Technical Review of

Google Environmental Insights Explorer

Data for

Local Greenhouse Gas Inventories

ICLEI USA
Local Governments for Sustainability

August 2019

Written by ICLEI - Local Governments for Sustainability USA

Authors
Michael Steinhoff, (former) ICLEI USA
Eli Yewdall, ICLEI USA
Tom Herrod, ICLEI USA

Google supported the preparation of this document, in an effort to improve the Environmental Insights Explorer as a tool for climate action.

Executive Summary

This document provides a technical review of the city-level greenhouse gas (GHG) inventory data provided by the Google Environmental Insights Explorer (EIE), in comparison to other data sources commonly used in local community-scale GHG inventories. Recognizing that at this time EIE only targets a subset of GHG inventory activity categories (albeit inclusive of the largest sectors: transportation and buildings), EIE has strong potential for streamlining GHG inventory processes, while also providing some data advantages for planning and performance measurement over existing data sources.

Specifically, EIE can provide:

- Transportation data based on continuous observation: EIE data provides a more reliable and readily available indication of year-to-year change in on-road transportation activity than do most existing data sources, which must rely on lengthy model runs and recalibration cycles.
- Multi-modal transportation data: EIE includes data not only regarding on-road vehicles, but also pedestrian, bicycle, and transit usage. This more complete picture allows for more-effective planning for a transportation system that prioritizes active and transit modes, and the ability to measure shifts in mode share over time.
- Data on building size and usage: EIE provides the data for total floor space of residential and non-residential buildings, which can be useful indicators for understanding emissions changes.

At present, EIE must be combined with other data sources in order to create a complete inventory of their city emissions, such as under either the U.S. Community Protocol or GPC. In addition, most cities will find that utility-provided data on building energy usage is currently more accurate than that provided by EIE, although continued improvements in EIE are expected.

However, since transportation and buildings are usually responsible for the majority of community emissions, for cities just getting started with climate action, the data contained within EIE can provide a starting point for cities beginning to explore climate action. Other cities more established in their GHG inventory and planning will find value in combining or overlaying EIE outputs, particularly those for transportation, with other inventory data sources.

Introduction

The primary purposes of local community GHG inventories are to prioritize emissions-reduction actions, and to support performance tracking of emissions over time. While existing inventory approaches and data sources do provide considerable valuable information for both of these goals, there are also limitations to what the data can show. Further, even advanced communities with well-established processes and resources for tracking encounter challenges.

Much of the current state of practice evolved around creative use of data sets that were close enough for estimating emissions, even if they lacked the ability to meaningfully describe the GHG generating processes in a useful way. Just within the U.S. context, highly modeled VMT data that was designed to inform forward-looking planning decisions became a standard source for backward-looking inventory data because it was readily available in many places.

Past efforts by state and federal governments to supply estimated data to ease the burden of inventory preparation can be beneficial for informing a general idea of the scale of the challenge as a starting point for cities—but they can be shortsighted in terms of their utility for informing local decision making and getting cities on pathways to performance management using data that is capable of signaling local trends.

Within this context, there are few examples of mechanisms to improve data quality over time and at a scale that benefits a wide range of cities across geographies and typologies. The Google Environmental Insights Explorer and related initiatives have the opportunity to begin to change an otherwise intractable problem. Recognizing that at this time EIE only targets a subset of GHG inventory activity categories and that further refinement is always desirable, EIE has strong potential for streamlining many GHG inventory processes, allowing local government staff to focus more attention on driving action, and reducing the cost and personnel burden of inventory preparation for disclosure.

Beyond the benefits to local government users, EIE can also benefit the consumers of inventory data. As a standardized input data source, any shortcomings about how well this particular resource accurately captures local variation can be more easily communicated and interpreted by anyone looking to analyze aggregate disclosure data.

Findings from the Technical Review

The state of information for community-scale GHG performance data has remained static from the first wave of methodologies applied to this emerging field. The Google Environmental Insights Explorer represents the biggest leap forward in new approaches to the development of activity data for GHG performance management and climate action planning. Currently, there exists no equal in it's potential to have a consistent measure of transportation data across geographic and temporal scales. The contextual information provided by measures of the size of the built environment and active transportation modes has many potential applications for improved local decision making around climate-related topics.

The following two sections detail the findings of the technical review as it pertains to using EIE for creating an emissions inventory, and the potential that EIE provides for action planning and performance tracking.

Opportunities for inventories

The "traditional" data used by many communities did not have to go through much deep questioning by virtue of being first available, although, as illustrated in this review, those data sources are far from perfect. While any particular data source may have additional benefits and drawbacks not listed here, the overall sense of the attributes of the available options should be useful to practitioners making general decisions about the approaches they wish to take for their GHG inventories in light of their goals.

Within the many different combinations of stakeholders and data sources that cities have available, the data in the Environmental Insights Explorer today offers advantages over many other options. The most compelling aspect of EIE is that it has significant evolution potential where other data sources generally do not.

Transportation: EIE offers transportation inventory data with the policy relevance of origin destination data, comparable to typical outputs of an agent-based transportation model. But EIE has the advantage of continuous measurement, providing a more accurate indicator of year-to-year changes in transportation activity.

Comparison of small-area estimates from EIE to road counts at specific locations have shown that EIE is accurate in terms of its ability to measure within small areas. There is no reason to believe that this accuracy would diminish when scaling up to the geographic extent of a city.

Buildings: EIE buildings data provides a useful addition for many cities by providing square footage of residential and non-residential buildings (some cities already have this data, but many do not, particularly for residential buildings). In addition, EIE building-energy-usage data could be used as a backup for cities unable to get local utility data on energy use. If the source of performance data could be improved, EIE building-energy-usage data would be valuable for a much wider range of cities.

The current limitation of EIE building data is that it relies on building performance data from CURB: Climate Action for Urban Sustainability (developed by the World Bank in 2016), which may vary widely from actual local building performance. City benchmarking and disclosure activities offer one of the most promising sources of more-accurate local performance data. Benchmarking and disclosure activities in cities are unlikely to provide complete coverage of cities in the foreseeable future. However, those activities could be designed to obtain representative samples of city performance that captures the unique aspects of the building stock and climate for a location. EIE could be used to “upscale” locally sourced performance data to achieve completeness. Using the two approaches in a complementary way could produce an ideal data set for both climate action planning as well as performance monitoring. To do so would likely require some additional ability to further classify building typologies within EIE.

Incorporating EIE into city inventory processes:
A city can directly use the data found in EIE as a baseline inventory, with the appropriate documentation and notation. Further, given the different goals a city may be trying to achieve, EIE data may be adjusted, combined with local sources of data, and/or incorporated into other tools.

Advanced communities with well-established processes and resources will face challenges when switching to new data sources such as EIE. Comparison to past reported inventory data is a comparison to largely estimated information. Such a comparison can only illustrate the difference between a previous estimate of unknown accuracy. This type of comparison will undoubtedly be made—particularly by cities considering the use of EIE for their own analyses—and results should be interpreted with this in mind.

One option is for a city to report previously calculated inventories alongside the new data, with an explanation as to the differences in underlying data and methodologies. This would provide the flexibility for the different approaches to achieve different goals, and facilitate the move toward new sources of data.

Opportunities for action planning and performance tracking

Much of the data that is used today was at some point an experimental application of something that was not necessarily designed for purpose. The Google EIE Team has the purpose of driving local action; while GHG inventories may be the initial use case envisioned by this data, experimentation by city practitioners to apply it to a range of questions they grapple with on a daily basis may reveal a number of additional ways EIE can support local action. Below are some examples of the opportunities for action planning and performance tracking that are enabled by the EIE tool.

Determinants of mode choice
EIE has a unique data set for spatially explicit mode choice across a range of locations and related factors. This data could be applied to ask critical questions that would inform

multi-modal transportation decision making and performance measurement. Spatial analysis of mode choice depending on social, demographic, and physical infrastructure data could reveal the extent to which city investments in different modes have potential to shift behavior. Some of those relationships might be fairly universal, and predictive models or simple estimate methods could be of value to all cities, even those without specific EIE profiles.

Planning with Building Areas
Many mitigation planning calculations take a bottom-up approach when considering the incremental energy and emissions savings that could be achieved through retrofits, improved building codes, tenant engagement, and other actions. In many cases, these are done without the benefit of knowing the size of the existing building stock. EIE already presents metrics for building area, which can help address:

- Decisions about the amount of space that must be retrofitted per year to improve the whole building stock.
- The physical limits to reductions from these measures, knowing the total stock that currently exists.
- Whether measures of density and space-use efficiency can assist with zoning and other planning measures.

Normalizing for Growth
Recent developments, such as contribution analysis techniques, have given new insights to cities with multiple inventories by correcting for changes in weather, emission factors, and population. However population-based normalization is essentially a proxy, considering that changes in building energy are more strongly driven by changes to the amount of building area as opposed to the number of people. This is data that EIE already produces; the only current limitation is in having it available as a time series, which will happen over time.

Evaluation Detail

The accounting principles that are common to all established inventory protocols are useful starting points to organizing the evaluation of EIE in comparison to other inventory data sources. These principles are:

- Completeness: The inventory should be as comprehensive as possible.
- Consistency: The inventory should allow for comparisons between years and across cities.
- Transparency: All assumptions, data sources, and calculations should be clearly documented to enable review and replication.
- Accuracy: Methods should not systematically overestimate or underestimate emissions.
- Relevance: Information should be useful for local government decision making to reduce emissions.

- Measurability: Data should be updated at regular intervals, and be obtainable with reasonable cost and effort.

It is important to keep in mind that inventory compilers face tradeoffs within these principles for any decision which relies on data. Compilers should make decisions about these tradeoffs in light of the long-term goals that they have for institutionalizing GHG management work within the local government. The following sections will evaluate sources for each of the major pieces of inventory data with respect to these principles.

Transportation Data

The most-common data sources currently used for local GHG inventory transportation data are agent-based models and road-count extrapolations. Table 1 illustrates a broad scoring in terms of how well many common data sources for on-road transportation conform with the reporting principles. The rationale for the scoring is provided below.

Common transportation data sources:

- Agent-based model: Most larger regional planning agencies in the United States run agent-based models for the purpose of planning future transportation infrastructure needs (examples include San Francisco Bay Area Metropolitan Transport Commission, Metropolitan Washington Council of Governments). The Vehicle Miles Traveled (VMT) outputs from these models have been adapted in many cases for use in local GHG inventories. The models have the ability to identify VMT from [modeled] trips that begin or end in a jurisdiction, making the data particularly useful for connecting land-use patterns to VMT and emissions.
- Downscaling from averages: Data may be downscaled from the county or state level based on population, or, in the case of freight transportation, based on job counts in specific industries such as retail, manufacturing, and warehousing.
- Road-count extrapolation: These data sets are typically managed by state transportation departments, and are based on vehicle counts at specific points on specific roadways. Estimation methods are needed to scale from these point measurements to the entire road system. Typically, this data results in an estimate of in-boundary VMT.
- Fuel use: Fuel-use data is collected for the vehicle fueling stations within a jurisdiction. Information collected for sales-tax purposes provides the most common source of this data.
- Other massive mobile: These sources, like EIE, aggregate location data from mobile devices.

Table 1: Transportation data comparison

Data Source	Completeness	Consistency	Transparency	Accuracy	Relevance	Measurability
Environmental Insights Explorer (EIE)	Good	Good	Fair	Good	Good	Excellent
Network-Based Model	Good	Good	Fair	Fair	Fair	Poor
Agent-Based Model	Fair	Fair	Fair	Fair	Excellent	Fair
Downscaled from Averages	Good	Good	Fair	Poor	Poor	Poor
Road-Count Extrapolation	Good	Fair	Fair	Good	Fair	Good
Fuel Use	Poor	Fair	Fair	Poor	Poor	Fair
Other Massive Mobile	Fair	Good	Fair	Good	Good	Excellent

Key

Poor	Fair	Good	Excellent

Google Environmental Insights Explorer

- Completeness
  - A single source covering all transportation modes is a unique offering.
- Consistency
  - A source of consistent data, assuming that currently methodologies provide complete coverage, and the underlying aggregation and other transformations do not change unpredictably.
- Transparency
  - The nature of the resource will always face challenges with regard to transparency. These can be mitigated with other validation efforts.
- Accuracy
  - Validation efforts thus far show a good match with other ground-truthed data.
- Relevance
  - There are a couple advantages of EIE data for policy development. The first is that, in common with agent-based models, but unlike other transportation data sources, EIE captures the impact of trips that begin or end within the jurisdiction, showing the contribution of local land uses to regional transportation flows. The ability to show flows between adjacent cities would further boost the value of EIE data for policy development. The second advantage does not directly relate to the inventory, but the ability to show

transit, bicycle, and walking trips is very valuable for transportation policy development and evaluation.

- Measurability
    - Transportation data in EIE is from a source of continuous observation, which has unique benefits beyond those provided by other sources of data. Many other potential applications can occur with data provided at shorter intervals that would be highly valuable to city planners. Through direct observation, the impact of acute events, anomalous years of transportation activity, and policy could be inferred, none of which is currently possible.

It should be noted that other data sources derived from phone, connected vehicles, and other “massive mobile” technologies share many of the same quality attributes as those which are in EIE. Two differences between EIE and other generic massive mobile sources stand out. Other sources have known limitations regarding how well they cover complete on-road transportation activity, whereas EIE appears to have enough coverage across both passenger and freight to contain both. A limitation of EIE is that it cannot differentiate between the portions that are passenger traffic versus commercial freight. In practice, the fuel economy and thus emissions intensity of the activity should reflect how differing levels of that activity affect the end result in emissions.

Other purchasable massive mobile data can have substantially better information for policy relevance. This is particularly true in the resolution of trip start and end points. EIE, on the other hand, has a unique ability to view transportation data across modes. This data on modes such as pedestrian and bicycling is very valuable for planning purposes, and could allow performance measurement of mode-shift efforts.

Agent-Based Model

- Completeness
    - Typically focused only on passenger vehicles.
- Consistency
    - Model parameters improved on a regular basis, which may improve accuracy, but at the detriment of developing consistent time series.
- Transparency
    - Fully documented, but opaque to non-experts.
- Accuracy
    - Representative of activity generation rates for various land uses and spatial configuration of a community, enhanced by calibration. May not pick up local variation or congestion patterns.
- Relevance
    - Detailed origin-destination information provides important insights into the relationship between regional development patterns, transportation infrastructure, and management decisions for transportation demand management.
- Measurability

    - Not a measurement of actual vehicle activity. Year-to-year trends cannot be attributed to driving factors. Relies on additional data like fleet makeup and fuel economy to determine emissions.

Road-Count Extrapolation

Example: State Department of Transportation Statistics
Several states, including New York and Minnesota:

- Completeness
    - These estimates are theoretically complete in that they capture both passenger and freight traffic indiscriminately.
- Consistency
    - These methods are generally consistent year to year, although occasionally a state may change its methodology. As states are not producing this information necessarily with GHG inventories in mind, any changes that happen are not necessarily communicated to the data end users, and there aren't ways of understanding the magnitude or direction of the differences introduced.
- Transparency
    - These data sets can be transparent in the methods used to generate starting points from a sampling procedure. The procedures used to scale up to the community-wide level are generally opaque.
- Accuracy
    - Accuracy at the community scale is difficult to assess for this type of data, although presumably it has passed a level of quality assurance by state agencies before being published. Some data separates types of vehicles based on weight and axle count, and this vehicle classification may be inaccurate.
- Relevance
    - Typically these values are cut along jurisdiction boundaries, reflecting exclusively Scope 1 transportation activity. There is little to observe in terms of how this data relates to transportation demand management decisions.
- Measurability
    - Since these values are driven at least initially by road-count measurements, this source does capture inter-annual variability and trends, even if the overall accuracy is unknown.

Fuel Use

- Completeness
    - It is difficult to evaluate fuel use for completeness, particularly for the measurement of induced activity that a city creates in out-of-boundary locations where fuel purchases are made beyond the region. What is more likely is that more energy will be attributed to a community regarding transportation than what that community is responsible for. Pass-through traffic, as well as the use of these fuels for stationary energy and other modes, is highly likely.
- Consistency

    - Individual survey instruments of retailers within a community are unlikely to be responded to consistently. The measurement may be consistent when these values are obtained from sales-tax receipts; however, if volumes are estimated from the revenue generated, fuel-price fluctuation may be a source of inconsistency.
- Transparency
    - Reporting by individual retailers and the processes that this kind of data go through in order to reach jurisdiction-wide numbers may not always be very transparent. Sales-tax data often includes refunds for tax-exempt entities, exports, and other externalities that are not readily transparent.
- Accuracy
    - To the extent that the data picks up non-target fuel uses and uneven geographic representation will limit accuracy.
- Relevance
    - Fuel use provides very little information that is relevant to managing the transportation demand that drives fuel use.
- Measurability
    - This approach does rely on measurement as at least a starting point before scaling or other refinements are needed. As such, it should be able to pick up inter-annual variability, assuming the data-collection mechanism stays consistent year to year.

Data on Stationary Energy Use

Data on stationary energy use is often seen as being straightforward in emissions accounting, but in reality it poses many challenges for both inventory and subsequent action planning. Consistent categorization of this data across data providers and consumers is one of the biggest challenges from all sides. Data providers organize their information along customer classes, which are generally straightforward; but all have slight variations in the definition of customer classes. Reporting categories are defined around national accounts, which is irrelevant for local practitioners. Both on the production and receiving end of activity data, neither is particularly well suited to understanding end use, which is critical to policy development. This is particularly challenging around places with any amount of industrial process energy. Whether this is aggregated with other non-residential uses is unknown in many instances, making interpretation difficult.

Potential data sources for stationary energy use:

- Utility energy: This is the most common data source used for local GHG inventories. Utilities provide data drawn from customer-metered usage, and aggregated by customer class to preserve privacy.
- Downscaled: Another source is usage data downscaled from the state or national level. An example is the State and Local Energy Data from DOE.

- Benchmarking: This source of usage is data reported through state or local building benchmarking and disclosure requirements. This data is very high quality and can allow analysis of energy use by building size and type. Data coverage is typically above 80% for buildings subject to the requirement, but the size cutoff is typically 35,000-50,000 ft^2, with no data on smaller buildings.

While EIE data has some limitations in the way it is estimated, it does have advantages in being more transparent than the measured data that cities can obtain directly from their utilities. The table below and following summaries assess where EIE data stands among other general options cities that have before them.

Table 2: Stationary-energy-use data comparison

Data Source	Completeness	Consistency	Transparency	Accuracy	Relevance	Measurability
EIE						
Utility Energy						
Downscaled						
Benchmarking						

EIE

EIE currently suffers many of the same limitations that downscaling suffers, with the exception that it is inherently slightly less transparent and slightly more relevant, since it provides good information on the quantity of available building space. However, the EIE is not downscaling so much as laterally transferring performance data from CURB to other contexts. If the source of performance data could be improved, EIE could provide a very useful service.

Benchmarking and disclosure activities in cities are not likely to provide complete coverage of cities in the foreseeable future. However, those activities could be designed to obtain representative samples of city performance that captures the unique aspects of the building stock and climate for a location. EIE could be used to “upscale” locally sourced performance data to achieve completeness. Using the two approaches in a complementary way could produce an ideal data set for both climate action planning as well as performance monitoring. To do so would likely require some additional ability to further classify building typologies within EIE.

Utility Energy

- Completeness
  - Energy-use data obtained directly from utilities is considered the gold standard for inventories today, but it does suffer some inherent limitations in many attributes. Often due to aggregation rules, some data may be omitted, such as large industrial users. This by itself would not be such a problem, as long as the

omitted users were rolled up into another class—allowing for more accuracy in total use, but less accuracy in user sector. As many utilities limit transparency to the data-collection methodology, completeness can never be truly assessed. In addition, there may be challenges if utility spatial data on customers does not align with jurisdictional boundaries (for example, the utility may classify customers by zip code or by their own unique zones, which may cross jurisdictional boundaries). This can result in incomplete coverage, or, more frequently, inclusion of usage for customers outside the jurisdiction, and it may not be obvious that this is the case.

- Consistency
  - Related to the point above, cities with users that are near a threshold may see the level of aggregation from one year to the next vary considerably, which complicates the ability to track trends. Utility definitions of customer types and determination of customers that are within the community boundary can also change from year to year, leading to misleading results.
- Transparency
  - Again, because cities are not privy to privacy-driven adjustments by the utility, it is often difficult to know exactly what the data represents.
- Accuracy
  - In theory, this data should be highly accurate, but there is no way to assess that.
- Relevance
  - When data is available in ways that help to understand end use and the efficiency of the overall building stock, it is highly relevant. Lack of detail on building type (e.g. single family vs. multifamily) and age is a limitation.
- Measurability
  - In theory, utility data is a direct measurement, but the quality of that measurement is subject to the challenges described above.

Downscaled State and National

- Completeness
  - Downscaled data is typically incomplete, in that many sources only focus on building energy and do not capture process energy. However, the ways in which it is incomplete is typically knowable from the specific approach used.
- Consistency
  - Downscaled data is typically consistent in how it is created, but, depending on the specific approach, it may be difficult to produce consistent results over time if the underlying source data changes to attempt to capture variability.
- Transparency
  - Downscaling is typically transparent when done with simple methods. However, some more-sophisticated approaches, such as those produced by federal agencies and based partially on proprietary data sets, lack this quality.
- Accuracy

    - This approach is generally recognized as inaccurate. Accuracy is not necessarily the end goal for using this approach as it is often the choice of last resort.
- Relevance
    - Because there is some control available in terms of how data is downscaled and what it represents, it can be more relevant than other approaches by focusing attention on segments of the built environment that are the target of interventions.
- Measurability
    - Use of downscaled data is not appropriate for performance measurement

Benchmark

- Completeness
    - Because benchmark data is often limited to a small selection of buildings grouped by size and usage, it is far from complete.
- Consistency
    - Voluntary benchmarking is limited by participation and can provide inconsistency from year to year. Ordinance-driven benchmarking can provide more consistency, but is still reliant on user input.
- Transparency
    - Benchmark programs initiated by a public entity can be more transparent, either through public reporting or through open records requests.
- Accuracy
    - Benchmark data can be influenced by user input. There is rarely an enforcement or data verification aspect to the benchmarking process unless combined with other ordinance-driven building improvements.
- Relevance
    - The highly detailed data from benchmarking is valuable for policy development, allowing differentiation of energy use by building type, size, and age.

Technical Review of **Google Environmental Insights Explorer** Data for
Local Greenhouse Gas Inventories

Written by ICLEI - Local Governments for Sustainability USA