Deliverable C.2.1 (Task C.2.1 and C.2.2): Type-approval data of vehicle models currently on the market and vehicle models of previous years still in customer hands

Version October 2021

MILE21

MORE
INFORMATION
LESS
EMISSIONS

List of authors
Georg Bieker – ICCT
Sonsoles Díaz – ICCT

1. Introduction

The present report describes MILE21 tasks C.2.1 and C.2.2, provides background information to put the tasks into context, and summarizes how the tasks were implemented. Additional details regarding select implementation steps are provided in the four annexes. The bulk of the tasks was completed by ICCT between October 2018 and November 2019. However, work related to task C.2.1 will continue throughout the project as data will be updated on a regular basis.

2. Description of the tasks

Task C.2.1, which is the first sub-action of Task C.2, is to gather and harmonize type-approval CO_2 emissions and fuel consumption data of vehicle models currently on sale on the EU market and models of previous years still in customers' hands. Task C.2.2 is to gather official real driving emissions (RDE) data on air pollutant emissions for these vehicles and is meant to complement Task C.2.1.

Task C.2.1 also involves collecting data on key vehicle characteristics such as build year, make, or fuel type of each model. Besides data collection, the task further involves transforming the raw data for MILE21 needs, including extensive data quality checks. To account for new vehicle models, the data is updated regularly during the project duration.

The above-mentioned data, as well as the data transformation, play a crucial role in the project. The collected vehicles official data is the backbone of the MILE21 vehicles' database. It determines, for example, which vehicle models will be covered by the website and the level of information that will or can be made available to final users. Also, further data such as model specific MILE21 real-world fuel consumption and CO_2 emission estimates resulting from Tasks C.3 and C.4 as well as fuel consumption data input by MILE21 users build on the official data.

3. Implementation of the tasks

To accomplish **Task C.2.1**, we first performed a scoping exercise of available type-approval data sources in the EU. This step involved listing all data sources and collecting information on their usability, reliability, relevance, and availability, including information on their license fees too. Several possible ways for collecting the wanted data were identified. Initially, the preferred option was to retrieve the data directly from EU type-approval authorities or the European Environmental Agency (EEA). The latter records and publishes type-approval monitoring CO_2 emissions data annually. However, ways of obtaining the data with greater frequency were also explored. For example, the German KBA or the UK VCA offer vehicle type-approval data on their websites for free if provided on an annual basis and for a negotiable fee if provided more regularly. Moreover, we approached other organizations that have set up similar databases in the past, such as the German car club ADAC or the Belgian platform Ecoscore.be, to seek cooperation with them while ensuring to not harm any business models built around the databases.

In a second step, after choosing the most appropriate data source and coming to an agreement with the data provider, the data was acquired and processed to meet the project requirements. One particularly important step at this stage was defining the unit of analysis, which depends on the level of detail of the data available, on the one hand, and the level of granularity deemed convenient for users, on the other hand.

To accomplish **Task C.2.2**, ICCT first approached type-approval authorities as they are requested to make RDE type approval test information available to interested parties by Commission Regulation (EU) 2017/1151[1]. However, all type approval authorities contacted essentially refused to provide the RDE data, pointing to the very high cost of (manually) extracting the data from type-approval certificates. Since the regulation only allows to charge a proportionate and non-discouraging fee for the data, we informed the European Commission (DG Grow) about this issue in June 2019. After two positive responses from DG Grow in July and August 2019, in which they confirmed having received our request and stated their intention to provide help, we did not hear back from them.

ICCT also reached out to all main vehicle manufacturers to request the RDE data. This approach worked out in most cases, although the quality of the data provided by manufacturers varied significantly. In parallel, in April 2019, we sought legal support to assess whether there were any legal impediments to the publication of manufacturers RDE data on the MILE21 website. The evaluation carried out by the expert lawyer Prof. Remo Klinger concluded that publishing the RDE data entailed a significant risk of lawsuits by manufacturers. His clear advice, therefore, was to not include the RDE data on the MILE21 website.

Besides, from a technical perspective, manufacturers RDE data imposed several challenges:

- RDE data is only provided for Euro 6d-TEMP and newer vehicles, i.e., there is no data available for older vehicle models.
- RDE data is only available at the vehicle family level but not for individual vehicle model variants.
- The testing conditions for the RDE data are not standardized.
- Since the results are modified by correction factors, if the testing conditions are outside a certain temperature range, for instance, they do not reflect real driving conditions.

As a potential alternative to publishing RDE data, ICCT explored a possible cooperation with the TRUE (The Real Urban Emissions) initiative. TRUE is a rating system which rates vehicle families on real-world NOx emissions, as measured by remote sensing. The initiative is led by the FIA Foundation, which expressed general interest in providing real-world NOx data to the MILE21 project. Further, the option of using data from real-world tests by the Green NCAP initiative was also assessed. However, these two initiatives have limited market coverage, which means that they could only provide data for a minority of the models covered by MILE21.

During the project meeting in November 2019, the consortium – given all the challenges described above – agreed on only publishing the Euro standard of each vehicle as an indicator of its pollutant emissions levels. The Euro standard was regarded as relevant information for consumers, given that modern vehicles are expected to comply with air pollutant requirements also under real-world driving conditions.

3.1 Scoping Exercise of Type-Approval Data Sources

Given the importance of the Task C.2.1 data, we performed an exhaustive scoping exercise of type-approval data sources available in the EU. The scoping involved 1) extensive research to list all data sources available, 2) contacting data providers and collecting information on the usability, reliability, cost etc. of each data source, and 3) gathering and evaluating data samples from the suitable

[1] Based on the provisions of paragraph 3.1.3.4. of Annex IIIA of Regulation (EU) 2017/1151 amended by (EU) 2017/1347 (13th of July 2017) and on the basis of point 3.1.3.4 of Annex IIIA of regulation (EU) 2017/1151 amended by (EU) 2018/1832 (5th of November 2018), EU type-approval authorities are required to make RDE type approval test information as defined in paragraph 3.1.3.1. and 3.1.3.2. of Annex IIIA of these regulations available to the public.

sources. A total of 15 sources of type-approval data was considered for the project, including type-approval authorities, EEA, the German car club ADAC or the Belgian platform Ecoscore.be. See annex 1 for a brief description of the data sources.

The scoping exercise was a laborious task due to the high number of data sources that had to be evaluated to find appropriate candidates as well as due to the intensive exchange with the different data providers, which in many cases spread over several weeks or months and required a considerable number of emails and conference calls. A notable obstacle in this process was the transition from the NEDC certification procedure to the new WLTP. Some data providers were still on the process of getting a grip on the new type-approval values and were not able to provide reliable fuel consumption and CO2 emissions data.

To assess the suitability of a data source systematically, we set the following list of criteria: completeness, data quality, geographic scope, historical scope, cost, sustainability, and vehicle categories. See annex 2 for a summary of the data source assessment (only data sources that were shortlisted are included). The summary also provides a description of the evaluation criteria. Based on these criteria, we selected the ADAC.

In March 2019, we came to an agreement with the ADAC and received the first type-approval dataset in early April. The ADAC data will be updated every 6 months but can be used only during the project duration, according to the license agreement. The first dataset included the following vehicle information: make, model series and name, model and build year, powertrain and fuel type, body type, engine power and displacement, transmission, emission standard, NEDC and/or WLTP fuel consumption and CO_2 emission values, and the electric range of electric models of about 49,000 vehicle model variants on sale in Germany in year 2013 and after. Regarding the age of the vehicle models covered in the database, a threshold of five years (i.e., starting in registration year 2013) was agreed on with the project partners.

3.2 Aggregation and Processing of the Type-Approval Dataset

From April 2019 on, we worked on cleaning and aggregating ADAC data to make it usable for the upcoming website tools. For this work, we followed the so-called ETL procedure (Extract – Transform – Load). Data extraction involves extracting data from the files provided by ADAC; data transformation means cleaning and transforming the data into a proper format and storage structure that fits MILE21 purposes; finally, data loading describes the insertion of data into the final target database, which in our case is managed by Emisia.

Data transformation is the most laborious step. Examples of transformation steps are cleaning make and model names and converting dates to ISO format. A key transformation step is data aggregation. In agreement with the project partners, it was decided to aggregate the data by make, model series, powertrain, fuel type, displacement, power, transmission, emission standard, and build year. A detailed description of the aggregated dataset is given in annex 3.

In July 2019, to provide further data for the development of models M2 and M3, we acquired data on start-stop system, number of gears, drive wheel, and height and width of the vehicle model variants from ADAC. Drive wheel and potentially also start-stop system are also foreseen as grouping variables to further narrow the number of car model variants within a group.

3.3 Estimating the Market Coverage of the Type-Approval Dataset

In July and August 2019, we estimated the EU market coverage of the ADAC data by comparing it with EEA registrations data. Across all years and countries, 79% of registrations in the (valid) EEA

data can be mapped to car models covered by the ADAC dataset. As a reference, depending on the year, this number is also 79-89% for the German market. Therefore, the result of this study is positive in the sense that the ADAC data does not seem particularly skewed toward Germany. Countries like Cyprus, Ireland, and Lithuania have considerably higher sales coverage than Germany in the ADAC data. The EEA data quality – not the limited coverage of the ADAC data – was the limiting factor for establishing coverage with higher accuracy. See annex 4 for further details.

When checking whether the most popular car models were included in the ADAC data, we found that the 80 most popular models (according to the sales numbers in the EEA data) are covered.

4. Conclusion

After an exhaustive scoping exercise of type-approval data sources available in the EU and an intensive exchange with the different data providers, we identified a list of potential data sources. By evaluating each of these data sources systematically, we selected the data provided by the German car club ADAC.

The first ADAC dataset included about 49,000 vehicle model variants on sale in Germany in year 2013 and after. By comparing with EEA registration data, we found that the ADAC dataset largely covers the EU market and is not particularly skewed toward Germany. To include new models, the database is updated every six months.

The ADAC dataset contains the following vehicle characteristics: make, model series and name, model and build year, powertrain and fuel type, body type, engine power and displacement, transmission, emission standard, NEDC and/or WLTP fuel consumption and CO_2 emission values, and the electric range of electric models. To provide data for the development of model M2 and M3, we further acquired data on start-stop system, number of gears, drive wheel, and height and width.

The quality, completeness, and regular updates of the ADAC raw data in combination with our systematic and rigorous data aggregation and transformation steps, including cross-checking with EEA data, provides a solid basis for all functionalities of the MILE21 website.

Annex 1 - Overview of type-approval data sources

The following table provides a description of the sources of type-approval data considered for the project to date.

Source name	Source description	Data manager	Funder	Status	Data origin	Geog. scope	Vehicle categories	Market coverage	Update frequency	Comments	Web page/References
Ecoscore.be	online rating tool; well-to-wheel rating on environmental performance of passenger cars based on official data; rating takes into account direct and indirect emissions, including greenhouse gases, air pollutants, and noise; score between 0 and 100 for each model	Flemish Institute for Technological Research (Vlaamse Instelling voor, VITO) and Vrije Universiteit Brussel (VUB)	Regional Belgian governments (Brussels, Wallonia, Flanders)	active	Belgian vehicle registration service (Dienst Inverkeersstelling, DIV) database, matched with data from Dutch counterpart (Rijksdienst voor Wegverkeer, RDW) to complement vehicle technical data with environmental data	Belgium	passenger cars and some light commercial vehicles	all new and second-hand models available in the Belgian market	bi-monthly	supports VIN-based search; database will be expanded in the near future	ecoscore.be/vehicles?fuel=F2&page=1; ecoscore.be/files/2018%20Ecoscore%20Report%20for%20the%20Belgian%20New%20and%20Re-registered%20car%20fleet%20in%202017.pdf; ecoscore.be/files/Comparison%20of%20Clean%20Vehicles%20Portal%20with%20Ecoscore_0.pdf
Clean Vehicle Portal (CVP)	online information platform; calculates Operational Lifetime Cost of a vehicle based on its tank-to-wheel official energy consumption, CO2 emissions, and pollutant emissions data; meant to support public authorities implementing the Clean Vehicles Directive	TÜV Nord	European Commission	launched in 2010, now inactive	unknown	EU	light-duty and heavy-duty vehicles	new vehicle models	inactive	has a forum to share information between car buyers	cleanvehicle.eu; ecoscore.be/files/Comparison%20of%20Clean%20Vehicles%20Portal%20with%20Ecoscore_0.pdf; clean-fleets.eu/fileadmin/files/documents/Publications/Clean_Fleets_Guide_-_Final__Nov_2014.pdf
Allgemeiner Deutscher Automobil-Club (ADAC)	largest automobile club in Europe, based in Germany; maintains publicly accessible database	ADAC Technikzentrum, Abteilung Fahrzeugtechnik	self-funded	active	OEMs sales information	Germany	passenger cars	new and old vehicle models ever sold in the German market (>100,000 in total)	flexible/ monthly	>200 specs/model; consumer oriented; mobile application available	adac.de/infotestrat/autodatenbank/default.aspx; adac.de/infotestrat/autodatenbank/default.aspx
Kraftfahrt-Bundesamt (KBA)	German type-approval authority	Fahrzeugstatistik	Federal Ministry of Transport and Digital Infrastructure	active since 1951	Type approval certificates/ Zentrales Fahrzeugregister (ZFZR)	Germany	light-duty and heavy-duty vehicles	all vehicle models sold in the German market	flexible/ monthly	CO2 data only recorded since Oct. 2005 (tbc); 2018 new registrations statistics of models with WLTP type	kba.de; kba.de/DE/Statistik/Fahrzeuge/fz_methodische_erlaueterungen_201805_pdf.pdf?__blob=publicationFile&v=6

Source name	Source description	Data manager	Funder	Status	Data origin	Geog. scope	Vehicle categories	Market coverage	Update frequency	Comments	Web page/References
										approval are based on NEDC correlated values, from 2019 only WTLP values will be considered in the statistics	
Deutsche Automobil Treuhand GmbH (DAT)	market research company focused on the automotive industry; publishes guide on vehicle efficiency of new passenger car models on a quarterly basis according to Directive 1999/94/EG on behalf of VDA and VDIK	DAT	VDA/VDIK	active since 2004	OEMs	Germany	passenger cars	new vehicles models for sale on the German market at the time of publication (about 4,000 per publication)	quarterly	free guides; further data available upon request for a fee	dat.de/co2/; dat.de/news/dat-ueberarbeitet-ihren-leitfaden-fuer-die-pkw-verbrauchswerte/
The Real Urban Emissions (TRUE) Initiative	online rating tool; rates vehicle families on real-world NOx, as measured by remote sensing; objective is to raise awareness of the magnitude and scope of excessive real-world vehicle emissions	Diverse partners	FIA Foundation, Bloomberg Philanthropies	active since 2018	unknown	EU	passenger cars	90% Euro 3 to Euro 6 models (about 760 vehicle families)	unknown	vehicle families defined as a unique combination of fuel type, Euro standard, manufacturer group, and engine displacement	trueinitiative.org; theicct.org/sites/default/files/TRUE_vehicle_rating_factsheet_20180604_0.pdf; theicct.org/sites/default/files/publications/TRUE_Remote_sensing_data_20180606.pdf
European Environment Agency (EEA)	EEA publishes dataset on the CO2 performance of new EU passenger car registrations annually; used by the European Commission to monitor and evaluate whether manufacturers are in compliance with mandatory CO2 emission targets for passenger cars	EEA	EU	active since 2010	OEMs	EU	passenger cars and light commercial vehicles	all new vehicle models	annually	free data source	eea.europa.eu/data-and-maps/data/co2-cars-emission-14; theicct.org/sites/default/files/TRUE_vehicle_rating_factsheet_20180604_0.pdf; theicct.org/sites/default/files/publications/TRUE_Remote_sensing_data_20180606.pdf
Rijksdienst voor Wegverkeer (Netherlands Vehicle Authority, RDW)	Dutch vehicle registry by the Netherlands Vehicle Authority	RDW	Ministry of Infrastructure and Water Management.	active	Dutch vehicle registry	Netherlands	light-duty and heavy-duty vehicles	all vehicles currently registered in the Netherlands	daily	open data source	opendata.rdw.nl; rdw.nl/-/media/rdw/rdw/pdf/sitecollectiondocuments/over-rdw/naslagwerk/beschrijving-dataset-typegoedkeuring-v10.pdf; rdw.nl/-

Source name	Source description	Data manager	Funder	Status	Data origin	Geog. scope	Vehicle categories	Market coverage	Update frequency	Comments	Web page/References
											/media/rdw/rdw/pdf/sitecollectiondocuments/over-rdw/brochures-en-folders/rdw_strategy_en_def.pdf
French Environment & Energy Management Agency (ADEME)	French agency publishing annual handbooks on CO2 performance of new models according to Directive 1999/94/EG	ADEME	French Minister for the Ecological and Inclusive Transition	active since 2007	OEMs	France	passenger cars	new vehicles models for sale on the French market at the time of publication (about 7,000 per publication)	annually	free guides	carlabelling.ademe.fr/index/; ademe.fr/sites/default/files/assets/documents/consommation-emissions-vehicules-particuliers-2018_8521.pdf
IHS Markit	global supplier of technical information regarding the vehicle market, among other industries	IHS Markit	private company	active	OEMs and type-approval authorities	EU/world wide	light-duty and heavy-duty vehicles	new and old vehicle models ever sold in the EU and other markets	flexible		ihsmarkit.com/products/automotive-market-data-analysis.html
JATO Dynamics Limited	global supplier of vehicle market information	JATO	private company	active	OEMs, car dealers, and type-approval authorities	EU/world wide	light-duty vehicles	new and old vehicle models ever sold in the EU and other markets	flexible	>1.200 specs/model	jato.com
Emissions Analytics (EA)	independent vehicle testing organization; set up the EQUA Index, a rating system developed to inform the public about the on-road performance of vehicles and available for consultation online	EA	private company	active since 2012	unknown	UK	passenger cars (but not battery-electric vehicles)	90% of car sales	unknown		equaindex.com/equa-fuel-economy-index/; equaindex.com/about-us/
km77.com	Spanish automobile website	km77.com	private company	active since 1999	unknown	Spain	passenger cars and light commercial vehicles	new and old vehicle models sold in the Spanish market	unknown		km77.com

Source name	Source description	Data manager	Funder	Status	Data origin	Geog. scope	Vehicle categories	Market coverage	Update frequency	Comments	Web page/References
Autoverbrauch.at	online database of fuel consumption and CO2 emissions of new models according to Directive 1999/94/EG	Unclear	Ministerium für ein lebenswertes Österreich (BMLFUW), Bundesgremium Fahrzeughandel und Arbeitskreis der Autoimporteure	active	OEMs	Austria	passenger cars	new vehicle models for sale on the Austrian market at the time of publication	continuously updated	Possibility to switch between NEDC and WLTP values	autoverbrauch.at; wien.arbeiterkammer.at/service/newsletter/UV_Fakten_und_Positionen_2018_03.pdf; bmnt.gv.at/umwelt/luft-laerm-verkehr/verkehr-laermschutz/autoverbrauch-at.html
Ministero dello sviluppo economico (Italian Ministry of Economy, MISE)	online handbook of fuel consumption and CO2 emissions of new models according to Directive 1999/94/EG	MISE	MISE	active since 2012	OEMs	Italy	passenger cars	new vehicle models for sale on the Italian market at the time of publication	low, uncertain	last handbook published in 2016, new handbook in preparation	mise.gov.it/index.php/it/mercato-e-consumatori/qualita-di-prodotti-e-servizi/auto-ed-emissioni-co2/archivio

Annex 2 - Summary of data source assessment

The following table provides a summary of the assessment of the type-approval data sources considered for the project to date. Please note that only shortlisted data sources are included. The assessment was made based on information and data samples provided by each of the data providers.

Criteria	Description	ADAC	km77.com	Ecoscore.be	EEA	IHS	RDW
1. Completeness	a. To what extent does the dataset cover all variables needed for the website?[2]	a. Complete. Further, it covers both NEDC and WLTP values for part of the recent models	a. The dataset covers all variables except the Euro standard letter designation and electricity consumption values. At the moment, only NEDC values provided.	a. The current database misses at least transmission and the Euro standard letter designation. At the moment, only NEDC values provided. Note: the Ecoscore team is currently working on integrating the missing variables in the dataset	a. The dataset covers all variables except transmission, Euro standard, and fuel consumption values. It is expected to cover NEDC and WLTP CO2 values in the future. Note: it covers number of registrations.	a. Complete. NEDC or WLTP values will be provided (not both). Notes: -IHS is only able to provide the minimum WLTP fuel consumption value of a trim level (not a range, not the maximum value). -IHS sales data available from ICCT's Pocketbook	a. Seems to cover all variables except transmission.
2. Data quality	a. How much work will data cleanup require? b. How likely is it that data will change retrospectively?	a. Little cleanup work expected b. Relatively unlikely	a. Moderate to high amount of work + work for retrieving the data using km77's API b. Unknown	a. Amount of work depends on whether access would be granted to DIV and/or RDW raw data or to clean data from VITO b. Unknown	a. High amount of work b. There is a provisional and a final data set per year	a. Moderate b. Relatively unlikely	a. Extremely high b. Unknown
3. Geographic scope	a. What is the geographic scope of the data source? b. Is the dataset reasonably representative of the EU vehicle market?	a. Germany b. Reasonably representative, Germany being the largest car market in the EU	a. Spain b. Questionable	a. Belgium b. Questionable - according to VITO, dataset is reasonably representative as long as information on trim level is not provided	a. EU b. Yes	a. Biggest five EU markets (DE, UK, FR, IT, ES) b. Yes	a. Netherlands b. Questionable

[2] The variables needed for the website are: make, model, build or model year, body type (optional), engine power and displacement, mass, powertrain and fuel type, transmission, euro standard incl. letter designation, type-approval fuel consumption and CO_2 emission values.

Criteria	Description	ADAC	km77.com	Ecoscore.be	EEA	IHS	RDW
4. Historical scope	a. How many years does the dataset cover?[3]	a. Build year ~ 1950 to date	a. Covers at least all build years needed for the website	a. Seems to cover at least all build years needed for the website	a. Registration year 2010 to date (M1 vehicles) and 2012 to date (N1 vehicles)	a. Covers all build years needed for the website	a. Covers all vehicles currently registered in the Netherlands
4. Cost	a. How much does the dataset cost? b. Are there any additional licensing fees?	a. Moderate (costs for historic data from build year 2013 to date and updates every six months) b. No - However, according to current offer, data should be deleted on completion of the project	a. Unknown but expected to be low - km77 has so far given ICCT free access to their database b. Unknown/probably not	a. Unknown but expected to be low - should be negotiated with Belgian governments b. Unknown	a. Free b. No	a. Upper budget limit (three-year contract) b. No	a. Free b. No
5. Sustainability	a. How often can we retrieve updated data from the data source? b. Will the data source continue to exist throughout and after the project?	a. Agreement on six-monthly updates b. Highly likely	a. As frequently as needed b. Likely	a. VITO updates Ecoscore dataset twice a month b. Likely (funders are currently investing in improving the platform)	a. Annually - provisional data set is published in spring. Publication date of final data varies/uncertain b. Highly likely	a. Agreement on six-monthly updates b. Highly likely	a. Seems like we could retrieve the data as frequently as needed b. Highly likely
6. Vehicle categories	a. Does the dataset cover N1 vehicles?	a. No (only vans that could be M1)	a. Yes	a. Dataset includes some N1 vehicles	a. Yes, in a separate dataset	a. No	a. Yes

[3] The dataset should cover at least vehicles models from build year 2013 and younger

Annex 3 - README: ADAC database for the MILE21 project

Code

README: ADAC database for the MILE21 project

Uwe Tietge

2019-09-11

1 Introduction

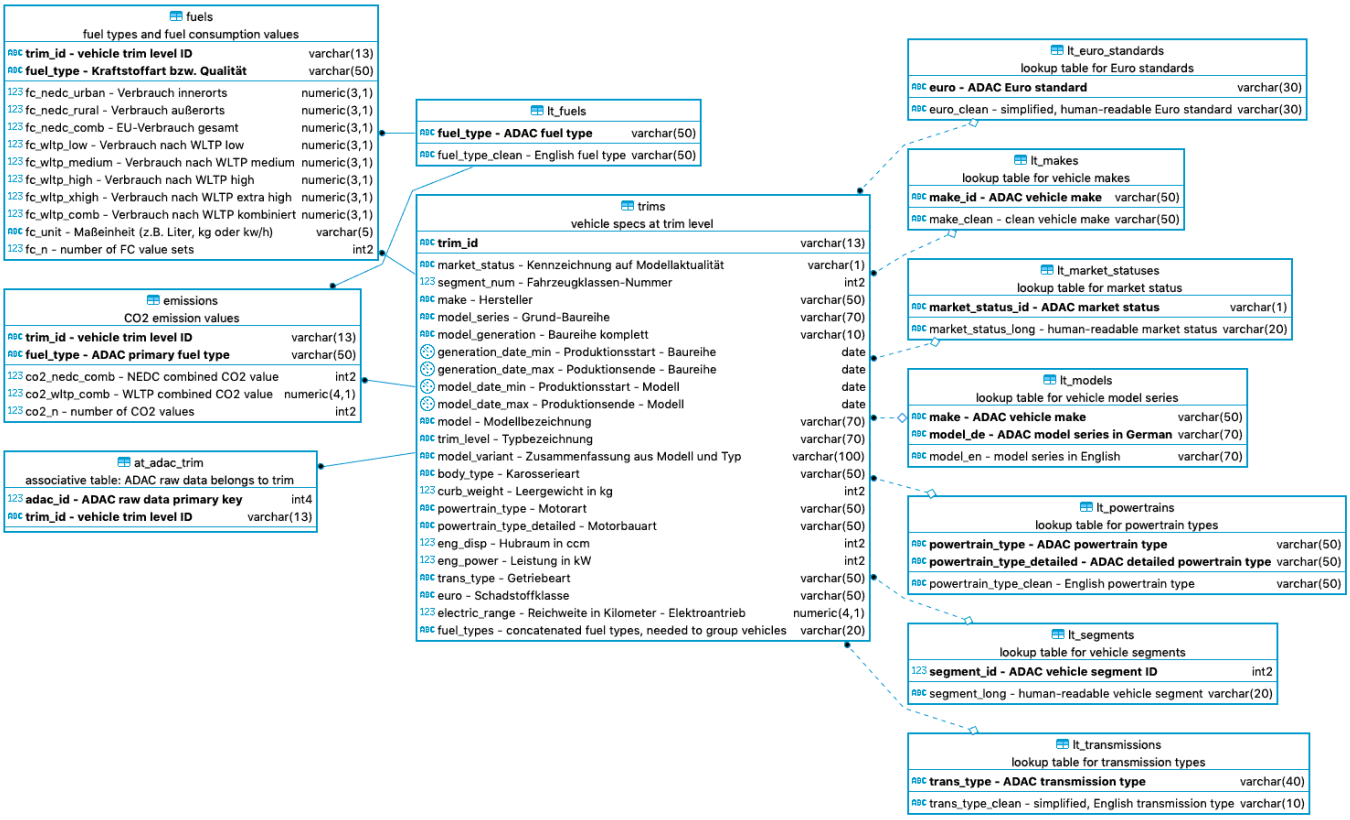
The ADAC database in its current form conducts five main tasks:

1. It imports the raw ADAC data.
2. It performs some database normalization.
3. It uses lookup tables and constraints for QC.
4. It uses lookup tables to translate, clean, or simplify variables to make the data suitable for a non-technical, English audience.
5. It aggregates the ADAC data and exports the aggregated data.

2 Documentation

2.1 ERD Diagram

The following entity relationship diagram depicts the ADAC data and corresponding lookup tables. Short descriptions of tables and columns are provided in the diagram.



2.2 Tables & views

Raw ADAC data is imported into a temporary table. A handful of errors in the ADAC data are corrected using UPDATE statements.

The database consists of three main tables, which are described below.

2.2.1 *trims*

The core technical vehicle specifications are stored in the *trims* table. This table has fewer rows than the raw ADAC data because the raw data lists individual trims using different fuel types (e.g., gasoline and LPG) in different rows. This would lead to problems because vehicles that are otherwise identical but that can use multiple fuels would not be identifiable as the same vehicles.

2.2.2 *fuels*

The *fuels* table stores fuel types and fuel consumption values. For each vehicle trim, there may be multiple fuel types (e.g., gasoline and LPG or gasoline and electricity) with different fuel consumption values and units.

2.2.3 *emissions*

The *emissions* table stores CO2 emission values. For each trim and for each *primary* fuel in the ADAC data, there is one CO2 emission value.

2.2.4 Table overview

The following table lists all tables in the ADAC database.

Show 10 entries Search:

table_name	type	description	rows_n	cols_n	size_bytes
All	All	All	All	All	All
at_adac_trim	table	associative table: ADAC raw data belongs to trim	51,129	2	2,277,376
emissions	table	CO2 emission values	51,123	5	2,736,128
fuels	table	fuel types and fuel consumption values	51,533	12	4,390,912
lt_euro_standards	table	lookup table for Euro standards	0	2	8,192
lt_makes	table	lookup table for vehicle makes	67	2	8,192
lt_market_statuses	table	lookup table for market status	0	2	8,192
lt_models	table	lookup table for vehicle model series	544	3	32,768
lt_powertrains	table	lookup table for powertrain types	0	3	8,192
lt_segments	table	lookup table for vehicle segments	0	2	8,192
lt_transmissions	table	lookup table for transmission types	0	2	8,192

Showing 1 to 10 of 12 entries Previous 1 2 Next

2.2.5 Column overview

All columns in all tables and views are listed below.

Show 20 entries Search:

table_name	column_name	data_type	column_description
All	All	All	All
at_adac_trim	adac_id	integer	ADAC raw data primary key
at_adac_trim	trim_id	character varying	vehicle trim level ID
emissions	co2_n	smallint	number of CO2 values
emissions	co2_nedc_comb	smallint	NEDC combined CO2 value
emissions	co2_wltp_comb	numeric	WLTP combined CO2 value
emissions	fuel_type	character varying	ADAC primary fuel type
emissions	trim_id	character varying	vehicle trim level ID
fuels	fc_n	smallint	number of FC value sets
fuels	fc_nedc_comb	numeric	EU-Verbrauch gesamt
fuels	fc_nedc_rural	numeric	Verbrauch außerorts
fuels	fc_nedc_urban	numeric	Verbrauch innerorts
fuels	fc_unit	character varying	Maßeinheit (z.B. Liter, kg oder kw/h)
fuels	fc_wltp_comb	numeric	Verbrauch nach WLTP kombiniert
fuels	fc_wltp_high	numeric	Verbrauch nach WLTP high
fuels	fc_wltp_low	numeric	Verbrauch nach WLTP low
fuels	fc_wltp_medium	numeric	Verbrauch nach WLTP medium
fuels	fc_wltp_xhigh	numeric	Verbrauch nach WLTP extra high

table_name	column_name	data_type	column_description
fuels	fuel_type	character varying	Kraftstoffart bzw. Qualität
fuels	trim_id	character varying	vehicle trim level ID
lt_euro_standards	euro	character varying	ADAC Euro standard

Showing 1 to 20 of 94 entries Previous 1 2 3 4 5 Next

3 Open issues

All lookup tables need to be thoroughly reviewed. How we translate or simplify values in the lookup tables affects how data is aggregated, and thus affects the quality of the aggregated data. Comments and questions on lookup tables are listed below:

- segments:
  - Do we use segments on the MILE21 website?
  - The ADAC data does not single out SUVs. Do we want to try to create one or more SUV segments via body type?
  - The “micro” segment is unusual and basically just consists of smart vehicles. Merge with “mini” segment?
- transmission type
  - ADAC reports the transmission type for fully electric vehicles as “automatic.” Should arguably be set to “N/A”?
  - 11 transmission types were “-” for hybrids, were set to “automatic” on import.
- powertrain/fuel types
  - Fine-tune MILE21 powertrain/fuel types if necessary.
- Euro standards
  - Letter designations (e.g., Euro 6*c*) are available in most, but not all, of the ADAC data. I removed letter designations in the lookup table, which harmonizes the data but diminishes its fidelity. Do we want users to be able to select letter designations?
  - Some Euro standards are “electric” or “fuel cell.” How to translate to conventional standards?
- engine displacement:
  - Engine displacement is NULL in case of fully electric powertrain. Does anyone prefer “N/A”?
- languages
  - Some lookup tables translate German terms into English (e.g., transmission types, model names, powertrain types, fuel types). It is conceivable to add more languages to the lookup table, but the interface with the website would have to be able to pick the correct language depending on user preference. Probably a question/suggestion for Emisia.

Annex 4 - Market coverage of the MILE21 database

Code

Market coverage of the MILE21 database

Uwe Tietge

2019-09-06

1 Introduction

This document combines the MILE21 ADAC data and European Environment Agency (EEA) CO2 monitoring data to determine the market coverage of the MILE21 database.

2 Methology

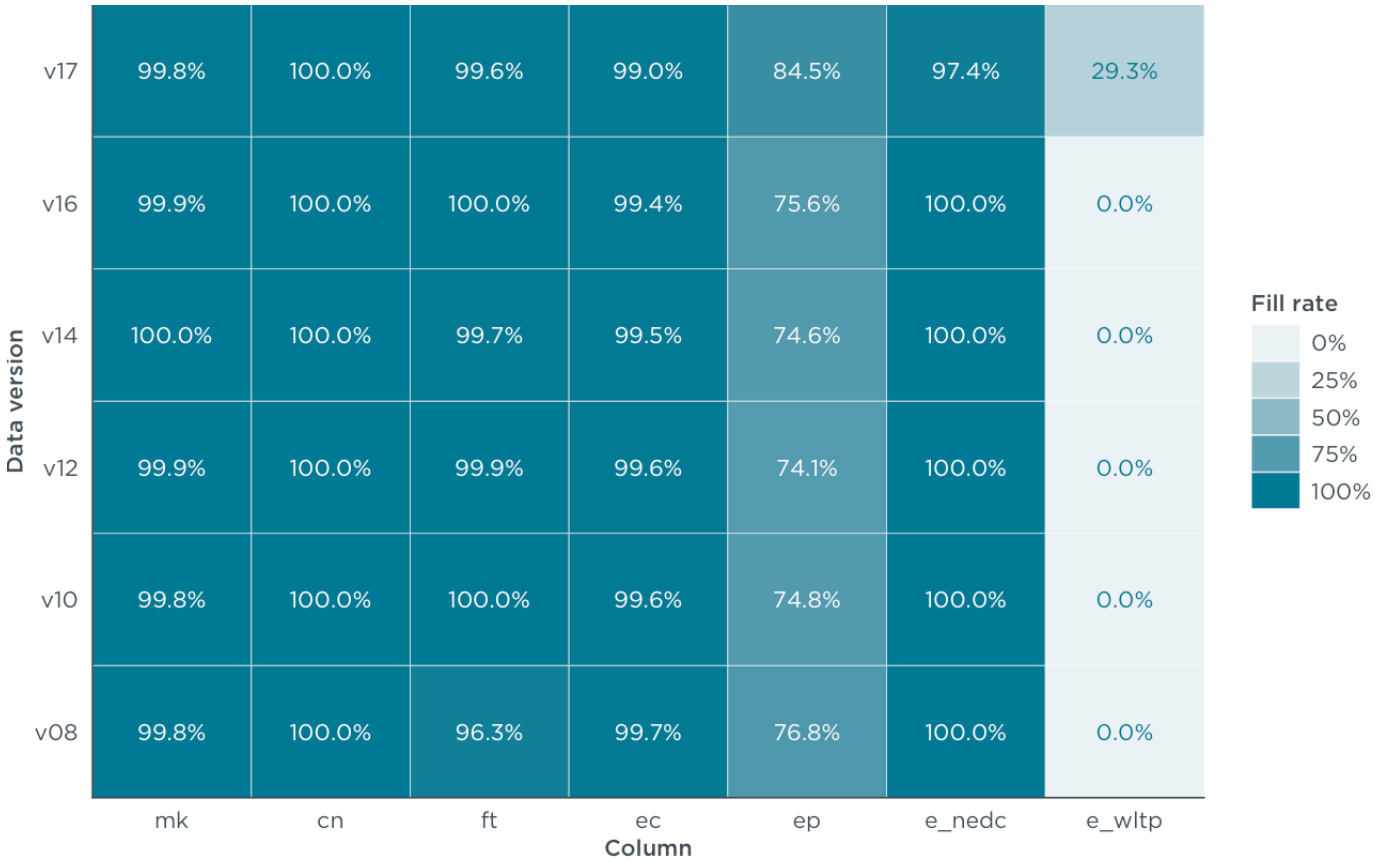
2.1 Overview of EEA data

The EEA CO2 monitoring data are documented on the EEA website (https://www.eea.europa.eu/data-and-maps/data/co2-cars-emission-16).

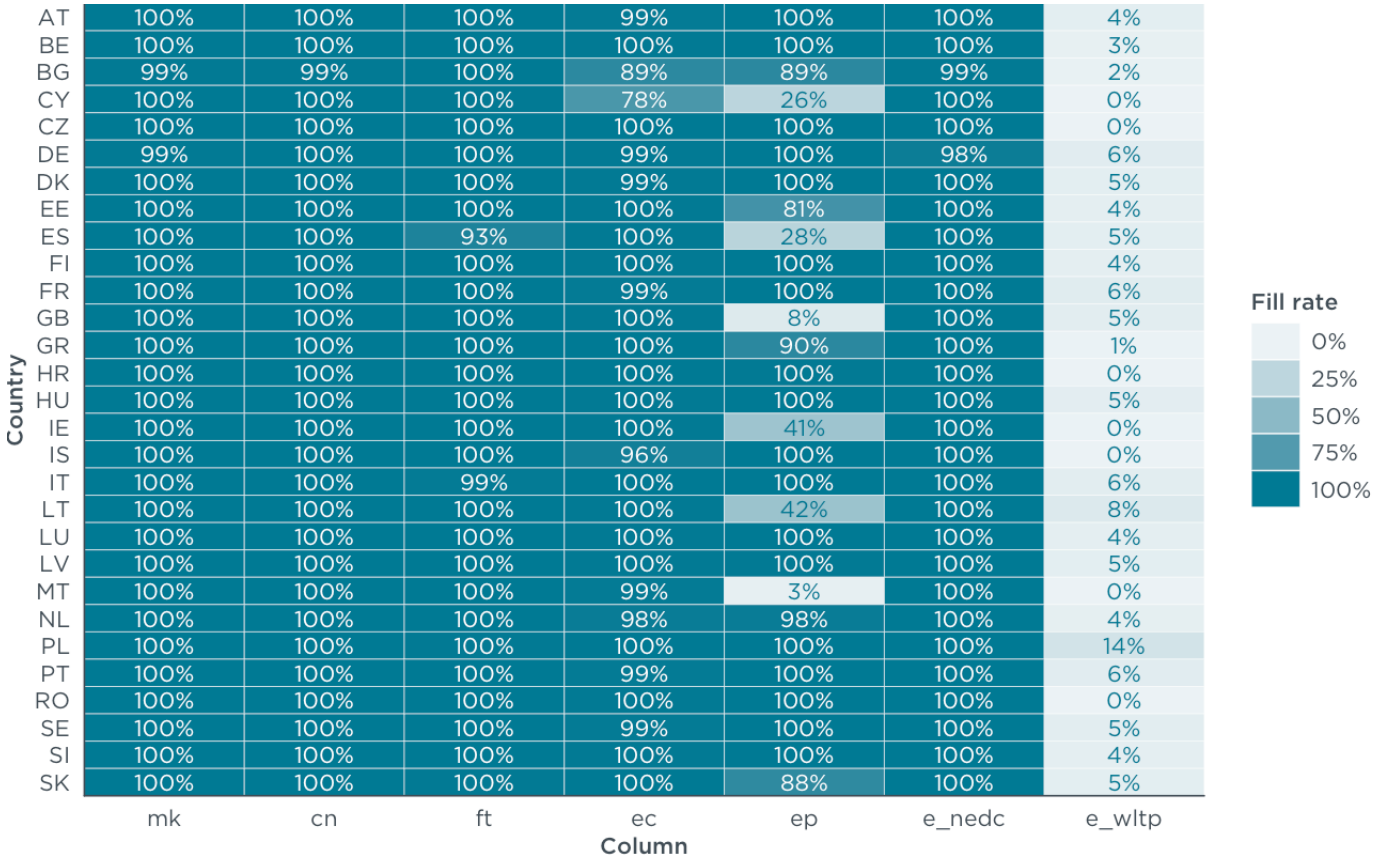
EEA data is released biannually: preliminary monitoring data for the preceding calendar year is released in spring and final monitoring data is released around fall. The datasets are versioned consecutively: v01 refers to preliminary 2010 data, v02 refers to final 2010 data, and so on. For this exercise, we use final data where available—we only use preliminary data for 2018. We only focus on data from 2013 forward.

The sales-weighted fill rate of essential columns are presented below. Columns are defined as follows: *mk* = make, *cn* = model, *ft* = fuel type, *ec* = engine displacement, *ep* = engine power, *r* = registrations, *e_nedc* = CO2 over NEDC, *e_wltp* = CO2 over WLTP.

The first plot shows the column fill rate—the sales-weighted share of non-missing values—per column and data version (year). The plot indicates that fill rates are generally close to 100%, but that engine power has a significant share of missing values.



The second plot shows the column fill rate per country. The plot indicates that missing engine power values in UK data are the primary cause of the poor engine power fill rate.



2.2 Join data

The ADAC and EEA data are joined on the following conditions:

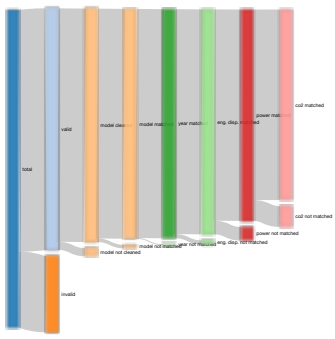
- Make and model are equal.
- Powertrain/fuel type are equal.
- The EEA year of registration is in range of ADAC build years.
- The engine displacement is equal.
- The engine power is equal.
- The NEDC and/or the WLTP CO2 value in the EEA data is in range of ADAC min/max values.
- Other MILE21 database grouping variables, namely transmission type and emission standard, can not be included in the join as these variables are not available in the EEA data.

Before the datasets were joined, make and model names in the EEA data were cleaned up.

3 Join data

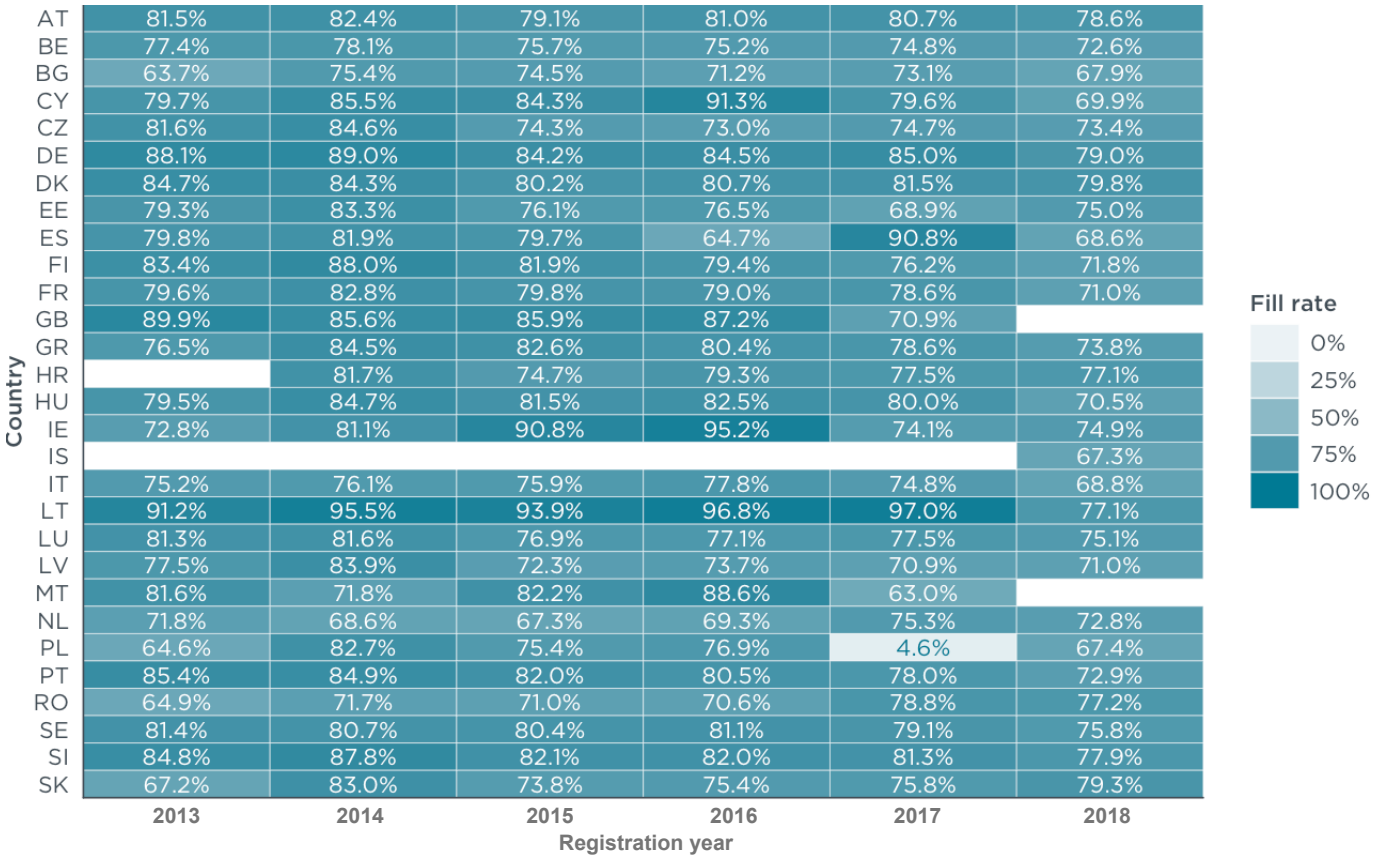
4 Results

The chart below shows the sales-weighted portion of EEA data successfully matched to ADAC data (top branches) and the portion of data that could not be matched to ADAC data (bottom branches) per added join variable.



In total, 76.0% of registered vehicles in the EEA data were valid, i.e., not marked as duplicates nor containing missing values in join variables. The significant portion of invalid data was primarily driven by missing engine power values in UK data. 78.2% of the valid data (59.4% of the total registrations) were matched in the ADAC data.

The figure below shows the market coverage per year and country. The values refer to percentage of registrations in valid EEA data matched in the ADAC dataset.



5 Interim conclusions

For the MILE21 project, interim results are positive in the sense that the ADAC data does not seem particularly skewed toward Germany: Countries like Cyprus, Italy, and Lithuania have considerably higher sales coverage than Germany in the ADAC data. The bad news for the MILE21 project is that EEA data quality—not the limited coverage of the ADAC data—seem to be the limiting factor for establishing coverage.