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

Member States Experience in Assessing EU-CEG data

Authors: ANSES
September 2024

Doc. Ref. N°: D5.8
Type: Document (R)
Dissemination: Public (P)

JATC 2 JOINT ACTION ON TOBACCO CONTROL

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

Version	Date	Authors	Comments
1	30 Sept. 2024	ANSES	First version

Table of contents

Introduction

The EU Common Entry Gate (EU-CEG) serves as a centralized data submission system where manufacturers of tobacco and related products notify national authorities about the characteristics and ingredients of their products.

Member States Competent Authorities (MSCA) are responsible for managing, analyzing, and ensuring compliance with these submissions according to both national and EU regulatory frameworks.

The purpose of this document is to provide an overview of the key aspects of EU-CEG data handling by MSCA, highlighting practices related to data integration, validation, analysis, quality control, and public dissemination. For confidentiality, this document does not reference specific countries or organizations.

1 Webinar workshop on EU-CEG data

As part of Work Package 5 (WP5) of the JATC-2 project, a webinar workshop was organized on April 27, 2022, with the EU Member States Competent Authorities (MSCA) to share the experiences of the most advanced users in managing EU-CEG data and to address the needs of the MSCA that were less advanced.

Prior to the discussions that took place during this webinar, a questionnaire was sent to participants asking them to describe how they handle EU-CEG data and their future processing needs.

2 Key aspects of EU-CEG data handling

2.1 Data Integration and Storage

MSCA across the EU utilize a variety of approaches to integrate and store EU-CEG data. The main data integration process involves extracting XML and Excel-based submissions from the EU-CEG system and uploading them into local databases. The following key points summarize common practices:

- **Data Import and Transformation:**
  - Most authorities use custom scripts and software solutions (e.g., Microsoft Access, Macro Excel, and local SQL databases like Oracle) to convert raw EU-CEG submissions into a standardized format suitable for further processing.
  - Some MSCA employ **Robotic Process Automation (RPA)** for automated data extraction and integration, reducing the manual workload associated with repetitive tasks.
- **Centralized Data Storage:**
  - Data is typically housed in a **local SQL database**, which serves as the backbone for data retrieval and analysis.
  - Temporary databases are often used to hold incoming data before full integration into a production database, allowing for initial validation checks and formatting adjustments.
- **Data Synchronization:**
  - Some MSCA use automated scheduling tools to keep their local databases synchronized with the EU-CEG central repository. This ensures that the local data remains up-to-date and reflects any modifications made at the EU level.

2.2 Data Curation and Quality Control

Ensuring the integrity and consistency of EU-CEG data is a top priority for MSCA. Various curation and quality control processes are implemented to address issues such as incomplete submissions, incorrect data formatting, and inconsistent ingredient nomenclature.

- **Format Curation:**
  - Most MSCA employ tools like **Microsoft Access** to convert data fields from text-based formats into appropriate datatypes (e.g., numerical, date, or categorical) to ensure consistency and facilitate downstream analysis.
  - Semi-automated correction procedures are used to address common errors, such as improperly formatted CAS numbers or ingredient names.
- **Ingredient Mapping and Validation:**
  - Authorities often maintain **reference tables** for ingredient data, which are used to map submitted ingredient information to standardized identifiers (e.g., CAS numbers, IUPAC names).
  - Chemicals are manually reviewed and validated using external databases (e.g., PubChem) to ensure that all ingredient entries are scientifically accurate.
- **Error Reporting and Handling:**
  - Custom scripts and macros are used to identify and flag discrepancies, such as missing data fields or incorrect values, which are then reviewed and corrected either automatically or manually by data curators.

2.3 Data Analysis and Reporting

MSCA leverage EU-CEG data for various analytical purposes, ranging from compliance checks to market trend analyses. The insights gained from these analyses are used to support both regulatory decisions and public health initiatives.

- **Market Surveillance and Compliance Analysis:**
  - Authorities analyze product characteristics (e.g., ingredient composition, nicotine levels, packaging details) to ensure that submissions comply with regulatory standards.
  - Trend analyses are conducted to track changes in product formulations, market entries, and the use of specific additives over time.
- **Data Mining and Visualization:**
  - Some MSCA use **R programming** in conjunction with Microsoft Excel for more advanced data mining, including clustering analysis, product similarity assessments, and ingredient prevalence studies.
  - Visualization tools are employed to reveal patterns in ingredient usage and detect outliers that may indicate non-compliance.
- **Public Health Research:**
  - EU-CEG data are used to evaluate the impact of new products on public health, such as studying the prevalence and concentration of harmful ingredients.
  - MSCA also focus on specific research questions, such as understanding the role of different nicotine formulations (e.g., nicotine salts) and their effects on user behavior.

2.4 Publication and Data Dissemination

MSCA aim to balance transparency with the need to protect proprietary information. Many have established public portals to share selected EU-CEG data with stakeholders and the general public, in line with EU directives.

- **Public Data Portals:**
  - Public-facing websites are developed to provide access to aggregated product and ingredient information. This promotes transparency while ensuring compliance with data protection regulations.
  - The data available typically include basic product characteristics, ingredient lists, and compliance status, excluding any confidential business information (CBI).
- **Customized Reporting:**
  - Authorities provide tailored reports in response to specific queries from other governmental

agencies, public health organizations, or industry stakeholders.
  - Examples include lists of approved products, summaries of compliance status, and detailed ingredient breakdowns.

2.5 Software and Technical Infrastructure

MSCA utilize a diverse set of software tools to manage and analyze EU-CEG data, including both commercial and open-source solutions. Key software tools include:

- **Microsoft Access and Excel:** Widely used for data integration, curation, and basic analysis. Custom macros and scripts are often developed for automation.
- **SQL Databases:** Serve as the primary storage and retrieval platform for EU-CEG data.
- **BI software:** Used for tailored data analysis (eg. SAP/Qlik).
- **R Programming:** Used for complex data analysis, modeling, and visualization.
- **Robotic Process Automation (RPA):** Automates repetitive tasks, such as data extraction and report generation.
- **Custom Interfaces and APIs:** Developed for seamless integration with EU-CEG and other external systems.

Conclusion: challenges and future directions

Despite significant efforts to establish robust data management systems, MSCA face ongoing challenges related to resource limitations, data volume, and evolving regulatory requirements.

- **Scalability and Data Volume:**
  - The sheer volume of EU-CEG data continues to grow, putting pressure on existing storage and processing infrastructure.
- **Harmonization and Standardization:**
  - Variations in data handling practices across MSCA can lead to inconsistencies in reporting and compliance assessments.
  - Ongoing efforts to harmonize data standards at the EU level aim to streamline processes and enhance data comparability.
- **Integration of New Data Types:**
  - With the introduction of emerging new tobacco and related product, MSCA are adapting their systems to accommodate new data types and analytical needs.
  - Future developments will likely include the use of machine learning algorithms for predictive modeling and automated compliance checks.

The management of EU-CEG data by MSCA involves complex workflows that span data integration, curation, quality control, analysis, and dissemination. While the specific practices vary, a common focus on data integrity, regulatory compliance, and public health underpins these efforts.

By leveraging advanced software tools and collaborative frameworks, MSCA continue to refine their data handling practices to meet evolving regulatory challenges and support informed decision-making.