



UNOCOM

Project acronym: UNICOM

Project full title: Up-scaling the global univocal identification of medicines
in the context of Digital Single Market strategy

Call identifier: H2020-SC1-DTH-2019

Deliverable D2.2: Guidance document on how MPDs should map to substance data

What is the need of MPDs for substance data from EU-SRS/SMS and how to make this work

Version: 1.0
Status: Final
Dissemination Level¹: PU
Due date of deliverable: 30.11.2022
Actual submission date: 21.12.2022
Work Package: WP2: IDMP – Substance Management in Europe
Lead partner for this deliverable: CBG
Partner(s) contributing: IHD, Z-INDEX, WHO-UMC, IDMP1
Deliverable type²: R

Main author(s):

Annet Rozema	CBG
Anne Rixt Molema-Buursma	CBG

Other author(s):

Leonora van Dorp-Grandia	Z-INDEX
Ursula Tschorn	IDMP1
Julie James	IHD
Olof Lagerlund	WHO-UMC
Anja van Haren	CBG
Marcel Hoefnagel	CBG
Pedro Batista	EMA

¹ Dissemination level: PU: Public; CO: Confidential, only for members of the consortium (including the Commission Services); EU-RES: Classified Information: RESTREINT UE (Commission Decision 2005/444/EC); EU-CON: Classified Information: CONFIDENTIEL UE (Commission Decision 2005/444/EC); EU-SEC Classified Information: SECRET UE (Commission Decision 2005/444/EC)

² Type of the deliverable: R: Document, report; DEM: Demonstrator, pilot, prototype; DEC: Websites, patent fillings, videos, etc.; OTHER; ETHICS: Ethics requirement; ORDP: Open Research Data Pilot

Revision history

Version	Date	Changes made	Author(s)
0.1	24.11.2022	Initial release for team review	Annet Rozema, Anne Rixt Molema-Buursma
0.2	05.12.2022	Team review, comments processed	Full team contribution
0.3	06.12.2022	Changes based on discussions during team meeting	Full team contribution
0.4	13.12.2022	Release of final draft for team review	Full team contribution
1.0	21.12.2022	Finalization of the D2.2 report	Anne Rixt Molema-Buursma, Annet Rozema and various reviewers of the report, including alignment conclusions with Robert VanderStichele (WP8)

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Deliverable abstract

Medicinal Product Dictionaries (MPDs) for clinical care are facing difficulties to retrieve substance data needed to include in their dictionary. Data are not available in a structured way, or not accessible to the MPD organization. This results in inefficiencies and sometimes causes incorrect data in the MPDs. The team working on D2.2 investigated the need for substance data (fields) to feed MPDs, and analysed if and how EU-SRS and SMS could provide these data.

It was concluded that combining SMS data with EU-SRS data and release these combined data into the public domain could greatly benefit MPD organizations in maintaining substance data in their MPD.

Based on the analysis performed, the following steps are recommended:

- ▶ Investigate - together with the EMA SMS team, the possibility to publish a combined set of SMS and EU-SRS data in the public domain. Selection based on substances used in authorized medicinal products may be of benefit in preventing confidentiality issues. SMSID, names, substance category could be retrieved from SMS. Other fields such as additional identifiers, molecular weight and molecular formula could be retrieved from EU-SRS (with SMSID as linking identifier)
 - ▶ A staged approach could be an option, where quick wins can be achieved (e.g. publishing substance identifiers, molecular weight and molecular formula)
 - ▶ The need for access to substance-type specific data fields is clearly expressed, but is more complex to arrange and therefore requires further analysis. This refers to, for example: relationships, hierarchy, amino acid sequences and SSG1 information related to a substance.
 - ▶ Adding product-related information (# of products using the substance, role of substance in products) is also very valuable but may require more investigation on how to make this possible. This information is not captured in EU-SRS/SMS, but in UPD & PMS.
- ▶ Investigate how to map substance records of SNOMED with substances captured in the SPOR systems (EU-SRS/SMS).
- ▶ Investigate the possibility – similar to the approach taken by FDA – to implement a public version of EU-SRS, containing non-confidential substance information.

Keywords: MPD, IDMP, SPOR, SMS, EU-SRS, SNOMED, Substance Data

This document contains material, which is the copyright of the members of the UNICOM consortium listed above, and may not be reproduced or copied without their permission.

The commercial use of any information contained in this document may require a license from the owner of that information.

This document reflects only the views of the authors, and the European Commission is not liable for any use that may be made of its contents. The information in this document is provided “as is”, without warranty of any kind, and accept no liability for loss or damage suffered by any person using this information.

© 2019-2023. The participants of the UNICOM project.

TABLE OF CONTENTS

Revision history	2
Deliverable abstract.....	3
List of abbreviations.....	6
Executive summary	8
1 Introduction.....	9
1.1 Problem statement	9
1.2 MPDs.....	9
1.3 SMS and EU-SRS – data sources for substances.....	10
2 Analysis of required data fields	12
2.1 Data requirements from MPD organizations	12
2.2 General impressions.....	12
2.3 High-level substance information	13
2.3.1 High level substance information in SMS.....	13
2.3.2 High-level substance information in EU-SRS.....	14
2.3.3 Matching the needs of MPDs	14
2.4 Substance characteristics.....	15
2.4.1 Substance characteristics in SMS	16
2.4.2 Substance characteristics in EU-SRS	16
2.4.3 Matching the needs of MPDs	18
2.5 Required identifiers & references	18
2.5.1 Identifiers & references available in SMS.....	18
2.5.2 Identifiers and references available in EU-SRS	19
2.5.3 Matching the needs of MPDs	20
3 Conclusions	21
3.1 Importance of quality of substance data.....	21
3.2 Importance of accessibility to substance data fields	21
3.3 Improving efficiency	21
4 Recommendations	23
Appendix A: Example of inconsistencies in substance data	24
Cetirizine - SmPC	24
Cetirizine data - SMS.....	24
Cetirizine data - EU-SRS.....	25

LIST OF FIGURES

Figure 1. The name field of Folic acid in EU-SRS.....	14
Figure 2. Need for groupers for non-specific drug verbatims (source: Unicom deliverable D8.7: IDMP coding principles and guidance for ICSRs.).....	16
Figure 3. Record in EU-SRS - Overview	17
Figure 4. Example of the 'Chemical Structure' field in EU-SRS.....	17
Figure 5. Example of mixture components field in EU-SRS.....	18
Figure 6. Codes/identifiers included in EU-SRS for Folinic acid	19
Figure 7. SmPC extracts for two products using different descriptions for the same precise active ingredient substance.	24
Figure 8. Public SMS export data; search for "Cetirizine" in Substance_Name	24
Figure 9. Search for "Cetirizine" in (test) EU-SRS	25
Figure 10. The record for Cetirizine (di)hydrochloride in (test) EU-SRS.....	25

LIST OF TABLES

Table 1. Identifying elements per substance type, with courtesy of Tyler Peryea (FDA, G-SRS team)	11
Table 2. MPD organizations included in this research	12
Table 3. Fields included in public export of SMS data (CSV on SPOR portal) and example values....	13

List of abbreviations

Abbreviation	Complete form
BoSS	Basis of Strength Substance
CAS	Chemical Abstracts Service
CSV	Comma Separated Values
DE-SRS	Deutschland Substance Registration System
eCTD	electronic Common Technical Document
EDQM	European Directorate for the Quality of Medicines & HealthCare
EMA	European Medicines Agency
EPAR	European public assessment report
EU-SRS	European Substance Registration System
FDA	Food and Drug Administration
GSID	Global Substance Identifier
G-SRS	Global Substance Registration System
HALMED	Agency for Medicinal Products and Medical Devices of Croatia
ICSR	Individual Case Safety Report
IDMP	Identification of Medicinal Products
IMPD	Investigational Medicinal Products
INN	International Non-proprietary Name
ITIS	Integrated Taxonomic Information System
ISO	International Organization of Standardization
KEGG	Kyoto Encyclopedia of Genes and Genomes
MHRA	Medicines and Healthcare products Regulatory Agency
MPD	Medicinal Product Dictionary
MPNS	Medicinal Plant Names Services
MW	Molecular Weight
NCA	National Competent Authority
NCATS	National Center for Advancing Translational Sciences
NCBI	National Center for Biotechnology Information
NCI	National Cancer Institute

Abbreviation	Complete form
NoMA	Norwegian Medicines Agency
PAI	Precise Active Ingredient
PMS	Product Management Services
PhPID	Pharmaceutical Product Identifier
PT	Preferred Term
SmPC	Summary of Medicinal Product Characteristics
SMS	Substance Management Services
SMSID	Substance Management Services Identifier
SNOMED	Systematic Nomenclature of Medicine
SNOMED-CT	Systematic Nomenclature of Medicine Clinical Terms
SPOR	Substances, Products, Organisations & Referentials
SSG1	Specified Substance Group 1
SVG	Substance Validation Group
SVGID	Substance Validation Group Identifier
UMC-SRS	Uppsala Monitoring Centre Substance Registration System
UNII	Unique Ingredient Identifier
UPD	Union Product Database (for authorized veterinary medicines)
WHO	World Health Organization
ATC	Anatomical Therapeutic Chemical
VATC	Veterinary ATC
WP	Work Package
xEVMPD	Extended EudraVigilance Medicinal Product Dictionary

Executive summary

UNICOM WP2 aims to create an EU regulatory network wide accessible, structured database, referred to as EU substance registration system (EU-SRS), for the identification of substances in medicinal products based on their scientific properties in accordance with ISO IDMP standard 11238 and ISO implementation standard 19844.

The need for substance data to feed Medicinal Product Dictionaries for clinical care (referred to as MPDs), was investigated by analysing the required data fields and comparing these to the data available in SPOR Substance Management Services (SMS) and in EU-SRS. SMS was used as a starting point, since data from SMS are published daily into the public domain. Questionnaires were sent to a number of MPD organizations. The feedback provided input on which fields are required but at this moment not available in the public domain through SMS. These gaps were further analysed, by investigating how substance data from EU-SRS could meet the needs of MPDs where SMS cannot. Next to processing the feedback from the questionnaire, a discussion took place by web conferences with colleagues from the Irish National Health Product Database and VIDAL (France).

EU-SRS, however, is not available in the public domain. Therefore, during the analysis, discussions took place to what extent confidentiality would prevent the required data from EU-SRS to be published next to the SMS data.

Also, different options for extracting and sharing substance data with MPDs (or other stakeholders without access to EU-SRS) were discussed, resulting in advice on next steps.

Conclusions:

- ▶ Having access to high quality substance data is important to MPDs. This applies to both data in SMS and EU-SRS, as well as to the quality of substance information in the SmPC
- ▶ The substance data fields required in the public domain is more than what is currently offered through SMS. EU-SRS contains most of these additional data fields, but the system is not publicly accessible. A solution could be to publish (some of) the non-confidential data fields into the public domain.
- ▶ The feedback from MPD organizations showed the substantial amount of double work is performed; every MPD owner creates their substance records based on various data sources including SMS, SNOMED, national SmPC's and various literature sources and websites.

Based on the analysis performed, the following steps are recommended:

- ▶ Investigate - together with the EMA SMS team, the possibility to publish a combined set of SMS and EU-SRS data in the public domain. Selection based on substances used in authorized medicinal products may be of benefit in preventing confidentiality issues. SMSID, names, substance category could be retrieved from SMS. Other fields such as additional identifiers, molecular weight and molecular formula could be retrieved from EU-SRS (with SMSID as linking identifier)
 - ▶ A staged approach could be an option, where quick wins can be achieved (e.g. publishing substance identifiers, molecular weight and molecular formula)
 - ▶ The need for access to substance-type specific data fields is clearly expressed, but is more complex to arrange and therefore requires further analysis. This refers to, for example: relationships, hierarchy, amino acid sequences and SSG1 information related to a substance.
 - ▶ Adding product-related information (# of products using the substance, role of substance in products) is also very valuable but may require more investigation on how to make this possible. This information is not captured in EU-SRS/SMS, but in UPD & PMS.
- ▶ Investigate how to map substance records of SNOMED with substances captured in the SPOR systems (EU-SRS/SMS).
- ▶ Investigate the possibility – similar to the approach taken by FDA – to implement a public version of EU-SRS, containing non-confidential substance information.

1 Introduction

1.1 Problem statement

Substances are the core of medicinal products. High quality substance (master) data is therefore crucial. In WP2, the Substance Validation Group (SVG) works together with the colleagues from EMA to increase data quality of the substance data, by cleansing the available data in the SPOR Substances Management Services (SMS). Next to that, the team is implementing EU-SRS; EU-SRS can be seen as scientific backbone of SMS.

Some of the data fields from SMS are being published daily on the SPOR portal, in CSV format which can then be loaded into an Excel file. Data included in the export are high-level substance data fields, as well as a flag indicating the cleansing status of the record. Although it may be useful that the data are now publicly available, on various occasions (e.g. Unicom consortium meeting early September 2022), users of substance data have indicated that they need more detailed substance data to meet their needs. Medicinal Product Dictionary (MPD) organizations also expressed their need for more detailed substance data to feed their databases.

In absence of structured data in the public domain, often, the Summary of Product Characteristics (SmPC) is used as a source to determine which substances are present in a product for entering in the MPDs. For the characteristics of a substance, several other (literature) sources and the public substance database, published by NCATS (GSRs), is used as source.

Colleagues of WP2 have investigated the needs of MPDs in the area of substance data fields required to feed their MPD, with the purpose of increasing efficiency and quality of the maintenance of MPDs. The team investigated the following:

- ▶ The need for substance data to feed MPDs
- ▶ The availability of these data fields in SMS and/or EU-SRS
- ▶ Which fields are ready to be published into the public domain
- ▶ Which fields are not available or are not suitable to be published in the public domain
- ▶ Advice on next steps

1.2 MPDs

Medicinal Product Dictionaries are dictionaries widely used in different areas, but in this report we focus on the use of MPDs in clinical care where these dictionaries provide medicinal product information in a structured way. Examples of MPDs are the G-Standaard (NL), WHODrug Global, VIDAL, National Health Product Catalogue and RxNorm (used in the US). MPDs offer an overview (identification and accurate description) of medicinal products and enable an efficient exchange of information between healthcare parties. Healthcare professionals such as clinicians in hospitals or pharmacists, use MPDs in prescribing, dispensing and administration of medicines, for the purpose of pharmacovigilance and declaring invoices to healthcare insurers. Collecting and capturing of correct and complete substance information in sufficient granularity is a challenge. With the implementation of IDMP across Europe, enabled by the UNICOM program, the MPD organizations expect to be able to increase the efficiency of their processes to maintain the data in their MPD as well as improve the data quality in their MPD. In some countries the MPD is used as the source for electronic prescriptions.

For MPD organizations, to create and maintain product descriptions in their MPD, many sources are consulted. This is an inefficient process, with ambiguous information found, for example, in SmPCs. A high-quality data source with substance (and product) data is expected to greatly improve efficiency and quality of the data in the MPDs. Examples of ambiguous descriptions of a substance in an SmPC are:

Arikayce:

- ▶ Name: Arikayce liposomal 590mg nebuliser dispersion
- ▶ Section 2 of the SmPC contains the following text:
 - Each vial contains amikacin sulphate equivalent to 590 mg amikacin in a liposomal formulation. The mean delivered dose per vial is approximately 312 mg of amikacin.

Based on this text, it is not clear if the 590 mg refers to amikacin, or to amikacin liposomal. A liposomal formulation of a substance is a unique substance in itself. Encapsulation of substances alters the therapeutic activity of the substance.

Methotrexate 2.5 mg Tablets Advanz:

Substance strength may be expressed in different ways in two different SmPCs (Section 2):

- ▶ Each tablet contains methotrexate 2.5 mg as methotrexate sodium, or;
- ▶ Each tablet contains methotrexate sodium 2.74 mg (which is equivalent to 2.5 mg of methotrexate)

Are these two products exactly equivalent in terms of their Precise Active Ingredient (PAI) and Basis of Strength Substance (BoSS) or only equivalent in terms of their BoSS?

1.3 SMS and EU-SRS – data sources for substances

Often, the Summary of Product Characteristics (SmPC) is used to extract substance data for entering in the MPDs. Since SmPCs are mostly available as text documents (PDFs), this means in practice that the SmPC is read, the relevant information extracted and entered into the MPD database. It cannot be avoided that sometimes the information in the SmPC needs to be interpreted, e.g. in case the salt/water is not precise, and especially also the excipients if the name is different from the name in the organisation's MPD; interpretation is time consuming and sensitive to errors. Next to the SmPC, the US public substance database, published by NCATS (GSRs), is regularly used as source, as well as SNOMED.

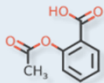
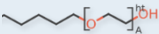
In Europe, substance data are stored in the SPOR database SMS (Substance Management Services), which is hosted and maintained by EMA. SMS contains mainly the substance identifier (SMSID) and names (Preferred Terms, Aliases, Translations). Also, there is need for increased data quality of the SMS data (source: [HMA approved implementation plan of EU-SRS \(deliverable 2.1 of UNICOM WP2\)](#)).

Prior to the start of UNICOM, SMS was considered not to be sufficient for MPDs to build their substance records in their database. Reasons for this were:

- ▶ Missing information in SMS; for example, information on relationships, molecular formula, molecular weight, and identifiers such as CAS numbers are not (yet) available in SMS. For each substance class, the same fields are captured in SMS, whereas different substance classes have different identifying elements, see **Table 1**.
- ▶ SMS (data) is not accessible in the public domain. During the course of the UNICOM program, EMA has started to publish on a daily basis the substance data from SMS, including a flag to indicate if the Substance Validation Group (SVG) has reviewed/confirmed the record.

The quality of the SMS data – prior to the start of cleansing work by the SVG – was considered not sufficient.

Table 1. Identifying elements per substance type, with courtesy of Tyler Peryea (FDA, G-SRS team)

Substance Type	Chemical	Polymer	Protein	Nucleic Acid	Structurally Diverse												
Defined by	Chemical Structure	Structural repeat unit(s)	Amino acid sequence(s)	Nucleobase sequence	Taxonomic information + part												
Example			<pre>>A35X00TA2K RCPGCGGVQAGCPGGCVVEE EDGGSPAEGCAEAEGLRRE GQECGVYTPNCAPGLQCHFP ...</pre>	<pre>>303159CVH9 TAAACGTTATAACGTT ATGACGTCAT</pre>	<table border="1"> <tr> <td>Organism Family</td> <td>CANNABACEAE</td> </tr> <tr> <td>Organism Genus</td> <td>CANNABIS</td> </tr> <tr> <td>Organism Species</td> <td>SATIVA</td> </tr> <tr> <td>Author</td> <td>L.</td> </tr> <tr> <td>Intraspecific Type</td> <td>SUBSPECIES</td> </tr> <tr> <td>Intraspecific Name</td> <td>SUBSP. SATIVA</td> </tr> </table>	Organism Family	CANNABACEAE	Organism Genus	CANNABIS	Organism Species	SATIVA	Author	L.	Intraspecific Type	SUBSPECIES	Intraspecific Name	SUBSP. SATIVA
Organism Family	CANNABACEAE																
Organism Genus	CANNABIS																
Organism Species	SATIVA																
Author	L.																
Intraspecific Type	SUBSPECIES																
Intraspecific Name	SUBSP. SATIVA																

As part of the UNICOM program, the team working in WP2 is cleansing the SMS substance data. Cleansing feedback is provided to the EMA SMS team for processing and implementing in SMS. On the other hand, EU-SRS is implemented for use by the European Regulatory Network. The system will become available for use at the end of January 2023, containing IDMP-based substance data, of high quality.

The system will be accessible to the EU Regulatory network, implying that:

- ▶ The system access will be setup through EudraNet, which means that MPD organizations outside of NCA's, will not have access to the database
- ▶ It is not envisaged to release a public version of EU-SRS in the near future; this will be further investigated in a later stage
- ▶ SMS could load certain data fields from EU-SRS. The SMS team has confirmed their intention to load CAS numbers, molecular formulas and possibly molecular weight data from EU-SRS into SMS in 2023. So far, only the UNII (US substance identifier) is loaded into SMS; the field has recently also been added to the public daily export of SMS data (CSV on SPOR portal).

2 Analysis of required data fields

2.1 Data requirements from MPD organizations

Input was gathered from organizations managing MPDs for clinical care to establish their requirements for substance-related data fields to include in their database.

A questionnaire was sent to various MPDs, containing the following questions:

- ▶ What is the name of the MPD you are working on?
- ▶ What is the main purpose of the MPD / What is the MPD consulted for?
- ▶ Which type of substance data do you use in your MPD? E.g. molecular structures, ingredient role. Please provide examples, that will be of much help to us. How are substance data currently collected (where from) and how (degree of granularity) are the data recorded in the MPD?
- ▶ Gap analysis: Which substance data are currently missing in the sources you consult? Could you mention (an) example(s)? What consequences could this 'gap' possibly have for the clinic / health care culture that your MPD serves?
- ▶ The availability of which substance data have priority, from the perspective of the purpose of your MPD?

Next to the questionnaires, two MPD organizations were interviewed, during which the same questions were discussed and the interim conclusions were validated.

MPD organizations participating in this research (either through input via the questionnaire or with the interviews) are listed in **Table 2**.

Table 2. MPD organizations included in this research

Organization	Country	Feedback
MHRA Drugs Dictionary	United Kingdom	Completed questionnaire
NoMA	Norway	Email response to questionnaire
Halmed	Hungary	Email response to questionnaire
WHODrug Global and UMC-SRS	Sweden	Completed questionnaire
G-standaard	Netherlands	Completed questionnaire
National Health Product Catalogue	Ireland	Interview
VIDAL	France	Interview
Unicom WP8 / Pilots (T8.1-T8.3)	Belgium	Interview / consult

Next to that, feedback was received from SNOMED.

Some MPDs are managed by the EU NCA, which implies they have access to EU-SRS. MPD organizations not managed by an EU NCA will not have access to EU-SRS and are therefore in higher need to access good quality substance data.

2.2 General impressions

The overall impression from feedback received is that the needs for substance data differs considerably between the various MPDs. The type of data fields required differs, as well as the sources used to capture substance data in the MPD. For example, a number of MPD organizations have indicated to use SNOMED as a source of information to feed their MPD, whereas others use their national SmPC or SPOR data sources.

All MPD organizations emphasized the need for unambiguous, high quality substance data to rely on when maintaining the MPD.

What also stood out was that MPDs all spend considerable efforts in finding substance information, either in databases or on public websites. This revealed a substantial amount of double work performed across MPDs to maintain substance data in the various MPD databases.

2.3 High-level substance information

The following high-level substance information has been found to be important for MPDs:

- ▶ Names (preferred terms + aliases and translations)
- ▶ Substance type (e.g. chemical, protein)
- ▶ Domain (veterinary versus human)
- ▶ Substance used in authorized products (yes/no)
- ▶ Active substance versus ingredient
- ▶ Role of the active ingredient or excipient when used in authorized MPs (colouring, coating, etc)
- ▶ Defining information, such as chemical structure or amino acid sequences

2.3.1 High level substance information in SMS

SMS is suitable for high-level characterization of substances, especially as source for substance names. **Table 3** shows the fields included in the public export of SMS data and example values. Currently, these names are jointly being cleansed by the SVG and SMS teams. In the daily export, a flag is included to indicate whether the record (including names) is cleansed/confirmed; the code 1 means that the substance record is confirmed, code 0 means the record will be deprecated. Records without a code were not yet cleansed.

Note: for the following tables and figures, the chemical 'Folinic acid' is used as the example substance.

Table 3. Fields included in public export of SMS data (CSV on SPOR portal) and example values.

Data Field	Example Value
SMSID	100000089017
Substance Name	Folinic Acid
Preferred Term Indication	TRUE
Language	English
Name Source	SPC
Domain	Human use
Status	Current
Substance Type	Chemical
Created Date	25-7-2019 11:55
Last Updated Date	6-6-2022 11:03
XEVMPD Code	SUB13910MIG
Cleansing Flag	1
FDA UNII	Q573I9DVLP

Confidential names are not included in the export file of SMS; it is the expectation that this is not a problem for MPDs.

2.3.2 High-level substance information in EU-SRS

In EU-SRS, each record has a display name, which is also known as the Preferred Term (PT). For PT and aliases, name types such as ‘systemic name’, ‘common name’ and ‘official name’ can be selected (see **Figure 1**). Every name has at least one reference.

Name	Type	Language	Details	References
Folinic acid ✓ ✓	Official Name	English	View	View
Leucovorin	Common Name	English	View	View
N-[4-[[[2-Amino-5-formyl-3,4,5,6,7,8-hexahydro-4-oxo-6-pteridiny]methyl]amino]benzoyl]-L-glutamic acid	Systematic Name	English	View	View

Figure 1. The name field of Folic acid in EU-SRS

2.3.3 Matching the needs of MPDs

The advice is to use the high-level substance information from SMS. The CSV file currently available in the public domain already provides most of these fields. Feedback from MPDs is that the quality of data is important. Therefore, it is advised to continue with cleansing of SMS data, especially SMS names of substances used in authorized products. In WP2, by the end of November 2022, about 80% of the substances used in products were cleansed in SMS.

There is no flag in SMS nor in EU-SRS to indicate that a substance is used in an authorized product. This information could possibly be derived from other databases. The advice is to further investigate if it would be possible to add an indicator for each substance whether or not these are used in one or more authorized medicinal products (binary field, since for MPDs it is not necessary to know in how many products the substance is used). Also, an indication on the extent in which a substance is used (# medicinal products making use of a substance) is of interest. This field is also not yet available in SMS nor in EU-SRS, but could be a derived field making use of other (EMA) data sources. This field is in principle considered a “nice to have”, the reason for including this information is the ability to prioritise maintenance work on the MPD by the MPD organizations.

In the G-SRS software, a ‘product’ extension has been developed, that is currently not in use in EU-SRS. It could be investigated if PMS information is suitable for this extension and can be included in EU-SRS. Then, export of this combined information into the public domain would be an option. There should be a flag for excipient (including adjuvants) yes/no as well as for active ingredient yes/no. It may add value to include an indicator of the number of products in which a substance is used. This information would have to be pulled from PMS and UPD (rather than from SMS/EU-SRS).

In certain cases, the substance characteristics on the level of the SSG1 (Specified Substance Group 1) are important for healthcare, e.g. for prescribing the right substance. Examples:

- ▶ Substances where the modifier influences the adverse effects of a substance, e.g. the liposomal or colloidal forms of amphotericin B, doxorubicin/daunorubicin.
- ▶ Substances that share the same name but have different dosages, like octocog alfa from Bayer and Baxter.

2.4 Substance characteristics

Unique substance identification is very important to MPDs and their users. In the light of that, the following general characteristics have been found to be important:

- ▶ Molecular Weight
- ▶ Molecular Formula
- ▶ Structure, including information on modifiers, hydrates, etc
- ▶ Relationships (including for example salts, hydrates, crystal forms) & hierarchy:
 - ▶ Stoichiometric relationships within a salt, for, for example Omeprazole Magnesium (2:1) or between components in a mixture such as Mepyramin 8-bromtheophyllinat, which consists of 50% Mepyramin and 50% 8-Bromtheophyllin (CAS Number is 606-05-3)

For MPDs, it is important to perform a “duplicate check”: is a substance a unique substance, or it is a synonym for a term already in the MPD, or does it refer to a substance that would be considered a variant of something already available in the MPD? In order to judge the uniqueness of a substance, the MPD need to know details, such as structure, crystal forms, modifications, bacterial or viral strains, host cell lines etc.

Comparing structures could be of help to pharmacists to understand the level of similarity between substances (e.g. in the light of allergies, where a part of a structure would be responsible for allergies).

Substance characteristics are used by MPDs to create MPD-identifiers, something similar to the Pharmaceutical product ID (PhPID) which is described in the ISO 11616 standard and is based on the substance, strength and dose form. Substance characteristics also form the basis for allergy checking. In this use case, especially information on grouping of or relationships between substances is important, e.g. substance is a salt-form of a certain substance.

Information on hierarchy or grouping of substances is currently not available to MPDs. A good example of this is Ibuprofen. If Ibuprofen could be part of a group, Ibuprofen would be the grouper, and all Ibuprofen substances would be part of this group. The need for PhPID grouping is explained and illustrated by taking the examples of Amlodipine and Ibuprofen in deliverable [D8.7 IDMP Coding Principles and Guidance for ICSRs](#). **Figure 2** provides a schematic overview of the Amlodipine and Ibuprofen groupers.

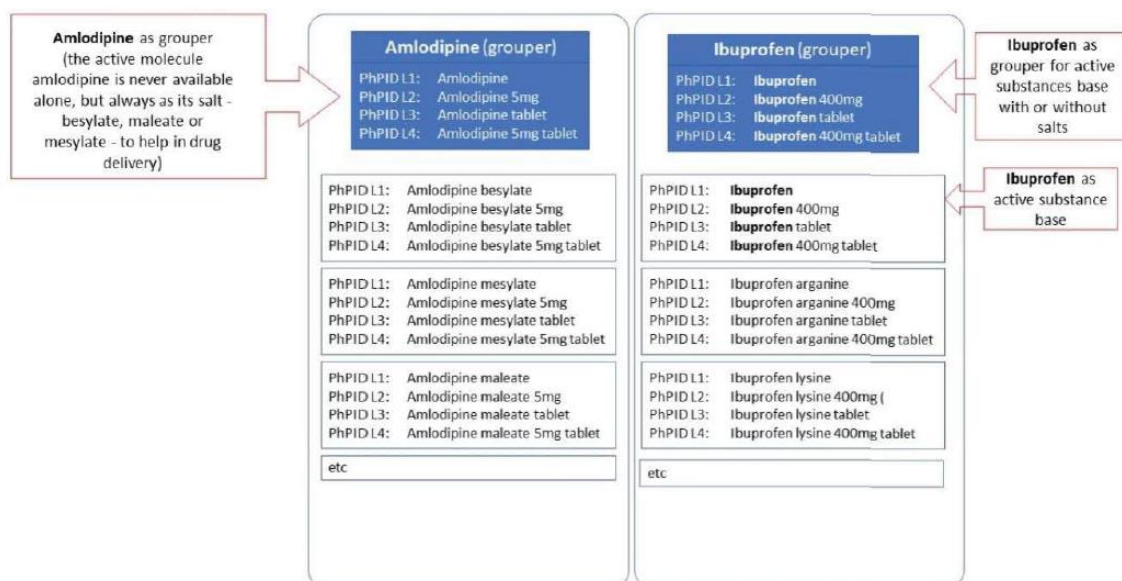


Figure 2. Need for groupers for non-specific drug verbatims (source: [Unicom deliverable D8.7: IDMP coding principles and guidance for ICSRs.](#))

More detailed, substance-type specific data field requirements were mentioned in the questionnaire feedback, including, for example cell line information, glycosylation and other physical properties. The detailed feedback can be found in the appendix.

WHO-UMC has indicated the wish to import full EU-SRS records in their SRS-database if allowed, since they are working on the implementation of UMC-SRS. This system would in the future be used to capture a global substance identifier (GSID). The GSID would be used for the generation of global PhPIDs. Records or parts of records (public data) from UMC-SRS could also be exported to enrich the EU-SRS with for example the GSID.

2.4.1 Substance characteristics in SMS

SMS is limited to high-level characterization of substances, where each substance type is defined with the same limited set of data fields. The needs from MPDs are, however, also to retrieve more detailed data on substance attributes.

In SMS, the molecular formula is available as data field but this is not in use yet. Note that the SMS team has indicated that in 2023 they intend to start using the molecular formula field by loading data from EU-SRS into the SMS field. This would allow to include the field in the public export of SMS. The field molecular weight is not available in SMS, however, EMA is considering adding this field to their database in 2023.

2.4.2 Substance characteristics in EU-SRS

In EU-SRS, substances are fully characterized. The general fields indicated by MPDs as required are all available in EU-SRS, for example molecular formula & molecular weight, structure and relationships.

In EU-SRS there are many relationships captured. The Parent/Salt relationship is currently creating the hierarchy, but other relationships can also be set to define e.g., prodrug and active metabolite, impurities etcetera if needed. Mixture components are also used to create hierarchy, but this is not a relation, but a mixture property. Possibly, the SSG1 can be used to indicate the amount that is used. For mixtures and SSG1 the relationship is integrated in the substance types (also to include the ratio if available).

The use of classifications could be a solution to the need from MPDs to introduce the grouper concept. This would allow, for example, to group all diclofenac substances into the same classification. This is searchable.

An overview of the EU-SRS record for Folic acid is presented in **Figure 3**.

Figure 3. Record in EU-SRS - Overview

From the Overview, the different fields (e.g. chemical structure, relationships) can be viewed further. For example, the chemical structure (**Figure 4**) and mixture components can be displayed (**Figure 5**)

Figure 4. Example of the 'Chemical Structure' field in EU-SRS.

Note that the Molecular Formula and Molecular Weight are calculated from the structure.

The screenshot displays a web interface for EU-SRS. On the left is a navigation menu with items like Overview, Substance Hierarchy, Mixture Components (4), Names And Synonyms (7), Codes - Classifications (27), Codes - Identifiers (16), Relationships Visualization, Relationships: Active Moiety (1), Notes (10), References (23), and Audit Information. The main content area is divided into two sections: 'Substance Hierarchy' and 'Mixture Components'. Under 'Substance Hierarchy', 'POLYMYXIN B' is selected, showing its ID 'J2VZ07J96K' and '(ACTIVE MOIETY)'. The 'Mixture Components' section states 'All of the following components must be present:' and lists four components with their IDs and chemical structures: POLYMYXIN B2 (99S3009VH2), POLYMYXIN B1 (B35589J3RU), POLYMYXIN B1.I (5XUD48XK49), and POLYMYXIN B3 (L194HGJ2LD).

Figure 5. Example of mixture components field in EU-SRS

EU-SRS is versatile. There are many choices you can make. For example, a relationship of 3:1 can be defined in different ways in different places in the system. In order to be able to share information in the public domain, it is important to be consistent in the way this type of information is captured in EU-SRS. This should be documented in the EU-SRS guides used by the SVG and where possible the system should be more guiding/or forcing this consistency. It is advised to align thoughts on this with the FDA colleagues.

Different types of relationships would also be very helpful in the healthcare domain. If you can use the characteristics in these cases, it would be easier to make the right decisions in healthcare.

2.4.3 Matching the needs of MPDs

Available data in SMS is not adequate to sufficiently support MPDs in characterizing substances. EU-SRS does contain the necessary data fields and therefore could be a valuable source for MPDs in characterizing substances.

In certain situations, the information in EU-SRS is confidential. Therefore, since MPDs require data in the public domain, the feasibility to share this data into the public domain needs to be evaluated further from a confidentiality perspective.

Advice:

- ▶ Further analyse the needs to have access to information on hierarchy of substances
- ▶ Next to hierarchy, analyse the possible relations between substances, for example if a substance is both a chemical as well as homeopathic substance, how we handle this in EU-SRS.

2.5 Required identifiers & references

The following identifiers were indicated to be important for MPDs:

- ▶ SMSID
- ▶ SVGID
- ▶ CAS
- ▶ UNII
- ▶ xEVMPD
- ▶ Other available identifiers

2.5.1 Identifiers & references available in SMS

In SMS, the following identifiers are currently available:

- ▶ SMSID
- ▶ xEVMPD code

► UNII

In the public export of SMS data, the SMSID and xEVMPD code are included, and recently also the UNII is added to the published dataset. Other identifiers such as the SVGID or CAS numbers are not yet included.

There is no additional functionality offered in the SMS export with regard to references or identifiers.

2.5.2 Identifiers and references available in EU-SRS

In EU-SRS, a substantial number of codes/identifiers will become available, for example:

- FDA UNII
- CAS
- ChEMBL
- ITIS
- Merck Index
- MPNS
- NCBI Taxonomy
- NCI_Thesaurus
- PubChem
- WHO-ATC (level 5)
- WHO-VATC
- Wikipedia
- EDQM monograph number
- INN
- DRUGBANK
- UNIPROT
- KEGG

In an EU-SRS record, for each substance identifier/code, a Code System and Code System Type is selected. Each code is coupled to its reference. This is presented in **Figure 6**. In EU-SRS, the URL available for identifiers can be selected, allowing the user to directly access the associated source on internet.

Code System	Code	Type	Comments	References
SMSID	10000089017	PRIMARY	View	View
CAS SCIFINDER	58-05-9	PRIMARY	View	View
WIKIPEDIA	FOLINIC ACID	PRIMARY	View	View
PUBCHEM	135403648	PRIMARY	View	View
MERCK INDEX	M5519	PRIMARY	View	View
CHEMBL	CHEMBL1679	PRIMARY	View	View
xEVMPD	SUB13910MIG	PRIMARY	View	View
FDA-UNII	Q57319DVLP	PRIMARY	View	View
SVGID	000076	PRIMARY	View	View
NCI_THESAURUS	C71631	GENERIC (FAMILY)	View	View
NCI_THESAURUS	C2078	GENERIC (FAMILY)	View	View

Figure 6. Codes/identifiers included in EU-SRS for Folinic acid

2.5.3 Matching the needs of MPDs

Any identifier available in EU-SRS could be of use to MPDs. At this moment, the ID of SNOMED-CT is not yet available in EU-SRS. It could be useful to also include this identifier in EU-SRS (and SMS). This requires further investigation; there is a link with PubChem and INN, possibly allowing to load the SNOMED-CT identifiers into EU-SRS automatically. When this could be arranged, this would allow MPDs to map to the SMSID based on the link with the SNOMED identifier.

Identifiers are in principle not confidential. Therefore, publishing substance identifiers from EU-SRS into the public domain should not cause any confidentiality conflicts. A realistic option to further investigate is to combine the currently available SMS substance data with identifiers from EU-SRS, where the SMSID is the link between the databases.

The UNII and the CAS numbers are considered the most important / required identifiers from an MPD perspective next to the SMSID and the SNOMED substance ID. Therefore, it is advised to prioritize these in making these identifiers publicly available.

3 Conclusions

3.1 Importance of quality of substance data

Unique identification of substances is of utmost importance to MPDs. One of the main reasons is to ensure correct dosing to patients. In order to ensure unique identification of substances, MPD organisations need more information than what is currently available in SMS (SPOR). Frequently, the national SmPC is used as a basis to extract substance information for feeding into the national MPD. This process is sensitive to errors, and may result in ambiguity in the description of substances, with implications on generic prescriptions and patient safety, and hindering the ability for cross-border e-prescriptions.

Control mechanisms should ensure that accurate substance information is reflected in SmPCs, with consistency between EU-countries.

At the start of the work of WP2, it was acknowledged that the quality of data in SMS needed improvement. The SVG, together with the EMA SMS team worked hard to cleanse the data in SMS. As a result, by the end of 2022, more than 80% of substances used in authorised products have already been cleansed and stakeholders will benefit from this work.

In EU-SRS, the substance records will be created based on the available regulatory (dossier & IMPD) information. With the process of peer review, the validity of data will be monitored. The SVG will also, supported by the functionality in EU-SRS, be monitoring for duplicate substance records.

EU-SRS is versatile. There are many choices you can make. Relationship 3:1 can be defined in different ways in different places in the system. This needs to be aligned also with FDA/NCATS, to ensure we capture/treat the data in the same way, so that we use the system in a consistent way. The alignment with the US colleagues is already happening. How to build human vaccines was discussed and agreed, as well as how to build polymers in EU-SRS was agreed with the US colleagues. The recently (pre-)released guides on substance management as a deliverable from WP2 will also be reviewed by the colleagues in the US.

3.2 Importance of accessibility to substance data fields

SMS data fields alone are not sufficient to support MPDs with characterizing substances in their databases. This applies to the SMS substance data fields already available in the public domain, but also to the SMS data fields not yet in use nor published. EU-SRS is not available in the public domain but can provide these additional data fields, either feeding SMS or publish directly on the internet.

Many of the substance data fields in EU-SRS that were indicated to be of need for MPD organizations are not confidential. For example, codes/identifiers, molecular formula, molecular weight. The focus of MPD organizations is mainly on substances used in authorized medicinal products; it may be relatively easy to make this data available in the public domain, either next to SMS data or integrated with SMS data. Both contain the SMSID, so merging the data would be based on the SMSID.

Also, adding a field with the number of products in which a substance is used and the role the substance played (active vs excipient vs both) is considered useful, however it is “nice to have”. One option would be to look into the use of the Product module available in the G-SRS software. This option is not yet used in Europe, but may be considered for the future; this implies a link with EMA’s PMS.

3.3 Improving efficiency

The feedback from MPD organizations showed the substantial amount of double work performed; every MPD owner creates their substance records based on various data sources including SMS, SNOMED, national SmPC’s and various literature sources and websites. The transition to IDMP & structured data in EU-SRS, governed by the SVG allows for a “one stop shop” approach on substance data. SMS and EU-SRS could jointly function as broker for substance data in Europe. With trustworthy data, meeting the needs of stakeholders including MPD organisations and NCAs, the efficiency of substance data

maintenance would be greatly improved, and data quality and consistency would also improve significantly.

Partners such as WHO-UMC who are using the same software as EU-SRS (the G-SRS software) would benefit a lot of being able to upload substance records from EU-SRS. It needs to be investigated further if and how this can be arranged. In line with this, the option to implement a public version of EU-SRS, containing non-confidential substance information should be further investigated.

4 Recommendations

Based on the analysis performed, the following steps are recommended:

- ▶ Investigate - together with the EMA SMS team, the possibility to publish a combined set of SMS and EU-SRS data in the public domain. Selection based on substances used in authorized medicinal products may be of benefit in preventing confidentiality issues. SMSID, names, substance category could be retrieved from SMS. Other fields such as additional identifiers, molecular weight and molecular formula could be retrieved from EU-SRS (with SMSID as linking identifier)
 - ▶ A staged approach could be an option, where quick wins can be achieved (e.g. publishing substance identifiers, molecular weight and molecular formula)
 - ▶ The need for access to substance-type specific data fields is clearly expressed, but is more complex to arrange and therefore requires further analysis. This refers to, for example: relationships, hierarchy, amino acid sequences and SSG1 information related to a substance.
 - ▶ Adding product-related information (# of products using the substance, role of substance in products) is also very valuable but may require more investigation on how to make this possible. This information is not captured in EU-SRS/SMS, but in UPD & PMS.
- ▶ Investigate how to map substance records of SNOMED with substances captured in the SPOR systems (EU-SRS/SMS).
- ▶ Investigate the possibility – similar to the approach taken by FDA – to implement a public version of EU-SRS, containing non-confidential substance information.

Appendix A: Example of inconsistencies in substance data

Below, an example is explained on inconsistencies found in substance data (see **Figure 7**, **Figure 8**, **Figure 9** and **Figure 10**). The example is on Cetirizine, an example already quoted in D9.1 (source: Unicom Deliverable D9.1: An analysis of the IDMP medicinal product identification data provided by NCAs (and SPOR) compared to that needed in MPD for clinical care and for secondary uses), but now being elaborated in the context of SMS/EU-SRS. Many more examples could be provided similar to this.

Note: The D9.1 report is not yet available in the public domain.

Cetirizine - SmPC

Different versions of the SmPC are using different descriptions for the same precise active ingredient substance. **Figure 7** shows that the active ingredient Cetirizine is recorded in different ways (section 2 of the SmPC):

<p>1. Name of the medicinal product</p> <p>Cetirizine Hydrochloride 10mg Tablets</p> <p>Tesco Allergy and Hayfever Relief 10mg Tablets</p> <p>EM Pharma Allergy & Hayfever Relief 10mg Tablets</p> <p>Pollenase Allergy & Hayfever Relief 10mg Tablets</p> <p>Lloyds Pharmacy Antihistamine Hayfever Relief 10mg Tablets</p> <p>Careway Antihistamine Hayfever Relief 10mg Tablets</p> <hr/> <p>2. Qualitative and quantitative composition</p> <p>Each tablet contains Cetirizine hydrochloride 10mg.</p>	<p>1. NAAM VAN HET GENEESMIDDEL</p> <p>Zyrtec 10 mg filmomhulde tabletten</p> <p>2. KWALITATIEVE EN KWANTITATIEVE SAMENSTELLING</p> <p>Elke filmomhulde tablet bevat 10 mg cetirizine dihydrochloride.</p>
--	--

Figure 7. SmPC extracts for two products using different descriptions for the same precise active ingredient substance.

At first sight, these products would appear to contain different precise active ingredient substances, since normally “hydrochloride” is used when the modification consists of a single hydrochloride (monohydrochloride). So, the information in the SmPC is not sufficient to unambiguously define the active ingredient.

Cetirizine data - SMS

Figure 8 shows the substance data available in SMS (based on public export):

	A	B	C	D	E
1	#SMS_ID	Substance_Name	Is_PREFERRED_Name	Language	Name_Source
26759	100000081513	Cetirizine	TRUE	English	INN
74207	100000091468	CETIRIZINE DIHYDROCHLORIDE	TRUE	English	INN
74208	100000091468	CETIRIZINE HYDROCHLORIDE	FALSE	English	INN
74212	100000091468	CETIRIZINE DIHYDROCHLORIDE	FALSE	Dutch	
74218	100000091468	DICHLORHYDRATE DE CETIRIZINE	FALSE	French	COMPANY SPECIFICATION

Figure 8. Public SMS export data; search for “Cetirizine” in Substance_Name

Interpretation of **Figure 8**:

Cetirizine hydrochloride as well as Cetirizine dihydrochloride are displayed, with the same SMS_ID. This indicates that it is one and the same substance.

When consulting EU-RS, the following is found:

Cetirizine data - EU-SRS

Figure 9 shows the search results when searching “cetirizine” in (test) EU-SRS.

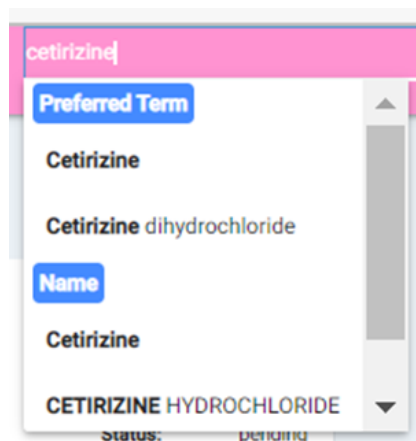
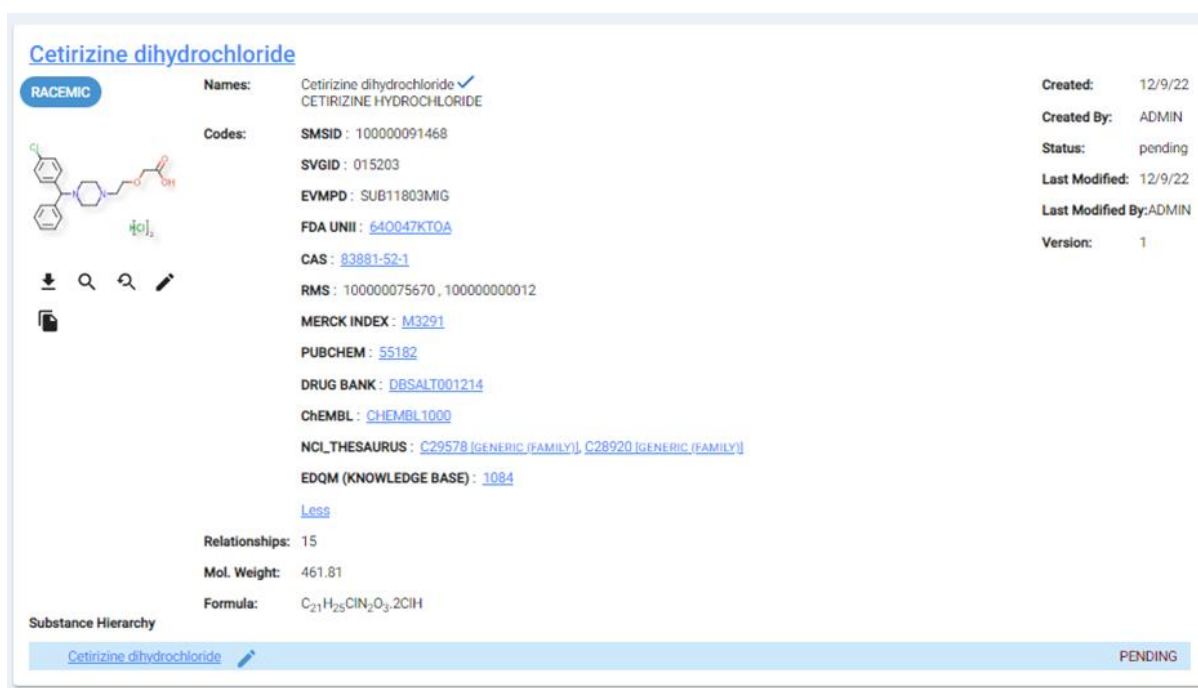


Figure 9. Search for “Cetirizine” in (test) EU-SRS

Figure 10 shows that the Cetirizine dihydrochloride is the search result of both Cetirizine hydrochloride and Cetirizine dihydrochloride.



Cetirizine dihydrochloride

RACEMIC

Names: Cetirizine dihydrochloride ✓
CETIRIZINE HYDROCHLORIDE

Codes: SMSID : 100000091468
SVGID : 015203
EVMPD : SUB11803MIG
FDA UNII : [640047KTOA](#)
CAS : [83881-52-1](#)
RMS : 100000075670, 1000000000012
MERCK INDEX : [M3291](#)
PUBCHEM : [55182](#)
DRUG BANK : [DBSALT001214](#)
ChEMBL : [CHEMBL1000](#)
NCI_THESAURUS : [C29578 \[GENERIC/FAMILY\]](#), [C28920 \[GENERIC/FAMILY\]](#)
EDQM (KNOWLEDGE BASE) : [1084](#)

[Less](#)

Relationships: 15
Mol. Weight: 461.81
Formula: C₂₁H₂₅ClN₂O₃·2ClH

Created: 12/9/22
Created By: ADMIN
Status: pending
Last Modified: 12/9/22
Last Modified By: ADMIN
Version: 1

Substance Hierarchy
[Cetirizine dihydrochloride](#)

PENDING

Figure 10. The record for Cetirizine (di)hydrochloride in (test) EU-SRS

Cetirizine hydrochloride and Cetirizine dihydrochloride both refer to the same substance record; where 'Cetirizine dihydrochloride' is the preferred term (PT) and 'Cetirizine hydrochloride' an alias.

From the EU-SRS data it can thus be concluded that Cetirizine hydrochloride and Cetirizine dihydrochloride are synonymous and both refer to the dihydrochloride.

It is shown that a substantial number of codes/identifiers is available and each code is coupled to its reference (refer to section 2.5.2 of this Deliverable), with clickable links.