

Guidelines for the documentation of data processing in the RECC model:

ODYM model data processes (ODP)

	a1	a2	a3
b1			
b2			
b3			

Stefan Pauliuk, RECC team, November 2023

ODYM model data processes (ODP), rationale and scope

ODYM – The Open Dynamic Material Systems Model is an open source framework for material systems modeling programmed in Python. The description of systems, processes, stocks, flows, and parameters is object-based, which facilitates the development of modular software and testing routines for individual model blocks (Pauliuk and Heeren 2020). ODYM MFA models can handle any depth of flow and stock specification: products, components, sub-components, materials, alloys, waste, and chemical elements can be traced simultaneously.

ODYM features a new data structure for material flow analysis, based on a newly developed data model for industrial ecology research (Pauliuk et al. 2019). All input and output data are stored in a standardized file format and can thus be exchanged across projects.

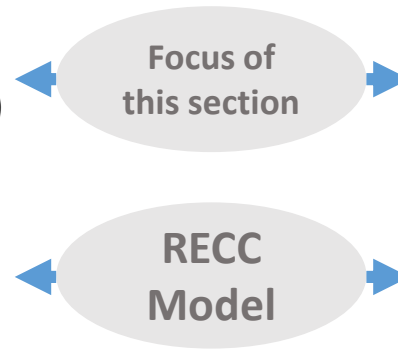
The databases and model calculations for ODYM-based projects can be quite complex, such as in the RECC (Resource Efficiency – Climate Change mitigation framework) model for circular economy scenarios (Pauliuk et al. 2021), which is why procedures for complete traceability of workflows and reproducibility of results were designed.

The figure on the next slide visualizes the traceability steps for the RECC model. While the workflow for data and result transparency goes from the original data sources to the final result documented in papers or reports, the tracing of results and data back to the original sources or model configuration goes the other way.

Bigger picture: Make each numerical result traceable and reproducible.

Documentation steps:

- Inventory of raw data files/sources
- Inventory of data parsing/processing scripts
- Inventory of data processes:
from raw data to model parameter (version)
- RECC database (version) plus config file
- RECC results (version) in separate folders
- Aggregated results: Report or paper supplement
- Figures or individual (representative or rounded)
numbers in paper or report



Tracing and reproducibility steps:

- Inventory of raw data files/sources
- Inventory of data parsing/processing scripts
- Inventory of data processes:
from raw data to model parameter (version)
- RECC database (version) plus config file
- RECC results (version) in separate folders
- Aggregated results: Report or paper supplement
- Figures or individual (representative or rounded)
numbers in paper or report

The data transparency workflow must be designed from the right side in the figure, by asking, at each research step: “Have I documented all the information to trace the results back to their original source and to reproduce my findings?” and documenting the information accordingly. Then, in the research practice, the workflow is the opposite, as shown on the left side of the figure.

Five steps for complete traceability of workflows & reproducibility of results in the RECC model

Step 1: Link data to their original sources when archiving them

Each model parameter file has a reference sheet, where the links to the DOI or other identifiers of the used data are listed and described (like 'Table 3 in DOI xxx' or 'Supplementary Figure X for article DOI xxx'). Where suitable, the original data files are archived in a separate raw data archive, which is part of the project's hosting institutions research data management, and the relative path of the archived raw data file is documented on the reference sheet.

All these sources need to be listed and intermediate steps (such as excel workbooks supplied by colleagues or difficult to access supplementary material) be archived. See the figure below for a screenshot of the reference sheet of a RECC model parameter file.

literature_id	literature_key, dataset or sciedc_dataset	iedc_dataset_ver	authors	title	year	journal_outlet	city	DOI (for paper) URL or link to document, no copyright	notes	archive location
1	Hotmaps building stock database		Hotmaps building st Simon Pezzutto, Stefano Z	HOTMAPS D2.3 WP2	2018				for EU28	archived under: \RECC_D
2	Hotmaps and JRC_Nemry data extraction	EU_NonResBuildin	Pauliuk						for EU28	archived under: \RECC_D
3			CBECs Survey	Stock of non-residentia	2016			https://www.eia.gov/consumption/com	for USA	archived under: \RECC_D
4	USA nrb stock data workbook		PB, LP	USA nrb stock data workbook					for USA	archived under: \RECC_D
5			Lixuan Hong, Nan Zhou, D	Modeling China's Buildi	2014	Lawrence Berkeley National Laboratory		https://www.aceee.org/file	©2014 ACEE for China	
6	Survey of Commercial and Institutional Energy Use (SCIEU) - Buildings 2014			Stock of non-residentia	2014			https://oee.nrcan.gc.ca/corporate/status/	for Canada	
7	Canada nrb stock data workbook, sheet SCIEU-2014		LP, SP	Canada nrb stock data 2021/22					for Canada	archived under: \RECC_D
8			Kumar et al.	Estimating India's comm	2019	Building Research & Inform		https://doi.org/10.1080/09613218.2018	for India	
9			SP	Sheet 'India' in this file	2022				for India	this file
10			GIZ		2019			https://www.giz.de/de/downloads/giz20	for Turkey/R5,2MNF_Other	
11			SP	Sheet 'Single_Number_					for Turkey/R5 this file	
12			Chatellier-Lorentzen		2020	Sustainable Cities and Society		https://doi.org/10.1016/j.scs.2020.1021	for Mexico/R5.2LAM_Other	
13			SP	Sheet 'Single_Number_					for Mexico/R5 this file	
14	RECC parameter 2_P_RECC_Population_SSP_32R, V2.2, ODP		UUID 76481f7c-db79-4d96-af8d-d48842f7c215						for R5.2REF_Other	
15			Okumura, K., Ikaga, T., Ke	Development of the daf	2012	AIJ J. Technol. Des. Vol. 1	DOI: 10.31	https://www.scopus.com/record/display	for Japan	
16			SP	Sheet 'Japan' in this file	2022				for Japan	this file

Figure: Screenshot of the ref sheet of a RECC parameter file (example).

Five steps for complete traceability of workflows & reproducibility of results in the RECC model

Step 2: Linking original data to processed data with the ODYM Data Processes (ODP)

The ODYM Data Processes (ODP) represents a specific implementation of linking each single data point to its original data sources and the documented data processing steps in between.

In the RECC model, the more than 100 individual parameter files consist of lists or tables with a multi-index to describe the meaning of the numbers.

The figure below, for example, shows the data table for a building stock parameter, where the building stock (in million m²) is specified by year and region and broken down into ago-cohorts and different building types. The data comes in table format but it has a hierarchical multi-index for both rows and columns. In a parameter that gives the stock for many different regions and types, a number of different data sources will be used to obtain the different numbers. The figure shows a (slightly exaggerated) situation where the individual numbers are marked with different colors, depending on what sources they come from.

For the different colors, not only the sources vary but also the way the data were processed. For example, it may be that one source has the data in the right resolution so they just need to be copied over. Another data source may only contain aggregate information so that proxy data have to be used in addition to arrive at the required level of resolution level. For still other figures, no matching data source may be present at all, so that proxy data have to be used or assumptions have to be made.

Five steps for complete traceability of workflows & reproducibility of results in the RECC model

Figure: Screenshot of the data sheet of a RECC parameter file, with the different colors indicating the different data sources and data processes they are linked to.

That means that the color code (or any other reference linked to each individual data point) cannot just simply point to a data source but points to a procedure, which in turn may involve a number of data source, ancillary calculation, and assumptions. Since the data format of the RECC model is that of the ODYM MFA software, the traceability documentation applies to all data in ODYM format, and we define the ODYM data process (ODP) as follows (see next slide).

Five steps for complete traceability of workflows & reproducibility of results in the RECC model

An ODYM data process (ODP) is a research procedure that links data from one or more data sources together with ancillary calculations, aggregation or disaggregation, and/or assumptions to output data that are contained and formatted in a parameter file in ODYM format.

The following requirements were central for designing the ODP and implementing it in the RECC project:

- Need for a traceable and distributed (across different team members) workflow for different raw data that flow into RECC parameters.
- The data documentation workflow needs to allow different parameter versions to be branched and later merged from certain departing file.
- Each single number must be traceable.
- Flexible procedure from large datasets (100000+ data points from a single ODP down to individual numbers)
- Model parameter files will be a patch of data from different data processes, plus modifications and corrections for individual numbers.
- The link between each single number and data source and the data processing must be documented.
- The whole documentation process should be lean and not create much overhead.
- The ODPs need a machine-readable documentation that is also easy to read and edit for humans.

Five steps for complete traceability of workflows & reproducibility of results in the RECC model

ODPs are documented in three parts with the following details:

- (1) The ref sheet of an ODYM parameter file (see Figure above for a screenshot) lists traceable links to all data sources (main and ancillary) as well as links to all additional calculation tools such as spreadsheets or scripts. These links can be GitHub links, paths to a local archive, or links to other sheets contain in the same .xlsx ODYM parameter file.
- (2) The log sheet of an ODYM parameter file (see the figure below for a screenshot) lists a definition and description of the different ODYM data process (ODP) used to compile the entire dataset (see details below). Each ODP is defined by a UUID and by a color.
- (3) The data sheet of an ODYM parameter file (see figure above) links the individual data points to the respective ODPs using machine-readable color codes.

With these three parts, the parameter data points are linked to the original sources (workflow direction, top of the figure on the next slide). In the retrieval direction (bottom of the figure on the next slide), the lookup direction is the exact opposite of the archiving direction.

ODYM model data processes (ODP): General scheme

From the workflow perspective:

Input, documented in *ref* sheet

Raw data files/sources *UUID*,
checksum
Processing script/workbook
archive:
Script name and version no.

ODP for parameter [name]: $UUID(\text{version } n) \rightarrow UUID(\text{version } n+1)$
Name of ODP: 'Update xy parameter', Different types of data
processes with specific documentation for each of them.

Output:

Parameter *data* sheet

	a1	a2	a3
b1			
b2			
b3			

Data process descriptions (ODP) archive, Documented on *log* sheet

ODP1: $UUID1 \rightarrow UUID 2$ Details...
ODP2: $UUID 2 \rightarrow UUID 3$ Details...
ODP3: $UUID 3 \rightarrow UUID 4$ Details...

From the data tracing perspective:

Starting point:

Parameter *data* sheet

	a1	a2	a3
b1			
b2			
b3			

Look up the relevant data processes for selected parameter
values.
Link to data sources, processing scripts etc. and description of
data processing

Data sources & scripts

Exact list of used raw data
files/sources and data
processing scripts

Five steps for complete traceability of workflows & reproducibility of results in the RECC model

The central documentation place is the log sheet of the ODYM parameter file (see figure below for a screenshot). Here, the different ODPs are defined (name, type, UUID, color code), linked to metadata (date stamp, parameter version, name of researcher), and a detailed description is provided together with a link to all references used (data sources, scripts, etc.) as listed on the ref sheet. All together, this information must be detailed enough and correct to allow the users of the data to link each data point to the correct ODP via the color code, the color code to the correct UUID and version number of the parameter, the ODP to all references (data, scripts, workbooks, ...) used, and provide enough verbal detail and ancillary calculation to make the data processing transparent.

Different types of ODPs were defined to make documentation easy and systematic:

To structure the documentation process that provides the verbal detail and ancillary calculations, different types of ODPs are used depending on various situations for the compilation of model input data from a wide variety of sources with different methods. These situations include data transfer via script, manual transfer of a few individual numbers, data transfer with ancillary calculations in spreadsheets, assumptions, and more. See the Table below for details of the specific documentation guidelines.

Five steps for complete traceability of workflows & reproducibility of results in the RECC model

ODP_Short_name	Date of ODP	Version number before ODP	Version number after ODP	old UUID	new UUID and reference color	Who	What - detailed description	ODP type	List items of ref sheet used
Create dataset and Transfer EU country data	21.11.2019		V1.0		a123e321-93de-453a-aa33-888f294df0a	sp	<p>dataset created and filed with EU data from intermediate file RECC_Dataset_NonResBuildingArea_EU_9R extracted with script</p> <p>EU in 9 regions: For Standard nonres. buildings: from Hotmaps project, ref. [1], converted and interpolated to all years by script ref. [2]</p>	0, 1	[1,2]
Transfer USA data	06.04.2022	V1.0	V1.0	a123e321-93de-453a-aa33-888f294df0a	eab5be57-1963-40ab-8bbb-6eaaad808a7	Peter Berill, LP, SP	<p>For the USA, detailed nonres building stock data are available for 2012 (10 age-cohort groups x 13 purposes, ref [3]) These data were aggregated to the 6 building groups used here and the age-cohort groups were split into the individual years. All data are multiplied by 0.093 to convert from sqft to m². We use the 2012 stock data for 2015 as proxy with same age structure but a slightly lower total stock (82000 instead of 87000 Msqft). To fill the 2013-2015 gap, we replicated age-cohort 2012 values for 2013-2015 to obtain a rough estimate. This process is documented in ref. [4].</p>	3	[3,4]
Estimate China data	06.04.2022	V1.0	V1.0	eab5be57-1963-40ab-8bbb-6eaaad808a7	34233199-85cf-4c15-8802-61a98b8cc68	sp	<p>Fig. 2 in ref. [5] reports ~14750 million m³ in 2015 non-res buildings, with the following breakdown: (units: million m³) (estimated from user)</p> <p>These data are allocated to standard efficiency buildings. Most of these buildings are assumed to be less than 30 years old. Hence, the totals are simply split evenly over the 1985-2015 age-cohorts.</p>	2	[5]
Transfer Canada data	06.04.2022	V1.0	V1.0	34233199-85cf-4c15-8802-61a98b8cc68	69dce5c5-dc42cf059f-ca40-444c-823e-9774ecb5223e	Peter Berill, LP, SP	<p>For Canada, detailed nonres building stock data are available for 2014 (8 age-cohort groups x 10 purposes, ref [6]).</p> <p>Assumptions for 2014 table</p> <ol style="list-style-type: none"> All years contribute equally in a time period Office buildings (non-medical) aggregated with Medical office building in "Offices" Commercial is comprised of food or beverage store and non-food retail store Education is comprised of elementary and secondary schools only as universities are included in other in the Canadian data Health is comprised of Assisted daily residential and hospital Hotels and restaurants is comprised of Hotels, motels or lodges as there is no data on restaurants in the Canadian data "Other" gathers Other from the Canadian data and Warehouse <p>To fill the 2015 gap, we replicated age-cohort 2014 values for 2015 to obtain a rough estimate. This process is documented in ref. [7].</p>	3	[6,7]
Estimate India data, use as proxy for R5.2ASIA	06.04.2022	V1.0	V1.0	69dce5c5-dc42cf059f-ca40-444c-823e-9774ecb5223e	SP		<p>Estimate data for India from 2017 data in journal paper, split into different age-cohorts, documented on sheet "India" in this file. Also, use the Scale to Other_AStA with population ratio 963/1305</p>	3	[8,9]
Estimate Turkey data and use as proxy for R5.1	10.04.2022	V1.0	V1.0	42cf059f-ca40-444c-823e-9774ecb5223e	99fc4efa-047e-428d-8847-03b29d04041f	SP	<p>"Non-residential, commercial and public buildings cover a total of 0.5 billion m2 area (in 2015)" From ref. [10]. Estimate data for Turkey and use as proxy for R5.2MNF_Other and R5.2OECD_Other from 2019 data in Giz report (ref. [10], split into diff</p>	2	[10,11]
Estimate Mexico data and use as proxy for R5.1	10.04.2022	V1.0	V1.0	99fc4efa-047e-428d-8847-03b29d04041f	SP		<p>"with an estimated floor space of 272 million m2, accounting only for restaurants, offices, shops, supermarkets, theatres and hospitals." Fro Estimate data for Mexico and use as proxy for R5.2LAM_Other from data in journal paper by Chatelier-Lorentzen (ref [12], split into different age-cohorts, documente</p>	2	[12,13]
Use R32EU12-M as proxy for R2.5REF_Other	10.03.2022	V1.0	V1.0	81dd5056-271556029-e8ce-4878-8a79-611789f11ce	SP		<p>Use R32EU12-M as proxy for R2.5REF_Other on a per Capita basis, use population from RECC parameter 2_P_RECC_Population_SSP, Data for R32EU12-M, SSP2, IASA GDP scenario, Million:</p>	2	[14]

Figure: Screenshot of the documentation of the ODP on the log sheet of a RECC parameter Excel template. The figure shows the documentation of the data processes, with UUID, type, color code, links to references, and documentation.

Five steps for complete traceability of workflows & reproducibility of results in the RECC model

Table: Overview of the different ODP types defined so far.

No.	ODP Type	Documentation details
0	Create Parameter	UUID and initial version number (1.0)
1	Data-script-Parameter (for 'large' data)	Script (list with path in ref sheet) parses raw data (list with DOI etc. in ref sheet), structures them (select, aggregate, ...) and saves resulting data to parameter file
1a	From RECC scenario target table	For the RECC model, a scenario target table with an interpolation script is used (Fishman et al. 2021). Document the version number of the interpolation script and of the scenario target table.
2	Small data	Few individual numbers are extracted from raw data (list with DOI etc. in ref sheet) and processed locally (documented directly in log sheet, with additional assumptions)
3	Medium data	Substantial reformatting of data is needed, documented on a separate workbook or worksheet (list with path in ref sheet).
4	Assumption	Mere assumption without additional references.
5	Correction	Correct dataset, e.g., calculations or labels.
6	Reformatting	Reformat dataset, typically with a spreadsheet calculator. E.g., add or remove new aspect, change aspect order, etc., without additional data or aggregation/disaggregation. Used also for manually reformatted large datasets (Excel...) imported from other models.
7	By definition	Given by definition, for example, a correspondence table indicating that the provinces of a country are part of the country.

ODYM model data processes (ODP), Details

Parameter value sheet

	a1	a2	a3
b1			
b2			
b3			

ODP List (log sheet)

None → UUID 1
UUID1 → UUID 2
UUID 2 → UUID 3
UUID 3 → UUID 4

ODP type // Status // Data sources

- (1) // Overwritten // [1], [2], [3]
- (2) // Active // [1], [2], [4]
- (3) // Active // [5]
- (4) // Active // [6], [7]

+ For each new parameter or update, define and document the corresponding data processes:

"Update xy parameter", "Define and populate ab parameter".

- + Data process: *Parameter Name, Version_old, UUID_old, Action (Name of data process), Version_new, UUID_new*
- + Specifiers: *UUID_new (is also the UUID of the ODP) / Parameter Name / Version_old / Version_new / Action_description*
- + User info: Who, when, what?
- + Raw Data items list: --> Link to master data inventory! (avoid duplications, but not crucial)
- + Code/software items list: --> Link to master data processing software inventory (RECC data git)
- + Description: Text (can be different fields), with separate link to data and software list items
- + Result: Define color: link between data and log sheet in Parameter file

→ With the introduction of the ODP, we make only minimal and necessary changes compared to existing routines!

Five steps for complete traceability of workflows & reproducibility of results in the RECC model

In step 3, the parameter list in the RECC model's config file is defined to select the right set of parameters and their exact versions. The config file lists the parameter file names so that there is a documented link between each model configuration and a specific model database.

In step 4, the model documents its own run, by creating a UUID and a separate result folder for each single model run, where the valid model config file is saved (as a copy), the used model code version is documented (via the Git commit ID), and the result summary is archived. A scenario definition file links all scenario runs that are defined and executed by the RECC model scripts with their corresponding result folders, so that a 1:1 match between each scenario definition (consisting of model version, database version, and model config settings) and each result folder is established.

In step 5, the result evaluation and documentation, model result figures and tables are created by scripts and scenario evaluation config files that are part of the RECC Git and Zenodo repositories. These RECC results evaluation scripts have built-in features that document (in RECC results evaluation log files) exactly which model results were analysed by which scripts and what the results (figures and tables) are. Each report or paper has a log table (part of the .xlsx workbook with all the plot and table data) that exactly documents/traces how the figures and tables were created (from the RECC results evaluation log) and how each single quantitative statement in the paper was derived from the detailed quantitative results. As the RECC model is being developed, more and more steps are automated to reduce manual and error-prone work.

Appendix: Specifications for the different ODYM data process (ODP) types

ODYM model data processes (ODP), Details for type 0 (creation)

Parameter value sheet

	a1	a2	a3
b1			
b2			
b3			

ODP List (log sheet)

None → UUID 1
UUID1 → UUID 2
UUID 2 → UUID 3
UUID 3 → UUID 4

ODP type // Status // Data sources

(1) // Overwritten // [1], [2], [3]
(2) // Active // [1], [2], [4]
(3) // Active // [5]
(4) // Active // [6], [7]

ODP type (0): Create Parameter. UUID and initial version number (1.0)

Corresponding data processes name: ,Define parameter.'

- + Document the ODP directly in the log sheet of the parameter file:
 Make a note on the *log sheet* of the parameter Excel file, see existing parameters as example.
- + Can be omitted to keep documentation simple, and the first data-inserting ODP can be used instead.

Example: 3_SHA_NonResBuildingRenovationScaleUp_V1.0

ODP_Short_name	Date of ODP	Version number before ODP	Version number after ODP	old UUID	new UUID and reference color	Who	What - detailed description	ODP type	List items of ref sheet used
Create dataset and populate dataset from scenario target table	29.03.2022	none	V1.0	none	849102e7-a61f-4c24-85ec-e1f09e86ccf7	sp	created parameter file, added log and ref sheets and completed the latter.	0	
							Use Data from 3_SHA_BuildingRenovationScaleUp_V1.0 as proxy for first m Interpolated from 3_SHA_BuildingRenovationScaleUp in target table V2.5 [1] and interpolation script V2.5 [2].	1a	[1,2]

ODP short name: *Create dataset*

ODP type: *0*

First version number: *V1.0*

Description: *created parameter file, added log and ref sheets and completed the latter.*

Can be combined with data process description for compiling an initial dataset

ODYM model data processes (ODP), Details for type 1 (script)

Parameter value sheet

	a1	a2	a3
b1			
b2			
b3			

ODP List (log sheet)

None → UUID 1
UUID1 → UUID 2
UUID 2 → UUID 3
UUID 3 → UUID 4

ODP type // Status // Data sources

- (1) // Overwritten // [1], [2], [3]
- (2) // Active // [1], [2], [4]
- (3) // Active // [5]
- (4) // Active // [6], [7]

ODP type (1): Data-script-Parameter.

Corresponding data processes name: ,Populate parameter [name] from data source [name] and script‘.

+ Document the ODP directly in the *log sheet* of the parameter file:

See provided template as example.

Required info:

Specifiers: UUID_new (is also the UUID or the ODP)/Par_Name_affected/Version_old/Version_new/Action_description

User info: Who, when?

ID/Link of script that was used to parse raw data (IDs!)

Description (single line/short remark is sufficient)

Result: Define color/link to log sheet in Parameter file

Example: 3_IO_NonResBuildings_UsePhase_V1.1

ODP_Short_name	Date of ODP sign-off before ODP	Version number after ODP	Version number before ODP	old UUID	new UUID and reference color	Who	What - detailed description	ODP type	List items of ref sheet used
Create dataset	23.11.2019	none	V1.0	none	849102e7-a61f-4c24-85ec-e1f09e86ccf7	sp	Datafile created	0	
Transfer EU country data	23.11.2019	V1.0	V1.0	849102e7-a69533adef-f54a-4fa1-580a-bb2685d061a6		sp	Datafile filled with EU data from intermediate file RECC_Dataset_NonResBuildingEnergyUseTypeSplit_EU_9R extracted with script.	1	[1,2]
							EU in 9 regions: For Standard nonres. buildings: from Hotmaps project, ref. [1], converted and interpolated to all years by script ref. [2]		

ODP short name:

Transfer ...

ODP type:

1

Description:

describe data process and mention, with their IDs in square brackets, which data sources from the ref sheet were used, e.g., [1]

Data sources:

data and script IDs (from ref sheet) listed in square brackets (for machine-parsing and to distinguish them from ODP type in the column to the left)

ODYM model data processes (ODP), Details for type 1a (from scenario target table)

Parameter value sheet

	a1	a2	a3
b1			
b2			
b3			

ODP List (log sheet)

None → UUID 1
UUID1 → UUID 2
UUID 2 → UUID 3
UUID 3 → UUID 4

ODP type // Status // Data sources

- (1) // Overwritten // [1], [2], [3]
- (2) // Active // [1], [2], [4]
- (3) // Active // [5]
- (4) // Active // [6], [7]

ODP type (1a): Scenario target table → interpolation script → Parameter.

Corresponding data processes name: ,Populate parameter [name] from scenario target table sheet [name+version] and interpolation script [version]‘.

+ Make a note on the *log sheet* of the parameter Excel file

See provided template as example.

Required info:

Specifiers: UUID_new (is also the UUID or the ODP)/Par_Name_affected/Version_old/Version_new/Action_description

User info: Who, when?

ID/Link of scenario target table interpolation script that was used to parse raw data (IDs!)

Description (single line/short remark is sufficient)

Result: Define color/link to log sheet in Parameter file

Example: 3_SHA_NonResBuildingRenovationScaleUp_V1.0

ODP_Short_name	Date of ODP	Version number before ODP	Version number after ODP	old UUID	new UUID and reference color	Who	What - detailed description	ODP type	List items of ref sheet used
Create dataset and populate dataset from scenario target table	29.03.2022	none	V1.0	none	849102e7-a61f-4c24-85ec-e1f09e86ccf7	sp	created parameter file, added log and ref sheets and completed the latter.	0	
							Use Data from 3_SHA_BuildingRenovationScaleUp_V1.0 as proxy for first m Interpolated from 3_SHA_BuildingRenovationScaleUp in target table V2.5 [1] and interpolation script V2.5 [2].	1a	[1,2]

ODP short name:

Populate dataset from scenario target table

ODP type:

1a

Description:

explain which version of scenario target table and interpolation script were used, link to the source numbers on the ref sheet, e.g., [1]

Data sources:

version of scenario target table and interpolation script (from ref sheet) listed in square brackets (for machine-parsing and to distinguish them from ODP type in the column to the left)

Dataset	literature_id	literature_key	iedc_datase	iedc_datase	authors	title	year	journal_outlet	city	DOI (for paper)	URL or link to copyright	archive locat	notes
3_SHA_NonResBuildingRenovatio	1	scenario target table output from the sheet, "3_SHA_BuildingRenovationScaleUp".								RECC scenario target table version 2.5			
3_SHA_NonResBuildingRenovatio	2	scenario target table interpolation script, version 2.5								RECC scenario target table interpolation script, version 2.5			

ODYM model data processes (ODP), Details for type 2 (small data)

Parameter value sheet

	a1	a2	a3
b1			
b2			
b3			

ODP List (log sheet)

None → UUID 1
UUID1 → UUID 2
UUID 2 → UUID 3
UUID 3 → UUID 4

ODP type // Status // Data sources

- (1) // Overwritten // [1], [2], [3]
- (2) // Active // [1], [2], [4]
- (3) // Active // [5]
- (4) // Active // [6], [7]

ODP type (2): Small data. Few individual numbers are extracted from raw data and processed locally.
Corresponding data processes name: 'Add data from small data source'.

+ Make a note on the *log sheet* of the parameter Excel file
 See provided template as example.

Required info:

- Specifiers: UUID_new (is also the UUID or the ODP)/Par_Name_affected/Version_old/Version_new/Action_description
- User info: Who, when?
- List of raw data sources (IDs!) and display of data used
- Description: Very detailed: Exactly how are the raw data processed, which additional assumptions are made, and exactly which result are generated?
- Result: Define color/link to log sheet in Parameter file

ODYM model data processes (ODP), Details for type 3 (medium data)

Parameter value sheet

	a1	a2	a3
b1			
b2			
b3			

ODP List (log sheet)

None → UUID 1
UUID1 → UUID 2
UUID 2 → UUID 3
UUID 3 → UUID 4

ODP type // Status // Data sources

(1) // Overwritten // [1], [2], [3]
(2) // Active // [1], [2], [4]
(3) // Active // [5]
(4) // Active // [6], [7]

ODP type (3): Medium data. Substantial reformatting of data, needs separate workbook or worksheet for proper documentation.
Corresponding data processes name: ‘Extract, process, and add data from medium-size data source’.

+ Use the *Small AND medium versions* for documenting the ODP: Make a note on the *log sheet* of the parameter Excel file
 See provided template as example.

Required info for the small documentation:

Specifiers: UUID_new (is also the UUID or the ODP)/Par_Name_affected/Version_old/Version_new/Action_description

User info: Who, when?

Result: Define color/link to log sheet in Parameter file

Link to medium version of documentation: Link to separate worksheet or workbook!

Required info for the medium documentation:

List of raw data sources (IDs!) and display of data used

Description: Very detailed: Exactly how are the raw data processed, which additional assumptions are made, and exactly which result are generated?

From *log* sheet of parameter file:

ODP type 3 with link to separate workbook (ref. [7]):

06.04.2022	V1.0	V1.0	34233199-8569dce5c5-dde1-4308-a660-99955b75d67	Peter Berill, LP, SP	For Canada, detailed nonres building stock data are available for 2014 (8 age-cohort groups x 10 purposes, ref [6]).	3 [6,7]
					Assumptions for 2014 table	
					1. All years contribute equally in a time period	
					2. Office buildings (non-medical) aggregated with Medical office building in "Offices"	
					3. Commercial is comprised of food or beverage store and non-food retail store	
					4. Education is comprised of elementary and secondary schools only as universities are included in oher in the Canadian data	
					5. Health is comprised of Assisted daily/residential and hospital	
					6. Hotels and restaurants is comprised of Hotels, motels or lodges as there is no data on restaurants in the Canadian data	
					7. "Other" gathers Other from the Canadian data and Warehouse	
					To fill the 2015 gap, we replicated age-cohort 2014 values for 2015 to obtain a rough estimate.	
					This process is documented in ref. [7].	

From *ref* sheet of parameter file: Link to separate workbook (ref. [7]) with details and storage location.

Datasources used in this workbook are also listed here ([6]):

Dataset	literature_id	literature_key, dataset or sc	iedc_dataset	iedc_dataset_vers	authors	title	year	journal_outlet	city	DOI (for pap	URL or link to document, n	copyright	archive location
2_S_RECC_FinalProducts_2015_nonresbuildings	1	Hotmaps building stock database	Hotmaps building st	Simon Pezzutto, Stefano Z	HOTMAPS D2.3 WP2	2018							archived under: \RECC_Database\Raw_Dr
2_S_RECC_FinalProducts_2015_nonresbuildings	2	Hotmaps and JRC_Nemry data extraction	EU_NonResBuildin	Pauliuk									archived under: \RECC_Database\Raw_Dr
2_S_RECC_FinalProducts_2015_nonresbuildings	3			CBECS Survey	Stock of non-residentia	2016				https://www.eia.gov/consumption/comr		archived under: \RECC_Database\Raw_Dr	
2_S_RECC_FinalProducts_2015_nonresbuildings	4	USA nrb stock data workbook		PB, LP	USA nrb stock data workbook								archived under: \RECC_Database\Raw_Dr
2_S_RECC_FinalProducts_2015_nonresbuildings	5			Lixuan Hong, Nan Zhou, D	Modeling China's Buildi	2014	Lawrence Berkeley National Laboratory			https://www.aceee.org/file	©2014 ACEEE Summer Study on Energy Efficiency in Bt		
2_S_RECC_FinalProducts_2015_nonresbuildings	6	Survey of Commercial and Institutional Energy Use (SCIEU) - Buildings 2014		LP, SP	Canada nrb stock data 2021/22	2014				https://oe.nrcan.gc.ca/corporate/statist			
2_S_RECC_FinalProducts_2015_nonresbuildings	7	Canada nrb stock data workbook, sheet SCIEU-2014		LP, SP	Canada nrb stock data 2021/22	2019	Building Research & Inform			https://doi.org/10.1080/09613218.2018			archived under: \RECC_Database\Raw_Dr
2_S_RECC_FinalProducts_2015_nonresbuildings	8			Kumar et al.	Estimating India's comr	2019				https://doi.org/10.1080/09613218.2018			
2_S_RECC_FinalProducts_2015_nonresbuildings	9			SP	Sheet 'India' in this file	2022							this file
2_S_RECC_FinalProducts_2015_nonresbuildings	10			GIZ		2019				https://www.giz.de/de/downloads/giz2019-en-turkish-building-sector.pdf			
2_S_RECC_FinalProducts_2015_nonresbuildings	11			SP	Sheet 'Single_Number_								this file
2_S_RECC_FinalProducts_2015_nonresbuildings	12			Chatellier-Lorentzen		2020	Sustainable Cities and Society			https://doi.org/10.1016/j.scs.2020.102165			
2_S_RECC_FinalProducts_2015_nonresbuildings	13			SP	Sheet 'Single_Number_								this file
2_S_RECC_FinalProducts_2015_nonresbuildings	14	RECC parameter 2_P_RECC_Population_SSP_32R, V2.2, ODP UUID 7648117c-db79-4d96-af8d-d48842f7c215											

Detailed calculations in separate workbook (sample screenshot):

5. Health is comprised of Assisted daily/residential and hospital														
6. Hotels and restaurants is comprised of Hotels, motels or lodges as there is no data on restaurants in the Canadian data														
7. "Other" gathers Other from the Canadian data and Warehouse														
Year	Age-cohort	Offices (non medical)	Offices (medical)	Offices	Food or beverage store	Non-food retail store	Commercial	Education	Assisted daily/residential care facilities	Hospital	Health			
2014	1900	0,955	0,005	0,96	0,02	0,05	0,07	0,115	0,06	0,055	0,12			
2014	1901	0,955	0,005	0,96	0,02	0,05	0,07	0,115	0,06	0,055	0,12			
2014	1902	0,955	0,005	0,96	0,02	0,05	0,07	0,115	0,06	0,055	0,12			
2014	1903	0,955	0,005	0,96	0,02	0,05	0,07	0,115	0,06	0,055	0,12			

ODYM model data processes (ODP), Details for type 4 (assumption)

Parameter value sheet

	a1	a2	a3
b1			
b2			
b3			

ODP List (log sheet)

None → UUID 1
UUID1 → UUID 2
UUID 2 → UUID 3
UUID 3 → UUID 4

ODP type // Status // Data sources

(1) // Overwritten // [1], [2], [3]
(2) // Active // [1], [2], [4]
(3) // Active // [5]
(4) // Active // [6], [7]

ODP type (4): Assumption. Set data in dataset based on assumption.

E.g., xyz values of parameter abc are set to 0 (zero).

Or, xyz values of parameter abc for region ijk are used as proxy for region mno.

Corresponding data processes name: ‚Assumption‘

+ Document the ODP directly in the log sheet of the parameter file:

Make a note on the *log sheet* of the parameter Excel file, see existing parameters as example.

+ No link to references needed, assumption is documented directly on *log sheet*.