universität innsbruck
Arbeitsbereich für Energieeffizientes Bauen

universität innsbruck
Arbeitsbereich für Energieeffizientes Bauen

R package "annex"

Aggregating and harmonizing IAQ measurement data

compatible with an existing repository containing IAQ data from ~10.000 residential homes

Gabriel Rojas
Constanza Molina Carvallo, Núria Casquero-Modrego, Marcel Loomans, Maria Justo Alonso, Francesco Babich, Aurora Monge Barrio, Gabriel Bekö, Núria Casquero-Modrego, Marie Coggins, Bart Cremers, Andreas Frei, Timm Freundorfer, Eol Geffre, Sascha Hammes, Hala Hassan, Dusan Licina, Marcel Loomans, Lena Mayr, Gráinne McGill, James McGrath, Constanza Molina, Charles-Florian Picard, Mohsen Pourkiaei, Sasan Sadrizadeh, Reto Stauffer, Linda Toledo, Kevin Verniers

AIVC Webinar "IAQ databases", online 20.10.2026

EBC
Bundesministerium Verkehr, Innovation und Technologie
IEA FORSCHUNGS KOOPERATION
FFG Forschung wirkt.

1

Agenda

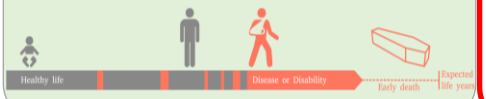
- Motivation / IEA EBA Annex 86
- Approach for aggregating and harmonizing IAQ measurement data
- Implementation of this approach in open source R package
- Application of R package to create large dataset
- Exemplary results from this large dataset

2

IEA EBC Annex 86 – Energy Efficient Indoor Air Quality Management in Residential Buildings

ST 1: Metrics

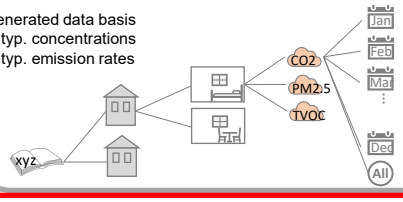
Develop health based rating system for IAQ performance
DALY based rating, see Morantes, G et al., 2023,
https://doi.org/10.1021/acs.est.3c07374



Von Original: Planemad Vektor, https://commons.wikimedia.org/w/index.php?curid=20278903

ST 2: Source characterization

Generated data basis
- typ. concentrations
- typ. emission rates

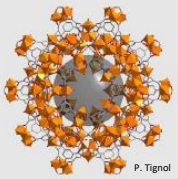


ST 5: Big Data

evaluate potentials
- Cloud-data
- IoT-technologies

ST 3: Smart materials

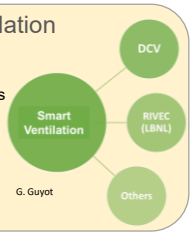
investigate „IAQ-active“ materials, e.g. MOF-based



P. Tignol

ST 4: Smart ventilation

investigate smart ventilation strategies
- comparison of IAQ, comfort and energy in operation
- robustness, long-term performance



G. Guyot

ST 6: Dissemination

EBC

Final reports available under: https://annex86.iea-ebc.org/ in particular see D3 & D4

Motivation

Problem

Need for open IAQ measurement data

- modelling / simulation: e.g. What are typical concentrations / emission rates?
- benchmarking: e.g. How do "my" measurements compare to other homes?
- (harm-based) assessment frameworks: e.g. What is an acceptable risk?
- policymaking: e.g. Which policies / technologies have most impact?

Many fragmented datasets out there, but available analysis/results often not comparable

Potential solution

Harmonize datasets by applying a standardized structure / processing

Potential challenges

Data protection, scalability / continuity

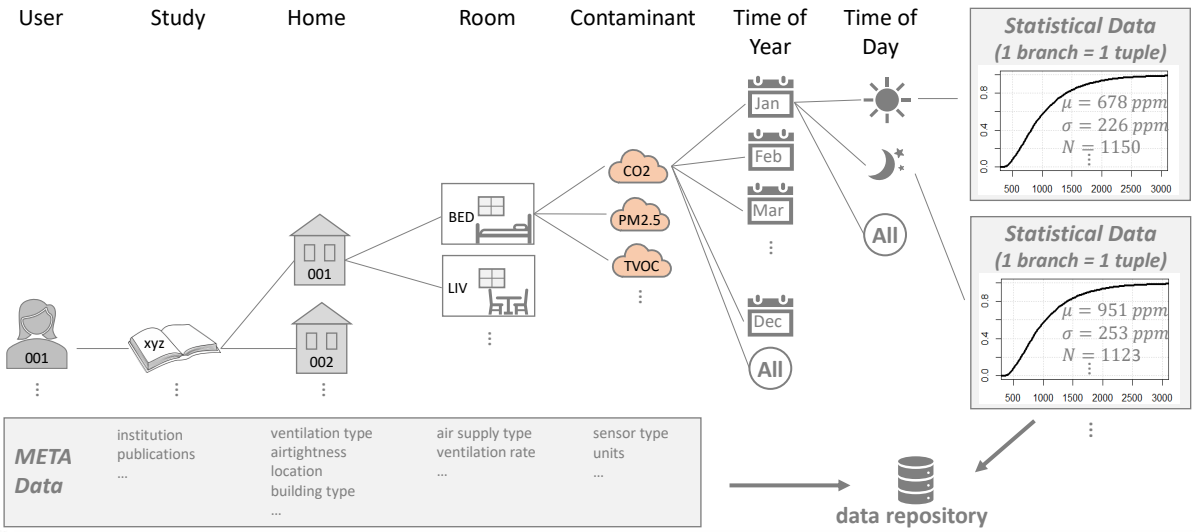
Our approach / Our vision

- decentral processing (data can stay with owner)
- aggregation on monthly basis
- maintain distributional information (eCDF/ePDF)
- implement in an easy to use script (well documented, open source)
- share aggregated data (repositories with DOI, IAQ database ...)
- IAQ science community continues to build upon and extend datasets

https://iea-ebc-annex86.github.io/annex/index.html

https://www.r-project.org/

Data structure



can accommodate time series data AND single value measurements (e.g. active/passive sampling)

Data processing

- Invalid non-numeric entries or data gaps removed
- For each tuple:
  - n≥30 -> all statistical outputs
  - n<30 -> reduced output
- lower and upper plausibility bounds -> warnings

Measured Variable	Additional info. req'd	Lower bound	Upper bound	Default (allowed) units
CO_2	No	350	1.0E+04	ppm (%)
CO	No	0	1.0E+02	µg/m³ (mg/m³, ppm, ppb)
HCHO	No	0	1.0E+03	µg/m³(mg/m³, ppm, ppb)
UFP	No	0	1.0E+09	#/cm³
PM_1	No	0	1.0E+04	µg/m³ (mg/m³)
$PM_{2.5}$	No	0	1.0E+04	µg/m³ (mg/m³)
PM_{10}	No	0	1.0E+04	µg/m³ (mg/m³)
PM_{Other}	Yes	n.a.	n.a.	n.a.
NO_2	No	0	1.0E+04	µg/m³ (mg/m³)
NO_x	No	0	1.0E+04	µg/m³ (mg/m³)
TVOC	Yes	0	1.0E+04	µg/m³ (mg/m³)
VOC	Yes	0	1.0E+04	µg/m³ (mg/m³)
O_3	No	0	1.0E+02	µg/m³ (mg/m³, ppm, ppb)
RH	No	5	105	%, -
Radon	No	0	1.0E+04	Bq/m³ (pCi/L)
Fungi	No	0	1.0E+09	CFU/m³
Ions	No	0	1.0E+09	1/cm³
Other	Yes	n.a.	n.a.	n.a.
T	No	-50	50	C, K, F
Pressure	No	500	1100	hPa, Pa, mbar, mmHg
SolRad	Yes	0	1500	W/m2
Flow	No	0	500	l/s, m3/h, cfm

Data processing - Output of statistical analysis (of time series data)

- arithmetic mean
- standard deviation
- percentiles (in 1% steps, plus 0.5%, 2.5%, 97.5% and 99.5%)

- various quality indicators
  e.g. start- and end-date, intervals, (estimated) number of data points, exceedance of plausibility bounds

user	study	home	room	year	month	tod	variable	quality_lower	quality_upper	quality_start	quality_end	interval_Min	interval_Q1	interval_Median	
0010	NL_ML	H0100	BED2	2012	11	all	CO2	0	0	28-11-2012	30-11-2012	58	180	180	
0010	NL_ML	H0100	BED2	2012	11	all	RH	0.08	0	28-11-2012	30-11-2012	58	180	180	
0010	NL_ML	H0100	BED2	2012	11	all	T	0	0	28-11-2012	30-11-2012	58	180	180	•••

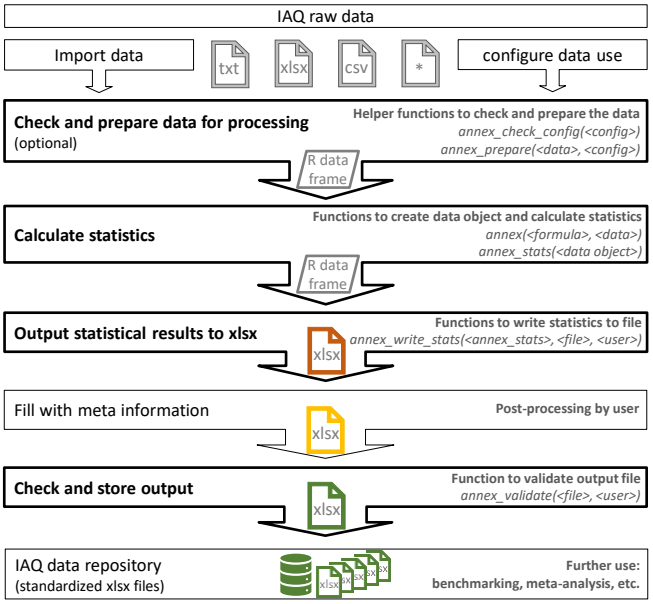
1 "data tuple"

interval_Median	interval_Mean	interval_Q3	interval_Max	Nestim	N	NAs	Mean	Sd	p00	p00.5	p01	
180	155.6	180	180	1440	1191	1	518.9146	92.66	368.4967	370.6336	372.9723	
180	155.6	180	180	1440	1191	0	39.7517	3.3566	0	34.73	34.825	
•••	180	155.6	180	180	1440	1191	0	19.8708	0.4882	4.69	19.39	19.39 •••

Implementation in R package

- decentral processing (data protection)
- set of easy to use functions
- open-source (continued improvement)
- flexible with different data (formats)
- output easy to integrate in database (but also human-processable: xlsx)

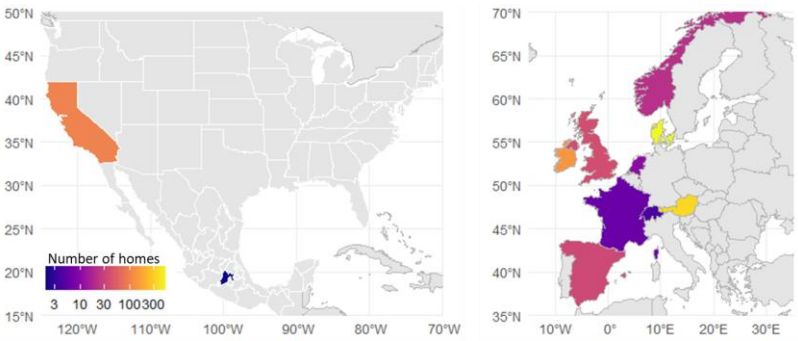
https://github.com/IEA-EBC-Annex86/annex



Rojas et al. (2026). Towards a harmonized database ... https://doi.org/10.1016/j.buildenv.2026.114297

Pooled available IAQ measurements

- 11 countries
- 20 studies
- 1160 homes (IAQ studies)
- +8500 homes (operational data)
- 220 000 room-month tuples
- 44 IAQ parameters



Country	N	Study acronyms	Ventilation type fractions [%]				Building type fractions [%]			
			Win	Nat	Hyb	Mech	SFH	TH	MFH	AB
AUT	332	Lod_Kuf, Lue3, Sinfonia	30	0	0	70	25	0	68	8
CHE	4	Büren	50	0	0	50	0	0	100	0
DNK	499	IECH	69	16	13	2	68	0	18	12
ESP	33	CISC, CliReady	36	30	12	21	0	0	73	27
FRA	6	ALLO	0	0	100	0	0	0	0	0
GBR	37	VentStd, AMR, SC1	0	22	0	78	0	24	0	76
IRL	129	VALIDate, ALIVE, ARDEN	10	21	5	64	74	26	0	1
MEX	2	MX1	0	50	50	0	0	0	0	100
NLD	10	NL_ML	0	0	0	100	0	100	0	0
NOR	21	HO_COV	0	52	24	24	33	24	10	33
USA	93	HENGH, LIA, SF1	0	0	0	100	75	0	0	25

N: Number of homes; Win: Window airing (not designed), Nat: Natural ventilation (designed), Hyb: Hybrid/mixed mode ventilation, Mech: Mechanical ventilation; SHF: single family house, TH: terraced house, MFH: multi family house, AB: apartment block

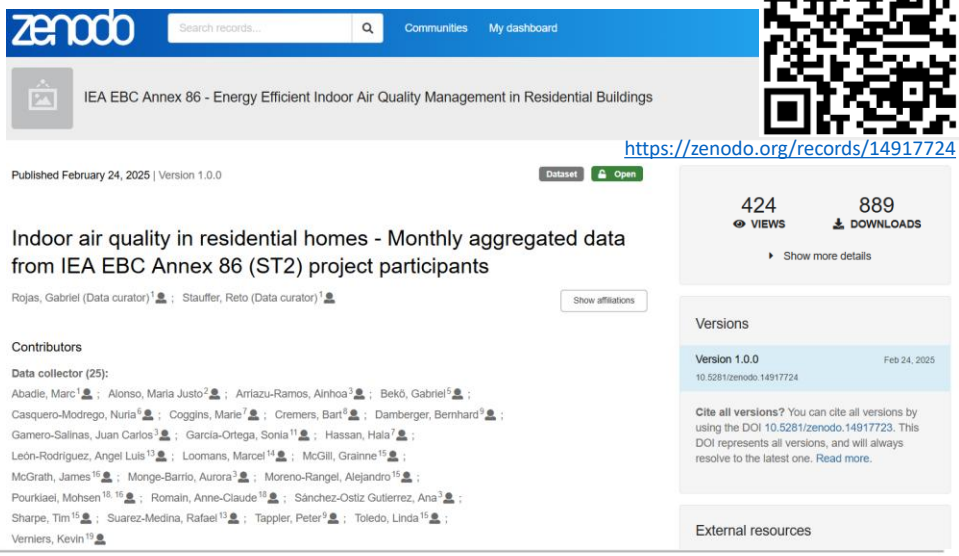
Rojas et al. (2026). Towards a harmonized database ... https://doi.org/10.1016/j.buildenv.2026.114297

Harmonized dataset created ("IAQ database")

Dataset available under:

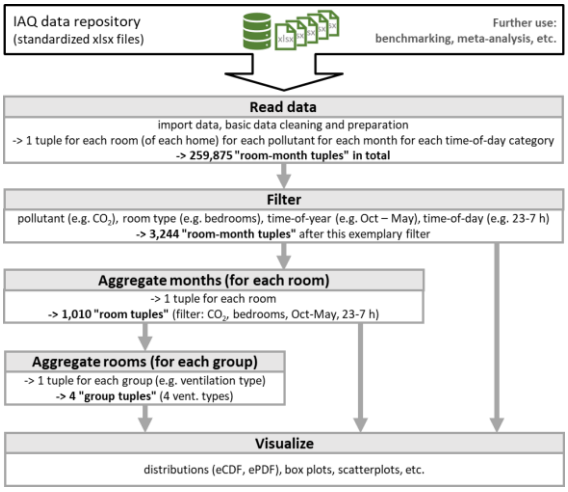
https://zenodo.org/records/14917724

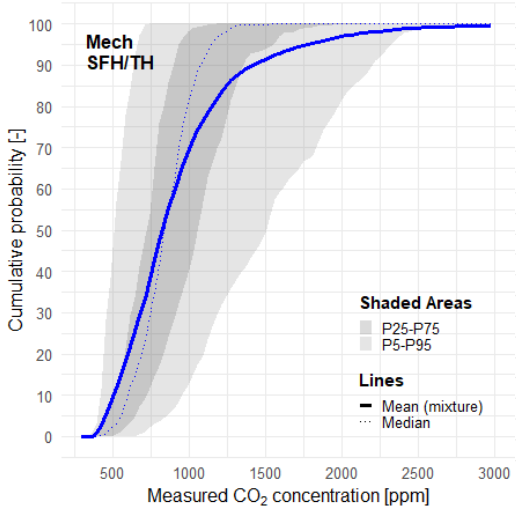
- 11 countries
- 20 studies
- 1160 homes (IAQ studies)
- +8500 homes (operational data)
- 220 000 room-month tuples
- 44 IAQ param.



Rojas et al. (2026). Towards a harmonized database ... https://doi.org/10.1016/j.buildenv.2026.114297

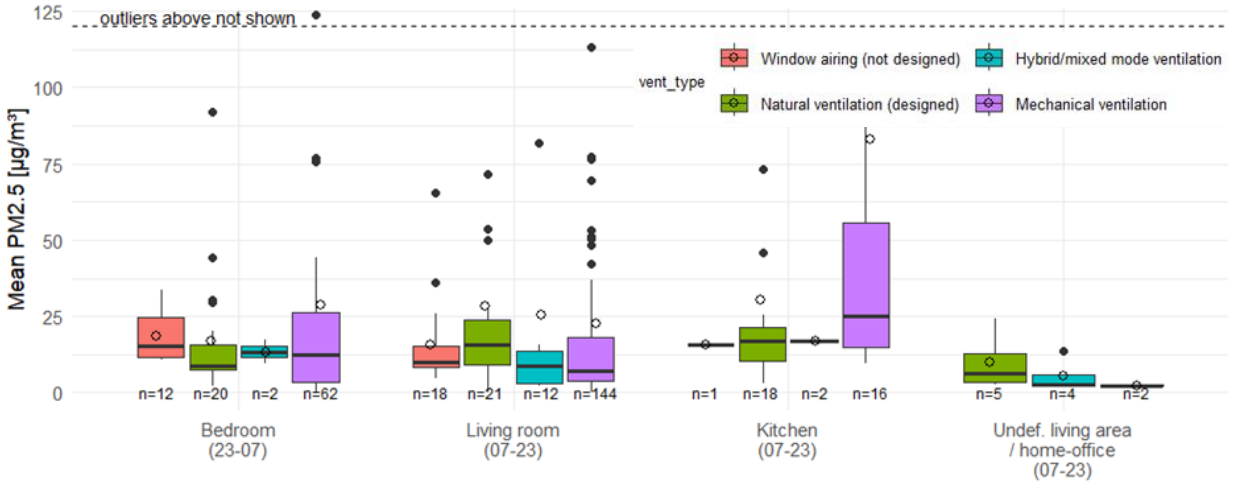
Exemplary results: CO_2 of ~1000 bedrooms





Rojas et al. (2026). Towards a harmonized database ... https://doi.org/10.1016/j.buildenv.2026.114297

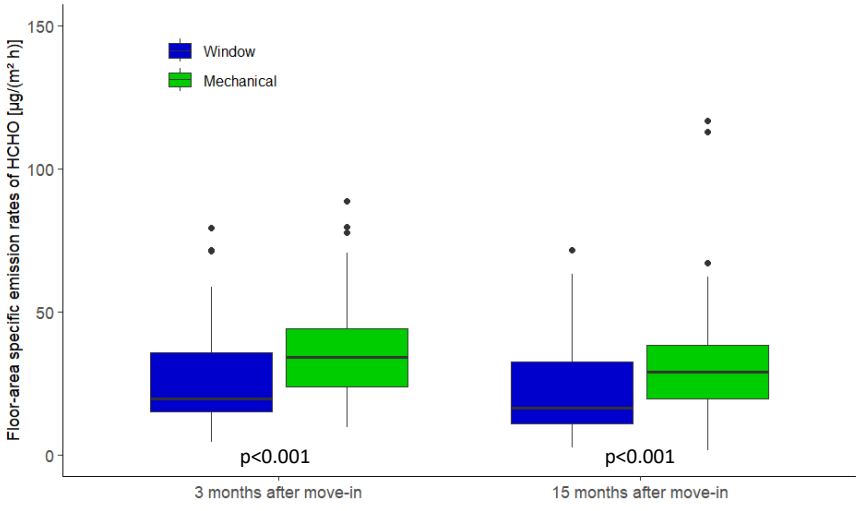
Exemplary results: PM2.5 in different rooms



Laverge, J. et al. (Ed.) (2025). *D3 - Methods and tools for the rating of IAQ management. IEA EBC Annex 86 Report.* https://annex86.iea-ebc.org/publications

Exemplary results: Formaldehyde emission rates

from measurements in Lüftung3.0 dataset

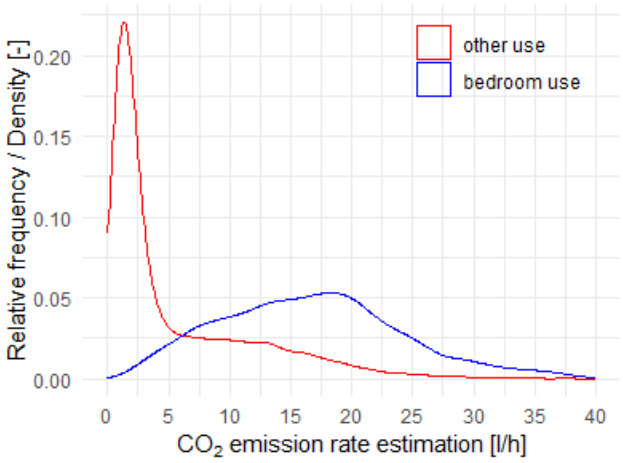


Median [µg/(m²h)]	Win	Mech
3 mo	19.8	34.5
15 mo	16.4	29.0

Compare with Pandora analysis:
Approach 1: 28.5 µg/(m²h)

Rojas, G. et al. (2026). Empirical Determination of Emission Rates of Formaldehyde and VOC Sources from over 120 Bedrooms. In *Proceedings of the 15th REHVA HVAC World Congress—CLIMA 2025* (Vol. 788, pp. 480–488). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-10546-2_45

Exemplary results: Empirical CO_2 emission rates
from operational data (8500 homes)



Rojas, G. Mayr, L., Verniers, K., Pollet, I. (2026), Estimating CO2 peak concentration in residential bedrooms using simple statistical analysis of large time series data sets, Indoor Air conference

Summary

We developed analysis method for IAQ data
- to harmonize measurement and meta data
- to aggregate ts-data to monthly summaries (maintaining distrib. information)
- implemented in simple-to-use R package

It allows a decentralized processing for wider/public use
- owner only hands out aggregated data (avoiding pot. data protection issues)
- yet ensures compatibility and scalability

You have IAQ measurement data?
Test our R package and provide feedback (for upcoming updates/improvements)
Use for e.g. benchmarking your IAQ data with up to ~10.000 other homes

Thank you!

Further information / instructions

Github repository https://github.com/IEA-EBC-Annex86/annex

Paper B&E https://doi.org/10.1016/j.buildenv.2026.114297

Annex 86 – D3 & D4 report https://annex86.iea-ebc.org/publications

Video instructions for R package use available:

1. Installation (update) and Documentation (4 min)
2. Importing (reading) data (11 min)
3. Prepare data for analysis (12 min)
4. Calculating statistics (9 min)
5. Write/save results (12min)
6. Validate output file, entering meta data (13min)0

Questions?

gabriel.rojas@uibk.ac.at