, ⁶Environmental Change Institute, University of Oxford, Oxford, United Kingdom, ⁷Department of Climate and Environmental Studies, Sookmyung Women's University, Seoul, Republic of Korea

Strategy planning for global climate goals requires structured, multisectoral data linking environmental pressures with socioeconomic drivers across time and geography. However, internationally harmonized, machine-actionable datasets integrating waste generation, waste-related greenhouse gas (GHG) emissions, and socioeconomic indicators remain scarce. This study provides a harmonized, AI-ready dataset to support global analyses of municipal solid waste (MSW) and associated emissions. This FAIR² dataset provides historical (1990–2020) and forecasted (2021–2050) national-level data for 43 countries, covering MSW generation, CO₂, CH₄, and N₂O emissions, GDP *per capita* (PPP), and population. Forecasts were generated using an ensemble of fixed-effects regression models and artificial neural networks informed by economic and demographic trends. By linking MSW, emissions, and socioeconomic drivers within a standardized structure, the dataset enables analyses including benchmarking, equity assessments, and decoupling analysis. While limited to national aggregates and subject to scenario uncertainty, the dataset complies with FAIR² principles, supporting reuse and traceability.

KEYWORDS

AI-ready, climate data, GDP (PPP), greenhouse gas emissions, municipal solid waste (MSW), national-scale, planetary boundaries, population

1 Introduction

In the pursuit of sustainable development and climate resilience, it is increasingly necessary to quantify national-level environmental pressures in a manner that is temporally explicit, geographically comparable, and analytically reusable. GHG emissions, resource consumption, and economic output are core metrics used to track progress toward global targets such as the Paris Agreement and the Sustainable Development Goals (SDGs). Several prior studies have applied data-driven or AI-based approaches for municipal solid waste (MSW) estimation and forecasting, including city-scale case studies based on administrative data (Abbasi and El Hanandeh, 2016), cross-sectional analyses combining socioeconomic

indicators with national statistics (Olawore et al., 2024), and large compilations of MSW-related records assembled from international databases and the literature (He et al., 2022). While these studies provide valuable insights into MSW generation dynamics and modeling approaches, they typically rely on limited temporal coverage per country, heterogeneous spatial units (cities vs. countries), or datasets assembled for specific modeling tasks rather than long term and harmonized reuse.

In parallel, major international repositories provide valuable but fragmented inputs for global waste and emissions analyses. For example, the World Bank, OECD, and United Nations statistical databases offer extensive coverage of economic, demographic, and selected waste indicators, yet waste generation, treatment pathways, and waste-sector-specific GHG emissions are rarely integrated within a single and consistent time-series structure. UNFCCC National Communications and Biennial Update Reports report emissions inventories designed primarily for compliance reporting, with substantial variation in methodological detail and limited linkage to socioeconomic drivers. As a result, integrated analyses of waste generation, associated emissions, and economic or population dynamics typically require substantial manual harmonization across multiple sources, constraining reproducibility, scalability, and downstream data-driven applications.

This dataset addresses these challenges by compiling historical (1990–2020) and forecasted (2021–2050) data for key national-level indicators: population, GDP *per capita* (purchasing power parity [PPP]-adjusted), MSW generation, and associated emissions of CO₂, CH₄, and N₂O. Population and GDP *per capita* capture socioeconomic drivers of consumption, while MSW generation and associated GHG emissions quantify resulting environmental pressures, enabling integrated analyses of national-level sustainability performance. The historical data were obtained from authoritative public sources and merged with projected values derived from model-based estimates. Data harmonization and exploratory validation were carried out to ensure consistency across countries and years. The dataset is structured according to FAIR² principles, with standardized variables, comprehensive metadata, and formats suited for AI-readiness and reproducibility.

The structured design enables researchers to conduct time-series and cross-sectional analyses of environmental burden, decoupling trends, and per-capita or per-GDP emissions efficiency. It is particularly well-suited for integration with planetary boundary-based life cycle assessments (PB-LCA), sustainability benchmarking, and data-driven modeling of climate policy impacts. While it does not resolve all limitations inherent to national-scale environmental datasets, it provides a transparent and reusable foundation for further research in climate science, sustainability assessment, and responsible AI development.

2 Methods summary

2.1 Study design

This dataset provides annual estimates of MSW generation and associated GHG emissions (CO₂, CH₄, and N₂O) for 43 countries from 1990 to 2050 (see Appendix A for the full list of countries).

Throughout this paper, the term ‘country’ refers to any territory reporting separate social or economic statistics, as per the World Bank classification. This usage does not imply political independence. The methodology combines bottom-up empirical data collection with econometric modeling and neural network-based forecasting. The historical component (1990–2020) draws on government-reported and internationally maintained databases (e.g., World Bank, OECD, Eurostat), while future projections (2021–2050) follow the Shared Socioeconomic Pathways (SSP) from the SSP Database (Version 2.0 December 2018).

2.2 Historical data compilation (1990–2020)

A comprehensive historical dataset was compiled for each of the 43 countries, covering the years 1990–2020. Data sources included the World Bank’s “What a Waste 2.0” (World Bank, 2018), national environmental reports, and submissions to the United Nations Framework Convention on Climate Change (BURs and NCs). For each year and country, total MSW generation (in tonnes) was recorded where available, or subsequently reconstructed using a model-based approach described in Section 2.3. Additional macroeconomic variables—namely total population and GDP *per capita* (PPP)—were obtained from the World Bank World Development Indicators (World Bank, 2022).

MSW composition was harmonized across nine material types: food, garden, paper, plastic, glass, metal, rubber and leather, textile, and other. Treatment types were categorized into six pathways: recycling, composting, incineration, managed landfill, uncategorized landfill, and open dump.

The compiled historical dataset represents an unbalanced panel, reflecting the uneven availability of reported MSW statistics across countries and over years. In total, 545 country-year observations across 43 countries from 1990 to 2020 are based on directly reported MSW generation data. For percentage-based variables (specifically MSW composition and treatment pathway shares), interpolation was applied to fill internal gaps where partial time series existed. These interpolations followed fixed assumptions: constant values before the first available year, linear interpolation between reported observations, and stabilization after the last observed year.

Emissions were calculated for each GHG and treatment pathway based on the IPCC methodologies (IPCC, 2006; IPCC, 2019). CH₄ emissions from landfills were calculated using the IPCC Tier 2 First Order Decay method, which requires estimates of waste degradation rates and methane recovery. Country-specific parameters were applied whenever available; otherwise, global default values were used. Incineration and open burning emissions were calculated based on CH₄, CO₂, and N₂O emissions, incorporating dry matter, fossil carbon content, and oxidation efficiency using Tier 2a parameters. Composting emissions included only CH₄ and N₂O, using Tier 1 emission factors. For recycling, gross CO₂ emissions were estimated using literature-based life cycle emission factors applied to the volume of recyclable materials treated. These estimates produced an annual record of GHG emissions by type of gas and country for 1990–2020 (Turner et al., 2015).

2.3 Panel regression modeling of MSW generation

To address missing historical data and enable forward projection, a panel data regression framework was adopted to exploit both cross-country and temporal variation in MSW generation. This approach follows the general strategy adopted in World Bank (2018), where MSW generation *per capita* is estimated as a function of GDP *per capita* using a single observation per country. However, whereas World Bank (2018) relies on cross-sectional regressions, this study extends the methodology by exploiting a panel structure with multiple years per country and explicitly controlling for unobserved country- and time-specific effects. Panel regression is particularly suitable for this application because it allows control for unobserved, time-invariant country characteristics (e.g., consumption patterns, institutional structures, and waste management practices) that are not directly observable but systematically affect waste generation. A pooled ordinary least squares (OLS) approach was therefore considered inappropriate, as it would ignore this unobserved heterogeneity and risk biased coefficient estimates.

A fixed-effects specification was selected over a random-effects alternative based on formal diagnostic testing. The Hausman test rejected the null hypothesis that country-specific effects are uncorrelated with GDP *per capita* (PPP), indicating that the random-effects estimator would be inconsistent, while a heteroscedasticity likelihood-ratio test rejected the homoscedasticity assumption. The model was specified as:

$$y_{i,t} = e^{0.395 \ln x_{i,t} + 2.029 + a_i + a_t}$$

where $y_{i,t}$ is the MSW generation *per capita* in country i and year t , $x_{i,t}$ is the GDP *per capita* (PPP), and a_i and a_t are country and year fixed effects, respectively. Calibration included 545 observations across 43 countries and 31 years, and the model exhibited high explanatory power ($R^2 = 0.947$), supporting its use for reconstructing and projecting MSW generation trajectories.

For each missing or interpolated MSW data point, the predicted *per capita* value was adjusted by scaling it to the actual base-year observation using a ratio of modeled values. This anchoring approach preserved the relative growth dynamics modeled by regression while maintaining empirical fidelity to known data points (Hoy et al., 2023).

2.4 Forecasting population and GDP (2021–2050)

Future values for population and GDP *per capita* (PPP) were extracted from the SSP Database, which provides global socioeconomic trajectories through 2100. These projections were available at 5-year intervals and required annual interpolation to match the temporal resolution of the MSW model. To ensure continuity between historical observations and future projections, a proportional adjustment was applied at the country level by aligning SSP values to the observed World Bank data in the base year (2020).

Specifically, for each country i and variable v (population or GDP *per capita*), the adjusted projected value $\tilde{X}_{i,t}^v$ was calculated as:

$$\tilde{X}_{i,t}^v = X_{i,t}^{v,SSP} \times \frac{X_{i,2020}^{v,WB}}{X_{i,2020}^{v,SSP}}$$

where $X_{i,t}^{v,SSP}$ is the original SSP-projected value for year t , and $X_{i,2020}^{v,WB}$ and $X_{i,2020}^{v,SSP}$ denote the observed World Bank and SSP values in the base year, respectively. This adjustment preserves relative growth trajectories from the SSP scenarios while ensuring consistency with observed historical data at the transition point.

Linear interpolation filled the gaps between quinquennial projections, producing annual time series for each country from 2021 to 2050. These time series were then fed into the fixed-effects regression model described in Section 3 to derive projected *per capita* MSW generation. The resulting *per capita* values were multiplied by the projected population to estimate total MSW generation for each year. Where needed, these values were further scaled using base-year anchoring to ensure comparability with the historical dataset.

2.5 Estimating and forecasting GHG emissions

Historical emissions (1990–2020) were computed using IPCC-approved methods tailored to each treatment type and gas, as detailed in Section 2. For future emissions (2021–2050), a machine learning approach was adopted. Three separate artificial neural network (ANN) models—one for each gas (CO₂, CH₄, N₂O)—were developed for each of the 43 countries. This machine learning framework follows the approach developed and validated in Hoy et al. (2023).

Each ANN used three inputs: annual population, GDP *per capita* (PPP), and MSW generation. The output was the corresponding GHG emission value for that year. Historical data were partitioned into training (70%), validation (15%), and testing (15%) sets. Bayesian optimization was used to tune the number of neurons and the learning rate. For each gas-country pair, the model was trained 10 times with different random seeds, forming a 10-member ensemble. The median of the ensemble was used as the central forecast, which is included in the final dataset. Note that formal uncertainty intervals (e.g., 5th–95th percentiles) are not provided; users should consider that the projections reflect median values only, which may limit interpretation for policy-relevant assessments. These forecasts extended the emissions dataset through 2050, preserving consistency with observed historical trends and model input structure.

2.6 Emissions normalization and budget alignment

Annual emissions were converted to a common metric of climate impact using carbon dioxide warming equivalents (CO₂-we). This was necessary to enable comparison against global emissions budgets associated with limiting warming to 1.5 °C or 2.0 °C. CO₂ and N₂O were converted using GWP-100 factors from the IPCC Sixth Assessment Report (IPCC, 2021). CH₄ emissions were transformed using the GWP* method, which incorporates both the current emission level and its 20-year rate of change, accounting for its short-lived climate effect.

For each country-year, the resulting CO₂-we values were summed annually to produce cumulative emissions curves. These were compared against scaled global budgets for the MSW sector (1.7% of total), using thresholds aligned with 50% and 67% likelihoods of achieving climate targets (Clark et al., 2020). This normalization enabled the quantification of whether current and projected emissions remained within safe operating space for the sector.

2.7 Dataset assembly and output structuring

All country-year records were merged into a single harmonized dataset, with complete coverage from 1990 to 2050. Columns included population, GDP *per capita* (PPP), MSW generation, and associated emissions of CO₂, CH₄, and N₂O in tonnes of carbon dioxide-equivalent (t CO₂-eq). Forecasts were reported as ensemble medians. A structured data dictionary was prepared to annotate each variable with units, data types, and provenance steps. The resulting dataset was formatted in tabular structure and prepared for FAIR²-aligned reuse.

3 Data overview

3.1 Data summary

The dataset provides harmonized, time-series records for key national-level sustainability indicators, covering both historical (1990–2020) and forecasted (2021–2050) periods. It integrates demographic, economic, waste generation, and associated GHG emissions data for multiple countries and is structured to support environmental performance assessments at national scales. The dataset contains annual values for:

- Population (number of people)
- GDP *per capita* (PPP, current international \$)
- MSW generation (t/year)
- GHG emissions from MSW activities:
 - CO₂ emissions (t CO₂-eq/year)
 - CH₄ emissions (t CO₂-eq/year)
 - CH₄ emissions (t CO₂-we/year), derived using GWP* methodology to reflect short-lived climate impacts
 - N₂O emissions (t CO₂-eq/year)

Each row in the dataset corresponds to a country-year combination. The data are compiled from harmonized sources and forecast models to ensure longitudinal consistency and machine-actionable formatting. The dataset is organized with standardized variable names and is intended to be FAIR²-compliant, enabling integration into sustainability benchmarking frameworks, *per capita* burden assessments, and planetary boundary downscaling.

Data validation was conducted through exploratory statistical analysis, revealing strong correlations between population and emissions, and highlighting distinct emission intensities across countries. Country-level trends in GHG emissions and MSW generation reflect both socioeconomic development trajectories and policy effects, enabling comparative analysis of sustainability transitions.

3.2 Quantitative summary of the dataset

This section describes the dataset's composition, structure prediction metrics, feature annotations, model performance, and associated computational costs.

3.2.1 Dataset composition

The dataset encompasses:

- Temporal coverage: 1990 to 2050, with seamless integration of historical and forecasted records.
- Geographic coverage: 43 countries across various income and development levels (see Appendix A for the full list of countries).
- Total number of records: Approximately 2,440 entries (61 years × 43 countries).
- Variables: 9 key variables (including country and year).
- Missing data: The dataset contains values for all variables across all country-year combinations. Some entries are interpolated (for historical gaps) or model-derived (for future projections), and thus not all values are directly observed.

3.2.2 Descriptive statistics (across all years and countries)

- Population: Ranges from 1.5×10^7 to over 1.9×10^9 people.
- GDP (PPP) *per capita*: Spans from approximately 340 to over 109,000 USD PPP.
- MSW generation: Ranges from 1.9×10^6 to 6.5×10^8 t/year.
- CO₂ emissions: Varies from 1.9×10^3 to over 1.6×10^8 t CO₂-eq/year.
- CH₄ emissions (using GWP-100): Extends from 5.8×10^5 to 1.2×10^8 t CO₂-eq/year.
- CH₄ emissions (using GWP*): Ranges from -2.9×10^7 to 2.6×10^8 t CO₂-we/year.
- N₂O emissions: Between 55 and 5.1×10^6 t CO₂-eq/year.


Observed trends across countries indicate differentiated development pathways, with some exhibiting steady growth in emissions and GDP, while others demonstrate plateauing or declining intensities relative to population or output. These patterns support a broad range of applications, including sustainability performance comparisons, decoupling diagnostics, and emissions equity assessments.

The dataset's structure facilitates disaggregation and normalization *per capita* or per unit GDP, enabling cross-country comparisons and integration with planetary boundary-based allocation models. Forecasted values maintain consistency with historical trends and are aligned for use in long-range scenario analyses.

3.3 FAIR² compliance certification

The dataset supporting the findings of this study is available through a FAIR² Data Portal (<https://doi.org/10.71728/senscience.k2f7-p5v9>), which ensures that the data adhere to the principles of Findability, Accessibility, Interoperability, and Reusability (FAIR), with additional emphasis on including detailed Contextual metadata and AI-Readiness and Responsible AI practices (see Table 1). All raw data, metadata, and supplementary materials, including detailed protocols

TABLE 1 The FAIR2 Compliance Certification presented here was generated through a Human-in-the-Loop (HITL) process combining automated FAIR2 system checks with author-supplied inputs. While certain metadata fields and validations (e.g., DOI registration, schema adherence, file accessibility) are verified automatically by the FAIR2 platform, other elements—such as domain-specific documentation quality and Responsible AI considerations—reflect expert curation by the dataset authors.

Criteria	Assessment
	
Findability (F)	
F1. Unique identifier	The dataset is assigned a DOI: https://doi.org/10.71728/senscience.k2f7-p5v9 , ensuring persistent and citable global identification
F2. Metadata	Metadata includes the dataset title, author affiliations, country-level scope, temporal range (1990–2050), variable definitions, units, and data generation methods. It also specifies the modeling assumptions used in forecasted indicators. Metadata could be enhanced by including variable-level uncertainty estimates
F3. Metadata includes data identifiers	The metadata explicitly includes the dataset DOI and links to all versioned releases, enabling traceability between data files and their descriptors
F4. Searchable metadata	Metadata uses controlled vocabularies, including QUDT for units, ISO country codes for geography, and ISO 8601 for dates, supporting structured queries and semantic search. The dataset is described using the dataset type from schema.org , enabling enhanced discoverability through google dataset search and other rich metadata aggregators
Indexed in repositories	The dataset is indexed in the FAIR ² data portal and zenodo, and is discoverable through DataCite, supporting both domain-specific and cross-domain data discovery
Accessibility (A)	
A1. Open access	The dataset is openly accessible without registration or restriction, supporting open science practices
A2. Long-term access	Archival redundancy is provided through versioned releases on zenodo and institutional backups within the FAIR ² infrastructure. Each version is independently citable
API access	Programmatic access is provided via the FAIR ² portal, supporting JSON and CSV downloads. Interactive visualizations are also available for exploring country-level trends, correlations, and projections
Interoperability (I)	
I1. Standardized formats	Data is provided in open, machine-readable formats (CSV and JSON). The structure aligns with MLCommons croissant and FAIR ² metadata specifications
I2. Controlled vocabularies	Variables use controlled vocabularies including QUDT (units), ISO 3166-1 (countries), and consistent naming across emissions and economic indicators
I3. Cross-platform integration	The dataset structure supports integration with modeling environments (R, python), as well as environmental and life cycle tools (e.g., SimaPro, OpenLCA)
Reusability (R)	
R1. Comprehensive documentation	Accompanied by a detailed README, data dictionary, and full methods section, including source references and harmonization steps across historical and forecast data
R1.1 . License	Licensed under the open data commons attribution license (ODC-by v1.0), enabling reuse and redistribution with attribution
R1.2. Detailed provenance	Provenance is described at the dataset and variable level, linking raw sources to final outputs, with references to national statistics, environmental databases, and transformation scripts
R1.3. Domain-relevant standards	Dataset structure is aligned with emerging environmental data standards (e.g., UNEP waste indicators, ISO 14040 family). Variables match common sustainability reporting frameworks
Versioning and updates	Version control is implemented through semantic versioning. Each release includes changelogs and retains access to prior versions through DOI-based archival
AI-readiness (AIR)	
Structured for machine learning	The dataset is stored in flat tabular structures with labeled columns and standardized types, enabling direct ingestion into ML pipelines
Scalable	The dataset can be efficiently used in batch processing and cloud environments. It supports scaling to additional geographies or years without structural changes

(Continued on following page)

TABLE 1 (Continued) The FAIR² Compliance Certification presented here was generated through a Human-in-the-Loop (HITL) process combining automated FAIR² system checks with author-supplied inputs. While certain metadata fields and validations (e.g., DOI registration, schema adherence, file accessibility) are verified automatically by the FAIR² platform, other elements—such as domain-specific documentation quality and Responsible AI considerations—reflect expert curation by the dataset authors.

Criteria	Assessment
Training and validation sets	Enables emissions forecasting, decoupling analysis, and cross-country profiling using structured indicators. Suitable for supervised learning and exploratory modeling. Emissions are expressed in CO ₂ -we and aligned with global budgets, supporting integration with LCA models and planetary boundary frameworks
Responsible AI (RAI)	
Ethical standards and misuse	Intended for environmental and sustainability research. Use in unrelated or high-risk domains (e.g., personal finance, healthcare) is discouraged to avoid inappropriate extrapolation
Biases in the dataset	Geographic coverage is uneven, with higher data density in larger economies. Forecasting assumptions may embed socioeconomic biases if applied without scenario context
Data privacy and security	No personal or sensitive data is included. All content is fully aggregated at the national level
Fairness and non-discrimination	No protected or identity-based variables are included. However, interpretations should avoid attributing environmental trends to cultural or political factors without context
Explainability and interpretability	All variables are clearly defined, with units, data sources, and transformations documented in the metadata and data dictionary
Data provenance and accountability	All variables are traceable to primary sources, including national statistics, international agencies, and modeled estimates. Contributors and curators are listed in the metadata, with accountability roles documented
Transparency and reporting	Users are encouraged to cite the dataset DOI. All known transformations and interpolations are described in the metadata, and changelogs accompany each versioned release
AI safety and fairness	Known risks include overfitting to site or injury-specific features. The inclusion of harmonized and raw volumes enables safety testing against dataset-induced model artifacts
Ethical and social impact	The dataset supports global equity research on waste, emissions, and development. It is useful for identifying disparities, but should be used alongside contextual socioeconomic data
Human-in-the-loop (HITL) considerations	Forecast modeling and data harmonization involved manual quality checks. HITL workflows are encouraged when applying the data in policy-critical systems

and methods, are accessible via the FAIR² Data Portal (<https://doi.org/10.71728/senscience.k2f7-p5v9>). The portal also includes interactive visualizations, such as global maps of MSW generation, distributions of country-level MSW, and the relationship between MSW generation and GDP *per capita*. Users can export the full dataset to generate additional customized plots for other indicators as needed.

The dataset has been structured to ensure compliance with FAIR² standards, enabling easy integration with other datasets and promoting reuse in future research. Researchers can access the dataset in multiple formats, and appropriate documentation is provided to facilitate transparency and reproducibility. Variable-level uncertainty descriptors are not included in the dataset; users should be aware that values reflect the reported or modeled estimates without explicit confidence intervals. Any updates or corrections to the dataset will also be managed and tracked through the portal, ensuring long-term accessibility and version control.

3.3.1 Overall FAIR² badge compliance

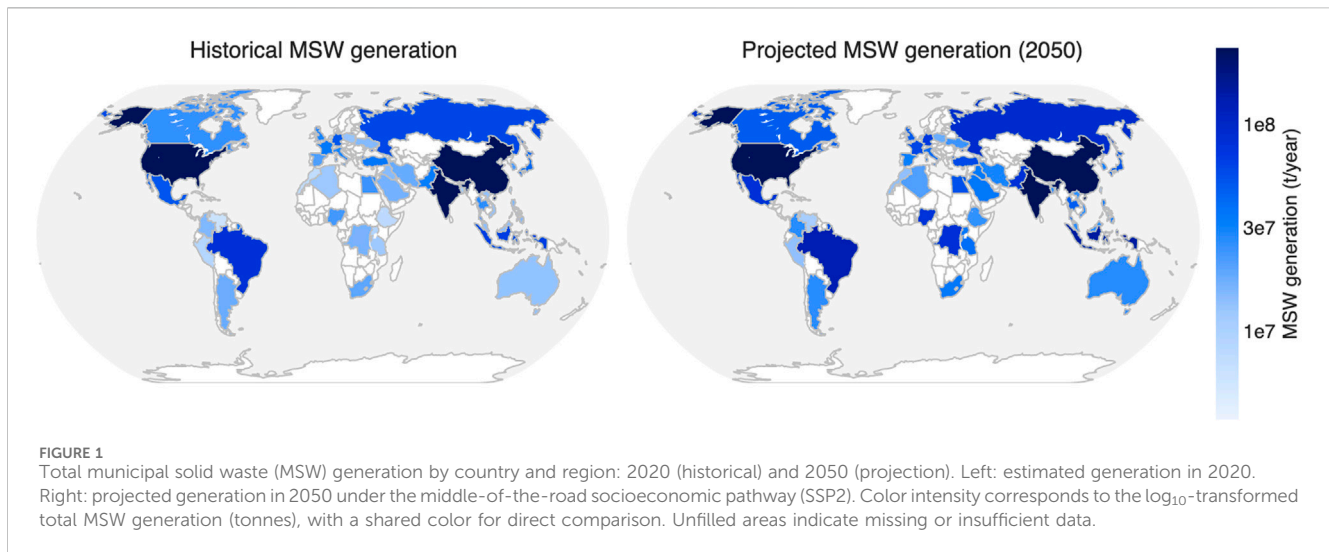
Compliant—The dataset qualifies for the FAIR² Badge, meeting all criteria across Findability, Accessibility, Interoperability, Reusability, AI-Readiness, and Responsible AI. It is assigned a DOI, described using schema.org metadata, and published in open, interoperable formats. The structure supports machine learning workflows, with clearly defined variables, documented provenance, and responsible use guidance. Its national-level

scope and transparent methodology make it a robust resource for sustainability science, environmental modeling, and AI applications in policy and planning.

4 Visual overview

This section provides a visual synthesis of key patterns in the dataset across space and time, focusing on national-level indicators relevant to environmental sustainability. The figures included here highlight trends in emissions contributed by the waste sector, economic activity, waste generation, and population, providing insight into how these factors interact across countries and over historical trajectories.

Figure 1 presents a comparative global overview of total MSW generation by country in 2020 (historical estimate) and in 2050 (projected under SSP2, selected for illustration due to its plausibility and alignment with median development trends (Hoy et al., 2023)). These maps reflect country-level MSW generations in tonnes, scaled using a unified log₁₀ color scale to enhance interpretability across several orders of magnitude. While waste generation correlates with economic activity, the spatial patterns also reveal important regional disparities driven by infrastructure, consumption intensity, and waste management systems. Notably, several rapidly industrializing economies are projected to experience substantial increases in total



MSW volumes by 2050, underscoring the need for investment in sustainable waste systems. Countries shown in gray lacked sufficient data for inclusion or projection.

Figure 2 illustrates the evolution of total MSW generation across ten most populous countries from 1990 through 2050. The figure combines historical estimates with projections based on SSP1–SSP5, representing divergent global development trajectories from sustainability-oriented (SSP1) to fossil fuel-intensive (SSP5). While most countries display steady growth in MSW generation, the magnitude and trajectory differ markedly. India and China show steep increases over time, reflecting both population growth and rising consumption. The United States maintains high absolute levels throughout, despite a slower growth rate in recent decades. In contrast, several countries in sub-Saharan Africa (e.g., Nigeria, Democratic Republic of Congo) and South Asia (e.g., Pakistan) are projected to undergo accelerated MSW growth after 2020, potentially outpacing their historical trends. This visualization highlights the diverse temporal dynamics of waste accumulation and emphasizes the growing burden of waste management infrastructure, particularly in emerging economies.

Figure 3 explores the temporal relationship between national wealth and MSW generation across ten countries with the highest GDP per capita (PPP). GDP per capita (PPP) is plotted against total MSW generation (tonnes/year), with bubble size indicating national population. A clear positive trend is observed for most countries, reflecting the well-established link between economic development and material consumption (IPCC, 2021). However, significant inter-country variation persists: for example, the United States generates consistently high volumes of waste across a broad income spectrum, while countries like India and Indonesia show rapid MSW growth at relatively lower income levels, driven by large and growing populations. The figure captures both absolute and per-capita patterns, illustrating different trajectories of waste generation intensity and the potential for future decoupling of waste growth from GDP.

Figure 4 presents the correlation structure among key demographic, economic, and environmental variables related to MSW systems and their associated GHG emissions. The matrix includes population size, GDP per capita (PPP), total MSW

generation, and national-level emissions of CO₂, CH₄, and N₂O contributed by the waste sector. MSW generation exhibits strong positive correlations with all three GHG emissions, particularly methane (CH₄, $r = 0.86$), underscoring its role as a major driver of short-lived climate pollutants. CO₂ and N₂O emissions also correlate closely ($r = 0.71$), reflecting overlapping energy and waste combustion sources. Interestingly, GDP per capita (PPP) shows a stronger relationship with CO₂ ($r = 0.69$) than with CH₄ or N₂O, suggesting that carbon intensity scales more directly with economic output, while CH₄ and N₂O emissions are more tightly linked to population and total waste. These correlations provide a quantitative foundation for understanding cross-variable interactions and prioritizing mitigation strategies within national waste sectors.

5 Discussion

5.1 The value of the dataset

This dataset provides a structured, longitudinal record of national-scale environmental pressures and socioeconomic drivers spanning both historical and forecasted periods (1990–2050). Its design responds to the increasing need for integrated, AI-ready data resources that link GHG emissions, economic growth, and population dynamics in a harmonized format. By including both historical and projected values, the dataset offers a continuous trajectory of waste-related sustainability indicators relevant to climate modeling, policy benchmarking, and environmental planning.

A key strength of the dataset is its alignment with planetary boundary frameworks, particularly its suitability for downscaling global environmental limits to the national or product level. The variables included—CO₂, CH₄, and N₂O emissions, MSW generation, GDP per capita (PPP), and population—support analyses ranging from emissions equity assessments to decoupling diagnostics and resource efficiency studies. Each variable is traceable to a defined methodological step and supported by metadata and unit standardization (e.g., QUDT-compliant), ensuring interoperability across analytical platforms.

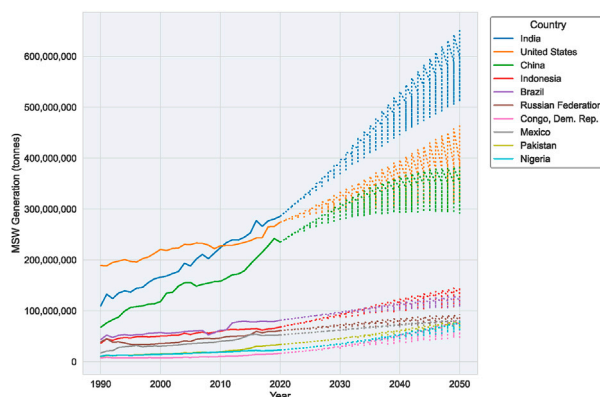


FIGURE 2 Historical and projected trends in total municipal solid waste (MSW) generation for selected high-population countries (1990–2050). Solid lines represent historical estimates of total MSW generation (tonnes per year), while dotted lines show future projections under five Shared Socioeconomic Pathways (SSP1–SSP5). The ten most populous countries in the dataset are displayed to highlight baseline levels, temporal growth, and long-term trajectories.

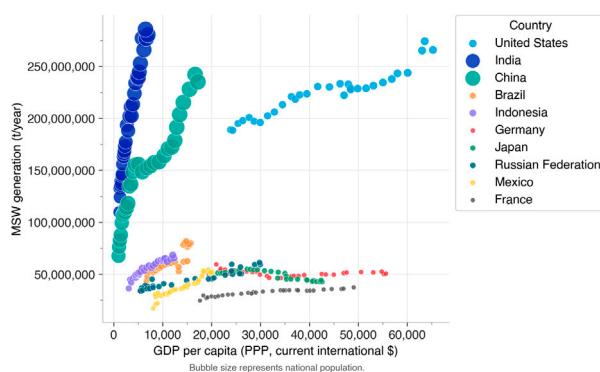


FIGURE 3 Relationship between GDP *per capita* (PPP) and total municipal solid waste (MSW) generation across selected countries. Each point represents a country-year observation for ten major economies, with bubble size proportional to population. The figure highlights both scale effects (population-driven totals) and efficiency patterns (waste generation per unit wealth).

The dataset’s modularity allows it to be extended or linked with other data sources, including life cycle inventories, policy datasets, or regional environmental indicators. Visual exploration of the data reveals cross-national patterns in emissions intensity, economic development, and *per capita* burdens, highlighting its utility for comparative research and communication of sustainability trade-offs.

5.2 The limitations of the dataset

While the dataset is designed to facilitate broad reuse and analytical flexibility, several limitations should be acknowledged to ensure appropriate interpretation. First, the dataset is restricted to national aggregates and does not provide subnational or sectoral disaggregation. This limits its utility for fine-grained spatial analysis or attribution of impacts to specific economic sectors or infrastructure systems, and may constrain its direct applicability to local or sector-specific policy planning.

Second, while historical values are based on curated statistical sources, forecast values (2021–2050) are derived from external projections under the SSP and may not fully capture future policy shifts, technological transitions, or exogenous shocks. Users should exercise caution when interpreting trends beyond 2020, particularly for comparative studies or decision-making under high uncertainty. Policy conclusions drawn solely from these projections should account for these inherent limitations.

Third, the dataset may exhibit geographic bias, as it includes only countries with consistently available and reconcilable data across all variables. This could lead to underrepresentation of certain countries, especially low-income or data-scarce countries, potentially skewing analyses of global waste and emissions trends or equity considerations in policy assessments. Similarly, even when variables are harmonized, their alignment may reflect differences in original data collection methodologies, which could introduce latent inconsistencies despite best-effort normalization.

Lastly, the dataset lacks uncertainty metrics at the individual data point level, and it does not model potential errors introduced

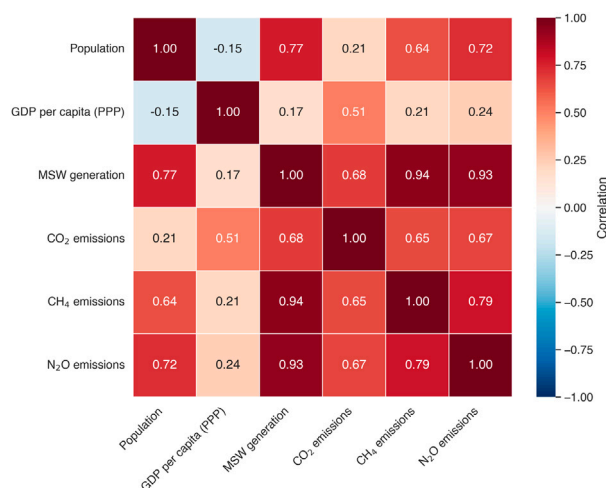


FIGURE 4

Pairwise correlations among demographic, economic, waste, and emissions variables. Pearson correlation coefficients between six national-level indicators—population, GDP *per capita* (PPP), municipal solid waste (MSW) generation, and emissions of CO₂, CH₄, and N₂O—aggregated across all countries and years. Positive values (red) indicate direct associations, while negative values (blue) reflect inverse relationships. Strongest correlations are observed between MSW generation and CH₄ emissions, and between CO₂ and N₂O emissions. Note that these correlations are fully aggregated; temporal and regional heterogeneity may be masked by this aggregation.

during source merging or projection harmonization. In addition, while the dataset is disseminated via the FAIR² platform, this manuscript does not include a formal evaluation of any synthetic data generated for AI-readiness purposes. Users should be aware that synthetic data may require independent assessment (e.g., outlier analysis, schema checks, or predictive utility testing) before use in downstream applications. While this is partly mitigated by structured metadata and known provenance, users requiring robust uncertainty quantification may need to supplement the dataset with external error models or sensitivity analyses.

6 Conclusion

The dataset described in this article provides harmonized, national-level records of population, GDP *per capita* (PPP), MSW generation, and associated GHG emissions (CO₂, CH₄, N₂O) from 1990 to 2050. Integrating historical statistics with consistent forecast projections supports longitudinal assessments of environmental pressure and sustainability performance across countries. The inclusion of both demographic and emissions-related variables enables diverse forms of analysis, including *per capita* and per GDP normalization, decoupling trends, and evaluation against planetary boundary-based thresholds.

Structured to meet FAIR² principles, the dataset is well-suited for use in data-driven environmental assessments, policy modeling, and responsible AI workflows. Its design allows for integration into planetary boundary life cycle assessments, national reporting frameworks, and emissions equity analyses. Users should be aware that the dataset focuses on national-level indicators and does not provide formal uncertainty ranges for forecasts. Despite this, it offers a robust and reusable foundation for comparative and exploratory research.

Overall, this dataset contributes to the growing ecosystem of open, structured, and reproducible data resources supporting global

sustainability science. All raw data, metadata, and supplementary materials, including detailed protocols and methods, are accessible via the FAIR² Data Portal (<https://doi.org/10.71728/senscience.k2f7-p5v9>).

Data availability statement

The dataset is FAIR²-certified and publicly available under the Open Data Commons Attribution License (ODC-By v1.0), permitting unrestricted reuse with appropriate attribution. Access is provided through two coordinated components: an interactive FAIR² Data Portal enabling visual exploration, and a downloadable FAIR² Data Package containing raw data, structured metadata, and detailed methodological documentation. The FAIR² Data Package and Data Portal are accessible via <https://doi.org/10.71728/senscience.k2f7-p5v9>.

Author contributions

ZH: Conceptualization, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review and editing. KW: Conceptualization, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing – review and editing. WC: Methodology, Supervision, Validation, Writing – review and editing. YF: Methodology, Validation, Writing – review and editing. SY: Methodology, Validation, Writing – review and editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was supported by the

Ministry of Higher Education Malaysia (Fundamental Research Grant Scheme FRGS/1/2020/TK0/XMU/02/2); Xiamen University Malaysia (Xiamen University Malaysia Research Fund XMUMRF/2023-C11/IENG/0057 and IENG/0069); and the Korea Environment Industry & Technology Institute (KEITI) of the Korea Ministry of Environment (MOE) (grant 2022003560007).

Acknowledgements

The authors would like to express their sincere gratitude to Z. X. Phuang, M. Y. Chin, J. K. Ooi, and W. L. Ng for providing comments on improving the work. This document and the accompanying data package were prepared by the authors with assistance from the SENSICENCE FAIR² Data Publishing platform (v0.11 β), which uses generative AI technology.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Abbasi, M., and El Hanandeh, A. (2016). Forecasting municipal solid waste generation using artificial intelligence modelling approaches. *Waste Management* 56, 13–22. doi:10.1016/j.wasman.2016.05.018
- Clark, M. A., Domingo, N. G. G., Colgan, K., Thakrar, S. K., Tilman, D., Lynch, J., et al. (2020). Global food system emissions could preclude achieving the 1.5°C and 2°C climate change targets. *Science* 370 (6517), 705–708. doi:10.1126/science.aba7357
- He, R., Sandoval-Reyes, M., Scott, I., Semeano, R., Ferrao, P., Matthews, S., et al. (2022). Global knowledge base for municipal solid waste management: framework development and application in waste generation prediction. *J. Clean. Prod.* 377, 134501. doi:10.1016/j.jclepro.2022.134501
- Hoy, Z. X., Woon, K. S., Chin, W. C., Fan, Y. V., and Yoo, S. J. (2023). Curbing global solid waste emissions toward net-zero warming futures. *Science* 382 (6675), 797–800. doi:10.1126/science.adg3177
- IPCC (2006). 2006 IPCC guidelines for national greenhouse gas inventories. *Prepared by the national greenhouse gas inventories programme*. Japan: Institute for Global Environmental Strategies IGES.
- IPCC (2019). in *2019 refinement to the 2006 IPCC guidelines for national greenhouse gas inventories*. Editor Calvo Buendia, E. (Japan: IGES).
- IPCC (2021). *Climate change 2021: the physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press.
- Olawore, A. S., Wong, K. Y., and Oladosu, K. O. (2024). Prediction of municipal waste generation using multi-expression programming for circular economy: a data-driven approach. *Environ. Sci. Pollut. Res.*, 1–16. doi:10.1007/s11356-024-35388-y
- Turner, D. A., Williams, I. D., and Kemp, S. (2015). Greenhouse gas emission factors for recycling of source-segregated waste materials. *Resour. Conservation Recycl.* 105, 186–197. doi:10.1016/j.resconrec.2015.10.026
- World Bank (2018). *What a waste 2.0: a global snapshot of solid waste management to 2050*. Washington, DC: World Bank. Available online at: <http://hdl.handle.net/10986/30317>.
- World Bank (2022). World development indicators. Available online at: <https://datacatalog.worldbank.org/dataset/world-development-indicators> (Accessed June 15, 2025).

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. Generative AI was used in the preparation of this manuscript to improve readability and language. The authors have reviewed and edited the content as needed and take full responsibility for the content of the publication

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Appendix A

TABLE A1 Country coverage of the dataset by continent.

Continent	Countries included
Africa	Algeria; Congo, dem. Rep.; Egypt, arab rep.; Ethiopia; Morocco; Nigeria; South Africa; Tanzania; Uganda
Asia	Bangladesh; China; India; Indonesia; Iran, islamic rep.; Iraq; Japan; Korea, rep.; Malaysia; Pakistan; Philippines; Saudi Arabia; Taiwan, China; Thailand; vietnam
Europe	France; Germany; Italy; Netherlands; Poland; Russian federation; Spain; Turkey; Ukraine; United Kingdom
North America	Canada; Mexico; United States
South America	Argentina; Brazil; Colombia; Peru; Venezuela
Oceania	Australia

