scientific data

OPEN

DATA DESCRIPTOR

The SESAME Human-Earth Atlas

Abdullah Al Faisal[1]✉, Maxwell Kaye[1,2], Maimoonah Ahmed[1] & Eric D. Galbraith[1,3]

Human activities such as food production, mining, transportation, and construction have extensively modified Earth's land and marine environments, causing biodiversity loss, water pollution, soil erosion, and climate change. However, studying spatial aspects of the relationships that link the global human system with non-human parts of the Earth-system is hampered by data fragmentation. Here we present the Surface Earth System Analysis and Modeling Environment (SESAME) Human-Earth Atlas, which includes hundreds of variables capturing both human and non-human aspects of the Earth system on two common spatial grids of 1- and 0.25-degree resolution. The Atlas is structured by common spheres, and many variables resolve changes over time. Machine learning is used selectively to interpolate data in undersampled regions. Many of the national-level tabular human system variables are downscaled to spatial grids using dasymetric mapping, accounting for country boundary changes over time. Raster, point, line, polygon, and tabular jurisdictional (i.e., country) data were mapped onto a standardized spatial grid at the desired resolution. The Atlas facilitates data discovery and modeling of human-Earth system dynamics.

Background & Summary

Earth system science is an emerging interdisciplinary field focused on understanding the structure and functioning of the Earth as a complex adaptive system[1], driven by varied interactions among energy, matter, and living organisms[2–5]. Human activities have disrupted many aspects of Earth system functioning, elevating the importance of understanding the processes within the coupled human-Earth system[6]. Earth system science provides conceptual and computational tools to investigate planetary level impacts and their far-reaching consequences, including climate change, loss of biodiversity, and excessive nutrient inputs[4,6]. However, studies concerning humans have long been conducted independently from the study of the non-human components of the Earth system[7]. This separation has led to divergent perspectives, goals, and approaches, making it challenging to integrate mechanistic understanding of the global human system within the rest of the Earth system[8].

Because of the persistent divide between studies of human and non-human parts of the Earth system, the production and collection of data relevant to each has evolved independently. Whereas data on the physical Earth system are frequently compiled in standardized, gridded formats, data on human system variables are frequently collected by governments at national or regional scales and archived in tabular formats. Human system data are often collected through surveys, which can lead to limitations on sample sizes and privacy, restricting their reporting in moderate and high-resolution grids. In contrast, many natural system variables are detectable by satellite remote sensing, which facilitates gridded data representation, while others are modeled on a grid. With a few notable exceptions (e.g., night lights[9], fishing[10], and crop production[11]), most human activities are not directly observable by satellites in a gridded manner. The lack of easily comparable datasets across diverse human-Earth components hinders the analysis of relationships between anthropogenic processes and Earth's physical variables and makes it difficult to integrate human processes into Earth system models[12,13]. Data on the non-human Earth system can also be challenging to work with, as relevant data are spread across websites, provided in diverse formats, and use different map projections, spatial resolution, and non-standardized units. Both beginner and experienced researchers struggle to access and use these valuable scientific datasets due to their heterogeneous data types[14].

An example of the challenge of bringing human data into a spatially resolved Earth system data framework is illustrated in Fig. 1. While much of the existing human data is only available at the country-level (e.g., building material masses, as shown in Fig. 1a), Earth system variables—such as wet bulb temperature (Fig. 1b)—are frequently gridded, revealing much more detailed geospatial trends and variations, with large gradients within

[1]Department of Earth and Planetary Sciences, McGill University, Montreal, Québec, H3A 0E8, Canada. [2]Department of Mathematics and Statistics, McGill University, Montreal, Québec, H3A 0B9, Canada. [3]Institut de Ciència i Tecnologia Ambientals (ICTA-UAB), Universitat Autònoma de Barcelona, Barcelona, Spain. ✉e-mail: abdullah-al-faisal@localpathways.org

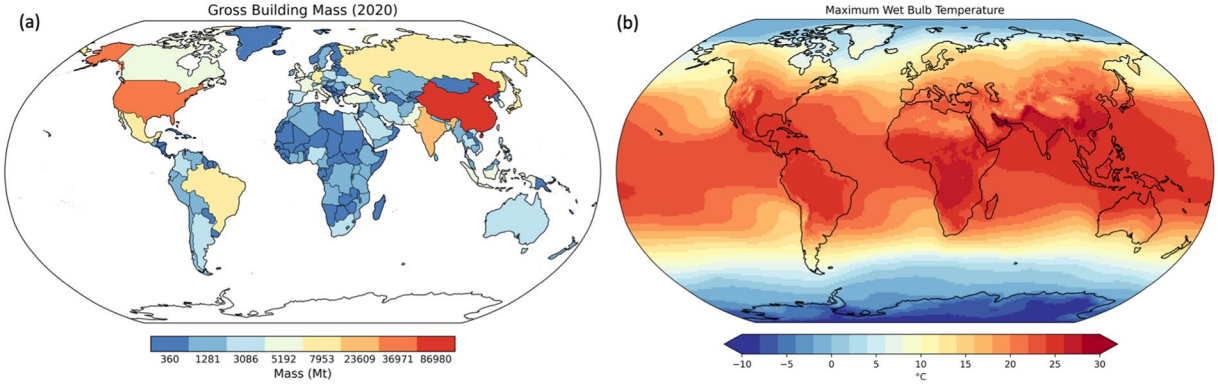

Fig. 1 Typical human and Earth-system data disparities. Panel (**a**) shows a map of country-level data, with a single value per country, in this case the total mass of buildings[70]. In contrast, panel b shows globally gridded maximum wet bulb temperature, which exhibits significant spatial variability within countries. The ability to directly compare between these datasets is very limited.

many individual countries. As a result, we currently face limitations in the ability to conduct insightful local (grid-level) geospatial analyses that bridge human and Earth-system variables.

The importance of developing integrated, multidimensional data products that follow FAIR (Findable, Accessible, Interoperable and Reusable) principles has been recognized in recent years[15–19]. For example, GriddingMachine is a recently developed database and software infrastructure for Earth System Models that includes a gridding software package to standardize diverse global datasets[14]. Another example builds on the concept of Earth System Data Cubes[17], intended to collect diverse Earth system data in a network Common Data Form (netCDF) format that can be accessed through an online interface (https://www.earthsystemdatalab.net). However, although these datasets have the potential to explicitly support Earth system research, we are not aware of similar efforts that emphasize the inclusion of human system data alongside the non-human Earth system data.

The conversion of jurisdictional data to spatially resolved gridded data requires the application of some form of downscaling, which can be achieved using the dasymetric mapping technique. Dasymetric mapping aims to improve spatial accuracy by reallocating jurisdictional data under the assumption of covariance with one or more spatially resolved surrogate variables[20,21]. It is a popular technique that was originally developed to distribute administrative-level population data based on land cover distribution[22]. Dasymetric mapping has already been used in a variety of fields, including sea level rise vulnerability assessment[23], examining urban spatial features from points-of-interest data using population, land use patterns, and socioeconomic characteristics[24], and understanding crime distribution[25]. This approach has also gained popularity in epidemiology and health research, including but not limited to the identification of fine-scale risk patterns for infectious and chronic diseases[26], and population distribution in health exposure research[27]. However, there is a lack of ready-to-use tools that can be applied at the global scale, and many of the existing dasymetric tools are limited to population distribution analysis and do not leverage broader applicability to other sectors.

Here, we aim to contribute to improving the accessibility, discoverability, and reusability of global human-Earth system data, by presenting a Human-Earth Atlas[28]. The Atlas is based on the following key objectives: (a) to provide a compact, downloadable, easily navigated data atlas that comprehensively spans data from many traditional disciplines, fostering holistic understanding of human-Earth interactions, and (b) to enable users to easily create spatially gridded datasets from diverse human and non-human sources, including dasymetric mapping for jurisdictional data. The dataset is constructed using the open-source Surface Earth System Analysis and Modelling Environment (SESAME) software package. The output is provided in netCDF format, which allows individual files to store multidimensional time series data. To our knowledge, the Atlas represents a novel approach to comprehensively integrate diverse human and non-human variables into a spatially resolved grid for rapid plotting, comparative analysis and re-use. For scientific data management, we follow the FAIR guiding principles[29]. Users can use the open-source standardized global data Atlas and run the open-source SESAME software tools, to generate, convert heterogeneous data types, download, and visualize data with only a few lines of code in a typical Python environment.

The Human-Earth Atlas[28] is intended to help advance interdisciplinary research, particularly to bridge the gap between human and Earth system sciences. We have placed special emphasis on the functions for gridding jurisdictional (country-level) tabular data onto global spatial grids, which overcomes a barrier to comparing human data with other Earth system data. The dataset is designed to bring together diverse data on all aspects of our planetary surface environment in a common format that can be easily accessed, navigated, and manipulated. We also hope that the additional tools provided with the dataset will prove useful to many researchers, particularly vector-to-raster grid conversion tools, because there are few tools available that work with vector data. The package also includes a number of ancillary tools for data conversion, summarization, visualization, and other operations. We have provided a few examples of variables included in the database here, but this is only a glimpse – the intention is that users download the database to explore, compare, and discover new relationships.

In the future, we plan to continue growing and updating the Atlas. The SESAME software package will also expand to include a foundation of modeling capabilities, with the intent to enable future researchers to conveniently explore, manipulate, model, and visualize human-Earth system data in a single platform. The Atlas is open-source and users are encouraged to contribute both software branches and datasets, as well as suggestions for improvements and additional functionality to encourage trans-disciplinary collaboration.

Methods

Human-earth atlas overview. The Human-Earth Atlas[28] is designed to provide a large number of geographically gridded human-Earth variables in a single standardized format for easy exploration, analysis, and discovery. To parsimoniously organize human-Earth data in an easily-navigated structure, we categorize datasets according to five spheres. While the spheres are all profoundly interconnected, continually exchanging fluxes of material and energy, such partitioning is largely intuitive and is frequently used[30]. The definitions of the spheres are based on the types of material of which they are composed, following Galbraith *et al.*[31] as given in Table 1.

Variables reflecting the state of an entity or material (e.g., masses, chemical compositions, categorizations, numerical abundances) are placed within the sphere corresponding to the entity or material, or in the sphere in which the property would be measured (e.g., albedo is placed in the atmosphere). Version 1 of the Atlas provides more than 250 human Earth system variables across the five planetary spheres. Most variables are stored as 2D spatial arrays, and many include a few vertical dimensions (i.e., depth) as well as time dimension. Missing values are indicated by Not a Number (NaN). Each variable includes additional attributes such as a long name, units, and sources. We standardized variable units where feasible (e.g., g m^{-2} for mass per grid area, W m^{-2} for energy flux per grid area, and fraction of grid cells for categorical data such as land cover, biomes, and rocks) and stored them by following a standardized naming convention (see Data Records) in the Atlas.

We provide the Atlas at two spatial resolutions, using consistent coordinates of latitude [−90 to 90] and longitude [−180 to 180]. The first resolution is 1° × 1°; so that each grid cell represents an area of approximately 110 kilometers by 110 kilometers (with the longitudinal dimension decreasing with distance from the equator). This resolution offers a reasonable level of spatial detail, allowing for the identification of regional and sub-regional patterns. At the same time, its resolution is sufficiently coarse, such that, when mapped, it does not overwhelm the viewer with local detail, aiding in the identification of large-scale global patterns. There are approximately 64800 values stored globally per variable. This coarse resolution is more computationally and theoretically manageable, especially for higher order interaction, and enables feasible polynomial time analysis of algorithms such as pairwise grid cell comparisons. This resolution also allows for the inclusion of hundreds of variables, while remaining sufficiently compact that it can be downloaded and maintained as a local copy on a personal computer for ready access. We also provide a limited collection of data at 0.25° × 0.25° resolution. High-resolution variables are only provided if the raw data or surrogate variable is found at a resolution of 0.25° or finer. There are approximately 1,036,800 values stored per variable at 0.25° spatial resolution, 16 times more than in 1-degree grids. Furthermore, the SESAME software package is designed to support higher-resolution ancillary datasets, allowing users to re-grid variables to meet the specific needs of their applications.

To align our research principles with the FAIR, we have systematically structured our data to enhance findability, accessibility, interoperability, and reusability. Each netCDF file is assigned with a unique and persistent identification (F1) through unique naming convention, storing rich metadata (F2) for each variable (e.g., long name, units and sources of raw data). Our datasets are easily accessible (A1) through Figshare with supporting open-source software packages available on GitHub and PyPI, which are free and universally implementable without authentication for public use (A1.1). The metadata remains accessible indefinitely though Figshare, even if the datasets are not (A2). Data is provided in netCDF format – a widely accepted standard for scientific data that supports compatibility with a broad range of tools, programming languages, and software (I1), which incorporates standardized vocabularies for the relevant scientific communities (I2). Furthermore, we ensured the inclusion of qualified references to other metadata (I3) in each variable and dataset's documentation files, enhancing integration and reusability of our data. Our data contains rich descriptions (R1), clear licensing (R1.1), detailed provenance (R1.2) and maintains domain-relevant community standards to ensure effective usability and reusability in diverse scientific contexts.

Atlas categories. To aid navigation, the Atlas is organized hierarchically by spheres. Tables 2–6 show a summary of the datasets available in SESAME Human-Earth Atlas[28] v1.0, along with their respective sources.

Atmosphere. We obtained monthly data on precipitation (mm), temperature (°C), surface pressure (Pa), 10-meter U and V wind components (m s^{-1}), relative humidity (%), cloud coverage, and cloud opacity (dimensionless) from ERA5[32,33] and MERRA-2[34] reanalysis models for the period from 2000 to 2022. The ERA5 dataset was obtained from the Climate Data Store (https://cds.climate.copernicus.eu/datasets), while the MERRA-2 dataset was retrieved from GES DISC (https://disc.gsfc.nasa.gov/) and My NASA Data (https://mynasadata.larc.nasa.gov/). The raw data, provided in netCDF format, were re-gridded to a 1-degree resolution by averaging values of 0.25-degree grid cells within each larger cell and integrated into the Atlas using SESAME toolboxes. Precipitation data (mean total precipitation rate, kg m^{-2} s^{-1}) were converted to total monthly precipitation (mm). The most commonly used surface weather variables, e.g., temperature and precipitation, are provided for a longer time range (1940–2024). We calculated the wet-bulb temperature (Tw), a key metric for assessing human survivability in climate science, using Stull's formula mentioned in equation S1[35]. Although MERRA-2 provides Tw at 2-meter height, its reliability is currently in question[36]. Instead, we calculated the monthly averaged maximum wet-bulb temperature using ERA5 reanalysis products of air temperature and humidity at 1000 hPa. These were first re-gridded using the SESAME grid_2_grid tool to 1-degree spatial grids, after the Tw was computed for the

Sphere	Definition	Example variables
Lithosphere	The non-living mineral Earth, composed of crustal rock and the non-organic component of soil and regolith.	Surface rock type; ore deposit compositions; volcanic gas efflux; earthquake frequency; copper ore extraction rates
Atmosphere	The blanket of gas that envelopes the Earth and extends to the edge of space, including air, precipitation, clouds, and particulate aerosols.	Precipitation rate; 2 m air temperature; net radiation at top of atmosphere; surface wind velocity; surface albedo
Hydrosphere	Liquid water and ice at the Earth's surface and in the ground, including lakes, rivers, ice sheets, groundwater, and the ocean.	Sea surface temperature; dissolved nitrate concentration; advection velocity; streamflow
Biosphere	All living organisms, including humans.	Taxonomic biomass density; net primary production; community metabolic rate; species diversity; sinking organic carbon flux; human activities; education level
Technosphere	All non-living human-creations.	Mass of roads; mass of steel in buildings; length of fibre optic cable; rate of plastic manufacturing; rate of particulate carbon emission

Table 1. Definition of spheres in Earth system science and some example variables.

NetCDF filename	Description	References
A.surface_weather.2000–2022.a.nc A.surface_weather.2000–2022.clim_m.nc	9 data variables describing climatological averages (i.e., average annual and seasonal cycle over multiple years) including temperature, precipitation, could coverage, wet bulb temperature, relative humidity, etc.	Refs. [32,33,38,96,97]
A.surface_weather.temp_prec.19402024.clim_m.nc A.surface_weather.temp_prec.1940–2024.a.nc	2 additional extensively used data variables in each netCDF file describing climatological averages for longer time period including precipitation and temperature	Ref. [33]
A.energy_balance.2000–2022.clim_m.nc A.energy_balance.2000–2022.a.nc	11 data variables in each netCDF file describing climatological averages of atmospheric energy fluxes, including surface longwave flux down, surface shortwave flux down, incoming solar flux, top of the atmosphere shortwave flux, etc.	Refs. [40,97,98]
A.aerosols.2000–2022.clim_m.nc A.aerosols.2000–2022.a.nc	Climatological aerosol averages in annual and seasonal cycle over multiple years	Ref. [39]

Table 2. Atmosphere data available in SESAME Human-Earth Atlas (version 1).

NetCDF filename	Description	References
Bh.population.2000–2020.a.nc	5 data variables describing population count, density, population weighted population density from two different sources	Ref. [41]
Bh.monetary.2000–2020.a.nc	GDP (PPP 2005 U.S. dollars)	Ref. [99]
B.land.biomes.nc	14 data variables describing fraction of different biomes in grid cells	Ref. [50]
B.food.nc	39 data variables describing crop production, consumption, livestock, fish catch, vegetables and other edible foods consumed by human and livestock	Refs. [47,100–110]
B.ocean.primary_production.1998–2023.clim_m.nc B.primary_production.2001–2023.a.nc	3 data variables such as terrestrial, ocean and combined NPP product available on climatological annual average and seasonal average is only available on ocean NPPs	Refs. [48,49]
B.land.species.2024.nc	10 data variables describing percentage of species richness and threats for amphibians, birds, mammals, and reptiles.	Ref. [111]
Bh.human_footprint.2000–2022.a.nc	2 indices of human footprints from different sources and time periods	Refs. [43,112]
B.land.forest.nc	6 data variables on forestry e.g., aboveground and belowground biomass, canopy heights, forest cover, gain and loss for certain periods	Refs. [52,113,114]
B.land.cover.2001–2023.a.nc	17 data variables of fraction of different land cover classes (e.g., open shrublands, woody savannas, grasslands, permanent wetlands, croplands, urban and built-up lands, and so on)	Ref. [115]

Table 3. Biosphere data available in SESAME Human-Earth Atlas (version 1).

period 2000–2024. We then calculated the monthly climatology and identified the peak Tw by comparing maximum grid values across the twelve months (Fig. 1b).

In addition to standard meteorological variables, we also collected atmospheric energy-related variables, including surface longwave (flux up and down), surface shortwave (flux up and down), top of the atmosphere shortwave (flux and net flux), and top of the atmosphere longwave flux under all sky conditions. We also collected additional energy related data[37,38] such as incoming solar flux, total latent energy flux, and sensible heat flux from turbulence, all in units of W m^{-2}. In addition, we collected data on aerosol optical depth[39] and albedo[40]. We prioritize the MERRA-2 reanalysis data for most of the atmospheric variables to leverage its extensive temporal coverage, recognizing that this choice prioritizes temporal depth over spatial resolution. All data were collected at a monthly resolution and subsequently provided as climatological averages (i.e., annual and seasonal cycles) at a 1-degree resolution (Table 2).

Biosphere. This category includes data on living species, including biological and socio-cultural features of the human population (Table 3). To organize this complex domain, we divide the variables into two broad subcategories: human variables (e.g., demography, gross domestic product (GDP), and labor) and ecological variables (e.g., net primary production (NPP), biomes, footprint, threatened species, and species richness).

NetCDF filename	Description	References
H.ocean.temperature.1870–2023.clim_m.nc H.ocean.temperature.1870–2023.a.nc	climatological sea surface temperature averages in annual and seasonal cycle over multiple years	Ref. [55]
H.ocean.chemistry.1965–2022.clim_m.nc	6 data variables describing climatological mean of global ocean variables, including dissolved oxygen, nitrate, silicate, phosphate, etc. at various depths	Refs. [56–59]
H.ocean.mixed_layer.clim_m.nc	4 data variables describing mixed layer properties (e.g., absolute salinity, depth, potential density, and conservative temperature) of ocean as climatological means	Ref. [60]
H.ocean.color.2012–2025.clim_m.nc H.ocean.color.2012–2025.a.nc	Climatological aerosol averages in annual and seasonal cycle of chlorophyll-a concentration and diffuse attenuation coefficient at 490 nm over multiple years	Refs. [61,62]
H.hydrowaste.nc	8 data variables describing population served by various level wastewater treatment plants and water discharges.	Ref. [66]
H.freshwater.river_lake.nc	5 data variables representing river, lake, reservoir and river discharge at standardized spatial grids.	Refs. [63,64]
H.freshwater.streamflow.1960–2015.a.nc	Mean annual streamflow re-gridded at standardized grids	Ref. [65]

Table 4. Hydrosphere data available in SESAME Human-Earth Atlas (version 1).

NetCDF filename	Description	References
L.rock.surface_lithology.nc	16 data variables describing fraction of different rock types (e.g., basic volcanic rocks, siliciclastic sedimentary rocks, mixed sedimentary rocks, and so on)	Ref. [116]
L.soil.nc	11 data variables that store the physical, chemical, and organic properties of soil (e.g., organic carbon density, pH, sand particles, clay particles, etc.) in six depth dimensions	Ref. [68]

Table 5. Lithosphere data available in SESAME Human-Earth Atlas (version 1).

NetCDF filename	Description	References
T.transportation.merchant_fleet.2011–2021.a.nc	12 data variables describing gross and steel mass of different types of ships (e.g., general cargo, oil tankers, etc.)	Refs. [70,75,117]
T.transportation.air.nc	7 data variables describing the number of airplanes and material masses (e.g., iron, steel, aluminum, titanium, etc.)	Refs. [72,70,118]
T.transportation.pipelines.nc	3 data variables including oil and gas pipeline length, density and steel mass	Refs. [92,93,119]
T.transportation.railways.nc	4 data variables such as railtrack length, density, steel mass and gross mass of railway infrastructure	Refs. [70,83,120–122]
T.transportation.rolling_stocks.nc	17 data variables including multiple material masses of wagon, locomotive, railcar and coach.	Refs. [70,88–90,123]
T.buildings.2000–2020.a.nc	24 data variables describing both residential and non-residential built-up volumes, building masses (e.g., aggregates, aluminum, concrete, copper) for both urban and rural areas	Refs. [70,79,124]
T.urban_extent.2000–2020.a.nc	1 data variable describing fraction of urban area in grid cells	Ref. [77]
T.transportation.roads.nc	12 data variables of road material masses (e.g., aggregates sub-bas, asphalt, concrete, gravel, etc.) with road length and density	Refs. [70,83,84]
T.transportation.vehicles_land.2022.a.nc	32 data variables describing passenger, commercial and trailer counts and masses (e.g., aluminum, copper, glass, iron, plastic, gross mass, etc.)	Refs. [70,71,84,85,125,126]
T.food_provision.agricultural_machinery.2001–2021.a.nc	1 data variable describing the tractor mass for agricultural uses	Refs. [70,71,127]
T.nightlight_time.1992–2022.a.nc	1 data variable showing annual DMSP and VIIRS nighttime light data	Ref. [128]
T.2021.nc	9 common technosphere variables collected and stored for a single year from multiple SESAME Atlas netCDF files	Refs. [70–72,74,75,81–84,87,93,120,125]

Table 6. Technosphere data available in SESAME Human-Earth Atlas (version 1).

We collected and re-gridded human populations[41] from WorldPop (https://www.worldpop.org/) and estimated population densities at a global scale from 2000 to 2020. The population data were available in high resolution (1 km) with the units of population per grid-cell. These were gridded into standardized 0.25- and 1-degree grids by summing all smaller cells within each large cell, thereby obtaining the corresponding population density (people m-2). However, since the population density of 1-degree spatial grids does not accurately represent populations in a large area, we estimated the population-weighted population density (PWPD). To achieve this, we calculated high-resolution density by dividing each pixel's population by its area using the high-resolution (1 km) gridded population data, then re-gridded the results to a 1-degree spatial grids. Additionally, we processed GDP data (in PPP 2005 U.S. dollars) from 2000 to 2020, which were also re-gridded to 0.25- and 1-degree grids[42]. These demographic and economic indicators provide essential context for analyzing socioeconomical dynamics.

The atlas also stores a re-gridded human footprint index to assess the impact of human activities on the environment, biodiversity, and ecological conservation. We downloaded human footprint[13] raster TIFF files from 2000 to 2022 and the global human modification[43] index (CSP gHM) for 2016 and re-gridded to 1-degree spatial grids. Human-related activity also includes food production and consumption. We used estimates of demersal and pelagic fish catches from the BOATSv2 model[44]. Ref. [45] provides production data for crops, fruits, and vegetables such as banana, barley, cassava, groundnut, maize, millet, rice, soybean, sugarcane, wheat, tobacco, and so on. The number of livestock[46] is also recorded in gridded format. We calculated the energy production and consumption of humans and livestock in W m-2 for 2015[47].

We collected the global MODIS land cover type data, which includes 17 classes such as croplands, various categories of forests (e.g., evergreen, deciduous needleleaf forests), closed and open shrublands, savannas, permanent snow & ice, barren, and water bodies on an annual basis from 2001 to 2023. The data were downloaded from Google Earth Engine in MODIS Sinusoidal projection system as raster TIFFs. We reprojected to WGS 1984, converted into polygon shapefiles, and then used to compute land cover fractions with SESAME's poly_2_grid function.

NPP, for both ocean and terrestrial environments, is crucial for understanding Earth's systems[48,49]. We collected estimates of ocean NPP from five different models at eight-day intervals. These were then converted to monthly values, averaged among the five models, and re-gridded. We calculated climatological averages for annual (g C m^{-2} y^{-1}) and seasonal monthly cycles (g C m^{-2} m^{-1}) for ocean NPP. Terrestrial NPP were collected annually from MODIS Terra products[48], re-gridded to the same spatial resolution, and the units were converted to g C m^{-2} y^{-1}. We combined both terrestrial and ocean NPP by adjusting the land fractions to achieve global coverage of NPP. This provided a comprehensive view of NPP worldwide on an annual scale from 2001 to 2023.

Ecological context is further enriched by biome classifications. Ecological biomes[50] (e.g., boreal forests/taiga, deserts & xeric shrublands, flooded grasslands & savannas, mangroves, mediterranean forests, woodlands & scrub, etc.) were obtained from (https://ecoregions.appspot.com/), and using the poly_2_grid tool, we estimated the fraction of each biome in each grid cell. Biodiversity indicators such as species richness and the number of threatened species[51] for organisms including amphibians, birds, mammals, and reptiles were collected from IUCN (https://www.iucnredlist.org/resources/other-spatial-downloads). This information was initially collected at 900 km^2 raster tiff files[51], then re-gridded into 0.25- and 1-degree grids. Lastly, we gathered some key vegetation structure variables from Google Earth Engine (https://developers.google.com/earth-engine/datasets), including forest gain, loss, tree cover[52], canopy heights[53], aboveground biomass and belowground biomass[54]. These datasets were processed and re-gridded to provide insight into terrestrial ecosystem health and carbon storage potential.

Hydrosphere. The atlas includes a wide range of marine and freshwater hydrosphere variables. We included various ocean variables (e.g., including sea surface temperature, dissolved oxygen, salinity, and chlorophyll concentration) from World Ocean Atlas (https://www.ncei.noaa.gov/access/world-ocean-atlas-2023/), and freshwater variables (e.g., river, lake properties, discharge, streamflow, and hydrowaste) from HydroSHEDS (https://www.hydrosheds.org/products). Climatological averages (e.g., annual and monthly) of global sea surface temperatures[55] from 1870 to 2023, as well as monthly climatological averages of ocean chemistry variables such as dissolved oxygen, nitrate, phosphate, silicate, and oxygen saturation fraction, are also stored at various depths[56–59]. In addition, we have added ocean mixed layer properties[60] (e.g., absolute salinity, conservative temperature, depth and potential density), ocean color properties[61,62] such as chlorophyll-a concentration, and diffuse attenuation coefficient, as well as river variables including river discharge[63], river length, fraction of reservoir, lake[64], mean annual streamflow[65], and hydowaste[66] data (Table 4).

Lithosphere. We collected lithology polygons from Ref. [67] and created fractional coverage of 15 rock types (e.g., unconsolidated sediments, basic volcanic rocks, metamorphic rocks, acid plutonic rocks, etc.), which we then stored in the Atlas using the poly_2_grid tool. Various soil properties[68,69], such as organic carbon density, stocks, carbon content, total nitrogen, bulk density of the fine earth fraction, cation exchange capacity, volumetric fraction of coarse fragments, proportion of clay particles, and pH are collected from SoilGrids (https://soilgrids.org/), re-gridded, and stored in the atlas at various depth intervals (Table 5).

Technosphere. The technosphere refers as all non-food matter extracted from other spheres of the Earth system and transformed to novel states that can provide end-uses to humans[31]. We collected agricultural machinery, commercial aircraft, merchant ships, rolling stocks, pipelines, roads, railways, and commercial and residential building stocks at the national level[70] and distributed them proportionally to local surrogate variables via dasymetric mapping for each grid cell (Table 6).

Agricultural machinery data were obtained from the Food and Agriculture Organization Statistics (FAOSTAT) (https://www.fao.org/statistics/en), including four-wheel (4 W), two-wheel (2 W) tractors, and combine tractors from 1963 to 2009 for 134 countries. For combined category tractors, a 50% steel mass assumption was made and the methodology is detailed in Ref. [70]. A random forest machine learning model was developed to predict tractor mass using variables like GDP, total population, urban population percentage, food production, harvested area, yield, income class, and year of production (Fig. S1). The model performed well, with R^2 values of 0.995 (train) and 0.977 (test), and showed consistent accuracy through residual and validation plots. The most influential predictors were GDP, total population, and harvested area (Fig. S1d). Finally, the agricultural machinery mass data from 2001 to 2023 was distributed proportionally to each country's cropland area[71] into 0.25- and 1 degree spatial grids using SESAME's table_2_grid tool.

National level commercial aircraft material compositions[70], including iron, steel, aluminum, plastic, and titanium, were distributed based on the number and types of airports[72] in grid cells. Estimating the ratio of plane capacities at different airports is difficult, but a rough estimate is—seaplane base: small Airport: medium airport: large airport = 1:5:30:100. SESAME's point_2_grid tool was used to convert airport location data into spatial grids. The mean of material intensity[73] for five types of commercial aircraft was used to calculate the average material composition of an aircraft using the CIA World Factbook (https://www.cia.gov/the-world-factbook/) data. Merchant fleet data was sourced from the United Nations Conference on Trade and Development (UNCTAD) (https://unctadstat.unctad.org/datacentre/) at country-level from 2011 to 2020. Ref. [70] compiled gross tonnage for five ship categories, focusing on vessels over 11,000 tons. Steel mass per gross tonnage sourced from Ref. [74] was used to compute the steel mass of the fleets. The global fleet mass data was later distributed proportionally to the global shipping traffic density[75] at 0.25- and 1 degree global spatial grids.

We gridded the national-level residential and non-residential material masses based on global built-up volume[76] in rural and urban regions[77] from 2000 to 2020 for multiple materials. Ref. [70] compiled country-level building material stock data for multiple materials such as concrete, aggregate, aluminum, copper, steel, and wood from Ref. [78] and Ref. [79]. GHSL built-up volume[76] for total and residential were downloaded, pre-processed, and separated into urban and rural areas based on urban extent dataset[80]. The built-up volume data was only found at 5-year intervals from 2000 to 2020, and the missing years volume data were interpolated linearly (Fig. S2). These preprocessed gridded volumes were used as surrogate variables to distribute the country-level material stock data. We also gridded and stored an additional gross building mass data obtained from Ref. [81] and dasymetrically distributed using the gridded Ref. [82] data.

Road materials, motor vehicles, railways, rolling stocks, and oil and gas pipelines data were similarly distributed through a combination of dasymetric mapping, vector-to-grid conversion, and other re-gridding methods. We used two sources of road infrastructure mass to include in the atlas. First, gross road materials including tunnels and bridges, were collected as gridded format from Ref. [83]. We re-gridded the gridded road masses to our standardized format using the grid_2_grid tool. The global sum of road materials is estimated at approximately 294.5 billion tonnes. Second, a detailed material specific road mass data were collected at country-level from Ref. [70] by compiling the Global Roads Inventory Project (GRIP) and CIA World Factbook data. Aggregates, asphalts, concerts, gravels, and gross mass were estimated and gridded based on GRIP[84] road density data (https://www.globio.info/download-grip-dataset). This approach yielded a slightly higher global estimate of approximately 349.2 billion tonnes. To distribute the country-level road mass data, we collected the GRIP Road polyline shapefile and using SESAME's line_2_grid tool, we created and stored gridded road length and density data at 0.25 and 1-degree spatial grids. The gridded road density data was used to distribute country-level material masses on roads onto spatial grids. The International Organization of Motor Vehicle Manufacturers (OICA) (https://www.oica.net/category/vehicles-in-use/) provided country-level data on the number of passenger and commercial vehicles from 2005 to 2015. Three distinct vehicle categories (e.g., passenger, commercial, and trailers) and their respective material compositions were estimated[70]. We use three random forest machine learning algorithms to estimate worldwide vehicle counts in three categories based on GDP per capita, total road length, urban population percentage, and year predictors. The models demonstrated high accuracy, with R^2 values greater than 0.94. The material composition of vehicles was approximated using total curb weight, with steel, cast iron, plastic composites, aluminum, rubber, glass, and copper accounting for 80% of the curb weight. Country-level vehicle mass for 2022 was distributed into spatial grids by assuming[85] that 5% of vehicles remained on the road[84], with the remaining 95% dispersed based on population density[86].

For railways and rolling stocks (locomotives, railcars, wagons, and coaches), we used vector railway line data sourced from Ref. [87], which were converted to density grids at 0.25 and 1 degree resolutions using the SESAME's line_2_grid tool. This gridded data was used to distribute the national-level material masses[70] for rail track steel mass, and rolling stocks. Like road infrastructure, we sourced railway mass from two sources. First, the country-level railway lengths data were collected from the United Nations Economic Commission for Europe (UNECE), the CIA World Factbook, and the World Bank's data from the International Union of Railways (UIC). Ref. [70] compiled the data and estimated country-level steel mass for rail track but did not consider materials like wood, steel, or concrete used as railway sleepers, or gravel used as ballast. Later, we distributed the country-level steel mass onto railway[87] densities using the table_2_grid tool. Second, gross mass of railway infrastructure, including railways, underground, elevated and ground level subways, trams, and others like monorails or mountains is sourced from Ref. [83] as raster format and re-gridded into standardized grids. Ref. [70] compiled mass and material compositions of the rolling stocks by reviewing three peer-reviewed studies[88–90] using the number of registered locomotives, railcars, wagons, and coaches were collected from a Union International des Chemins de Fer (UIC) report and their data portal. We proportionally distributed the country-level material compositions (e.g., steel, aluminum, plastic, glass and gross mass) onto railway densities.

Finally, the pipeline for gas and oil lengths was collected from Global Energy Monitor (GEM) and the CIA World Factbook, and final pipeline lengths were estimated using the arithmetic mean between the geometric and arithmetic means[70]. Standard pipeline specifications (e.g., X70/X80 pipelines with an outer diameter of 89 cm and thickness of 1.9 cm; steel density: 7,900 kg/m^3) were used to estimate steel mass per kilometer[91]. The country-level pipeline steel mass was then distributed proportionally to the Global Gas Infrastructure Tracker and Global Oil Infrastructure Tracker pipeline densities, sourced from the GEM (https://globalenergymonitor.org/)[92,93]. The line_2_grid tool from SESAME was used to convert the line shapefile to spatial global grids.

Data processing pipeline. We converted and harmonized 5 major types of spatial data: grid/raster, point, line, polygon and tabular. Four of these involved geospatial-to-geospatial transformations, while the fifth (table to grid) required a downscaling approach. A visual overview of the methods is shown in Fig. 2. The key features of the data processing include:

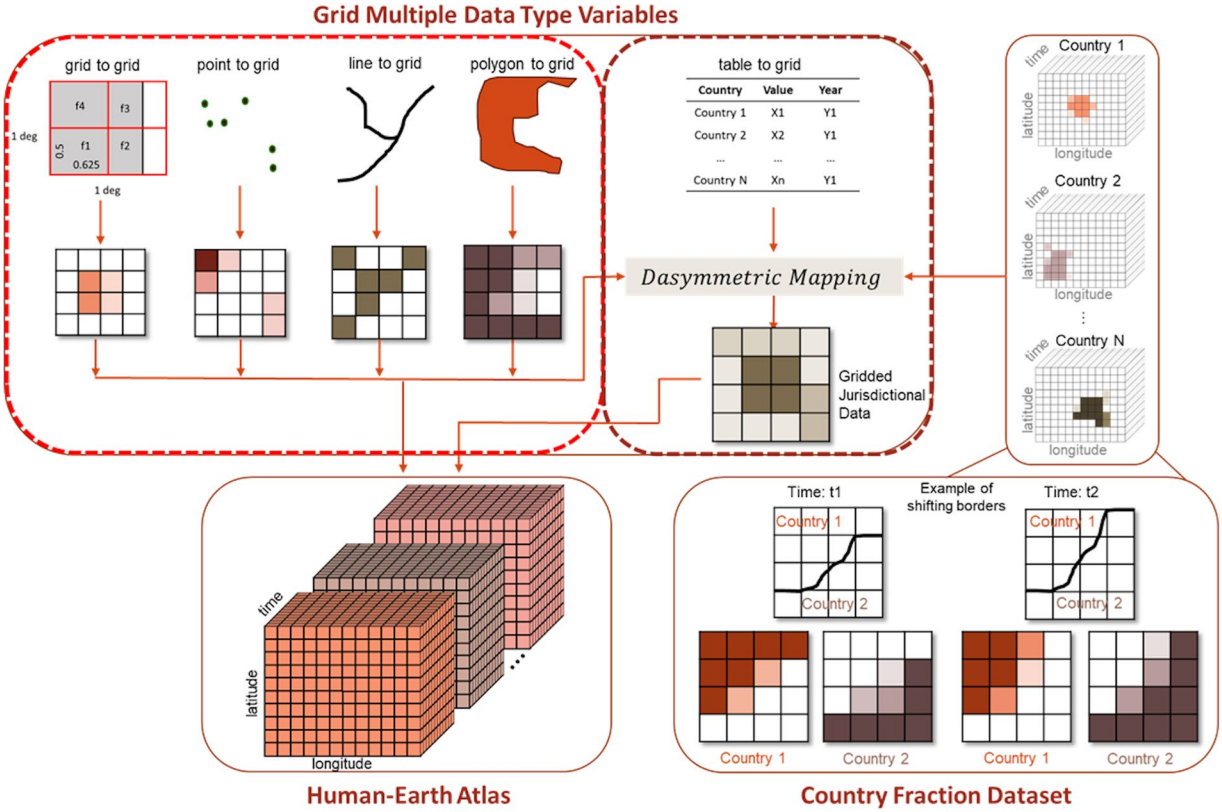

Fig. 2 The overall SESAME data processing flow. An initial variable, which can be one of various data types, is converted to a globally gridded variable, using dasymetric mapping if necessary. Four types of data conversions (grid, point, line, and polygon to grid) involve geospatial to geospatial conversion (light red dotted area), whereas the dasymetric table to grid involves a downscaling approach that requires a country fraction netCDF file and a surrogate variable.

- **Data input:** we used a wide range of data inputs, including spatial (e.g., temperature, land use, precipitation) and tabular human data (e.g., demographic, economic, social) in various structures and resolution.
- **Spatial gridding:** we converted the data automatically into a standardized spatial grid of arbitrary resolution. Jurisdictional (e.g., country-level) tabular human data were allocated to appropriate grid cells based on surrogate variables.
- **Data scrutiny:** we applied subroutines to calculate global statistics (i.e., global sum) to check for conversion errors that might have been caused by inconsistencies in the input data.
- **netCDF standardization:** final output was stored in netCDF format, which is a widely accepted standard for multidimensional time-series data. While preserving the multidimensional nature, this file format allows users to access, analyze, and manipulate large gridded datasets.

Geospatial-to-geospatial gridding. Spatially explicit data can be found in multiple formats, including multidimensional netCDF format (e.g., ERA5 reanalysis products, MERRA-2 climate dataset, and various climate models), raster TIFF format (e.g., WorldPop population data, WorldClim climate data, and global human settlement built-up layers), point format (e.g., mine locations, power plant locations, and various industrial locations), polyline format (e.g., road network, railway tracks, and oil-gas network data), and polygon format (e.g., ecoregions and mining areas). Raster data were converted to the new grid resolution by checking the input's projections, extending the coverage globally, and filling extra/missing grids with NaNs. Grid values were aggregated in each output cell by methods such as sum, mean, max, min, or standard deviation. If the input latitude and longitude resolutions differed (e.g. MERRA-2 is 0.5 by 0.625 degrees), or did not evenly divide the target resolution, the area-weighted fraction was distributed to each grid cell at the specified resolution (see SI). Point values were gridded by counting the number of points in each desired grid cell, summing or averaging their associated values, or grouping them by class to create multi-variable datasets. If a point lay exactly on a grid boundary (which occurs due to round-off), it was shifted by 0.0001 degree in the positive latitude and positive longitude directions to avoid double counting. For lines, the portion of each line that falls within a grid cell was calculated by intersecting the line shapefile with the desired grids. Similarly, for polygons we calculated the area each covers in individual grid cells through intersection, either as a fraction or total area.

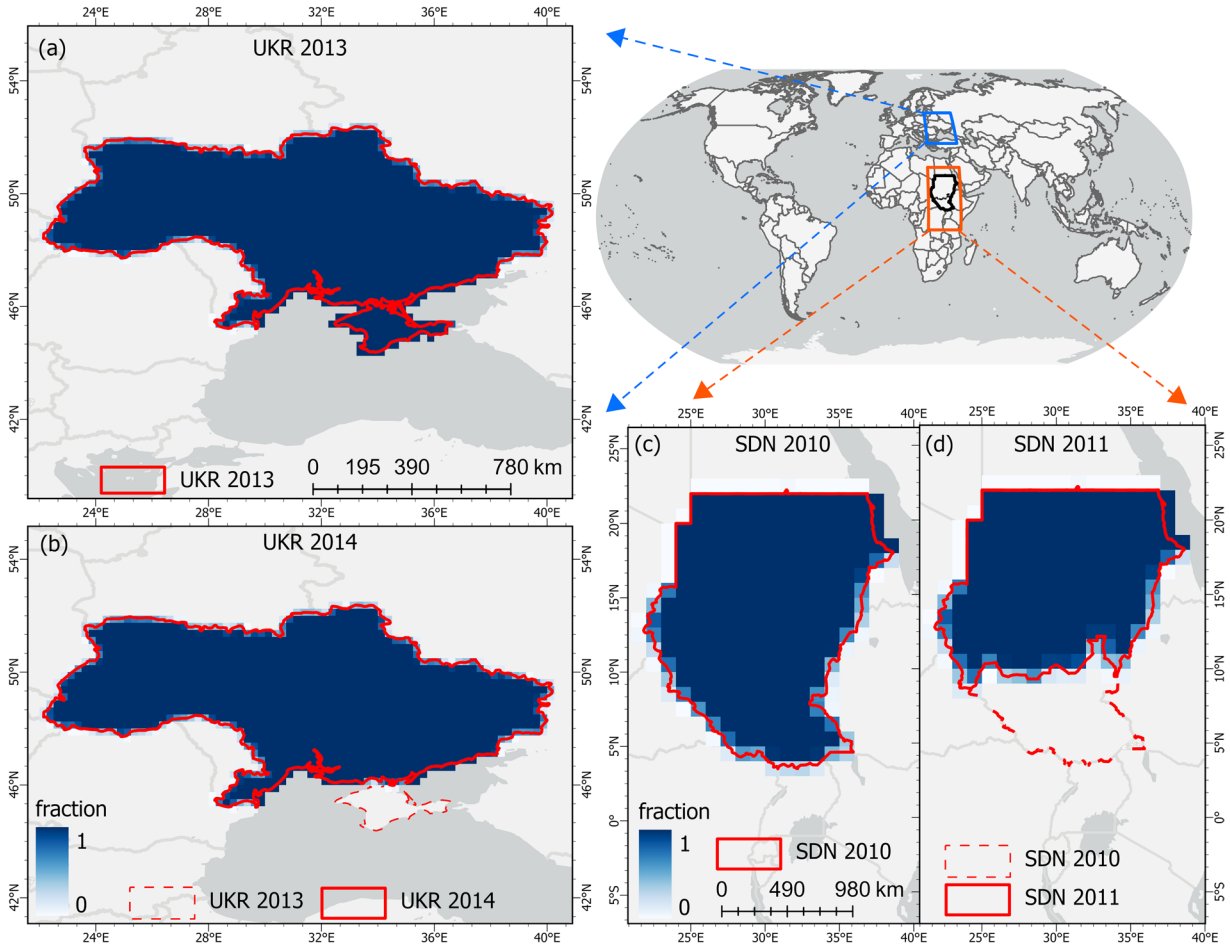

Fig. 3 The country fractions change over time as administrative boundaries change. Crimea was part of Ukraine (UKR) until 2013, as shown in (**a**), but Russia occupied it in 2014 (**b**). South Sudan gained independence in 2011, splitting Sudan (SDN) into two countries. Until 2010, there was only one country SDN (**c**), which became Sudan and South Sudan in 2011 (**d**). Maps in panels (**a,b**) have a resolution of 0.25 degrees, while panels (**c,d**) have a resolution of 1 degree. The country fraction dataset tracks administrative boundary changes over time.

Downscaling tabular jurisdictional data to spatial grids. One of the difficulties of dealing with jurisdictional data is that the jurisdiction names and boundaries change over time. Although these are infrequent occurrences for individual countries, changes in names and/or boundaries of one or more countries have occurred somewhere in the world during most years of the past century. This presents a major hurdle to the construction of long-term gridded time series from tabular country-level data. The SESAME software provides a flexible solution to the problem of boundary changes at the country-level by converting time series of boundary polygons into a netCDF dataset as fractions of historical or extant country names.

The current version makes use of the Schvitz *et al.*[94] administrative boundary shapefile from 1886 to 2019. By applying the poly_2_grid function to this shapefile, we created a netCDF dataset corresponding to the areas associated with each of 246 countries, annually resolved from 2000 to 2023 (not including any administrative boundary changes since 2019). Some country boundaries, such as Sudan, South Sudan, Kosovo, Serbia, Montenegro, Crimea, and others, underwent changes during 2000 to 2023 time period. The country locations are stored as 246 variables, where each variable is named by its ISO3 code. To increase ease of use, an additional function, provided with the software package, called "add_iso3_column" can be used to convert country names to the corresponding ISO3 code. This function recognizes multiple spellings or names for many countries (for example, United States of America, USA, United States, or US) and converts them to the unique ISO3 code (e.g. "USA"). Each variable of the country fraction dataset represents the fraction of the land area within the grid cell that belongs to the country, ignoring non-land area. Thus, where a single country occupies half a grid cell and the other half is ocean, the corresponding country fraction would be 1 and the land-sea fraction 0.5 (Fig. 3). The country fraction netCDF file is supplied with the SESAME package, together with a land-sea mask, which records the fraction of each grid cell that is occupied by land, based on the Copernicus Digital Elevation Model (DEM)[95].

To achieve the downscaling required to produce spatially gridded raster datasets from jurisdiction-level data, we employ the dasymetric mapping method. The dasymetric equation allocates data from jurisdictions to

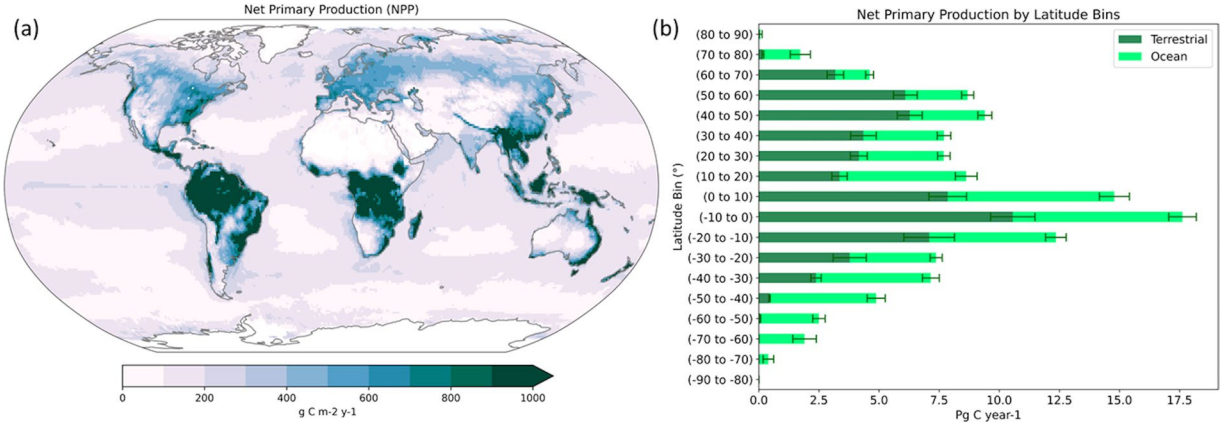

Fig. 4 Standardized terrestrial and ocean net primary productivity (NPP) re-gridded using SESAME's grid_2_grid tool. Terrestrial NPP data was collected from MODIS[48], while ocean NPP was averaged over five model outputs[49]. Panel (**a**) presents both NPPs standardized to the same spatial resolution grid, averaged for the year 2023. Panel (**b**) illustrates the distribution of NPP across 10-degree latitude bins, along with standard deviation, averaged over 2001 to 2023.

grid-cells based on reference variables (referred to as surrogate variables) that are selected to predict the spatial distribution of the quantity at hand. A surrogate variable is a mandatory input variable during the conversion of tabular data to gridded format using SESAME software packages. Based on the surrogate variable, the jurisdictional data is distributed proportionally over the jurisdictional area. If there are no surrogate variable values found for a particular country, the country-level jurisdictional data will be distributed evenly across the grid cells within the jurisdictional border, which is often an acceptable solution for small countries.

The generic dasymetric equation was formulated following Eq. 1. Let X_C be the tabular data associated with jurisdiction C, which we refer to as the target variable. Let $C_{i,j} \in [0,\ 1]$ be the fraction of area of grid-cell *i, j* (i.e. at longitude i and latitude j) under jurisdiction C. Let $x_{i,j}$ be the spatially distributed gridded values of target variable X, over the spatial coordinates *i, j*, which are the desired results of the dasymetric mapping. Let $s_{i,j}$ be the gridded surrogate data variable. The original global surrogate variable is $s_{i,j,obs}$. If tabular data for jurisdiction C (X_C) is available, but no surrogate variable is available within that particular country domain (i.e. $s_{i,j,obs}$ does not exist where $C_{i,j} > 0$), we set the surrogate variable for that jurisdiction everywhere to 1, evenly distributing the tabular value across the jurisdiction to prevent data loss.

We assume that the target jurisdictional variable is proportionally correlated with a specific surrogate variable. Therefore, under the assumption that $x_{i,j}$ is spatially distributed proportionally to $s_{i,j}$, we have $x_{i,j} \propto s_{i,j}$. To estimate grid level jurisdictional data $x_{i,j}$, we perform generalized dasymetric mapping to approximate the spatial distribution of X_C according to the dasymetric equation (1).

for i, j ∈ C:

$$x_{i,j} = X_c \times \frac{s_{i,j}}{\sum_{i,j \in C} C_{i,j} s_{i,j}} \tag{1}$$

where

$$s_{i,j} = \begin{cases} 1 & \text{if } s_{i,j,obs} \text{ does not exit and } X_C \text{ does exit} \\ s_{i,j,obs} & \text{otherwise} \end{cases}$$

As an example, we distributed country-level tabular building mass data using gridded built-up volumes as the surrogate variable. The table_2_grid is a feature of the SESAME toolbox, which utilizes the dasymetric mapping technique. The function requires the use of a country fraction netCDF file that matches with the desired resolution. The SESAME contains a country fraction dataset, representing as $C_{i,j}$, (available for 1, 0.5-, and 0.25-degree resolution) that represents the country's boundary changes over time. Using a surrogate variable from a netCDF or xarrayDataset, the function proportionally distributes tabular country-level data into spatial grids (see SI).

Example datasets. Here, we look at a few examples of Human-Earth Atlas variables, how they are gridded from diverse raw data types, and an example of how the Altas dataset can help us understand the human-Earth system. The code example is available on both SI and Figshare.

An example of Human-Earth Atlas is the data on NPP for both the land and ocean. We collected annual terrestrial NPP data from Running and Zhao (2021)[48] as raster TIFF files, which needed to be re-gridded to the standard 1-degree grid cells of the Atlas. The raw raster metadata contained information about NoData values

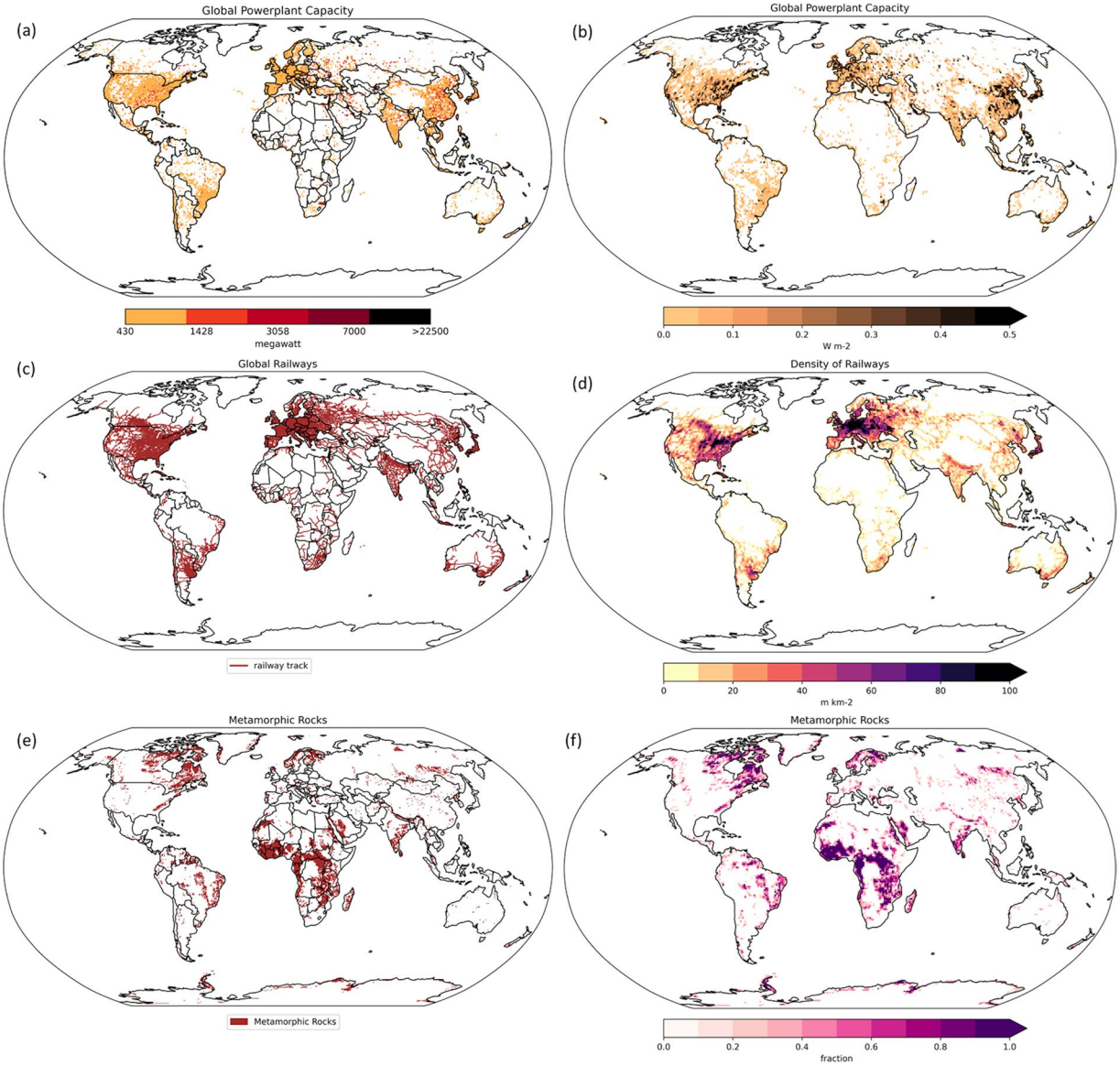

Fig. 5 Examples of various types of vector data stored in the Atlas. Using SESAME's point_2_grid function, point data on (**a**) global power plant locations and capacity[129] is converted to (**b**) standardized gridded power plant capacities in W m-2. Global railways are found in (**c**) polylines format[87], which is then converted to gridded railway track densities in m km-2 using the line_2_grid tool. Different types of rocks (for example, metamorphic rock) are found in (**e**) polygon format[67], which is then converted to the same standardized grids in fractions of rock cover per grid cell (**f**) using the poly_2_grid tool of SESAME. All of the gridded datasets have the same standardized grids, spatial resolution, and projection system, making it simple to perform multi domain research with different data types.

and scale factors, which were checked and converted during the raster data pre-processing step. The raw raster values were checked whether all raster values fall within the valid −30000 to 32700 range. When this function is run after defining additional attribute information such as variable name, long name, units and sources of the data, the grid_2_grid function calculates the mean terrestrial NPP of all high-resolution grid cells within each target grid cell, resulting in the desired 1-degree grid cell resolution dataset (see SI).

The terrestrial NPP data were then combined with annual ocean NPP data, which were collected, re-gridded, and unit-standardized using five different models (as described in Ref. [49]). Before storing the re-gridded raster as a netCDF file, we changed the unit from the original kgC m^{-2} y^{-1} to gC m^{-2} y^{-1} and the grid values were multiplied by the scale factor of 0.0001 to align the storage format of the raw data. Because we have the same spatial grids and units, we combined terrestrial and ocean NPP by multiplying by the land sea fractions to avoid over-estimates in coastal grid cells, as shown in Fig. 4a. The combined data also helps us to quantify the magnitude of NPP contribution in different latitude bins (e.g., a large portion of NPP in +10 to −10), as shown in Fig. 4b.

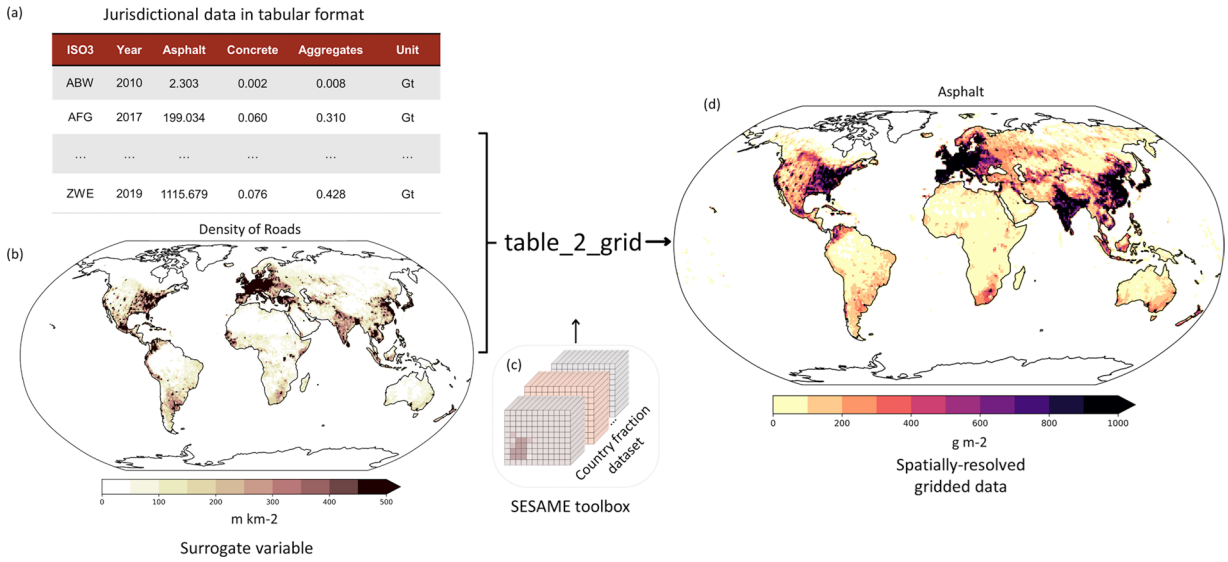

Fig. 6 Schematic diagram of dasymetric mapping using the table_2_grid function of SESAME. The table_2_grid function requires two mandatory files: (**a**) jurisdictional-level tabular data in CSV or pandas DataFrame format, with each country documented as its ISO3 code and associated column values (in this case asphalt), and (**b**) road density dataset as a surrogate variable. The tool uses already generated (**c**) country fraction data stored in SESAME to distribute tabular jurisdictional level asphalt data onto spatial grids proportional to road densities, resulting in (**d**) standardized gridded asphalt data.

We also show example conversion of vector geospatial data to spatial grids, including gridded global power plant capacity, railway density, and fraction of rocks, which were created from point, polyline and polygon dataset, respectively (Fig. 5).

We used country-level road materials data as an example of storing tabular data in spatial grids[70] by utilizing the dasymetric operation. SESAME's table_2_grid function automatically performs this task and requires two mandatory inputs: tabular data and a surrogate variable. We distributed country-level road materials, specifically the mass of asphalt. Since the country fraction data is only stored under ISO3 codes for each country, we converted the country names to their respective ISO3 codes. Then the pre-processed tabular data was provided to the function as pandas DataFrame (Fig. 6a); alternatively, a CSV file can also be used as tabular data. The road density data was used as a surrogate variable (Fig. 6b), which can either be loaded as a netCDF file or as an xarray dataset for this operation. The "tabular_column" argument specified that the DataFrame contains a column named "asphalt" to create gridded asphalt and the units were later converted from gigatons to grams per square meter (Fig. 6c).

Finally, to illustrate the types of relationships that could help our current and future models, we return to the example introduced in Fig. 1. We used the SESAME table_2_grid function to convert the tabular building mass data into standardized grids, using the global human settlement built-up volume as the surrogate variable[76] (Fig. 7a, compare with Fig. 1a). Figure 7b depicts the percentage of building mass that has changed over the last 20 years, revealing how, where, and to what extent urbanization was expanding. Given human (gross building mass) and geophysical data (average maximum wet bulb temperature from Fig. 1b) in the same spatial grid, we can easily compare between the data variables and reveal patterns. The highest wet bulb temperature was determined by first calculating the average monthly climatology of the wet bulb temperature over the period from 2000 to 2024. After obtaining the monthly averages, the peak value was identified by comparing the maximum grid values across all twelve months (see SI). The 3D plot (Fig. 7c) suggests that a considerable number of buildings are being constructed in areas where wet bulb temperature is approaching the threshold of human habitability (~32–35 °C wet bulb temperature), regions that are at risk of extreme heat waves as global temperature rises.

Data Records

The Human-Earth Atlas is publicly available at Figshare[28] (https://doi.org/10.6084/m9.figshare.28432499). It is provided as a series of netCDF files with standardized filenames and follows a hierarchical structure in order to help navigate the database. The first element of each filename is the first letter of the corresponding sphere, e.g., L for lithosphere. We aim to avoid duplicating the same variable in multiple spheres, though at times this may be unavoidable, e.g., moisture content is an important feature of soils, but could also contribute to groundwater, part of the hydrosphere. The subsequent elements indicate the group and sub-group(s) of variables. We also have separate letters for human (Bh) and non-human (B) variables in the Biosphere. A uniform temporal discretization is used for the variables in each file, which is indicated by the last one to two elements of the filename prior to the extension (.nc). For variables that do not change significantly over centennial timescales and are not representative of a particular timeframe, such as lithology, no time is indicated, e.g., L.rock.surface_lithology.nc.

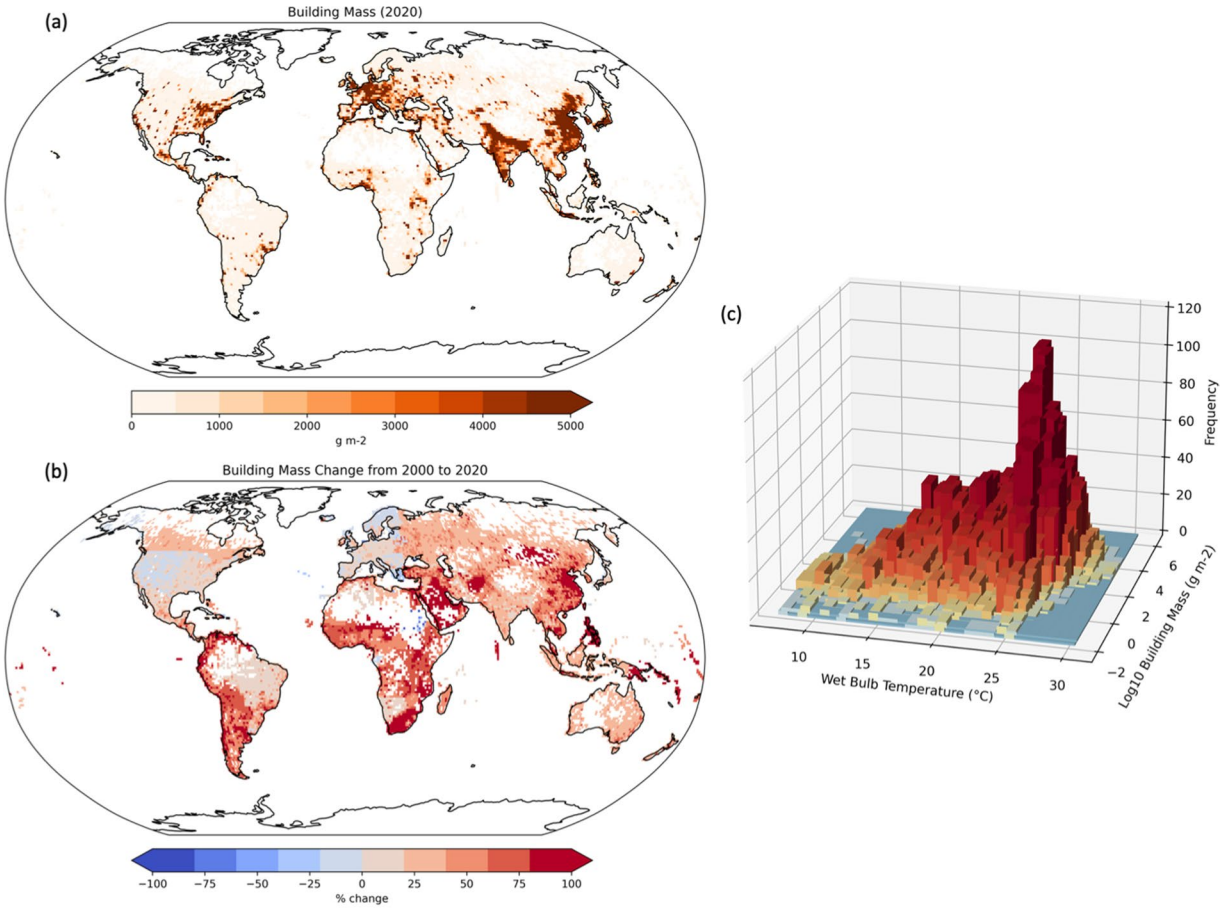

Fig. 7 Unlocking the potential of gridded data. The spatially resolved gridded building mass data (**a**) is derived from dasymetric mapping of GHSL built-up volume, while panel (**b**) shows the change of gridded building mass from 2000 to 2020. The relationship between a human data variable (gross building mass in log10 gm^{-2}) and an Earth system variable (maximum wet bulb temperature, Fig. 1b) is presented in panel c.

For variables that are indicative of a single year, that year is listed, and if more than one value is available for that year it is indicated as daily (d), weekly (w) or monthly (m), e.g., T.technosphere.2021.nc, B.primary_production.2015.w.nc. For files that include values spanning multiple years, the range of years is listed, and it is specified whether they are annual (a), weekly (w) or monthly (m), e.g., T.transportation.merchant_fleet.2011–2021.a.nc. For climatological averages (i.e., average seasonal cycle over multiple years) this is indicated by "clim" followed by the specification of daily (d) weekly (w) or monthly (m), e.g. A.surface_weather.2000–2020.clim_m.nc. A short summary of available files included at the time of writing are listed in Tables 2–6, and detailed description of variable sources, preprocessing, operations are found in SI.

Technical Validation

After downloading the data, we conducted an initial quality check by visually inspecting the mapped raw data. Heterogeneous representations for missing values (e.g., −9999, or any "FillValue" mentioned in the raw data) were assessed and the missing values were converted to NaN. After re-gridding, we calculated and compared global summary statistics for all grid cells between the original data sources and the re-gridded output to check for unintended data loss or processing errors. We required that the global sum be exactly reproduced after re-gridding within floating-point precision. Subsequently, we visually inspected gridded products, and in rare cases where unrealistic extreme values were present (e.g., very high NPP in two grid-cells in Sub-Sharan Africa, very high nighttime light values in very high latitude and random ocean areas), the extreme values were excluded. While dasymetric mapping is a well-established concept, the assessment of uncertainty remains challenging, and there do not yet exist standardized methods. The gridding process cannot easily integrate the uncertainties inherent in each dataset.

Usage Notes

We provide the Human-Earth Atlas data in netCDF format along with the scripts used to generate the dataset. The detailed documentation files describing the current version of the Atlas are available on Figshare[28].

Code availability

The open-source SESAME Python software, which was used to generate the Human-Earth Atlas[28], is available on GitHub (https://github.com/A2Faisal/SESAME) and PyPI. The SESAME toolkit offers five major functions—grid_2_grid, point_2_grid, line_2_grid, poly_2_grid, and table_2_grid—which can conveniently be used to grid a wide range of input data formats. In addition to the core functions, SESAME provides an extensive sets of utility tools to support data processing workflows. These include the add_iso3_column function to interpret diverse country names as standard ISO3 codes, and the grid_2_table function which reverses the gridding process by aggregating gridded data back into summary tables by regions or countries. SESAME also includes quick visualization tools (e.g., plot_histogram, plot_scatter, plot_time_series, plot_hexbin), and mapping functions (plot_map, plot_country), enabling users to generate exploratory charts and publication-ready gridded or choropleth maps with minimal effort. Furthermore, SESAME provides several simple operations, such as summation, subtraction, and averaging, which can be applied to grid variables. The SESAME GitHub repository provides full documentation, installation steps, and worked examples, making it a comprehensive and user-friendly solution for transforming and visualizing global datasets in a unified geographic format.

Received: 3 September 2024; Accepted: 29 April 2025;
Published online: 12 May 2025

References

1. Steffen, W. *et al*. The emergence and evolution of Earth System Science. *Nature Reviews Earth & Environment* **1**, 54–63 (2020).
2. Jacobson, M., Charlson, R. J., Rodhe, H. & Orians, G. H. *Earth System Science: From Biogeochemical Cycles to Global Changes*. (Academic Press, 2000).
3. Reichstein, M. *et al*. Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).
4. Reid, W. V. *et al*. Earth system science for global sustainability: grand challenges. *Science* **330**, 916–917 (2010).
5. Shao, Y. *et al*. Dust cycle: An emerging core theme in Earth system science. *Aeolian Research* **2**, 181–204 (2011).
6. Schellnhuber, H. J. Earth system' analysis and the second Copernican revolution. *Nature* **402**, C19–C23 (1999).
7. Snow, C. P. *The Rede Lecture 1959*. (Cambridge University Press Cambridge, 1959).
8. Galbraith, E. D. Earth system economics: a biophysical approach to the human component of the Earth system. *Earth System Dynamics* **12**, 671–687 (2021).
9. Earth Science Data Systems, N. Nighttime Lights | NASA Earthdata. https://www.earthdata.nasa.gov/topics/human-dimensions/nighttime-lights (2024).
10. Kroodsma, D. A. *et al*. Tracking the global footprint of fisheries. *Science* **359**, 904–908 (2018).
11. Qin, X., Wu, B., Zeng, H., Zhang, M. & Tian, F. Global Gridded Crop Production Dataset at 10 km Resolution from 2010 to 2020. *Sci Data* **11**, 1377 (2024).
12. Newth, D. & Lenton, A. Human-Earth-System framework for assessing Carbon Dioxide Removal strategies – CarbonLock. https://research.csiro.au/carbonlock/human-earth-system/.
13. Mu, H. *et al*. A global record of annual terrestrial Human Footprint dataset from 2000 to 2018. *Sci Data* **9**, 176 (2022).
14. Wang, Y. *et al*. GriddingMachine, a database and software for Earth system modeling at global and regional scales. *Scientific data* **9**, 258 (2022).
15. Baumann, P. *et al*. Big data analytics for earth sciences: the EarthServer approach. *International journal of digital earth* **9**, 3–29 (2016).
16. Gorelick, N. *et al*. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment* **202**, 18–27 (2017).
17. Mahecha, M. D. *et al*. Earth system data cubes unravel global multivariate dynamics. *Earth System Dynamics* **11**, 201–234 (2020).
18. Atakan, K. *et al*. National EPOS initiatives and participation to the EPOS integration plan. *Annals of Geophysics* (2022).
19. De Smedt, K., Koureas, D. & Wittenburg, P. FAIR digital objects for science: From data pieces to actionable knowledge units. *Publications* **8**, 21 (2020).
20. Baynes, J., Neale, A. & Hultgren, T. Improving intelligent dasymetric mapping population density estimates at 30 m resolution for the conterminous United States by excluding uninhabited areas. *Earth Syst Sci Data* **14**, 2833–2849 (2022).
21. Mennis, J. Generating surface models of population using dasymetric mapping. *The Professional Geographer* **55**, 31–42 (2003).
22. Eicher, C. L. & Brewer, C. A. Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography and Geographic Information Science* **28**, 125–138 (2001).
23. Mitsova, D., Esnard, A.-M. & Li, Y. Using enhanced dasymetric mapping techniques to improve the spatial accuracy of sea level rise vulnerability assessments. *Journal of Coastal Conservation* **16**, 355–372 (2012).
24. Liu, L., Peng, Z., Wu, H., Jiao, H. & Yu, Y. Exploring urban spatial feature with dasymetric mapping based on mobile phone data and LUR-2SFCAe method. *Sustainability* **10**, 2432 (2018).
25. Poulsen, E. & Kennedy, L. W. Using dasymetric mapping for spatially aggregated crime data. *Journal of Quantitative Criminology* **20**, 243–262 (2004).
26. Barrozo, L. V., Pérez-Machado, R. P., Small, C. & Cabral-Miranda, W. Changing spatial perception: dasymetric mapping to improve analysis of health outcomes in a megacity. *Journal of Maps* **12**, 1242–1247 (2016).
27. Requia, W. J., Koutrakis, P. & Arain, A. Modeling spatial distribution of population for environmental epidemiological studies: Comparing the exposure estimates using choropleth versus dasymetric mapping. *Environment international* **119**, 152–164 (2018).
28. Faisal, A. A., Kaye, M., Ahmed, M. & Galbraith, E. The SESAME Human-Earth Atlas. *figshare* https://doi.org/10.6084/m9.figshare.28432499 (2025).
29. Wilkinson, M. D. *et al*. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016).
30. *Next Generation Earth Systems Science at the National Science Foundation*. https://doi.org/10.17226/26042 (National Academies Press, Washington, D.C., 2022).
31. Galbraith, E. *et al*. Resolving the Technosphere. *EGUsphere* 1–25, https://doi.org/10.5194/egusphere-2024-1133 (2024).
32. ERA5. Copernicus Climate Change Service (C3 S)(2017). Copernicus Climate Change Service Climate Data Store: Fifth generation of ECMWF atmospheric reanalyses of the global climate. (2017).
33. Hersbach, H. *et al*. ERA5 monthly averaged data on single levels from 1940 to present [Dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/cds.f17050d7 (2023).
34. Global Modeling and Assimilation Office (GMAO). GES DISC Dataset: MERRA-2 tavgU_2d_flx_Nx: 2d,diurnal,Time-Averaged,Single-Level,Assimilation, Surface Flux Diagnostics V5.12.4 (M2TUNXFLX 5.12.4). https://doi.org/10.5067/LUHPNWAKYIO3.
35. Stull, R. Wet-Bulb Temperature from Relative Humidity and Air Temperature. *Journal of Applied Meteorology and Climatology* **50**, 2267–2269 (2011).

36. Errors in MERRA-2 T2MWET (wet-bulb temperature) - Earthdata Forum. https://forum.earthdata.nasa.gov/viewtopic.php?t=3333.
37. Global Modeling and Assimilation Office (GMAO). MERRA-2 tavgM_2d_flx_Nx: 2d, Monthly mean,Time-Averaged, Single-Level, Assimilation, Surface Flux Diagnostics V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [2024-03-13], https://doi.org/10.5067/0JRLVL8YV2Y4 (2015).
38. Nasa/Larc/Sd/Asdc. CERES Energy Balanced and Filled (EBAF) TOA and Surface Monthly means data in netCDF Edition 4.2. (2023).
39. Nasa/Larc/Sd/Asdc. MISR Level 3 Component Global Aerosol product in netCDF format covering a month V004. https://doi.org/10.5067/Terra/MISR/MIL3MAEN_L3.004 (2008).
40. Nasa/Larc/Sd/Asdc. CERES Energy Balanced and Filled (EBAF) TOA and Surface Monthly means data in netCDF Edition 4.1. (2019).
41. WorldPop. Global 1km Population. University of Southampton https://doi.org/10.5258/SOTON/WP00647 (2018).
42. Wang, T. & Sun, F. Global gridded GDP data set consistent with the shared socioeconomic pathways. *Scientific data* **9**, 221 (2022).
43. Kennedy, C. M., Oakleaf, J. R., Theobald, D. M., Baruch-Mordo, S. & Kiesecker, J. Managing the middle: A shift in conservation priorities based on the global human modification gradient. *Global Change Biology* **25**, 811–826 (2019).
44. Guiet, J., Bianchi, D., Scherrer, K. J. N., Heneghan, R. F. & Galbraith, E. D. BOATSv2: New ecological and economic features improve simulations of High Seas catch and effort. *Geoscientific Model Development Discussions* 1–48, https://doi.org/10.5194/gmd-2024-26 (2024).
45. Grogan, D. *et al*. Global gridded crop harvested area, production, yield, and monthly physical area data circa 2015. *Scientific Data 2022 9:1* **9**, (2022).
46. Gilbert, M. *et al*. Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. *Sci Data* **5**, 180227 (2018).
47. Kaye, M., MacDonald, G. K. & Galbraith, E. Energetic closure of the spatially resolved global food system. Preprint at https://doi.org/10.48550/arXiv.2412.10421 (2024).
48. Running, S. W. & Zhao, M. MODIS/Terra Net Primary Production Gap-Filled Yearly L4 Global 500m SIN Grid V061. https://doi.org/10.5067/MODIS/MOD17A3HGF.061 (2021).
49. Ryan-Keogh, T. J., Thomalla, S. J., Chang, N. & Moalusi, T. A new global oceanic multi-model net primary productivity data product. *Earth System Science Data Discussions* **2023**, 1–34 (2023).
50. Dinerstein, E. *et al*. An ecoregion-based approach to protecting half the terrestrial realm. *BioScience* **67**, 534–545 (2017).
51. IUCN. The IUCN Red List of Threatened Species. Version 2024-2. (2024).
52. Hansen, M. C. *et al*. High-Resolution Global Maps of 21st-Century Forest Cover Change. *American Association for the Advancement of Science* https://doi.org/10.1126/science.1244693 (2013).
53. Simard, M., Pinto, N., Fisher, J. B. & Baccini, A. Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research: Biogeosciences* **116**, (2011).
54. Spawn, S. A., Sullivan, C. C., Lark, T. J. & Gibbs, H. K. Harmonized global maps of above and belowground biomass carbon density in the year 2010. *Sci Data* **7**, 112 (2020).
55. Rayner, N. *et al*. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres* **108**, (2003).
56. Locarnini, R. A. *et al*. World Ocean Atlas 2023, Volume 1: Temperature.
57. Reagan, J. R. *et al*. World Ocean Atlas 2023, Volume 2: Salinity.
58. Garcia, H. E. *et al*. World Ocean Atlas 2023, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, Dissolved Oxygen Saturation and 30-year Climate Normal.
59. Garcia, H. E. *et al*. World Ocean Atlas 2023, Volume 4: Dissolved Inorganic Nutrients (Phosphate, Nitrate, and Silicate).
60. Johnson, G. C. & Lyman, J. M. A global ocean surface mixed layer monthly climatology: Means, percentiles, skewness, and kurtosis.
61. O'Reilly, J. E. *et al*. Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research: Oceans* **103**, 24937–24953 (1998).
62. Wang, M., Son, S. & Harding, L. W. Jr. Retrieval of diffuse attenuation coefficient in the Chesapeake Bay and turbid ocean regions for satellite ocean color applications. *Journal of Geophysical Research: Oceans* **114**, (2009).
63. Lehner, B. & Grill, G. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes* **27**, 2171–2186 (2013).
64. Messager, M. L., Lehner, B., Grill, G., Nedeva, I. & Schmitt, O. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nat Commun* **7**, 13603 (2016).
65. Barbarossa, V. *et al*. FLO1K, global maps of mean, maximum and minimum annual streamflow at 1 km resolution from 1960 through 2015. *Sci Data* **5**, 180052 (2018).
66. Ehalt Macedo, H. *et al*. Distribution and characteristics of wastewater treatment plants within the global river network. *Earth System Science Data* **14**, 559–577 (2022).
67. Hartmann, J. & Moosdorf, N. The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems* **13**, (2012).
68. Poggio, L. *et al*. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *SOIL* **7**, 217–240 (2021).
69. Turek, M. E. *et al*. Global mapping of volumetric water retention at 100, 330 and 15 000 cm suction using the WoSIS database. *International Soil and Water Conservation Research* **11**, 225–239 (2023).
70. Matitia, T. Global end-use inventory of the anthropogenic mass. *Earth & Planetary Sciences* vol. Masters of Science (McGill University, Montréal, Québec, Canada, 2022).
71. Friedl, M. & Sulla-Menashe, D. MODIS/Terra+ Aqua land cover type yearly L3 global 500m SIN grid V061. *NASA EOSDIS Land Processes DAAC* (2022).
72. Open data @ OurAirports. https://ourairports.com/data/.
73. Jemiolo, W. Life cycle assessment of current aned future passenger air transport in Switzerland. (Universitetet i Nordland, 2015).
74. Kong, X. *et al*. Steel stocks and flows of global merchant fleets as material base of international trade from 1980 to 2050. *Global Environmental Change* **73**, 102493 (2022).
75. Global Shipping Traffic Density | Data Catalog. https://datacatalog.worldbank.org/search/dataset/0037580/Global-Shipping-Traffic-Density.
76. Schiavina, M. *et al*. GHSL data package 2023. *Publications Office of the European Union: Luxembourg* (2023).
77. Zhao, M. *et al*. A global dataset of annual urban extents (1992–2020) from harmonized nighttime lights. *Earth System Science Data Discussions* **2021**, 1–25 (2021).
78. Deetman, S., de Boer, H. S., Van Engelenburg, M., van Der Voet, E. & van Vuuren, D. P. Projected material requirements for the global electricity infrastructure - generation, transmission and storage. *Resour Conserv Recy* **164**, (2021).
79. Marinova, S., Deetman, S., van der Voet, E. & Daioglou, V. Global construction materials database and stock analysis of residential buildings between 1970-2050. *Journal of Cleaner Production* **247**, 119146 (2020).
80. Zhao, M. *et al*. A global dataset of annual urban extents (1992-2020) from harmonized nighttime lights. *Earth Syst Sci Data* **14**, 517–534 (2022).
81. Wiedenhofer, D. *et al*. From Extraction to End-uses and Waste Management: Modelling Economy-wide Material Cycles and Stock Dynamics Around the World. SSRN Scholarly Paper at https://doi.org/10.2139/ssrn.4794611 (2024).

82. Haberl, H. *et al*. Weighing the global built environment: High-resolution mapping and quantification of material stocks in buildings. *Journal of Industrial Ecology* https://doi.org/10.1111/jiec.13585.
83. Wiedenhofer, D. *et al*. Mapping and modelling global mobility infrastructure stocks, material flows and their embodied greenhouse gas emissions. *Journal of Cleaner Production* **434**, 139742 (2024).
84. Meijer, J. R., Huijbregts, M. A. J., Schotten, K. C. G. J. & Schipper, A. M. Global patterns of current and future road infrastructure. *Environ. Res. Lett.* **13**, 064006 (2018).
85. Morris, D. Z. Want to know why Uber and automation really matter? Here's your answer. *Fortune* https://fortune.com/2016/03/13/cars-parked-95-percent-of-time/.
86. Bondarenko, M. Global 1km Population. (2018).
87. Programme, W. F. Global railways (WFP SDI-T - Logistics Database). (2017).
88. Delogu, M., Del Pero, F., Berzi, L., Pierini, M. & Bonaffini, D. End-of-Life in the railway sector: Analysis of recyclability and recoverability for different vehicle case studies. *Waste management* **60**, 439–450 (2017).
89. Harvey, L. D. Reconciling global iron and steel mass flow datasets, with an update to 2011–2015 and an assessment of uncertainty in global end-of-life scrap flow. *Resources, Conservation and Recycling* **182**, 106281 (2022).
90. Kaewunruen, S. & Rungskunroch, P. A through-life evaluation of end-of-life rolling stocks considering asset recycling, energy recovering, and financial benefit. *Journal of cleaner production* **212**, 1008–1024 (2019).
91. Bai, Q. & Bai, Y. *Subsea Pipeline Design, Analysis, and Installation*. (Gulf Professional Publishing, 2014).
92. Global Oil Infrastructure Tracker. *Global Energy Monitor* https://globalenergymonitor.org/projects/global-oil-infrastructure-tracker/ (2022).
93. Global Gas Infrastructure Tracker. *Global Energy Monitor* https://globalenergymonitor.org/projects/global-gas-infrastructure-tracker/ (2023).
94. Schvitz, G. *et al*. Mapping the International System, 1886–2019: The CShapes 2.0 Dataset. *J Conflict Resolut* **66**, 144–161 (2022).
95. Copernicus Digital Elevation Model. *Copernicus Contributing Missions Online* https://spacedata.copernicus.eu/collections/copernicus-digital-elevation-model.
96. Muñoz-Sabater, J. *et al*. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. https://doi.org/10.24381/cds.68d2bb30 (2021).
97. My NASA Data. Data Collections: Earth System Data Explorer | My NASA Data. My NASA Data (2019).
98. Global, M., Assimilation, O. & Pawson, S. MERRA-2 tavgM_2d_flx_Nx: 2d,Monthly mean,Time-Averaged,Single-Level,Assimilation, Surface Flux Diagnostics V5.12.4. [object Object] https://doi.org/10.5067/0JRLVL8YV2Y4 (2015).
99. Wang, T. & Sun, F. Gross domestic product (GDP) downscaling: a global gridded dataset consistent with the Shared Socioeconomic Pathways. *Zenodo* https://doi.org/10.5281/zenodo.5880037 (2022).
100. Gilbert, M. *et al*. Global chickens distribution in 2015 (5 minutes of arc). *Harvard Dataverse* https://doi.org/10.7910/DVN/SXHLF3 (2022).
101. Gilbert, M. *et al*. Global sheep distribution in 2015 (5 minutes of arc). *Harvard Dataverse* https://doi.org/10.7910/DVN/VZOYHM (2022).
102. Gilbert, M. *et al*. Global pigs distribution in 2015 (5 minutes of arc). *Harvard Dataverse* https://doi.org/10.7910/DVN/CIVCPB (2022).
103. Gilbert, M. *et al*. Global horses distribution in 2015 (5 minutes of arc). *Harvard Dataverse* https://doi.org/10.7910/DVN/JJGCTX (2022).
104. Gilbert, M. *et al*. Global goats distribution in 2015 (5 minutes of arc). *Harvard Dataverse* https://doi.org/10.7910/DVN/YYG6ET (2022).
105. Gilbert, M. *et al*. Global ducks distribution in 2015 (5 minutes of arc). *Harvard Dataverse* https://doi.org/10.7910/DVN/S9ONXV (2022).
106. Gilbert, M. *et al*. Global buffaloes distribution in 2015 (5 minutes of arc). *Harvard Dataverse* https://doi.org/10.7910/DVN/I1WCAB (2022).
107. Gilbert, M. *et al*. Global cattle distribution in 2015 (5 minutes of arc). *Harvard Dataverse* https://doi.org/10.7910/DVN/LHBICE (2022).
108. Frolking, S., Wisser, D., Grogan, D., Proussevitch, A. & Glidden, S. GAEZ+_2015 Crop Harvest Area. *Harvard Dataverse* https://doi.org/10.7910/DVN/KAGRFI (2020).
109. Frolking, S., Wisser, D., Grogan, D., Proussevitch, A. & Glidden, S. GAEZ+_2015 Crop Production. *Harvard Dataverse* https://doi.org/10.7910/DVN/KJFUO1 (2020).
110. Guiet, J. *et al*. BOATSv2 and dataset for 'BOATSv2: New ecological and economic features improve simulations of High Seas catch and effort'. *Zenodo* https://doi.org/10.5281/zenodo.11043334 (2024).
111. IUCN. The IUCN Red List of Threatened Species.
112. Mu, H. *et al*. An annual global terrestrial Human Footprint dataset from 2000 to 2018. https://doi.org/10.6084/m9.figshare.16571064.v7 (2021).
113. Spawn, S. A. & Gibbs, H. K. Global Aboveground and Belowground Biomass Carbon Density Maps for the Year 2010. https://doi.org/10.3334/ORNLDAAC/1763 (2020).
114. Simard, M., Pinto, N., Fisher, J. B. & Baccini, A. Global Forest Canopy Height, 2005 | Earth Engine Data Catalog.
115. Friedl, M. & Sulla-Menashe, D. MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V061. *NASA EOSDIS Land Processes Distributed Active Archive Center* https://doi.org/10.5067/MODIS/MCD12Q1.061 (2022).
116. Hartmann, J. & Moosdorf, N. Global Lithological Map Database v1.0 (gridded to 0.5° spatial resolution). *Supplement to: Hartmann, Jens; Moosdorf, Nils (2012): The new global lithological map database GLiM: A representation of rock properties at the Earth surface. Geochemistry, Geophysics, Geosystems, 13, Q12004*, 10.1029/2012GC004370 PANGAEA https://doi.org/10.1594/PANGAEA.788537 (2012).
117. UNCTADstat Data Centre. https://unctadstat.unctad.org/datacentre/.
118. National air transport system - The World Factbook. https://www.cia.gov/the-world-factbook/field/national-air-transport-system/.
119. Pipelines - The World Factbook. https://www.cia.gov/the-world-factbook/field/pipelines/.
120. Railways - The World Factbook. https://www.cia.gov/the-world-factbook/field/railways/.
121. Total Length of Lines Operated (km) at 31 December by Type of Track,Country and Year. *United Nations Economic Commission for Europe* https://w3.unece.org/PXWeb2015PXWeb2015/pxweb/en/STAT/STAT__40-TRTRANS__11-TRINFRA/ZZZ_en_TRRailInfra1_r.px/.
122. World Bank Open Data. *World Bank Open Data* https://data.worldbank.org.
123. RAILISA STAT UIC. https://uic-stats.uic.org/select/.
124. Deetman, S. *et al*. Modelling global material stocks and flows for residential and service sector buildings towards 2050. *Journal of Cleaner Production* **245**, 118658 (2020).
125. Vehicles in use | www.oica.net. https://www.oica.net/category/vehicles-in-use/.
126. OICA. https://www.oica.net/production-statistics/.
127. FAOSTAT - Machinery. https://www.fao.org/faostat/en/#data/RM.
128. Li, X., Zhou, Y., zhao, M. & Zhao, X. Harmonization of DMSP and VIIRS nighttime light data from 1992–2021 at the global scale. https://doi.org/10.6084/m9.figshare.9828827.v8 (2020).
129. Global Energy Observatory, G., KTH Royal Institute of Technology in Stockholm, Enipedia, World Resources Institute. Global power plant database. Published on Resource Watch and Google Earth Engine. (2018).

Acknowledgements
We thank Tanya Matitia for her compilation of material masses of the technosphere and Maya Willard-Stepan for her contributions to the early stage of SESAME development. We also thank research organizations, particularly NASA, Copernicus, Google Earth Engine, and individuals for their contributions to the research community through open-source data. We acknowledge the funding agencies, including the Canada Research Chairs Program, the Vanier Canada Graduate Scholarships (Vanier CGS) program supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the NSERC Discovery Grant. We acknowledge ChatGPT and Microsoft Bing AI (copilot) for their assistance with a variety of needs.

Author contributions
A.A.F. — conceptualization, visualization, developed toolboxes, methodology, formal analysis, writing - original draft, writing - review and editing, and documentation. M.K. — conceptualization, editing toolboxes, writing - review and editing, and documentation. M.A. — editing and reviewing toolboxes and documentation. E.D.G. – conceptualization, supervision, data curation, writing - original draft, writing - review and editing.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-025-05087-5.

Correspondence and requests for materials should be addressed to A.A.F.