3 April 2025

SMS Guidance for external users

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000 An agency of the European Union

Table of contents

1. Introduction 3

2. SMS data and data fields 5

3. Substance confidentiality 8

3.1. General substance confidentiality principles 8

3.2. Confidentiality for Human substances 10

3.3. Confidentiality for Veterinary substances 10

4. Accessing substance data 11

4.1. SMS User Interface 12

4.2. Downloading data 13

4.3. Application Programming Interface 15

5. Consuming systems 19

5.1. Data flow 19

5.2. Systems details 20

6. Change requests 22

7. Customer Service 27

8. Data Stewardship 28

8.1. Data Mappings 28

8.2. Data Enrichments 29

8.3. Data Cleansing 30

8.4. Data Quality monitoring 35

9. Document versions 37

Annex 1: Business rules per substance type 38

Annex 2: Frequently Asked Questions 43

Annex 3: Glossary 45

Annex 4: RMS Lists/IDs used in SMS 46

Annex 5: Data fields to FHIR mappings 49

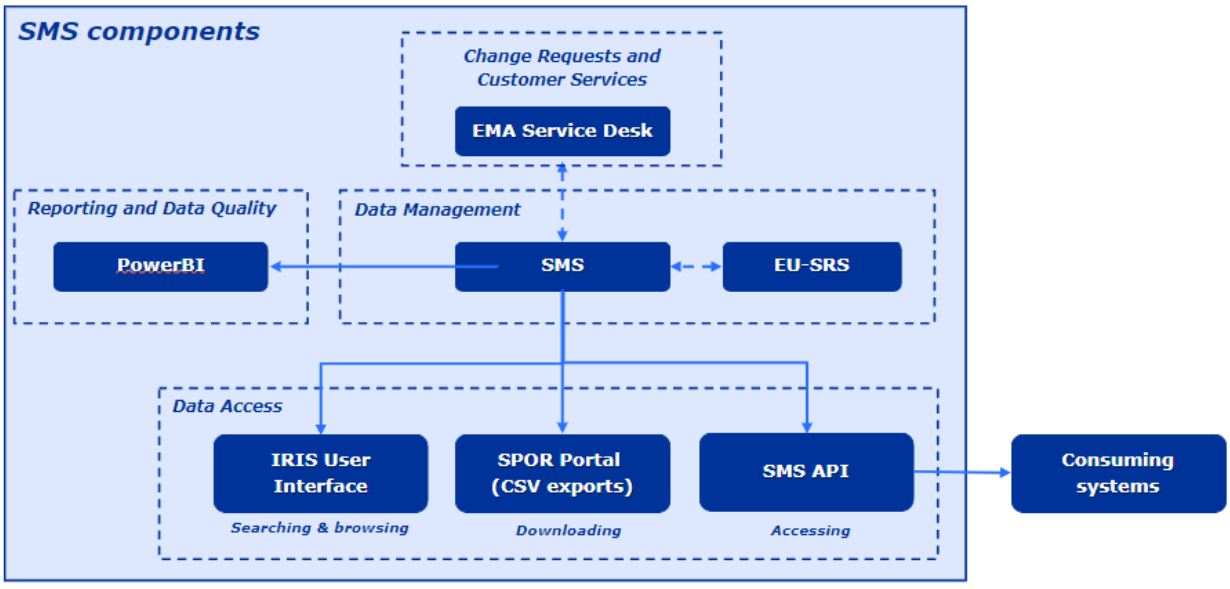
1. Introduction

Substances Management Service (SMS) is one of the four domains of Substance, Product, Organisation and Referential (SPOR) master data to be used in pharmaceutical regulatory processes. The SPOR services support the implementation of ISO IDMP standards in the European Union (EU) and European Economic Area (EEA). Although ISO IDMP standards relate to Human medicinal products, SMS manages both human and veterinary substance data.

SMS provides a central source of substance data that supports selection in regulatory processes and therefore enables to distinguish between two or more substances. It is managed by EMA data stewards for the benefit of the EU/EEA Network and Industry. EU-SRS is a complimentary system managed by the Substance Validation Group, with additional substance details, in accordance with ISO IDMP. It is currently accessible to NCAs only.

As illustrated in **Figure *1.*** SMS components, users can request changes in SMS (see Chapter *6. Change requests*) or ask questions about its processes (see *Chapter 7. Customer Service*) via the EMA Service Desk. SMS reports and data quality are analysed with PowerBI (see *Chapter 8.4 Data Quality monitoring*). SMS public data is accessible to external stakeholders via the IRIS User Interface (see *Chapter 4.1 SMS User Interface* and CSV exports in the SPOR Portal (see *Chapter 4.2 Downloading data*). SMS public data is also available the SMS Application Programming Interface (API) for Industry users, while NCAs have access to the full SMS data (see *Chapter 4.3 Application Programming Interface* . SMS data is also made available in several consuming systems such as XEVMPD, CTIS, UPD, etc. (see *Chapter 5. Consuming systems*).

Figure 1. SMS components



SMS currently supports the regulatory procedures and activities described in **Table 1**.

Table 1. Supported domain for regulatory procedures and activities

Regulatory Procedure	Domain
Innovation Task Force	Human & Veterinary
Scientific Advice	Human & Veterinary
PRIME	Human
Orphan Designation	Human
ATMP Classification & Certification	Human
Paediatric Investigation Plan	Human
Clinical Trial Application	Human
Eligibility for Centralised Procedure[1]	Human & Veterinary
Marketing Authorisation Application	Human & Veterinary
Maximum Residue Limits[1]	Veterinary
Article 58 - Medicines for use outside the EU	Human
Renewal & Variation	Human & Veterinary
Good Manufacturing Practices Certificates	Human & Veterinary
Art. 57/xEVMPD legal requirements	Human
Safety reporting	Human & Veterinary
Medicines Shortages	Human

[1]The application form for the regulatory procedure is free text, however, the information is then recorded in SIAMED, (the EMA product information and application tracking system for centrally authorised products) which consumes substance data from SMS.

2. SMS data and data fields

SMS data is managed in SMS IDD, which stands for "Informatica Data Director", a master data management tool. IDD is the back-end interface that enables specialised EMA data stewards to perform SPOR data management duties. Access to IDD is restricted to EMA users with the role "SMS data steward", which is authorised only after an appropriate extensive training.

External users do not have access to IDD and therefore cannot view, search or manage data in SPOR directly. However, external users can still access SMS data in several ways and can request data changes (i.e. creation or update of substance data) via the EMA Service Desk (see *Chapter 6. Change requests*). The accepted changes are then reflected in all systems using SMS data and in the SMS API, while the public data is also made available in the SMS exports from the SPOR portal and the SMS User Interface (UI) in IRIS. In the future, submissions of change requests are expected to be possible via a user interface, similarly to the OMS and RMS.

When SMS went live, substance data was imported into SMS IDD from xEVMPD, EUTCT and SIAMED. In addition, several data enrichment exercises were subsequently performed both based on internal EMA databases (e.g. Orphan Designation database, Scientific Advice Veterinary database) and external sources (e.g. INN and USAN). These resulted into a large dataset which is continuously undergoing cleansing and enrichment activities.

The most relevant data fields in SMS IDD are listed in **Table 2**, described by section.

Table 2. Relevant data fields in SMS IDD

Section	Data field	Conditionality	Content	Comments
Substance	SMS ID	System-generated		Unique for each substance record. SMS ID is also known as the EUTCT ID.
	Domain	Mandatory	RMS List "Domain"[2]	The choice between the "Human" or "Veterinary" domains trigger a different feedback loop to the concerned consuming systems. "Human" substances can be used in both human and veterinary products, while "Veterinary" substances can be exclusively used in veterinary products.[3] Further details are available in *Chapter 5. Consuming systems*.
	Data classification (Substance)	Mandatory	RMS List "Data Classification"[2]	SMS only uses the terms "Public" and "Restricted"
	Authorisation State	System-generated	RMS List "Substance Authorisation Status"[2]	All new substances are registered as "Authorised", which is known as "Approved" in xEVMPD, regardless of the approval status of the product. Legacy "Development" substances are kept for historical purposes only. Further details are available in *Annex 2: Frequently Asked Questions*.

[2] The respective RMS IDs and terms are listed in *Annex 4*.
[3] Substance registered with domain "Veterinary" are mainly vaccines for animal-exclusive diseases, monoclonal antibodies that target animal proteins and cell therapy substances made from animal cells for veterinary use.

	Substance Status	Mandatory	RMS List "Record Status" [2]	SMS only uses the terms "Current" and "Non-current". Users can only use "Current" records, while "Non-current" ones reflect the nullified substance records and cannot and shall not be used.
	Substance Name	System-generated	Automatically populated based on the preferred term set on Substance Name section	
	Comment	Optional	Free text	This field is mandatory when changing a substance to "non-current" (i.e. for substance nullification). It is used to flag duplicate or invalid substances.
	Substance Type	Mandatory	RMS List "Substance type"[2]	This data field has no regulatory impact.
	Molecular formula	Optional	Free text	
	Molecular weight	Optional	Numeric field (2 decimals)	
	InChIKey	Optional	Free text (27 characters)	International Chemical Identifier, which is an alphanumeric representation of the substance systematic name (mainly for chemicals). It can be generated from Opsin or other systematic name parsers.
	Last Update Date	System-generated		
Substance Name	Substance Name	Mandatory		Each substance must have at least one name, the preferred term.
	Language	Mandatory	RMS List "Language"[2]	The preferred term must be in English or Latin. Aliases (synonyms and translations) can be in any official EU/EEA language.
	Is Preferred	Mandatory	Binary	Only one preferred term can be set per record.
	Name Status	Mandatory	RMS List "Record Status"[2]	
	Data Classification (Name)	Mandatory	RMS List "Data Classification"[2]	Implications on confidentiality. Further details are available in *Chapter 3.1.General substance confidentiality principles*.
Substance Name Source	Source (Name Source)	Mandatory	RMS List "Source of Information"[2]	The source of information is required for all English and Latin terms.
Substance External Code	Source (External Code)	Optional	RMS List "Source of Information" – External codes (All with attribute "SMS Code")[2]	System-generated for EV Codes for Human substances. External codes (e.g. SVG flag, UNII, etc.) can be manually added.

	ID/Code Provided by Source	System-generated for EV Codes (Human) Optional for other codes	External code	
Substance Custom Attribute	Attribute Term	Optional	RMS List "SMS Custom Attribute" [2]	Active substance in valid EEA AMP Critical medicine
	Attribute Value	Optional	Free text	
Substance Relationship	Related substance	Optional	SMS ID	SMS ID of the related substance
	Substance Relationship Combination	Conditional	RMS List "Substance Relationship Combination"[2]	
Substance Current	Substance Current	Optional	SMS ID	SMS ID of the replacement substance

3. Substance confidentiality

3.1. General substance confidentiality principles

The confidentiality of substance data is managed in SMS IDD via the data fields "Data classification", both at substance and name level:

- **Data classification (substance)**: this data field allows the management, at substance record level, of the confidentiality of the molecular formula, molecular weight, InChIKey and enables the registration of restricted substance names in the Substance Name section;
- **Data classification (name)**: this data field allows the management, at substance name level, of the confidentiality of each individual name.

The data field "Data classification" can be set, both at substance and name level, as:

- **Public**: the concerned data is made publicly visible to all users (via consuming systems, substance exports in SPOR Portal and SMS API);
- **Restricted**: the concerned data is only made visible to EMA SMS Data Stewards, EMA IRIS users and NCA users (via the SMS API).

The tables below (**Table 3** and

Table *4*) further describe the management of the confidentiality of data via the two "Data Classification" data fields and the implemented action at system level as per the selected value (i.e. "Public" vs "Restricted"). Examples of public and restricted substances can be seen in

Figure 2. Example of Public (SMS ID: 100000128703) and Restricted (SMS ID: 300000032819) substances in the Substance export published in the SPOR Portal

Table 3. Data classification at substance level

Data Classification (substance)	Molecular formula	Molecular weight	InChIKey	Restricted names
Public	Published (if registered)	Published (if registered)	Published (if registered)	Not allowed for any name
Restricted	Not published	Not published	Not published	Allowed for aliases

Table 4. Data classification at name level

Data Classification (name)	Name
Public	Published
Restricted	Not published

Figure 2. Example of Public (SMS ID: 100000128703) and Restricted (SMS ID: 300000032819) substances in the Substance export published in the SPOR Portal

A	B	C	D	I	K	O	P	Q
#SMS_ID	Substance_Name	Is_Pref	Language	Molecular_Formula	Inchikey	External_Code_XEVMPD	External_Code_SVG	External_Code_UNI
100000128703	Olodaterol	TRUE	English	C21H26N2O5	COUYJEVMBVSIHV-SFHVURJKSA-N	SUB36104	1	VD2YSN1AFD
100000128703	BI 1744	FALSE	English	C21H26N2O5	COUYJEVMBVSIHV-SFHVURJKSA-N	SUB36104	1	VD2YSN1AFD
100000128703	Olodaterool	FALSE	Estonian	C21H26N2O5	COUYJEVMBVSIHV-SFHVURJKSA-N	SUB36104	1	VD2YSN1AFD
100000128703	OLODATEROL	FALSE	German	C21H26N2O5	COUYJEVMBVSIHV-SFHVURJKSA-N	SUB36104	1	VD2YSN1AFD
100000128703	OLODATEROL	FALSE	Hungarian	C21H26N2O5	COUYJEVMBVSIHV-SFHVURJKSA-N	SUB36104	1	VD2YSN1AFD
100000128703	Olodaterol	FALSE	Polish	C21H26N2O5	COUYJEVMBVSIHV-SFHVURJKSA-N	SUB36104	1	VD2YSN1AFD
300000032819	BI 1819479	TRUE	English			SUB222429	1	

It is important to be noted that even when the data classification is set as "Restricted", the following limited information **will always be made public**:

- SMS ID and associated EV code.
- preferred term and the respective name source.
- official names (e.g. INN, USAN) and the respective name source.

However, the restricted information will not be visible to industry users, including the requestor, and it will only be visible to EMA SMS data stewards, EMA IRIS users and NCA users accessing SMS data via the SMS API.

Depending on the concerned regulatory procedure, the applicant may designate a company code as the preferred term. In case a company code is requested as the substance preferred term, a systematic name or a common name describing the substance must still be provided. Any additional names will be registered as "restricted" and will not be displayed to the public nor published; this data will only be made visible to EMA SMS Data Stewards, EMA IRIS users and NCA users (via the SMS API and IRIS).

While these general substance confidentiality principles apply to both human and veterinary substance data, there are several particularities implied for each domain that will be further addressed and detailed in the next chapters.

3.2. Confidentiality for Human substances

The information in this chapter is provided for informational purpose only as it is currently under review by various relevant stakeholders, and it is therefore subject to change and further revision during 2025.

As per the general principles, each substance record shall have a public preferred term. However, when selecting the appropriate preferred term, it is essential to consider the regulatory procedure for which the substance record is required.

While for substances requested for use in PRIME, scientific advice or clinical trial applications it is possible for the company code to be set as the preferred term, with all the remaining names being registered as restricted, for Orphan Designations and Paediatric Investigation Plans the company code cannot be used as the substance preferred term. The use of a company code as the preferred term, in such instances, will block the regulatory procedure until a substance request for a preferred term update is submitted and processed. Instead, as required by the European Commission, a systematic name, for chemicals, or a common name, for other substance types, must be used.

In **Table 5**, information on setting the preferred term, taking into account the concerned regulatory purpose, is further detailed.

Table 5. Selection of the preferred term for human domain

Regulatory procedure	Company code allowed as preferred term?
PRIME	Yes, if no official name (e.g. INN, Ph. Eur.) is available
Scientific Advise	
Clinical Trials	
Marketing Authorisation Application	No, a systematic name, for chemicals, or a common name, for other substance types, must be used.
Orphan Designation	
Paediatric Investigation Plan	

3.3. Confidentiality for Veterinary substances

For veterinary substances, the impact of the concerned regulatory procedure on the selection of the preferred term is also essential. In **Table 6**, further details are made available regarding the main regulatory procedures using veterinary substances and the possibilities of using a company code as the substance preferred term, with all the available names being registered as restricted.

Table 6. Selection of the preferred term for the veterinary domain

Regulatory procedure	Company code allowed as preferred term?
Scientific Advice	Yes, if all conditions apply simultaneously: • The substance is not used in any approved Veterinary medicinal product; • The substance does not have an official name (e.g. INN, USAN, Ph. Eur.). • The systematic name or common name is requested as "Restricted" and the company code is requested as "Public" by the Applicant in the Substance Request Form.
Marketing Authorisation Application (Centralised)	
Marketing Authorisation Application (MRP/DCP/National)	
Variation	No, a common name must be used.

4. Accessing substance data

SMS substance data can be accessed through several methods, depending on the user type. The concerned system will display the appropriate data, public or restricted, based on the user's authorisation level:

- SMS IDD - access is possible for EMA SMS data stewards who can view and manage all, both public and restricted, substance data.
- SMS User Interface (UI) in IRIS – access is possible for all users who can view limited public data; EMA IRIS users and can view additional restricted information.
- SMS tab in the SPOR portal – access is possible for all users, who can download almost all public data.
- SMS Fast Healthcare Interoperability Resources Application Programming Interface (FHIR API) – access is possible for EMA and NCA users who can access public and restricted substance data, and to any other users who can only access public data.

Table 7. Substance data access

User	**SMS IDD** *Browse* *Search* *Edit* *Export*	**IRIS (SMS UI)** *Browse* *Search*	**EUTCT** *Browse* *Search*	**SPOR Portal** *Export*	**API** *Browse* *Search*
EMA	Public and restricted data	Public and restricted data	Public data	Public data	Public and restricted data
NCA	N.A.	Public data	Public data	Public data	Public and restricted data
Any other	N.A.	Public data	Public data	Public data	Public data

4.1. SMS User Interface

The SMS User Interface (UI) is available to the public via IRIS, no login or credentials being required to access it. The SMS UI makes available to the public a set of public SMS substance data which is updated via the SMS API on a daily basis.

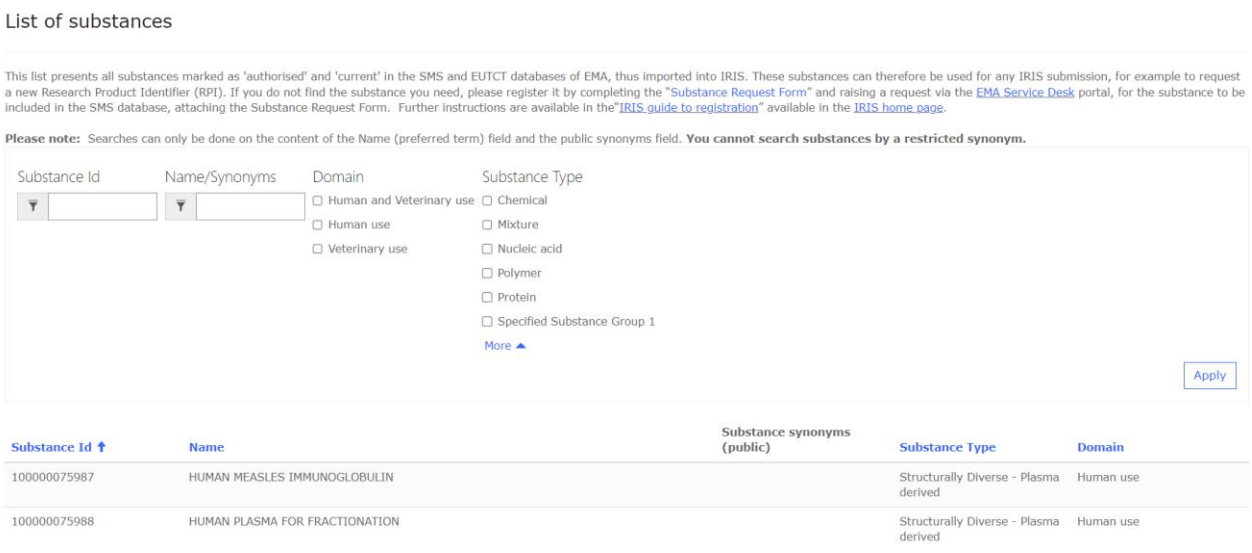
The following criteria is applied for the published SMS substance records:

- **Substance Authorisation State** in SMS IDD [RMS list ID 200000004906 "Substance Authorisation Status"] is set as "Authorised" (i.e. approved substance records) [RMS term ID 200000004972].
- **Substance Status** in SMS IDD [RMS list ID 200000005003 "Record Status"] is set as "Current" [RMS term ID 200000005004].

The following SMS data fields are currently available, as illustrated in **Figure *3.*** SMS User interface in IRIS:

- SMS ID.
- Preferred term.
- Public aliases (i.e. synonyms in English, with Data Classification at name level set as "Public").
- Substance type.
- Domain.

Figure 3. SMS User interface in IRIS



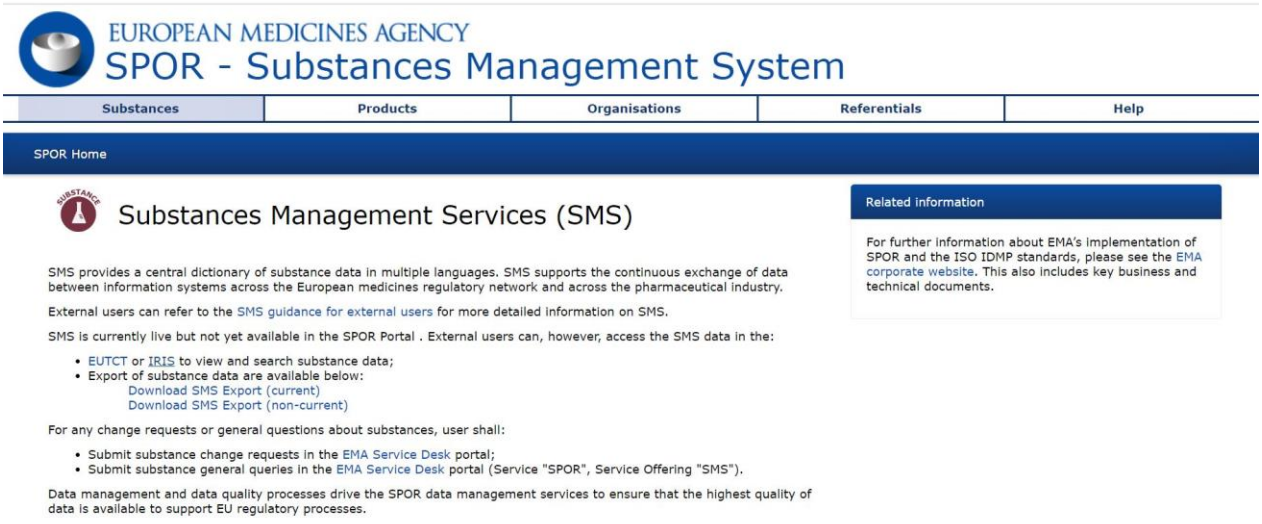
Additional data fields are expected to be added in the future:

- Translations.
- External codes (e.g. EV Code, SVG flag, UNII, etc.).
- Replacement substance ID (for identified duplicates).

4.2. Downloading data

In the SMS tab in the SPOR Portal, two comma-separated (.CSV) substance export files are made available. One file contains public information of all the substances with the status set as "Current" and the second one contains the public data of those records with the status set as "Non-current". These exports are refreshed every night and are available to all users, without the need for credentials.

Figure 4. The SMS tab in the SPOR Portal



Users can click on the appropriate link to download either the current or non-current .CSV export. In order to open the .CSV substance exports and access the substance data, the file should be opened and imported into a spreadsheet software. The following steps can be taken in Excel:

1. Click on the respective link to download the current or non-current export
2. Open Excel
3. Click "Blank workbook", "Data", "From Text/CSV"
4. Select the downloaded CSV file
5. Click load
6. Select column A, then change from "General" to "Number"
7. Remove the decimal cases by clicking twice on the button below

Both exports will reference the data fields described in **Table 8**.

Table 8. Data fields available in the public substance exports

Data field	Details
#SMS_ID	System-generated ID associated with a substance name
Substance_Name	Substance name, which can be the preferred term, a "Public" alias or a translation
Is_Preferred_Name	Information on the preferred name: the preferred term is indicated by the label "Yes"; the aliases and translations are marked with the label "No"
Language	Language of the concerned substance name as per the respective RMS list; the preferred term and aliases are in English or Latin, while the translations are in one of the official EU/EEA languages
Name_Source	Name of the source of information for the substance name as per the respective RMS list
Substance_Domain	Appropriate domain as per the respective RMS list (i.e. "Human" or "Veterinary")
Status	Status of the substance name as per the respective RMS list (i.e. "Current" or "Non-current"); each export only contains one status
Substance_Type	Substance type as per the respective RMS list
Molecular_Formula	Molecular formula, if applicable (only if the substance record is set as "Public")
Molecular_Weight	Molecular weight (field not yet populated; only applicable if the substance record is set as "Public")
InchIKey	InchIKey (field only populated for chemicals; Only applicable if the substance record is set as "Public")
Comment	Comment associated with the concerned term; it may reference the SMS ID/EV Code of an appropriate replacement substance for duplicate records. Additional information provided in *Chapter 8.3. Data Cleansing*
Created_Date	Date when the substance was created
Last_Updated_Date	Date when the substance was last updated
EV Code	Substance EV code from XEVMPD (only applicable to substances with domain "Human")
SVG flag	The label "1" indicates a valid substance, while the label "0" reflects an invalid/duplicated substance, empty data field not yet reviewed
UNII	Unique Ingredient Identifier from FDA
INN_number	International Nonproprietary Name number from WHO
EC_List/Number	European Commission list or number
Parent_Substance	SMS ID of the parent substance
Replacement_substance	Recommended replacement substance (only applicable for duplicates in the "Non-current" export)

4.3. Application Programming Interface

SMS exposes an Application Programming Interface (API) for system-to-system integration in order to query and retrieve substance data.

The SMS API is implemented using the Fast Healthcare Interoperability Resources (FHIR) standard. The API is currently on FHIR version 4.4.0 (R5 preview 2) and supports the SubstanceDefinition resource SubstanceDefinition - FHIR v4.4.0 (hl7.org).

The API is available for both NCAs and general users. However, general users will be able to view only public data fields. Endpoints 201 and 202 from the SPOR API Specification V2 R5 (europa.eu) are currently in scope.

Authentication to the API is using the OAuth2 standard Client credentials flow.

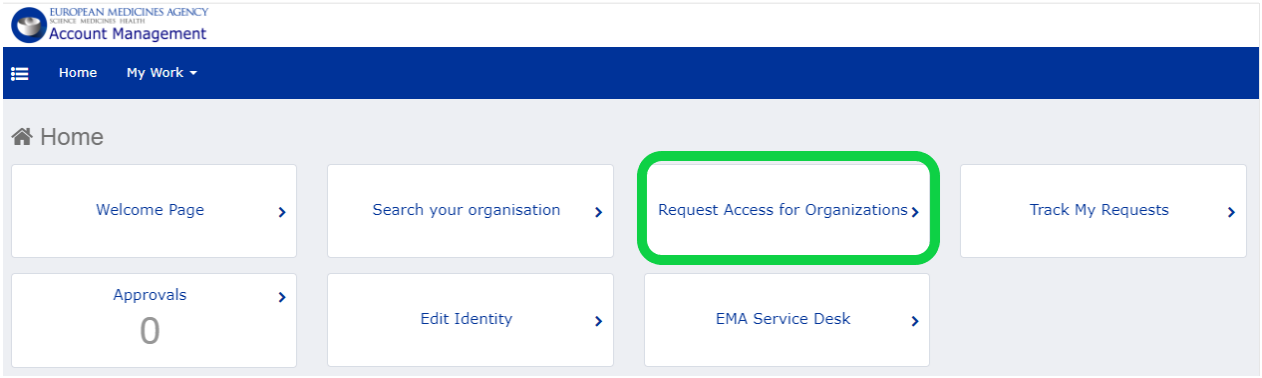
Table 9. API properties

API Property	Value
Base Path	https://spor.azure-api.net/sms/api/v3
Protocol	HTTPS
Authentication	OAuth2 Client Credentials flow
FHIR Version	4.4.0 (R5 Preview 2)

4.3.1. API registration process

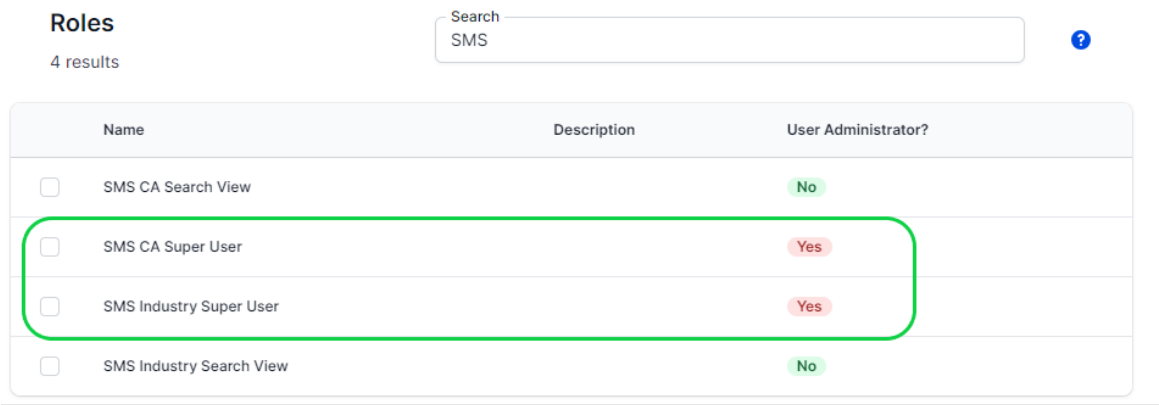
In order to generate client credentials, users should request their relevant SMS Super User role. Roles can be requested by registering on EMA Account Management portal. Once registered, users can login and request access in the context of the Organisation(s) they represent.

Figure 5. EAM portal link to Request access



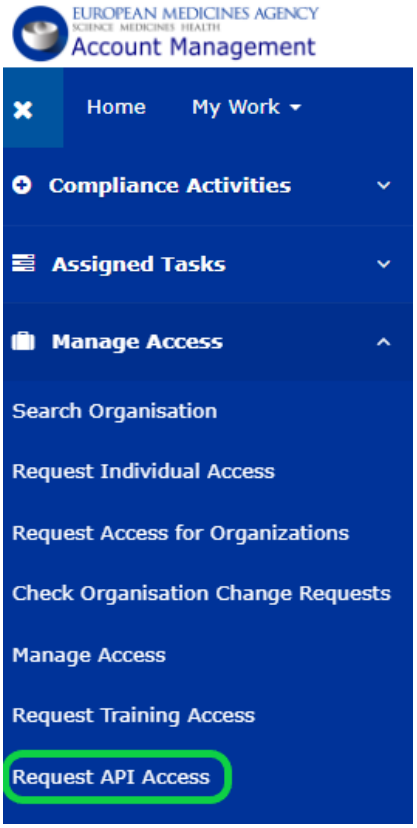
Once an Organisation is chosen, the user shall search for SMS roles (**Figure 6**) and select the relevant Super User role.

Figure 6. SMS Super User Roles



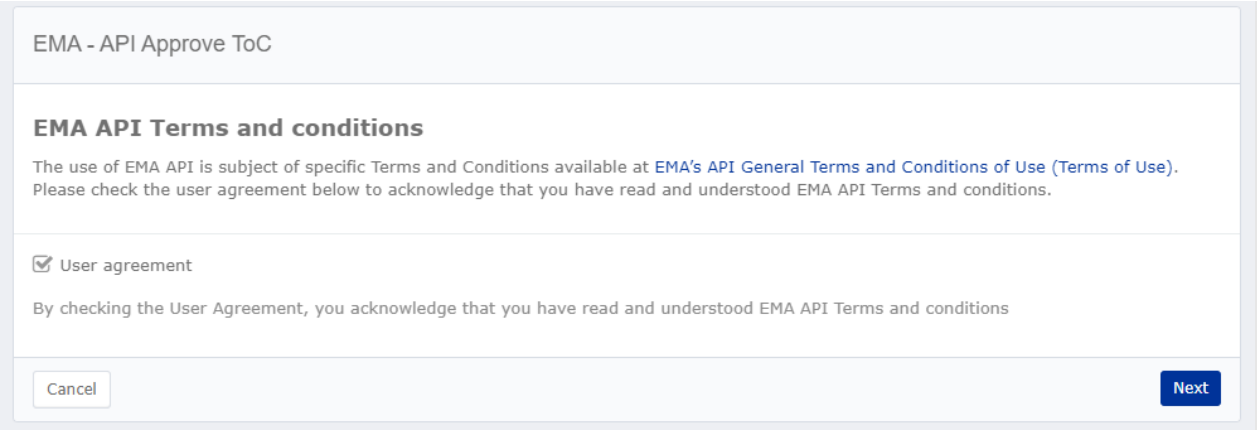
Once submitted, the request will be made available for approval by the system owner. Once the role is approved, the user will be able to request API Access from the side menu (**Figure *7***).

Figure 7. EAM Request API Access menu item



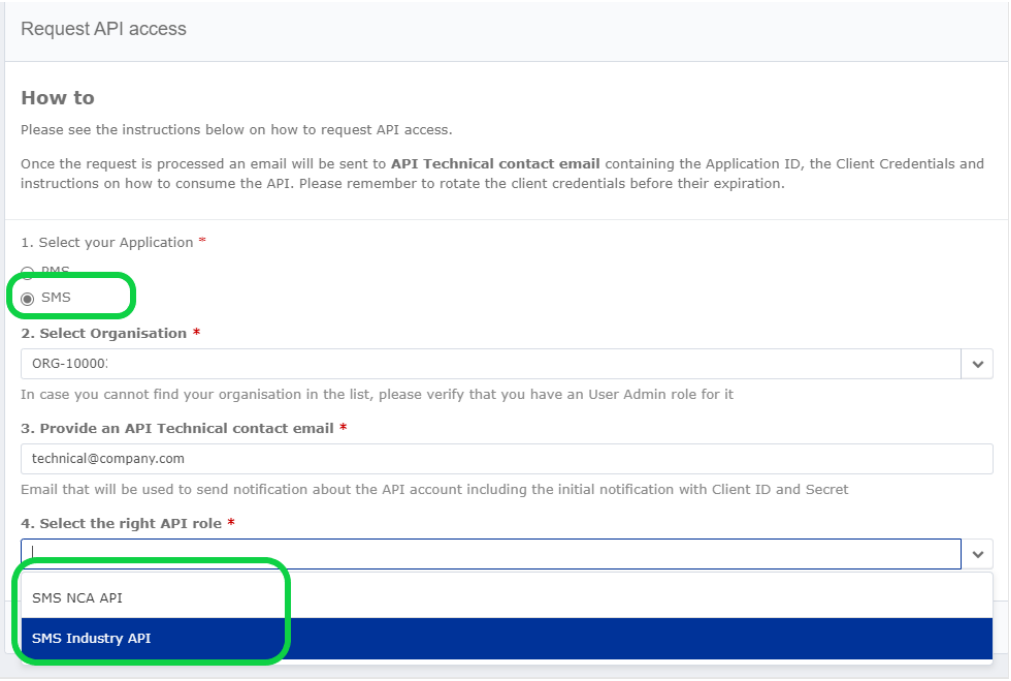
In order to proceed with the request for the API Access, the user shall first accept the API terms of use.

Figure 8. API Terms and conditions screen



In the next screen, the user shall select SMS and the relevant Organisation for which they would like to request access. An e-mail address shall be provided in order to receive the client credentials details and further select the relevant API role.

Figure 9. Request API access screen



Finally, an e-mail with all the information necessary for OAuth2 Client credentials flow authentication will be sent to the provided e-mail address. The following information is included in the e-mail:

- Client ID.
- Client Secret.
- OAuth2 Token Endpoint.
- Scope.

Figure 10. Example of e-mail with OAuth2 information

Application Registration	
Object ID	8e379c2a-1385-409a-a22c-a467c79e643a
Client ID	e1e68a03-6b1f-47d1-a158-
Display Name	spor-apim-s2s-client

Application Secret	
Key ID	39ece8a8-7c7a-41a0-b63c-3220410926ff
Display Name	spor-apim-s2s-client-54
Expiration Date	10/20/2025 12:16:13
Secret	pKZ8(

OAuth 2.0 - Client Code Grant Flow

OAuth 2.0 token endpoint (v2)	GET https://login.microsoftonline.com/4efbf65c-4a81-4f2d-835a-e8630de67663/oauth2/v2.0/token
Scope	api://euemanp.

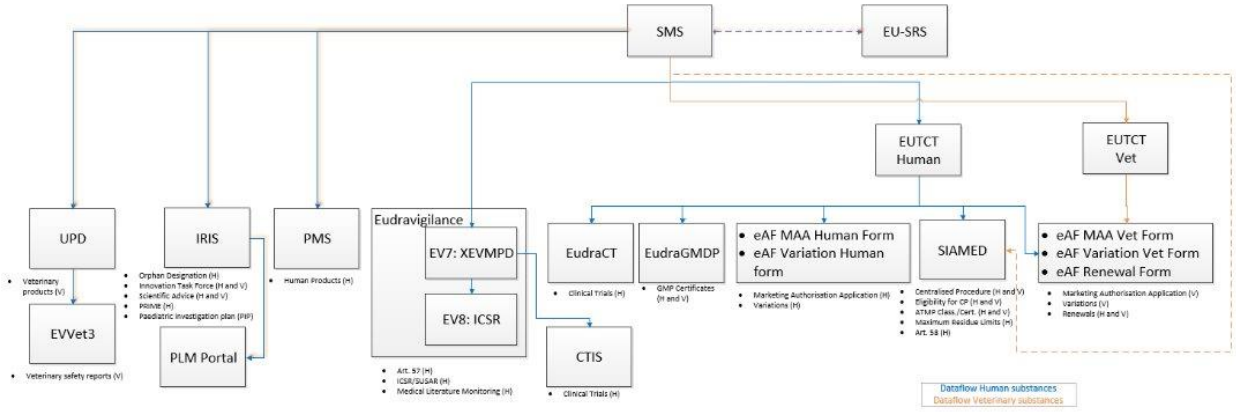
5. Consuming systems

5.1. Data flow

Substances created or updated in SMS are automatically synchronised with all the appropriate consuming systems, depending on the substance's domain (i.e. Human or Veterinary), as presented in **Figure 11**. Like it was mentioned in *Chapter 2. SMS data and data fields,* this data field has no regulatory impact, and it only impacts the consuming systems that will display the concerned substance.

Substances registered as "Human" are made available in all consuming systems (represented by blue arrows), while substances registered as "Veterinary" are only made available in certain systems (represented by orange arrows). Substance registered with domain "Veterinary" are mainly vaccines for animal-exclusive diseases, monoclonal antibodies that target animal proteins and cell therapy substances made from animal cells for veterinary use; all other substances are registered as "Human" in SMS.

Figure 11. Substance data flow



Once a *human substance record* is registered in SMS, the substance data is sent to the Extended EudraVigilance Medicinal Product Dictionary (xEVMPD) via a feedback loop. After several minutes, the substance record is automatically created in the xEVMPD, and an EV code is assigned to the new approved substance record. The substance EV Code is then provided to SMS, with the SMS record being automatically updated with the new external code. From the xEVMPD, substances can be used in the whole EudraVigilance module (particularly in pharmacovigilance activities) and in CTIS (for use in clinical trials).

In parallel, the new human substance record is also sent from SMS to EUTCT Human list, within seconds, where the public information can be browsed and viewed by any guest user. From EUTCT Human, the substance record is further made available in:

- EudraCT: for clinical trials.
- EudraGMDP: for GMP certificates.
- eAF forms:
  - Marketing Authorisation Application Human.
  - Marketing Authorisation Application Veterinary.
  - Variation Human.

  - Variation Veterinary.
  - Renewal.
- SIAMED: for EMA management of Centralised Procedure, Eligibility, ATMP Classification/Certification, Maximum Residue Limits and Art. 58 procedures (medicines for use in outside EU).

Both human and veterinary substance records are provided, via the SMS API, to the following systems:

- IRIS: for PRIME, Innovation Task Force, Orphan Designation, Innovation Task Force and Scientific Advice.
- UPD: for the registration of Veterinary Medicinal Products.
- EVVet3: for the pharmacovigilance of Veterinary Medicinal Products.

For *veterinary substance records*, the substance data is made available in EUTCT Veterinary list, for further use in the eAF for the "Marketing Authorisation Application - Veterinary", "Variation Veterinary" and "Renewal" forms. Since there is no automatic synchronisation of veterinary substances to SIAMED, they are kept aligned via a manual process (marked by dashed lines in the data flow).

SMS and EU-SRS are currently being aligned via data fixes/batches of data.

The presented data flow and principles also apply for substance updates, with the respective substance changes being synchronized with all the concerned consuming systems accordingly.

5.2. Systems details

SMS substance data is made available in multiple systems, various data fields being displayed depending on the particular needs and requirements of the concerned consuming system. This data is then maintained up to date in all systems, as it is synchronised at a specific interval. Detailed information on the consuming systems is further presented in **Table 10**.

Table 10. Consuming systems details

Consuming System	Substance Domain	Substances available	Data fields available	Synchronisation Frequency
IRIS	Human and Veterinary	Auth. State: Authorised Status: Current	• SMS ID • Preferred term • Public aliases • Domain • Substance type	Daily
UPD	Human and Veterinary	Auth. State: Authorised Status: Current	• Latest updated substance name (not necessary the Preferred term)	Daily
EudraVigilance/ xEVMPD	Human	Auth. State: Authorised (Approved) and Development (legacy data) Status: Current and Non-current (Nullified)	• EV Code • Authorisation state • Status • Preferred term (with Source) • Substance type • Molecular formula (if public) • Comments (if public) • Public aliases (with Source)	Within minutes

Consuming System	Substance Domain	Substances available	Data fields available	Synchronisation Frequency
			• Translations	
CTIS	Human	Auth. State: Authorised and Development Status: Current	• EV Code • Preferred term	Daily
EUTCT Human	Human	Auth. State: Authorised and Development Status: Current and Non-current	• SMS ID • Authorisation state • Status • Preferred term (with Source) • Public aliases (with Source) • Translations • Molecular formula (if public) • EV Code	Within seconds
EUTCT Veterinary	Veterinary			
EudraCT	Human	Auth. State: Authorised and Development Status: Current	• EV Code • Preferred term (with Source) • Public aliases • Translations • Authorisation state • Status	Within seconds
eAF MAA Human	Human	Auth. State: Authorised Status: Current	• Preferred term	Within seconds
eAF Variation Human				
eAF MAA Veterinary	Human and Veterinary			
eAF Variation Veterinary				
eAF Renewal				
SIAMED	Human and Veterinary	Auth. State: Authorised and Development Status: Current and Non-current	• SMS ID • Authorisation state • Status • Preferred term (with source) • Public aliases • Translations	Daily

6. Change requests

Once the creation or update of a substance record has been deemed necessary, a change request shall be submitted via the EMA Service Desk. In such cases, the following steps shall be followed for submitting a substance request:

1. *Check if the concerned substance data is already available.*
2. *Complete the Substance Request Form.*
3. *Prepare the required supporting documentation.*
4. *Submit the ticket via the EMA Service Desk.*
5. *Wait for the resolution of the request.*

1. Check if the concerned substance data is already available

Before creating a substance request, the applicant should first check if the substance is already available. The rationale for this step is that the concerned substance might already be registered, as needed or with a different preferred term, and can therefore be used by the applicant, even if the update of the substance record might still be required.

This can be done by searching in the concerned consuming system, in the SMS UI in IRIS or in the substance exports published in the SPOR Portal. Further information on accessing substance data and substance consuming systems is available in *Chapter 4. Accessing substance data* and *Chapter 5. Consuming systems,* respectively.

2. Complete the Substance Request Form

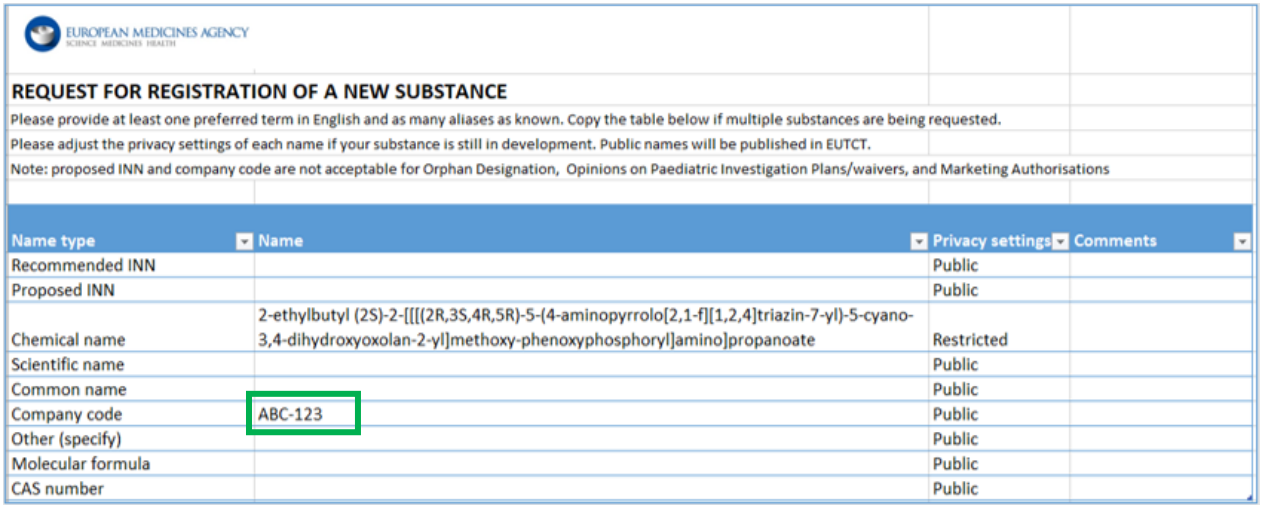
The applicant shall download and fill out the Substance Request Form. All available substance names shall be provided and, if needed, the "Privacy settings" can be set to "Restricted" individually. However, *at least one substance name must always be set as "Public" because every substance record shall have a public preferred term*, as detailed in *Chapter 3. Substance confidentiality*.

The choice for the substance preferred term will be decided by the SMS Data Stewards, based on the names provided, the concerned regulatory procedure and the naming rules defined in Annex 1 (to be published in 2024).

Some examples of completed Substance Request Forms can be found below, in the **Figures 10-13**. The preferred term is highlighted in each example and this represents the term which will be displayed to the public.

In **Figure 12**, the applicant has submitted a request for registration of a chemical for a clinical trial application and has provided the chemical name, set as restricted, and the company code, set as public. Since in the context of clinical trials, the company code can be registered as the preferred term, this term will be registered accordingly as the preferred term and the chemical name will be recorded as "Restricted".

Figure 12. Request for the registration of a Chemical for a Clinical Trial



EUROPEAN MEDICINES AGENCY

REQUEST FOR REGISTRATION OF A NEW SUBSTANCE

Please provide at least one preferred term in English and as many aliases as known. Copy the table below if multiple substances are being requested.

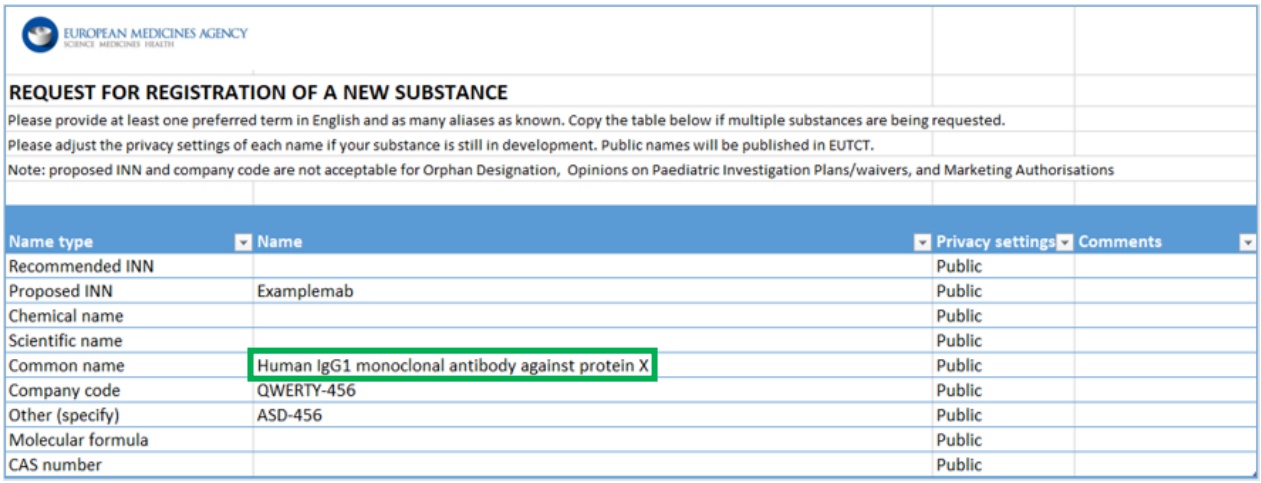
Please adjust the privacy settings of each name if your substance is still in development. Public names will be published in EUTCT.

Note: proposed INN and company code are not acceptable for Orphan Designation, Opinions on Paediatric Investigation Plans/waivers, and Marketing Authorisations

Name type	Name	Privacy settings	Comments
Recommended INN		Public	
Proposed INN		Public	
Chemical name	2-ethylbutyl (2S)-2-[[[(2R,3S,4R,5R)-5-(4-aminopyrrolo[2,1-f][1,2,4]triazin-7-yl)-5-cyano-3,4-dihydroxyoxolan-2-yl]methoxy-phenoxyphosphoryl]amino]propanoate	Restricted	
Scientific name		Public	
Common name		Public	
Company code	ABC-123	Public	
Other (specify)		Public	
Molecular formula		Public	
CAS number		Public	

In **Figure 13**, the applicant has submitted a request for registration of a protein for Orphan Designation/Paediatric Investigation Plan and has provided the proposed INN, common name, company code and another name, all set as public. In this context, the company code and the proposed INN cannot be registered as the preferred term and the common name will be registered accordingly as the preferred term, while all the other terms will be entered as public aliases.

Figure 13. Request for the registration of a Protein for Orphan Designation/Paediatric Investigation Plan



EUROPEAN MEDICINES AGENCY

REQUEST FOR REGISTRATION OF A NEW SUBSTANCE

Please provide at least one preferred term in English and as many aliases as known. Copy the table below if multiple substances are being requested.

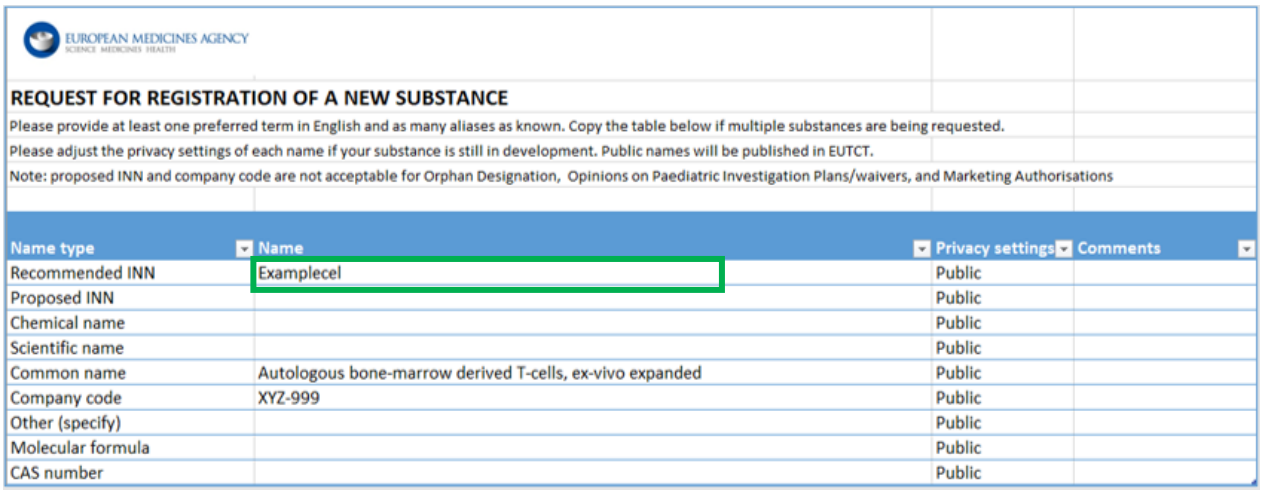
Please adjust the privacy settings of each name if your substance is still in development. Public names will be published in EUTCT.

Note: proposed INN and company code are not acceptable for Orphan Designation, Opinions on Paediatric Investigation Plans/waivers, and Marketing Authorisations

Name type	Name	Privacy settings	Comments
Recommended INN		Public	
Proposed INN	Examplemab	Public	
Chemical name		Public	
Scientific name		Public	
Common name	Human IgG1 monoclonal antibody against protein X	Public	
Company code	QWERTY-456	Public	
Other (specify)	ASD-456	Public	
Molecular formula		Public	
CAS number		Public	

In **Figure 14**, the applicant has submitted a request for registration of a cell therapy for Marketing Authorisation Application and has provided the recommended INN, common name and company code, all set as public. In this context, the recommended INN will be registered as the preferred term and all the other terms will be entered as public aliases.

Figure 14. Request for the registration of a Cell therapy for Marketing Authorisation Application



EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

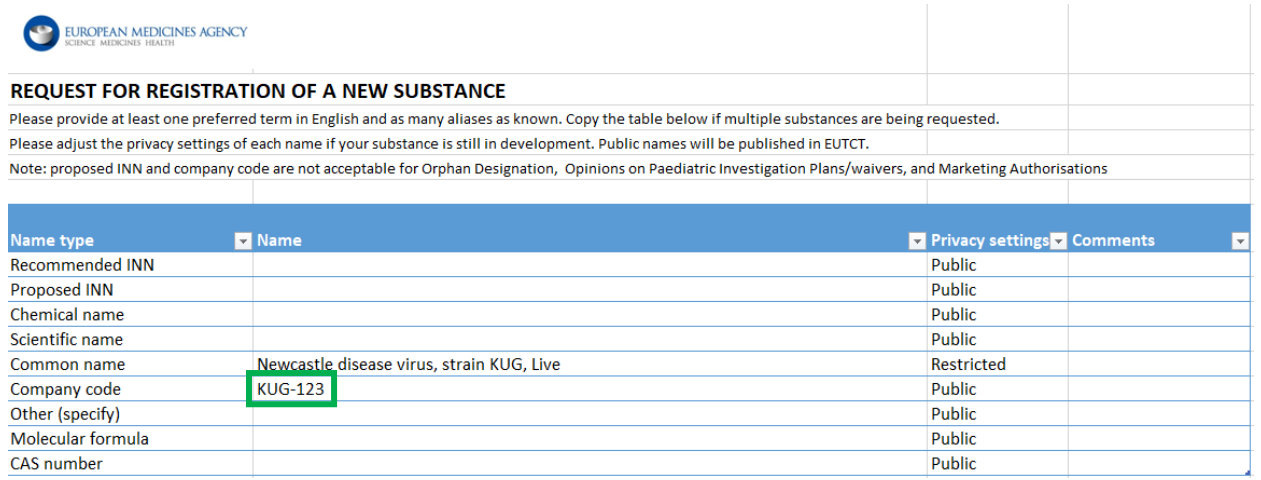
REQUEST FOR REGISTRATION OF A NEW SUBSTANCE

Please provide at least one preferred term in English and as many aliases as known. Copy the table below if multiple substances are being requested.
Please adjust the privacy settings of each name if your substance is still in development. Public names will be published in EUTCT.
Note: proposed INN and company code are not acceptable for Orphan Designation, Opinions on Paediatric Investigation Plans/waivers, and Marketing Authorisations

Name type	Name	Privacy settings	Comments
Recommended INN	Examplecel	Public	
Proposed INN		Public	
Chemical name		Public	
Scientific name		Public	
Common name	Autologous bone-marrow derived T-cells, ex-vivo expanded	Public	
Company code	XYZ-999	Public	
Other (specify)		Public	
Molecular formula		Public	
CAS number		Public	

In **Figure 15**, the applicant has submitted a request for registration of a veterinary vaccine for Scientific Advice and has provided the common name, set as restricted, and the company code, set as public. In this context, the company code will be registered as the preferred term and the common name will be registered accordingly as a restricted alias.

Figure 15. Request for the registration of a Veterinary vaccine for Scientific Advice



EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

REQUEST FOR REGISTRATION OF A NEW SUBSTANCE

Please provide at least one preferred term in English and as many aliases as known. Copy the table below if multiple substances are being requested.
Please adjust the privacy settings of each name if your substance is still in development. Public names will be published in EUTCT.
Note: proposed INN and company code are not acceptable for Orphan Designation, Opinions on Paediatric Investigation Plans/waivers, and Marketing Authorisations

Name type	Name	Privacy settings	Comments
Recommended INN		Public	
Proposed INN		Public	
Chemical name		Public	
Scientific name		Public	
Common name	Newcastle disease virus, strain KUG, Live	Restricted	
Company code	KUG-123	Public	
Other (specify)		Public	
Molecular formula		Public	
CAS number		Public	

For bulk requests concerning translations, an excel file containing the substance SMS IDs/EV codes and the concerned translations can be provided instead of the Substance Request Form.

3. Prepare the required supporting documentation

For all substance requests, supporting documentation describing the substance structure/nomenclature must be provided. The only exceptions are when an official name (e.g. INN, USAN) or a UNII is available. In these cases, reference to the official name/UNII is deemed sufficient.

The documents listed below, including the draft versions of such documents, are some examples of accepted documentation:

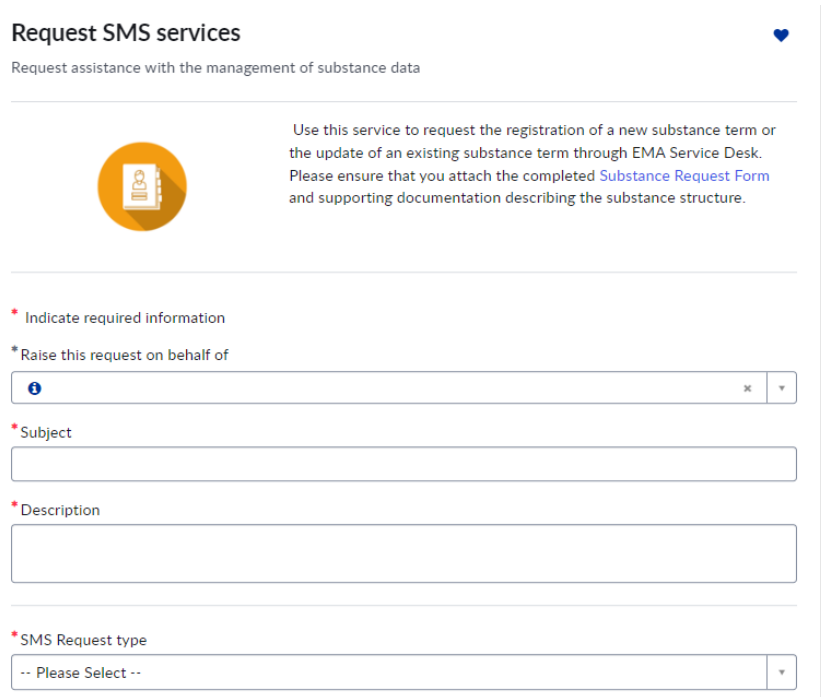
- Investigator's Brochure.
- Investigator's Medicinal Product Dossier.
- Summary of Product Characteristics.
- Package Leaflet.
- Section A.3. Medical Plausibility of Orphan Designation Application.

The provided supporting documentation will never be published and it will only be used in the context of the creation or update of a substance record. Documents are kept for at least 5 years, in accordance with the EMA Data Protection policy.

4. Submit the ticket via the EMA Service Desk

Access the form Request SMS Services in the EMA Service Desk (**Figure 16**) and fill out the mandatory fields, depending on the request type.

Figure 16. Request SMS Services form in the EMA Service Desk



For requests to create new substance record(s), the following fields are mandatory:

- **Subject**: enter a relevant subject (e.g. New substance(s) request).
- **Description**: provide a brief description of the requested substance(s).
- **SMS Request type**: select the most appropriate type that aligns with the regulatory procedure for the substance's subsequent use; please note that this will have an impact on the selection of the preferred term.
- **Attachments**: attach the filled-out Substance Request Form and appropriate supporting documentation.

For requests to update substance record(s) (e.g. to change the current preferred term, add one or more aliases, add one or more translations), the following fields should be filled-out:

- **Subject**: enter a relevant subject (e.g. Update substance(s) request).
- **Description**: provide the SMS ID(s) or EV Code(s) of the concerned substance record(s) and the new name(s)/translation(s);
- **SMS Request type**: select the most appropriate type that aligns with the regulatory procedure for the substance's subsequent use; please note that this will have an impact on the selection of the preferred term.
- **Attachments**: attach the filled-out Substance Request Form and appropriate supporting documentation.

5. Wait for the resolution of the request

After the request has been submitted, it will be further processed in accordance with the agreed Service-Level Agreements (SLAs), as described in **Table 11**. Applicants should therefore be mindful of the SLAs when planning a regulatory application submission.

Table 11. SMS change requests SLAs

Type of request	75% of requests	90% of requests	Expected outcome
Creation of substance, addition of an alias or change in preferred term (up to 20 terms/records)	Resolved within 5 working days	Resolved within 10 working days	The substance/alias is available for use, if the request is approved.
Addition of translation(s) (up to 20 terms/records)	Resolved within 10 working days	Resolved within 15 working days	The translation is available for use if the request is approved.
Creation of substance, addition of an alias, change in preferred term or addition of a translation(s) (over 20 terms/records)	No guaranteed SLA	No guaranteed SLA	The substances/aliases/translations are available for use if the request is approved.

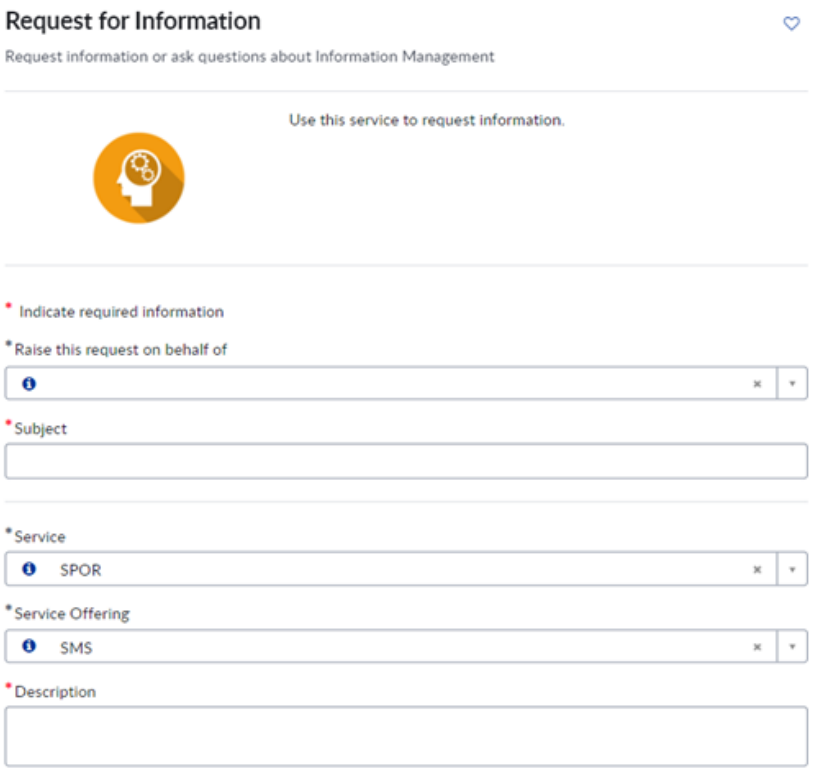
The requestor can withdraw the request at any time before the substance is registered in SMS. Once the substance is registered in SMS, it cannot be deleted, however, it can still be updated.

If additional clarifications or documentation is required during the request assessment, the SLA will be paused until the requestor replies. The requestor has 5 working days to provide a reply, otherwise the request will be closed. Any closed ticket can be reopened within 5 additional working days. If a new ticket is still required, reference to the original ticket should be provided.

7. Customer Service

Users can also submit substance related queries by accessing the Request for Information form in the EMA Service Desk (**Figure 17**) and setting the Service as "SPOR" and the Service Offering as "SMS".

Figure 17. Request for Information form in the EMA Service Desk



After the mandatory information is provided and the request for information concerning a substance related query has been submitted, it will be further processed in accordance with the agreed SLA, as described below, in **Table 12**.

Table 12. Request for Information SLAs

Type of request	75% of requests	90% of requests	Expected outcome
Request for information	Resolved within 22 working days	Resolved within 44 working days	The query is addressed

A list of frequently asked questions is available in *Annex 2: Frequently Asked Questions* to better support users with their queries on substances.

8. Data Stewardship

EMA SMS Data Stewards are continuously involved in data-centred activities, such as data mappings, data enrichments and data cleansing activities in cooperation with the SVG. These activities are essential for ensuring the quality, accuracy, and consistency of the data that is used by various stakeholders.

8.1. Data Mappings

Data mapping activities consist in matching substance data in EMA/NCA legacy systems to existing SMS substance data, resulting in the subsequent update or creation of relevant data in SMS. These exercises support data migrations and implementation of SMS in telematics systems.

The following substance data mappings have already been completed after SMS went live:

- Active and Withdrawn Orphan Designations for IRIS.
- Veterinary Scientific Advice for IRIS.
- Innovation Task Force for IRIS.
- PRIME for IRIS.
- Minor Use Minor Species (MUMS) for IRIS.
- EVVet substances for SMS.
- Veterinary active substances from 27 NCAs for UPD.
- xEVMPD development substances for SMS.
- Relevant Veterinary substances shared with European Food and Safety Authority (EFSA)
- Relevant Human substances shared with European Chemicals Agency (ECHA)

As an outcome of the substance mapping exercises, substance records might be created or updated (e.g. new aliases added), based on supporting documentation from the related regulatory procedures and information from the public domain.

New data mappings can be performed when needed to integrate a new system with SMS or to support new activities with other international organisations.

8.2. Data Enrichments

Data enrichment activities consist in proactively adding new and relevant information to existing SMS data, according to public reference information. These exercises improve data accuracy, reliability and value, ultimately increasing the data quality in SMS and, by facilitating substance identification and subsequently preventing the submission of change requests, they minimise the burden on industry, NCA and EMA users.

There are currently 8 data enrichment exercises regularly performed by the SMS team, listed in **Table *13*.** Data enrichment exercises

Table 13. Data enrichment exercises

Enrichment	Trigger	Frequency	Data fields impacted	Outcome
Proposed INN (pINN)	Publication of a new pINN list by WHO	Twice a year	• Substance names • External code: UNII	Substances created or updated
Recommended INN (rINN)	Publication of a new rINN list by WHO	Twice a year	• Data Classification • Substance names • External code: UNII	Substances created or updated
United States Adopted Name (USAN)	Publication of a new batch of USANs by USAN Council	Ad-hoc	• Substance names	Substances created or updated
Unique Ingredient Identifier (UNII)	Refresh of UNII exports by FDA	Ad-hoc	• External code: UNII	Substances updated
European Pharmacopoeia (Ph. Eur.)	Publication of a new Ph. Eur. supplement by the European Directorate for the Quality of Medicines & HealthCare (EDQM)	Three times a year	• Substance names	Substances created or updated
Custom Attribute: Critical Medicines	Publication of a new version of the Union List of Critical Medicines by the European Commission	Yearly	• Custom Attribute: Critical Medicine	Substances updated
Custom Attribute: Active substance in EEA human medicinal products	Internal process	Quarterly	• Custom Attribute: Active substance in valid EEA human AMP	Substances updated
Custom Attribute: Excipient/Adjuvant in EEA human medicinal products	Internal process	Quarterly	• Custom Attribute: Excipient/Adjuvant in valid EEA human AMP	Substances updated

8.3. Data Cleansing

Data cleansing activities consist of proactively reviewing SMS data to identify and eliminate duplicates, as well as to verify and to standardise SMS data against established public reference information.

Data cleansing improves the data quality and, by preventing the submission of change requests, minimises the burden on industry users.

The Substance Validation Group (SVG) is a European group of experts involved in the cleansing of the legacy substance data, as prioritised by the Network. As of March 2025, the following substance types/groups have now been cleansed:

- Veterinary active substances used in veterinary authorised products (all substance types)
- Chemicals
- Proteins
    - Proteins used in approved medicinal products
    - All insulins (Proteins and Specified Substance Group 1)
- Polymers most used in approved medicinal products
- Proprietary mixtures of excipients from Colorcon® (Specified Substance Group 1)
- Solutions (Specified Substances Group 1)
- Substances with pharmacopoeia references (Specified Substances Group 3)

An overview of the cleansing status of all substance types, as of March 2025, can be found in the **Table *14***.

Table 14. Cleansing status by substance type, as of March 2025

Substance type	% cleansed	Comments
Chemical	100%	
Specified Substance Group 3	95%	
Structurally Diverse - Vaccine	54%	Veterinary vaccines cleansed, Human vaccines ongoing
Polymer	34%	
Mixture	23%	
Structurally Diverse - Herbal	20%	
Specified Substance Group 1	16%	Includes flavours, proprietary mixtures of excipients, solutions, homeopathics, herbal extracts
Structurally Diverse - Plasma derived	12%	
Protein	11%	
Structurally Diverse - Other	10%	Includes gene therapy and other organism parts
Nucleic acid	4%	
Structurally Diverse - Allergen	1%	
Structurally Diverse - Cell therapy	0%	
Overall	**46%**	

This represents 100% of active substances used in veterinary authorised medicinal products and 79% of substances used in human authorised medicinal products.

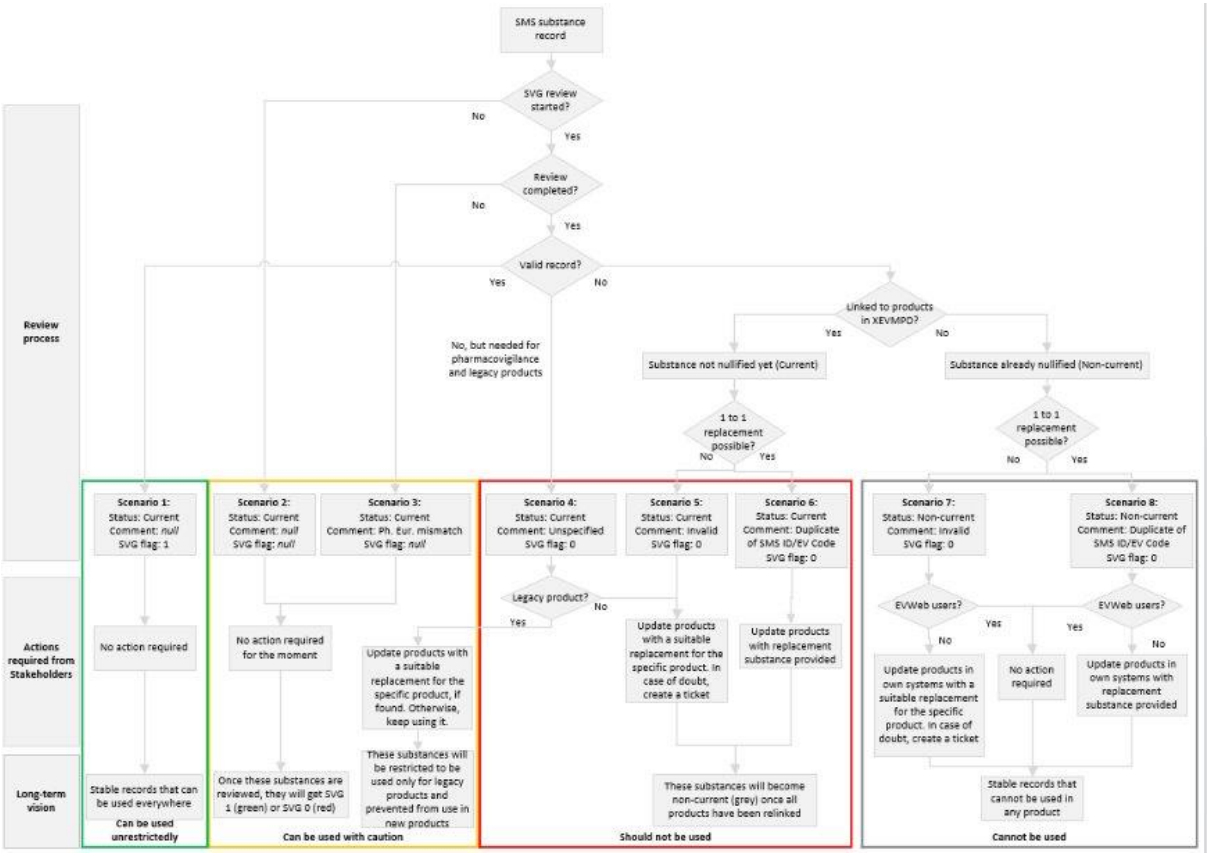
The most relevant data fields to highlight the substance cleansing status are described in **Table 15**:

Table 15. SMS data fields relevant for cleansing outcomes

Data field	Value	Meaning
Substance Status	Current	Record that can be used
	Non-current	Record that cannot be used
Comments	Ph. Eur. mismatch	Record cannot be cleansed due to a mismatch in Ph. Eur.
	Unspecified	Record is very general and should be avoided
	Duplicate of *SMS ID*/EV *Code*	Record is a duplicate of another record and should not be used
	Invalid	Record is not a substance
SVG flag	1	Record deemed valid by SVG
	0	Record deemed not valid by SVG
	Null	Record not yet reviewed by SVG
Substance (Current)	SMS ID of the replacement substance	Only added in case of Duplicates

Depending on the substance cleansing status and the respective products linked, there are currently 8 identified scenarios, with different combinations of the data fields above. The different scenarios expected actions from stakeholders and long-term vision are illustrated in **Figure *18.*** Request for Information form in the EMA Service Desk and further explained below.

Figure 18. Request for Information form in the EMA Service Desk



Scenario 1

These are substance records that have been fully reviewed and were deemed valid. These records have status "Current", SVG flag 1 and no comments nor replacement substance. These records are not expected to have major changes in the future, with the exception of regular updates as part of the substance lifecycle (e.g. addition of an INN, an UNII, translations, etc.). They do not require any action from stakeholders and can be used without restrictions in any products or regulatory procedures.

Scenario 2

These are substance records that have not yet been reviewed. They have status "Current" and no comments, SVG flag or replacement substance. These substances will be reviewed in the future taking into account the priorities defined by the Network. Once reviewed, they will then become a scenario 1 if deemed valid or any other scenario if deemed not valid. For the moment no action is required on them from stakeholders, and they can be used with caution in products or regulatory procedures. However, the status/SVG flag might change in the future and an action might then become required.

Scenario 3

These are substance records that the review has started but could not be completed since a mismatch between the European Pharmacopoeia monograph title (usually pointing to an anhydrous substance) and definition (usually pointing to a hydrated substance) was detected, which makes the substance record ambiguous. They have status "Current", Comment "Ph. Eur. mismatch" but no SVG flag or replacement substance. There is ongoing discussion between the Substance Validation Group and EDQM to review these monographs. For the moment, no action is required from stakeholders, and they can be used with caution in products or regulatory procedures. Once the monographs are reviewed, these substances will be fully cleansed and will follow the same outcome as from scenario 2.

Scenario 4

These are substance records that have been fully reviewed and were deemed not valid, since they refer to very general substances. However, these records are still needed to support pharmacovigilance activities and to be used in legacy products (e.g. veterinary products that were approved without a strain in the SmPC). These records have status "Current", SVG flag 0, Comment "Unspecified" but no replacement substance. When possible, stakeholders should try to refer to a more specific substance in their legacy products. However, if that is not possible, these records can still be used.

For new products these records should not be used, and a more specific substance should be used instead. For these cases, if no suitable replacement is found, then users should create a ticket in EMA Service Desk asking for advice. In the future, it is expected that these substances will be technically blocked from being used in new products and their use will be restricted to only legacy products (i.e. already approved with these Unspecified substances).

Scenario 5

These are substance records that have been fully reviewed and were deemed not valid, since they are invalid according to the SMS business rules. However, since they are linked to product records in XEVMPD, they cannot yet be nullified. They have status "Current", SVG flag 0, Comment "Invalid" but the data field "Substance (Current)" is null, since the replacement SMS ID will depend on a case-by-case situation. Stakeholders should not use these records in any products or regulatory products and, instead, refer to a suitable replacement substance for their specific products. If no suitable replacement is found, then users should create a ticket in EMA Service Desk asking for advice. Once all products in XEVMPD have been relinked, these substances will be made non-current, and they will no longer be able to be used (i.e. scenario 7).

Scenario 6

These are substance records that have been fully reviewed and were deemed not valid, since they are duplicates of another substance. However, since they are linked to product records in XEVMPD, they cannot yet be nullified. They have status "Current", SVG flag 0, Comment "Duplicate of *SMS ID of replacement substance */*EV Code of replacement substance*" and the data field "Substance (Current)" is also populated with the SMS ID of replacement substance. Stakeholders should not use these records in any products or regulatory products and, instead, refer to the replacement substance provided in the "Comments" and "Substance (Current)" data fields. Once all products in XEVMPD have been relinked, these substances will be made non-current, and they will no longer be able to be used (i.e. scenario 8).

Scenario 7

These are substance records that have been fully reviewed and were deemed not valid, since they are invalid according to the SMS business rules. They were not linked to any product in XEVMPD, so these substances were already nullified. They now have status "Non-current", no SVG flag, Comment "Invalid" but the data field "Substance (Current)" is null, since the replacement SMS ID will depend on a case-by-case situation. These substances can no longer be used in any product or regulatory procedures. Non-EVWeb (i.e. RIM or UPD) users should ensure that their own systems no longer refer to these substances and, instead, refer to a suitable replacement substance for their specific products. If no suitable replacement is found, then users should create a ticket in EMA Service Desk asking for advice. The non-current status is permanent and cannot be reverted back to current.

Scenario 8

These are substance records that have been fully reviewed and were deemed not valid, since they are duplicates of another substance. They were not linked to any product in XEVMPD, so these substances were already nullified. They now have status "Non-current", no SVG flag Comment "Duplicate of *SMS ID of replacement substance */*EV Code of replacement substance*" and the data field "Substance (Current)" is also populated with the SMS ID of replacement substance. These substances can no longer be used in any product or regulatory procedures. Non-EVWeb (i.e. RIM or UPD) users should ensure that their own systems no longer refer to these substances and, instead, refer to the respective replacement substance. The non-current status is permanent and cannot be reverted back to current.

8.3.1. Cleansing follow-up actions

8.3.1.1. SVG flag 0 and comment "Duplicate"

Known duplicated substance records have SVG flag 0, so they are no longer available in eAF and UPD (for new products). However, they are still available in XEVMPD and can technically be used in the creation or update of product records. It is not technically possible to flag these substances in XEVMPD in the product view, so it is not easy for MAH to identify them and correct them. In order to minimise the burden on MAHs, EMA is preparing a data fix to relink the impacted products in XEVMPD/PMS during 2025. Before the data fix, EMA will publish an export of impacted product and substance EV codes. EMA will then nullify the duplicated substances, and they will no longer be usable in any consuming system. MAH that are Gateway users should update their systems with the replacement substances provided ahead of the data fix to prevent negative acknowledgements. If desired, MAH can optionally update their products in XEVMPD before the data fix.

8.3.1.2. Pending changes in Preferred Terms

The SVG has identified substances that have an incorrect PT in SMS. The substance name that should be the preferred term is highlighted with the name source "Substance Validation Group". However, changing PTs has an impact on ICSR recoding, so these updates have not been implemented yet. There is a technical development ongoing to prevent disruptions in the ICSR recoding. After this is completed, the respective substance preferred terms will be updated in SMS.

8.3.1.3. SVG flag 0 and comment "Invalid" or "Unspecified"

Known "Invalid" or "Unspecified" substance records have SVG flag 0, so they are no longer available in eAF and UPD (for new products). However, they are still available in XEVMPD and can technically be used in the creation or update of product records. These are complex scenarios that require a product-by-product analysis and will be only addressed in late 2025, after the duplicates are addressed.

8.4. Data Quality monitoring

Data quality encompasses the overall condition and usability of data assets within an organisation. The following data quality dimensions are monitored and maintained in SMS:

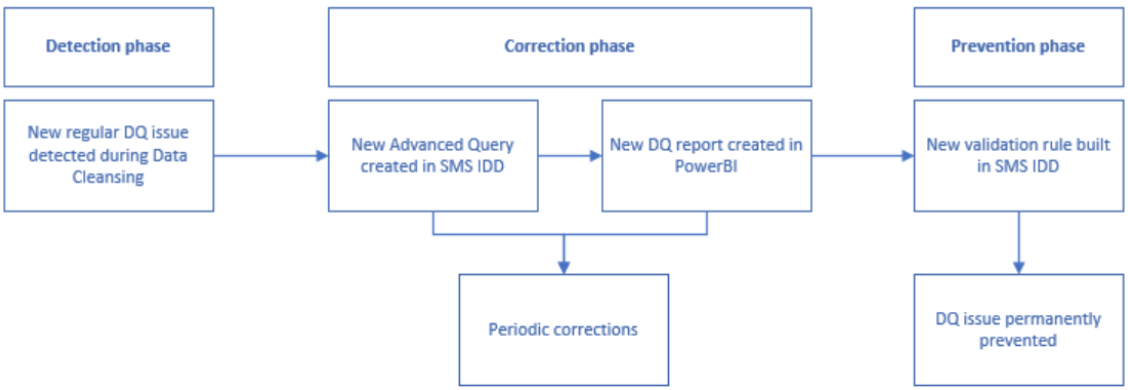
- Uniqueness
- Accuracy
- Completeness
- Consistency
- Conformity
- Currency

Uniqueness refers to the quality of data being distinct and devoid of duplicates. In SMS, each record should represent a unique substance. Accuracy denotes the correctness of data. In SMS, each record should reflect the actual values the substance it represents. Completeness measures the extent to which all required data elements are present in a dataset. In SMS, each record should have all the required values populated, in accordance with the business rules of each substance type and domain. Consistency refers to the coherence and uniformity of data across different data fields. In SMS, it means that values from different sections are consistent with each other. Conformity relates to the adherence of data to predefined standards. In SMS, it means that the concerned data fields are populated according to the defined business rules. Currency pertains to the timeliness and relevance of data in relation to the present moment. In SMS, it means that the concerned data fields are still up to date.

In order to prevent data quality issues, SMS IDD has built-in validation rules covering the data quality dimensions mentioned above. In addition, there are data quality profiles available in SMS IDD and in PowerBI to monitor quality issues that can't be prevented (yet) by validation rules. There are then defined processes to correct the identified issues in a timely manner.

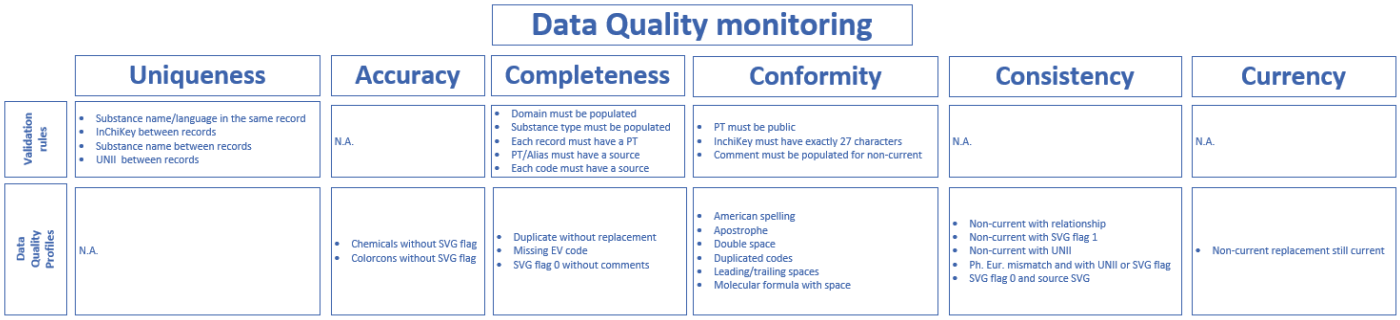
As illustrated in **Figure *19.*** Development of data quality profiles and validation rules, when regular DQ issues are identified during data cleansing exercises, a new advanced query is developed in SMS IDD and a corrective process is developed and implemented on a daily, weekly or monthly basis. Later, the same report will be built in PowerBI to allow an easier monitoring and correction. Finally, if possible, a new validation rule is built in SMS IDD directly to prevent the data quality issue from happening again.

Figure 19. Development of data quality profiles and validation rules



In **Figure *20.*** Validation rules and Data Quality Profiles it is possible to see the current validation rules in SMS and the data quality profiles. When a new issue is detected in a data quality profile, it tiggers a corrective action in SMS to address it.

Figure 20. Validation rules and Data Quality Profiles



9. Document versions

A major version is considered when as a new chapter is published or an existing chapter is radically changed, while a minor version is considered when an already existing chapter is updated.

Version	Date	Main changes
3.0	3 April 2025	Chapter 8.2 - Data enrichments revised Annex 1 – Business rules published
2.0	25 November 2024	Chapter 4.3 - Application Programming Interface revised Chapter 8.3 - Data Cleansing revised Chapter 8.4 - Data quality monitoring published Annex 5 - Data fields to FHIR mappings published
1.0	1 December 2023	First version

Annex 1: Business rules per substance type

1. General naming rules

A substance record may reference different name types which, based on concerned regulatory procedure, will impact the selection of the preferred term and influence the confidentiality selection:

- **Systematic (chemical) names** follow a set of rules that are defined by an international body, such as the International Union of Pure and Applied Chemistry (IUPAC). Systematic names are based on the structure and composition of the chemical, and they uniquely identify one chemical.

  This name type can be either public or restricted and can therefore be set either as the preferred term or as an alias.

- **Common names** are names that are widely used, but do not follow the systematic naming convention.

  This name type can be either public or restricted and can therefore be set either as the preferred term or as an alias.

- **Official names** are names that are referenced in an official public source (e.g. Ph. Eur., INN, USAN, JAN).

  This name type will always be public and can therefore be set as the preferred term (for recommended INNs) or as an alias (for proposed INNs, USAN, etc.).

- **Company or laboratory codes** are combinations of letters and numbers given by companies developing a substance and are used to identify the substance while it is in development. These types of codes might have prefixes that correspond to the company name.

The Preferred Term of a substance should be selected according to the priority ranking of the following reference sources and name types:

1. Pharmacopoeia Europea name[4]
2. (Modified) Recommended International Non-Proprietary Name
3. Other official name type with EU/EEA jurisdiction (e.g. INCI)
4. Common name mentioned in the SmPC or PiL
5. Systematic/scientific name
6. Company code

In addition to the sources/name types used for preferred terms, the following sources can also be used for aliases:

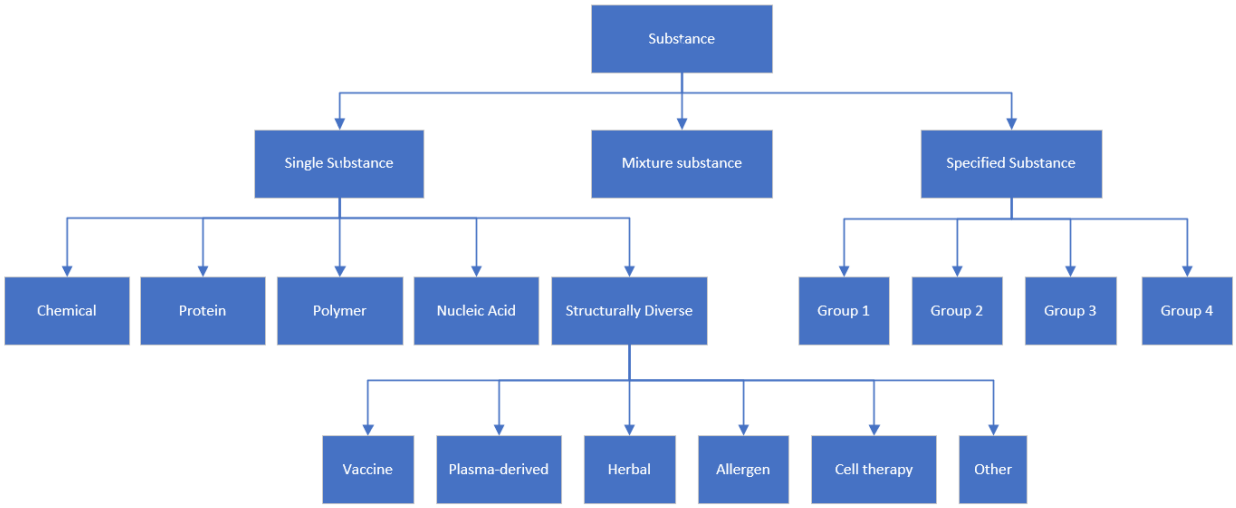
- Proposed INN
- United States Approved Name
- United States Pharmacopoeia
- Japanese Approved Name
- Official name in other jurisdictions (e.g. BAN, AAN)

[4] Some older Ph. Eur. monographs have a mismatch between the name and its definition, leading to ambiguous substance records. These situations are exceptions and are between discussed between the SVG and EDQM. Further information is provided in chapter 8.3 Data Cleansing – Section 3.

2. Substance type-specific naming rules

The substance type is defined by data stewards, and it is only used for data management purposes in SMS and has no impact in any regulatory procedure. Like mentioned in chapter 2, the substance types in SMS are a controlled vocabulary list from RMS and are listed in **Figure 21.** Substance type values in SMS.

Figure 21. Substance type values in SMS



The business rules described below have been reviewed by the SVG and are considered stable. However, minor adjustments are still possible.

2.1. Chemicals

Chemicals are single substances with a discrete molecular structure.

- Systematic name must always be provided, unless an official name or UNII or CAS is provided
- Molecular formula, molecular weight and InChIKey will be generated based on systematic name
- If a salt is requested, the free base (active moiety) will also be registered for the purposes of grouping/hierarchy, if missing. If only a company code is provided, then ABC-123 will be used for the free base and ABC-123 sodium will be used for the salt. If different company codes are provided, then they will be used accordingly. Similarly, when a radioactive substance is requested, the unlabelled substances will also be registered, if missing.
- UNII will be added as an external code, if available at the time of registration

2.2. Veterinary vaccines

The following data elements must be provided:

- Microorganism
- Serotype/Serogroup/Serovar (if applicable to the concerned microorganism)
- Strain
- Status (Live or Inactivated) or Antigen

Some examples are listed below:
300000040441 - Bovine viral diarrhoea virus, strain BK-1/B-1, Live
300000040172 - Infectious bursal disease virus, strain 89/03, Inactivated

Exceptions for veterinary vaccines without strain:

- Legacy veterinary vaccines for which a product was approved without a strain mentioned, are kept without strain, but with SVG flag 0 to prevent their use in new products
- For the purposes of Scientific Advice, it is also possible to register new veterinary vaccines without strain, if not yet defined at that stage. However, the strain must be added, before a marketing authorisation application can be provided

2.3. Specified Substances Group 1 (SSG1)

SSG1 include several different groups of substances, which are listed below:

- Proprietary mixtures of excipients
- Coatings
- Flavours
- Substances with certain physical properties defined in Ph. Eur
- Solutions or dispersions with a defined concentration defined in Ph. Eur.
- Homoeopathic substances; as they define dilutions
- Veterinary vaccines with different antigens from different strains
- Herbal extracts
- Isophane insulins

2.3.1. Proprietary mixtures of excipients

The most common manufacturer of propriety mixtures of excipients is Colorcon®. The most common excipients from this manufacturer are listed below:

· Opadry®
· Opalux®
· Opacode®
· Opaspray®
· Opatint®
· Pigment blend®
· Lake blend®
· Acryl-EZE®

The Colorcon® substance naming convention is as follows:
type *subtype* *code* *colour*
Example: Opadry II Y-30-18037 white

The business rules described below have not yet been reviewed by the SVG and are subject to considerable changes in the future. However, they can serve as guidance in the interim.

2.4. Proteins

Proteins are single substances with a structure based on a sequence of amino acids. Small proteins (peptides) can also be scientifically described as a chemical. According to the SMS business rules, any peptide with more than 3 amino acids will be registered as a protein. Proteins modified with a polymer or a chemical will also be registered as a protein.

2.4.1. Peptide

Peptides can be registered as proteins or chemicals. Peptide with three or less amino acids are to be considered chemicals, while peptides with four or more are to be considered proteins.

- Amino acid sequence must be provided

2.4.2. Antibody

The following data elements must be provided:

- Source of the immunoglobulin (e.g. Human, Humanised, etc.)
- Immunoglobulin type (e.g. IgG1, IgG4, etc.)
- Immunoglobulin modifications (if applicable)
- Target(s)

In case of antibody-drug conjugates, information on the payload (usually a chemical) must also be provided. The payload might also be registered as a separate substance record, taking into account the business rules for the concerned substance type.

2.4.3. Fusion protein

- Individual protein components must be provided

2.5. Polymers

Polymers are substances with a multiple repeating unit. Substances composed of a chemical covalently linked to a polymer will be registered as a polymer.

2.6. Nucleic acid

Nucleic acids are single substances with a structure based on a sequence of nucleotides.

2.7. Mixture

Mixtures are combinations of related single substances that are extracted/synthesised together without defined proportions.

2.8. Structurally Diverse – Vaccine

These are substance records with multiple unknown molecular entities, that are used to stimulate the body's immune response against one or more microorganisms. Vaccines based on a protein antigen are also registered as Structurally Diverse – Vaccine.

2.9. Structurally Diverse - Plasma derived

These are substance records with multiple unknown molecular entities that are derived from plasma. It mainly includes immunoglobulins for passive immunity.

2.10. Structurally Diverse - Herbal

These are substance records with multiple unknown molecular entities that are derived from a plant (e.g. root, herb, flower, etc.). Herbal extracts are not included here and, instead they are registered as Specified Substance Group 1.

2.11. Structurally Diverse - Cell therapy

These are substance records with multiple unknown molecular entities, composed of multiple cells.

The following data elements must be provided:

- Organism source (e.g. Autologous, Allogenic, etc.)
- Tissue source (e.g. adipose-derived)
- Cells
- Relevant modifications

2.12. Structurally Diverse - Other

All remaining structurally diverse substances (i.e. that do not fit any of the types listed above). For the moment this also includes gene therapy substance but, in the future they will have a separate substance type.

The following data elements must be provided:

- Microorganism
- Serotype (if applicable)
- Gene encoded

2.13. Specified Substances Group 2 (SSG2):

SSG2 substances are not currently in use in SMS.

2.14. Specified Substances Group 3 (SSG3):

SSG3 substances are not currently in use in SMS.

2.15. Specified Substances Group 4 (SSG4):

SSG4 substances are not currently in use in SMS.

Annex 2: Frequently Asked Questions

I didn't request the registration of my substance, but it is already available. Who has registered it?

Substances can be requested by any user. The concerned substance might have been requested by a partner company or a Contract Research Organisation (CRO) acting on behalf of a Sponsor. The substance might have also been reported in an ICSR and was registered based on publicly available information. Additionally, the substance might have been registered as part of a data mapping based on a legacy regulatory procedure or a data enrichment exercise based on INN, USAN or FDA public data.

Why is a substance registered as "approved" if it is still in clinical development?

Since SMS went live in July 2019, all substances have been registered by EMA SMS data stewards with the authorisation state "Authorised"; this substance data is subsequently made available in the xEVMPD list of "Approved" substances. This has no relationship with the approval status of the respective medicinal product.

Can I delete my substance from SMS?

The requestor can withdraw the request for the creation of a substance record at any time before the substance is registered in SMS. Once the substance is registered in SMS, it cannot be deleted, however, it can still be updated. Substances are kept in SMS even if development has stopped, since the substance can be linked to ICSRs/SUSARs or be used in the future by a different sponsor.

Why has the preferred term been published as public if I set all the data as restricted?

The substance preferred term is always registered as "public". This is a technical validation rule in SMS, and it ensures that the requestor/sponsor can find its substance.

Why can't the company code be registered as the preferred term?

Company codes cannot be used as the substance preferred term for the purposes of Orphan Designation, Paediatric Investigation plan. This is a requirement from the European Commission. Additionally, company codes can also not be used as the preferred term for Marketing Authorisation Application for Human medicinal products.

Why has my substance been published as "public" since I have requested it to be registered as "restricted"?

The substance has been registered as "public" since it is considered to be in the public domain, most likely due to the fact that an official name (e.g. INN, USAN) has been published.

Why is the substance unavailable after requesting its registration?

Depending on the concerned system, the SMS substance data is made available in different systems after a determined period of time, as described in **Table 10**. If the referenced period has passed and the substance is still unavailable in the concerned system, then the user shall contact the EMA via the EMA Service Desk.

Why has my substance been updated as "public" after its initial registration as "restricted"?

The substance has been initially registered as "restricted" according to the information provided by the applicant and information available in the public domain. As part of regular data enrichments (see section 8.2), substance records are reviewed periodically and might be updated based on new information made available in the public domain. Therefore, even if a substance record is initially registered as "restricted" it might be later updated and made "public", once new information is published, or the substance record is required for another regulatory procedure which requires an update to be performed.

Annex 3: Glossary

Acronym	Name
API	Application Programming Interface
Art. 57	Article 57 of Regulation (EU) 726/2004, which requires marketing authorisation holders to electronically submit to the Agency information on all medicinal products for human use authorised in the EU
CAP	Centrally Authorised Product
CR	Change Request
CTIS	Clinical Trials Information System
DMP	Development Medicinal Product
DCP	De-centralised Procedure
DQ	Data Quality
eAF	Electronic Application Form
EU-SRS	European Substance Reference System
EDQM	European Directorate for the Quality of Medicines & HealthCare
EUTCT	European Union Telematics Controlled Terms
FHIR	Fast Healthcare Interoperability Resources
ICSR	Individual Case Safety Report
IDMP	The ISO IDMP standards specify the use of standardised definitions for the identification and description of medicinal products for human use
INN	International Nonproprietary Names
InChIKey	International Chemical Identifier in "key" format
IRIS	A secure online platform for handling product-related scientific and regulatory procedures with EMA (iris.ema.europa.eu)
KUG	Key User Group
KPI	Key Performance Indicator
MAA	Marketing Authorisation Application
MAH	Marketing Authorisation Holder
MRP	Mutual Recognition Procedure
NAP	Nationally Authorised Product
NCA	National Competent Authority
OD	Orphan Designation
OMS	Organisation Management Service
Ph. Eur.	European Pharmacopoeia
pINN	Proposed International Non-proprietary Name
PIP	Paediatric Investigation Plan
PMS	Product (Data) Management Service
PT	Preferred Term
rINN	Recommended International Non-proprietary Name
RMS	Referential Management Service
SIAMED	An Information System for the management of regulatory procedure for centrally authorised products
SLA	Service Level Agreement
SPOR	Substance, Product, Organisation and Referential
SmPC	Summary of product characteristics
SMS	Substance Management Service
SVG	Substance Validation Group
UNII	Unique Ingredient Identifier
USAN	United States Adopted Name
WHO	World Health Organisation
xEVMPD	eXtended EudraVigilance Medicinal Product Dictionary

Annex 4: RMS Lists/IDs used in SMS

RMS List ID	RMS List name	RMS terms used	SMS Data field
100000000004	Domain	100000000012 - Human use 100000000013 - Veterinary use	Domain
200000004983	Data Classification	200000004985 - Public 200000004986 - Restricted	Data Classification (Substance) Data Classification (Name)
200000004906	Substance Authorisation Status	200000004972 - Authorised 200000004975 - Development	Authorisation State
200000005003	Record Status	200000005004 - Current 200000005006 - Non-current	Substance Status Name Status
100000075826	Substance Type	100000075670 - Chemical 200000005023 - Mixture 200000005035 - Nucleic acid 200000005022 - Polymer 200000005020 - Protein 200000005031 - Specified Substance Group 1 200000005032 - Specified Substance Group 2 200000005033 - Specified Substance Group 3 200000005034 - Specified Substance Group 4 200000005026 - Structurally Diverse - Allergen 200000005029 - Structurally Diverse - Cell therapy 200000005025 - Structurally Diverse - Herbal 200000005030 - Structurally Diverse - Other 200000005024 - Structurally Diverse - Plasma derived 200000005027 - Structurally Diverse - Vaccine	Substance Type
100000072057	Language	100000072142 - Bulgarian 100000072258 - Croatian 100000072167 - Czech 100000072168 - Danish 100000072169 - Dutch 100000072147 - English 100000072172 - Estonian 100000072149 - Finnish 100000072175 - French 100000072178 - German 100000072181 - Greek, Modern (1453-) 100000072187 - Hungarian 100000072156 - Icelandic 100000072179 - Irish 100000072194 - Italian 100000072226 - Latin 100000072205 - Latvian 100000072206 - Lithuanian 100000072236 - Maltese 100000072243 - Norwegian 100000072217 - Polish 100000072251 - Portuguese 100000072254 - Romanian 100000072259 - Slovak 100000072260 - Slovenian 100000072264 - Spanish 100000072288 - Swedish	Language

100000000009	Source of Information	Substance names (All with attribute "SMS Name"): 200000025184 - ADISINSIGHT 200000032270 - Anthroposophic Pharmaceutical Codex 100000075734 - BAN 100000115264 - BRITISH HERBAL PHARMACOPOEIA 100000075723 - BRITISH PHARMACOPOEIA 100000152989 - CHEBI 100000075787 - CHEMICAL ABSTRACT SERVICE 100000125793 - CHEMICALBOOK 100000124125 - CHEMIDPLUS 200000005827 - CHINA FOOD AND DRUG ADMINISTRATION 100000125797 - CHINESE PHARMACOPOEIA 100000133971 - COLORCON 100000125798 - COMMUNITY HERBAL MONOGRAPHS 100000125799 - COMMUNITY REGISTER OF THE EUROPEAN COMMISSION 100000075712 - COMPANY SPECIFICATION 100000170926 - COSING 100000167113 - CZECH PHARMACOPOEIA 100000075697 - DEUTSCHES ARZNEIBUCH 100000075824 - EMEA RECOMMENDATIONS FOR INFLUENZA VACCINATIONS PLANS 100000159429 - ENCYCLOPEDIA OF LIFE (EOL) 100000170925 - EUROPEAN CHEMICALS AGENCY (ECHA) 200000005829 - EUROPEAN FOOD SAFETY AUTHORITY (EFSA) 100000075790 - EUROPEAN PHARMACOPOEIA 100000134409 - FDA INACTIVE INGREDIENTS DATABASE 100000151869 - FDA SUBSTANCE REGISTRATION SYSTEM 100000075745 - FRENCH PHARMACOPOEIA 100000075820 - GLOBAL DIVERSITY INFORMATION FACILITY 100000134592 - GREEK PHARMACOPOEIA 100000144867 - HAB 100000075692 - HANDBOOK OF PHARMACEUTICAL EXCIPIENT 100000127906 - HOMEOPATHIC PHARMACOPOEIA OF THE UNITED STATES 200000005828 - INDEX FUNGORUM 100000075715 - INN 100000075699 - INNM 100000125815 - INTERNATIONAL CODE OF NOMENCLATURE OF BACTERIA (ICNB) 100000125816 - INTERNATIONAL CODE OF VIRUS CLASSIFICATION AND NOMENCLATURE 100000125813 - INTERNATIONAL CODE OF ZOOLOGICAL NOMENCLATURE (ICZN) 100000133377 - INTERNATIONAL NUMBERING SYSTEM (INS) FOR FOOD ADDITIVES 100000075807 - INVESTIGATOR'S BROCHURE 100000075748 - ITALIAN PHARMACOPOEIA 100000124117 - ITIS (INTEGRATED TAXONOMIC INFORMATION SYSTEM) 100000075688 - JAN 100000075690 - JAPANESE PHARMACOPOEIA 100000125807 - KEW GARDEN- IPLANTS DATABASE 100000124119 - MARTINDALE 100000133378 - PHARMACOPOEIA HELVETICA	Source (Name Source) Source (External Code)

		100000075797 - POLISH PHARMACOPOEIA 100000130608 - PORTUGUESE PHARMACOPOEIA 200000005826 - PROTEIN DATA BANK (PDB) 100000153275 - PUBCHEM 100000075742 - SPC 200000025197 - Substance Validation Group 100000075717 - THE MERCK INDEX 100000125810 - THE PLANT LIST 100000075736 - USAN 100000075738 - USP 100000075726 – WHO <u>External codes (All with attribute "SMS Code"):</u> 100000075787 - CHEMICAL ABSTRACT SERVICE 200000032418 - EC/List number 100000075665 - Extended EudraVigilance Medicinal Product Dictionary 100000075715 - INN 100000146035 - SIAMED - EMA CP management system 200000025197 - Substance Validation Group 200000018817 - Unique Ingredient Identifier 100000075736 - USAN	

Annex 5: Data fields to FHIR mappings

Section	Data field	FHIR Path
Substance	SMS ID	SubstanceDefinition.id
	Domain	SubstanceDefinition.domain
	Data classification (Substance)	SubstanceDefinition.extension url="https://ema.europa.eu/fhir/dataClassification"
	Authorisation State	N/A
	Substance Status	SubstanceDefinition.status
	Substance Name	SubstanceDefinition.name
	Comment	SubstanceDefinition.note
	Substance Type	SubstanceDefinition.type
	Molecular formula	SubstanceDefinition.structure.molecularFormula
	Molecular weight	SubstanceDefinition.structure.molecularWeight
	InChIKey	SubstanceDefinition.structure.representation.representation
	Last Update Date	SubstanceDefinition.meta.lastUpdated
Substance Name	Substance Name	SubstanceDefinition.name.name
	Language	SubstanceDefinition.name.language
	Is Preferred	SubstanceDefinition.name.preferred
	Name Status	SubstanceDefinition.name.status
	Data Classification (Name)	SubstanceDefinition.name.extension url="https://ema.europa.eu/fhir/dataClassification"
Substance Name Source	Source (Name Source)	SubstanceDefinition.name.source

Substance External Code	Source (External Code)	SubstanceDefinition.code.code.coding.system
	ID/Code Provided by Source	SubstanceDefinition.code.code.coding.code
Substance Custom Attribute	Attribute Term	SubstanceDefinition.property.code
	Attribute Value	SubstanceDefinition.property.amountString
Substance Relationship	Related substance	SubstanceDefinition.relationship.substanceDefinitionReference
	Substance Relationship Combination	SubstanceDefinition.relationship.type
Substance Current	Substance Current	SubstanceDefinition.extension url="https://ema.europa.eu/fhir/currentSubstance"