

Work Package 5 EU-CEG data and
enhanced laboratory capacity for
regulatory purposes

EU-CEG Public Data Sharing

Author: ANSES
September 2024

Doc. Ref. N°: D5.1
Type: Document (R)
Dissemination: Public (P)



Co-funded by the European Union's Health
Programme under Grant Agreement n°: 101035968
- JA-01-2020 - HP-JA-2020 / HP-JA-2020-2

The content of this publication represents the views of the author only and is his/
her sole responsibility; it cannot be considered to reflect the views of the European
Commission and/or the Consumers, Health, Agriculture and Food Executive Agency
or any other body of the European Union. The European Commission and the Agency
do not accept any responsibility for use that may be made of the information it
contains.

Version	Date	Authors	Comments
1	27 Sept. 2024	ANSES	Submitted version



Table of contents

1	First considerations about data sharing	4
2	The need for ready-to-publish EU-CEG data	4
3	Technical specifications of the EU-CEG public data files	5
4	Structure of {MS}-EC-Public-{yyyy-mm-dd}-List.CSV	6
5	Structure of {MS}-EC-Public-{yyyy-mm-dd}-Comp.CSV	7
6	Structure of {MS}-TP-Public-{yyyy-mm-dd}-List.CSV	8
7	Structure of {MS}-TP-Public-{yyyy-mm-dd}-Comp.CSV	9
8	Technical implementation and sharing of EU-CEG public data	10
9	Conclusion	11
10	ANNEX: proposal for publication thresholds of ingredients other than tobacco, according to Ingredient Function and Product Type (based on provisions laid on art.6(2) of Decision of execution (EC) 2015/2186).	12

1 First considerations about data sharing

At the outset of the project, establishing a data sharing agreement between the Member State Competent Authority (MSCA) and third-party research organizations (eg JATC-2 partners) was envisioned as a means to facilitate scientific research and promote transparency within the consortium.

However, after a thorough review of the applicable legal frameworks, it became clear that sharing trade secrets without meeting strict criteria would not be legally permissible. Directive (EU) 2016/943, which governs the protection of undisclosed know-how and business information, and the Tobacco Products Directive (EU) 2014/40 (TPD) which addresses the confidentiality of some submitted data (such as full recipes or sales volumes data), impose stringent conditions on the acquisition, use, and disclosure of sensitive information.

In particular, such information may only be shared when there is a legitimate legal basis, explicit consent from the trade secret holder, or a clear public interest justification. Moreover, any disclosure must be limited to the specific purpose for which it was authorized, and appropriate safeguards, such as Non-Disclosure Agreements (NDAs), must be in place to ensure confidentiality.

Given these restrictions, it was concluded that the proposed data sharing agreement could not be implemented as initially intended.

Nevertheless, given that MSCA have a legal obligation to disclose certain information to the public under the Tobacco Products Directive (EU) 2014/40, and considering that the criteria for determining which data are non-confidential are clearly outlined in the implementing decisions (EU) 2015/2183 for electronic cigarettes and (EU) 2015/2186 for tobacco products, there is a legal basis to share specific data elements. In particular, ingredients present above specified concentration thresholds are classified as non-confidential under these decisions.

Consequently, the data in question are considered public in nature, allowing for their communication to third parties without breaching confidentiality rules. This ensures compliance with transparency requirements while upholding the Directive's intention to provide consumers and stakeholders with access to relevant product information.

2 The need for ready-to-publish EU-CEG data

Pursuant to TPD¹ Art 5(4), Art 20(8) and Art 22(2), Member States shall ensure that the information submitted on tobacco, e-cigarettes & refill containers and herbal smoking products, is made publicly available on a website. The Member States shall take the need to protect trade secrets duly into account when making that information publicly available.

Moreover, Art. 6(2) of Commission Implementing Decision (EU) 2015/2186 (tobacco products)² and Art. 6(2) of Commission Implementing Decision (EU) 2015/2183 (e-cigarettes and refill containers)³ provide some concentration thresholds above which the information on ingredients is not considered as a trade secret.

During the first joint action (hereafter 'JATC1'), the issue of determining precisely what data is public and non-confidential in EU-CEG was addressed through the deliverable D5.1⁴. In this report, any field of the data dictionary is assessed either as being public, confidential or conditionally public (provided an additional criterion). This report has been presented in 2020 during a meeting of the Expert group on tobacco policy.

Although there has been neither common nor formal agreement regarding the whole criteria set out

1 Tobacco products Directive (2014/40/EU): https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AJOL_2014_127_R_0001

2 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:JOL_2015_312_R_0003

3 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015D2183>

4 <https://jaotc.eu/wp-content/uploads/2021/02/WP5-D5.1-Report-on-what-data-is-public-and-non-confidential-in-EU-CEG.pdf>

in this report, the need to implement it has become more and more urgent from the Member States.

On its side, the DG SANTE is still missing a more precise information about how to implement it under MS-Rep for the benefit of the Member States.

The purpose of this document is not to discuss once again the criteria developed in JATC1-D5.1 but more to give a short and straightforward guidance for the DG SANTE IT Services (SANTE EUCEG Team) to implement this provision into ready-to-publish⁵ files available from MS-Rep.

3 Technical specifications of the EU-CEG public data files

Each month and for each Member State, 2 CSV files (field names in the first row, UTF-8-BOM encoding, semicolon/double-quotes separator, dot for decimal separator and ISO 8601 date-time format) are generated for both TP (tobacco and herbal) and EC products.

Each set of 2 files is compressed into a distinct ZIP archive which is made available under the 'JATC reports for sharing' section into the 'Report repository' for both TP and EC. Each MS can share its national public files to any other MS through the sharing feature of MS-Rep.

The choice of ZIP compressed CSV is considered as the best fit between the ease of reuse (CSV opens into Excel, open data format) and the reduced file size for hosting into MS-Rep and downloading from it.

Product Category	ZIP File Location	ZIP file name	CSV file name	CSV file description
ECigarette products (EC)	Report repository for EC (JATC reports for sharing)	{MS}-EC-Public-{yyyy-mm-dd}.ZIP	{MS}-EC-Public-{yyyy-mm-dd}-List.CSV	List of EC products/presentations with additional characteristics (one row per ECID/presentation)
			{MS}-EC-Public-{yyyy-mm-dd} Comp.CSV	Ingredients with quantities (multiple rows per ECID)
Tobacco products (TP) (including Herbals)	Report repository for TP (JATC reports for sharing)	{MS}-TP-Public-{yyyy-mm-dd}.ZIP	{MS}-TP-Public-{yyyy-mm-dd}-List.CSV	List of TP products/presentations with additional characteristics (one row per TPID/presentation)
			{MS}-TP-Public-{yyyy-mm-dd} Comp.CSV	Tobacco and other ingredients with quantities (multiple rows per TPID)

{MS} refers to the ISO 3166 Alpha-2 code of the National_Market.
 {yyyy-mm-dd} is the ISO 8601 Date cut-off for data extraction.

⁵ Alternatively, at least, publishable after removal of some fields and/or rows by each Member State, in case of partial endorsement of the JATC1-D5.1 report.

4 Structure of {MS}-EC-Public-{yyyy-mm-dd}-List.CSV

This file contains the list of all active presentations of EC products on a national market, as of #YYYYMMDD# (date cut-off for data extraction).

Products and presentations are sorted by descending order of LastUpdate then LaunchDate.

Pseudo-SQL Statements

SELECT

Col#	Table	Field	Comment [output data type]
1	PRESENTATION	NationalMarket	'FR', 'BE', 'DK'... [TEXT]
2	SUBMISSION	LastUpdate	This the SubmissionDate of the last known submission for the National Market. Data are extracted from this submission. [DATETIME yyyy-mm-dd hh:mm:ss]
3	PRODUCT	ProductTypeEN	English description for ProductType [TEXT]
4	PRODUCT	ProductID	EC-ID: 'NNNNN-NN-NNNNN' [TEXT]
5	SUBMITTER	SubmitterName	[TEXT]
6	SUBMITTER	SubmitterCountry	[TEXT]
7	PRESENTATION	BrandName	[TEXT]
8	PRESENTATION	BrandSubtypeName	[TEXT]
9	DESIGN	Description	[TEXT]
10	DESIGN	LiquidVolumeCapacity	[NUMERIC]
11	PRODUCT	Volume	[NUMERIC]
12	DESIGN	NicotineConcentration	[NUMERIC]
13	PRESENTATION	ProductNumberType	'EAN', 'GTIN', 'SKU', 'UPC', 'SUBMITTER'. [TEXT]
14	PRESENTATION	ProductNumber	[TEXT]
15	PRESENTATION	LaunchDate	[DATE yyyy-mm-dd]
16	DESIGN	BatteryType	[TEXT]
17	DESIGN	BatteryCapacity	[NUMERIC]
18	DESIGN	EcigAdjustableEN	English description [TEXT]
19	DESIGN	Voltage	[NUMERIC]
20	DESIGN	Wattage	[NUMERIC]
21	DESIGN	AirflowAdjustable	[BOOLEAN TEXT]
22	DESIGN	WickChangeable	[BOOLEAN TEXT]
23	DESIGN	Microprocessor	[BOOLEAN TEXT]
24	DESIGN	CoilComposition	[TEXT]
25	DESIGN	CoilResistance	[NUMERIC]

WHERE LaunchDate < #YYYYMMDD# **AND** List = 'Active' **AND** (WithdrawalDate **IS NULL OR** WithdrawalDate >= #YYYYMMDD#)

ORDER BY LastUpdate **DESC**, LaunchDate **DESC**

5 Structure of {MS}-EC-Public-{yyyy-mm-dd}-Comp.CSV

This file contains the list of ingredients with reported quantities and computed concentrations (multiple rows per EC-ID) of all EC-ID selected for publication (<EC-ID selection>).

Only ingredients with concentration above 0.1% by weight are extracted. If the total weight resulting from the sum of ingredient quantities for each (ECID, IdentificationRefillContainerCartridge) [so-called RecipeWeight] is negative or is equal to zero, concentrations cannot be computed and ingredients are not extracted.

Products are sorted by ascending order of EC-ID and IdentificationRefillContainerCartridge then ingredients are sorted by descending order of ConcentrationPPM.

Pseudo-SQL Statements

1/ In this 1st step, recipe quantities of ingredients are summed up to afford total recipe weights then only strictly positive values are kept. The intermediate RECIPE_WEIGHT table is thus created.

```
SELECT ProductID, SUM(RecipeQuantity) AS RecipeWeight INTO RECIPE_WEIGHT
FROM INGREDIENT
GROUP BY ProductID , IdentificationRefillContainerCartridge
HAVING SUM(RecipeQuantity) > 0
```

2/ In this 2nd step, the ingredients list is generated, concentrations in parts per million (ppm, 0.1% w/w = 1 000 ppm) are computed and ingredients with concentration above 1 000 ppm are filtered.

SELECT

Col#	Table	Field	Comment
1	<EC-ID selection>	ProductID	[TEXT]
2	INGREDIENT	IdentificationRefillContainerCartridge	[TEXT]
3	INGREDIENT	FunctionEN	In case of multiple functions, they are aggregated (comma separator); English description, with replacement by Other value if any [TEXT]
4	INGREDIENT	CasNumber	[TEXT]
5	INGREDIENT	IngredientName	[TEXT]
6	<computed field>	ConcentrationPPM	= ROUND(1 000 000 * RecipeQuantity / RecipeWeight , 0) [NUMERIC]
7	INGREDIENT	RecipeQuantity	[NUMERIC]
8	<RECIPE_WEIGHT>	RecipeWeight	[NUMERIC]
9	PRODUCT	Weight	[NUMERIC]
10	PRODUCT	Volume	[NUMERIC]

```
WHERE ConcentrationPPM >= 1 000
```

```
ORDER BY ProductID, IdentificationRefillContainerCartridge, ConcentrationPPM DESC
```

6 Structure of {MS}-TP-Public-{yyyy-mm-dd}-List.CSV

This file contains the list of all active presentations of TP products on a national market, as of #YYYYMMDD# (date cut-off for data extraction).

Products and presentations are sorted by descending order of LastUpdate then LaunchDate.

Pseudo-SQL Statements

SELECT

Col#	Table	Field	Comment
1	PRESENTATION	NationalMarket	'FR', 'BE', 'DK'... [TEXT]
2	SUBMISSION	LastUpdate	This the SubmissionDate of the last known submission for the National Market. Data are extracted from this submission. [DATETIME yyyy-mm-dd hh:mm:ss]
3	PRODUCT	ProductTypeEN	English description for ProductType [TEXT]
4	PRODUCT	ProductID	TP-ID: 'NNNNN-NN-NNNNN' [TEXT]
5	SUBMITTER	SubmitterName	[TEXT]
6	SUBMITTER	SubmitterCountry	[TEXT]
7	PRESENTATION	BrandName	[TEXT]
8	PRESENTATION	BrandSubtypeName	[TEXT]
9	PRODUCT	Filter	[BOOLEAN TEXT]
10	PRESENTATION	PackageUnits	[NUMERIC]
11	PRESENTATION	ProductPackageEN	English description [TEXT]
12	PRESENTATION	PackageNetWeight	[TEXT]
13	PRESENTATION	ProductNumberType	'EAN', 'GTIN', 'SKU', 'UPC', 'SUBMITTER' [TEXT]
14	PRESENTATION	ProductNumber	[TEXT]
15	PRESENTATION	LaunchDate	[DATE yyyy-mm-dd]
16	PRODUCT	Length	[NUMERIC]
17	PRODUCT	Diameter	[NUMERIC]
18	PRODUCT	Weight	[NUMERIC]
19	PRODUCT	TobaccoWeight	[NUMERIC]
20	PRODUCT	FilterLength	[NUMERIC]
21	CIGARETTE_SPECIFIC	CigaretteFilterVentilation	[NUMERIC]
22	CIGARETTE_SPECIFIC	CigaretteFilterDropPressureClosed	[NUMERIC]
23	CIGARETTE_SPECIFIC	CigaretteFilterDropPressureOpen	[NUMERIC]
24	TNCO_EMISSION	Tar	[NUMERIC]
25	TNCO_EMISSION	Nicotine	[NUMERIC]
26	TNCO_EMISSION	Co	[NUMERIC]
27	RYO_PIPE_SPECIFIC	RyoPipeTotalNicotineContent	[NUMERIC]
28	RYO_PIPE_SPECIFIC	RyoPipeUnionisedNicotineContent	[NUMERIC]
29	SMOKELESS_SPECIFIC	SmokelessTotalNicotineContent	[NUMERIC]
30	SMOKELESS_SPECIFIC	SmokelessUnionisedNicotineContent	[NUMERIC]
31	SMOKELESS_SPECIFIC	SmokelessPh	[NUMERIC]
32	SMOKELESS_SPECIFIC	SmokelessTotalMoisture	[NUMERIC]

WHERE LaunchDate < #YYYYMMDD# **AND** List = 'Active' **AND** (WithdrawalDate **IS NULL OR** WithdrawalDate >= #YYYYMMDD#)

ORDER BY LastUpdate **DESC**, LaunchDate **DESC**

7 Structure of {MS}-TP-Public-{yyyy-mm-dd}-Comp.CSV

This file contains the list of ingredients with reported quantities and computed concentrations (multiple rows per TP-ID) of all TP-ID selected for publication (<TP-ID selection>).

Only ingredients with concentration above thresholds⁶ by weight are extracted. In case of multiple functions provided for an ingredient, the lowest threshold is kept.

If the product weight is not strictly positive, concentrations cannot be computed and ingredients are not extracted.

Two queries are performed on both ingredients and tobacco ingredients then merged into a unique view.

Products are sorted by ascending order of TP-ID then ingredients are sorted by descending order of ConcentrationPPM.

Pseudo-SQL Statements

1/

SELECT

Col#	Table	Field	Comment
1	<TP-ID selection>	ProductID	[TEXT]
2	INGREDIENT	FunctionEN	In case of multiple functions, they are aggregated (comma separator); English description, with replacement by Other value if any [TEXT]
3	INGREDIENT	CategoryEN	If Category is 15, description 'Other' is replaced by CategoryOther [TEXT]
4	INGREDIENT	CasNumber	[TEXT]
5	INGREDIENT	IngredientName	[TEXT]
6	<computed field>	ConcentrationPPM	= ROUND(1 000 000 * RecipeQuantity / Weight , 0) [NUMERIC]
7	INGREDIENT	RecipeQuantity	[NUMERIC]
8	PRODUCT	Weight	[NUMERIC]
9	PRODUCT	TobaccoWeight	[NUMERIC]

WHERE (ConcentrationPPM >= 1 000 **AND** FirstFunction **IN** (12,22) **AND** ProductType **IN** (1,4))

OR (ConcentrationPPM >= 5 000 **AND** FirstFunction **IN** (12,22) **AND** ProductType **NOT IN** (1,4))

OR (ConcentrationPPM >= 5 000 **AND** FirstFunction **IN** (2,4,8,9,10,11,17,19,23,26))

OR (ConcentrationPPM > 0 **AND** FirstFunction **NOT IN** (12,22,2,4,8,9,10,11,17,19,23,26))

⁶ Thresholds differ according to product type and ingredient function. See Erreur ! Source du renvoi introuvable. & Annex of JATC1-D5.1 report (<https://jaotc.eu/wp-content/uploads/2021/02/WP5-D5.1-Report-on-what-data-is-public-and-non-confidential-in-EU-CEG.pdf>)

2/

SELECT

Col#	Table	Field	Comment
1	<TP-ID selection>	ProductID	[TEXT]
2	<computed field>	FunctionEN	= ' ' [TEXT]
3	<computed field>	CategoryEN	= ' ' [TEXT]
4	<computed field>	CasNumber	= ' ' [TEXT]
5	TOB_INGREDIENT	IngredientName	Concatenation of LeafType PartType CureMethod with replacement by Other value if any [TEXT]
6	<computed field>	ConcentrationPPM	= ROUND(1 000 000 * Quantity / Weight , 0) [NUMERIC]
7	TOB_INGREDIENT	Quantity	[NUMERIC]
8	PRODUCT	Weight	[NUMERIC]
9	PRODUCT	TobaccoWeight	[NUMERIC]

WHERE ConcentrationPPM >= 5 000

3/

SELECT 1/

UNION

2/

ORDER BY ProductID, ConcentrationPPM **DESC**

8 Technical implementation and sharing of EU-CEG public data

These technical specifications were taken up by the DG SANTE EU-CEG Team to produce the requested public data extractions. Several cycles of testing, corrections, and revisions took place between 2022 and 2023. Finally, the definitive solution was deployed in MS-Rep during Spring 2023.

Now, the MSCA receive a monthly extraction of EU-CEG data that is both publishable and shareable among themselves. Indeed, following the procedure developed during JATC-1, the publication files are placed in an area of the MS-Rep repository that allows sharing between MSCA.

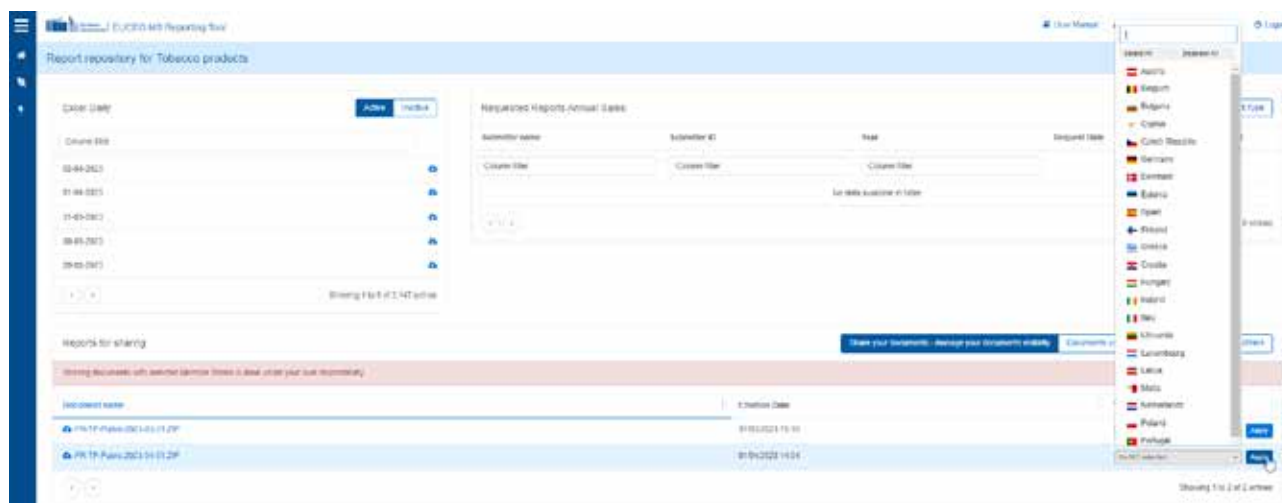


Figure. Screenshot of the MS-Rep French repository with the data sharing feature.

The authorized user can select one or more Member States with whom they wish to share their public data. This data will then appear in the MS-Rep interface of the selected Member States.

This functionality has been used in the project to share public datasets for analysis purposes.

9 Conclusion

Although it was not possible to establish a sufficiently robust legal framework for sharing complete EU-CEG data, due to significant constraints related to trade secrets concerning certain information, we were able to build on the work carried out during JATC-1 to convert the functional specifications defined in this first project (which information should or could be made public) into technical specifications to be implemented by the European Commission's IT services (DG SANTE EU-CEG Team).

Now, the Member States Competent Authorities (MSCA) receive a monthly extraction of richer data than what they previously had, particularly concerning product composition. For those who had difficulties in implementing the obligation to make this information from declarations public, we hope that this will make their task easier.

Moreover, it is now possible to share these public data files to conduct cross-sectional data analyses. It is hoped that, in the future, MSCA will make more data available to the scientific community to advance knowledge in the field of tobacco and related products.

Finally, this work demonstrates that while national initiatives have enabled certain Member States to carry out advanced processing of their EU-CEG data for scientific research or regulatory enforcement purposes, collaboration with DG SANTE EU-CEG Team to improve the EU-CEG portal's back-office interface (MS-Rep) allows the development of functionalities useful to all.

In the future, it would be worthwhile to develop true data visualization tools (such as BI-type tools) aimed at the general public. A revision of the directive and the data reporting format should include consideration of securing the legal framework to define the nature of public data and the conditions for making it available for research purposes.

10 ANNEX: proposal for publication thresholds of ingredients other than tobacco, according to Ingredient Function and Product Type (based on provisions laid on art.6(2) of Decision of execution (EC) 2015/2186)

Value	Name	Publication Threshold (% w/w)
1	Addictive Enhancer	always public
2	Adhesive	0.5%
3	Binder	always public
4	Carrier	0.5%
5	Colour	always public
6	Combustion Modifier	always public
7	Casing	always public
8	Fibre	0.5%
9	Filler	0.5%
10	Filter Component	0.5%
11	Filtration Material	0.5%
12	Flavour and/or Taste Enhancer	0.1% (Cigarette, RYO) 0.5% (other TP types)
13	Humectant	always public
14	pH Modifier	always public
15	Plasticiser	always public
16	Preservative	always public
17	Solvent - Processing Aid	0.5%
18	Reduced Ignition Propensity Agent	always public
19	Sizing Agent	0.5%
20	Smoke Enhancer	always public
21	Smoke Colour Modifier	always public
22	Smoke Odour Modifier	0.1% (Cigarette, RYO) 0.5% (other TP types)
23	Wrapper	0.5%
24	Water-Wetting Agents	always public
25	Viscosity Modifier	always public
26	Other	0.5%