IMM Ambient Exposome database Application Procedure

Institute of Environmental Medicine (IMM)

Karolinska Institutet

Table of Contents

Karolinska Institutet

1. Aim of this document

The *IMM ambient exposome* database is a source of environmental exposure data available for research and risk assessment. The current document targets researchers exploring the possibility to link ambient environmental exposure data from the IMM ambient exposome database to their own study data for research purposes. It also targets risk assessors who want to estimate the health impact of environmental exposures on a population level. The document includes descriptions of available data, how to apply and prepare your data, conditions for usage and sharing, application forms, and a template for data transfer agreement.

2. IMM ambient exposome database

For the last decades, the Institute of Environmental Medicine (IMM) together with Center for Occupational and Environmental Medicine (CAMM) have developed methods to assess individual exposure to various environmental factors. These include air pollution, transportation noise, green/blue spaces and temperature which have been combined with other data on an individual or population level. Subsequently, the combined data were used to conduct epidemiological studies or health risk assessments resulting in scientific publications and reports.

IMM ambient exposome database aims to develop, describe, validate, and keep up to date all these environmental data pieces into one comprehensive database that is easy to use in current and future projects, as well as to keep track of previous research and corresponding methods of exposure assessment.

3. Description of the exposure assessment

In the next sections, we briefly describe the different exposure variables of air pollution, noise, natural spaces and ambient air temperature available in *the IMM ambient exposome database* (table 1). We further give a brief overview of the different exposure assessment methodologies applied for each indicator; a more detailed exposure assessment description can be found elsewhere: https://ki.se/en/imm/the-imm-ambient-exposome-database.

Table 1: List of spatial exposure variables available in the IMM exposome database

topic	indicator	Models	Source
Air pollution	$PM_{2.5}$, PM_{10}, Black carbon (BC), NO_2 and NO_X	Dispersion models	Model developed in collaboration with SLB-analys, City of Stockholm.
Air pollution	$PM_{2.5}$, PM_{10}, $PM_{2.5-10}$, NO_2, O_3	Spatiotemporal models	Models developed by IMM including monitoring sites, satellite data, dispersion models, land cover, meteorology, population, road traffic, etc.
Noise (road and railway)	$L_{Aeq,24h}$, L_{den}	Nordic Prediction Model (simplified)	Model developed in collaboration with Gothenburg University
Noise (aircraft)	L_{den}	Integrated Noise Model 7.0	Swedavia
Natural spaces	Greenness	Normalized difference vegetation index (NDVI)	Landsat 5, Landsat 7, and Landsat 8
Natural spaces	% Blue spaces	Coverage/proportion of water	Landsat 5, Landsat 7, and Landsat 8

topic	indicator	Models	Source
Temperature	Daily mean temperature (°C)	Spatiotemporal models	Models developed by IMM including monitoring sites, satellite data, land cover, meteorology, population, among others

In the data request form, the applicant will need to indicate which data from the IMM exposome database that are requested, and specify the following aspects:

- **Area coverage.** The applicant should carefully check the area coverage of each model. For example, the air pollution dispersion model data are only available for Stockholm and Uppsala County, whereas the spatiotemporal model data are available for the whole of Sweden. If the model covers larger areas, it goes at the expense of lower spatial resolution.
- **Time periods of the exposures.** Multiple time periods of exposures are available, and these vary depending on the exposure factor of interest. Carefully check the time period of availability in the descriptions below.
- **Temporal resolution of the exposures.** Depending on the aim of study, the applicant should differentiate short-term exposure studies (such as daily exposure levels) and long-term studies (annual exposure levels).

If needed, more guidance can be provided by the researchers of the IMM exposome database. Please contact exposome-imm@ki.se for further information.

Karolinska Institutet

3.1. Air pollution

3.1.1. Dispersion model

Dispersion modelling	
Model	*The assessment entails use of an emission inventory together with a Gaussian dispersion model to estimate outdoor levels of air pollutants from local sources and data on long-range contributions are added based on measurements at regional background stations. Meteorological input to the dispersion modelling includes measurements of wind velocity and direction, solar radiation, and temperature. Optionally, a street canyon contribution is added for addresses in the most polluted street segments in the inner city of Stockholm with multi-story houses on both sides.* *References:* *Segersson D et al. Int J Environ Res Public Health 2017, 14:7, https://doi.org/10.3390/ijerph14070742* *Ljungman PLS et al. Environ Health Perspect 2019, 127:10, https://doi.org/10.1289/EHP4757*
Indicators	*$PM_{2.5}$, PM_{10}, NO_2 and NO_X, as well as specific source contributions (e. g. road traffic exhaust, traffic road wear, residential heating)*
Source	*Model developed in collaboration with SLB-analys, City of Stockholm.*
Temporal resolution	*Annual averages. Using data from continuous measurement stations adjustments of the annual averages can be made to estimate levels during shorter time periods.*
Years of data	*Levels modelled 2000, 2011, 2018, and 2020. Interpolation is used to estimate levels between these years.* *Calculations for 1990 and 1995 are also available, but have to go through SLB-analys.*
Spatial resolution	*The calculations are performed on a 50-m resolution grid for addresses in the more densely populated areas of Stockholm County, such as urban areas, and a 500-m grid in less densely populated areas.*
Area coverage	*Stockholm County from 1990 and onwards, with other counties in eastern Sweden added after 2000 (such as Uppsala County).*
Map availability	*Yes*
What data does the applicant need to provide?	*Geocodes and years of follow-up*
Expected time to provide data	*1-2 months.*
Other	*Assistance can be provided on individual air pollution exposure assessment, such as calculation of time-weighted exposure during specific time windows.*
Responsible contact	*exposome-imm@ki.se*

3.1.2. Monitoring stations

Data on monitoring stations across Stockholm are available through SLB-analys (in Swedish Stockholms Luft- och Bulleranalys). We can provide guidance on how to obtain the data, but the data request will need to go through SLB-analys. For more information check their website: https://www.slb.nu/slbanalys/

3.1.3. Satellite-based spatiotemporal models

Spatio-temporal model	
Model	*This multi-stage methodology is based on a machine learning method – random forests – to estimate daily mean concentrations of 5 air pollutants. This model uses data from multiple sources, including satellite retrievals, dispersion models, land cover, meteorology, population, road traffic, among others. A methodology was developed to calibrate such data on measurements of air pollutants from existing monitoring networks. The calibration model was evaluated through cross-validation procedures, by holding out individual monitors, training the model on the others, and checking the performance of the model fit on the left-out sites. Then, the model was applied to all places and days with no observations, to obtain spatiotemporal predictions of air pollutants for all 1x1-km grid cells of Sweden. The performance of the model was fair to good (variable R^2 depending on the pollutants)* *Reference:* *Stafoggia et al. Atmosphere 2020, 11:239, doi:10.3390/atmos11030239*
Indicators	*Daily mean concentrations of $PM_{2.5}$, PM_{10}, $PM_{2.5-10}$, NO_2, O_3 ($\mu g/m^3$)*
Source	*Monitoring sites, satellite data, dispersion models, land cover, meteorology, population, road traffic, etc.*
Temporal resolution	*Daily*
Years of data	*From 2005 until 2019 (soon to be updated to 2020)*
Spatial resolution	*1km x 1km*
Area coverage	*The whole Sweden*
Map availability	*Yes*
What data does the applicant need to provide?	*Geocodes and years of follow-up*
Expected time to provide data	*1-2 months, if data requested are retrospective (2005-2019). Longer time, if they are after 2020 (models to be developed)*
Other	*Assistance can be provided on individual ambient air pollution exposure assessment, such as calculation of time-weighted exposure during specific time windows.*
Responsible contact	*exposome-imm@ki.se*

3.2. Noise

3.2.1. Transportation/railway/aircraft noise

Road, railway and aircraft noise at the façade	
Model	*For estimation of road and railway noise a database has been developed including information on terrain, ground surface, building density, traffic flows on roads (> 1000 vehicles/24 h) and railway lines, speed limits and to some extent noise barriers. Information on aircraft noise exposure is available as contours around Stockholm Arlanda and Bromma airports based on radar flight tracks, aircraft types, engine types, and runway distribution as well as spatial and temporal spread of the aircraft traffic.* *References:* *Ögren M, Barregård L. PloS One 2016, 11:5, https://doi.org/10.1371/journal.pone.0155328* *Pyko A et al. Environ Health Perspect 2017, 125:11, https://doi.org/10.1289/EHP10745*
Indicators	*For road and railway noise the 24-h A-weighted equivalent continuous sound pressure level ($L_{Aeq,24h}$) is estimated which can be recalculated to L_{den} (day-evening-night level). The L_{den} level is provided for aircraft noise.*
Source	*The assessment of exposure to road traffic and railway noise is based on a simplified version of the Nordic Prediction Model developed in collaboration with Gothenburg University. Swedavia provided aircraft noise estimates based on the Integrated Noise Model version 7.0.*
Temporal resolution	*Annual average*
Years of data	*Road traffic noise levels modelled 1990, 1995, 2000, 2005, 2010, 2018 and 2020, and less frequently for railway and aircraft noise.*
Spatial resolution	*Up to 10 x 10 m in the more densely populated areas*
Area coverage	*Road traffic noise maps for Stockholm and Uppsala region. Aircraft and railway noise for Stockholm County*
Map availability	*Yes, for road and aircraft noise, no for railway noise*
What data does the applicant need to provide?	*Geocodes and years of follow-up*
Expected time to provide data	*1-2 months.*
Other	*Assistance can be provided on individual noise exposure assessment, such as calculation of time-weighted exposure during specific time windows.*
Responsible contact	*exposome-imm@ki.se*

3.3.   Natural spaces (green and blue spaces)

Normalized difference vegetation index (NDVI) and blue space	
Model	*Data on spatial greenness as a normalized difference vegetation index (NDVI), with 25 x 25m resolution, were obtained from satellite images (Landsat 5, Landsat 7 and Landsat 8) for the period 1 May-30 September 1988–2019. To avoid underestimating NDVI values due to random cloud contamination, the annual value of each 25 x 25m pixel was replaced with a 5-year average value (2 years before, the current year, 2 years after). Because NDVI for water is represented by negative values, we replaced all water surfaces (lakes, streams, oceans) with "data-free" (no-data) surfaces.* *References:* *Chander G et al. Remote Sensing of Environment 2009, 113:5, https://doi.org/10.1016/j.rse.2009.01.007* *Persson A et al. Environment International 2018, 121:1, doi:10.1016/j.envint.2018.10.009*
Indicators	*Address based NDVI values (0-1) are calculated within corresponding buffers, for example circles of 100m, 250m, 500m or 1000m around coordinates.* *Coverage/proportion of water i.e. blue space could be calculated for each corresponding buffer.*
Source	*Processed satellite images (Landsat 5, Landsat 7 and Landsat 8)*
Temporal resolution	*Annual average*
Years of data	*Annual levels 1990-2017, extended to 1988, 1989, 2018, 2019 assuming no changes in floating averages.*
Spatial resolution	*25x25m*
Area coverage	*The whole Sweden*
Map availability	*Yes, Tiff*
What data does the applicant need to provide?	*Geocodes and years of follow-up*
Expected time to provide data	*1-2 months.*
Other	*Assistance can be provided on individual greenness exposure assessment, such as calculation of time-weighted exposure during specific time windows.*
Responsible contact	*exposome-imm@ki.se*

3.4. Temperature

3.4.1. Monitoring stations

Data on monitoring stations across Stockholm are available through SLB-analys (in Swedish Stockholms Luft- och Bulleranalys). We can provide guidance on how to obtain the data, the data request will need to go through SLB-analys. For more information check their website: https://www.slb.nu/slbanalys/

3.4.2. Satellite-based spatiotemporal models

Spatio-temporal hybrid model	
Model	*This multi-stage methodology is based on a machine learning method – random forests - to estimate daily mean ambient air temperature. This model uses data from multiple sources, including satellite retrievals, land cover, meteorology, population, among others. A methodology was developed to calibrate such data on measurements of temperature from existing monitoring networks. The calibration model was evaluated through cross-validation procedures, by holding out individual monitors, training the model on the others, and checking the performance of the model fit on the left-out sites. Then, the model was applied to all places and days with no observations, to obtain spatiotemporal predictions for all 1 km x km grid cells in Sweden. The performance of the model was excellent for both days with and without available temperature observations (R^2 ~ 0.95)* *Reference:* *de Bont, J et al. Environmental Research, 213, https://doi.org/10.1016/J.ENVRES.2022.113586*
Indicators	*Daily mean temperature (°C).*
Source	*Monitoring sites, satellite data, land cover, meteorology, population, among others*
Temporal resolution	*Daily*
Years of data	*From 2005 until 2019 (soon to be updated to 2020)*
Spatial resolution	*1km x 1km.*
Area coverage	*The whole Sweden*
Map availability	*Yes*
What data does the applicant need to provide?	*Geocodes and years of follow-up*
Expected time to provide data	*1-2 months, if data requested are retrospective (2005-2019). Longer time, if they are after 2020 (models to be developed)*
Other	*Assistance can be provided on individual ambient air temperature exposure assessment, such as calculation of time-weighted exposure during specific time windows.*

10

Responsible contact	*exposome-imm@ki.se*

4. Policy and conditions

The general conditions for collaboration with the IMM Ambient Exposome database (A-EDB) can be found in Annex A. Start the process by reading the agreement document so that all conditions can be met. The IMM A-EDB Review Committee meets four times per year and decides on all requests.

In order to process your request, a complete application must be received by the Review Committee no later than one week before a scheduled Committee meeting (check website for details). For more information, please contact us by e-mail to exposome-imm@ki.se.

5. How to apply?

To be complete, an application must contain the following:

Step 1: download the Data request - IMM Exposome database template (Annex B)

Step 2: prepare documents
1. Fill out the data request template. The applicant will need to provide a description of the project including:
   a) applicant information
   b) specific aims
   c) general background
   d) methods and statistical analyses
   e) requested exposure variables

2. Provide 2-page CV.

Step 3: send the data request and CV to exposome-imm@ki.se.

Step 4: after approval. Upon approval, both parties (IMM and the applicant) sign the Data Transfer Agreement (Annex A) and the applicant will be contacted for further steps and provided with a timeline.

6. What data do we need from the applicant?

To apply the environmental data to a study dataset, applicants will need to provide **geocoded addresses** (typically residential histories) of study participants with the exact **latitude** and **longitude** location (for example, IMM geocoded address = 671946, 671946, WSG84). Geocoding an address is the process of taking a text-based description of a location, such as an address or the name of a place, and returning the exact latitude and longitude location, to identify a location on the Earth's surface. In the data request form, the applicant needs to specify if the addresses are geocoded and the source of the addresses (registry data,

Karolinska Institutet

questionnaire data/self-reported, address data is not available or others). The applicant also needs to provide information about the **coordinate system** and **projection** used. The coordinate system used in by the IMM Exposome Group is Sweref 99 TM (EPSG 3006).

IMPORTANT: The IMM Exposome group does <u>not</u> provide a service to geocode addresses. However, we can give support and contacts how to geocode and to perform an accurate manual cleaning, which is needed to provide sufficient quality of the geocoding process. Please contact us for further guidance.

7. How to obtain addresses for study subjects and how to geocode them?

1) The Researcher transfers a file containing personal identification numbers (personnummer) of the study participants to the Swedish Tax Agency (Skatteverket) and unique study id numbers.
2) The Tax Agency extracts information from the Population Registry (Folkbokföringsregistret) for every individual and for selected years on for example the following variables (as agreed on beforehand):
    a. Street address (Gatuadress)
    b. Zip code (Postnummer)
    c. Town (Postort)
    d. County code (Länskod)
    e. Municipality code (Kommunkod)
    f. Registration date (Folkbokföringsdatum), the date the person moved to the address
3) The Tax Agency removes the personal identification numbers and only the study id numbers are kept as identifiers.
4) The Tax Agency transfers the file back to the Researcher.
5) The Researcher transfers the file to the company Metria. Addresses are matched with address coordinates from the Property Registry (Fastighetsregistret). The variables added are X and Y SWEREF coordinates and an indicator of the quality of the match, called rank.
6) The file is transferred back to the Researcher. The quality of the match, indicated by the rank variable, ranges from 0-1, depending on the precision of the rank. Below 0.8 is considered low rank, and these coordinates may be imprecise or inaccurate. The Researcher needs to decide how to handle these (exclude, correct manually, etc).
7) Once the file has been quality-checked and imprecise and inaccurate coordinates have been handled, the Researcher may send the coordinates to the IMM exposome team.

Karolinska Institutet

8. Bibliography

Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., Tsai, M.-Y., Künzli, N., Schikowski, T., Marcon, A., Eriksen, K. T., Raaschou-Nielsen, O., Stephanou, E., Patelarou, E., Lanki, T., Yli-Tuomi, T., Declercq, C., Falq, G., Stempfelet, M., … de Hoogh, K. (2013). Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project. *Atmospheric Environment*, *72*, 10–23. https://doi.org/10.1016/j.atmosenv.2013.02.037

Chander, G., Markham, B. L., & Helder, D. L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment*, *113*(5), 893–903. https://doi.org/10.1016/J.RSE.2009.01.007

de Bont, J., Stafoggia, M., Nakstad, B., Hajat, S., Kovats, S., Part, C., Chersich, M., Luchters, S., Filippi, V., Stephansson, O., Ljungman, P., & Roos, N. (2022). Associations between ambient temperature and risk of preterm birth in Sweden: A comparison of analytical approaches. *Environmental Research*, *213*, 113586. https://doi.org/10.1016/J.ENVRES.2022.113586

de Hoogh, K., Chen, J., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., van Donkelaar, A., Hvidtfeldt, U. A., Katsouyanni, K., Klompmaker, J., Martin, R. v., Samoli, E., Schwartz, P. E., Stafoggia, M., Bellander, T., Strak, M., Wolf, K., Vienneau, D., … Hoek, G. (2018). Spatial PM2.5, NO2, O3 and BC models for Western Europe – Evaluation of spatiotemporal stability. *Environment International*, *120*, 81–92. https://doi.org/10.1016/J.ENVINT.2018.07.036

Eeftens, M., Beelen, R., de Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., Declercq, C., Dėdelė, A., Dons, E., de Nazelle, A., Dimakopoulou, K., Eriksen, K., Falq, G., Fischer, P., Galassi, C., Gražulevičienė, R., Heinrich, J., Hoffmann, B., Jerrett, M., … Hoek, G. (2012). Development of Land Use Regression Models for PM $_{2.5}$, PM $_{2.5}$ Absorbance, PM $_{10}$ and PM $_{coarse}$ in 20 European Study Areas; Results of the ESCAPE Project. *Environmental Science & Technology*, *46*(20), 11195–11205. https://doi.org/10.1021/es301948k

Ljungman, P. L. S., Andersson, N., Stockfelt, L., Andersson, E. M., Sommar, J. N., Eneroth, K., Gidhagen, L., Johansson, C., Lager, A., Leander, K., Molnar, P., Pedersen, N. L., Rizzuto, D., Rosengren, A., Segersson, D., Wennberg, P., Barregard, L., Forsberg, B., Sallsten, G., … Pershagen, G. (2019). Long-Term Exposure to Particulate Air Pollution, Black Carbon, and Their Source Components in Relation to Ischemic Heart Disease and Stroke. *Environmental Health Perspectives*, *127*(10). https://doi.org/10.1289/EHP4757

Ögren, M., & Barregard, L. (2016). Road Traffic Noise Exposure in Gothenburg 1975–2010. *PLOS ONE*, *11*(5), e0155328. https://doi.org/10.1371/JOURNAL.PONE.0155328

Persson, Pyko, A., Lind, T., Bellander, T., Östenson, C. G., Pershagen, G., Eriksson, C., & Lõhmus, M. (2018). Urban residential greenness and adiposity: A cohort study in Stockholm County. *Environment International*, *121*(Pt 1), 832–841. https://doi.org/10.1016/J.ENVINT.2018.10.009

Pyko, A., Eriksson, C., Lind, T., Mitkovskaya, N., Wallas, A., Ögren, M., Östenson, C. G., & Pershagen, G. (2017). Long-Term Exposure to Transportation Noise in Relation to Development of Obesity—a Cohort Study. *Environmental Health Perspectives*, *125*(11), 117005. https://doi.org/10.1289/EHP1910

Segersson, D., Eneroth, K., Gidhagen, L., Johansson, C., Omstedt, G., Nylén, A. E., & Forsberg, B. (2017). Health impact of PM10, PM2.5 and black carbon exposure due to different source sectors in Stockholm, Gothenburg and Umea, Sweden. *International Journal of Environmental Research and Public Health*, *14*(7). https://doi.org/10.3390/ijerph14070742

Stafoggia, M., Johansson, C., Glantz, P., Renzi, M., Shtein, A., Hoogh, K. de, Kloog, I., Davoli, M., Michelozzi, P., & Bellander, T. (2020). A random forest approach to estimate daily particulate matter, nitrogen dioxide, and ozone at fine spatial resolution in Sweden. *Atmosphere*, *11*(3), 239. https://doi.org/10.3390/atmos11030239

9. Annex A: Data Transfer Agreement IMM Ambient Exposome Database and Recipient

DATA TRANSFER AGREEMENT

This Data Transfer Agreement (the "Agreement") is concluded between:
(1) Institute of Environmental Medicine, Karolinska Institutet (IMM), 171 77 Stockholm; and
(2) <mark>XXX</mark> (the "Recipient") regarding Exposome data from IMM; below jointly called the "Parties".
Contact at IMM Petter Ljungman, telephone: 073-762 15 37, e-mail: petter.ljungman@ki.se; and at the Recipient <mark>XXX</mark>.

IMM has, in collaboration with, the Centre for Occupational and Environmental Medicine, Region Stockholm (CAMM), Gothenburg University and SLB-analys, City of Stockholm, collected, processed and analyzed data regarding exposure to various environmental factors, so-called Exposome data. Within the framework of the Agreement, Exposome data specified in Appendix 1 are transferred from IMM to the Recipient. The Parties agree on the following conditions for the data transfer:

1. Use of data

   Data shall be used in accordance with the specification in Annex 1. This may mean:

   a. Exposure estimation in the framework of epidemiological investigations of relationships between environmental factors and different health outcomes. This means that Exposome data are handled for individuals based on address information. No exposure information for individuals will be disclosed in the publications/reports, and all data will be reported at group level. The Recipient is responsible for ensuring that ethical permission exists for all data processing and that personal data are handled in accordance with the EU Data Protection Directive.
   b. Risk assessment at the population level. Based on Exposome data, estimations can be made of the exposure of the population to various environmental factors. Together with knowledge of exposure-response relationships, quantitative risk assessment can be carried out in relevant population groups. No individual-level information is handled in these calculations.

   No commercial use may be made of data transferred within the framework of the Agreement.

Karolinska Institutet

2. Access to data

The Recipient is responsible for ensuring that Exposome data within the framework of the Agreement are only used in accordance with paragraph 1 and that IMM must be involved in all cooperation where the transferred data are utilized. Access to Exposome data is given based on a standardized application procedure and approval after evaluation in a designated review group.

No onward transfer from the Recipient of data within the framework of the Agreement may take place.

3. Publication of data

Results from epidemiological investigations and risk assessments based on transferred Exposome data within the framework of the Agreement are expected to be published in international scientific journals and various reports. A representative of IMM as well as of CAMM and/or SLB-analys or in the case of road traffic and railway noise Mikael Ögren from Gothenburg University, depending on the type of Exposome data transferred, shall be offered co-authorship on all these publications, and given the opportunity to review prior to publication. When dealing with the media regarding studies or online dissemination of results were transferred Exposome data have been used, e.g. in press releases, the corresponding institutions should be mentioned.

4. Validity of the Agreement

The Agreement is valid for three years (36 months) from the time of signature by the Parties. It can subsequently be extended. The contract may be terminated by either party upon written request at least 30 days in advance. No part of the Agreement may be amended or modified without written agreement between the Parties.

Institute of Environmental Medicine, Karolinska Institutet   XXX
Date: Date:
Location: Location:

Anna Bergström, Head and director of IMM   XXX

Petter Ljungman, Associate professor   XXX

15

10. Annex B: Data Request

Institute of Environmental Medicine

Data request - IMM Exposome database template

The applicant should read the document IMM Ambient Exposome database Application Procedure before filling in the data request form. This clarifies which exposures can be provided by IMM and what data are needed to estimate the exposures.

The data request form needs to be completed by the principal investigator of the project. The IMM Ambient Exposome Database Review Committee meets approximately every 3 months to discuss the feasibility of each project (check the dates for Review Committee meetings: https://ki.se/en/imm/the-imm-ambient-exposome-database). A formal answer will be sent to the applicant regarding the use of IMM Exposome data.

Date of data request form (DD/MM/YYYY):

Personal information of the principal investigator (PI):
(The PI must have a PhD, if not, please provide a collaborator with a PhD)

Work institution:	
PI Surname:	PI First name:
E-mail:	Telephone:
Working address:	
Postcode, town, and country:	

Description of the project *(give a short overview of the aims of the study, the background, and the methodology):*

Title of the project:

The background and scientific rationale of the proposed research project (up to 5000 characters or 300 words):

Aims of the study:

A brief description of the methods including data source, study design and statistical analyses (up to 5000 characters or 300 words):

Exposure assessment *(give information about the addresses and the requested environmental exposures)*

Do you have address information during the relevant follow-up time?	
☐	Yes
☐	No

	No, but data will be available in the near future Please, specify:
☐	

What is the source of the addresses available in your study?	
☐	Registry data
☐	Questionnaire data/self-reported
☐	Address data are not available
☐	Others Please, specify:

Are the addresses geocoded (e.g., information on the exact longitude and latitude)? *(The IMM Exposome group does not provide service to geocode the addresses. However, we can give support and contacts how to geocode and to perform an accurate manual cleaning, which is needed to provide sufficient quality of the address geocodes)*

☐	Yes
☐	No

What is the coordinate system and projection used when geocoding the data?

Please indicate which environmental exposures from the IMM exposome database you are requesting? **(Multiple answers are possible)** *(check the IMM Exposome protocol for more information)*

☐	Air pollution - Dispersion model
☐	Air pollution - Spatiotemporal model
☐	Air pollution - Monitoring stations (only available through SLB-analys)
☐	Noise – Road traffic
☐	Noise - Railway
☐	Noise - Aircraft
☐	Green spaces
☐	Blue spaces
☐	Ambient air Temperature - Spatiotemporal model
☐	Ambient air Temperature - Monitoring stations (only available through SLB-analys)

What is the area coverage of the applicant's data? *(Example: Stockholm County, the whole of Sweden, etc.)*

For which time periods should the exposures be estimated? *(Example from 2010 until 2019, or only specific years)*

At which temporal resolution should the exposures be estimated? *(Example: daily, weekly, monthly, or annually)*

Please, clarify any other specific requests:

When would the applicant need the environmental exposures *(Check the document IMM Ambient Exposome database Application Procedure on the expected time to provide data).*

Karolinska Institutet

Ethical approvals:

Remember that linkage of geographical data to health data where individuals can be identified are typically subject to requirement of ethical approval. The applicant PI is responsible for obtaining any ethical approval.

Thank you for completing the data request form.

For any urgent matters, please contact exposome-imm@ki.se